When do exact and powerful p-values and e-values exist?

Zhenyuan Zhang* Aaditya Ramdas† Ruodu Wang‡

29th May 2023

Abstract

Given a composite null $\mathcal{P}$ and composite alternative $\mathcal{Q}$, when and how can we construct a p-value whose distribution is exactly uniform under the null, and stochastically smaller than uniform under the alternative? Similarly, when and how can we construct an e-value whose expectation exactly equals one under the null, but its expected logarithm under the alternative is positive? We answer these basic questions, and other related ones, when $\mathcal{P}$ and $\mathcal{Q}$ are convex polytopes (in the space of probability measures). We prove that such constructions are possible if and only if (the convex hull of) $\mathcal{Q}$ does not intersect the span of $\mathcal{P}$. If the p-value is allowed to be stochastically larger than uniform under $P \in \mathcal{P}$, and the e-value can have expectation at most one under $P \in \mathcal{P}$, then it is achievable whenever $\mathcal{P}$ and $\mathcal{Q}$ are disjoint. The proofs utilize recently developed techniques in simultaneous optimal transport. A key role is played by coarsening the filtration: sometimes, no such p-value or e-value exists in the richest data filtration, but it does exist in some reduced filtration, and our work provides the first general characterization of when or why such a phenomenon occurs. We also provide an iterative construction that explicitly constructs such processes, that under certain conditions finds the one that grows fastest under a specific alternative Q. We discuss implications for the construction of composite nonnegative (super)martingales, and end with some conjectures and open problems.

Contents

*Department of Mathematics, Stanford University, USA. Email: zzy@stanford.edu.

†Departments of Statistics and Data Science, and Machine Learning, Carnegie Mellon University, USA. Email: aramdas@cmu.edu.

‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca.

1 Introduction

Consider a universe of distributions Π on a sample space $(\mathfrak{X}, \mathcal{F})$, where $\mathfrak{X}$ is a Polish space. The data are generated according to some $\mathbb{P} \in \Pi$. Let $\mathcal{P}$ and $\mathcal{Q}$ be disjoint subsets of Π. When we say we are testing $\mathcal{P}$, we mean that we are testing the null hypothesis $\mathbb{P} \in \mathcal{P}$. When we say we are testing $\mathcal{P}$ against $\mathcal{Q}$, we mean additionally that the alternative hypothesis is $\mathbb{P} \in \mathcal{Q}$.

We ask (and answer) two central questions in this paper. The first one is:

> **Q1.** Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, when can we find an *exact* p-value for $\mathcal{P}$ that has nontrivial power under $\mathcal{Q}$? To elaborate, we would like to find a $[0,1]$-valued random variable T that is exactly uniform for every $P \in \mathcal{P}$, but is stochastically smaller than uniform under every $Q \in \mathcal{Q}$.

The second central question in this paper is the following:

> **Q2.** Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, does there exist an *exact* e-value for $\mathcal{P}$ that has nontrivial power under $\mathcal{Q}$? To elaborate, we would like to find a nonnegative random variable X such that $\mathbb{E}^P[X] = 1$ for every $P \in \mathcal{P}$, but $\mathbb{E}^Q[\log X] > 0$ for every $Q \in \mathcal{Q}$.

We will provide a complete answer to both questions in this paper, when $\mathcal{P}$ and $\mathcal{Q}$ are convex polytopes in the space of probability measures on $\mathfrak{X}$. The solution is surprisingly clean: we prove that constructions for **Q1** and **Q2** exist if and only if (the convex hull of) the alternative $\mathcal{Q}$ does not intersect the span of the null $\mathcal{P}$.

We also answer the non-exact versions of both problems, where we only require T to be stochastically larger than uniform under any $P \in \mathcal{P}$:

> **Q3.** Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, does there exist a p-value for $\mathcal{P}$ that has nontrivial power against $\mathcal{Q}$?

Or, we require that $\mathbb{E}^P[X] \leqslant 1$ for any $P \in \mathcal{P}$:

> **Q4.** Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, does there exist an e-value for $\mathcal{P}$ that has nontrivial power against $\mathcal{Q}$?

For these non-exact problems, the solutions are still very clean: constructions for **Q3** and **Q4** exist if and only if the (convex hulls of) $\mathcal{P}$ and $\mathcal{Q}$ are disjoint.

These appear to be rather fundamental results, and will be proved using recent techniques in simultaneous optimal transport, combined with classical convex geometric arguments. Note that in the characterizations for **Q1** and **Q3** above, a technical condition of joint non-atomicity will be assumed, which is essentially equivalent to allowing for extra randomization. Our proofs are constructive, and yields a simple iterative construction, called SHINE, that can explicitly build these objects and calculate their values on a given dataset, but it seems only computationally feasible for low-dimensional settings.

Towards the end of the paper, we show how answers to the above two questions help answer a final related question:

> **Q5.** Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, can we determine if there is a nonnegative (super)martingale M for $\mathcal{P}$ that grows to infinity under $\mathcal{Q}$? In other words, when can we find a process M that is a nonnegative (super)martingale simultaneously under every $P \in \mathcal{P}$ but almost surely grows to infinity under every $Q \in \mathcal{Q}$?

Before proceeding, we introduce the most important terminology used throughout the paper.

Terminology. We define pivotal, exact, and nontrivial e- and p-variables below.

1. A random variable X is *pivotal* for $\mathcal{P}$ if X has the same distribution under all $P \in \mathcal{P}$.
2. A nonnegative random variable X is a *e-variable* for $\mathcal{P}$ if $\mathbb{E}^P[X] \leqslant 1$ for all $P \in \mathcal{P}$. An e-variable X for $\mathcal{P}$ is *exact* if $\mathbb{E}^P[X] = 1$ for all $P \in \mathcal{P}$. We say X is *nontrivial* for $\mathcal{Q}$ if $\mathbb{E}^Q[X] > 1$ for all $Q \in \mathcal{Q}$. An e-variable X for $\mathcal{P}$ is said to have *nontrivial e-power* against $\mathcal{Q}$ if for each $Q \in \mathcal{Q}$, $\mathbb{E}^Q[\log X] > 0$.
3. A nonnegative random variable X is a *p-variable* for $\mathcal{P}$ if $P(X \leqslant \alpha) \leqslant \alpha$ for all $\alpha \in (0, 1)$ and $P \in \mathcal{P}$, and a p-variable X is *exact* if $P(X \leqslant \alpha) = \alpha$ for all $\alpha \in (0, 1)$ and $P \in \mathcal{P}$. A p-variable X for $\mathcal{P}$ is *nontrivial* against $\mathcal{Q}$ if, for each $Q \in \mathcal{Q}$, $Q(X \leqslant \alpha) \geqslant \alpha$ for all $\alpha \in (0, 1)$ with strict inequality for some $\alpha \in (0, 1)$. Without loss of generality, p-variables can be restricted to the range $[0, 1]$ by truncation, without changing any of their properties.

Note that an exact p-variable is always pivotal, but not vice versa. An exact e-variable need not be pivotal, and a pivotal e-variable need not be exact. Using $x - 1 \geqslant \log x$, an e-variable that has nontrivial e-power against $\mathcal{Q}$ is also nontrivial for $\mathcal{Q}$. We will often omit $\mathcal{P}$ and $\mathcal{Q}$ in our subsequent mentions of p/e-variables when they are clear from the context.

Remark 1.1. For the majority of this paper, we suppress the raw data that is observed and used to form the p-values or e-values. One may simply assume that we have observed one data point Z from $\mathbb{P}$. This Z could itself be a random vector of some size $n \geqslant 1$ lying in (say) $\mathbb{R}^d$ for some $d \geqslant 1$ (which means $\mathbb{P}$ may be μ^n for some μ on $\mathbb{R}^d$), but we leave all this implicit. Thus our p-values and e-values can be treated as "single-period" statistics. We briefly return to the multi-period (sequential) case briefly later in the paper.

Related results. The most directly related work is that of Grünwald et al. [2023], which focuses primarily on e-values, and in particular **Q4**. To paraphrase one of their main results, consider any $\mathcal{P}$ and $\mathcal{Q}$ with a common reference measure, whose convex hulls do not intersect. They show that as long as a particular "worst case prior" exists, then one can construct an e-value for $\mathcal{P}$ which maximizes the worst case e-power for $\mathcal{Q}$. This is a topic we return to later in the paper, when we provide a more detailed geometric study of **Q3** and **Q4** together. We need fewer technical conditions to establish our results, but their additional assumptions allow them to derive an analogous result for **Q4** even when $\mathcal{P}, \mathcal{Q}$ are not convex polytopes.

A second related work is that of Ramdas et al. [2022b]. Here, the authors work in the sequential setting, and ask when nontrivial nonnegative (super)martingales for $\mathcal{P}^\infty := \{P^\infty : P \in \mathcal{P}\}$ exist. We can paraphrase their geometric solution: assuming a common reference measure, nontrivial nonnegative (super)martingales cannot exist if the "fork-convex hull" of $\mathcal{P}^\infty$ intersects $\mathcal{Q}^\infty$.

Thus, the above two papers both hinted at a deeper underlying geometric picture, and our work elaborates significantly on this theme, completely characterizing the case of convex polytopes. One key point is that the earlier works did not give a systematic and thorough treatment of what one can do in reduced filtrations, while this is a central aspect of our paper. Informally, we will (optimally) transport $\mathcal{P}$ to a single measure μ, while transporting $\mathcal{Q}$ to a single measure ν, and this collapse of the null and alternative corresponds exactly to working in a coarser σ-algebra.

The above idea of transport from multiple measures to specified measures is addressed in the framework of simultaneous transport studied by Wang and Zhang [2023]. We borrow techniques from this framework to provide answers to our questions, in particular, **Q1-Q3** and **Q5**.

Background on e-values. E-values are an alternative to p-values, and they have recently been actively studied in statistical testing by Shafer [2021], Vovk and Wang [2021], Grünwald et al. [2023] and Howard et al. [2021] under various names. Tests with e-values are often based on martingale techniques, which date back to Wald [1945], and they emphasize optional stopping or continuation of experiments. The notion of e-processes generalizes that of likelihood ratios to composite hypotheses. Some advantages of testing with e-values are summarized in Wang and Ramdas [2022, Section 2]. The idea of testing with e-values is intimately connected to the game-theoretic probability and statistics of Shafer and Vovk [2001, 2019]. For a review on e-values and game-theoretic statistics, see Ramdas et al. [2022a].

Notation. We collect the notation we use throughout this paper.

1. *Topology.* For a set $A \subseteq \mathbb{R}^d$, A° (resp. $\overline{A}$, ∂A, A^c, $\mathrm{Conv}A$) is the interior (resp. closure, boundary, complement, convex hull) of A and $\mathrm{aff}\, A$ is the smallest affine subspace of $\mathbb{R}^d$ containing A. For an affine subspace $S \subseteq \mathbb{R}^d$, we denote by $\mathrm{ri}(A; S)$ is the relative interior of A in S, that is, the interior of A in the relative topology on S.

2. *Probability and measure.* All measures we consider will be finite and have a finite first moment, i.e., $\int |x| \mu(\mathrm{d}x) < \infty$. For a Polish space $\mathfrak{X}$, we let $\mathcal{M}(\mathfrak{X})$ be the set of all finite measures on $\mathfrak{X}$ and $\Pi(\mathfrak{X})$ be the set of probability measures on $\mathfrak{X}$. For $\mu \in \mathcal{M}(\mathbb{R}^d)$, we denote its barycenter by $\mathrm{bary}(\mu) := \int_{\mathbb{R}^d} x\, \mu(\mathrm{d}x)/\mu(\mathbb{R}^d)$. For a finite set $\mathcal{A}$ of random variables or probability measures on the same space, we define $\mathrm{Conv}\mathcal{A}$ and $\mathrm{Span}\mathcal{A}$ in the usual sense of convex hull and span. We write $X \overset{\mathrm{law}}{\sim}_P \mu$, or simply $X \overset{\mathrm{law}}{\sim} \mu$, if the random variable X has distribution μ under P. We say "a probability measure μ is supported on a set A" if $\mu(A) = 1$. This does not imply that A is closed or $A = \mathrm{supp}\, \mu$. The product measure is denoted by $P \otimes Q$.

3. *Stochastic orders.* For $F, G \in \Pi(\mathbb{R})$, we write $F \preceq_{\mathrm{st}} G$ if $F((-\infty, a]) \geqslant G((-\infty, a])$ for all $a \in \mathbb{R}$. Also, $F \prec_{\mathrm{st}} G$ if $F \preceq_{\mathrm{st}} G$ and $F \neq G$. For $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$, we denote by $\mu \preceq_{\mathrm{cx}} \nu$ if $\int \phi\, \mathrm{d}\mu \leqslant \int \phi\, \mathrm{d}\nu$ for every convex function ϕ, in which case we say μ is smaller than ν in convex order.[1] If μ, ν are probability measures and $X \overset{\mathrm{law}}{\sim} \mu$, $Y \overset{\mathrm{law}}{\sim} \nu$, we sometimes abuse notation and write $X \preceq_{\mathrm{cx}} Y$ instead of $\mu \preceq_{\mathrm{cx}} \nu$. We write $\mu \leqslant \nu$ if $\mu(A) \leqslant \nu(A)$ for every Borel set A.

4. *Other notation.* Bold symbols such as $\mathbf{x}$ and $\boldsymbol{\alpha}$ will typically denote vectors. Write $\mathbf{1}_d = (1, \ldots, 1) \in \mathbb{R}^d$, $\mathbf{0}_d = (0, \ldots, 0) \in \mathbb{R}^d$, $\mathcal{I}_d = \{\mathbf{x} \in \mathbb{R}^d |\, x_1 = \cdots = x_d\} = \mathbb{R}\mathbf{1}_d$, and $\mathcal{I}_d^+ = \{\mathbf{x} \in \mathbb{R}^d |\, x_1 = \cdots = x_d \geqslant 0\} = \mathbb{R}_+\mathbf{1}_d$. When the dimension d is clear, we may omit the subscript d and write $\mathbf{1}, \mathbf{0}, \mathcal{I}, \mathcal{I}^+$ instead. We let U_1 denote the Lebesgue measure on $[0, 1]$. Denote the Euclidean norm by $\|\cdot\|$.

Outline of the paper. The rest of this paper is organized as follows. Section 2 provides the necessary mathematical background regarding convex order and simultaneous optimal transport. We then collect a few straightforward relations between the existences of p-values and e-values in Section 3, and further reduce our main problems to the case where $\mathcal{P}$ and $\mathcal{Q}$ are finite. The easier

[1] This is sometimes called the Choquet order in the mathematical literature, e.g., Simon [2011].

case with a simple alternative ($|\mathcal{Q}| = 1$) will be solved first in Section 4. Under suitable conditions, we solve the maximization problem of the e-power in Section 5 and illustrate the SHINE construction for finding a reasonably powerful e-variable in Section 6. We answer **Q1-Q4** in full in Section 7, where we deal with a general composite (and even infinite) alternative $\mathcal{Q}$. Finally, an application to composite test (super)martingales related to **Q5** will be discussed in Section 8, followed by a summary in Section 9. Appendix A contains a few technical results that are used in our proofs.

2 Preliminaries on convex order and simultaneous transport

In this section we collect results related to convex order and simultaneous transport for future use. We rely on some results from Shaked and Shanthikumar [2007] and Wang and Zhang [2023].

We recall from Strassen [1965] that $\mu \preceq_{\mathrm{cx}} \nu$ if and only if there exists a martingale transport from μ to ν, i.e., a martingale coupling (X, Y) such that $X \overset{\text{law}}{\sim} \mu$ and $Y \overset{\text{law}}{\sim} \nu$. This result is called Strassen's theorem. The relation $\preceq_{\mathrm{cx}}$ is a partial order on $\Pi(\mathbb{R}^d)$. Given a subset $\mathcal{N} \subseteq \Pi(\mathbb{R}^d)$, we say μ is a (Pareto) *maximal* element in $\mathcal{N}$ if there exists no $\nu \in \mathcal{N}$ such that $\nu \neq \mu$ and $\mu \preceq_{\mathrm{cx}} \nu$; we say μ is the *maximum* element in $\mathcal{N}$ if $\nu \preceq_{\mathrm{cx}} \mu$ for each $\nu \in \mathcal{N}$.

In the following, we collect a few properties of the convex order. These results can be found in Shaked and Shanthikumar [2007, Section 3.A].

Lemma 2.1. *Assuming all random variables below take values in $\mathbb{R}$ and are integrable, the following statements hold.*

(i) If $\mathbb{E}[X] = \mathbb{E}[Y]$, then $X \preceq_{\mathrm{cx}} Y$ if and only if

$$\mathbb{E}[(X-a)_+] \leqslant \mathbb{E}[(Y-a)_+] \text{ for all } a \in \mathbb{R}.$$

(ii) If $\{X_n\}$ is a sequence of random variables that converge weakly to X and $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$, then

$$X_n \preceq_{\mathrm{cx}} Y \implies X \preceq_{\mathrm{cx}} Y.$$

For $d \geqslant 1$ and two $\mathbb{R}^d$-valued measures $\boldsymbol{\mu}, \boldsymbol{\nu}$ on Polish spaces $\mathfrak{X}, \mathfrak{Y}$ (denoted by $\boldsymbol{\mu} \in \mathcal{M}(\mathfrak{X})^d$ and $\boldsymbol{\nu} \in \mathcal{M}(\mathfrak{Y})^d$) such that $\boldsymbol{\mu}(\mathfrak{X}) = \boldsymbol{\nu}(\mathfrak{Y})$, let $\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ denote the set of all *simultaneous transport maps* and *plans* from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$ respectively, i.e.,

$$\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \{T : \mathfrak{X} \to \mathfrak{Y} \mid \boldsymbol{\mu} \circ T^{-1} = \boldsymbol{\nu}\},$$

and $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is the set of all stochastic kernels κ such that

$$\kappa_{\#}\boldsymbol{\mu}(\cdot) := \int_{\mathfrak{X}} \kappa(x; \cdot)\boldsymbol{\mu}(\mathrm{d}x) = \boldsymbol{\nu}(\cdot).$$

When $d = 1$, $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is often represented as the set of all joint distributions on $\mathfrak{X} \times \mathfrak{Y}$ whose marginals are $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ respectively, but for $d > 1$, we prefer the above representation. To further characterize the existence of simultaneous transport maps and plans, we need the notion of joint non-atomicity.

Definition 2.2. Consider a tuple of probability measures $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ on a Polish space $\mathfrak{X}$. We say that $\boldsymbol{\mu}$ is *jointly atomless* if there exist $\mu \gg \sum_{i=1}^d \mu_i$ and a random variable ξ such that under μ, ξ is atomless and independent of $(\mathrm{d}\mu_1/\mathrm{d}\mu, \dots, \mathrm{d}\mu_d/\mathrm{d}\mu)$.

As a simple example, $(\mu_1 \times \mathrm{U}_1, \dots, \mu_d \times \mathrm{U}_1)$ on $\mathfrak{X} \times [0, 1]$ is jointly atomless for each collection $(\mu_1, \dots, \mu_d)$ on $\mathfrak{X}$. We refer to Shen et al. [2019] and Wang and Zhang [2023] for more discussions on this notion.

In statistical terms, the hypothesis $\{P_1, \dots, P_L\}$ as a tuple being jointly atomless is equivalent to allowing for additional randomization, i.e., simulating a uniform random variable independent of $(\mathrm{d}P_1/\mathrm{d}P, \dots, \mathrm{d}P_L/\mathrm{d}P)$ for some $P \in \Pi(\mathfrak{X})$. It suffices if simulating a uniform random variable independent of existing random variables is always allowed. Such an assumption is common in statistical methods based on resampling or data splitting.

Proposition 2.3. *Consider $\boldsymbol{\mu} \in \Pi(\mathfrak{X})^d$ and $\boldsymbol{\nu} \in \Pi(\mathfrak{Y})^d$. Let $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ satisfy $\|\boldsymbol{\lambda}\|_1 = 1$, $\mu_j \ll \mu := \boldsymbol{\lambda}^\top \boldsymbol{\mu}$, and $\nu_j \ll \nu := \boldsymbol{\lambda}^\top \boldsymbol{\nu}$ for each $1 \leqslant j \leqslant d$.*

(i) The set $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is non-empty if and only if

$$\left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}, \dots, \frac{\mathrm{d}\mu_d}{\mathrm{d}\mu}\right)\Big|_{\mu} \succeq_{\mathrm{cx}} \left(\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu}, \dots, \frac{\mathrm{d}\nu_d}{\mathrm{d}\nu}\right)\Big|_{\nu},$$

where $X|_P$ means the distribution of a random variable X under a measure P.

(ii) Assume that $\boldsymbol{\mu}$ is jointly atomless. The set $\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is non-empty if and only if

$$\left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}, \dots, \frac{\mathrm{d}\mu_d}{\mathrm{d}\mu}\right)\Big|_{\mu} \succeq_{\mathrm{cx}} \left(\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu}, \dots, \frac{\mathrm{d}\nu_d}{\mathrm{d}\nu}\right)\Big|_{\nu}.$$

Proof. Theorem 3.4 of Wang and Zhang [2023] implies that the statements hold with $\boldsymbol{\lambda} = (1/d, \dots, 1/d)$. The more general case follows from Lemma 3.5 of Shen et al. [2019], in the direction (iii) $\Longrightarrow$ (ii) there. □

We briefly describe the intuition behind this result, which is crucial for our paper. The pushforward $\kappa_\# \boldsymbol{\mu}$ mixes the ratios between different coordinates of the (vector-valued) masses of $\boldsymbol{\mu}$ at different places of $\mathfrak{X}$; see Figure 2. The "ratios" can be recognized as Radon-Nikodym derivatives. The "mix" effect can be interpreted as a "backward martingale transport", because looking in the backward direction gives rise to a martingale coupling of the Radon-Nikodym derivatives. Strassen's theorem then gives the convex order constraint on the Radon-Nikodym derivatives. In Wang and Zhang [2023], such an observation leads also to the MOT-SOT parity that relates the simultaneous transport to the underlying martingale transport, which will be useful for our purpose when constructing explicitly an e/p-variable. We state a weak form of the MOT-SOT parity below, which can be proved similarly to Corollary 3 of Wang and Zhang [2023]. In the sequel, a coupling (X, Y) is *backward martingale* if $\mathbb{E}[X|Y] = Y$; that is, (Y, X) forms a martingale. It is *Monge* if Y is a measurable function of X.

Proposition 2.4. *Let $\boldsymbol{\mu} \in \Pi(\mathfrak{X})^d$ and $\boldsymbol{\nu} \in \Pi(\mathfrak{Y})^d$ satisfy $\boldsymbol{\mu} \ll \mu_d$, $\boldsymbol{\nu} \ll \nu_d$, and $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ non-empty. Suppose that $\boldsymbol{\mu}$ is jointly atomless and $(\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\mu_d)|_{\mu_d}$ is atomless. Then there exists a backward martingale coupling between $(\mathrm{d}\boldsymbol{\mu}/\mathrm{d}\mu_d)|_{\mu_d}$ and $(\mathrm{d}\boldsymbol{\nu}/\mathrm{d}\nu_d)|_{\nu_d}$ that is also Monge. Moreover, if we denote by h the map that induces this Monge transport, then there exists a simultaneous transport map $Y \in \mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ satisfying*

$$\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\nu}(Y(x)) = h\left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}\mu}(x)\right), \ x \in \mathfrak{X}.$$

Finally, we recall the following basic fact on Radon-Nikodym derivatives.

Lemma 2.5. *Let $d \in \mathbb{N}$ and τ be a probability measure supported on $\mathbb{R}_+^d$ with mean $\mathbf{1}$. Then there exist probability measures $F_1, \dots, F_d$ supported on $[0, 1]$ such that*

$$\left(\frac{\mathrm{d}F_1}{\mathrm{d}\mathrm{U}_1}, \dots, \frac{\mathrm{d}F_d}{\mathrm{d}\mathrm{U}_1}\right)\Big|_{\mathrm{U}_1} = \tau.$$

Proof. Since U_1 is atomless, $\mathcal{T}(\mathrm{U}_1, \tau) \neq \emptyset$. Pick $(f_1, \dots, f_d) \in \mathcal{T}(\mathrm{U}_1, \tau)$, and define F_i by $\mathrm{d}F_i/\mathrm{d}\mathrm{U}_1 = f_i$ for $1 \leqslant i \leqslant d$. This is well-defined since f_i is nonnegative a.e. and $\mathbb{E}^{\mathrm{U}_1}[f_i] = 1$ for each $1 \leqslant i \leqslant d$. □

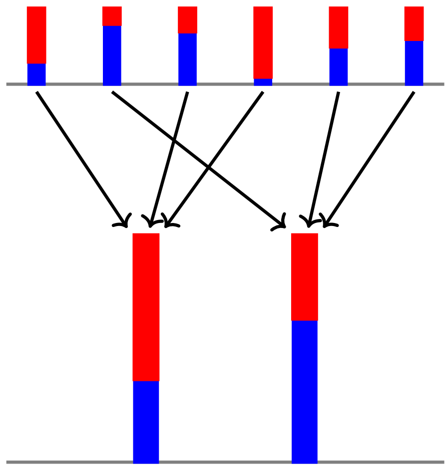


Figure 1: A showcase of simultaneous transport and mixing of the Radon-Nikodym derivatives; red and blue represent the two dimensions of the vector $\boldsymbol{\mu}$, with the height of a bar indicating its mass. This figure is taken from Wang and Zhang [2023, Figure 1].

3 General relations on the existence of p- and e-variables

For convex polytopes $\mathcal{P}$ and $\mathcal{Q}$ in Π, we may write $\mathcal{P} = \mathrm{Conv}\widetilde{\mathcal{P}}$ and $\mathcal{Q} = \mathrm{Conv}\widetilde{\mathcal{Q}}$ where $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{Q}}$ are finite. The following result helps us to reduce the problems to the case where $\mathcal{P}$ and $\mathcal{Q}$ are finite.

Proposition 3.1. *Suppose that $\mathcal{P} = \mathrm{Conv}\widetilde{\mathcal{P}}$ and $\mathcal{Q} = \mathrm{Conv}\widetilde{\mathcal{Q}}$.*

(i) There exists an (exact) nontrivial p-variable for $\mathcal{P}$ and $\mathcal{Q}$ if and only if the same exists for $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{Q}}$.[2]

(ii) There exists a (pivotal, exact, bounded) e-variable that is nontrivial for (or has nontrivial e-power against) $\mathcal{Q}$ for $\mathcal{P}$ and $\mathcal{Q}$ if and only if the same exists for $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{Q}}$.

Proof. This is clear from definitions of p/e-variables. □

As a result of the above proposition, in what follows, we can concern ourselves, without loss of generality, with the case where $\mathcal{P}$ and $\mathcal{Q}$ are finite subsets of $\Pi(\mathfrak{X})$ (except for Section 7.2).

Proposition 3.2. *Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are both finite. Let X be a (pivotal and exact) bounded e-variable for $\mathcal{P}$ that is nontrivial for $\mathcal{Q}$. Then there exists a (pivotal and exact) bounded e-variable for $\mathcal{P}$ that has nontrivial e-power against $\mathcal{Q}$.*

Proof. The following fact is crucial: by the Taylor expansion of the log function, for every $\varepsilon > 0$, there exists $\delta > 0$, such that for each $x \in [1-\delta, 1+\delta]$, $(1-\varepsilon)(x-1) \leqslant \log x \leqslant (1+\varepsilon)(x-1)$. Note that each $Y = (1-b) + bX$ with $b > 0$ is clearly an e-variable. On the other hand, since X is bounded, the range of Y can be chosen arbitrarily close to 1 by picking b small enough. Using $\min_{Q \in \mathcal{Q}} \mathbb{E}^Q[X] > 1$ we see that with b small enough, Y is an e-variable that has nontrivial e-power against $\mathcal{Q}$. Note that pivotality and exactness are preserved under this transformation. □

[2]Here and later, we mean that the statement holds regardless of whether the bracketed constraint exists, i.e., the current sentence contains two (similar) statements.

Remark 3.3. Proposition 3.2 also holds true without the boundedness assumption on X, i.e., there exists a (pivotal and exact) e-variable nontrivial for $\mathcal{Q}$ if and only if there exists a (pivotal and exact) e-variable that has nontrivial e-power against $\mathcal{Q}$. Indeed, this is a direct consequence of our main results, Theorems 7.3 and 7.4 below. However, we are not aware of a simpler proof of this fact.

In the sequel, when the equivalence of the existences is clear, we may write "there exists a nontrivial e-variable" instead. When $\mathcal{Q}$ is infinite, these two definitions are in general different, as shown by the following example.

Example 3.4. Let Z_μ denote the law $\mathrm{N}(\mu, 1)$ for $\mu \in \mathbb{R}$, and consider $\mathcal{P} = \{Z_0\}$ and $\mathcal{Q} = \{Z_\mu \mid \mu > 0\}$. Clearly, $X(\omega) = 1/2 + \mathbb{1}_{\{\omega>0\}}$ is a bounded e-variable that is nontrivial for $\mathcal{Q}$. Suppose for contradiction that Y is a bounded e-variable that has nontrivial e-power against $\mathcal{Q}$. Since Y cannot be a constant, $\mathbb{E}^{Z_0}[\log Y] < \mathbb{E}^{Z_0}[Y] - 1 = 0$. Since $Z_\mu \to Z_0$ in total variation as $\mu \to 0$, we have for $\mu > 0$ small enough that $\mathbb{E}^{Z_\mu}[\log Y] < 0$, contradicting $\mathbb{E}^{Q}[\log Y] > 0$ for all $Q \in \mathcal{Q}$.

The following calibration result is in place to help us construct an e-variable based on a p-variable.

Proposition 3.5. *Suppose that $\mathcal{Q}$ is finite.*

(i) If there exists an exact (hence pivotal) and nontrivial p-variable, then there exists a pivotal, exact, and bounded e-variable with nontrivial e-power against $\mathcal{Q}$.

(ii) If there exists a nontrivial p-variable, then there exists an e-variable with nontrivial e-power against $\mathcal{Q}$.

Proof. (i) Suppose that X is an exact nontrivial p-variable. It follows that $E := 2 - 2X$ is a pivotal, exact, and bounded e-variable, and $\mathbb{E}^{Q}[E] > 1$ for each $Q \in \mathcal{Q}$. Proposition 3.2 then finishes the proof. (ii) is similar. □

Our next simple result provides general necessary conditions for the existence of p/e-variables, hence answering the trivial parts of **Q1-Q4**.

Proposition 3.6. *Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are arbitrary subsets of $\Pi(\mathfrak{X})$.*

(i) If there exists an e-variable nontrivial for $\mathcal{Q}$, then $\mathrm{Conv}\mathcal{P} \cap \mathrm{Conv}\mathcal{Q} = \emptyset$.

(ii) If there exists an exact e-variable nontrivial for $\mathcal{Q}$, then $\mathrm{Span}\mathcal{P} \cap \mathrm{Conv}\mathcal{Q} = \emptyset$.

Proof. For (i), suppose that $R \in \mathrm{Conv}\mathcal{P} \cap \mathrm{Conv}\mathcal{Q}$, then since $\mathbb{E}^P[X] \leqslant 1$ for all $P \in \mathcal{P}$, $\mathbb{E}^R[X] \leqslant 1$. But $\mathbb{E}^Q[X] > 1$ for all $Q \in \mathcal{Q}$ implies $\mathbb{E}^R[X] > 1$, contradiction. For (ii), suppose that $R \in \mathrm{Span}\mathcal{P} \cap \mathrm{Conv}\mathcal{Q}$, then $\mathbb{E}^P[X] = 1$ for all $P \in \mathcal{P}$ gives that $\mathbb{E}^R[X] = 1$. But $\mathbb{E}^Q[X] > 1$ for all $Q \in \mathcal{Q}$ gives $\mathbb{E}^R[X] > 1$, contradiction. □

Let us end this section by incorporating the following important result, which sometimes helps us remove the jointly atomless condition when pivotality is not involved.

Proposition 3.7. *Let $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta_0}$ and $\mathcal{Q} = \{Q_{\theta'}\}_{\theta'\in\Theta_1}$. If there exists an (exact) e-variable (defined on $(\mathfrak{X} \times [0,1], \mathcal{F} \otimes \mathcal{B}([0,1]))$) that is nontrivial for $\{Q_{\theta'} \times \mathrm{U}_1\}_{\theta'\in\Theta_1}$ with null $\{P_\theta \times \mathrm{U}_1\}_{\theta\in\Theta_0}$, then there exists an (exact) e-variable that is nontrivial for $\mathcal{Q}$ with null $\mathcal{P}$.*

Proof. Let Y be an exact e-variable that is nontrivial for $\{Q_{\theta'} \times \mathrm{U}_1\}_{\theta'\in\Theta_1}$ with null $\{P_\theta \times \mathrm{U}_1\}_{\theta\in\Theta_0}$. Define $X = \mathbb{E}^{\mathrm{U}_1}[Y]$ by taking expectation of Y over the second coordinate. Then $\mathbb{E}^{P_i}[X] = \mathbb{E}^{P_i\times\mathrm{U}_1}[Y] = 1$ and $\mathbb{E}^{Q_j}[X] = \mathbb{E}^{Q_j\times\mathrm{U}_1}[Y] > 1$, meaning that X is an exact e-variable nontrivial for $\mathcal{Q}$. The non-exact case is similar. □

4 Composite null and simple alternative

In this section, we characterize the existence of exact and pivotal p-variables and e-variables for composite null and simple alternative (singleton). Although our results in this case are covered by the more general result for composite alternatives treated in Section 7, studying this setting first helps with building intuition behind our proof techniques. Moreover, the concept of e-power studied in Section 5 is defined for a single Q in the alternative hypothesis. We fix $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q\}$ in $\Pi(\mathfrak{X})$ and assume that $(P_1, \dots, P_L, Q)$ is jointly atomless (unless otherwise stated). The main results are Theorems 4.3 and 4.4 below. When $P_1, \dots, P_L \ll Q$, we write the measure $\gamma = (\mathrm{d}P_1/\mathrm{d}Q, \dots, \mathrm{d}P_L/\mathrm{d}Q)|_Q$ on $\mathbb{R}^L$.

Lemma 4.1. *Suppose that $Q \notin \mathrm{Span}(P_1, \dots, P_L)$ and $P_1, \dots, P_L \ll Q$. There exists a disjoint collection of closed balls $B_1, \dots, B_k$ in $\mathbb{R}^L$ of positive measure (under γ) not containing $\mathbf{1}$ such that denoting by t_j the point of B_j closest to $\mathbf{1}$, we have $\mathbf{1} \in \mathrm{Conv}(\{t_1, \dots, t_k\})^\circ$.*

Proof. Since $Q \notin \mathrm{Span}(P_1, \dots, P_L)$, the measure γ cannot have support contained in a hyperplane in $\mathbb{R}^L$ by definition. In other words, $\mathrm{aff}\,\mathrm{supp}\,\gamma = \mathbb{R}^L$. By Lemma A.1(ii), $\mathbf{1} = \mathrm{bary}(\gamma) \in (\mathrm{Conv}\,\mathrm{supp}\,\gamma)^\circ$. Therefore, there exist $s_1, \dots, s_k \in \mathrm{supp}\,\gamma$ such that $\mathbf{1} \in (\mathrm{Conv}\{s_1, \dots, s_k\})^\circ$. Let B_j be the ball centered at s_j with radius $r > 0$ for $1 \leqslant j \leqslant k$. For r small enough, these balls will be disjoint from $\mathbf{1}$, and the closest points $t_1, \dots, t_k$ satisfy $\mathbf{1} \in \mathrm{Conv}(\{t_1, \dots, t_k\})^\circ$. □

Proposition 4.2. *We have $Q \notin \mathrm{Span}(P_1, \dots, P_L)$ if and only if there exist probability measures $G \neq F$ such that*

$$\mathcal{K}((P_1, \dots, P_L, Q), (F, \dots, F, G)) \neq \emptyset.$$

If moreover $(P_1, \dots, P_L, Q)$ is jointly atomless, then $Q \notin \mathrm{Span}(P_1, \dots, P_L)$ if and only if there exist probability measures $G \neq F$ such that

$$\mathcal{T}((P_1, \dots, P_L, Q), (F, \dots, F, G)) \neq \emptyset.$$

In addition, in both cases above, we may pick $F = \mathrm{U}_1$ and G atomless.

Proof. The "if" is clear. For "only if", let $F = \mathrm{U}_1$ and consider first the case where $P_1, \dots, P_L \ll Q$. Then using Proposition 2.3, it suffices to prove that there exists some $G \gg F$ such that

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}, \frac{\mathrm{d}Q}{\mathrm{d}Q}\right)\Big|_Q \succeq_{\mathrm{cx}} \left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G}, \frac{\mathrm{d}G}{\mathrm{d}G}\right)\Big|_G.$$

Equivalently,

$$\gamma = \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}\right)\Big|_Q \succeq_{\mathrm{cx}} \left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G}\right)\Big|_G. \tag{1}$$

We will first consider a special type of density $\mathrm{d}F/\mathrm{d}G$ which allows us to construct G such that (1) holds. Suppose that

$$\frac{\mathrm{d}G}{\mathrm{d}F}(x) = \begin{cases} 1 & \text{if } 0 \leqslant x \leqslant 1 - \varepsilon; \\ 1 + \varepsilon & \text{if } 1 - \varepsilon < x \leqslant 1 - \frac{\varepsilon}{2}; \\ 1 - \varepsilon & \text{if } 1 - \frac{\varepsilon}{2} < x \leqslant 1, \end{cases}$$

where $\varepsilon > 0$ is a small number. Clearly, G is atomless. Moreover, $(\mathrm{d}F/\mathrm{d}G)|_G$ is concentrated on $[(1+\varepsilon)^{-1}, (1-\varepsilon)^{-1}]$ and $\mathbb{P}^G[\mathrm{d}F/\mathrm{d}G = 1] = 1 - \varepsilon$. Therefore, the measure $(\mathrm{d}F/\mathrm{d}G, \dots, \mathrm{d}F/\mathrm{d}G)|_G$ is supported on the line segment $\{\mathbf{x} \in \mathbb{R}^L \mid x_1 = \dots = x_L \in [(1+\varepsilon)^{-1}, (1-\varepsilon)^{-1}]\}$, with mean $\mathbf{1}$ and $\mathbb{P}^G[\mathrm{d}F/\mathrm{d}G \neq 1] = \varepsilon$. We will find a measure $(\mathrm{d}F/\mathrm{d}G, \dots, \mathrm{d}F/\mathrm{d}G)|_G$ that satisfies the condition above and also (1).

Consider a disjoint collection of closed balls $\{B_j\}_{1 \leqslant j \leqslant k}$ in $\mathbb{R}^L$ as constructed in Lemma 4.1. By Lemma A.2, there is $\delta > 0$ and a segment $\{\mathbf{x} \in \mathbb{R}^L \mid x_1 = \dots = x_L \in [1-\delta, 1+\delta]\}$ containing

$\mathbf{1}$, such that any measure of total mass δ supported on it will be smaller in extended convex order than some $\widetilde{\gamma}$ such that $\widetilde{\gamma} \leqslant \gamma|_{\bigcup_{j=1}^k B_j}$. We choose $\varepsilon > 0$ so that $(1-\varepsilon)^{-1} < 1+\delta$. As a result, the measure G constructed in the above paragraph satisfies

$$\omega := \left(\left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G} \right) |_G \right) \Big|_{\mathbb{R}^L \setminus \{\mathbf{1}\}} \preceq_{\mathrm{cx}} \widetilde{\gamma}.$$

The measure $(\mathrm{d}F/\mathrm{d}G, \dots, \mathrm{d}F/\mathrm{d}G)|_G - \omega$ is concentrated at $\mathbf{1}$, which is smaller in convex order than any measure with barycenter $\mathbf{1}$ and the same total mass. Since $\mathrm{bary}(\gamma) = \mathrm{bary}(\widetilde{\gamma}) = \mathbf{1}$, we conclude

$$\left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G} \right) \Big|_G \preceq_{\mathrm{cx}} \gamma.$$

If $P_1, \dots, P_L \ll Q$ does not hold, then we define $Q' = Q/2 + (P_1 + \dots + P_L)/(2L)$, and repeat the above arguments, so that there is κ sending $(P_1, \dots, P_L, Q')$ to some $(F, \dots, F, G')$ where $G' \neq F$. By linearity, κ also sends $(P_1, \dots, P_L, Q)$ to $(F, \dots, F, G)$ where $G = 2G' - F \neq F$. □

Theorem 4.3. *Suppose that we are testing $\mathcal{P} = \{P_1, \dots, P_L\}$ against $\mathcal{Q} = \{Q\}$. The following are equivalent:*

(a) there exists an exact (hence pivotal) and nontrivial p-variable;

(b) there exists a pivotal, exact, and bounded e-variable that has nontrivial e-power against $\mathcal{Q}$;

(c) there exists an exact e-variable that is nontrivial against $\mathcal{Q}$;

(d) there exists a random variable X that is pivotal for $\mathcal{P}$ but has a different distribution under Q, where the laws of X under both are atomless;

(e) it holds that $Q \notin \mathrm{Span}(P_1, \dots, P_L)$.

Proof. The direction $(a) \Rightarrow (b)$ is precisely Proposition 3.5, $(b) \Rightarrow (c)$ is clear from definition, $(c) \Rightarrow (e)$ being precisely Proposition 3.6, and $(e) \Rightarrow (d)$ is Proposition 4.2. To show $(d) \Rightarrow (a)$, let X be a random variable that has a common law F under $P \in \mathcal{P}$, and law G under Q. Let ϕ be as given in Lemma A.3. It follows immediately that $\phi \circ X$ is an exact p-variable. □

Theorem 4.4. *Suppose that we are testing $\mathcal{P} = \{P_1, \dots, P_L\}$ against $\mathcal{Q} = \{Q\}$. The following are equivalent:*

(a) there exists a nontrivial p-variable;

(b) there exists an e-variable that is nontrivial for $\mathcal{Q}$;

(c) it holds that $Q \notin \mathrm{Conv}(P_1, \dots, P_L)$.

Proof. That $(a) \Rightarrow (b)$ is precisely Proposition 3.5; $(b) \Rightarrow (c)$ is clear from Proposition 3.6. For $(c) \Rightarrow (a)$, we define the set $U = U_L = (-\infty, 1)^L \cup (1, \infty)^L \cup \{\mathbf{1}\}$. We claim that it suffices to find a measure $\mu \preceq_{\mathrm{cx}} \gamma$ that is supported on U and not equal to $\delta_{\mathbf{1}}$. Given such μ, we apply Lemma 2.5 to find measures $F_1, \dots, F_L$ such that $(\mathrm{d}F_1/\mathrm{dU}_1, \dots, \mathrm{d}F_L/\mathrm{dU}_1)|_{\mathrm{U}_1} = \mu$. Since μ is supported on U, we may without loss assume that there is a threshold $\beta \in (0,1)$ such that for each $1 \leqslant i \leqslant L$, $\mathrm{d}F_i/\mathrm{dU}_1 \leqslant 1$ on $[0, \beta)$ and $\mathrm{d}F_i/\mathrm{dU}_1 \geqslant 1$ on $(\beta, 1]$. In particular, $F_i \succ_{\mathrm{st}} \mathrm{U}_1$. Proposition 2.3 then yields a random variable $X \in \mathcal{T}((P_1, \dots, P_L, Q), (F_1, \dots, F_L, \mathrm{U}_1))$. Let Ψ be as given in Lemma A.4(ii). By definition, $\Psi \circ X$ is a nontrivial p-variable.

To find a measure $\mu \preceq_{\mathrm{cx}} \gamma$ that is supported on U and not equal to $\delta_{\mathbf{1}}$, for simplicity we translate U by $\mathbf{1}$, and from now on $U = (-\infty, 0)^L \cup (0, \infty)^L \cup \{\mathbf{0}\}$ and γ has mean $\mathbf{0}$. Our goal is to find a measure μ supported on U such that $\mu \preceq_{\mathrm{cx}} \gamma$ and $\mu \neq \delta_{\mathbf{0}}$. We apply induction on L. Suppose that $L = 2$. Then since $Q \notin \mathrm{Conv}(P_1, P_2)$, the measure γ is not supported on any line that has a negative slope and contains $\mathbf{0}$. There are two cases.

- If γ is not supported on any line (hyperplane in $\mathbb{R}^2$), then $Q \notin \mathrm{Span}(P_1, P_2)$. By Theorem 4.3, a nontrivial p-variable exists.

- If γ is supported on a line, then such a line must contain $\mathbf{0}$ and have a positive slope, and hence is contained in U.

Now suppose that $L > 2$. We say a set $K \subseteq \mathbb{R}^L$ is a linear cone if it is the union of two symmetric convex cones at $\mathbf{0}$ in $\mathbb{R}^L$. Clearly, U is a linear cone, and

(i) the intersection of a subspace and a linear cone is a linear cone;

(ii) if S is a subspace of $\mathbb{R}^L$ and K is a linear cone, then $\{\mathbf{0}\} \subsetneq S \cap K$ if and only if there exists a one-dimensional subspace T of S such that $T \subseteq K$.

Since $Q \notin \mathrm{Conv}(P_1, \dots, P_L)$, the measure γ is not supported on any hyperplane that is contained in $U^c \cup \{\mathbf{0}\}$ and contains $\mathbf{0}$. If γ is not supported on any hyperplane, then using Theorem 4.3, a nontrivial p-variable exists. Thus we may assume that γ is supported on some hyperplane S such that $\{\mathbf{0}\} \subsetneq S \cap U$. We then lower the dimension by one and identify $S = \mathbb{R}^{L-1}$. There are two cases.

- If γ is not supported on any hyperplane in S, then there exist points $x_1, \dots, x_{L-1} \in \mathrm{supp}\,\gamma$ such that $\mathrm{ri}\,\mathrm{aff}\{x_1, \dots, x_{L-1}\} = S$. Let T be a one-dimensional subspace of S such that $T \subseteq U$. Then $T \cap \mathrm{Conv}\{x_1, \dots, x_{L-1}\}$ is nonempty. Using Lemma A.2, we may find a measure μ supported on the bounded set $T \cap \mathrm{Conv}\{x_1, \dots, x_{L-1}\}$ (and thus supported on U) such that $\mu \preceq_{\mathrm{cx}} \gamma$ and $\mu \neq \delta_{\mathbf{0}}$. It follows that a nontrivial p-variable exists.

- If γ is supported on a hyperplane S' of S, then by Mazur's separation theorem (Conway [1990, Corollary 3.4]) and since γ is not supported on any hyperplane that intersects with U only at $\mathbf{0}$, we must have $\{\mathbf{0}\} \subsetneq S' \cap U$. In this case, we have reduced the dimension by one. Thus induction works for this case.

By reducing the problem iteratively in the above manner, we eventually arrive at the problem with $L = 2$, which we already showed above. □

Remark 4.5. The directions $(c) \Leftrightarrow (e)$ in Theorem 4.3 and $(b) \Leftrightarrow (c)$ in Theorem 4.4 also hold without the condition that $(P_1, \dots, P_L, Q)$ is jointly atomless, in view of Proposition 3.7.

Example 4.6. (i) Let $P_1 \sim \mathrm{Ber}(0.1)$, $P_2 \sim \mathrm{Ber}(0.2)$, and $Q \sim \mathrm{Ber}(0.3)$. It follows that $Q \in \mathrm{Span}(P_1, P_2) \setminus \mathrm{Conv}(P_1, P_2)$. By Theorems 4.3 and 4.4, a nontrivial e-variable (or p-variable) exists, but an exact nontrivial e-variable (or p-variable) does not exist.

(ii) Let $P_1 \sim \mathrm{N}(-1, 1)$, $P_2 \sim \mathrm{N}(1, 1)$, and $Q \sim \mathrm{N}(0, 1)$. By Theorem 4.3, there exists a pivotal exact nontrivial e-variable (or p-variable).

5 Constructing a powerful e-variable

We focus on e-variables in this section. Provided the existence, our next step is to maximize the e-power of an e-variable that is pivotal and exact. The e-power of an e-variable X can be measured by $\mathbb{E}^Q[\log X]$, which has long been a popular criterion; see for example Kelly [1956], Shafer [2021], Grünwald et al. [2023], Waudby-Smith and Ramdas [2022].[3] It has been recently called the *e-power* of X (Vovk and Wang [2022]), a term we continue to use for simplicity. For $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q\}$ such that $P_1, \dots, P_L \ll Q$ and $(P_1, \dots, P_L, Q)$ is jointly atomless, our goal is to solve

$$\max\{\mathbb{E}^Q[\log X] : X \text{ is an pivotal exact e-variable}\}. \tag{2}$$

[3]In short, it captures the rate of growth of the test martingale under the alternative Q; see Section 8.

This optimization problem turns out to be a special case of a more general problem that is illustrated by (6) below. Such a connection will be explained in Section 5.1. We describe an equivalent condition for the existence of a maximal element for (6) in Section 5.2. A further sufficient condition in the case $L = 2$ is illustrated in Section 5.3. Finally, we provide several examples in Section 5.5. In this section, we denote by γ the law of $(\mathrm{d}P_1/\mathrm{d}Q, \dots, \mathrm{d}P_L/\mathrm{d}Q)$ under Q. In particular, γ is a probability measure on $\mathbb{R}_+^L$ with mean $\mathbf{1}$.

5.1 E-power maximization and convex order

We first recall the maximizer of e-power in the case of a simple null versus a simple alternative, which has an explicit form. This fact is used frequently in the above literature.

Example 5.1. Let us first illustrate an example with simple null $\mathcal{P} = \{P\}$ $(L = 1)$ and simple alternative $\mathcal{Q} = \{Q\}$. Clearly, any e-variable is pivotal. We arrive at the optimization problem

$$\max\{\mathbb{E}^Q[\log X] : X \geqslant 0,\, \mathbb{E}^P[X] = 1\}. \tag{3}$$

By Gibbs' inequality, the maximum value is attained by the likelihood ratio, i.e., when $X = \mathrm{d}Q/\mathrm{d}P$.

Below we illustrate the solution to (3) using our theory, which sheds light on the composite null case. For simplicity, we assume $P \ll Q$ and (P, Q) jointly atomless. Denote by $\gamma := (\mathrm{d}P/\mathrm{d}Q)|_Q$. Consider the set $\mathcal{M}_\gamma$ of probability measures μ such that $\mu \preceq_{\mathrm{cx}} \gamma$. Using Lemma 2.5, every $\mu \in \mathcal{M}_\gamma$ corresponds to a probability measure F such that $(\mathrm{d}F/\mathrm{dU}_1)|_{\mathrm{U}_1} = \mu$. By Proposition 2.3, there exists a random variable Y that has law F under P and law U_1 under Q. In the next step, we assert in (3) that X is of the form $X = (\mathrm{dU}_1/\mathrm{d}F)(Y)$, and optimize $\mathbb{E}^Q[\log X]$ over $F \in \mathcal{M}_\gamma$. It is clear that the constraint

$$\mathbb{E}^P[X] = \mathbb{E}^P\Big[\frac{\mathrm{dU}_1}{\mathrm{d}F}(Y)\Big] = \mathbb{E}^F\Big[\frac{\mathrm{dU}_1}{\mathrm{d}F}\Big] = 1$$

is satisfied, and the objective in (3) becomes

$$\mathbb{E}^Q[\log X] = \mathbb{E}^Q\Big[\log\Big(\frac{\mathrm{dU}_1}{\mathrm{d}F}(Y)\Big)\Big] = \mathbb{E}^{\mathrm{U}_1}\Big[\log\frac{\mathrm{dU}_1}{\mathrm{d}F}\Big] = \mathbb{E}^{\mathrm{U}_1}\Big[-\log\frac{\mathrm{d}F}{\mathrm{dU}_1}\Big].$$

We have thus arrived at the optimization problem

$$\max\left\{\mathbb{E}^{\mathrm{U}_1}\Big[-\log\frac{\mathrm{d}F}{\mathrm{dU}_1}\Big] : \frac{\mathrm{d}F}{\mathrm{dU}_1}\Big|_{\mathrm{U}_1} \in \mathcal{M}_\gamma\right\}. \tag{4}$$

The value (4) gives a lower bound on (3). Since the set $\mathcal{M}_\gamma$ has a maximum element γ in convex order, the problem (4) has a trivial solution $\mathbb{E}^Q[-\mathrm{d}P/\mathrm{d}Q]$. This corresponds to the solution to (3) using Gibbs' inequality.

The fact that the two values (3) and (4) are the same is not a coincidence and holds more generally for composite null, which we will prove in Proposition 5.2. With a composite null, the main difficulty arises from solving (4), because the set $\mathcal{M}_\gamma$ has a complicated structure, and may not contain a maximum element in convex order.

As explained in Example 5.1, the first step to solving (2) is to impose the further condition that X is of the form $(\mathrm{d}G/\mathrm{d}F)(Y)$ for some F, G, Y. As a consequence of Gibbs' inequality, this does not affect the optimal value of (2).

Proposition 5.2. *There exists a maximizer X to (2) of the form $X = (\mathrm{d}G/\mathrm{d}F)(Y)$, where $F, G \in \Pi(\mathbb{R})$, and $Y \in \mathcal{T}((P_1, \dots, P_L, Q), (F, \dots, F, G))$.*

Proof. Suppose that Z is a maximizer to (2). Since Z is a pivotal e-variable, we denote by F' as the common distribution of Z under P_i, $1 \leqslant i \leqslant L$, and G' the distribution of Z under Q. Let $\widetilde{Z}$ be the identity random variable on $\mathbb{R}$, we have $\mathbb{E}^{F'}[\widetilde{Z}] = 1$. By Gibbs' inequality,

$$\mathbb{E}^Q[\log Z] = \mathbb{E}^{G'}[\log \widetilde{Z}] \leqslant \mathbb{E}^{G'}\Big[\log\frac{\mathrm{d}G'}{\mathrm{d}F'}\Big] = \mathbb{E}^Q\Big[\log\Big(\frac{\mathrm{d}G'}{\mathrm{d}F'}(Z)\Big)\Big].$$

Thus, $X = (\mathrm{d}G'/\mathrm{d}F')(Z)$ must also be a maximizer to (3). $\square$

Given $X = (\mathrm{d}G/\mathrm{d}F)(Y)$ where $Y \in \mathcal{T}((P_1, \dots, P_L, Q), (F, \dots, F, G))$, we may rewrite

$$\mathbb{E}^Q[\log X] = \mathbb{E}^Q\Big[\log\Big(\frac{\mathrm{d}G}{\mathrm{d}F}(Y)\Big)\Big] = \mathbb{E}^G\Big[-\log\frac{\mathrm{d}F}{\mathrm{d}G}\Big].$$

As a consequence of Proposition 2.3, the optimization problem (2) is equivalent to finding

$$\max\left\{\mathbb{E}^G\Big[-\log\frac{\mathrm{d}F}{\mathrm{d}G}\Big] : \left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G}\right)\Big|_G \preceq_{\mathrm{cx}} \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}\right)\Big|_Q\right\}. \tag{5}$$

More generally, since $x \mapsto -\log x$ is convex on its domain, we may formulate the problem of optimizing $\mathbb{E}^G[\phi(\mathrm{d}F/\mathrm{d}G)]$ for all convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$. In other words, let γ be the law of $(\mathrm{d}P_1/\mathrm{d}Q, \dots, \mathrm{d}P_L/\mathrm{d}Q)$ under Q and introduce the set $\mathcal{M}_\gamma$ of probability measures supported on $\mathcal{I}_L^+$ that is dominated by γ in convex order, and our goal is

$$\text{to maximize } \mu \text{ in } \preceq_{\mathrm{cx}}, \text{ subject to } \mu \in \mathcal{M}_\gamma. \tag{6}$$

This will the goal of the present section. The reader should keep in mind that unfortunately, even if (6) allows a unique maximum element, it does not necessarily solve (2) uniquely when the logarithm in (2) is replaced by other concave functions. This is because Proposition 5.2 requires Gibbs' inequality, where the logarithm plays a crucial role.

5.2 Existence of the maximum element in convex order

To ease our presentation, we will assume further that

$$\gamma \text{ does not give positive mass to any hyperplane in } \mathbb{R}^L. \tag{N}$$

That is, for every half-space $\mathbb{H} \subseteq \mathbb{R}^L$, $\gamma(\partial\mathbb{H}) = 0$. This is a technical assumption which greatly simplifies our proofs (as we will explain in Remarks 5.7 and 5.11), and we expect that analogous results hold without such an assumption.

Proposition 5.3. *In the above setting, consider $x \geqslant 0$. There exists a closed half-space $\mathbb{H}_x$ of $\mathbb{R}^L$ and a measure μ_x supported on $\mathbb{H}_x$, such that*

(i) the positive diagonal $\mathcal{I}_L^+ \not\subseteq \mathbb{H}_x$;

(ii) $-\mathbf{1} \in \mathbb{H}_x$;

(iii) $x\mathbf{1} \in \partial\mathbb{H}_x = \mathbb{H}_x \cap \mathbb{H}_x^c$, where $\mathbb{H}_x^c$ is the closed complement of $\mathbb{H}_x$;

(iv) the measure $\mu_{\mathbb{H}_x^c} := \gamma - \mu_x$ is supported on $\mathbb{H}_x^c$, and the barycenters of μ_x and $\mu_{\mathbb{H}_x^c}$ both lie on $\mathcal{I}^+$.

Moreover, if (N) holds, there exists a unique measure μ_x satisfying the above conditions. In this case, we call $\partial\mathbb{H}_x$ a separating hyperplane at x.

Remark 5.4. In fact, it also holds that $\mu_x = \gamma|_{\mathbb{H}_x}$ and $\mu_{\mathbb{H}_x^c} = \gamma|_{\mathbb{H}_x^c}$. If moreover, γ has a strictly positive density on $\mathbb{R}_+^d$, then $\mathbb{H}_x$ is also unique.

Proof. We induct on L. The base case is $L = 1$, where the claims follow simply by picking $\mathbb{H}_x = (-\infty, x]$.

Fix an arbitrary $L \geqslant 2$ and $x \geqslant 0$. Consider the plane $\mathcal{P}_L = \{\mathbf{x} \in \mathbb{R}^L \mid x_1 = x_2\} \subseteq \mathbb{R}^L$, so that $\mathcal{I}^+ \subseteq \mathcal{P}_L$. The collection of lines in $\mathcal{P}_L$ through $x\mathbf{1}$ will be denoted by $\mathcal{L}_\theta$, $\theta \in [0, 2\pi)$, where $\mathcal{I}^+ \subseteq \mathcal{L}_0$. For each $\mathcal{L}_\theta$, consider the projection of γ on the hyperplane P_θ to which $\mathcal{L}_\theta$ is normal.

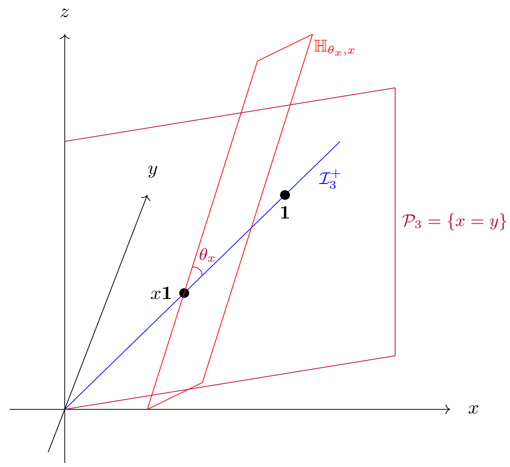


Figure 2: Scheme of the proof of Proposition 5.3 for $L = 3$

It follows from our induction hypothesis that there is some half-space $\mathbb{H}_{\theta,x}$ of $\mathbb{R}^L$ on which some measure $\mu_{\theta,x} \leqslant \gamma$ is supported, such that $\mathrm{bary}(\mu_{\theta,x}) \in \mathcal{P}_L$ and $\mathrm{bary}(\gamma - \mu_{\theta,x}) \in \mathcal{P}_L$, as well as $\mathcal{L}_\theta \subseteq \partial\mathbb{H}_{\theta,x}$.

Suppose that (i) does not hold. Then γ is supported on a hyperplane S in $\mathbb{R}^L$ containing $\mathcal{I}_L$. By the induction hypothesis, we may find a closed half-space $\mathbb{H}'_x$ of S satisfying the conditions (i)-(iv). Clearly, any closed half-space $\mathbb{H}_x$ of $\mathbb{R}^L$ containing $\mathbb{H}'_x$ also satisfies the same conditions.

Therefore, we may assume (i) and that γ is not supported on any hyperplane in $\mathbb{R}^L$. In particular, $\mu_{0,x}$ and $\mu_{\pi,x}$ are non-zero. In this case, $\mathrm{bary}(\mu_{0,x})$ and $\mathrm{bary}(\mu_{\pi,x})$ lie in the two different half-planes in $\mathcal{P}_L$ separated by $\mathcal{I}^+$. By continuity of the measure, there exists some θ_x such that $\mathrm{bary}(\mu_{\theta_x,x}) \in \mathcal{I}^+$. This establishes (iii) and (iv). Finally, by replacing $\mathbb{H}_x$ by $\mathbb{H}_x^c$, we may assume that (ii) holds as well.

Suppose that (N) holds and μ_x, ν_x are distinct measures satisfying the above conditions. Then $\mu_x - \nu_x$ is a nontrivial signed measure supported on a hyperplane in $\mathbb{R}^L$, contradicting (N). □

Recall from (6) that our goal is to find the maximum element in $\mathcal{M}_\gamma$ in convex order.

Theorem 5.5. *Assuming (N), the following are equivalent.*

(a) There exists a unique maximum element μ in convex order in $\mathcal{M}_\gamma$, i.e., $\mu \preceq_{\mathrm{cx}} \gamma$ and for each ν supported on $\mathcal{I}^+$ with $\nu \preceq_{\mathrm{cx}} \gamma$, it holds that $\nu \preceq_{\mathrm{cx}} \mu$.

(b) The class of measures $\{\mu_x\}_{x \geqslant 0}$ from Proposition 5.3 is monotone (in the usual order), i.e., for all $x \leqslant y$, $\mu_x \leqslant \mu_y$.

Example 5.6. Suppose that $L = 1$. It is clear from the proof of Proposition 5.3 that condition (b) is always satisfied. Therefore, the maximum element μ in $\mathcal{M}_\gamma$ always exists. This agrees with Example 5.1, where the likelihood ratio maximizes the e-power.

Remark 5.7. The only place we used our assumption (N) is on the uniqueness of the measure μ_x in Proposition 5.3. When there is no uniqueness, the condition (b) in Theorem 5.5 needs to be replaced by the existence of a monotone selection of measures $\{\mu_x\}_{x \geqslant 0}$, each of them satisfying the conditions in Proposition 5.3.

The condition (b) in Theorem 5.5 is in general not easy to check, especially in higher dimensions.[4] Later, we supply a sufficient condition in Section 5.3, and a few examples in Section 5.5.

[4]In this paper when we mention "dimension" we always mean the dimension of null/alternative, but not dimension of the space $\mathfrak{X}$.

Lemma 5.8. *Let ν be a probability measure on $\mathcal{I}_L^+$ such that $\nu \preceq_{\mathrm{cx}} \gamma$. For $\mathbb{H}_x$ and μ_x defined in Proposition 5.3, denote by $b_x\mathbf{1}$ the barycenter of μ_x and ξ distributed as the first marginal of ν. Then $\mathbb{E}[(\xi - x)_-] \leqslant (x - b_x)\mu_x(\mathbb{R}^L)$ for all $x \geqslant 0$. Moreover, equality holds for x if and only if for every martingale coupling (X, Y) such that $X \overset{\text{law}}{\sim} \nu$ and $Y \overset{\text{law}}{\sim} \gamma$, it holds $(Y \mid X \leqslant x) \overset{\text{law}}{\sim} \mu_x/\mu_x(\mathbb{R}^L)$.*

Proof. Let v_x be a unit normal vector to $\partial\mathbb{H}_x$, such that the angle θ_x between the vectors v_x and $\mathbf{1}$ satisfies $0 \leqslant \theta_x < \pi/2$. For $y \in \mathbb{R}^L$ we write $a_y = \langle y, v_x\rangle$ with Euclidean inner product. Define $\phi_x : \mathbb{R}^L \to \mathbb{R}$ by

$$\phi_x(y) := \begin{cases} a_y/\cos\theta_x & \text{if } a_y \leqslant 0; \\ 0 & \text{if } a_y > 0. \end{cases}$$

Since ϕ_x is concave and $\nu \preceq_{\mathrm{cx}} \gamma$, it follows that

$$(b_x - x)\mu_x(\mathbb{R}^L) = \int \phi_x \mathrm{d}\gamma \leqslant \int \phi_x \mathrm{d}\nu = -\mathbb{E}[(\xi - x)_-].$$

This completes the proof. The rest is clear. $\square$

Proof of Theorem 5.5. We first prove $(b) \Rightarrow (a)$. We first characterize the measure μ by the cumulative density function of its first marginal (recall that μ is supported on the nonnegative diagonal $\mathcal{I}^+$). For $x \geqslant 0$, pick μ_x as in Proposition 5.3. Note that $x \mapsto \mu_x(\mathbb{R}^L)$ is nondecreasing in x and continuous by (b). Define μ by the unique probability measure on $\mathcal{I}^+$ such that $\mu([0, x]^L) = \mu_x(\mathbb{R}^L)$.

We next show that $\mu \preceq_{\mathrm{cx}} \gamma$. By Strassen's theorem, it suffices to find a martingale coupling (X, Y) such that $X \overset{\text{law}}{\sim} \mu$ and $Y \overset{\text{law}}{\sim} \gamma$. Let us fix $X \overset{\text{law}}{\sim} \mu$ and let

$$(Y \mid x < X \leqslant x') \overset{\text{law}}{\sim} \frac{\mu_{x'} - \mu_x}{\mu_{x'}(\mathbb{R}^L) - \mu_x(\mathbb{R}^L)},$$

where we identify the random variable X supported on $\mathcal{I}$ with its first coordinate. This defines a coupling (X, Y) since $x \leqslant x' \implies \mu_x \leqslant \mu_{x'}$. Let $a \geqslant 0$ be arbitrary. On the event $\{X \leqslant a\}$, Y is distributed as $\mu_a/\mu_a(\mathbb{R}^L)$. By Proposition 5.3, we have $\mathbb{E}[Y \mathbb{1}_{\{X \leqslant a\}}] \in \mathcal{I}^+$. In addition,

$$\mathbb{E}\left[X \mathbb{1}_{\{X \leqslant a\}}\right] = \left(\int_0^a x \,\mathrm{d}(\mu_x(\mathbb{R}^L))\right) \times \mathbf{1},$$

which is exactly $\mathrm{bary}(\mu_a)$ projected to $\mathcal{I}$. Therefore, we must have $\mathbb{E}[X \mathbb{1}_{\{X \leqslant a\}}] = \mathbb{E}[Y \mathbb{1}_{\{X \leqslant a\}}]$, so that (X, Y) is indeed a martingale. Thus $\mu \preceq_{\mathrm{cx}} \gamma$.

Now by Lemma 2.1(i), it suffices to show that for each $\nu \preceq_{\mathrm{cx}} \gamma$ and ξ_ν, ξ_μ denoting the first marginals of ν, μ, it holds that $\mathbb{E}[(\xi_\nu - x)_-] \leqslant \mathbb{E}[(\xi_\mu - x)_-]$ for all $x \geqslant 0$. This is indeed a consequence of Lemma 5.8, since

$$\mathbb{E}[(\xi_\mu - x)_-] = x\mu_x(\mathbb{R}^L) - \int x \,\mathrm{d}(\mu_x(\mathbb{R}^L)) = (x - b_x)\mu_x(\mathbb{R}^L).$$

We next show $(a) \Rightarrow (b)$. Suppose that $x < y$ but $\mu_x \not\leqslant \mu_y$. In particular, $\operatorname{supp}\mu_x \not\subseteq \operatorname{supp}\mu_y$. We may assume that $\mu_x(\mathbb{R}^L)$ and $\mu_y(\mathbb{R}^L)$ are positive. Suppose for contradiction that (a) holds with a maximum element μ. We define the measures

$$\nu_x = \mu_x(\mathbb{R}^L)\delta_{\mathrm{bary}(\mu_x)} + (1 - \mu_x(\mathbb{R}^L))\delta_{\mathrm{bary}(\gamma - \mu_x)}$$

and similarly ν_y. Then with the usual notation,

$$\mathbb{E}[(\xi_\mu - x)_-] \geqslant \mathbb{E}[(\xi_{\nu_x} - x)_-] = (x - b_x)\mu_x(\mathbb{R}^L) \tag{7}$$

and

$$\mathbb{E}[(\xi_\mu - y)_-] \geqslant \mathbb{E}[(\xi_{\nu_y} - y)_-] = (y - b_y)\mu_y(\mathbb{R}^L). \tag{8}$$

By Lemma 5.8, equalities hold for (7) and (8), and for every martingale coupling (X, Y) such that $X \overset{\text{law}}{\sim} \mu$ and $Y \overset{\text{law}}{\sim} \gamma$, it holds $(Y \mid X \leqslant x) \overset{\text{law}}{\sim} \mu_x/\mu_x(\mathbb{R}^L)$ and $(Y \mid X \leqslant y) \overset{\text{law}}{\sim} \mu_y/\mu_y(\mathbb{R}^L)$. This contradicts $\operatorname{supp}\mu_x \not\subseteq \operatorname{supp}\mu_y$. $\square$

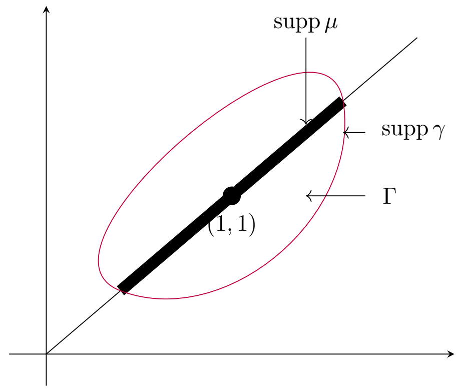


Figure 3: Illustration of Theorem 5.9. The convex set Γ is enclosed by the red contour $\partial\Gamma$ on which γ (the law of $(\mathrm{d}P_1/\mathrm{d}Q, \mathrm{d}P_2/\mathrm{d}Q)$ under Q) is supported. The measure μ is supported on the thick segment on the diagonal $\mathcal{I}$.

5.3 A sufficient condition in case $|\mathcal{P}| = 2$

When $L = 2$, we provide a sufficient condition for the class of measures $\{\mu_x\}_{x\geqslant 0}$ to satisfy the monotonicity condition $x \leqslant y \implies \mu_x \leqslant \mu_y$. In view of Theorem 5.5, this condition implies the existence of the maximum element μ. We keep the same setting as in Section 5.2 and assume (N), with the exception that $L = 2$.

Theorem 5.9. *Suppose that there exists a convex set $\Gamma \subseteq \mathbb{R}^2$ such that $\gamma(\partial\Gamma) = 1$.*[5] *Then there exists a unique maximum element μ in convex order in $\mathcal{M}_\gamma$. Moreover, μ is the unique probability measure on the $\mathcal{I}_2^+$ with $\mu([0, x]^2) = \mu_x(\mathbb{R}^2)$, where μ_x was given in Proposition 5.3 applied with $L = 2$. In particular, there exist distinct measures $F, G \in \Pi(\mathbb{R})$ such that $(\mathrm{d}F/\mathrm{d}G, \dots, \mathrm{d}F/\mathrm{d}G)|_G = \mu$, attaining the maximum in (5).*

Lemma 5.10. *Suppose that there exists a convex set $\Gamma \subseteq \mathbb{R}^2$ such that γ is supported on $\partial\Gamma$. For $x \geqslant 0$, let $\mathbb{H}_x$ and μ_x be defined as in Proposition 5.3. Then for $0 \leqslant x \leqslant x'$, $\mu_x \leqslant \mu_{x'}$.*

Proof. Fix $0 \leqslant x < x'$. Define $C_1 = \mathbb{H}_x \setminus \mathbb{H}_{x'}^\circ$ and $C_2 = \mathbb{H}_{x'} \setminus \mathbb{H}_x^\circ$. It follows that the positive part of $\mu_x - \mu_{x'}$ is supported on C_1. Let us define

$$S = \mathbb{R}\mathbf{1} + \partial\mathbb{H}_x \cap \partial\mathbb{H}_{x'}.$$

The line S separates $\mathbb{R}^2$ into two (closed) half-spaces, and we denote by $\mathbb{H}'$ the one that does not contain C_1. Since the barycenters of μ_x and $\mu_{x'}$ lie on $\mathcal{I}^+$, it suffices to show that $\gamma(C_2 \setminus (C_1 \cup \mathbb{H}')) = 0$. Suppose not. Then there exist $z_1 \in \partial\Gamma \cap C_1$ and $z_2 \in \partial\Gamma \cap C_2 \setminus (C_1 \cup \mathbb{H}')$. Since Γ is convex, it cannot hold that both $x\mathbf{1}$ and $x'\mathbf{1}$ belong to Γ°. Suppose that $x\mathbf{1} \notin \Gamma^\circ$. Then by convexity of Γ and our assumption (N), we have $\mu_x = 0$, thus $\mu_x \leqslant \mu_{x'}$ holds trivially. The case $x'\mathbf{1} \notin \Gamma^\circ$ is similar. $\square$

Proof of Theorem 5.9. The first claim follows from Theorem 5.5 and Lemma 5.10. The existence of F, G follows from Lemma 2.5, by setting $G = \mathrm{U}_1$. $\square$

Remark 5.11. With essentially the same arguments, we may remove assumption (N) from Theorem 5.9. With the presence of atoms, selecting any monotone collection $\{\mu_x\}_{x\geqslant 0}$ would be enough; see Remark 5.7.

[5]This assumption is far from being necessary, but might be convenient to verify.

5.4 On multiple observations

Before we proceed, let us remark on the case with multiple data points. Suppose that instead of one data point, we observe n iid data points $Z_1, \dots, Z_n$ in the space $\mathfrak{X}$ from the experiment. The e-variable is built based on the n data points together instead of a single data point. In other words, given $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q\}$, we build an e-variable that is pivotal, exact, and has nontrivial power against $\mathcal{Q}^n := \{Q^n\}$ for $\mathcal{P}^n := \{P_1^n, \dots, P_L^n\}$. We first see that, as long $Q \notin \mathcal{P}$, at most two observations are needed to build a pivotal and exact e-variable based on Theorem 4.3. Note that no joint non-atomicity or absolute continuity needs to be assumed for this result.

Proposition 5.12. *Suppose that $P_1, \dots, P_L \ll Q$ and $Q \notin \mathcal{P}$. Then either $Q \notin \mathrm{Span}\mathcal{P}$ or $Q^2 \notin \mathrm{Span}\mathcal{P}^2$.*

The proof of Proposition 5.12 is put in Appendix A. Note that the absolute continuity condition cannot be removed.

Let us denote by ℓ_n the maximum e-power with n data points for $\mathcal{P}^n$ against $\mathcal{Q}^n$ using a pivotal and exact e-variable, similarly as in (2).

Proposition 5.13. *In the setting above, suppose that $P_1, \dots, P_L \ll Q$ and $(P_1, \dots, P_L, Q)$ is jointly atomless, and $Q \notin \mathcal{P}$. For any $n, m \in \mathbb{N}$, $\ell_{n+m} \geqslant \ell_n + \ell_m$. In particular, ℓ_n/n converges to a positive limit less than or equal to $\min_{P\in\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$.*

Proof. We first show that for any $n, m \in \mathbb{N}$, $\ell_{n+m} \geqslant \ell_n + \ell_m$. Suppose that $X^{(n)}$ attains the maximum e-power among pivotal and exact e-variables against $\mathcal{Q}^n$ for $\mathcal{P}^n$, and $X^{(m)}$ against $\mathcal{Q}^m$ for $\mathcal{P}^m$. Define $X^{(n+m)}(\omega_1, \omega_2) = X^{(n)}(\omega_1)X^{(m)}(\omega_2)$, where $\omega_1 \in \mathfrak{X}^n$ and $\omega_2 \in \mathfrak{X}^m$. Clearly, $X^{(n+m)}$ is pivotal and exact against $\mathcal{Q}^{n+m}$ for $\mathcal{P}^{n+m}$. Its e-power is $\mathbb{E}^{Q^{n+m}}[\log X^{(n+m)}] = \mathbb{E}^{Q^n}[\log X^{(n)}] + \mathbb{E}^{Q^m}[\log X^{(m)}]$, thus $\ell_{n+m} \geqslant \ell_n + \ell_m$ holds. It then follows from Fekete's lemma that ℓ_n/n converges to some limit in $\mathbb{R} \cup \{\infty\}$. Since this e-power is bounded from above by the e-power for $\{P_1^n\}$ against $\{Q^n\}$, we have for all $1 \leqslant i \leqslant L$ that $\ell_n/n \leqslant \mathbb{E}^{Q^n}[\log(\mathrm{d}Q^n/\mathrm{d}P_i^n)]/n = \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P_i)]$ as we see in Example 5.1. That the limit is positive follows from Proposition 5.12, Theorem 4.3, and $\ell_{n+m} \geqslant \ell_n + \ell_m$. $\square$

Asymptotically, is there a loss of power caused by imposing exactness or pivotality? By Proposition 5.13, $\ell_{2^n}/2^n$ is increasing in n. This means the loss of power caused by imposing pivotality and exactness is getting weaker as the number of data points grows. It remains an open question whether $\ell_n/n \to \min_{P\in\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$. If this holds true, then the loss of power vanishes asymptotically, by noting that $n \min_{P\in\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$ is the theoretical best e-power for testing $\mathcal{P}^n$ against $\mathcal{Q}^n$ (see Example 5.1). In Example 5.15, we present a setting of normal distributions in which $\ell_n/n \to \min_{P\in\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$ holds true. We conjecture that this limit holds true in general, but we did not find a proof.

5.5 Examples

The condition in Theorem 5.9 that $\gamma = (\mathrm{d}P_1/\mathrm{d}Q, \mathrm{d}P_2/\mathrm{d}Q)|_Q$ is supported on the boundary of a convex set is not very restrictive. When $P_1, P_2, Q \in \Pi(\mathbb{R})$, the vector of density functions $((\mathrm{d}P_1/\mathrm{d}Q)(x), (\mathrm{d}P_2/\mathrm{d}Q)(x))$ forms a parameterized curve in $\mathbb{R}^2$ by $x \in \mathbb{R}$. In certain nice cases, such a curve lies on the boundary of a convex set. We illustrate with a few examples below.

Example 5.14. Consider $P_1 \sim \mathrm{N}(-1, 1)$, $P_2 \sim \mathrm{N}(1, 1)$, and $Q \sim \mathrm{N}(0, 1)$. It follows from direct computation that

$$\gamma = \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)\Big|_Q = \left(e^{-\xi-1/2}, e^{\xi-1/2}\right)\Big|_{\xi \overset{\text{law}}{\sim} \mathrm{N}(0,1)},$$

which is supported on the hyperbola $\{(x_1, x_2) \in \mathbb{R}_+^2 | x_1x_2 = 1/e\}$, the boundary of the convex set $\{(x_1, x_2) \in \mathbb{R}_+^2 \mid x_1x_2 \geqslant 1/e\}$. By Theorem 5.9, there exists a unique maximal element μ in $\mathcal{M}_\gamma$ in convex order.

Using the above notation, it is easy to see that $\mathbb{H}_x = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leqslant 2x\}$ and $\mu_x = \gamma|_{\mathbb{H}_x}$. Moreover, Theorem 5.9 yields that μ is the unique probability measure on $\mathcal{I}^+$ with

$$\mu([0,x]^2) = 2\Phi\left(\log(\sqrt{e}x + \sqrt{ex^2 - 1})\right) - 1 \text{ for } x \geqslant \frac{1}{\sqrt{e}}, \tag{9}$$

where Φ is the Gaussian cumulative density function. It can be directly seen from the figure below that points $x, y \in \mathbb{R}$ are shrunk to a single point precisely when the points $(e^{x-1/2}, e^{-x-1/2})$ and $(e^{y-1/2}, e^{-y-1/2})$ are symmetric around $\mathcal{I}$. This happens if and only if $x = -y$. In other words, the most powerful pivotal e-variable is a function of $|Z|$, where Z is the observed data point. Using Example 5.1 on testing the simple hypothesis $|Z| \overset{\text{law}}{\sim} |\xi + 1|$ against $|Z| \overset{\text{law}}{\sim} |\xi|$, this e-variable is given by $X = 2e^{1/2}/(e^Z + e^{-Z}) = e^{1/2}\cosh(Z)^{-1}$, and the e-power is $\mathbb{E}^Q[\log X] \approx 0.125$. In the sequential setting where iid observations $Z_1, \dots, Z_n$ are available (treated in the next example), we effectively reduce the filtration generated by $Z_1, \dots, Z_n$ to the one generated by $|Z_1|, \dots, |Z_n|$. This corresponds to the intuition that taking absolute value transports P_1, P_2 to the same measure but not for Q, and indeed this is the optimal solution to (5).

Example 5.15. We consider the setting in Example 5.14 but instead of one data point, we observe n iid data points $Z_1, \dots, Z_n$ in the experiment. Here, we build an e-variable based on the n data points together instead of building an e-variable for each data point; this allows for more flexibility than Example 5.14. In this setting, $P_1 = \mathrm{N}(-\mathbf{1}_n, I_n)$, $P_2 = \mathrm{N}(\mathbf{1}_n, I_n)$, and $Q = \mathrm{N}(\mathbf{0}_n, I_n)$, where $\mathbf{1}_n = (1, \dots, 1) \in \mathbb{R}^n$, $\mathbf{0}_n = (0, \dots, 0) \in \mathbb{R}^n$, and I_n is the $n \times n$ identity matrix. It follows from direct computation that

$$\gamma = \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)\Big|_Q = \left(e^{-\xi - n/2}, e^{\xi - n/2}\right)\Big|_{\xi \overset{\text{law}}{\sim} \mathrm{N}(0,n)},$$

which is very similar to Example 5.14. Using a similar argument as in Example 5.14, the most powerful pivotal e-variable is given by $E_n = e^{n/2}\cosh(\sum_{i=1}^n Z_i)^{-1}$. Note that this is different from the sequential one built in Example 5.14 which is $E_n^* = e^{n/2}\prod_{i=1}^n \cosh(Z_i)^{-1}$. The contrast between E_n and E_n^* is interesting to discuss. On the one hand, E_n has better e-power than E_n^* since $E_n \geqslant E_n^*$ due to log-convexity of the cosh function. This is intuitive, as E_n^* effectively tests more null hypotheses such as $\mathrm{N}(\boldsymbol{\mu}, I_n)$ for $\boldsymbol{\mu} \in \{-1, 1\}^n$ than E_n. On the other hand, $n \mapsto E_n^*$ is a martingale under both P_1 and P_2, but we can check that $n \mapsto E_n$ is not a martingale under either P_1 or P_2. In Section 6, we will compare the e-power of the two approaches numerically, and in Section 8, we further discuss test martingales. Finally, we note that $\ell_n/n = \mathbb{E}^Q[\log E_n]/n \to 1/2 = \min_{i=1,2}\mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P_i)]/n$, and hence the upper bound in Proposition 5.13 is sharp. On the other hand, $\mathbb{E}^Q[\log E_n^*]/n = 1/2 - \mathbb{E}^Q[\log\cosh(Z_1)] \approx 0.125$.

Example 5.16. Let us examine some further sufficient conditions with $L = 2$. Consider $P_1, P_2, Q \in \Pi(\mathbb{R})$ such that $P_1, P_2 \ll Q$ and $\mathrm{d}P_i/\mathrm{d}Q \in C^2(\mathbb{R})$ for $i = 1, 2$. Recall that a simple C^2 parameterized curve $(x(t), y(t))$ in $\mathbb{R}^2$ lies on the boundary of a convex set if and only if its curvature

$$k = \frac{x'y'' - y'x''}{((x')^2 + (y')^2)^{3/2}}$$

is always nonnegative or always nonpositive (Theorem 2.31 of Kühnel [2015]). Therefore, $(\mathrm{d}P_1/\mathrm{d}Q, \mathrm{d}P_2/\mathrm{d}Q)$ lies on the boundary of a convex set if

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}\right)'\left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)'' - \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}\right)''\left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)'$$

remains of a constant sign. As a simple example, this is the case if P_1, P_2, Q are Gaussian distributions on $\mathbb{R}$ with different means but the same variance, or with the same mean but different variances. In particular, this recovers Example 5.14.

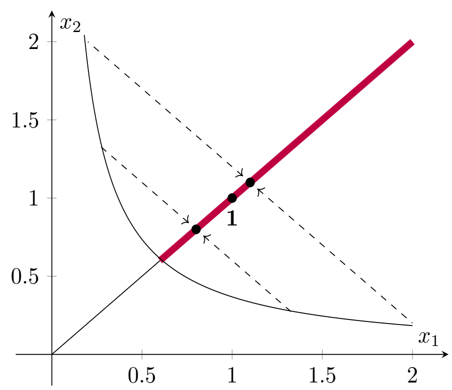


Figure 4: An illustration of Example 5.14: γ is supported on the hyperbola $x_1x_2 = e^{-1}$, the optimal μ is supported on the red ray. Dashed arrows indicate the reduction of filtration.

More generally, suppose that P_1, P_2, Q have densities $p_1, p_2, q \in C^2(\mathbb{R})$ where q is strictly positive, and denote by $W(f_1, \dots, f_n)$ the Wronskian of $f_1, \dots, f_n$. Then we have the further sufficient condition that

$$W\left((p_1/q)', (p_2/q)'\right) \neq 0 \text{ everywhere},$$

or equivalently, $W(p_1, p_2, q)(x) \neq 0$ for all $x \in \mathbb{R}$. By the Abel-Liouville identity (Teschl [2012]), this is the case if p_1, p_2, q form a fundamental system of solutions of the ODE

$$y^{(3)} = a_2(x)y^{(2)} + a_1(x)y^{(1)} + a_0(x)y$$

for some continuous functions $a_i : \mathbb{R} \to \mathbb{R},\ 0 \leqslant i \leqslant 2$.

In higher dimensions with more than two nulls, Theorem 5.9 is often not applicable. Nevertheless, given enough symmetry, we may directly compute $\{\mu_x\}$ from Theorem 5.5 and prove that they are monotone. Surprisingly, many intuitively straightforward tests are suboptimal.

Example 5.17. Consider probability measures $P_1 \sim \mathrm{N}((0,1), I)$, $P_2 \sim \mathrm{N}((-\sqrt{3}/2, -1/2), I)$, $P_3 \sim \mathrm{N}((\sqrt{3}/2, -1/2), I)$, and $Q \sim \mathrm{N}((0,0), I)$. Note that Theorem 5.9 is not directly applicable here. It is natural to guess from Example 5.14 that the optimal solution is the Euclidean norm, i.e., the distance from 0 in $\mathbb{R}^2$. On the contrary, we show this is not the case. A routine computation gives that

$$\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \frac{\mathrm{d}P_2}{\mathrm{d}Q}, \frac{\mathrm{d}P_3}{\mathrm{d}Q}\right)\Big|_Q = \left(e^{\xi_1 - 1/2}, e^{(-\xi_1 - \sqrt{3}\xi_2 - 1)/2}, e^{(-\xi_1 + \sqrt{3}\xi_2 - 1)/2}\right)\Big|_{\xi_1, \xi_2 \overset{\text{law}}{\sim} \mathrm{N}((0,0), I) \text{ independent}}.$$

Note that this forms an exchangeable random vector. The support is contained in $\{(x,y,z) \in \mathbb{R}^3_+ \mid xyz = e^{-3/2}\}$. By symmetry and exchangeability, the unique optimal solution shrinks the set $\{(x,y,z) \mid x+y+z = a, xyz = e^{-3/2}\}$ into a single point. In the coordinate (ξ_1, ξ_2), this is equivalent to

$$h(\xi_1, \xi_2) := e^{\xi_1} + e^{(-\xi_1 - \sqrt{3}\xi_2)/2} + e^{(-\xi_1 + \sqrt{3}\xi_2)/2} = \sqrt{e}\, a. \tag{10}$$

In other words, we reduce the filtration generated by the sequence of observations $Z_1, \dots, Z_n$ to the one generated by $h(Z_1), \dots, h(Z_n)$. It is clear that (10) does not agree with $\xi_1^2 + \xi_2^2 = a'$ for any $a' \in \mathbb{R}$, so taking the Euclidean distance from 0 instead of h is suboptimal. Using a general technique of constructing the most powerful e-variable in Section 6.2 below, one can show that the e-variable takes the form $X = 3(e^{Z^{(1)} - 1/2} + e^{(-Z^{(1)} - \sqrt{3}Z^{(2)} - 1)/2} + e^{(-Z^{(1)} + \sqrt{3}Z^{(2)} - 1)/2})^{-1} = 3\sqrt{e}h(Z)^{-1}$, where $Z = (Z^{(1)}, Z^{(2)})$ forms a single observation. This example also generalizes to more than three nulls (Gaussian with the same variance) whose means form a regular polygon centered at 0.

Finally, we supply the following example illustrating an explicit calculation of $\{\mu_x\}_{x\geqslant 0}$ with the presence of atoms in γ. This example also shows that without pivotality, the maximum value of (2) increases.

Example 5.18. Let $a, b, c, d \in (0,1)$ such that $\max(a+c, b+d) < 1$. On the probability space $\Omega = [0,3]$, we define atomic measures P_1, P_2, Q where

(i) P_1 has density $a\mathbb{1}_{[0,1]} + c\mathbb{1}_{[1,2]} + (1-a-c)\mathbb{1}_{[2,3]}$;

(ii) P_2 has density $b\mathbb{1}_{[0,1]} + d\mathbb{1}_{[1,2]} + (1-b-d)\mathbb{1}_{[2,3]}$;

(iii) Q is uniformly distributed.

It is clear that $P_1, P_2 \ll Q$ and (P_1, P_2, Q) is jointly atomless. The measure $\gamma = (\mathrm{d}P_1/\mathrm{d}Q, \mathrm{d}P_2/\mathrm{d}Q)|_Q$ has

$$\gamma = \frac{1}{3}\left(\delta_{(3a,3b)} + \delta_{(3c,3d)} + \delta_{(3(1-a-c),3(1-b-d))}\right).$$

Note that assumption (N) is not satisfied. In the following, we specify the choices of $a = 0.2,\ b = 0.3,\ c = 0.5,\ d = 0.6$. The triangle connecting the points $(0.6, 0.9), (1.5, 1.8), (0.9, 0.3)$ intersects with $\mathcal{I}^+$ at the points $(0.7, 0.7)$ and $(1.3, 1.3)$. Note that the measures

$$\frac{1}{3}\delta_{(3a,3b)} + \frac{1}{6}\delta_{(3(1-a-c),3(1-b-d))} \quad \text{and} \quad \frac{1}{3}\delta_{(3c,3d)} + \frac{1}{6}\delta_{(3(1-a-c),3(1-b-d))}$$

have barycenters equal to $(0.7, 0.7)$ and $(1.3, 1.3)$ respectively. In particular, we may pick

$$\mu_x = \begin{cases} 0 & \text{for } x < 0.7; \\ \frac{1}{3}\delta_{(3a,3b)} + \frac{1}{6}\delta_{(3(1-a-c),3(1-b-d))} & \text{for } 0.7 \leqslant x < 1.3; \\ \gamma & \text{otherwise.} \end{cases}$$

so that $\{\mu_x\}_{x\geqslant 0}$ is monotone and satisfies the four conditions in Proposition 5.3.

In view of Theorem 5.9 and Remark 5.11, the maximum is attained in (5) and hence in (2), by the choice $X(\omega) = 0.7\mathbb{1}_{\{\omega\in[0,1]\cup[2,2.5]\}} + 1.3\mathbb{1}_{\{\omega\in[1,2]\cup[2.5,3]\}}$. The optimal value is ≈ 0.047. On the other hand, if we remove the constraint that X is pivotal, then with $X \approx 1.63\delta_{[0,1]} + 1.15\delta_{[1,2]} + 0.66\delta_{[2,3]}$, we have $\mathbb{E}^Q[\log X] \approx 0.07$, showing that the maximum in (2) increases.

6 The SHINE construction

The current section develops the SHINE construction (Separating Hyperplanes Iteration for Nontrivial and Exact e/p-variables), that effectively produces a pivotal nontrivial exact e/p-variable via separating hyperplanes (see Proposition 5.3, which is the key to our construction). Unless otherwise stated, we follow the setup of Section 5 and assume (N).

The first goal of the SHINE construction is to solve the optimization problem (6). In the case where the condition in Theorem 5.5 is satisfied, the construction outputs the maximum element. When the maximum element μ does not exist or when the condition (b) in Theorem 5.5 is hard to check, we provide a reasonable *maximal* element μ in convex order. In the second part of the SHINE construction, we recover the corresponding e/p-variable from the output μ in the first part. The two parts are respectively illustrated in Sections 6.1 and 6.2. We end this section by providing examples and simulation results in Section 6.3.

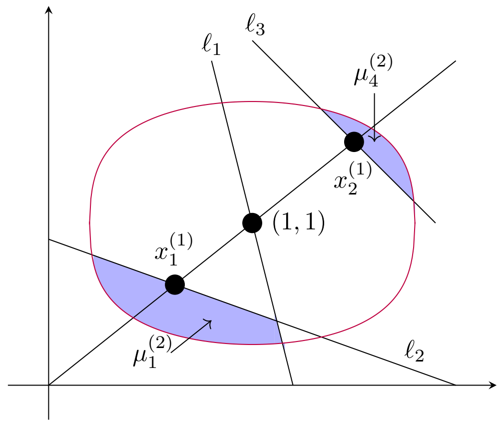


Figure 5: An illustration of the SHINE construction in dimension $L = 2$. The measure γ is supported on the region enclosed by the red contour, where $\mathrm{bary}(\gamma) = (1, 1)$. In the first step of the SHINE construction, we use Proposition 5.3 to find a line ℓ_1 through $(1, 1)$ that partitions γ into two parts $\mu_1^{(1)}$ and $\mu_2^{(1)}$, each of whose barycenters lies on the diagonal. In the second step, we find a line ℓ_2 through $x_1^{(1)} = \mathrm{bary}(\mu_1^{(1)})$ that partitions $\mu_1^{(1)}$ into two measures $\mu_1^{(2)}$ and $\mu_2^{(2)}$, each of whose barycenters lies on the diagonal, and similarly a line ℓ_3. We then proceed iteratively.

6.1 Description of the SHINE construction

Start with $\mu^{(0)} = \delta_{\mathbf{1}}$, $x_1^{(0)} = \mathbf{1}$, and $\mu_1^{(0)} = \gamma$. At step $n \geqslant 0$, we are given $\mu^{(n)}$, $\{x_k^{(n)}\}_{1 \leqslant k \leqslant 2^n}$, and $\{\mu_k^{(n)}\}_{1 \leqslant k \leqslant 2^n}$. For each k, we apply Proposition 5.3 to the sub-probability measure $\mu_k^{(n)}$ at the point $x_k^{(n)}$. This yields a unique decomposition of $\mu_k^{(n)}$ into two measures, each having barycenter on $\mathcal{I}^+$. Denote them by $\mu_{2k-1}^{(n+1)}$ and $\mu_{2k}^{(n+1)}$. For $1 \leqslant k \leqslant 2^{n+1}$, define $x_k^{(n+1)} = \mathrm{bary}(\mu_k^{(n+1)})$. Finally, let $\mu^{(n+1)}$ be the probability measure having mass $\mu_k^{(n+1)}(\mathbb{R}^L)$ on $x_k^{(n+1)}$ for every k, i.e.,

$$\mu^{(n+1)} := \sum_{k=1}^{2^{n+1}} \mu_k^{(n+1)}(\mathbb{R}^L)\delta_{x_k^{(n+1)}}. \tag{11}$$

The output of the SHINE construction at step n is the measure $\mu^{(n)}$.

It is easy to see that each $\mu^{(n)}$ is centered at $\mathbf{1}$, supported on $\mathcal{I}^+$, and satisfies $\mu^{(n)} \preceq_{\mathrm{cx}} \gamma$. Thus the sequence $\{\mu^{(n)}\}$ is tight and allows a weak limit. In fact, an even stronger assertion can be made. Define $\{X_n\}$ the coupling of the first coordinate of $\{\mu^{(n)}\}$ such that $X_0 = 1$ and at each $n \geqslant 0$, for $j = 2k-1, 2k$,

$$\mathbb{P}\left[X_{n+1} = x_j^{(n+1)} \mid X_n = x_k^{(n)}\right] = \frac{\mu_j^{(n+1)}(\mathbb{R}^L)}{\mu_{2k-1}^{(n+1)}(\mathbb{R}^L) + \mu_{2k}^{(n+1)}(\mathbb{R}^L)}.$$

It holds that $\{X_n\}$ forms a nonnegative martingale, and hence converges a.s. to some X_∞ by the martingale convergence theorem. Denote by μ the law of $X_\infty \mathbf{1} = (X_\infty, \dots, X_\infty)$. Then $\mu \preceq_{\mathrm{cx}} \gamma$ by Lemma 2.1(ii).

Remark 6.1. The first step of the construction, i.e., after finishing step $n = 0$, already contains a proof of Proposition 4.2, because $\delta_{\mathbf{1}} \neq \mu^{(1)} \preceq_{\mathrm{cx}} \gamma$. Nevertheless, the ideas behind the original proof of Proposition 4.2 extend to the composite alternative scenario.

Example 6.2. Suppose that $L = 1$, i.e., we have simple null versus simple alternative. In this case, Proposition 5.3 applies trivially: for each $n \geqslant 0$ and $1 \leqslant k \leqslant 2^n$, the measure $\mu_k^{(n)}$ is decomposed

into
$$\mu_k^{(n)} = \mu_{2k-1}^{(n+1)} + \mu_{2k}^{(n+1)} := \mu_k^{(n)}\big|_{[0,\mathrm{bary}(\mu_k^{(n)}))} + \mu_k^{(n)}\big|_{[\mathrm{bary}(\mu_k^{(n)}),\infty)}.$$

As in (11), this results in a sequence of laws $\{\mu^{(n)}\}_{n\geqslant 0}$ on $\mathbb{R}$ that are increasing in convex order and dominated by γ. This is closely related to a martingale decomposition theorem by Simons [1970]: if we denote by $\{Z_n\}_{n\geqslant 0}$ the natural martingale coupling of $\{\mu^{(n)}\}_{n\geqslant 0}$, then $Z_n \to Z$ a.s. for some Z that has law γ.

Theorem 6.3. *Assume* (N)*. The above construction always produces in the limit a maximal element μ in convex order in $\mathcal{M}_\gamma$.*

When we apply the construction we need to stop at finitely many steps, so we will not always obtain a maximal element. However, Theorem 6.3 shows that our output of the construction is quite close to being maximal. We start the proof with a few simple observations.

Lemma 6.4. *Suppose that ρ is a finite measure on $\mathbb{R}$, and I is a nonempty bounded open interval. Assume that there exists a sequence of decreasing intervals $I_n \downarrow I$, such that* $\mathrm{bary}(\rho|_{I_n}) \notin I$ *for every n where the barycenter is well-defined. Then $I \subseteq (\operatorname{supp}\rho)^c$.*

Proof. This is a direct consequence of continuity of the measure. We omit the details. □

Lemma 6.5. *Assume* (N)*. Any maximal element ν in $\mathcal{M}_\gamma$ is atomless.*

Proof. Suppose that ν has an atom at $x_0\mathbf{1}$. Then a martingale coupling of ν and γ transports the mass at $x_0\mathbf{1}$ to some measure γ' on $\mathbb{R}^L$. In particular, $\mathrm{bary}(\gamma') = x_0\mathbf{1}$ and $\gamma' \leqslant \gamma$. By assumption (N), $\operatorname{supp}\gamma'$ is not contained in any hyperplane. Using a similar argument in the proof of Proposition 4.2, we conclude that there exists a measure $\nu' \preceq_{\mathrm{cx}} \gamma'$ supported on $\mathcal{I}^+$ satisfying $\nu' \neq \gamma'(\mathbb{R}^L)\delta_{x_0}$. The new measure $\nu - \gamma'(\mathbb{R}^L)\delta_{x_0} + \nu'$ then dominates ν in convex order, contradicting the maximality of ν. □

Proof of Theorem 6.3. Suppose that μ is the measure from the construction, $\mu \preceq_{\mathrm{cx}} \nu$, and $\nu \preceq_{\mathrm{cx}} \gamma$ with ν supported on $\mathcal{I}^+$. Our goal is to show $\mu = \nu$. Let ξ_μ, ξ_ν denote the first coordinate of μ, ν. Consider the collection $\mathcal{X}$ of the first coordinates of all points $x_k^{(n)}$, $n \geqslant 0$, $1 \leqslant k \leqslant 2^{n+1}$ defined in the middle of the construction. We first show that for each $x \in \mathcal{X}$,
$$\mathbb{P}[\xi_\mu \geqslant x] = \mathbb{P}[\xi_\nu \geqslant x] \quad \text{and} \quad \mathbb{E}[\xi_\mu \mid \xi_\mu \geqslant x] = \mathbb{E}[\xi_\nu \mid \xi_\nu \geqslant x]. \tag{12}$$

Note that given the first equality, the second equality in (12) is equivalent to $\mathbb{E}[(\xi_\mu - x)_+] = \mathbb{E}[(\xi_\nu - x)_+]$. The proof is similar to the "⇒" direction of Theorem 5.5. Let π_μ be any martingale coupling of (μ, γ) and π_ν be any martingale coupling of (ν, γ). For $x = x_1^{(0)} = \mathbf{1}$, by (a symmetric version of) Lemma 5.8, ξ_μ attains the maximum value of $\mathbb{E}[(\xi_\mu - x)_+]$, and thus $\mathbb{E}[(\xi_\mu - x)_+] = \mathbb{E}[(\xi_\nu - x)_+]$. Lemma 5.8 further implies that[6]
$$\pi_\mu([0,x) \times \mathbb{H}_x) = \pi_\nu([0,x) \times \mathbb{H}_x) = 1 - \pi_\mu([x,\infty) \times \mathbb{H}_x^c) = 1 - \pi_\nu([x,\infty) \times \mathbb{H}_x^c).$$

In particular, $\mathbb{P}[\xi_\mu \geqslant x] = \mathbb{P}[\xi_\nu \geqslant x]$, proving (12). In the general case, consider $x = x_k^{(n)}$. There exists an interval J whose endpoints are the two neighbor points of $x_k^{(n)}$ in $\{x_k^{(m)}\}_{m<n,\,1\leqslant k\leqslant 2^{m+1}} \cup \{0,\infty\}$. By definition, ν maximizes $\mathbb{E}[(\xi_\mu - x)_+]$, and μ maximizes $\mathbb{E}[(\xi_\mu - x)_+ \mathbb{1}_{\{\xi_\mu \in J\}}]$. Our induction hypothesis (12) applied to the right endpoint of J meanwhile implies that the two optimization problems are the same. Thus, there exists a similar block decomposition of the supports of π_μ, π_ν where the total masses coincide on the blocks, and (12) holds for $x = x_k^{(n)}$. We leave the details to the reader.

[6]By our assumption on γ, μ cannot have an atom at x.

We now finish the proof given (12). We claim that the set $\mathcal{X}$ is dense in $\operatorname{supp}\nu$. Indeed, suppose that I is an open connected component of the open set $\mathbb{R}\setminus\overline{\mathcal{X}}$. By construction and (12), each $x_k^{(n)}$ is the barycenter of ν restricted to the interval formed by two neighbor points of $x_k^{(n)}$ in $\{x_k^{(m)}\}_{m<n,\,1\leqslant k\leqslant 2^{m+1}}$. In particular, there exist intervals $I_n\downarrow I$ where the endpoints of each I_n belong to $\mathcal{X}$ and $\operatorname{bary}(\nu|_{I_n})\notin I$. By Lemma 6.4, $I\subseteq(\operatorname{supp}\nu)^c$, establishing the claim.

Therefore, the distribution functions of μ and ν coincide on a dense subset $\mathcal{X}$ of the support of the atomless measure ν. This implies $\mu=\nu$. □

We note in particular that Lemma 6.5 together with Theorem 6.3 yield that the construction always gives an atomless measure μ in the limit.

When the maximum element exists, a maximal element must be maximum. This has the following consequence.

Corollary 6.6. *Suppose that the condition in Theorem 5.5 is satisfied, i.e., there exists a maximum element μ_0. Then our construction produces the same maximum element μ_0.*

With the presence of atoms, the decomposition given by Proposition 5.3 is not necessarily unique when applied to our construction. The degree of freedom of each μ_x is the measure on the hyperplane $\partial\mathbb{H}_x$. To describe a well-defined construction, we need to specify $\mu_x|_{\partial\mathbb{H}_x}$ uniquely for each x. Analyzing the maximality of the output remains a technical task, which we do not discuss in this paper.

6.2 Recovering explicitly an e/p-variable

We aim first to recover our e-variable X, which we recall from Proposition 5.2 is of the form $X=(\mathrm{d}G/\mathrm{d}F)(Y)$, where $Y\in\mathcal{T}((P_1,\dots,P_L,Q),(F,\dots,F,G))$ and F,G come from our SHINE construction. We have seen that at the n-th step, our construction leads to a canonical martingale coupling of $\mu^{(n)}$ and γ that couples the mass $\mu_k^{(n+1)}(\mathbb{R}^L)\delta_{x_k^{(n+1)}}$ with $\mu_k^{(n+1)}$. We denote the martingale coupling by (Λ_n,Γ_n), which is a random vector of dimension $2L$. Under assumption (N), we know further that the measures $\{\mu_k^{(n)}\}_{1\leqslant k\leqslant 2^n}$ are mutually singular, and hence (Λ_n,Γ_n) is backward Monge, i.e., in the backward direction we have $\Lambda_n=h(\Gamma_n)$ for some h. Since $(P_1,\dots,P_L,Q)$ is jointly atomless, we may apply Proposition 2.4 to find a simultaneous transport map $Y\in\mathcal{T}((P_1,\dots,P_L,Q),(F,\dots,F,G))$ such that for each $x\in\mathfrak{X}$,

$$\frac{\mathrm{d}F}{\mathrm{d}G}(Y(x))\times\mathbf{1}=h\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(x),\dots,\frac{\mathrm{d}P_L}{\mathrm{d}Q}(x)\right).$$

This leads to

$$(X(x))^{-1}\times\mathbf{1}=h\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(x),\dots,\frac{\mathrm{d}P_L}{\mathrm{d}Q}(x)\right),\ x\in\mathfrak{X}.$$

For example, the n-th step of the construction gives explicitly

$$(X(x))^{-1}\times\mathbf{1}=h(x_k^{(n+1)})\quad\text{if}\quad\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(x),\dots,\frac{\mathrm{d}P_L}{\mathrm{d}Q}(x)\right)\in\operatorname{supp}\mu_k^{(n+1)},\ x\in\mathfrak{X}.\tag{13}$$

Note that the measures F,G can meanwhile be reconstructed from Lemma 2.5, and further Lemma A.4 (i) if one requires $F=\mathrm{U}_1$. In this case Y is the valid p-variable as desired, which can be effectively described by the MOT-SOT parity of Wang and Zhang [2023].

Example 6.7. Suppose that we are in the setting of Example 5.14, with $P_1\sim\mathrm{N}(-1,1)$, $P_2\sim\mathrm{N}(1,1)$, and $Q\sim\mathrm{N}(0,1)$. Recall that

$$\gamma=\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q},\frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)\Big|_Q=\left(e^{-Z-1/2},e^{Z-1/2}\right)\Big|_{Z\overset{\text{law}}{\sim}\mathrm{N}(0,1)}.$$

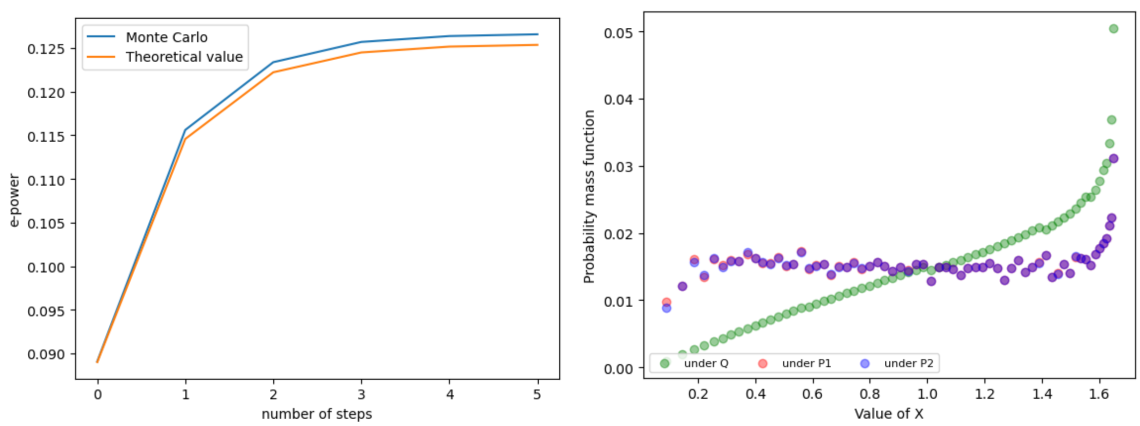
By symmetry of γ, it is clear that the separating hyperplanes $\mathbb{H}_x$ in the SHINE construction are given by $\mathbb{H}_x = \{(a,b) : a+b \leqslant 2x\}$. In the first step of the construction, we locate the barycenters of the measures $\gamma|_{\mathbb{H}_1}$ and $\gamma|_{\mathbb{H}_1^c}$. By direct calculation, we obtain $\mathrm{bary}(\gamma|_{\mathbb{H}_1}) \approx 0.713 \times \mathbf{1}$ and $\mathrm{bary}(\gamma|_{\mathbb{H}_1^c}) \approx 1.743 \times \mathbf{1}$. Using (13), the corresponding e-variable has the form

$$X(x) = \begin{cases} 1.403 & \text{if } |x| \geqslant \log(\sqrt{e} + \sqrt{e-1}); \\ 0.574 & \text{if } |x| < \log(\sqrt{e} + \sqrt{e-1}). \end{cases}$$

The resulting e-power $\mathbb{E}[\log X]$ is approximately 0.089.[7] In general, we may construct X in multiple steps.

6.3 Simulation results

We first consider the setting of Example 5.14, where we recall that $P_1 \sim \mathrm{N}(-1,1)$, $P_2 \sim \mathrm{N}(1,1)$, and $Q \sim \mathrm{N}(0,1)$. In Figure 6, we provide two figures illustrating the e-power at each step in the SHINE construction and the corresponding laws of the e-variable under P_1, P_2, and Q. In the left panel, we compute the e-power in two ways: from the analytic expression (9) and by Monte Carlo simulations. The e-powers are reasonably close with only 2×10^4 samples and converge quickly to their limits, where it is straightforward to compute from (9) that the theoretical maximum e-power is approximately 0.12543. In the right panel, we show the distributions of the e-variable under P_1, P_2, and Q at step $n = 5$ of the SHINE construction, again by simulating 2×10^4 samples of each distribution. The pivotality of the e-variable implies that the laws of X under P_1 and P_2 are the same, while the marginal errors shown by the figure are due to our Monte Carlo simulation.[8] Note that within finitely many steps, the SHINE construction always yields a discrete e-variable.



(a) growth of the e-power

(b) distributions of the e-variable X

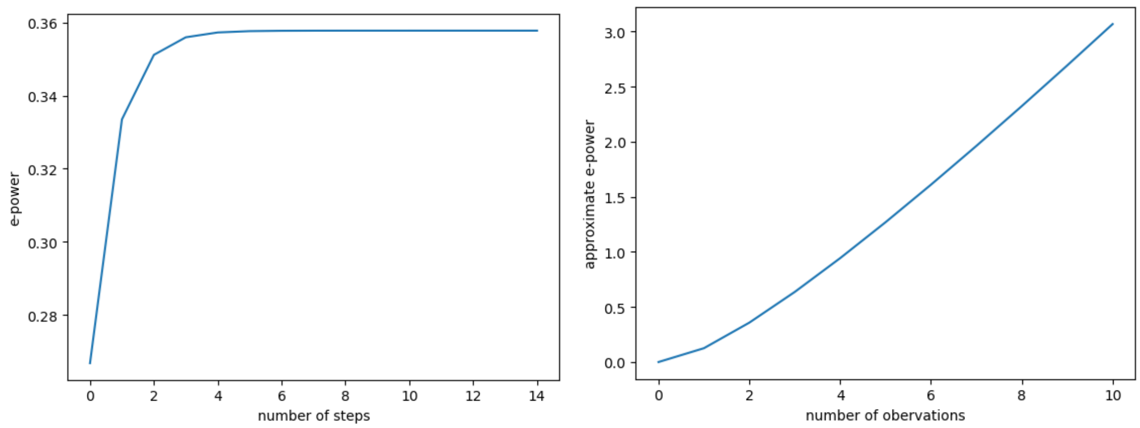
Figure 6: The SHINE construction for Example 5.14

In Figure 7, we complement the discussions in Example 5.15 regarding multiple data points. Recall that $P_1 = \mathrm{N}(-\mathbf{1}_n, I_n)$, $P_2 = \mathrm{N}(\mathbf{1}_n, I_n)$, and $Q = \mathrm{N}(\mathbf{0}_n, I_n)$, where $n \in \mathbb{N}$. Panel (a) computes the theoretical e-power developed after a number of steps with two data points ($n = 2$), which is approximately 0.35775, significantly higher than 0.25086, which is twice the e-power with a single data point. Panel (b) plots the theoretical e-power at the seventh step of the SHINE construction, for various numbers of observations. Observe that the curve is convex and tends to linear, reflecting

[7] One may compare this to the maximum e-power 0.12543, which can be directly computed from (9).

[8] With Monte Carlo, our e-variable is *approximately* pivotal since the measure γ is atomic, which violates Assumption (N).

the fact that taking multiple data points increases the average e-power, while the normalized e-power converges as shown in Proposition 5.13.



(a) growth of e-power with two data points

(b) e-power versus number of observations

Figure 7: Maximum e-power with multiple data points for Example 5.15.

The implementation of the SHINE construction in dimensions greater than two has the obstacle that it is difficult in general to find the separating hyperplanes. We leave this to future work, as well as generalizations of the SHINE construction when (N) does not hold.

7 Composite null and composite alternative

Our goal in this section is to extend Theorems 4.3 and 4.4 to composite alternative, i.e., when $|\mathcal{P}|, |\mathcal{Q}| > 1$. A full characterization of the existence of (exact and pivotal) nontrivial p/e-variables is provided in the case where both $\mathcal{P}$ and $\mathcal{Q}$ are finite. We also discuss the general case where $\mathcal{P}, \mathcal{Q}$ are infinite, including a few open problems.

7.1 Existence of an exact and pivotal p/e-variable for the finite case

We start with the case where $\mathcal{P}, \mathcal{Q}$ are both finite. That is, given $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q_1, \dots, Q_M\}$ such that $(P_1, \dots, P_L, Q_1, \dots, Q_M)$ is jointly atomless (with the exception of Remark 7.5), we characterize equivalent conditions for the existence of an (exact and) nontrivial e-variable (or p-variable). We build upon the ideas from Proposition 4.2.

Lemma 7.1. *Let $V \subseteq \mathbb{R}^d$ be a subspace containing $\mathbf{1} \in \mathbb{R}^d$ and $\mathscr{S}$ be a collection of affine hyperplanes in $\mathbb{R}^d$ containing $\mathbf{1}$ such that whenever $S \in \mathscr{S}$ and an affine subspace T satisfies $T \cap V \subseteq S$, it holds $T \subseteq S'$ for some $S' \in \mathscr{S}$. Then for each measure μ centered at $\mathbf{1}$ whose support is not a subset of S for any $S \in \mathscr{S}$, there exists a measure ν on V such that $\nu \preceq_{\mathrm{cx}} \mu$ and the support of ν is not a subset of S for any $S \in \mathscr{S}$.*

Proof. The proof is similar to the proof of Proposition 4.2. Define $T = \operatorname{aff} \operatorname{supp} \mu$. By Lemma A.2, it suffices to find points $s_1, \dots, s_k \in \operatorname{supp} \mu$ such that $\operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_k\}; T)$ contains $\mathbf{1}$ and intersects with V not on a single $S \in \mathscr{S}$.

Suppose that the contrary holds. That is, any $s_1, \dots, s_k \in \operatorname{supp} \mu$ satisfies $\mathbf{1} \notin \operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_k\}; T)$ or $\operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_k\}; T) \cap V \subseteq S$ for some $S \in \mathscr{S}$. By Lemma A.1(i), any $\operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_k\}; T)$ is contained in $\operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_K\}; T)$ for some $k \leqslant K$ and $s_1, \dots, s_K \in \operatorname{supp} \mu$ such that $\mathbf{1} \in \operatorname{ri}(\operatorname{Conv}\{s_1, \dots, s_K\}; T)$. This implies for all $s_1, \dots, s_k \in \operatorname{supp} \mu$

that $\mathrm{ri}(\mathrm{Conv}\{s_1,\dots,s_k\};T)\cap V\subseteq S$ for some $S\in\mathscr{S}$. Consequently, there exists $S\in\mathscr{S}$ such that $\mathrm{ri}(\mathrm{Conv}\,\mathrm{supp}\,\mu;T)\cap V\subseteq S$. By Lemma A.1(ii), $T\subseteq\mathrm{aff}\,\mathrm{ri}(\mathrm{Conv}\,\mathrm{supp}\,\mu;T)$. Since V,S are affine spaces, it holds that $T\cap V\subseteq S$. Moreover, $\mathrm{supp}\,\mu\subseteq T\subseteq S'$ for some $S'\in\mathscr{S}$ by our assumption. Hence, the support of μ is contained in S', contradicting our assumption. □

Proposition 7.2. *Let $L,M\in\mathbb{N}$ and $(P_1,\dots,P_L,Q_1,\dots,Q_M)$ be a jointly atomless tuple of probability measures on $\mathfrak{X}$ such that $\mathrm{Span}(P_1,\dots,P_L)\cap\mathrm{Conv}(Q_1,\dots,Q_M)=\emptyset$. Then there exist probability measures $F,G_1,\dots,G_M$ on $\mathbb{R}$ such that $F\notin\mathrm{Conv}(G_1,\dots,G_M)$ and*

$$\mathcal{T}((P_1,\dots,P_L,Q_1,\dots,Q_M),(F,\dots,F,G_1,\dots,G_M))\neq\emptyset.$$

Proof. Let μ be a dominating measure for $(P_1,\dots,P_L,Q_1,\dots,Q_M)$, say $\mu=(P_1+\dots+P_L+Q_1+\dots+Q_M)/(L+M)$, and $\nu=\mathrm{U}_1$. Then $\mathrm{Span}(P_1,\dots,P_L)\cap\mathrm{Conv}(Q_1,\dots,Q_M)\neq\emptyset$ is equivalent to the existence of $\{\alpha_i\}_{1\leqslant i\leqslant L}$ and $\{\beta_j\}_{1\leqslant j\leqslant M}$ such that

$$\sum_{i=1}^L\alpha_i\frac{\mathrm{d}P_i}{\mathrm{d}\mu}=\sum_{j=1}^M\beta_j\frac{\mathrm{d}Q_j}{\mathrm{d}\mu},\ \beta_j\geqslant 0,\ \sum_{i=1}^L\alpha_i=\sum_{j=1}^M\beta_j=1.$$

Similarly, $F\in\mathrm{Conv}(G_1,\dots,G_M)$ is equivalent to the existence of $\{\lambda_j\}_{1\leqslant j\leqslant M}$ such that

$$\frac{\mathrm{d}F}{\mathrm{d}\nu}=\sum_{j=1}^M\lambda_j\frac{\mathrm{d}G_j}{\mathrm{d}\nu},\ \lambda_j\geqslant 0,\ \sum_{j=1}^M\lambda_j=1.$$

To this end, we define

$$\mathscr{S}=\left\{S_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mid\beta_j\geqslant 0,\ \sum_{i=1}^L\alpha_i=\sum_{j=1}^M\beta_j=1\right\},$$

where

$$S_{\boldsymbol{\alpha},\boldsymbol{\beta}}:=\left\{(x_1,\dots,x_p,y_1,\dots,y_q)\mid\sum_{i=1}^L\alpha_ix_i=\sum_{j=1}^M\beta_jy_j\right\}.$$

We now claim that for each measure γ such that $\mathrm{supp}\,\gamma$ is not a subset of some $S\in\mathscr{S}$, there exists $\tau\preceq_{\mathrm{cx}}\gamma$ such that τ is supported on $V:=\{(\mathbf{x},\mathbf{y})\in\mathbb{R}^{L+M}\mid x_1=\dots=x_L\}$ but not concentrated on a single $V_{\boldsymbol{\lambda}}:=\{(\mathbf{x},\mathbf{y})\in\mathbb{R}^{L+M}\mid x_1=\dots=x_L=\sum_{j=1}^M\lambda_jy_j\}$ for all $\lambda_j\geqslant 0,\sum_{j=1}^M\lambda_j=1$. Provided the claim is true, we construct using Lemma 2.5 the measures $F,G_1,\dots,G_M$ such that

$$\left(\frac{\mathrm{d}F}{\mathrm{d}\nu},\dots,\frac{\mathrm{d}F}{\mathrm{d}\nu},\frac{\mathrm{d}G_1}{\mathrm{d}\nu},\dots,\frac{\mathrm{d}G_M}{\mathrm{d}\nu}\right)\Big|_\nu=\tau.$$

Since $\nu\preceq_{\mathrm{cx}}\gamma$ and γ is supported on the hyperplane $\{(\mathbf{x},\mathbf{y})\mid\sum_{i=1}^Lx_i+\sum_{j=1}^My_j=L+M\}$, our measure ν will be supported on the same hyperplane, thus $\nu=(LF+\sum_{j=1}^MG_j)/(L+M)$. In other words, μ,ν allow the same linear combination of the measure tuples $(P_1,\dots,P_L,Q_1,\dots,Q_M)$ and $(F,\dots,F,G_1,\dots,G_M)$. By Proposition 2.3,

$$\mathcal{T}((P_1,\dots,P_L,Q_1,\dots,Q_M),(F,\dots,F,G_1,\dots,G_M))\neq\emptyset$$

holds as desired.

To prove the above claim, we apply Lemma 7.1 with $d=L+M$, $\mu=\gamma$, $\nu=\tau$, and $V,\mathscr{S}$ defined as above. Note that if the support of τ is contained in V but not in a certain S, then it cannot be contained in a certain $V_{\boldsymbol{\lambda}}$. Thus the conclusion of Lemma 7.1 suffices for our purpose.

It then suffices to check the condition in Lemma 7.1 that whenever $S \in \mathscr{S}$ and an affine subspace T satisfies $T \cap V \subseteq S$, it holds $T \subseteq S'$ for some $S' \in \mathscr{S}$. To this end, we consider $S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathscr{S}$ and first assume T is a hyperplane containing $S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V = \{(\mathbf{x},\mathbf{y}) \mid x_1 = \cdots = x_L = \sum_{j=1}^M \beta_j y_j\}$. In this case, a normal vector to T (which is unique up to a multiplicative constant) must also be a normal vector of $S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V$, and hence must be of the form $(t_1,\dots,t_L,-\beta_1\sum_{i=1}^L t_i,\dots,-\beta_M\sum_{i=1}^L t_i)$ for some $t_1,\dots,t_L \in \mathbb{R}$. If $\sum_{i=1}^L t_i = 0$, then $T \supseteq V$, and thus $V \subseteq S_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, which is impossible. Thus $\sum_{i=1}^L t_i \neq 0$. It follows that for some $t_1,\dots,t_L$ with $\sum_{i=1}^L t_i \neq 0$,

$$T = \left\{ (x_1,\dots,x_L,y_1,\dots,y_M) \mid \sum_{i=1}^L t_i x_i = \sum_{i=1}^L t_i \sum_{j=1}^M \beta_j y_j \right\}.$$

Therefore, $T \in \mathscr{S}$. More precisely, $T = S_{\mathbf{t}/\sum_{i=1}^L t_i,\boldsymbol{\beta}}$.

Next, we prove the general case of an affine subspace T that satisfies $T \cap V \subseteq S_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. Note that $V \nsubseteq S_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ for each $S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathscr{S}$, and that V is of dimension $M+1$. Thus $S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V$ is of dimension M. Let $T' = T + (S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V) + V^{\perp}$. Then T' is a hyperplane, $T \subseteq T'$, and

$$T' \cap V \subseteq (T + (S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V)) \cap V = (T \cap V) + (S_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cap V) \subseteq S_{\boldsymbol{\alpha},\boldsymbol{\beta}}.$$

This completes the proof. ☐

Theorem 7.3. *Suppose that we are testing $\mathcal{P} = \{P_1,\dots,P_L\}$ against $\mathcal{Q} = \{Q_1,\dots,Q_M\}$. The following are equivalent:*

(a) there exists an exact (hence pivotal) and nontrivial p-variable;

(b) there exists a pivotal, exact, and bounded e-variable that has nontrivial e-power against $\mathcal{Q}$;

(c) there exists an exact e-variable that is nontrivial for $\mathcal{Q}$;

(d) there exists a random variable X that is pivotal for $\mathcal{P}$ and satisfies $F \notin \mathrm{Conv}(G_1,\dots,G_M)$, where F is the law of X under $P \in \mathcal{P}$ and G_j is the law of X under Q_j for $1 \leqslant j \leqslant M$;

(e) it holds that $\mathrm{Span}(P_1,\dots,P_L) \cap \mathrm{Conv}(Q_1,\dots,Q_M) = \emptyset$.

Proof. The direction $(a) \Rightarrow (b)$ is Proposition 3.5, $(b) \Rightarrow (c)$ is clear, $(c) \Rightarrow (e)$ being precisely Proposition 3.6, and $(e) \Rightarrow (d)$ is Proposition 7.2. To show $(d) \Rightarrow (a)$, we let ϕ be a nontrivial p-variable with null $\{G_1,\dots,G_M\}$ and alternative $\{F\}$, whose existence is guaranteed by Theorem 4.4. Then by definition, $\phi \circ X$ has a common law that is $\prec_{\mathrm{st}} \mathrm{U}_1$ under each P_i, and has law that is $\succeq_{\mathrm{st}} \mathrm{U}_1$ under each Q_j. Applying Lemma A.4(i) then yields a random variable Ψ such that $\Psi \circ \phi \circ X$ is an exact and nontrivial p-variable as desired. ☐

Theorem 7.4. *Suppose that we are testing $\mathcal{P} = \{P_1,\dots,P_L\}$ against $\mathcal{Q} = \{Q_1,\dots,Q_M\}$. The following are equivalent:*

(a) there exists a nontrivial p-variable;

(b) there exists an e-variable that is nontrivial for $\mathcal{Q}$;

(c) it holds that $\mathrm{Conv}(P_1,\dots,P_L) \cap \mathrm{Conv}(Q_1,\dots,Q_M) = \emptyset$.

Proof. The proofs are similar to Theorem 4.4, where in the direction $(c) \Rightarrow (a)$ we replace the linear cone $(-\infty,0)^L \cup (0,\infty)^L \cup \{\mathbf{0}\}$ by the linear cone $(-\infty,0)^L \times (0,\infty)^M \cup (0,\infty)^L \times (-\infty,0)^M \cup \{\mathbf{0}\}$. ☐

Remark 7.5. By Proposition 3.7, the directions $(c) \Leftrightarrow (e)$ in Theorem 7.3 and $(b) \Leftrightarrow (c)$ in Theorem 7.4 also hold without the condition that $(P_1,\dots,P_L,Q_1,\dots,Q_M)$ is jointly atomless.

Remark 7.6. The equivalence of (b) and (c) in Theorem 7.4 is a special case of Kraft's theorem [Kraft, 1955], which states that for each $\varepsilon > 0$, there exists an X with

$$\inf_{Q\in\mathcal{Q}} \mathbb{E}^Q[X] \geqslant \varepsilon + \sup_{P\in\mathcal{P}} \mathbb{E}^P[X] \tag{14}$$

if and only if the total variation distance $d_{\mathrm{TV}}(\mathrm{Conv}\mathcal{P}, \mathrm{Conv}\mathcal{Q}) \geqslant \varepsilon$.[9] Kraft's theorem serves as a starting point for impossible inference; see Bertanha and Moreira [2020]. To see that this implies the equivalence of (b) and (c) in Theorem 7.4, suppose that (b) holds. It follows that $\mathbb{E}^Q[X] > 1 \geqslant \mathbb{E}^P[X]$ for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$. Kraft's theorem implies the existence of some $\varepsilon > 0$ such that $d_{\mathrm{TV}}(\mathrm{Conv}\mathcal{P}, \mathrm{Conv}\mathcal{Q}) \geqslant \varepsilon$, and in particular, (c) holds. On the other hand, if (c) is true, then Kraft's theorem yields $\varepsilon > 0$ and X satisfying (14). A suitable linear transformation Y of X then satisfies $\mathbb{E}^Q[Y] > 1 \geqslant \mathbb{E}^P[Y]$ for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$, and the rest follows from Proposition 3.2.

Corollary 7.7. *Suppose that we are testing $\mathcal{P}$ against $\mathcal{Q}$, where $\mathcal{P}$ and $\mathcal{Q}$ are convex polytopes in Π. Denote by $\{P_1, \ldots, P_L\}$ (resp. $\{Q_1, \ldots, Q_M\}$) the vertices of the polytope $\mathcal{P}$ (resp. $\mathcal{Q}$) and assume that $(P_1, \ldots, P_L, Q_1, \ldots, Q_M)$ is jointly atomless. Precisely the same conclusions in Theorems 7.3 and 7.4 hold.*

Proof. This follows immediately from Proposition 3.1. □

Example 7.8. Fix $0 < q_1 < q_2 < 1$ and let $\mathcal{P} = \{\mathrm{Ber}(q_1)\}$ and $\mathcal{Q} = \{\mathrm{Ber}(p) \mid q_2 \leqslant p \leqslant 1\}$. Corollary 7.7 then provides an exact nontrivial e-variable (or p-variable). Nevertheless, such an exact nontrivial e-variable (or p-variable) would not exist if we replace $\mathcal{P}$ by $\{\mathrm{Ber}(p) \mid 0 \leqslant p \leqslant q_1\}$.

Due to the complication of convex order in higher dimensions, it remains a challenging task how to generalize Theorem 5.9 and the SHINE construction to the composite alternative case.

7.2 Infinite null and alternative

We first state a weaker version of Theorem 7.3 when both $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta_0}$ and $\mathcal{Q} = \{Q_\theta\}_{\theta\in\Theta_1}$ may be infinite but allow a common reference measure.

Proposition 7.9. *Assume that there exists a common reference measure $R \in \Pi(\mathfrak{X})$ such that $P_\theta \ll R$ for $\theta \in \Theta_0$ and $Q_\theta \ll R$ for $\theta \in \Theta_1$. Then there exists an exact bounded e-variable X for $\mathcal{P}$ against $\mathcal{Q}$ satisfying $\inf_{Q\in\mathcal{Q}} \mathbb{E}^Q[\log X] > 0$ if and only if $0 \notin \overline{\overline{\mathrm{Span}}\mathcal{P} + \overline{\mathrm{Conv}}\mathcal{Q}}$ where the closure is taken wrt the total variation distance. If $\mathcal{Q}$ is tight, then we have the further equivalence to $\overline{\mathrm{Span}}\mathcal{P} \cap \overline{\mathrm{Conv}}\mathcal{Q} = \emptyset$.*

Note that we have put a stronger assumption on the e-variable X ($\inf_{Q\in\mathcal{Q}} \mathbb{E}^Q[\log X] > 0$) than having nontrivial e-power against $\mathcal{Q}$ (for all $Q \in \mathcal{Q}$, $\mathbb{E}^Q[\log X] > 0$). Proposition 7.9 can thus be seen as a sufficient condition for the existence of an exact e-variable that has nontrivial e-power against $\mathcal{Q}$. Dealing with pivotal p-variables is beyond the scope of this paper and left open.

Proof of Proposition 7.9. We first show the "only if" direction. Suppose that $0 \in \overline{\overline{\mathrm{Span}}\mathcal{P} + \overline{\mathrm{Conv}}\mathcal{Q}}$ and X is an exact and bounded e-variable satisfying $\inf_{Q\in\mathcal{Q}} \mathbb{E}^Q[\log X] > 0$. In particular, since X is bounded, for a sequence of distributions converging in total variation, the expectations of X also converge. Let $P^{(n)} \in \overline{\mathrm{Span}}\mathcal{P}$ and $Q^{(n)} \in \overline{\mathrm{Conv}}\mathcal{Q}$ be such that $P^{(n)} + Q^{(n)} \to 0$ in total variation. It follows that $\mathbb{E}^{P^{(n)}}[X] = 1$ and $\mathbb{E}^{Q^{(n)}}[X] \geqslant \inf_{\theta\in\Theta_1} \mathbb{E}^{Q_\theta}[X] > 1$. But then $\liminf \mathbb{E}^{P^{(n)}+Q^{(n)}}[X] \geqslant \inf_{\theta\in\Theta_1} \mathbb{E}^{Q_\theta}[X] - 1 > 0$, a contradiction.

Next, we show the "if" direction. Let $R \in \Pi(\mathfrak{X})$ be as given. We may abuse notation and identify each P_θ and Q_θ with its density wrt R. Clearly, convergence in total variation is equivalent to convergence in $L^1(R)$. Thus, $S := \overline{\mathrm{Span}}\{P_\theta : \theta \in \Theta_0\}$ is a closed subspace of $L^1(R)$. By assumption,

[9] Here, we do not require that $\mathcal{P}$ and $\mathcal{Q}$ are finite.

the set $C := \overline{\mathrm{Conv}}\{Q_\theta : \theta \in \Theta_1\}$ satisfies that $\overline{C+S}$ is closed, convex, and disjoint from 0 in the quotient space $L^1(R)/S$. By the Hahn-Banach separation theorem, there is $\overline{h} : L^1(R)/S \to \mathbb{R}$ such that $\overline{h}|_{\overline{C+S}} > \varepsilon > 0$. Composing with the quotient map we obtain a linear functional $h : L^1(R) \to \mathbb{R}$, and it is easy to check that h vanishes on S and $h|_C > \varepsilon$. By duality, we may recognize $h \in L^\infty(R)$. It follows that the bounded random variable $X = h+1$ satisfies $\mathbb{E}^P[X] = 1+\mathbb{E}^P[h] = 1+\int hP\mathrm{d}R = 1$ for each $P \in \mathcal{P}$ and $\mathbb{E}^Q[X] = 1 + \int hQ\mathrm{d}R > 1 + \varepsilon$ for each $Q \in \mathcal{Q}$. Proposition 3.2 then concludes the proof.

Suppose that $\mathcal{Q}$ is tight and that there exist $P^{(n)} \in \overline{\mathrm{Span}}\mathcal{P}$ and $Q^{(n)} \in \overline{\mathrm{Conv}}\mathcal{Q}$ such that $P^{(n)} + Q^{(n)} \to 0$. By Prokhorov's theorem, $\overline{\mathrm{Conv}}\mathcal{Q}$ is weakly compact. This implies for some subsequence $\{n_k\}$, $Q^{(n_k)}$ is convergent. The limit then belongs to $\overline{\mathrm{Span}}\mathcal{P} \cap \overline{\mathrm{Conv}}\mathcal{Q}$. The other direction is obvious. $\square$

We pose the open problem of characterizing the existence of pivotal, exact, and nontrivial p/e-variables with $\mathcal{P}, \mathcal{Q}$ infinite. For instance, in a very close direction, we pose the following conjecture, strengthening Proposition 7.9. We expect that the theory of simultaneous transport between infinite collections of measures will be helpful.

Conjecture 1. Suppose that $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta_0}$ and $\mathcal{Q} = \{Q_{\theta'}\}_{\theta' \in \Theta_1}$ are probability measures on $\mathfrak{X}$ and that there exist $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ such that $P_\theta \ll P_{\theta_0}$ and $Q_{\theta'} \ll Q_{\theta_1}$ for all θ, θ'. Assume also that $(P_\theta, Q_{\theta'})_{\theta \in \Theta_0, \theta' \in \Theta_1}$ is jointly atomless.[10] There exists a pivotal and exact e-variable X satisfying $\inf_{Q \in \mathcal{Q}} \mathbb{E}^Q[\log X] > 0$ if and only if $0 \notin \overline{\overline{\mathrm{Span}}\mathcal{P} + \overline{\mathrm{Conv}}\mathcal{Q}}$, where the closure is taken wrt the total variation distance.

Our next result shows that surprisingly, even in simple settings where $\mathcal{P}$ and $\mathcal{Q}$ are seemingly distant, an e-variable may not exist.

Proposition 7.10. *Let P be an infinitely divisible distribution on $\mathbb{R}^d$ with a density p. Consider $\mathcal{P} := \{P_\theta\}_{\theta \in \mathbb{R}^d}$ that are the shifts of the measure P, where P_θ has density $p(x-\theta)$. Let Q be any distribution on $\mathbb{R}^d$ with a density q. Then for each $\mathcal{Q}$ that contains Q, there exists no exact e-variable for $\mathcal{P}$ that is nontrivial for $\mathcal{Q}$.*

Note that here we have reached a slightly stronger conclusion than the forward direction of Proposition 7.9, that even an unbounded e-variable would not exist. The absolute continuity of Q cannot be removed. For instance, if Q has a mass at $x \in \mathbb{R}^d$, $X = 1+\delta_x$ would be an exact e-variable that is nontrivial for $\{Q\}$.

A particular instance of interest is when Q is Gaussian. In this case, Gangrade et al. [2023] effectively proved that for the set of all Gaussians (of all means and all covariances), there does not exist an e-variable with nontrivial power, even non-exact. Thus, our result is stronger in that it allows for a much smaller $\mathcal{P}$ that just includes all translations of any single Gaussian, but it is weaker in that it only shows that an *exact* e-variable with nontrivial power does not exist.

Proof of Proposition 7.10. By Sato [1999, Lemma 7.5], the Fourier transform of the density p of an infinitely divisible distribution has no real zeros. By Wiener's Tauberian theorem (Theorem 8 of Wiener [1933]), the linear span of the set of translates $\{p(\cdot - \theta)\}_{\theta \in \mathbb{R}}$ is dense in $L^1(\mathbb{R})$. Therefore, there is $P \in \mathrm{Span}\{P_\theta : \theta \in \mathbb{R}\}$ with density $\widetilde{p}$ such that $d_{\mathrm{TV}}(P, Q) = (\int |\widetilde{p}(x) - q(x)| \,\mathrm{d}x)/2 < \varepsilon$. In other words, $Q \in \overline{\mathrm{Span}}\{P_\theta : \theta \in \mathbb{R}\}$, say we have $Q = \lim_{k\to\infty} P^{(k)}$, where $P^{(k)} \in \mathrm{Span}\{P_\theta : \theta \in \mathbb{R}\}$.

Suppose that X is an exact e-variable that is nontrivial for $\{Q\}$. Then there exists a large number $K > 0$ such that $\widetilde{X} := X\mathbb{1}_{\{X \leqslant K\}}$ satisfies $\mathbb{E}^Q[\widetilde{X}] > 1$. Since $\widetilde{X}$ is bounded, we have

$$1 < \mathbb{E}^Q[\widetilde{X}] = \lim_{k\to\infty} \mathbb{E}^{P^{(k)}}[\widetilde{X}] \leqslant \limsup_{k\to\infty} \mathbb{E}^{P^{(k)}}[X] = 1.$$

This leads to a contradiction. $\square$

[10] If Θ_0 or Θ_1 is infinite, this can be defined in the natural way as in Definition 2.2.

8 On the existence of nontrivial test (super)martingales

From here on, for $t \in \{1, 2, \dots\}$, let Z^t denote $(Z_1, \dots, Z_t)$, which represents data on $\mathfrak{X}^t$, and let $\mathcal{F}$ by default represent the data filtration, meaning that $\mathcal{F}_t = \sigma(Z^t)$.

A sequence of random variables $Y \equiv (Y_t)_{t \geqslant 0}$ is called a *process* if it is adapted to $\mathcal{F}$, that is, if Y_t is measurable wrt $\mathcal{F}_t$ for every t. However, Y may also be adapted to a coarser filtration $\mathcal{G}$; for example $\sigma(Y^t)$ could be strictly smaller than $\mathcal{F}_t$. Such situations will be of special interest to us. Henceforth, $\mathcal{F}$ will always denote the data filtration, $\mathcal{G}$ will denote a generic subfiltration (which could equal $\mathcal{F}$, or be coarser). An $\mathcal{F}$-stopping time τ is a nonnegative integer valued random variable such that $\{\tau \leqslant t\} \in \mathcal{F}_t$ for each $t \geqslant 0$. Denote by $\mathbb{T}_{\mathcal{F}}$ the set of all $\mathcal{F}$-stopping times, excluding the constant 0 and including ones that may never stop. Note that if $\mathcal{G} \subseteq \mathcal{F}$, then $\mathbb{T}_{\mathcal{G}} \subseteq \mathbb{T}_{\mathcal{F}}$. In this section, $\mathcal{P}$ is a set of measures on the sample space $\mathfrak{X}^\infty$.

Test (super)martingales. A process M is a *martingale* for P wrt $\mathcal{G}$ if

$$\mathbb{E}^P[M_t \mid \mathcal{G}_{t-1}] = M_{t-1} \tag{15}$$

for all $t \geqslant 1$. M is a *supermartingale* for P if it satisfies (15) with "=" relaxed to "$\leqslant$". A (super)martingale is called a *test (super)martingale* if it is nonnegative and $M_0 = 1$. A process M is called a test (super)martingale for $\mathcal{P}$ if it is a test (super)martingale for every $P \in \mathcal{P}$. The process M is then called a *composite test (super)martingale*. We say that M is *nontrivial for* $\mathcal{Q}$ if $\mathbb{E}^Q[M_\tau] > 1$ under all $Q \in \mathcal{Q}$ and for all finite stopping times $\tau \in \mathbb{T}_{\mathcal{F}}$, and M has *nontrivial power against* $\mathcal{Q}$ if $\mathbb{E}^Q[\log M_\tau] > 0$ under all $Q \in \mathcal{Q}$ and for all finite stopping times $\tau \in \mathbb{T}_{\mathcal{F}}$.

It is easy to construct test martingales for singletons $\mathcal{P} = \{P\}$: we can pick any $Q \ll P$, and then the likelihood ratio process $(\mathrm{d}Q/\mathrm{d}P)(X^t)$ is a test martingale for P (and its reciprocal is a test martingale for Q). In fact, every test martingale for P takes the same form, for some Q.

Composite test martingales M are simultaneous likelihood ratios, meaning that they take the form of a likelihood ratio simultaneously for every element of $\mathcal{P}$. Formally, for every $P \in \mathcal{P}$, there exists a distribution Q^P that is absolutely continuous wrt P and satisfies $M_t = (\mathrm{d}Q^P/\mathrm{d}P)(X^t)$. Trivially, the constant process $M_t = 1$ is a test martingale for each $\mathcal{P}$, and any decreasing process taking values in $[0, 1]$ is a test supermartingale for each $\mathcal{P}$. We call a test (super)martingale *nondegenerate* if it is not always a constant (or decreasing) process. Nondegenerate test supermartingales do not always exist: whether they exist or not depends on the richness of $\mathcal{P}$.

On the existence of nondegenerate test (super)martingales. If $\mathcal{P}$ is too large, there may be no nondegenerate test martingales wrt $\mathcal{F}$. To explain the situation, suppose that $\mathcal{P}$ contains only measures of iid sequences with marginal distributions in a set $\mathcal{P}^{\mathrm{mar}} \subseteq \Pi(\mathfrak{X})$. Examples of the non-existence phenomenon include the case when $\mathcal{P}^{\mathrm{mar}}$ is the set of all subGaussian distributions [Ramdas et al., 2020], all log-concave distributions [Gangrade et al., 2023], or all Bernoulli distributions [Ramdas et al., 2022b]. In all these cases, nondegenerate test martingales have been proven to not exist, at least in the original filtration $\mathcal{F}$. Sometimes, nondegenerate test *super*martingales may still exist, as in the subGaussian case. But if $\mathcal{P}^{\mathrm{mar}}$ is too large or rich (as in the exchangeable and log-concave cases), even nondegenerate test supermartingales do not exist.

However, the situation is subtle: in the above situations, there could still exist nontrivial test (super)martingales in some $\mathcal{G} \subseteq \mathcal{F}$; note that a nontrivial one must be nondegenerate. Indeed, for the exchangeable setting described above, Vovk [2021] constructs exactly such a test martingale in a reduced filtration. It is a priori not obvious exactly when shrinking the filtration allows for nontrivial test (super)martingales to emerge, and how exactly should one shrink $\mathcal{F}$ (the relevant filtration $\mathcal{G}$ is not evident at the outset).

Our results for (exact) e-variables have direct implications for the existence of test (super)martingales. For simplicity, consider the iid case, where each $Z_i \sim P$ for some $P \in \mathcal{P}^{\mathrm{mar}}$ or $P \in \mathcal{Q}^{\mathrm{mar}}$; that is, $\mathcal{P} = \{P^\infty \mid P \in \mathcal{P}^{\mathrm{mar}}\}$ and $\mathcal{Q} = \{P^\infty \mid P \in \mathcal{Q}^{\mathrm{mar}}\}$.

Corollary 8.1. *Let $\mathcal{P}^{\mathrm{mar}}$ and $\mathcal{Q}^{\mathrm{mar}}$ be convex polytopes in $\Pi(\mathfrak{X})$. If $\mathcal{P}^{\mathrm{mar}}$ and $\mathcal{Q}^{\mathrm{mar}}$ are disjoint, then there exists a test supermartingale for $\mathcal{P}$ that is nontrivial for $\mathcal{Q}$. If $(\mathrm{Span}\mathcal{P}^{\mathrm{mar}}) \cap \mathcal{Q}^{\mathrm{mar}} = \emptyset$, then there exists a test martingale for $\mathcal{P}$ that is nontrivial for $\mathcal{Q}$.*

The proof is simple, and does not require joint non-atomicity. The conditions on $\mathcal{P}$ and $\mathcal{Q}$ imply that an (exact) e-variable (based on t sample points for any t) exists for $\mathcal{P}$ that is powerful against $\mathcal{Q}$ by Corollary 7.7. We can form our (super)martingale by simply multiplying these e-values (thus constructively proving the corollary).

We conjecture that the converse direction in the above corollary also holds, perhaps with some additional conditions; in other words, we conjecture that if a test martingale for $\mathcal{P}$ is nontrivial for $\mathcal{Q}$, then the span of $\mathcal{P}^{\mathrm{mar}}$ does not intersect $\mathcal{Q}^{\mathrm{mar}}$. (To explain why we cannot directly invoke the reverse directions of our theorems, it is possible that the construction of the e-variable at step t can use information about the distribution gained in the first $t-1$ steps. In short, there (of course) exist test (super)martingales that are not simply the products of independent e-values, and ruling those out requires further arguments.)

The first (supermartingale) part of Corollary 8.1 is closely related to the main result by Grünwald et al. [2023], albeit they require some extra technical conditions in their theorem statement, while relaxing the polytope requirement. The second (martingale) part is new to the best of our knowledge, and is a key addition to the emerging literature on game-theoretic statistics [Ramdas et al., 2022a].

Remark 8.2. Let $\mathcal{P}^{\mathrm{mar}} = \mathrm{Conv}(\{P_1, \dots, P_L\})$ with L finite and suppose $Q \in \mathrm{Span}\mathcal{P}^{\mathrm{mar}}$ but $Q \notin \mathcal{P}^{\mathrm{mar}}$. By using Theorem 4.3, there does not exist a nontrivial test martingale for $\mathcal{P}$ against $\{Q^\infty\}$ wrt the original filtration. On the other hand, if $P_1, \dots, P_L \ll Q$, then by Proposition 5.12, there exists a reduced filtration — in particular formed by combining data points — with respect to which a nontrivial test martingale exists.

E-processes and fork-convex hulls. E-processes are a generalization of test supermartingales, that have emerged as a powerful general concept in the recent literature when testing composite $\mathcal{P}$. We briefly introduce the concept below and relate it to the above results.

A family $(M^P)_{P\in\mathcal{P}}$ is a *test martingale family* if M^P is always a test martingale for P (wrt $\mathcal{G}$). A nonnegative process E is called an *e-process* for $\mathcal{P}$ (wrt $\mathcal{G}$) if there is a test martingale family $(M^P)_{P\in\mathcal{P}}$ such that

$$E_t \leqslant M_t^P \text{ for every } P \in \mathcal{P}, t \geqslant 0. \tag{16}$$

This type of definition was used by Howard et al. [2020], who used the name "sub-ψ process". In parallel, Grünwald et al. [2023] implicitly defined an e-process for $\mathcal{P}$ (wrt $\mathcal{G}$), also without using the name "e-process", as a nonnegative process E such that

$$\mathbb{E}^P[E_\tau] \leqslant 1 \text{ for every } \tau \in \mathbb{T}_{\mathcal{G}}, P \in \mathcal{P}.$$

In words, E_τ must be an e-value at any $\mathbb{T}_{\mathcal{G}}$-stopping time. Ramdas et al. [2020] proved that the two definitions are (under mild technical conditions) equivalent.

By the optional stopping theorem for nonnegative supermartingales, it is easy to see that test supermartingales for $\mathcal{P}$ are e-processes for $\mathcal{P}$ (wrt $\mathcal{G}$). But the latter class is much larger: there are many problems for which an e-process can be designed wrt $\mathcal{F}$, but no test supermartingale exists wrt $\mathcal{F}$ (e.g., the earlier mentioned cases of testing exchangeability and log-concavity; also see Ruf et al. [2022]).

The difference between test supermartingales and e-processes is connected to the geometric concept of *fork convexity*. Ramdas et al. [2022b] proved that if M is a test supermartingale for $\mathcal{P}$, then it is a test supermartingale for $\mathrm{fConv}(\mathcal{P})$, the fork convex hull of $\mathcal{P}$ (a much larger set). The same is not true for e-processes, which are only closed under convex hulls, not fork-convex hulls: if M is an e-process for $\mathcal{P}$ (wrt some $\mathcal{G}$), then it is an e-process for $\mathrm{Conv}(\mathcal{P})$ (wrt $\mathcal{G}$), the convex hull of $\mathcal{P}$, while it is generally not an e-process for $\mathrm{fConv}(\mathcal{P})$.

It is an important open problem to fully characterize the geometric conditions needed for the (non)existence of e-processes and test (super)martingales wrt $\mathcal{F}$ or some reduced $\mathcal{G}$. Comparing the e-powers obtained for the different solution concepts in different filtrations seems nontrivial, and also a key open direction.

9 Summary

This paper uses tools from convex geometry and simultaneous optimal transport to shed light on some fundamental questions in statistics: when can one construct an exact p/e-value for a composite null, which is nontrivially powerful against a composite alternative? The answer, in the case where the null and alternative hypotheses are convex polytopes in the space of probability measures, is cleanly characterized by convex hulls and spans of the null and alternative sets of distributions. Several other related properties, like pivotality under the null, end up being central, where the technical property of joint non-atomicity is assumed in some of our results.

Our proofs are constructive when the alternative is simple, and in simple cases, we provide corroborating empirical evidence of the correctness of our theory. A key role is played by the shrinking of the data filtration (accomplished by the transport map which maps the composite null to a single uniform). Implications for the existence of composite test (super)martingales are also briefly discussed.

We mention several open problems along the way, and anticipate that our results can be generalized beyond convex polytopes with the development of newer technical tools. For instance, it is of great interest to generalize the SHINE construction to the composite alternative setting, and extend our results to general convex subsets of probability measures that are not polytopes.

Acknowledgements

We thank Martin Larsson and Johannes Wiesel for helpful discussions. Codes used to generate simulation and numerical results can be found at https://github.com/Hungryzzy/SHINE.

References

Marinho Bertanha and Marcelo J Moreira. Impossible inference in econometrics: Theory and applications. *Journal of Econometrics*, 218(2):247–270, 2020.

John B Conway. *A Course in Functional Analysis*, volume 96. Springer, 1990.

Aditya Gangrade, Alessandro Rinaldo, and Aaditya Ramdas. A sequential test for log-concavity. *arXiv preprint arXiv:2301.03542*, 2023.

P. Grünwald, Rianne De Heide, and Wouter Koolen. Safe testing. *Journal of the Royal Statistical Society, Series B (to appear)*, 2023.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

J.L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, pages 917–926, 1956.

Charles Kraft. Some conditions for consistency and uniform consistency of statistical procedures. *University of California Publication in Statistics*, 2:125–141, 1955.

Wolfgang Kühnel. *Differential Geometry*, volume 77. American Mathematical Soc., 2015.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv:2009.03167*, 2020.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *arXiv preprint arXiv:2210.01948*, 2022a.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter M Koolen. Testing exchangeability: fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 141:83–109, 2022b.

R Tyrrell Rockafellar. *Convex Analysis*, volume 18. Princeton University Press, 1970.

Johannes Ruf, Martin Larsson, Wouter M Koolen, and Aaditya Ramdas. A composite generalization of ville's martingale theorem. *arXiv:2203.04485*, 2022.

Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, 1999.

Glenn Shafer. Testing by betting: a strategy for statistical and scientific communication (with discussion and response). *Journal of the Royal Statistic Society A*, 184(2):407–478, 2021.

Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game.* Wiley, New York, 2001.

Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance.* Wiley, Hoboken, New Jersey, 2019.

Moshe Shaked and J George Shanthikumar. *Stochastic Orders.* Springer, 2007.

Jie Shen, Yi Shen, Bin Wang, and Ruodu Wang. Distributional compatibility for change of measures. *Finance and Stochastics*, 23(3):761–794, 2019.

Barry Simon. *Convexity: an Analytic Viewpoint*, volume 187. Cambridge University Press, 2011.

Gordon Simons. A martingale decomposition theorem. *The Annals of Mathematical Statistics*, 41 (3):1102–1104, 1970.

Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140. American Mathematical Soc., 2012.

Vladimir Vovk. Testing randomness online. *Statistical Science*, 36(4):595–611, 2021.

Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination, and applications. *Annals of Statistics*, 49:1736–1754, 2021.

Vladimir Vovk and Ruodu Wang. Efficiency of nonparametric e-tests. *arXiv preprint arXiv:2208.08925*, 2022.

Abraham Wald. Sequential tests of statistical hypotheses. *Ann. Math. Statist.*, 16(2):117–186, 1945.

Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2022.

Ruodu Wang and Zhenyuan Zhang. Simultaneous optimal transport. *arXiv preprint arXiv:2201.03483 v2*, 2023.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2022. (accepted, to appear with discussion).

Norbert Wiener. *The Fourier integral and certain of its applications.* Cambridge University Press, 1933.

A Some technical lemmas

We start with a few elementary results from convex analysis. We refer the readers to Rockafellar [1970] and Simon [2011] for more background.

Lemma A.1. *Let $A \subset \mathbb{R}^d$ be a closed set, and $\mu \in \mathcal{M}(\mathbb{R}^d)$ with $\operatorname{supp}\mu = A$. Then the followings hold.*

(i) $\operatorname{aff} A = \operatorname{aff}\operatorname{ri}(\operatorname{Conv}A; \operatorname{aff} A)$*;*

(ii) $\operatorname{bary}(\mu) \in \operatorname{ri}(\operatorname{Conv}A; \operatorname{aff} A)$*.*

Proof. (i) The $\supseteq$ direction is obvious. To prove $\subseteq$, we may replace A by $\operatorname{Conv}A$ and without loss of generality assume A is also convex. Let $a \in A$ and $b \in \operatorname{ri}(A; \operatorname{aff} A)$, then elementary geometric arguments show that $(a+b)/2 \in \operatorname{ri}(A; \operatorname{aff} A)$; see Theorem 6.1 of Rockafellar [1970]. Thus $A \in \operatorname{aff}\operatorname{ri}(A; \operatorname{aff} A)$.

(ii) We may without loss of generality assume $\operatorname{aff} A = \mathbb{R}^d$ and replace the relative interior by interior. An application of the Hahn-Banach separation theorem yields $\operatorname{bary}(\mu) \in \overline{\operatorname{Conv}A}$. Suppose $\operatorname{bary}(\mu) \notin \operatorname{ri}(\overline{\operatorname{Conv}A}; \operatorname{aff} A) = (\overline{\operatorname{Conv}A})^\circ$, then the Hahn-Banach separation theorem implies the existence of a closed hyperplane $\mathbb{H} \subseteq \mathbb{R}^d$ such that $\operatorname{bary}(\mu) \in \mathbb{H}$ and $(\overline{\operatorname{Conv}A})^\circ \subseteq \mathbb{R}^d \setminus \mathbb{H}$; see Theorem 11.2 of Rockafellar [1970]. Therefore, $A \subseteq \partial\mathbb{H}$, contradicting $\operatorname{aff} A = \mathbb{R}^d$. By Theorem 6.3 of Rockafellar [1970], $\operatorname{ri}(\overline{\operatorname{Conv}A}; \operatorname{aff} A) = \operatorname{ri}(\operatorname{Conv}A; \operatorname{aff} A)$. This completes the proof. □

We also prove the following variant of the Choquet-Meyer theorem.

Lemma A.2. *Suppose that μ is a finite measure on $\mathbb{R}^d$, $x_1, \dots, x_k \in \operatorname{supp}\mu$, and $x \in \operatorname{ri}(\operatorname{Conv}\{x_1, \dots, x_k\}; \operatorname{aff}\operatorname{supp}\mu)$. Then there exists $\delta > 0$ such that any measure γ with total mass $\gamma(\mathbb{R}^d) \leqslant \delta$, supported on $B(x;\delta) \cap (\operatorname{aff}\operatorname{supp}\mu)$, satisfies $\gamma \preceq_{\mathrm{cx}} \widetilde{\mu}$ for some $\widetilde{\mu} \leqslant \mu$.*

Proof. First, we may assume without loss of generality that $\operatorname{aff}\operatorname{supp}\mu = \mathbb{R}^d$, and replace the relative interior by interior. In this case, we must have $\operatorname{aff}\{x_1, \dots, x_k\} = \operatorname{aff}\operatorname{supp}\mu = \mathbb{R}^d$, otherwise $(\operatorname{Conv}\{x_1, \dots, x_k\})^\circ = \emptyset$ and the statement is vacuously true.

Since $x \in (\operatorname{Conv}\{x_1, \dots, x_k\})^\circ$, there exists $\varepsilon > 0$ such that the distance of x from $\partial\operatorname{Conv}\{x_1, \dots, x_k\}$ is larger than ε. Let μ_N for $N \in \mathbb{N}$ be the conditional distribution of μ given the σ-field generated by cubes with coordinates in $\mathbb{Z}^d/N$. The smallest cubes have size $(1/N)^d$. For each $j = 1, \dots, k$, pick a cube D_j^N of size $(1/N)^d$ in $\mathbb{R}^d$ containing x_j (possibly on its boundary) that has a positive μ-measure, which is possible since x_j is in the support of μ. Let $y_j^N = \operatorname{bary}(\mu|_{D_j^N})$. It is then clear that $\mu_N(\{y_j^N\}) > 0$ and $\mu_N(\{y_j^N\})\delta_{y_j^N} \preceq_{\mathrm{cx}} \mu|_{D_j^N}$.

For $N > d^{3/2}/\varepsilon$, $\|y_j^N - x_j\| < \varepsilon/d$. Therefore, $x \in \operatorname{ri}(\operatorname{Conv}\{y_1^N, \dots, y_k^N\}; \operatorname{aff}\{y_1^N, \dots, y_k^N\})$. Fix $N > d^{3/2}/\varepsilon$ such that the boxes $\{D_j^N\}_{1 \leqslant j \leqslant k}$ are disjoint. Write $(y_1, \dots, y_k) = (y_1^N, \dots, y_k^N)$ and $D_j = D_j^N$. Note that the distance between x and $\partial\operatorname{Conv}\{y_1, \dots, y_k\}$ is positive by the triangle inequality, and hence $\operatorname{aff}\{y_1, \dots, y_k\} = \mathbb{R}^d$, so that $x \in (\operatorname{Conv}\{y_1, \dots, y_k\})^\circ$.

Pick $\delta > 0$ small enough such that $B(x;\delta) \subseteq (\mathrm{Conv}\{y_1,\dots,y_k\})^\circ$ and that $\delta < \min\{\mu_N(\{y_1\}),\dots,\mu_N(\{y_k\})\}$. By Choquet's theorem (Theorem 10.7(ii) of Simon [2011]), for each $y \in B(x;\delta)$, there exists a probability measure γ_y supported on $\{y_1,\dots,y_k\}$ such that $\mathrm{bary}(\gamma_y) = y$, and γ_y is continuous in y.

Consider an arbitrary measure γ with total mass $\gamma(\mathbb{R}^d) \leqslant \delta$ and supported on $B(x;\delta)$. Define

$$\widetilde{\gamma} = \int \gamma_y \gamma(\mathrm{d}y).$$

Observe that $\gamma \preceq_{\mathrm{cx}} \widetilde{\gamma}$ and $\widetilde{\gamma}$ is supported on $\{y_1,\dots,y_k\}$ with

$$\widetilde{\gamma}(\{y_j\}) = \int \gamma_y(\{y_j\})\gamma(\mathrm{d}y) \leqslant \gamma(\mathbb{R}^d) \leqslant \delta \leqslant \mu_N(\{y_j\}),\ 1 \leqslant j \leqslant k.$$

Define

$$\widetilde{\mu} = \sum_{j=1}^k \left(\frac{\widetilde{\gamma}(\{y_j\})}{\mu_N(\{y_j\})}\right)\mu|_{D_j}.$$

It follows that

$$\gamma \preceq_{\mathrm{cx}} \widetilde{\gamma} = \sum_{j=1}^k \left(\frac{\widetilde{\gamma}(\{y_j\})}{\mu_N(\{y_j\})}\right)\mu_N(\{y_j\})\delta_{y_j} \preceq_{\mathrm{cx}} \sum_{j=1}^k \left(\frac{\widetilde{\gamma}(\{y_j\})}{\mu_N(\{y_j\})}\right)\mu|_{D_j} = \widetilde{\mu}.$$

Since $\{D_j\}_{1\leqslant j\leqslant k}$ are disjoint,

$$\widetilde{\mu} \leqslant \sum_{j=1}^k \mu|_{D_j} \leqslant \mu,$$

as desired. □

We next state and prove a few elementary results regarding the stochastic order $\preceq_{\mathrm{st}}$. These are useful when proving existence of p-variables.

Lemma A.3. *Suppose that $F, G \in \Pi(\mathbb{R})$ are atomless and $F \neq G$. Then there exists a bounded random variable ϕ on $\mathbb{R}$ such that its law under F is U_1 and its law under G is $\preceq_{\mathrm{st}} \mathrm{U}_1$ but distinct from U_1.*

Proof. We pick a random variable ϕ that has law U_1 and is comonotone with $\mathrm{d}G/\mathrm{d}F$ under the law F. In particular, ϕ and $\mathrm{d}G/\mathrm{d}F$ are positively associated. Therefore, for $\alpha \in [0,1]$,

$$G(\phi \leqslant \alpha) = \int \frac{\mathrm{d}G}{\mathrm{d}F}\mathbb{1}_{\{\phi\leqslant\alpha\}}\mathrm{d}F \geqslant \int \frac{\mathrm{d}G}{\mathrm{d}F}\mathrm{d}F \int \mathbb{1}_{\{\phi\leqslant\alpha\}}\mathrm{d}F = F(\phi \leqslant \alpha) = \alpha.$$

Since $\mathrm{d}G/\mathrm{d}F$ is not a constant under F, there exists $\alpha \in (0,1)$ such that the inequality is strict. □

Lemma A.4. *Suppose that $F_1,\dots,F_L,G$ are atomless probability measures on $[0,1]$.*

(i) If $F_i \succeq_{\mathrm{st}} \mathrm{U}_1$ for all i and $G \prec_{\mathrm{st}} \mathrm{U}_1$, then there exists a random variable $\Psi : [0,1] \to [0,1]$ such that $\Psi|_{F_i} \succ_{\mathrm{st}} \mathrm{U}_1$ for all i and $\Psi|_G = \mathrm{U}_1$.

(ii) If there exists $\beta \in (0,1)$ such that $\mathrm{d}F_i/\mathrm{dU}_1 \leqslant 1$ on $[0,\beta)$ and $\mathrm{d}F_i/\mathrm{dU}_1 \geqslant 1$ on $(\beta,1]$, and $F_i \succ_{\mathrm{st}} \mathrm{U}_1$ for all i and $G = \mathrm{U}_1$, then there exists a random variable $\Psi : [0,1] \to [0,1]$ such that $\Psi|_{F_i} \succeq_{\mathrm{st}} \mathrm{U}_1$ for all i and $\Psi|_G \prec_{\mathrm{st}} \mathrm{U}_1$.

Proof. (i) Let $F = \max F_i$ and Id be the identity on $[0,1]$. Let $\widetilde{F}$, $\widetilde{G}$ and $\widetilde{F_i}$ be the corresponding cdfs. $\widetilde{G} > \mathrm{Id} \geqslant \widetilde{F} \geqslant \widetilde{F_i}$ and $\widetilde{F_i}|_{F_i} \succeq_{\mathrm{st}} \mathrm{Id}|_{F_i} \overset{\mathrm{law}}{=} \mathrm{U}_1$ for each i. Hence $\widetilde{G}$ follows U_1 under G and it dominates U_1 under each F_i by Theorem 1.A.3.(a) of Shaked and Shanthikumar [2007].

(ii) Denote by $\alpha = \min\{\beta - F_i([0,\beta))\} > 0$. Pick τ such that $\max F_i([\beta, \beta+2\tau)) < \alpha$. Define

$$\Psi(x) = \begin{cases} x & \text{if } x \in [0, \beta+\tau] \cup [\beta+2\tau, 1]; \\ x - \tau & \text{otherwise.} \end{cases}$$

By construction, it is then easy to check that $\Psi|_{\mathrm{U}_1} \prec_{\mathrm{st}} \mathrm{U}_1$ and $\Psi|_{F_i} \succeq_{\mathrm{st}} \mathrm{U}_1$. $\square$

Finally, we prove Proposition 5.12, which we believe is well-known but have not found a proof in the literature.

Proof of Proposition 5.12. Suppose that $Q \in \mathrm{Span}\mathcal{P}$ and $Q^2 \in \mathrm{Span}\mathcal{P}^2$. We may assume that $\mathrm{Span}\mathcal{P} = \mathrm{Span}(P_1, \dots, P_\ell)$ for some $\ell \leqslant L$ and that $P_1, \dots, P_\ell$ are linearly independent. Denote by $f_j = \mathrm{d}P_j/\mathrm{d}Q$, so that $f_1, \dots, f_\ell$ are linearly independent as functions in $L^1(Q)$. By construction, there exists a unique tuple of nonzero numbers $(a_1, \dots, a_\ell)$ such that $\sum_{j=1}^{\ell} a_j P_j = Q$. In particular, $\sum_{j=1}^{\ell} a_j = 1$ and

$$1 = a_1 f_1 + \cdots + a_\ell f_\ell \quad Q\text{-a.e.} \tag{17}$$

Since $Q^2 \in \mathrm{Span}\mathcal{P}^2$, there exist $b_1, \dots, b_L$ such that for any set $A \in \mathcal{F}$,

$$\int_{A\times A} 1 Q(\mathrm{d}x) Q(\mathrm{d}y) = \int_{A\times A} \sum_{j=1}^{L} b_j f_j(x) f_j(y) Q(\mathrm{d}x) Q(\mathrm{d}y).$$

By symmetry of the integrand wrt x, y, we must have for any $A, B \in \mathcal{F}$,

$$\int_{A\times B} 1 Q(\mathrm{d}x) Q(\mathrm{d}y) = \int_{A\times B} \sum_{j=1}^{L} b_j f_j(x) f_j(y) Q(\mathrm{d}x) Q(\mathrm{d}y).$$

By Carathéodory's extension theorem, it holds

$$\sum_{j=1}^{L} b_j f_j(x) f_j(y) = 1 \quad Q^2\text{-a.e.} \tag{18}$$

By considering the a.e. set $y \in \mathfrak{X}$ where (18) holds and comparing (17) and (18), we have for any $1 \leqslant j \leqslant \ell$, f_j is Q-a.e. constant. This implies $P_1 = Q$. Therefore, $Q \in \mathcal{P}$. $\square$