Testing hypotheses generated by constraints

Martin Larsson[*]   Aaditya Ramdas[†]   Johannes Ruf[‡]

August 6, 2025

Abstract

E-variables are nonnegative random variables with expected value at most one under any distribution from a given null hypothesis. Every nonasymptotically valid test can be obtained by thresholding some e-variable. As such, e-variables arise naturally in applications in statistics and operations research, and a key open problem is to characterize their form. We provide a complete solution to this problem for hypotheses generated by constraints—a broad and natural framework that encompasses many hypothesis classes occurring in practice. Our main result is an abstract representation theorem that describes all e-variables for any hypothesis defined by an arbitrary collection of measurable constraints. We instantiate this general theory for three important classes: hypotheses generated by finitely many constraints, one-sided sub-ψ distributions (including sub-Gaussian distributions), and distributions constrained by group symmetries. In each case, we explicitly characterize all e-variables as well as all admissible e-variables. Numerous examples are treated, including constraints on moments, quantiles, and conditional value-at-risk (CVaR). Building on these, we prove existence and uniqueness of optimal e-variables under a large class of expected utility-based objective functions used for optimal decision making, in particular covering all criteria studied in the e-variable literature to date.

1 Introduction

Fix a measurable space $\mathcal{X}$ and let $\mathcal{M}_1$ denote the set of all probability measures on $\mathcal{X}$. Suppose we observe a random datum X with values in $\mathcal{X}$, and consider the null hypothesis that the distribution of X belongs to some set of probability measures $\mathcal{P} \subset \mathcal{M}_1$. An e-variable for $\mathcal{P}$ is a nonnegative (possibly infinite) random variable whose expected value under every distribution in $\mathcal{P}$ is at most one. The set of all e-variables for $\mathcal{P}$ is denoted by $\mathcal{E}$:

$$\mathcal{E} = \left\{ \text{all measurable } h \colon \mathcal{X} \to [0, \infty] \text{ such that } \int_{\mathcal{X}} h \, d\mu \leq 1 \text{ for all } \mu \in \mathcal{P} \right\}.$$

E-variables have recently been recognized as fundamental objects in a variety of hypothesis testing and inference problems. A rapidly growing body of work uses e-variables as the basis for solving a wide range of problems in statistics and operations research, such as

---

[*]Department of Mathematical Sciences, Carnegie Mellon University, `larsson@cmu.edu`
[†]Departments of Statistics and ML, Carnegie Mellon University, `aramdas@cmu.edu`
[‡]Department of Mathematics, London School of Economics, `j.ruf@lse.ac.uk`

multiple testing, A/B testing, and sequential anytime-valid inference, just to mention a few. Some recent papers include Wasserman et al. (2020); Vovk and Wang (2021); Shafer (2021); Grünwald et al. (2024); several other key references appear later in this paper.

To see why e-variables are fundamentally connected to hypothesis testing, observe that every e-variable for $\mathcal{P}$ yields a nonasymptotic level-α test for $\mathcal{P}$: we reject the null when the e-variable exceeds $1/\alpha$. Markov's inequality implies that the type-I error of such a test is at most α. Conversely, it is known that every level-α test for $\mathcal{P}$ can be recovered by thresholding some e-variable at $1/\alpha$. Consequently, a description of all e-variables leads to a description of all valid tests. Further, it is desirable to identify e-variables that lead to *powerful* tests. Standard notions of power for e-variables are based on expected utility under an alternative hypothesis (Shafer, 2021; Grünwald et al., 2024; Larsson et al., 2025), and finding powerful e-variables amounts to solving optimization problems over the set $\mathcal{E}$. In addition to their importance for statistical analysis, solutions to these problems serve as input for optimal decision making.

For these reasons, it is of interest to characterize $\mathcal{E}$. Indeed, a characterization of $\mathcal{E}$ is effectively a characterization of the set of all tests for $\mathcal{P}$. Furthermore, a low-dimensional parameterization of $\mathcal{E}$ simplifies the task of finding optimal e-variables.

In this paper we characterize all e-variables for sets $\mathcal{P}$ that are described by constraints. This is a very natural class, both in parametric and in nonparametric settings. In parametric settings, for example in exponential families, it is common to test whether a parameter lies in some range; this is a constraint. In nonparametric settings, it is common to specify classes of distributions whose moments or supports are restricted in some way; these are also constraints. Thus, the classes $\mathcal{P}$ considered in this paper are quite general, and special cases of such "constrained hypotheses" have been frequently considered in the literature. Our work characterizes all possible e-variables for such classes, without restrictions on the constraints (there could be uncountably many, they can be discontinuous functions, etc.) and without restrictions on the underlying measure space (we do not require any topological properties such as compactness, closedness, finite dimensionality, etc.).

While we will cite papers that study special cases of $\mathcal{P}$ in later sections of the paper that instantiate our general results, we note that the only general attempt to characterize e-variables for constrained hypotheses appears in a recent work by Clerico (2024a). Among our contributions are greatly generalized versions of the results in that paper. We would also like to highlight the nonasymptotic nature of our work. This is in contrast to, for example, Romano et al. (2014), who design asymptotically valid tests for finitely constrained hypotheses.

Lastly, although we frame our results as being about a single observation from the sample space $\mathcal{X}$, we haste to emphasize two things. First, this "single observation" could be a finite or infinite sample, say if $\mathcal{X} = \mathbb{R}^n$ or $\mathcal{X} = \mathbb{R}^\infty$, and our results still apply. Second, given an e-variable $h(x)$ for $\mathcal{P}$, the product $\prod_{i=1}^{n} h(X_i)$ is an e-variable for any i.i.d. sequence $X_1, \ldots, X_n$ sampled from a distribution in $\mathcal{P}$. More generally, it is an e-variable provided that for all i, the conditional distribution of X_{i+1} given $X_1, \ldots, X_i$ belongs to $\mathcal{P}$. Thus our results immediately yield explicit e-variables in these settings as well, although we do not characterize *all* e-variables in this case. This and other challenging open problems are discussed in Section 10.

Paper outline. Section 2 formally defines hypotheses generated by constraints, introduces the mathematical setting needed to analyze them, and presents the main result of this paper, Theorem 2.2, which gives an abstract description of the set of all e-variables for any hypothesis generated by constraints. An important role is played by the theory of dual pairs of vector spaces, and in particular the notion of weak closure; we review these in the appendix.

Sections 3–5 instantiate the abstract theorem for three particular hypothesis classes of practical interest. In Section 3, hypotheses generated by finitely many constraints are considered. The main results are Theorem 3.1 and Corollary 3.3, and several special cases are discussed thereafter. Section 4 tackles an important and general class of distributions generated by uncountably many constraints, that of one-sided sub-ψ distributions, which includes the well-studied sub-Gaussian case. The main result is Theorem 4.3. Section 5 is devoted to another nontrivial and general class of distributions, those that are constrained to remain invariant under a group of symmetries (such as exchangeable distributions). The main results are Theorems 5.6 and 5.7.

Up to this point in the paper, the focus is on obtaining explicit descriptions of the set of *all* e-variables for a given hypothesis. However, in statistical applications one wishes to work with *admissible* e-variables whenever possible. This is the topic of Section 6, where *minimal complete classes* of (admissible) e-variables are introduced. It is shown that such classes do not exist in general, but do exist in the setting of this paper under mild additional assumptions. The main result is Theorem 6.5.

Section 7 tackles the problem of finding *optimal* e-variables for finitely generated and sub-ψ hypotheses. For a general class of objective functions, existence of optimizers is established in Theorems 7.3 and 7.4, and a simple uniqueness criterion is provided in Theorem 7.5. These results cover all objective functions studied in the e-variable literature to date, as well as criteria such as the continuous variational preferences of Maccheroni et al. (2006). The results in this section depend crucially on the representation theorems developed earlier.

The last part of the paper contains further discussion and amplifications. Section 8 studies hypotheses that are unions of hypotheses generated by constraints. A nontrivial example involving conditional value-at-risk (CVaR) illustrates its utility. Section 9 considers hypotheses with a relaxed integrability condition. It is shown that for finitely generated hypotheses, this relaxation makes no difference to the set of associated e-variables, but leaves an open question in the infinite case. Section 10 concludes with a discussion of further open problems. Appendix A contains some basic results from topology and functional analysis that are used in the paper, for example providing background for our convergence results that rely on *nets* instead of sequences, and reviewing a key bipolar theorem that underlies our main results.

Notation. We denote by $\mathcal{L}$ and $\mathcal{M}$ the space of all real-valued measurable functions and finite signed measures, respectively, on our measurable space $\mathcal{X}$. We write $|\mu|$ for the total variation measure of any $\mu \in \mathcal{M}$. Given a subset $\mathcal{P} \subset \mathcal{M}_1$ (the probability measures on $\mathcal{X}$), we say that a measurable set $A \subset \mathcal{X}$ is $\mathcal{P}$-negligible if $\mu(A) = 0$ for all $\mu \in \mathcal{P}$. A pointwise property of a function $f \in \mathcal{L}$ holds $\mathcal{P}$-quasi-surely, abbreviated $\mathcal{P}$-q.s., if the set where it fails is $\mathcal{P}$-negligible. A function $f \in \mathcal{L}$ is called a $\mathcal{P}$-version of another function $g \in \mathcal{L}$ if $f = g$, $\mathcal{P}$-q.s. For subsets A, B of a vector space, we write $A - B = \{a - b \colon a \in A, b \in B\}$. The set of natural numbers is $\mathbb{N} = \{1, 2, \ldots\}$.

2 Hypotheses generated by constraints

Definition 2.1. A *constraint set* is any nonempty set of functions $\Phi \subset \mathcal{L}$. The elements of Φ are called *constraint functions*. The *hypothesis generated by* Φ is the (possibly empty) set $\mathcal{P}$ of probability measures given by

$$\mathcal{P} = \left\{ \mu \in \mathcal{M}_1 \colon \int_{\mathcal{X}} |f| d\mu < \infty \text{ and } \int_{\mathcal{X}} f d\mu \leq 0 \text{ for all } f \in \Phi \right\}. \tag{2.1}$$

Although the hypothesis $\mathcal{P}$ is defined through inequality constraints, it is easy to encode equality constraints by letting Φ contain both f and $-f$. Given a constraint set Φ and the hypothesis $\mathcal{P}$ that it generates, we define the vector spaces

$$\mathcal{L}^{\Phi} = \left\{ f \in \mathcal{L} \colon \int_{\mathcal{X}} |f| d\mu < \infty \text{ for all } \mu \in \mathcal{P} \right\},$$

$$\mathcal{M}^{\Phi} = \left\{ \mu \in \mathcal{M} \colon \int_{\mathcal{X}} |f| d|\mu| < \infty \text{ for all } f \in \mathcal{L}^{\Phi} \right\}.$$

These spaces serve as a useful arena for our theory because $\mathcal{L}^{\Phi}$ contains all bounded measurable functions, all constraint functions, and all (finite) e-variables. Moreover, $\mathcal{M}^{\Phi}$ contains the hypothesis $\mathcal{P}$ and all Dirac measures, and its elements integrate all the functions in $\mathcal{L}^{\Phi}$ by construction. It is also convenient to introduce the *quasi-surely positive cone*,

$$\mathcal{L}_p^{\Phi} = \{ f \in \mathcal{L}^{\Phi} \colon f \geq 0, \ \mathcal{P}\text{-q.s.}\},$$

which may in general differ from the set $\mathcal{L}_+^{\Phi}$ of functions in $\mathcal{L}^{\Phi}$ that are nonnegative *everywhere*. Finally, we define

$$\mathcal{C} = \text{cone}(\Phi) - \mathcal{L}_p^{\Phi}. \tag{2.2}$$

This is the convex cone of all functions that are quasi-surely dominated by a conic combination of constraint functions. In symbols, $\mathcal{C}$ consists of all $f = g - h$ with $g \in \text{cone}(\Phi)$ and $h \in \mathcal{L}_p^{\Phi}$, or equivalently, $f \leq g$, $\mathcal{P}$-q.s.

The elements of $\mathcal{C}$ have nonpositive expectation under every measure in $\mathcal{P}$. The following result shows that the *weak closure* of $\mathcal{C}$ actually consists of *all* functions in $\mathcal{L}^{\Phi}$ with this property. This immediately leads to a description of the set of all e-variables for $\mathcal{P}$. Here the weak closure refers to the topology $\sigma(\mathcal{L}^{\Phi}, \mathcal{M}^{\Phi})$ induced by the dual pairing $\langle f, \mu \rangle = \int_{\mathcal{X}} f d\mu$; see Appendix A.2.

Theorem 2.2.   (i) *A function $f \in \mathcal{L}^{\Phi}$ satisfies $\int_{\mathcal{X}} f d\mu \leq 0$ for all $\mu \in \mathcal{P}$ if and only if f belongs to $\overline{\mathcal{C}}$, the weak closure of $\mathcal{C}$.*

  (ii) *In particular, the set $\mathcal{E}$ of all e-variables for $\mathcal{P}$ consists precisely of those $[0, \infty]$-valued measurable functions that are $\mathcal{P}$-q.s. equal to $1 + f$ for some $f \in \overline{\mathcal{C}}$.*

Proof. Let us first confirm that the bilinear form $\langle f, \mu \rangle = \int_{\mathcal{X}} f d\mu$ separates points. Indeed, if $\mu \in \mathcal{M}^{\Phi}$ is fixed and $\langle f, \mu \rangle = 0$ for all $f \in \mathcal{L}^{\Phi}$, then by taking $f = \mathbf{1}_A$ for any measurable set $A \subset \mathcal{X}$, we see that $\mu = 0$. If instead $f \in \mathcal{L}^{\Phi}$ is fixed and $\langle f, \mu \rangle = 0$ for all $\mu \in \mathcal{M}^{\Phi}$, we may take $\mu = \delta_x$ for any $x \in \mathcal{X}$ to see that $f = 0$. With this out of the way we may proceed with the proof of the theorem.

(i): The bipolar theorem states that $\mathcal{C}^{\circ\circ} = \overline{\mathcal{C}}$; see Theorem A.2. This is actually the desired conclusion because, as we show next,

$$\mathcal{C}^{\circ\circ} = \left\{ f \in \mathcal{L}^{\Phi} : \int_{\mathcal{X}} f d\mu \le 0 \text{ for all } \mu \in \mathcal{P} \right\}.$$

This representation of $\mathcal{C}^{\circ\circ}$ follows directly from the identity $\mathcal{C}^{\circ} = \mathbb{R}_{+}\mathcal{P}$, which we now prove. (Here $\mathbb{R}_{+}\mathcal{P}$ refers to the set of nonnegative multiples of elements of $\mathcal{P}$.) For the forward inclusion, consider an element $\mu \in \mathcal{C}^{\circ}$. Since $-\mathbf{1}_A \in \mathcal{C}$ for every measurable set A we have $\mu \in \mathcal{M}_{+}$, and since $\Phi \subset \mathcal{C}$ we have $\int_{\mathcal{X}} f d\mu \le 0$ for all $f \in \Phi$. Thus μ is a nonnegative multiple of an element of $\mathcal{P}$. Conversely, if μ is a nonnegative multiple of an element of $\mathcal{P}$, then for any function $f \in \mathcal{C}$, say $f = g - h$ with $g \in \text{cone}(\Phi)$ and $h \in \mathcal{L}_p^{\Phi}$, we have $\int_{\mathcal{X}} f d\mu \le \int_{\mathcal{X}} g d\mu \le 0$, and hence $\mu \in \mathcal{C}^{\circ}$. Thus $\mathcal{C}^{\circ} = \mathbb{R}_{+}\mathcal{P}$, and the proof is complete.

(ii): Since every e-variable is $\mathcal{P}$-q.s. equal to a finite e-variable, and since $\mathcal{L}^{\Phi}$ contains all finite e-variables, the claim is immediate from (i). $\qquad\square$

Remark 2.3. Although Φ is assumed to be nonempty, $\mathcal{P}$ may be empty, and Theorem 2.2 still applies. In this case, $\mathcal{L}^{\Phi}$ is the set $\mathcal{L}$ of all measurable functions f. The condition $\int_{\mathcal{X}} f d\mu \le 0$ for all $\mu \in \mathcal{P}$ is vacuously satisfied by *every* such f, so the theorem states that $\overline{\mathcal{C}}$ equals all of $\mathcal{L}$. This can also be seen directly: if $\mathcal{P} = \emptyset$ then every measurable set $A \subset \mathcal{X}$ is negligible, thus $\mathcal{L}_p^{\Phi} = \mathcal{L}^{\Phi} = \mathcal{L}$, and $\mathcal{C} = \text{cone}(\Phi) - \mathcal{L}_p^{\Phi} = \mathcal{L}$ (recall that Φ is nonempty).

Remark 2.4. A hypothesis generated by constraints will not admit a reference measure in general. However, if a desired reference measure $\bar{\mu}$ is given, one can ensure that every $\mu \in \mathcal{P}$ satisfies $\mu \ll \bar{\mu}$ simply by augmenting Φ with all functions $\mathbf{1}_A$ where A is a $\bar{\mu}$-nullset. At the end of Section 3 we discuss this construction in the context of finitely generated hypotheses, see especially Theorem 3.10.

For later use we record the following basic property of the quasi-surely positive cone.

Lemma 2.5. $\mathcal{L}_p^{\Phi}$ *is weakly closed.*

Proof. We claim that

$$\mathcal{L}_p^{\Phi} = \left\{ f \in \mathcal{L}^{\Phi} : \int_{\mathcal{X}} f d\mu \ge 0 \text{ for all } \mu \in \mathcal{M}_{+}^{\Phi} \text{ with } \mu \ll \mathcal{P} \right\}, \qquad (2.3)$$

where $\mathcal{M}_{+}^{\Phi}$ is the set of nonnegative elements of $\mathcal{M}^{\Phi}$, and $\mu \ll \mathcal{P}$ means that $\mu(A) = 0$ for every $\mathcal{P}$-negligible set A. The forward inclusion "$\subset$" is clear. For the reverse inclusion "$\supset$", consider some $f \notin \mathcal{L}_p^{\Phi}$. Then the set $A_0 = \{f < 0\}$ is not $\mathcal{P}$-negligible, and hence $\nu(A_0) > 0$ for some $\nu \in \mathcal{P}$. Define $\mu = \nu(\cdot \cap A_0)$. Then μ belongs to $\mathcal{M}_{+}^{\Phi}$ and $\mu \ll \mathcal{P}$ (indeed, $\mu \ll \nu$). But $\int_{\mathcal{X}} f d\mu < 0$, so f does not belong to the right-hand side of (2.3). This establishes (2.3) and shows that $\mathcal{L}_p^{\Phi}$ is an intersection of sets of the form $\{f \in \mathcal{L}^{\Phi} : \int_{\mathcal{X}} f d\mu \ge 0\}$. Since $f \mapsto \int_{\mathcal{X}} f d\mu$ is weakly continuous, all these sets are weakly closed, and thus so is $\mathcal{L}_p^{\Phi}$. $\qquad\square$

3   Finitely generated hypotheses

Consider a finite nonempty constraint set,

$$\Phi = \{g_1, \ldots, g_d\},$$

and let $\mathcal{P}$ be the hypothesis generated by Φ. The constraint functions must be real-valued and measurable, but can otherwise be completely arbitrary.

Theorem 3.1. *A function $f \in \mathcal{L}^\Phi$ satisfies $\int_{\mathcal{X}} f d\mu \leq 0$ for all $\mu \in \mathcal{P}$ if and only if*

$$f \leq \sum_{i=1}^d \pi_i g_i, \quad \mathcal{P}\text{-}q.s.$$

for some $\pi = (\pi_1, \ldots, \pi_d) \in \mathbb{R}^d_+$. In particular, the set $\mathcal{E}$ of all e-variables for $\mathcal{P}$ consists precisely of those $[0, \infty]$-valued measurable functions which are $\mathcal{P}$-q.s. dominated by

$$1 + \sum_{i=1}^d \pi_i g_i$$

for some $\pi = (\pi_1, \ldots, \pi_d) \in \mathbb{R}^d_+$.

Before giving the proof, we introduce some terminology. The *support* of a vector $\rho \in \mathbb{R}^d_+$ is the set $\mathrm{supp}(\rho) = \{i \colon \rho_i > 0\}$. An index set $I \subset \{1, \ldots, d\}$ is called *redundant* if there exists some nonzero $\rho \in \mathbb{R}^d_+$ with $\mathrm{supp}(\rho) \subset I$ such that $\sum_{i=1}^d \rho_i g_i = 0$, $\mathcal{P}$-q.s. Note in particular that the empty set $I = \emptyset$ is *not* redundant. Note also that the set of vectors whose supports are not redundant,

$$K = \left\{ \pi \in \mathbb{R}^d_+ \colon \mathrm{supp}(\pi) \text{ is not redundant} \right\}, \tag{3.1}$$

is closed. Indeed, if $\pi_n \in K$ converges to some $\pi \in \mathbb{R}^d_+$, then $\mathrm{supp}(\pi) \subset \mathrm{supp}(\pi_n)$ for all sufficiently large n. Since the latter are not redundant, neither is the former, so $\pi \in K$. Finally, recall the set $\mathcal{C} = \mathrm{cone}(\Phi) - \mathcal{L}^\Phi_p$ introduced in (2.2).

Lemma 3.2. *Every $g' \in \mathrm{cone}(\Phi)$ has a $\mathcal{P}$-version $g = \sum_{i=1}^d \pi_i g_i$ for some $\pi \in K$. Thus every $f \in \mathcal{C}$ is of the form $f = g - h$ with $g = \sum_{i=1}^d \pi_i g_i$ for some $\pi \in K$ and $h \in \mathcal{L}^\Phi_p$.*

Proof. Consider any $g' = \sum_{i=1}^d \pi'_i g_i \in \mathrm{cone}(\Phi)$ with $\pi' \in \mathbb{R}^d_+$. To see that g' has a $\mathcal{P}$-version $g = \sum_{i=1}^d \pi_i g_i$ for some $\pi \in K$, suppose π' does not already belong to K, meaning that $\mathrm{supp}(\pi')$ is redundant. Let $\rho \in \mathbb{R}^d_+$ be as in the definition of redundant. Then there exists $\varepsilon > 0$ such that $\pi'' = \pi' - \varepsilon\rho$ belongs to $\mathbb{R}^d_+$ and satisfies $\mathrm{supp}(\pi'') \subsetneq \mathrm{supp}(\pi')$. Note that $\sum_i \pi''_i g_i = \sum_i \pi'_i g_i$, $\mathcal{P}$-q.s. If $\mathrm{supp}(\pi'')$ is not redundant, we take $\pi = \pi''$. Otherwise we repeat the process, each time reducing the size of the support. Since the empty set is not redundant, we must eventually reach a representation in terms of a vector π whose support is not redundant. This proves the first statement of the lemma. For the second statement, consider any $f = g' - h' \in \mathcal{C}$ with $g' \in \mathrm{cone}(\Phi)$ and $h' \in \mathcal{L}^\Phi_p$. Let $g = \sum_{i=1}^d \pi_i g_i$ be a $\mathcal{P}$-version of g' with $\pi \in K$ and note that $f = g - h$ where $h = h' + g - g'$ still belongs to $\mathcal{L}^\Phi_p$. $\square$

Proof of Theorem 3.1. We will show that the set $\mathcal{C}$ in (2.2) is already weakly closed; the result then follows from Theorem 2.2. We will prove closedness by showing that the limit of any *convergent net* in $\mathcal{C}$ is again an element of $\mathcal{C}$. Nets are generalizations of sequences, and are required for checking closedness in certain topological spaces. For the benefit of readers who do not work with nets regularly, we review the basic definitions and properties in Appendix A.1. Readers who are not familiar with nets may replace 'net' with 'sequence' and 'α' with 'n'

6

everywhere below without losing any essential ideas. This modification of the proof would show that $\mathcal{C}$ is *sequentially closed*, but this does not imply closedness in general. For this reason the actual proof uses nets. We now turn to the details.

Consider a net (f_α) in $\mathcal{C}$ that converges weakly to some $f \in \mathcal{L}^\Phi$. We must show that $f \in \mathcal{C}$. Thanks to Lemma 3.2, for each α we have $f_\alpha = g_\alpha - h_\alpha$, where $g_\alpha = \sum_{i=1}^d \pi_{\alpha,i} g_i$ for some $\pi_\alpha \in K$ and $h_\alpha \in \mathcal{L}_p^\Phi$. We claim that the real-valued net $\|\pi_\alpha\| = \pi_{\alpha,1} + \cdots + \pi_{\alpha,d}$ cannot converge to infinity. Assume for contradiction that it does, and write

$$\sum_{i=1}^d \frac{\pi_{\alpha,i}}{1 + \|\pi_\alpha\|} g_i - \frac{f_\alpha}{1 + \|\pi_\alpha\|} = \frac{h_\alpha}{1 + \|\pi_\alpha\|}. \tag{3.2}$$

Since $\pi_\alpha/(1 + \|\pi_\alpha\|)$ is a bounded net in $\mathbb{R}_+^d$, we may pass to a subnet and assume that it converges to some limit $\rho \in \mathbb{R}_+^d$, which then satisfies $\|\rho\| = 1$. Each $\pi_\alpha/(1 + \|\pi_\alpha\|)$ belongs to the closed set K, so ρ does too. Since the vector space operations are weakly continuous, the first term on the left-hand side of (3.2) converges weakly to $\sum_{i=1}^d \rho_i g_i$. Next, for any $\mu \in \mathcal{M}^\Phi$ we have $\langle f_\alpha, \mu \rangle \to \langle f, \mu \rangle$ and hence

$$\left\langle \frac{f_\alpha}{1 + \|\pi_\alpha\|}, \mu \right\rangle = \frac{1}{1 + \|\pi_\alpha\|} \langle f_\alpha, \mu \rangle \to 0.$$

Thus the second term on the left-hand side of (3.2) converges weakly to zero. Overall, the left-hand side converges to $\sum_{i=1}^d \rho_i g_i$, and we conclude using Lemma 2.5 that this quantity is nonnegative, $\mathcal{P}$-q.s. On the other hand, we have

$$\int_{\mathcal{X}} \sum_{i=1}^d \rho_i g_i d\mu = \sum_{i=1}^d \rho_i \int_{\mathcal{X}} g_i d\mu \leq 0$$

for every $\mu \in \mathcal{P}$. Thus $\sum_{i=1}^d \rho_i g_i = 0$, $\mathcal{P}$-q.s., which contradicts the fact that $\mathrm{supp}(\rho)$ is not redundant. We conclude that $\|\pi_\alpha\|$ cannot converge to infinity.

Since $\|\pi_\alpha\|$ does not converge to infinity, it admits a convergent subnet, which we again denote by π_α. Denote the limit by $\pi \in \mathbb{R}_+^d$. It follows that g_α converges to $g = \sum_{i=1}^d \pi_i g_i$, and then that $h_\alpha = g_\alpha - f_\alpha$ converges to $h = g - f$. The latter belongs to $\mathcal{L}_p^\Phi$ since this set is weakly closed thanks to Lemma 2.5. We conclude that $f = g - h \in \mathcal{C}$, showing that $\mathcal{C}$ is weekly closed, as required. $\square$

An immediate consequence of Theorem 3.1 is the following generalization of a result of Clerico (2024a, Theorem 1). Here an e-variable h is called *admissible* if whenever another e-variable h' satisfies $h' \geq h$, $\mathcal{P}$-q.s., we actually have $h' = h$, $\mathcal{P}$-q.s. (See also Section 6 for further discussion of admissibility.)

Corollary 3.3. *Every admissible e-variable is $\mathcal{P}$-q.s. equal to*

$$1 + \sum_{i=1}^d \pi_i g_i$$

for some π in the set

$$\Pi^\Phi = \left\{ \pi \in \mathbb{R}_+^d : 1 + \sum_{i=1}^d \pi_i g_i \geq 0 \ \mathcal{P}\text{-}q.s. \right\}.$$

Conversely, every function of the above form is an admissible e-variable provided Φ satisfies the following constraint qualification:

$$\text{If } g, g' \in \text{cone}(\Phi) \text{ and } g \leq g', \mathcal{P}\text{-q.s., then } g = g', \mathcal{P}\text{-q.s.} \tag{3.3}$$

Proof. The first part is immediate from Theorem 3.1. For the second part, fix an e-variable of the form $1 + g$, where $g = \sum_{i=1}^{d} \pi_i g_i$ for some $\pi \in \Pi^{\Phi}$, and consider any e-variable h that $\mathcal{P}$-q.s. dominates $1 + g$. We must show that $h = 1 + g$, $\mathcal{P}$-q.s. By Theorem 3.1, h is $\mathcal{P}$-q.s. dominated by an e-variable of the form $1 + g'$, where $g' = \sum_{i=1}^{d} \pi_i' g_i$ for some $\pi' \in \mathbb{R}_+^d$. We thus have $1 + g \leq h \leq 1 + g'$, $\mathcal{P}$-q.s. The constraint qualification (3.3) now yields $g = g'$, and hence $h = 1 + g$, $\mathcal{P}$-q.s. This shows that h is admissible. $\square$

Remark 3.4. In Corollary 3.3 we could further restrict π to the set $K_0 = K \cap \Pi^{\Phi}$ with K given in (3.1); see Lemma 3.2. This is noteworthy because, as we show in Lemma 7.2 below, the set K_0 is compact. This significantly simplifies the problem of finding *optimal* e-variables, and is discussed in detail in Section 7. It is worth mentioning that, is practice, Π^{Φ} itself is often already compact.

As another corollary, we are able to describe all admissible nonasymptotically valid tests for $\mathcal{P}$. Here a *level-α test* for $\mathcal{P}$, where $\alpha \in (0, 1)$, is a $[0, 1]$-valued measurable function φ such that $\int_{\mathcal{X}} \varphi d\mu \leq \alpha$ for all $\mu \in \mathcal{P}$. The test φ is *admissible* if whenever another such test φ' satisfies $\varphi' \geq \varphi$, $\mathcal{P}$-q.s., we actually have $\varphi' = \varphi$, $\mathcal{P}$-q.s.

Corollary 3.5. *Fix $\alpha \in (0, 1)$. Every admissible level-α test φ for $\mathcal{P}$ admits the representation $\varphi = 1 \wedge (\alpha + \sum_{i=1}^{d} \pi_i g_i)$, $\mathcal{P}$-q.s., for some $\pi \in \Pi^{\Phi}$.*

Proof. Note that φ/α is an e-variable for $\mathcal{P}$, and is thus $\mathcal{P}$-q.s. dominated by some e-variable of the form $h = 1 + \sum_{i=1}^{d} \pi_i' g_i$ with $\pi' \in \Pi^{\Phi}$. Define $\varphi' = 1 \wedge (\alpha + \sum_{i=1}^{d} \alpha \pi_i' g_i)$. Then $\varphi' \geq \varphi$, $\mathcal{P}$-q.s., and, thanks to the e-variable property, φ' is a level-α test for $\mathcal{P}$. Since φ was admissible, $\varphi' = \varphi$, $\mathcal{P}$-q.s. Setting $\pi = \alpha \pi'$ we get the claimed representation. $\square$

The representation in Corollary 3.3 is useful because the set Π^{Φ} can be described explicitly in various cases of interest. The following simple, yet interesting, example illustrates this; see Agrawal et al. (2020, 2021a); Clerico (2024a); Wang and Ramdas (2023); Fan et al. (2025) for more details.

Example 3.6. *We take $\mathcal{X} = \mathbb{R}$ and let $\mathcal{P}$ consist of all zero mean distributions with standard deviation bounded by a positive number σ. This hypothesis is generated by the constraint set $\Phi = \{x, -x, x^2 - \sigma^2\}$. Thus Π^{Φ} consists of all $(\pi_1, \pi_2, \pi_3) \in \mathbb{R}_+^3$ such that $1 + (\pi_1 - \pi_2)x + \pi_3(x^2 - \sigma^2) \geq 0$ for all $x \in \mathbb{R}$. (The only $\mathcal{P}$-negligible set is the empty set, which is why the inequality must hold for* all *x.) It is natural to re-parameterize in terms of $\alpha = \pi_1 - \pi_2 \in \mathbb{R}$ and $\beta = \pi_3 \sigma^2 \in \mathbb{R}_+$, constrained to satisfy $1 + \alpha x + \beta(x^2/\sigma^2 - 1) \geq 0$ for all $x \in \mathbb{R}$. Minimizing over x and requiring that the minimum value be nonnegative, one arrives at the constraint $\sigma^2 \alpha^2 + (2\beta - 1)^2 \leq 1$ on α, β. We conclude that every admissible e-variable is of the form*

$$1 + \alpha x + \beta \left(\frac{x^2}{\sigma^2} - 1 \right), \quad x \in \mathbb{R}, \tag{3.4}$$

for some (α, β) inside the ellipse determined by $\sigma^2 \alpha^2 + (2\beta - 1)^2 \leq 1$. Note we do not have to impose $\beta \geq 0$ separately, since this is already implied by the ellipse constraint. Finally, Φ

satisfies the constraint qualification (3.3). *Indeed, if* $\alpha x + \beta(x^2/\sigma^2 - 1) \geq \alpha' x + \beta'(x^2/\sigma^2 - 1)$ *for all* $x \in \mathbb{R}$*, we first take* $x = \pm\sigma$ *to get* $\alpha = \alpha'$*, and then (say)* $x = 0$ *and* $x = 2\sigma$ *to get* $\beta = \beta'$*. Consequently, every function of the form* (3.4) *is an admissible e-variable.*

The fact that the constraint functions are not required to satisfy any kind of continuity or other regularity conditions beyond measurability is sometimes useful, for instance in the context of quantiles.

Example 3.7. *We continue to take* $\mathcal{X} = \mathbb{R}$*. Fix* $\alpha \in (0,1)$ *and* $q \in \mathbb{R}$*, and let* $\mathcal{P}$ *consist of all distributions* μ *whose* α*-quantile is at most* q*, meaning that* $\mu((-\infty, q]) \geq \alpha$*. This hypothesis is generated by the single constraint function* $\alpha - \mathbf{1}_{(-\infty,q]}(x)$*. Since this function takes both positive and negative values, the constraint qualification* (3.3) *holds. Thus the admissible e-variables are the functions* $1 + \pi_1(\alpha - \mathbf{1}_{(-\infty,q]}(x))$ *with* $\pi_1 \in [0, (1-\alpha)^{-1}]$ *to ensure nonnegativity.*

The following example is common in the recent literature involving the mean of a bounded random variable.

Example 3.8. *Take* $\mathcal{X} = [0,1]$ *and let* $\mathcal{P}$ *consist of all distributions whose mean is at most a given constant* $m \in (0,1)$*. This hypothesis is generated by the constraint function* $x - m$*, and the constraint qualification* (3.3) *holds. Thus, the admissible e-variables are the functions* $1 + \pi_1(x - m)$ *with* $\pi_1 \in [-1/(1-m), 1/m]$ *to ensure nonnegativity.*

In particular, this recovers the class of e-variables used in Waudby-Smith and Ramdas (2024); Larsson et al. (2025); Orabona and Jun (2023); Clerico (2024b). A minor variant of Example 3.6 shows that without the boundedness assumption, there do not exist any nontrivial e-variables.

Example 3.9. *Take* $\mathcal{X} = \mathbb{R}$ *and let* $\mathcal{P}$ *consist of all distributions whose mean exists and equals zero. This hypothesis is generated by the constraint functions* x *and* $-x$*, and the constraint qualification* (3.3) *holds. Thus, the admissible e-variables are the functions* $1 + (\pi_1 - \pi_2)x = 1 + \alpha x$*, where we reparameterize in terms of* $\alpha = \pi_1 - \pi_2 \in \mathbb{R}$ *as in Example 3.6. We must choose* α *so that* αx *is nonnegative for any* $x \in \mathbb{R}$*. This immediately implies* $\alpha = 0$*, showing that the e-variable equal to one is the only admissible e-variable in this class (and all other e-variables must be less than or equal to one). The same argument shows that the hypothesis of a well-defined and nonpositive mean, generated by the single constraint function* x*, also only admits trivial e-variables.*

We end this section with a brief discussion of reference measures, building on Remark 2.4. In addition to the constraint set $\Phi = \{g_1, \ldots, g_d\}$, suppose a desired reference measure $\bar{\mu}$ is given. We are interested in the hypothesis

$$\mathcal{P}_{\bar{\mu}} = \{\mu \in \mathcal{P} : \mu \ll \bar{\mu}\}.$$

Note that $\mathcal{P}_{\bar{\mu}}$ is generated by the augmented constraint set $\Phi_{\bar{\mu}} = \Phi \cup \{\mathbf{1}_A : \bar{\mu}(A) = 0\}$. Moreover, functions which are zero $\bar{\mu}$-a.e. are actually $\mathcal{P}_{\bar{\mu}}$-negligible, including all conic combinations of the constraint functions $\mathbf{1}_A$ with $\bar{\mu}(A) = 0$. Therefore the set $\mathcal{C}$ in (2.2), with $\Phi_{\bar{\mu}}$ in place of Φ, consists of all functions of the form $f = \sum_{i=1}^{d} \pi_i g_i - h$ with $\pi_1, \ldots, \pi_d \in \mathbb{R}_+$ and $h \geq 0$, $\mathcal{P}_{\bar{\mu}}$-q.s. With this observation in hand we obtain the following result, whose proof is word for word the same as that of Theorem 3.1 (including the proof of Lemma 3.2) and Corollary 3.3, replacing only Φ with $\Phi_{\bar{\mu}}$ and $\mathcal{P}$ with $\mathcal{P}_{\bar{\mu}}$.

9

Theorem 3.10. *The statements of Theorem 3.1 and Corollary 3.3 remain true with $\Phi_{\bar{\mu}}$ and $\mathcal{P}_{\bar{\mu}}$ in place of Φ and $\mathcal{P}$.*

4   One-sided sub-ψ distributions

Fix a closed convex function $\psi\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ whose effective domain $\mathrm{dom}(\psi)$ is either $[0, \lambda_{\max})$ for some $\lambda_{\max} \in (0, \infty]$, or $[0, \lambda_{\max}]$ for some $\lambda_{\max} \in (0, \infty)$.[1] We assume that ψ is nonnegative and that $\psi(0) = 0$. Key examples include cumulant generating functions of zero-mean distributions, modified to take the value infinity on the negative half-line. More generally, ψ could be a *CGF-like function* in the terminology of Howard et al. (2020, 2021). Our goal is to describe the e-variables for the hypothesis consisting of all distributions of the following kind, whose usage stems back to the work of Cramér (1994) (originally 1938) and Chernoff (1952).

Definition 4.1. A probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$ is called *(one-sided) sub-ψ* if its cumulant generating function is bounded above by ψ, that is,

$$\int_{\mathbb{R}} e^{\lambda x} \mu(dx) \le e^{\psi(\lambda)} \text{ for all } \lambda \in \mathrm{dom}(\psi).$$

For example, when $\psi(\lambda) = \sigma^2 \lambda^2 / 2$ and $\lambda_{\max} = \infty$, the measure is called σ-sub-Gaussian in the sense that its (right) tail is lighter than that of a centered Gaussian with variance σ^2. The sub-ψ property in Definition 4.1 is "one-sided" in the sense that no condition is imposed for negative values of λ. Several of the proofs below make use of this property. To simplify terminology we will usually omit the qualifier "one-sided".

The convex conjugate of ψ is the convex function ψ^* given by

$$\psi^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \psi(\lambda)\}, \quad x \in \mathbb{R}.$$

Since $\psi(0) = 0$, ψ^* takes values in $[0, \infty]$. Moreover, because ψ is nonnegative and $\psi(\lambda) = \infty$ for $\lambda < 0$, we have $\psi^*(x) = 0$ for $x \le 0$. Thus the effective domain $\mathrm{dom}(\psi^*)$ contains the negative half-line.

Lemma 4.2. *Every sub-ψ distribution is concentrated on $\mathrm{dom}(\psi^*)$.*

Proof. This follows from the well-known fact that any sub-ψ distributed random variable X satisfies the Chernoff tail bound $\mathsf{P}(X \ge x) \le e^{-\psi^*(x)}$ for all x. Indeed, if $\mathrm{dom}(\psi^*) = \mathbb{R}$ there is nothing to prove. Otherwise, let $\bar{x} < \infty$ denote the right endpoint of the interval $\mathrm{dom}(\psi^*)$. If $\bar{x} \notin \mathrm{dom}(\psi^*)$, then $\psi^*(\bar{x}) = \infty$ and hence $\mathsf{P}(X \ge \bar{x}) \le e^{-\psi^*(\bar{x})} = 0$, so that X is concentrated on $\mathrm{dom}(\psi^*)$. If $\bar{x} \in \mathrm{dom}(\psi^*)$, then $\mathsf{P}(X > \bar{x}) = \lim_{x \downarrow \bar{x}} \mathsf{P}(X \ge x) = 0$, showing that X is concentrated on $\mathrm{dom}(\psi^*)$ in this case too. $\square$

Thanks to Lemma 4.2, any sub-ψ distribution can be regarded as a probability measure on $\mathrm{dom}(\psi^*)$. We thus take $\mathcal{X} = \mathrm{dom}(\psi^*)$ with its Borel σ-algebra. Consider the infinitely many (even uncountably many) constraint functions

$$g_\lambda(x) = e^{\lambda x - \psi(\lambda)} - 1, \quad x \in \mathrm{dom}(\psi^*),$$

---

[1] The effective domain is the set $\mathrm{dom}(\psi) = \{\lambda \in \mathbb{R}\colon \psi(\lambda) < \infty\}$ where ψ is finite. That ψ is closed means that its epigraph $\{(\lambda, y) \in \mathbb{R} \times \mathbb{R}\colon \psi(\lambda) \le y\}$ is closed. For our ψ, this just says that ψ is continuous at $\lambda_{\max}$ (if this is finite) and at 0.

indexed by $\lambda \in \mathrm{dom}(\psi)$. The hypothesis generated by these functions is

$$\mathcal{P} = \{\mu \in \mathcal{M}_1 \colon \mu \text{ is sub-}\psi\}. \tag{4.1}$$

If $\lambda_{\max}$ is not already in $\mathrm{dom}(\psi)$, we include the additional constraint function

$$g_{\lambda_{\max}}(x) = \lim_{\lambda \uparrow \lambda_{\max}} g_\lambda(x), \quad x \in \mathrm{dom}(\psi^*).$$

The limit exists and is finite because $\lambda \mapsto \lambda x - \psi(\lambda)$ is concave and $-1 \le g_\lambda(x) \le e^{\psi^*(x)} - 1$. Fatou's lemma implies that $\int_{\mathcal{X}} g_{\lambda_{\max}} d\mu \le 0$ for any sub-ψ distribution μ, so $g_{\lambda_{\max}}$ is redundant in the sense that including it does not alter the generated hypothesis $\mathcal{P}$ in (4.1). It does, however, play a role in the representation theorem below. Note that the constraint functions g_λ are now indexed by the compact set

$$\Lambda = [0, \lambda_{\max}]$$

(where we stress that $\lambda_{\max}$ may be infinity), and the maps $\lambda \mapsto g_\lambda(x)$ are continuous on Λ for every $x \in \mathrm{dom}(\psi^*)$. Our full constraint set is

$$\Phi = \{g_\lambda \colon \lambda \in \Lambda\}.$$

Theorem 4.3.   (i) *A function $f \in \mathcal{L}^\Phi$ satisfies $\int_{\mathcal{X}} f d\mu \le 0$ for all $\mu \in \mathcal{P}$ if and only if*

$$f(x) \le \int_\Lambda g_\lambda(x) \pi(d\lambda), \quad x \in \mathrm{dom}(\psi^*),$$

for some $\pi \in \mathcal{M}_+(\Lambda)$.

(ii) *The set $\mathcal{E}$ of all e-variables for $\mathcal{P}$ consists precisely of those $[0, \infty]$-valued measurable functions that are pointwise dominated on $\mathrm{dom}(\psi^*)$ by*

$$\int_\Lambda e^{\lambda x - \psi(\lambda)} \pi(d\lambda)$$

for some $\pi \in \mathcal{M}_1(\Lambda)$. (Note that π is a probability measure here. Moreover, if $\lambda_{\max} \notin \mathrm{dom}(\psi)$, the integrand for this value of λ is understood as $\lim_{\lambda \uparrow \lambda_{\max}} e^{\lambda x - \psi(\lambda)}$.)

Just as in the finitely generated case (see Corollary 3.5), we obtain a description of all admissible tests for $\mathcal{P}$. We do not repeat the proof.

Corollary 4.4. *Fix $\alpha \in (0,1)$. Every admissible level-α test φ for $\mathcal{P}$ admits the representation $\varphi(x) = 1 \wedge \int_\Lambda \alpha e^{\lambda x - \psi(\lambda)} \pi(d\lambda)$ for $x \in \mathrm{dom}(\psi^*)$ for some $\pi \in \mathcal{M}_1(\Lambda)$.*

We now give a brief roadmap of the proof of Theorem 4.3, introducing some notation along the way. The most involved part of the proof is to show part (i). Once this has been done, the proof of part (ii) is straightforward.

To deduce Theorem 4.3(i) from Theorem 2.2(i), it suffices to show that the weak closure $\overline{\mathcal{C}}$ of the set $\mathcal{C} = \mathrm{cone}(\Phi) - \mathcal{L}_p^\Phi$ in (2.2) is equal to $\mathcal{G} - \mathcal{L}_+^\Phi$, where we define the convex cone

$$\mathcal{G} = \left\{ \int_\Lambda g_\lambda \pi(d\lambda) \colon \pi \in \mathcal{M}_+(\Lambda) \right\}. \tag{4.2}$$

11

By considering finitely supported measures π, one sees that $\mathcal{G}$ contains cone(Φ). Moreover, it is shown in Lemma 4.6 below that the only $\mathcal{P}$-negligible subset of dom(ψ^*) is the empty set, and hence $\mathcal{L}_p^\Phi = \mathcal{L}_+^\Phi$. Thus $\mathcal{C} \subset \mathcal{G} - \mathcal{L}_+^\Phi$. On the other hand, for any $g = \int_\Lambda g_\lambda \pi(d\lambda) \in \mathcal{G}$ and $\mu \in \mathcal{P}$ we have from Tonelli's theorem that

$$\int_\mathcal{X} g(x)\mu(dx) = \int_\Lambda \int_\mathcal{X} g_\lambda(x)\mu(dx)\pi(d\lambda) \leq 0.$$

This shows, first, that $\mathcal{G}$ is indeed a subset of $\mathcal{L}^\Phi$. It also shows, via the forward implication of Theorem 2.2(i), that $\mathcal{G} - \mathcal{L}_+^\Phi \subset \overline{\mathcal{C}}$. In summary, we have

$$\mathcal{C} \subset \mathcal{G} - \mathcal{L}_+^\Phi \subset \overline{\mathcal{C}}.$$

Therefore, to show that $\mathcal{G} - \mathcal{L}_+^\Phi = \overline{\mathcal{C}}$ it is enough to show that

$$\mathcal{G} - \mathcal{L}_+^\Phi \text{ is } \sigma(\mathcal{L}^\Phi, \mathcal{M}^\Phi)\text{-closed.} \tag{4.3}$$

This is the heart of the matter, and the proof relies on a closedness criterion for convex subsets of Banach space duals known as the Krein–Šmulian theorem (see Appendix A.3).

Unfortunately the Krein–Šmulian theorem cannot be applied directly, because $\mathcal{L}^\Phi$ is not the dual of a Banach space. Instead, we first endow $\mathcal{L}^\Phi$ with a slightly weaker topology than $\sigma(\mathcal{L}^\Phi, \mathcal{M}^\Phi)$, which allows us to embed it into a larger space that *is* the dual of a Banach space. Checking closedness in the weaker topology can now be done using the Krein–Šmulian theorem. As we show below, this amounts to checking that for each $r \in \mathbb{R}_+$, the subset $\mathcal{G}_r = \{g \in \mathcal{G}: g \geq -r\} \subset \mathcal{G}$ of elements uniformly bounded below by r is compact. This turns out to be fairly straightforward, because $\mathcal{G}_r$ is a continuous image of the compact set $\{\pi \in \mathcal{M}_+(\Lambda): \pi(\Lambda) \leq r\}$ equipped with the usual weak topology coming from duality with the continuous functions on Λ.

The details of this argument depend on several preliminary results.

Lemma 4.5. *Let $x_0 \in$ dom(ψ^*). There exists $y_0 \geq 0$ such that for any $p \in [0, \frac{1}{2}e^{-\psi^*(x_0)}]$ and $y \geq y_0$, the probability measure $\nu = p\delta_{x_0} + (1-p)\delta_{-x_0-y}$ is sub-ψ.*

Proof. Define

$$f(\lambda, p, y) = \int_\mathcal{X} e^{\lambda x - \psi(\lambda)}\nu(dx) = pe^{\lambda x_0 - \psi(\lambda)} + (1-p)e^{-\lambda(x_0+y)-\psi(\lambda)}.$$

If $x_0 \leq 0$, the right-hand side is bounded by one for any $p \in [0, 1]$, $y \geq -x_0$, and $\lambda \in \mathbb{R}$, recalling that ψ is nonnegative and that $\psi(\lambda) = \infty$ for $\lambda < 0$. Thus ν is sub-ψ, and we may take $y_0 = -x_0$.

Consider now the case $x_0 > 0$. We have $f(\lambda, \frac{1}{2}, 1) = \frac{1}{2}(e^{\lambda x_0} + e^{-\lambda(x_0+1)})e^{-\psi(\lambda)}$, which is strictly decreasing with respect to λ in a right neighborhood of zero. This is because ψ is nondecreasing on $[0, \infty)$, being a nonnegative convex function, and $\frac{d}{d\lambda}|_{\lambda=0}(e^{\lambda x_0} + e^{-\lambda(x_0+1)}) = -1$. Moreover, $f(0, \frac{1}{2}, 1) = 1$, so there exists some $\lambda_0 > 0$ such that $f(\lambda, \frac{1}{2}, 1) \leq 1$ for all $\lambda \in [0, \lambda_0]$. For $p \in [0, \frac{1}{2}]$ and $y \geq 1$, $f(\lambda, p, y)$ is nondecreasing in p and nonincreasing in y, so it follows that $f(\lambda, p, y) \leq 1$ for all such p and y, provided $\lambda \in [0, \lambda_0]$.

For $\lambda > \lambda_0$, $p \in [0, \frac{1}{2}e^{-\psi^*(x_0)}]$, and $y \geq 1$, we have the bound $f(\lambda, p, y) \leq \frac{1}{2} + e^{-\lambda_0(x_0+y)-\psi(\lambda_0)}$. This uses that $x_0 > 0$ and that ψ is nondecreasing on $[0, \infty)$. The right-hand side is bounded

by one for all $y \geq y_0$, where we may take $y_0 = \max\{1, \lambda_0^{-1}(\log(2) - \psi(\lambda_0)) - x_0\}$. In summary, we have the sub-ψ inequality $f(\lambda, p, y) \leq 1$ for all $\lambda \in \mathbb{R}$, provided $p \in [0, \frac{1}{2}e^{-\psi^*(x_0)}]$ and $y \geq y_0$. $\qquad\square$

Lemma 4.6. *The only subset of* $\mathrm{dom}(\psi^*)$ *that is* $\mathcal{P}$*-negligible is the empty set.*

Proof. Let A be any nonempty measurable subset of $\mathrm{dom}(\psi^*)$ and pick $x_0 \in A$. Lemma 4.5 yields a sub-ψ distribution that charges x_0. Thus A is not $\mathcal{P}$-negligible. $\qquad\square$

Lemma 4.7. *Every* $f \in \mathcal{L}^\Phi$ *satisfies* $\sup_{x \in \mathrm{dom}(\psi^*)} |f(x)|e^{-\psi^*(x)} < \infty$.

Proof. We prove the contrapositive. Let f be measurable with $\sup_{x \in \mathrm{dom}(\psi^*)} |f(x)|e^{-\psi^*(x)} = \infty$. Then for each $n \in \mathbb{N}$, there exists $x_n \in \mathrm{dom}(\psi^*)$ such that $|f(x_n)| \geq 2^n e^{\psi^*(x_n)}$. Furthermore, Lemma 4.5 yields $y_n \geq 0$ such that the probability measure $\nu_n = p_n \delta_{x_n} + (1-p_n)\delta_{-x_n-y_n}$ with $p_n = \frac{1}{2}e^{-\psi^*(x_n)}$ is sub-ψ. Then so is the mixture $\mu = \sum_{n \in \mathbb{N}} 2^{-n} \nu_n$. On the other hand,

$$\int_{\mathcal{X}} |f(x)|\mu(dx) = \sum_{n \in \mathbb{N}} 2^{-n} \int_{\mathcal{X}} |f(x)|\nu_n(dx) \geq \sum_{n \in \mathbb{N}} 2^{-n} p_n |f(x_n)| \geq \sum_{n \in \mathbb{N}} p_n e^{\psi^*(x_n)} = \infty.$$

This shows that f does not belong to $\mathcal{L}^\Phi$, and completes the proof. $\qquad\square$

Proof of Theorem 4.3(i). We will make use of the space

$$E = \{\mu \in \mathcal{M} : \int_{\mathcal{X}} e^{\psi^*} d|\mu| < \infty\},$$

which is a Banach space with the weighted total variation norm $\|\mu\| = \int_{\mathcal{X}} e^{\psi^*} d|\mu|$. The Banach space dual E' is itself a Banach space with the dual norm $\|\varphi\|' = \sup\{\varphi(\mu) : \mu \in E, \|\mu\| \leq 1\}$, and admits the weak* topology $\sigma(E', E)$; see Appendix A.3. The positive cones of E and E' are $E_+ = \{\mu \in E : \mu \geq 0\}$ and $E'_+ = \{\varphi \in E' : \varphi(\mu) \geq 0 \text{ for all } \mu \in E_+\}$.

For any $f \in \mathcal{L}^\Phi$ and $\mu \in E$ we have

$$\int_{\mathcal{X}} |f| d|\mu| = \int_{\mathcal{X}} |f| e^{-\psi^*} e^{\psi^*} d|\mu| \leq c_f \|\mu\|, \tag{4.4}$$

where $c_f = \sup |f| e^{-\psi^*}$ is finite thanks to Lemma 4.7. From (4.4) it follows that every $f \in \mathcal{L}^\Phi$ defines a bounded linear functional $\varphi_f(\mu) = \int_{\mathcal{X}} f d\mu$ on E, and we may thus regard $\mathcal{L}^\Phi$ as a subspace of E'. In particular, E' contains $\mathcal{G}$, which is defined in (4.2) and is a subset of $\mathcal{L}^\Phi$. We will show below that

$$\mathcal{G} - E'_+ \text{ is } \sigma(E', E)\text{-closed.} \tag{4.5}$$

Once this has been done, the proof of Theorem 4.3(i) is completed as follows. Observe that $\mathcal{L}^\Phi \cap (\mathcal{G} - E'_+) = \mathcal{G} - \mathcal{L}^\Phi \cap E'_+ = \mathcal{G} - \mathcal{L}^\Phi_+$, where the equality $\mathcal{L}^\Phi \cap E'_+ = \mathcal{L}^\Phi_+$ holds because a function $f \in \mathcal{L}^\Phi$ is nonnegative if and only if $\int_{\mathcal{X}} f d\mu \geq 0$ for all $\mu \in E_+$. Consequently, (4.5) implies that $\mathcal{G} - \mathcal{L}^\Phi_+$ is closed in $\sigma(\mathcal{L}^\Phi, E)$, which is the trace of $\sigma(E', E)$ on $\mathcal{L}^\Phi$. Now, thanks to (4.4), E is a subset of $\mathcal{M}^\Phi$. Thus the topology $\sigma(\mathcal{L}^\Phi, E)$ is weaker than $\sigma(\mathcal{L}^\Phi, \mathcal{M}^\Phi)$, and we conclude that $\mathcal{G} - \mathcal{L}^\Phi_+$ is closed in the latter topology as well. This establishes (4.3) and proves the first part of the theorem.

We are left with proving (4.5). Thanks to the Krein–Šmulian theorem (see Theorem A.3), we only need to show that

$$(\mathcal{G} - E'_+) \cap B'_r \text{ is } \sigma(E', E)\text{-closed for every } r \in (0, \infty),$$

where $B'_r = \{\varphi \in E' : \|\varphi\|' \leq r\}$ is the centered closed dual ball of radius r. Fix any $r \in (0, \infty)$. Let $\varphi = g - \eta \in (\mathcal{G} - E'_+) \cap B'_r$ be arbitrary, and observe that for any $\mu \in E_+$ with $\|\mu\| \leq 1$ we have

$$-r \leq -\|\varphi\|' \leq \varphi(\mu) = \int_\mathcal{X} g d\mu - \eta(\mu) \leq \int_\mathcal{X} g d\mu.$$

By taking $\mu = \delta_x$ we find that $g \geq -r$ pointwise. We conclude from this that

$$(\mathcal{G} - E'_+) \cap B'_r = (\mathcal{G}_r - E'_+) \cap B'_r,$$

where $\mathcal{G}_r = \{g \in \mathcal{G} : g \geq -r\}$. We will argue that $\mathcal{G}_r$ is $\sigma(E', E)$-compact. This will conclude the proof because E'_+ and B'_r are both $\sigma(E', E)$-closed, the sum of a compact set and a closed set is closed, and the intersection of two closed sets is closed.

To show that $\mathcal{G}_r$ is $\sigma(E', E)$-compact, we define

$$K_r = \{\pi \in \mathcal{M}_+(\Lambda) : \pi(\Lambda) \leq r\},$$

where we recall that $\Lambda = [0, \lambda_{\max}]$ is compact. Then K_r is a compact subset of $\mathcal{M}_+(\Lambda)$, equipped with the usual weak topology induced by duality with the set of real-valued continuous functions on Λ. Next, define the map

$$T : \pi \mapsto T(\pi) = \int_\Lambda g_\lambda \pi(d\lambda)$$

from K_r to $\mathcal{L}^\Phi \subset E'$. Here we identify $T(\pi)$ with the linear functional $\varphi_{T(\pi)}(\mu) = \int_\mathcal{X} T(\pi) d\mu$ on E. We claim that

$$\mathcal{G}_r = T(K_r). \tag{4.6}$$

The inclusion '$\supset$' is clear since $g_\lambda \geq -1$ for all λ. For the inclusion '$\subset$', consider any $g \in \mathcal{G}_r$, that is, $g = \int_\Lambda g_\lambda \pi(d\lambda) \geq -r$ for some $\pi \in \mathcal{M}_+(\Lambda)$. Now, for all $x < 0$ and $\lambda \in (0, \lambda_{\max}]$, we have $-1 \leq g_\lambda(x) \leq 0$ and $\lim_{x \to -\infty} g_\lambda(x) = -1$, while $g_0(x) = 0$ for all x. The dominated convergence theorem then yields $-r \leq \lim_{x \to -\infty} g(x) = -\pi((0, \lambda_{\max}])$. Thus the measure $\pi' = \pi(\cdot \cap (0, \lambda_{\max}])$ belongs to K_r, and we have $g = \int_\Lambda g_\lambda \pi'(d\lambda)$. This completes the proof of (4.6).

Next, we claim that T is continuous when E' is equipped with $\sigma(E', E)$. This is the initial topology generated by the maps $\varphi \mapsto \varphi(\mu)$, $\mu \in E$, so to show continuity it suffices to show that the composition

$$\pi \mapsto \varphi_{T(\pi)}(\mu) = \int_\mathcal{X} T(\pi) d\mu = \int_\Lambda \int_\mathcal{X} g_\lambda(x) \mu(dx) \pi(d\lambda) \tag{4.7}$$

from $\mathcal{M}_+(\Lambda)$ to $\mathbb{R}$ is continuous for every $\mu \in E$. (We used Fubini's theorem to interchange the integrals on the right-hand side of (4.7).) The map $\lambda \mapsto \int_\mathcal{X} g_\lambda(x) \mu(dx)$ is continuous on the compact set Λ. This follows from the dominated convergence theorem because g_λ is continuous in λ and dominated in absolute value by e^{ψ^*} which is μ-integrable by definition of E. Thus by definition of the weak topology on $\mathcal{M}_+(\Lambda)$, the map in (4.7) is continuous. We conclude that T is continuous, and hence that $\mathcal{G}_r = T(K_r)$ is $\sigma(E', E)$-compact. This completes the proof of Theorem 4.3(i). $\qquad \square$

Proof of Theorem 4.3(ii). Consider a $[0, \infty]$-valued measurable function h that is pointwise dominated on $\mathrm{dom}(\psi^*)$ by $\int_\Lambda e^{\lambda x - \psi(\lambda)} \pi(d\lambda)$ for some $\pi \in \mathcal{M}_1(\Lambda)$. Tonelli's theorem and the definition of $\mathcal{P}$ then yields $\int_{\mathcal{X}} h d\mu \leq 1$ for all $\mu \in \mathcal{P}$, showing that h is an e-variable.

Conversely, let h be an e-variable and set $f = h - 1$. Then by part (i) of the theorem, there is some $\pi' \in \mathcal{M}_+(\Lambda)$ such that $f(x) \leq \int_\Lambda g_\lambda(x) \pi'(d\lambda)$ for all $x \in \mathrm{dom}(\psi^*)$. Since $f \geq -1$, the argument after (4.6) with $r = 1$ yields $\pi'((0, \lambda_{\max}]) \leq 1$. Thus the measure $\pi = \pi'(\cdot \cap (0, \lambda_{\max}]) + (1 - \pi'((0, \lambda_{\max}]))\delta_0$ belongs to $\mathcal{M}_1(\Lambda)$. Since $g_0(x) = 0$ for all x, we have $\int_\Lambda g_\lambda(x) \pi'(d\lambda) = \int_\Lambda g_\lambda(x) \pi(d\lambda)$, and thus $h(x) = 1 + f(x) \leq \int_\Lambda (1 + g_\lambda(x)) \pi(d\lambda) = \int_\Lambda e^{\lambda x - \psi(\lambda)} \pi(d\lambda)$, for all $x \in \mathrm{dom}(\psi^*)$. $\qquad\square$

5 Distributions invariant under a group of symmetries

Let Σ be a compact topological group acting (from the left) on the measurable space $\mathcal{X}$. This means that every group element $\sigma \in \Sigma$ induces a map $x \mapsto \sigma x$ from $\mathcal{X}$ to itself, the identity element of Σ induces the identity map, and one has $(\sigma_1 \sigma_2)x = \sigma_1(\sigma_2 x)$ for all $\sigma_1, \sigma_2 \in \Sigma$ and $x \in \mathcal{X}$. We assume that the group action is measurable, meaning that the map $(\sigma, x) \mapsto \sigma x$ is jointly measurable, where Σ is equipped with its Borel σ-algebra. Since Σ is compact, it admits a unique left Haar probability measure π. Here are two examples of such group actions.

Example 5.1. (i) *The symmetric group on n elements $\Sigma(n)$ acts on vectors in $\mathbb{R}^n$ by permuting the components. Its Haar probability measure is the normalized counting measure on $\Sigma(n)$.*

(ii) *The special orthogonal group $SO(n)$ acts on $\mathbb{R}^n$ by rotations. Its Haar probability measure is the uniform distribution on $SO(n)$.*

We use the left Haar probability measure π to symmetrize measures and to average functions. First, for any measurable function f bounded below, we define its *orbit average* function f_π by

$$f_\pi(x) = \int_\Sigma (\sigma^* f)(x) \pi(d\sigma),$$

where $(\sigma^* f)(x) = f(\sigma x)$ is the *pullback* of f under the map $x \mapsto \sigma x$. Thus $f_\pi(x)$ is indeed the average of f over the orbit $O_x = \{\sigma x : \sigma \in \Sigma\}$ of x. Next, there is a dual operation on measures (we focus on probability measures for simplicity). For any $\mu \in \mathcal{M}_1$ we define its *symmetrization* $\mu_\pi \in \mathcal{M}_1$ by

$$\mu_\pi(A) = \int_\Sigma (\sigma_* \mu)(A) \pi(d\sigma),$$

where $(\sigma_* \mu)(A) = \mu(\sigma^{-1} A)$ is the *pushforward* of μ under the map $x \mapsto \sigma x$. Here $\sigma^{-1} A = \{\sigma^{-1} x : x \in A\}$. The fact that μ_π has unit mass is seen by taking $A = \mathcal{X}$ and using that μ and π both have unit mass. The following lemma records some basic properties of the orbit averaging and symmetrization operations.

Lemma 5.2. *Let f be a measurable function bounded below and let $\mu \in \mathcal{M}_1$.*

(i) *The symmetrization μ_π is Σ-invariant in the sense that $\sigma_* \mu_\pi = \mu_\pi$ for all $\sigma \in \Sigma$.*

15

(ii) *One has the adjoint identity*

$$\int_{\mathcal{X}} f_\pi d\mu = \int_{\mathcal{X}} f d\mu_\pi. \tag{5.1}$$

Proof. (i): For any $\sigma \in \Sigma$ and any measurable set $A \subset \mathcal{X}$, one has $\int_\Sigma \mu((\sigma\rho)^{-1}A)\pi(d\rho) = \int_\Sigma \mu(\rho^{-1}A)\pi(d\rho)$ thanks to the left-invariance of π. The left-hand side equals $(\sigma_*\mu_\pi)(A)$ and the right-hand side equals $\mu_\pi(A)$, showing that the two are equal.

(ii): Linearity and the definition of pushforward yield $\int_{\mathcal{X}} f d\mu_\pi = \int_\Sigma \int_{\mathcal{X}} f(\sigma x)\mu(dx)\pi(d\sigma)$ for every simple function f, and then for every bounded measurable f by the monotone class theorem. On the other hand, Fubini's theorem yields $\int_{\mathcal{X}} f_\pi d\mu = \int_\Sigma \int_{\mathcal{X}} f(\sigma x)\mu(dx)\pi(d\sigma)$ for bounded measurable f. This shows (5.1) for all such f. For f unbounded above, just apply (5.1) with $f \wedge n$ in place of f, send n to infinity, and use monotone convergence. $\square$

Remark 5.3. Given the Σ-invariance property (i) of μ_π, it is perhaps surprising that the analogous property does *not* hold for f_π in the sense that $\sigma^* f_\pi$ and f_π are not equal in general. They are however equal if π is a *right* Haar measure, since then $f_\pi(\sigma x) = \int_\Sigma f_\pi(\rho\sigma x)\pi(d\rho) = \int_\Sigma f_\pi(\rho x)\pi(d\rho) = f_\pi(x)$, using the right-invariance of π in the second step. If the group Σ is unimodular, for example, if it is a discrete group, then π is both a left and right Haar measure, and thus $\sigma^* f_\pi = f_\pi$.

We are interested in describing the set of e-variables for the hypothesis consisting of all Σ-invariant distributions,

$$\mathcal{P} = \{\mu \in \mathcal{M}_1 \colon \mu = \sigma_*\mu \text{ for all } \sigma \in \Sigma\}.$$

Such classes, or infinite-sample versions of them, have been studied in many recent works. For example, testing exchangeability (Vovk, 2021; Ramdas et al., 2022b; Saha and Ramdas, 2024), two-sample and independence testing (Shekhar and Ramdas, 2023; Podkopaev et al., 2023; Podkopaev and Ramdas, 2023), but there are also papers that study this class in an abstract and general manner (Koning, 2023; Pandeva et al., 2024) like we do above. We note the subtle fact that our setting is different from the case where $\mu \neq \sigma_*\mu$, but $\sigma_*\mu \in \mathcal{P}$ whenever $\mu \in \mathcal{P}$, for which the term 'group invariance' is also used (Pérez-Ortiz et al., 2024).

Example 5.4. *A finite sequence of random variables $X_1, \ldots, X_n$ is exchangeable if and only if its joint distribution on $\mathbb{R}^n$ is invariant under the symmetric group $\Sigma(n)$. An infinite sequence $(X_n)_{n \in \mathbb{N}}$ is exchangeable if and only if its joint distribution on $\mathbb{R}^\mathbb{N}$ is invariant under the group Σ of finite permutations of the elements of $\mathbb{R}^\mathbb{N}$.*

The following lemma is the key to characterizing the set of e-variables for $\mathcal{P}$.

Lemma 5.5. *A distribution $\mu \in \mathcal{M}_1$ belongs to $\mathcal{P}$ if and only if $\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f_\pi d\mu$ for all measurable functions f bounded below.*

Proof. Fix $\mu \in \mathcal{M}_1$. We have the following chain of equivalences:

$$\mu \in \mathcal{P} \quad \Leftrightarrow \quad \mu = \mu_\pi \quad \Leftrightarrow \quad \int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f d\mu_\pi \text{ for all measurable } f \text{ bounded below,}$$

where the first equivalence follows from the definition of $\mathcal{P}$ and of μ_π for the forward implication, and Lemma 5.2(i) for the reverse implication. Thanks to (5.1), the third statement in the above display is in turn equivalent to $\int_{\mathcal{X}} f d\mu = \int_{\mathcal{X}} f_\pi d\mu$ for all measurable functions f bounded below. This completes the proof. $\square$

16

Theorem 5.6. *The set $\mathcal{E}$ of all e-variables for $\mathcal{P}$ consists precisely of those $[0, \infty]$-valued measurable functions that are pointwise dominated by an e-variable of the form*

$$1 + f - f_\pi$$

for some $[-1, \infty]$-valued measurable f such that $f_\pi \leq 0$. Moreover, every function of this form is an exact *e-variable, meaning that the e-variable property holds with equality for all $\mu \in \mathcal{P}$.*

Proof. Let f be a $[-1, \infty]$-valued measurable function such that $f_\pi \leq 0$, and note that, in addition, $f_\pi \geq -1$. This ensures that $h = 1 + f - f_\pi$ is a well-defined $[0, \infty]$-valued function, and that we may compute $\int_{\mathcal{X}} h d\mu = 1 + \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f_\pi d\mu = 1$ for every $\mu \in \mathcal{P}$, using Lemma 5.5 in the last step. This shows that every function of this form is an exact e-variable.

Conversely, let h be an e-variable and set $f = h - 1$. Fix any $x_0 \in \mathcal{X}$ and consider the symmetrization $\mu = (\delta_{x_0})_\pi$ of the Dirac mass at x_0, which belongs to $\mathcal{P}$ due to Lemma 5.2(i). Thanks to (5.1) and the e-variable property of h we have $f_\pi(x_0) = \int_{\mathcal{X}} f_\pi(x) \delta_{x_0}(dx) = \int_{\mathcal{X}} f(x)(\delta_{x_0})_\pi(dx) = \int_{\mathcal{X}} f d\mu \leq 0$. Thus $f_\pi \leq 0$, and we have $h \leq 1 + f - f_\pi$ pointwise. $\square$

So far we have not made use of the abstract characterization of e-variables, Theorem 2.2. Indeed, we were able to describe $\mathcal{E}$ completely without it. We may however use the abstract theorem in a different way: our next result provides a general method of identifying constraint sets Φ that generate $\mathcal{P}$, and the abstract theorem then ensures that any e-variable can be approximated in the weak sense using conic combinations of the constraint functions.

A set $\mathcal{F}$ of bounded measurable functions is a *separating set* for $\mathcal{M}$ if, for any $\mu \in \mathcal{M}$, one has $\mu = 0$ if and only if $\int_{\mathcal{X}} f d\mu = 0$ for all $f \in \mathcal{F}$. Such a set separates any two distinct measures μ_1, μ_2 in the sense that there is some $f \in \mathcal{F}$ such that $\int_{\mathcal{X}} f d\mu_1 \neq \int_{\mathcal{X}} f d\mu_2$. Next, a *generating set* for Σ is a subset Σ_0 such that any $\sigma \in \Sigma$ can be expressed as $\sigma = \sigma_1 \sigma_2 \cdots \sigma_n$ for some $n \in \mathbb{N}$ and $\sigma_1, \ldots, \sigma_n \in \Sigma_0$.

Theorem 5.7. *Let $\mathcal{F}$ be a separating set for $\mathcal{M}$, and Σ_0 a generating set for Σ. A distribution μ belongs to $\mathcal{P}$ if and only if $\int_{\mathcal{X}} (f(\sigma x) - f(x)) \mu(dx) = 0$ for all $\sigma \in \Sigma_0$ and $f \in \mathcal{F}$. In other words, $\mathcal{P}$ is generated by the constraint set*

$$\Phi = \{\sigma^* f - f : \sigma \in \Sigma_0, \ f \in \mathcal{F} \cup (-\mathcal{F})\}.$$

Proof. The forward implication follows because for any $\mu \in \mathcal{P}$ and any bounded measurable function f, $\int_{\mathcal{X}} f(\sigma x) \mu(dx) = \int_{\mathcal{X}} f(x)(\sigma_* \mu)(dx) = \int_{\mathcal{X}} f(x) \mu(dx)$. We thus focus on the reverse implication and assume that

$$\int_{\mathcal{X}} (f(\sigma x) - f(x)) \mu(dx) = 0 \tag{5.2}$$

for all $\sigma \in \Sigma_0$ and $f \in \mathcal{F}$, where $\mu \in \mathcal{M}_1$ is fixed. We let $\mathcal{L}_b$ denote the space of all bounded measurable functions on $\mathcal{X}$.

Because $\mathcal{F}$ is separating, its span is $\sigma(\mathcal{L}_b, \mathcal{M})$-dense in $\mathcal{L}_b$. Indeed, if the span were not dense, the Hahn–Banach theorem would yield a nonzero measure vanishing on the span, contradicting that $\mathcal{F}$ is separating. Furthermore, the map $f \mapsto \int_{\mathcal{X}} f(\sigma x) \mu(dx) = \int_{\mathcal{X}} f d(\sigma_* \mu)$ is $\sigma(\mathcal{L}_b, \mathcal{M})$-continuous by definition of the topology. Combining these two facts, we deduce that (5.2) holds for all $\sigma \in \Sigma_0$ and all $f \in \mathcal{L}_b$.

Next, fix any $\sigma \in \Sigma$ and $f \in \mathcal{L}_b$, write $\sigma = \sigma_1 \sigma_2 \cdots \sigma_n$ for some $\sigma_1, \ldots, \sigma_n \in \Sigma_0$, and set $f_i(x) = f(\sigma_1 \cdots \sigma_i x)$ for $i = 1, \ldots, n-1$ as well as $f_0(x) = f(x)$. We then have $f_i(x) = f_{i-1}(\sigma_i x)$ and $f_{n-1}(\sigma_n x) = f(\sigma x)$, and hence also the telescoping sum

$$f(\sigma x) - f(x) = \sum_{i=0}^{n-1}(f_i(\sigma_i x) - f_i(x)).$$

Since each f_i belongs to $\mathcal{L}_b$, we may use that (5.2) holds for functions in $\mathcal{L}_b$ and group elements in Σ_0 to obtain

$$\int_{\mathcal{X}}(f(\sigma x) - f(x))\mu(dx) = \sum_{i=0}^{n-1}\int_{\mathcal{X}}(f_i(\sigma_i x) - f_i(x))\mu(dx) = 0.$$

This shows that (5.2) actually holds for all $\sigma \in \Sigma$ and all $f \in \mathcal{L}_b$. By integrating over Σ and using Fubini's theorem, we obtain $\int_{\mathcal{X}} f_\pi d\mu = \int_{\mathcal{X}} f d\mu$ for all $f \in \mathcal{L}_b$, and then by monotone convergence for all measurable f bounded below. Lemma 5.5 now yields $\mu \in \mathcal{P}$. $\qquad\square$

6   Admissible e-variables and complete classes

Consider a hypothesis $\mathcal{P}$ and its set of e-variables $\mathcal{E}$. Recall that an e-variable $h \in \mathcal{E}$ is *admissible* if any other $h' \in \mathcal{E}$ with $h' \geq h$, $\mathcal{P}$-q.s., actually satisfies $h' = h$, $\mathcal{P}$-q.s. The results in this paper, in particular Theorem 3.1, Corollary 3.3, Theorem 3.10, Theorem 4.3, and Theorem 5.6, lead to conditions for admissibility. We now elaborate on this aspect of our results and show that we actually obtain *minimal complete classes*, defined below, for the hypotheses under consideration. This generalizes Clerico (2024a, Section 3).

Definition 6.1. A *complete class* of e-variables for $\mathcal{P}$ is a subset $\mathcal{E}' \subset \mathcal{E}$ such that every $h \in \mathcal{E}$ is $\mathcal{P}$-q.s. dominated by some $h' \in \mathcal{E}'$. A complete class $\mathcal{E}'$ is *minimal* if removing an e-variable (and all its $\mathcal{P}$-versions) from $\mathcal{E}'$ renders the class non-complete.

For statistical applications it suffices to work with complete classes, and minimal complete classes are preferable. Clearly $\mathcal{E}$ itself is a complete class, but it is not minimal. Existence of a minimal complete class has some useful implications, recorded in the following result which extends Clerico (2024a, Lemma 3.3) to our more general class of hypotheses.

Lemma 6.2. *A minimal complete class $\mathcal{E}'$ exists if and only if every e-variable is $\mathcal{P}$-q.s. dominated by an admissible e-variable. In this case, every $h' \in \mathcal{E}'$ is admissible, and every admissible $h \in \mathcal{E}$ has a version $h' \in \mathcal{E}'$. In particular, a minimal complete class is unique provided different $\mathcal{P}$-versions of any e-variable are identified.*

Proof. We drop the "$\mathcal{P}$-q.s." qualifier for brevity. If every e-variable is dominated by an admissible e-variable, then $\mathcal{E}' = \{h' \in \mathcal{E} : h'$ is admissible$\}$ is a complete class. It is also minimal, simply because admissible e-variables cannot be strictly dominated. Conversely, if a minimal complete class $\mathcal{E}'$ exists, then by completeness, every $h \in \mathcal{E}$ is dominated by some $h' \in \mathcal{E}'$, but by minimality, h' cannot be further strictly dominated, so it is admissible. This reasoning also yields the remaining statements. $\qquad\square$

Any hypothesis with a reference measure admits a minimal complete class. This extends Clerico (2024a, Lemma 3.6), with a similar proof.

Theorem 6.3. *Suppose $\mathcal{P}$ admits a reference measure $\bar{\mu} \in \mathcal{M}_1$, i.e., $\mu \ll \bar{\mu}$ for all $\mu \in \mathcal{P}$. Then a minimal complete class exists.*

Proof. We use an argument from Ramdas et al. (2022a) based on transfinite induction to show that every e-variable is dominated by an admissible e-variable. The result then follows from Lemma 6.2. We pick any $h \in \mathcal{E}$ and recursively define an increasing transfinite sequence of e-variables $h_\beta \in \mathcal{E}$ indexed by the countable ordinals β. First, $h_0 = h$. For a successor ordinal $\beta = \alpha + 1$ such that h_α has already been defined, if h_α is admissible we set $h_\beta = h_\alpha$, otherwise we choose any $h_\beta \in \mathcal{E}$ that strictly dominates h_α. (I.e., $h_\beta \geq h_\alpha$, $\bar{\mu}$-a.s., and $h_\beta > h_\alpha$ with positive $\bar{\mu}$-probability.) For any countable limit ordinal $\beta = \lim_n \alpha_n$ such that $h_\alpha \in \mathcal{E}$ has been defined for all $\alpha < \beta$, we let $h_\beta = \lim_n h_{\alpha_n}$. By the monotone convergence theorem, $h_\beta \in \mathcal{E}$. By transfinite induction, $h_\beta \in \mathcal{E}$ for every countable ordinal β. Consider now the decreasing $[0,1]$-valued transfinite sequence $c_\beta = \int_{\mathcal{X}} e^{-h_\beta} d\bar{\mu}$. This sequence must become constant for all β beyond some countable ordinal β_0; otherwise we would have an uncountable strictly decreasing sequence in $[0,1]$, which is impossible. By construction, h_{β_0} is then admissible and dominates h. $\square$

Unfortunately, as the following example shows, there are hypotheses for which a minimal complete class does not exist. A different example, involving a certain non-measurable set, is given by Clerico (2024a, Appendix C).

Example 6.4. *Let $\mathcal{X} = [0,1]$ and let $\mathcal{P}$ consist of the standard uniform distribution together with all measures of the form $(\delta_0 + \delta_x)/2$ for $x \in [0,1]$. Then $\mathcal{E}$ consists of all nonnegative measurable functions h with $\int_0^1 h(x)dx \leq 1$ and $h(0) + h(x) \leq 2$ for all $x \in [0,1]$. Note that $x = 0$ is included here, so that $h(0) \leq 1$.*

We claim that if h is an e-variable with $h(0) < 1$, then it cannot be admissible. To show this, assume $h(0) < 1$ and let $y \in (0,1]$ be such that $h(y) < 2 - h(0)$; such y must exist, because otherwise $\int_0^1 h(x)dx = 2 - h(0) > 1$, contradicting the e-variable property. Then, the function g with $g(y) = 2 - h(0)$ and $g = h$ elsewhere is still an e-variable, and it strictly dominates h under $(\delta_0 + \delta_y)/2$. So h is not admissible.

We deduce from this that the only way an e-variable h can be admissible is if $h(0) = 1$. But then $h(x) \leq 2 - h(0) = 1$ for all $x \in (0,1]$ so that, as a result, the only admissible e-variable is the trivial e-variable equal to one. Thus if $h(x) > 1$ for some $x \in (0,1]$ (e.g. $h(x) = 2$ and zero otherwise) then h cannot be admissible, and it cannot be dominated by an admissible e-variable. Consequently, due to Lemma 6.2, no minimal complete class exists for this hypothesis.

In view of this example, it is remarkable that the hypotheses considered in Sections 3–5 do admit minimal complete classes under mild additional assumptions. (We conjecture that minimal complete classes exist even without these assumptions.)

Theorem 6.5.   (i) *With the setup and notation of Section 3, the set of nonnegative functions of the form $1 + \sum_{i=1}^d \pi_i g_i$, with $\pi \in \mathbb{R}_+^d$, is a complete class for the hypothesis generated by $g_1, \ldots, g_d$. If (3.3) holds, this is a minimal complete class.*

(ii) *With the setup and notation of Section 4, let $\mathcal{E}'$ be the set of functions of the form $\int_\Lambda h_\lambda(x)\pi(d\lambda)$ for $x \in \operatorname{dom}(\psi^*)$, where $h_\lambda(x) = e^{\lambda x - \psi(\lambda)}$ and $\pi \in \mathcal{M}_1(\Lambda)$.[2] Then $\mathcal{E}'$ is a complete class for the one-sided sub-ψ hypothesis. Assume now that on $\mathbb{R}_+$, ψ is the cumulant generating function of some zero-mean distribution $\bar{\mu}$ whose support has a cluster point in the interior of $\operatorname{dom}(\psi^*)$. There are two cases depending on whether $\operatorname{dom}(\psi^*)$ contains its right endpoint:*

   (1) *if $\operatorname{dom}(\psi^*) = (-\infty, \bar{x}]$ with $\bar{x} \in [0, \infty)$, then $\mathcal{E}'$ is a minimal complete class.*

   (2) *if $\operatorname{dom}(\psi^*) = (-\infty, \bar{x})$ with $\bar{x} \in [0, \infty]$, then $\mathcal{E}'_0$ is a minimal complete class, where $\mathcal{E}'_0 \subset \mathcal{E}'$ is the set of all functions of the form $\int_{\Lambda_0} h_\lambda(x)\pi(d\lambda)$ for $x \in \operatorname{dom}(\psi^*)$, with $\pi \in \mathcal{M}_1(\Lambda_0)$ and $\Lambda_0 = [0, \lambda_{\max})$.*

(iii) *With the setup and notation of Section 5, the set of e-variables of the form $1 + f - f_\pi$, with f being $[-1, \infty]$-valued measurable and such that $f_\pi \le 0$, is a minimal complete class for the hypothesis of Σ-invariant distributions.*

Proof. (i): This follows directly from Corollary 3.3 (or Theorem 3.10 in the setting with a reference measure) and Lemma 6.2.

(ii): Theorem 4.3(ii) yields that $\mathcal{E}'$ is a complete class. Assume now that ψ is the cumulant generating function of some distribution $\bar{\mu}$ as in the statement. We first consider case (1). The minimality of $\mathcal{E}'$ will follow from Lemma 6.2 once we show that all e-variables in $\mathcal{E}'$ are admissible. To this end, pick $h \in \mathcal{E}'$ specified by some $\pi \in \mathcal{M}_1(\Lambda)$, meaning that $h(x) = \int_\Lambda h_\lambda(x)\pi(d\lambda)$ for $x \in \operatorname{dom}(\psi^*)$. Let h' be an e-variable that dominates h. By completeness, h' is further dominated by some $h'' \in \mathcal{E}'$, specified by some $\pi'' \in \mathcal{M}_1(\Lambda)$. We thus have $h'' - h \ge 0$ on $\operatorname{dom}(\psi^*)$. We claim that $\int_\mathcal{X}(h'' - h)d\bar{\mu} = 0$. To see this, note that $\psi(\lambda)$ is the cumulant generating function of $\bar{\mu}$ at each $\lambda \in [0, \lambda_{\max})$, so that $\int_\mathcal{X} h_\lambda(x)\bar{\mu}(dx) = 1$. Moreover, because $h_\lambda(x) \le h_\lambda(\bar{x}) \le e^{\psi^*(\bar{x})} < \infty$ for $x \in \operatorname{dom}(\psi^*)$ and $\lambda \in [0, \lambda_{\max})$, the dominated convergence theorem yields $\int_\mathcal{X} h_{\lambda_{\max}}(x)\bar{\mu}(dx) = 1$ as well. Thus by Tonelli's theorem,

$$\int_\mathcal{X}(h''(x) - h(x))\bar{\mu}(dx) = \int_\Lambda \int_\mathcal{X} h_\lambda(x)\bar{\mu}(dx)(\pi'' - \pi)(d\lambda) = \pi''(\Lambda) - \pi(\Lambda) = 0.$$

Hence $h''(x) = h(x)$ for $\bar{\mu}$-a.e. x, and then, by continuity, for all $x \in \operatorname{supp}(\bar{\mu})$. Since h and h'' are real analytic in $(-\infty, \bar{x})$, and since $\operatorname{supp}(\bar{\mu})$ has a cluster point there, the identity theorem for real analytic functions (see e.g. Krantz and Parks (2002, Corollary 1.2.7)) implies that $h'' = h$ on $(-\infty, \bar{x})$ and then by continuity on $\operatorname{dom}(\psi^*)$. It follows that $h' = h$ on $\operatorname{dom}(\psi^*)$, showing that h is admissible.

Consider now case (2). Then $h_{\lambda_{\max}}(x) = 0$ for all $x \in \operatorname{dom}(\psi^*)$. Indeed, if $\lambda_{\max} < \infty$, this follows from the fact that $\psi(\lambda) \to \infty$ as $\lambda \uparrow \lambda_{\max}$ by virtue of being a cumulant generating function. If $\lambda_{\max} = \infty$, we pick $y \in (x, \bar{x})$ and note that $h_\lambda(x) \le e^{-\lambda(y-x)+\psi^*(y)}$, which tends to zero as $\lambda \to \infty$. As a result, any $h \in \mathcal{E}'$, specified by $\pi \in \mathcal{M}_1(\Lambda)$, is dominated by $h' \in \mathcal{E}'_0$ specified by $\pi' = \pi(\cdot \cap \Lambda_0) + (1 - \pi(\{\lambda_{\max}\}))\delta_0 \in \mathcal{M}_1(\Lambda_0)$. Thus $\mathcal{E}'_0$ is a complete class. The proof of minimality works as in case (1) on replacing $\mathcal{E}$ by $\mathcal{E}_0$ and Λ by Λ_0. (Note that since Λ_0 does not contain $\lambda_{\max}$, there is no need to establish $\int_\mathcal{X} h_{\lambda_{\max}}(x)\bar{\mu}(dx) = 1$, which does not hold in the setting of case (2).)

---

[2]Recall that $\Lambda = [0, \lambda_{\max}]$ and that $h_{\lambda_{\max}}(x) = \lim_{\lambda \uparrow \lambda_{\max}} e^{\lambda x - \psi(\lambda)}$ for $x \in \mathcal{X}$, where $\mathcal{X} = \operatorname{dom}(\psi^*)$.

(iii): This follows directly from Theorem 5.6 and Lemma 6.2, noting that exact e-variables (see Theorem 5.6) are necessarily admissible. $\qquad\square$

7   Optimal e-variables

We now consider the problem of finding *optimal* e-variables in the set $\mathcal{E}$ of all e-variables for a given hypothesis $\mathcal{P}$. We work with a general class of objective functions defined in terms of the following data:

- *Alternative hypothesis*: any family $\mathcal{Q} \subset \mathcal{M}_1$ of probability measures such that $\mathcal{Q} \ll \mathcal{P}$, meaning that every $\mathcal{P}$-negligible set is also $\mathcal{Q}$-negligible.

- *Utility function*: a nondecreasing upper semicontinuous function $U \colon [0, \infty) \to [-\infty, \infty)$ that is linearly bounded in the sense that there exist $a \in \mathbb{R}$ and $b > 0$ such that
$$U(x) \le a + bx \text{ for all } x \in [0, \infty).$$

- *Penalty function*: an arbitrary function $\chi \colon \mathcal{Q} \to \mathbb{R}$.

The objective function is defined in terms of these data as
$$\mathbb{U}(h) = \inf_{\mathsf{Q} \in \mathcal{Q}} \left\{ \int_{\mathcal{X}} U(h) d\mathsf{Q} - \chi(\mathsf{Q}) \right\}, \quad h \in \mathcal{E},$$

with the convention that the integral is $-\infty$ whenever $\int_{\mathcal{X}} U(h)^- d\mathsf{Q} = \infty$. This covers several criteria appearing in the literature:

Example 7.1. (i) *With a simple hypothesis $\mathcal{Q} = \{\mathsf{Q}\}$, log-utility $U(x) = \log(x)$, and vanishing penalty function $\chi = 0$, one maximizes* e-power *(Ramdas and Wang, 2024, Ch. 3), thus obtaining (assuming the optimal value is finite) the log-optimal* GRO e-variable *or* numeraire *of Grünwald et al. (2024); Larsson et al. (2025).*

  (ii) *More generally, log-utility can be replaced with any proper concave function U. By concavity, such a function is automatically linearly bounded. Examples include the power utilities considered by Larsson et al. (2025, Section 6) and those used by Koning (2025).*

  (iii) *If, instead, we keep log-utility, let $\mathcal{Q}$ be composite, assume that both $\mathcal{P}$ and $\mathcal{Q}$ admit a common reference measure, and maintain a zero penalty function, we obtain the* GROW e-variable *problem of Grünwald et al. (2024). Of course, we can again work with general concave utility functions.*

  (iv) *Still in the composite setting with reference measure and log-utility $U(x) = \log(x)$, allowing a bounded penalty function χ yields the* REGROW e-variable *problem of Grünwald et al. (2024). The main example is $\chi(\mathsf{Q}) = \sup_{h \in \mathcal{E}} \int_{\mathcal{X}} U(h) d\mathsf{Q}$, the optimal e-power for the point alternative Q. The objective function $\mathbb{U}(h)$ can then be interpreted as (the negative of) the worst-case* regret *of using h instead of the e-variable with optimal e-power for the particular alternative $\mathsf{Q} \in \mathcal{Q}$ that actually generated the data. Note that our setup does not require the penality function χ to be bounded.*

(v) *Any* continuous variational preference *in the sense of* Maccheroni et al. (2006) *is represented by a utility* $\mathbb{U}(h)$ *of the above form, with U affine and χ a nonpositive function which, when extended to take the value $-\infty$ outside $\mathcal{Q}$, is concave and upper semicontinuous with weakly compact super-level sets; see* (Maccheroni et al., 2006, Sections 3.2 and 3.4) *for details.*

Because U is nondecreasing, so is $\mathbb{U}$ in the sense that $\mathbb{U}(h') \geq \mathbb{U}(h)$ whenever $h' \geq h$, $\mathcal{P}$-q.s. and hence $\mathcal{Q}$-q.s. For this reason it suffices to maximize $\mathbb{U}$ over a complete class of e-variables; see Definition 6.1. Several of the results in this paper yield complete classes $\mathcal{E}_0$ which are *compactly parameterized* in the sense that there is a compact set K_0 and a continuous map $\pi \mapsto h(\pi)$ from K_0 onto $\mathcal{E}_0$. The problem of maximizing $\mathbb{U}(h)$ over $\mathcal{E}$ thus reduces to maximizing $\mathbb{U}(h(\pi))$ over K_0. Thanks to this reduction, the existence of an optimal e-variable boils down to the standard fact that an upper semicontinuous function on a compact set achieves its supremum. In contrast, maximizing $\mathbb{U}(h)$ directly over $\mathcal{E}$ seems intractable in general, especially when $\mathcal{Q}$ is non-dominated or U is non-concave. We elaborate below on the details of this reduction in two cases: finitely generated hypotheses and one-sided sub-ψ hypotheses. After that we discuss uniqueness of the optimizer.

Finitely generated hypotheses. Consider a finite constraint set $\Phi = \{g_1, \ldots, g_d\}$ and let $\mathcal{P}$ be the hypothesis it generates. We know from Theorem 6.5(i), or directly from Theorem 3.1, that a complete class is given by all e-variables of the form

$$h = 1 + \sum_{i=1}^{d} \pi_i g_i$$

with $\pi \in \mathbb{R}_+^d$ belonging to the set Π^Φ defined in Corollary 3.3. Thanks to Lemma 3.2 we may always restrict π to the set K in (3.1). Defining $K_0 = K \cap \Pi^\Phi$, the problem of maximizing $\mathbb{U}(h)$ over $\mathcal{E}$ thus reduces to

$$\sup_{\pi \in K_0} u(\pi), \quad \text{where} \quad u(\pi) = \inf_{\mathsf{Q} \in \mathcal{Q}} \left\{ \int_{\mathcal{X}} U\left(1 + \sum_{i=1}^{d} \pi_i g_i\right) d\mathsf{Q} - \chi(\mathsf{Q}) \right\}. \tag{7.1}$$

The key advantage of this reduction is captured by the following lemma.

Lemma 7.2. *The set $K_0 = K \cap \Pi^\Phi$ is compact.*

Proof. Both K and Π^Φ are closed, so it suffices to prove that K_0 is bounded. The argument is similar to the proof of Theorem 3.1. Suppose for contradiction that there is a sequence $(\pi^n)_{n \in \mathbb{N}}$ in K_0 with $\|\pi^n\| \to \infty$, where $\|\pi^n\| = \sum_{i=1}^{d} \pi_i^n$. After passing to a subsequence and using that K_0 is closed and contains $\pi^n/\|\pi^n\|$ whenever $\|\pi^n\| \geq 1$, we obtain $\pi^n/\|\pi^n\| \to \rho$ for some $\rho \in K_0$ with $\|\rho\| = 1$. We then get $\sum_{i=1}^{d} \rho_i g_i = \lim_{n \to \infty} (1 + \sum_{i=1}^{d} \pi_i^n g_i)/\|\pi^n\| \geq 0$, $\mathcal{P}$-q.s. On the other hand, $\int_{\mathcal{X}} \sum_{i=1}^{d} \rho_i g_i d\mu \leq 0$ for all $\mu \in \mathcal{P}$. Thus $\sum_{i=1}^{d} \rho_i g_i = 0$, $\mathcal{P}$-q.s., contradicting the fact that $\rho \in K$. We deduce that K_0 is bounded. $\qquad \square$

It is now straightforward to show existence of an optimal e-variable.

Theorem 7.3. *Assume the constraint set is finite, $\Phi = \{g_1, \ldots, g_d\}$, and that*

$$\int_{\mathcal{X}} (|g_1| + \cdots + |g_d|) \, d\mathsf{Q} < \infty \text{ for all } \mathsf{Q} \in \mathcal{Q}.$$

Then there exists an e-variable $h^* = 1 + \sum_{i=1}^{d} \pi_i^* g_i$ with $\pi^* \in K_0$ that achieves the supremum $\sup_{h \in \mathcal{E}} \mathbb{U}(h)$.

Proof. We need to find $\pi^* \in K_0$ that maximizes u in (7.1). Since K_0 is compact by Lemma 7.2, it suffices to show that u is upper semicontinuous. Set $c = \sup\{\|\pi\| : \pi \in K_0\}$, which is finite since K_0 is compact, and note that $U(1 + \sum_{i=1}^{d} \pi_i g_i) \leq a + b + bc(|g_1| + \cdots |g_d|)$ for any $\pi \in K_0$. Since the upper bound is integrable under any fixed $\mathsf{Q} \in \mathcal{Q}$, it follows from Fatou's lemma and upper semicontinuity of U that the map $\pi \mapsto \int_{\mathcal{X}} U(1 + \sum_{i=1}^{d} \pi_i g_i) d\mathsf{Q} - \chi(\mathsf{Q})$ from K_0 to $[-\infty, \infty]$ is upper semicontinuous. Thus u is upper semicontinuous too, being the pointwise infimum of a family of upper semicontinuous functions. $\square$

One-sided sub-ψ hypotheses. We now consider the hypothesis $\mathcal{P}$ associated with a function ψ as in Section 4. We then get from Theorem 6.5(ii), or directly from Theorem 4.3(ii), that a complete class is given by all e-variables of the form

$$h(x) = \int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi(d\lambda), \quad x \in \operatorname{dom}(\psi^*),$$

with $\pi \in \mathcal{M}_1(\Lambda)$. The problem of maximizing $\mathbb{U}(h)$ over $\mathcal{E}$ thus reduces to maximizing $u(\pi)$ over $\mathcal{M}_1(\Lambda)$, where

$$u(\pi) = \inf_{\mathsf{Q} \in \mathcal{Q}} \left\{ \int_{\mathcal{X}} U\left(\int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi(d\lambda) \right) d\mathsf{Q} - \chi(\mathsf{Q}) \right\}. \tag{7.2}$$

Just as in the finitely generated case, this reduction immediately leads to a general existence result.

Theorem 7.4. *Assume $\mathcal{P}$ is the one-sided sub-ψ hypothesis associated with a function ψ as in Section 4, and that*

$$\int_{\mathcal{X}} e^{\psi^*} d\mathsf{Q} < \infty \text{ for all } \mathsf{Q} \in \mathcal{Q}.$$

Then there exists an e-variable $h^*(x) = \int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi^*(d\lambda)$ with $\pi^* \in \mathcal{M}_1(\Lambda)$ that achieves the supremum $\sup_{h \in \mathcal{E}} \mathbb{U}(h)$.

Proof. Since $\Lambda = [0, \lambda_{\max}]$ is compact, so is $\mathcal{M}_1(\Lambda)$ thanks to Prokhorov's theorem; see, e.g., Billingsley (1999, Theorem 5.1). Here $\mathcal{M}_1(\Lambda)$ is equipped with the topology of weak convergence. Moreover, if a sequence $(\pi_n)_{n \in \mathbb{N}}$ in $\mathcal{M}_1(\Lambda)$ converges weakly to a limit π, then the integrals $\int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi_n(d\lambda)$ converge to $\int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi(d\lambda)$ for every $x \in \operatorname{dom}(\psi^*)$. Since

$$U\left(\int_{\Lambda} e^{\lambda x - \psi(\lambda)} \pi(d\lambda) \right) \leq a + b e^{\psi^*(x)}$$

for all $\pi \in \mathcal{M}_1(\Lambda)$, and since the upper bound is integrable under any fixed $\mathsf{Q} \in \mathcal{Q}$, the same argument as in the proof of Theorem 7.3 now yields upper semicontinuity of u in (7.2). It follows that an optimal e-variable h^* of the stated form exists. $\square$

23

Uniqueness. Under suitable strict concavity conditions one expects that the e-variable attaining the supremum is unique. However, since the objective value $\mathbb{U}(h)$ is unchanged if h is replaced by a $\mathcal{Q}$-version, one can only expect uniqueness up to $\mathcal{Q}$-negligible sets. In what follows, uniqueness will always be understood in this sense. It is clear that if U is concave, then so is $\mathbb{U}$. We say that $\mathbb{U}$ is *strictly concave* if $\mathbb{U}(th + (1-t)h') > t\mathbb{U}(h) + (1-t)\mathbb{U}(h')$ holds for all $t \in (0,1)$ and all $h, h' \in \mathcal{E}$ that are not $\mathcal{Q}$-q.s. equal. Strict concavity of $\mathbb{U}$ implies uniqueness of the optimal e-variable. However, in the general composite case, strict concavity of $\mathbb{U}$ does not follow from strict concavity of U. We now present a simple sufficient condition for uniqueness in terms of U and $\mathcal{Q}$.

Theorem 7.5. *Assume that U is concave, $\sup_{h \in \mathcal{E}} \mathbb{U}(h) \in \mathbb{R}$, and the following condition holds:*

if $h_1, h_2 \in \mathcal{E}$ and $\displaystyle \inf_{\mathsf{Q} \in \mathcal{Q}} \int_{\mathcal{X}} \left(U\left(\frac{h_1 + h_2}{2} \right) - \frac{U(h_1)}{2} - \frac{U(h_2)}{2} \right) d\mathsf{Q} = 0,$ *then $h_1 = h_2$, $\mathcal{Q}$-q.s.*

Then the optimal e-variable is unique. The above condition holds, in particular, if U is strictly concave and $\mathcal{Q}$ is contained in the convex hull of finitely many mutually absolutely continuous probability measures.

Proof. Let $h_1, h_2 \in \mathcal{E}$ be any optimal e-variables, and set $\overline{h} = (h_1 + h_2)/2$. We then have

$$
\begin{aligned}
\mathbb{U}(\overline{h}) &= \inf_{\mathsf{Q} \in \mathcal{Q}} \left\{ \int_{\mathcal{X}} U(\overline{h}) d\mathsf{Q} - \chi(\mathsf{Q}) \right\} \\
&= \inf_{\mathsf{Q} \in \mathcal{Q}} \left\{ \frac{1}{2} \left(\int_{\mathcal{X}} U(h_1) d\mathsf{Q} - \chi(\mathsf{Q}) \right) + \frac{1}{2} \left(\int_{\mathcal{X}} U(h_2) d\mathsf{Q} - \chi(\mathsf{Q}) \right) \right. \\
&\qquad \left. + \int_{\mathcal{X}} \left(U(\overline{h}) - \frac{1}{2} U(h_1) - \frac{1}{2} U(h_2) \right) d\mathsf{Q} \right\} \\
&\geq \frac{1}{2} \mathbb{U}(h_1) + \frac{1}{2} \mathbb{U}(h_2) + \inf_{\mathsf{Q} \in \mathcal{Q}} \int_{\mathcal{X}} \left(U(\overline{h}) - \frac{1}{2} U(h_1) - \frac{1}{2} U(h_2) \right) d\mathsf{Q}.
\end{aligned}
$$

The first line is just the definition of $\mathbb{U}(\overline{h})$. To justify the second line, note that concavity of U yields $U(\overline{h}) - U(h_1)/2 - U(h_2)/2 \geq 0$, and that the negative parts $U(h_i)^-$, $i = 1, 2$, are integrable under every $\mathsf{Q} \in \mathcal{Q}$ since $\sup_{h \in \mathcal{E}} \mathbb{U}(h) > -\infty$. The last line follows from the definition of $\mathbb{U}(h_i)$, $i = 1, 2$. We deduce that $\overline{h}$ is optimal, and then, since the optimal value is finite, that

$$
\inf_{\mathsf{Q} \in \mathcal{Q}} \int_{\mathcal{X}} \left(U(\overline{h}) - \frac{1}{2} U(h_1) - \frac{1}{2} U(h_2) \right) d\mathsf{Q} = 0.
$$

The assumption of the theorem now yields $h_1 = h_2$, $\mathcal{Q}$-q.s., showing uniqueness. $\qquad\square$

8 Hypotheses expressed as unions

Some hypotheses of interest are naturally expressed as unions of simpler hypotheses. This happens, for instance, if one imposes an upper bound on the *conditional value-at-risk*; see Example 8.2 below. Hypotheses of this form are not generated by constraints in a natural way, but our results can still sometimes be used to describe their e-variables. This is a consequence of the following simple but very general result.

Lemma 8.1. *Consider an arbitrary collection of hypotheses $\mathcal{P}_\theta$, $\theta \in \Theta$, and let $\mathcal{E}_\theta$ be the set of e-variables for $\mathcal{P}_\theta$. Define the union $\mathcal{P} = \bigcup_{\theta \in \Theta} \mathcal{P}_\theta$, and let $\mathcal{E}$ be the associated set of e-variables. Then $\mathcal{E} = \bigcap_{\theta \in \Theta} \mathcal{E}_\theta$.*

Proof. A measurable function $h \colon \mathcal{X} \to [0, \infty]$ belongs to $\bigcap_{\theta \in \Theta} \mathcal{E}_\theta$ if and only if $\int_\mathcal{X} h \, d\mu \leq 1$ for all $\mu \in \mathcal{P}_\theta$ and all $\theta \in \Theta$. But this just says that $f \in \mathcal{E}$. $\qquad \square$

It is worth mentioning that there is no analog of Lemma 8.1 allowing to reconstruct the set $\mathcal{E}$ of e-variables for an *intersection* $\mathcal{P} = \bigcap_{\theta \in \Theta} \mathcal{P}_\theta$ from the individual sets $\mathcal{E}_\theta$. Such an analog would be of interest, because any hypothesis generated by a constraint set Φ can be expressed as the intersection $\bigcap_{f \in \Phi} \mathcal{P}_f$, where $\mathcal{P}_f$ is the hypothesis generated by the single function f. To see the obstruction, consider the constraint function $f(x) = x$. We know from Example 3.9 that $\mathcal{E}_f$ is trivial, and thus contains no information about the constraint function that generated it. It follows that there is no way to use $\mathcal{E}_f$ and, say, $\mathcal{E}_g$ where $g(x) = x^2 - 1$, to construct the set of e-variables for the hypothesis generated by f *and* g, whose e-variables are characterized in Example 3.6.

We now illustrate how Lemma 8.1 can be used by examining a hypothesis involving a bound on conditional value-at-risk (CVaR). Interestingly, this example shows that e-variables can be used to test CVaR without imposing any moment bounds as is done, for instance, by Agrawal et al. (2021b). Backtesting of CVaR using e-variables has been studied by Wang et al. (2024). The representation derived below extends the one in their Theorem 5 by removing the restriction to e-variables that are increasing in x imposed in that theorem.

Example 8.2. *Let $\mathcal{X} = \mathbb{R}$. For any $\alpha \in (0, 1)$, the* conditional value-at-risk *at level α of a distribution μ is given by*

$$\mathrm{CVaR}_\alpha(\mu) = \min_{\theta \in \mathbb{R}} \left\{ \theta + \frac{1}{1 - \alpha} \int_\mathcal{X} (x - \theta)^+ \mu(dx) \right\}. \tag{8.1}$$

This formula can be interpreted in terms of the value-at-risk *of μ at level α, which is the quantile $\mathrm{VaR}_\alpha(\mu) = \min\{x \in \mathbb{R} \colon \mu((-\infty, x]) \geq \alpha\}$. Assuming that μ does not have an atom at $\mathrm{VaR}_\alpha(\mu)$, one has $\mathrm{CVaR}_\alpha(\mu) = \mathsf{E}[X \mid X \geq \mathrm{VaR}_\alpha(\mu)]$ in probabilistic notation, where X is a random variable with distribution μ. A slightly more involved formula holds in the general case. We refer to Rockafellar and Uryasev (2002, Definition 3) for details.*

For a given level $\alpha \in (0, 1)$ and threshold $c \in \mathbb{R}$, we consider the hypothesis that the conditional value-at-risk at level α does not exceed c, i.e.,

$$\mathcal{P} = \{\mu \in \mathcal{M}_1 \colon \mathrm{CVaR}_\alpha(\mu) \leq c\}.$$

In view of (8.1), $\mathcal{P}$ is the union of the hypotheses $\mathcal{P}_\theta$, each generated by the single constraint function $\theta - c + (x - \theta)^+/(1 - \alpha)$. For $\theta > c$, this function is strictly positive, so $\mathcal{P}_\theta$ is empty and can be excluded from the union. We thus restrict to $\theta \leq c$. For $\theta = c$, $\mathcal{P}_\theta$ consists of all distributions on $(-\infty, c]$, and admits the single admissible e-variable $1 + \infty \mathbf{1}_{(c, \infty)}(x)$. For $\theta < c$, we get from Corollary 3.3 that every e-variable for $\mathcal{P}_\theta$ is pointwise dominated by $1 + \pi_\theta(\theta - c + (x - \theta)^+/(1 - \alpha))$ for some $\pi_\theta \in [0, (c - \theta)^{-1}]$. We then deduce from Lemma 8.1 that every e-variable for $\mathcal{P}$ is dominated by a function of the form

$$h(x) = 1 + \inf_{\theta < c} \pi_\theta \left(\theta - c + \frac{(x - \theta)^+}{1 - \alpha} \right) \wedge \left(\infty \mathbf{1}_{(c, \infty)}(x) \right),$$

for some $\pi_\theta \in [0, (c-\theta)^{-1}]$. Here the only role played by the term $\infty\mathbf{1}_{(c,\infty)}(x)$ is to ensure that $h(c) = 1$. Indeed, for $x < c$ the infimum is negative, and for $x > c$ it is, of course, finite. Thus the minimum with $\infty\mathbf{1}_{(c,\infty)}(x)$ has no effect in these cases. But it does have an effect for $x = c$, since $h(c) = 1 + \frac{\alpha}{1-\alpha} \inf_{\theta < c} \pi_\theta(c-\theta) \wedge 0 = 1$. It is natural to reparameterize in terms of $w_\theta = \pi_\theta(c-\theta)$, and we finally arrive at the form

$$h(x) = \begin{cases} 1 + \inf_{\theta < c} w_\theta \left(\frac{1}{1-\alpha} \left(\frac{x-\theta}{c-\theta} \right)^+ - 1 \right), & x \neq c, \\ 1, & x = c, \end{cases} \tag{8.2}$$

where $w_\theta \in [0,1]$. Since the pointwise infimum of a family of continuous functions is upper semicontinuous and hence measurable, any h of the form (8.2) is measurable. As a result, h is itself an e-variable for $\mathcal{P}$ and, therefore, the functions of this form constitute a complete class for $\mathcal{P}$ in the sense of Definition 6.1.

It is remarkable that CVaR, unlike the mean, can be tested without any a priori moment bounds. Indeed, taking $c = 0$ for simplicity, we obtain a nontrivial e-variable by setting $w_\theta = 1$ for all $\theta < 0$ in (8.2), namely

$$h(x) = \begin{cases} 0, & x < 0, \\ 1, & x = 0, \\ (1-\alpha)^{-1}, & x > 0. \end{cases}$$

Although α is required to lie in $(0,1)$ above, it is worth noting that with $\alpha = 0$, CVaR formally reduces to the mean. In this case any choice of $w_\theta \in [0,1]$ in (8.2) yields a trivial e-variable $h(x) \leq 1$, which is consistent with Example 3.9.

9 Hypotheses with relaxed integrability

We now elaborate on the integrability requirement in the definition of $\mathcal{P}$ in (2.1). To illustrate the issue, consider a single constraint function f_0 which is bounded above but unbounded below. We can then find $\mu \in \mathcal{M}_1$ such that $\int_{\mathcal{X}} f_0 d\mu$ is well-defined but equal to $-\infty$. *This measure does not qualify for membership in $\mathcal{P}$.* More generally, given a general constraint set Φ, it is natural to consider the larger, relaxed, hypothesis

$$\widetilde{\mathcal{P}} = \left\{ \mu \in \mathcal{M}_1 : \int_{\mathcal{X}} f^+ d\mu < \infty \text{ and } \int_{\mathcal{X}} f d\mu \in [-\infty, 0] \text{ for all } f \in \Phi \right\}. \tag{9.1}$$

What is the set $\widetilde{\mathcal{E}}$ of e-variables for $\widetilde{\mathcal{P}}$? It is certainly included in $\mathcal{E}$, but can the two be different? Relatedly, suppose we start with a constraint set Φ and include a single additional *negative* function f_0 to form $\Phi_0 = \Phi \cup \{f_0\}$. How, if at all, does this affect the hypotheses and their sets of e-variables? Intuitively one might expect that including a negative constraint function would not change the hypothesis. In the following discussion we indicate the constraint sets explicitly by writing $\mathcal{P}(\Phi)$, $\mathcal{P}(\Phi_0)$, $\widetilde{\mathcal{P}}(\Phi)$, and $\widetilde{\mathcal{P}}(\Phi_0)$ for the hypotheses, and $\mathcal{E}(\Phi)$, etc., for the corresponding sets of e-variables.

It is clear that the relaxed hypotheses $\widetilde{\mathcal{P}}(\Phi)$ and $\widetilde{\mathcal{P}}(\Phi_0)$ are equal, and so are their sets of e-variables. However, as the following example shows, $\mathcal{P}(\Phi)$ can differ from $\mathcal{P}(\Phi_0)$, and their sets of e-variables can also be different.

Example 9.1. *Let $\mathcal{X} = \mathbb{N}$. Consider the constraint set Φ consisting of the functions f_n, $n \in \mathbb{N}$, defined by $f_n(x) = 1$ for $x \neq n$ and $f_n(n) = 1 - 2^n$. Let $\Phi_0 = \Phi \cup \{f_0\}$, where the negative function f_0 is given by $f_0(x) = -2^x$. For any $\mu = \sum_{x \in \mathbb{N}} p_x \delta_x$ in the hypothesis $\mathcal{P}(\Phi)$, the condition $\int_{\mathbb{N}} f_n d\mu \leq 0$ translates to the inequality $1 - p_n + p_n(1 - 2^n) \leq 0$, or $p_n \geq 2^{-n}$. It follows that $\mathcal{P}(\Phi)$ consists of the single measure $\mu = \sum_{x \in \mathbb{N}} 2^{-x} \delta_x$, and that $\mathcal{E}(\Phi)$ consists of all nonnegative functions h such that $\sum_{x \in \mathbb{N}} h(x) 2^{-x} \leq 1$. However, $\int_{\mathbb{N}} f_0 d\mu = -\infty$, so $\mu \notin \mathcal{P}(\Phi_0)$. Thus $\mathcal{P}(\Phi_0) = \emptyset$, and $\mathcal{E}(\Phi_0)$ consists of all nonnegative functions on $\mathbb{N}$. In contrast, the relaxed hypotheses $\widetilde{\mathcal{P}}(\Phi)$ and $\widetilde{\mathcal{P}}(\Phi_0)$ both coincide with $\mathcal{P}(\Phi)$, and their (common) set of e-variables coincides with $\mathcal{E}(\Phi)$.*

In this example the constraint sets are infinite. We now give a positive result showing that with a finite constraint set, the above issue cannot occur. It is an interesting open problem to characterize $\widetilde{\mathcal{E}}(\Phi)$ for a general infinite constraint set Φ.

Theorem 9.2. *Consider a finite nonempty constraint set $\Phi = \{g_1, \ldots, g_d\}$. Let $\mathcal{P}$ be the hypothesis it generates, and let $\widetilde{\mathcal{P}}$ be the relaxed hypothesis defined in (9.1). Then the set $\widetilde{\mathcal{E}}$ of e-variables for $\widetilde{\mathcal{P}}$ coincides with the set $\mathcal{E}$ of e-variables for $\mathcal{P}$.*

The proof relies on the following lemma.

Lemma 9.3. *Let $\mu \in \mathcal{M}_1$ and let $f_1, \ldots, f_m$ be measurable and μ-integrable. Then there is a finitely supported probability measure ν such that $\int_{\mathcal{X}} f_i d\nu = \int_{\mathcal{X}} f_i d\mu$ for $i = 1, \ldots, m$.*

Proof. Define $f \colon \mathcal{X} \to \mathbb{R}^m$ by $f(x) = (f_1(x), \ldots, f_m(x))$ and set $z_0 = \int_{\mathcal{X}} f d\mu \in \mathbb{R}^m$. To prove the lemma it is enough to show that $z_0 \in \mathrm{conv}(\mathrm{range}(f))$, the convex hull of the range of f. To this end, consider the pushforward $\gamma = f_* \mu$ on $\mathbb{R}^m$ and define

$$C = \mathrm{conv}(\mathrm{supp}(\gamma) \cap \mathrm{range}(f)).$$

Suppose for contradiction that $z_0 \notin \mathrm{ri}(C)$, the relative interior of C. Then there is some nonzero $u \in \mathbb{R}^m$ such that

$$u \cdot (z - z_0) > 0 \text{ for all } z \in \mathrm{ri}(C). \tag{9.2}$$

This implies that $u \cdot (z - z_0) \geq 0$ for all $z \in C$, hence γ-a.e. Since also

$$\int_{\mathbb{R}^m} u \cdot (z - z_0) \gamma(dz) = u \cdot \left(\int_{\mathbb{R}^m} z \gamma(dz) - z_0 \right) = 0,$$

we deduce that $u \cdot (z - z_0) = 0$, γ-a.e. This shows that $\mathrm{supp}(\gamma)$ is contained in the set $\{z \in \mathbb{R}^m \colon u \cdot (z - z_0) = 0\}$, which is disjoint from $\mathrm{ri}(C)$ in view of (9.2). We have established that $C \cap \mathrm{ri}(C) = \emptyset$, which contradicts the fact that every nonempty convex set in $\mathbb{R}^m$ has nonempty relative interior. Thus $z_0 \in \mathrm{ri}(C) \subset \mathrm{conv}(\mathrm{range}(f))$. $\square$

Proof of Theorem 9.2. Since $\mathcal{P} \subset \widetilde{\mathcal{P}}$, we always have $\widetilde{\mathcal{E}} \subset \mathcal{E}$. We therefore only have to prove the opposite inclusion, so we fix any $h \in \mathcal{E}$. Assume first that h is bounded. Consider any $\mu \in \widetilde{\mathcal{P}}$ and let c be a sufficiently large negative constant to ensure that $\int_{\mathcal{X}} (g_i \vee c) d\mu \leq 0$ for

those $i \in \{1, \ldots, d\}$ such that $\int_{\mathcal{X}} g_i^- d\mu = \infty$. Using Lemma 9.3 we obtain a finitely supported probability measure ν such that

$$\int_{\mathcal{X}} h d\nu = \int_{\mathcal{X}} h d\mu, \tag{9.3}$$

$$\int_{\mathcal{X}} g_i d\nu = \int_{\mathcal{X}} g_i d\mu, \qquad\qquad i \in \{1, \ldots, d\} \text{ with } \int_{\mathcal{X}} g_i^- d\mu < \infty, \tag{9.4}$$

$$\int_{\mathcal{X}} (g_i \vee c) d\nu = \int_{\mathcal{X}} (g_i \vee c) d\mu, \qquad i \in \{1, \ldots, d\} \text{ with } \int_{\mathcal{X}} g_i^- d\mu = \infty. \tag{9.5}$$

Since ν is finitely supported, all g_i are ν-integrable. We thus have from (9.4) and (9.5) that $\nu \in \mathcal{P}$. Then (9.3) and the fact that $h \in \mathcal{E}$ yield $\int_{\mathcal{X}} h d\mu \leq 1$. Since $\mu \in \widetilde{\mathcal{P}}$ was arbitrary, this shows that $h \in \widetilde{\mathcal{E}}$. If $h \in \mathcal{E}$ is unbounded, then by what we just proved, $h \wedge n \in \widetilde{\mathcal{E}}$ for each $n \in \mathbb{N}$, and thus $h \in \widetilde{\mathcal{E}}$ by monotone convergence. This completes the proof. $\qquad\square$

10 Summary and open problems

In various applications, statistical hypotheses are naturally described through constraints on the data-generating distribution. In this paper, we use the duality theory for locally convex topological vector spaces, in particular the bipolar theorem, to give an abstract representation of all e-variables for hypotheses of this kind in a general setting. We then show how the abstract representation instantiates in several concrete cases. As such, our results facilitate the use of e-variables for testing hypotheses generated by constraints.

There are, nonetheless, many important and challenging open problems. One of them is to describe the set of e-variables for the i.i.d. n-sample hypothesis $\mathcal{P}^{\otimes n}$ consisting of all product distributions $\mu^{\otimes n}$ on $\mathcal{X}^n$ with $\mu \in \mathcal{P}$, for some hypothesis $\mathcal{P}$ of distributions on $\mathcal{X}$. Any function of the form $h(x_1, \ldots, x_n) = \prod_{i=1}^n h_i(x_i)$ with $h_i \in \mathcal{E}$ is an e-variable for $\mathcal{P}^{\otimes n}$, as are all convex combinations of such functions, and their limits. However, not all e-variables arise in this way in general. For example, suppose $\mathcal{P}$ is the set of *all* distributions on $\mathbb{R}$, so that $\mathcal{P}^n$ is the set of *all* laws of real i.i.d. n-samples. Then any $h_i \in \mathcal{E}$ satisfies $h_i \leq 1$, and thus so do e-variables of the product form above, their convex combinations, etc. But, taking $n = 2$ for illustration, the function $h(x_1, x_2) = 2\mathbf{1}_{\{x_1 < x_2\}} + \mathbf{1}_{\{x_1 = x_2\}}$ is an example of a nontrivial e-variable for $\mathcal{P}^{\otimes n}$. Can all nontrivial e-variables be characterized?

Another interesting future direction is to extend our results to the sequential setting, where testing and inference based on e-variables have been proven to be particularly effective. Let us mention two specific problems whose solutions would be of interest. We formulate these in a probabilistic language, rather than the analytic language we otherwise use in this paper.

Suppose we observe data X_t, $t \in \mathbb{N}$, taking values in a measurable space $\mathcal{X}$. We model $(X_t)_{t \in \mathbb{N}}$ as the canonical process on $\mathcal{X}^{\mathbb{N}}$, and let $(\mathcal{F}_t)_{t \in \mathbb{N}}$ with $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$ be the filtration generated by the data. We let $\mathcal{F}_0$ be the trivial σ-algebra. Consider now a finite constraint set $\Phi = \{g_1, \ldots, g_d\}$ of functions on $\mathcal{X}$, which generates a family $\mathcal{P}$ of distributions on $\mathcal{X}$ as in Definition 2.1. We may then examine those distributions P on $\mathcal{X}^{\mathbb{N}}$ under which the conditional distributions of the data belong to $\mathcal{P}$. That is,

$$\mathsf{P}(X_t \in \cdot \mid \mathcal{F}_{t-1}) \text{ belongs to } \mathcal{P} \text{ for all } t \in \mathbb{N}, \ \mathsf{P}\text{-a.s.,}$$

or, equivalently,

$$\mathsf{E}_\mathsf{P}[g_i(X_t) \mid \mathcal{F}_{t-1}] \leq 0, \ \mathsf{P}\text{-a.s., for all } i = 1, \ldots, d \text{ and all } t \in \mathbb{N}.$$

Let $\overline{\mathcal{P}}$ denote the collection of all such P. Note that $\overline{\mathcal{P}}$ is a family of distributions over the entire data sequence, whereas $\mathcal{P}$ is a family of distributions over a single observation.

An *e-process* for $\overline{\mathcal{P}}$ is a nonnegative adapted process $(Y_t)_{t \in \mathbb{N}_0}$ (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) such that $\mathsf{E}_\mathsf{P}[Y_\tau] \leq 1$ for all $\mathsf{P} \in \overline{\mathcal{P}}$ and all finite stopping times τ for the given filtration. E-processes are the natural analog of e-variables in the sequential setting, and it is of interest to describe their structure. One can easily construct a large class of such e-processes by setting $Y_0 = 1$ and

$$Y_t = \prod_{s=1}^{t} \left(1 + \sum_{i=1}^{d} \pi_{i,s} g_i(X_s) \right), \quad t \in \mathbb{N}, \tag{10.1}$$

where $\pi_t = (\pi_{1,t}, \ldots, \pi_{d,t})$, $t \in \mathbb{N}$, is a predictable process taking values in the set Π^Φ defined in Corollary 3.3. The predictability property means that π_t is a measurable function of $X_1, \ldots, X_{t-1}$. Any process $(Y_t)_{t \in \mathbb{N}_0}$ as in (10.1) above is actually a $\overline{\mathcal{P}}$-supermartingale, i.e., a supermartingale under every $\mathsf{P} \in \overline{\mathcal{P}}$. The e-process property then follows from the stopping theorem. It is now natural to ask whether *every* admissible e-process for $\overline{\mathcal{P}}$ is of the form (10.1), at least assuming the constraint qualification (3.3). Settling this question would be of great interest.

The second open problem concerns the smaller hypothesis $\mathcal{P}^\infty \subset \overline{\mathcal{P}}$ consisting of all i.i.d. distributions $\mathsf{P} = \mu^{\otimes \infty}$ on $\mathcal{X}^\mathbb{N}$ with $\mu \in \mathcal{P}$. This hypothesis is significantly smaller than $\overline{\mathcal{P}}$, so its family of admissible e-processes is potentially larger. Can it be described explicitly? How, if at all, is it related to the family of e-processes for $\overline{\mathcal{P}}$?

Lastly, one can of course formulate versions of these questions for infinite constraint sets Φ as well, and this also leads to interesting open problems.

A   Some results from topology and functional analysis

A.1   Closedness, compactness, and continuity using nets

Let X be a topological space. In many cases, for example if X is a metric space, properties such as closedness, compactness, and continuity can be characterized using sequences. For example, a set $C \subset X$ is closed if and only if C contains the limit of every convergent sequence $(x_n)_{n \in \mathbb{N}} \subset C$. In general topological spaces, this characterization may fail. However, it can be restored by replacing sequences with the more general concept of *nets*, also known as *Moore–Smith sequences*. We review the basic definitions and facts below, and refer the reader to Kelley (1975, Chapter 2), Willard (1970, Chapter 4), and Aliprantis and Border (2006, p. 32) for more details.

Instead of only using the natural numbers $\mathbb{N}$ as index set, a net can be indexed by a general *directed set* A. This is a nonempty set with a binary relation $\geq$ that is symmetric ($\alpha \geq \alpha$ for all $\alpha \in A$), transitive ($\alpha \geq \beta$ and $\beta \geq \gamma$ implies $\alpha \geq \gamma$ for all $\alpha, \beta, \gamma \in A$), and enjoys the directedness property that for any $\alpha, \beta \in A$ there is $\gamma \in A$ with $\gamma \geq \alpha$ and $\gamma \geq \beta$. The natural numbers with the standard ordering is an example of a directed set; another is the family of all neighborhoods U of a given point $x \in X$, with $U \geq V$ if $U \subset V$. (Recall that

a *neighborhood* of a point $x \in X$ is any set that contains an open set containing x.) A *net* is a map from some directed set A to X, denoted by $(x_\alpha)_{\alpha \in A}$ in analogy with the notation for sequences. For brevity we often write (x_α) or even just x_α for the net $(x_\alpha)_{\alpha \in A}$. The net *converges* to a point $x \in X$ if it is eventually in any neighborhood of x; that is, if for any neighborhood U of x, there is some $\alpha \in A$ such that $x_\beta \in U$ for all $\beta \in A$ with $\beta \geq \alpha$. We express this by saying that "x_α converges to x", or just "$x_\alpha \to x$". Lastly, a *subnet* of $(x_\alpha)_{\alpha \in A}$ is a net of the form $(x_{\varphi(\beta)})_{\beta \in B}$, where $\varphi \colon B \to A$ is increasing ($\gamma \geq \beta$ implies $\varphi(\gamma) \geq \varphi(\beta)$) and cofinal (for every $\alpha \in A$ there is $\beta \in B$ such that $\varphi(\beta) \geq \alpha$).

Theorem A.1.   (i) *A set $C \subset X$ is closed if and only if it contains all limits of nets in C.*

(ii) *A set $C \subset X$ is compact if and only if every net in C has a subnet with a limit in C.*

(iii) *A function f from X to a topological space Y is continuous if and only if $x_\alpha \to x$ implies $f(x_\alpha) \to f(x)$. To be precise, the latter property means that for every $x \in X$ and every net $(x_\alpha)_{\alpha \in A}$ in X that converges to x, the net $(f(x_\alpha))_{\alpha \in A}$ in Y converges to $f(x)$.*

Proof. Parts (i) and (iii) are Theorem 11.7 and Theorem 11.8 of Willard (1970). Part (ii) follows from Theorem 11.5 and Theorem 17.4 of Willard (1970). □

We now specialize some of the above to Euclidean space $\mathbb{R}^d$. The Heine–Borel theorem states that any closed and bounded subset of $\mathbb{R}^d$ is compact. Therefore, Theorem A.1(ii) implies that any bounded net in $\mathbb{R}^d$ has a convergent subnet. We use this fact in the proof of Theorem 3.1.

For a net $(x_\alpha)_{\alpha \in A}$ in $\mathbb{R}$ one can define the limsup and liminf exactly as for sequences,

$$\liminf_\alpha x_\alpha = \lim_\alpha \inf_{\beta \geq \alpha} x_\beta \quad \text{and} \quad \limsup_\alpha x_\alpha = \lim_\alpha \sup_{\beta \geq \alpha} x_\beta.$$

That is, $\liminf_\alpha x_\alpha$ is the limit of the net $y_\alpha = \inf_{\beta \geq \alpha} x_\beta$ in the extended reals $[-\infty, \infty]$. This net is increasing in the sense that $\gamma \geq \alpha$ implies $y_\gamma \geq y_\alpha$, and this ensures that the limit exists. We say that x_α converges to infinity if $\liminf_\alpha x_\alpha = \infty$, meaning that x_α is eventually larger than any real number. The case of limsup is analogous. If a net in $\mathbb{R}_+$ does not converge to infinity, meaning that its liminf is finite, it has a bounded subnet. Hence, by Heine–Borel and Theorem A.1(ii) as above, it has a further subnet that converges to a limit in $\mathbb{R}_+$. This is again something we make use of in the proof of Theorem 3.1.

A.2   Dual pairs and the bipolar theorem

Here we review some concepts and facts from the classical duality theory of locally convex spaces. All the required material can be found in Schaefer and Wolff (1999), see in particular Chapter IV.

Two real vector space F and G form a *dual pair* (or *dual system*) if there is a bilinear form $\langle \cdot, \cdot \rangle$ on $F \times G$ that separates points in the following sense: if $x \in F$ and $\langle x, y \rangle = 0$ for all $y \in G$, then $x = 0$; and if $y \in F$ and $\langle x, y \rangle = 0$ for all $x \in F$, then $y = 0$. One also says that $\langle \cdot, \cdot \rangle$ places F and G in *(separated) duality*, and writes $\langle F, G \rangle$ as shorthand for the tuple $(F, G, \langle \cdot, \cdot \rangle)$.

Given a dual pair $\langle F, G \rangle$, one defines the *weak topology* $\sigma(F, G)$ as the initial topology generated by the maps $x \mapsto \langle x, y \rangle$, $y \in G$. That is, $\sigma(F, G)$ is the weakest topology on F such

that the map $x \mapsto \langle x, y \rangle$ from F to $\mathbb{R}$ is continuous for every $y \in G$. With this topology F is a locally convex space; see Schaefer and Wolff (1999), Chapter II, Section 5.

For any subset $C \subset F$, the *polar* of C (sometimes called the *one-sided polar*) is the set

$$C^\circ = \{y \in G \colon \langle x, y \rangle \leq 1 \text{ for all } x \in C\}.$$

The *bipolar* of C is the polar of the polar,

$$C^{\circ\circ} = \{x \in F \colon \langle x, y \rangle \leq 1 \text{ for all } y \in C^\circ\}.$$

The polar and bipolar are always convex. If C is a cone, meaning that $\lambda C \subset C$ for every $\lambda \in [0, \infty)$, then the polar and bipolar are also cones and can be written

$$C^\circ = \{y \in G \colon \langle x, y \rangle \leq 0 \text{ for all } x \in C\},$$
$$C^{\circ\circ} = \{x \in F \colon \langle x, y \rangle \leq 0 \text{ for all } y \in C^\circ\}.$$

We make extensive use of the following fundamental result, which we state here for convex cones. This follows from Theorem 1.5 of Section IV in Schaefer and Wolff (1999) along with the fact that the convex hull of $\{0\} \cup C$ is just C itself when C is a convex cone.

Theorem A.2 (Bipolar theorem). *If $C \subset F$ is a convex cone, then the bipolar $C^{\circ\circ}$ is the $\sigma(F, G)$-closure of C.*

A.3 The Krein–Šmulian theorem

Let E be a Banach space with norm $\|\cdot\|$. Its dual space E' consists of all bounded linear functionals on E, and is equipped with the dual norm given by $\|\varphi\|' = \sup\{\varphi(x) \colon x \in E, \|x\| \leq 1\}$. The weak* topology on E' is the initial topology generated by the maps $\varphi \mapsto \varphi(x)$ from E' to $\mathbb{R}$, where $x \in E$. This is also the topology $\sigma(E', E)$ coming from the dual pair $\langle E', E \rangle$ with bilinear form $\langle \varphi, x \rangle = \varphi(x)$; see Section A.2. The following result plays a crucial role in Section 4. For a proof, see Theorem 12.1 in Conway (1990).

Theorem A.3 (Krein–Šmulian). *Let $(E, \|\cdot\|)$ be a Banach space with dual space $(E', \|\cdot\|')$. A convex subset $C \subset E'$ is weak* closed if and only its intersection with every dual ball is weak* closed, that is, $C \cap \{\varphi \in E' \colon \|\varphi\|' \leq r\}$ is weak* closed for all $r \in (0, \infty)$.*

References

Shubhada Agrawal, Sandeep Juneja, and Peter Glynn. Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory*, pages 61–110. PMLR, 2020. 8

Shubhada Agrawal, Sandeep K Juneja, and Wouter M Koolen. Regret minimization in heavy-tailed bandits. In *Conference on Learning Theory*, pages 26–62. PMLR, 2021a. 8

Shubhada Agrawal, Wouter M Koolen, and Sandeep Juneja. Optimal best-arm identification methods for tail-risk measures. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 25578–25590. Curran Associates, Inc., 2021b. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d69c7ebb6a253532b266151eac6591af-Paper.pdf. 25

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis*. Springer, Berlin, third edition, 2006. ISBN 978-3-540-32696-0; 3-540-32696-0. A Hitchhiker's Guide. 29

Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. URL https://doi.org/10.1002/9780470316962. A Wiley-Interscience Publication. 23

Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952. 10

Eugenio Clerico. Optimal e-value testing for properly constrained hypotheses, 2024a. URL https://arxiv.org/abs/2412.21125. 2, 7, 8, 18, 19

Eugenio Clerico. On the optimality of coin-betting for mean estimation. *arXiv preprint arXiv:2412.02640*, 2024b. 9

John B. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990. ISBN 0-387-97245-5. 31

Harald Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. In *Collected Works II*, pages 895–913. Springer, 1994. 10

Yixuan Fan, Zhanyi Jiao, and Ruodu Wang. Testing the mean and variance by e-processes. *Biometrika*, 112(1):asae049, 2025. 8

Peter Grünwald, Rianne de Heide, and Wouter M Koolen. Safe testing. *Journal of the Royal Statistical Society, Series B (Methodology), with discussion*, 2024. 2, 21

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 2020. 10

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 2021. 10

John L. Kelley. *General Topology*, volume No. 27 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.]. 29

Nick W Koning. Post-hoc and anytime valid permutation and group invariance testing. *arXiv preprint arXiv:2310.01153*, 2023. 16

Nick W. Koning. Continuous testing: Unifying tests and e-values, 2025. URL https://arxiv.org/abs/2409.05654. 21

Steven G. Krantz and Harold R. Parks. *A primer of real analytic functions*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston, Inc., Boston, MA, second edition, 2002. ISBN 0-8176-4264-1. doi: 10.1007/978-0-8176-8134-0. URL https://doi.org/10.1007/978-0-8176-8134-0. 20

Martin Larsson, Aaditya Ramdas, and Johannes Ruf. The numeraire e-variable and reverse information projection. *The Annals of Statistics*, 2025. 2, 9, 21

Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006. ISSN 0012-9682,1468-0262. doi: 10.1111/j.1468-0262.2006.00716.x. URL https://doi.org/10.1111/j.1468-0262.2006.00716.x. 3, 22

Francesco Orabona and Kwang-Sung Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*, 70(1): 436–455, 2023. 9

Teodora Pandeva, Patrick Forré, Aaditya Ramdas, and Shubhanshu Shekhar. Deep anytime-valid hypothesis testing. In *International Conference on Artificial Intelligence and Statistics*, pages 622–630. PMLR, 2024. 16

Muriel Felipe Pérez-Ortiz, Tyron Lardy, Rianne de Heide, and Peter D Grünwald. E-statistics, group invariance and anytime-valid testing. *The Annals of Statistics*, 52(4):1410–1432, 2024. 16

Aleksandr Podkopaev and Aaditya Ramdas. Sequential predictive two-sample and independence testing. *Advances in Neural Information Processing Systems*, 36:53275–53307, 2023. 16

Aleksandr Podkopaev, Patrick Blöbaum, Shiva Kasiviswanathan, and Aaditya Ramdas. Sequential kernelized independence testing. In *International Conference on Machine Learning*, pages 27957–27993. PMLR, 2023. 16

Aaditya Ramdas and Ruodu Wang. Hypothesis testing with e-values. *arXiv preprint arXiv:2410.23614*, 2024. 21

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales, 2022a. URL https://arxiv.org/abs/2009.03167. 19

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter M Koolen. Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 141:83–109, 2022b. 16

R.Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002. ISSN 0378-4266. URL https://www.sciencedirect.com/science/article/pii/S0378426602002716. 25

Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/24029299. 2

Aytijhya Saha and Aaditya Ramdas. Testing exchangeability by pairwise betting. In *International Conference on Artificial Intelligence and Statistics*, pages 4915–4923. PMLR, 2024. 16

H. H. Schaefer and M. P. Wolff. *Topological Vector Spaces*, volume 3 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1999. ISBN 0-387-98726-6. doi: 10.1007/978-1-4612-1468-7. URL https://doi.org/10.1007/978-1-4612-1468-7. 30, 31

Glenn Shafer. Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society Series A: Statistics in Society, with discussion*, 184 (2):407–431, 2021. 2

Shubhanshu Shekhar and Aaditya Ramdas. Nonparametric two-sample testing by betting. *IEEE Transactions on Information Theory*, 70(2):1178–1203, 2023. 16

Vladimir Vovk. Testing randomness online. *Statistical Science*, 36(4):595–611, 2021. 16

Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736–1754, 2021. 2

Hongjian Wang and Aaditya Ramdas. Catoni-style confidence sequences for heavy-tailed mean estimation. *Stochastic Processes and Their Applications*, 163:168–202, 2023. 8

Qiuqi Wang, Ruodu Wang, and Johanna Ziegel. E-backtesting, 2024. URL https://arxiv.org/abs/2209.00991. 25

Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890, 2020. 2

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1): 1–27, 2024. 9

Stephen Willard. *General Topology*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1970. 29, 30