The Annals of Statistics
2024, Vol. 52, No. 5, 2241–2267
https://doi.org/10.1214/24-AOS2434

ON THE EXISTENCE OF POWERFUL P-VALUES AND E-VALUES FOR COMPOSITE HYPOTHESES

BY ZHENYUAN ZHANG[1,a], AADITYA RAMDAS[2,b] AND RUODU WANG[3,c]

[1]*Department of Mathematics, Stanford University,* [a]*zzy@stanford.edu*
[2]*Department of Statistics & Data Science, Carnegie Mellon University,* [b]*aramdas@cmu.edu*
[3]*Department of Statistics and Actuarial Science, University of Waterloo,* [c]*wang@uwaterloo.ca*

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 $\mathcal{Q}$ 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. More generally, even when $\mathcal{P}$ and $\mathcal{Q}$ are not polytopes, we characterize the existence of a bounded nontrivial e-variable whose expectation exactly equals one under any $P \in \mathcal{P}$. 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 this phenomenon. We also provide an iterative construction that explicitly constructs such processes, and under certain conditions it 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.

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) several central questions in this paper. The first one is:

> *(Q-exact-p).* 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:

> *(Q-exact-e).* Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, when 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 $\mathbb{E}^Q[\log X] > 0$ (or $\mathbb{E}^Q[X] > 1$) 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 and will be explained soon below.

Received July 2023; revised May 2024.
MSC2020 subject classifications. 62B15, 62G10, 49Q22, 60G42.
Key words and phrases. p-values, e-values, composite hypothesis testing, convex order, simultaneous optimal transport, nonnegative martingale.

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

> *(Q-general-p).* Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, when does there exist a p-value for $\mathcal{P}$ that has nontrivial power against $\mathcal{Q}$? (see Terminology below.)

Or, for the e-value, we require that $\mathbb{E}^P[X] \le 1$ for any $P \in \mathcal{P}$:

> *(Q-general-e).* Given a null $\mathcal{P}$ and an alternative $\mathcal{Q}$, when does there exist an e-value for $\mathcal{P}$ that has nontrivial "e-power" against $\mathcal{Q}$? (see Terminology below.)

For these nonexact problems, we can still provide a clean characterization of the existence for both *(Q-general-p)* and *(Q-general-e)*. An immediate follow-up question is:

> *(Q-power).* Suppose that we know the p-values or e-values in the above questions do exist. How can we algorithmically construct powerful, or even optimal, ones?

This question is important for the application of our ideas in hypothesis testing.

These appear to be rather fundamental questions, and our answers will be proved using recent techniques in simultaneous optimal transport, combined with classical convex geometric arguments. A natural motivation for exactness of p-values and e-values comes from the trivial observation that, in the case of a simple null hypothesis, any nonexact p-value or e-value can be strictly improved. Although this is not necessarily true for composite hypotheses, the existence of such exact p-values and e-values, as well as the trade-off between exactness and power, is useful for the design of tests.

Note that in the characterizations for *(Q-exact-p)* and *(Q-general-p)* above, a technical condition of joint nonatomicity will be assumed, which is essentially equivalent to allowing for external randomization. Our proofs are constructive and yield a simple iterative construction addressing *(Q-power)*, called SHINE (Separating Hyperplanes Iteration for Nontrivial and Exact e/p-variables), that can in principle explicitly build these objects and calculate their values on a given dataset, but it is 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:

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

Before proceeding, we introduce 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 $\mathbb{E}^P[X] \le 1$ 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 $P(X \le \alpha) \le \alpha$ for all $\alpha \in (0,1)$ and $P \in \mathcal{P}$, and a p-variable X is *exact* if $P(X \le \alpha) = \alpha$ for all $\alpha \in (0,1)$ and $P \in \mathcal{P}$. A p-variable X for $\mathcal{P}$ is *nontrivial* (or has *nontrivial power*) against $\mathcal{Q}$ if, for each $Q \in \mathcal{Q}$, $Q(X \le \alpha) \ge \alpha$ 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 $[0,1]$ by truncation, without changing 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. Since $x - 1 \ge \log x$, 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. Realizations of e-variables are called e-values. Like many other authors, we do not distinguish these terms when there is no confusion; the same applies to p-values and p-variables.

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 $n \geq 1$ lying in (say) $\mathbb{R}^d$ for some $d \geq 1$ (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 calculated on a batch of data. We return to the multi-period (sequential) case briefly later in the paper.

Summary of contributions. We briefly summarize the main results of this paper below. With the help of techniques from simultaneous transport, the existence of p/e-values for a convex polytope $\mathcal{P}$ and a simple alternative $\mathcal{Q} = \{Q\}$ is fully characterized in Theorems 3.1 and 3.4: under a natural condition of nonatomicity, we show that pivotal, exact, and powerful p/e-values exist if and only if $Q \notin \operatorname{Span}\mathcal{P}$; powerful p/e-values exist if and only if $Q \notin \mathcal{P}$. Theorems 6.1 and 6.2 extend these earlier results to the case of composite alternatives that are polytopes: for convex polytopes $\mathcal{P}$ and $\mathcal{Q}$, similar conclusions as before hold with the condition $Q \notin \operatorname{Span}\mathcal{P}$ being replaced by $\operatorname{Span}\mathcal{P} \cap \mathcal{Q} = \varnothing$, and the condition $Q \notin \mathcal{P}$ being replaced by $\mathcal{P} \cap \mathcal{Q} = \varnothing$. Theorem 6.8 extends these results to the case of general (nonpolytope) infinite $\mathcal{P}$, $\mathcal{Q}$, where the situation is more complicated: we now additionally need a common reference measure and a closure with respect to the total variation distance.

For the particular case of a simple alternative ($\mathcal{Q} = \{Q\}$), we can speak of maximizing the e-power under Q among all exact e-variables. The exact e-variable with the largest e-power is studied in a series of results including Theorems 4.4 and 4.7, and this finally leads to the SHINE construction, with maximality of the constructed e-variable shown in Theorem 5.3, providing an answer to *(Q-power)*.

Finally, the above results directly give rise to an answer to *(Q-martingale)* by obtaining sufficient conditions for the existence of a powerful e-process (Corollary 7.1).

Related results. The most directly related work is that of Grünwald, de Heide and Koolen (2024), which focuses primarily on e-values, and in particular *(Q-general-e)*. 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 *(Q-general-p)* and *(Q-general-e)* together. We need fewer technical conditions to establish our results, but their additional assumptions allow them to handle general $\mathcal{P}$, $\mathcal{Q}$ that are not polytopes. See also Harremoës, Lardy and Grünwald (2023) for a very recent follow-up work by the same group, which relaxes some of the original technical conditions.

A second related work is that of Ramdas et al. (2022). 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 $\mathcal{Q}^\infty$.

The above papers hint 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 accomplish 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 $\nu \neq \mu$, 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 several techniques from their work and build on them significantly to provide answers to our questions. In particular, our work tightly connects arguably basic testing problems with the modern theory of optimal transport.

A third classical yet fundamental related work is Kraft's theorem (Kraft (1955), Theorem 5), which states that if there is a σ-finite reference measure R that dominates every distribution in $\mathcal{P} \cup \mathcal{Q}$, then for each $\varepsilon > 0$ there exists a [0, 1]-valued random variable X with

$$\inf_{Q \in \mathcal{Q}} \mathbb{E}^{Q}[X] \geq \varepsilon + \sup_{P \in \mathcal{P}} \mathbb{E}^{P}[X] \tag{1}$$

if and only if the total variation distance $d_{\mathrm{TV}}(\mathrm{Conv}\mathcal{P}, \mathrm{Conv}\mathcal{Q}) \geq \varepsilon$. Kraft's theorem serves as a starting point for distinguishing sets of distributions (Hoeffding and Wolfowitz (1958)) and impossible inference (Bertanha and Moreira (2020)). In particular, in Remark 6.3 below we will see how Kraft's theorem can answer *(Q-general-e)* above.

Finally, likelihood ratios play an intimate role throughout our paper, but in rather different ways than classical hypothesis testing results. For general composite nulls and alternatives, generalized likelihood ratio-based methods require certain regularity conditions in order for Wilks' theorem Wilks (1938) to apply, which in turn yields an asymptotically exact p-value. An alternative, recent e-value approach is taken by universal inference Wasserman, Ramdas and Balakrishnan (2020). Our paper takes a very different approach, designing nonasymptotically exact p-values *(Q-exact-p)* or nonasymptotically conservative p-values *(Q-general-p)*, and also doing the same for e-values *(Q-exact-e, Q-general-e)*. We do not impose the regularity conditions required for Wilks' theorem to hold (our assumptions are different and quite mild), and we are interested in when such p-values or e-values exist and how one can construct them *(Q-power)*. As a rough, but instructive, intuition for how likelihood ratios play a role in our work, when deriving exact p-values or e-values, our method tries to find a transport map that can simultaneously transport the entire composite $\mathcal{P}$ into a single uniform U, while simultaneously transporting the composite $\mathcal{Q}$ into some distribution $F \neq U$. Now, having effectively converted the given composite problem into a point null U and a point alternative F, one can use simple likelihood ratios to design either the p-values or e-values.

Background on e-values. E-values are an alternative to p-values, and they have recently been actively studied in statistical testing by Wasserman, Ramdas and Balakrishnan (2020), Shafer (2021), Vovk and Wang (2021), Grünwald, de Heide and Koolen (2024), and Howard et al. (2021) under various names. Tests based on e-values are closely related to nonnegative supermartingale techniques for testing and estimation, which date back to work by Robbins Darling and Robbins (1967), Robbins and Siegmund (1974), and they emphasize continuous monitoring, optional stopping or continuation of experiments. The notion of e-processes generalizes that of likelihood ratios to composite hypotheses Ramdas et al. (2022). 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 game-theoretic probability Shafer and Vovk (2001, 2019). For a recent review on e-values and game-theoretic statistics, see Ramdas et al. (2023).

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, that is, $\int |x|\mu(\mathrm{d}x) < \infty$. 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 $X \overset{\text{law}}{\sim}_P \mu$, or simply $X \overset{\text{law}}{\sim} \mu$, 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 $A = \operatorname{supp}\mu$. The product measure is denoted by $P \otimes Q$. If $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q_1, \dots, Q_M\}$ are two sets of probability measures on $\mathfrak{X}$, we sometimes denote the tuple $(P_1, \dots, P_L, Q_1, \dots, Q_M)$ by $(\mathcal{P}, \mathcal{Q})$. For $P, Q \in \mathcal{M}(\mathfrak{X})$ we write $P \ll Q$ if P is absolutely continuous with respect to Q (sometimes we say Q dominates P), and $P \approx Q$ if $P \ll Q \ll P$.

3. *Stochastic orders*. For $F, G \in \Pi(\mathbb{R})$, we write $F \preceq_{\mathrm{st}} G$ if $F((-\infty, a]) \geq G((-\infty, a])$ 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 $\int \phi\, \mathrm{d}\mu \leq \int \phi\, \mathrm{d}\nu$ 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 $X \overset{\text{law}}{\sim} \mu$, $Y \overset{\text{law}}{\sim} \nu$, we sometimes abuse notation and write $X \preceq_{\mathrm{cx}} Y$ instead of $\mu \preceq_{\mathrm{cx}} \nu$. We write $\mu \leq \nu$ if $\mu(A) \leq \nu(A)$ for every Borel set A.

4. *Other notation*. Bold symbols such as $\mathbf{x}$ and $\boldsymbol{\alpha}$ will typically denote vectors. Write $\mathbf{1}_d = (1, \dots, 1) \in \mathbb{R}^d$, $\mathbf{0}_d = (0, \dots, 0) \in \mathbb{R}^d$, $\mathcal{I}_d = \{\mathbf{x} \in \mathbb{R}^d \mid x_1 = \cdots = x_d\} = \mathbb{R}\mathbf{1}_d$, and $\mathcal{I}_d^+ = \{\mathbf{x} \in \mathbb{R}^d \mid x_1 = \cdots = x_d \geq 0\} = \mathbb{R}_+\mathbf{1}_d$. 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. The easier case with a simple alternative ($|\mathcal{Q}| = 1$) will be solved first in Section 3. Under suitable conditions, we solve the maximization problem of the e-power in Section 4 and illustrate the SHINE construction for finding a powerful e-variable in Section 5 for a simple alternative, thus answering *(Q-power)*. We answer *(Q-exact-p)*, *(Q-exact-e)*, *(Q-general-p)* and *(Q-general-e)* in full in Section 6, where we deal with a general composite (and even infinite) alternative $\mathcal{Q}$. Finally, an application to composite test (super)martingales related to *(Q-martingale)* will be discussed in Section 7, followed by a summary in Section 8.

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).

In the setting of classical optimal transport theory, one usually starts with two measures $\mu \in \Pi(\mathfrak{X})$, $\nu \in \Pi(\mathfrak{Y})$ on Polish spaces $\mathfrak{X}, \mathfrak{Y}$, and a typical goal would be optimizing a certain functional over (X, Y) with respective marginals μ, ν (such (X, Y) are called couplings). The set of such couplings is also referred to as *transport plans*. In certain cases, one is interested in a special class of transport plans where Y is required to be a function of X. Such couplings are called *transport maps*. See Santambrogio (2015) and Villani (2009) for background on optimal transport.

A coupling (X, Y) on $\mathbb{R}^d \times \mathbb{R}^d$ is called a *martingale coupling* if $\mathbb{E}[Y|X] = X$. Given $\mu, \nu \in \Pi(\mathbb{R}^d)$, a *martingale transport (plan)* from μ to ν is a martingale coupling (X, Y) such that $X \overset{\text{law}}{\sim} \mu$ and $Y \overset{\text{law}}{\sim} \nu$. We recall from Strassen (1965) that there exists a martingale

[1] This is sometimes called the Choquet order in the mathematical literature, for example, Simon (2011).

transport from μ to ν if and only if $\mu \preceq_{\mathrm{cx}} \nu$ (see point 3 in the notation subsection for a definition). 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 $\nu \in \mathcal{N}$. These next facts can be found in Shaked and Shanthikumar (2007), Section 3.A.

LEMMA 2.1. *The followings hold for all integrable real-valued random variables*:

(i) *If* $\mathbb{E}[X] = \mathbb{E}[Y]$, *then* $X \preceq_{\mathrm{cx}} Y$ *if and only if* $\mathbb{E}[(X-a)_+] \le \mathbb{E}[(Y-a)_+]$ *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* $X_n \preceq_{\mathrm{cx}} Y \implies X \preceq_{\mathrm{cx}} Y$.

The recent work of Wang and Zhang (2023) proposed the notion of simultaneous optimal transport as an extension of classical optimal transport. As explained above, classical optimal transport theory concerns a coupling between two measures. In the setting of simultaneous optimal transport, one starts from two d-tuples of probability measures $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ on $\mathfrak{X}$ and $\boldsymbol{\nu} = (\nu_1, \dots, \nu_d)$ on $\mathfrak{Y}$, and requires that the transport plan (or map) sends μ_j to ν_j *simultaneously* for all $1 \le j \le d$. If $d = 1$, this coincides with the classical optimal transport.

Let us give a formal definition. For $d \ge 1$ 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 $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ denote the set of all *simultaneous transport plans* and *maps* from $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$ respectively, that is, $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is the set of all stochastic kernels κ such that

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

and

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

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. The mathematical structure of simultaneous optimal transport is very different from classical optimal transport, and the existence of simultaneous transport plans (or maps) is a nontrivial question. To further characterize the existence of simultaneous transport maps and plans, we need the notion of joint nonatomicity.

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 exists $\mu \gg \sum_{i=1}^d \mu_i$ 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, that is, simulating a uniform random variable independent of the Radon–Nikodym derivatives $(\mathrm{d}P_1/\mathrm{d}P, \dots, \mathrm{d}P_L/\mathrm{d}P)$ 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.

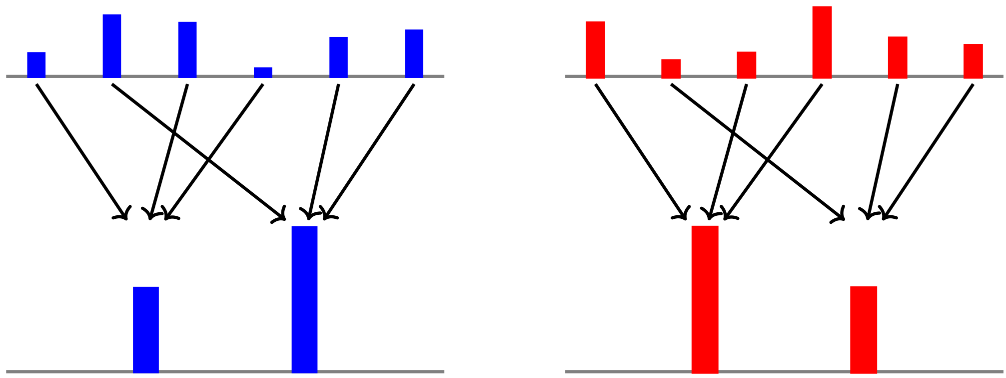

FIG. 1. *A showcase of simultaneous transport: here the input vector* $\boldsymbol{\mu}$ *is two-dimensional, as is the output vector* $\boldsymbol{\nu}$*. The two input distributions are discrete distributions over the same alphabet of size six and are drawn in different colors in the top row, with the height of a bar indicating its mass. The two target distributions are binary, indicated on the bottom row. The* simultaneous *transport requires that the maps that transport from* μ_1 *to* ν_1 *(left) and from* μ_2 *to* ν_2 *(right) are identical. This map is achieved by mixing (averaging) the Radon–Nikodym derivatives. Denoting* $\bar{\nu} = \nu_1 + \nu_2$ *and* $\bar{\mu} = \mu_1 + \mu_2$*, we have* $\frac{\mathrm{d}\nu_1}{\mathrm{d}\bar{\nu}}(1) = \frac{1}{3}(\frac{\mathrm{d}\mu_1}{\mathrm{d}\bar{\mu}}(1) + \frac{\mathrm{d}\mu_1}{\mathrm{d}\bar{\mu}}(3) + \frac{\mathrm{d}\mu_1}{\mathrm{d}\bar{\mu}}(4))$*, and analogously for the other coordinate.*

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* $\|\boldsymbol{\lambda}\|_1 = 1$*, and define* $\mu := \boldsymbol{\lambda}^\top \boldsymbol{\mu}$*, and* $\nu := \boldsymbol{\lambda}^\top \boldsymbol{\nu}$*. Assume that* $\mu_j \ll \mu$ *and* $\nu_j \ll \nu$ *for each* $1 \le j \le d$*. Then,*

(i) *The set* $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ *is nonempty 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 nonempty 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) $\Rightarrow$ (ii) there. $\square$

We briefly describe the intuition behind this result, which is crucial for our paper. 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. The key observation is that 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 1. The "ratios" can be recognized as Radon–Nikodym derivatives. The "mix" effect can be interpreted as a backward martingale transport, because reversing the transport arrows (or equivalently, looking at the transport 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[2] parity that relates the simultaneous transport to the underlying backward martingale transport, which will be useful for our purpose when constructing explicitly an e/p-variable. We state a weak form of the

[2]Here, MOT stands for martingale optimal transport, and SOT stands for simultaneous optimal transport.

MOT-SOT parity below, which can be proved similarly to Corollary 3 of Wang and Zhang (2023).

PROPOSITION 2.4. *Let $\boldsymbol{\mu} \in \Pi(\mathfrak{X})^d$ and $\boldsymbol{\nu} \in \Pi(\mathfrak{Y})^d$ satisfy $\boldsymbol{\mu} \ll \mu_d$, $\boldsymbol{\nu} \ll \nu_d$ (where we recall that μ_d is the dth component of the vector-valued measure $\boldsymbol{\mu}$), and $\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ nonempty. 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 $T \in \mathcal{T}(\boldsymbol{\mu}, \boldsymbol{\nu})$ satisfying*

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

In the above proposition, we have picked the dth entry μ_d, ν_d to evaluate the Radon–Nikodym derivatives. One could as well use μ_j, ν_j for any $1 \le j \le d$, or even $\bar{\mu}$, $\bar{\nu}$. When applying this result, we have in mind that the last entry of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ will be given by the alternative and the rest by the null, which makes it convenient to evaluate the Radon–Nikodym derivatives using the dth entry.

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*

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

PROOF. Since U_1 is atomless, $\mathcal{T}(\mathrm{U}_1, \tau) \neq \varnothing$.[3] 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 $1 \le i \le d$. This is well-defined since f_i is nonnegative a.e. and $\mathbb{E}^{\mathrm{U}_1}[f_i] = 1$ for each $1 \le i \le d$. □

3. 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 6, studying this setting first helps with building intuition behind our proof techniques. Moreover, the concept of e-power studied in Section 4 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 will assume that

(JA) $$(\mathcal{P}, \mathcal{Q}) \quad \text{is jointly atomless,}$$

unless otherwise stated. The main results are Theorems 3.1 and 3.4 below. When

(AC) $$P_1, \dots, P_L \ll Q$$

holds, we define the measure $\gamma = (\mathrm{d}P_1/\mathrm{d}Q, \dots, \mathrm{d}P_L/\mathrm{d}Q)|_Q$ on $\mathbb{R}^L$.

THEOREM 3.1. *Suppose that we are testing $\mathcal{P} = \{P_1, \dots, P_L\}$ against $\mathcal{Q} = \{Q\}$ and (JA) holds. The following are equivalent:*

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

[3] It is a standard fact in optimal transport that a Monge transport map from μ to ν exists if μ is atomless.

(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* $Q \notin \mathrm{Span}(P_1, \dots, P_L)$.

To prove Theorem 3.1, we need the following preparation.

LEMMA 3.2. *Suppose that* $Q \notin \mathrm{Span}(P_1, \dots, P_L)$ *and* (AC) *holds. 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 H.1(ii) of the Supplementary Material (Zhang, Ramdas and Wang (2024)), abbreviated as the SM henceforth, $\mathbf{1} = \mathrm{bary}(\gamma) \in (\mathrm{Conv}\,\mathrm{supp}\,\gamma)^\circ$. 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 $1 \le j \le k$. 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 3.3. *We have* $Q \notin \mathrm{Span}(P_1, \dots, P_L)$ *if and only if there exist probability measures* $G \neq F$ *such that*

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

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

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

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

PROOF. The "if" is clear since $Q \in \mathrm{Span}(P_1, \dots, P_L)$ would imply $G \in \mathrm{Span}(F) = \{F\}$. For "only if", let $F = \mathrm{U}_1$ and consider first the case where (AC) holds. Then using Proposition 2.3 with $d = L + 1$, it suffices to prove that there exists some $G \gg F$ such that

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

Equivalently, we need to show that

$$\gamma = \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}\right)\bigg|_Q \succeq_{\mathrm{cx}} \left(\frac{\mathrm{d}F}{\mathrm{d}G}, \dots, \frac{\mathrm{d}F}{\mathrm{d}G}\right)\bigg|_G. \tag{2}$$

We will first consider a special type of density $\mathrm{d}F/\mathrm{d}G$ which allows us to construct G such that (2) holds. Suppose that

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

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 $\{\mathbf{x} \in \mathbb{R}^L \mid x_1 = \cdots = x_L \in$

$[(1+\varepsilon)^{-1},(1-\varepsilon)^{-1}]\}$, 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 (2).

Consider a disjoint collection of closed balls $\{B_j\}_{1\le j\le k}$ in $\mathbb{R}^L$ as constructed in Lemma 3.2. By Lemma H.2 of the SM, there is $\delta > 0$ and a segment $\{\mathbf{x} \in \mathbb{R}^L \mid x_1 = \cdots = x_L \in [1-\delta, 1+\delta]\}$ 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 $\widetilde{\gamma} \le \gamma|_{\bigcup_{j=1}^k B_j}$. We choose $\varepsilon > 0$ so that $(1-\varepsilon)^{-1} < 1+\delta$. As a result, the measure G constructed in the above paragraph satisfies

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

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 (AC) does not hold, then we define $Q' = Q/2 + (P_1 + \cdots + P_L)/(2L)$, 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 $G = 2G' - F \neq F$. □

PROOF OF THEOREM 3.1. The direction (a) ⇒ (b) is proved as Proposition A.5 of the SM, (b) ⇒ (c) is clear from definition, (c) ⇒ (e) is proved as Proposition A.6 of the SM, and (e) ⇒ (d) is Proposition 3.3. To show (d) ⇒ (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 given in Lemma H.3 of the SM. It follows immediately that $\phi \circ X$ is an exact p-variable. □

THEOREM 3.4. *Suppose that we are testing* $\mathcal{P} = \{P_1, \dots, P_L\}$ *against* $\mathcal{Q} = \{Q\}$ *and* (JA) *holds. The following are equivalent*:

(a) *there exists a nontrivial p-variable*;
(b) *there exists a bounded e-variable that has nontrivial e-power against* $\mathcal{Q}$;
(c) *there exists an e-variable that is nontrivial for* $\mathcal{Q}$;
(d) *it holds that* $Q \notin \mathrm{Conv}(P_1, \dots, P_L)$.

REMARK 3.5. The directions (c) ⇔ (e) in Theorem 3.1 and (c) ⇔ (d) in Theorem 3.4 also hold without (JA), in view of Proposition A.7 of the SM.

EXAMPLE 3.6.

(i) Let $P_1 \overset{\text{law}}{\sim} \mathrm{Ber}(0.1)$, $P_2 \overset{\text{law}}{\sim} \mathrm{Ber}(0.2)$, and $Q \overset{\text{law}}{\sim} \mathrm{Ber}(0.3)$. It follows that $Q \in \mathrm{Span}(P_1, P_2) \setminus \mathrm{Conv}(P_1, P_2)$. By Theorems 3.1 and 3.4, a nontrivial e-variable (or p-variable) exists, but an exact nontrivial e-variable (or p-variable) does not exist.

(ii) Let $P_1 \overset{\text{law}}{\sim} \mathrm{N}(-1, 1)$, $P_2 \overset{\text{law}}{\sim} \mathrm{N}(1, 1)$, and $Q \overset{\text{law}}{\sim} \mathrm{N}(0, 1)$. By Theorem 3.1, there exists a pivotal exact nontrivial e-variable (or p-variable).

REMARK 3.7. When the sample space $\mathfrak{X}$ is finite (say $|\mathfrak{X}| = d$) and $Q \notin \mathrm{Span}\mathcal{P}$, it is easy to construct nontrivial exact e-variables. We can associate the distributions with their Radon–Nikodym derivatives, which are just d-dimensional vectors, and one can consider an e-variable of the form $1 + Y$ where Y is proportional to the orthogonal part of Q relative to the span of $\mathcal{P}$ (so that Y integrates to zero under any $P \in \mathcal{P}$, but has positive expectation under Q). In case $\mathcal{P} = \{P\}$, taking P as the reference measure, this construction yields $Y = \alpha(\mathrm{d}Q/\mathrm{d}P - 1)$ for any $\alpha \in [0, 1]$, and the e-variable is precisely $\mathrm{d}Q/\mathrm{d}P$ when $\alpha = 1$. For infinite $\mathfrak{X}$, such a direct construction exploiting orthogonality is no longer possible because the Radon–Nikodym derivatives do not live in a Hilbert space.

Next, Section 4 constructs a powerful exact e-variable by additionally imposing pivotality.

4. Constructing a powerful exact 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), Breiman (1960), Bell and Cover (1988), Shafer et al. (2011), Grünwald, de Heide and Koolen (2024), Waudby-Smith and Ramdas (2024).[4] It has been recently called the *e-power* of X (Vovk and Wang (2024)), a term we continue to use for simplicity. In this section, we will fix $\mathcal{P} = \{P_1, \dots, P_L\}$ and $\mathcal{Q} = \{Q\}$. Our goal is to solve

$$\text{(3)} \qquad \begin{aligned} \max \quad & \mathbb{E}^Q[\log X], \\ \text{s.t.} \quad & X : X \text{ is a pivotal exact e-variable.} \end{aligned}$$

This optimization problem turns out to be a special case of a more general problem that is illustrated by (8) below. Such a connection will be explained in Section 4.1. We describe an equivalent condition for the existence of a maximal element for (8) in Section 4.2. A further sufficient condition in the case $L = 2$ is illustrated in Section 4.3. Section 4.4 contains a few discussions regarding batching multiple data points and how it affects the e-power. Finally, we provide several examples in Section 4.5. In this section, we let

$$\text{(4)} \qquad \gamma := \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \dots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}\right)\bigg|_Q.$$

In particular, γ is a probability measure on $\mathbb{R}_+^L$ with mean $\mathbf{1}$.

4.1. *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 4.1. 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. Thus (3) reduces to

$$\text{(5)} \qquad \begin{aligned} \max \quad & \mathbb{E}^Q[\log X], \\ \text{s.t.} \quad & X : X \geq 0, \quad \mathbb{E}^P[X] = 1. \end{aligned}$$

By Gibbs' inequality, the maximum value is attained by the likelihood ratio, that is, when $X = \mathrm{d}Q/\mathrm{d}P$ (see Shafer (2021) for this simple setting).

Below we illustrate the solution to (5) using our theory, which sheds light on the composite null case. For simplicity, we assume (JA) and (AC). Denote by $\gamma := (\mathrm{d}P/\mathrm{d}Q)|_Q$. Consider the set $\mathcal{M}_\gamma$ of probability measures μ such that $\mu \preceq_{\mathrm{cx}} \gamma$:

$$\mathcal{M}_\gamma := \{\mu \in \Pi(\mathbb{R}) : \mu \preceq_{\mathrm{cx}} \gamma\}.$$

[4] In short, it captures the rate of growth of the test martingale under the alternative Q; see Section 7.

Using Lemma 2.5, every $\mu \in \mathcal{M}_\gamma$ (in fact also for $\mu \notin \mathcal{M}_\gamma$) 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. Next, consider X of the form $X = (\mathrm{dU}_1/\mathrm{d}F)(Y)$, and we optimize $\mathbb{E}^Q[\log X]$ over F satisfying $(\mathrm{d}F/\mathrm{dU}_1)|_{\mathrm{U}_1} \in \mathcal{M}_\gamma$. It is clear that the constraint

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

is satisfied, and the objective in (5) becomes

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

We have thus arrived at the optimization problem

$$\begin{aligned} &\max \quad \mathbb{E}^{\mathrm{U}_1}\left[-\log\frac{\mathrm{d}F}{\mathrm{dU}_1}\right], \\ &\text{s.t.} \quad F \in \Pi(\mathbb{R}) : F \ll \mathrm{U}_1, \quad \frac{\mathrm{d}F}{\mathrm{dU}_1}\bigg|_{\mathrm{U}_1} \in \mathcal{M}_\gamma. \end{aligned} \tag{6}$$

The value (6) gives a lower bound on (5). Since the set $\mathcal{M}_\gamma$ has a maximum element γ in convex order, the problem (6) has a trivial solution $\mathbb{E}^Q[-\log(\mathrm{d}P/\mathrm{d}Q)]$. This corresponds to the solution to (5) using Gibbs' inequality.

The fact that the two values (5) and (6) are the same is not a coincidence and holds more generally for composite nulls, which we will prove in Theorem 4.2. With a composite null, the main difficulty arises from solving (6), because the set $\mathcal{M}_\gamma$ has a complicated structure, and may not contain a maximum element in convex order.

As explained in Example 4.1, the first step to solving (3) 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 (3), as shown in the following result.

THEOREM 4.2. *Assume* (JA) *and* (AC). *There exists a maximizer X to* (3) *of the form $X = (\mathrm{d}G/\mathrm{d}F)(Y)$, where $F, G \in \Pi(\mathbb{R})$, and $Y \in \mathcal{T}((P_1, \ldots, P_L, Q), (F, \ldots, F, G))$.*

The fact that the log-optimal pivotal and exact e-variable is a likelihood ratio is quite aesthetically appealing, a phenomenon that is known to be true without the restrictions of pivotality and exactness Grünwald, de Heide and Koolen (2024), Larsson, Ramdas and Ruf (2024), but in this more general case F could be a *sub*-probability distribution.

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

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

As a consequence of Proposition 2.3, the optimization problem (3) is equivalent to finding

$$\begin{aligned} &\max \quad \mathbb{E}^G\left[-\log\frac{\mathrm{d}F}{\mathrm{d}G}\right], \\ &\text{s.t.} \quad F, G \in \Pi(\mathbb{R}) : \left(\frac{\mathrm{d}F}{\mathrm{d}G}, \ldots, \frac{\mathrm{d}F}{\mathrm{d}G}\right)\bigg|_G \preceq_{\mathrm{cx}} \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \ldots, \frac{\mathrm{d}P_L}{\mathrm{d}Q}\right)\bigg|_Q. \end{aligned} \tag{7}$$

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 smaller than γ in convex order, and our goal is to

$$\begin{aligned} \max \quad & \mu \quad \text{in } \preceq_{\mathrm{cx}}, \\ \text{s.t.} \quad & \mu \in \mathcal{M}_\gamma. \end{aligned} \tag{8}$$

This will be the goal of the present section. The reader should keep in mind that unfortunately, even if (8) allows a unique maximum element, it does not necessarily solve (3) uniquely when the logarithm in (3) is replaced by other concave functions. This is because Theorem 4.2 requires Gibbs' inequality, where the logarithm plays a crucial role.

4.2. *Existence of the maximum element in convex order.* To ease our presentation, we will assume further that

(N) γ from (4) does not give positive mass to any hyperplane in $\mathbb{R}^L$.

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 4.6 and 4.8), and we expect that analogous results hold without such an assumption.

PROPOSITION 4.3. *Let γ be a probability measure on $\mathbb{R}_+^L$ with mean $\mathbf{1}$. Consider $x \geq 0$. 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 $\mathcal{I}^+$.*

In this case, we call $\partial\mathbb{H}_x$ a separating hyperplane at x. Moreover, if (N) holds, there exists a unique measure μ_x satisfying the above conditions, in which case it also holds that $\mu_x = \gamma|_{\mathbb{H}_x}$ and $\mu_{\mathbb{H}_x^c} = \gamma|_{\mathbb{H}_x^c}$.

We remark that if γ has a strictly positive density on $\mathbb{R}_+^L$, then the above $\mathbb{H}_x$ is unique.
Recall from (8) that our goal is to find the maximum element in $\mathcal{M}_\gamma$ in convex order.

THEOREM 4.4. *Assuming (N), the following are equivalent.*

(a) *There exists a unique maximum element μ in convex order in $\mathcal{M}_\gamma$, that is, $\mu \preceq_{\mathrm{cx}} \gamma$ 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 $\{\mu_x\}_{x\geq 0}$ from Proposition 4.3 is monotone (in the usual order), that is, for all $x \leq y$, $\mu_x \leq \mu_y$.*

EXAMPLE 4.5. Suppose that $L = 1$. It is clear from the proof of Proposition 4.3 that condition (b) in Theorem 4.4 is always satisfied. Therefore, the maximum element μ in $\mathcal{M}_\gamma$ always exists. This agrees with Example 4.1, where the likelihood ratio maximizes the e-power.

REMARK 4.6. The only place we used our assumption (N) is on the uniqueness of the measure μ_x in Proposition 4.3. When there is no uniqueness, the condition (b) in Theorem 4.4 needs to be replaced by the existence of a monotone selection of measures $\{\mu_x\}_{x\geq 0}$, each of them satisfying the conditions in Proposition 4.3.

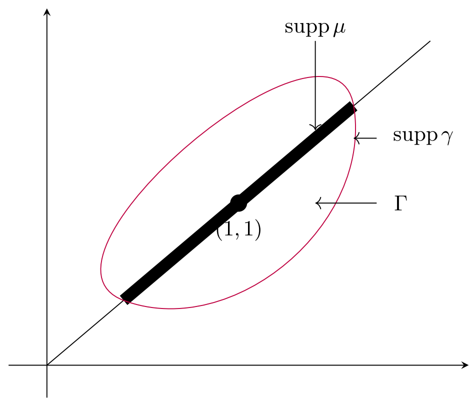

FIG. 2. *Illustration of Theorem 4.7. The convex set* Γ *is enclosed by the red contour* $\partial\Gamma$ *on which* γ (*the law of* $(\mathrm{d}P_1/\mathrm{d}Q, \mathrm{d}P_2/\mathrm{d}Q)$ *under* Q) *is supported. The measure* μ *is supported on the thick segment on the diagonal* $\mathcal{I}$.

The condition (b) in Theorem 4.4 is in general not easy to check, especially in higher dimensions.[5] Later, we supply a sufficient condition in Section 4.3, and a few examples in Section 4.5.

4.3. *A sufficient condition in case* $|\mathcal{P}| = 2$. When $L = 2$, we provide a sufficient condition for the class of measures $\{\mu_x\}_{x\geq 0}$ to satisfy the monotonicity condition $x \leq y \implies \mu_x \leq \mu_y$. In view of Theorem 4.4, this condition implies the existence of the maximum element μ. We keep the same setting as in Section 4.2 and assume (N), with the exception that $L = 2$.

THEOREM 4.7. *Assume* (N), (JA), (AC). *Suppose that there exists a convex set* $\Gamma \subseteq \mathbb{R}^2$ *such that* $\gamma(\partial\Gamma) = 1$.[6] *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 4.3 applied with* $L = 2$. *In particular, there exist distinct measures* $F, G \in \Pi(\mathbb{R})$ *such that* $(\mathrm{d}F/\mathrm{d}G, \mathrm{d}F/\mathrm{d}G)|_G = \mu$, *attaining the maximum in* (7).

A pictorial illustration of Theorem 4.7 is given by Figure 2.

REMARK 4.8. With essentially the same arguments, we may remove assumption (N) from Theorem 4.7. With the presence of atoms, selecting any monotone collection $\{\mu_x\}_{x\geq 0}$ would be enough; see Remark 4.6.

4.4. *On multiple observations.* Before we proceed, let us discuss the case with multiple data points. Suppose that instead of one data point, we observe n i.i.d. 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 for $\mathcal{P}^n := \{P_1^n, \dots, P_L^n\}$ that is pivotal, exact, and has nontrivial e-power against $\mathcal{Q}^n := \{Q^n\}$. We first see that, as long as $Q \notin \mathcal{P}$ and $\mathcal{P}$ is linearly independent, at most two observations are needed to build a pivotal and exact e-variable based on Theorem 3.1. Furthermore, without linear independence of $\mathcal{P}$, a finite number of observations would suffice when the underlying space $\mathfrak{X}$ is Euclidean.

[5] In this paper when we mention "dimension" we typically refer to the dimension of the null, but not the dimension of the underlying space $\mathfrak{X}$.

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

THEOREM 4.9. *Suppose that $\mathfrak{X}$ is an Euclidean space and $P_1, \dots, P_L$ are distinct probability measures on $\mathfrak{X}$. If $Q \in \Pi(\mathfrak{X})$ satisfies $Q \notin \mathcal{P} = \{P_1, \dots, P_L\}$, then there exists $k \geq 1$ such that $Q^k \notin \mathrm{Span}\mathcal{P}^k$ (and in particular $Q^k \notin \mathrm{Conv}\mathcal{P}^k$). Moreover, if we also assume that Q satisfies (AC) and that $P_1, \dots, P_L$ are linearly independent, then either $Q \notin \mathrm{Span}\mathcal{P}$ or $Q^2 \notin \mathrm{Span}\mathcal{P}^2$ (or both); in particular, either $Q \notin \mathrm{Conv}\mathcal{P}$ or $Q^2 \notin \mathrm{Conv}\mathcal{P}^2$ (or both).*

In the last claim above, one can show that neither the linear independence condition nor (AC) can be removed. The proof of Theorem 4.9 is put in Section C of the SM, which relies on the following fundamental fact: If $\mathfrak{X}$ is an Euclidean space and $P_1, \dots, P_L$ are distinct probability measures on $\mathfrak{X}$, then there exists $k \geq 1$ (possibly large) such that $P_1^k, \dots, P_L^k$ are linearly independent. This fact may be known, but we are not aware of a proof in the literature, and we present it as Lemma C.3 in the SM. The weaker statement that there exists k for which $Q^k \notin \mathrm{Conv}\mathcal{P}^k$ also follows from Lemma 2 of Berger (1951).

EXAMPLE 4.10. Suppose that $P_1 \overset{\text{law}}{\sim} \mathrm{Ber}(0.1)$, $P_2 \overset{\text{law}}{\sim} \mathrm{Ber}(0.2)$, and $Q \overset{\text{law}}{\sim} \mathrm{Ber}(0.3)$. We explained in Example 3.6 that an exact nontrivial e-variable does not exist. Nevertheless, Theorem 4.9 implies that $Q^2 \notin \mathrm{Span}(P_1^2, P_2^2)$, and hence an exact and nontrivial e-variable exists for a batch of two data points. An example of such an e-variable X is given by

$$X(\omega) \approx \begin{cases} 1.009 & \text{for } \omega = (0, 0); \\ 0.939 & \text{for } \omega = (0, 1), (1, 0); \\ 1.338 & \text{for } \omega = (1, 1). \end{cases}$$

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 (3).

PROPOSITION 4.11. *In the setting above, suppose that (AC) and (JA) hold, and $Q \notin \mathcal{P} = \{P_1, \dots, P_L\}$. For any $n, m \in \mathbb{N}$, $\ell_{n+m} \geq \ell_n + \ell_m$. In particular, ℓ_n/n converges to a positive limit less than or equal to $\min_{P \in \mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$.*

The above proposition formalizes the straightforward observation that constructing an e-value using $m + n$ points is potentially more powerful than multiplying two e-values together that were constructed separately using m and n points, respectively.

Is there a loss of e-power caused by imposing exactness or pivotality?. The superadditivity property established in Proposition 4.11 implies in particular that $\ell_{2^n}/2^n$ is increasing in n. The intuitive reason of the increase in the average e-power is partly due to the fact that the pivotality constraint becomes less restrictive for a higher number of observations. To see this, imagine laws $P \in \mathcal{P}$ with a complicated entangled overlapping structure. To achieve pivotality, we need to send all laws P simultaneously to a single distribution F, the ways of which may be quite limited due to the overlapping structure.[7] On the other hand, with multiple observations, the laws P^n, $P \in \mathcal{P}$ have much fewer overlapping parts than $P \in \mathcal{P}$ do (for instance, P^∞, $P \in \mathcal{P}$ are mutually singular), meaning that there are more ways to achieve pivotality.

It remains an open question whether $\ell_n/n \to \min_{P \in \mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$. Note that this conjectural limit can be different from $\min_{P \in \mathrm{Conv}\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$ because the linearity structure is lost after taking powers. If the above question is answered in the affirmative,

[7]For instance, if $P \in \mathcal{P}$ all have disjoint support (no overlap), the transport map can be picked independently on the disjoint supports to send P to F, but this is not possible of $P \in \mathcal{P}$ all have the same support.

then the loss of e-power vanishes asymptotically, by noting that $n \min_{P\in\mathcal{P}} \mathbb{E}^Q[\log(\mathrm{d}Q/\mathrm{d}P)]$ is an upper bound on the theoretical best e-power for testing $\mathcal{P}^n$ against $\mathcal{Q}^n$ (see Example 4.1). In Example 4.13, we present a setting of Gaussian 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.[8]

4.5. *Examples.* The condition in Theorem 4.7 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 4.12. Consider $P_1 \overset{\text{law}}{\sim} \mathrm{N}(-1, 1)$, $P_2 \overset{\text{law}}{\sim} \mathrm{N}(1, 1)$, and $Q \overset{\text{law}}{\sim} \mathrm{N}(0, 1)$. It follows from a direct computation that

$$\gamma = \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}, \frac{\mathrm{d}P_2}{\mathrm{d}Q}\right)\bigg|_Q = \left(e^{-\xi-1/2}, e^{\xi-1/2}\right)\big|_{\xi \overset{\text{law}}{\sim} \mathrm{N}(0,1)}, \tag{9}$$

which is supported on the hyperbola $\{(x_1, x_2) \in \mathbb{R}_+^2 \mid x_1x_2 = 1/e\}$, the boundary of the convex set $\{(x_1, x_2) \in \mathbb{R}_+^2 \mid x_1x_2 \geq 1/e\}$; see Figure 3 for an illustration. By Theorem 4.7, there exists a unique maximal element μ in $\mathcal{M}_\gamma$ in convex order.

Using the notation from Proposition 4.3, it is easy to see that $\mathbb{H}_x = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leq 2x\}$ and $\mu_x = \gamma|_{\mathbb{H}_x}$. Moreover, Theorem 4.7 yields that μ is the unique probability measure on $\mathcal{I}^+$ with

$$\mu([0, x]^2) = 2\Phi\left(\log\left(\sqrt{e}x + \sqrt{ex^2 - 1}\right)\right) - 1 \quad \text{for } x \geq \frac{1}{\sqrt{e}}, \tag{10}$$

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 4.1 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 i.i.d. 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 (7).

EXAMPLE 4.13. We consider the setting in Example 4.12 but instead of one data point, we observe n i.i.d. 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 4.12. In this setting, $P_1 = \mathrm{N}(-\mathbf{1}_n, I_n)$, $P_2 =$

[8]The argument in Example 4.13 is analytical. On the other hand, numerical verification of this conjecture remains a challenging task due to drastic extremal values of the Radon–Nikodym derivatives (in high dimensions, almost all mass of $(\mathrm{d}P^n/\mathrm{d}Q^n)|_{Q^n}$ concentrates near 0 or ∞), exponential time complexity, and the slow convergence of ℓ_n/n. We leave it as an open problem to design a more efficient iterative algorithm for this problem (or more generally, computing numerically the best e-power in high dimensions), or to prove that one cannot exist.

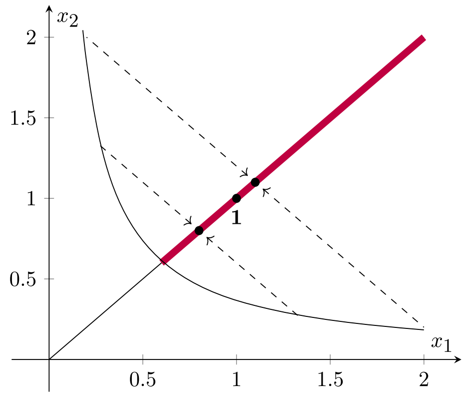

FIG. 3. *An illustration of Example 4.12: γ is supported on the hyperbola $x_1x_2 = e^{-1}$, the optimal μ is supported on the red ray. Dashed arrows indicate the reduction of filtration.*

$\mathrm{N}(\mathbf{1}_n, I_n)$, 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 a 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 (9). Using a similar argument as in Example 4.12, 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 4.12 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 $E_n \geq E_n^*$ due to the 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 5, we will compare the e-power of the two approaches numerically, and in Section 7, 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 4.11 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 4.14. 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 4.12.

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

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

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 $a_i : \mathbb{R} \to \mathbb{R},\ 0 \le i \le 2$.

In Section D of the SM, we show how to instantiate and employ Theorem 4.4 to an instructive example where $\mathcal{P} = \{P_1, P_2, P_3\}$ and $\mathcal{Q} = \{Q\}$, where $P_1 \overset{\text{law}}{\sim} \mathrm{N}((0,1), I)$, $P_2 \overset{\text{law}}{\sim} \mathrm{N}((-\sqrt{3}/2, -1/2), I)$, $P_3 \overset{\text{law}}{\sim} \mathrm{N}((\sqrt{3}/2, -1/2), I)$, and $Q \overset{\text{law}}{\sim} \mathrm{N}((0,0), I)$, as well as some other simple examples (such as distributions with bounded support and Q being uniform).

5. 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 4.3, which is the key to our construction). Unless otherwise stated, we follow the setup of Section 4.

The first goal of the SHINE construction is to solve the optimization problem (8). In the case where the condition in Theorem 4.4 is satisfied, the construction outputs the maximum element. When the maximum element μ does not exist or when the condition (b) in Theorem 4.4 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 5.1 and 5.2. Relevant simulation results will be provided in Section F of the SM.

5.1. *Description of the SHINE construction.* Start with $\mu^{(0)} = \delta_{\mathbf{1}}$, $x_1^{(0)} = \mathbf{1}$, and $\mu_1^{(0)} = \gamma$ from (4). At step $s \ge 0$, we are given $\mu^{(s)}$, $\{x_k^{(s)}\}_{1\le k\le 2^s}$, and $\{\mu_k^{(s)}\}_{1\le k\le 2^s}$. For each k, we apply Proposition 4.3 to the sub-probability measure $\mu_k^{(s)}$ at the point $x_k^{(s)}$. This yields a unique decomposition of $\mu_k^{(s)}$ into two measures, each having a barycenter on $\mathcal{I}^+$. Denote them by $\mu_{2k-1}^{(s+1)}$ and $\mu_{2k}^{(s+1)}$. For $1 \le k \le 2^{s+1}$, define $x_k^{(s+1)} = \mathrm{bary}(\mu_k^{(s+1)})$. Finally, let $\mu^{(s+1)}$ be the probability measure having mass $\mu_k^{(s+1)}(\mathbb{R}^L)$ on $x_k^{(s+1)}$ for every k, that is,

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

The output of the SHINE construction at step s is the measure $\mu^{(s)}$. We refer to Figure 4 for an illustration in dimension $L = 2$.

It is easy to see that each $\mu^{(s)}$ is centered at $\mathbf{1}$ and supported on $\mathcal{I}^+$. Moreover, $\mu^{(s)} \preceq_{\mathrm{cx}} \gamma$ by Strassen's theorem because $\mu^{(s)}$ is the aggregation of barycenters of different components in the decomposition of γ. By Markov's inequality, the sequence $\{\mu^{(s)}\}$ is tight and allows a weak limit. In fact, an even stronger assertion can be made. Define $\{X_s\}$ as the coupling of the first coordinate of $\{\mu^{(s)}\}$ such that $X_0 = 1$ and at each $s \ge 0$, for $j = 2k-1, 2k$,

$$\mathbb{P}\big(X_{s+1} = x_j^{(s+1)} \big| X_s = x_k^{(s)}\big) = \frac{\mu_j^{(s+1)}(\mathbb{R}^L)}{\mu_{2k-1}^{(s+1)}(\mathbb{R}^L) + \mu_{2k}^{(s+1)}(\mathbb{R}^L)}. \tag{12}$$

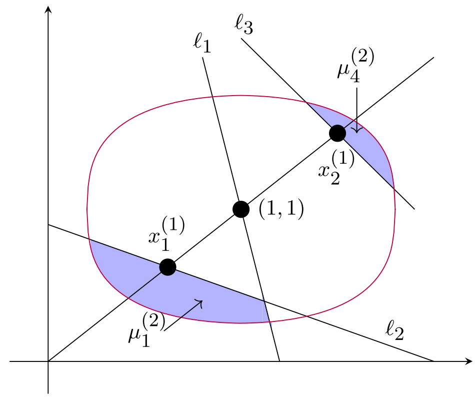

FIG. 4. *An illustration of the SHINE construction in dimension $L = 2$. Suppose that the measure γ is supported on the region enclosed by the red contour, where $\mathrm{bary}(\gamma) = (1, 1)$. In the first step of the SHINE construction, we use Proposition 4.3 to find a line ℓ_1 through $(1, 1)$ that 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.*

By construction,

$$x_{2k-1}^{(s+1)} \mu_{2k-1}^{(s+1)}(\mathbb{R}^L) + x_{2k}^{(s+1)} \mu_{2k}^{(s+1)}(\mathbb{R}^L) = x_k^{(s)} \mu_k^{(s)}(\mathbb{R}^L) = x_k^{(s)} \big(\mu_{2k-1}^{(s+1)}(\mathbb{R}^L) + \mu_{2k}^{(s+1)}(\mathbb{R}^L)\big).$$

It can thus be checked by direct calculation that $\mathbb{E}[X_{s+1} \mid X_s = x_k^{(s)}] = x_k^{(s)}$, meaning that $\{X_s\}$ forms a nonnegative martingale, and hence converges a.s. to some X_∞ by the martingale convergence theorem. We call $\{X_s\}$ the SHINE martingale (associated with γ). 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 5.1. The first step of the construction, that is, after finishing step $s = 0$, already contains a proof of Proposition 3.3, because $\delta_{\mathbf{1}} \neq \mu^{(1)} \preceq_{\mathrm{cx}} \gamma$. Nevertheless, the ideas behind the original proof of Proposition 3.3 extend to the composite alternative scenario.

EXAMPLE 5.2. Suppose that $L = 1$, that is, we have simple null P versus simple alternative Q, where $P \approx Q$. In this case, Proposition 4.3 applies trivially: for each $s \geq 0$ and $1 \leq k \leq 2^s$, the measure $\mu_k^{(s)}$ is decomposed into

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

As in (11), this results in a sequence of laws $\{\mu^{(s)}\}_{s \geq 0}$ on $\mathbb{R}$ that are increasing and smaller than γ in convex order. This is closely related to a martingale decomposition theorem by Simons (1970): if we denote by $\{Z_s\}_{s \geq 0}$ the natural martingale coupling of $\{\mu^{(s)}\}_{s \geq 0}$, then $Z_s \to Z$ a.s. for some Z that has law γ. In other words, the e-variable obtained from the SHINE construction converges to $\mathrm{d}Q/\mathrm{d}P$ a.s. under both P and Q.

THEOREM 5.3. *Assume (JA) and (AC). For any s, we have $\mu^{(s)} \preceq_{\mathrm{cx}} \mu^{(s+1)}$, and if $\mu^{(s)} \neq \mu^{(s+1)}$, then the inequality is strict, meaning that the above SHINE construction makes progress at each step. Further, assuming (N), it produces a sequence of measures that converges almost surely to a maximal element μ in convex order in $\mathcal{M}_\gamma$. In this case, if there exists a maximum element μ_0, then the output of our SHINE construction converges to μ_0.*

When we apply the construction in practice, we need to stop at finitely many steps, so we will not always obtain an exactly maximal element. Later in Section 5.3, we show that the e-power from the kth step in SHINE converges exponentially to the optimal value produced by SHINE, with a rate that can be made explicit given mild moment conditions.

We note in particular that Lemma E.2 of the SM together with Theorem 5.3 yield that the construction always gives an atomless measure μ in the limit.

With the presence of atoms, the decomposition given by Proposition 4.3 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.

5.2. *Recovering explicitly an e/p-variable.* We aim first to recover our e-variable X, which we recall from Theorem 4.2 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 from (11) and (12) that at the sth step, our construction leads to a canonical martingale coupling of $\mu^{(s)}$ and γ that couples the mass $\mu_k^{(s+1)}(\mathbb{R}^L)\delta_{x_k^{(s+1)}}$ with $\mu_k^{(s+1)}$. We denote the martingale coupling by (Λ_s, Γ_s), which is a random vector of dimension $2L$. Under assumption (N), we know further that the measures $\{\mu_k^{(s)}\}_{1\le k\le 2^s}$ are mutually singular, and hence (Λ_s, Γ_s) is backward Monge, that is, in the backward direction we have $\Lambda_s = h(\Gamma_s)$ 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 $x \in \mathfrak{X}$,

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

This leads to

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

For example, the sth step of the construction gives explicitly

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

Note that the measures F, G can meanwhile be reconstructed from Lemma 2.5, and further Lemma H.4(i) of the SM 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 5.4. Suppose that we are in the setting of Example 4.12, with $P_1 \overset{\text{law}}{\sim} \mathrm{N}(-1, 1)$, $P_2 \overset{\text{law}}{\sim} \mathrm{N}(1, 1)$, and $Q \overset{\text{law}}{\sim} \mathrm{N}(0, 1)$. Recall (9). By symmetry of γ, it is clear that the separating hyperplanes $\mathbb{H}_x$ in the SHINE construction are given by $\mathbb{H}_x = \{(a, b) : a + b \le 2x\}$. 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 $\operatorname{bary}(\gamma|_{\mathbb{H}_1}) \approx 0.713 \times \mathbf{1}$ and $\operatorname{bary}(\gamma|_{\mathbb{H}_1^c}) \approx 1.743 \times \mathbf{1}$. Using (13), the corresponding e-variable has the form

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

The resulting e-power $\mathbb{E}^Q[\log X]$ is approximately 0.089. (One may compare this to the maximum e-power 0.12543, which can be directly computed from (10).) In general, we may construct X in multiple steps.

5.3. *Convergence rate of SHINE.* We complement Theorem 5.3 with the following result on the convergence rate of the e-power given by the SHINE construction. Recall from (7) that the e-power is given by $\mathbb{E}^Q[-\log X_k]$, where $\{X_k\}$ is the SHINE martingale.

THEOREM 5.5. *Assume the same conditions as in Theorem 5.3. Suppose that there exists $\varepsilon > 0$ such that*

$$\mathbb{E}^Q\left[\left(\frac{\mathrm{d}P_j}{\mathrm{d}Q}\right)^{2+\varepsilon}\right] < \infty \quad \text{for some } j, \tag{14}$$

and

$$\mathbb{E}^Q\left[\left(\frac{\mathrm{d}P_{j'}}{\mathrm{d}Q}\right)^{-2}\right] < \infty \quad \text{for some } j'. \tag{15}$$

Consider the e-power $\mathrm{EP}_k := \mathbb{E}^Q[-\log X_k]$ where $\{X_k\}$ is the SHINE martingale. Then there exist $r \in (0, 1)$ and $C > 0$ such that

$$\mathrm{EP}_\infty - \mathrm{EP}_k \le Cr^k,$$

where $\mathrm{EP}_\infty = \mathbb{E}[-\log X_\infty]$ and $X_k \to X_\infty$ a.s.

Our result relies on a particular feature of the SHINE martingale $\{X_k\}$ produced by (12). Intuitively, the martingale $\{X_k\}$ has a binary tree representation, and the legs in the tree never intersect with other legs at all levels. In this way, one gains control of the fluctuations of the martingale from its values at previous times. The key step to proving Theorem 5.5 is the following convergence rate of the L^2 Wasserstein distance. In particular, this exponential convergence applies to the Simons martingale introduced by Simons (1970); see also Example 5.2.

LEMMA 5.6. *Suppose that the SHINE martingale $\{X_k\}_{k\ge 0}$ satisfies $\mathbb{E}[|X_\infty|^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$ where $X_k \to X_\infty$ a.s. Then there exist $r = r(\varepsilon) < 1$ and a constant $C > 0$ (where C may depend on ε, the law of $|X_0 - X_1|$, and $\mathbb{E}[|X_\infty|^{2+\varepsilon}]$) such that*

$$\mathbb{E}\left[(X_k - X_\infty)^2\right] \le Cr^k.$$

If X is uniformly bounded, one can pick $r < 0.827$.

Lemma 5.6 immediately implies $\mathrm{EP}_\infty - \mathrm{EP}_k \le Cr^k$ for some $r < 0.827$ if $\mathrm{d}P_j/\mathrm{d}Q$ for each j is bounded above and bounded away from 0. The proofs of Theorem 5.5 and Lemma 5.6 are collected in Section E of the SM.

6. Composite null and composite alternative. Our goal in this section is to extend Theorems 3.1 and 3.4 to composite alternative, that is, 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.

6.1. *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 (JA) holds, we characterize equivalent conditions for the existence of an (exact and) nontrivial e-variable (or p-variable).

THEOREM 6.1. *Assume (JA). 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*, *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 every* $P \in \mathcal{P}$ *and* G_j *is the law of* X *under* Q_j *for* $1 \le j \le M$;
(e) *it holds that* $\mathrm{Span}(P_1, \dots, P_L) \cap \mathrm{Conv}(Q_1, \dots, Q_M) = \varnothing$.

Furthermore, *the equivalence of* (c) *and* (e) *does not require* (JA).

THEOREM 6.2. *Assume* (JA). *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 a bounded e-variable that has nontrivial e-power against* $\mathcal{Q}$;
(c) *there exists an e-variable that is nontrivial for* $\mathcal{Q}$;
(d) *it holds that* $\mathrm{Conv}(P_1, \dots, P_L) \cap \mathrm{Conv}(Q_1, \dots, Q_M) = \varnothing$.

Furthermore, *the equivalence of* (c) *and* (d) *does not require* (JA) *or finiteness of* $\mathcal{P}$ *and* $\mathcal{Q}$.

REMARK 6.3. The equivalence of (c) and (d) in Theorem 6.2 is a special case of Kraft's theorem, which we recall from (1). Note that here we do not require that $\mathcal{P}$ and $\mathcal{Q}$ are finite, but only the existence of a reference measure R dominating $\mathcal{P} \cup \mathcal{Q}$. To see that Kraft's theorem implies the equivalence of (c) and (d) in Theorem 6.2 in case $\mathcal{P}$ or $\mathcal{Q}$ is infinite, suppose that (c) holds. It follows that $\mathbb{E}^Q[X] > 1 \ge \mathbb{E}^P[X]$ for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$. Kraft's theorem implies the existence of some $\varepsilon > 0$ such that $d_{\mathrm{TV}}(\mathrm{Conv}\mathcal{P}, \mathrm{Conv}\mathcal{Q}) \ge \varepsilon$, and in particular, ($d$) holds. On the other hand, if (d) is true, then Kraft's theorem yields $\varepsilon > 0$ and X satisfying (1). A suitable linear transformation Y of X then satisfies $\mathbb{E}^Q[Y] > 1 \ge \mathbb{E}^P[Y]$ for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ (we may assume Y is positive since X is bounded by construction), and the rest follows from the following Proposition 6.4, which will be proved as Proposition A.2 in the SM.

PROPOSITION 6.4. *Assume the same setting as Theorem* 6.2. *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* $\mathcal{Q}$.

COROLLARY 6.5. *Suppose that we are testing* $\mathcal{P}$ *against* $\mathcal{Q}$, *where* $\mathcal{P}$ *and* $\mathcal{Q}$ *are convex polytopes in* Π. *Denote by* $\{P_1, \dots, P_L\}$ (*resp.* $\{Q_1, \dots, Q_M\}$) *the vertices of the polytope* $\mathcal{P}$ (*resp.* $\mathcal{Q}$) *and assume that* $(P_1, \dots, P_L, Q_1, \dots, Q_M)$ *is jointly atomless. Precisely the same conclusions in Theorems* 6.1 *and* 6.2 *hold.*

PROOF. This follows immediately from Proposition A.1 of the SM. □

COROLLARY 6.6. *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* $\mathcal{Q}$.

PROOF. This is a direct consequence of Theorems 6.1 and 6.2, and Proposition A.5 of the SM. □

EXAMPLE 6.7. Fix $0 < q_1 < q_2 < 1$ and let $\mathcal{P} = \{\mathrm{Ber}(q_1)\}$ and $\mathcal{Q} = \{\mathrm{Ber}(p) \mid q_2 \le p \le 1\}$. Corollary 6.5 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 $\{\mathrm{Ber}(p) \mid 0 \le p \le q_1\}$.

Due to the complication of convex order in higher dimensions, it remains a challenging task how to generalize Theorem 4.7 and the SHINE construction to the composite alternative case.

6.2. *Infinite null and alternative.* We first state a weaker version of Theorem 6.1 when both $\mathcal{P}$ and $\mathcal{Q}$ may be infinite but allow a common reference measure.

THEOREM 6.8. *Assume that there exists a common reference measure $R \in \Pi(\mathfrak{X})$ such that $P \ll R$ for $P \in \mathcal{P}$ and $Q \ll R$ for $Q \in \mathcal{Q}$. 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 with respect to the total variation distance. If $\mathcal{Q}$ is tight, then we have the further equivalence to $\overline{\mathrm{Span}}\mathcal{P} \cap \overline{\mathrm{Conv}}\mathcal{Q} = \varnothing$.*

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 $Q \in \mathcal{Q}$, $\mathbb{E}^Q[\log X] > 0$). Theorem 6.8 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 appears beyond the techniques of this paper.

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 Theorem 6.8. We expect that the theory of simultaneous transport between infinite collections of measures will be helpful.

CONJECTURE 6.9. Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are collections of probability measures on $\mathfrak{X}$ with a common reference measure. Assume also that $(\mathcal{P}, \mathcal{Q})$ is jointly atomless.[9] 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 with respect to the total variation distance.

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

PROPOSITION 6.10. *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 Theorem 6.8, 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, Rinaldo and Ramdas (2023) proved that for the set of all Gaussians (of all means and all covariances), there does not exist an e-variable with nontrivial e-power, even nonexact. 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 e-power does not exist.

We conclude this section with the following example that shows pivotal and exact p/e-values exist for a classic statistical problem. Technically, the construction below does not require any of our previous results, but it leads to a SOT of infinite dimensions.

[9] If $\mathcal{P}$ or $\mathcal{Q}$ is infinite, this can be defined in a natural way as in Definition 2.2.

EXAMPLE 6.11. Let $\mathcal{P}$ be the class of all symmetric distributions on $\mathbb{R}$ with no mass at 0, and $\mathcal{Q}$ the class of distributions Q on $\mathbb{R}$ satisfying $Q(\mathbb{R}_+) > 1/2$. A typical case in applications is to test whether the difference $Y - X$ of pre-treatment measurement X and post-treatment measurement Y is symmetric about 0. Many possible pivotal e-values for n observations can be built based on the signs, the ranks, and the sizes of the data; see Ramdas et al. (2020), Vovk and Wang (2024). For instance, with one observation, a simple e-value is given by $X(\omega) = 3/2$ if $\omega > 0$ and $X(\omega) = 1/2$ if $\omega < 0$, which is exact. Note that X is also pivotal since X simultaneously maps $\mathcal{P}$ to the uniform distribution on $\{1/2, 3/2\}$. If one allows for additional randomization using a uniform distribution on [0, 1], then

$$X(\omega, u) = \begin{cases} u/2 & \text{if } \omega > 0; \\ (u+1)/2 & \text{if } \omega < 0, \end{cases} \qquad u \in [0, 1],$$

is a nontrivial exact p-value.

7. 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 $Y \equiv (Y_t)_{t \geq 0}$ is called a *process* if it is adapted to $\mathcal{F}$, that is, if Y_t is measurable with respect to $\mathcal{F}_t$ 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, and $\mathcal{G}$ 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 $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$. 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. An integrable process M is a *martingale* for P with respect to $\mathcal{G}$ if

$$\mathbb{E}^P[M_t \mid \mathcal{G}_{t-1}] = M_{t-1} \tag{16}$$

for all $t \geq 1$. M is a *supermartingale* for P if it satisfies (16) with "=" relaxed to "$\leq$". 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 has *power one against* $\mathcal{Q}$ if $\mathbb{E}^Q[\log M_t] \to \infty$ under all $Q \in \mathcal{Q}$.

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 $Q^P \ll P$ 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: their existence 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 with respect to $\mathcal{F}$. To explain the situation, suppose that $\mathcal{P}$ contains only measures of iid sequences with marginal distributions in a set $\mathcal{P}^{\text{mar}} \subseteq \Pi(\mathfrak{X})$. Examples of the nonexistence phenomenon include the case when $\mathcal{P}^{\text{mar}}$ is the set of all mean-zero subGaussian distributions Ramdas et al. (2020), all log-concave distributions Gangrade, Rinaldo and Ramdas (2023), or all Bernoulli distributions Ramdas et al. (2022). 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 sub-Gaussian 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 nondegenerate (or power one) test (super)martingales in some $\mathcal{G} \subseteq \mathcal{F}$. 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 one should 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 i.i.d. case, where each $Z_i \overset{\text{law}}{\sim} P$ 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 7.1. *Let $\mathcal{P}^{\mathrm{mar}}$ and $\mathcal{Q}^{\mathrm{mar}}$ be subsets of $\Pi(\mathfrak{X})$ allowing for a common reference measure $R \in \Pi(\mathfrak{X})$. If $\overline{\mathrm{Conv}}\mathcal{P}^{\mathrm{mar}} \cap \overline{\mathrm{Conv}}\mathcal{Q}^{\mathrm{mar}} = \varnothing$, then there exists a test supermartingale for $\mathcal{P}$ that has power one against $\mathcal{Q}$. If $0 \notin \overline{(\overline{\mathrm{Span}}\mathcal{P}^{\mathrm{mar}} + \overline{\mathrm{Conv}}\mathcal{Q}^{\mathrm{mar}})}$, then there exists a test martingale for $\mathcal{P}$ that has power one against $\mathcal{Q}$.*

The proof is immediate from Kraft's theorem (see Remark 6.3) and Theorem 6.8, and does not require the joint nonatomicity condition (JA). 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 6.5. We can form our (super)martingale by simply multiplying these e-values for $t = 1$ (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 $\mathcal{P}$ has power one against $\mathcal{Q}$, 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, for example, presented in the subGaussian setting by Ramdas et al. (2020).)

The first (supermartingale) part of Corollary 7.1 is closely related to the main result by Grünwald, de Heide and Koolen (2024), 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. (2023).

REMARK 7.2. 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 Theorem 3.1, there does not exist a nontrivial test martingale for $\mathcal{P}$ against $\{Q^\infty\}$ with respect to the original filtration. On the other hand, if (AC) holds, then by Theorem 4.9, there exists a reduced filtration—in particular, formed by combining data points—with respect to which a nontrivial test martingale exists.

8. 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. For general null and alternative hypotheses (which are not polytopes) that allow a common reference measure, we provide a further characterization of the existence of an exact bounded e-variable that has a uniformly positive e-power.

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 some open problems along the way (see Conjecture 6.9 and Sections 4.4 and 7). For instance, it is of great interest to extend the SHINE construction to the composite alternative setting.

Acknowledgments. We thank Peter Grunwald, Martin Larsson and Johannes Wiesel for helpful discussions. We are also grateful to three referees for their thoughtful comments and for pointing towards the convergence rate of the SHINE construction.

Codes used to generate simulation and numerical results can be found at https://github.com/Hungryzzy/SHINE.

SUPPLEMENTARY MATERIAL

Supplement to "On the existence of powerful p-values and e-values for composite hypotheses" (DOI: 10.1214/24-AOS2434SUPP; .pdf). The supplementary material Zhang, Ramdas and Wang (2024) is devoted to the proofs and extra discussions. Section A contains a few general results on the existence of p/e-values, followed by proofs of our main results in Sections B–G, with the exceptions that Section D contains supplementary examples for Section 4 and Section F contains simulation results for the SHINE construction. Section H contains a few technical results that are used in our proofs.

REFERENCES

BELL, R. and COVER, T. M. (1988). Game-theoretic optimal portfolios. *Manage. Sci.* **34** 724–733. MR0943277 https://doi.org/10.1287/mnsc.34.6.724

BERGER, A. (1951). On uniformly consistent tests. *Ann. Math. Stat.* **22** 289–293. MR0042653 https://doi.org/10.1214/aoms/1177729649

BERTANHA, M. and MOREIRA, M. J. (2020). Impossible inference in econometrics: Theory and applications. *J. Econometrics* **218** 247–270. MR4149226 https://doi.org/10.1016/j.jeconom.2020.04.016

BREIMAN, L. (1960). Optimal gambling systems for favorable games. In *Proc. 4th Berkeley Sympos. Math. Statist. And Prob., Vol. I* 65–78. Univ. California Press, Berkeley. MR0135630

DARLING, D. A. and ROBBINS, H. (1967). Confidence sequences for mean, variance, and median. *Proc. Natl. Acad. Sci. USA* **58** 66–68. MR0215406 https://doi.org/10.1073/pnas.58.1.66

GANGRADE, A., RINALDO, A. and RAMDAS, A. (2023). A sequential test for log-concavity. arXiv preprint. Available at arXiv:2301.03542.

GRÜNWALD, P. D., DE HEIDE, R. and KOOLEN, W. (2024). Safe testing. *J. Roy. Statist. Soc. Ser. B*.

HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions. (The case of independent and identically distributed chance variables). *Ann. Math. Stat.* **29** 700–718. MR0095555 https://doi.org/10.1214/aoms/1177706531

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

KELLY, J. L. JR. (1956). A new interpretation of information rate. *Bell Syst. Tech. J.* **35** 917–926. MR0090494 https://doi.org/10.1002/j.1538-7305.1956.tb03809.x

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–141. MR0073896

KÜHNEL, W. (2015). *Differential Geometry*: *Curves—Surfaces—Manifolds*, 3rd ed. *Student Mathematical Library* **77**. Am. Math. Soc., Providence, RI. Translated from the 2013 German edition by Bruce Hunt, with corrections and additions by the author. MR3443721 https://doi.org/10.1090/stml/077

LARDY, T., GRÜNWALD, P. and HARREMOËS, P., (2023). Universal reverse information projections and optimal e-statistics. To appear in *IEEE Transactions on Information Theory*. Available at arXiv:2306.16646.

LARSSON, M., RAMDAS, A. and RUF, J. (2024). The numeraire e-variable and reverse information projection. arXiv preprint. Available at arXiv:2402.18810.

RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist*. *Sci*. **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. Available at arXiv:2009.03167.

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. M. (2022). Testing exchangeability: Fork-convexity, supermartingales and e-processes. *Internat*. *J*. *Approx*. *Reason*. **141** 83–109. MR4364897 https://doi.org/10.1016/j.ijar.2021.06.017

ROBBINS, H. and SIEGMUND, D. (1974). The expected sample size of some tests of power one. *Ann*. *Statist*. **2** 415–436. MR0448750

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *A* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

SHAFER, G., SHEN, A., VERESHCHAGIN, N. and VOVK, V. (2011). Test martingales, Bayes factors and p-values. *Statist*. *Sci*. **26** 84–101. MR2849911 https://doi.org/10.1214/10-STS347

SHAFER, G. and VOVK, V. (2001). *Probability and Finance*: *It's Only a Game!*. *Wiley Series in Probability and Statistics*. *Financial Engineering Section*. Wiley-Interscience, New York. MR1852450 https://doi.org/10.1002/0471249696

SHAFER, G. and VOVK, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley, Hoboken, NJ.

SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

SHEN, J., SHEN, Y., WANG, B. and WANG, R. (2019). Distributional compatibility for change of measures. *Finance Stoch*. **23** 761–794. MR3968284 https://doi.org/10.1007/s00780-019-00393-4

SIMON, B. (2011). *Convexity*: *An Analytic Viewpoint*. *Cambridge Tracts in Mathematics* **187**. Cambridge Univ. Press, Cambridge. MR2814377 https://doi.org/10.1017/CBO9780511910135

SIMONS, G. (1970). A martingale decomposition theorem. *Ann*. *Math*. *Stat*. **41** 1102–1104. MR0261678 https://doi.org/10.1214/aoms/1177696991

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann*. *Math*. *Stat*. **36** 423–439. MR0177430 https://doi.org/10.1214/aoms/1177700153

TESCHL, G. (2012). *Ordinary Differential Equations and Dynamical Systems*. *Graduate Studies in Mathematics* **140**. Am. Math. Soc., Providence, RI. MR2961944 https://doi.org/10.1090/gsm/140

VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

VOVK, V. (2021). Testing randomness online. *Statist*. *Sci*. **36** 595–611. MR4323055 https://doi.org/10.1214/20-sts817

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann*. *Statist*. **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

VOVK, V. and WANG, R. (2024). Nonparametric e-tests of symmetry. *N*. *Engl*. *J*. *Statist*. *Data Sci*. 1–10.

WANG, R. and RAMDAS, A. (2022). False discovery rate control with e-values. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **84** 822–852. MR4460577

WANG, R. and ZHANG, Z. (2023). Simultaneous optimal transport. arXiv preprint. Available at arXiv:2201.03483v2.

WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann*. *Math*. *Stat*. **9** 60–62.

ZHANG, Z., RAMDAS, A. and WANG, R. (2024). Supplement to "On the existence of powerful p-values and e-values for composite hypotheses." https://doi.org/10.1214/24-AOS2434SUPP