NORTHWESTERN UNIVERSITY

Bayesian-robust Algorithms Analysis with Applications in Mechanism Design

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Aleck Christopher Johnsen

EVANSTON, ILLINOIS

September 2021

ABSTRACT

Bayesian-robust Algorithms Analysis
with Applications in Mechanism Design

Aleck Christopher Johnsen

This thesis studies *Bayesian-robustness* of algorithm design. The main perspective requires for a single fixed algorithm that its performance is an approximation of the optimal performance when its inputs are independent and identical draws (i.i.d.) from every unknown distribution which is an element of a known, large class of distributions. Formally this information framework is the *prior independent setting*. Generally this thesis studies design structure that is common to arbitrary algorithms problems for which the prior independent setting is appropriate. The questions addressed by this thesis were largely motivated by questions within mechanism design and thus application of its general results focuses on mechanism design problems.

As a major contribution, this thesis gives a method – *the Blends Technique* – that is agnostic to algorithm problem setting for proving lower bounds on the prior independent approximation factor of any algorithm. The method constructs a correlated distribution over inputs that can be generated both as a distribution over i.i.d. good-for-algorithms

distributions and as a distribution over i.i.d. bad-for-algorithms distributions. Prior independent algorithms are upper-bounded by the optimal algorithm for the latter distribution even when the true distribution is the former. Thus, the ratio of the expected performances of the Bayesian optimal algorithms for these two decompositions is a lower bound on the prior independent approximation ratio. The structure of the Blends Technique connects prior independent algorithm design, Yao's Minimax Principle, information design, tensor decomposition, and benchmark design for the prior free information setting (i.e., worst-case over inputs / competitive analysis). This framework is applied to give novel lower bounds on canonical prior independent mechanism design problems.

Another main contribution of this thesis is to use the objective of Bayesian-robustness to inform prior free benchmark design. Benchmarks are free parameters in worst-case algorithm design and choice of benchmark is of critical concern for algorithm analysis. This thesis gives a framework for optimal benchmark design from a requirement that approximation of a prior free benchmark must further hold as a prior independent approximation guarantee. Subsequently, it shows that benchmark design from this framework is equivalent to optimal prior independent algorithm design.

This thesis includes the solution to a central open question in prior independent mechanism design, namely it identifies the prior independent revenue-optimal mechanism for selling a single item to two agents with i.i.d. values from a regular distribution. This optimal mechanism is used in the construction of an optimal solution to the corresponding benchmark design problem.

Acknowledgements

I want to thank all of the following people from the bottom of my heart: for bottomless support and belief in me, especially in the times when I felt like I was unable to find that support for myself. I might spend the rest of my life repaying my debt of gratitude.

In the world of math, in perturbed-chronological order, thank you to:

... my early math influences, Walter Dodge in high school and professors Bruce Reznick and A.J. Hildebrand (UIUC, Mathematics) for introducing me to research, specifically through a REU in Number Theory at UIUC in 2002.

... past-and-present Northwestern professors Nicole Immorlica, Ming-Yang Kao, Lance Fortnow, Ronen Gradwohl, Doug Downey, and Aravindan Vijayaraghavan for various support of me ranging over: welcoming me into Northwestern's computer science theory group, providing funding for my early years, and serving as committee members for my Quals and Proposal presentation milestones – you believed in my potential from the beginning and made it possible for me to advance.

... Shinnosuke Seki for inviting and funding me to visit him at Aalto University (Helsinki) in Summer 2013; and Avrim Blum and Jamie Morgenstern for hosting my time at Carnegie Mellon University (Pittsburgh) in Spring 2014.

... all of the people at Booking.com (Amsterdam) for hosting my project with them in Summer 2016, including the full Ranking Team and all authors of our presentation at MIT-CODE (2017): Andrea Ciancone, Shirish Kasa, Igor Malikov, Melanie Mueller, Vladyslav

Noryk, Steve Oliemans, Roberto Pagano, Nam Pham, Eric Soetrisno, Leonardo Soto, Marcell Szathmari, and Onno Zoeter; and also with Jason Hartline and Denis Nekipelov.

... past Northwestern students for many nostalgic memories of lighthearted social events back when life was simpler: Yorgos Askalidis, Huck Bennett (post-doc), Rachel Cummings, Abrhatanu Dutta, Bach Ha, Nima Haghpanah, Darrell Hoy, Manolis Pountourakis, Greg Stoddard, and Sam Taggart – one of my favorite things about conferences is bumping into you and catching up and sharing old stories.

... recent and current Northwestern students and general members of the Online Markets lab: Alireza Amanihamedani, Hedyeh Beyhaghi, Modibo Camara, Charles Cui, Jinshuo Dong, Yiding Feng, Jessica Hullman, Yingkai Li, Sheng Long, Michalis Mamakos, Rad Niazadeh, Anant Shah, Liren Shan, Matthew von Allmen, Yifan Wu, and Chenhao Zhang – your attention and perspectives during lab meeting unquestionably forced me always to better organize my ideas and consequently take forward best versions of my work.

More broadly, I have benefited from conversations with Jake Abernethy, Amartya Banerjee, Avrim Blum, Simina Branzei, Xue Chen, Costis Daskalakis, Anindye De, Yingni Guo, Katya Khmelnitskaya, Samir Khuller, Kostya Makarychev, and Madhav Suresh.

(I defer special acknowledgement of my PhD committee and co-authors to the last.)

Thank you to my advisor, Jason Hartline, who served in just about every role described above, and who opened all the doors to make it possible for me to write this thesis and produce my whole body of work. It has been a long time since, "a solution so weird that it is interesting." You let me work on hard problems and together we slowly produced phenomenal results. With your support and encouragement, I was able to travel the world and collaborate with – and present my work to – some of the best minds in computer

science anywhere. Always, both elegantly and painfully, you forced me to be a better researcher and a better communicator. Thank you Dr. Hartline.

I want to thank Carl Plicner, maven of Peet's (Clark and Barry, Chicago) and well-wishing coffee-patron to every flagging graduate student who ever walked in that door.

To my family – the moments of blessing, grace, and pure love that I have received over decades from you are uncountably-infinite. I want to thank forever:

... my beautiful wife, Cara Mullen, PhD, because your own words back to you: "let's do math problems;" and my wonderful children: Adelein, Zachary, and Samuel; to my family for daily, ... hourly, ... continuous love, support, and always, always, goofball smiles.

... my parents: Amalia and David Johnsen, because obviously there are no words to explain all the things you have done for every aspect of my life.

... my in-laws: Janet and Stephen Mullen, for Cara and who she is, and so much more.

... my grandparents: E. Rosemary Johnsen (RIP) and Aleck Johnsen; Zacharoula Mitsis and Gus Mitsis (RIP); for the risks that you took and the families, opportunities, and legacies that you built.

... my brothers and sisters, *as my best friends*: Dean Johnsen and Julia; Mark Johnsen and Elaina; Chris Mullen and Laura, PsyD.

... my aunts and uncles: my "Chicago second-moms" Christine and her husband John, and Frances and her husband Mark; Steven who told me to do engineering undergrad and then "I would be able to do whatever I wanted afterwards" and his wife Kathy; Nancy (RIP) and her husband Dieter, PhD, who always asked about my progress.

... and all my family.

And lastly yet unequivocally, I want to thank the people without whom I would not be able to do it at all: my PhD defense Committee and my co-authors, *published, currently submitting, and unlisted-but-undoubtedly-included-here future.* The *Ender's Game* quotes on the next page are for you, an open invitation, regarding the "story of this thesis" as a continuing pursuit of joint-work. The *true story* remains to be solved.

You each know what I owe you for our joint work. What more could I say here? You are all brilliant and it is my honor to work with you. I hope we never stop doing math.

Jason Hartline: as my advisor (one more time); and alphabetically:

Modibo Camara: Northwestern University

Yiding Feng: Northwestern University

Ronen Gradwohl: Ariel University;
previously Northwestern University Kellogg

Ming-Yang Kao: Northwestern University, emeritus

Yingkai Li: Northwestern University

Denis Nekipelov: University of Virginia

Shinnosuke Seki: University of Electro-Communications;
previously Aalto University and Academy of Finland

Aravindan Vijayaraghavan: Northwestern University

Zihe Wang: Renmin University of China;
previously Shanghai University of Finance and Economics

Onno Zoeter: Booking.com

and Sam Taggart (Oberlin College, previously NU) for many incipient, invaluable discussions about the problem that became the Blends Technique of this thesis.

"But you gotta dig deep to the Heaven above:

Sit, Learn, and Create ...

we all have a Black Rock in our mind and I go there all the time."

$\sim$ O.A.R., *Black Rock* (1997)

"An equation for me has no meaning, unless it expresses a thought of God."

$\sim$ Srinivasa Ramanujan

$\Sigma\tilde{\omega}\sigma o\nu,\ K\acute{\upsilon}\rho\iota\epsilon,\ \tau\grave{o}\nu\ \lambda\alpha\acute{o}\nu\ \sigma o\upsilon\ \kappa\grave{\alpha}\iota\ \epsilon\grave{\upsilon}\lambda\acute{o}\gamma\eta\sigma o\nu\ \tau\grave{\eta}\nu\ \kappa\lambda\eta\rho o\nu o\mu\acute{\iota}\alpha\nu\ \sigma o\upsilon$

$X\alpha\acute{\iota}\rho\epsilon\ \nu\acute{\upsilon}\mu\phi\eta\ \grave{\alpha}\nu\acute{\upsilon}\mu\phi\epsilon\upsilon\tau\epsilon$

$\Sigma\epsilon\ \gamma\nu\omega\rho\acute{\iota}\zeta\omega\ldots\ X\alpha\acute{\iota}\rho\epsilon,\ \omega\ \chi\alpha\acute{\iota}\rho\epsilon,\ \epsilon\lambda\epsilon\upsilon\theta\epsilon\rho\iota\acute{\alpha}$

"The story itself, the true story, is the one the audience members create in their minds, guided and shaped by my text, but then transformed, elucidated, expanded, edited, and clarified by their own experience, their own desires, their own hopes and fears."

$\sim$ Orson Scott Card in the Introduction to *Ender's Game* (1991)

"Remember – the enemy's gate is *down*."

$\sim$ Ender Wiggin

Table of Contents

List of Figures

CHAPTER 1

Introduction

This thesis studies *Bayesian-robust algorithms analysis*. Stochastic models are enabling theoretical understanding of algorithms beyond those provided by classical worst-case treatments (see Roughgarden, 2019). These models are especially interesting for algorithm design problems with information theoretic constraints such as online algorithms, mechanism design, streaming algorithms, etc.

Within the stochastic framework, the overarching setting for thesis is *prior independent algorithm design*: n inputs are independent and identical draws (i.i.d.) from an *unknown distribution* F, which comes from a *known class* of distributions $\mathcal{F}$. Informally, "inputs are each drawn from a common distribution but we do not know what the distribution is." Prior independent algorithms are measured by their worst-case (over distributions) approximation to the performance of the optimal algorithm which knows the distribution. Thus, the goal of this setting is to identify the algorithm A^* that minimizes the worst-case ratio of expected optimal performance to the expected performance of A^* for the same distribution, thereby optimizing a measure of *Bayesian-robustness*.

Application Settings. The main application of this thesis is mechanism design. One or more items are allocated to n strategic agents with *private types*; design of mechanisms is subject to incentivizing the agents to participate, without knowing exactly how their types inform their behavior. A minor application of this thesis is online algorithms. An

algorithm designer receives a sequence of n inputs and must commit to a decision before seeing each new individual input, *without knowing the future*. Prior independent analyses in mechanism design (e.g., Dhangwatnotai et al., 2015) and online learning (e.g., Auer et al., 2002) are i.i.d. respectively over values of agents in a mechanism and rounds of online inputs.

Thus in both settings, an algorithm's information about the input is structurally limited, and there is no single algorithm that is optimal given all realizations of the unknown inputs. Typically, mechanism design and online algorithms are treated as separate realms of research. This thesis gives many results that are domain-agnostic, i.e., from a unified framework that abstracts a common environment of incomplete information.

The Blends Technique for Prior Independent Lower Bounds. As a major contribution, this thesis develops a novel, algorithm-agnostic method for establishing lower bounds on the performance of prior independent algorithms (for classes of i.i.d. distributions). The method is based on Yao's Minimax Principle (Yao, 1977). As just described, the prior independent setting asks for the designer to pick one algorithm that is good on an adversary's chosen worst-case distribution. Yao's Minimax Principle allows the order of moves of the designer and adversary to be swapped. Thus, the prior independent optimal approximation ratio can be equivalently identified by an adversary choosing a distribution over prior distributions and then the designer choosing an optimal algorithm in response. Note, *the class of i.i.d. distributions is not closed under convex combination* (which adds a critical degree of technical complexity to the analysis). Consequently, the adversary's distribution over distributions – called a *blend* – gives generally a symmetric, correlated distribution over inputs.

The main object of study from this minimax approach is *dual blends*, which are pairs of distinct distributions over i.i.d. distributions of inputs that induce the same correlated distribution. To establish a prior independent lower bound, the method considers dual blends where one side of the dual blend mixes over good-for-algorithms distributions and the other side mixes over bad-for-algorithms distributions. The adversary can choose the mix over good-for-algorithms distributions in which case the expectation over Bayesian optimal performances for this mix will define the benchmark of the prior independent framework. On the other hand, the algorithm cannot tell the two blends apart and thus its expected performance is upper bounded by the expectation over performances of the Bayesian optimal algorithms for the bad-for-algorithms mix. The gap that results from this *Blends Technique* between these expected performances is a lower bound on prior independent approximation.

As a simple example, consider the mechanism design problem of posting a price to a single agent with value $v \in [1, h]$. (Here the restriction to i.i.d. distributions is trivial as there is only one agent.) A class of good-for-algorithms distributions is given by point masses. Note that the Bayesian optimal pricing mechanism for a point mass is to post identically the same price as the value (at which the agent always buys). A class of bad-for-algorithms distributions is given by the equal revenue distribution with cumulative distribution $F(v) = 1 - 1/v$ and a point mass of $1/h$ at h. The equal revenue distribution has the property that the expected revenue from any posted price is 1 (the agent buys if value is at least the price). Now consider the dual blend where the good-for-algorithms side is chosen to be the equal revenue distribution over point masses and the bad-for-algorithms side is chosen to be a point mass on the equal revenue distribution. The expected revenue

over Bayesian optimal algorithms (in response to point mass distributions) from the good-for-algorithms side is the expected value of the equal revenue distribution on $[1, h]$, i.e., $1 + \ln h$. The expected revenue from the bad-for-algorithms side is 1. This establishes a lower bound of $1 + \ln h$ on the approximation factor of single-agent posted pricing. (In fact, this example analysis is tight due to a matching upper bound from Hartline and Roughgarden (2014).)

There are two challenges in establishing lower bounds for prior independent algorithms via the blends method. The first challenge is in sufficiently understanding the Bayesian optimal algorithm for the class of distributions under consideration. In several of the central studied areas of Bayesian algorithms, this first challenge is solved in closed form. Bayesian optimal mechanisms are identified broadly by Myerson (1981). For online learning with payoffs that are i.i.d. across rounds, the Bayesian optimal algorithm is trivial, it selects the action with the highest expected payoff (which is the same in each round). Of course, when closed forms are not available, bounds on the Bayesian optimal performance can be employed instead. An important observation of the method of dual blends is that not only are Bayesian optimal algorithms used to define the benchmark, but they can also be *solely-sufficient* to get non-trivial bounds on any algorithm's prior independent approximation. I.e., they play this role without any reference at all to a specific algorithm.

The second challenge of the blends method is in identifying dual blends where the expected Bayesian-optimal performances for good-for-algorithms and bad-for-algorithms distributions are significantly separated. In pursuit of this challenge, this thesis gives two general approaches for constructing dual blends for inputs of size two. (Many of the challenge problems in prior independent mechanism design are for inputs of size two,

e.g., this thesis builds on Allouah and Besbes (2018).) The first approach is based on the observation that when the density function of a correlated distribution on inputs of size two can be written as a separable product of independent functions *per order statistic* of the inputs, then it can be decomposed into two distinct distributions over i.i.d. distributions. The second approach considers one side of the dual blend constructed from any scaled class of distributions with the other side given by the inverse-distributions of these (for which, as a class, the roles of values and scales are reversed in comparison to the original class).

Overall, the Blends Technique is an elegant method that brings together prior independent algorithm design, Yao's Minimax Principle, prior free benchmark design (see next), the economics topic of information design, and the mathematics/ computer science topic of tensor decomposition. This thesis applies the Blends Technique to two canonical problems in mechanism design. Discussion of these applications is given below as part of *Results in Mechanism Design*.

Optimal Benchmark Design. There is a second leading approach for robust optimization of algorithm design which measures performance in worst-case over inputs against a benchmark function. This approach to algorithm design is typically named *prior free* in mechanism design (which this thesis adopts) and *competitive analysis* in online algorithms. Critically, the benchmark function is a *free parameter* in prior free algorithm design and choice of benchmark is material. In fact, in comparing prior independent and prior free design, a determination of which setting is *more-robust* depends on the prior free benchmark function (which theoretically may be arbitrarily small). This thesis formalizes a first problem for designing a good benchmark for prior free approximation

and connects this benchmark design problem to prior independent optimization in a specific way – for a large class of distributions (of interest to a prior independent algorithm problem), approximation of a benchmark *must induce a guarantee* of the same prior independent approximation. (This property of benchmarks is *normalization*, (Hartline and Roughgarden, 2008)). Benchmark design is motivated from the following philosophy.

The choice of prior free benchmark impacts the ability of approximation with respect to the benchmark to distinguish between good and bad algorithms. On one hand a benchmark should be an upper bound on what is achievable by an algorithm; otherwise, approximating it does not necessarily mean that an algorithm is good. On the other hand it should not be too loose an upper bound; otherwise, neither good nor bad algorithms can obtain good approximations and the degree to which good and bad algorithms can be distinguished via the benchmark is limited.

As a simple example, consider comparing the revenue of two mechanisms (below) for selling digital goods[1] to n agents with values $\boldsymbol{v} = (v_1, \ldots, v_n)$ bounded on $[1, h]$ to one of two benchmarks, the *sum-of-values* benchmark $\sum_i v_i$ and the *price-posting-revenue* benchmark $\max_i i\, v_{(i)}$ where $v_{(i)}$ is the ith highest value. To ensure benchmarks give an upper bound on revenue, Hartline and Roughgarden (2008) suggested that the benchmark satisfies the property that, for any distribution on inputs from a given family of distributions, the expected benchmark (over the same distribution of inputs) be at least the expected performance of the optimal mechanism that knows the distribution. This property is exactly normalization above, thus inducing prior independent guarantees. Both

[1] In a digital goods auction, there are n copies of an item for sale and each agent has one-unit-demand.

the sum-of-values and price-posting-revenue benchmarks are normalized (Hartline and Roughgarden, 2008).

Not all normalized benchmarks are equally good at discriminating between good and bad mechanisms. Consider the following two mechanisms. The *random-sampling* mechanism partitions the agents at random and offers the optimal price from each part to the other part (Goldberg et al., 2006). The *random-power-pricing* mechanism posts a take-it-or-leave-it price drawn from the uniform distribution on powers of two in $[1, h]$ (Goldberg and Hartline, 2003). These mechanisms and the benchmarks of the preceding paragraph are related as follows. On all inputs $\boldsymbol{v}$, the sum-of-values benchmark exceeds the random-power-pricing mechanism by a $\Theta(\log h)$ factor. On all inputs $\boldsymbol{v}$, the random sampling mechanism and the price-posting-revenue benchmark are $\Theta(1)$. Moreover, the performance of the latter benchmark and mechanism are always sandwiched between the former benchmark and mechanism and can equal either of them up to $\Theta(1)$. Thus, the loose benchmark of sum-of-values does not discriminate between good mechanisms like random-sampling and bad mechanisms like random-power-pricing.

The preceding discussion suggests a benchmark design problem of identifying the normalized benchmark to which the tightest approximation is possible. The tightest approximation possible for a benchmark is its *best-response resolution*, as up to this factor the benchmark cannot distinguish between good and bad algorithms. The sum-of-values benchmark has logarithmic resolution while the price-posting-revenue benchmark has constant resolution.

Returning to the general algorithms setting, the outline of the approach so far to benchmark design is summarized as follows:

- A class of distributions over inputs induces a class of normalized benchmarks.
- The benchmark admitting the tightest approximation is optimal, i.e., the benchmark having smallest best-response resolution.
- The algorithm achieving this approximation is the prior free optimal algorithm for the benchmark.
- Normalization of the benchmark implies that the prior free approximation factor of an algorithm (to the benchmark) is at least its prior independent approximation factor (for the normalizing class of distributions).

A natural question is how this prior free approach – and its optimal algorithm – compares to directly identifying the prior independent optimal algorithm, i.e., the one with the best worst-case-over-distributions approximation to the Bayesian optimal algorithm. A main contribution of this thesis is the general result that optimal benchmark design is equivalent to prior independent optimization. Specifically, the prior free optimal algorithm for the optimal benchmark is the prior independent optimal algorithm and further, the optimal benchmark is simply the prior independent optimal algorithm scaled up by its approximation factor (so as to satisfy the normalization constraint).[2] Consequently, it is not possible to identify optimal benchmarks and their corresponding optimal algorithms when the solution is unknown for the prior independent optimization algorithm problem.

There is an important negative interpretation of this equivalence result (i.e., that the prior free benchmark optimization problem and the prior independent optimization problem give the same answer). One reason to prefer prior free analysis over prior independent

[2] The analysis that proves this equivalence is straightforward and, perhaps, obvious in hindsight. As will be described next, there is an alternative measure of resolution which, in hindsight, may be viewed as a relaxation of best-response resolution. The prior literature strongly suggested that benchmark design from either definition was equivalent; this thesis includes the result that the relaxation is sometimes lossy.

analysis is that, intuitively, prior free should be more robust by measuring approximation pointwise per input vector $\boldsymbol{v}$ rather than in expectation per distribution. (This is implemented formally from normalization.) However, the equivalence result described above indicates that there is no added robustness from the prior free approximation of the optimal benchmark (above the robustness of prior independent analysis).

Generally, there are environments where increased robustness can informally be observed from prior free approximation of alternative benchmarks. An example is the *best-in-hindsight* benchmark of the expert learning problem. This thesis analyzes expert learning to make explicit the challenge to equivalence of prior independent design and the given framework for benchmark design.

Towards a possible improvement to benchmark design, observe that the approach of normalized-benchmark optimization via resolution can be generalized by abstracting to allow consideration of *other techniques* which establish provable lower bounds on approximation. A second such method to establish a lower bound is measured in worst-case-over-distributions as the approximation of the expected benchmark (over inputs from a distribution) by the optimal algorithm which knows the distribution (Goldberg et al., 2006)). In fact, Hartline and McGrew (2005) used the prominent *equal revenue distribution* (EQR) of mechanism design to prove that this approach is tight for a large class of benchmarks and $n = 3$ agents; and Chen et al. (2014) used EQR to prove that this approach is tight for a large class of benchmarks and a general number n of agents. This measure of lower bound on approximation from Goldberg et al. (2006) is a benchmark's *normalization-symmetric resolution*, as it is measured "symmetrically" to normalization constraints. This thesis includes comparison of best-response versus normalization-symmetric resolution measures

by which it realizes: optimizing the latter yields smaller "absolute" resolution but the optimal algorithm approximation to its benchmark must be larger (tautologically by design of best-response resolution).

Results in Mechanism Design. A last main contribution of this thesis is identification of the optimal, *scale-invariant, truthful* prior independent mechanism for the setting of maximizing revenue from the sale of a single item to two agents for the standard class of i.i.d. regular value distributions.[3] Its optimal approximation factor is ≈ 1.907. Thus, per the previous discussion of an equivalence between prior independent algorithm design and (best-response resolution) benchmark design, the optimal prior free benchmark is also identified by corollary (and with identical optimal resolution ≈ 1.907).

This prior independent mechanism design result answers a major question left open from Dhangwatnotai et al. (2015), Fu et al. (2015), and Allouah and Besbes (2018). The optimal mechanism is a mixture between the Second Price Auction (SPA), where each agent is offered a price equal to the highest of the other agents' values, and the auction where these prices are scaled up by a factor of about ≈ 2.44. This solution is the first example of a prior independent optimal mechanism that arises as the solution to a non-trivial optimization problem and is not a standard mechanism from the literature (though the mechanism does fall into the class of lookahead mechanisms described by Ronen (2001) and has the same form as mechanisms used to prove upper bounds in Fu et al. (2015) and Allouah and Besbes (2018)). Further, the result identifies the worst-case class of distributions (when values have support $[0, \infty)$), specifically by showing that the optimal

[3] Allouah and Besbes (2018) conjecture that the restriction to scale-invariance within this setting is without loss. The author of this thesis endorses this conjecture.

mechanism and worst-case distributions are mutual-best-responses (and are thus a mixed Nash equilibrium).

Some context follows regarding the restriction to mechanisms that are (a) dominant strategy incentive compatible (DSIC), i.e., where truthtelling is a good strategy for each agent regardless of the strategies of other agents, and (b) scale-invariant. It is known that there are environments for prior independent mechanism design where DSIC is not without loss (Feng and Hartline, 2018). The restricted class of DSIC mechanisms is interesting even if it is with loss. It is not known whether scale-invariance is without loss or not.[3] Without assuming scale-invariance, the question is open since Dhangwatnotai et al. (2010).

As previously mentioned, this thesis further includes novel results stating lower bounds within mechanism design from implementation of the Blends Technique. The settings for these results are similar but not identical to the setting of the optimal mechanism just identified. All three settings have the following in common: they are two-agent, single-item auction environments. Otherwise, distinct parameters from the optimal result above are *italicized*.

Both blends-induced lower bound results are applied within *finite value support* $[1, h]$ and thus *drop the assumption of scale-invariance*. (Note, the worst-case distributions of the optimal analysis above have positive density in both limits approaching 0 and ∞ and thus do not fit into this setting.)

Within the new context of these modifications, the first result considers the objective of revenue maximization for the standard class of i.i.d. regular value distributions. (i.e.,

same as above). This thesis uses the Blends Technique to establish a lower bound for this setting of $^{23}/_{18} \approx 1.2777$.

While in absolute terms, this lower bound on the optimal prior independent approximation factor does not compare favorably to the tight, scale-invariant factor above of ≈ 1.907, consider the comparison. Intuitively, holding everything else fixed, the prior independent approximation of an optimal mechanism should be smaller for scale-conscious mechanisms when values have finite support $[1, h]$ than for scale-invariant mechanisms when values have infinite support $[0, \infty)$. Certainly the optimal mechanism which knows bounded support will not commit a priori to a markup mechanism (other than possibly the benign-1-mark-up of the SPA). Thus in the finite-support setting, the 1.907-lower-bound is unlikely to persist and a lower bound of 1.2777 is meaningful. Comparing approximation factors indicates the possibility of a *price of scale-anonymity*.

The second lower-bound result from the Blends Technique considers the objective of *residual surplus maximization* (i.e., maximizing: the expected value of the agent who receives the single-item minus any payments made) for the class of *all distributions*. For this residual surplus objective, an upper bound of $^4/_3$ exists as a corollary of Hartline and Roughgarden (2014). This thesis establishes a lower bound of 1.00623 (no previous lower bound was known other than the trivial lower bound at 1).

As a point of interest, these two applications of the Blends Technique (a) use the same example of a dual blend solution to prove their lower bounds, but (b) *choose different sides of the dual blend to effectively set the prior independent benchmark*. This observation has a natural interpretation in terms of information design and a consequence of this dichotomy is that there is no Blackwell ordering between the two sides of the example dual blend.

Historical Context: Online Algorithms and Mechanism Design. Online algorithms have been analyzed via a worst-case competitive analysis since Sleator and Tarjan (1985) with textbooks on the subject, e.g., Borodin and El-Yaniv (1998). In competitive analysis, the performance of an online algorithm is measured as its worst case ratio to the optimal offline algorithm. Good online algorithms are known with respect to this measure for many problems. For some problems this measure is too pessimistic, occurring when no good ratio is achievable by any algorithm and therefore good algorithms are not meaningfully separated from bad algorithms. The two approaches for resolving this issue are to either (a) restrict the offline algorithm to which the performance of the online algorithm is compared or (b) restrict the class of inputs that are considered. For this thesis, the most relevant example of (a) comes from expert learning where an algorithm's regret is measured with respect to the best fixed action in hindsight (Littlestone and Warmuth, 1994; Freund and Schapire, 1997). The most relevant example of (b) for this thesis is the diffuse adversary model of Koutsoupias and Papadimitriou (2000) which evaluates an algorithm as the ratio between its expected performance and the optimal offline performance in worst-case over a class of distributions on inputs.

Less immediately related to this thesis is a systematic study of analysis frameworks for online algorithms conducted by Boyar et al. (2015). They focus on a canonical *two server problem* and show how specific analysis frameworks result in different rankings of standard algorithms for the problem. These frameworks tend to vary the inputs compared between the algorithm and the benchmark rather than varying the benchmark itself. For example, the competitive ratio is the worst case over inputs of the optimal offline performance to the online algorithm's performance; while the max-max ratio compares the worst optimal

offline performance to the worst algorithm performance, with both worst-cases taken over normalized inputs. They also consider analyses that directly compare algorithms rather than comparing the algorithms indirectly via a benchmark.

To place in context the development of robust analyses of mechanisms it is helpful to consider the predominant method for the analysis of mechanisms in economics. In economics, the preferences of the agents are assumed to be drawn from a known distribution and the design framework asks for the mechanism that maximizes performance (and satisfies incentive constraints) in expectation over this prior distribution over preferences. This area is known as *Bayesian mechanism design*, e.g., see the survey by Hartline (2013).

Competitive analysis was introduced to the design of mechanisms by Goldberg et al. (2006). While in online algorithms, the information theoretic barrier to good performance is the lack of information that comes from the online arrival of the input, for dominant strategy incentive compatible mechanism design, the information theoretic barrier to good performance comes from the need to satisfy the incentive constraints of the agents, namely that truthtelling is an equilibrium. The equivalent of the optimal offline algorithm, namely the optimal performance without incentives, is rarely an interesting benchmark as no good ratio is achievable. Both approaches (a) and (b) described above for online algorithms have been taken. For an example of (a), and with parallels to online learning, Goldberg et al. (2006) compares mechanisms for selling a digital good to the optimal revenue from posting a single price to all agents (i.e., the price-posting-revenue benchmark discussed in the introduction). They and the subsequent literature developed the area of prior free mechanism design. In this area, the performance of a mechanism is compared to a benchmark performance in worst-case over inputs.

Lower bounds are important in the study of good mechanisms. Goldberg et al. (2006) introduced the following approach for establishing a lower bound on the prior free approximation of any mechanism to a given benchmark. The approach considers a distribution over inputs for which all (undominated) mechanisms perform the same. For the objective of revenue maximization this distribution is known as the *equal revenue distribution* and has cumulative distribution function $F(v) = 1 - 1/v$ on support $[1, \infty)$. An agent with value drawn from this distribution offered any price $p \geq 1$ accepts with probability $1/p$ and yields expected revenue 1. Goldberg et al. (2006) show a lower bound on the prior free approximation to a posted-price-revenue based benchmark of 2.42 in the limit with the number of agents going to infinity. For the special case of $n = 3$ agents, the lower bound is 3.25; Hartline and McGrew (2005) proved that this lower bound in the $n = 3$ case is indeed tight by giving a mechanism that achieved it. Chen et al. (2014) proved that the lower bound of 2.42 is tight with a non-constructive proof and, moreover, that the lower-bounding method gives a tight bound for a large class of benchmarks that, like the one of Goldberg et al. (2006), are constant in the value of the highest agent.

Hartline and Roughgarden (2008) revisited the choice of benchmark of Goldberg et al. (2006) and identified the normalization constraint. Recall, they observe that if a benchmark satisfies the normalization constraint then mechanisms that approximate the benchmark in worst case also approximate the Bayesian optimal mechanism for any distribution in the class (with no worse an approximation factor). This consequence is known as the *prior independent corollary of approximation of a normalized benchmark.*

A specific normalized benchmark given by Hartline and Roughgarden (2008) is defined by the supremum of the performance of the Bayesian optimal mechanisms for each distribution in the class. For a digital goods auction with a revenue objective,[1] this benchmark is the price-posting-revenue benchmark (from the benchmark design example above). For expert learning, this benchmark is the best-in-hindsight benchmark. For auction settings, Devanur et al. (2015) give a simpler normalized benchmark based on relaxing the incentive constraints to constraints of envy-freedom.

The following gives an outline of the origins of prior independent analysis and compare it to the diffuse adversary approach in online algorithms. The prior independent corollary of prior free approximation of the benchmarks of Hartline and Roughgarden (2008) motivated the consideration of relaxing the assumption of worst-case inputs in a similar fashion to approach (b) above. A key difference, however, between the diffuse adversary model for online algorithms (Koutsoupias and Papadimitriou, 2000) and the prior independent model is that the diffuse adversary model compares algorithms against the optimal offline benchmark (which relaxes the information theoretic constraints of the online optimization problem) where as, in prior independent mechanism design, mechanisms are compared to the optimal mechanism for the distribution (that satisfies the incentive constraints). The advantage of considering distributions in the diffuse adversary model is more in the spirit of smoothed analysis. Dhangwatnotai et al. (2015) considered prior independent mechanism design as a first-order goal and since then it has been the subject of a flourishing area of research.

The prior independent mechanism design framework gives a natural question of identifying the optimal mechanism. This question is framed by a restriction to a class of distributions, but is not subject to an ad hoc performance benchmark as is prior free mechanism design. Previous literature has only identified optimal prior independent mechanisms in environments that are special cases of the fully general problem. Hartline and Roughgarden (2014) gave the prior independent optimal mechanism for revenue maximization in the sale of a single item to a single agent with value from a bounded support where the prior independent optimal mechanism posts a randomized price.

For revenue maximization in the sale of an item to one of two agents with values drawn from an i.i.d. regular distribution, Dhangwatnotai et al. (2015) show that the Second Price Auction is a 2-approximation (i.e., as an upper bound). Fu et al. (2015) gave a randomized mechanism showing that this factor of 2 is not tight. Upper and lower bounds on this canonical problem were improved by Allouah and Besbes (2018) to be within $[1.80, 1.95]$. Recall, a main result of this thesis is to identify the prior independent optimal mechanism for the scale-invariant environment with approximation factor ≈ 1.907.[4] For this two agent problem with i.i.d. values from a distribution in the subset of regular distributions that further satisfy a monotone hazard rate condition, Allouah and Besbes (2018) show that the Second Price Auction is optimal. It can now be observed that the optimality of the Second Price Auction results from constraints binding that are loose when the class

[4] The Allouah and Besbes (2018) lower bound of 1.80 was proved under the same additional assumption of scale invariance as the optimal lower bound of 1.91. Further, they reduce the scale-invariant assumption to an assumption that the allocation for every fixed ratio of values v_1/v_2 exists in the limit as the values themselves approach 0. Technically, they show that if $\lim_{\alpha \to 0} x_i(\alpha \cdot \boldsymbol{v})$ always exists, then the optimal prior-independent mechanism is scale-invariant.

of distributions includes a special type of distribution (Triangle revenue curves, which do not satisfy monotone hazard rate).

Final Introductory Notes. An outline of the rest of the chapters of this thesis is deferred to Section 2.3 where it may generally reference the Preliminaries in Chapter 2. Additional *Related Work* sections for topics are presented locally where each topic is introduced. *Future Work* discussions are included in the Conclusion in Chapter 7.

The author of this thesis is a co-author of **bolded** paper references.

CHAPTER 2

Preliminaries

This thesis considers two directions regarding information settings of robust algorithm design. Predominantly, we study the prior independent setting which takes n inputs independently and identically (i.i.d.) drawn from an unknown distribution from a known class of distributions (Definition 2). Also, as a topic, we study a connection to benchmark design within the prior free setting which takes n inputs arbitrarily from an input space with known support. Our main results are generally *algorithm-agnostic*, i.e., the results describe analyses that do not depend on the underlying algorithm setting.

As an application, this thesis further states and solves applied prior independent results within *mechanism design* which fit in variously to our main, agnostic results. In particular, our application is single-item, truthful auction environments with n single-dimensional agents who have linear utility functions. The predominant objective for the auction designer is to maximize either revenue or residual surplus.

This chapter provides a thorough foundation for the rest of the thesis.

2.1. Setup of Prior Independent Design

Let $\mathcal{F}$ be a class of probability distributions with known fixed support $\mathcal{V}$ (e.g., $[0, \infty)$). In the *prior independent algorithm design setting* (PI), there is a distribution F which is known to come from the class $\mathcal{F}$ and n inputs are drawn i.i.d. from F (thus input space is $\mathcal{V}^n$). Critically, the algorithm designer does not know the specific $F \in \mathcal{F}$. The notation

F is overloaded to be the cumulative distribution function (CDF), and its probability density function (PDF) is f.

(On a technical note, this thesis includes a number of parameterized distributions as components of various analyses. To support notation for these, we an include an explanation of our distribution naming schemes in Appendix B.1.)

Fix an algorithm design problem that takes n i.i.d. inputs. Denote a class of feasible algorithms by $\mathcal{A}$ and an algorithm in this class by A with expected performance $A(\boldsymbol{v})$ for inputs $\boldsymbol{v}$. When evaluating the performance in expectation over inputs drawn from a distribution F, we adopt the notation $A(F) = \mathbf{E}_{\boldsymbol{v}\sim F}[A(\boldsymbol{v})]$. An algorithm's performance for an unknown distribution F is measured against the performance of the optimal algorithm which knows F. With these abstractions, we formally define the Bayesian and prior independent (PI) optimization problems.

Definition 1. *The* Bayesian optimal algorithm design problem *is given by a distribution F and class of algorithms $\mathcal{A}$; and solves for the algorithm* OPT_F *with the maximum expected performance:*

$$(\mathrm{OPT}_F) \qquad \mathrm{OPT}_F = \mathrm{argmax}_{A\in\mathcal{A}}\, A(F).$$

Note that OPT_F is an algorithm. Given a distribution F, the expected performance of the optimal algorithm is $\mathrm{OPT}_F(F)$ and is the *benchmark* that we use for prior independent algorithms:

Definition 2. *The* prior independent algorithm design problem *is given by a class of algorithms $\mathcal{A}$ and a class of distributions $\mathcal{F}$; and searches for the algorithm that minimizes*

its worst-case approximation:

$$(\alpha^{\mathcal{F}}) \qquad \alpha^{\mathcal{F}} = \min_{A \in \mathcal{A}} \left[\max_{F \in \mathcal{F}} \frac{\mathrm{OPT}_F(F)}{A(F)} \right] = \min_{A \in \mathcal{A}} \left[\alpha_A^{\mathcal{F}} \right]$$

where the value of the program α^F is the optimal prior independent approximation factor *for class $\mathcal{F}$ and class $\mathcal{A}$ (which we leave implicit). The bracketed term is the* prior independent approximation guarantee *of a fixed algorithm A given $\mathcal{F}$ and is denoted by $\alpha_A^{\mathcal{F}}$ (as shown on the right).*

There is a well-understood interpretation of $\min - \max$ optimization problems as being 2-player games between (1) in our case, an algorithm designer who "plays first" and commits to an algorithm in the outer min-program, and (2) an *adversary* who has oracle-knowledge of the worst-case arguments to "play second" within the inner max-program.[1]

Subsequently in Chapter 4, this thesis will study lower bounds on prior independent approximation by fixing adversary strategies up front. Given a space Ω, denote the set of all possible distributions by $\Delta(\Omega)$ – i.e., the *probability simplex*. Denote a distribution over elements $\omega \in \Omega$ by $\gamma \in \Delta(\Omega)$. Given a function $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ where Ω_1 and Ω_2 have arbitrary dimensions, we denote the expectation of f over arguments $\boldsymbol{\omega}_i \in \Omega_i$ according to $\gamma_i \in \Delta(\Omega_i)$ as $f(\gamma_i, \boldsymbol{\omega}_{j \neq i}) = \mathbf{E}_{\boldsymbol{\omega}_i \sim \gamma_i} [f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j)]$, e.g., in Theorem 9. In particular, we study adversary distributions $\delta \in \Delta(\mathcal{F})$ which are pre-committed distributions over the distribution-elements of $\mathcal{F}$. (Formally, see Definition 18 for *blends*.) Thus, inputs will be drawn using a two-step process: (1) a random distribution $\hat{F} \sim \delta$ is drawn, and then (2) n inputs are drawn i.i.d. from $\hat{F}$.

[1] Because the abstract concept of an adversary is an inherent consequence of nature rather than being a tangible real-world player, the adversary is generally not subject to algorithm running time constraints in terms of identifying its optimal play.

2.2. Application: Mechanism Design Preliminaries

This section gives a formal introduction to mechanism design as the highlighted application of our general results in the thesis, which otherwise do not depend on the algorithm setting. This section is included in order to provide full support for reference as we prove our technical mechanism design results later (namely, Theorem 5, Theorem 11, and Theorem 12).

2.2.1. Mechanism Design Basics

We consider mechanism design as it relates to *auctions*, i.e., an algorithmic setting of requesting bids from strategic agents, and subsequently allocating items to the agents and charging them monetary payments. The canonical auction consists of maximizing revenue (i.e., agent payments) by selling one item to one of n agents (possibly randomly) who each have a private value for the item drawn i.i.d. from a common probability distribution, with the distribution known by the auction designer (i.e, Definition 1 applied to this setting). The optimal auction to maximize revenue (or other simple objectives, we define common objectives later) in this setting was solved by Myerson (1981).

Each agent $i \in \{1, \ldots, n\}$ has value v_i in a range of known support, e.g., $v_i \in \mathcal{V} = [0, \infty)$ or $v_i \in \mathcal{V} = [1, h]$ for which $\mathcal{V}$ is one agent's value space. Values are private to the agent and are not known by the mechanism. A profile of n agent values is denoted $\boldsymbol{v} = (v_1, \ldots, v_n)$; the profile with agent i's value replaced with z is $(z, \boldsymbol{v}_{-i}) = (v_1, \ldots, v_{i-1}, z, v_{i+1}, \ldots, v_n)$. The list of agent values *in decreasing order* is $v_{(1)}, \ldots, v_{(n)}$.[2]

[2] Re-arranging agents to be labeled in order is typically without loss of generality. For where it is helpful, we further abstractly define $v_{(n+1)} = 0$ to be a default, "sentinel" value.

A *mechanism* collects reports from each agent as *bids* and maps them to (possibly randomized) allocations and payments. A *truthful mechanism* is a special case which takes values $\boldsymbol{v}$ as input rather than arbitrary bids (and must be designed to incentivize agents to report their values truthfully, see Myerson's characterization below in Theorem 1).

Specifically, a *stochastic social choice function* $\boldsymbol{x}$ and a *payment function* $\boldsymbol{p}$ map a profile of values $\boldsymbol{v}$ respectively to a profile of allocation probabilities, and a profile of expected payments. Thus, a truthful mechanism is denoted $M = (\boldsymbol{x}^M, \boldsymbol{p}^M)$. Where the mechanism is clear from context, we will use the simpler notation $M = (\boldsymbol{x}, \boldsymbol{p})$. We may also overload notation and write a mechanism's expected performance as a function $M : \boldsymbol{v}^n \to \mathbb{R}$. A mechanism design problem may restrict the design space to a class of mechanisms $\mathcal{M}$.

For allocation probability x_i and expected payment p_i, the agent's expected utility is linear as $v_i\, x_i - p_i$ and agents maximize utility in expectation. We give the most common objectives for mechanism design as a formal definition:

Definition 3. *The most common objectives for mechanism design are:*

- Revenue *is the sum-total over agent payments:* $\sum_i p_i$.
- Residual surplus *is the sum-total over agent utilities:* $\sum_i x_i v_i - p_i$.
- Total welfare *is the sum-total over agent expected-value-of-allocation:* $\sum_i x_i v_i$*; note that this total respects:* [revenue + residual surplus = total welfare].

With definitions to follow, we restrict attention to mechanisms that are *feasible*, *dominant strategy incentive compatible* (DSIC/ truthful), and *individually rational* (IR), properties which become formal *constraints* for mechanism design. The feasibility constraint for

single-item mechanisms requires that for all inputs $\boldsymbol{v}$, the profile of expected allocations across all agents sums to at most 1. The following DSIC and IR constraints must hold for all agents i, values v_i, and other agent values $\boldsymbol{v}_{-i}$. The DSIC constraint requires that it is always optimal for an agent i to "bid" value true v_i. In this sense, DSIC mechanisms are *truthful*. The IR constraint requires that an agent i always gets non-negative utility by truthfully bidding v_i. The rest of this Section 2.2.1 presents pertinent structures from the mechanism design literature in order to support main results of this thesis which appear in later subsections.

2.2.2. Characterization of Truthful Equilibrium

The following theorem of Myerson (1981) characterizes social choice functions $\boldsymbol{x}$ that can be implemented by truthful (DSIC) mechanisms (i.e., all agents reporting truthfully is a Nash equilibrium).

Theorem 1 (Myerson, 1981)**.** *Allocation and payment rules* $(\boldsymbol{x}, \boldsymbol{p})$ *are induced by a dominant strategy incentive compatible mechanism if and only if for each agent* i*,*

(1) (monotonicity) allocation rule $x_i(v_i, \boldsymbol{v}_{-i})$ *is monotone non-decreasing in* v_i*, and*

(2) (payment identity) payment rule $p_i(\boldsymbol{v})$ *satisfies*

$$p_i(\boldsymbol{v}) = v_i\, x_i(\boldsymbol{v}) - \int_0^{v_i} x_i(z, \boldsymbol{v}_{-i})\, \mathrm{d}z + p_i(0, \boldsymbol{v}_{-i}), \tag{2.1}$$

where the payment of an agent with value zero is often zero, i.e., $p_i(0, \boldsymbol{v}_{-i}) = 0$*.*

Unless stated specifically otherwise in this thesis, we do fix $p_i(0, \boldsymbol{v}_{-i}) = 0$.

2.2.3. Standard Mechanisms

This section describes a number of common auctions. The *Second Price Auction* (SPA) is a special case of the VCG Mechanism which has a number of nice properties: it is a truthful auction, it optimizes total welfare, and it also optimizes revenue when used in conjunction with a correct "reserve price" (which is a minimum price that any agent must pay to be allocated).

Definition 4. *A* reserve price *is a minimum price for allocation regardless of any other considerations, e.g. auction parameters or the realized values of other agents.*

Definition 5. *The single-item* Second Price Auction (SPA) *with n agents allocates the item to an agent with largest value $v_{(1)}$ at a price equal to the second-largest value $v_{(2)}$.*

The SPA is in fact an example of a *k-lookahead auction* (Ronen, 2001) which defines an important class of auctions restricted to those that only ever allocate to the k largest bidders (after ordering and breaking value-ties uniformly at random). The SPA is a 1-lookahead auction.

Definition 6 (Ronen, 2001)**.** *The class of single-item k-*lookahead mechanisms (k-$\mathcal{LA}$) *with $n \geq k$ agents is defined by restriction to mechanisms that only ever give positive allocation to the k agents with largest values $v_{(1)}, \ldots, v_{(k)}$.*

Note, the allocations to large-valued agents may condition on the realized values of the un-allocated, small-valued agents with values $v_{(k+1)}, \ldots v_{(n)}$.

A *markup mechanism* is a special case of 1-lookahead that commits to a markup scalar $r \geq 1$ in advance and offers the price $r \cdot v_{(2)}$ to the largest-valued agent. The SPA is the edge-case markup mechanism with $r^{\text{spa}} = 1$.

Definition 7. *The* r-markup mechanism M_r *offers the price* $r \cdot v_{(2)}$ *to the agent with the largest value* $v_{(1)}$. *A* randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ *draws random* $\hat{r}$ *from a given distribution* $\boldsymbol{\xi}$. *The class of randomized markup mechanisms is* $\mathcal{M}^{\text{mark}}$.

Definition 8. *An* anonymous price posting – *denoted* AP_π – *posts a take-it-or-leave-it common price* π *and randomly allocates to the agents who are willing to pay* π *(i.e., any agent* i *with* $v_i \geq \pi$*).*

Lastly, a *k-lottery* is another special case of k-lookahead mechanism.

Definition 9. *A* k-lottery – *denoted* LOT_k – *is a* k*-lookahead in which a price posting mechanism is used internally: set* $\pi = v_{(k+1)}$ *and allocate randomly to the top* k *agents. Most generally,* the Lottery *mechanism randomly gives away the item for free:* $\text{LOT}_n = \text{AP}_0$.

2.2.4. Myerson Virtual Values

This section introduces a monumentally important component of Myerson's analysis is the concept of *virtual value*. Myerson illustrates how mechanism design and optimization are greatly simplified by using an amortized analysis to calculate performance, specifically by adding up the "marginal" gain (or loss) from serving an agent over all possible agent types as the price is monotonically decreased (weighted by the agent's distribution over values),

according to the mechanism's allocation rule. For derivation of virtual value and further discussion of its intuition, see Myerson (1981) and Chapter 3.3.1 of Hartline (2020).

Fact 1. *Given an agent with value v drawn independently from distribution F, the agent's* virtual value function ϕ^F *(mapping value to virtual value) in an auction fixing each of the following objectives is given by:*

Virtual Values for Revenue: $\phi^F(v) = v - \frac{1-F(v)}{f(v)}$
Virtual Values for Residual Surplus: $\phi^F(v) = \frac{1-F(v)}{f(v)}$
Virtual Values for Total Welfare: $\phi^F(v) = v$

As observed, we let the definition for virtual value be overloaded across objectives. Some results given from the perspective of virtual value are constant across settings, exhibiting the power of virtual values as an analytical tool (e.g. Theorem 2, Theorem 3 below). We end this section with the following useful observation about virtual value functions, which states that virtual value $\phi^F(\hat{v})$ at $\hat{v} \geq z$ is unchanged when a draw from F is conditionally known to be at least z.

Fact 2. *Given a revenue, residual surplus, or total welfare objective, and a distribution F with domain $[a, b]$ (or $[a, b = \infty)$). Let $\overrightarrow{F}^z$ be the distribution resulting from conditioning one random draw $v \sim F$ by $v \geq z$ for $a \leq z \leq b$. Then for $\hat{v} \geq z$,*

$$\phi^F(\hat{v}) = \phi^{\overrightarrow{F}^z}(\hat{v}) \tag{2.2}$$

Fact 2 holds because the operation of conditioning $v \geq z$ applies the same multiplicative factor $1/(1 - F(z))$ to both the $(1 - F(v))$ and $f(v)$ terms appearing in the revenue and

residual surplus virtual value functions – which cancels. For total welfare it is trivially true.

2.2.5. Monotone Hazard Rate, Regular, and Irregular Distributions

This section describes important properties of distributions – namely Definition 10 for monotone hazard rate (MHR) and anti-monotone hazard rate (a-MHR); and Definition 11 for regularity which is related to the definitions of virtual value for various auction objectives. The properties define canonical analytical settings within mechanism design. They affect both the strength and complexity of result statements that can be obtained (by the mechanism design literature generally) by acting as natural restrictions on classes of distributions for robust mechanism design.

Definition 10. *Given a distribution F, its* hazard rate *function $\lambda^F(v) = {}^{f(v)}/_{(1-F(v))}$ describes an "instantaneous rate of failure" of draws from F.* Monotone hazard rate (MHR) *distributions have ${}^{d\lambda^F(v)}/_{dv} \geq 0$ for all inputs v. By comparison,* anti-monotone hazard rate (a-MHR) *distributions have ${}^{d\lambda^F(v)}/_{dv} \leq 0$.*

Let $\mathcal{F}^{\text{mhr}}$ be the class of all MHR distributions and $\mathcal{F}^{\text{a-mhr}}$ be the class of all a-MHR distributions (each within a context of known input support).

We make two observations relating to hazard rate functions. First, note that the classes $\mathcal{F}^{\text{mhr}}$ and $\mathcal{F}^{a\text{-}mhr}$ are disjoint excepting that they share a "boundary" when ${}^{d\lambda^F}/_{dv\,(v)} = 0$ for all inputs v. Second, note that the multiplicative-inverse of hazard rate $(\lambda^F(v))^{-1}$ appeared above in the virtual value function for both revenue and residual surplus objectives.

Definition 11. *For a virtual value function $\phi(\cdot)$ paramterized by a given auction objective, a distribution F is* regular *if $^{d\phi(v)}/_{dv} \geq 0$ for all inputs v. Otherwise it is* irregular*.*

Let $\mathcal{F}^{\text{reg}}$ be the class of all regular distributions (within the context of known input support and a given auction objective).

The following explains relationships between the property-based classes of this section for auctions with specific objectives.

Fact 3. *Given a revenue objective, the class of MHR distributions is a subset of the class of regular distributions, which is a subset of all distributions: $\mathcal{F}^{\text{mhr}} \subset \mathcal{F}^{\text{reg}} \subset \mathcal{F}^{\text{all}}$.*

Given a residual surplus objective, the class of a-MHR distributions and the class of regular distributions are equal: $\mathcal{F}^{\text{a-mhr}} = \mathcal{F}^{\text{reg}} \subset \mathcal{F}^{\text{all}}$.

2.2.6. Quantiles and "Revenue" Curves

For distribution F, the *quantile* q of an agent with value v denotes how weak that agent is relative to the distribution F, i.e., the probability that a random draw from F will be at least v.[3] Technically, quantiles are defined by the mapping $Q_F(v) = 1 - F(v) = \Pr[v' \geq v \mid v' \sim F]$. Denote the function mapping back to value space by V_F, i.e., $V_F(q) = F^{-1}(1-q)$ is the value of the agent with quantile q. Note that all functions defined for all inputs in *quantile space* have domain $[0, 1]$, and that a default random quantile $\hat{q}$ is a uniform draw from the range $[0, 1]$. The rest of this section describing "Revenue Curves" R_F adopts the standard nomenclature of the revenue perspective. However unless

[3] For the places we use it, a *percentile* of a value v is $1 - q$ to reflect how *strong* an agent is relative to distribution F, i.e., the output of the CDF function which is the probability that a random draw from F will be at most v.

otherwise stated, everything presented in this section *for revenue curves in quantile space* holds for alternative objectives if "revenue" was replaced with the correct "performance" measurement.

A single agent *revenue curve* $R_F : [0,1] \to \mathbb{R}$ gives the revenue from posting a price as a function of quantile q, i.e., of the probability that the agent accepts the price. For an agent with value distribution F, price $V_F(q)$ is accepted with probability q, so revenue is $R_F(q) = q\,V_F(q)$. We overload the function R_F to also take inputs from value space, defined by $R_F(v) = R_F(Q_F(v)) = Q_F(v) \cdot v = (1 - F(v))v$.[4] [5] The slope of the revenue curve R'_F is *marginal revenue.*

Fact 4 (Myerson, 1981)**.** *The slope of the revenue curve R'_F – i.e., the marginal revenue function – is equal to virtual value (in quantile space):*

$$\phi^F(q) = R'_F(q)$$

and regular distributions (Definition 11) are equivalent to (weakly) concave revenue curves in quantile space.

Towards analyzing irregular distributions, Myerson implements a second amortization technique called *ironing.* For continuous regions of quantile space where an agent's allocation function has constant output, expected marginal surplus R'_F – equivalently expected

[4] In fact, revenue curves for both quantile space and value space domains are potentially set-valued functions. For quantile space, set-valued outputs result from regions of the domain of F where the CDF is a constant smaller than 1, because revenue (or residual surplus) changes while quantile does not. For value space, set-valued outputs result from regions of quantile space corresponding to point masses, because quantile drops while value is constant.

[5] For residual surplus we will have the overloaded definition $R_F(q) = q \cdot (\mathbf{E}_{\hat{q}}\left[V_F(q) \mid \hat{q} \le q\right] - V_F(q)) = \int_0^q (V_F(\hat{q}) - V_F(q))\,d\hat{q}$. For total welfare it is $R_F(q) = q \cdot \mathbf{E}_{\hat{q}}\left[V_F(q) \mid \hat{q} \le q\right] = \int_0^q V_F(\hat{q})d\hat{q}$.

ϕ^F – can be treated as its average value of the region, at all points in the region. Technically, given an ironed region $[a, b] \in [0, 1]$, the *ironed marginal revenue* at all quantiles $q \in [a, b]$ is $R_F(a) + \frac{R_F(b)-R_F(a)}{b-a} \cdot (q - a)$.

We define a single agent *ironed revenue curve* within the context of optimal analysis (rather than allowing arbitrary choice of ironed regions). A single agent ironed revenue curve $\bar{R}_F : [0, 1] \to \mathbb{R}$ is defined only for the quantile space domain (and not for value space), and is defined as the *concave hull* of the original revenue curve (which is always possible to achieve for a single agent by ironing exactly all of the regions of the domain where the revenue curve and its concave hull are not already equal). The definition takes advantage of the following.

Fact 5. *Given a distribution F for a single agent, for fixed $\hat{q}$ as an* a priori *fixed* probability of sale*, the maximum revenue achievable given $\hat{q}$ is $\bar{R}_F(\hat{q})$.*

The regions where the revenue curve and ironed revenue curve are not equal are described as *strictly ironed*. Paralleling Fact 4, we have a corresponding definition for *ironed virtual value* $\bar{\phi}^F$, which is again equal to the slope of the ironed revenue curve as marginal ironed revenue.

Fact 6 (Myerson, 1981)**.** *The slope of the ironed revenue curve $\bar{R}'_F$ – i.e., the marginal ironed revenue function – is equal to ironed virtual value $\bar{\phi}^F$:*

$$\bar{\phi}^F(q) = \bar{R}'_F(q)$$

Continuing, note the following are both (weakly) concave functions: (1) revenue curves for regular distributions and (2) ironed revenue curves for irregular distributions. The remaining observations of this section identify consequences of this geometry.

Fact 7. *The concavity of (1) revenue curves for regular distributions and (2) ironed revenue curves for irregular distributions implies that both are non-increasing in quantile.*

In comparison to value space, the uniformity of the quantile space domain is relatively simpler for analysis, having already encapsulated the density function $f(\cdot)$. Thus, it is more intuitive to illustrate revenue curves using quantile space. For the revenue objective specifically, values have a geometric interpretation when the revenue curve is drawn in quantile space, which we describe in Fact 8.

Fact 8. *(For the revenue objective specifically, there is a bijection between values $v \in [0, \infty)$ and the slopes of rays coming out of the origin in the revenue curve graph with quantile space domain. Explicitly, a point $(\hat{q}, \hat{R} = R_F(\hat{q}))$ of a revenue curve R_F necessarily implies $R_F(V_F(\hat{q})) = V_F(\hat{q}) \cdot \hat{q} = \hat{R}$. Equivalently, $R_F(v) = \hat{R}$ if and only if $Q_F(v) = \hat{q}$.*

Lastly we consider distributions that include point masses. Interpreted within a revenue curve, a point mass at v corresponds to a continuous region $[a, b] \in [0, 1]$ of quantile space with measure equal to the point mass' discrete probability measure. A specific distribution F (incorporating v) maps v to the quantile-range $a = 1 - \lim_{x \to v^+} F(x)$ and $b = 1 - \lim_{x \to v^-} F(x)$.

We consider the formal interpretation of allocating with fixed probability $\hat{q}$ when it requires partial allocation to a value with a point mass, i.e., when we strictly have

$\hat{q} \in (a, b)$. The solution is to internally re-weight allocation at v by a (diminishing) multiplicative factor of $(\hat{q} - a)/(b - a)$. This becomes transparent in the geometry. We give some final information relating to the geometry of a point mass within a revenue curve:

Fact 9. *Consider a distribution F with point mass at value v with implicit measure $\lim_{x \to v^+} F(x) - \lim_{x \to v^-} F(x)$ and with definitions for a, b in the immediately preceding text.*

For the revenue objective specifically, the geometric interpretation of a point mass follows directly from Fact 8. A point mass is graphed into the revenue curve by: restricting the line segment between $(0, 0)$ and $(b, R_F(b))$ to inputs in $[a, b]$. A necessary identity is that the slope R'_F of this line segment is the value v.

This follows directly from $R'_F(v) = \phi^F(v) = v - (1 - F(v))/f(v) = v$ because $f(v) = \infty$.

As a further consequence for revenue: regular distributions can only incorporate point masses as the (closed) upper bounds of their domains. The contrapositive statement is: any distribution with positive measure of density strictly above a point mass is irregular.

For the residual surplus objective specifically, the geometric interpretation of a point mass is a horizontal line segment on $[a, b]$ (because all probabilities-of-sale $\hat{q} \in [a, b]$ require posting price v and all agent-quantiles in this range have value v for which the objective is 0 regardless if they are included or not, so the derivative here is 0).

This follows directly from $R'_F(v) = \phi^F(v) = (1 - F(v))/f(v) = 0$ because $f(v) = \infty$. Regular distributions must not have a lower bound on their support $a > 0$ where $F(a) = 0$. The contrapositive statement is: any distribution F with $F(a) = 0$ for $a > 0$ is irregular.

As a further consequence for residual surplus: regular distributions with positive density can only incorporate a point mass at $v = 0$ as the (closed) lower bound of their domains. The contrapositive statement is: any distribution with positive measure of density strictly below a point mass is irregular.

2.2.7. Optimal Bayesian Mechanisms and Foundational Results

This section summarizes pertinent results in mechanism design. Optimizing revenue from a single agent whose value v is drawn from a known distribution F is straightforward.

Fact 10. *For any distributions F, the optimal mechanism for a single agent posts the* monopoly price $V_F(q_m)$ *(Definition 8) corresponding to the* monopoly quantile $q_m = \text{argmax}_q R_F(q)$.

Next we work towards the solution for Bayesian settings with n agents (Theorem 3 below). First we show technically how mechanism performance can be measured using virtual values – a mechanism's revenue can be understood via the marginal revenue approach of Myerson (1981) and Bulow and Roberts (1989).

Simply, the expected revenue of a mechanism M with n agents is equal to its expected surplus of marginal revenue, equivalently, its expected surplus of virtual value. Theorem 2 gives two related statements and the differences are bolded.[6]

Theorem 2 (Myerson, 1981)**.** *Given any incentive-compatible mechanism M with* ***any*** *allocation rule $\boldsymbol{x}(\cdot)$, the expected revenue of mechanism M for agents with values*

[6] Per previous discussion regarding setting, the exact theorem statement of Theorem 2 holds for alternative objectives with their respective definitions of ϕ^F and $\bar{\phi}^F$.

drawn i.i.d from F is equal to its expected surplus of virtual value, i.e.,

$$M(F) = \sum\nolimits_i \mathbf{E}_{\boldsymbol{v}\sim F}\left[p_i(\boldsymbol{v})\right] = \sum\nolimits_i \left(\mathbf{E}_{\boldsymbol{v}\sim F}\left[\phi_i^F(v_i)\, x_i(\boldsymbol{v})\right] + R_{i,F}(q_i = 0)\right) \tag{2.3}$$

Alternatively, given any incentive-compatible mechanism M with allocation rule $\boldsymbol{x}(\cdot)$, the expected revenue of mechanism M for agents with values drawn i.i.d. from F is equal to its expected surplus of ***ironed*** *virtual value* ***if additionally $\boldsymbol{x}(\cdot)$ is constant for each agent i on regions that are strictly ironed by $\bar{R}_{i,F}$****:i.e., then*

$$M(F) = \sum\nolimits_i \mathbf{E}_{\boldsymbol{v}\sim F}\left[p_i(\boldsymbol{v})\right] = \sum\nolimits_i \left(\mathbf{E}_{\boldsymbol{v}\sim F}\left[\bar{\phi}_i^F(v_i)\, x_i(\boldsymbol{v})\right] + R_{i,F}(q_i = 0)\right) \tag{2.4}$$

The Bayesian optimal single-item auction OPT_F given F is the one that maximizes expected surplus of virtual value, or equivalently, the one that maximizes ironed virtual value.

Theorem 2 gives a description of mechanism performance using a reduction to virtual value. It includes an abstract description of the optimal mechanism following directly at the end of the theorem statement. The following corollary makes explicit the optimal structure for auctions within the setting of regular distributions $\mathcal{F}^{\mathrm{reg}}$:

Theorem 3 (Myerson, 1981)**.** *For i.i.d., regular, single-item auctions with any objective, the optimal mechanism OPT_F is the second-price auction with reserve price equal to the monopoly price.*

We will use the following Lemma 1 for the calculation of the performance of specific optimal mechanisms for our dual blends analyses (towards proving the revenue gap of Theorem 11 in Section 4.5.3 and the residual surplus gap of Theorem 12 in Section 4.5.4).

For the revenue objective and specifically $n = 2$, evaluating a mechanism's performance via virtual values has a natural, geometric interpretation. An extension of this lemma is given in Appendix A.2 for use there.

Lemma 1 (Dhangwatnotai et al., 2015)**.** *In i.i.d. two-agent single-item settings, the expected revenue of the second price auction is twice the area under the revenue curve and the expected revenue of the optimal mechanism is twice the area under the smallest monotone concave upperbound of the revenue curve.*

The last foundational result in this section is due to Bulow and Klemperer (1996), the structure of which later motivates some our specific directions of analysis. For the revenue objective and a regular distribution F, it relates the performance of the performance of the SPA with $n + 1$ agents to the optimal auction with n agents.

Theorem 4 (Bulow and Klemperer, 1996)**.** *For any auction objective, fix a regular distribution F. The (prior independent) SPA with $n + 1$ agents whose values are drawn i.i.d. from F has expected revenue that is at least the expected revenue from the optimal auction* OPT_F *for n agents which knows the distribution F.*

2.2.8. Distribution-Class Boundaries and Equal "Revenue" Distributions

This section gives technical description relating to structure and usage for some of the most pertinent distributions in mechanism design. The distributions discussed in this section are important because they have one or both of the following properties (in the context of one of the relevant auction objectives for this thesis): (a) the distribution

defines a *boundary* of the MHR/a-MHR or regular classes of distributions; and/or (b) the distribution has *constant virtual value* at all values of its domain.

Not surprisingly, a major theme from identifying these distributions is that particular *boundary* distributions which meet definitions of class-restrictions with equality are the same ones used to prove tightness in a variety of theorem statements. We give an example of this below in Fact 12 (and further, recall the example in the Introduction using equal revenue distribution / point mass distributions). To these ends, notice that MHR and regularity are both properties relating to monotonicity of functions parameterized by distributions, respectively hazard rate and virtual value function.

Definition 12. *A distribution is a* boundary distribution *for a given class when its characterizing derivative evaluates to a constant* 0 *for all (relevant) inputs.*

The following fact describes some of these characteristics for distributions that act as boundaries for classes requiring the MHR, a-MHR, or regularity properties.

Fact 11. *The following are true about distribution class boundaries.*

(1) For revenue, the boundary of the MHR class of distributions ($\mathcal{F}^{\text{mhr}}$) – requiring the derivative of hazard rate be equal to 0 *on upwards closed domain– are as follows:*

- *the general case where hazard rate is a constant (β):*

  is the shifted exponential distribution 'Sed' *parameterized by its (shifted)*

domain lower bound $a \geq 0$ and its hazard rate $\beta > 0$:

$$\mathrm{Sed}_{a,\beta}(v) = 1 - e^{-\beta(v-a)} \text{ for } v \in [a, \infty) \qquad \lambda^{\mathrm{Sed}_{a,\beta}}(v) = \beta$$

$$\mathrm{sed}_{a,\beta}(v) = \beta \cdot e^{-\beta(v-a)} \qquad d\lambda^{\mathrm{Sed}_{a,\beta}}/dv\ (v) = 0$$

- *a special case where hazard rate is infinite (∞):*

  is the point mass function 'Pmd' *(c.f., a Dirac function[7]) parameterized by constant output a; this is effectively derived from the shifted exponentials in the general case using in-the-limit analysis as ($\beta \to \infty$):*

  $$\mathrm{Pmd}_a(v) = 1 \text{ at } \ v = a \qquad \lambda^{\mathrm{Pmd}_a}(v) = \infty$$

  $$\mathrm{pmd}_a(v) := \infty \qquad d\lambda^{\mathrm{Pmd}_a}/dv\ (v) = 0$$

(2) For residual surplus,

- *the (common) strong-boundary of the MHR and a-MHR classes of distributions ($\mathcal{F}^{\mathrm{mhr}}$ and $\mathcal{F}^{\text{a-mhr}}$) is the specific exponential distribution* 'Exd' *(which requires fixing lower bound $a = 0$) parameterized only by its hazard rate*

[7] If formal definitions are necessary, we use the following Dirac function technique, which most naturally aligns with the formal definitions needed to evaluate integrals in the limit $dx \to 0$,

$\mathrm{Pmd}_a := \lim_{dx \to 0}[\mathrm{Pmd}_a^{dx}(x) = (x-a)/dx \text{ for } x \in [a, a+dx]]$, $\mathrm{pmd}_a := \lim_{dx \to 0}[\mathrm{pmd}_a^{dx}(x) = 1/dx]$

and if necessary, a point mass at the upper endpoint h of an integral is modified to $h + dx$ (which is inconsequential because it is inside the limit). However we trust that this thesis's computation of expected values of functions over inputs drawn from distributions incorporating point masses is clearly correct; which is: by separating out the contribution of the function value at the point mass as an additive term with probability weight equal to the point mass.

$\beta > 0$:

$$\mathrm{Exd}_\beta(v) = 1 - e^{-\beta v} \text{ for } v \in [0, \infty) \qquad \lambda^{\mathrm{Exd}_\beta}(v) = \beta$$

$$\mathrm{exd}_\beta(v) = \beta \cdot e^{-\beta v} \qquad d\lambda^{\mathrm{Exd}_\beta}/dv\ (v) = 0$$

- *in fact $\mathcal{F}^{\text{a-mhr}} = \mathcal{F}^{\text{reg}}$ (from Fact 3), so the boundary of the class of regular distributions $\mathcal{F}^{\text{reg}}$ is again the exponential distribution* Exd. *They are the same because the virtual value function (given distribution F) for residual surplus is equal to $1/(\lambda^F)$ and must be non-decreasing to be regular, and similarly the hazard rate (λ^F) must be non-increasing to be a-MHR. Clearly these are equivalent conditions. Specifically, we have:*

$$\phi^{\mathrm{Exd}_\beta}(v) = 1/\beta, \quad d\phi^{\mathrm{Exd}_\beta}/dv\ (v) = 0$$

(3) For revenue, the boundaries of the class of regular distributions $\mathcal{F}^{\text{reg}}$ – requiring the derivative of virtual value be equal to 0 *– are as follows:*

- *an important special case where virtual value is the constant* 0*:*

  *is the quadratic distribution '*Qud*' parameterized by its domain lower bound $a > 0$:*

$$\mathrm{Qud}_a(v) = 1 - a/v \text{ for } v \in [a, \infty) \qquad \phi^{\mathrm{Qud}_a}(v) = 0$$

$$\mathrm{qud}_a(v) = a/v^2 \qquad d\phi^{\mathrm{Qud}_a}/dv\ (v) = 0$$

- *the general case where virtual value is a constant (ϕ):*

  *is the shifted quadratic distribution '*Sqd*' parameterized by its domain lower*

bound $a \geq 0$ and its shift $\phi < a$ (for which there is no lower bound on ϕ and for which setting $\phi = 0$ gives the previous special case):

$$\text{Sqd}_{a,\phi}(v) = 1 - (a-\phi)/(v-\phi) \text{ for } v \in [a, \infty) \qquad \phi^{\text{Sqd}_{a,\phi}}(v) = \phi$$

$$\text{sqd}_{a,\phi}(v) = (a-\phi)/(v-\phi)^2 \qquad d\phi^{\text{Sqd}_{a,\phi}}/dv\ (v) = 0$$

- *a special case where virtual value is a positive constant ($a = vv > 0$): is the point mass function* 'Pmd' *(c.f., a Dirac function[7]) parameterized by constant output a; this is effectively derived from the general case using in-the-limit analysis as ($vv \to a^-$):*

$$\text{Pmd}_a(v) = 1 \text{ at } \ v = a \qquad \phi^{\text{Pmd}_a}(v) = a$$

$$\text{pmd}_a(v) := \infty \qquad d\phi^{\text{Pmd}_a}/dx\ (v) = 0$$

In cases (3b), the general class of shifted quadratics summarizes a very simple geometric interpretation for revenue curves in quantile space: for inputs in $(0, 1]$, it includes all non-negative line segments as outputs. Specifically, the revenue curve for $\text{Sqd}_{a,\phi}$ is a line segment with slope ϕ connecting an (open) point $(0, a-\phi)$ and a (closed) point $(1, a)$.

As previously mentioned, class-boundary distributions are frequently used to prove that theorem statements are tight and we give an example here.

Fact 12 (Bulow and Klemperer, 1996)**.** *Theorem 4 is tight for regular class-boundary distributions $\text{Sqd}_{0,\phi}$ for all $\phi < 0$, i.e., both the SPA with $n+1$ agents drawn i.i.d. from $\text{Sqd}_{0,\phi}$ and $\text{OPT}_{\text{Sqd}_{0,\phi}}$ with n i.i.d agents have expected revenue equal to $(-\phi) \cdot n > 0$.*

As alluded in Fact 11, quadratic distributions Qud_a are the $\phi = 0$ special case of the shifted quadratic distributions $\text{Sqd}_{a,0}$. Quadratic distributions play an important role in auction design for the revenue objective, where they are examples of *equal revenue distributions* (EQRs). Equal revenue distributions have the following definition and properties (Fact 13).

Definition 13. *A distribution is an* equal revenue distribution (EQR) *if* all *1-agent price posting auctions have the same expected revenue.*

Generally, the Quadratics Qud_a describe exactly the class of *regular* equal revenue distributions (and they maintain both the regularity and equal-revenue properties under top-truncation).

Fact 13. *The following are true about equal revenue distributions.*

- *A sufficient condition for a distribution with domain* $[a, \infty)$ *to be an* equal revenue distribution (EQR) *is that its virtual value function evaluates to* 0 *at all quantiles* $q \in (0, 1]$ *corresponding to values at least* a*. All quadratic distributions* Qud_a *meet this condition.*
- *Consider a Bayesian auction (of Definition 1) for a revenue objective with 1 agent whose value is drawn from a quadratic distribution* Qud_a*. The expected revenue of any price-posting auction with price* $\pi \in [a, \infty)$ *is* a*, i.e., posting any price* $\pi \geq a$ *gets* equal revenue.

There exists an indirect extension of "EQR" to the residual surplus setting which we name EQRS. An indirect extension of the EQR-concept is necessary because virtual values for

residual surplus are strictly positive everywhere except at $q = 0$ where they are 0 (see Fact 1) and on point mass regions of quantile space, so there is no analogous distribution (in an auction with 1 agent) that achieves equal residual surplus for all posted prices.

For residual surplus, the key adaptation towards establishing an EQRS is to require equal performance of price posting *critically within the context of knowing that the price will be accepted by at least 1 agent* (i.e., an agent with unknown larger value). Equivalently, the auction (a) must always sell to (b) an agent with unconditionally-constant virtual value.

For residual surplus, assuming a price will trade is fairly natural because heuristically this condition holds at price 0 where all virtual values are positive. Intuitively, given constant virtual values, only an irrational auction would increase a posted price to a level at which it might not trade. Further, exponential distributions are the natural class of EQRSs under this condition, which makes sense because exponentials are the unique boundary of regularity for residual surplus (cf. the equal revenue condition versus border-regularity condition for revenue). We reinforce these intuitions by **bolding** in Fact 14 the key differences of EQRS, in comparison to revenue and EQR.

Fact 14. *Consider a Bayesian residual surplus auction* ***with*** $\mathbf{n \geq 1}$ ***agents*** *whose values are drawn i.i.d. from an exponential distribution* Exd_β*. The expected residual surplus of any of the following k-lookahead auctions* ***which are all guaranteed to trade (without conditioning on the value of any winning agent)*** *is* $1/\beta$*: for values* $v_{(1)} \geq v_{(2)} \ldots \geq v_{(n)}$*, a k-lookahead auction for any* $k \in \{2, \ldots, n+1\}$ *which offers respectively prices* $\pi \in \{v_{(2)}, \ldots, v_{(n)}, 0\}$ *gets equal residual surplus.*

Sufficient conditions for a distribution to be an equal residual surplus distribution (EQRS) ***are that its virtual value function evaluates to a constant*** β ***at all quantiles*** $q \in (0, 1]$***, and the distribution has domain lower bound at 0.*** *The class of exponential distributions* Exd_β *meet this condition.*

We conclude with "canonical" definitions for equal revenue / residual surplus distributions, because these unique, simple forms are frequently sufficient for result statements.

Definition 14. *The* canonical equal revenue distribution – i.e., the equal revenue distribution – *is* $\text{Eqrd} = \text{Qud}_1$.
The canonical equal residual surplus distribution – i.e., the equal residual surplus distribution – *is* $\text{Eqrsd} = \text{Exd}_1$.

2.3. Thesis Outline

The rest of this thesis is laid out as follows. We start with applied, *optimal* prior independent design. Chapter 3 gives the solution for a long-standing open question in mechanism design going back to Dhangwatnotai et al. (2010) (and implicated by Hartline and Roughgarden (2008)), namely it identifies the prior independent revenue-optimal mechanism for selling a single item to two agents with i.i.d. values from a regular distribution (Theorem 5; under a scale-invariance assumption which is conjectured to be without loss).

Chapter 4 relaxes the focus to identifying *lower bounds* on prior independent approximation. We develop the Blends Technique (Definition 19) which is an explicit method for applying Yao's Minimax Principle (Theorem 9) to the prior independent setting. The construction of lower bounds from the Blends Technique will apply for any prior independent

algorithms problem and hold without characterizing the optimal algorithm. Rather, these lower bounds only depend on computation of expected *Bayesian optimal* performance for distributions in the prior independent reference class (Theorem 10). Chapter 4 includes examples of *dual blend solutions* for the approach of the Blends Technique. Further, applies these examples within mechanism design to state novel lower bounds (for settings similar but not identical to the setting of Chapter 3).

Chapter 5 presents an expanded analysis of the structure of the Blends Technique. The main dual blends example of Chapter 4 is generalized in two distinct ways to state the existence of two broad classes of such solutions that may fit into the Blends Technique. Theoretical optimization of the Blends Technique for establishing lower bounds – which still may be with loss compared to optimal prior independent algorithm design – is connected to *information design*. A natural perspective that arises from considering blends within information design is to assess whether or not the two "sides" of a dual blend have a relationship in terms of Blackwell ordering. Lastly, dual blends themselves are explained to be a special case of non-unique tensor decomposition.

Chapter 6 considers how the prior independent setting may connect to the prior free setting (i.e., worst-case over inputs / competitive analysis) via *benchmark design*. This is implemented by an idea to require that prior free benchmarks are design in a way that an algorithm's prior free approximation implies that the algorithm achieves the same prior independent approximation (a property called *normalization*). We show that if the optimal normalized benchmark is selected as the one which allows smallest approximation by an algorithm, then the problem of benchmark design is equivalent to the problem of prior independent approximation (Theorem 16). Benchmark design from this first

perspective is also analyzed for the online-algorithms problem of *expert learning* and this analysis raises certain challenges for this approach. A second measure to determine the "optimal" benchmark is also presented and then the two benchmark design problems of this chapter are compared.

Chapter 7 concludes this thesis by highlighting its main themes. Respecting the order of presentation in Chapter 7, these themes are: Benchmark Design, Adversarial Play (and the Blends Technique, and the Role of Scale-invariance. Included in the concluding discussion for each topic is an outline for directions of future work. One additional theme is featured that does not otherwise appear in this thesis to cover other works by the author: Economic Inference in Auctions **(Hartline et al., 2019, 2020b)**.

CHAPTER 3

A Revenue-optimal Prior Independent Mechanism

In this chapter we solve the following optimization problem in mechanism design. The setting is a prior independent single-item, 2-agent auction setting with a revenue objective. The class of distributions is regular $\mathcal{F}^{\text{reg}}$ and the class of mechanisms is truthful (DSIC), individually rational and *scale invariant* mechanisms $\mathcal{M}^{\text{si}}$, i.e., mechanisms which respect allocations $x_i(k \cdot \boldsymbol{v}) = x_i(\boldsymbol{v}) \ \forall \ \boldsymbol{v}, \ k > 0$. The consequences of scale-invariance are: mechanism performance is everywhere linear in inputs ($M(k \cdot \boldsymbol{v}) = k \cdot M(\boldsymbol{v}) \ \forall \ \boldsymbol{v}, \ k > 0$) and we may assume a default scale for analysis.

This chapter is organized as follows. Section 3.1 states the main result up front, which includes a description of the optimal mechanism and the optimal choice of distribution by the adversary in order to measure worst-case prior independent approximation. Section 3.2 gives definitions and supporting material as needed for this chapter.

Section 3.3 proves optimality of our solution (as mixed Nash, i.e., mechanism and worst-case distribution are mutual best-responses) using further restrictions on the mechanism space and distribution space in order to simplify to a local analysis (specifically, we first restrict to random markup mechanisms (Definition 7) and triangle revenue-curve distributions (Definition 16)). Finally, Section 3.4 reduces the original mechanism space $\mathcal{M}^{\text{si}}$ and distribution space $\mathcal{F}^{\text{reg}}$ of the broad problem to the local setting (in the context

of the locally-proposed actions of mechanism designer and adversary). This final reduction finally shows that the local analysis is sufficient to apply overall and establishes the general theorem statement.

Material in this chapter is largely drawn from **Hartline et al. (2020a)**.

3.1. Statement of the Optimal Mechanism Result

This section gives our optimal result (which makes the assumption of scale-invariance). In particular, we identify the optimal approximation factor to be $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$.

Theorem 5. *Given a single item, 2-agent auction with a revenue objective, the optimal truthful, scale-invariant mechanism (from the class $\mathcal{M}^{\text{si}}$) against regular distributions $\mathcal{F}^{\text{reg}}$ for the prior independent design program ($\alpha^{\mathcal{F}^{\text{reg}}}$) is $M_{\hat{r},\boldsymbol{\xi}}$ which randomizes according to $\boldsymbol{\xi}$ over the second-price auction M_1 with probability ξ_1 and r^*-markup mechanism M_{r^*} with probability $\xi_{r^*} = 1 - \xi_1$, where $\xi_1 \approx 0.806$ and $r^* \approx 2.447$. The worst-case regular distribution is $\text{Trvd}_{q_m^*}$ with its monopoly quantile $q_m^* \approx 0.093$ and its approximation ratio is $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$.*

Some notation to fully explain this statement is deferred to the next section. To summarize, the optimal mechanism is an a priori mixture over the SPA and a specific markup mechanism, effectively a mixture over markup factors of $r^{\text{spa}} = 1$ and $r^* \approx 2.447$ (see action definitions in Section 2.2.3). The adversary's optimal choice of $\delta^*(\mathcal{F}^{\text{reg}})$ is realized as a point mass on a single triangle revenue-curve distribution $F^* = \text{Trvd}_{q_m^*}$ with $q_m^* \approx 0.093$ (see Definition 16).

3.2. Supporting Definitions and Structures

This section covers new definitions and facts as needed by this chapter. The statement and proof of Theorem 5 depend on analysis of revenue curves R_F for every $F \in \mathcal{F}^{\text{reg}}$. Given the restriction to scale-invariant mechanisms, it will be sufficient to consider only distributions that are *anonymously-scaled*, and specifically, they are standardized so that the single-agent optimal revenue defaults to $\max_q R_F(q) = 1$.

Fact 15. *Given a distribution $F_{z=1}$ with default scaling parameter $z = 1$ and with domain $[a, b]$ (or domain $[a, \infty)$). The distribution F_1 can be arbitrarily re-scaled for $z \in (0, \infty)$ to $F_z(x) = F_1(x/z)$ with domain $[z \cdot a, z \cdot b]$ (respectively domain $[z \cdot a, \infty)$).*

Revenue curves scale linearly: $R_{F_k} = k \cdot R_F$. For example, if $k = 1/2$, $F = \text{Ud}_{0,2}$, and $F_{1/2} = \text{Ud}_{0,1}$, then $R_{\text{Ud}_{0,1}} = \frac{1}{2} \cdot R_{\text{Ud}_{0,2}}$. Further, the ratio of optimal performance $\text{OPT}_F(F)$ to the performance of any scale-invariant mechanism $M^{si}(F)$ is constant under such re-scaling.

The rest of this section leads up to the definitions of triangle and quadrilateral *revenue-curve* distributions. We emphasize that these distributions are named based on the shapes of their revenue-curves, rather than the shapes of their density functions.

Definition 15. *Given the class of Shifted Quadratics with CDF* $\text{Sqd}_{a,\phi}(x) = 1 - (a-\phi)/(x-\phi)$ *and PDF* $\text{sqd}_{a,\phi}(x) = (a-\phi)/(x-\phi)^2$ *for $x \in [a, \infty)$. For revenue auctions, a* constant negative virtual value distribution Cnvd_ϕ *with $\phi < 0$ is parameterized by its (negative) virtual value ϕ and is the* $\text{Sqd}_{0,\phi<0}$ *special case of shifted-quadratic distribution. It has CDF* $\text{Cnvd}_\phi(x) = 1 + \phi/(x-\phi)$ *and PDF* $\text{cnvd}_\phi(x) = -\phi/(x-\phi)^2$. *Define the sub-class of Constant Negative (revenue) Virtual Value distributions* $\text{CNVV} = \{\text{Cnvd}_\phi \ : \ \phi < 0\}$.

For a revenue objective, the (distribution) elements of the CNVV class are equivalently defined by all revenue curves (parameterized by negative ϕ) that are simple line segments connecting a point (in quantile revenue curve space) $(0, -\phi)$ to $(1, 0)$ with slope ϕ.

Recall from Fact 9 that point masses appear in revenue curves as (positively-sloped) line segments and thus, truncating a distribution (see Appendix B.1) at a specific value v (equivalently quantile $Q_F(v)$) will convert the revenue curve to a line segment between the origin $(0, 0)$ and the point $(Q_F(v), R_F(v))$. The theme of monopoly-truncation appears in the next two definitions for triangle and quadrilateral *revenue-curve* distributions. Triangles are top-truncated CNVVs. Quadrilaterals apply top-truncation to underlying revenue curves described by 2 regions of constant negative virtual values. See Figure 3.1 for the intuition of both classes of distributions.

Definition 16. *A* default-scaled triangle revenue-curve distribution Trvd_{q_m} *is parameterized by its monopoly quantile* q_m*, and is defined by the (set-valued) quantile function:*

$$Q_{\text{Trvd}_{q_m}}(v) = \begin{cases} \frac{1}{1+v(1-q_m)} & \text{for } 0 \leq v < 1/q_m \\ [0, q_m] & \text{for } v = 1/q_m \\ 0 & \text{otherwise} \end{cases}$$

Equivalently the revenue-curve distribution Trvd_{q_m} *is given by the truncation at* $v = 1/q_m$ *of the CNVV distribution* $\text{Sqd}_{0,(-1/1-q_m)}$*. The class of default-scaled Triangles is*

$$\mathcal{F}^{\text{trv}} = \{\text{Trvd}_{q_m} : q_m \in [0, 1]\}$$

and contains both the (untruncated) CNVV distribution $\mathrm{Sqd}_{0,-1}$ *and the (effectively "fully-truncated") point mass distribution* Pmd_1 *as edge cases.*[1] *Triangles are exhibited in Figure 3.1.*

The triangulation *of a default-scaled (regular) distribution* F *with monopoly quantile* q_m *is the conversion of* F *to the corresponding triangle* Trvd_{q_m}*.*

Definition 17. *A* default-scaled quadrilateral revenue-curve distribution $\mathrm{Qrvd}_{q_m,\bar{q},r}$ *is parameterized by its monopoly quantile* $q_m \in [0,1]$*, a second "inflection" quantile* $\bar{q} \in [q_m, 1]$*, and a (markup factor) ratio* r *with* $1/\bar{q} \leq r \leq (1-\bar{q})/\bar{q}(1-q_m)$ *which is the ratio between the monopoly slope and the slope through its inflection point,*[2] *and is defined by the (set-valued) quantile function:*

$$Q_{\mathrm{Qrvd}_{q_m,\bar{q},r}}(v) = \begin{cases} \frac{\bar{q}}{\bar{q}+vrq_m(1-\bar{q})} & \text{for } 0 \leq v \leq 1/rq_m \\ \frac{\bar{q}q_m(r-1)}{vrq_m(\bar{q}-q_m)+(rq_m-\bar{q})} & \text{for } 1/rq_m \leq v < 1/q_m \\ [0, q_m] & \text{for } v = 1/q_m \\ 0 & \text{otherwise} \end{cases}$$

[1] Consider the CNVV with virtual value equal to $k \to -\infty$ with effectively "all" density above value 1, and truncate it at value 1.

[2] Note that the quantile $\bar{q}$ together with a feasible markup factor r identifies exactly the inflection point (on the quantile-space revenue curve graph) to the right of the peak of a quadrilateral distribution. If we choose the parameters in the order $q_m, r, \bar{q}$ (rather than $q_m, \bar{q}, r$ as in the definition), the feasible ranges are $q_m \in [0,1]$, then $r \geq 1$, then $\frac{q_m r}{q_m r+(1-q_m)} \leq \bar{q} \leq \min\{rq_m, 1\}$.

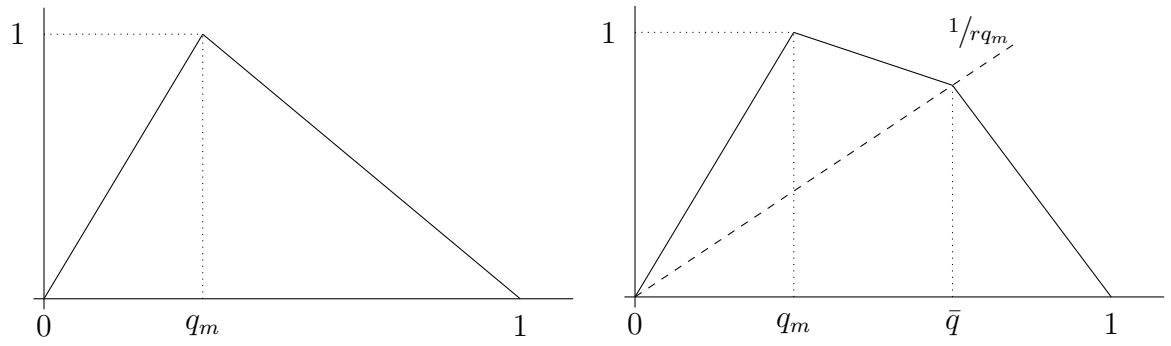

Figure 3.1. Revenue Curves for Triangle and Quadrilateral Distributions
The left hand side is the revenue curve for triangle distribution Trvd_{q_m} and the right hand side is the revenue curve for quadrilateral distribution $\text{Qrvd}_{q_m,\bar{q},r}$.

Equivalently the revenue-curve distribution $\text{Qrvd}_{q_m,\bar{q},r}$ *is given by monopoly truncation of a distribution that is 2-piecewise composed of Shifted Quadratics.*[3] *The class of default-scaled Quadrilaterals is*

$$\mathcal{F}^{\text{qrv}} = \{\text{Qrvd}_{q_m,\bar{q},r} : q_m \in [0,1],\ \bar{q} \in [q_m, 1],\ 1/\bar{q} \leq r \leq (1-\bar{q})/\bar{q}(1-q_m)\}$$

and contains edge cases that are analogous to those of Triangles (Definition 16, due to either being untruncated or a point mass), and edge cases that are Triangles themselves. Quadrilaterals are exhibited in Figure 3.1.

The quadrilateralization *of a default-scaled (regular) distribution F with monopoly quantile q_m is parameterized by (feasible) r and is the conversion of F to the corresponding quadrilateral* $\text{Qrvd}_{q_m,\bar{q}(q_m,r),r}$ *where the slope* $1/rq_m$ *fixes the resulting quadrilateral's "inflection point" by intersection with the revenue curve R_F, thereby identifying $\bar{q}$.*

[3] For technical completeness. On $[q_m, \bar{q}]$, the distribution is the shifted-quadratic $\text{Sqd}_{a,\phi}$ for $a = \frac{\bar{q}(1-q_m)-rq_m(1-\bar{q})}{rq_m(\bar{q}-q_m)}$ and $\phi = \frac{-(rq_m-\bar{q})}{rq_m(\bar{q}-q_m)}$. On $[\bar{q}, 1]$, the distribution is the shifted-quadratic $\text{Sqd}_{0,\phi}$ for $\phi = \frac{-\bar{q}}{rq_m(1-\bar{q})}$. The quantile range $[0, q_m]$ is truncated.

3.3. Local Proof for Markup Mechanisms, Triangle Revenue Curves

To recap, in this section and the next we prove Theorem 5 with the following main steps. First, in Theorem 6 (stated next) we characterize the prior independent optimal mechanism under the restriction to randomized markup mechanisms $\mathcal{M}^{\text{mark}}$ and triangle revenue-curve distributions $\mathcal{F}^{\text{trv}}$, cf. Alaei et al. (2018). This restricted program has the same solution as we claim in Theorem 5.[4] Second, we show that the optimal randomized markup mechanism and the class of Triangle distributions are mutual best responses among the more general classes of truthful, scale-invariant mechanisms $\mathcal{M}^{\text{si}}$ and regular distributions $\mathcal{F}^{\text{reg}}$. Combining these results gives the main result in Theorem 5.

Theorem 6. *Given a single item, 2-agent revenue auction, the optimal truthful, scale-invariant,* ***randomized markup*** *mechanism (from the class $\mathcal{M}^{\text{mark}}$) against* ***default-scaled Triangles*** $\mathcal{F}^{\text{trv}}$ *for the prior independent design program ($\alpha^{\mathcal{F}^{\text{trv}}}$) is $M_{\hat{r},\xi}$ which randomizes according to $\boldsymbol{\xi}$ over the second-price auction M_1 with probability ξ_1 and r^*-markup mechanism M_{r^*} with probability $\xi_{r^*} = 1 - \xi_1$, where $\xi_1 \approx 0.806$ and $r^* \approx 2.447$. The worst-case* ***triangle*** *distribution for this mechanism is $\text{Trvd}_{q_m^*}$ with $q_m^* \approx 0.093$ and* ***its approximation ratio is $\alpha^{\mathcal{F}^{\text{trv}}} \approx 1.907$.***

[4] Note the almost identical language of Theorem 6 in comparison to previous statement of Theorem 5: the only changes are the (bolded) definitions of the class of mechanisms and the distribution class.

The proofs of Theorem 5 and Theorem 6 depend on a sequence of sub-results. We start by giving the proof of Theorem 6 here,[5] then this section and the next give full lists of sub-result statements but otherwise omit a number of the formal proofs.

Proof. The approach of this proof is to identify the triangle $\text{Trvd}_{q_m^*}$ for which the designer is indifferent between the second price auction M_1 and at least one other (non-trivial) markup mechanism, denoted M_{r^*}, *both of which must be optimal*. As part of this proof we show that where this occurs, the markup factor r^* of alternative optimal mechanisms M_{r^*} is indeed unique.

For such a distribution $\text{Trvd}_{q_m^*}$, the mechanism designer is also indifferent (in minimizing the approximation ratio) between any mixture over M_1 (with probability ξ_1) and M_{r^*} (with probability $\xi_{r^*} = 1 - \xi_1$), and all other r-markup mechanisms for $1 \leq r \notin \{1, r^*\}$ are inferior (which we graph in Figure 3.2, in the context of $\text{Trvd}_{q_m^*}$). Critically though, the mechanism designer can strategically choose ξ_1^* to constrain the behavior of the adversary.

We then identify the ξ_1 for which the adversary's best response (in maximizing the approximation ratio) to $M_{\hat{r},\boldsymbol{\xi}} \in \mathcal{M}^{\text{mark}}$ is the distribution $\text{Trvd}_{q_m^*} \in \mathcal{F}^{\text{trv}}$. The solution "profile" composed of $M_{\hat{r},\boldsymbol{\xi}}$ and $\text{Trvd}_{q_m^*}$ is a (mixed) Nash equilibrium between the mechanism designer and the adversary, and therefore it solves the prior independent optimization problem. The parameters can be numerically identified as $\xi_1^* \approx 0.80564048$, $r^* \approx 2.4469452$, $q_m^* \approx 0.0931057$, and the approximation ratio is $\alpha^{\mathcal{F}^{\text{trv}}} \approx 1.9068943$.

---

[5] Theorem 6 appears in **Hartline et al. (2020a)** with the following explanation of its proof:
"The parameters of this optimal mechanism are the solution to an algebraic expression (cf. Lemma 4) that we are unable to solve analytically. Our proof will instead combine numeric calculations of select points in parameter space with theoretical analysis to rule out most of the parameter space. For the remaining parameter space, we can show that the expression is well-behaved and, thus, numeric calculation can identify near optimal parameters."

Per the outline above, first we identify the triangle distribution monopoly quantile q_m^* and the r^* for which M_1 and M_{r^*} obtain the same ratio. Denote the approximation ratio for the SPA M_1 as $\text{APX}_1(q_m) = 2 - q_m$ (the ratio of quantities in Lemma 2 to Lemma 3 below), which is continuous in q_m. Denote the approximation ratio of the optimal markup mechanism against distribution Trvd_{q_m} by

$$\text{APX}_*(q_m) = \sup_{r>1} \frac{\text{OPT}_{\text{Trvd}_{qm}}(\text{Trvd}_{q_m})}{M_r(\text{Trvd}_{q_m})} \tag{3.1}$$

By Lemma 4, the approximation ratio $\text{APX}_*(q_m)$ is continuous in q_m as well. It is easy to verify that $\text{APX}_1(0) = 2 > \text{APX}_*(0) = 1$ while $\text{APX}_1(1) = 1 < \text{APX}_*(1) = \infty$. By continuity, there exists a q_m^* where these two functions cross, i.e., $\text{APX}_*(q_m^*) = \text{APX}_1(q_m^*)$. See Figure 3.2. By numerical calculation, $q_m^* \approx 0.0931057$ for which

$$r^* = \text{argmax}_{r>1} \frac{\text{OPT}(\text{Trvd}_{q_m^*})}{M_r(\text{Trvd}_{q_m^*})} \approx 2.4469452 \tag{3.2}$$

Now fix r^* and let $\boldsymbol{\xi}_{r^*}$ be a distribution that is restricted to mixing over markup mechanisms M_1 and M_{r^*}. We search for $\xi_{r^*,1}^*$ – the probability on M_1 – for which the adversary maximizes the approximation ratio of mechanism $M_{\hat{r},\boldsymbol{\xi}_{r^*}}$ by specific selection of the triangle distribution $\text{Trvd}_{q_m^*}$ (i.e., with the desired monopoly quantile).[6] Denote by $q_m(\xi_{r^*,1})$ the monopoly quantile as a function of $\xi_{r^*,1}^*$ for the triangle distribution that maximizes

[6] Note the following technique: a (mechanism designer) player identifies a correct "mixture" over the set of designer-best-response actions in order to require the specific action-response by the (adversary) other player *which induces those designer-best-responses in the first place.* This is a common theme towards solving for mixed Nash equilibrium in 2-player, zero-sum games.

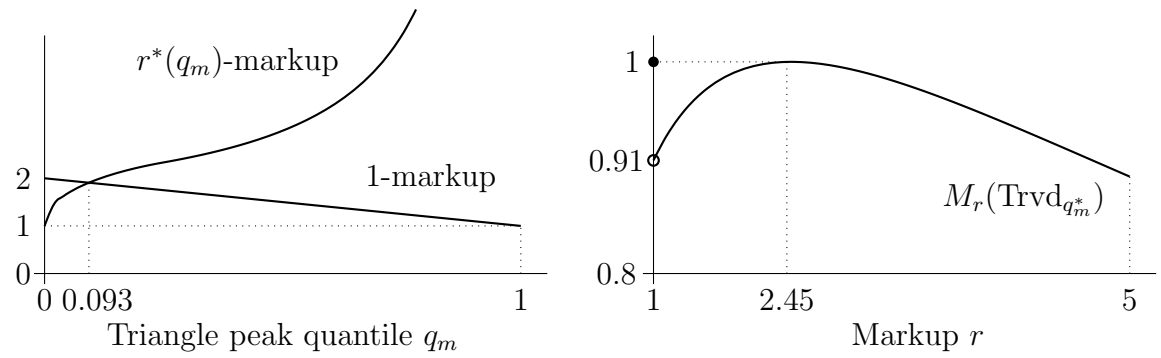

Figure 3.2. Performance of Optimal Markups against Worst-case Triangle
The figure on the left plots, as a function of q_m, the approximation ratio $\text{APX}_1(q_m)$ of $M_1 = \text{SPA}$ against triangle distribution Trvd_{q_m} (straight line), and the approximation ratio $\text{APX}_*(q_m)$ of the optimal non-trivial markup mechanism $M_{r^*(q_m)}$ against triangle distribution Trvd_{q_m} (curved line). Note explicitly here – the graph of APX_* embeds a locally optimal markup factor $r^*(q_m)$ as an argmax (per equation (3.1), with functional dependence on q_m) and therefore the r^* used to plot each point (of the curved line) is generally distinct from the optimal ratio for an alternative monopoly quantile q'_m. The functions cross at $q_m^* \approx 0.0931057$.
The figure on the right plots the revenue of the r-markup mechanism M_r on triangle distribution $\text{Trvd}_{q_m^*}$ as a function of markup r, i.e., $M_r(\text{Trvd}_{q_m^*})$. Notice how by choice of q_m^*, the optimal non-trivial markup mechanism has the same revenue as M_1.

the approximation ratio of mechanism $M_{\hat{r},\boldsymbol{\xi}_{r^*}}$, i.e.,

$$q_m(\xi_{r^*,1}) = \text{argmax}_{q_m} \frac{\text{OPT}(\text{Trvd}_{q_m})}{M_{\hat{r},\boldsymbol{\xi}_{r^*}}(\text{Trvd}_{q_m})}.$$

By numerical calculation, for any $r \in [2.445, 2.449]$, we observe $q_m(0.81) < q_m^* < q_m(0.8)$. Continuity of $q_m(\cdot)$ for $r \in [2.445, 2.449]$ and $\xi_{r^*,1} \in [0.8, 0.81]$ (with formal proof of continuity in the reference paper but omitted here) then implies that there exists $\xi^*_{r^*,1}$ such that $q_m(\xi^*_{r^*,1}) = q_m^*$. By numerical calculation, $\xi_1^* = \xi^*_{r^*,1} \approx 0.80564048$. □

The proofs of Lemma 2 and Lemma 3 (next) both follow directly from Lemma 1 (Dhangwatnotai et al., 2015). The proof of Lemma 4 is technical but straightforward.

Lemma 2. *Given a single item, i.i.d., 2-agent revenue auction, the optimal truthful mechanism against any regular distribution $\mathcal{F}^{\text{reg}}$ posts the monopoly price equivalent to q_m; and against distributions which further are default-scaled and truncated, this optimal mechanism obtains revenue $2 - q_m$ (where q_m is in fact the probability that an agent's value is equal to the monopoly price, from truncation).*

Lemma 3. *The revenue of the second-price auction M_1 for distribution Trvd_{q_m} is 1, i.e., $M_1(\text{Trvd}_{q_m}) = 1$.*

Lemma 4. *The revenue of the r-markup mechanisms M_r on triangle distribution Trvd_{q_m}, for $r \in (1, \infty)$ and $q_m \in [0, 1)$, is*

$$M_r(\text{Trvd}_{q_m}) = \frac{2r}{(1-q_m)(r-1)} \left(\frac{1-q_m}{(1-q_m)+rq_m} - \frac{\ln\left(\frac{r}{(1-q_m)+rq_m}\right)}{r-1} \right).$$

3.4. Final Proof for All Truthful Mechanisms, All Regular Distributions

We now show that the characterization of the previous section – i.e., optimal strategies for the mechanism designer restricted to markup mechanisms $\mathcal{M}^{\text{mark}}$ and the adversary restricted to default-scaled Triangles $\mathcal{F}^{\text{Trvd}}$ – remain optimal in the more general setting in which the mechanism class is all truthful scale invariant mechanisms $\mathcal{M}^{\text{si}} \supset \mathcal{M}^{\text{mark}}$ and the distribution class is all regulars $\mathcal{F}^{\text{reg}} \supset \mathcal{F}^{\text{tri}}$. We restate the full theorem statement here and are now able to give its proof; but to repeat, a number of formal proofs for its sub-results (below) are omitted.

Theorem 5. *Given a single item, 2-agent auction with a revenue objective, the optimal truthful, scale-invariant mechanism (from the class $\mathcal{M}^{\text{si}}$) against regular distributions $\mathcal{F}^{\text{reg}}$*

for the prior independent design program ($\alpha^{\mathcal{F}^{\text{reg}}}$) *is* $M_{\hat{r},\boldsymbol{\xi}}$ *which randomizes according to* $\boldsymbol{\xi}$ *over the second-price auction* M_1 *with probability* ξ_1 *and* r^**-markup mechanism* M_{r^*} *with probability* $\xi_{r^*} = 1 - \xi_1$*, where* $\xi_1 \approx 0.806$ *and* $r^* \approx 2.447$*. The worst-case regular distribution is* $\text{Trvd}_{q_m^*}$ *with its monopoly quantile* $q_m^* \approx 0.093$ *and its approximation ratio is* $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$*.*

Proof. The technique relies on a mutual-best-response argument that the optimal mechanism and adversarial distribution can be localized to "well-behaved" markup mechanisms on one side (which incorporate probability at least 2/3 on $M_1 = \text{SPA}$) and Triangles on the other. The proof fully proceeds from the following outline, with dependence on the referenced sub-results.

- our starting "guess-and-check" assumption is that the adversary will choose only monopoly-truncated distributions from class $\mathcal{F}^{\text{mt}}$ in mixed-Nash equilibrium;
- using the starting assumption, we reduce the mechanism search space from $\mathcal{M}^{\text{si}}$ to $\mathcal{M}^{\text{mark}}$, which completes one side of the overall reduction (see Theorem 7);
- from within $\mathcal{M}^{\text{mark}}$, propose the randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ as the distribution over markups according to $\boldsymbol{\xi}$ (per the theorem statement);
- finally, show that the adversary's best response to $M_{\hat{r},\boldsymbol{\xi}}$ among all regular distributions is an element of the Triangles $\mathcal{F}^{\text{trv}}$ (specifically to allow the edge-case Triangles represented by a point mass, i.e., Trvd_1), which completes the second side of the overall reduction (see Theorem 8);
- further, the previous point confirms the starting guess-and-check assumption, therefore it completes the proof. □

The outline of the rest of this section is as follows. The next two theorem statements each solve one side of the reduction in the proof of Theorem 5. Following that, we give a sequence of lemmas that are needed to support Theorem 8, and finally a proof sketch for Theorem 8 to give intuition for how the lemmas fit together. See the reference paper for omitted proofs and the full proof of Theorem 8 (Hartline et al., 2020a).

Theorem 7. *Given a single item, i.i.d., 2-agent, truthful, revenue auction, the class of scale-invariant mechanisms (from the class $\mathcal{M}^{\text{si}}$), and the class of regular distributions $\mathcal{F}^{\text{reg}}$.*

For every mechanism $M \in \mathcal{M}^{\text{si}}$, there is a randomized markup mechanism $M' \in \mathcal{M}^{\text{mark}} \subset \mathcal{M}^{\text{si}}$ with (weakly) higher revenue (and weakly lower approximation ratio) on every monopoly-truncated (and regular) distribution $F \in \mathcal{F}^{\text{mt}}$. I.e., $M'(F) \geq M(F)$.

Theorem 8. *Given a single item, i.i.d., 2-agent, truthful revenue auction, the class of regular distributions $\mathcal{F}^{\text{reg}}$ and any stochastic markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ that assigns probability $\xi_1 \in [2/3, 1]$ on the second-price auction M_1.*

For any regular distribution $F \in \mathcal{F}^{\text{reg}}$, either the triangulation to F^{Trvd} of F or the point mass Trvd_1 has (weakly) higher approximation ratio than F. I.e.,

$$\max\left\{\frac{\text{OPT}_{F^{\text{Trvd}}}(F^{\text{Trvd}})}{M(F^{\text{Trvd}})}, \frac{\text{OPT}_{\text{Trvd}_1}(\text{Trvd}_1)}{M(\text{Trvd}_1)}\right\} \geq \frac{\text{OPT}_F(F)}{M[(F)}$$

The point of Theorem 8 is that it shows that the adversary does not need to consider arbitrary regular distribution F to maximize the approximation ratio of the (optimal) mechanism $M_{\hat{r},\boldsymbol{\xi}}$ of Theorem 5, because it always prefers instead either its triangulation

or the point mass distribution. (Note, the point mass is an edge-case triangle revenue-curve distribution and its value does not matter because of scale-invariance).

Because these are both Triangles, all regular distributions are shown to be (weakly) "dominated" in the adversary's choice set by elements of the Triangles. From here we list the lemma statements needed to prove Theorem 8. We only give the proofs of the first two of these: Lemma 5 and the Truncation Lemma 6.

Lemma 5. *Given a single item, i.i.d., 2-agent, truthful, revenue auction, the class of randomized markup mechanisms (from the class $\mathcal{M}^{\text{mark}}$), and the class of regular distributions $\mathcal{F}^{\text{reg}}$.*

The optimal approximation factor $\alpha^{\mathcal{F}^{\text{reg}}}_{M_{\hat{r},\boldsymbol{\xi}}}$ of any $M_{\hat{r},\boldsymbol{\xi}} \in \mathcal{M}^{\text{mark}}$ is lower bounded by $1/\xi_1$ for ξ_1 the probability on the SPA.

Proof. By example, point masses are regular distributions and $1/\xi_1$ is exactly the approximation of $M_{\hat{r},\boldsymbol{\xi}}$ to the SPA which is the optimal mechanism. Therefore $\alpha^{\mathcal{F}^{reg}}_{M_{\hat{r},\boldsymbol{\xi}}}$ must be at least $1/\xi_1$. □

Lemma 6 (The Truncation Lemma)**.** *Given a single item, i.i.d., 2-agent, truthful, revenue auction, the class of randomized markup mechanisms (from the class $\mathcal{M}^{\text{mark}}$), and the class of regular distributions $\mathcal{F}^{\text{reg}}$. Consider any randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ that places probability $\xi_1 \in [1/2, 1]$ on the second-price auction M_1.*

For any regular distribution $F \in \mathcal{F}^{\text{reg}}$, either the monopoly-truncation of F to F' or the point mass distribution Trvd_1 *has (weakly) higher approximation ratio than F. I.e.,*

$$\max\left\{\frac{\text{OPT}_{F'}(F')}{M(F')}, \frac{\text{OPT}_{\text{Trvd}_1}(\text{Trvd}_1)}{M(\text{Trvd}_1)}\right\} \geq \frac{\text{OPT}_F(F)}{M(F)}$$

.

Proof. By Lemma 5, it can be assumed that the approximation of randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ on distribution F is at least $1/\xi_1$ (where ξ_1 is the probability on the SPA's markup factor of 1). First note that the expected revenue $M_{\hat{r},\boldsymbol{\xi}}$ against the point mass distribution Pmd_1 is ξ_1 and the optimal revenue on this distribution is 1 (from the SPA). If the approximation factor $\text{OPT}_F(F)/M_{\hat{r},\boldsymbol{\xi}}(F)$ is less than $1/\xi_1$ then Pmd_1 gives a higher approximation factor than F and the lemma follows. For the remainder of the proof, assume that the approximation factor of mechanism $M_{\hat{r},\boldsymbol{\xi}}$ against F is more than $1/\xi_1$.

View the randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ as a distribution over two mechanisms: the second-price auction M_1 with probability ξ_1, and $M_* = M_{\hat{r},\boldsymbol{\xi}_{-1}}$, a distribution over all non-trivial markup mechanisms M_r with $r > 1$, with probability $1 - \xi_1$. The optimal mechanism is OPT_F. Decompose the revenue from distribution F across the three mechanisms M_1, M_*, and OPT_F as follows. Denote the monopoly quantile of F by q_m.

- OPT_+ and OPT_- give the expected revenue of the optimal mechanism from agents with values above and below the monopoly price (respectively **below** and **above** the monopoly quantile q_m).
- $\text{SPA}_+ = \text{OPT}_+$ and SPA_- give the expected revenue of the second-price auction M_1 from agents with values above and below the monopoly price.
- MKUP_+ and MKUP_- give the expected revenue of the stochastic markup mechanism M_* from prices (strictly) above and (weakly) below the monopoly price.

Consider monopoly-truncating the distribution F at q_m to obtain $F' \in \mathcal{F}^{mt}$ (the class of monopoly-truncated regular distributions). Define analogous quantities (with identities):

- $\mathrm{OPT}'_+ < \mathrm{OPT}_+$ and $\mathrm{OPT}'_- = \mathrm{OPT}_-$.

  Identities follow from the geometric analysis of Lemma 1.

- $\mathrm{SPA}'_+ = \mathrm{OPT}'_+$ and $\mathrm{SPA}'_- = \mathrm{SPA}_-$.

  Identities follow from the geometric analysis of Lemma 1.

- $\mathrm{MKUP}'_+ = 0$ and $\mathrm{MKUP}'_- = \mathrm{MKUP}_-$.

  Values above the monopoly price are not supported by the truncated distribution, so the revenue from those prices is zero. On the other hand, prices (weakly) below the monopoly price are bought with the same probability because the cumulative distribution functions F' and F are the same for these prices.

The remainder of the proof follows a straightforward calculation, though as a technical note we do reference Lemma 22 deferred to Appendix B.2 to support the final sequence of inequalities below. Write the approximation ratio of $M_{\hat{r},\boldsymbol{\xi}}$ on distribution F (using the given identities) and rearrange:

$$\begin{aligned}\frac{\mathrm{OPT}_F(F)}{M_{\hat{r},\boldsymbol{\xi}}(F)} &= \frac{\mathrm{OPT}_+ + \mathrm{OPT}_-}{\xi_1\left(\mathrm{OPT}_+ + \mathrm{SPA}_-\right) + (1-\xi_1)\left(\mathrm{MKUP}_+ + \mathrm{MKUP}_-\right)} \\ &= \frac{\mathrm{OPT}_+ + [\mathrm{OPT}_-]}{\xi_1\ \mathrm{OPT}_+ + [\xi_1\,\mathrm{SPA}_- + (1-\xi_1)\left(\mathrm{MKUP}_+ + \mathrm{MKUP}_-\right)]}\end{aligned}$$

Since the approximation ratio on F is at least $1/\xi_1$, the ratio of the first term in the numerator and denominator is at most the ratio of the remaining terms [in brackets]:

$$\frac{1}{\xi_1} = \frac{\mathrm{OPT}_+}{\xi_1\ \mathrm{OPT}_+} \leq \frac{[\mathrm{OPT}_-]}{[\xi_1\,\mathrm{SPA}_- + (1-\xi_1)\left(\mathrm{MKUP}_+ + \mathrm{MKUP}_-\right)]}$$

Now write the approximation ratio of $M_{\hat{r},\boldsymbol{\xi}}$ on monopoly-truncation F' (using the given identities) and obtain a bound on it:

$$\begin{aligned}
\frac{\mathrm{OPT}_{F'}(F')}{M_{\hat{r},\boldsymbol{\xi}}(F')} &= \frac{\mathrm{OPT}'_+ + [\mathrm{OPT}_-]}{\xi_1\ \mathrm{OPT}'_+ + [\xi_1\ \mathrm{SPA}_- + (1-\xi_1)\ \mathrm{MKUP}_-]} \\
&\geq \frac{\mathrm{OPT}'_+ + [\mathrm{OPT}_-]}{\xi_1\ \mathrm{OPT}'_+ + [\xi_1\ \mathrm{SPA}_- + (1-\xi_1)\ (\mathrm{MKUP}_+ + \mathrm{MKUP}_-)]} \\
&\geq \frac{\mathrm{OPT}_+ + [\mathrm{OPT}_-]}{\xi_1\ \mathrm{OPT}_+ + [\xi_1\ \mathrm{SPA}_- + (1-\xi_1)\ (\mathrm{MKUP}_+ + \mathrm{MKUP}_-)]} \\
&= \frac{\mathrm{OPT}_F(F)}{M_{\hat{r},\boldsymbol{\xi}}(F)}.
\end{aligned}$$

The calculation shows that, for any distribution F, the monopoly-truncated distribution F' increases the approximation factor of the randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$. Therefore, the worst-case distribution is monopoly-truncated. □

Lemma 7 (Allouah and Besbes, 2018)**.** *Given a single item, i.i.d., 2-agent, truthful, revenue auction, a markup mechanism $M_r \in \mathcal{M}^{\mathrm{mark}}$ with specific r, and the class of regular distributions $\mathcal{F}^{\mathrm{reg}}$. For any $F \in \mathcal{F}^{\mathrm{reg}}$ with monopoly quantile q_m, let the distribution F' be the result of ironing F on the region $[\bar{q}, 1]$ for $\bar{q} = Q_F(1/rq_m)$.*[7]

The virtual surplus of $M_r(F)$ from quantiles in $[\bar{q}, 1]$ is at least the virtual surplus of $M_r(F')$ from quantiles in $[\bar{q}, 1]$.

Lemma 8. *Given a single item, i.i.d., 2-agent, truthful, revenue auction, a randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}} \in \mathcal{M}^{\mathrm{mark}}$ specifically with $\boldsymbol{\xi}$ a mixture over ξ_1 on the SPA and*

[7] Recall the definition of *ironing* from Section 2.2.6; note that the usage of "ironing" here is an inexact example (of non-optimal revenue curve ironing). In this case, the revenue curve is altered with the same effect – i.e., the original distribution F itself is converted (via moving the density functions) to F' such that the revenue curve of F' is the same as the ironed-as-indicated revenue curve of F.

$\xi_{\bar{r}} = 1 - \xi_1$ *on a single non-trivial markup factor* $\bar{r}$*, and the class of monopoly-truncated regular distributions* $\mathcal{F}^{\mathrm{mt}}$*.*

For every $F \in \mathcal{F}^{\mathrm{mt}}$*, there exists a quadrilaterial distribution* $F^{\mathrm{Qr}} \in \mathcal{F}^{\mathrm{qrv}}$ *for which optimal revenue is the same but* $M_{\hat{r},\boldsymbol{\xi}}$ *gets (weakly) lower revenue, both in comparison to the original* F*. I.e.,* $\mathrm{OPT}_{F^{\mathrm{Qr}}}(F^{\mathrm{Qr}}) = \mathrm{OPT}_F(F)$ *and* $M_{\hat{r},\boldsymbol{\xi}}(F^{\mathrm{Qr}}) \leq M_{\hat{r},\boldsymbol{\xi}}(F)$*.*

Lemma 9. *Given a single item, i.i.d., 2-agent, truthful, revenue auction, a randomized markup mechanism* $M_{\hat{r},\boldsymbol{\xi}} \in \mathcal{M}^{\mathrm{mark}}$ *specifically with* $\boldsymbol{\xi}$ *a mixture over* $\xi_1 \in [2/3, 1]$ *on the SPA and* $\xi_{\bar{r}} = 1 - \xi_1$ *on a single non-trivial markup factor* $\bar{r}$*, and the class of Quadrilaterals* $\mathcal{F}^{\mathrm{qrv}}$*.*

For every quadrilateral distribution $\mathrm{Qrvd}_{q_m,\bar{q},r} \in \mathcal{F}^{\mathrm{qrv}}$*, consider the triangulation of* $\mathrm{Qrvd}_{q_m,\bar{q},r}$ *to* $\mathrm{Trvd}_{q_m} \in \mathcal{F}^{\mathrm{trv}}$*. Optimal revenue is the same for the triangle but* $M_{\hat{r},\boldsymbol{\xi}}$ *gets (weakly) lower revenue on the Triangle* Trvd_{q_m}*, both in comparison to the original quadrilateral* $\mathrm{Qrvd}_{q_m,\bar{q},r}$*. I.e.,* $\mathrm{OPT}_{\mathrm{Trvd}_{q_m}}(\mathrm{Trvd}_{q_m}) = \mathrm{OPT}_{\mathrm{Qrvd}_{q_m,\bar{q},r}}(\mathrm{Qrvd}_{q_m,\bar{q},r})$ *and* $M_{\hat{r},\boldsymbol{\xi}}(\mathrm{Trvd}_{q_m}) \leq M_{\hat{r},\boldsymbol{\xi}}(\mathrm{Qrvd}_{q_m,\bar{q},r})$*.*

To end the section, as previously indicated, we include the following proof sketch for Theorem 8.

PROOF SKETCH. Start generally with an arbitrary regular distribution $F \in \mathcal{F}$ and an assumption that the mechanism is a markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ with $\xi_1 \in [2/3, 1]$. The sequence of lemmas show – via a series of steps – that the approximation factor of $M_{\hat{r},\boldsymbol{\xi}}$ is worse for the triangulation of F than for the original F. I.e., *for every such* $M_{\hat{r},\boldsymbol{\xi}}$, the following steps hold.

The Truncation Lemma 6 shows for $M_{\hat{r},\boldsymbol{\xi}}$ that either we can ignore F because the point mass Pmd_1 is a worse case, or else otherwise the approximation factor (of $M_{\hat{r},\boldsymbol{\xi}}$) gets worse when F is converted to F' which is the monopoly-truncation of F. I.e., the adversary's class of regulars $\mathcal{F}^{\text{reg}}$ is reduced to the class of monopoly-truncated (regulars) $\mathcal{F}^{\text{mt}}$ (which include point masses).

Lemma 7 and Lemma 8 work together to show that the quadrilateralization of F' – within the context of using *any* markup parameter $r > 1$ – makes the approximation ratio of the respective M_r worse. By implication, the approximation ratio of the randomized markup mechanism $M_{\hat{r},\boldsymbol{\xi}}$ becomes worse. I.e., the adversary's class of monopoly-truncated regulars $\mathcal{F}^{\text{mt}}$ is reduced to the class of Quadrilaterals $\mathcal{F}^{\text{qrv}}$.

Finally, Lemma 9 shows that the triangulation of any quadrilateral distribution again makes the approximation ratio worse. This gives the final reduction of the adversary's choice set to the Triangles $\mathcal{F}^{\text{trv}}$. □

CHAPTER 4

Lower Bounds for Prior Independent Algorithms

Chapter 3 gave the characterization of an optimal 2-agent, one-item, truthful mechanism for a revenue objective and the regular class of distributions $\mathcal{F}^{\text{reg}}$. This chapter shifts the analysis from characterization of optimal algorithms to a broad approach for establishing lower bounds. Specifically, it introduces a novel Blends Technique (Definition 19) that is agnostic to problem setting for proving lower bounds on the prior independent approximation factor of any algorithm.

The Blends Technique depends on solving a general equation that results from application of Yao's Minimax Theorem (Yao (1977), see Theorem 9) to the prior independent framework, thus we must identify solutions to the equation to prove lower bounds. We call these solutions *dual blends*. This chapter explicitly includes example dual blends. As an application, we use these example solutions within 2-agent mechanism design settings to prove lower bounds (similar to the setting of the previous chapter but with different parameters).

A first formal study of the *mathematical structure* of our solutions is deferred to Chapter 5, where we identify key properties of the examples from this chapter and use these properties to inform the design of broad classes of solutions. We will also show there that the search for an optimal dual blend – with "optimal" determined within the context of a given prior independent algorithm problem – is connected to the economics topic of *information design*, for which our dual blends are competing *signalling strategies.*

This chapter is organized as follows. Section 4.1 quickly states Yao's Minimax Principle to set up the analysis of our approach. Section 4.2 explains our Blends Technique as application of Theorem 10 for which the proof is deferred to Section 4.3. Section 4.4 explains in detail the procedure of crafting dual blends solutions to satisfy the key equation of the Blends Technique (see equation (4.2)) and simultaneously builds a first, simple example solution for reference through the presentation of the procedure. Section 4.5 gives a second example of a dual blend with its design motivated by common tools from mechanism analysis and then proceeds to state novel lower bounds for prior independent mechanism design problems.

Section 4.6 provides an outline of two additional examples of dual blends of interest for mechanism design but presentation of these highly technical examples is deferred to Appendix A. Finally, Section 4.7 summarizes the benefits of lower bounds from the Blends Technique in comparison to similar lower bounds which it implicates.

Material in this chapter is largely drawn from **Hartline and Johnsen (2021)**.

4.1. Theoretical Lower Bounds from Minimax

Yao's Minimax Principle (Theorem 9) illustrates the role of the adversary through a direct connection to a 2-player zero-sum game. Some definitions for terms used in Theorem 9 were introduced at the end of Section 2.1.

Theorem 9 (Yao, 1977)**.** [Yao's Minimax Theorem] *Given a 2-player zero-sum game $\mathcal{G}$ in which sequentially player 1 chooses mixed action $\gamma_1 \in \Delta(\Omega_1)$, then player 2 chooses action $\omega_2 \in \Omega_2$. The players are cost minimizers and the cost functions on pure actions are (any real-valued function) $C_1(\omega_1, \omega_2) \geq 0$ and $C_2 = -C_1$. Then the* value *of game $\mathcal{G}$*

(the left-hand side) satisfies:

$$\inf_{\gamma_1 \in \Delta(\Omega_1)} \sup_{\omega_2 \in \Omega_2} C_1(\gamma_1, \omega_2) \geq \sup_{\gamma_2 \in \Delta(\Omega_2)} \inf_{\omega_1 \in \Omega_1} C_1(\omega_1, \gamma_2) \tag{4.1}$$

4.2. A Technique for Prior Independent Lower Bounds: Blends

There is a detailed explanation of the high-level technique of lower bounds from Yao's Minimax Principle in the textbook by Borodin and El-Yaniv (1998). Regarding a worst-case perspective (competitive analysis), they write, "... to obtain a lower bound... it is sufficient to choose any probability distribution over [the adversary's choice set] and to bound from below... the ratio of average optimal ["offline" performance to average performance] of any *deterministic* online algorithm (where the expectations are taken with respect to [the adversary's fixed distribution])."

This section gives a minimax approach that is specific to prior independent design. To outline, we: (a) fix a randomization over adversary strategies in advance; (b) prove an upper bound on the performance of the best-response algorithm *from an alternative description of the adversary's induced correlated distribution over inputs*; and (c) measure the gap between the adversary's expected optimal performance and the upper bound on the expected performance of any algorithm. The key idea is the correlation in (b):

Definition 18. *A* blend *is a distribution-over-distributions $\delta \in \Delta(\mathcal{F})$. (Thus, $\delta(F)$ is the density at F.) A* blended distribution $\delta^n \in \Delta(\mathcal{V}^n)$ *is the induced density function of the correlated distribution resulting from n i.i.d. draws from a common distribution $\hat{F}$, with $\hat{F}$ drawn from δ.*

Two blends $\delta_1,\ \delta_2$ *are called* dual blends *if there exists correlated density function* g *such that:*

$$\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v}) \quad \forall\ \boldsymbol{v} \tag{4.2}$$

Each of $\delta_1,\ \delta_2$ *are a* side *of the dual blend. Finally, define* $\text{opt}_{n,i} = \mathbf{E}_{F\sim\delta_i}\left[\text{OPT}_F(F)\right]$ *to be the expected performance of an optimal algorithm which knows* F *over a blend* δ_i*.*

The point is: an arbitrary blend δ can be "flattened" to describe a specific (symmetric) correlated distribution $\delta^n = g$ over input space $\mathcal{V}^n$. Now suppose in fact two distinct blends δ_1 and δ_2 as choices of the PI adversary induce the same correlated distribution, i.e., they satisfy Definition 18. Because both induce the same description of input profiles, every algorithm is limited by the structure of either description. The lower bound of the technique has the following intuition: the adversary chooses δ_2 which fixes the benchmark of the current scenario to $\text{opt}_{n,2} = \mathbf{E}_{F\sim\delta_2}\left[\text{OPT}_F(F)\right]$;[1] δ_2 induces the correlated distribution g and the algorithm best responds to g; however the fact that δ_1 also induces g means that every algorithm is upper bounded by $\text{opt}_{n,1}$; if this upper bound is strictly smaller than the benchmark, then a strict gap necessarily ensues.

An important interpretation is: even when the algorithm knows that the adversary has committed to the mixed action δ_2 (and thus the process for how the inputs are generated is *fully revealed*), the algorithm can not help but be upper bounded by the information constraints of δ_1. The proof of Theorem 10 appears in Section 4.3.

[1] Lemma 10 in Section 4.3 shows that we can set the prior independent benchmark in this way.

Theorem 10 (The Dual Blends Theorem)**.** *Consider a prior independent setting with input space $\mathcal{V}^n$, class of algorithms $\mathcal{A}$, and class of distributions $\mathcal{F}$. Let $\mathcal{F}^{\text{all}}$ be all distributions. Assume there exist two distinct dual blends $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$ and $\delta_2 \in \Delta(\mathcal{F})$ and correlated density function g (of Definition 18) such that:*

$$\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v}) \quad \forall\ \boldsymbol{v}$$

The optimal prior independent approximation factor $\alpha^{\mathcal{F}}$ is at least the ratio ${}^{\text{opt}_{n,2}}/_{\text{opt}_{n,1}}$:

$$\alpha^{\mathcal{F}} = \min_{A \in \mathcal{A}} \max_{F \in \mathcal{F}} \frac{\text{OPT}_F(F)}{A(F)} \geq \frac{\text{opt}_{n,2}}{\text{opt}_{n,1}} \tag{4.3}$$

Definition 19. The Blends Technique *is the proof technique for approximation lower bounds which applies Theorem 10 to a specified prior independent design problem.*

A detailed outline of the necessary computations to confirm that descriptions of δ_1 and δ_2 are dual blends is given in Section 4.4, which also includes a first *non-trivial* $n = 2$ example of a dual blend. Construction of dual blends does not depend on problem domain – e.g., mechanism design or online algorithms – but which dual blend induces the largest lower bound does depend on domain. The Blends Technique is inherently an application of *non-unique tensor decomposition*. We discuss this connection – including the likely impossibility of applying the Blends Technique directly for $n > 3$ – in Section 5.5.

4.3. Proof of The Dual Blends Theorem

For use in this section, recall our notation $A(F) = \mathbf{E}_{\boldsymbol{v} \sim F}[A(\boldsymbol{v})]$ for the expected performance of algorithm A on n i.i.d. draws from a distribution F.

First we state and prove Lemma 10 which shows that for any fixed blend $\bar{\delta}$ (as implicit choice of the adversary), we can obtain a lower bound on prior independent approximation. (This lower bound is used as an interim step within the proof of Theorem 10.)

Lemma 10 states that we can replace the adversary's maximization problem within prior independent design (for reference see equation (4.4)). In its place, the adversary effectively sets a benchmark as the expectation of optimal performance over distributions drawn from $\bar{\delta}$ (thus, the benchmark is $\mathbf{E}_{F\sim\bar{\delta}}[\mathrm{OPT}_F(F)]$). Symmetrically, the algorithm's performance is its expected performance over distributions drawn from $\bar{\delta}$ (thus, its performance is $\mathbf{E}_{F\sim\bar{\delta}}[A(\boldsymbol{v})]$).

An algorithm's approximation of the benchmark is measured as the ratio of this benchmark to its performance, i.e., as *ratio-of-expectations* (ROE). The lower bound results from the minimum ratio achieved by any algorithm $A \in \mathcal{A}$. Practically, this interim lower bound is only an abstraction because we don't say anything about how to optimize the algorithm A.

Lemma 10 is an application of Yao's Minimax Principle (Yao, 1977) which treats distributions themselves as the inputs (rather than the subsequent i.i.d. draws). In fact from this perspective, the standard prior independent benchmark set by $\mathrm{OPT}_F(F)$ may be regarded as the expected performance of the *offline optimal* algorithm.

Lemma 10 (The Ratio-of-Expectations Benchmark Lemma)**.** *Consider a prior independent setting with input space $\mathcal{V}^n$, class of algorithms $\mathcal{A}$, and class of distributions $\mathcal{F}$. Let $\bar{\delta} \in \Delta(\mathcal{F})$ be any fixed blend, i.e., a fixed distribution over the distributions of $\mathcal{F}$. Then*

$$\alpha^{\mathcal{F}} = \min_{A\in\mathcal{A}} \max_{F\in\mathcal{F}} \frac{\mathrm{OPT}_F(F)}{A(F)} \geq \min_{A\in\mathcal{A}} \left[\frac{\mathbf{E}_{F\sim\bar{\delta}}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\bar{\delta}}\left[A(F)\right]}\right], \text{ for fixed } \bar{\delta} \tag{4.4}$$

Proof. We start with the prior independent design problem. Explanations for each step of this sequence are given following.

$$\begin{aligned}
\min_{A\in\mathcal{A}} \max_{F\in\mathcal{F}} \frac{\mathrm{OPT}_F(F)}{A(F)} &= \min_{A\in\mathcal{A}} \max_{\delta\in\Delta(\mathcal{F})} \frac{\mathbf{E}_{F\sim\delta}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta}\left[A(F)\right]} \\
&\geq \max_{\gamma\in\Delta(\Delta(\mathcal{F}))} \min_{A\in\mathcal{A}} \mathbf{E}_{\delta\sim\gamma}\left[\frac{\mathbf{E}_{F\sim\delta}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta}\left[A(F)\right]}\right] \\
&= \max_{\delta\in\Delta(\mathcal{F})} \min_{A\in\mathcal{A}} \left[\frac{\mathbf{E}_{F\sim\delta}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta}\left[A(F)\right]}\right] \\
&= \min_{A\in\mathcal{A}} \left[\frac{\mathbf{E}_{F\sim\bar{\delta}}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\bar{\delta}}\left[A(F)\right]}\right], \text{ for fixed } \bar{\delta}
\end{aligned}$$

- The first line above both relaxes the adersary's action space to allow a mixture of distributions – i.e., a blend $\delta \in \Delta(\mathcal{F})$ – and changes the benchmark (numerator) to be set by the expected optimal performance over distributions from the blend.

  It holds with equality because by Lemma 21 (given in Appendix B.2), the value of the inner maximization program before-and-after this step is the same for every A – the adversary gains no extra advantage because the ratio on the right-hand side must always be dominated anyway by the ratio achieved by some distribution F_+ in the support of any chosen δ. (To explain in further detail, the adversary could choose F_+ in the left-hand program and can still choose a point mass on F_+ in the right-hand program.)

- The second line applies Yao's Theorem Principle (Theorem 9). Note, the adversary's choice of actions $\gamma \in \Delta(\Delta(\mathcal{F}))$ represents the exact transformation using Minimax: the adversary now acts first and plays a distribution over actions in its support from the initial $\min - \max$ side. Then:
- The third line holds because the set of all blends over $\mathcal{F}$ – namely, $\Delta(\mathcal{F})$ – is closed under convex combination.
- The last line holds because fixing an argument of the outer program can only impair its objective (in this case by fixing $\delta = \bar{\delta}$ for any $\bar{\delta} \in \Delta(\mathcal{F})$ per the lemma statement). □

Given Lemma 10, we are prepared to restate and prove The Dual Blends Theorem 10.

Theorem 10 (The Dual Blends Theorem)**.** *Consider a prior independent setting with input space $\mathcal{V}^n$, class of algorithms $\mathcal{A}$, and class of distributions $\mathcal{F}$. Let $\mathcal{F}^{\text{all}}$ be all distributions. Assume there exist two distinct dual blends $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$ and $\delta_2 \in \Delta(\mathcal{F})$ and correlated density function g (of Definition 18) such that:*

$$\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v}) \quad \forall\ \boldsymbol{v}$$

The optimal prior independent approximation factor $\alpha^{\mathcal{F}}$ is at least the ratio ${\text{opt}_{n,2}}/{\text{opt}_{n,1}}$:

$$\alpha^{\mathcal{F}} = \min_{A \in \mathcal{A}} \max_{F \in \mathcal{F}} \frac{\text{OPT}_F(F)}{A(F)} \geq \frac{\text{opt}_{n,2}}{\text{opt}_{n,1}}$$

Proof. We start with the prior independent design problem and apply Lemma 10 (given above; by assigning $\bar{\delta} = \delta_2$). Justifications for the next steps are given afterwards.

$$
\begin{aligned}
\min_{A\in\mathcal{A}}\max_{F\in\mathcal{F}}\frac{\mathrm{OPT}_F(F)}{A(F)} &\geq \min_{A\in\mathcal{A}}\left[\frac{\mathbf{E}_{F\sim\delta_2}\left[\mathrm{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta_2}\left[A(F)\right]}\right] \\
&= \min_{A\in\mathcal{A}}\left[\frac{\mathrm{opt}_{n,2}}{\mathbf{E}_{\boldsymbol{v}\sim g}\left[A(\boldsymbol{v})\right]}\right] \\
&= \min_{A\in\mathcal{A}}\left[\frac{\mathrm{opt}_{n,2}}{\mathbf{E}_{F\sim\delta_1}\left[A(F)\right]}\right] \\
&\geq \min_{A\in\mathcal{A}}\left[\frac{\mathrm{opt}_{n,2}}{\mathbf{E}_{F\sim\delta_1}\left[\mathrm{OPT}_F(F)\right]}\right] = \frac{\mathrm{opt}_{n,2}}{\mathrm{opt}_{n,1}}
\end{aligned}
\tag{4.5}
$$

- The second and third lines substitute using the definition of $\mathrm{opt}_{n,i}$ and the assumption in the theorem statement that $\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v})$.

  Note, the adversary's choice of δ_2 is restricted to the set $\Delta(\mathcal{F})$ up front in the prior indepdent problem (i.e., the parameter $\mathcal{F}$ is fixed exogenously), and δ_2 induces $g = \delta_2^n$. However given g, there may exist any alternative description δ_1 with $g = \delta_1^n$, including a $\delta_1 \in \Delta(\mathcal{F}^{\mathrm{all}})$ that uses distributions outside the original class $\mathcal{F}$. This freedom to design δ_1 is an inherent *consequence of nature*.
- The fourth line inequality recognizes that expectation over locally optimal performances – each knowing the true F when realized – must weakly dominate the performance of a single algorithm run against all realizations of F (formally: Fact 16 after this proof).
- The final equality substitutes and realizes that the algorithm no longer appears in the function to be minimized, i.e., the objective is constant. □

The following holds because each OPT_F algorithm is optimal pointwise per F, whereas running A against each F is itself immediately upper bounded by OPT_F:

Fact 16. *Given an arbitrary prior independent algorithm design setting with class of distributions $\mathcal{F}$ and class of algorithms $\mathcal{A}$, and given $\delta \in \Delta(\mathcal{F})$. For any fixed algorithm $A \in \mathcal{A}$:*

$$\mathbf{E}_{F\sim\delta}\left[\mathrm{OPT}_F(F)\right] \geq \mathbf{E}_{F\sim\delta}\left[A(F)\right]$$

Towards increased understanding of the result and proof of Theorem 10, we include a second, informal proof from *linear programming* in Appendix A.1. It applies to a restricted family of algorithms problems due to the requirement that there be a linear programming description of the problem. Additional perspective gained from this approach is outlined in discussion interwoven with the proof.

4.4. Details of the Blends Technique and a First Simple Example: Shifted-Exponentials versus Uniforms

The goal of this section is to illustrate (a) the process of proving a dual blends structure from description of its elements in order to fit into Theorem 10, and then (b) the process of obtaining an algorithm-specific lower bound on prior independent approximation (which requires a specific algorithm setting). In addition to working through the process for (a) in detail in this section, it is fully outlined within Figure 4.1. This section includes some steps of the general process that do not apply in the case of our example here. As previously mentioned – both this process and the construction of concrete examples of dual blends exist independently of algorithm setting. For both (a) and (b) we use a dual blends example of Shifted-Exponentials-versus-Uniforms.

Looking ahead, the setting for (b) will be an application of mechanism design (which is introduced in Section 2.2.1). Specifically, the setting for (b) will use a 2-agent truthful

auction with a revenue objective, which is sufficient description to analyze a revenue gap; only at the very end will we identify relevant classes of distributions for which the revenue gap is meaningful and then formalize the gap with Proposition 1.

Dual Blends. We now describe the elements of our example and prove that they describe dual blends. We assume $n = 2$ and start with δ_1 and δ_2 as follows. The δ_1 blend is a mixture restricted to upward-closed Shifted-Exponentials. The shifted exponential distribution $\text{Sed}_{z,1}$ has PDF $\text{sed}_{z,1}(x) = e^{-(x-z)}$ and CDF $\text{Sed}_{z,1}(x) = 1-e^{-(x-z)}$ on $[z, \infty)$.

The δ_2 blend is a mixture restricted to downward-closed Uniforms. The uniform distribution $\text{Ud}_{0,z}$ has PDF $\text{ud}_{0,z}(x) = 1/z$ and CDF $\text{Ud}_{0,z}(x) = x/z$ on $[0, z]$. Note explicitly, each input has support $\mathcal{V} = [0, \infty)$ and input space is $\mathcal{V}^2$.

The weights o_F correspond to distributions in δ_1 and the weights ω_F correspond to distributions in δ_2. We require here that the total weight in the system is $W = 1$, though this could be relaxed for general blend examples:

Fact 17. *The total weight W of a dual blends analysis may be any positive constant as long as the total weight is finite – any finite weight will divide-out regardless in the right-hand side of line* (4.5)*). The total weight on each side of the dual blend must be equal.*

Further, blends must match up exactly for every measure of density, i.e., in a way that recognizes the difference between continuous density dv and point masses. The blend itself is included as a dimension if puts weight on a continuous mix over a parameter z.

Definition 20. *Define the* count of dimensional density *by the number of (axis-aligned) dimensions $i \in \{1, \ldots, n\}$ in which density is continuous: dv_i or dz.*

Overview observation:

Blends in this thesis always result from distributions over: i.i.d. draws from a distribution F. As such, continuous density at any fixed input (v_1, v_2) is measured per dv_i. Measurements of density must respect the difference between continuous density dv_i and point masses. I.e., regarding total density at a fixed input (v_1, v_2) (accounting for contributions from all elements of a blend), it could consist of continuous density in both dimensions (i.e., $dv_1 dv_2$), point mass density in both dimensions, or one dimension of each, e.g., point mass density of v_1 multiplied by a continuous density per dv_2.

Steps to confirm δ_1 and δ_2 as dual a blend:

1. Description: Explicitly enumerate the composition of the distributions δ_1 and δ_2.
2. Pointwise equality: For all inputs $\boldsymbol{v} = (v_1, v_2) \in \mathcal{V}^2$, compute the total density resulting from each blend for each type of measurement of density at the given $\boldsymbol{v}$.
3. Finiteness: (if applicable) Compute the total weight over all inputs for each blend to confirm they are finite; this computation doubles as a sanity check to help confirm that they are equal.

Identification of sufficient integral endpoints:

When a blend contains an integral over distributions parameterized by a bound z on the distributions' respective domains, distributions that contribute 0 at a point $\boldsymbol{v}$ may – without loss – be excluded by the integral computation of density at $\boldsymbol{v}$. This can be implemented by assignment of the integral endpoints because by observation, the ignored distributions are described by either an upward or downward-closed set over parameters z. E.g. for Uniforms at input $(4, 2)$, ignore $z < 4$ because only $\mathrm{Ud}_{0,z}$ with $z \geq 4$ contribute positive density at $(4, 2)$; cf. for $z = 3$, we have $\mathrm{ud}_{0,3}(4) \cdot \mathrm{ud}_{0,3}(1) = 0 \cdot 1/3 \; dz = 0$. Symmetric consequences apply for z as a parameter for a distribution's lower bound. See Figure 4.2 for illustration.

Steps to analyze resulting performance gap:

1. Optimal performances: Compute $\mathrm{OPT}_F(F)$ for every F with positive weight in either δ_1 or δ_2.
2. Blend performance: Compute $\mathrm{opt}_{2,1}$ and $\mathrm{opt}_{2,2}$ as the measures of blend-weighted expected optimal performance, accounting for both continuous density and point mass blends weights.
3. Identify lower bound: The ratio of blend performances (arranged to be at least 1) proves a necessary gap between an adversary's choice and the performance of any algorithm, and therefore lower bounds optimal approximation $\alpha^{\mathcal{F}}$.
4. Worst-case: (optional/ if applicable) If the analysis is parameterized, analyze worst-case assignment of the parameter – e.g., for value space $[1, h]$, consider $h \to \infty$.

Figure 4.1. Outline for a Dual Blend Computation

Given class of distributions $\mathcal{F}$ and $n = 2$, this offset provides outlines of required steps to (a) prove that two distributions over distributions δ_1 and δ_2 are dual blends (of Definition 18); and (b) prove a performance gap from Theorem 10 applied to δ_1 and δ_2.

Fact 18. *For distributions δ_1 and δ_2 to be dual blends, it is necessary at every input $\boldsymbol{v}$ that they match up density exactly* for every type of measurement of density*, in order to account for the difference between continuous density and point masses.*

The weights on the upward-closed Shifted-Exponentials blend (δ_1) are as follows:

- point mass of weight $o_{\text{pm}} = \frac{1}{2}$ on the distribution $\text{Sed}_{0,1}$.
- weights $o_{Ez} = \frac{1}{2}e^{-z}dz$ on all upward-closed distributions $\text{Sed}_{z,1}$ for $z \in [0, \infty)$.

The weights on the downward-closed Uniforms blend (δ_2) are as follows:

- (explicitly) we do not need a point mass;
- weights $\omega_{Uz} = \frac{1}{2}z^2e^{-z}dz$ on all downward-closed distributions $\text{Ud}_{0,z}$ for $z \in [0, \infty)$.

Here we observe that none of the distributions composing the blends incorporate point mass density. Therefore the only type of measurement of density that appears in the correlated distribution of this example is of the doubly-continuous form dv_1dv_2. I.e., we only need to check the 2-D density of $g(\boldsymbol{v})$ at each $\boldsymbol{v} \in [0, \infty)^2$. By symmetry we analyze density in the cone $v_1 \geq v_2 \geq 0$.

Determination of endpoints of integrals to compute the blends' densities are both (a) described in Figure 4.1 as part of procedure, and (b) illustrated in Figure 4.2. Intuitively, we start with an integral over all $z \in [0, \infty)$. However, we truncate the integral endpoints because not all elements of a blend will put positive density on an input ($v_1, v_2 \leq v_1$). The calculations of density – ignoring the continuous density terms dv_1dv_2 which are implied by the $2D$ subscript of g – are given by:

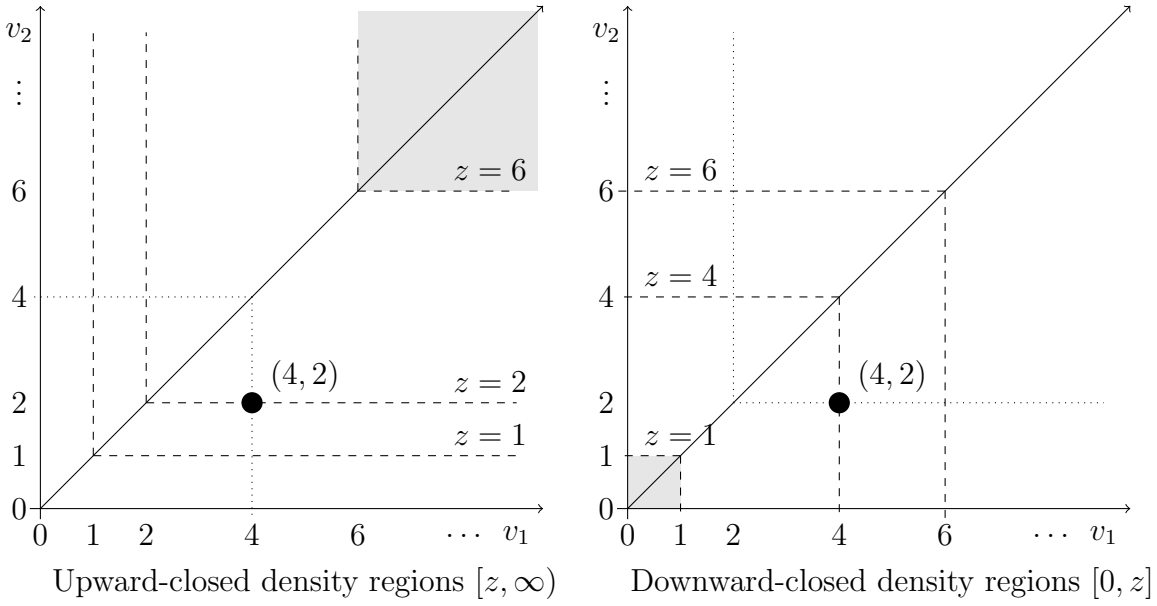

Figure 4.2. Identification of Positive-Weight Ranges of Blend-parameter z
We illustrate identification of integral endpoints for blends calculations when one side (left) is a blend over distributions with upward-closed domain $[z, \infty)$ for all z and the other side (right) is a blend over distributions with downward-closed domain $[0, z]$ for all z. Fix input values, e.g., $(v_1, v_2) = (4, 2)$. On each side, the shaded region illustrates a parameter that contributes 0 density (or mass) at the point $(4, 2)$. Observably, the effective ranges of integration for a blends calculation are respectively $[0, 2]$ and $[4, \infty)$. Generally, they are $[0, v_2]$ and $[v_1, \infty)$.

(4.6)

$$\begin{aligned}\text{result of } \mathrm{Sed}_{z,1} \text{ blend} &= \int_0^{v_2} o_{Ez} \cdot \mathrm{sed}_{z,1}(v_1) \cdot \mathrm{sed}_{z,1}(v_2) + o_{\mathrm{pm}} \cdot \mathrm{sed}_{0,1}(v_1) \cdot \mathrm{sed}_{0,1}(v_2) \\ &= \int_0^{v_2} \left(\frac{1}{2} e^{-z}\right) \cdot e^{-(v_1 - z)} \cdot e^{-(v_2 - z)}\, dz + \frac{1}{2} \cdot e^{-v_1} \cdot e^{-v_2} \\ &= \frac{1}{2} e^{-v_1} = g_{2D}(\boldsymbol{v})\end{aligned}$$

(4.7)

$$\begin{aligned}\text{result of } \mathrm{Ud}_{0,z} \text{ blend} &= \int_{v_1}^{\infty} \omega_{Uz} \cdot \mathrm{ud}_{0,z}(v_1) \cdot \mathrm{ud}_{0,z}(v_2) \\ &= \int_{v_1}^{\infty} \left(\frac{1}{2} z^2 e^{-z}\right) \cdot \frac{1}{z} \cdot \frac{1}{z}\, dz = \frac{1}{2} e^{-v_1} = g_{2D}(\boldsymbol{v})\end{aligned}$$

This concludes the blending of Shifted-Exponentials on one side and Uniforms on the other side into the same correlated distribution. The total weight of the system is 1 from the $\text{Sed}_{z,1}$ side:

$$o_{\text{pm}} + \int_0^\infty o_{Ez} = \frac{1}{2} + \int_0^\infty \frac{1}{2} \cdot e^{-z} dz = 1$$

Total weight is fairly obvious from the Shifted-Exponentials side. To confirm the total weight from the Uniforms side, we compute the total weight using two iterations of integration-by-parts:

$$\begin{aligned}
\int_0^\infty \omega_{Uz} = \int_0^\infty \frac{1}{2} z^2 e^{-z} dz &= \frac{1}{2} \left(\left[-z^2 e^{-z} \right]_0^\infty + \int_0^\infty 2z e^{-z} dz \right) \\
&= \frac{1}{2} \left(0 + 2 \left[\left[-z e^{-z} \right]_0^\infty + \int_0^\infty e^{-z} dz \right] \right) \\
&= \frac{1}{2} \left(2 \left[0 + 1 \right] \right) = 1
\end{aligned}$$

Lower Bound from Revenue Gap. Having Shifted-Exponentials-versus-Uniforms as an example of dual blends, we now illustrate how to plug them into Theorem 10 for a mechanism design setting to obtain a lower bound on prior independent approximation via calculation of $\text{opt}_{2,1}$ and $\text{opt}_{2,2}$. (Mechanism design is introduced in Section 2.2.1.) The setting is a 2-agent truthful auction with a revenue objective and the specific bound that we will obtain is $\alpha^{\mathcal{F}} \geq 1.1326$.

The first step is to compute optimal revenue for each distribution appearing in either side of the dual blend. Consider the δ_2 side – i.e., the $\text{Ud}_{0,z}$ side. The optimal revenue from $n = 2$ agents drawn i.i.d. from a uniform distribution $\text{U}[0, z]$ is known to be $\frac{5}{12} z$.

This solution completely covers the distributions used on the δ_2 side. Therefore we have

$$\text{(4.8)} \quad \text{opt}_{2,2} = \int_0^\infty \omega_{Uz} \cdot \text{OPT}_{\text{Ud}_{0,z}}(\text{Ud}_{0,z}) = \int_0^\infty \left(\frac{1}{2}z^2 e^{-z}\right) \cdot \frac{5}{12}z \; dz = \frac{1}{2} \cdot \frac{5}{12} \cdot 3! = \frac{5}{4}$$

where the factor of 3! results from noting that for positive integers k, after repeated integration by parts, $\int_0^\infty z^k e^{-z} dz = k!$, and in this case we have $k = 3$.

To analyze the δ_1 side – i.e. the $\text{Sed}_{z,1}$ side – it is sufficient to only look at the virtual value function for arbitrary $\text{Sed}_{z,1}$ because all distributions $\text{Sed}_{z,1}$ can be represented by draws from [$\text{Sed}_{0,1}$ conditioned on $v_2 \geq z > 0$]. Later, Fact 2 will state that given this setup, an observed value $\hat{v}$ will necessarily *have the same virtual value* regardless of which distribution-with-conditioning was used to generate it. In fact this property holds independently of the hazard rate parameter $\beta > 0$:

$$\text{(4.9)} \qquad \phi^{\text{Sed}_{z,\beta}}(v) = v - \frac{1 - \left(1 - e^{-\beta(v-z)}\right)}{\beta \cdot e^{-\beta(v-z)}} = \left(v - \frac{1}{\beta}\right) \; \forall \; z > 0, \; v \geq z$$

Fact 19. *For all $z \geq 0$ in the $\text{Sed}_{z,\beta}$ class of distributions including our specific case for $\beta = 1$, there exists a single mechanism which is optimal for them all: a Second Price Auction with a reserve price at $1/\beta$.*

This simplifies our task to calculate $\text{opt}_{2,1}$ because we can directly add up the revenue of the globally optimal mechanism across the correlated distribution $g(\boldsymbol{v}) = \frac{1}{2}e^{-v_1}$. Recalling

$v_1 \geq v_2$, we use symmetry across the line $v_1 = v_2$ and calculate this as:

(4.10)

$$\begin{aligned}
&\text{opt}_{2,1} \\
=&2\left[\int_0^1 0 \cdot dv_1 + \int_1^\infty \left(\int_0^1 \frac{1}{2}e^{-v_1}[\text{reserve price}]\ dv_2 + \int_1^{v_1} \frac{1}{2}e^{-v_1}[\text{second price}]\ dv_2\right) dv_1\right] \\
=&\ 2\left[\int_1^\infty \left(\int_0^1 \frac{1}{2}e^{-v_1} \cdot 1 \cdot dv_2 + \int_1^{v_1} \frac{1}{2}e^{-v_1} \cdot v_2 \cdot dv_2\right) dv_1\right] \\
=&\ \int_1^\infty e^{-v_1}\left(1 + \frac{1}{2}\left(v_1^2 - 1\right)\right) dv_1 \\
=&\ \left[\frac{-1}{2}e^{-v_1}\left(v_1^2 + 2v_1 + 3\right)\right]_1^\infty = \left[\left(\frac{1}{2} \cdot \frac{1}{e} \cdot 6\right) - 0\right] = \frac{3}{e} \approx 1.1036
\end{aligned}$$

Discussion. Having solved for $\text{opt}_{2,1}$ and $\text{opt}_{2,2}$, Theorem 10 gives us a necessary revenue gap and therefore a lower bound on approximation. The gap is[2]

$$\frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} = \frac{5/4}{3/e} = \frac{5e}{12} \approx 1.1326... \tag{4.11}$$

and reflects the following intuition. The adversary commits to an action up front, specifically δ_2 the Uniforms distribution. Interpreting this adversary-moves-first choice through Yao's Minimax Principle and our Benchmark Lemma 10, this choice sets the benchmark

[2] Dual blends analyses generally can depend on difficult and technically tricky computation. Here we exhibit a quick sanity check on the revenue of the Shifted-Exponentials side. As stated above, the optimal algorithm is the SPA with a reserve price of 1. Note then that the SPA with no reserve is sub-optimal, therefore its revenue lower bounds the optimal revenue:

The easiest way to calculate the revenue of the SPA is to note that for each uniform distribution $\text{Ud}_{0,z}$, the SPA gets $\frac{1}{3}z$. Compare this to the optimal auction per uniform distribution, which got $\frac{5}{12}z$. From this we see that the SPA simply gets $4/5$ths of the revenue of the optimal mechanisms for each distribution on the Uniforms side. Because $\text{opt}_{2,2} = 5/4$, it follows that the SPA gets 1. We confirm that this non-optimal revenue is less than the optimal revenue for the Shifted-Exponentials side which was $\text{opt}_{2,1} \approx 1.1036$.

(in the numerator) to $\text{opt}_{2,2}$. The designer knows the adversary's strategy and best responds. However $\delta_2^2 = g = \delta_1^2$ which shows that even when the designer knows that g was generated by δ_2, the designer can not do better than best responding directly to g, and further, can not do better than by making a particular assumption that g was generated by the Shifted-Exponentials blend δ_1. However critically by Fact 16, the performance of any mechanism is upper bounded by $\text{opt}_{2,1}$. Our intermediate conclusion is that our descriptions of Shifted-Expontials and Uniforms as a dual blend result in a revenue gap of any mechanism of at least 1.1326.

As previously mentioned, we were in fact able to proceed with analysis to this point without even specifying a class of allowable distributions $\mathcal{F}$. The lower bound holds for *any* class of distributions that includes all uniform distributions $\text{Ud}_{0,b}$. Let $\mathcal{F}^{\supseteq \text{UNIFORM}} = \{\mathcal{F} \mid \text{all } \text{Ud}_{0,b} \in \mathcal{F}\}$ be the (meta)-set of classes of distributions that contain all uniform distributions (as a subset).

Proposition 1. *Given a single-item, 2-agent, truthful auction setting with a revenue objective and with agent values in space* $[0, \infty)$*. For every class of distributions* $\mathcal{F} \in \mathcal{F}^{\supseteq \text{UNIFORM}}$*, the optimal prior independent approximation factor of any (truthful) mechanism is lower bounded as:*

$$\alpha^{\mathcal{F}} \geq 1.1326$$

As a corollary, the bound of equation (4.11) holds for two classes commonly used within mechanism design – it holds for both the regular $\mathcal{F}^{\text{reg}}$ (Definition 11) and monotone hazard rate $\mathcal{F}^{\text{mhr}}$ (Definition 10) distributions settings,[3] and this *depends only on* the adversary's

[3] For MHR and regular settings, our bound here is generally only illustrative – for $\mathcal{V} = [0, \infty)$, and with a restriction to *scale-invariant mechanisms* which is conjectured to be without loss, Allouah and

choice of The Uniforms side of the dual blend because all distributions $\text{Ud}_{a,b}$ are MHR and also regular (for the revenue objective). I.e., we have both that $\alpha^{\mathcal{F}^{\text{reg}}} \geq {}^{\text{opt}_{2,2}}/_{\text{opt}_{2,1}} \approx 1.1326$ and – as a stronger result because the MHR class is smaller per Fact 20– that $\alpha^{\mathcal{F}^{\text{mhr}}} \geq {}^{\text{opt}_{2,2}}/_{\text{opt}_{2,1}} \approx 1.1326$.

There is an important observation here (as alluded within the proof of Theorem 10 regarding the implicit relaxation to $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$). The Shifted-Exponentials comprising δ_1 are also both MHR and regular, *but this doesn't matter* – the upper bound on revenue of any mechanism as results from the Shifted-Exponentials blend-description (of the common correlated distribution) is a consequence of nature itself as follows solely from the adversary's choice of the Uniforms blend, and the structure of that "consequence" (namely, the revenue-upper-bound structure of the Shifted-Exponentials) faces no constraints at all. The adversary chooses its blend subject to a particular $\mathcal{F}$ but the other side of the dual blend can be composed of any subset of distributions in $\mathcal{F}^{\text{all}}$ (and examples exist for which this is the case).

4.5. A Dual Blend Inducing Novel Mechanism Design Results

The first goal of this section is to exhibit a second concrete example of dual blends. In contrast to the previous section, the example here proceeds in two steps: (1) we describe a relaxed solution that allows *infinite weight* which is not directly usable for lower bounds but has simpler algebraic form; and (2), we show that this relaxed solution can be modified to become proper dual blends.

Besbes (2018) show a tight bound for $\mathcal{F}^{\text{mhr}}$ of $\alpha^{\mathcal{F}^{\text{mhr}}} \approx 1.398$ and this thesis (Theorem 5) showed a tight bound for $\mathcal{F}^{\text{reg}}$ of $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$.

The second part of the section uses the dual blends example to state novel lower bounds for two distinct problems from mechanism design. Interestingly, the distinct objectives of these two problems results in *the two sides of the dual blends playing opposite roles* (as choice of the adversary versus gap-inducing upper bound). Later in Section 5.4, we discuss the implications of this observation in terms of precluding *Blackwell ordering* between the two sides of the dual blend.

4.5.1. A Dual Blends Example Starting with Infinite Weights

This section provides another explicit example of dual blends – with motivation for the chosen distributions from themes in mechanism design. First, we will describe a blends-type solution that has unbounded input support and infinite total weight (so it is not a probability distribution and it is not possible to re-normalize its weights to become one).[4] Second, we modify the infinite-weight solution to have finite weight in a finite input space (which can be normalized to 1 for any fixed weight).

For this running dual blends example, the δ_1 side will be parameterized by a base class of upward-closed Quadratics (called "equal revenue" in the mechanism design literature), with PDF given by $\text{qud}_z(x) = z/x^2$ and CDF given by $\text{Qud}_z(x) = 1 - z/x$ on $[z, \infty)$. The δ_2 side will be a base class of downward-closed Uniforms, with PDF given by $\text{ud}_{0,z}(x) = 1/z$ and CDF given by $\text{Ud}_{0,z}(x) = x/z$ on $[0, z]$. (Generally, let $\text{Ud}_{a,b}$ be the Uniform distribution on $[a, b]$.)

[4] The elements of a blend δ are technically densities but we generally refer to them as *weights*, i.e., the weight corresponding to a distribution within the mixture over $\mathcal{F}$ according to δ. We do this to accommodate a relaxed definition for *blend* which allows arbitrary total weight (including infinite).

Infinite-weight Blends. We start by describing the weights o_F corresponding to δ_1 and weights ω_F corresponding to δ_2. Because we first allow the total weight to be infinite, we only require the function g (relaxed to be a "correlated function" rather than a correlated distribution) to match up its output *mass* at every input (cf., density of a correlated distribution).

The weights of the upward-closed Quadratics blend (δ_1) are as follows:

- weights $o_{Qz} = \frac{2}{z}dz$ on all upward-closed distributions Qud_z for $z \in (0, \infty)$.

The weights of the downward-closed Uniforms blend (δ_2) are as follows:

- weights $\omega_{Uz} = \frac{2}{z}dz$ on all downward-closed distributions $\mathrm{Ud}_{0,z}$ for $z \in (0, \infty)$.

Using symmetry, we analyze mass in the cone $v_1 \geq v_2 \geq 0$. The calculations of total mass at any point $\boldsymbol{v} \in (0, \infty)^2$ are confirmed to be equal from either dual blends description of the common correlated function g.

(4.12)
$$\text{result of } \mathrm{Qud}_z \text{ blend} = \int_0^{v_2} o_{Qz} \cdot \mathrm{qud}_z(v_1) \cdot \mathrm{qud}_z(v_2) = \int_0^{v_2} \frac{2}{z} \cdot \frac{z}{v_1^2} \cdot \frac{z}{v_2^2}\, dz \;\; = \frac{1}{v_1^2} = g(\boldsymbol{v})$$

(4.13)
$$\text{result of } \mathrm{Ud}_{0,z} \text{ blend} = \int_{v_1}^{\infty} \omega_{Uz} \cdot \mathrm{ud}_{0,z}(v_1) \cdot \mathrm{ud}_{0,z}(v_2) = \int_{v_1}^{\infty} \frac{2}{z} \cdot \frac{1}{z} \cdot \frac{1}{z}\, dz \quad = \frac{1}{v_1^2} = g(\boldsymbol{v})$$

The setup of these calculations is expanded in detail in Section 4.4. As desired, each side of the dual blends describes exactly the same function g over $\mathcal{V}^2$. The remaining issue to be addressed is that the total weight of all included distributions is divergent: $\int_0^{\infty} \frac{2}{z}dz = \infty$.

Modification to Finite-weight Blends. Next we show how to modify the infinite-weight solution above to a proper dual blends solution with approximately the same elements. Consider input support $\mathcal{V} = [1, h]$ for $1 < h < \infty$. First we define the weights o_F and ω_F, largely informed by the infinite-weight solution. We let the total weight in the system be any constant and can assume that it gets normalized to 1 later. In fact the total weight will be: $1 + \int_1^h \frac{2}{z} dz = 1 + 2 \ln h$.

The Quadratics have the same general description as the infinite-weight case but are now top-truncated at h, with truncated density moved to a point mass at h.[5] Formally, Quadratics have PDF $\overleftarrow{\text{qud}}_z^{h'}(x) = {}^{z}\!/\!_{x^2}$ on $[1, h)$ and point mass $\overleftarrow{\text{qud}}_z^{h'}(h) = {}^{1}\!/\!_{h}$, correspondingly CDF $\overleftarrow{\text{Qud}}_z^{h'}(x) = 1 - {}^{z}\!/\!_{x}$ on $[1, h)$ and $\overleftarrow{\text{Qud}}_z^{h'}(h) = 1$.

The Uniforms have the same general description as the infinite-weight case but now have domain lower bound at 1 and *allow* top-truncation at h. Formally, Uniforms without truncation have PDF $\text{ud}_{1,z}(x) = {}^{1}\!/\!_{z-1}$ and CDF $\text{Ud}_{1,z}(x) = {}^{x-1}\!/\!_{z-1}$ on $[1, z]$. Uniforms with truncation have PDF $\overleftarrow{\text{ud}}_{1,b}^{h'}(x) = {}^{1}\!/\!_{b-1}$ on $[1, h)$ and point mass $\overleftarrow{\text{ud}}_{1,b}^{h'}(h) = {}^{b-h}\!/\!_{b-1}$, correspondingly $\overleftarrow{\text{Ud}}_{1,b}^{h'}(x) = {}^{x-1}\!/\!_{b-1}$ on $[1, h)$ and $\overleftarrow{\text{Ud}}^{h}(h) = 1$.

The weights of the upward-closed Quadratics blend (δ_1) are as follows:

- point mass of weight $o_{\text{pm}} = 1$ on (truncated) distribution $\overleftarrow{\text{Qud}}_1^{h'}$;
- weights $o_{Qz} = \frac{2}{z} dz$ on all upward-closed (truncated) distributions $\overleftarrow{\text{Qud}}_z^{h'}$ for $z \in [1, h]$.

The weights of the downward-closed Uniforms blend (δ_2) are as follows:

- point mass of weight $\omega_{\text{pm}} = \frac{(2h-1)^2}{h^2}$ on (truncated) distribution $\overleftarrow{\text{Ud}}_{1,2h}^{h'}$;

[5] We briefly explain notation of $\overleftarrow{F}^{h'}$. Let a left-over-arrow modify the domain-upper-bound of F to be h. The accent in $\overleftarrow{F}^{h'}$ indicates that density above h is *truncated* to h as a point mass, i.e., the original CDF jumps to 1 at h.

- weights $\omega_{Uz} = \frac{2(z-1)^2}{z^3}dz$ on all downward-closed distributions $\mathrm{Ud}_{1,z}$ for $z \in [1, h]$.

(In fact, we use only one uniform distribution with truncation: $\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}$.) With the introduction of pure point masses into underlying distributions, recall that dual blends must match up for every dimension count. For convenience we re-state Definition 20. Then we calculate and confirm all (un-normalized) densities from both sides.

Definition 20. *Define the* count of dimensional density *by the number of (axis-aligned) dimensions $i \in \{1, \ldots, n\}$ in which density is continuous: dv_i or dz.*

$$
\begin{aligned}
g_{2D}(\boldsymbol{v}) &= \int_1^{v_2} o_{Qz} \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(v_1) \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(v_2) + o_{\mathrm{pm}} \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(v_1) \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(v_2) \\
&= \int_1^{v_2} \frac{2}{z} \cdot \frac{z}{v_1^2} \cdot \frac{z}{v_2^2}\, dz + 1 \cdot \frac{1}{v_1^2} \cdot \frac{1}{v_2^2} = \frac{1}{v_1^2} \\
g_{2D}(\boldsymbol{v}) &= \int_{v_1}^{h} \omega_{Uz} \cdot \mathrm{ud}_{1,z}(v_1) \cdot \mathrm{ud}_{1,z}(v_2) + \omega_{\mathrm{pm}} \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(v_1)\right) \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(v_2)\right) \\
&= \int_{v_1}^{h} \frac{2(z-1)^2}{z^3} \cdot \frac{1}{(z-1)^2}\, dz + \frac{(2h-1)^2}{h^2} \cdot \left(\frac{1}{(2h-1)}\right)^2 = \frac{1}{v_1^2} \\
g_{0D}(h, h) &= \int_1^{v_2=h} o_{Qz} \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(h) \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(h) + o_{\mathrm{pm}} \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(h) \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(h) \\
&= \int_1^{h} \frac{2}{z} \cdot \frac{z}{h} \cdot \frac{z}{h}\, dz + 1 \cdot \frac{1}{h} \cdot \frac{1}{h} = 1 \\
g_{0D}(h, h) &= \int_{v_1=h}^{h} \omega_{Uz} \cdot \mathrm{ud}_{1,z}(h) \cdot \mathrm{ud}_{1,z}(h) + \omega_{\mathrm{pm}} \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(h)\right) \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(h)\right) \\
&= 0 + \frac{(2h-1)^2}{h^2} \cdot \left(\frac{h}{(2h-1)}\right)^2 = 1 \\
g_{1D}(h, v_2) &= \int_1^{v_2} o_{Qz} \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(h) \cdot \overleftarrow{\mathrm{qud}}^{h'}_z(v_2) + o_{\mathrm{pm}} \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(h) \cdot \overleftarrow{\mathrm{qud}}^{h'}_1(v_2)
\end{aligned}
$$

$$= \int_1^{v_2} \frac{2}{z} \cdot \frac{z}{h} \cdot \frac{z}{v_2^2}\, dz + 1 \cdot \frac{1}{h} \cdot \frac{1}{v_2^2} = \frac{1}{h}$$

$$g_{1D}(h, v_2) = \int_{v_1=h}^{h} \omega_{Uz} \cdot \mathrm{ud}_{1,z}(h) \cdot \mathrm{ud}_{1,z}(v_2) + \omega_{\mathrm{pm}} \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(h)\right) \cdot \left(\overleftarrow{\mathrm{ud}}^{h'}_{1,2h}(v_2)\right)$$

$$= 0 + \frac{(2h-1)^2}{h^2} \cdot \left(\frac{h}{2h-1}\right) \cdot \left(\frac{1}{2h-1}\right) = \frac{1}{h}$$

As desired, each side of the dual blends yields the same function $g = (g_{2D},\ g_{0D},\ g_{1D})$. Calculations to confirm total weight is equal from any description of g or δ_1 or δ_2 are given in Appendix B.3.

Observable Structure of Dual Blends. Having completed two finite-weight blends solutions (Shifted-Exponentials-versus-Uniforms in Section 4.4 and now Quadratics-versus-Uniforms here in Section 4.5.1), we identify the following common structure which is observed in our example dual blends so far (and which later will be implicated but not necessary within our general methods of Chapter 5).

Illustrated here on input support $\mathcal{V} = (0, \infty)$, the observed structure is: the input size is $n = 2$; distributions composing δ_1 are *upward-closed* and are parameterized by $z \in (0, \infty)$ with domain $[z, \infty)$; and distribution in δ_2 are *downward-closed* and are parameterized by $z \in (0, \infty)$ with domain $(0, z]$. In further detail:

- δ_1 is a distribution with weights o_z over realized values of a single distributional parameter for a given upward-closed distribution; e.g., δ_1 was a distribution over domain-lower-bounds z of the Shifted-Exponentials in Section 4.4;
- δ_2 is a distribution with weights ω_z over realized values of a single distributional parameter for a given x-closed distribution; e.g., δ_2 was a distribution over domain-upper-bounds z of the Uniforms in Section 4.4.

We conjecture that no dual blends exist for $n > 2$ from our Blends Technique. For the intuition of this conjecture, see our discussion of "algebraic consequences of the integral endpoints" in Section 5.1. We believe that this upward-closed/downward-closed dual structure is an important property that deserves further study.

4.5.2. Illustrative Results in Mechanism Design

We state two prior independent lower bounds in mechanism design with their proofs deferred to the following sections.

The results follow from the exact same dual blends solution (using Quadratics-versus-Uniforms with finite weight of Section 4.5.1 and the Blends Technique of Definition 19). *Revenue* and *residual surplus* are two objectives within mechanism design (Section 2.2.1). Theorem 11 (below, for a revenue objective) uses an adversarial distribution over the Uniforms side of the dual blend. By contrast, Theorem 12 (for a residual surplus objective) uses an adversarial distribution over the Quadratics side. This dichotomy of the respective adversaries' choices highlights how even a single example of dual blends can be distinctly applied to two algorithm settings in order to identify a PI approximation lower bound within each setting.

A fixed prior independent lower bound is stronger if it holds for a smaller class of distributions. Let $L^{\mathcal{F}}$ be a lower bound on the optimal approximation factor $\alpha^{\mathcal{F}}$ for a class $\mathcal{F}$. Fact 20 makes clear that $L^{\mathcal{F}}$ holds additionally for a superclass $\mathcal{E}$:

Fact 20. *Given two classes of distributions $\mathcal{E}$ and $\mathcal{F}$ such that $\mathcal{E} \supset \mathcal{F}$. Then $\alpha^{\mathcal{E}} \geq \alpha^{\mathcal{F}} \geq L^{\mathcal{F}}$.*

Thus, we give our results for the smallest classes of distributions in order to state the strongest bounds from our analysis. Define two sub-classes: Uniforms $\mathcal{F}^{\text{unif}}[1,h] = \{\overleftarrow{\text{Ud}}^{h'}_{1,b} \ : \ 1 \leq b\} \equiv$ uniforms on $[1,b]$ truncated at h; and Quadratics $\mathcal{F}^{\text{quad}}[1,h] = \{\overleftarrow{\text{Qud}}^{h'}_{a} \ : \ 1 \leq a \leq h\} \equiv$ quadratics on $[a,h]$ truncated at h. We explain the approach for both theorems but full proofs are deferred to Section 2.2.

Theorem 11. *Given a single-item, 2-agent, truthful auction setting with a revenue objective and with agent values restricted to support* $[1,h]$ *for* $h > 2$. *For the class of uniform distributions* $\mathcal{F}^{\text{unif}}$*, the optimal prior independent approximation factor of any (truthful) mechanism is lower bounded as:*

$$\alpha_h^{\mathcal{F}^{\text{unif}}} \geq \frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} = \frac{\frac{23h}{6} - \frac{7}{2} - \ln(h/2)}{3h-2} = L_h^{\mathcal{F}^{\text{unif}}} \tag{4.14}$$

The lower bound $L_h^{\mathcal{F}^{\text{unif}}} \to 23/18 \approx 1.2777$ *as* $h \to \infty$ *and this is the supremum of* $L_h^{\mathcal{F}^{\text{unif}}}$ *over* $h \geq 1$.

The canonical PI revenue maximization problem measures worst-case approximation with respect to the class of *regular* distributions $\mathcal{F}^{\text{reg}}$ (Definition 11). All of our Uniforms are regular: $\mathcal{F}^{\text{reg}} \supset \mathcal{F}^{\text{unif}}$. As a corollary, we get a lower bound for regular distributions: $\alpha_h^{\mathcal{F}^{\text{reg}}} \geq L_h^{\mathcal{F}^{\text{unif}}}$.

As already stated, the proof of Theorem 11 follows the script of the Blends Technique (Definition 19). We set $\delta_2 \in \Delta(\mathcal{F}^{\text{unif}})$ to be the Uniforms blend with finite weights (page 99) and we set $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$ to be the corresponding Quadratics dual blend. The Second Price Auction (SPA; Definition 5) is optimal for all Quadratics in $\mathcal{F}^{\text{quad}} \subset \mathcal{F}^{\text{all}}$; the lower bound $h > 2$ is necessary so that the SPA is not also optimal for all Uniform

distributions with positive weight in δ_2 (otherwise there is no gap: ${}^{\text{opt}_{2,2}}/_{\text{opt}_{2,1}} = 1$). Given these, the right-hand side of equation (4.14) is simply the result of evaluating ${}^{\text{opt}_{2,2}}/_{\text{opt}_{2,1}}$ (and recalling from Definition 18 that $\text{opt}_{n,i} = \mathbf{E}_{F\sim\delta_i}[\text{OPT}_F(F)]$).

Theorem 12. *Given a single-item, 2-agent, truthful auction setting with a residual surplus objective and with agent values restricted to support* $[1, h]$ *for* $h \geq 8.56$*. For the class of quadratic distributions* $\mathcal{F}^{\text{quad}}$*, the optimal prior independent approximation factor of any (truthful) mechanism is lower bounded as:*

$$\alpha_h^{\mathcal{F}^{\text{quad}}} \geq \frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} > \frac{4h^2 - 2h - h\ln h - e\ln h - e}{4h^2 - 3h - h\ln h} = L_h^{\mathcal{F}^{\text{quad}}} \tag{4.15}$$

The lower bound $L_h^{\mathcal{F}^{\text{quad}}} \to 1$ *as* $h \to \infty$*. As an example bound: for* $h \in \mathbb{N}$*, the maximum of* $L_h^{\mathcal{F}^{\text{quad}}}$ *is achieved at* $h = 18$ *with* $L_{18}^{\mathcal{F}^{\text{quad}}} \approx 1.00623$*.*

The canonical PI residual surplus maximization problem measures worst-case approximation with respect to the class of all distributions $\mathcal{F}^{\text{all}}$.[6] As a corollary, we get a lower bound for all distributions: $\alpha_h^{\mathcal{F}^{\text{all}}} \geq L_h^{\mathcal{F}^{\text{quad}}}$.

Once again, the proof of Theorem 12 uses the Blends Technique. This time we set $\delta_2 \in \Delta(\mathcal{F}^{\text{quad}})$ to be the Quadratics blend with finite weights and set $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$ to be the corresponding Uniforms. The Lottery (Definition 9) is optimal for all Uniforms in $\mathcal{F}^{\text{unif}} \subset \mathcal{F}^{\text{all}}$; the lower bound $h \geq 8.56$ is necessary so that the Lottery is not also optimal for all Quadratics with positive weight in δ_2 (otherwise there is no gap). Note,

[6] We note the contrast: $\mathcal{F}^{\text{all}}$ is standard for prior independent design with a residual surplus objective, whereas $\mathcal{F}^{\text{reg}}$ is standard with a revenue objective. As partial explanation: for the class $\mathcal{F}^{\text{all}}$, Hartline and Roughgarden (2014) show that constant-approximation is possible for residual surplus, and also show a super-constant lower bound for revenue. Revenue maximization restricts to regular distributions which satisfy a natural concavity property, and for which constant-approximation is possible (the first upper bound was from Dhangwatnotai et al. (2015)).

the right-hand side of equation (4.15) is a simplified lower bound on the ratio $^{\text{opt}_{2,2}}/_{\text{opt}_{2,1}}$ as shown in the statement.

Previously for 2-agent auctions for revenue and *unbounded value space*, with the additional restriction to *scale-invariant* mechanisms, Allouah and Besbes (2018) proved for monotone hazard rate distributions ($\mathcal{F}^{\text{mhr}}$; Definition 10) that the SPA is optimal and gave the optimal approximation $\alpha^{\mathcal{F}^{\text{mhr}}} \approx 1.398$ (Theorem 20); and also proved for regular distributions ($\mathcal{F}^{\text{reg}}$) the first-ever PI lower bound. This thesis gave the optimal mechanism and approximation $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$ (Theorem 5). For residual surplus, there is no previous lower bound. Our mechanism design results have not been optimized in order to identify best lower bounds from the Blends Technique.

4.5.3. Revenue Gap from Quadratics-versus-Uniforms

The goal of this section is to use the Blends Technique (Definition 19) to prove a revenue gap for the Quadratics-versus-Uniforms dual blend, resulting in a prior independent lower bound (summarized in Equation (4.14) in Theorem 11 and copied at the end of this section).

Recall value space is $\mathcal{V}^2 = [1, h]^2$ with an assumption that $h > 2$.[7] We still use symmetry to assume $v_1 \geq v_2 \geq 0$ in calculations and will then count permutations where

[7] The assumption of $h > 2$ is necessary to make the result interesting. Because of the assumption that value space has domain lower bound at 1, uniform distributions with domain upper bound at most 2 are trivially optimized by the SPA – which is the same as for the Quadratics in these dual blends – and thus do not induce an approximation gap (the ratio is 1).

necessary. For use in this section and the next, we re-summarize the function g:

$$
\begin{aligned}
(4.16) \qquad g_{2D}(\boldsymbol{v}) &= \text{result of } \overleftarrow{\text{Qud}}_z^{h'} \text{ blend} \\
&= \text{result of } \text{Ud}_{0,z} \text{ blend and } \overleftarrow{\text{Ud}}_{1,2h}^{h'} = \frac{1}{v_1^2} \\
g_{0D}(\boldsymbol{v}) &= \text{result of truncation in } \overleftarrow{\text{Qud}}_z^{h'} \text{ blend} \\
&= \text{no point mass density from } \text{Ud}_{0,z} \text{ blend, result of } \overleftarrow{\text{Ud}}_{1,2h}^{h'} = 1 \\
g_{1D}(\boldsymbol{v}) &= \text{result of } \textbf{exactly one} \text{ truncation in } \overleftarrow{\text{Qud}}_z^{h'} \text{ blend} \\
&= \text{no point mass density from } \text{Ud}_{0,z} \text{ blend, result of } \overleftarrow{\text{Ud}}_{1,2h}^{h'} = \frac{1}{h}
\end{aligned}
$$

Also for use in this section, we state the following fact regarding uniform order statistics.

Fact 21. *Given (unordered)* $\boldsymbol{v} = (v_1, \ldots, v_n)$ *which are* n *i.i.d. draws from the uniform distribution* $\text{Ud}_{0,1}$. *Let* $k = 1$ *be the first, largest order statistic, and count order statistics down to* $k = n$ *the last, smallest order statistic. The expected value of an order statistic* $v_{(k)}$ *is given by* $\mathbf{E}_{\boldsymbol{v} \sim \text{Ud}_{0,1}} \left[v_{(k)} \right] = \frac{n+1-k}{n+1}$.

4.5.3.1. Expected Optimal Revenue from Quadratics. We calculate the expected optimal revenue from the Quadratics side $\text{opt}_{2,1}$ using the $\boldsymbol{o}$ weights above. The revenue of the Quadratics blend is easy to calculate *because every distribution that is a component of the blend is an equal revenue distribution* for which offering every price in $[v_{(2)}, h]$ to the largest-valued agent gets the same revenue and is optimal (see Definition 13 and its surrounding discussion).

The immediate consequence is that there exists a single mechanism that is optimal against every distribution in the δ_1 Quadratics blend: the anonymous price-posting mechanism AP_h with constant price h is globally optimal. Therefore, $\mathrm{opt}_{2,1} = \mathrm{AP}_h(g)$ which is the revenue of posting price h against the entire correlated distribution g.

The revenue conditioned on selling is obviously h. The probability of selling can be obtained from the equations of line (4.16) to determine total density where at least one agent has value h, which is exactly the total of 0D and 1D density over all of value space: $g_{0D} \cdot 1 + 2 \cdot g_{1D} \cdot (h-1) = {}^{(3h-2)}/_{h}$. Revenue from the Quadratics blend δ_1 is given by

$$\mathrm{opt}_{2,1} = \mathrm{AP}_h(g) = h \cdot \frac{(3h-2)}{h} = 3h - 2 \tag{4.17}$$

4.5.3.2. Expected Optimal Revenue from Uniforms. Because of the lower bound of value space at 1, optimal revenue analyses for the Uniforms break down by both distribution and type of optimal reserve price (which is either the $\mathrm{Ud}_{0,z}$ monopoly reserve price, or the lower bound 1). The way we implement all Uniforms distributions here is equivalent to conditioning a random draw from $\mathrm{Ud}_{0,z}$ to be at least 1. This structure makes Fact 2 applicable to our distributions, thus we can use virtual values as if the values were drawn from $\mathrm{Ud}_{0,z}$ rather than its respective $\mathrm{Ud}_{1,z}$. By observation, all Uniforms with positive weight in δ_2 are regular. It is a well-known corollary to Theorem 3 that the optimal mechanism given n agents drawn i.i.d. from a uniform distribution $\mathrm{Ud}_{0,z}$ is the SPA with reserve price at the monopoly price ${}^{z}/_{2}$. We have the following summary of monopoly prices over the distributions in δ_1:

- monopoly price h for $\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}$ (the truncation at h observably does not affect the monopoly price);

- monopoly price 1 for $\mathrm{Ud}_{1,z}$ for $z \in [1,2]$ (effectively the SPA);
- monopoly price $z/2$ for $\mathrm{Ud}_{1,z}$ for $z \in [2,h]$ (which is the same as the monopoly price for $\mathrm{Ud}_{0,z}$).

We treat these cases in sequence to compute $\mathrm{opt}_{2,2}$, incorporating the weights $\boldsymbol{\omega}$. The contribution of the $\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}$ distribution is actually the same as the entire revenue $\mathrm{opt}_{2,2} = 3h-2$ of the Quadratics. This follows from: its optimal mechanism – post price h – is the same as the globally optimal mechanism for Quadratics; and, the distribution $\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}$ accounts for every possible realization in which an agent shows up with value h (when considering any distribution in δ_2).

The second and third cases represent revenue from the Uniform blend's random draw according to weights $\omega_{Uz} = \frac{2(z-1)^2}{z^3}dz$. For the case $z \leq 2$, the optimal mechanism is the SPA, therefore the optimal mechanism always sells and its expected revenue is exactly the expected value of v_2. Using Fact 21, the expected value of v_2 is [1 plus 1/3 of the width of the range $[1,z]$]. The total contribution from this case $z \leq 2$ is:

$$\begin{aligned}(4.18)\qquad &\mathrm{OPT}_{\mathrm{Ud}_{1,z}}(\mathrm{Ud}_{1,z})\\ &= \int_1^2 \frac{2(z-1)^2}{z^3}\cdot\left(1+\frac{z-1}{3}\right)\,dz = \int_1^2 \frac{2z^2-4z+2+\frac{2}{3}z^3-2z^2+2z-\frac{2}{3}}{z^3}\,dz\\ &= \int_1^2 \frac{2}{3}-\frac{2}{z^2}+\frac{4}{3z^3}\,dz = \frac{2}{3}+\left[\frac{2}{z}-\frac{2}{3z^2}\right]_1^2 = \frac{2}{3}+\left(1-\frac{1}{6}\right)-\left(2-\frac{2}{3}\right) = \frac{1}{6}\end{aligned}$$

For the remaining case $z \geq 2$, the expected revenue conditioned on selling depends on distribution-specific reserve prices (there is no reduction to the SPA). We separate uniform

draws of the values $\boldsymbol{v}$ into three natural sub-cases and calculate the optimal revenue afterwards (given a $\mathrm{Ud}_{1,z}$): [8]

- both values are smaller than the reserve; we fail to sell, with probability $\frac{(\frac{z-2}{2})^2}{(z-1)^2}$;
- (2 permutations of) one value is larger and one smaller than the reserve; we sell at the reserve price of $z/2$, each permutation with probability $\frac{(\frac{z-2}{2})(\frac{z}{2})}{(z-1)^2}$;
- both values are larger than the reserve; we sell at the conditional expected value of v_2 which is $\mathbf{E}_{\boldsymbol{v}\sim\mathrm{Ud}_{z/2,z}}[v_2] = 2z/3$, with probability $\frac{(\frac{z}{2})^2}{(z-1)^2}$.

The optimal revenue from a distribution $\mathrm{Ud}_{1,z}$ for $z \geq 2$ is

$$\mathrm{OPT}_{\mathrm{Ud}_{1,z}}(\mathrm{Ud}_{1,z}) = 2 \cdot \frac{(\frac{z-2}{2})(\frac{z}{2})}{(z-1)^2} \cdot \frac{z}{2} + \frac{(\frac{z}{2})^2}{(z-1)^2} \cdot \frac{2z}{3} = \frac{1}{(z-1)^2}\left(\frac{5}{12}z^3 - \frac{1}{2}z^2\right) \tag{4.19}$$

Analogous to equation (4.18), the total contribution from this last case $z \geq 2$ results from a weighted integral and is:

$$\int_2^h \frac{2(z-1)^2}{z^3} \cdot \mathrm{OPT}_{\mathrm{Ud}_{1,z}}(\mathrm{Ud}_{1,z})\, dz = \int_2^h \frac{5}{6} - \frac{1}{z}\, dz = \frac{5}{6}(h-2) - \ln h/2 \tag{4.20}$$

Total revenue from the Uniforms blend adds up as

$$\mathrm{opt}_{2,2} = (3h-2) + \frac{1}{6} + \frac{5}{6}(h-2) - \ln h/2 = \frac{23}{6} \cdot h - \frac{7}{2} - \ln h/2 \tag{4.21}$$

4.5.3.3. The Revenue Gap as Lower Bound. For 2-agent, 1-item prior independent auctions with a revenue objective, we have now established a necessary revenue gap via

[8] Note, the quantity $z/2 - 1 = (z-2)/2$ is the length of the (uniform) region below the monopoly price at $z/2$, versus $z/2$ is the length above it.

blends:

$$\frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} = \frac{\frac{23h}{6} - \frac{7}{2} - \ln(h/2)}{3h-2}, \qquad \lim_{h\to\infty} \frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} = \frac{23}{18} \approx 1.27777 \tag{4.22}$$

where the limit calculation is trivial from application of l'Hopital's rule. The applicability of the revenue gap as a prior independent lower bound is subject to the design problem's parameter $\mathcal{F}$ to describe the adversary's allowable choice set of distributions. By inspection, all of the Uniforms composing δ_2 are included in small-class $\mathcal{F}^{\text{unif}}[1,h] = \{\overleftarrow{\text{Ud}}_{1,b}^{h'} \ : \ 1 \le b\} \equiv$ uniforms on $[1,b]$ truncated at h. Putting this together with equation (4.22) and Theorem 10 gives Theorem 11:

Theorem 11. *Given a single-item, 2-agent, truthful auction setting with a revenue objective and with agent values restricted to support $[1,h]$ for $h > 2$. For the class of uniform distributions $\mathcal{F}^{\text{unif}}$, the optimal prior independent approximation factor of any (truthful) mechanism is lower bounded as:*

$$\alpha_h^{\mathcal{F}^{\text{unif}}} \ge \frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} = \frac{\frac{23h}{6} - \frac{7}{2} - \ln(h/2)}{3h-2} = L_h^{\mathcal{F}^{\text{unif}}}$$

The lower bound $L_h^{\mathcal{F}^{\text{unif}}} \to 23/18 \approx 1.2777$ as $h \to \infty$ and this is the supremum of $L_h^{\mathcal{F}^{\text{unif}}}$ over $h \ge 1$.

Further, all Uniforms are regular (for revenue, for which $\mathcal{F}^{\text{reg}}$ is the standard comparison class of distributions for prior independent design), thus as a corollary, our bound here holds for $\mathcal{F}^{\text{reg}}$.

Previously in Theorem 5, the *optimal* prior independent approximation factor was given for the version of this setting which allowed unbounded values in $[0,\infty)$. The optimal

factor was ~ 1.907. Note, we should expect the approximation factor of the restricted value space to be smaller than the unbounded value space, because the mechanism can specifically take advantage of information relating to scale in the latter case. The optimal mechanism for unbounded value space was an a priori mixture over v_2-markup prices of 1 (the SPA) and ~ 2.44. Certainly, we would not expect the optimal mechanism to have the same form for the value space $[1, h]$, because it will not commit a priori to posting a marked-up price of $2 \cdot v_2$ when it could be that $v_2 \in [h/2, h]$. This makes prior independent design in a finite value space setting a distinct problem in terms of analysis, in comparison to the unbounded value space setting.

4.5.4. Residual Surplus Gap from Quadratics-versus-Uniforms

Paralleling the previous section, the goal of this section is to use the Blends Technique to prove a residual surplus gap for the Quadratics-versus-Uniforms dual blend, resulting in a prior independent lower bound (summarized in equation (4.15) in Theorem 12 and copied at the end of this section). un-normalized densities (which will cancel at the end), and Fact 21 (which states that the expected value of n order statistics from a distribution $\mathrm{Ud}_{0,z}$ divide the range into $n+1$ equal parts). We include here a similar assumption to the one we had for the revenue gap – we assume $h \geq 8.56$ (otherwise our relaxed analysis does not show a gap).

The residual surplus gap presented here uses the same description of finite-weight Quadratics-versus-Uniforms dual blends, but the adversary will now in fact *choose the distribution over the Quadratics and use them to set the benchmark*, whereas the revenue-adversary set the benchmark via the Uniforms. (Thus, we reassign δ_1 to describe the

Uniforms side and δ_2 to describe the Quadratics side.) As a consequence, we now have calculation of residual surplus that is "easy" for the Uniforms rather than for the Quadratics (which is reversed in comparison to revenue calculations). This follows because, for residual surplus, it is now the Uniforms side for which a single dominant mechanism exists (the Lottery of Definition 9). We will calculate $\text{opt}_{2,1}$ for the easy Uniforms side first.

Afterwards, the calculation of $\text{opt}_{2,2}$ for expected optimal residual surplus of the Quadratics faces some technical complexities. For simplicity, we will calculate instead a lower bound on $\text{opt}_{2,2}$. This is sufficient because we are designing a residual surplus gap (via the Blends Technique) between the adversary's benchmark set by the Quadratics and an upper bound on expected performance of any algorithm as set by the Uniforms. By using a lower bound on the ratio's numerator, we will exhibit a weaker – but legitimate – non-trivial lower bound on prior independent approximation.

4.5.4.1. Expected Optimal Residual Surplus from Uniforms. We calculate the expected optimal residual surplus from the Uniforms side ($\text{opt}_{2,1}$) using the $\boldsymbol{\omega}$ weights above. The residual surplus of the Uniforms blend is easy to calculate because *every distribution has everywhere decreasing virtual value and therefore it is optimal to iron the entire region of value space.* The immediate consequence is that there exists a single mechanism that is optimal against every distribution in the δ_1 Uniforms blend: the 2-lottery mechanism $\text{LOT}_2 = \text{AP}_0$ is optimal (see definitions from page 39). I.e., $\text{opt}_{2,1}$ can be calculated directly from the expectation of one draw $v \sim g$ (and with price 0).

The probability of selling is obviously 1. In this case, the expected residual surplus given any distribution can be obtained as the mean of the distribution: $\text{OPT}_{\text{Ud}_{1,z}}(\text{Ud}_{1,z}) =$

$1+{}^{(z-1)}/_{2} = {}^{(z+1)}/_{2}$. Given the distribution $\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}$, expected residual surplus is calculated to be:

$$\mathrm{OPT}_{\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}}(\overleftarrow{\mathrm{Ud}}^{h'}_{1,2h}) = h \cdot \left(\frac{h}{(2h-1)}\right) + \left(1 + \frac{(h-1)}{2}\right) \cdot \left(\frac{(h-1)}{(2h-1)}\right) = \frac{(3h^2-1)}{(2(2h-1))}$$

Residual surplus from the Uniforms blend δ_1 (using un-normalized weights) gives:

$$\begin{aligned}
\mathrm{opt}_{2,1} &= \omega_{\mathrm{pm}} \cdot \frac{(3h^2-1)}{(2(2h-1))} + \int_1^h \omega_{Uz} \cdot \left(\frac{(z+1)}{2}\right) \\
&= \frac{(2h-1)^2}{h^2} \cdot \frac{(3h^2-1)}{(2(2h-1))} + \int_1^h \frac{2(z-1)^2}{z^3} \cdot \left(\frac{(z+1)}{2}\right)\, dz \\
&= \frac{(2h-1)\cdot(3h^2-1)}{2h^2} + \left[z - \ln z + \frac{1}{z} - \frac{1}{2z^2}\right]_1^h \\
&= \frac{6h^3-3h^2-2h+1}{2h^2} + \left[(h-1) - \ln h + \left(\frac{1}{h}-1\right) - \left(\frac{1}{2h^2}-\frac{1}{2}\right)\right] \\
&= \frac{4h^3-3h^2-h^2\cdot\ln h}{h^2} = 4h - 3 - \ln h \qquad (4.23)
\end{aligned}$$

4.5.4.2. Lower Bound on Expected Optimal Residual Surplus from Quadratics. The goal of the calculations in this section is to quantify a lower bound on the expected optimal residual surplus from the Quadratics blend δ_2. Thus, we want: $\mathrm{lb}_{2,2} < \mathrm{opt}_{2,2}$.

We do this in place of calculating $\mathrm{opt}_{2,2}$ which is more complicated technically. Further, the lower bound $\mathrm{lb}_{2,2}$ must be strictly larger than $\mathrm{opt}_{2,1}$ (equation (4.23) just above), in which case we can exhibit a prior independent approximation lower bound from the ratio ${}^{\mathrm{lb}_{2,2}}/_{\mathrm{opt}_{2,1}}$ (see the proof of Theorem 12 below).

This section only includes high-level introduction of the structures that are necessary to calculate $\mathrm{lb}_{2,2}$ and state that it is a lower bound. Therefore we only give here: the

residual surplus curve for the Quadratics (recall – as a function of quantile); the definition for the quantity $\mathrm{lb}_{2,2}$; and Lemma 11 which shows that $\mathrm{lb}_{2,2}$ is an appropriate lower bound for our purposes. Supporting material for this section – including explanations, sub-calculations, and proofs – is provided in Appendix A.2.

First we give the un-ironed residual surplus curve for the Quadratics disributions with CDF of the specific distribution $\overleftarrow{\mathrm{Qud}}^{h'}_{z=1}$ given by $\overleftarrow{\mathrm{Qud}}^{h'}_{1}(x) = 1 - 1/x$ on $x \in [1, h]$, and $\overleftarrow{\mathrm{Qud}}^{h'}_{1}(x) = 1$ for $x \geq h$. Explanation are given in Appendix A.2.1 (page 197). The residual surplus curve is

$$R_{\overleftarrow{\mathrm{Qud}}^{h'}_{1}}(q) = \begin{cases} 0 & \text{for } q \in [0, 1/h] \\ \ln(q \cdot h) & \text{for } q \in [1/h, 1) \\ [\ln h, 1 + \ln h] & \text{for } q = 1 \end{cases} \tag{4.24}$$

Next, recall, we have the definition $\mathrm{opt}_{2,2} = \mathbf{E}_{F \sim \delta_2}[\mathrm{OPT}_F(F)]$, which embeds the weights $\boldsymbol{o}$ (from δ_2). At a high level, the quantity $\mathrm{lb}_{2,2}$ is similarly a calculation of weighted residual surplus, according to weights $\boldsymbol{o}$. With explanation to follow, we formally define:

$$\begin{aligned} \mathrm{lb}_{2,2} &:= o_{\mathrm{pm}} \cdot M_{o_{\mathrm{pm}}}(\overleftarrow{\mathrm{Qud}}^{h'}_{1}) \quad + \left[\int_1^h o_z \cdot \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}^{h'}_{z}) \right] \\ &= \quad 1 \cdot M_{o_{\mathrm{pm}}}(\overleftarrow{\mathrm{Qud}}^{h'}_{1}) \quad + \left[\int_1^h \frac{2}{z} \cdot \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}^{h'}_{z})\ dz \right] \end{aligned} \tag{4.25}$$

Specifically, $\mathrm{lb}_{2,2}$ is calculated using the residual surplus of the 2-lottery (on the corresponding distributions) for all weights $o_z = 2/z \cdot dz$ making up the integral part of the δ_2

blend. (We do this for simplicity even though the 2-lottery is sub-optimal for a range of z-parameters of distributions within this component of the blend.)

The only element of δ_2 for which it does not use the 2-lottery is $\overleftarrow{\text{Qud}}_1^{h'}$ (with weight $o_{\text{pm}} = 1$) where it uses the residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ for a specially constructed mechanism $M_{o_{\text{pm}}}$ (for which we defer presentation to Definition 37 in Appendix A.2.2).

The point is that while the lottery is not necessarily optimal where we use its performance, this relaxed lower bound simplifies our calculation generally to only require calculating expected residual surplus for a single Quadratic distribution, in particular the performance of $M_{o_{\text{pm}}}$ on distribution $\overleftarrow{\text{Qud}}_1^{h'}$. Note, the total quantity $\text{lb}_{2,2}$ is for comparison only – there is no prior independent mechanism that can commit to this behavior (which varies by distribution) and achieve this precise performance.

The expected residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ is stated in Lemma 11, though its proof is also deferred to Appendix A.2.2.

Lemma 11. *The residual surplus of mechanisms $M_{o_{pm}}$ and LOT_2 given 2 agents with values drawn i.i.d. from $\overleftarrow{\text{Qud}}_1^{h'}$ are calculated as*

$$M_{o_{pm}}(\overleftarrow{\text{Qud}}_1^{h'}) = \frac{((2+\ln h)h - (1+\ln h)e)}{h}$$

$$\text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'}) = 1 + \ln h$$

The following lemma states that the quantity $\text{lb}_{2,2}$ is strictly upper bounded by $\text{opt}_{2,2}$ and strictly lower bounded by $\text{opt}_{2,1}$. Its proof is deferred to Appendix A.2.3.

Lemma 12. *Given* $\text{opt}_{2,1}$ *and* $\text{opt}_{2,2}$ *resulting from the finite-weight Quadratics-versus-Uniforms dual blends (along with the rest of the local assumptions of this section), and*

$\mathrm{lb}_{2,2}$ *as defined in equation* (4.25)*. Then we have*

$$\mathrm{opt}_{2,2} > \mathrm{lb}_{2,2} > \mathrm{opt}_{2,1}$$

We now have the outline and justification to calculate the quantity $\mathrm{lb}_{2,2}$ as a meaningful lower bound for $\mathrm{opt}_{2,2}$ towards using the Blends Technique to prove a non-trivial residual surplus gap. The calculation of $\mathrm{lb}_{2,2}$ is a simple adjustment from $\mathrm{opt}_{2,1}$ which runs the 2-lottery everywhere, versus, the quantity $\mathrm{lb}_{2,2}$ is calculated from running the 2-lottery everywhere except with weight $o_{\mathrm{pm}} = 1$ it measures performance $M_{o_{\mathrm{pm}}}(\overleftarrow{\mathrm{Qud}}_1^{h'})$ rather than $\mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}_1^{h'})$. Therefore we have:

$$\begin{aligned}
\mathrm{lb}_{2,2} &= \left[\mathrm{opt}_{2,1}\right] + o_{\mathrm{pm}}\left(M_{o_{\mathrm{pm}}}(\overleftarrow{\mathrm{Qud}}_1^{h'}) - \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}_1^{h'})\right) \\
&= [4h - 3 - \ln h] + 1 \cdot \left(\frac{((2+\ln h)h - (1+\ln h)e)}{h} - (1+\ln h)\right) \\
&= \frac{4h^2 - 2h - h\ln h - e\ln h - e}{h} \qquad (4.26)
\end{aligned}$$

4.5.4.3. The Residual Surplus Gap as Lower Bound. For 2-agent, 1-item prior independent revenue auctions, we have now established a necessary residual surplus gap via blends:

$$\frac{\mathrm{opt}_{2,2}}{\mathrm{opt}_{2,1}} > \frac{\mathrm{lb}_{2,2}}{\mathrm{opt}_{2,1}} = \frac{4h^2 - 2h - h\ln h - e\ln h - e}{4h^2 - 3h - h\ln h}, \qquad \lim_{h\to\infty}\frac{\mathrm{lb}_{2,2}}{\mathrm{opt}_{2,1}} = 1 \qquad (4.27)$$

where the limit calculation is obvious from observing highest-order terms (equivalently, from repeated application of l'Hopital's rule). Evaluation in the limit makes clear that ratio-gaps *from our loose calculations* for any finite h are the result only of differences in lower order terms.

The applicability of the residual surplus gap as a prior independent lower bound is subject to the design problem's parameter $\mathcal{F}$ to describe the adversary's allowable choice set of distributions. However for residual surplus problems, there is precedent to allow the full set of distributions $\mathcal{F}^{\text{all}}$. Putting this together with equation (4.27) and Theorem 10 gives the following theorem to exhibit an approximation lower bound, which parallels Theorem 11 for revenue.

Theorem 12. *Given a single-item, 2-agent, truthful auction setting with a residual surplus objective and with agent values restricted to support* $[1, h]$ *for* $h \geq 8.56$. *For the class of quadratic distributions* $\mathcal{F}^{\text{quad}}$*, the optimal prior independent approximation factor of any (truthful) mechanism is lower bounded as:*

$$\alpha_h^{\mathcal{F}^{\text{quad}}} \geq \frac{\text{opt}_{2,2}}{\text{opt}_{2,1}} > \frac{4h^2 - 2h - h\ln h - e\ln h - e}{4h^2 - 3h - h\ln h} = L_h^{\mathcal{F}^{\text{quad}}}$$

The lower bound $L_h^{\mathcal{F}^{\text{quad}}} \to 1$ *as* $h \to \infty$. *As an example bound: for* $h \in \mathbb{N}$*, the maximum of* $L_h^{\mathcal{F}^{\text{quad}}}$ *is achieved at* $h = 18$ *with* $L_{18}^{\mathcal{F}^{\text{quad}}} \approx 1.00623$.

As a corollary, our bound here holds for $\mathcal{F}^{\text{all}}$.

4.6. Further Example Dual Blends from Optimal Mechanism Design

Theorem 5 in Section 3.4 gave the optimal prior independent 2-agent, single-item, scale-invariant mechanism with a revenue objective for the class regular distributions $\mathcal{F}^{\text{reg}}$. An analogous result for the more-restricted prior independent setting of MHR distributions $\mathcal{F}^{\text{mhr}}$ – showing that the SPA is optimal – was proved by Allouah and Besbes (2018) and this result is given in Theorem 20 in Appendix A.3. This thesis includes

examples of dual blends based on the worst-case distribution of each distribution-class environment, i.e., dual blends (modified) from exclusively distributions which are used by *optimal adversaries* in these specific mechanism design settings.

We move presentation of this section to the appendix due to its technical complexity. Appendix A.3 provides infinite-weight and finite-weight dual blends motivated by the strategy of the optimal adversary of Theorem 20 (for $\mathcal{F}^{\text{mhr}}$). Appendix A.4 gives analogous dual blends descriptions from the adversary of Theorem 5 (for $\mathcal{F}^{\text{reg}}$).

4.7. Comparative Context of Lower Bounds from the Blends Technique

Given a prior independent algorithm design problem, the key inequality to describe the lower bound of the Blends Technique is copied here as:

$$\alpha^{\mathcal{F}} = \min_{A \in \mathcal{A}} \max_{F \in \mathcal{F}} \frac{OPT_F(F)}{A(F)} \geq \frac{\text{opt}_{n,2}}{\text{opt}_{n,1}}$$

In fact, the denominator $\text{opt}_{n,1}$ in the right-hand side may be the expected optimal performance over distributions which represent any decomposition of the correlated distribution $g(\cdot)$ as induced by δ_2. This generalization may indicate a stronger lower bound, *but only abstractly* – there still remain the questions of (1) identifying a better decomposition and (2) calculating its expected optimal performance to upper bound the performance of any algorithm. In particular, not every decomposition is useful. As an extreme example, the decomposition of g into point masses on every fixed profile $\boldsymbol{v}$ will almost certainly have expected optimal performance that is greater than $\text{opt}_{n,2}$ in the numerator, resulting in an innocuous "lower bound" which is smaller than the trivial bound at 1. The point is that good decompositions of g are difficult to find.

Two significant benefits of focusing the search over alternative-decompositions to distributions over i.i.d. product distributions (represented by the blend δ_1 as a mixture over the class $\mathcal{F}^{\text{all}}$) are:

- earlier in this chapter, we have identified dual blends for which the δ_1-side has

$$\text{opt}_{2,1} = \mathbf{E}_{F\sim\delta_1}\left[\text{OPT}_F(F)\right] = \int_{\mathcal{F}^{\text{all}}} \text{OPT}_F(F)\cdot\delta_i(F)\ dF$$

  which is *tractable*, in particular because we only need to solve one or two closed-form evaluations of $\text{OPT}_F(F)$ due to our ability to parameterize these performance functions by z;

- and, as we will be formalized in the next chapter, the (near)-symmetrical structure of our dual blends gives us access to tools in other established fields – we will emphasize connections to the economics topic of information design (Section 5.4); and the mathematics (and algorithms) topic of tensor decomposition (Section 5.5).

CHAPTER 5

Mathematical Structure of Dual Blends

This chapter describes a suite of broad approaches for infinite-weight dual blends solutions that may be useful for identifying good lower bounds for problems of interest, i.e., within a search over dual blends for the one that yields the best lower bound. To do so, this chapter analyzes the fundamental mathematical structure of *general dual blends construction and design*.

First, this chapter identifies and proves existence of classes of solutions based on key properties.One key property is order-statistic separability (Definition 21, Theorem 13) in Section 5.1, i.e., the common function g can be written as *multiplicatively-separable functions per order-statistic of the inputs* (for $n = 2$). Another pair of key properties with interwoven effect are inverse-distributions (Definition 23), Theorem 14) in Section 5.2, i.e., one side of the dual blend by parameterizing over scales of a fixed, base function F, and the other side is then automatically generated by parameterizing over scales of the *inverse-distribution* of F. The example of Section 4.5.1 is a special case of both approaches. Within the theme of general solutions, Section 5.3 presents the straightforward result that dual blends maintain their definition under additive-shifts of their domains. For simplicity, we describe these constructions allowing for infinite-weight blends. Similar methods as used in the example of Section 4.5.1 can modify them to proper probability distributions. Further examples of modification from infinite to finite-total weight are given in Appendix A.3 and Appendix A.4.

Second, for any given prior independent algorithm problem, this chapter considers the theoretical question of optimal lower bounds as may be generated by The Blends Technique. Section 5.4 shows that the search for an optimal dual blend may be separated into a high-level search for an optimal correlated distribution g, for the search for the optimal dual blend given any candidate g reduces to two independent questions of *information design*. Thus, the overall process is *information-design-design*. Section 5.4.3 assesses the comparative information content of the two sides of a dual blend in terms of *Blackwell ordering* (Blackwell, 1953), a perspective which arises naturally within the topic of information design.

Lastly, Section 5.5 provides a quick outline of the connection between blends and *tensors* with dual blends effectively as non-unique continuous-tensor decompositions.

Material in this chapter is largely drawn from **Hartline and Johnsen (2021)**.

5.1. A Class from Order-statistic Separability

This section introduces *order-statistic-separable* functions and subsequently describes a class of dual blends based on these functions. Fix $n = 2$ and our inputs in the cone $v_1 \geq v_2 \geq 0$ in which v_1 represents the first (largest) order statistic and v_2 the second (smaller) order statistic.

Definition 21. *Given $n = 2$. An* order-statistic-separable function *(with domain $\mathcal{V}^2$) is symmetric across the line $v_1 = v_2$ and for inputs subject to $v_1 \geq v_2 \geq 0$, has the form:*

$$g(\boldsymbol{v}) = g_1(v_1) \cdot g_2(v_2)$$

for which both g_1 and g_2 adopt the domain $\mathcal{V}$.

To be clear, the separate functions g_1 and g_2 are not independent factors of g because of the condition $v_1 \geq v_2$. The function g is correlated and is not a product itself. Let $G_1(z) = \int_z^\infty g_1(y)dy$ and $G_2(z) = \int_0^z g_2(y)dy$ be respectively upward-cumulative and downward-cumulative functions. (Intuitively, if $G_1(z)$ is finite, then a "normalized" function $g_1(x)/G_1(z)$ gives the PDF of a conditional probability distribution parameterized by z, on domain $[z, \infty)$; and the same is true for finite $G_2(z)$ on domain $(0, z]$.)

Before stating a formal result in Theorem 13 to construct dual blends, we show that the Quadratics-versus-Uniforms example of Section 4.5.1 exhibits order-statistic separability. The blends' correlated density at every point $\boldsymbol{v} \in \mathbb{R}^2_+$ for $v_1 \geq v_2$ was calculated in equations (4.12) and (4.13) to be $g(\boldsymbol{v}) = 1/v_1^2$. It is easy to verify that $g_1(v_1) = 1/v_1^2$ and $g_2(v_2) = 1$ satisfy Definition 21.

Theorem 13. *Consider non-negative functions $g_1(\cdot)$ and $g_2(\cdot)$ each with domain $(0, \infty)$. For every $z > 0$, let $g_{1,z}$ be g_1 restricted to the domain $[z, \infty)$ and $g_{2,z}$ be g_2 restricted to the domain $(0, z]$.*

*Each δ_i blend is a distribution over the set $\{g_{i,z} \ : \ z > 0\}$. Let $o_{g_1}(z)$ and $\omega_{g_2}(z)$ be functions (*as free parameters which we may design*) to describe weights corresponding respectively to each $g_{1,z}$ and to each $g_{2,z}$.*

First, assume $g_1(\cdot)$ and $g_2(\cdot)$ satisfy the following conditions:

(1) The function $\chi(z) = \frac{g_1(z)}{g_2(z)}$ evaluated in the limit at ∞ is 0, i.e., $\lim_{z\to\infty} \chi(z) = 0$;

(2) the function $\psi(z) = \frac{g_2(z)}{g_1(z)}$ evaluated in the limit at 0 is 0, i.e., $\lim_{z\to 0} \psi(z) = 0$;

(3) $\chi(z)$ must be weakly decreasing, equivalently, $\psi(z)$ must be weakly increasing;

Then the weights functions $o_{g_1}(z) = d\psi(z)$ *and* $\omega_{g_2}(z) = -d\chi(z)$ *give a dual blends solution with:*

$$g(\boldsymbol{v}) = g_1(v_1) \cdot g_2(v_2) \text{ for } \boldsymbol{v} = (v_1, v_2 \leq v_1)$$

If the following condition additionally holds:

(4) the integrals $G_1(z) = \int_z^\infty g_1(y)\, dy$ *and* $G_2(z) = \int_0^z g_2(y)\, dy$ *are positive and finite for all* $z \in (0, \infty)$*;*

then for the same function g*, there exists a dual blends solution (by modification from the original solution) for which all of the* $g_{1,z}$ *and* $g_{2,z}$ *functions are distributions.*

Proof. At a high level, the proof is constructive: it is possible to back out weights functions $o_{g_1}(z)$ and $\omega_{g_2}(z)$. Per the statement, let $\chi(z) = \frac{g_1(z)}{g_2(z)}$. Choose

$$\omega_{g_2}(z) = (-1) \cdot d\chi(z) \tag{5.1}$$

such that the upwards-closed integral over all $g_{2,z}$ (where $g_{2,z}(v_1)$ is positive) gives[1]

$$\begin{aligned} \int_{v_1}^\infty \omega_{g_2}(z) \cdot g_{2,z}(v_1) \cdot g_{2,z}(v_2) &= (g_2(v_1) \cdot g_2(v_2)) \int_{v_1}^\infty (-1) \cdot \ d\chi(z) \\ &= (g_2(v_1) \cdot g_2(v_2)) \left[(-1) \cdot \frac{g_1(z)}{g_2(z)} \right]_{v_1}^\infty = g_1(v_1) \cdot g_2(v_2) \end{aligned} \tag{5.2}$$

[1] Note within the sequence of equation (5.2) that function $g_{2,z}$ is used in the starting evaluation, where its domain informs the integral endpoints; after the endpoints are fixed however, we have $g_{2,z} = g_2$ everywhere. Thus the first step can simplify to the common function g_2 and pull multiplicative constants out of the integral.

where Condition (1) in the theorem statement is sufficient for the final equality. Similarly, let $\psi(z) = \frac{g_2(z)}{g_1(z)}$ and choose

$$o_{g_1}(z) = d\psi(z) \tag{5.3}$$

such that the downward-closed integral over all $g_{1,z}$ (where $g_{1,z}(v_2)$ is positive) gives

$$\begin{aligned} \int_0^{v_2} o_{g_1}(z) \cdot g_{1,z}(v_1) \cdot g_{1,z}(v_2) &= (g_1(v_1) \cdot g_1(v_2)) \int_0^{v_2} d\psi(z) \\ &= (g_1(v_1) \cdot g_1(v_2)) \left[\frac{g_2(z)}{g_1(z)}\right]_0^{v_2} = g_1(v_1) \cdot g_2(v_2) \end{aligned} \tag{5.4}$$

where Condition (2) is sufficient for the final equality.

By implicit assumption throughout this thesis, the weights o_{g_1} and ω_{g_2} (and the function g) must be non-negative everywhere. Observing weights definitions in equations (5.1) and (5.3), Condition (3) is sufficient to meet these high-level assumptions.[2] This completes the proof of the main theorem statement. To prove the distributions-special-case using Condition (4), we show how to use definitions in the theorem statement to modify the ω_{g_2}-side calculations above (and leave the o_{g_1}-side to follow from symmetry, similar to the symmetry above between the two sides). For this setting, we have a modified blends solution. Critically, we have $\tilde{g}_{1,z}(x) = {}^{g_{1,z}(x)}/_{G_1(z)}$ and $\tilde{g}_{2,z}(x) = {}^{g_{2,z}(x)}/_{G_2(z)}$. Condition (4) is sufficient to guarantee that all of the functions $\tilde{g}_{1,z}$ and $\tilde{g}_{2,z}$ are in fact probability distributions. Choose

$$\tilde{\omega}_{g_2}(z) = (-1) \cdot d\chi(z) \cdot (G_2(z))^2 \tag{5.5}$$

[2] Without our global assumptions on weights here, Condition (3) could be relaxed.

which "corrects for the normalization" within each $\tilde{g}_{i,z}$ by re-factoring the weights, such that the same effective calculation as before goes through. I.e., the following upward-closed integral gives

$$(5.6) \quad \int_{v_1}^{\infty} \tilde{\omega}_{g_2}(z) \cdot \tilde{g}_{2,z}(v_1) \cdot \tilde{g}_{2,z}(v_2) =$$

$$\int_{v_1}^{\infty} \tilde{\omega}_{g_2}(z) \cdot \left(\frac{g_{2,z}(v_1)}{G_2(z)}\right)\left(\frac{g_{2,z}(v_2)}{G_2(z)}\right) = (g_2(v_1) \cdot g_2(v_2)) \int_{v_1}^{\infty} (-1) \cdot d\chi(z)$$

$$= (g_2(v_1) \cdot g_2(v_2)) \left[(-1) \cdot \frac{g_1(z)}{g_2(z)}\right]_{v_1}^{\infty} = g_1(v_1) \cdot g_2(v_2)$$

again relying on Conditions (1) and (3). Condition (2) is sufficient for the o_{g_1}-side to work out symmetrically, which uses the modification $\tilde{o}_{g_1}(z) = d\psi(z) \cdot (G_1(z))^2$. □

We illustrate the math of Theorem 13 for our main example of Quadratics-versus-Uniforms in Appendix B.4. We give a second example in Appendix B.5, for which $G_2(z) = \int_0^z g_2(x)\, dx$ evaluates to ∞ and therefore the functions $g_{2,z}$ can not possibly be converted to probability distributions by trying to normalize their total weights.

Through the rest of this section, we discuss a number of intuitive observations regarding the structure of Theorem 13.

The g_i functions as "un-normalized" density functions. The proof of Theorem 13 makes clear how a function like $g_2(v_2) = 1$ is the common function representing *un-normalized density* of every downward-closed uniform distribution $\mathrm{Ud}_{0,z}$. I.e., a process to generate any downward-closed uniform distribution is to start with $g_2(v_2) = 1$ on $[0, \infty)$, restrict it to the domain $[0, z]$, and then divide by the total area under the curve $\int_0^z g_2(y)\, dy = z$. This gives the PDF $\mathrm{ud}_{0,z}(y) = 1/z$.

Similarly, the function $g_1(v_1)$ gives the un-normalized density of every upward-closed quadratic distribution Qud_z. To normalize g_1 to become distribution Qud_z, we divide by the tail area $\int_z^\infty g_1(y)\ dy = 1/z$ and the resulting PDF is exactly $\text{qud}_z(y) = z/y^2$.

Application to distributions requires finite tails. In the statement of Theorem 13, the special case for which we construct $\tilde{g}_{1,z}$ and $\tilde{g}_{2,z}$ to necessarily be probability distributions required additionally Condition (4) which states, "the integrals $G_1(z) = \int_z^\infty g_1(x)dx$ and $G_2(z) = \int_0^z g_2(x)dx$ are positive and finite for all z." This is necessary because, e.g., $\tilde{g}_{1,z}(x) = g_{1,z}(x)/G_1(z)$ would otherwise be not well-defined or 0. See the previous example in Appendix B.5.

The interpretation of Condition (4) is that g_1 must be everywhere "upward-finite" and g_2 must be everywhere "downward-finite."

Definition 22. *Given a non-negative function $g_i(x)$ with domain $(0, \infty)$. The function $g_i(\cdot)$ is* upward-finite *if $\int_z^\infty g_i(x)\ dx$ is finite for every z, and it is* downward-finite *if $\int_0^z g_i(x)\ dx$ is finite for every z.*

We identify a couple consequences of this structure. First, it makes permanent the setting of integral endpoints when calculating density at a fixed input $(v_1, v_2 \leq v_1)$ from each side of the dual blends (recall Figure 4.2 in Section 4.4). Second, it allows us to write any number of simple corollaries to state existence of classes of dual blends that have distributions as elements of the blends, for example:

Corollary 1. *Consider parameterized functions $g^\eta(x) = 1/x^\eta$ for any $\eta \in \mathbb{R}$. Setting $g_1 = g^{\eta_+}$ for $\eta_+ > 1$ and $g_2 = g^{\eta_-}$ for $\eta_- < 1$ will meet all conditions (1) through (4) of*

Theorem 13. Thus, there is a dual blends solution for which the elements are distributions from any $g^{\eta+}$ and $g^{\eta-}$.

Algebraic consequences of the integral endpoints in the construction. The assigned integral endpoints of the dual blends calculations – as resulting from Definition 22 – are critical to making the algebra work out. Specifically, each side observably employs an integral endpoint to "correct" the $g_i(v_{j\neq i})$ term which originally appears inside the integral, as a "constant" given the integration per dz.

E.g., the equation in line (5.6) at the end of the proof of Theorem 13 makes this clear: both $g_i(v_i)$ and $g_i(v_j)$ terms get pulled out. After this step, the evaluation of the integral *given its endpoints* is needed to both construct $g_i(v_i)$ and cancel $g_i(v_j)$ – there are no other algebraic tools available to construct the function g. In fact, we can't change $g_i(v_i)$ and it passes intact as a factor of g. Evaluation of the integral must replace the $g_i(v_j)$ term with a $g_j(v_j)$ term which is the second factor of g. Then the weights terms are designed to get the overall integrand correct so that the anti-derivative function evaluates the "extreme" endpoint (at 0 or ∞) to 0 and the other endpoint at v_j to convert an original $g_i(v_j)$ term to $g_i(v_i)$ as needed within the order-statistic-separable function g.

With this algebraic set up in mind, it should now be clear why we should not expect a direct extension of The Blends Technique (or general dual blends solutions) for $n \geq 3$. For example, consider trying to construct dual blends for the function

$$g(v_1, v_2, v_3) = g_1(v_1) \cdot g_2(v_2) \cdot g_3(v_3)$$

by direct analogy to the $n = 2$ case. The problem for generalization is that the design for $n = 2$ gives each side exactly two "degrees of freedom" to set g_1 and g_2. To attempt the same design for $n = 3$, let i, j be distinct elements of the set $\{1, 2, 3\}$. Each side of the (supposed) dual blend must be symmetric from a functional starting point:

$$\begin{aligned} g_i(v_1) \cdot g_i(v_2) \cdot g_i(v_3) \int_0^b (\cdot)\, dz = g(\boldsymbol{v}) &= g_1(v_1) \cdot g_2(v_2) \cdot g_3(v_3) \\ &= g_j(v_1) \cdot g_j(v_2) \cdot g_j(v_3) \int_a^\infty (\cdot)\, dz \end{aligned}$$

but there is no way to evaluate the integrals – no matter what their integrands are or what their endpoints are – to combine with each of $\prod_k g_i(v_k)$ and $\prod_k g_j(v_k)$ to get $g_1(v_1) \cdot g_2(v_2) \cdot g_3(v_3)$. The only solution is $g_i = g_j$.

The $G_i(\cdot)$ functions as continuous scalars. By inspection of equation (5.6), the (finite) functions $G_i(x)$ can in fact be set to any function that is strictly positive (or even more generally, non-zero) *as long as they are still offset by $G_i(\cdot)$ terms in the weights functions.* Therefore, the $g_i(\cdot)$ functions only need to be subject to the restrictions on $\chi(\cdot)$ and $\psi(\cdot)$ (and be non-zero and finite) for a blend $g(\boldsymbol{v}) = g_1(v_1) \cdot g_2(v_2)$ to be induced.[3]

Corollary 2. *Consider non-negative functions $g_1(\cdot)$ and $g_2(\cdot)$ which each have domain $(0, \infty)$. Let $g_{1,z}$ be g_1 restricted to the domain $[z, \infty)$ and $g_{2,z}$ be g_2 restricted to the domain $(0, z]$. Assume there exists a dual blends solution*

$$g(\boldsymbol{v}) = g_1(v_1) \cdot g_2(v_2) \text{ for } \boldsymbol{v} = (v_1, v_2 \leq v_1)$$

[3] The example applying Theorem 13 to the infinite-weight Quadratics-versus-Uniforms dual blends in Appendix B.4 exhibits this type of "forward construction" of its intended function g.

using weights $o_{g_1}(z)$ *and* $\omega_{g_2}(z)$.

Then for any finite, positive functions $G_1,\ G_2$, *the weights* $\tilde{o}_{g_1}(z) = o_{g_1}(z) \cdot G_1(z)$ *and* $\tilde{\omega}_{g_2}(z) = \omega_{g_2}(z) \cdot G_2(z)$ *applied to functions* $g_{i,z}/G_i(z)$ *describes the same dual blends solution from* $g_1,\ g_2$.

There exists a comparison here to nth-order tensors. The specific observation here is that this Corollary 2 is analogous to dividing a symmetric tensor's element-vector $\boldsymbol{a}_z$ by a factor κ_z and multiplying its scalar l_z by κ_z^n. In the same way that we can multiply-and-divide by the respective G_i with no effect on g, these multiplicative factors cancel and have no effect on nth-order tensor $T = l_z \cdot (\boldsymbol{a}_z \otimes \ldots \otimes \boldsymbol{a}_z)$. Exploring the connection between blends and tensors is the focus of Section 5.5.

5.2. A Class from Inverse-distributions

It is a remarkable feature of the infinite-weight Quadratics-versus-Uniforms dual blends *that both sides use the exact same weights parameters per* z, namely $o_{Qz} = \omega_{Uz} = 2/z \cdot dz$. This structure is not an anomaly – it is indicative of a class of infinite-weight dual blends solutions which we formalize in Theorem 14 (and give the key definitions and proof below).

The critical structure is the multiplicative inverse '$1/z$.' Its importance is highlighted from two perspectives: inverse-distributions and arbitrary distribution rescaling. Notably, Quadratics and Uniforms are inverse-distributions to each other, which we see directly from $\mathrm{Qud}_1(x) = 1 - 1/x$ on $[1, \infty)$ for which the inverse-distribution CDF is $1 - \mathrm{Qud}_1(1/x) = 1 - (1 - 1/1/x) = x = \mathrm{Ud}_{0,1}(x)$ on $[0, 1]$. Additionally, the Quadratics blend assigns weights to all rescalings of Qud_1 and the Uniform blend assigns weights to all rescalings of $\mathrm{Ud}_{0,1}$.

Fundamentally, Theorem 14 shows that there is a duality between distribution values and distribution scales, as can be observed in equation (5.7).

Theorem 14. *Given distribution* F*, define members* F_y *of its parameterized class of all possible rescalings* $y > 0$*, and its inverse-distribution* i-F *by*

$$\mathrm{F}_z(x) = \mathrm{F}(x/z) = 1 - \text{i-F}(z/x) = 1 - \text{i-F}_x(z) \tag{5.7}$$

For $n = 2$*,* F_z *and* i-F_z *give classes that are dual blends using weights* $o_z = \omega_z = 1/z$*, i.e., they describe a common function* g *at every* $\boldsymbol{v} = (v_1, v_2 \leq v_1)$*:*

$$\int_0^\infty \frac{1}{z} \cdot \mathrm{f}_z(v_1) \cdot \mathrm{f}_z(v_2)\ dz = g(\boldsymbol{v}) = \int_0^\infty \frac{1}{z} \cdot \text{i-f}_z(v_1) \cdot \text{i-f}_z(v_2)\ dz \tag{5.8}$$

Definition 23. *Given a distribution* F *with domain* $[a, b]$ *(or domain* $[a, \infty)$*), i.e.,* $F(a) = 0$ *and* $F(b) = 1$*. The* inverse-distribution *of* F *is defined by the CDF function* $\text{i-}F(x) = 1 - F(1/x)$ *on domain* $[1/b, 1/a]$ *(respectively domain* $(0, 1/a]$*). The PDF of the inverse-distribution is denoted* i-f*. (Fact: as an operation,* distribution inversion *is its own inverse, i.e., it respects the identity* i-(i-F) $= F$*.)*

Recall Fact 15, copied here from Section 3.2:

Fact 15. *Given a distribution* $F_{z=1}$ *with default scaling parameter* $z = 1$ *and with domain* $[a, b]$ *(or domain* $[a, \infty)$*). The distribution* F_1 *can be arbitrarily re-scaled for* $z \in (0, \infty)$ *to* $F_z(x) = F_1(x/z)$ *with domain* $[z \cdot a, z \cdot b]$ *(respectively domain* $[z \cdot a, \infty)$*).*

These concepts come together in Theorem 14 which proves that an infinite-weight blends solution always exists effectively from fixing symmetric weights $o_z = \omega_z = 1/z \cdot dz$ and

then choosing the g_1 and g_2 as inverse-distributions of each other. In Theorem 13 by comparison, g_1 and g_2 were (relatively) free parameters to be chosen first, for which weights could then be identified to complete a dual blends solution. We give a concise proof of Theorem 14 from the key ideas of this section (inverse-distributions and rescaling):

Proof. Given distribution F and its inverse-distribution i-F, the rescaled CDFs and PDFs are:

$$\mathrm{F}_z(x) = \mathrm{F}(x/z) \qquad \text{i-F}_z(x) = \text{i-F}(x/z) = 1 - \mathrm{F}(z/x)$$

$$\mathrm{f}_z(x) = \frac{1}{z} \cdot \mathrm{f}(x/z) \qquad \text{i-f}_z(x) = \frac{z}{x^2} \cdot \mathrm{f}(z/x)$$

Starting from the right-hand side of equation (5.8), the following sequence completes the proof:

$$\int_0^\infty \frac{1}{z} \cdot \text{i-f}_z(v_1) \cdot \text{i-f}(v_2)\ dz = \int_0^\infty \left[\frac{1}{z} \cdot dz\right] \cdot \left(\frac{z}{v_1^2} \cdot \mathrm{f}(z/v_1)\right) \cdot \left(\frac{z}{v_2^2} \cdot \mathrm{f}(z/v_2)\right)$$

(here we perform calculus-change-of-variables using $z = \zeta(y) = \frac{v_1 \cdot v_2}{y}$; recall that part of the substitution is $dz = \zeta'(y) \cdot dy$, and integral endpoints get mapped by $\zeta^{-1}(\cdot)$)

$$= \int_\infty^0 \left[\frac{1}{\frac{v_1 \cdot v_2}{y}} \cdot \left(\frac{-v_1 \cdot v_2}{y^2} \cdot dy\right)\right] \cdot \left(\frac{\frac{v_1 \cdot v_2}{y}}{v_1^2} \cdot \mathrm{f}(v_2/y)\right) \cdot \left(\frac{\frac{v_1 \cdot v_2}{y}}{v_2^2} \cdot \mathrm{f}(v_1/y)\right)$$

$$= \int_0^\infty \left[\frac{1}{y} \cdot dy\right] \cdot \left(\frac{1}{y} \cdot \mathrm{f}(v_2/y)\right) \cdot \left(\frac{1}{y} \cdot \mathrm{f}(v_1/y)\right) = \int_0^\infty \frac{1}{y} \cdot \mathrm{f}_y(v_2) \cdot \mathrm{f}_y(v_1)\ dy \qquad \square$$

An interesting property of (infinite-weight) dual blends from Theorem 14 that emerges from the proof is: *we don't have to solve for a closed-form expression* for the function g in order to prove equality of its dual descriptions. As a consequence, the process of

obtaining lower bounds from dual blends may reduce to computation of expectations over optimal performances $\mathrm{OPT}_F(F)$.

5.3. Additive Shifts of Dual Blends

This section identifies a straightforward technique to generalize the solution of any (finite-weight) dual blends solution when the input space is unbounded (positive) reals $\mathbb{R}_+$. The technique simply *additively shifts the density.* We state this formally with Proposition 2 and give a natural example of the technique following from Quadratics-versus-Uniforms as used in the previous two sections.

Proposition 2. *Given a dual blends solutions described $\delta_1^n = g = \delta_2^n$ for (symmetric) correlated distribution g and blends $\delta_1^n,\ \delta_2^n$ respectively described according to weights o_z on distributions F_{o_z} and weights ω_z on distributions F_{ω_z} with input domain a subset of $[a, \infty)$ for $a \geq 0$. For every $\epsilon \geq -a$, the following is also a dual blends solution with input domain a subset of $[0, \infty)$:*

$$\delta_{\epsilon,1}^n = g^\epsilon = \delta_{\epsilon,2}^n \tag{5.9}$$

for correlated distribution g^ϵ and blends $\delta_{\epsilon,1}^n,\ \delta_{\epsilon,2}^n$ respectively described according to weights $\breve{o}_z = o_{z-\epsilon}$ on distributions $F_{\breve{o}_z,\epsilon}$ defined by PDF $f_{\breve{o}_z,\epsilon}(x) = f_{o_z}(x-\epsilon)$ and weights $\breve{\omega}_z = \omega_{z-\epsilon}$ on distributions $F_{\breve{\omega}_z,\epsilon}$ defined by PDF $f_{\breve{\omega}_z,\epsilon}(x) = f_{\omega_z}(x-\epsilon)$.

We now outline the following positive example. Consider the 2-input Quadratics-versus-Uniforms solution of Section 4.5. We substitute $t+1$ into the maximum value h and then shift the coordinates Quadratics-versus-Uniforms solution using parameter

$\epsilon = -1$ via Proposition 2. This puts the (still finite-weight) solution into the value space $[0,t]^2$ for $t > 0$. With inputs shifted down by 1, the class of Quadratics $\overleftarrow{\text{Qud}}_z^{h'}$ becomes exactly a class of Shifted-Quadratics generally described by $\overleftarrow{\text{Sqd}}_{z,-1}^{t'}$ with weights $\breve{o}_z$. The original class of Uniforms generally used domain lower bound 1, which after the downward shift is more naturally realized at 0, and generally they become the class of $\text{Ud}_{0,z}$ distributions with weights $\breve{\omega}_z$. The weights are assigned to their respective appropriately-shifted distributions. Obviously the total weight remains the same and we can normalize later. To re-summarize the total weight here, we have $1 + \int_0^t \frac{2}{z+1} dz = 1 + 2\ln(t+1)$.

The weights on the upward-closed Shifted-Quadratics blend ($\delta_{\epsilon,2}^2$) are as follows:

- point mass of weight $\breve{o}_{\text{pm}} = 1$ on (truncated) distribution $\overleftarrow{\text{Sqd}}_{0,-1}^{t'}$;
- weights $\breve{o}_{Qz} = \frac{2}{z+1} dz$ on all upward-closed (truncated) distributions $\overleftarrow{\text{Sqd}}_{z,-1}^{t'}$ for $z \in [0,t]$.

The weights on the downward-closed Uniforms blend ($\delta_{\epsilon,1}^2$) are as follows:

- point mass of weight $\breve{\omega}_{\text{pm}} = \frac{(2t+1)^2}{(t+1)^2}$ on a distribution defined by $\overleftarrow{\text{Ud}}_{0,2t+1}^{t'}$;
- weights $\breve{\omega}_{Uz} = \frac{2z^2}{(z+1)^3} dz$ on all downward-closed distributions $\text{Ud}_{0,z}$ for $z \in [0,t]$.

We have the following types of density that will match: 2-D $dv_1 dv_2$, 0-D pure point mass at input (t,t), and 1-D dv_2 at inputs (t, v_2). Because these calculations are simple additive shifts of previous calculations, we believe it is sufficient to write down the solution from

one side only and without showing all steps (cf., calculations on page 101).

$$(5.10) \qquad g_{2D}(\boldsymbol{v}) = \text{result of } \overleftarrow{\text{Sqd}}^{t'}_{z,-1} \text{ blend} = \frac{1}{(v_1+1)^2}$$

$$(5.11) \qquad g_{0D}(\boldsymbol{v}) = \text{result of truncation in } \overleftarrow{\text{Sqd}}^{t'}_{z,-1} \text{ blend} = 1$$

$$(5.12) \qquad g_{1D}(\boldsymbol{v}) = \text{result of } \textbf{exactly one} \text{ truncation in } \overleftarrow{\text{Sqd}}^{t'}_{z,-1} \text{ blend} = \frac{1}{t+1}$$

5.4. Dual Blends Design is Information-Design-Design

This section connects *theoretical optimization* of the Blends Technique to the economics topic of information design, specifically as a procedure of *information-design-design* (IDD). For a given prior independent problem (parameterized by class of distributions $\mathcal{F}$), the main idea is to separate into modular problems the search for the optimal dual blend (which yields the largest lower bound of any dual blend). (1) An "outer" problem identifies an optimal correlated distribution $g^* \in \mathcal{G} = \{\delta^n \mid \delta \in \Delta(\mathcal{F})\}$. The outer problem searches over: (2) for any exogenous $g \in \mathcal{G}$, an "inner" problem identifies two blends that induce g – respectively from $\mathcal{F}$ and $\mathcal{F}^{\text{all}}$ – to maximally separate the ratio of optimal performances given each blend (cf., the Blends Technique).

Effectively, the distributions that compose each blend acts as *signals* to which each corresponding optimal algorithm OPT_F may respond. If signals can be designed as outputs of a mapping from underlying inputs as *fixed states*, then such signal-response games are called *information design*. (We can design signals in this way for our problems, see Lemma 13 based on Bayes Law.) We exhibit the separation of problems first and defer the presentation of information design.

Describing the sequence of inequalities below, the first line starts with a prior independent problem and its right-hand side optimizes over lower bounds from the Blends Technique. This step removes the algorithm design problem of the min-player and gives a new problem (which is constrained with respect to the original, possibly with loss).

Next where an adversary optimizes both steps of a sup − sup program, we rearrange these two successive choices to: (a) optimize the correlated distribution g which represents both (flattened) sides of the dual blends simultaneously; and then (b) optimize over sets of blends which induce g to maximize the numerator (using $\mathcal{F}$) and minimize the denominator (using $\mathcal{F}^{\text{all}}$).[4] The final line is a reorganization using independence of numerator and denominator which now each comprise a *sub-problem of the Blends Technique*.

$$\begin{aligned}
\alpha^{\mathcal{F}} = \min_{A\in\mathcal{A}} \max_{F\in\mathcal{F}} \frac{\text{OPT}_F(F)}{A(F)} &\geq \sup_{\delta_2\in\Delta(\mathcal{F})} \left[\sup_{\delta_1\in\{\delta \mid \delta\in\Delta(\mathcal{F}^{\text{all}}) \text{ and } \delta^n=g=\delta_2^n\}} \left[\frac{\mathbf{E}_{F\sim\delta_2}\left[\text{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta_1}\left[\text{OPT}_F(F)\right]} \right] \right] \\
&= \sup_{g\in\mathcal{G}} \left[\sup_{\substack{\delta_2\in\{\delta \mid \delta\in\Delta(\mathcal{F}) \text{ and } \delta^n=g\} \\ \delta_1\in\{\delta \mid \delta\in\Delta(\mathcal{F}^{\text{all}}) \text{ and } \delta^n=g\}}} \left[\frac{\mathbf{E}_{F\sim\delta_2}\left[\text{OPT}_F(F)\right]}{\mathbf{E}_{F\sim\delta_1}\left[\text{OPT}_F(F)\right]} \right] \right] \\
&= \sup_{g\in\mathcal{G}} \left[\frac{\sup_{\delta_2\in\{\delta \mid \delta\in\Delta(\mathcal{F}) \text{ and } \delta^n=g\}} \left(\mathbf{E}_{F\sim\delta_2}\left[\text{OPT}_F(F)\right]\right)}{\inf_{\delta_1\in\{\delta \mid \delta\in\Delta(\mathcal{F}^{\text{all}}) \text{ and } \delta^n=g\}} \left(\mathbf{E}_{F\sim\delta_1}\left[\text{OPT}_F(F)\right]\right)} \right] \qquad (5.13)
\end{aligned}$$

Definition 24. *The optimization problem of equation* (5.13) *is* Information-Design-Design. *Within the brackets, we refer to the optimizations respectively as the* Numerator *and* Denominator Games.

[4] This optimization may be non-trivial – for a single exogenous g, there are generally multiple candidate blends which induce g. Intuitively, this is true because the set $\{\delta \mid \delta^n = g\}$ is closed under convex combination. As illustration, first consider two distinct dual blends examples $g^a = \delta_1^n = \delta_2^n$ and $g^b = \delta_3^n = \delta_4^n$ as may be generated per the large class of Theorem 14. Then $g^{ab} = {}^{g^a}\!/_2 + {}^{g^b}\!/_2$ has four blends solutions: ${}^{\delta_i^n}\!/_2 + {}^{\delta_j^n}\!/_2$ for all $i \in \{1,2\}$, $j \in \{3,4\}$. (We count here the four combinations of "corner" descriptions of g^{ab}. We ignore that, e.g., the ${}^{\delta_i^n}\!/_2$ term may mix over ${}^{\delta_1^n}\!/_2$ and ${}^{\delta_2^n}\!/_2$ – an optimization never needs this mix by linearity of expectation.) To generalize, the convex set $\{\delta \mid \delta^n = g\}$ is generally a Hilbert space, e.g., if g is a continuous mixture over a continuum of dual blends.

Thus, when g is fixed exogenously by an outer maximization, there is a reduction to diametrically-opposite questions of constrained information design (Proposition 3 next). Constraining the design is the key step – informally information design is a *signalling game* and we require that signals be distributions $F \in \mathcal{F}$ (or $F \in \mathcal{F}^{\text{all}}$; which each induce an i.i.d. product distribution). Thus, (a) the marginal distribution over signals is a blend, and (b) an optimal algorithm can be run in response to a given signal $\hat{F}$ (cf., the use of distributions-as-signals in $\text{opt}_{n,i} = \mathbf{E}_{F\sim\delta_i}[\text{OPT}_F(F)]$).

Proposition 3. *Consider the prior independent design problem (Definition 2) given a class of distributions $\mathcal{F}$, a class of algorithms $\mathcal{A}$, and n inputs. Optimization of the Blends Technique approach to prior independent lower bounds is described by:*

$$\alpha^{\mathcal{F}} \geq \sup_{g\in\mathcal{G}} \left[\frac{\sup_{\delta_2\in\{\delta \mid \delta\in\Delta(\mathcal{F}) \text{ and } \delta^n=g\}} \left(\mathbf{E}_{F\sim\delta_2}[\text{OPT}_F(F)]\right)}{\inf_{\delta_1\in\{\delta \mid \delta\in\Delta(\mathcal{F}^{all}) \text{ and } \delta^n=g\}} \left(\mathbf{E}_{F\sim\delta_1}[\text{OPT}_F(F)]\right)} \right]$$

Further, its Numerator Game and its Denominator Game can be independently instantiated as problems of constrained information design.

This section additionally evaluates dual blends from the perspective of *Blackwell (partial) ordering*, which compares two designs of signalling strategies equivalently in terms of both a strong measure of their information content, and a strong measure of their usefulness for arbitrary optimization objectives. In our case, signalling strategies correspond to blends, and the IDD Numerator Game searches for the *best* signals using $\mathcal{F}$ while its Denominator Game searches for the *worst* signals using $\mathcal{F}^{\text{all}}$. We will show that our finite-weight Quadratics-versus-Uniforms dual blend (of Section 4.5.1) is an example for which there is no relationship according to Blackwell ordering (Corollary 3).

An outline for this section is: Section 5.4.1 gives an introduction to information design. Section 5.4.2 gives an intuitive explanation of the reduction of the Numerator and Denominator Games within equation (5.13) to information design; it includes Lemma 13 which shows that the crux of the reduction is a straightforward application of Bayes Law. Section 5.4.3 introduces Blackwell ordering and observes that our dual blends in the Quadratics-versus-Uniforms example of Section 4.5.1 do not have a Blackwell ordering.

Related Work for Information Design. The canonical model of information design with a single sender and single receiver was introduced by Rayo and Segal (2009) and Kamenica and Gentzkow (2011). A few points of context with this literature are as follows. In our setting the allowable posterior distributions are constrained. The early work of Glazer and Rubinstein (2004) – in which the sender can only present certain kinds of evidence – can be viewed as a posterior-constrained setting of information design. In our mechanism design applications, the receiver is a seller and faces a number of potential buyers. Bergemann et al. (2015) previously studied information design in such a scenario with only one buyer, with the goal of characterizing the feasible outcomes that a regulator (the sender) can obtain in terms of the tradeoff between revenue and residual surplus. While it is not directly related to the methods of this thesis, there is a literature starting with Dughmi et al. (2019) that shows that some problems of information design are computationally tractable. See Bergemann and Morris (2019) for a more complete survey of the breadth of literature on information design.

5.4.1. Introduction to Information Design

From economics, information design is a game between two players – a Sender and a Receiver – who have unaligned objective functions. There is an unknown state of the world θ from a set of states Θ. Realized state $\hat{\theta}$ is Bayesian and is drawn from a prior $\bar{\pi}$ that is common knowledge.

The Sender observes $\hat{\theta}$ and sends a signal s from the signal space $\mathcal{S}$. This is implemented by: up front, the Sender commits to a signalling strategy $\sigma : \Theta \to \Delta(\mathcal{S})$ that maps states to distributions over signals, with $\sigma \in \Sigma$ the space of (possibly restricted) strategies.[5] Strategies σ implicitly lead to *information structures* because of the existence of the prior $\bar{\pi}$ – information structures describe the ex ante correlated distribution over paired state-and-signal. The takeaway is that information structures represent strategic design by the Sender to convert the prior $\bar{\pi}$ into a structured system of posteriors (conditional for each s) for specific use by the Receiver.

Definition 25. *An* information structure $\mathcal{I} : \Theta \times \mathcal{S} \to [0, 1]$ *is a correlated probability distribution over state and signal.*

We make two critical observations: an information structure is induced from a given prior $\bar{\pi}$ over state and a signalling strategy σ; and in turn, an information structure induces a posterior distribution over states (conditioned on a realized output signal $\hat{s}$).

After the Sender commits to σ, the Sender observes $\hat{\theta}$ and sends a signal $\hat{s}$ to the Receiver as randomly drawn from $\sigma(\hat{\theta})$. The Receiver sees $\hat{s}$ and chooses an action ω from

[5] In the context of a fixed prior, there is a bijection between signalling strategies and information structures as we define them. The economics literature may use the term "information structure" for our signalling strategies.

its action space Ω. Finally, each player has utility functions respectively as $S : \Theta \times \Omega \to \mathbb{R}$ and $R : \Theta \times \Omega \to \mathbb{R}$.

It is standard to assume that the Receiver plays a best-response action: given the context of knowing $\bar{\pi}$ and σ, it uses $\hat{s}$ to get a posterior distribution over state and then simply optimizes against the posterior. This leaves the Sender's construction of σ as the unique strategic consideration, called *information design*. The utility functions S and R typically embed a degree of objectives being orthogonal – or adversarial. If the utility functions are aligned (which will be true for one of our cases), information design is trivial unless the Sender's signal space is restricted to not be able to fully reveal the realized state.

5.4.2. Reduction of Blends Technique Sub-problems to Information Design

This section explains the reduction from the Numerator and Denominator Game sub-problems within the reorganized Blends Technique in equation (5.13), to constrained information design. I.e., this section proves Proposition 3.

The key element of the reduction is to carefully constrain the Sender's space of signalling strategies to *blends-revelation signalling strategies*, defined as follows. Effectively, we implement a *Revelation Principle for information design* which states that the Sender's signal may as well be a correct posterior over state space – *which we further require to be a symmetric product distribution* – so that the Receiver only needs to best respond to the posterior-signal.

Definition 26. *Within information design, we define a* blends-revelation signalling strategy *(BRSS) to be a signalling strategy* σ^{br} *in which:*

- *signals are distributions;*
- *the marginal distribution over signal-distributions resulting from* σ^{br} *as a blend induces* g*;*
- *the Receiver's posterier given any signal-distribution* F *is in fact* F^n*.*

Fixing a prior independent design problem (PIP), the instantiation of its Numerator and Denominator Games (equation (5.13), Definition 24) as information design problems is now from the following reduction. The reductions are the same with the exception of the Sender's objective function (described in the last point).

- $\Theta = \mathcal{V}^n$; state space is the input space of the prior independent algorithm;
- $\bar{\pi} = g$; the prior over states is equal to the correlated distribution g (for any g as fixed by the outer program in equation (5.13));
- $\mathcal{S} = \mathcal{F}$; signal space is restricted to be the PIP's allowable class of distributions $\mathcal{F}$;
- $\Sigma = \{\sigma^{\text{br}} \mid \sigma^{\text{br}} \text{ is a BRSS}\}$; **the key element of the reduction:** signalling strategy space Σ is the set of blends-revelation signalling strategies of Definition 26; note that at least one such signalling strategy must exist because $\bar{\pi} = g$ was constructed up front from a blend and can in fact be implemented (see Lemma 13 below);
- $\Omega = \mathcal{A}$; the Receiver's action space is naturally the algorithm space $\mathcal{A}$ from the PIP, and
- $R(\boldsymbol{v}, A) = A(\boldsymbol{v})$; the Receiver's utility is equal to the objective of the algorithm designer in the PIP;

- the Sender's utility S is either perfectly aligned with the Receiver's objective (Numerator Game) or perfectly adversarial to it (Denominator Game):
  - Numerator Game: $S(\boldsymbol{v}, A) = A(\boldsymbol{v}) = R(\boldsymbol{v}, A)$;
  - Denominator Game: $S(\boldsymbol{v}, A) = -A(\boldsymbol{v}) = -R(\boldsymbol{v}, A)$;

  but note how in both cases, because the Sender's signal is always a *correct posterior* (per Definition 26), it must be that the marginal distribution over posteriors has exactly the structure of a blend.

The challenge for the Sender is how to produce and optimize strategies that meet Definition 26. From application of Bayes Law, it turns out that the Sender is able to design a signalling strategy in advance that, ex post observing state, simulates a random latent variable distribution to provide as signal to the Receiver, in a way that the Receiver will use the distribution-signal as if it is correct.

The Sender chooses a signalling strategy σ using the following outline. Similar to the PIP's adversary, the Sender optimizes over $\delta \in \Delta(\mathcal{F})$ such that $\delta^n = g$. It uses observed state $\boldsymbol{v}$ to do Bayesian updating on the distribution δ (over distributions $F \in \mathcal{F}$) and then randomly draws $\hat{F}$ from the posterior to send as the signal.

As the final key piece which we state next and prove, this Sender's choice of randomized σ yields a "correct" posterior (for every realized $\hat{F}$ as signal), thereby satisfying the last requirement of Definition 26. The proof makes clear that lem:iolemma is an application of Bayes Law.

Lemma 13. *Given state space Θ equal to algorithm input space $\mathcal{V}^n$ and prior $\bar{\pi}$ over stats as a blend $g = \delta \in \Delta(\mathcal{F})$. Given realized $\boldsymbol{v} \sim g$, let the Sender's signalling strategy*

draw distribution-signal $\hat{F}$ from the posterior of δ (given $\boldsymbol{v}$). Then the Receiver's induced posterior over state is $\hat{F}^n$ and the distribution over induced posteriors is δ.

Proof. Given correlated g, it is equivalent to assume that inputs were drawn from a two-step procedure: first draw $F \sim \delta \in \Delta(\mathcal{F})$ and then draw n inputs i.i.d. from F. Consider from this perspective that density in the original function g is further broken down for each input to reflect density of its latent variable F, i.e., consider correlated density g^+ over $\mathcal{V}^n \times \mathcal{F}$. (Note, we can recover the function g by fixing each $\boldsymbol{v}$ and integrating over $\mathcal{F}$.)

Given the definition of the Designer's signalling strategy, the correlated distribution over (state, signal) is exactly equal to g^+. The reason is that given g^+ the final correlated description over $\mathcal{V}^n \times \mathcal{F}$, for every $(\boldsymbol{v}, F)$, Bayes Law states that the following quantities are equal: $\Pr[\boldsymbol{v} \mid F] \cdot \Pr[F] = \Pr[F \mid \boldsymbol{v}] \cdot \Pr[\boldsymbol{v}]$. Thus our problem's equivalence is:

- The left-hand side of our Bayes-Law-equation gives an unfalsifiable description of how inputs were generated (see the first sentence of proof).
- The right-hand side describes how (state, signal) pairs are generated within the Information Design game: first there is a randomly revealed state; and second, per the pre-committed signalling strategy, there is a random mapping from state to signals using a posterior distribution (from updating δ given $\boldsymbol{v}$).

We claim that the Receiver's posterior over state from a realized signal-distribution $\hat{F}$ is $\hat{F}$. Consider the correlated distribution g^+ as an abstraction of a matrix: values $\boldsymbol{v}$ are rows and distributions F are columns. Given a true state of a realized $\hat{\boldsymbol{v}}$-row the Sender

produces a signal $\hat{F}$ from the posterior over distributions. Observing $\hat{F}$, the Receiver's posterior over values is exactly the column corresponding to $\hat{F}$, which is exactly $\hat{F}$ itself.

The final point is: receiving a specific signal $\hat{F}$, the Receiver's posterior over $\mathcal{V}^n$ is obtained from conditioning g^+ given $\hat{F}$ and then the Receiver's posterior is exactly $\hat{F}^n$ as desired. Note for this last point, the Receiver has access to g^+ because the Designer's choice of δ is known. □

5.4.3. Assessment of Blends' Blackwell Ordering

This section considers if dual blends have the property that one side of the dual blend is "strictly more informative" than the other side (per Blackwell (1953)), as part of understanding what is driving prior independent lower bounds that follow from dual blends. It introduces Blackwell (partial) ordering and shows that the two sides of a dual blend do not generally have a Blackwell ordering: for our main example of (Section 4.5.1), there is no informational relationship.

Blackwell Ordering: Global Usefulness and Garbles. Blackwell (1953) proposed a framework of *partial ordering* between the distributions over signals – and their respective systems of posteriors – of two information structures $\mathcal{I}_1$ and $\mathcal{I}_2$ to reflect a notion of information-dominance called Blackwell ordering. It is based on two strong properties being equivalent. According to one description, there is an ordered relationship between the information structures based on *global usefulness*, i.e., if one information structure $\mathcal{I}_1$ is preferred to $\mathcal{I}_2$ *for every possible* utility function (by an optimizer using a random signal). The equivalent descriptive property is called a *garbling* and it applies when the signals of

$\mathcal{I}_2$ can themselves be interpreted as obfuscating mixtures over the signal probabilities of $\mathcal{I}_1$ while maintaining exactly the same induced prior over state.

The next two definitions are presented within the context of the problem statement of information design: we are given an information design problem with fixed prior $\bar{\pi}$ over state space Θ, and two feasible information structures $\mathcal{I}_1 : \Theta \times \mathcal{S}_1 \to [0,1]$ and $\mathcal{I}_2 : \Theta \times \mathcal{S}_2 \to [0,1]$ for respective signal spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, and a class of allowable Receiver-algorithms $\mathcal{A}$.

Fact 22. *Given $\bar{\pi}$ and an information structure $\mathcal{I}_2$, the signalling strategy $\sigma_{\mathcal{I}_2}$ (which induces $\mathcal{I}_2$ starting from $\bar{\pi}$) can necessarily be reverse-engineered.*

Definition 27 (Blackwell, 1953)**.** *Let $\sigma_{\mathcal{I}_1}$, $\sigma_{\mathcal{I}_2}$ be the signalling strategies induced by $\bar{\pi}$ and the respective information structures (per Fact 22). Let A_1^*, A_2^* be optimal algorithms given respective information structures. Information structure $\mathcal{I}_1$ has greater* global usefulness *than $\mathcal{I}_2$ if <u>for every</u> (Borel-measurable) Receiver's utility function R, expected <u>optimal</u> utility is weakly greater when signals are drawn given $\sigma_{\mathcal{I}_1}$ compared to signals drawn given $\sigma_{\mathcal{I}_2}$ (i.e., if $\mathcal{I}_1$ is* preferred *to $\mathcal{I}_2$):*

$$\mathbf{E}_{\hat{\theta}\sim\bar{\pi},\ \hat{s}\sim\sigma_{\mathcal{I}_1}(\hat{\theta})}\left[R(\hat{\theta}, A_1^*(\hat{s}))\right] \geq \mathbf{E}_{\hat{\theta}\sim\bar{\pi},\ \hat{s}\sim\sigma_{\mathcal{I}_2}(\hat{\theta})}\left[R(\hat{\theta}, A_2^*(\hat{s}))\right] \tag{5.14}$$

Definition 28 (Blackwell, 1953)**.** *The information structure $\mathcal{I}_2$ is a* garble *of $\mathcal{I}_1$ if there exists a mapping $\eta : \mathcal{S}_1 \times \mathcal{S}_2 \to [0,1]$ such that $\int_{\mathcal{S}_2} \eta(s_i, s_j)\ ds_j = 1$ for all $s_i \in \mathcal{S}_1$; and for every $s_j \in \mathcal{S}_2$ and every state $\theta \in \Theta$ we have*

$$\mathcal{I}_2(\theta, s_j) = \int_{\mathcal{S}_1} \eta(s_i, s_j) \cdot \mathcal{I}_1(\theta, s_i)\ ds_i \tag{5.15}$$

i.e., we identify $\mathcal{I}_1$ *as being* more informative (per garbling order) *in comparison to* $\mathcal{I}_2$.[6]

We now give Blackwell's classic theorem which states equivalence of Definition 27 and Definition 28.

Theorem 15 (Blackwell, 1953)**.** *An information structure* $\mathcal{I}_1$ *has greater global usefulness than* $\mathcal{I}_2$ if and only if $\mathcal{I}_2$ *is a garble of* $\mathcal{I}_1$.

As contrapositive (in one direction), if there exist two distinct utility functions R_1 *and* R_2 *such that* $\mathcal{I}_1$ *is* strictly *preferred to* $\mathcal{I}_2$ *for* R_1 *but* $\mathcal{I}_2$ *is* strictly *preferred to* $\mathcal{I}_1$ *for* R_2*, then there can not be a garbling order relationship between the information structures.*

The "strictly different" preferences of distinct information structures given distinct utility functions is necessary to apply the contrapositive. Motivated by Theorem 15, the common ordering from global usefulness and garbles is called *Blackwell ordering*.

Dual Blends Do Not Generally Have Blackwell Ordering. Dual blends are represented by simple information structures when their common correlated distribution g (over inputs in $\mathcal{V}^n$) is interpreted as the prior $\bar{\pi}$ over state and when signalling strategies are designed as in Section 5.4.2 using Lemma 13.

When message space $\mathcal{S}$ is set equal to distribution class $\mathcal{F}$ (as support for elements of the blends), it is clear that blends properly define an information structure (of Definition 25) as a distribution over paired message-and-state. For example, describe a blend by $\mathcal{I}_2(F_z, \boldsymbol{v}) = o_z \cdot \prod_k f_z(v_k)$. These are "simple" because the blend already describes posteriors which are independent given a signal F_z as: n i.i.d. draws from F_z.

[6] An intuitive explanation of garbles is: each signal $s_j \in \mathcal{S}_2$ can be interpreted as a distribution over the signals of $s_i \in \mathcal{S}_1$. In response to each signal $s_j \in \mathcal{S}_2$, we respond with a single optimal algorithm for the posterior given s_j, which may not be optimal given each signal $s_i \in \mathcal{S}_1$ in the implicit distribution; hence, $\mathcal{I}_2$ has "garbled" $\mathcal{I}_1$.

Describing blends as information structures aligns exactly with the calculations of lower bounds in the Blends Technique. Recall for blends $\delta_1^n,\ \delta_2^n$, expected "optimal performance" within an algorithm setting is given by $\text{opt}_{n,i} = \mathbf{E}_{F\sim\delta_i^n,\ \boldsymbol{v}\sim F}\left[\text{OPT}_F(\boldsymbol{v})\right]$, and then a lower bound is given by ${}^{\text{opt}_{n,2}}/_{\text{opt}_{n,1}}$. Each blend is a possible information structure to represent the same underlying correlated distribution over states, and for each blend the quantity $\text{opt}_{n,i}$ is the optimal performance in expectation over state, as the algorithm knows the realized distribution-signal.

We are ready to state by counterexample that dual blends do not generally have Blackwell ordering, using the Quadratics-versus-Uniforms example of Section 4.5.1 and Definition 25). Recall Theorem 11 for revenue auctions used an adversary choosing the benchmark from the Uniforms side of the dual blends but Theorem 12 for residual surplus used the Quadratics side, and that the settings have distinct objective functions.

Theorem 11 and Theorem 12 each show strict performance gaps for their respective settings. Thus, these results give an immediate example meeting the condition of the contrapositive statement in Theorem 15, because a "Receiver" strictly prefers distinct information structures depending on the auction objective.

Corollary 3. *Finite-weight Quadratics-versus-Uniforms dual blends are an example for which there is no relationship according to Blackwell ordering.*

5.5. Dual Blends are Non-unique Tensor Decompositions

In this section we examine how the Blends Technique is an application of the mathematical technique of *tensor decomposition* which is of great importance in a number

of computer science fields, e.g., signal processing, machine learning, and computer vision. Usually in these fields, studies of tensor decomposition are interested in *uniqueness of decomposition – and especially uniqueness of approximate, constrained optimization – rather than multiple, exact solutions which can not be distinguished.* This puts our interest in tensor decomposition via blends at a cross-purpose to previous literature, and for an intuitive reason: existing studies are mostly interested in systems of "positive" communication (e.g., electronic signals and robust coding alphabets which desire that unique, correct messages get through) whereas we are interested in systems of *obfuscation*: our adversary uses dual blends to strictly impair interpretation of the true signal structure.

Specifically, our interest ignores the standard question of minimal description (as measured by rank) and focuses on identification versus non-identification of symmetric, continuous tensors. The goal of this section is to highlight an obvious connection between the Blends Technique and the existing, deep field of tensor analysis, whose tools may be utilized to further the study of our dual blends. We do not claim new results within tensor decomposition.

Related Work is included next but otherwise this simple, exploratory section is deferred to Appendix A.7.

Related Work for Tensor Decomposition. There exist an abundance of techniques for the special case of *matrix decomposition.* A survey is given by Plassman (2005). Whereas our blends describe decompositions of non-negative values in the specific case of mechanism design (and other algorithm domains), Donoho and Stodden (2004) give sufficient conditions for a geometric interpretation of *non-negative matrix factorization* in terms of unique, "simplicial cones." Relating to the *continuous-weights* description of our

blends, Townsend and Trefethen (2015) partially or fully extend the matrix-decomposition techniques of singular value decomposition (SVD), QR, LU and Cholesky factorizations to "cmatrices" (continuous in both dimensions).

Sidiropoulos et al. (2017) give a thorough primer to tensor decomposition, in which they reference Anandkumar et al. (2014) who study tensor decomposition from topic models, i.e., from a perspective of tensor scalars as *latent variables* over parameters of underlying distributions (e.g., Gaussians). Cf., our blends are effectively distributions over distribution-signals as latent variables. Whereas we may be able to exhibit dual blends with one side a mixture over Gaussians in one real dimension (likely from Theorem 14), there is a line of work in machine learning regarding mixtures of high-dimensional, axis-aligned Gaussians from Dasgupta (1999), Feldman et al. (2006), Hsu and Kakade (2013), and Bhaskara et al. (2014). Feldman et al. (2008) prominently give an algorithm to learn mixtures of product distributions, which compare to our blends defined from symmetric product distributions.

CHAPTER 6

Benchmark Design and Prior Independent Optimization

This chapter expands this thesis' study of Bayesian-robustness beyond optimal algorithm description (of a mechanism; Chapter 3) and lower bounds on prior independent algorithm design (per the Blends Technique; Chapter 4 and Chapter 5) to make a connection to the worst-case-over-inputs setting. *Prior free* is an alternative information setting in which to measure algorithm robustness – the n inputs are selected arbitrarily from a known support and an algorithm's performance is measured in worst-case over inputs as its ratio against a *benchmark performance function* which is evaluated for each input.

Critically within the prior free setting, choice of the benchmark function itself is effectively a free parameter. Consider a comparison between the prior independent setting which is measured in worst-case over distributions in expectation given the distribution, versus the prior free setting which is measured in worst-case over inputs pointwise. Despite the higher-precision measurement of prior free approximation (per input rather than per distribution), it is in fact not possible to determine which setting provides a stronger guarantee of robustness without reference to the prior free benchmark.

A formal approach to benchmark design will seek out the existence of guiding principles to inform and measure the design of the benchmark, motivated by the heuristic narrative that intuitively we must guard both against designing benchmarks that are too small or too big. A common heuristic benchmark across both mechanism design and online

algorithms is to use the *offline optimal* performance, i.e., the optimal performance if the inputs were known in advance.

Hartline and Roughgarden (2014) propose a property of benchmarks called *normalization* (paramterized by a reference class of distributions $\mathcal{F}$; see Definition 31), which ties prior free benchmark design back to Bayesian settings in a way that benchmarks are justifiably not too small, specifically, *it guarantees that if an algorithm A approximates a normalized prior free benchmark, then A is Bayesian-robust because it must have at most the same prior independent approximation (for distributions in $\mathcal{F}$).*

Hartline (2020) proposes a second property *resolution* for measurement of normalized benchmarks to describe intuitively: how well a benchmark may distinguish between good and bad algorithms. (Benchmarks may inherently induce lower bounds on prior free approximation from various lower-bounding techniques, and smaller lower bounds allow for magnified distinction between the approximation factors of candidate algorithms.) Of particular relevance, resolution is a measurement that can be subjected to *optimization.*

As the main result of this chapter – for a general algorithms setting and not restricted to e.g. mechanism design – we show an equivalence between the questions of prior independent design and benchmark design for the prior free setting. To do this, we will abstract the definition of Hartline (2020) for resolution[1] to a definition of *resolution measures* (Definition 32) and choose for our analysis a novel measure (*best-response resolution*; Definition 34).

[1] We give the definition for *normalization-symmetric resolution* as the original resolution measure of Hartline (2020) in Definition 35, which induces a lower bound from the intuition: it measures in worst-case over distributions how big a benchmark is – versus – how big it minimally needs to be to simply meet the definition of normalization.

Technically, consider a *benchmark design problem* (BDP; Definition 33) as the following min − max operation: for use in a prior free setting, choose the "optimal" benchmark as the normalized one with smallest resolution (minimization); with the resolution measure set as an exogenous parameter (worst-case maximization). We will show that with best-response resolution as the choice for this parameter, the solution to describe the optimal benchmark is "the same" as the solution for the problem of finding the optimal prior independent algorithm and approximation factor – for every input vector $\boldsymbol{v}$, the optimal benchmark function is set by the minimal scaling-up of the performance function of the optimal prior independent algorithm on $\boldsymbol{v}$ to be big enough to meet the definition of normalization (Theorem 16).

Subsequently however, we will use the canonical problem of *expert learning* within the broad field of online algorithms to identify a challenge to the approach to benchmark design which uses our best-response resolution measure. The objective within expert learning is to minimize average *regret* over n sequential online inputs. Specifically, the optimal algorithm for k-expert learning is the intuitive *follow-the-leader* algorithm (FTL; Definition 36; Theorem 18). Thus as a corollary from the equivalence in Theorem 16 (described above), the optimal benchmark using best-response resolution is the adjusted-up performance of FTL,[2] for which FTL is naturally the optimal prior free algorithm. By contrast, FTL is known to be a linear approximation in worst-case to the canonical prior free *best-in-hindsight* benchmark (BIH) for expert learning.[3]

[2] We use "adjust up" rather than "scale up" for expert learning because its approximation-objective regret is measured additively rather than multiplicatively; otherwise we largely ignore this distinction, our results go through for either approach with correct technical modification.

[3] We summarize some intuition because our benchmark design and algorithm behavior are alien to standard approaches of analysis for expert learning. Our approach here is prior free (i.e., worst-case) and the objective remains regret-minimization. The benchmark function within the calculation of regret

This chapter is laid out as follows. Section 6.1 gives preliminaries for the prior free setting, in particular, definitions for the prior free algorithm design problem, benchmark properties of normalization and resolution, and our benchmark design problems. Section 6.2 proves that for our best-response resolution measure, the optimal normalized benchmark is set as the adjusted-up performance of the corresponding optimal prior independent algorithm and thus the optimal benchmark's resolution is equal to optimal prior independent approximation. Lastly, Section 6.3 gives a sufficient introduction to the problem of expert learning for which it then shows that FTL is prior independent optimal and thus informs the optimal prior free benchmark (from best-response resolution); further, it discusses the conflicting assessment of FTL's prior free performance, as measured by competing prior free benchmarks for expert learning.

Material in this chapter is largely drawn from **Hartline et al. (2020a)**.

6.1. Prior Free Design and Benchmark Design

This section provides the preliminaries for the prior free information setting which measures approximation pointwise in worst-case over input space against a benchmark function. Benchmarks are a free parameter of prior free design and thus we also introduce preliminaries for an embedded question of benchmark design. A main goal of this chapter generally is to propose first principles of *optimal benchmark design*.

---

is evaluated pointwise over input histories. The idea is, our designed benchmark function (adjusted-up-FTL) will depend itself on the pointwise performance of the FTL algorithm. Thus, on sequences of inputs for which it is well-known that FTL does poorly when the benchmark is set according to BIH, our alternative prior free benchmark function sets a target performance for these sequences that "does-not-punish" the FTL algorithm, effectively by self-referential design given the context of characterization of our optimal benchmark.

Section 6.1.1 introduces the prior free algorithm design problem. In Section 6.1.2, we introduce and motivate *economic properties of benchmarks* in order to justify a framework in which to consider optimal benchmark design. One of the key properties that we identify is a benchmark's ability to guarantee a lower bound on the approximation of any algorithm to the benchmark. Abstractly we call this property *resolution*, however there exist multiple, *formally-definable*, computational techniques to guarantee a lower bound and we call these techniques *resolution measures*. Section 6.1.2 end with the general definition of a benchmark design problem (Definition 33). Section 6.1.3 and Section 6.1.4 define benchmark design problems paramaterized respectively by two such resolution measures.

6.1.1. Robust Algorithm Design Problem Statements

The goal of this section is to define the prior free algorithm design problem. For convenience and clarity of comparison, we will restate the algorithm design problems for Bayesian and prior independent settings from Section 2.1 as a lead up.

Definition 1. *The* Bayesian optimal algorithm design problem *is given by a distribution F and class of algorithms $\mathcal{A}$; and solves for the algorithm* OPT_F *with the maximum expected performance:*

$$(\text{OPT}_F) \qquad \text{OPT}_F = \text{argmax}_{A \in \mathcal{A}} A(F).$$

Settings of incomplete information like prior independent and prior free measure the performance of algorithms by their *worst case approximation* to the benchmark, on any input of the benchmark's domain (thereby guaranteeing an approximation against every

input and justifying the description of the analysis as "robust"). A function $B : \Omega \to \mathbb{R}_+$ is the *benchmark* function. The *abstract* definition for approximation to a benchmark is:

Definition 29. *For a benchmark function $B : \Omega \to \mathbb{R}_+$ (with domain Ω equal to either an input space $\mathcal{V}^n$, or a class of distributions $\mathcal{F}$), an algorithm A is an* α-approximation *to B if*

$$(\alpha) \qquad \alpha \geq \frac{B(\omega)}{A(\omega)} \quad \forall\ \omega \in \Omega$$

Where a benchmark function is defined on inputs, i.e. for $\Omega = \mathcal{V}^n$, we use notation that is analogous to notation for algorithm performance in Section 2.1: let $B(\boldsymbol{v})$ be an expected performance of a benchmark (which may rely on internal randomization[4]) and let $B(F) = \mathbf{E}_{\boldsymbol{v}\sim F}[B(\boldsymbol{v})]$. (We only use a benchmark function defined on distributions, i.e. $\Omega = \mathcal{F}$, for the prior independent problem for which we reduce: $\mathrm{OPT}_F(F) = B(F)$.)

Definition 2. *The* prior independent algorithm design problem *is given by a class of algorithms $\mathcal{A}$ and a class of distributions $\mathcal{F}$; and searches for the algorithm that minimizes its worst-case approximation:*

$$(\alpha^{\mathcal{F}}) \qquad \alpha^{\mathcal{F}} = \min_{A\in\mathcal{A}} \left[\max_{F\in\mathcal{F}} \frac{\mathrm{OPT}_F(F)}{A(F)}\right] = \min_{A\in\mathcal{A}} \left[\alpha_A^{\mathcal{F}}\right]$$

[4] Algorithms are evaluated in expectation over possible internal randomization. Benchmark functions may be defined by similar processes, in particular when they depend on sub-functions with internal randomization, e.g., *they may depend on an algorithm's performance.*

where the value of the program α^F *is the* optimal prior independent approximation factor *for class* $\mathcal{F}$ *and class* $\mathcal{A}$ *(which we leave implicit). The bracketed term is the* prior independent approximation guarantee *of a fixed algorithm* A *given* $\mathcal{F}$ *and is denoted by* $\alpha_A^{\mathcal{F}}$ *(as shown on the right).*

Definition 30. *The* prior free algorithm design problem *is given by a class of algorithms* $\mathcal{A}$ *and a benchmark* B*; and searches for the* argmin *of the* $\min - \max$ *program*

$$(\alpha^B) \qquad \alpha^B = \min_{A \in \mathcal{A}} \left[\max_{\boldsymbol{v} \in \mathcal{V}} \frac{B(\boldsymbol{v})}{A(\boldsymbol{v})} \right] = \min_{A \in \mathcal{A}} \left[\alpha_A^B \right]$$

where the value of the program α^B *is the* optimal prior free approximation factor *for benchmark* B *and class* $\mathcal{A}$ *(which we leave implicit). The bracketed term is the* tight prior free approximation guarantee *of a fixed algorithm* A *given* B *and is denoted by* α_A^B *(as shown on the right).*

6.1.2. Benchmark Properties: Normalization and Resolution

Benchmark design should follow from principles which (a) represent economic justification, and (b) efficacy to distinguish good algorithms from bad. This section starts with definitions for *normalization* and *resolution* as such principles. Normalization provides a guarantee that benchmarks are not too small, in a way that directly connects the prior independent and prior free settings. Resolution measures benchmarks in a way that they can be optimized to be small (constrained by normalization). From these simple motivations, we will subsequently formulate two benchmark design problems.

Prior Free Related Work. Specifically for mechanism design, in Appendix A.5 we describe some commonly used benchmarks from the historical literature and also give some of the most pertinent existing results from related work.

Normalization. The first part of this section relates the prior independent and prior free settings – which are both searching for algorithms with robust performance guarantees – in a way that motivates a framework for benchmark design. In theory, prior free guarantees can provide more robustness than prior independent guarantees because the approximation is required to hold pointwise on all inputs rather than in expectation according to each distribution. Intuitively, an algorithm with a prior free guarantee can not "cheat" and ignore an input and make it up elsewhere given the weighting of a distribution.

This observation only holds "in theory" because, for the prior free setting, the degree to which its guarantee is meaningful depends on the choice of benchmark. Practically, both the prior free benchmark B and the prior independent comparison class $\mathcal{F}$ are free parameters in separate problems (and they are independent parameters), and there is no immediate connection between the prior independent and prior free settings. We desire a property that guarantees: prior free approximation necessarily implies prior independent approximation.

With this in mind, Hartline and Roughgarden (2008) recommend restricting attention to benchmarks that satisfy the following *normalization* property which requires that the

benchmark be lower-bounded by the performances of optimal algorithms.[5] By intention, the consequence of this definition is Proposition 4 below.

Definition 31 (Hartline and Roughgarden, 2008)**.** *A benchmark B is* normalized *for a class of distributions $\mathcal{F}$ and class of algorithms $\mathcal{A}$ if for every distribution in the class the expected benchmark is at least the optimal expected performance, i.e.,*

$$B(F) \geq \mathrm{OPT}_F(F), \quad \forall F \in \mathcal{F}.$$

Denote the class of normalized benchmarks for $\mathcal{F}$ by $\mathcal{NB}^{\mathcal{F}}$.

A benefit of the structure of Definition 31 is that OPT_F takes into account the constraints on the algorithm, i.e., that $\mathrm{OPT}_F(F) \in \mathcal{A}$. As previously suggested, the normalization property implies a strong guarantee. An algorithm A that is an α-approximation to a normalized benchmark $B \in \mathcal{NB}^{\mathcal{F}}$ guarantees a prior independent α-approximation for the class $\mathcal{F}$ (for "approximation" of Definition 29). Thus we have $\alpha_A^{\mathcal{F}} \leq \alpha_A^B$. Formally:

Proposition 4 (Hartline and Roughgarden, 2008)**.** *If algorithm A is a prior free α-approximation of a benchmark B normalized to distributions $\mathcal{F}$, then its tight prior independent approximation for distributions in $\mathcal{F}$ is at most α, i.e., $\alpha_A^{\mathcal{F}} \leq \alpha_A^B \leq \alpha$.*

Note, given B, the first inequality holds in particular from the optimal prior free algorithm A^* and optimal approximation factor α^B, i.e., $\alpha_{A^*}^{\mathcal{F}} \leq \alpha_{A^*}^B = \alpha^B$.

[5] Notably, the normalization property does not measure optimal performance pointwise per input (which is how prior free algorithms are measured against benchmarks). Rather, Hartline and Roughgarden (2008) recommend measuring this optimal performance in expectation with respect to any distribution in a class of distributions $\mathcal{F}$ – theoretically the same $\mathcal{F}$ as used for the base prior independent design problem for which approximation-by-extension is desired.

Resolution. The ultimate goal of algorithm design is a principled method for choosing one algorithm over another. Normalized benchmarks of Definition 31 embed an economic justification that they are not too small, however, arbitrarily large benchmarks will qualify as normalized. No algorithm will have good approximation against a prior free benchmark that is too large, and thus good algorithms will not necessarily be separated from bad ones. Generically, the ability of a benchmark to discriminate between good and bad algorithms is its resolution, by which we will define something specific. A *resolution property* of a benchmark is a formal measurement of an approximation lower bound *that is guaranteed to hold from a proven, specified description of a lower-bounding technique* on any algorithm in comparison to B. Consistent with our purposes, we give a definition specifically within the context of a *normalized* benchmark B:

Definition 32 (Hartline, 2020)**.** *Given a class of distributions $\mathcal{F}$ and the space of normalized benchmark functions $\mathcal{NB}^{\mathcal{F}}$, a resolution measure $\mu^B : \mathcal{NB}^{\mathcal{F}} \to \mathbb{R}_+$ is a function on benchmarks $B \in \mathcal{NB}^{\mathcal{F}}$ to formally measure a* guaranteed, functional lower bound *on the approximation of B by any algorithm.*

Thus, resolution is a necessary performance gap over a domain Ω. The lower bound on approximation is not necessarily tight. Regardless, given a defined measure of resolution, we can optimize over normalized benchmarks: small measures of resolution represent good, "high-quality-resolution." (Our Chapter 1 described philosophical and technical justification for resolution.)

Benchmark Design. Having identified and motivated properties of normalization and (abstract) resolution, the objective and formulation of *benchmark design* becomes straightforward.

Definition 33 (Hartline, 2020)**.** *A* benchmark design problem (BDP) *is given by a class of distributions $\mathcal{F}$, a class of algorithms $\mathcal{A}$, and a resolution measure μ^B, and searches for the normalized benchmark with minimum resolution as*

$$(\gamma_\mu^{\mathcal{F}}) \qquad \gamma_\mu^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\mu^B\right]$$

(with an embedded assumption that the resolution measure μ^B is a lower bound on algorithm approximation to B, from Definition 32).

The next two sections each define a resolution measure to parameterize a corresponding benchmark design problem.

6.1.3. Resolution and Design from Best Response Algorithms

One way to quantify the ability of a benchmark to discriminate is by considering *the approximation factor of the optimal algorithm as best response to the benchmark.* For any benchmark B, we refer to ρ^B as the *best-response resolution* of benchmark B, and it has a special structure. Because it simply measures optimal approximation, we have $\rho^B = \alpha^B$ from the prior free algorithm design problem (Definition 30). The following program formally optimizes ρ^B subject to normalization (by minimizing over $B \in \mathcal{NB}^{\mathcal{F}}$).

Definition 34. *The* best-response resolution benchmark design problem (brBDP) *is given by a class of distributions $\mathcal{F}$, a class of algorithms $\mathcal{A}$, input space $\mathcal{V}^n$, and* best-response resolution measure $\rho^B = \alpha^B$*, and searches for the* argmin *of the* $\min - \min - \max$ *program*

$$(\gamma_\rho^{\mathcal{F}}) \qquad \gamma_\rho^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\rho^B\right] = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\alpha^B\right] = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\min_{A \in \mathcal{A}} \max_{\boldsymbol{v} \in \mathcal{V}^n} \frac{B(\boldsymbol{v})}{A(\boldsymbol{v})}\right]$$

Lemma 14 (Hartline, 2020)**.** *The best-response resolution measure of benchmark B gives a lower bound on the approximation of any algorithm to B.*

Proof. By setting resolution measure $\rho^B = \alpha^B$, the brBDP inherits the (tight) lower bound on approximation of any algorithm that is implicitly present in the value α^B of the prior free design problem given B. □

Remarkably, Theorem 16 next states that this best-response resolution benchmark design problem (brBDP) is equivalent to the prior independent algorithm design problem (Definition 2). We defer proof and discussion of Theorem 16 to be the focus of Section 6.2.

Theorem 16. *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, the best-response resolution benchmark design problem (brBDP) is equivalent to the prior independent algorithm design problem.*

Specifically, consider optimal benchmark B^* with optimal resolution $\gamma_\rho^{\mathcal{F}}$ and optimal prior independent algorithm A^* with approximation $\alpha^{\mathcal{F}}$. Then $B^*(\cdot) = \alpha^{\mathcal{F}} \cdot A^*(\cdot)$ and $\gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}}$.

6.1.4. Resolution and Design from Symmetry to Normalization

Before getting into the details of this section, we note: the approach of this section is inherently connected to our Blends Technique for lower bounds on prior independent approximation. An explicit study of this connection is deferred to Appendix A.6 along with some supporting material for this section.

Philosophically, the brBDP of the previous section embeds "accommodative" design by choosing a benchmark within the context of existence of a good algorithm for approximation – benchmark and algorithm are optimized together. This can be observed from the successive min − min optimizations in Definition 34. By contrast, a disinterested benchmark designer might do optimization in a vacuum, and leave algorithm designers to optimize responses against the resulting benchmark in a second, independent step.

Hartline (2020) proposed a definition for *resolution measurement* of normalized benchmarks to describe intuitively: how big they are – versus – how big they need to be to simply meet the definition of normalization. I.e., when benchmarks are normalized, a second way to quantify the ability of a benchmark to discriminate is by considering *in worst-case over distributions the expectation of the benchmark over each distribution* $\mathcal{F}$*, versus optimal algorithm performance* $\mathrm{OPT}_F(F)$ *which appears in a constraint of the normalization property.*

Thus, we intuitively consider the perspective of a myopic benchmark designer, who defines and minimizes resolution as a standalone endeavor. Given our existing framework for benchmark design, if normalization is all that is necessary to provide economic justification that a benchmark is not too small, then loose "normalization constraints" represent potentially excessive performance targets and the "worst" excesses should be reduced if

possible. As we explain next, loose constraints induce lower bounds and thus inform a resolution measure. The algorithm designer is still implicitly benefited when we minimize resolution up front (as a lower bound on approximation) but this is not the priority of our benchmark design here. For this second measure of resolution, we formulate a second benchmark optimization problem.

For any benchmark B, we refer to σ^B as the *normalization-symmetric resolution* of benchmark B. Like best-response resolution of the previous section, σ^B also has special structure. The technical definition is $\sigma^B = \max_{F \in \mathcal{NB}^{\mathcal{F}}} \frac{B(F)}{\mathrm{OPT}_F(F)}$ (given formally in Definition 35 below). Because σ^B is calculated in worst-case over *distributions* – and in particular, with direct symmetry to the normalization constraints – when the resolution-optimizing program is written as a linear program, its system of constraints exhibits an elegant symmetry (see Linear Program 6 in Appendix A.6.2).

Similar to the brBDP, the following program formally optimizes σ^B subject to normalization. In contrast to the brBDP, σ^B is calculated in worst-case over distributions (the brBDP was worst-case pointwise over inputs $\boldsymbol{v}$).

Definition 35 (Hartline, 2020)**.**[6] *The* normalization-symmetric resolution benchmark design problem (nsBDP) *is given by a class of distributions $\mathcal{F}$, a class of algorithms $\mathcal{A}$, and resolution measure σ^B; and searches for the* argmin *of the* $\min - \max$ *program*

$$(\gamma_\sigma^{\mathcal{F}}) \qquad \gamma_\sigma^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\sigma^B\right] = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\max_{F \in \mathcal{F}} \frac{B(F)}{\mathrm{OPT}_F(F)}\right]$$

[6] This benchmark design problem depends on $\mathcal{V}$ implicitly through $\mathcal{F}$.

The following lemma is sufficient to prove that normalization-symmetric resolution is a lower bound on approximation (in order to meet Definition 32 for a resolution measure; a version of this lemma was first given by Goldberg et al. (2006)).

Lemma 15 (Goldberg et al., 2006)**.** *For any benchmark B, the class of distributions $\mathcal{F}^{\text{all}}$ (given input space $\mathcal{V}^n$), and class of algorithms $\mathcal{A}$ which induce OPT_F, the optimal prior free approximation α^B is at least*

$$(\breve{\alpha}^B) \qquad \breve{\alpha}^B = \max_{F \in \mathcal{F}^{\text{all}}} \frac{B(F)}{\text{OPT}_F(F)}$$

Proof. Let A^* be the prior free optimal algorithm for benchmark B with optimal approximation α^B, and let F^* be the distribution that optimizes the lower bound program $(\breve{\alpha}^B)$ with distribution support $\mathcal{V}(F^*)$. With the second inequality here following from Fact 27 in Appendix B.2, we have

$$\alpha^B = \max_{\boldsymbol{v} \in \mathcal{V}} \frac{B(\boldsymbol{v})}{A^*(\boldsymbol{v})} \geq \max_{\boldsymbol{v} \in \mathcal{V}(F^*)} \frac{B(\boldsymbol{v})}{A^*(\boldsymbol{v})} \geq \frac{B(F^*)}{A^*(F^*)} \geq \frac{B(F^*)}{\text{OPT}(F^*)} = \breve{\alpha}^B \qquad \square$$

Remarkably for the nsBDP, its value $\gamma_\sigma^{\mathcal{F}}$ necessarily lower bounds $\gamma_\rho^{\mathcal{F}}$ as the value of the brBDP in Definition 34. We state this next in Proposition 5 (with its proof given as part of proving a stronger statement: Proposition 7 in Appendix A.6.3). Further, we are able to identify a setting for which the inequality is strict (Theorem 17 below).

Proposition 5. *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, the value of the nsBDP $\gamma_\sigma^{\mathcal{F}}$ lower bounds the value of the brBDP $\gamma_\rho^{\mathcal{F}}$ and the approximation*

factor of the prior independent optimal algorithm for $\mathcal{F}$:

$$\gamma_\sigma^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}} \tag{6.1}$$

Proposition 5 illustrates an interesting divergence between resolution measures. On one hand, $\gamma_\rho^{\mathcal{F}}$ and its corresponding optimal benchmark B_{rs}^* from the brBDP are specifically designed to minimize the optimal algorithm's approximation of B_{rs}^* in response. On the other hand, the nsBDP achieves weakly smaller (absolute) resolution measure $\gamma_\sigma^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$. From this we observe: despite the optimal benchmark B_{ns}^* of the nsBDP having smaller resolution than B_{rs}^*, its design precludes algorithms with strictly smaller approximation (otherwise B_{ns}^* and the optimal algorithm against it as arguments to the brBDP would strictly improve the objective and replace B_{rs}^*).

We defer broader study of normalization-symmetric resolution to Appendix A.6. There we explain how the nsBDP is an interesting design implicated by Theorem 25 (Chen et al., 2014) and describe a natural extension of the nsBDP. We write the nsBDP as a linear program and show its connection to the Blends Technique. Finally, as mentioned above, we give the proof of Proposition 7 in Appendix A.6.3.

6.2. Best-response Resolution is Equivalent to Prior Independent Approximation and Dominates Normalization-symmetric Resolution

Theorem 16 was first stated in Section 6.1.3 where we pointed out that this statement holds for an arbitrary algorithm setting (with the assumption that the same class of distributions $\mathcal{F}$ parameterizes both problems). It should be noted that this result exists

independent of the fact that it may not be understood how to generally solve these problems for arbitrary algorithm settings.

Theorem 16. *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, the best-response resolution benchmark design problem (brBDP) is equivalent to the prior independent algorithm design problem.*

Specifically, consider optimal benchmark B^* with optimal resolution $\gamma_\rho^{\mathcal{F}}$ and optimal prior independent algorithm A^* with approximation $\alpha^{\mathcal{F}}$. Then $B^*(\cdot) = \alpha^{\mathcal{F}} \cdot A^*(\cdot)$ and $\gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}}$.

Theorem 16 follows from the next two results:

- $(\alpha^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}})$ from Corollary 4 which follows from Proposition 4 (page 157), and
- $(\gamma_\rho^{\mathcal{F}} \leq \alpha^{\mathcal{F}})$ from Lemma 16.

Corollary 4 (Hartline, 2020)**.** *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, the prior independent optimal ratio $\alpha^{\mathcal{F}}$ is at most the optimal benchmark ratio $\gamma_\rho^{\mathcal{F}}$, i.e., $\alpha^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$.*

Proof. From Definition 34 for the brBDP with value $\gamma_\rho^{\mathcal{F}}$, the optimal algorithm A^* for the optimal benchmark achieves a prior free approximation factor of $\gamma_\rho^{\mathcal{F}}$. By Proposition 4 and the normalization of the optimal benchmark, A^* also achieves a prior independent approximation factor of $\gamma_\rho^{\mathcal{F}}$. Approximation of the *optimal* prior independent algorithm is at least as good, i.e., $\alpha^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$. □

Lemma 16. *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, the optimal benchmark ratio $\gamma_\rho^{\mathcal{F}}$ is at most the prior independent optimal ratio $\alpha^{\mathcal{F}}$, i.e., $\gamma_\rho^{\mathcal{F}} \leq \alpha^{\mathcal{F}}$.*

Proof. Consider the prior independent optimal algorithm A^* (i.e., the argmin of the prior independent $\min - \max$ program) with optimal approximation $\alpha^{\mathcal{F}}$. Define the benchmark

$$B^*(\boldsymbol{v}) = \alpha^{\mathcal{F}} \cdot A^*(\boldsymbol{v}) \tag{6.2}$$

i.e., the benchmark is the performance of the prior independent optimal algorithm scaled up by its approximation factor. Taking the expectation of $\boldsymbol{v}$ drawn from any distribution F, we have

$$B^*(F) = \alpha^{\mathcal{F}} \cdot A^*(F) \tag{6.3}$$

First, notice that B^* is normalized. Because A^* is a prior independent $\alpha^{\mathcal{F}}$-approximation, it follows that $A^*(F) \geq \frac{1}{\alpha^{\mathcal{F}}} \operatorname{OPT}_F(F)$ for all $F \in \mathcal{F}$ the class of distributions. Multiplying through by $\alpha^{\mathcal{F}}$ and applying equation (6.3) shows that the benchmark meets the definition of normalization.

Second, equation (6.2) implies that A^* is a prior free $\alpha^{\mathcal{F}}$-approximation of the benchmark B^*. Therefore, (A^*, B^*) is in fact a solution to the brBDP $(\gamma_\rho^{\mathcal{F}})$ with ratio $\alpha^{\mathcal{F}}$. The optimal solution to the program is no larger. Therefore $\gamma_\rho^{\mathcal{F}} \leq \alpha^{\mathcal{F}}$. □

Specifically note that the proof technique of Lemma 16 was to strategically define a benchmark using the performance function of the argmin A^* of the prior independent program, and then assign the resulting benchmark along with algorithm A^* into the arguements of the brBDP. Combined with Corollary 4, this further shows that the benchmark $B^* = \alpha^{\mathcal{F}} \cdot A^*$ is in fact *optimal*.

Corollary 5. *Given a class of distributions $\mathcal{F}$ and a class of algorithms $\mathcal{A}$, and consider the best-response resolution benchmark design problem (brBDP). The optimal benchmark B^* is given by $\alpha^{\mathcal{F}} \cdot A^*$ which is the performance of the prior independent optimal algorithm scaled up by its optimal approximation $\alpha^{\mathcal{F}}$.*

The relaxation from best-response resolution to normalization-symmetric is not generally tight. In particular, the optimal benchmark for the nsBDP can have $\gamma_\sigma^{\mathcal{F}} < \gamma_\rho^{\mathcal{F}}$. The following theorem states this strict inequality for an example from mechanism design using the Triangle revenue-curve distributions $\mathcal{F}^{trv}$ from Theorem 6 in Section 3.3. The technical proof of Theorem 17 is deferred to the reference paper.

Theorem 17. *Given a single item, 2-agent auction with a revenue objective, the class of truthful and scale-invariant mechanisms $\mathcal{M}^{\text{mark}}$, and the class of default-scaled Triangles $\mathcal{F}^{\text{trv}}$. The nsBDP ($\gamma_\sigma^{\mathcal{F}}$) has a strictly smaller objective value than the brBDP ($\gamma_\rho^{\mathcal{F}}$) and the optimal prior independent approximation $\alpha^{\mathcal{F}}$, i.e., $\gamma_\sigma^{\mathcal{F}} < \gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}}$.*

6.3. Prior Free versus Prior Independent Expert Learning

The previous section showed that optimal benchmark design using best-response resolution is equivalent to prior independent optimization. Moreover, the optimal prior free algorithm for the optimal benchmark is the optimal prior independent algorithm. A consequence of these results is that there is no added robustness from the prior free framework over the prior independent framework. In this section we observe, from an example of *expert learning*, that this potential lack of robustness is serious and the optimal prior independent algorithm can perform much worse than the standard algorithms that are known to approximate the standard prior free *best-in-hindsight* benchmark.

These observations are straightforward from the perspective of the expert learning literature. We discuss them and outline key results statements so as to map them onto the framework of Section 6.1. Otherwise however, we omit technical proofs and some supporting lemma statements (for full proofs see the reference paper: **Hartline et al. (2020a)**). As the last in our list of results, we include a statement that the relaxation of benchmark design (from brBDP to nsBDP) is not without loss for expert learning (which is thus a parallel to Section 6.2).

Introduction to Expert Learning. We consider the binary-reward variant of the canonical online expert learning problem. A single player plays a repeated game against Nature for n rounds. In each round $t \in \{1, \ldots, n\}$, each expert j from a discrete set $\{1, \ldots, k\}$ will receive a binary reward $v_{t,j} \in \{0, 1\}$. Thus, each input has support $\{0, 1\}^k$ and the input space is $\mathcal{V} = [\{0, 1\}^k]^n$.

Before rewards are realized in each round, the player chooses to "follow" a (possibly randomized) expert for the round, and receives a reward (possibly in expectation) equal to the reward of the followed expert. When the round concludes, the player gets to observe the rewards of all experts, including those not followed by the player. The player's algorithm is A which outputs distributions $A_t(\boldsymbol{v})$ over experts using only the history $(v_1, \ldots, v_{t-1})$ in each round t. The class of all such *online* algorithms is denoted by $\mathcal{A}^{\text{onl}}$ and the performance of an online algorithm $A \in \mathcal{A}^{\text{onl}}$ on input $\boldsymbol{v}$ is:

$$A(\boldsymbol{v}) = \sum\nolimits_{t=1}^{n} \mathbf{E}_{j \sim A_t(\boldsymbol{v})} \left[v_{t,j} \right].$$

To simplify analysis of the example in this section, we restrict the prior independent comparison class of distributions. Consider the following family of *binary independent stationary* distributions $\mathcal{F}^{\text{bis}}$ for the Bayesian variant of the expert learning problem. For a distribution $F \in \mathcal{F}^{\text{bis}}$, each expert j's reward in each round is a Bernoulli random variable with mean $\mathbb{F}_j = \mathbf{E}[v_{t,j}]$. The class $\mathcal{F}^{\text{bis}}$ is composed of all possible k-permutations over means $\mathbb{F}_j \in [0, 1]$ for each expert j. In each round t, the rewards are drawn independently from each other and from other rounds. Importantly, the distribution and mean of each expert's reward is identical across rounds.

As outlined in Section 6.1.1 we can define Bayesian, prior independent, and prior free versions of the expert learning problem. We summarize as follows:

- In the Bayesian model, the optimal algorithm is $\text{OPT}_F = \text{argmax}_{A \in \mathcal{A}^{\text{onl}}} A(F)$. For a known, binary independent stationary distribution $F \in \mathcal{F}^{\text{bis}}$, the optimal algorithm picks the expert with the highest ex ante probability $j^* = \text{argmax}_j \mathbb{F}_j$ and follows expert j^* in each round; its expected performance (over n rounds) is

$$\text{OPT}_F(F) = n \max_j \mathbb{F}_j.$$

- In the prior free model, the optimal algorithm is the one that minimizes regret in worst-case over inputs $\boldsymbol{v} \in \mathcal{A}$ against a given benchmark B defined as

$$\alpha^B = \min_{A \in \mathcal{A}^{\text{onl}}} \max_{\boldsymbol{v} \in \mathcal{A}} [B(\boldsymbol{v}) - A(\boldsymbol{v})].$$

The *best-in-hindsight* benchmark B^{BIH} for any reward profile $\boldsymbol{v} \in \mathcal{A}$ is the total reward of any fixed expert, i.e.,

$$B^{\text{BIH}}(\boldsymbol{v}) = \max_{j \in \{1,\ldots,k\}} \sum\nolimits_{t=1}^{n} v_{t,j}.$$

The canonical approach to online expert learning measures performance in terms of worst-case regret with respect to this prior free benchmark: B^{BIH}.

- In the prior independent model, the optimal algorithm is the one that minimizes regret in worst-case over distributions $F \in \mathcal{F}$ against the optimal algorithm for the distribution

$$\alpha^{\mathcal{F}} = \min_{A \in \mathcal{A}^{\text{onl}}} \max_{F \in \mathcal{F}} [\text{OPT}_F(F) - A(F)].$$

We will be considering this question for the class of binary independent stationary distributions $\mathcal{F}^{\text{bis}}$ where $\text{OPT}_F(F)$ is from the Bayesian model above.

Specifically, the natural *follow-the-leader* algorithm, which in round t chooses a uniform random expert from the set of experts with highest total reward from the first $t-1$ rounds, is the prior independent optimal algorithm (Theorem 18).

Definition 36. *The* follow-the-leader algorithm (FTL) *selects an expert uniformly at random from the set of experts with highest total reward from previous rounds:*

$$A_t^{\text{ftl}}(\boldsymbol{v}) = \text{Uniform over} \quad \left\{ \text{argmax}_j \sum\nolimits_{t'<t} v_{t',j} \right\}$$

Theorem 18. *For binary independent stationary distributions, the follow-the-leader algorithm is the prior independent optimal online learning algorithm.*

FTL further directly informs the optimal bsBDP benchmark (from Theorem 16). Recall that by design, adjusting-up the performance of FTL by its prior independent regret is the minimal adjustment needed to satisfy normalization.

Corollary 6. *The optimal benchmark from best-response resolution is calculated as the performance of FTL adjusted up additively by its regret-approximation of the prior independent benchmarks set by the respective performances* $\mathrm{OPT}_F(F)$ *for each* $F \in \mathcal{F}^{bis}$. *The optimal prior free algorithm against this benchmark is FTL.*

The rest of this section covers two key ideas: (1) FTL *does quite poorly* against the BIH benchmark which is the standard prior free benchmark for expert learning and thus we recognize that our model requires further study; (2) we state strict inequality between optimal resolution measures of the brBDP and nsBDP for expert learning.

Follow-the-leader is a Linear Approximation of Best-in-hindsight. We observe that the BIH benchmark is normalized. In fact, it is analogous to the optimal-Bayesian-optimal benchmark (OBO) for evaluating prior free mechanisms (Hartline and Roughgarden (2008), see page 243). Thus, an algorithm that is a prior free approximation of BIH is also a prior independent approximation algorithm (with the same bound on regret, cf. Proposition 4).

Lemma 17. *For inputs from binary independent stationary distributions* $\mathcal{F}^{bis}$, *the best-in-hindsight benchmark* B^{BIH} *is normalized, i.e.,* $B^{\mathrm{BIH}}(F) \geq \mathrm{OPT}_F(F)$, $\forall F \in \mathcal{F}^{bis}$.

From Corollary 6, the optimal prior free benchmark is the performance of the follow-the-leader algorithm adjusted up by its prior independent regret-approximation. Moreover, the optimal algorithm for the optimal benchmark is the follow-the-leader algorithm

itself. While we may have hoped for the prior free analysis to lead to strictly more robust algorithms than the prior independent analysis, by optimizing benchmarks in the framework provided in Section 6.1.3, we have lost all of this potential robustness. Specifically, the standard expert learning algorithms that have low worst-case regret against the best-in-hindsight benchmark exhibit robustness that the follow-the-leader lacks. This observation is formalized in the following lemma which contrasts with the optimal regret of standard algorithms like randomized weighted-majority (Littlestone and Warmuth, 1994). The optimal worst-case regret against best-in-hindsight is $\Theta(\sqrt{n \ln k})$ for k experts, n rounds, and binary rewards (Haussler et al. (1995); or $[0, 1]$ rewards: Kalai and Vempala (2005)).

By comparison, FTL gets linear regret $\Theta(n)$ against the BIH benchmark (Fact 23). This follows from a well-known example of "bad" worst-case inputs designed to alternate according to:

- odd round payoffs: $(1, 0, \ldots, 0) \in \{0, 1\}^k$
- even round payoffs: $(0, 1, \ldots, 1) \in \{0, 1\}^k$

FTL chooses a uniform random expert for odd rounds and obtains expected payoff $1/k$ and chooses expert 1 for even rounds and obtains expected payoff of 0, however, the BIH benchmark increases linearly on average (by 1 every 2 rounds).

Fact 23. *The prior free regret of follow-the-leader against the best-in-hindsight benchmark with n rounds is $\Theta(n)$.*

The observations of this section suggest that further study of the formulation of the benchmark optimization problem is necessary to better understand the trade-offs between

prior free and prior independent robustness. This assessment in turn serves as justification for our inclusion of the definition of normalization-symmetric resolution as an alternative measure of normalized lower-bounds in Section 6.1.4. Recall, there is extended analysis of its nsBDP in Appendix A.6 (which in particular includes a direct connection to our Blends Technique).

Benchmark Design Gap for Expert Learning. Finally, we simply state that the inequality $\gamma_\sigma^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$ of Proposition 5 is strict for expert learning (cf., the same observation for an example from mechanism design in Theorem 17). I.e., the nsBDP has strictly smaller value than the brBDP for expert learning (albeit, the nsBDP may be the more-economic benchmark design framework, see the discussion after Proposition 5 on page 164).

Theorem 19. *Given the expert learning problem with the class of with k experts, finite time horizon $n \geq 2$, and the class of binary independent stationary reward distributions $\mathcal{F}^{\text{bis}}$. The nsBDP ($\gamma_\sigma^{\mathcal{F}}$) has a strictly smaller objective value than the brBDP ($\gamma_\rho^{\mathcal{F}}$), i.e., $\gamma_\sigma^{\mathcal{F}} < \gamma_\rho^{\mathcal{F}}$.*

CHAPTER 7

Conclusion

We start this Conclusion with an overview of the holistic, thematic perspectives which couched this thesis, towards identifying promising directions of future research as both *implicated and inspired* by its main results. Subsequently we summarize the main contributions of this thesis together with next objectives for new work by theme.

The theories of economics and computer science are inherently and essentially related. On one side solutions for models and games within economics require computational tractability to be practically relevant. On the other side there is a reduction from the analysis of robustness for arbitrary algorithms questions to a fundamental two-player game (effectively, Yao's Minimax Theorem 3). The result is the broad field of algorithmic game theory. We largely considered agnostic analyses that are appropriate for at least two major branches of algorithms questions: mechanism design and online learning.

The main concern of this thesis has been Bayesian-robustness of algorithms though a spectrum of information settings have received attention (see Section 6.1.1). In studying robustness, a number of natural perspectives arise and lead to entire fundamental subtopics. This Conclusion will organize around the following themes of this thesis:

(1) **Foundations of Benchmark Design.** *Benchmark functions* describe target performance. The benchmark can theoretically be set too low, or too high. Benchmarks themselves can and should be subject to scientific inquiry.

(2) **Adversarial Play as Information-Design-Design.** Benchmark approximation is measured against a worst "test case" as if chosen by an omniscient *adversary*, playing against the algorithm in a zero-sum game. Independent of algorithm setting, the Blends Technique follows from analysis of the adversary's play with connections to *information design* and *non-unique tensor decomposition*.

(3) **Scalability of Inputs.** Any distribution can be re-scaled by composing its CDF with $f_k(x) = x/k$. When inputs come from the unbounded set of (positive) reals, algorithms must handle all scales; versus, algorithms can take advantage of inputs having bounded support.

(4) **Econometric Inference in Auctions.** Loosely related to robustness, we add one more theme related to private information in mechanism design:[1] there is a question of *econometric inference*, i.e., identifying auction inputs from its rules and its outcomes. *Counterfactual estimation* considers the question offline, and *dashboard mechanisms* online.

Benchmark Design. A formal approach to benchmark design will seek out the existence of guiding principles to inform and measure the design of the benchmark, motivated by the heuristic narrative that intuitively we must guard both against designing benchmarks that are too small or too big. Hartline and Roughgarden (2014) proposed *normalization* (Definition 31), which ties prior free benchmark design back to Bayesian settings in a way that benchmarks are justifiably not too small. Hartline (2020) proposes a second property *resolution* (Definition 32).

[1] This direction follows from two of the author's other published joint-works which were otherwise excluded from this thesis: **Hartline et al. (2019)** and **Hartline et al. (2020b)**.

For a general algorithms setting, Theorem 16 connected the prior independent and prior free settings together. Further, the optimal benchmark is set by scaling up the performance of the optimal prior independent algorithm to be big enough to meet the definition of normalization.

Future work. Whereas Theorem 16 solves the best-response benchmark design problem (brBDP; Definition 34) for the optimal benchark – insofar as it describes it in terms of the prior independent optimal algorithm – the normalization-symmetric benchmark design program (nsBDP; Definition 35) remains unsolved. This question is of increased interest because it relates directly to optimization of lower bounds from The Blends Technique (more from dual blends below), and remains interesting for each distinct algorithms problem because the quantities $\mathrm{OPT}_F(F)$ are distinct to each problem.

A second, orthogonal direction for next research on benchmark fundamentals is the following. The combination of the normalization and best-response resolution properties is elegant, but has known drawbacks. First, there is the concern in this thesis that the brBDP does not increase robustness in the prior free setting beyond the prior independent setting. Further, there is actually no constraint on the benchmark design (for either brBDP or nsBDP) to prevent the optimal benchmark from being set larger than the offline optimal performance at a specific input $\boldsymbol{v}$, effectively a pointwise unachievable target.[2] Thus, the optimal benchmark may be increased (above offline optimal) on inputs where it is not tight to resolution, simply in order to help meet the requirement of normalization. The question is to identify what pertinent new property on benchmarks can address this most effectively, and what new scientific benchmarks result.

[2]For a single-item auction with a revenue objective and 2 agents, consider setting the benchmark greater than 1 when the value-profile input is $(1, 1)$.

Adversarial Play. Lower bounds on approximation may be identified from fixed adversary strategies (Yao's Minimax Principle, Theorem 9). With an *optimal* such adversarial strategy, we can measure the tight approximation factor of an optimal prior independent algorithm. This thesis developed the following novel method to prove lower bounds for an arbitrary algorithm question, in the prior independent setting.

The adversary picks a blend $\delta_2 \in \Delta(\mathcal{F})$ which induces a correlated distribution g over input space $\mathcal{V}^n$ and thus sets the prior independent benchmark to be the expected optimal performance over this blend (Lemma 21). By design, there exists a second blend to describe the same correlated distribution, namely $\delta_1 \in \Delta(\mathcal{F}^{\text{all}})$ (which may use any distribution as an inherent consequence of nature). I.e., the same correlated distribution is described by n inputs drawn either i.i.d. as $v_i \sim (F_2 \sim \delta_2)$, or drawn i.i.d. as $v_i \sim (F_1 \sim \delta_1)$. From this set up, the ratio of optimal performances (in expectation over distributions of the respective blends) $\text{opt}_{n,2}/\text{opt}_{n,1}$ is a necessary lower bound on approximation of the original prior independent problem.

The technique is information design of a different flavor than is standard (Section 5.4). In our analysis, the adversary fully reveals the choice of δ_2 but the designer is powerless to use it directly. Instead the correlated distribution is like a continuous-dimension tensor, symmetric, with factors as probability distributions and weights on factors also as a probability distribution (Section 5.5).[3] In this stylized tensor setting, it is the non-uniqueness of decomposition of the tensor that leads to necessary gaps in performance between the benchmark and any algorithm.

[3] This thesis presented a number of these *dual blends*, e.g., in Section 4.4 and Section 4.5.1.

Future work. There are many research directions following from blends, such that they are presented only at a high level. The Blends Technique itself ties into the dual program description of the nsBDP (as mentioned above). A first pressing question within prior independent analysis is to determine for what classes of algorithms questions does there exist a lower bounds from blends that is tight to the optimal prior independent approximation. More generally there is a question of how to optimize lower bounds from dual blends (cf. Section 5.4). The Blends Technique as exhibited in this thesis applies directly for $n = 2$ inputs, but for more inputs the Blends Technique seems to face a likely impossibility result (alternatively from algebra of dual blends for order-statistic separability (Section 5.1) or from general uniqueness of third-order tensor decomposition (Appendix A.7.1).) We propose here a partial extension for $n \geq 3$ using an idea to *project away* all but the two largest order statistics and only requiring that the "common correlated distribution" $g^{(1),(2)}$ match up at each input when reduced to two (largest) coordinates.[4]

Individual algorithms problems have the potential to realize first-ever or improved lower bounds on approximation from the Blends Technique, after which characterizations of classes of algorithms can be studied. The method generally can be framed as information-design-design and connections to existing work on information design can be strengthened. Finally, it may give rise to new special-case questions within the study of non-uniqueness of matrix decomposition, for example, we described a nice class of $n = 2$

[4] In fact *this direction succeeds* and is currently being developed. Extensions to $n \geq 3$ would benefit in particular settings such as online algorithms.

dual blends solutions which were characterized as having their two sides of blends generated each by one of a pair of distributions which are *inverse-distributions* to each other (Theorem 14 in Section 5.2).

Scale-Invariant Design (or lack thereof). This thesis solved a simple variant of the longstanding open question of the optimal prior independent truthful mechanism for 2 agents with unbounded value support, a revenue objective, and all regular distributions as the comparison class (Theorem 5). Specifically, the variant adds a restriction to only consider *scale-invariant* mechanisms. Remarkably, the description of the optimal prior independent mechanism is not the result of some complicated equation, as might be feared. It is a simple mix over the Second Price Auction, and an auction posting approximately 2.447-times the second price to the largest-valued agent.

Future work. A top priority is to establish an argument that it is sufficient to restrict attention to scale-invariant algorithms; and further, the argument should hopefully exist from first principles, and apply very broadly to algorithms questions with ideally no assumptions on the setting. The rest of this discussion of future work regarding the scale-invariance property focuses on mechanism design as the main application of this thesis.

Our optimal prior independent mechanism solution is specifically for 2 agents. Initial attempts to extend to $n > 2$ agents show that further technical tools will be needed for a solution. Additionally, there is the question of solving for the optimal prior independent mechanism in the single-item, 2-agent setting for the *residual surplus* objective, with or without scale-invariance.

(We return the discussion to the revenue objective.) Perhaps more interestingly, scale-invariance of the optimal mechanism is likely to be a sufficient assumption when values are unbounded, and becomes possibly its most prevalent feature. Two follow up questions in the same prior independent space become quite interesting. First, note that within the context of scale-invariant mechanisms, the optimal auction (of Theorem 5) only posts two prices to the largest agent, conditioned on the second price: a price "marked up" by factors of 1 or 2.447. At this point, the mechanism is taking quite specific actions with small total support. It makes sense to additionally consider the robustness theme of *minimizing regret* with respect to these markup factors. Related, note that the standard question of prior independence allows the benchmark performance to be set by an optimal mechanism that knows the distribution. In the case that we solved, the optimal auction knows the scale of the distribution. The first follow up perspective considers changing the benchmark to only be set by the optimal performance of a respectively scale-invariant mechanism. Thus, the distribution's scale is meaningless and only its shape matters.

The second perspective considers an assumption of hard bounds on the support of value space, for example $[1, h]^n$. In this case, the idea of scale-invariant design has no relevance, as inputs observed to be close to the boundaries will have material, scale-dependent effects on optimal design. A gap in prior independent approximation between unbounded and finite-support cases would represent a *price of scale-anonymity.* Of note, lower bounds on approximation from the Blends Technique of the Adversarial Play section may become directly relevant. An intuition for this possibility is: (a) input space and thus adversary-blends-space being compact and convex Hilbert spaces suggests that mixed Nash may necessarily exist from fixed-point theorems in which case, (b) dual blends which have one

side as a blend over specifically *equal revenue distributions* (Definition 13) would make the optimal mechanism design indifferent over all lookahead-posted-prices (Definition 6) which is a common theme of equilibrium (make the opponent indifferent over all played actions) and thus worth investigating.

Applications of Inference in Auctions. We end with a summary of another direction of the author's research in auction theory. The predominant setting is prior free: agents' values are fixed constants, unknown by the designer. The theme is *econometric inference*, including results for both "offline" estimation and "online" implementation. Generally implicit in the setting is a repeated auction: observations of historical agent behavior inform decisions by the auction designer in the future.

Hartline et al. (2020b) addresses a first question of offline econometric inference – the goal is to identify the agents' private values. The setting is truthful single-item auctions with proportional weights allocation rules, and charging winner-pays prices to (stochastic) winners. The perspective is that of an *outsider* to the auction who can not observe the agents' reports, but rather can observe the payments made by winners. This work's two main results are: (1) that Myerson's payment function (Theorem 1) in this setting is one-to-one and therefore it can be theoretically inverted; and (2) a given algorithm to compute the inversion arbitrarily to within ϵ.[5] Its supporting proofs include new results of independent interest. In particular, it extends uniqueness of pure Nash equilibrium in "concave games" of Rosen (1965), using a related definition of concavity based on Gale and Nikaido (1965).

[5]The algorithm has running time that is polynomial in the number of agents n, the upper bound of the support of value space h, $\ln 1/\epsilon$, and the number of calls to the agents' weights functions (embedding dependence on the complexity of computing them).

Hartline et al. (2019) explores the potential use of a *bidding dashboard* by a designer in a repeated *non-truthful* auction setting (with all-pay semantics) to facilitate agents' understanding and computation of optimal play. For simplicity, assume a benevolent designer who wants to maximize the total surplus of the auction. Having assumed a non-truthful setting, the VCG-mechanism is not available.

Two key complications are introduced with non-truthful settings. First, equilibrium is likely to have a complicated description of agents' jointly-optimal strategies, which implies that real world agents are unlikely to be able to compute and execute the bids of theoretical equilibrium, regardless of other considerations. Second, there might be multiple equilibria, with the *price of anarchy* ratio (Koutsoupias and Papadimitriou, 1999; Papadimitriou, 2001; Roughgarden, 2009) between welfare of the worst equilibrium and the offline-optimal welfare being possibly quite large asymptotically (Dütting and Kesselheim, 2015). Thus, even if agents are able to self-direct themselves towards an equilibrium, it might have decidedly poor welfare performance.

Bidding dashboards are estimated *price-allocation* curves published by the platform in advance of an auction, with one individually-tailored curve given privately to each agent. Their use is considered in a many-round analysis, for which the main result is to show that a *sequential-non-truthful* auction gets almost exactly the same performance as a comparative *sequential-truthful* auction, in the long-term.

Future work. **Hartline et al. (2020b)** solved price inversion under a number of restrictive assumptions (single-item, proportional weights allocation, fixed agent values), each of which can potentially be generalized. Its results related to Rosen's concave games connect to the emerging field of "learning in games as a computational tool" (e.g.,

Daskalakis and Panageas (2018); Abernethy et al. (2021)) and some of its pursuits like fast convergence (Syrgkanis et al., 2015), and the possibility of convergence of learning in games to smaller classes of equilibrium than is guaranteed by the existing literature (Blum et al., 2008).**Hartline et al. (2019)** describes a mechanism that could potentially run in the real world; its theoretical analysis would benefit from empirical study of its implementation – even initially with offline simulation – to help understand its inherent convergence properties and parameter tuning.

References

Abernethy, J., Lai, K. A., and Wibisono, A. (2021). Last-iterate convergence rates for min-max optimization: Convergence of hamiltonian gradient descent and consensus optimization. In *Algorithmic Learning Theory*, pages 3–47. PMLR.

Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, New York.

Alaei, S., Hartline, J., Niazadeh, R., Pountourakis, E., and Yuan, Y. (2018). Optimal auctions vs. anonymous pricing. *Games and Economic Behavior*.

Allouah, A. and Besbes, O. (2018). Prior-independent optimal auctions. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 503–503. ACM.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–57.

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95.

Bhaskara, A., Charikar, M., Moitra, A., and Vijayaraghavan, A. (2014). Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272.

Blum, A., Hajiaghayi, M., Ligett, K., and Roth, A. (2008). Regret minimization and the price of total anarchy. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, page 373–382, New York, NY, USA. Association for Computing Machinery.

Borodin, A. and El-Yaniv, R. (1998). *Online Computation and Competitive Analysis*. Cambridge University Press, Cambridge, UK.

Boyar, J., Irani, S., and Larsen, K. S. (2015). A comparison of performance measures for online algorithms. *Algorithmica*, 72(4):969–994.

Bulow, J. and Klemperer, P. (1996). Auctions versus negotiations. *American Economic Review*, 86(1):180–194.

Bulow, J. and Roberts, J. (1989). The simple economics of optimal auctions. *The Journal of Political Economy*, 97:1060–90.

Chen, N., Gravin, N., and Lu, P. (2014). Optimal competitive auctions. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, page 253–262, New York, NY, USA. Association for Computing Machinery.

Dasgupta, S. (1999). Learning mixtures of gaussians. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 634–644. IEEE.

Daskalakis, C. and Panageas, I. (2018). The limit points of (optimistic) gradient descent in min-max optimization. In *Advances in Neural Information Processing Systems 31*, pages 9236–9246. Curran Associates, Inc.

Devanur, N. R., Hartline, J. D., and Yan, Q. (2015). Envy freedom and prior-free mechanism design. *Journal of Economic Theory*, 156:103–143.

Dhangwatnotai, P., Roughgarden, T., and Yan, Q. (2010). Revenue maximization with a single sample. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 129–138.

Dhangwatnotai, P., Roughgarden, T., and Yan, Q. (2015). Revenue maximization with a single sample. *Games and Economic Behavior*, 91:318–333.

Donoho, D. and Stodden, V. (2004). When does non-negative matrix factorization give a correct decomposition into parts? In *17th Annual Conference on Neural Information Processing Systems, NIPS 2003*. Neural information processing systems foundation.

Dughmi, S., Niazadeh, R., Psomas, A., and Weinberg, S. M. (2019). Persuasion and incentives through the lens of duality. In *International Conference on Web and Internet Economics*, pages 142–155. Springer.

Dütting, P. and Kesselheim, T. (2015). Algorithms against anarchy: Understanding non-truthful mechanisms. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 239–255. ACM.

Feldman, J., O'Donnell, R., and Servedio, R. A. (2008). Learning mixtures of product distributions over discrete domains. *SIAM Journal on Computing*, 37(5):1536–1564.

Feldman, J., Servedio, R. A., and O'Donnell, R. (2006). Pac learning axis-aligned mixtures of gaussians with no separation assumption. In *International Conference on Computational Learning Theory*, pages 20–34. Springer.

Feng, Y. and Hartline, J. D. (2018). An end-to-end argument in mechanism design (prior-independent auctions for budgeted agents). In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 404–415. IEEE.

Fiat, A., Goldgerg, A., Hartline, J., and Karlin, A. (2002). "competitive generalized auction". In *Proceeding of 34th Annual ACM symposium on Theory of computing*.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Fu, H., Immorlica, N., Lucier, B., and Strack, P. (2015). Randomization beats second price as a prior-independent auction. In *Proceedings of the Sixteenth ACM Conference*

on Economics and Computation, EC '15, page 323, New York, NY, USA. Association for Computing Machinery.

Gale, D. and Nikaido, H. (1965). The Jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159(2):81–93.

Glazer, J. and Rubinstein, A. (2004). On optimal rules of persuasion. *Econometrica*, 72(6):1715–1736.

Goldberg, A. and Hartline, J. (2003). Envy-free auctions for digital goods. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, pages 29–35. ACM.

Goldberg, A. V., Hartline, J. D., Karlin, A., Saks, M., and Wright, A. (2006). Competitive auctions. *Games and Economic Behavior*, 55:242–269.

Goldberg, A. V., Hartline, J. D., Karlin, A. R., and Saks, M. (2004). A lower bound on the competitive ratio of truthful auctions. In Diekert, V. and Habib, M., editors, *STACS 2004*, pages 644–655, Berlin, Heidelberg. Springer Berlin Heidelberg.

Harshman, R. A. (1972). Determination and proof of minimum uniqueness conditions for parafac1. *UCLA working papers in phonetics*, 22(111-117):3.

Hartline, J. (2011-2020). Mechanism design and approximation. *Working manuscript*. `jasonhartline.com/MDnA/`.

Hartline, J. (2013). Bayesian mechanism design. *Foundations and Trends® in Theoretical Computer Science*, 8(3):143–263.

Hartline, J. (2014). Prior-free analysis of digital-good auctions and beyond. *Turing's Invisible Hand,* 4 September 2014, `agtb.wordpress.com/2014/09/04/prior-free-analysis-of-digital-good-auctions-and-beyond/`.

Hartline, J. and Johnsen, A. (2021). Lower bounds for prior independent algorithms. *Working paper*. `arxiv.org/pdf/2107.04754.pdf`.

Hartline, J., Johnsen, A., and Li, Y. (2020a). Benchmark design and prior-independent optimization. In *61st Annual Symposium on Foundations of Computer Science*, IEEE FOCS '20. Institute of Electrical and Electronics Engineers.

Hartline, J., Johnsen, A., Nekipelov, D., and Wang, Z. (2020b). Inference from auction prices. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms*, SODA '20, pages 2472–2491. SIAM.

Hartline, J., Johnsen, A., Nekipelov, D., and Zoeter, O. (2019). Dashboard mechanisms for online marketplaces. In *Proceedings of the 2019 ACM 20th Conference on Economics and Computation*, ACM EC '19. Association for Computing Machinery.

Hartline, J. and McGrew, R. (2005). From optimal limited to unlimited supply auctions. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 175–182.

Hartline, J. and Roughgarden, T. (2008). Optimal mechanism design and money burning. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 75–84".

Hartline, J. and Roughgarden, T. (2014). Optimal platform design. *CoRR*, `arxiv.org/abs/1412.8518`.

Hartline, J. D., Kleinberg, R., and Malekian, A. (2015). Bayesian incentive compatibility via matchings. *Games and Economic Behavior*, 92:401–429.

Håstad, J. (1989). Tensor rank is np-complete. In *International Colloquium on Automata, Languages, and Programming*, pages 451–460. Springer.

Haussler, D., Kivinen, J., and Warmuth, M. K. (1995). Tight worst-case loss bounds for predicting with expert advice. In Vitányi, P., editor, *Computational Learning Theory*, pages 69–83, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hsu, D. and Kakade, S. M. (2013). Learning mixtures of spherical gaussians: moment methods and spectral decompositions. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 11–20.

Kalai, A. and Vempala, S. (2005). Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615.

Koutsoupias, E. and Papadimitriou, C. (1999). Worst-case equilibria. In Meinel, C. and Tison, S., editors, *STACS 99*, pages 404–413, Berlin, Heidelberg. Springer Berlin Heidelberg.

Koutsoupias, E. and Papadimitriou, C. H. (2000). Beyond competitive analysis. *SIAM Journal on Computing*, 30(1):300–317.

Kruskal, J. B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138.

Littlestone, N. and Warmuth, M. K. (1994). The weighted majority algorithm. *Information and computation*, 108(2):212–261.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

Papadimitriou, C. (2001). Algorithms, games, and the internet. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 749–753.

Plassman, G. E. (2005). A survey of singular value decomposition methods and performance comparison of some available serial codes. Technical report, Raytheon Technical Services, Hampton, Virginia.

Rabanser, S., Shchur, O., and Günnemann, S. (2017). Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*.

Rayo, L. and Segal, I. (2009). Optimal information disclosure. *Journal of Political Economy*, 118:949–949.

Ronen, A. (2001). On approximating optimal auctions. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, EC '01, page 11–17, New York, NY, USA. Association for Computing Machinery.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534.

Roughgarden, T. (2009). Intrinsic robustness of the price of anarchy. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, page 513–522, New York, NY, USA. Association for Computing Machinery.

Roughgarden, T. (2019). Beyond worst-case analysis. *Communications of the ACM*, 62(3):88–96.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582.

Sleator, D. D. and Tarjan, R. E. (1985). Amortized efficiency of list update and paging rules. *Communications of the ACM*, 28(2):202–208.

Syrgkanis, V., Agarwal, A., Luo, H., and Schapire, R. E. (2015). Fast convergence of regularized learning in games. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, page 2989–2997, Cambridge, MA, USA. MIT Press.

Townsend, A. and Trefethen, L. N. (2015). Continuous analogues of matrix factorizations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2173):20140585.

Vohra, R. V. (2011). *Mechanism design: a linear programming approach*, volume 47. Cambridge University Press.

Yao, A. C. (1977). Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pages 222–227.

APPENDIX A

Deferred Analyses and Proofs

A.1. An Alternative Proof of The Blends Technique from Linear Programming

(from page 87) We give a second, informal proof of The Blends Technique (which implements Theorem 10) for algorithms settings in which it is possible to explicitly model the prior independent problem (Definition 2) as a linear program, in particular in which the algorithm's performance is a linear combination over variables. We use a specific example of truthful auctions within mechanism design but it will be clear where algorithm-specific considerations "disappear" and we are left with an alternative proof for the Blends Technique. The techniques and principles of linear programming that we apply here follow from Vohra (2011).

To summarize, this section re-proves the Blends Technique using an example problem (a simple auction) in a restricted analytical setting (linear programming). We identify two prominent structures:

(1) The Blends Technique describes lower bounds by measuring the prior independent approximation of an "algorithm" that – rather than choosing assignments of problem-specific variables – can directly choose its *pseudo-performance* outcome on every input $\boldsymbol{v}$ independently of problem-specific constraints, as long as for every distribution $F \in \mathcal{F}^{\text{all}}$, its expected pseudo-performance on inputs drawn

from F does not exceed the optimal algorithm's performance $\text{OPT}_F(F)$. This structure is observed in Linear Program 4 below, which is a relaxation of the initial problem's LP.

(2) Linear Program 5 is the dual program of the primal in the previous point. Critically, our dual blends (of Definition 18) give feasible solutions for this dual program. The Blends Technique for obtaining lower bounds on prior independent approximation then follows from the inequality between the optimal value of the primal program and the value of the objective of the dual for feasible solutions.

Regarding specifics of mechanism design: we write a program to describe the prior independent truthful mechanism design problem, for which it is sufficient to use virtual value maximization and characterization of truthful mechanisms (Theorem 2, Myerson (1981)). Note that we can write the program once and it applies for each objective using the corresponding virtual value function. Further, the linear programming approach – in conjunction with Myerson's characterization – uses the fact that optimization over truthful mechanisms $M = (\boldsymbol{x}, \boldsymbol{p})$ reduces to optimization over implementable allocations $\boldsymbol{x}$ (cf. Theorem 1). Thus, the arguments of the initial linear program are (monotone) allocations $\boldsymbol{x}$. Let $M(F)$ be the expected performance of mechanism M on n i.i.d. draws from F.

In order to write the problem as a linear program, we define $\check{\alpha}^{\mathcal{F}} = 1/\alpha^{\mathcal{F}}$ to be the multiplicative inverse of our standard approximation factor. Thus, we may think of $\check{\alpha}^{\mathcal{F}} \in [0, 1]$ as the largest $(\max - \min)$ fraction of OPT_F that optimal M^* can guarantee in worst-case (i.e., $M^*(F) \geq \check{\alpha}^{\mathcal{F}} \cdot \text{OPT}_F(F) \; \forall \; F \in \mathcal{F}$).

We need to write a linear program with a single objective. The technique to "unravel" the $\max - \min$ formulation (of prior independent design) in order to remove the embedded

adversarial-min-objective relies on moving it into a constraint (see the "approximation" line below) and optimizing an approximation-ratio variable $\check{\alpha}$ as the value of the program. The optimal factor $\check{\alpha}^{\mathcal{F}}$ is necessarily at most 1 and we copy this fact into the objective function line. Let $\boldsymbol{f}(\boldsymbol{v}) = \prod_i f(v_i)$.

Linear Program 1 (The Prior Independent Truthful Mechanism Design Program)**.** *Given a class of distributions $\mathcal{F}$ and any auction objective – along with its corresponding definition of the virtual value function – the optimal single-item, n-agent truthful mechanism (described by $\boldsymbol{x}^*$) and its optimal approximation factor $\alpha^{\mathcal{F}} = 1/\check{\alpha}^{\mathcal{F}}$ are given by the* argmax *of the following program:*[1]

(A.1)
$$\check{\alpha}^{\mathcal{F}} = \max_{\boldsymbol{x},\ \check{\alpha}} \check{\alpha} \ \leq 1$$

s.t.

$$\int_{\mathcal{V}^n} \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v}) \right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \geq \check{\alpha} \cdot \mathrm{OPT}_F(F) \quad \forall\ F \in \mathcal{F} \qquad \text{(approximation)}$$

$$\sum\nolimits_i x_i(\boldsymbol{v}) \leq 1 \qquad \forall\ \boldsymbol{v} \in \mathcal{V}^n \qquad \text{(single-item feasibility)}$$

$$x_i(v_i, \boldsymbol{v}_{-i}) \leq x_i(v_i', \boldsymbol{v}_{-i}) \qquad \forall\ i,\ v_i,\ v_i' > v_i,\ \boldsymbol{v}_{-i} \quad \text{(monotonicity)}$$

$$x_i(\boldsymbol{v}) \geq 0 \qquad \forall\ i,\ \boldsymbol{v} \qquad \text{(non-negativity)}$$

[1] The left-hand side of each approximation constraint is equivalently described as *expected sum-total virtual value* because:

$$\int_{\mathcal{V}^n} \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v}) \right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} = \mathbf{E}_{\boldsymbol{v} \sim F,\ \boldsymbol{x}} \left[\phi_i^F(\boldsymbol{v}) \right]$$

We expand the expectation to its integral form because it gives an explicit description of the constraint in terms of the variables of the program.

(From now on, we assume non-negativity without writing it.) Starting from Linear Program 1, we provide a sequence of modifications in order to reprove Theorem 10 for linear prior independent algorithm design problems. The goal from here is to obtain a linear program for which we can assign weights of a dual blend to its variables as a feasible solution, and then analysis of an identifiable bound on the objective function implies the desired inequality: $\alpha^{\mathcal{F}} \geq {}^{\text{opt}_{n,2}}/_{\text{opt}_{n,1}}$.

The key observation for the first modification step is that *without loss* we can add to the program a constraint of *non-super-optimality*, and not only with respect to $\mathcal{F}$ but with respect to all distributions (represented by the class $\mathcal{F}^{\text{all}}$):

Linear Program 2 (The Appended Program)**.** *This program adds a* non-super-optimality *constraint to Linear Program 1 without loss. We give only the new constraint:*

$$\int_{\mathcal{V}^n} \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v}) \right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \leq \text{OPT}_F(F) \quad \forall\ F \in \mathcal{F}^{\text{all}} \quad \text{(non-super-optimality)}$$

The new constraint is without loss because no prior independent algorithm can do strictly better given F than the optimal algorithm OPT_F which knows F (Fact 16), and further, this is true regardless of any restrictions imposed on the distribution by the class $\mathcal{F}$. The next step is to in fact *drop all of the setting-specific constraints* within the linear program, giving us a program whose optimal value $\check{\alpha}_{\text{lax}}^{\mathcal{F}}$ upper bounds the previous program (i.e., the maximum may now be larger):

Linear Program 3 (The Appended-Relaxed Program)**.** *This program relaxes Linear Program 2 by dropping its mechanism-design-setting-specific constraints. We are left with:*

$$\text{(A.2)} \qquad \check{\alpha}^{\mathcal{F}} \leq \check{\alpha}^{\mathcal{F}}_{\text{lax}} = \max_{\boldsymbol{x},\ \check{\alpha}} \check{\alpha} \ \leq 1$$

s.t.

$$\int_{\mathcal{V}^n} \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v}) \right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \geq \check{\alpha} \cdot \text{OPT}_F(F) \quad \forall\ F \in \mathcal{F} \qquad \text{(approximation)}$$

$$\int_{\mathcal{V}^n} \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v}) \right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \leq \text{OPT}_F(F) \qquad \forall\ F \in \mathcal{F}^{\text{all}} \quad \text{(non-super-optimality)}$$

Of course, the bound $\check{\alpha}^{\mathcal{F}} \leq \check{\alpha}^{\mathcal{F}}_{lax}$ holds if and only $\alpha^{\mathcal{F}} = 1/\check{\alpha}^{\mathcal{F}} \geq 1/\check{\alpha}^{\mathcal{F}}_{lax}$, therefore $1/\check{\alpha}^{\mathcal{F}}_{lax}$ is a lower bound on the prior independent approximation factor of the original problem. The next step is to notice that without coordinate-specific constraints on the variables $\boldsymbol{x}$, each parenthetical term may in fact be replaced by a pair of variables $\tilde{A}(\boldsymbol{v})$ and $\tilde{B}(\boldsymbol{v})$, which together represent a measure of algorithm pseudo-performance on input $\boldsymbol{v}$ that is locally unconstrained. (We use $\tilde{A}(\boldsymbol{v}) - \tilde{B}(\boldsymbol{v})$ everywhere, effectively as one variable that may be positive or negative.)

The only remaining constraint on the assignment of the variables $\tilde{\boldsymbol{A}} = \{\tilde{A}(\boldsymbol{v}) \ :\ \boldsymbol{v} \in \mathcal{V}^n\}$ and $\tilde{\boldsymbol{B}} = \{\tilde{B}(\boldsymbol{v}) \ :\ \boldsymbol{v} \in \mathcal{V}^n\}$ is: the expectation of pseudo-performance on any distribution F must not exceed the optimal algorithm given F (which retains all constraints), i.e., per the non-super-optimality constraint which remains.

We make one more modification to the linear program in this step: we multiply its objective by a positive constant κ. For now, we leave κ to-be-defined but we will use it later to help short-cut the analysis. This modification is obviously benign in terms of

the argmax. (Note that if we want to ignore κ, we set $\kappa = 1$ and the objective line here satisfies $\breve{\alpha}^{\mathcal{F}} \leq \breve{\alpha}^{\mathcal{F}}_{lax} = \max_{\tilde{\boldsymbol{A}},\ \tilde{\boldsymbol{B}},\ \breve{\alpha}} \breve{\alpha}\ \leq 1$.)

Linear Program 4 (The Appended-Relaxed-Simplified Program (ARS))**.** *This program simplifies the variable-space of Linear Program 3 without loss by replacing the original allocation variables* $\boldsymbol{x}$ *with algorithm pseudo-performance variables* $\tilde{\boldsymbol{A}}$ *and* $\tilde{\boldsymbol{B}}$*, i.e., by substituting* $\tilde{A}(\boldsymbol{v}) - \tilde{B}(\boldsymbol{v}) = \left(\sum_i \phi_i^F(\boldsymbol{v}) \cdot x_i(\boldsymbol{v})\right)$*:*[2]

$$\kappa \cdot \breve{\alpha}^{\mathcal{F}} \leq \kappa \cdot \breve{\alpha}^{\mathcal{F}}_{\text{lax}} = \max_{\tilde{\boldsymbol{A}},\ \tilde{\boldsymbol{B}},\ \breve{\alpha}} \kappa \cdot \breve{\alpha}\ \leq \kappa \tag{A.3}$$

s.t.

$$\int_{\mathcal{V}^n} \left(\tilde{A}(\boldsymbol{v}) - \tilde{B}(\boldsymbol{v})\right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \geq \breve{\alpha} \cdot \text{OPT}_F(F) \quad \forall\ F \in \mathcal{F} \qquad \text{(approximation)}$$

$$\int_{\mathcal{V}^n} \left(\tilde{A}(\boldsymbol{v}) - \tilde{B}(\boldsymbol{v})\right) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \leq \text{OPT}_F(F) \qquad \forall\ F \in \mathcal{F}^{\text{all}} \quad \text{(non-super-optimality)}$$

At this point, *no structure of the original mechanism design setting remains in Linear Program 4* – thus, any algorithm setting may continue from this point if its prior independent program can drop setting-specific constraints and write pseudo-performance as a single variable (because also: any algorithm setting may add non-super-optimality).

We now convert Linear Program 4 to its dual program. (The value of the dual program is at least the value of the primal program and we write this into the objective line.) Each constraint-line of the dual is assigned an intuitive label to describe its behavior within the program; and the dual has the following variables (one per primal constraint):

[2] The left-hand side of each approximation constraint is equivalently described as *expected pseudo-performance*, cf. the explanation in footnote 1 on page 191.

approximation:: $\omega_F \; \forall \; F \in \mathcal{F}$

non-super-optimality:: $o_F \; \forall \; F \in \mathcal{F}^{\text{all}}$

Linear Program 5 (The Dual of the ARS Program)**.**

$$\kappa \cdot \check{\alpha}^{\mathcal{F}}_{\text{lax}} \leq \min_{\boldsymbol{\omega},\, \boldsymbol{o}} \int_{\mathcal{F}^{\text{all}}} o_F \cdot \text{OPT}_F(F) \; dF \tag{A.4}$$

s.t.

$$\int_{\mathcal{F}} \omega_F \cdot \text{OPT}_F(F) \; dF \geq \kappa \qquad (\text{for } \check{\alpha}) \qquad \text{(scale-setting)}$$

$$\int_{\mathcal{F}^{\text{all}}} (o_F - \omega_F) \cdot f^n(\boldsymbol{v}) \; dF \geq 0 \qquad \forall \; \boldsymbol{v} \in \mathcal{V}^n \; (\text{for } \tilde{A}(\boldsymbol{v})) \qquad \text{(density-matching-A)}$$

$$\int_{\mathcal{F}^{\text{all}}} (-o_F + \omega_F) \cdot f^n(\boldsymbol{v}) \; dF \geq 0 \qquad \forall \; \boldsymbol{v} \in \mathcal{V}^n \; (\text{for } \tilde{B}(\boldsymbol{v})) \qquad \text{(density-matching-B)}$$

The final point is to choose (a) dual arguments $\boldsymbol{\omega} = \{\omega_F \; : \; F \in \mathcal{F}\}$ and $\boldsymbol{o} = \{o_F \; : \; F \in \mathcal{F}^{\text{all}}\}$ such that *these variables describe a finite-weight dual blend* with $\boldsymbol{\omega}$ the weights for distributions in a blend $\delta_2 \in \mathcal{F}$ and $\boldsymbol{o}$ the weights for distributions in a blend $\delta_1 \in \mathcal{F}^{\text{all}}$; and (b) choose $\kappa = \int_F \omega_F \cdot \text{OPT}_F(F) dF$. Making all of these substitutions into Linear Program 5, we see that this assignment of dual arguments gives a feasible solution to the dual constraints:

- this assignment meets density-matching with equality by definition of a dual blend which is in fact a necessary structure to satisfy both constraints (and further, note that equality is necessary per complimentary slackness wherever we need to allow strictly positive assignment to the corresponding primal variables $\tilde{A}(\boldsymbol{v})$ and $\tilde{B}(\boldsymbol{v})$);

- and, it meets scale-setting with equality by choice of κ (which makes it is easy to verify).

An assignment to variables that satisfies all constraints gives an upper bound on the optimal value of a minimization LP. Thus, substituting, re-arranging the objective of Linear Program 5, and incorporating relationships stated previously gives

$$\check{\alpha}^{\mathcal{F}}_{\text{lax}} \leq \frac{\int_{\mathcal{F}^{\text{all}}} o_F \cdot \text{OPT}_F(F)\ dF}{\int_{\mathcal{F}} \omega_F \cdot \text{OPT}_F(F)\ dF} = \frac{\text{opt}_{n,1}}{\text{opt}_{n,2}}$$

$$\Downarrow$$

$$\frac{\text{opt}_{n,2}}{\text{opt}_{n,1}} \leq \frac{1}{\check{\alpha}^{\mathcal{F}}_{\text{lax}}} \leq \frac{1}{\check{\alpha}^{\mathcal{F}}} = \alpha^{\mathcal{F}}$$

which finishes the re-proof of Theorem 10 for linear algorithm settings.

A.2. Supporting Work for Quadratics-versus-Uniforms Residual Surplus Gap

This section provides material to support Section 4.5.4.2. The presentation generally assumes its terms, assumptions, and context while only restating the most important definitions here.

- Appendix A.2.1 justifies the residual surplus curve of $\overleftarrow{\text{Qud}}^{h'}_1$, stated previously in equation (4.24).
- Appendix A.2.2 defines $M_{o_{\text{pm}}}$ and calculates its residual surplus on $\overleftarrow{\text{Qud}}^{h'}_1$.
- Appendix A.2.3 calculates the residual surplus of the 2-lottery on $\overleftarrow{\text{Qud}}^{h'}_1$ to show that $M_{o_{\text{pm}}}$ has better performance, and concludes that we have strictly $\text{opt}_{2,2} > \text{lb}_{2,2} > \text{opt}_{2,1}$.

- Appendix A.2.4 gives further supporting analysis for completeness, for example it explains the choice to use $\text{lb}_{2,2}$ rather than calculating $\text{opt}_{2,2}$ and describes the un-ironed residual surplus curve for $\overleftarrow{\text{Qud}}_1^{h'}$ and the design of $M_{o_{\text{pm}}}$.

One critical assumption that we do repeat is $h \geq 8.56$. Copying equation (4.25) for local reference, we have

$$\text{lb}_{2,2} = 1 \cdot M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) \quad + \quad \left[\int_1^h \frac{2}{z} \cdot \text{LOT}_2(\overleftarrow{\text{Qud}}_z^{h'})\ dz\right]$$

A.2.1. The Residual Surplus Curve for the Quadratic on $[1, h]$

This section explains the un-ironed residual surplus curve for $\overleftarrow{\text{Qud}}_{z=1}^{h'}$. Recall, the CDF of the specific distribution $\overleftarrow{\text{Qud}}_{z=1}^{h'}$ is given by $\overleftarrow{\text{Qud}}_1^{h'}(x) = 1 - 1/x$ on $x \in [1, h]$, and $\overleftarrow{\text{Qud}}_1^{h'}(x) = 1$ for $x \geq h$. I.e., the CDF has a vertical line segment at $x = h$ where it maps to the set-value $[1 - 1/h, 1]$, because the distribution is top-truncated with a point mass at h. With explanation to follow, we restate the residual surplus curve:

$$R_{\overleftarrow{\text{Qud}}_1^{h'}}(q) = \begin{cases} 0 & \text{for } q \in [0, 1/h] \\ \ln(q \cdot h) & \text{for } q \in [1/h, 1) \\ [\ln h, 1 + \ln h] & \text{for } q = 1 \end{cases}$$

This residual surplus curve $R_{\overleftarrow{\text{Qud}}_1^{h'}}$ is illustrated in Figure A.1. We proceed to justify this equation.

The residual surplus curve $R_{\overleftarrow{\text{Qud}}_1^{h'}}(\cdot)$ is defined piece-wise including (a) a piece that is identically 0 for quantiles $q \in [0, 1/h]$ (from top-truncation); and (b) a piece that is a vertical line segment at $q = 1$ of length $z = 1$ (from consideration of price-posting in

$[0, 1]$). The lower endpoint of this vertical line segment is identified by the residual surplus curve at quantile $q = 1$ corresponding to the lower bound on the distribution's domain in value space, in this case value $v = 1$. We will show next that the height of the residual surplus curve corresponding to this point is in fact $\ln h$. The expected residual surplus from one agent value drawn from $\overleftarrow{\text{Qud}}_1^{h'}$ with a posted price of 1 is:

$$\text{AP}_1(\overleftarrow{\text{Qud}}_1^{h'}) = \mathbf{E}_{v \sim \overleftarrow{\text{Qud}}_1^{h'}}[v - 1] = \left[\frac{1}{h} \cdot h + \int_1^h \frac{1}{z^2} \cdot z \; dz\right] - 1 = \ln h$$

Thus, the exact description of the vertical line segment is the set-valued output range of $[\ln h, 1 + \ln h]$ at input $q = 1$. We now show that generally, the residual surplus curve for quantiles $q \in [1/h, 1]$ is described by $R_{\overleftarrow{\text{Qud}}_1^{h'}}(q) = \ln(q \cdot h)$. (Combined with line-segment-pieces (a) and (b) which have already been explained, this completes the description of the residual surplus curve of $\overleftarrow{\text{Qud}}_1^{h'}$.)

Given the distribution $\overleftarrow{\text{Qud}}_1^{h'}$ and the residual surplus objective, as functions of value inputs $v \in [1, h]$, and then as functions of quantiles $q \in [1/h, 1]$, the virtual value function and quantile/value functions are

$$\phi^{\overleftarrow{\text{Qud}}_1^{h'}}(v) = \frac{1 - \overleftarrow{\text{Qud}}_1^{h'}(v)}{\overleftarrow{\text{qud}}_1^{h'}(v)} = \frac{1 - (1 - 1/v)}{1/v^2} = v$$

$$Q_{\overleftarrow{\text{Qud}}_1^{h'}}(v) = 1 - \overleftarrow{\text{Qud}}_1^{h'}(v) = 1 - (1 - 1/v) = 1/v$$

$$\phi^{\overleftarrow{\text{Qud}}_1^{h'}}(q) = 1/q, \qquad V_{\overleftarrow{\text{Qud}}_1^{h'}}(q) = 1/q$$

Using the identity $\phi^F(q) = R'_F(q)$ from Fact 4 and then integrating the function $\phi^{\overleftarrow{\text{Qud}}_1^{h'}}$ just given ($\int_{1/h}^q 1/z \; dz$), we confirm the case for $q \in [1/h, 1]$ of equation (4.24) that states:

the residual surplus curve is described by $R_{\overleftarrow{\text{Qud}}_1^{h'}}(q) = \ln(q \cdot h)$ on this sub-domain of quantile space.

A.2.2. Definition and Residual Surplus of the "Two-piece-iron" Mechanism

The goal of this section is to define the special mechanism $M_{o_{\text{pm}}}$ as needed by Section 4.5.4.2 for the calcuation of $\text{lb}_{2,2}$, and calculate its expected residual surplus given 2 agents with values drawn i.i.d. from $\overleftarrow{\text{Qud}}_1^{h'}$ within the deferred proof of Lemma 11. Note, the presentation is within the context of proving that the residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ is strictly more than the residual surplus $\text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'})$ of the 2-lottery.

For use in this section, we need to extend the definition of an *ironed residual surplus curve* to allow arbitrary ironing.[3] Let $\mathcal{Q}$ be any *possibly-non-optimal* set of non-overlapping ranges to be ironed (thus each element of $\mathcal{Q}$ is a subset of $[0,1] \in \mathbb{R}$). Given a distribution F, define $\bar{R}_{F,\mathcal{Q}}$ to be the residual surplus curve given base-distribution F and its residual surplus curve R_F, and then accounting for $\mathcal{Q}$ as the given list of ironed regions.

Call the mechanism $M_{o_{\text{pm}}}$ the *two-piece-iron* mechanism. Its definition depends on a critical quantile $q^* = {}^{e}/_{h}$. Specifically motivated by $\overleftarrow{\text{Qud}}_1^{h'}$, the mechanism irons two regions: (1) large values below and (2) small values above this quantile. Ultimately, the mechanism $M_{o_{\text{pm}}}$ runs an *ironed* second price auction on the two inferred types (one common type for each ironed region). For illustration of $M_{o_{\text{pm}}}$ applied to $R_{\overleftarrow{\text{Qud}}_1^{h'}}$, see Figure A.1. Formally, define the following sets of ironed ranges which will be used by our subsequent analysis:

- $\mathcal{Q}^* = \{[0, q^* = {}^{e}/_{h}]\}$; corresponding to value-range $[{}^{h}/_{e}, h]$;

[3] The original definition for an ironed residual surplus curve was derived from the definition of an ironed revenue curve, specifically in the context of *optimal ironing*. See page 44.

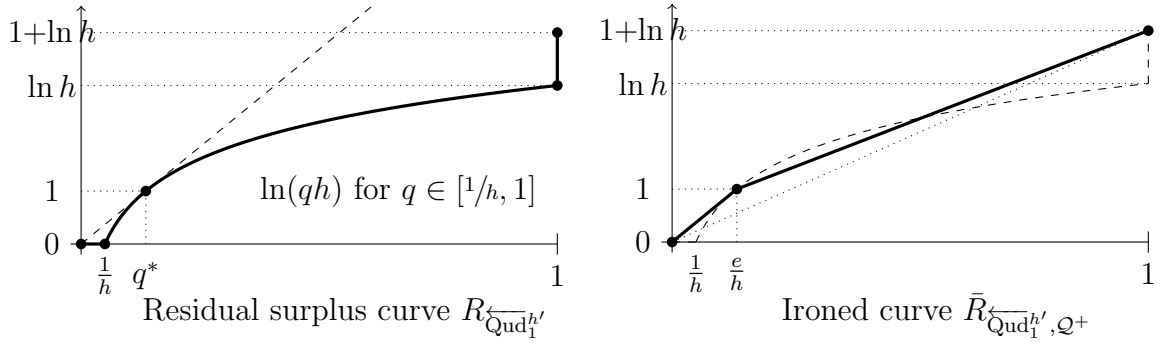

Figure A.1. Residual Surplus Curve and Ironing of $[1, h]$ Truncated-Quadratic
We illustrate the effect of ironing on our truncated quadratic distribution (defined by equation (4.24) below). The left-hand side shows the residual surplus curve $R_{\overleftarrow{\text{Qud}}_1^{h'}}$. The right-hand side shows the ironing of $R_{\overleftarrow{\text{Qud}}_1^{h'}}$ according to the description of mechanism $M_{o_{\text{pm}}}$, resulting in $\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}$. The example is graphed for $h = 20$. Note: $q^* = {}^{e}/_{h}$; the left figure indicates that ironing the interval $q \in [0, {}^{e}/_{h}]$ is tangent to the original curve and therefore optimal on this region; the right figure makes clear that the point $({}^{e}/_{h}, 1)$ is above the line that represents ironing everywhere (for sufficiently large h).

- $\mathcal{Q}^+ = \{[0, q^*],\ [q^*, 1]\}$; corresponding to value-ranges $[{}^{h}/_{e}, h]$ and $[0, {}^{h}/_{e}]$ (where identifying the lower bound of the second value range to be 0 is a necessary distinction because $R_{\overleftarrow{\text{Qud}}_1^{h'}}$ is set-valued at $q = 1$);[4]
- $\mathcal{Q}^1 = \{[0, 1]\}$ corresponding to value-range $[0, h]$ (which is effectively the lottery).

Definition 37. *Define the* two-piece-iron *mechanism $M_{o_{pm}}$ for $n = 2$ agents to be the* ironed second price auction *which respects the ironed ranges $\mathcal{Q}^+ = \{[0, q^*],\ [q^*, 1]\}$.*

Equivalently, $M_{o_{pm}}$ irons the regions of value space $[{}^{maxval}/_{e}, h]$ and $[0, {}^{h}/_{e}]$ and runs the second price auction on these two inferred types.

Thus, when agent values are drawn i.i.d. from $\overleftarrow{\text{Qud}}_1^{h'}$, the residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ may be calculated using the ironed residual surplus curve $\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}(\cdot)$. (See Figure A.1.)

[4] We get $v^* = {}^{h}/_{e}$ as the solution of $q^* = {}^{e}/_{h} = 1 - \overleftarrow{\text{Qud}}_1^{h'}(v^*) = 1 - (1 - {}^{1}/_{v^*})$.

In fact for $h \geq 8.56$, ironing the region $[0, q^* = e/h]$ is optimal given the underlying distribution $\overleftarrow{\text{Qud}}_1^{h'}$; and ironing the region $[e/h, 1]$ is strictly suboptimal (for intuition for this, see Figure A.2).

We conclude this section with the deferred proof of Lemma 11, which depends on Lemma 18 below as an extension of Lemma 1 (Dhangwatnotai et al., 2015).

Lemma 11. *The residual surplus of mechanisms $M_{o_{pm}}$ and LOT_2 given 2 agents with values drawn i.i.d. from $\overleftarrow{\text{Qud}}_1^{h'}$ are calculated as*

$$M_{o_{pm}}(\overleftarrow{\text{Qud}}_1^{h'}) = \frac{((2+\ln h)h - (1+\ln h)e)}{h}$$
$$\text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'}) = 1 + \ln h$$

Proof. Using Lemma 18 below (which extends Dhangwatnotai et al. (2015) to allow ironing and any auction objective, in our case residual surplus) and the definition of $M_{o_{\text{pm}}}$, the residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ is calculated as twice the area under the ironed residual surplus curve $\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}$. This area is calculated from

$$\begin{aligned}
\text{area under } \bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+} &= \text{area in quantile range } [0, e/h] + \text{area in quantile range } [e/h, 1] \\
&= 1/2 \cdot \bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}(e/h) \cdot \left(\frac{e}{h} - 0\right) \\
&\quad + \frac{1}{2} \cdot \left(\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}(e/h) + \bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^+}(1)\right)\left(1 - \frac{e}{h}\right) \\
&= \frac{1}{2} \cdot 1 \cdot \frac{e}{h} + \frac{1}{2} \cdot (1 + (1 + \ln h)) \cdot \frac{h-e}{h} \\
&= \frac{1}{2} \cdot \frac{(2+\ln h)h - (1+\ln h)e}{h}
\end{aligned}$$

where the residual surplus at the endpoints of the ironed ranges – namely, quantiles $q = {}^{e}/_{h}$ and $q = 1$ – are from the definition of $R_{\overleftarrow{\text{Qud}}_1^{h'}}$ (equation (4.24) earlier). Therefore the residual surplus of the mechanism is $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) = {}^{((2+\ln h)h-(1+\ln h)e)}/_{h}$.

Using Lemma 18 and equation (4.24) again – this time applied to ironed revenue curve $\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'},\mathcal{Q}^1}$, i.e., with respect to the lottery's ironing $\mathcal{Q}^1$ – the residual surplus of the 2-lottery is $\text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'}) = 1 + \ln h$. □

For completeness, we prove the extension of Lemma 1 (Dhangwatnotai et al., 2015) to apply both (a) for an arbitrary auction objective, and (b) to allow arbitrary ironing. In the case of ironing, the SPA must be interpreted as treating each ironed range as a single value space type – it allocates all agents in an ironed range uniformly.[5] Define this mechanism as the *Ironed Second Price Auction*.

Let $\mathcal{Q}$ be a set of ironed ranges (in quantile space; as defined on page 199), and let $\bar{\phi}^{F,\mathcal{Q}}$ be the ironed virtual value function given an underlying distribution F that has been ironed on ranges according to $\mathcal{Q}$.

Lemma 18. *In i.i.d. two-agent single-item settings given distribution F,* for any auction objective *let R_F be the performance curve in quantile space and $\bar{R}_{F,\mathcal{Q}}$ be an ironed performance curve given a set of ironed ranges $\mathcal{Q}$.*

The expected performance of the ISPA subject to $\bar{R}_{F,\mathcal{Q}}$ – assuming uniform allocation to agents within each ironed range as if the range was one type – is twice the area under the curve $\bar{R}_{F,\mathcal{Q}}$.

[5] Recall, this type of treatment is a necessary condition to apply the technique of ironing – see the introduction of ironing in discussion on page 43 and its conditional use in Theorem 3.

Proof. We note the following up front. Without loss of generality, our 2 agents have ordered values $v_{(1)} \geq v_{(2)}$, equivalently, ordered quantiles $q_{(1)} \leq q_{(2)}$. The ISPA mechanism of the statement is symmetric.

The technique of this proof is to sum up the performance of the ISPA mechanism (for arbitrary ironing) by calculating expected performance over the distribution of the smaller quantile-order-statistic $q_{(1)}$. To outline, we: identify this conditional performance as a function of virtual value; and then insert this quantity into the existing proof of Lemma 1 (Dhangwatnotai et al., 2015).

For agents labeled according to order statistic i, define $x_{(i)} = x^{\text{ISPA}}_{(i)}(V_F(q_{(i)}), V_F(q_{(j\neq i)}))$. Define $\mathcal{R}(q_1)$ to be the expectation of the winning agent's virtual value conditioned on the smaller quantile being q_1. (Note, the winner is not necessarily the agent $i = 1$.) Thus, we have:

$$\mathcal{R}(q_1) = \mathbf{E}_{q_2 \sim \text{Ud}_{q_1,1}} \left[x_{(1)} \cdot \bar{\phi}^{F,\mathcal{Q}}(q_1) + x_{(2)} \cdot \bar{\phi}^{F,\mathcal{Q}}(q_2) \right] = \bar{\phi}^{F,\mathcal{Q}}(q_1) = \bar{R}'_{F,\mathcal{Q}}(q_1)$$

We get the second equality here because the following holds for all inputs q_1 into function $\mathcal{R}$: *pointwise within the expectation*: either $x_{(1)} = 1$, or otherwise $x_{(1)} + x_{(2)} = 1$ and $\bar{\phi}^{F,\mathcal{Q}}(q_1) = \bar{\phi}^{F,\mathcal{Q}}(q_2)$. In any case, $\mathbf{E}_{q_2 \sim \text{Ud}_{q_1,1}} \left[x_{\{1\}} \cdot \bar{\phi}^{F,\mathcal{Q}}(q_1) + x_{\{2\}} \cdot \bar{\phi}^{F,\mathcal{Q}}(q_2) \right] = \bar{\phi}^{F,\mathcal{Q}}(q_1)$. We can now effectively implement the proof of Lemma 1 which did not accommodate ironing and which was stated for the specific objective of revenue.

Let $\text{osd}(q) = 2(1-q)$ be the density function of the smallest order-statistic $q_{(1)}$ out of 2 agents' quantiles drawn i.i.d. from $\text{Ud}_{0,1}$. (Note, for simplicity, we dropped all parameters from the distribution name 'osd.')

We are now prepared to evaluate ISPA(F) using $\bar{\mathcal{R}}_{F,\mathcal{Q}}$ and osd:

$$\begin{aligned}
\text{ISPA}(F) &= \int_0^1 \text{osd}(q) \cdot \bar{R}'_{F,\mathcal{Q}}(q)\ dq \\
&= \int_0^1 2 \cdot 1 \cdot (1-q) \cdot \bar{R}'_{F,\mathcal{Q}}(q)\ dq \\
&= 2 \cdot \bar{R}_{F,\mathcal{Q}}(1) - 2 \cdot \int_0^1 q \cdot \bar{R}'_{F,\mathcal{Q}}(q)\ dq \\
&= 2 \cdot \bar{R}_{F,\mathcal{Q}}(1) - 2 \cdot \left[q \cdot \bar{R}_{F,\mathcal{Q}}(q)\right]_0^1 + 2 \cdot \int_0^1 \bar{R}_{F,\mathcal{Q}}(q)\ dq = 2 \cdot \int_0^1 \bar{R}_{F,\mathcal{Q}}(q)\ dq \quad \square
\end{aligned}$$

A.2.3. The "Two-piece-iron" Mechanism is Sufficient for a Lower Bound

In this section we show that the residual surplus $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) = {((2+\ln h)h - (1+\ln h)e)}/{h}$ of Lemma 11 is strictly worse than optimal and strictly better than the 2-lottery given $h \geq 8.56$. The main goal is to give the deferred proof of Lemma 12 which states that $\text{opt}_{2,2} > \text{lb}_{2,2} > \text{opt}_{2,1}$.

Lemma 19. *Given $h \geq 8.56$. The mechanism $M_{o_{pm}}$ is strictly sub-optimal:* $\text{OPT}_{\overleftarrow{\text{Qud}}_1^{h'}}(\overleftarrow{\text{Qud}}_1^{h'}) > M_{o_{pm}}(\overleftarrow{\text{Qud}}_1^{h'})$*; and the mechanism $M_{o_{pm}}$ strictly dominates the lottery:* $M_{o_{pm}}(\overleftarrow{\text{Qud}}_1^{h'}) > \text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'})$.

Proof. First, we prove the lower bound on $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ in the lemma statement. We use $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) = {((2+\ln h)h - (1+\ln h)e)}/{h}$ and $LOT_2(\overleftarrow{\text{Qud}}_1^{h'}) = 1 + \ln h$ from Lemma 11.

We reduce the condition that the difference is positive:

$$M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) - \text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'}) = \frac{(2+\ln h)h - (1+\ln h)e}{h} - (1+\ln h)$$
$$= \frac{h - (1+\ln h)e}{h} > 0$$
$$\Leftrightarrow \quad h - (1+\ln h)e > 0 \tag{A.5}$$

Treating the left-hand side of the inequality in line (A.5) as a function of h, it is negative and decreasing for $h \in [1, e]$, it is increasing for all $h > e$, and it has a 0 within the range $h \in [8.55, 8.56]$ (and then is positive for $h > 8.56$ because it is increasing). Therefore the lower bound $h \geq 8.56$ is sufficient for the lowerbound on $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$.

For the upper bound in the lemma statement, the analysis and discussion surrounding equation (A.9) below in Appendix A.2.4 are sufficient to show that the ironing by $M_{o_{\text{pm}}}$ of the range $[e/h, 1]$ is strictly suboptimal (given $h \geq 8.56$, which infers the lower endpoint of this range is upper bounded as $e/h < 0.31$). It is dominated specifically in comparison to ironing the quantile-space upward-closed range $[q^{\&}, 1]$ for the optimal value of $q^{\&} \in [0.31, 0.32]$ defined and proved in Appendix A.2.4. □

Lemma 12. *Given* $\text{opt}_{2,1}$ *and* $\text{opt}_{2,2}$ *resulting from the finite-weight Quadratics-versus-Uniforms dual blends (along with the rest of the local assumptions of this section), and* $\text{lb}_{2,2}$ *as defined in equation* (4.25)*. Then we have*

$$\text{opt}_{2,2} > \text{lb}_{2,2} > \text{opt}_{2,1}$$

Proof. Lemma 19 states that if $h \geq 8.56$, then (a) $\text{OPT}_{\overleftarrow{\text{Qud}}_1^{h'}}(\overleftarrow{\text{Qud}}_1^{h'}) > M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'})$ and (b) $M_{o_{\text{pm}}}(\overleftarrow{\text{Qud}}_1^{h'}) > \text{LOT}_2(\overleftarrow{\text{Qud}}_1^{h'})$. Following directly from these and from definitions

we have:

$$\begin{aligned}
& \mathrm{opt}_{2,2} = && o_{\mathrm{pm}} \cdot \mathrm{OPT}_{\overleftarrow{\mathrm{Qud}}_1^{h'}}(\overleftarrow{\mathrm{Qud}}_1^{h'}) && + && \left[\int_1^h o_z \cdot \mathrm{OPT}_{\overleftarrow{\mathrm{Qud}}_1^{h'}}(\overleftarrow{\mathrm{Qud}}_z^{h'})\right] \\
> \quad & \mathrm{lb}_{2,2} = && o_{\mathrm{pm}} \cdot M_{o_{\mathrm{pm}}}(\overleftarrow{\mathrm{Qud}}_1^{h'}) && + && \left[\int_1^h o_z \cdot \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}_z^{h'})\right] \\
> \quad & \mathrm{opt}_{2,1} = && o_{\mathrm{pm}} \cdot \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}_1^{h'}) && + && \left[\int_1^h o_z \cdot \mathrm{LOT}_2(\overleftarrow{\mathrm{Qud}}_z^{h'})\right]
\end{aligned}$$

where the definition of $\mathrm{opt}_{2,1}$ may use the weights $\boldsymbol{o}$ and the distributions in δ_2 rather than its original definition which respectively used $\boldsymbol{\omega}$ and δ_1. This last point holds because it runs the constant lottery mechanism on all inputs anyway and $\delta_1^2 = g = \delta_2^2$. □

A.2.4. Complexities and Technicals of the Quadratics Residual Surplus Blend

The definition of $M_{o_{\mathrm{pm}}}$ was given in Definition 37. For completeness, here we build up the motivation for it – effectively reverse-engineering it to be sufficient for the residual surplus gap which is our goal (of equation (4.27) of Section 4.5.4.3). Here is an outline of this section:

(1) identify technical difficulties of residual surplus curves $R_{\overleftarrow{\mathrm{Qud}}_z^{h'}}(\cdot)$ for arbitrary z and motivate the relaxation to $\mathrm{lb}_{2,2}$ and the assumption of $h \geq 8.56$ for simplicity;

(2) given $h \geq 8.56$, determine that $[0, {}^{e}/_{h}]$ is an element of the set of optimal ranges to iron, by analyzing slopes of possible quantile-downward-closed ironed ranges;

(3) from slopes of respective ironed ranges of $M_{o_{\mathrm{pm}}}$ and LOT_2, re-confirm the statement of Lemma 19 with dependence on equation (A.5) in its proof;

(4) identify the optimal quantile-downward-closed range to iron, which is independent of h.

(1) Explanation of the choice to simplify from $\mathbf{opt}_{2,2}$ to $\mathbf{lb}_{2,2}$ and $h \geq 8.56$. This is a discussion of the difficulties of revenue curves for the class of Quadratics $\overleftarrow{\text{Qud}}_z^{h'}$ (with positive weight in δ_2). Generalizing equation (4.24), residual surplus curves for Quadratics and arbitrary z are described by:

$$R_{\overleftarrow{\text{Qud}}_z^{h'}}(q) = \begin{cases} 0 & \text{for } q \in [0, z/h] \\ z \ln(q \cdot \frac{h}{z}) & \text{for } q \in [z/h, 1) \\ [z \ln \frac{h}{z}, z + z \ln \frac{h}{z}] & \text{for } q = 1 \end{cases} \tag{A.6}$$

The first challenge is that for "large" $z \to h$, the optimal mechanism for $\overleftarrow{\text{Qud}}_z^{h'}$ is the lottery. There is a threshold for z above which this becomes true (see point (4) below and also Figure A.2 which illustrates the threshold-change in the ironing structure of the residual surplus curve).

We greatly simplify this complication as follows: with an assumption of $h \geq 8.56$, then for $z = 1$, the lottery mechanism for $\overleftarrow{\text{Qud}}_1^{h'}$ is strictly not optimal. The calculation of the lower bound quantity $\text{lb}_{2,2}$ uses the performance of the lottery on all other distributions, even though the lottery is sub-optimal for many of these distributions. Critically however, for "small" z – and specifically for $z = 1$ where there is a point mass o_{pm}, and relying on our assumption of "large" $h \geq 8.56$ – the optimal mechanism in response to distribution $\overleftarrow{\text{Qud}}_z^{h'}$ will <u>not</u> use the lottery to iron the entire region of values, and rather, a distinct mechanism is strictly preferred.

The calculation of expected performance for the optimal mechanism for $\overleftarrow{\text{Qud}}_z^{h'}$ is itself complicated. To simplify, we relax the optimal mechanism to $M_{o_{\text{opm}}}$ which irons on just

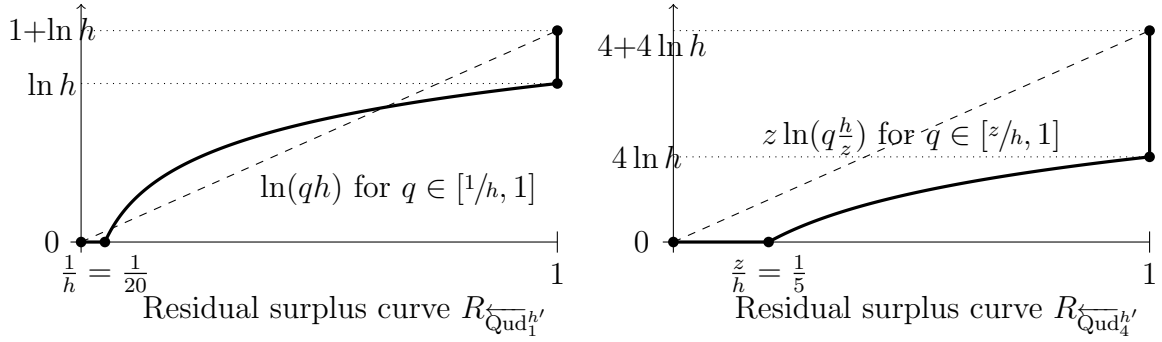

Figure A.2. Effect of $z \to h$ for $[z, h]$ Truncated-Quadratics
The dashed lines show the ironing of the lottery mechanism. As $z \to h$, there is a threshold beyond which the lottery mechanism becomes optional. Both graphics depict $h = 20$, but note their vertical scales are not equal. The left side uses $z = 1$ for which the lottery is not optimal. The right side uses $z = 4$ for which the lottery is optimal. Setting $h \geq 8.56$ is sufficient to guarantee that at least for the relevant corner case which has $z = 1$, the lottery is not optimal.

two regions, an optimal region over small quantiles and "all other large" quantiles (see Figure A.1 for illustration).

(2) The optimal region for partial (downward-closed) ironing. Now we find the *optimal* value $v \in [1, h]$ to iron all values above it, equivalently, the *optimal downward-closed region of quantile space*. This is a step towards motivating the definition of $M_{o_{\text{pm}}}$ as chosen in Definition 37.

Given the graph of the residual surplus curve, we find this optimal quantile range by considering a line segment with one endpoint as the origin ($q = 0, R_{\overleftarrow{\text{Qud}}_1^{h'}}(0) = 0$), and the other endpoint on the revenue curve at $R_{\overleftarrow{\text{Qud}}_1^{h'}}(q) = \ln(q \cdot h)$) for $q \in [1/h, 1]$. We search for the line segment of this type with largest slope. Directly from "change in y over change

in x," the slope function and its derivative are given by: [6]

$$\zeta(q) = \frac{\ln(q \cdot h)}{q} \quad \text{on } [1/h, 1] \tag{A.7}$$

$$q^2 \cdot \zeta'(q) = q \cdot \frac{1}{q} - (\ln(q \cdot h) \cdot 1 = 1 - \ln(q \cdot h)$$

such that the derivative shows that the slope function achieves its maximum at $q^* = e/h$:

$$\zeta'(q^*) = \zeta'(e/h) = 0 \tag{A.8}$$

The optimal range for ironing of small quantiles is $[0, e/h]$. Letting ζ^* be the optimal slope of the ironed region and recalling $\phi^{\overleftarrow{\text{Qud}}_1^{h'}}(q) = 1/q$, we have

$$\zeta^* = \zeta(q^*) = \zeta(e/h) = \frac{\ln(e/h \cdot h)}{e/h} = \frac{h}{e}$$

(3) Ironed slopes confirm equation (A.5) **is sufficient for Lemma 19.** Lemma 18 in Appendix B.4 shows that residual surplus is proportional to area under an *ironed residual surplus curve.* From the geometry of the ironed curves used respectively by $M_{o_{\text{pm}}}$ and LOT_2, it is clear that the question of which has larger area under the curve reduces to the question of which has the larger slope on the range $[0, e/h]$.

Consider comparing (a) the ironed slope $\zeta^* = h/e$ just calculated in (3) as used by $M_{o_{\text{pm}}}$; to (b) the slope $\zeta^1 = 1 + \ln h$ of the 2-lottery which irons everywhere. Per the reduction just mentioned, we have $M_{o_{\text{pm}}} > \text{LOT}_2$ if $h/e > 1 + \ln h$ which is equivalent to equation (A.5).

[6] In fact, by continuity of the derivative of the residual surplus curve in this region, the line segment will be tangent to the residual surplus curve if the optimal quantile is interior, i.e., in $(1/h, 1)$.

(4) The optimal region for partial (downward-closed) ironing. To end this section, we show that the optimal set of ironed ranges for $\overleftarrow{\text{Qud}}_1^{h'}$ is $\mathcal{Q}^{\&} = \{[0, {}^{e}\!/\!_{h}],\ [q^{\&}, 1]\}$, with $q^{\&}$ identified below, and the optimality of the set $\mathcal{Q}^{\&}$ self-evident from inspection of the geometry of the residual surplus curve $\bar{R}_{\overleftarrow{\text{Qud}}_1^{h'}}$. We do this by finding the quantile $q^{\&}$ at which the tangent line intersects the point $(1, 1 + \ln h)$ in the residual surplus curve graph (i.e., the top right corner point). The correct quantile $q^{\&}$ is the one – *observably independent of* h – that satisfies the equality:

$$\ln\left(q^{\&} \cdot h\right) + (1 - q^{\&}) \cdot \frac{1}{q^{\&}} = 1 + \ln h$$

$$\Leftrightarrow \qquad \ln q^{\&} + \frac{1}{q^{\&}} = 2 \tag{A.9}$$

which is a unique $q^{\&} \in ({}^{e}\!/\!_{8.56}, 0.32]$ (where we chose the lower endpoint as motivated by $h \geq 8.56$).

We end with the following notes. Naturally for $h^{\&}$ representing its threshold value in $[8.55, 8.56]$ at which $\text{lb}_{2,2} = \text{opt}_{2,1}$, this critical $h^{\&}$ sets $q^* = {}^{e}\!/\!_{h^{\&}} = q^{\&}$. Increasing h above $h^{\&}$, we still have that $q^{\&}$ is constant but $q^* = {}^{e}\!/\!_{h}$ is decreasing. Therefore the optimal ironing leaves the range $[{}^{e}\!/\!_{h}, q^{\&}]$ un-ironed for $h > h^{\&}$.

A.3. Monotone Hazard Rate Dual Blends

The goal of this section and the next one is to give further examples of dual blends. The examples are both motivated by optimal prior independent characterizations of single-item, 2-agent, truthful and scale-invariant mechanisms for the revenue objective. This section considers the prior independent class of MHR distributions $\mathcal{F}^{\text{mhr}}$ (Definition 10).

The optimal mechanism for this class is given below and this section constructs a finite-weight dual blend using its worst-case distribution to parameterize one side.

(Appendix A.4 will consider the larger prior independent class of regular distributions $\mathcal{F}^{\text{reg}}$. The optimal mechanism for this class was given in Theorem 5 and its worst-case distribution was a specific instance of a Triangle revenue curve distribution (Definition 16). Triangles are truncations of the class of constant negative virtual value distributions (CNVVs, Definition 15; which are themselves a sub-class of the Shifted-quadratics). Appendix A.4 will construct finite-weight dual blends using the class of CNVVs.)

Allouah and Besbes (2018) showed that the SPA is the optimal prior independent mechanism for the class of MHR distributions $\mathcal{F}^{\text{mhr}}$. The worst-case distribution for the problem is a *truncation* of the standard exponential distribution Exd_1 which has CDF $\text{Exd}_1(x) = 1 - e^{-x}$ on $[0, \infty)$ and PDF $\text{exd}_1(x) = e^{-x}$. The quantile $1/e$ is the monopoly reserve quantile of Exd_1 (without truncation; see Figure A.3). The optimal adversarial truncation is identified by Allouah and Besbes (2018) at quantile $q^*_m \approx 0.42659$, which is the monopoly quantile of the new distribution post-truncation (as a result of regularity and $q^*_m > 1/e$). In value space, the truncation point – equivalently, the monopoly price – is $V_{\text{Exd}_1}(q^*_m) \approx 0.851935$. Therefore the worst-case distribution is $\overleftarrow{\text{Exd}}_1^{.852'}$ (within which we allow the approximation 0.852 of $V_{\text{Exd}_1}(q^*_m)$ to represent the exact optimal truncation).

Theorem 20 (Allouah and Besbes, 2018)**.** *Given a single item, 2-agent auction with a revenue objective, the optimal truthful, scale-invariant mechanism (from the class $\mathcal{M}^{\text{si}}$) against MHR distributions $\mathcal{F}^{\text{mhr}}$ for the prior independent design program ($\alpha^{\mathcal{F}^{\text{mhr}}}$) is the Second Price Auction. The worst-case MHR distribution for this mechanism is the*

truncated-exponential $\overleftarrow{\text{Exd}}_1^{852'}$ *with its monopoly quantile* $q_m^* \approx 0.426$ *and its approximation ratio is* $\alpha^{\mathcal{F}^{\text{mhr}}} \approx 1.398$.

The following discussion motivates construction of blends as informed by $\overleftarrow{\text{Exd}}_1^{852'}$ which is the worst-case distribution of Theorem 20

Because the SPA is scale-invariant, all re-scalings of the worst-case distribution $\overleftarrow{\text{Exd}}_1^{852'}$ are elements of the worst-case set of distributions (per Fact 15). Further, while Allouah and Besbes (2018) do not formally state the following, the techniques of their proof of Theorem 20 indicate that the class of all scalings of $\overleftarrow{\text{Exd}}_1^{852'}$ represents an exhaustive description of worst-case distributions. I.e., ignoring scale, the *shape* of the resulting revenue curve is uniquely worst-case for MHR distributions.

At this point, there exists a connection to the property of mixed Nash equilibrium described in Corollary 7 (and generally cf., Yao's Minimax Principle (Theorem 9) in Section 4.1).

Corollary 7 (Yao, 1977)**.** *Given a* 2*-player game* $\mathcal{G}$ *in which player* 1 *chooses action* $\omega_1 \in \Omega_1$*, then player* 2 *chooses action* $\omega_2 \in \Omega_2$*, and respective cost functions* C_1 *and* C_2*. The game* $\mathcal{G}$ *is subject to Yao's Minimax Principle:*

$$\inf_{\delta_1 \in \Delta_1(\Omega_1)} \sup_{\omega_2 \in \Omega_2} C_1(\delta_1, \omega_2) \geq \sup_{\delta_2 \in \Delta_2(\Omega_2)} \inf_{\omega_1 \in \Omega_1} C_1(\omega_1, \delta_2)$$

If mixed Nash strategies exist for $\mathcal{G}$ *– equivalently, if there exists a profile of mixed actions* $(\delta_1^*, \delta_2^*) \in \Delta_1(\Omega_1) \times \Delta_2(\Omega_2)$ *which are respectively optimal for the left-hand and right-hand outer objective functions – then the set of all actions with positive support in* δ_1^* *is necessarily a subset of the inner, non-empty* arginf *on the right-hand side, and vice*

versa, the set of all actions with positive support in δ_2^* *is necessarily a subset of the inner, non-empty* argsup *on the left-hand side.*

Corollary 7 states that if mixed Nash exists, then all actions with non-zero probability of being played are best responses to the opponent's mix over actions, and in particular, all actions with positive support have the same expected performance. The consequence for the MHR result of Allouah and Besbes (2018) here in Theorem 20 is: if mixed Nash exists and if $\overleftarrow{\text{Exd}}_1^{852'}$ is the unique worst-case distribution, then the adversary's mixed Nash actions (which are distributions) must all come from the class of scalings of $\overleftarrow{\text{Exd}}_1^{852'}$.

Motivated by this ideal description of an optimal adversary (for an auction setting with unbounded value support and an assumption of scale-invariant mechanisms), we use the dual blends outline of Theorem 14 – in particular its *inverse-distributions* of Definition 23, its rescalings of Fact 15, and its $1/z \cdot dz$ weights – to initially describe infinite-weight blends for value support $\mathcal{V} = [0, \infty)$, and subsequently modify them to finite-weight blends in value space $\mathcal{V}^2 = [1/h, h]^2$.

To set up the blends below, we describe the base distributions for which we consider all scales. Without truncation, given the exponential distribution Exd_1, its inverse-distribution is naturally the *inverse-exponential* distribution i-Exd_1 defined with CDF $\text{i-Exd}_1(x) = e^{-1/x}$ and with PDF $\text{i-exd}_1(x) = \frac{1}{x^2} \cdot e^{-1/x}$. Revenue curves for Exd_1 and i-Exd_1 are given in Figure A.3. The set of inverse-exponentials as a class are a special case of the inverse-gamma distribution (with its shape parameter set to 1).[7] The case of inverse-exponentials with general "hazard rate" β is defined by CDF $\text{i-Exd}_\beta(x) = e^{-\beta/x^2}$

[7] For shape parameter $k > 0$ and the standard Gamma Function $\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz$, the Gamma distribution is defined by PDF $\gamma\text{d}_{\beta,k}(x) = \frac{\beta^k}{\Gamma(k)} \cdot x^{k=1} \cdot e^{-\beta x}$ and the Inverse-Gamma distribution is defined by PDF $\text{i-}\gamma\text{d}_{\beta,k}(x) = \frac{\beta^k}{\Gamma(k)} x^{-k-1} e^{-\beta/x}$, both on domain $[0, \infty)$.

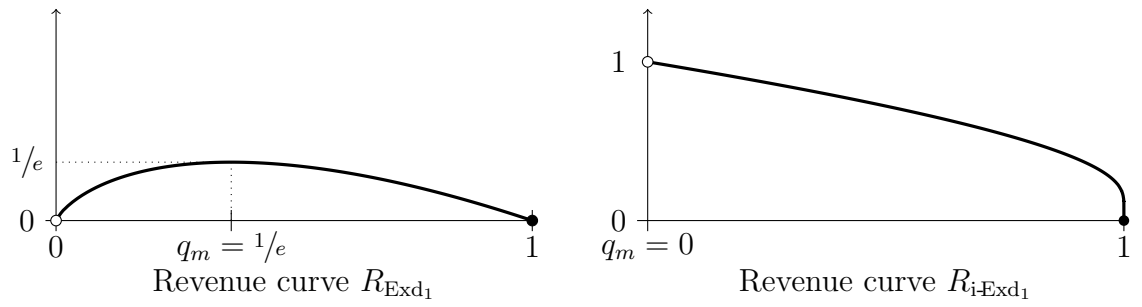

Figure A.3. Exponential and Inverse-exponential Revenue Curves
The left figure shows the revenue curve $R_{\text{Exd}_1}(q) = q \cdot \ln 1/q$. The right figure shows the revenue curve $R_{\text{i-Exd}_1}(q) = q/-\ln(1-q)$. In quantile space, the endpoints of both curves R_{Exd_1} and $R_{\text{i-Exd}_1}$ correspond to posted prices of ∞ and 0 and may require evaluation in the limit.

and PDF $\text{i-exd}_\beta(x) = \frac{1}{x^2} \cdot e^{-\beta/x^2}$ on $[0, \infty)$.[8] In this section, we plan to parameterize over domain bounds by using deliberate truncation, rather than using deliberate shifts as in Section 4.4 or inherent shifts from re-scaling of the blend elements as in Section 4.5. Thus, towards implementing the blends of this section, we use top-truncation to modify exponentials into our downward-closed class of Exponentials, and *bottom-truncation* to modify inverse-exponentials into our upward-closed class of Inverse-Exponentials.

Consider top-truncating an arbitrary-hazard-rate exponential distribution Exd_β at a *fixed* quantile $\bar{q}$ regardless of β. Two examples of top-truncated revenue curves for base distribution Exd_1 are given in the left-hand side graph of Figure A.4. As we will show below, there exists an exact bijection between exponential hazards rates $\beta \in (0, \infty)$ and positive, real-valued top-truncations $z \in (0, \infty)$. In this way, it can be confirmed that the distributions included in the Exponentials blend are re-scalings of each other.

[8] We will persistently refer to the parameter β of an inverse-exponential distribution as its "hazard rate" due to the obvious analogy to exponential distributions.

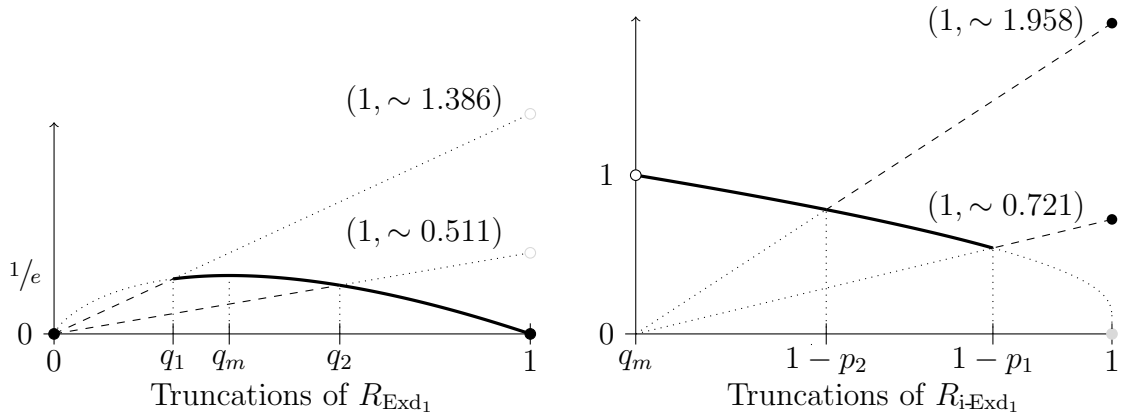

Figure A.4. Example Ironings of Exponential and Inverse-Exponential
The left figure shows the results of two examples of (dashed) top-truncation of the distribution Exd_1, respectively at $q_1 = 0.25 < q_m$ and $q_2 = 0.6 > q_m$. The right figure shows the results of the corresponding two examples of (dashed) bottom-truncation of the distribution i-Exd_1, respectively at $1 - p_1 = 1 - q_1 = 0.75$ and $1 - p_2 = 1 - q_2 = 0.4$; the (approximate) heights of the points on the right describe the *values* of the respective truncations.

For the Inverse-Exponentials, consider bottom-truncating an arbitrary-hazard-rate inverse-exponential distribution i-Exd_β at a *fixed* percentile $\bar{p}$ regardless of β. Explicitly, $\bar{p}$ as a percentile is intended to be a point mass density of equal size to a corresponding $\bar{q}$ according to $\bar{p} = \bar{q}$ (and cf., it represents bottom-truncation versus top-truncation). Two examples of bottom-truncated revenue curves for base distribution i-Exd_1 are given in the right-hand side graphic of Figure A.4. As we will show below, there exists an exact bijection between inverse-exponential hazards rates $\beta \in (0, \infty)$ and positive, real-valued bottom-truncations $z \in (0, \infty)$. In this way, it can again be confirmed that the distributions included in the Inverse-Exponentials blend are re-scalings of each other.

For the purpose of designing a dual blend, we set up specific classes of (truncated) Exponentials and Inverse-Exponentials. Fix in advance top-truncation quantile $\bar{q}$ and

bottom-truncation percentile $\bar{p} = \bar{q}$. Thus, perecentile $\bar{p}$ has *quantile* $1 - \bar{q}$. We design a blend for every *arbitrary pair* of truncation points $\bar{q}$, and $\bar{p} = \bar{q}$. Thus, the example of this section describes an entire class of dual blends, one for every fixed $\bar{q} \in (0, 1)$.

The set up of our Exponentials-versus-Inverse-Exponentials blends meets the conditions of Theorem 14, therefore we immediately have the following corollary without need for a proof. However, in order to illustrate the behavior of the point masses resulting from truncation in an infinite-weight blend, and as a technically complete step on the path to describing finite-weight blends, we do explicitly calculate the function $g(\cdot)$ for the infinite-weight blend.

Informed by Theorem 14, the (infinite-total) weights are set to be $o_z = \omega_z = {}^1\!/_z \cdot dz$ and are formally described within the analysis below. We slightly abuse notation and describe distributions as parameterized by both β and z even though either one of these is sufficient – there is an exact functional relationship between them for each side of the dual blend, which is also described in the respective analyses below.[9]

Corollary 8. *Fix $\bar{q} = \bar{p} \in (0, 1)$. Given the class of downward-closed Exponentials with members $\overleftarrow{\text{Exd}}^{z'}_{\beta}$ and the class of upward-closed Inverse-Exponentials with members $\overrightarrow{\text{i-Exd}}^{z'}_{\beta}$, each class including all $z > 0$ (equivalently all $\beta > 0$). For $o_z = \omega_z = {}^1\!/_z \cdot dz$ and $n = 2$, we have the following dual blends matching up at every $\boldsymbol{v} = (v_1, v_2 \leq v_1)$ to describe a common function g:*

$$\int_0^{v_2} \frac{1}{z} \cdot \overleftarrow{\text{exd}}^{z'}_{\beta}(v_1) \cdot \overleftarrow{\text{exd}}^{z'}_{\beta}(v_2)\ dz = g(\boldsymbol{v}) = \int_{v_1}^{\infty} \frac{1}{z} \cdot \overrightarrow{\text{i-exd}}^{z'}_{\beta}(v_1) \cdot \overrightarrow{\text{i-exd}}^{z'}_{\beta}(v_2)\ dz \tag{A.10}$$

[9] Specifically, we use β in the subscript where base distribution parameters "always appear," and use z in the superscript to indicate the truncation value where truncation parameters "always appear." We keep this notation throughout for simplicity, i.e., to avoid properly replacing either parameter with its more-complicated substitution function in terms of the other.

Some technical calculations to support this section that hamper its presentation are deferred to Appendix B.6.

A.3.1. Infinite-weight Blends

Calculations here give a re-proof Corollary 8 (which already holds from Theorem 14).

These calculations further describe supporting structure for calculations of finite weight blends. Because it is of critical importance for our work here, we state one more time: we have a fixed $\bar{q} \in (0,1)$ and $\bar{p} = \bar{q}$. In contrast to past presentations of dual blends, the presence of additional technical complexity here leads us to analyze each blend sequentially (and thus, in relative isolation) and only confirm that they match at the end.

Downward-closed quantile-$\bar{q}$ top-truncated Exponentials. Recall, an exponential distribution without truncation has CDF $\text{Exd}_\beta(x) = 1 - e^{-\beta x}$ on $(0,\infty)$ and PDF $\text{exd}_\beta(x) = \beta \cdot e^{-\beta x}$ on $(0,\infty)$. We want to truncate each exponential (with varying β) at fixed quantile $\bar{q} \in (0,1)$, and change the indexing to be purely by truncation value z. This makes the downward-closed property clear and sufficiently organizes the integral calculations of mass at every $\boldsymbol{v}$ to match our previous blends analyses throughout this work (i.e., via integration dz; the takeaway idea is, we don't have a direct interest in the hazard rate parameter β, so we replace it). Given fixed $\bar{q}$ (and keeping in mind that $\bar{q} < 1 \Rightarrow -\ln \bar{q} > 0$), we get a substitution identity to replace β with a function of z:

$$\begin{aligned} \bar{q} &= 1 - \text{Exd}_\beta(z) = e^{-\beta z} \\ \beta &= \frac{-\ln \bar{q}}{z} \end{aligned} \tag{A.11}$$

Thus, given a top-truncated exponential distribution parameterized by its truncation value z, we fix its hazard rate parameter to be $\frac{-\ln q}{z}$. The final description of an element with parameter z of the Exponentials Blend is

(A.12) (CDF) $\overleftarrow{\text{Exd}}_\beta^{z'}(x) = 1 - e^{\frac{x}{z}\ln \bar{q}}$ on $(0, z)$, $\overleftarrow{\text{Exd}}_\beta^{z'}(z) = [1 - \bar{q}, 1]$

(A.13) (PDF) $\overleftarrow{\text{exd}}_\beta^{z'}(x) = \frac{-\ln \bar{q}}{z} \cdot e^{\frac{x}{z}\ln \bar{q}}$ on $(0, z)$, $\overleftarrow{\text{exd}}_\beta^{z'}(z) = \bar{q}$ point mass

Note, we explicitly allow the CDF to be set-valued as a consequence of the point mass; and, we have implemented here the abuse of notation mentioned above, with both z and β parameters appearing in order to help simplify the notation.

The weights per z on the Exponentials are $\omega_z = {}^1\!/_z \cdot dz$ for $z \in (0, \infty)$. We need to match up mass on each side of the dual blend for three measurements of mass, partly breaking down by dimension count (Definition 20), and partly from introducing a new dimension: point masses within a distribution are identically realized to have value equal to the parameter of the distribution z. Thus, they are measured dz (as inherited from the individual distribution weight) rather than being evaluated by integrating across distributions (as in previous blends examples, e.g., at h). The three measures of mass are:

- first, a nonstandard 1-D mass measured dz at every point $(v_1, v_2 = v_1)$, which is a pure point mass when considering only the dimensions of value space; mass of this measure can only occur when $v_1 = v_2 = z$, values which can only be drawn from the distribution $\overleftarrow{\text{Exd}}_\beta^{z'}$ with a point mass at z;
- second, standard 2-D mass measured $dv_1 dv_2$ at every point $\boldsymbol{v} \in \mathcal{V}^2 = (0, \infty)^2$;

- third, a non-standard 2-D mass measured $dzdv_i = dzdv_1 = dzdv_2$ at every point $\boldsymbol{v}$; this represents mass at a point $(v_1, v_2 \leq v_1)$ that results from $v_1 = z$ as a realized draw in the $\bar{q}$-point-mass quantile of the distribution $\overleftarrow{\text{Exd}}^{z'}_{\beta}$, and v_2 as a realized draw from the continuous range below z;
- addendum: later for the dual blend composed of Inverse-Exponentials, this third point will match a non-standard 2-D mass at a point $(v_1, v_2 \leq v_1)$ that results from v_1 as a realized draw from the continuous range of the distribution $\overrightarrow{\text{i-Exd}}^{z'}_{\beta}$ above z, and $v_2 = z$ as a realized draw in the $\bar{p}$-point-mass percentile range (as reference, below see the description of the Inverse-Exponentials according to parameterized CDF (equation (A.19)) and PDF (equation (A.20));
- final note: while it is not an exact fit – because total weight is infinite – it might be easier to understand the breakdown into the different measures of mass here using the idea of *deferred realization of random variables*. This perspective is accessible because of the presence of the point mass densities $\bar{q} = \bar{p}$ as *constant probabilities of point mass draws within each distribution*: for a distribution with anonymous z, it is possible to first pre-draw each v_i as either being from the point mass or being from continuous density, in this case deferring realization of z until after this step.[10]

[10] The natural perspective of blends following from its application to prior independent lower bounds is to assume that the domain-bound z is determined first, i.e., the distribution is drawn first and fixes a parameter z; and second to draw $\boldsymbol{v}$ i.i.d. from the parameter-z distribution, in particular using the following *two-step* algorithm: randomize with Bernoulli probabilities $\bar{q}$ and $1 - \bar{q}$ to determine if a value is realized as the point mass value or drawn from the continuous region, and then draw the final value conditional on the first Bernoulli outcome.

Alternatively as suggested here, *we could run the Bernoulli draw first*, before drawing parameter z; for our analysis in this section, this ordering of random draws – which defers the fixed assignment of z – more naturally aligns with the three measures of densities that we have itemized.

As a reference, Hartline et al. (2015) illustrates how *deferred realization* is used in mechanism design.

Next per the Blends Technique, we add up mass at every ordered point $(v_1, v_2 \leq v_1)$. First consider 1-D mass dz. Calculations at every point (z, z) for $z > 0$ are given by:[11]

$$\begin{aligned} g_{1D}(z,z) &= \text{resulting mass of double-point-mass draws from } \overleftarrow{\text{Exd}}_\beta^{z'} \\ &= \omega_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftarrow{\text{Exd}}_\beta^{z'}\right]\right)^2 = \frac{1}{z} \cdot \bar{q}^2 \cdot dz \end{aligned} \tag{A.14}$$

for which $\overleftarrow{\text{Exd}}_\beta^{z'}$ is the unique distribution within the Exponentials to contribute mass of this measure at this point (z, z). Second, consider standard 2-D mass $dv_1 dv_2$. Calculations at every point use integration-by-parts and are given by:

$$\begin{aligned} g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws across the } \overleftarrow{\text{Exd}}_\beta^{z;} \text{ blend} \\ &= \int_{v_1}^{\infty} \omega_z \cdot \overleftarrow{\text{exd}}_\beta^{z'}(v_1) \cdot \overleftarrow{\text{exd}}_\beta^{z'}(v_2) \\ &= \int_{v_1}^{\infty} \frac{1}{z} \cdot \left(\frac{-\ln \bar{q}}{z} \cdot e^{\frac{v_1}{z} \ln \bar{q}}\right) \cdot \left(\frac{-\ln \bar{q}}{z} \cdot e^{\frac{v_2}{z} \ln \bar{q}}\right) \, dz \\ &= (\ln \bar{q})^2 \int_{v_1}^{\infty} \left(\frac{1}{z}\right) \left[\frac{1}{z^2} \cdot e^{\frac{v_1+v_2}{z} \ln \bar{q}} \, dz\right] \qquad \text{(A.15)} \\ &= (\ln \bar{q})^2 \left(\left[\left(\frac{1}{z}\right) \cdot \left(\frac{-1}{\ln \bar{q} \cdot (v_1+v_2)} \cdot e^{\frac{v_1+v_2}{z} \ln \bar{q}}\right)\right]_{v_1}^{\infty} \right. \\ &\qquad \left. - \int_{v_1}^{\infty} \left(\frac{-1}{\ln \bar{q} \cdot (v_1+v_2)} e^{\frac{v_1+v_2}{z} \ln \bar{q}}\right) \left(\frac{-1}{z^2} \cdot dz\right) \right) \\ &= \left[\frac{\ln \bar{q}}{v_1} \cdot \frac{1}{v_1+v_2} e^{\frac{v_1+v_2}{v_1} \ln \bar{q}}\right] - \frac{\ln \bar{q}}{v_1+v_2} \left[\left(\frac{-1}{\ln \bar{q}(v_1+v_2)}\right) e^{\frac{v_1+v_2}{z} \ln \bar{q}}\right]_{v_1}^{\infty} \\ &= \left[\frac{\ln \bar{q}}{v_1} \cdot \frac{1}{v_1+v_2} \cdot \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}}\right] + \frac{1}{(v_1+v_2)^2} \left(1 - \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}}\right) > 0 \qquad \text{(A.16)} \end{aligned}$$

[11] Note, as has been standard presentation so far throughout this work, we write dz explicitly but rather suppress dv_i terms in blends calculations.

for which the final term is definitively positive for $\bar{q} \in (0,1)$ (with its first additive term negative and its second term positive; see Appendix B.6.1). Third, consider 2-D mass $dzdv_i$ for which we use the functional notation g_{2Dz} to describe its contribution to the function g. For this measure of mass, calculations for the Exponentials require that the larger value v_1 be drawn equal to z and value v_2 be a weakly smaller draw from continuous density. At every point, these calculations are given by:

$$\begin{aligned} g_{2Dz}(v_1 = z, v_2) &= \text{resulting mass of ``one each'' draws from } \overleftarrow{\text{Exd}}^{z'}_{\beta} \\ &= \omega_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftarrow{\text{Exd}}^{z'}_{\beta} \right] \right) \cdot \overleftarrow{\text{exd}}^{z'}_{\beta}(v_2) \\ &= \omega_{v_1} \cdot \bar{q} \cdot \overleftarrow{\text{exd}}^{v_1'}_{\beta}(v_2) = \frac{\bar{q}}{v_1} \cdot \left(\frac{-\ln \bar{q}}{v_1} e^{\frac{v_2}{v_1} ln \bar{q}} \right) \cdot dz \end{aligned} \tag{A.17}$$

We now turn to the other side of the dual blend: the Inverse-Exponentials.

Upward-closed percentile-$\bar{p}$ bottom-truncated Inverse-Exponentials. Recall, an inverse-exponential distribution without truncation has CDF i-Exd$_\beta(x) = e^{-\beta/x}$ on $(0, \infty)$ and PDF i-exd$_\beta(x) = \frac{\beta}{x^2} \cdot e^{-\beta/x}$ on $(0, \infty)$. We want to truncate each inverse-exponential (with varying β) at fixed percentile $\bar{p} = \bar{q}$ for $\bar{q}$ as fixed for the Exponentials side of the dual blend above, and change the indexing to be purely by truncation value z. This makes the upward-closed property clear and sufficiently organizes the integral calculations of mass at every $\boldsymbol{v}$ to match our previous blends analyses throughout this work (i.e., via integration dz; again the takeaway idea is, we don't have a direct interest in the hazard rate parameter β, so we replace it). Given fixed $\bar{p} = \bar{q}$ (and keeping in mind that $\bar{p} < 1 \Rightarrow -\ln \bar{p} > 0$),

we get a substitution identity to replace β with a function of z:

$$\bar{p} = \text{i-Exd}_\beta(z) = e^{-\beta/z}$$

$$\beta = -z \ln \bar{p} \tag{A.18}$$

To be clear, because this truncation is at the bottom of the distribution rather than the top, we now have that $\bar{p}$ is the probability of the point mass and $1 - \bar{p}$ is the probability of the continuous draw (now above the point mass). Thus, given a bottom-truncated inverse-exponential distribution parameterized by its truncation value z, we fix its hazard rate parameter to be $-z \ln \bar{p}$. The final description of an element with parameter z of the Inverse-Exponentials Blend is

$$\text{(CDF)} \quad \overrightarrow{\text{i-Exd}}^{z'}_\beta(z) = [0, \bar{p}], \qquad \overrightarrow{\text{i-Exd}}^{z'}_\beta(x) = e^{\frac{z}{x} \ln \bar{p}} \qquad \text{on } (z, \infty) \tag{A.19}$$

$$\text{(PDF)} \quad \overrightarrow{\text{i-exd}}^{z'}_\beta(z) = \bar{p} \text{ point mass}, \quad \overrightarrow{\text{i-exd}}^{z'}_\beta(x) = \frac{-z \ln \bar{p}}{x^2} \cdot e^{\frac{z}{x} \ln \bar{p}} \quad \text{on } (z, \infty) \tag{A.20}$$

Note again, the CDF is explicitly allowed to be set-valued as a consequence of the point mass; and, we keep here the abuse of notation mentioned previously, with both z and β parameters appearing in order to help simplify the notation.

The weights per z on the Inverse-Exponentials are $o_z = {}^1\!/_z \cdot dz$ for $z \in (0, \infty)$. The goal of the rest of this analysis is to match up the three measures of mass identified from the Exponentials side of the dual blend, i.e., we specifically show that the blend over Inverse-Exponentials has overall mass to match each of: 1-D mass in equation (A.14), standard 2-D mass in equation (A.16), and 2-D mass-measure $dzdv_i$ in equation (A.17).

First, considering 1-D mass dz, calculations at every point (z, z) for $z > 0$ are given by:

$$\begin{aligned} g_{1D}(z,z) &= \text{resulting mass of double-point-mass draws from } \overrightarrow{\text{i-Exd}}_{\beta}^{z'} \\ &= o_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overrightarrow{\text{i-Exd}}_{\beta}^{z'}\right]\right)^2 = \frac{1}{z} \cdot \bar{p}^2 \cdot dz \end{aligned} \tag{A.21}$$

for which $\overrightarrow{\text{i-Exd}}_{\beta}^{z'}$ is the unique distribution within the Inverse-Exponentials to contribute mass of this measure at this point (z, z), and which matches equation (A.14) (because $\bar{p} = \bar{q}$). Second, considering standard 2-D mass $dv_1 dv_2$, calculations at every point use integration-by-parts and are given by:

$$\begin{aligned} g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws across the } \overrightarrow{\text{i-Exd}}_{\beta}^{z;} \text{ blend} \\ &= \int_0^{v_2} o_z \cdot \overrightarrow{\text{i-exd}}_{\beta}^{z'}(v_1) \cdot \overrightarrow{\text{i-exd}}_{\beta}^{z'}(v_2) \\ &= \int_0^{v_2} \frac{1}{z} \cdot \left(\frac{z \cdot (-\ln \bar{p})}{v_1^2} e^{\frac{z}{v_1} \ln \bar{p}}\right) \cdot \left(\frac{z \cdot (-\ln \bar{p})}{v_2^2} e^{\frac{z}{v_2} \ln \bar{p}}\right) \, dz \\ &= \left(\frac{\ln \bar{p}}{v_1 v_2}\right)^2 \int_0^{v_2} \left(\frac{1}{z} \cdot z^2\right) \left[e^{\left(\frac{v_1+v_2}{v_1 v_2}\right) z \ln \bar{p}} \, dz\right] \qquad \text{(A.22)} \\ &= \left(\frac{\ln \bar{p}}{v_1 v_2}\right)^2 \left(\left[(z)\left(\frac{v_1 v_2}{\ln \bar{p} \cdot (v_1 + v_2)} e^{\left(\frac{v_1+v_2}{v_1 v_2}\right) z \ln \bar{p}}\right)\right]_0^{v_2} \right. \\ &\qquad \left. - \int_0^{v_2} \left(\frac{v_1 v_2}{\ln \bar{p} \cdot (v_1 + v_2)} \cdot e^{\left(\frac{v_1+v_2}{v_1 v_2}\right) z \ln \bar{p}}\right) (1 \cdot dz)\right) \\ &= \left[\frac{\ln \bar{p}}{v_1} \cdot \frac{1}{v_1 + v_2} e^{\left(\frac{v_1+v_2}{v_1}\right) \ln \bar{p}}\right] - \left(\frac{\ln \bar{p}}{v_1 v_2}\right)^2 \cdot \left(\frac{v_1 v_2}{\ln \bar{p} \cdot (v_1 + v_2)}\right)^2 \left[e^{\left(\frac{v_1+v_2}{v_1 v_2}\right) z \ln \bar{p}}\right]_0^{v_2} \\ &= \left[\frac{\ln \bar{p}}{v_1} \cdot \frac{1}{v_1 + v_2} \cdot \bar{p} \cdot e^{\frac{v_2}{v_1} \ln \bar{p}}\right] + \frac{1}{(v_1 + v_2)^2} \left(1 - \bar{p} \cdot e^{\frac{v_2}{v_1} \ln \bar{p}}\right) \qquad \text{(A.23)} \end{aligned}$$

which matches equation (A.16). Third, consider 2-D mass $dzdv_i$ for which we use the notation g_{2Dz} to describe its contribution to the function g. For this measure of mass, calculations for the Inverse-Exponentials require that the smaller value v_2 be drawn equal to z and value v_1 be a weakly larger draw from continuous density. At every point, these calculations are given by:

$$\begin{aligned} g_{2Dz}(v_1, v_2 = z) &= \text{resulting mass of ``one each'' draws from } \overrightarrow{\text{i-Exd}}_{\beta}^{z'} \\ &= o_z \cdot \overrightarrow{\text{i-exd}}_{\beta}^{z'}(v_1) \cdot \left(\Pr\left[X = z \text{ for } X \sim \overrightarrow{\text{i-Exd}}_{\beta}^{z'}\right]\right) \\ &= o_{v_2} \cdot \overrightarrow{\text{i-exd}}_{\beta}^{v_2'}(v_1) \cdot \bar{p} = \frac{\bar{p}}{v_2} \cdot \left(\frac{-v_2 \ln \bar{p}}{v_1^2} e^{\frac{v_2}{v_1} \ln \bar{p}}\right) \cdot dz \end{aligned} \tag{A.24}$$

which matches equation (A.17). This line effectively finishes a re-proof of Corollary 8.

A.3.2. Notes for Modification to Finite-weight

Before modifying the infinite-weight blends solution of the previous section to a finite-weight solution, we highlight a number of features of the modification for intuition and completeness. As a preliminary note, modification of the infinite-weight blends solution of Appendix A.3.1 to finite-weight is designed to fit it into value space as a finite box: $\mathcal{V}^2 = [1/h, h]^2$.

Introducing blend-elements as distributions with conditioning. For the first time throughout this work, blends will employ doubly-conditioned modifications of base-distributions as elements. (Truncated-and-conditioned distributions also appear, see below.) In particular, Exponentials that were otherwise top-truncated *above* h (the upper

bound on value support here) are doubly-conditioned (above and below) to restrict draws to be from the range $[1/h, h]$. These blend elements are thus of the type $\overleftrightarrow{\text{Exd}}_{\beta}^{1/h,h}$.

The Inverse-Exponentials that were otherwise bottom-truncated *below* $1/h$ (the lower bound on value support here) are similarly doubly-conditioned to restrict draws to be from the range $[1/h, h]$. These blend elements are thus of the type $\overleftrightarrow{\text{i-Exd}}_{\beta}^{1/h,h}$.

The truncated-and-conditioned distributions are respectively $\overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'}$ and $\overleftrightarrow{\text{i-Exd}}_{\beta}^{z',h}$. Consider for these in particular, e.g., the distributions composing the Exponentials side which have top-truncation at a fixed point mass with density $\bar{q}$ (which is the same as the infinite-weight solution above). When we also condition the distribution to be above $1/h$ and then re-normalize, the underlying distribution stays the same (i.e., the top-truncation point mass stays at value z and the distribution keeps the same hazard parameter β). Furthermore, for κ the total remaining measure after conditioning, the point mass density becomes $\bar{q}/\kappa$. Similarly the *total* density on the continuous density region of the distribution becomes $(1-\bar{q}-\kappa)/\kappa$ (because all of the density removed from the conditioning is from the continuous density region and is subtracted up front). Modifications here to Inverse-Exponentials have symmetric interpretation.

Finite but *unknown* total weight. For this section, we allow the total weight to be anything finite per Fact 17. This *verified assumption* that we can normalize the weight to 1 later is actually somewhat critical – because in fact we are not going to calculate the exact weight due to technical complexity.

Instead, we only run calculations to confirm that total weights are equal as defined by each side of the dual blends, and we further confirm that weights are finite by calculating

and analyzing an upper bound. Both of these sets of calculations (equality, finiteness) will be deferred to Appendix B.6.2.

Elements of finite-weight Exponentials blend maintain the MHR property post-modification. One of the goals of modifying infinite-weight Exponentials into a "usable" finite weight blend is that the modified distributions remain feasible within the setting of an adversary restricted to the class of MHR distributions, towards application of the Blends Technique. The finite-weight blends solution of Appendix A.3.3 is presented within this context. The confirmation that distributions of the finite-weight Exponentials are MHR is deferred to Appendix B.6.3.

Outlining the technique of the modification. At a high level, the observable goal of designing the finite-weight blends in Appendix A.3.3 is simple: re-achieve the exact mathematical calculations of the infinite-weight solution from Appendix A.3.1. The implemented technique follows a three-point outline to design the finite-weight blends:

(1) recalling that all infinite-weight blends elements (on both the Exponentials side and the Inverse-Exponentials side) are parameterized by z which respectively represents top-trunction or bottom-truncation, for all $z \in [1/h, h]$ and for both sides, we further modify the distribution by conditioning it at a bound of value support in order to have domain that is a subset of value support (Exponentials are bottom-conditioned at $1/h$, Inverse-Exponentials top-conditioned at h); this provides a formulaic starting point to recover the exact calculations of the infinite-weight blends, only needing the following two "fix ups;"

(2) densities of distributions described in point (1) no longer match the infinite-weight solution because of the re-normalization step within conditioning; we fix this innately by modifying the choice of weights o_z and ω_z to include a multiplicative factor to exactly cancel the re-normalization factor;

(3) the integrals used to calculate density at an input $\boldsymbol{v}$ have endpoints that reflect the restriction to finite ranges of values in $[1/h, h]$; we modify distributions with parameters z outside this range in a way that their contributions exactly offset evaluation of the integrals at these "problem" endpoints (and thus the combined contribution is 0; cf., the role of the o_{pm}-term in line (4.6) in Section 4.4 which effectively cancels evaluation of the o_{Ez}-integral at its endpoint 0).

A.3.3. Modifying to Finite-weight Blends

We now modify the infinite-weight blends solution of Appendix A.3.1 to a finite-weight solution, specifically within value space as a finite box: $\mathcal{V}^2 = [1/h, h]^2$. As indicated above, this section continues to make heavy use of truncated distributions (and now also conditioned distributions) as formally explained in Appendix B.1. Where truncation is employed, we still fix in advance truncation at the quantile $\bar{q}$ for the Exponentials and at the percentile $\bar{p} = \bar{q}$ for the Inverse-Exponentials. We have a working assumption that the weights described here result in a total weight that is finite, and we will confirm this assumption later.

The rest of this section generally parallels the presentation of the infinite-weight solution in Appendix A.3.1. Supplementary material to support the rest of this section is deferred as appropriate to Appendix B.6.

Quantile-$\bar{q}$ Exponentials. There are two types of distributions which have weights ω_z, each modified from the Exponentials distributions in the infinite-weight solution which were described in equations (A.12) and (A.13) and embedding again the identity $\beta = {}^{-\ln \bar{q}}/_{z}$. *For simplicity of the equations and following analysis, we only present the relevant PDFs and within their definitions, we also make use of "un-substituted" base CDF functions for the conditioning.* For every $z \in [{}^{1}/_{h}, h]$, we include the bottom-conditioned, top-truncated Exponential distribution

(A.25)
$$\overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(x) = \frac{\frac{-\ln \bar{q}}{z} \cdot e^{\frac{x}{z}\ln \bar{q}}}{1 - \text{Exd}_{\beta}({}^{1}/_{h})} \text{ on } [{}^{1}/_{h}, z), \quad \overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(z) = \frac{\bar{q}}{1 - \text{Exd}_{\beta}({}^{1}/_{h})} \text{ point mass}$$

and for every $z > h$ (with $\beta = {}^{-\ln \bar{q}}/_{z}$), we include the doubly-conditioned distribution

(A.26)
$$\overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}(x) = \frac{\frac{-\ln \bar{q}}{z} \cdot e^{\frac{x}{z}\ln \bar{q}}}{\text{Exd}_{\beta}(h) - \text{Exd}_{\beta}({}^{1}/_{h})} \text{ on } [{}^{1}/_{h}, h]$$

The weights ω_z are:

- for every $z \in [{}^{1}/_{h}, h]$, the exponential distribution $\overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'}$ has weight $\omega_z = \frac{1}{z}\left(1 - \text{Exd}_{\beta}({}^{1}/_{h})\right)^2 \cdot dz$;
- for every $z \in (h, \infty)$, the exponential distribution $\overleftrightarrow{\text{Exd}}_{\beta}^{1/h,h}$ has weight $\omega_z = \frac{1}{z}\left(\text{Exd}_{\beta}(h) - \text{Exd}_{\beta}({}^{1}/_{h})\right)^2 \cdot dz$.

At the end of the previous section, we explained how these (finite-total) weights were specifically designed to recover the infinite-weight calculation forms of Appendix A.3.1. With this direct comparison to infinite-weights, the finite-weights are calculated as follows. First consider 1-D density dz. For every $z \in [{}^{1}/_{h}, h]$, calculations at point (z, z) are given by

$$
\begin{aligned}
g_{1D}(z,z) &= \text{resulting mass of double-point-mass draws from } \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'} \\
&= \omega_z \cdot \left(\frac{\bar{q}}{1 - \text{Exd}_{\beta}(1/h)} \right)^2 = \frac{1}{z} \cdot \bar{q}^2 \cdot dz
\end{aligned} \tag{A.27}
$$

Second, consider standard 2-D mass $dv_1 dv_2$. Calculations are solved using reference to infinite-weights equations for simplicity – our construction is specifically design for this simplification, and we do this in one step when we substitute in the PDF definitions. At every point $\boldsymbol{v}$ with $h \geq v_1 \geq v_2 \geq 1/h$, calculations are given by

$$
\begin{aligned}
g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws from } \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'} \\
&\quad + \text{resulting mass of double-continuous draws from } \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,h} \\
&= \int_{v_1}^{h} \omega_z \cdot \overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(v_1) \cdot \overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(v_2) + \int_{h}^{\infty} \omega_z \cdot \overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}(v_1) \cdot \overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}(v_2) \\
&= \int_{v_1}^{\infty} \frac{1}{z} \cdot \left(\frac{\ln \bar{q}}{z} \right)^2 \cdot e^{\frac{v_1+v_2}{z} \ln \bar{q}} dz = \text{ line (A.15)} \\
&= \text{ line (A.16)} = \left[\frac{\ln \bar{q}}{v_1} \cdot \frac{1}{v_1 + v_2} \cdot \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}} \right] + \frac{1}{(v_1 + v_2)^2} \left(1 - \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}} \right)
\end{aligned} \tag{A.28}
$$

Third, consider 2-D mass $dz dv_i$ for which we use the functional notation g_{2Dz} to describe its contribution to the function g. Recall for this measure of mass, calculations for the Exponentials require that the larger value v_1 be drawn equal to z, and now value v_2 must be a weakly smaller draw from continuous density in $[1/h, v_1]$. Calculations at every point

are given by

$$\begin{aligned} g_{2Dz}(v_1 = z, v_2) &= \text{resulting mass of ``one each'' draws from } \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'} \\ &= \omega_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,z'} \right] \right) \cdot \overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(v_2) \\ &= \omega_{v_1} \cdot \frac{\bar{q}}{1 - \text{Exd}_{\beta}(1/h)} \cdot \overleftrightarrow{\text{Exd}}_{\beta}^{1/h,v_1'}(v_2) \qquad \text{(A.29)} \\ &= \frac{\bar{q} \cdot (1 - \text{Exd}_{\beta}(1/h))^2}{v_1 \cdot (1 - \text{Exd}_{\beta}(1/h))} \cdot \frac{\left(\frac{-\ln \bar{q}}{v_1} e^{\frac{v_2}{v_1} ln \bar{q}} \right)}{1 - \text{Exd}_{\beta}(1/h)} \cdot dz = \frac{\bar{q}}{v_1} \cdot \left(\frac{-\ln \bar{q}}{v_1} e^{\frac{v_2}{v_1} ln \bar{q}} \right) \cdot dz \qquad \text{(A.30)} \end{aligned}$$

Percentile-$\bar{p}$ Inverse-Exponentials. On the other side of the dual blend, again there are two types of distributions which have weights o_z, each now modified from the Inverse-Exponentials distributions in the infinite-weight solution which are described in equations (A.19) and (A.20) and embedding again their identity $\beta = -z \ln \bar{p}$. Recall, percentile $\bar{p} = \bar{q}$ is now the measure of the point mass at the bottom-truncation point z. Again for simplicity, we only give PDFs with dependence on base CDFs. For every $z \in [1/h, h]$, we include the bottom-truncated, top-conditioned Inverse-Exponential distribution

$$\overleftrightarrow{\text{i-exd}}_{\beta}^{z',h}(z) = \frac{\bar{p}}{\text{i-Exd}_{\beta}(h)} \text{ point mass}, \quad \overleftrightarrow{\text{i-exd}}_{\beta}^{z',h}(x) = \frac{\frac{-z \ln \bar{p}}{x^2} \cdot e^{\frac{z}{x} \ln \bar{p}}}{\text{i-Exd}_{\beta}(h)} \quad \text{on } (z, h] \qquad \text{(A.31)}$$

and for every $0 < z < 1/h$ (with $\beta = -z \ln \bar{p}$), we include the doubly-conditioned distribution

$$\overleftrightarrow{\text{i-exd}}_{\beta}^{1/h,h}(x) = \frac{\frac{-z \ln \bar{p}}{x^2} \cdot e^{\frac{z}{x} \ln \bar{p}}}{\text{i-Exd}_{\beta}(h) - \text{i-Exd}_{\beta}(1/h)} \text{ on } [1/h, h] \qquad \text{(A.32)}$$

The o_z weights are

- for every $z \in [1/h, h]$, the inverse-exponential distribution $\overleftrightarrow{\text{i-Exd}}_\beta^{z',h}$ has weight $o_z = \frac{1}{z}\left(\text{i-Exd}_\beta(h)\right)^2 \cdot dz$;
- for every $z \in (0, 1/h)$, the inverse-exponential distribution $\overleftrightarrow{\text{i-Exd}}_\beta^{1/h,h}$ has weight $o_z = \frac{1}{z}\left(\text{i-Exd}_\beta(h) - \text{i-Exd}_\beta(1/h)\right)^2 \cdot dz$.

Now the density calculations for the Inverse-Exponentials. First, considering 1-D mass dz, for every $z \in [1/h, h]$, calculations at point (z, z) are given by

$$\begin{aligned} g_{1D}(z,z) &= \text{resulting mass of double-point-mass draws from } \overleftrightarrow{\text{i-Exd}}_\beta^{z',h} \\ &= o_z \cdot \left(\frac{\bar{p}}{\text{i-Exd}_\beta(h)}\right)^2 = \frac{1}{z} \cdot \bar{p}^2 \cdot dz \end{aligned} \tag{A.33}$$

We confirm, the quantity in equation (A.33) matches the Exponentials blend quantity in equation (A.27). Second, considering standard 2-D mass $dv_1 dv_2$, calculations at every point use reference to infinite-weights equations for simplicity – recall, they are constructed specifically to implement this simplification, and we do this in one step when we substitute in PDF definitions. Calculations at every point $\boldsymbol{v}$ with $h \geq v_1 \geq v_2 \geq 1/h$ are given by

$$\begin{aligned} g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws from } \overleftrightarrow{\text{i-Exd}}_\beta^{z',h} \\ &\quad + \text{resulting mass of double-continuous draws from } \overleftrightarrow{\text{i-Exd}}_\beta^{1/h,h} \\ &= \int_{1/h}^{v_2} o_z \cdot \overleftrightarrow{\text{i-exd}}_\beta^{z',h}(v_1) \cdot \overleftrightarrow{\text{i-exd}}_\beta^{z',h}(v_2) + \int_0^{1/h} o_z \cdot \overleftrightarrow{\text{i-exd}}_\beta^{1/h,h}(v_1) \cdot \overleftrightarrow{\text{i-exd}}_\beta^{1/h,h}(v_2) \\ &= \int_0^{v_2} \frac{1}{z}\left(\frac{z \ln \bar{p}}{v_1 v_2}\right)^2 e^{\left(\frac{v_1+v_2}{v_1 v_2}\right) z \ln \bar{p}} dz = \text{ line (A.22)} \\ &= \text{ line (A.23)} = \left[\frac{\ln \bar{p}}{v_1} \cdot \frac{1}{v_1+v_2} \cdot \bar{p} \cdot e^{\frac{v_2}{v_1} \ln \bar{p}}\right] + \frac{1}{(v_1+v_2)^2}\left(1 - \bar{p} \cdot e^{\frac{v_2}{v_1} \ln \bar{p}}\right) \end{aligned} \tag{A.34}$$

We confirm, the quantity in equation (A.34) matches the Exponentials blend quantity in equation (A.28). Third, consider 2-D mass $dzdv_i$ for which we use the notation g_{2Dz} to describe its contribution to the function g. Again for this measure of mass, calculations for the Inverse-Exponentials require that the smaller value v_2 be drawn equal to z and value v_1 be a weakly larger draw from continuous density in $[v_2, h]$. Calculations at every point are given by

$$\begin{aligned} g_{2Dz}(v_1, v_2 = z) &= \text{resulting mass of ``one each'' draws from } \overleftrightarrow{\text{i-Exd}}_{\beta}^{z',h} \\ &= o_z \cdot \overleftrightarrow{\text{i-exd}}_{\beta}^{z',h}(v_1) \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftrightarrow{\text{i-Exd}}_{\beta}^{z',h} \right] \right) \\ &= o_{v_2} \cdot \overleftrightarrow{\text{i-Exd}}_{\beta}^{v_2',h}(v_1) \cdot \frac{\bar{p}}{\text{i-Exd}_{\beta}(h)} \qquad \text{(A.35)} \\ &= \frac{\bar{p} \cdot (\text{i-Exd}_{\beta}(h))^2}{v_2 \cdot \text{i-Exd}_{\beta}(h)} \cdot \frac{\left(\frac{-v_2 \ln \bar{p}}{v_1^2} e^{\frac{v_2}{v_1} \ln \bar{p}} \right)}{\text{i-Exd}_{\beta}(h)} \cdot dz = \bar{p} \cdot \left(\frac{-\ln \bar{p}}{v_1^2} e^{\frac{v_2}{v_1} \ln \bar{p}} \right) \cdot dz \qquad \text{(A.36)} \end{aligned}$$

We confirm, the quantity in equation (A.36) matches the Exponentials blend quantity in equation (A.30), which completes the "proof" that the finite-weight blends match everywhere for every measure of density, and therefore are proper dual blends.

At this point, two technical analyses are deferred to the appendix. First, calculations to confirm that (a) weights are equal, and (b) weights are finite are given in Appendix B.6.2. To show that weights are finite, in fact we *upper bound them* for technical simplicity, rather than calculating them exactly. The finite weights can be used within the Blends Technique even leaving them "unknown," because they are equal and the total divides out of the ratio of optimal performances (Fact 17).

Second, as a point of interest, we want to be able to use the finite-weight Exponentials blend as the adversary's choice within prior independent settings in which the class of distributions is $\mathcal{F}^{\text{mhr}}$, i.e., MHR distributions (Definition 10). The base distributions Exd_β (un-truncated and un-conditioned) are well-known to be MHR (recall Fact 11). Proofs to confirm that the distributions $\overleftrightarrow{\text{Exd}}_\beta^{1/h,z'}$ and $\overleftrightarrow{\text{Exd}}_\beta^{1/h,h}$ used here remain MHR even after their respective truncation and conditioning operations are given in Appendix B.6.3. We formalize this final statement as a proposition.

Proposition 6. *For every fixed $\bar{q} \in (0,1)$ and $h > 1$, all distributions of the types $\overleftrightarrow{\text{Exd}}_\beta^{1/h,z'}$ (of equation (A.25)) and $\overleftrightarrow{\text{Exd}}_\beta^{1/h,h}$ (of equation (A.26)) are MHR. Therefore they are all elements of the class of distributions $\mathcal{F}^{\text{mhr}}$ and are accessible to a prior independent design problem's adversary when restricted to $\mathcal{F}^{\text{mhr}}$.*

A.4. Regular (Revenue) Dual Blends

This section repeats only the infinite-weight calculations of the previous section for value support $\mathcal{V} = [0, \infty)$, this time for the setting of regular distributions, i.e., restriction to the class $\mathcal{F}^{\text{reg}}$. The setting corresponds to Theorem 5 which was proved in this thesis, which we restate here.

Theorem 5. *Given a single item, 2-agent auction with a revenue objective, the optimal truthful, scale-invariant mechanism (from the class $\mathcal{M}^{\text{si}}$) against regular distributions $\mathcal{F}^{\text{reg}}$ for the prior independent design program ($\alpha^{\mathcal{F}^{\text{reg}}}$) is $M_{\hat{r},\boldsymbol{\xi}}$ which randomizes according to $\boldsymbol{\xi}$ over the second-price auction M_1 with probability ξ_1 and r^*-markup mechanism M_{r^*} with probability $\xi_{r^*} = 1 - \xi_1$, where $\xi_1 \approx 0.806$ and $r^* \approx 2.447$. The worst-case regular*

distribution is $\text{Trvd}_{q_m^*}$ *with its monopoly quantile* $q_m^* \approx 0.093$ *and its approximation ratio is* $\alpha^{\mathcal{F}^{\text{reg}}} \approx 1.907$.

Thus, the worst-case distribution for this setting to be used in this section is a Triangle revenue-curve, in particular, the $q_m^* \approx 0.093$-truncation of the standard *constant negative virtual value* (CNVV)[12] which has CDF $\text{Cnvd}_{-1}(x) = 1 - {}^{1}\!/_{x+1}$ on $[0, \infty)$ and PDF $\text{cnvd}_{-1}(x) = {}^{1}\!/_{(x+1)^2}$ and constant negative virtual value $\phi^{\text{Cnvd}_{-1}}(v) = -1$. With ${}^{1}\!/_{q_m^*} \approx 10.74$ and ${}^{1}\!/_{(1-q_m^*)} \approx 1.103$, this worst-case distribution is denoted $\overleftarrow{\text{Cnvd}}_{1.103}^{10.74'}$.

While *Default-scaled* Triangle revenue curves – with peak revenue fixed to be 1 independent of monopoly quantile – were defined in Definition 16, we need to generalize usage of Triangle revenue curves here. Whereas Default-scaled Triangles fix the height of the monopoly-peak and leave the monopoly quantile free, we want to flip these: we will want to consider instead all scalings of Triangles given a fixed monopoly quantile $\bar{q}$. These are exactly the $\bar{q}$-truncations of distributions in the set $\text{CNVV} = \{\text{Cnvd}_\phi \ : \ \phi < 0\}$. Recall per Definition 15, an arbitrary distribution of this class has CDF $\text{Cnvd}_\phi(x) = 1 + {}^{\phi}\!/_{(x-\phi)}$ and PDF $\text{cnvd}_\phi(x) = {}^{-\phi}\!/_{(x-\phi)^2}$.

CNVVs have three interesting properties. One we have mentioned previously in Fact 12: they are the class of tight distributions for Theorem 4 (Bulow and Klemperer, 1996). The second is to recall that they are shifted-quadratics which are always shifted to-the-left. The point is that while $\int_0^\infty \frac{1}{x^2} dx$ is originally divergent, a leftward-shift for

[12] Recall, constant negative virtual value distributions were defined in Definition 15 to be the set $\text{CNVV} = \{\text{Sqd}_{0,\phi} \ : \ \phi < 0\}$. Thus equivalently, this worst case distribution is described as a special-case of shifted-quadratic distribution $\text{Sqd}_{0,-1}$ which has CDF $\text{Sqd}_{0,-1}(x) = 1 - {}^{1}\!/_{x+1}$ on $[0, \infty)$ and PDF $\text{sqd}_{0,1}(x) = {}^{1}\!/_{(x+1)^2}$ and constant negative virtual value $\phi^{\text{Sqd}_{0,-1}}(v) = -1$.

Graphically, their revenue curves were described as "all revenue curves that are simple line segments connecting a point (in quantile revenue curve space) $(0, -\phi)$ to $(1, 0)$ with slope ϕ (and for all $\phi < 0$)." Additionally note here, from the geometry of CNVV revenue curves, top-truncation at $\bar{q}$ is always the monopoly truncation q_m.

every $\epsilon > 0$ will induce an integral that converges, i.e., $\int_0^\infty \frac{1}{(x+\epsilon)^2} dx$ is finite for every $\epsilon > 0$ and these describe the class of CNVVs.

The third interesting property is that *applying the inverse-distribution operation to all CNVVs as a class is an identity operation*. Specifically, the inverse-distribution of a CNVV distribution with virtual value $\phi < 0$ yields the CNVV distribution with virtual value $1/\phi$, as can be directly observed from the following:

$$\begin{aligned} \text{Cnvd}_\phi(x) &= 1 + \frac{\phi}{x - \phi} && \text{on } (0, \infty) \\ \text{i-Cnvd}_\phi(x) &= 1 - \text{Cnvd}_\phi(1/x) = \frac{-\phi}{1/x - \phi} = \frac{-\phi x}{1 - \phi x} && \text{on } (0, \infty) \\ &= 1 - \frac{1}{1 - \phi x} = 1 + \frac{1/\phi}{x - 1/\phi} = \text{Cnvd}_{1/\phi}(x) \end{aligned} \tag{A.37}$$

Obviously, mapping all negative ϕ to $1/\phi$ is a self-bijection with range equal to domain. Therefore the goal of this section is *to describe (infinite-weight) dual blends with CNVVs on both sides, with only the distinction that one side is top-truncated and the other side is bottom-truncated*. Analogous to the previous section, we use notation from Appendix B.1 to indicate truncation.

On the downward-closed side with weights ω_z, we therefore have a fixed (monopoly) top-truncation quantile $\bar{q} = q_m$; versus, on the upward-closed side with weights o_z, we have a fixed (in fact *revenue-curve-minimizer*) bottom-truncation percentile $\bar{p} = \bar{q}$. Figure A.5 illustrates revenue curves for two distributions that are both CNVVs and also inverse-distributions to each other; and Figure A.6 shows the effects of respective top-truncation and bottom truncation (cf., Figure A.3 and Figure A.4 of the previous section).

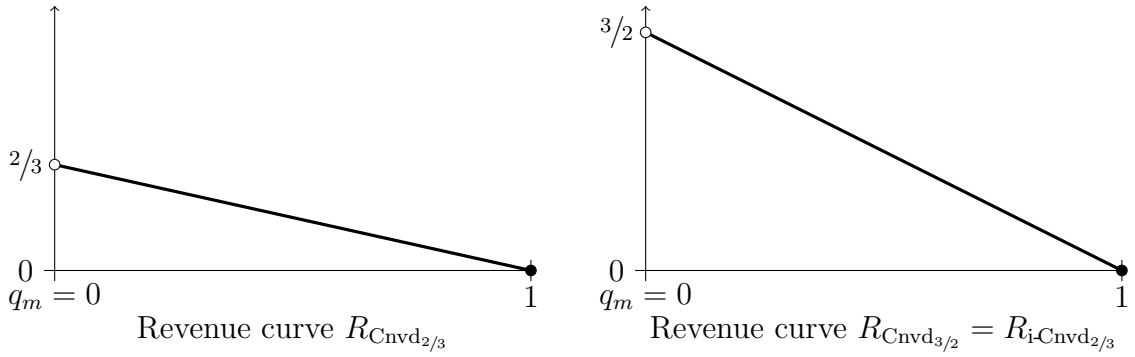

Figure A.5. Shifted-Quadratic and (Same-Class) Inverse Revenue Curves
The left figure shows the revenue curve $R_{\text{Cnvd}_{2/3}}(q) = 2/3 \cdot (1 - q)$. The right figure shows the revenue curve $R_{\text{i-Cnvd}_{3/2}}(q) = 3/2 \cdot (1 - q)$. (The inverse-distributions are the same class with the scales inverted.) In quantile space, the endpoints of both curves $R_{\text{Cnvd}_{2/3}}$ and $R_{\text{i-Cnvd}_{3/2}}$ correspond to posted prices of ∞ and 0.

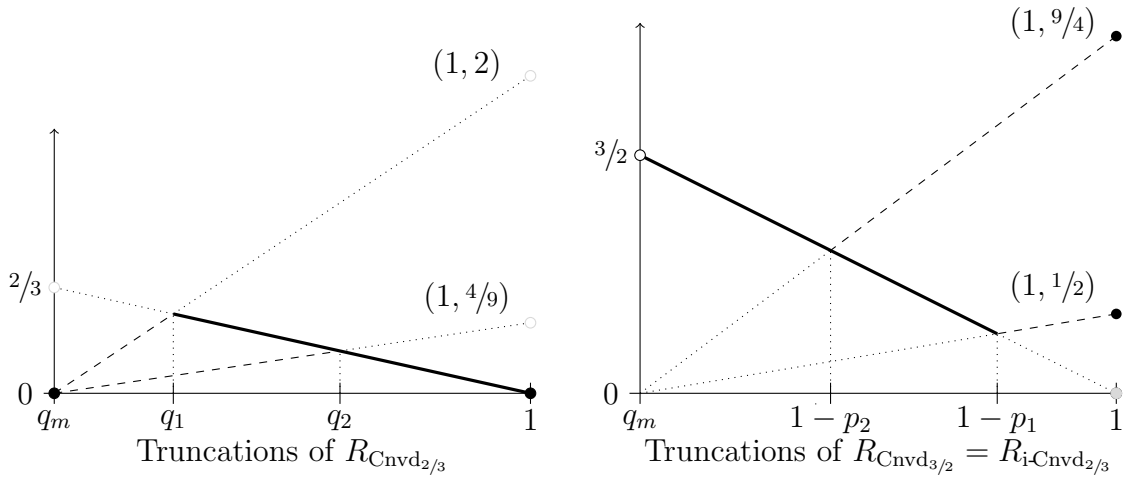

Figure A.6. Example Ironings of Shifted-Quadratics and (Same-Class) Inverses
The left figure shows the results of two examples of (dashed) top-truncation of the distribution $\text{Cnvd}_{2/3}$, respectively at $q_1 = 0.25$ and $q_2 = 0.6$. The right figure shows the results of the corresponding two examples of (dashed) bottom-truncation of the distribution $\text{i-Cnvd}_{2/3} = \text{Cnvd}_{3/2}$, respectively at $1 - p_1 = 1 - q_1 = 0.75$ and $1 - p_2 = 1 - q_2 = 0.4$. The heights of the points on the right describe the *values* of the respective truncations.

It will be convenient to describe the CNVV distributions by truncation parameters z rather than by virtual value ϕ. (This is also an exact analogy to the previous section which used z in place of hazard rate β). For the purpose of designing a dual blend, we set up respective classes of Top-Truncated CNVVs and Bottom-Truncated CNVVs. Fix in advance top-truncation quantile $\bar{q}$ and bottom-truncation percentile $\bar{p} = \bar{q}$ (so perecentile $\bar{p}$ has *quantile* $1 - \bar{q}$). We design a blend for every *arbitrary pair* of truncation points $\bar{q}$ and $\bar{p} = \bar{q}$. Thus the example of this section describes an entire class of dual blends, one for every fixed $\bar{q} \in (0, 1)$.

The set up of our Top-Truncated-CNVVs-versus-Bottom-Truncated-CNVVs blends meets the conditions of Theorem 14, therefore we immediately have the following corollary without need for a proof. However, we still work out the exact solution in order to illustrate the unique structure of this special example in which distributions of the first side of the dual blend (CNVVs as a class) are their own class of inverse-distributions. We also again consider this a step for technical completeness on the path to describing finite-weight blends. Therefore we calculate the function $g(\cdot)$ for the infinite-weight blend. The weights are $o_z = \omega_z = {}^1\!/_z \cdot dz$.[13]

Corollary 9. *Fix $\bar{q} = \bar{p} \in (0, 1)$. Given the class of downward-closed Top-Truncated-CNVVs with members $\overleftarrow{\text{Cnvd}}_\phi^{z'}$ and the class of upward-closed Bottom-Truncated-CNVVs with members $\overrightarrow{\text{i-Cnvd}}_\phi^{z'}$, each class including all $z > 0$ (equivalently all $\phi < 0$). For*

[13] The following note is critical to explain our choice regarding use of notation here. Similar to the previous section, we abuse notation slightly and describe distributions as parameterized by both ϕ and z even though there is an exact functional relationship between them for each side of the dual blend. As noted above, both sides of the blends are composed of distributions that are modified from CNVVs. *Each side of the blend will use a different functional identity to map between ϕ and z. To make this clear, we will explicitly use* i-Cnvd *distributions on the second side of the blend,* even though we could technically write down each side using CNVVs directly. The second side is the o_z-weights, bottom-truncated, upward-closed-distributions side.

$o_z = \omega_z = 1/z \cdot dz$ and $n = 2$, we have the following dual blends matching up at every $\boldsymbol{v} = (v_1, v_2 \leq v_1)$ to describe a common function g:

$$\text{(A.38)} \quad \int_0^{v_2} \frac{1}{z} \cdot \overleftarrow{\text{cnvd}}_\phi^{z'}(v_1) \cdot \overleftarrow{\text{cnvd}}_\phi^{z'}(v_2)\ dz = g(\boldsymbol{v}) = \int_{v_1}^{\infty} \frac{1}{z} \cdot \overrightarrow{\text{i-cnvd}}_\phi^{z'}(v_1) \cdot \overrightarrow{\text{i-cnvd}}_\phi^{z'}(v_2)\ dz$$

We next give the (infinite-weight) mass calculations of the dual blends for arbitrary fixed $\bar{q} \in (0, 1)$ and $\bar{p} = \bar{q}$, which effectively re-prove Corollary 9.

Downward-closed quantile-$\bar{q}$ Top-Truncated-CNVVs. For fixed quantile $\bar{q}$, first we show the substitution identity to replace the (negative) virtual value parameter ϕ with truncation parameter z, after which we present the truncated distributions as components of the Top-Truncated-CNVVs blend.

$$\bar{q} = 1 - \text{Cnvd}_\phi(z) = \frac{-\phi}{z - \phi}$$
$$-\phi = \left(\frac{q}{1-q}\right) \cdot z$$

The description of the distributions is

$$\text{(A.39)} \quad \text{(CDF)} \quad \overleftarrow{\text{Cnvd}}_\phi^{z'}(x) = 1 - \frac{\bar{q}z}{((1-\bar{q})x + \bar{q}z)} \quad \text{on } (0, z), \quad \overleftarrow{\text{Cnvd}}_\phi^{z'}(z) = [1 - \bar{q}, 1]$$

$$\text{(A.40)} \quad \text{(PDF)} \quad \overleftarrow{\text{cnvd}}_\phi^{z'}(x) = \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})x + \bar{q}z)^2} \quad \text{on } (0, z), \quad \overleftarrow{\text{cnvd}}_\phi^{z'}(z) = \bar{q} \text{ point mass}$$

The weights are ω_z. There are three measures of mass which are the same as the previous section: 1-D mass dz described by function g_{1D}; 2-D mass $dv_1 dv_2$ described by function g_{2D}; and 2-D mass $dz dv_i$ described by function g_{2Dz}. First, considering 1-D mass dz,

calculations at every point (z, z) for $z > 0$ are given by:

$$\begin{aligned} g_{1D}(z,z) &= \text{resulting mass of double-point-mass draws from } \overleftarrow{\text{Cnvd}}_{\phi}^{z'} \\ &= \omega_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftarrow{\text{Cnvd}}_{\phi}^{z'} \right] \right)^2 = \frac{1}{z} \cdot \bar{q}^2 \cdot dz \end{aligned} \tag{A.41}$$

Second, considering standard 2-D mass $dv_1 dv_2$, calculations at every point are described as an integral form by:

$$\begin{aligned} g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws across the } \overleftarrow{\text{Cnvd}}_{\phi}^{z'} \text{ blend} \\ &= \int_{v_1}^{\infty} \omega_z \cdot \overleftarrow{\text{cnvd}}_{\phi}^{z'}(v_1) \cdot \overleftarrow{\text{cnvd}}_{\phi}^{z'}(v_2) \\ &= \int_{v_1}^{\infty} \frac{1}{z} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_1 + \bar{q}z)^2} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_2 + \bar{q}z)^2} \, dz \end{aligned} \tag{A.42}$$

for which technical evaluation is deferred to Appendix B.7, where we also confirm that it exactly matches this mass-measure for the inverse-CNVVs side. Third, considering 2-D mass $dz dv_i$, calculations at every point are given by:

$$\begin{aligned} g_{2Dz}(v_1 = z, v_2) &= \text{resulting mass of "one each" draws from } \overleftarrow{\text{Cnvd}}_{\phi}^{z'} \\ &= \omega_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overleftarrow{\text{Cnvd}}_{\phi}^{z'} \right] \right) \cdot \overleftarrow{\text{cnvd}}_{\phi}^{z'}(v_2) \\ &= \omega_{v_1} \cdot \bar{q} \cdot \overleftarrow{\text{cnvd}}_{\phi}^{v_1'}(v_2) = \frac{\bar{q}}{v_1} \cdot \left(\frac{\bar{q}(1-\bar{q})v_1}{((1-\bar{q})v_2 + \bar{q}v_1)^2} \right) \cdot dz \end{aligned} \tag{A.43}$$

Upward-closed percentile-$\bar{p}$ Bottom-Truncated-CNVVs. For fixed percentile $\bar{p} = \bar{q}$, first we show the substitution identity to replace the (negative) virtual value parameter ϕ with truncation parameter z, after which we present the truncated distributions as

components of the Bottom-Truncated-CNVVs blend.

$$\bar{p} = \text{i-Cnvd}_{\phi}(z) = \frac{-\phi z}{1-\phi z}$$
$$-\phi = \left(\frac{\bar{p}}{1-\bar{p}}\right) \cdot \frac{1}{z}$$

The description of the distributions is

(A.44)

$$\text{(CDF)} \quad \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi}(z) = [0, \bar{p}], \qquad \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi}(x) = \frac{\bar{p}x}{((1-\bar{p})z + \bar{p}x)} \quad \text{on } (z, \infty)$$

(A.45)

$$\text{(PDF)} \quad \overrightarrow{\text{i-cnvd}}^{z'}_{\phi}(z) = \bar{p} \text{ point mass}, \quad \overrightarrow{\text{i-cnvd}}^{z'}_{\phi}(x) = \frac{\bar{p}(1-\bar{p})z}{((1-\bar{p})z + \bar{p}x)^2} \quad \text{on } (z, \infty)$$

The weights are o_z. There are three measures of mass which are the same as the previous section: 1-D mass dz described by function g_{1D}; 2-D mass $dv_1 dv_2$ described by function g_{2D}; and 2-D mass $dz dv_i$ described by function g_{2Dz}. First, considering 1-D mass dz, calculations at every point (z, z) for $z > 0$ are given by:

$$g_{1D}(z,z) = \text{resulting mass of double-point-mass draws from } \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi}$$

(A.46)
$$= o_z \cdot \left(\Pr\left[X = z \text{ for } X \sim \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi}\right]\right)^2 = \frac{1}{z} \cdot \bar{p}^2 \cdot dz$$

which matches equation (A.41). Second, considering standard 2-D mass $dv_1 dv_2$, calculations at every point are described as an integral form by:

$$\begin{aligned} g_{2D}(\boldsymbol{v}) &= \text{resulting mass of double-continuous draws across the } \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi} \text{ blend} \\ &= \int_0^{v_2} o_z \cdot \overrightarrow{\text{i-cnvd}}^{z'}_{\phi}(v_1) \cdot \overrightarrow{\text{i-cnvd}}^{z'}_{\phi}(v_2) \\ &= \int_0^{v_2} \frac{1}{z} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\,z + \bar{p}v_1\right)^2} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\,z + \bar{p}v_2\right)^2}\, dz \end{aligned} \tag{A.47}$$

for which technical evaluation is deferred to Appendix B.7, where we also confirm that it exactly matches equation (A.42). Third, considering 2-D mass $dz dv_i$, calculations at every point are given by:

$$\begin{aligned} g_{2Dz}(v_1, v_2 = z) &= \text{resulting mass of "one each" draws from } \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi} \\ &= o_z \cdot \overrightarrow{\text{i-cnvd}}^{z'}_{\phi}(v_1) \cdot \left(\Pr\left[X = z \text{ for } X \sim \overrightarrow{\text{i-Cnvd}}^{z'}_{\phi}\right]\right) \\ &= o_{v_2} \cdot \overrightarrow{\text{i-cnvd}}^{v_2'}_{\phi}(v_1) \cdot \bar{p} = \frac{\bar{p}}{v_2} \cdot \left(\frac{\bar{p}(1-\bar{p})v_2}{\left((1-\bar{p})\,v_2 + \bar{p}v_1\right)^2}\right) \cdot dz \end{aligned} \tag{A.48}$$

which exactly matches equation (A.43).[14] With Appendix B.7, this completes the blending of quantile-$\bar{q}$ Top-Truncated CNVVs against percentile-$\bar{p}$ Bottom-Truncated CNVVs.

[14] We make special note of a comparison of calculations of mass measure $dz dv_i$ between this CNVVs example and the Exponentials example earlier. We look at the "computational effect" in the blends calculations of the choice of weights $\omega_z = o_z = {}^1\!/_z \cdot dz$. *Before inserting weights functions here*, we have $\bar{q} = \bar{p}$ and know that we need weights functions in order that line (A.43) and line (A.48) are equal:

$$\bar{q} \cdot \omega_{v_1} \cdot \frac{\bar{q}(1-\bar{q})v_1}{\left((1-\bar{q})\,v_2 + \bar{q}v_1\right)^2} = \bar{p} \cdot o_{v_2} \cdot \frac{\bar{p}(1-\bar{p})v_2}{\left((1-\bar{p})\,v_2 + \bar{p}v_1\right)^2}$$

so in fact, the assignment of $\omega_z = o_z = \frac{1}{z}dz$ was necessary, as these weights are the only possible free parameters to fix the existing mismatch in numerators. Thus, weights functions here are used on both sides of the dual blend to cancel existing mismatched terms. This is in fact qualitatively different than what was observed in Exponentials line (A.30) and line (A.36) in which respectively the first ω_z-side used the weight function to add a factor of ${}^1\!/_{v_1}$ and the second o_z-side used the weight function to cancel an existing factor of v_2. These distinct effects of the common weights of Theorem 14 deserve further study.

A.5. Prior Free Benchmarks and Related Work in Mechanism Design

This section introduces common benchmark functions and gives a short survey of the mechanism design literature to highlight the most pertinent results in prior free analysis. Competitive analysis was introduced to the design of mechanisms by Goldberg et al. (2006). Each result includes the benchmark for which its statement holds.

A.5.1. Common Benchmark Functions in Mechanism Design

The concept of benchmark design and optimization is relatively new. This section presents commonly employed benchmarks as they exist in the historical literature for prior free mechanism design. Historical results are deferred to the next section. Past benchmarks were somewhat ad hoc – they were instantiated with natural structures from best efforts and ended up with some nice properties, but were not optimized for such. The four benchmarks presented are offline optimal, optimal-Bayesian-optimal, ex-feasible price-posting, and ex-feasible price-posting-(l). Of these, the first two benchmarks have feasibility constraints of the setting built into the benchmark, but the last two ignore them.[15] Their descriptions are given next, along with related context.

(1) The *offline optimal benchmark* (OO) is defined by the (deterministic) performance of the optimal auction that knows the agents' values. For single-item revenue, $\text{OO}(\boldsymbol{v}) = v_{(1)}$, the largest value. This function also maximizes welfare, and is

[15] The "ex-feasible" benchmarks ignore implicit feasibility constraints of any setting and make the most sense for "digital goods" settings (in which an auction with n unit-demand agents has n items to sell). We define these benchmarks even though work in this thesis focuses exclusively on the single-item case. However benchmarks are simply functions, so benchmarks designed for digital goods settings are of course still computable in the single-item case. Some of the results that we will give in Appendix A.5.2 are for the digital goods settings, in order to round out (a) historical context of the prior free setting and (b) illustration of the critical behavior of equal revenue distributions.

also named the *welfare benchmark*. For single-item residual surplus, $\text{OO}(\boldsymbol{v}) = \max_{k\in\{2,\dots,n+1\}} \left(\sum_{j=1}^{k-1} v_{\{j\}} - v_{\{k\}}\right)/(k-1)$ by optimizing k.

(2) The *optimal-Bayesian-optimal benchmark* (OBO) is parameterized by a class of distributions $\mathcal{F}$ and is defined for an input $\boldsymbol{v}$ as the **optimal** (maximum) expected performance of any mechanism that is **(Bayesian) optimal** for a distribution in $\mathcal{F}$ (Hartline and Roughgarden, 2008). This definition is otherwise independent of auction objective, and is given by $\text{OBO}^{\mathcal{F}}(\boldsymbol{v}) = \max_{F\in\mathcal{F}} \text{OPT}_F(\boldsymbol{v})$.

(3) The *ex-feasible price-posting benchmark* is set by the (offline) optimal performance of posting any single, constant price to all agents and wholly selling to all agents who accept. This benchmark is also frequently named the *envy-free optimal benchmark* (EFO) in the literature (Hartline and McGrew, 2005; Devanur et al., 2015).[16] We adopt this notation 'EFO' but use both names based on context (Chapter 1 gave an example using the price-posting-revenue benchmark). For revenue, $\text{EFO}(\boldsymbol{v}) = \max_{k\in\{1,\dots,n\}} k \cdot v_{\{k\}}$. For residual surplus, $\text{EFO}(\boldsymbol{v}) = \sum_k v_k$, which reduces to a trivial *sum-of-values benchmark*.

(4) The *ex-feasible price-posting-*(l) *benchmark* – alternatively *envy-free-*(l) *benchmark* ($\text{EFO}^{(l)}$) – has $l \geq 2$ (and most typically $l = 2$) and is the EFO benchmark with a quota or restriction l on the number of items sold. For revenue, $\text{EFO}^{(l)}(\boldsymbol{v}) = \max_{k\in\{l,\dots,n\}} k \cdot v_{\{k\}}$ maximizes EFO when required to sell to at least l agents. For residual surplus, $\text{EFO}^{(l)}(\boldsymbol{v}) =$
$\max_{k\in\{l+1,\dots,n+1\}} l \cdot \left(\sum_{j=1}^{k-1} v_{\{j\}} - v_{\{k\}}\right)/(k-1)$ maximizes EFO (using price k)

[16] An "envy-free" outcome-based computation requires restriction to theoretical allocations in which no agent i prefers the outcome (x_j, π_j) of any other agent j to (x_i, π_i). With "ex-feasible," this reduces to digital goods anonymous price-posting.

when required to (wholly) sell to at most l agents (and inherently always sells to l agents).

Fact 24 (Hartline and Roughgarden, 2008)**.** *For revenue, the offline optimal (OO), optimal-Bayesian-optimal (OBO), and and envy-free (EFO) benchmarks are all* normalized *benchmarks (Definition 31).*

A.5.2. A Selection of Historical Prior Free Results

This section presents a list of results from the historical literature on prior free auction design. Each statement fully describes its setting and result, with minimal overall context. The list reflects the references listed in a blog post on *Turing's Invisible Hand* (Hartline, 2014) which motivated much of the work in this thesis. In the theorem statements that follow, benchmark references include an index corresponding to the list in the previous section if applicable. The common theme across the results is that *equal revenue distributions* can be used to prove that their respective approximation guarantees are tight. This observation is summarized in discussion culminating with Fact 25 at the end of the section.

Theorem 21 (Hartline and Roughgarden, 2014)**.** *Given the single-item setting with* 1 *agent with value support* $[1, h]$*, a revenue objective, and the offline optimal benchmark* (1)*. The optimal auction posts a random price from the distribution with cdf* $F(x) = (1 + \ln x)/(1 + \ln h)$ *(which embeds a point mass at* $x = 1$*). Its approximation factor is* $1 + \ln h$*.*

Theorem 22 (Fiat et al., 2002)**.** *Given the digital goods setting with* 2 *agents with value support* $[1, \infty)$*, a revenue objective, and the EFO*$^{(2)}$ *benchmark* (4)*. The optimal auction is the second price auction. Its approximation factor is* 2*.*

Theorem 23 (Hartline and McGrew, 2005)**.** *Given the digital goods setting with* 3 *agents with value support* $[0, \infty)$*, a revenue objective, and the EFO*$^{(2)}$ *benchmark* (4)*. The optimal auction offers a random price to each agent* i *based on the values of other agents* v_j *and* $v_k \leq v_j$ *without loss of generality. The following description is* scale-invariant *based on re-scaling* $v_{(1)} = v_j/v_k$ *and* $v_{(2)} = 1$*. The "final" price offered to each agent* i *is a draw from the following conditional density function* f *(for which sub-function* $\eta(x) = \frac{2/13}{(x-1)^3}$*), re-scaled back up by the factor* v_k*.*

- *If* $v_{(1)} \leq 3/2$*,* f *has a point mass at* $x = 1$ *with weight* $9/13$*, and otherwise has density* $\eta(x)$ *on range* $x \in (3/2, \infty)$*.*
- *If* $v_{(1)} > 3/2$*,* f *has:*
  - *a point mass at* $x = 1$ *with weight* $9/13 - \int_{3/2}^{v_{(1)}} z \cdot \eta(z)dz$*;*
  - *a point mass at* $x = v_{(1)}$ *with weight* $\int_{3/2}^{v_{(1)}} (z+1)\eta(z)dz$*;*
  - *and otherwise has density* $\eta(x)$ *on range* $x \in (v_{(1)}, \infty)$

Its optimal approximation factor is $13/6$*.*

Theorem 24 (Hartline and Roughgarden, 2014)**.** *Given a* 1*-item auction with* $n = 2$ *agents with value support* $[0, \infty)$*, a residual surplus objective, and the OBO benchmark* (2) *parameterized by the class of all distributions* $\mathcal{F}^{\text{all}}$*. The optimal auction is as follows.*

- *If* $v_{(1)} \leq 2 \cdot v_{(2)}$*, then allocate each agent with probability* $1/2$*;*

- *Otherwise, allocate the larger-valued agent (with $v_{(1)}$) with probability* $3/4$ *and the smaller-valued agent with probability* $1/4$.

Its optimal approximation factor is $4/3$.

The following abstract result characterizes approximation factors for a *large class* of benchmarks, without explicit specification of the auction corresponding to a benchmark. As Chen et al. (2014) write, their "characterization is provided by the set of inequalities [given in (A.49) below] that only involves a [benchmark B and a ratio α], but does not describe an actual auction. This is similar in spirit to the characterization of truthfully implementable allocation functions [described in Theorem 1 of this thesis]... that only specifies an allocation function but ... without describing any payment function, one can determine whether there is a truthful auction with a specified allocation."

Theorem 25 (Chen et al., 2014)**.** *Given the digital goods setting with n agents with value support* $[1, \infty)$*, a revenue objective, and* any non-negative and increasing benchmark function B. *There is a truthful auction that achieves approximation factor α if and only*

$$\int_{\mathcal{S}} B(\boldsymbol{v}) \cdot \text{eqrd}(\boldsymbol{v}) d\boldsymbol{v} \leq \alpha \cdot \sum_{k=1}^{n} \int_{\mathcal{S}_{\downarrow k}} \text{eqrd}(\boldsymbol{v}_{-k}) d\boldsymbol{v}_{-k} \tag{A.49}$$

where $\mathcal{S} \in \mathcal{V}$ is any upward closed subset of value space, $\mathcal{S}_{\downarrow k}$ is the projection of $\mathcal{S}$ into value space excluding dimension k, and $\text{eqrd}(\boldsymbol{v}) = \prod_k 1/v_k^2$ *is the density at $\boldsymbol{v}$ as if the inputs were drawn i.i.d. from the* equal revenue distribution *(of Definition 14).*

Theorem 25 makes possible the following corollary, thereby positively confirming a conjecture of Goldberg et al. (2004).

Corollary 10 (Chen et al., 2014)**.** *Given the digital goods setting with n agents with value support $[1, \infty)$, a revenue objective, and the $EFO^{(2)}$ benchmark. The optimal prior free approximation factor as $n \to \infty$ is ~ 2.42 (with the worst-case (largest) approximation factor of any n achieved in this limit).*

For a fixed benchmark, the problem of identifying the optimal mechanism is generally not well understood. However there is a canonical method for identifying lower bounds on the approximation ratio of the optimal mechanism which motivated this list of results. The sequence of varied results above exhibits the critical role of the equal revenue distribution in optimal mechanism design, illustrating a technique first highlighted by Hartline and McGrew (2005) and Goldberg et al. (2006). They suggested that the prior free approximation of any mechanism for a benchmark can be lower bounded by identifying a distribution over inputs for which all "non-dominated" mechanisms have the same (expected) performance. These "neutralizing" distributions are exactly the EQRs for revenue and EQRSs for residual surplus that were introduced in Section 2.2.8.

We make this formal as follows, using revenue as an example. The ratio of the expected benchmark to the revenue of any mechanism for the EQR distribution gives a lower bound on the prior free approximation of the benchmark, i.e., for the optimal prior free mechanism there must exist an input where the ratio of benchmark to mechanism revenue is at least the ratio of their expectations for the distribution. E.g., for digital goods revenue auctions and a large class of benchmarks, in order to prove Theorem 25 above, (Chen et al., 2014) give a non-constructive proof that this lower bound is tight. In fact, we generalized this lower bound for arbitrary i.i.d. distributions in Lemma 15 in Section 6.1.4.

Fact 25. *The optimal approximation factors guaranteed in the statements of Theorem 21, Theorem 22, Theorem 23, Theorem 24, Theorem 25 can be proved tight by using Lemma 15, i.e., by assuming that value inputs* $\boldsymbol{v}$ *are drawn i.i.d. from* $F = \text{Eqrd}$ *(respectively* $F = \text{Eqrsd}$ *for the residual surplus result); and subsequently showing that for each, the (respectively) optimal mechanism's expected approximation given these inputs is equal to a (respective) proven upper bound guarantee on its* α^B*.*

A.6. Fundamentals of Normalization-Symmetric Resolution

This section expands our foundational analysis of the normalization-symmetric resolution measure and its corresponding nsBDP. The definition of its benchmark design problem is copied here from Section 6.1.4 for direct reference.

Definition 35 (Hartline, 2020)**.** *The* normalization-symmetric resolution benchmark design problem (nsBDP) *is given by a class of distributions* $\mathcal{F}$*, a class of algorithms* $\mathcal{A}$*, and resolution measure* σ^B*; and searches for the* argmin *of the* $\min - \max$ *program*

$$(\gamma_\sigma^{\mathcal{F}}) \qquad \gamma_\sigma^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\sigma^B\right] = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\max_{F \in \mathcal{F}} \frac{B(F)}{\text{OPT}_F(F)}\right]$$

Appendix A.6.1 motivates the nsBDP from the historical literature and identifies a natural relaxation to a related benchmark design problem using: *normalization-assymetric-2* resolution measure (na2BDP; measure σ_2). Appendix A.6.2 connects both of these problems to dual blends and bounds from the Blends Technique by writing them as linear programs. Appendix A.6.3 relates both problems to the brBDP by proving Proposition 7 which states that $\gamma_\sigma^{\mathcal{F}} \leq \gamma_{\sigma_2}^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$ (and which subsumes Proposition 5 first given in Section 6.1.4).

A.6.1. Historical Perspective and Extension to Normalization-asymmetric-2

Recall Lemma 15 from Section 6.1.4.

Lemma 15 (Goldberg et al., 2006)**.** *For any benchmark B, the class of distributions $\mathcal{F}^{\text{all}}$ (given input space $\mathcal{V}^n$), and class of algorithms $\mathcal{A}$ which induce OPT_F, the optimal prior free approximation α^B is at least*

$$(\breve{\alpha}^B) \qquad \breve{\alpha}^B = \max_{F \in \mathcal{F}^{\text{all}}} \frac{B(F)}{\text{OPT}_F(F)}$$

With respect to the lower bound of Lemma 15, (a corollary of) the Hartline and McGrew (2005) result stated as Theorem 23 and the Chen et al. (2014) result stated as Theorem 25 showed in fact for digital goods auctions that $\alpha^B = \breve{\alpha}^B$ for a large class of benchmarks (in particular using $F^* = \text{Eqrd}$ as the Equal Revenue Distribution). Note further from Theorem 16, the optimal best-response resolution benchmark also satisfies a slightly-modified version of Lemma 15 with equality (when the comparison class $\mathcal{F}^{\text{all}}$ is replaced in Lemma 15 to be possibly-restricted to the brBDP's fixed-parameter class $\bar{\mathcal{F}} \subseteq \mathcal{F}^{\text{all}}$).

Recall from Section 6.3 that we identified challenges to indiscriminate use of the brBDP, which suggests that we might need a less-precise measure of resolution. A natural relaxation of the brBDP is – instead of optimizing benchmarks that admit the best prior free approximation which tailors too much to the interests of the subsequent algorithm design – to optimize benchmarks that admit the best lower bound of Lemma 15.

First we consider such a relaxation from a simpler version of Lemma 15 which only searches over a class $\bar{F}$ rather than F^{all}. The idea is: if a prior independent approximation guarantee against a specific class of distributions $\bar{\mathcal{F}}$ is such a focused objective of a given

prior free benchmark design and algorithm design task, then perhaps it makes sense to implement the relaxed benchmark via resolution lower bound only with respect to the class $\bar{\mathcal{F}}$, rather than with respect to $\mathcal{F}^{\text{all}}$. The resulting lower-bound-technique in this case induces exactly the normalization-symmetric resolution σ^B, and the resulting benchmark design program is exactly the nsBDP for $\bar{\mathcal{F}}$.[17] Further, this relaxed design induces a natural and elegant symmetry in the linear program description of the nsBDP (see Linear Program 6).

However, an incidental benchmark design – which need not use normalization to guarantee approximation to distributions $F^{\text{out}} \notin \mathcal{F}$ but which wants to prohibit possibly-spurious benchmark designs by regulating the expected benchmark over F^{out} to not be too large – might instead keep the entire class of distributions $\mathcal{F}^{\text{all}}$ against which to measure resolution per the lower bound of Lemma 15. This motivates a new, benchmark design program with measure normalization-asymmetric-2 resolution σ_2^B. Changes from the nsBDP are **bolded**. (In the equation line, the only change is that the inner-maximization problem changes its feasible space from $F \in \mathcal{F}$ to $F \in \mathcal{F}^{\text{all}}$. Thus, the nsBDP and na2BDP are the same for nsBDP problems that are canonically defined for the class $\mathcal{F}^{\text{all}}$, e.g., our residual surplus auctions.)

Definition 38. *The* normalization-asymmetric-2 resolution benchmark design problem (na2BDP) *is given by* ***a value space $\mathcal{V}^n$ and the class of all distributions $\mathcal{F}^{\text{all}}$ (over $\mathcal{V}^n$)****, a class of distributions $\mathcal{F}$ for normalization, a class of algorithms $\mathcal{A}$, and*

[17] Historically, there was intuition (following from Chen et al. (2014)) that universally $\gamma_\sigma^{\mathcal{F}} = \alpha^{\mathcal{F}}$, i.e., that the nsBDP might always have its optimal resolution equal to optimal prior independent approximation (cf. Theorem 16). This thesis identified the best-response resolution measure and generally assessed this intuition to be misguided (Theorem 17 and discussion). There still may exist restricted settings for which $\gamma_\sigma^{\mathcal{F}} = \alpha^{\mathcal{F}}$ holds.

resolution measure σ_2^B*; and searches for the* argmin *of the* $\min - \max$ *program*

$$(\gamma_{\sigma_2}^{\mathcal{F}}) \qquad \gamma_{\sigma_2}^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\sigma_2^B\right] = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\max_{F \in \mathcal{F}^{\text{all}}} \frac{B(F)}{\text{OPT}_F(F)}\right]$$

A.6.2. The Linear Programs and Connection to the Blends Technique

This section uses techniques of Vohra (2011) to write linear programs for the nsBDP (Definition 35) and the na2BDP (Definition 38 immediately above). The two programs are almost identical to each other but we write both for clarity and comparison.

The following applies for both programs. Let $B(\boldsymbol{v})$ be the assigned output of the benchmark function B at $\boldsymbol{v}$, thus the full set of variables for assignment is $\boldsymbol{B} = \{B(\boldsymbol{v}) \ : \ \boldsymbol{v} \in \mathcal{V}^n\}$ The normalization property is naturally a linear constraint on benchmark design. The inner-minimization of resolution is moved to be a constraint and we optimize the resolution measure directly as a variable (cf. Linear Program 1). It is impossible to have resolution smaller than 1 and this fact is copied into the objective function line.

The dual variables associated with each constraint are given along with constraint descriptors. *For the nsBDP only*, the dual variables are $\boldsymbol{\omega} = \{\omega_F \ : \ F \in \mathcal{F}\}$ and $\boldsymbol{o} = \{o_F \ : \ F \in \mathcal{F}\}$.

Linear Program 6 (The Normalization-symmetric Benchmark Design Program[18])**.** *Given a prior free algorithm design setting, a class of distributions* $\mathcal{F}$*, a class of algorithms* $\mathcal{A}$ *(which induce* OPT_F *per* $F \in \mathcal{F}$*), and the resolution measure* σ^B*, the optimal normalized benchmark and its optimal resolution* $\gamma_\sigma^{\mathcal{F}}$ *are given by the* argmax *of the following*

[18] Observe the symmetry of the constraints which motivates the name of the resolution measure.

program:[19]

$$\gamma_{\sigma}^{\mathcal{F}} = \min_{\boldsymbol{B},\ \sigma} \sigma \ (\geq 1) \tag{A.50}$$

s.t.

$$\int_{\mathcal{V}^n} B(\boldsymbol{v}) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \geq \mathrm{OPT}_F(F) \qquad \forall\ F \in \mathcal{F} \qquad (\omega_F,\ \text{normalization})$$

$$\int_{\mathcal{V}^n} B(\boldsymbol{v}) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \leq \sigma \cdot \mathrm{OPT}_F(F) \qquad \forall\ F \in \mathcal{F} \qquad (o_F,\ \text{ns-resolution})$$

The following linear programs describe the na2BDP. In comparison to the nsBDP, the only alterations are as follows. The resolution constraint in the primal is generalized to hold for each distribution $F \in \mathcal{F}^{\text{all}}$. Correspondingly, there exist generalized dual variables $\boldsymbol{o} = \{o_F \ :\ F \in \mathcal{F}^{\text{all}}\}$.

Linear Program 7 (The Normalization-asymmetric-2 Benchmark Design Program)**.** *Given a prior free mechanism design setting, a value space $\mathcal{V}$ with $\mathcal{F}^{\text{all}}$, a normalization class of distributions $\mathcal{F}$, and a class of mechanisms which induce* OPT_F*, the optimal normalized benchmark with respect to resolution measure σ_2 and its value $\gamma_{\sigma_2}^{\mathcal{F}}$ are given by the* argmax *of the following program:*

$$\gamma_{\sigma_2}^{\mathcal{F}} = \min_{\boldsymbol{B},\ \sigma_2} \sigma_2 \ (\geq 1) \tag{A.51}$$

s.t.

[19] The left-hand side of each approximation constraint is equivalently described as *expected benchmark*: $\int_{\mathcal{V}^n} B(\boldsymbol{v}) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} = \mathbf{E}_{\boldsymbol{v}\sim F}\left[B(\boldsymbol{v})\right]$. The variables of the program are more explicit from the integral form (cf. footnote 1 on page 191).

$$\int_{\mathcal{V}^n} B(\boldsymbol{v}) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \geq \mathrm{OPT}_F(F) \qquad \forall\ F \in \mathcal{F} \qquad (\omega_F,\ \text{normalization})$$

$$\int_{\mathcal{V}^n} B(\boldsymbol{v}) \cdot \boldsymbol{f}(\boldsymbol{v})\ d\boldsymbol{v} \leq \sigma_2 \cdot \mathrm{OPT}_F(F) \qquad \forall\ F \in \mathcal{F}^{all} \qquad (o_F,\ \text{ns-resolution})$$

Without going into detail, both Linear Program 6 and Linear Program 7 are clear analogies to Linear Program 4 on page 194. Linear Program 4 was the Appended-Relaxed-Simplified (ARS) modification of Linear Program 1 for the canonical description of optimal prior independent single-item mechanism design, in particular, Linear Program 4 had stripped out the constraints and details of mechanism design and replaced the variables with pseudo-performance variables. Here we have non-negative benchmark functions in place of pseudo-performance, a multiplicatively-inverted objective, and a possible restriction of both sides to $\mathcal{F}$ (only in the nsBDP). However the dual programs have essentially identical structure and behavior to the dual-of-the-ARS program in Linear Program 5, for which our dual blends Definition 18 were the exact class of feasible solutions.

Moreover, the programs in this section are original formulations of benchmark design problems (i.e., they are not "Appended-Relaxed-Simplified" from a more-involved problem description). Thus, not only do dual blends give lower bounds on Linear Program 6 and Linear Program 7 from the dual programs (Corollary 11), we should expect them to provide (or approach within a limiting sequence) *tight lower bounds* to the values of the primal programs.

In facxt, the nsBDP (Linear Program 6) is an interesting case in comparison to the na2BDP (Linear Program 7) and the ARS-program (Linear Program 4) because it is required to only use distributions in $\mathcal{F}$ to construct *both sides of its dual blends.* This

represents a possibly-material restriction on the nsBDP adversary. Consider the following dual blends examples specifically for an auction with a revenue objective restricted to regular distributions $\mathcal{F}^{\text{reg}}$. Cf., our finite weight Quadratics-versus-Uniforms example of Section 4.5 may be used by any of the programs because all of the Quadratics and all of the Uniforms are regular and thus equally available within $\mathcal{F}^{\text{reg}}$ and $\mathcal{F}^{\text{all}}$. Alternatively cf., the example dual blends of Appendix A.3 and Appendix A.4 both incorporated bottom-truncated irregular distributions and thus may only be used by the na2BDP and ARS programs (and not by the nsBDP program). This gap may be significant because irregular distributions are known to be difficult to approximate in prior independent settings (indeed this is why auctions for revenue restrict to $\mathcal{F}^{\text{reg}}$). Within the nsBDP, the adversary may not design the dual blend to use irregulars for the algorithm-upper-bound that is inherent to the calculation of the denominator $\text{opt}_{n,1}$ in the Blends Technique.

Two interesting open questions are:

(1) What is the characterization of the set of algorithm problems – if any – for which the nsBDP and the na2BDP are equivalent?

(2) Where the programs are not equivalent, how does behavior of the optimal algorithm differ to take advantage of the restricted adversary in the nsBDP program?

We end this section with the statement that dual blends give lower bounds to our benchmark design programs here.

Corollary 11. *Consider the nsBDP. Assume there exist two distinct dual blends* $\delta_1 \in \Delta(\mathcal{F})$ *and* $\delta_2 \in \Delta(\mathcal{F})$ *and correlated density function* g *(of Definition 18) such that:*

$$\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v}) \quad \forall\ \boldsymbol{v}$$

Then the value of the nsBDP ($\gamma_\sigma^{\mathcal{F}}$) *is at least the ratio* $^{\text{opt}_{n,2}}/_{\text{opt}_{n,1}}$*:*

$$\gamma_\sigma^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\max_{F \in \mathcal{F}} \frac{B(F)}{\text{OPT}_F(F)} \right] \geq \frac{\text{opt}_{n,2}}{\text{opt}_{n,1}}$$

Alternatively, consider the na2BDP setting. Assume there exist two distinct dual blends $\delta_1 \in \Delta(\mathcal{F}^{all})$ *and* $\delta_2 \in \Delta(\mathcal{F})$ *and correlated density function* g *such that:*

$$\delta_1^n(\boldsymbol{v}) = g(\boldsymbol{v}) = \delta_2^n(\boldsymbol{v}) \quad \forall\ \boldsymbol{v}$$

Then the value of the na2BDP ($\gamma_{\sigma_2}^{\mathcal{F}}$) *is at least the ratio* $^{\text{opt}_{n,2}}/_{\text{opt}_{n,1}}$*:*

$$\gamma_{\sigma_2}^{\mathcal{F}} = \min_{B \in \mathcal{NB}^{\mathcal{F}}} \left[\max_{F \in \mathcal{F}^{\text{all}}} \frac{B(F)}{\text{OPT}_F(F)} \right] \geq \frac{\text{opt}_{n,2}}{\text{opt}_{n,1}}$$

A.6.3. Formal Relationship to Best-response Resolution

Ostensibly, the goal of this section is to prove Proposition 5 from Section 6.1.4. Actually, we state a stronger version here as Proposition 7, specifically to include $\gamma_{\sigma_2}^{\mathcal{F}}$ in the ordering. The statement and proof effectively subsume Proposition 5.

Proposition 7. *Given a value space* $\mathcal{V}^n$ *and all distributions* $\mathcal{F}^{\text{all}}$*, a class of distributions* $\mathcal{F}$ *(for normalization) and a class of algorithms* $\mathcal{A}$*, the value of the nsBDP* $\gamma_\sigma^{\mathcal{F}}$ *lower bounds the value of the na2BDP* $\gamma_{\sigma_2}^{\mathcal{F}}$ *which lower bounds the value of the brBDP* $\gamma_\rho^{\mathcal{F}}$ *and the approximation factor of the prior independent optimal mechanism for* $\mathcal{F}$*:*

$$\gamma_\sigma^{\mathcal{F}} \leq \gamma_{\sigma_2}^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}} \tag{A.52}$$

Proof. Recall that the equality $\gamma_\rho^{\mathcal{F}} = \alpha^{\mathcal{F}}$ was shown in Theorem 16. It is included here (a) for completeness, and (b) because below we will actually show $\gamma_{\sigma_2}^{\mathcal{F}} \leq \alpha^{\mathcal{F}}$ rather than $\gamma_{\sigma_2}^{\mathcal{F}} \leq \gamma_\rho^{\mathcal{F}}$.

The first inequality follows directly from inspection of Linear Program 6 for $\gamma_\sigma^{\mathcal{F}}$ and Linear Program 7 for $\gamma_{\sigma_2}^{\mathcal{F}}$, where we see that the latter minimization problem faces exactly a super set of constraints (from the resolution-line constraints) in comparison to the former. Thus its objective can not be smaller.

To show the remaining inequality $\gamma_{\sigma_2}^{\mathcal{F}} \leq \alpha^{\mathcal{F}} = \gamma_\rho^{\mathcal{F}}$, we effectively copy the proof of Lemma 16 which stated $\gamma_\rho^{\mathcal{F}} \leq \alpha^{\mathcal{F}}$ in the first place. Specifically, let A^* be the optimal prior independent algorithm for class $\mathcal{F}$ with optimal approximation factor $\alpha^{\mathcal{F}} \geq 1$. Then assigning $\boldsymbol{B} = \bar{B}$ and $\sigma_2 = \alpha^{\mathcal{F}}$ as the arguments of the na2BDP gives a feasible solution to Linear Program 7 and thus the optimal measure $\gamma_{\sigma_2}^{\mathcal{F}}$ can not be larger than $\alpha^{\mathcal{F}}$. □

A.7. Deferred Connection between Blends and Tensor Decomposition

This section makes explicit the connection between our blends and continuous tensors as introduced in Section 5.5 on page 147. Appendix A.7.1 gives the technical basics of tensors. Appendix A.7.2 adapts the framework of tensors to interpret blends.

A.7.1. Introduction to Tensor Decomposition

This section introduces *tensors* and describes the standard question of *tensor decomposition*. An *n-th order* tensor T is the n-dimensional analogue of a 2-dimensional $m \times n$-matrix (which is a 2nd-order special-case tensor).

Consider the standard *outer product* operation '$\otimes$' of two vectors $\boldsymbol{a},\ \boldsymbol{b}$ as $\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a} \cdot \boldsymbol{b}^\top$ to output a matrix H with elements $h_{i,j} = a_i \cdot b_j$. A *rank-1* tensor can be written as[20] $\boldsymbol{a}_{1,1} \otimes \ldots \otimes \boldsymbol{a}_{1,j} \otimes \ldots \otimes \boldsymbol{a}_{1,n} = T_{\boldsymbol{A}_1}$ for a given list of vectors $\boldsymbol{a}_{1,1}, \ldots, \boldsymbol{a}_{1,j}, \ldots, \boldsymbol{a}_{1,n}$ with respective sizes $m_1, \ldots, m_n$. Thus, $T_{\boldsymbol{A}_1}$ is nth-order with dimensions-sizes corresponding to the m_i.

The *rank* of a tensor T is the minimum number of rank-1 tensors needed in a summation to yield T. Analogous to standard matrix operations, tensors can accommodate scalars as multipliers of every entry, e.g., $T' = l_1 \cdot T_{\boldsymbol{A}_1} + l_2 \cdot T_{\boldsymbol{A}_2}$. (Scalars are not strictly necessary because they can be multiplied into their respective tensors up front, so they are usually employed in situations when there are constraints on the input vectors $\boldsymbol{a}_{i,j}$ and/or the scalars themselves, e.g., positive and sum to 1.) A tensor T of rank r is expressed by $T = \sum_{z=1}^{r} l_z \cdot T_{\boldsymbol{A}_z}$, and the sum-expression is referred to as the *decomposition* of tensor T.

All of our examples of dual blends have $n = 2$ and thus are not restricted by the challenges of higher-order tensors. In fact in Section 5.1, we conjectured that there is no generalization of our classes of dual blends to $n > 3$ due to an algebraic analysis of the structure of our dual blends (in particular, from the role of the integral endpoints in the calculations of density at every $\boldsymbol{v}$). Our impossibility-conjecture is naturally analogous to uniqueness of higher-order tensor decomposition. However, for 2nd-order tensors –

[20] An explanation of tensor indexing variables: a vector $\boldsymbol{a}_{z,j}$ is effectively a column-j vector used within the computation of tensor $T_{\boldsymbol{A}_z}$; therefore, the entries of $\boldsymbol{a}_{z,j}$ have row indexes i; finally, we use the z index of a rank-1 tensor $T_{\boldsymbol{A}_z}$ to set up analogy to its usage as a blends parameter.

simply, *standard matrices* – decomposition is known to be highly non-unique due to what is called the Rotation Problem.[21]

The study of tensors in computer science is typically motivated by the existence of a tensor-object $\tilde{T}$ as a representation of high-dimensional, real-world data. As such, a tensor $\tilde{T}$ frequently incorporates "noisy" entries $\tilde{t}_{i_1,\ldots,i_n}$ and it is neither well-understood nor appropriate how to reinterpret the tensor via decomposition as an exact sum over rank-1 tensors. (Despite decomposition being generally unique (Harshman, 1972; Kruskal, 1977), determination of tensor rank is generally NP-Hard for 3rd-order (three-dimensional) tensors (Håstad, 1989).) A decomposition is desirable to act effectively as a *factor analysis* of the tensor. With an analogy to statistical factor analysis in mind, it is standard in the computer science literature to seek *approximate decompositions* of a (real-world-generated) tensor $\tilde{T}$, with an objective to optimize the approximation with respect to an accuracy measurement (e.g., the Frobenius norm), and *subject to a maximum rank* $\breve{r}$ to avoid overfitting.

A.7.2. Blends as Special-case Tensors

This section explains how our blends are special-case tensors, requiring both restriction and relaxation from the standard model of tensors as sums over rank-1 tensors, i.e., $T = \sum_{z=1}^{r} l_z \cdot T_{\boldsymbol{A}_z}$. Starting from this standard model, consider the following restrictions:

[21] For an accessible presentation of the Rotation Problem, see Section 2.2 of Rabanser et al. (2017) who further write, "matrix decompositions are only unique under very stringent conditions, such as orthogonality constraints which are imposed by the singular value decomposition (SVD)." Rank-1 tensors also have unique decomposition.

(1) Require symmetry: the tensor has constant size m in every dimension and further, component rank-1 tensors have the form $T_{\boldsymbol{A}_z} = \boldsymbol{a}_z \otimes \ldots \otimes \boldsymbol{a}_z$, i.e., they are each the outer product of n copies of a common element-vector $\boldsymbol{a}_z$.

(2) Require the tensor to have overall structure to describe a blend, i.e., to describe a distribution-over-distributions; thus:

- Require the vectors $\boldsymbol{a}_z$ to be probability distributions (for each $\boldsymbol{a}_z$, we have $a_{z,i} \geq 0$ and $\sum_i a_{z,i} = 1$).
- The set of scalar factors l_z must be a probability distribution ($l_z > 0$ and $\sum_{z=1}^{r} l_z = 1$).

Thus, the tensor acts as a (symmetric, correlated) probability distribution over its indexes, with component rank-1 tensors as *latent variables* according to the l_z scalars as a disrcrete probability distribution $\boldsymbol{l}$. A random draw from the tensor-correlated distribution can be effected by drawing a random rank-1 tensor $\hat{\boldsymbol{A}}_z \sim \boldsymbol{l}$ and then drawing each coordinate i.i.d. according to the probabilities of the corresponding $\hat{\boldsymbol{a}}_z$ (cf. the procedure for drawing n inputs in the prior independent information setting).

(3) For use within the Blends Technique (as applied to the prior independent design setting), for the δ_2 side as chosen by the adversary, we require restricting the (probability) vectors $\boldsymbol{a}_z$ to a given class of distributions $\mathcal{F}$. (The δ_1 side may use arbitrary distributions from $\mathcal{F}^{\text{all}}$.)

Already, this restricted case of tensors corresponds to a special case application of the Blends Technique, in which the underlying algorithm setting is defined to only have input from a finite, discrete set – e.g., an auction setting in which values have support as a finite

set of integers (and there would be a fixed (abstract) mapping between discrete inputs and tensor indexing).

Definition 39. *A* latent-distribution, symmetric probability tensor *is a convex combination over rank-*1 *tensors with each component constructed as the outer-product of successive copies of a single* distribution-vector *(defined by non-negative entries that sum to* 1*).*

From here, consider the following relaxations to extend our *latent-distribution, symmetric probability tensor* special-case:

(1) Relax the standard indexing of tensor entries to a general setting in which they may be continuous and unbounded (though we keep an assumption of non-negativity). I.e., considering a tensor as a function defined for discrete (highly-structured) domain, this point relaxes to arbitrary domain.

(2) Relax from a sum over rank-1 tensors to allow continuous integration over rank-1-tensors. Correspondingly, we relax the discreteness of the scalars to allow them to be continuous quantities (measured dz). In this case we have $T = \int_{\boldsymbol{z}} T_{\boldsymbol{A}_z} \cdot l(z)\ dz$ and $\int_{\boldsymbol{z}} l(z)\ dz = 1$ for $l(z) > 0$ for all z.[22]

We state here without proof that the tensor setting which aggregates all restrictions and relaxations just described is generally equivalent to Definition 18 for our blends, specifically for use within the Blends Technique.

[22] Recall, avoiding overfitting is one of the justifications to prefer tensors with small rank – at some point, simplicity is better than spurious accuracy. One justification of our relaxation to allow infinite rank (with cardinality equal to the reals) is that short *functional* descriptions are themselves "simple" descriptions – as desired. Informally, this idea suggests replacing upper bounds on rank with upper bounds on *the structure of bases with (potentially) full rank*; e.g., our finite-weight Quadratics-versus-Uniforms dual blend (page 99) was sufficiently presented by just two points for each side of the dual blend, with each point a description of weight-at-z/distribution-per-z/over-z-range.

APPENDIX B

Contextual and Technical Notes

B.1. Distribution Naming Conventions, Including Exogenous Restrictions

We include here a technical note to describe naming schemes for distributions. First, we give two approaches to naming to reflect two distinct approaches of analyses in this thesis – one line from standard probabilistic approaches (which start by defining CDF and PDF, etc.) and another line in which distributions are motivated by geometric properties of their revenue curves. Second, we describe notation to represent modification of a base distribution via standard operations (truncation and conditioning).

Within the setup of Theorem 5 in Section 3.2, we have a restriction to scale-invariant mechanisms for which it was sufficient to consider only distributions that have been: "standardized so that the single-agent optimal revenue defaults to $\max_q R_F(q) = 1$." Subsequently, Triangle and Quadrilateral *revenue-curve* distributions are described (Definition 16 and Definition 17) with their names chosen specifically to reflect the geometry of their revenue curves. In a similar line, Constant Negative Virtual Value (Definition 15) are named from revenue curves.

All other distributions – largely employed to describe dual blends to inform prior independent lower bounds per the Blends Technique – have motivation more generically from distribution properties and will be "named" functions written in un-italicized lettering, using the following scheme. These distribution names will:

- use letters corresponding to the beginning letters of their standard names in the math community (or natural attempts to copy such); and end with the last letter 'd' for "distribution;"
- use the first letter capitalized to reference the distribution itself as an object and to represent its CDF; and use all letters lowercase to reference the PDF;
- use a lower-case-i prefix to identify an inverse-distribution (per Definition 23);
- e.g., these should clearly distinguish the exponential function $\exp(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{x!}$; versus an exponential distribution 'Exd_1' with PDF $\text{exd}_1(x) = e^{-x}$.
- an exception to this scheme is *local definition and usage of a distribution* $\boldsymbol{\xi}$.

We give further notation to represent operations to modify a given distribution F to a related form. For absolute clarity, we first explicitly explain these standard operations. *Truncation* cuts off a distribution (either at the top or at the bottom) and re-allocates the deleted probability measure of the discarded support to a point mass at the truncation point. *Conditioning* cuts off a distribution and re-normalizes the densities in the remaining domain by dividing by its remaining total probability measure.

Given a distribution F, we introduce the following formal notation. Everything that follows applies to a distribution name F, its CDF F, and its PDF f. Denote a bottom-conditioning and re-normalization of F at input a by $\overrightarrow{F}^a$, top-conditioning and re-normalization at input b by $\overleftarrow{F}^b$, and both operations at a and $b > a$ respectively by $\overleftrightarrow{F}^{a,b}$.[1] If the distribution F instead becomes truncated on one side with a point mass (rather than being conditioned and re-normalized), we accent the endpoint to indicate

[1] The arrows, where present, indicate the deleted density's direction of movement on the real line. This includes the use of 'left-right-arrow' to indicate a both-top-and-bottom domain restriction which "smushes" the density towards the middle of the original domain.

the point mass, e.g., $\overleftrightarrow{F}^{a,b'}$ represents conditioning above a and top-truncation at b. If the original distribution is described by one or more parameters, e.g. F_z, these naturally persist as subscripts, for example $\overleftrightarrow{F}_z^{a,b'}$.

With respect to modification of distributions with truncation or conditioning (as defined in Appendix B.1), we have the following lemma to describe when distribution properties are necessarily preserved.

Lemma 20. *Given a distribution F with the MHR property and/or the regularity property in a revenue auction setting, its properties are preserved under modification to $\overrightarrow{F}^a$, $\overleftarrow{F}^{b'}$, and $\overleftrightarrow{F}^{a,b'}$.*

Proof. The statement for $\overrightarrow{F}^a$ follows directly from Fact 2. The statement for $\overleftarrow{F}^{b'}$ holds because calculations of hazard rate and virtual value (for revenue) for inputs less than b are unaffected by top-truncation to a point mass at input b, and at input b the hazard rate becomes ∞ and virtual value becomes b which are both automatically sufficient to preserve the respective original properties. The statement for $\overleftrightarrow{F}^{a,b'}$ holds by sequential application of the first two cases. □

Whether or not properties are preserved under top-conditioning with re-normalization of the density (rather than moving to point mass as in Lemma 20) is dependent on the distribution in question.

B.2. Ordinal and Inequality Properties of Re-weighted Fractions

(from page 84) Lemma 21 supports the proof of Lemma 10. It states that for a ROE objective as used in Lemma 10, a point mass on an element of the mixture must achieve

at least the value of the overall ratio. This statement is similar to a standard statement from the probabilistic method. Lemma 21 is proved using this Fact 26.

Fact 26. *Given a distribution F with PDF $f : \Omega \to \mathbb{R}_+$ and with expected value μ_F. There exists an element $\omega_+ \in \Omega$ for which $\omega_+ \geq \mu_F$.*

Lemma 21. *Consider a domain Ω and two positive functions $a : \Omega \to \mathbb{R}_+$ and $b : \Omega \to \mathbb{R}_+$. For every distribution γ over the elements of Ω, there exists ω_+ in the support of γ for which*

$$\frac{a(\omega_+)}{b(\omega_+)} \geq \frac{\mathbf{E}_{\omega\sim\gamma}\left[a(\omega)\right]}{\mathbf{E}_{\omega\sim\gamma}\left[b(\omega)\right]} \tag{B.1}$$

Proof. Set $\alpha = \mathbf{E}_{\omega\sim\gamma}\left[a(\omega)\right]$ and $\beta = \mathbf{E}_{\omega\sim\gamma}\left[b(\omega)\right]$. The first line uses these definitions and the second line is a simple re-organization:

$$\frac{\alpha}{\beta} = \frac{\mathbf{E}_{\omega\sim\gamma}\left[a(\omega)\right]}{\mathbf{E}_{\omega\sim\gamma}\left[b(\omega)\right]}$$

$$0 = \mathbf{E}_{\omega\sim\gamma}\left[\beta \cdot a(\omega) - \alpha \cdot b(\omega)\right]$$

Applying the probabilistic method (explained immediately before this lemma) to the last line, there must exist ω_+ for which $\beta \cdot a(\omega_+) - \alpha \cdot b(\omega_+) \geq 0$ which is equivalent to ${}^{a(\omega_+)}/_{b(\omega_+)} \geq {}^{\alpha}/_{\beta}$. □

(from page 163) The following fact is an application of Lemma 21 above which is appropriate when the supremum over ratios can in fact be achieved as the maximum by an element of the space Ω, which is true when Ω is a probability simplex over finite elements. Specifically, it states that when a fraction has numerator and denominator

composed of the same convex combination over positive ordered pairs (over the elements of each pair respectively), the mininizing (alternatively, maximizing) convex combination is a point mass on a single pair-element, and therefore, there exists an element-pair for which the fractional ratio is at least as large (respectively, as small) as the weighted fraction resulting from any convex combination.

Fact 27. *Given a set of n pairs $\{(a_i, b_i)\}_i$, with $0 < a_i, b_i \in \mathbb{R}\ \forall\ i$, applying any common probability distribution vector $\boldsymbol{\xi}$ (i.e., over a discrete, finite list of elements) and taking expectation separately (over the indexes of vectors $\mathbf{a}, \mathbf{b}$) is an operation minimized (or maximized) by selecting any index i which individually minimizes (respectively maximizes) the ratio:*

$$\min_{\boldsymbol{\xi} \in \Delta(\{1,\ldots,n\})} \left(\frac{[\boldsymbol{\xi}]^\top \cdot [\mathbf{a}]}{[\boldsymbol{\xi}]^\top \cdot [\mathbf{b}]} \right) = \min_i \left(\frac{a_i}{b_i} \right) \tag{B.2}$$

(from page 74) The following lemma supports proof of the Truncation Lemma 6 in Section 3.4.

Lemma 22. *Given a fractional quantity*

$$Q = \frac{\kappa \cdot A' + B'}{\kappa \cdot A'' + B''}$$

for $\kappa, A', A'', B', B'' > 0$. Let $A = A'/A'' > 0$ and $B = B'/B'' > 0$, then the following hold:

(1) if $A \geq B$ (respectively equal to), then $A \geq Q$ (respectively $A = Q$);

(2) if $A \geq B$ (respectively equal to, at most), then Q is increasing in κ (respectively constant, decreasing).

Proof. Both statements are true by simple algebraic manipulation. Note we rearrange the assumption $A \geq B$ by $B' \leq (A'/A'') \cdot B''$. For (1):

$$Q = \frac{\kappa \cdot A' + B'}{\kappa \cdot A'' + B''} \leq \frac{\kappa \cdot A' + \frac{A'}{A''}B''}{\kappa \cdot A'' + B''}$$
$$= \frac{\kappa \cdot A'A'' + A'B''}{A''(\kappa \cdot A'' + B'')} = \frac{A'(\kappa \cdot A'' + B'')}{A''(\kappa \cdot A'' + B'')} = \frac{A'}{A''} = A$$

For (2), first we take the derivative of a function $Q(\kappa)$ with respect to κ and work from there, with the last inequality equivalent to the assumption $A \geq B$:

$$(\kappa \cdot A'' + B'')^2 \cdot \frac{\partial Q}{\partial \kappa} = (\kappa \cdot A'' + B'') \cdot A' - (\kappa \cdot A' + B') \cdot A''$$
$$= B'' \cdot A' - B' \cdot A'' \geq 0$$ □

B.3. Confirmation of Total Weights for Quadratics-versus-Uniforms

(from page 101) The goal of this section is to prove cross-checks of finite-weight blends math presented in Section 4.5. The total weight from the Quadratics side (o_F) was easily shown to be $1+\int_1^H \frac{2}{z}dz = 1+2\ln H$ from the defined weights o_{pm} and o_{Qz}. We confirm this from the description of the correlated distribution (which matches from both Quadratics and Uniforms):

$$\int_1^h \frac{1}{z^2} \cdot (z-1) \cdot 2dz + 2 \cdot (h-1) \cdot \frac{1}{h} + 1 = \int_1^h \frac{2}{z} - \frac{2}{z^2}dz + \frac{3h-2}{h} = \int_1^h \frac{2}{z}dz + 1$$

where the additive terms are respectively the correlated distribution's totals of 2D probability mass, total 1D, and total 0D. To confirm this from the ω_F weights on Uniforms,

we have

$$\left[1+\int_1^h \frac{2}{z}dz\right]-\int_1^h \frac{2(z-1)^2}{z^3}dz = 1+\int_1^h \frac{4}{z^2}-\frac{2}{z^3}dz = 4-\frac{4}{h}+\frac{1}{h^2}=\frac{(2h-1)^2}{h^2}$$

so on the left, we start with the total weight (in the brackets) minus the integral $\int_1^h \omega_{Uz}$; and this is equal to ω_{pm} on the right, which is confirmed as the point mass on $\boldsymbol{\xi}$.

B.4. Quadratics-versus-Uniforms from Order-statistic Independence

(from page 125, page 128) We show how our main example of Quadratics-versus-Uniforms fits into Theorem 13. We use the distribution-version of the theorem which includes its Condition (4). Motivated by the Quadratics, let $g_1(x) = 1/x^2$ inducing $G_1(x) = 1/x$. Motivated by the Uniforms, let $g_2(x) = 1$ inducing $G_2(x) = x$. Recall we assume $v_1 \geq v_2 > 0$. Therefore on the Uniforms side we have the following. Note that in fact, these calculations apply for arbitrary upward-finite g_2 because we can wait until the end to substitute.

$$\chi(z) = \frac{g_1(z)}{g_2(z)} = \frac{1}{z^2 \cdot g_2(z)}, \qquad d\chi(z) = d\left(\frac{1}{z^2 \cdot g_2(z)}\right) = (-1)\cdot \frac{2g_2(z)+z\cdot g_2'(z)}{z^3\cdot (g_2(z))^2}\cdot dz$$

(where the evaluation of $d\chi(z)$ doesn't matter but we write it for completeness). We further have

$$\omega_{g_2}(z) = (-1)\cdot d\left(\frac{1}{z^2\cdot g_2(z)}\right)\cdot (G_2(z))^2 = \left(\frac{2g_2(z)+z\cdot g_2'(z)}{z^3\cdot (g_2(z))^2}\right)\cdot (G_2(z))^2\cdot dz \geq 0$$

with the final inequality included to illustrate that it is non-negative.[2] From here we have

$$\begin{aligned}
&\int_{v_1}^{\infty} \omega_{g_2} \cdot \left(\frac{g_2(v_1)}{G_2(z)}\right)\left(\frac{g_2(v_2)}{G_2(z)}\right) \\
&= \int_{v_1}^{\infty} (-1) \cdot d\left(\frac{1}{z^2 \cdot g_2(z)}\right) \cdot (G_2(z))^2 \cdot \left(\frac{g_2(v_1)}{G_2(z)}\right)\left(\frac{g_2(v_2)}{G_2(z)}\right) \\
&= (g_2(v_1) \cdot g_2(v_2)) \cdot \int_{v_1}^{\infty} (-1) \cdot d\left(\frac{1}{z^2 \cdot g_2(z)}\right) = (g_2(v_1) \cdot g_2(v_2)) \left[\frac{1}{z^2 \cdot g_2(z)}\right]_{\infty}^{v_1} = \frac{g_2(v_2)}{v_1^2}
\end{aligned}$$

as desired, because $g_2(x) = 1$ and $g(\boldsymbol{v}) = {}^1\!/_{v_1^2}$ is correct for infinite-weight Quadratics-versus-Uniforms dual blends of Section 4.5.1. On the Quadratics side, symmetric to the analysis above, we have

$$\psi(z) = \frac{g_2(z)}{g_1(z)} = z^2, \qquad d\psi(z) = d\left(z^2\right) = 2z \cdot dz$$

$$\omega_{g_1}(z) = d\left(z^2\right) \cdot (G_1(z))^2 = 2z \cdot \left(\frac{1}{z}\right)^2 \cdot dz = \frac{2}{z} \cdot dz \geq 0$$

Finally we have

$$\begin{aligned}
&\int_0^{v_2} o_{g_1} \cdot \left(\frac{g_1(v_1)}{G_1(z)}\right)\left(\frac{g_1(v_2)}{G_1(z)}\right) = \int_0^{v_2} d\left(z^2\right) \cdot (G_1(z))^2 \cdot \left(\frac{g_1(v_1)}{G_1(z)}\right)\left(\frac{g_1(v_2)}{G_1(z)}\right) \\
&= (g_1(v_1) \cdot g_2(v_2)) \cdot \int_0^{v_2} d(z^2) = (g_1(v_1) \cdot g_2(v_2)) \left[\ z^2\ \right]_0^{v_2} = \frac{1}{v_1^2} \cdot \frac{1}{v_2^2} \cdot v_2^2 = \frac{1}{v_1^2}
\end{aligned}$$

B.5. Blends from Order-statistic Separability that are not Distributions

(from page 125) We give a simple second example which illustrates Theorem 13. (The first example is immediately above in Appendix B.3.)

[2] Here we can also already confirm that $\omega_{g_2} = \left(\frac{2g_2(z)+z\cdot g_2'(z)}{z^3\cdot(g_2(z))^2}\right) \cdot (G_2(z))^2 \cdot dz = \left(\frac{2\cdot 1+z\cdot 0}{z^3\cdot(1)^2}\right) \cdot (z)^2 \cdot dz = {}^2\!/_{z\,\cdot\, dz}$ as we expect, given the example.

The dual blends have one side as Quadratics and the other side as *Cubics*. In this case, the Quadratics have downward-closed domain and can not be normalized to distributions because the function $G_2(x) = \int_0^z 1/y^2 \, dy = \infty$. Without further comment, we write down the evaluations of all necessary elements using the definitions of Theorem 13:

$$g_1(x) = \frac{1}{x^3} \text{ for } x \in (0, \infty) \qquad g_2(x) = \frac{1}{x^2} \text{ for } x \in (0, \infty)$$

$$g_{1,z}(x) = \frac{1}{x^3} \text{ for } x \in [z, \infty) \qquad g_{2,z}(x) = \frac{1}{x^2} \text{ for } x \in (0, z]$$

$$\psi(z) = z \text{ for } z \in (0, \infty) \qquad \chi(z) = \frac{1}{z} \text{ for } z \in (0, \infty)$$

$$o_{g_1}(z) = 1 \cdot dz \text{ for } z \in (0, \infty) \qquad \omega_{g_2}(z) = \frac{1}{z^2} \cdot dz \text{ for } z \in (0, \infty)$$

$$\int_0^\infty o_{g_1}(z) \cdot g_{1,z}(v_1) \cdot g_{1,z}(v_2) = \int_0^{v_2} o_{g_1}(z) \cdot g_1(v_1) \cdot g_1(v_2) = \int_0^{v_2} 1 \cdot \frac{1}{v_1^3} \cdot \frac{1}{v_2^3} \, dz$$

$$= \frac{1}{v_1^3} \cdot \frac{1}{v_2^2} = g_1(v_1) \cdot g_2(v_2) = g(\boldsymbol{v})$$

$$= \int_0^\infty \omega_{g_2}(z) \cdot g_{2,z}(v_1) \cdot g_{2,z}(v_2) = \int_{v_1}^\infty \omega_{g_2}(z) \cdot g_2(v_1) \cdot g_2(v_2) = \int_{v_1}^\infty \frac{1}{z^2} \cdot \frac{1}{v_1^2} \cdot \frac{1}{v_2^2} \, dz$$

Note – this solution concept would fail if we assigned the Quadratics to be upward-closed and the Cubics to be downward-closed because the monotonicity conditions of Theorem 13 would be violated.

B.6. Supporting Material for Exponentials-versus-Inverse-Exponentials

(from page 217) This section provides material to supplement Appendix A.3; the presentation generally assumes its terms, assumptions, and context while only restating

the most important definitions here. There are two subsections. Appendix B.6.2 shows that dual blends weights are equal in total and finite. Appendix B.6.3 shows that the set of distributions composing the Exponentials side of the finite-weight dual blend are MHR.

B.6.1. Top-truncated Exponentials Blend: Confirmation of Positive 2-D Mass

(from page 221) The goal of this section is to show that the standard 2-D blends mass $dv_1 dv_2$ which is common to both equation (A.16) and equation (A.23) is positive. Recalling $v_1 \geq v_2 > 0$ and $\bar{q} \in (0, 1)$, starting from equation (A.16) without loss of generality, we have

$$\left[\frac{\ln \bar{q}}{v_1} \cdot \frac{1}{v_1 + v_2} \cdot \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}}\right] + \frac{1}{(v_1 + v_2)^2} \left(1 - \bar{q} \cdot e^{\frac{v_2}{v_1} \ln \bar{q}}\right) > 0$$

$$\Updownarrow$$

$$v_1 + \bar{q}^{\frac{v_2}{v_1}} \cdot (\bar{q} \ln \bar{q} \cdot (v_1 + v_2 - \bar{q} v_1) > 0$$

$$\Uparrow$$

$$v_1 + \bar{q} \cdot (\bar{q} \ln \bar{q} \cdot (v_1 + v_1) - \bar{q} v_1) > 0$$

$$\Updownarrow$$

$$1 + 2\bar{q}^2 \ln \bar{q} - \bar{q}^2 > 0$$

Regarding the last line, the derivative of the left-hand side is $4\bar{q} \ln \bar{q}$ which is negative for all $\bar{q} \in (0, 1)$. Therefore its worst-case $\bar{q}$ is evaluated in the limit $\bar{q} \to 1$. By observation, we confirm that the left-hand side is 0 in this limit $\bar{q} \to 1$ and positive for $\bar{q} < 1$. Note,

it is not surprising that this blends mass goes to 0 as $\bar{q} \to 1$, because $\bar{q}$ represents the "point mass at z" within every distribution of the blend: if $\bar{q} = 1$, then 2-D mass $dv_1 dv_2$ does not exist.

B.6.2. Total Dual Blends Weight is Equal and Finite

(from page 232) We start by writing out the non-closed-form total weights from each side of the finite-weight Exponentials-versus-Inverse-Exponentials dual blend of Appendix A.3. Then we first we run a sanity cross-check that total weight is equal, in light of the technical complexity of the analysis. The total weight of the Exponentials blend is

$$\begin{aligned}\int_{1/h}^{h} \omega_z + \int_{h}^{\infty} \omega_z &= \int_{1/h}^{h} \frac{1}{z} \left(1 - \text{Exd}_\beta(1/h)\right)^2 \, dz + \int_{h}^{\infty} \frac{1}{z} \left(\text{Exd}_\beta(h) - \text{Exd}_\beta(1/h)\right)^2 \, dz \\ &= \int_{1/h}^{h} \frac{e^{\frac{2}{zh} \ln \bar{q}}}{z} \, dz + \int_{h}^{\infty} \frac{\left(e^{\frac{1}{zh} \ln \bar{q}} - e^{\frac{h}{z} \ln \bar{q}}\right)^2}{z} \, dz \end{aligned} \tag{B.3}$$

and the total weight of the Inverse-Exponentials side of the blend is

$$\begin{aligned}\int_{1/h}^{h} o_z + \int_{0}^{1/h} o_z &= \int_{1/h}^{h} \frac{1}{z} \left(\text{i-Exd}_\beta(h)\right)^2 \, dz + \int_{0}^{1/h} \frac{1}{z} \left(\text{i-Exd}_\beta(h) - \text{i-Exd}_\beta(1/h)\right)^2 \, dz \\ &= \int_{1/h}^{h} \frac{e^{\frac{2z}{h} \ln \bar{q}}}{z} \, dz + \int_{0}^{1/h} \frac{\left(e^{\frac{z}{h} \ln \bar{q}} - e^{hz \ln \bar{q}}\right)^2}{z} \, dz \end{aligned} \tag{B.4}$$

We show the two sides have equal weight through calculus-change-of-variables with each of the additive terms matching up individually. Using $y = \zeta(z) = 1/z$ and $-dy = 1/z^2 \cdot dz$

for both calculus substitutions, we get:

$$\begin{aligned}\int_{1/h}^{h} \frac{e^{\frac{2}{zh}\ln\bar{q}}}{z}\,dz &= \int_{1/h}^{h} z\cdot\left(e^{\frac{2}{zh}\ln\bar{q}}\right)\cdot\left(\frac{dz}{z^2}\right)\\ &= \int_{h}^{1/h} \frac{1}{y}\cdot\left(e^{\frac{2y}{h}\ln\bar{q}}\right)\cdot(-dy)\\ &= \int_{1/h}^{h} \frac{e^{\frac{2y}{h}\ln\bar{q}}}{y}\,dy\end{aligned}$$

and

$$\begin{aligned}\int_{h}^{\infty} \frac{\left(e^{\frac{1}{hz}\ln\bar{q}}-e^{\frac{h}{z}\ln\bar{q}}\right)^2}{z}\,dz &= \int_{h}^{\infty} z\cdot\left(e^{\frac{1}{hz}\ln\bar{q}}-e^{\frac{h}{z}\ln\bar{q}}\right)^2\cdot\left(\frac{dz}{z^2}\right)\\ &= \int_{1/h}^{0} \frac{1}{y}\cdot\left(e^{\frac{y}{h}\ln\bar{q}}-e^{hy\ln\bar{q}}\right)^2\cdot(-dy)\\ &= \int_{0}^{1/h} \frac{\left(e^{\frac{y}{h}\ln\bar{q}}-e^{hy\ln\bar{q}}\right)^2}{y}\,dy\end{aligned}$$

Therefore we confirm that total weight is equal. Now we show that total weight is finite. We work from line (B.4) to show that it is finite for any $h \in (1,\infty)$ and $\bar{q} \in (0,1)$. First we note that within the first additive integral term, the function $e^{\frac{2z}{h}\ln\bar{q}}/z$ is both strictly positive and bounded above on the finite domain $[1/h, h]$, so this term is finite. For the second additive integral term in line (B.4), it is not obvious that the integrand is bounded. In order to show it is bounded, we start by simplifying with

$$\int_{0}^{1/h} \frac{\left(e^{\frac{z}{h}\ln\bar{q}}-e^{hz\ln\bar{q}}\right)^2}{z}\,dz < \int_{0}^{1/h} \frac{e^{\frac{z}{h}\ln\bar{q}}-e^{hz\ln\bar{q}}}{z}\,dz \tag{B.5}$$

which is true because $\left(e^{\frac{z}{h}\ln\bar{q}} - e^{hz\ln\bar{q}}\right) < 1$ is the (positive) difference between the cdf function for inverse-exponentials $e^{\frac{z}{x}\ln q}$ evaluated at H and $1/H$ (with $\ln q$ negative; and also the integrand is now confirmed to be positive everywhere).

It is sufficient to show the right-hand side of (B.5) is finite. We would like to additively separate the integral into two integrals by "splitting" it with respect to the "minus sign." In order to do this, for technical reasons, we need to explicitly evaluate the lower endpoint of the integral in the limit going to 0. If we do not undertake this technical adaption up front, then subsequently otherwise, splitting the integral into its additive terms results in both being divergent, and it is impossible to put them back together.[3]

$$\begin{aligned}\int_0^{1/h} \frac{e^{\frac{z}{h}\ln\bar{q}} - e^{hz\ln\bar{q}}}{z}\, dz &= \lim_{\delta\to 0}\int_\delta^{1/h} \frac{e^{\frac{z}{h}\ln\bar{q}} - e^{hz\ln\bar{q}}}{z}\, dz \\ &= \lim_{\delta\to 0}\left[\int_\delta^{1/h} \frac{e^{\frac{z}{h}\ln\bar{q}}}{z}\, dz - \int_\delta^{1/h} \frac{e^{hz\ln\bar{q}}}{z}\, dz\right] \end{aligned} \tag{B.6}$$

Now we use calculus-change-of-variables on each integral within the brackets. We set $y_1 = \zeta_1(z) = {}^{-z\ln\bar{q}}/_{h}$, $dy_1 = \frac{-\ln\bar{q}}{h}\cdot dz$; and $y_2 = \zeta_2(z) = -hz\ln\bar{q}$, $dy_2 = -h\ln\bar{q}\cdot dz$. (Note we use y_1 and y_2 for the respective substitutions, after which we change both

[3] This specifically plays out when – after doing calculus-change-of-variables separately for each additive integral term – the evaluations at the lower endpoint of each resulting integral will "cancel" when they should not, leaving a negative quantity.

variables of integration to a common y.) Continuing we get

$$\begin{aligned}
\text{line (B.6)} &= \lim_{\delta\to 0}\left[\int_{\delta}^{1/h}\frac{e^{\frac{z}{h}\ln\bar{q}}}{z}\,dz - \int_{\delta}^{1/h}\frac{e^{hz\ln\bar{q}}}{z}\,dz\right] \\
&= \lim_{\delta\to 0}\left[\int_{\frac{-\delta\ln\bar{q}}{h}}^{\frac{-\ln\bar{q}}{h^2}}\frac{e^{-y_1}}{y_1}\,dy_1 - \int_{-h\delta\ln\bar{q}}^{-\ln\bar{q}}\frac{e^{-y_2}}{y_2}\,dy_2\right] \\
&= \lim_{\delta\to 0}\left[\int_{\frac{-\delta\ln\bar{q}}{h}}^{\infty}\frac{e^{-y}}{y}\,dy - \int_{\frac{-\ln\bar{q}}{h^2}}^{\infty}\frac{e^{-y}}{y}\,dy - \int_{-h\delta\ln\bar{q}}^{\infty}\frac{e^{-y}}{y}\,dy + \int_{-\ln\bar{q}}^{\infty}\frac{e^{-y}}{y}\,dy\right] \\
&= \lim_{\delta\to 0}\left[-\mathrm{Ei}\left(\frac{\delta\ln\bar{q}}{h}\right) + \mathrm{Ei}\left(\frac{\ln\bar{q}}{h^2}\right) + \mathrm{Ei}\left(h\delta\ln\bar{q}\right) - \mathrm{Ei}\left(\ln\bar{q}\right)\right] \\
&= \lim_{\delta\to 0}\left[\mathrm{E}_1\left(-\frac{\delta\ln\bar{q}}{h}\right) - \mathrm{E}_1\left(-h\delta\ln\bar{q}\right)\right] + \mathrm{Ei}\left(\frac{\ln\bar{q}}{h^2}\right) - \mathrm{Ei}\left(\ln\bar{q}\right) \qquad \text{(B.7)}
\end{aligned}$$

where the "Ei" function is the (irreducible) "exponential-integral" function defined for $x \neq 0$ by $\mathrm{Ei}(x) = -\int_{-x}^{\infty}\frac{e^{-y}}{y}dy$. Further for $x > 0$, $-\mathrm{Ei}(-x) = \mathrm{E}_1(x)$, with "$\mathrm{E}_1$" the (irreducible) exponential-integral function defined by $\mathrm{E}_1(x) = \int_{x}^{\infty}\frac{e^{-y}}{y}dy$ for the domain of positive reals and converging to finite real output for this domain.[4]

Although the exponential-integral $\mathrm{E}_1(x) = \int_{x}^{\infty}\frac{e^{-y}}{y}dy$ itself is irreducible, we can analyze the quantity using its Taylor series. Starting with e^{-y}, we get:

$$\begin{aligned}
e^{-y} &= \sum_{k=0}^{\infty}\frac{(-y)^k}{k!} = 1 - y + \frac{y^2}{2!} - \frac{y^3}{3!} + \frac{y^4}{4!} - \dots \\
\frac{e^{-y}}{y} &= \frac{1}{y} - 1 + \frac{y}{2!} - \frac{y^2}{3!} + \frac{y^3}{4!} - \dots \\
\int\frac{e^{-y}}{y}\,dy &= C + \ln y - y + \frac{y^2}{2\cdot 2!} - \frac{y^3}{3\cdot 3!} + \frac{y^4}{4\cdot 4!} - \dots
\end{aligned}$$

[4] The function E_1 as a function of z is actually defined for the complex numbers (other than the non-positive reals) and requires $|\mathrm{Arg}(z)| < \pi$ but we simplify the identity here to the reals. For being irreducible, this function is generally well-understood.

$$\int_x^\infty \frac{e^{-y}}{y}\, dy = -\gamma_{\text{EM}} - \ln x + x - \frac{x^2}{2 \cdot 2!} + \frac{x^3}{3 \cdot 3!} - \frac{x^4}{4 \cdot 4!} - \dots \tag{B.8}$$

$$\text{E}_1(x) = \int_x^\infty \frac{e^{-y}}{y}\, dy = -\gamma_{\text{EM}} - \ln x - \sum_{k=1}^{\infty} \frac{(-x)^k}{k \cdot k!} \tag{B.9}$$

where it is known that the integral's constant term C is the Euler-Mascheroni constant $\gamma_{\text{EM}} \approx 0.57721$ when the integral endpoints are x and ∞ (Abramowitz and Stegun, 1964).[5] We substitute the equation of line (B.9) twice into line (B.7) to get

$$\begin{aligned}
&\lim_{\delta \to 0} \left[\text{E}_1 \left(-\frac{\delta \ln \bar{q}}{h} \right) - \text{E}_1 \left(-h\delta \ln \bar{q} \right) \right] + \text{Ei}\left(\frac{\ln \bar{q}}{h^2} \right) - \text{Ei}\left(\ln \bar{q} \right) \\
&= \lim_{\delta \to 0} \left[\left(-\gamma_{\text{EM}} - \ln \left(\frac{-\delta \ln \bar{q}}{h} \right) - \sum_{k=1}^{\infty} \frac{\left(-\left(\frac{-\delta \ln \bar{q}}{h} \right) \right)^k}{k \cdot k!} \right) \right. \\
&\quad \left. - \left(-\gamma_{\text{EM}} - \ln \left(-h\delta \ln \bar{q} \right) - \sum_{k=1}^{\infty} \frac{\left(-\left(-h\delta \ln \bar{q} \right) \right)^k}{k \cdot k!} \right) \right] + \text{Ei}\left(\frac{\ln \bar{q}}{h^2} \right) - \text{Ei}\left(\ln \bar{q} \right) \\
&= \lim_{\delta \to 0} \left[\ln \left(\frac{-h\delta \ln \bar{q}}{\frac{-\delta \ln \bar{q}}{h}} \right) - \left(\sum_{k=1}^{\infty} \frac{\left(\frac{\delta \ln \bar{q}}{h} \right)^k}{k \cdot k!} \right) + \left(\sum_{k=1}^{\infty} \frac{\left(h\delta \ln \bar{q} \right)^k}{k \cdot k!} \right) \right] + \text{Ei}\left(\frac{\ln \bar{q}}{h^2} \right) - \text{Ei}\left(\ln \bar{q} \right)
\end{aligned} \tag{B.10}$$

$$= \ln(h^2) + \text{Ei}\left(\frac{\ln \bar{q}}{h^2} \right) - \text{Ei}\left(\ln \bar{q} \right)$$

where the last step cancels within the ln-term such that δ drops completely, and then it can trivially evaluate the limit point of $\delta = 0$. For every fixed $h \in (1, \infty)$ and $\bar{q} \in (0, 1)$, the total in the last line (B.10) is indeed finite. However the sum of the second and third "Ei" terms is negative. Thus as a last step, we confirm that the total is positive as a

[5] Note where the signs flipped after specific endpoints of the integral are added, this is because x is the lower endpoint, e.g., $\int_x^\infty f(x)dx = -F(x) + F(\infty)$. Line (B.9) is the standard identity of the E_1 function (Abramowitz and Stegun, 1964).

sanity check.

$$
\begin{aligned}
&\text{line (B.10)} = \ln(h^2) + \mathrm{Ei}\left(\frac{\ln \bar{q}}{h^2}\right) - \mathrm{Ei}\left(\ln \bar{q}\right) \\
&= \ln(h^2) - \mathrm{E}_1\left(\frac{-\ln \bar{q}}{h^2}\right) + \mathrm{E}_1\left(-\ln \bar{q}\right) \\
&= \ln(h^2) - \left[-\gamma_{\mathrm{EM}} - \ln\left(\frac{-\ln \bar{q}}{h^2}\right) - \sum_{k=1}^{\infty} \frac{\left(\frac{\ln \bar{q}}{h^2}\right)^k}{k \cdot k!}\right] + \left[-\gamma_{\mathrm{EM}} - \ln\left(-\ln \bar{q}\right) - \sum_{k=1}^{\infty} \frac{\left(\ln \bar{q}\right)^k}{k \cdot k!}\right] \\
&= \sum_{k=1}^{\infty} \frac{\left(\frac{\ln \bar{q}}{h^2}\right)^k}{k \cdot k!} - \sum_{k=1}^{\infty} \frac{\left(\ln \bar{q}\right)^k}{k \cdot k!} \\
&= \sum_{k=1}^{\infty} \frac{(-1) \cdot \left(\ln \bar{q}\right)^k \cdot \left(1 - \frac{1}{h^2}\right)^k}{k \cdot k!} \\
&= \sum_{\text{odd } k \geq 1} \left[\left(\frac{\left(-\ln \bar{q}\right) \cdot \left(\ln \bar{q}\right)^{k-1} \left(1 - \frac{1}{h^2}\right)^k}{k \cdot k!}\right) - \left(\frac{\left(\ln \bar{q}\right)^2 \cdot \left(\ln \bar{q}\right)^{k-1} \left(1 - \frac{1}{h^2}\right)^{k+1}}{(k+1) \cdot (k+1)!}\right)\right] \\
&= \sum_{\text{odd } k \geq 1} \left[\left(\frac{\left(-\ln \bar{q}\right) \cdot \left(\ln \bar{q}\right)^{k-1} \left(1 - \frac{1}{h^2}\right)^k}{k \cdot k!}\right) \cdot \left(1 - \frac{\left(-\ln \bar{q}\right)\left(1 - \frac{1}{h^2}\right)}{\frac{(k+1)^2}{k}}\right)\right] > 0
\end{aligned}
$$

for which the terms of the last summation are definitively positive for every k by inspection (for every fixed $h \in (1, \infty)$ and $\bar{q} \in (0, 1)$).

B.6.3. All Modified-Exponentials with positive weight ω_z are MHR

(from page 233) The goal of this section is to prove Proposition 6, i.e., to show that the distributions that compose the finite-weight Exponentials blend of Appendix A.3 are MHR, taking into consideration truncation and conditioning. An exhaustive description of these distributions is copied here as follows. As previously stated: for every $z \in [1/h, h]$,

we include the bottom-conditioned, top-truncated Exponential distribution

$$\overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(x) = \frac{\frac{-\ln \bar{q}}{z} \cdot e^{\frac{x}{z}\ln \bar{q}}}{1 - \text{Exd}_{\beta}(1/h)} \text{ on } [1/h, z), \quad \overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}(z) = \frac{\bar{q}}{1 - \text{Exd}_{\beta}(1/h)} \text{ point mass}$$

and for every $z > h$ (with $\beta = {}^{-\ln \bar{q}}/_{z}$), we include the doubly-conditioned distribution

$$\overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}(x) = \frac{\frac{-\ln \bar{q}}{z} \cdot e^{\frac{x}{z}\ln \bar{q}}}{\text{Exd}_{\beta}(h) - \text{Exd}_{\beta}(1/h)} \text{ on } [1/h, h]$$

As our starting point, the base distributions Exd_{β} are all MHR. For the first of these cases – for every $\overleftrightarrow{\text{exd}}_{\beta}^{1/h,z'}$ – that these distributions are MHR follows directly from Lemma 20, restated here for convenience.

Lemma 20. *Given a distribution F with the MHR property and/or the regularity property in a revenue auction setting, its properties are preserved under modification to $\overrightarrow{F}^{a}$, $\overleftarrow{F}^{b'}$, and $\overleftrightarrow{F}^{a,b'}$.*

The second case – for every $\overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}$ – is more complicated because conditioning given an upper bound (in this case h) is not guaranteed to maintain the MHR property for arbitrary base distribution. That the distributions $\overleftrightarrow{\text{exd}}_{\beta}^{1/h,h}$ are MHR follows as a special case of Lemma 23 next.

Lemma 23. *Let F be an arbitrary probability distribution with the following properties: (1) F is MHR; and (2) F has non-increasing density, i.e., f as the PDF of F is non-increasing on its domain. Then for $a,\ b > a$ in the domain of F, the doubly-conditioned distribution $\overleftrightarrow{F}^{a,b}$ is also MHR.*

Proof. For the proof, it is sufficient to show that the distribution $\overleftarrow{F}^b$ remains MHR after top-conditioning. It is sufficient because if this sub-claim is true, then sequentially applying the bottom-conditioning (at a) rule of Lemma 20 to $\overleftarrow{F}^b$ completes the proof.

Given top-conditioning, we need to re-normalize the density by dividing by $F(b)$, which is the remaining measure of density. This same re-normalization factor applies to both the PDF f and CDF F of the original distribution. The hazard rate function of the newly top-conditioned distribution $\overleftarrow{F}^b$ is given by

$$\lambda^{\overleftarrow{F}^b}(x) = \frac{\frac{f(x)}{F(b)}}{1 - \frac{F(x)}{F(b)}} = \frac{f(x)}{F(b) - F(x)}$$

and the derivative is

$$\begin{aligned}
\frac{d\lambda^{\overleftarrow{F}^b}}{dx}(x) &= \frac{(F(b) - F(x))\,(f'(x)) - (f(x))\,(-f(x))}{(F(b) - F(x))^2} \\
&= \frac{((F(b) - 1) + (1 - F(x)))\,(f'(x)) - (f(x))\,(-f(x))}{(F(b) - F(x))^2} \\
&\geq \frac{(F(b) - 1)\,(f'(x))}{(F(b) - F(x))^2} \geq 0
\end{aligned}$$

where the the third line follows by assumption that f is MHR and the conclusion that the final quantity is positive follows by assumption that $f' \leq 0$. □

Proposition 6. *For every fixed $\bar{q} \in (0, 1)$ and $h > 1$, all distributions of the types $\overleftrightarrow{\text{exd}}_\beta^{1/h, z'}$ (of equation (A.25)) and $\overleftrightarrow{\text{exd}}_\beta^{1/h, h}$ (of equation (A.26)) are MHR. Therefore they are all elements of the class of distributions $\mathcal{F}^{\text{mhr}}$ and are accessible to a prior independent design problem's adversary when restricted to $\mathcal{F}^{\text{mhr}}$.*

Proof. As already discussed, the proof follows from Lemma 20 and Lemma 23. □

B.7. Continuous Density Calculations for Double-sided CNVV Blends

(from page 239 and page 241) The goal of this section is to provide the technical calculations to show that the dual-blends of Appendix A.4 match up in terms of 2-D density, specifically, the integral in line (A.42) on one side of the dual blend is equal to the integral in line (A.47) on the other side. Explicitly, the objective is to show

$$\text{(B.11)} \qquad \int_{v_1}^{\infty} \frac{1}{z} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_1 + \bar{q}z)^2} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_2 + \bar{q}z)^2}\, dz = \int_{0}^{v_2} \frac{1}{z} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\, z + \bar{p}v_1\right)^2} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\, z + \bar{p}v_2\right)^2}\, dz$$

Elegantly, if we were to replace the left-hand side with $r = 1-\bar{q}$ we would see exact symmetry of the functions to be integrated, but with different evaluation endpoints. However we deal with them as-is. First we evaluate each side of this equality to show that the resulting densities are the same, with explicit statement of each anti-derivative: G^{ω}_{2D} and G^{o}_{2D}. Afterwards, for technical completeness, we calculate the derivative of the anti-derivative G^{ω}_{2D} (equation (B.12)) as used on the ω_z-side, in order to confirm its correctness. Write the functional form of the anti-derivative of the integrand of the left-hand side as

$$\text{(B.12)} \qquad G^{\omega}_{2D}(z \mid \boldsymbol{v}) = \frac{1}{(v_1 - v_2)^3} \cdot \left[\frac{(1-\bar{q})v_1(v_1 - v_2)}{((1-\bar{q})v_1 + \bar{q}z)} + \frac{(1-\bar{q})v_2(v_1 - v_2)}{((1-\bar{q})v_2 + \bar{q}z)} + (v_1 + v_2) \cdot \ln\left(\frac{(1-\bar{q})v_2 + \bar{q}z}{(1-\bar{q})v_1 + \bar{q}z} \right) \right]$$

To simplify "in advance," we multiply equation (B.11) through by constant $(v_1 - v_2)^3$. Then calculations to evaluate the left-hand side integral are given by

$$
\begin{aligned}
&(v_1 - v_2)^3 \cdot \int_{v_1}^{\infty} \frac{1}{z} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_1 + \bar{q}z)^2} \cdot \frac{\bar{q}(1-\bar{q})z}{((1-\bar{q})v_2 + \bar{q}z)^2}\, dz \\
&= (v_1 - v_2)^3 \cdot \left(\lim_{\delta \to \infty} G_{2D}^{\omega}(\delta) - G_{2D}^{\omega}(v_1) \right) \\
&= 0 - \left((1-\bar{q})(v_1 - v_2) + \frac{(1-\bar{q})v_2(v_1 - v_2)}{((1-\bar{q})v_2 + \bar{q}v_1)} + (v_1 + v_2) \ln \left(\frac{(1-\bar{q})v_2 + \bar{q}v_1}{v_1} \right) \right) \qquad \text{(B.13)}
\end{aligned}
$$

Write the functional form of the anti-derivative of the integrand of the right-hand side as

$$
\begin{aligned}
G_{2D}^{o}(z \mid \boldsymbol{v}) = \frac{1}{(v_1 - v_2)^3} \cdot \Bigg[& \frac{\bar{p}v_1(v_1 - v_2)}{(\bar{p}v_1 + (1-\bar{p})z)} \\
& + \frac{\bar{p}v_2(v_1 - v_2)}{(\bar{p}v_2 + (1-\bar{p})z)} + (v_1 + v_2) \cdot \ln \left(\frac{\bar{p}v_2 + (1-\bar{p})z}{po v_1 + (1-\bar{p})z} \right) \Bigg]
\end{aligned} \qquad \text{(B.14)}
$$

Again we have pre-multiplication equation (B.11) through by constant $(v_1 - v_2)^3$. Then calculations to evaluate the left-hand side integral are given by

$$
\begin{aligned}
&(v_1 - v_2)^3 \cdot \int_{0}^{v_2} \frac{1}{z} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\, z + \bar{p}v_1\right)^2} \cdot \frac{\bar{p}(1-\bar{p})z}{\left((1-\bar{p})\, z + \bar{p}v_2\right)^2}\, dz \\
&= (v_1 - v_2)^3 \cdot (G_{2D}^{o}(v_2) - G_{2D}^{o}(0)) \\
&= \left(\frac{\bar{p}v_1(v_1 - v_2)}{((1-\bar{p})v_2 + \bar{p}v_1)} + \bar{p}(v_1 - v_2) + (v_1 + v_2) \ln \left(\frac{v_2}{((1-\bar{p})v_2 + \bar{p}v_1)} \right) \right) \\
&\quad - \left(2(v_1 - v_2) + (v_1 + v_2) \ln \left(\frac{v_2}{v_1} \right) \right) \\
&= \left(\frac{\bar{p}v_1(v_1 - v_2)}{((1-\bar{p})v_2 + v_1)} - (1-\bar{p})(v_1 - v_2) + (v_1 + v_2) \ln \left(\frac{v_1}{((1-\bar{p})v_2 + \bar{p}v_1)} \right) \right) \\
&\quad - \left((v_1 - v_2) \cdot \frac{((1-\bar{p})v_2 + v_1)}{((1-\bar{p})v_2 + v_1)} \right)
\end{aligned}
$$

$$\text{(B.15)} \quad = \left(\frac{-(1-\bar{p})v_2(v_1-v_2)}{((1-\bar{p})v_2+v_1)} - (1-\bar{p})(v_1-v_2) + (v_1+v_2)\ln\left(\frac{v_1}{((1-\bar{p})v_2+\bar{p}v_1)}\right)\right)$$

which exactly matches equation (B.12). Lastly, we calculate the derivative of G^{ω}_{2D} to confirm that it recovers the ω_z-side integrand of line (B.11). Again ignoring the constant of $(v_1-v_2)^3$ (and accounting for it later), starting from $\frac{dG^{\omega}_{2D}}{dz}$, we get

$$\begin{aligned}
&\frac{d}{dz}\left(\left[\frac{(1-\bar{q})v_1(v_1-v_2)}{((1-\bar{q})v_1+\bar{q}z)} + \frac{(1-\bar{q})v_2(v_1-v_2)}{((1-\bar{q})v_2+\bar{q}z)} + (v_1+v_2)\cdot\ln\left(\frac{(1-\bar{q})v_2+\bar{q}z}{(1-\bar{q})v_1+\bar{q}z}\right)\right]\right)\\
&= -\frac{(1-\bar{q})v_1(v_1-v_2)\bar{q}}{((1-\bar{q})v_1+\bar{q}z)^2} - \frac{(1-\bar{q})v_2(v_1-v_2)\bar{q}}{((1-\bar{q})v_2+\bar{q}z)^2}\\
&\quad + (v_1+v_2)\left[\frac{\bar{q}}{((1-\bar{q})v_2+\bar{q}z)} - \frac{\bar{q}}{(1-\bar{q})v_1+\bar{q}z)}\right]\\
&= \frac{\left(-\bar{q}(1-\bar{q})(v_1-v_2)\left[v_1\left((1-\bar{q})v_2+\bar{q}z\right)^2 + v_2\left((1-\bar{q})v_1+\bar{q}z\right)^2\right]\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&\quad + \frac{(\bar{q}(v_1+v_2)\left((1-\bar{q})v_1+\bar{q}z\right)\left((1-\bar{q})v_2+\bar{q}z\right)\left[(1-\bar{q})v_1-(1-\bar{q})v_2\right])}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&= \frac{\bar{q}(1-\bar{q})(v_1-v_2)\left(-v_1(1-\bar{q})^2v_2^2 - v_2(1-\bar{q})^2v_1^2 - 4\bar{q}(1-\bar{q})v_1v_2z - (v_1+v_2)\left[\bar{q}^2z^2\right]\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&\quad + \frac{\bar{q}(1-\bar{q})(v_1-v_2)\left((v_1+v2)\left[(1-\bar{q})^2v_1v_2 + \bar{q}(1-\bar{q})z(v_1+v_2) + \bar{q}^2z^2\right]\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&= \frac{\bar{q}(1-\bar{q})(v_1-v_2)\left(-4\bar{q}(1-\bar{q})v_1v_2z + (v_1+v_2)\left[\bar{q}(1-\bar{q})z(v_1+v_2)\right]\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&= \frac{\bar{q}^2(1-\bar{q})^2(v_1-v_2)\left(-4v_1v_2z + z\left[v_1^2+2v_1v_2+v_2^2\right]\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}\\
&= \frac{\bar{q}^2(1-\bar{q})^2(v_1-v_2)\left(z(v_1-v_2)^2\right)}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2} = (v_1-v_2)^3\cdot\frac{\bar{q}^2(1-\bar{q})^2z}{((1-\bar{q})v_1+\bar{q}z)^2\left((1-\bar{q})v_2+\bar{q}z\right)^2}
\end{aligned}$$

which indeed is the desired integrand form (after adjusting the constant).

THE SENTINEL