The Annals of Statistics
2024, Vol. 52, No. 4, 1410–1432
https://doi.org/10.1214/24-AOS2394

E-STATISTICS, GROUP INVARIANCE AND ANYTIME-VALID TESTING

BY MURIEL FELIPE PÉREZ-ORTIZ[1,a], TYRON LARDY[2,b], RIANNE DE HEIDE[3,c] AND PETER D. GRÜNWALD[4,d]

[1]*Statistics, Probability and Operations Research Cluster, Eindhoven University of Technology,* [a]*m.f.perez.ortiz@tue.nl*
[2]*Mathematical Institute, Leiden University,* [b]*t.d.lardy@math.leidenuniv.nl*
[3]*Department of Mathematics, Vrije Universiteit,* [c]*r.de.heide@vu.nl*
[4]*Centrum Wiskunde & Informatica,* [d]*pdg@cwi.nl*

We study worst-case-growth-rate-optimal (GROW) e-statistics for hypothesis testing between two group models. It is known that under a mild condition on the action of the underlying group G on the data, there exists a maximally invariant statistic. We show that among all e-statistics, invariant or not, the likelihood ratio of the maximally invariant statistic is GROW, both in the absolute and in the relative sense, and that an anytime-valid test can be based on it. The GROW e-statistic is equal to a Bayes factor with a right Haar prior on G. Our treatment avoids nonuniqueness issues that sometimes arise for such priors in Bayesian contexts. A crucial assumption on the group G is its amenability, a well-known group-theoretical condition, which holds, for instance, in scale-location families. Our results also apply to finite-dimensional linear regression.

1. Introduction. We develop e-statistics and anytime-valid methods (Ramdas et al. (2023)) for composite hypothesis testing problems where both null and alternative models remain unchanged under a group of transformations. Assume that the parameter of interest is a function $\delta = \delta(\theta)$ that is invariant under these transformations. Here, $\theta \in \Theta$ is the parameter of a probabilistic model $\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$ on an observation space $\mathcal{X}$. In the simplest case that we address, we are interested in testing whether the invariant parameter δ takes one of two values, that is,

$$\mathcal{H}_0 : \delta(\theta) = \delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \delta(\theta) = \delta_1. \tag{1}$$

A prototypical example is the one-sample t-test where $\mathcal{P} = \{N(\mu, \sigma) : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}$ and the parameter of interest is the effect size $\delta(\mu, \sigma) = \mu/\sigma$, an invariant function of the model parameters under changes of scale. Other examples include tests about the correlation coefficient, which is invariant under affine transformations, and the variance of the principal components, an invariant under rotations around the origin (for more examples, see Berger, Pericchi and Varshavsky (1998)). Data can be reduced by only considering its invariant component. Roughly speaking, by replacing the data $X^n = (X_1, \dots, X_n)$ with an invariant statistic $M_n = m_n(X^n)$, one discards all information that is not relevant to the parameter δ (see the formal definitions in Section 2). For example, for the one-sample t-test, we can set M_n equal to the t-statistic $M_{\mathcal{S},n} \propto \hat{\mu}_n/\hat{\sigma}_n$ but also to $M_n = (X_1/|X_1|, \dots, X_n/|X_1|)$. Both are invariant functions under rescaling of all data points by the same factor that retain, as we will see, as much information as possible about the data.

By reducing the data through an invariant function, an invariant test can be obtained. Through the lens of the invariance-reduced data M_n, the composite hypotheses about θ simplify and (1) becomes simple-vs.-simple in terms of δ. Indeed, because M_n is an invariant

Received October 2023; revised April 2024.
MSC2020 subject classifications. Primary 62L10; secondary 62F03, 62B99.
Key words and phrases. E-values, anytime-valid methods, group invariance, Bayes factors, right Haar priors.

function, its density depends only on δ. Let us denote p^{M_n} and q^{M_n} the densities of M_n under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Both fixed-sample-size and sequential tests can be based on assessing the value of the likelihood ratio

$$T^{M_n} := \frac{q^{M_n}(m_n(X^n))}{p^{M_n}(m_n(X^n))}. \tag{2}$$

However, it is not clear a priori whether this reduction affects the optimality of the resulting tests. In other words, does the family of invariant tests, that is, tests that can be written as a function of (2), contain the best ones?

For fixed-sample size tests, with power as a criterion, the answer is positive: a celebrated theorem of Hunt and Stein shows that, when looking for a test that has max-min power, no loss is incurred by looking only among group-invariant tests (Lehmann and Romano (2005), Section 8.5). In classic sequential testing, the principle of invariance has been used (Cox (1952), Hall, Wijsman and Ghosh (1965)), but no optimality results are known. In this article, we address this question and provide an analogue of the Hunt–Stein theorem within the setting of *anytime-valid* tests. We replace power by GROW (see again below), the natural optimality criterion in this context, and we show that, under some regularity conditions, T^{M_n} is the *optimal e-statistic* for testing (1).

The *e*-statistic (also known as *e*-variable or *e*-value) is a central concept within the theory of *anytime-valid* testing (Grünwald, de Heide and Koolen (2024), Ramdas et al. (2020), Shafer (2021), Vovk and Wang (2021)), interest in which has recently exploded—Ramdas et al. (2023) provide a comprehensive overview. The main objective that is achieved by testing with *e*-statistics is finite-sample type-I error control in two common situations: when experiments are optionally stopped—sampling is stopped at a data-dependent sample size—, and when aggregating the evidence of interdependent experiments. In the latter case, called optional continuation (Grünwald, de Heide and Koolen (2024), GHK from now on), the decision to start a new experiment may depend in unknowable ways on the outcome of previous experiments (Vovk and Wang (2021)). We will use the qualifier anytime-valid as an umbrella term that covers both optional stopping and continuation, and study invariance reductions for anytime-valid tests; we stress that, as elaborated in Section S3 of the Supplementary Material (Pérez-Ortiz et al. (2024)), anytime-valid testing, while taking place in a sequential setting, is different from classical, Wald-style sequential testing, in which power *is* meaningful. While *e*-statistics have also found applications beyond anytime-validity, for example, in multiple testing (Ren and Barber (2024), Wang and Ramdas (2022)) and when not just the stopping time but also the relevant loss function or significance level may depend in unknowable ways on the data itself (decision-theoretic robustness, Grünwald (2023)), our results focus on optimality in the anytime-valid context. In this context, power is not a meaningful measure of optimality (see Section 2.4). A natural replacement of power is the *GROW* criterion, which stands for *growth rate optimal in the worst case*. Informally, among all *e*-statistics, those that are GROW accumulate evidence against the null as fast as possible (in terms of sample size). Some other authors refer to GROW as "maximal *e*-power" (Zhang, Ramdas and Wang (2023)) or as "optimizing the Kelly criterion" (Ramdas et al. (2023)). Sometimes, it is beneficial to consider instead the growth rate relative to an oracle that knows the distribution of the data, not in absolute terms. *e*-statistics that are optimal in this relative sense are called relatively GROW. Especially this relative criterion (or closely related variations of it) has often been used to design *e*-statistics; recent examples include the work of Henzi et al. (2023) and Waudby-Smith and Ramdas (2024)—see Ramdas et al. (2023) for a more comprehensive list.

Under regularity conditions, the GROW *e*-statistic can be found by minimizing the Kullback–Leibler (KL) divergence between the convex hull of the null and alternative models

(GHK). Indeed, the likelihood ratio of the distributions that achieve this minimum KL is a GROW e-statistic, and the GROW e-statistic is then essentially unique in the sense that any two GROW e-statistics agree almost surely under all distributions in $\mathcal{H}_0$ and $\mathcal{H}_1$. As such, e-statistics can be seen as composite generalizations of likelihood ratios. In particular, any likelihood ratio of a statistic that has the same distribution under all elements of the null and another single distribution under the alternative is an e-statistic (GHK). As a consequence, for any invariant function of the data M_n, the likelihood ratio statistic T^{M_n} from (2) is an e-statistic for the testing problem (1). As our main contribution, we show that, under regularity conditions, if M_n is a maximally invariant statistic of the data or of a sufficient statistic for θ, then the KL divergence between q^{M_n} and p^{M_n} equals the minimum KL divergence between the convex hulls of the null and alternative models. By the result of GHK mentioned above that links KL minimization to GROW e-statistics, T^{M_n} is GROW. A maximally invariant statistic, informally, loses as little information as possible about the data while being invariant. For example, with $V_n = (X_1/|X_1|, \dots, X_n/|X_1|)$, setting $M_n := V_n$ as in the beginning of the introduction for the t-test gives a maximal invariant, while using $M_n' := V_{n-1}$ gives an invariant that is not maximal. Furthermore, the t-statistic is not maximally invariant for the raw data, but it is a maximally invariant function of $(\hat{\mu}_n, \hat{\sigma}_n)$ which is a sufficient statistic. As we will see, the likelihood ratio statistic $T^{M_{\mathcal{S},n}}$, where $M_{\mathcal{S},n}$ is the t-statistic and T^{M_n} with $M_n = V_n$ coincide (see Section S1 of the Supplementary Material (Pérez-Ortiz et al. (2024))), and it will follow from our results that both are GROW.

Additionally, we show that any GROW e-statistic is also relatively GROW e-statistic in the group-invariant setting. Hence, T^{M_n} is relatively GROW as well. This growth rate optimality motivates the use of T^{M_n} in optional continuation settings. As a further contribution, we show that every time that M_n is a maximal invariant, the sequence $T = (T^{M_n})_{n \in \mathbb{N}}$ is a nonnegative martingale. This extends its use and optimality to optional stopping.

The rest of this article is organized as follows. In Section 2, we introduce notation, formally lay the groundwork for group-invariant testing, review e-statistics and their optimality criteria, and discuss related work. Section 3 is devoted to stating our main results: showing that the e-statistic T^{M_n} for a maximally invariant function $M_n = m_n(X^n)$ is both GROW and relatively GROW, proving that T^{M_n} is suited for both optional continuation and optional stopping, and extending these results to composite hypotheses, that is, sets Δ_1 and Δ_0 of δ's, both with and without a prior distribution imposed on them (for general discussion on how to choose δ_j, Δ_j or such priors, we refer to GHK, Section 6). Next, in Section 4, we apply our results to two examples. We end this article with Section 5, where we discuss further the technical conditions that our results require and related work on group-invariant testing. Section 6 contains all proofs that were omitted earlier.

2. Preparation for the main results. This section is structured as follows. We first introduce notation. Then, in Section 2.2, we introduce the formal setup and our running example, the t-test. In Section 2.3, we define e-statistics, our main objects of study, and in Section 2.4 we define our optimality criteria. Finally, Section 2.5 highlights previous work.

2.1. *Notation.* All spaces that we consider are assumed to be topological spaces with an additional measurable structure given by the respective σ-algebra of Borel sets. We write X for a random variable taking values in the observation space $\mathcal{X}$, and $X^n := (X_1, \dots, X_n)$ for n independent copies of X under the distributions that are to be considered.

Statistics of the data X^n are denoted as $T = t(X^n)$, where t is a measurable map $t : \mathcal{X}^n \to \mathcal{T}_n$. We use letters $\mathbf{P}$ and $\mathbf{Q}$ to refer to distributions of X. For a statistic $T = t(X^n)$, we write $\mathbf{P}^T$ for the image measure of $\mathbf{P}$ under t, that is, $\mathbf{P}^T\{T \in A\} = \mathbf{P}\{t(X^n) \in A\}$ for measurable $A \subseteq \mathcal{T}_n$ (note that we may think of T as a random variable on the space $\mathcal{X}^n$).

When writing conditional expectations, we write $\mathbf{E}^{\mathbf{P}}[f(X)|Y]$, and $\mathbf{P}^{X|y}$ for the conditional distribution of X given $Y = y$. We only deal with situations where such conditional distributions exist. If we are considering a set of distributions parameterized in terms of a parameter space Θ, we write $\mathbf{E}^{\mathbf{P}}_{\theta}[f(X)]$ rather than $\mathbf{E}^{\mathbf{P}_\theta}[f(X)]$ for the sake of readability. Furthermore, for a prior distribution $\mathbf{\Pi}$ on Θ, we write $\mathbf{\Pi}^{\theta}\mathbf{P}_\theta$ for the marginal distribution that assigns probability $\mathbf{\Pi}^{\theta}\mathbf{P}_\theta\{X \in B\} = \int \mathbf{P}_\theta\{X \in B\}\,\mathrm{d}\mathbf{\Pi}(\theta)$ to any measurable set $B \subseteq \mathcal{X}$. For the posterior distribution of θ given X we write $\mathbf{\Pi}^{\theta|X}$. The Kullback–Leibler (KL) divergence between $\mathbf{Q}$ and $\mathbf{P}$ is denoted by $\mathrm{KL}(\mathbf{Q}, \mathbf{P}) = \mathbf{E}^{\mathbf{Q}}[\ln(\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P})]$ (Kullback and Leibler (1951)) whenever the Radon–Nikodym derivative $\mathrm{d}\mathbf{Q}/\mathrm{d}\mathbf{P}$ exists. Given two subsets H, K of a group G we write $HK = \{hk : h \in H, k \in K\}$ for the set of all products between elements of H and elements of K. Similarly, for $g \in G$ and $K \subseteq G$, we write $gK = \{gk : k \in K\}$ for the translation of K by g, and $K^{-1} = \{k^{-1} : k \in K\}$ for the set of inverses of K. We say that K is symmetric if $K = K^{-1}$. If G acts on $\mathcal{X}$, then we denote the action of G on $\mathcal{X}$ by $(g, x) \mapsto gx$ for $g \in G$ and $x \in \mathcal{X}$, and extend the action to $\mathcal{X}^n$ component-wise; that is, $(g, x^n) \mapsto gx^n := (gx_1, \dots, gx_n)$ for $g \in G$ and $x^n \in \mathcal{X}^n$. We write $gB = \{gb : b \in B\}$ for the left translate of a subset $B \subseteq \mathcal{X}$ by g. If G acts on Θ, the notation is completely analogous.

2.2. *Group invariance.* We consider a group G that acts freely on both the observation space $\mathcal{X}$ and the parameter space Θ. Recall that G acts freely on a set $\mathcal{Z}$ if anytime that $gz = z$ for some $g \in G$ and $z \in \mathcal{Z}$, then g is the identity element of the group G. A probabilistic model $\mathcal{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$ on $\mathcal{X}$ is said to be invariant under the action of G if the distribution $\mathbf{P}_\theta$ satisfies

$$\mathbf{P}_\theta\{X \in B\} = \mathbf{P}_{g\theta}\{X \in gB\} \tag{3}$$

for any $g \in G$, any measurable $B \subseteq \mathcal{X}$, and any $\theta \in \Theta$. Furthermore, a function $m(x)$ is said to be invariant under the action of G if $m(gx) = m(x)$ for all $x \in \mathcal{X}$ and all $g \in G$; in other words, m is constant on the orbits of G. Moreover, m is said to be maximally invariant if it indexes the orbits of $\mathcal{X}$ under the action of G; that is, $m(x) = m(x')$ for $x, x' \in \mathcal{X}$ if and only if there exists a $g \in G$ such that $x = gx'$. A statistic is called (maximally) invariant if the corresponding function is. These definitions are completely analogous for functions defined on Θ. In particular, we study situations where the parameter of interest $\delta = \delta(\theta)$ is a maximally invariant function of the parameter θ. We then say that δ is a maximally invariant parameter.

We now reparametrize the problem described in (1) using the group G. Using that the action of the group on the parameter space is free, we can reparametrize each orbit in Θ/G with G. Indeed, we can pick an arbitrary but fixed element in the orbit $\theta_0 \in \delta^{-1}(\delta_0)$ and, for any other element $\theta \in \delta^{-1}(\delta_0)$, we can identify θ with the group element $g(\theta) \in G$ that transports θ_0 to θ, that is, such that $g(\theta)\theta_0 = \theta$. Hence, with a slight abuse of notation, we can identify $\theta \in \delta^{-1}(\delta_0)$ with $g = g(\theta) \in G$ and identify $\mathbf{P}_\theta = \mathbf{P}_{g(\theta)\theta_0}$ with $\mathbf{P}_g$. Define $\mathbf{Q}_g$ using the same construction in the alternative model by an analogous choice of $\theta_1 \in \delta^{-1}(\delta_1)$. The starting problem (1) may now be rewritten in the form

$$\mathcal{H}_0 : X^n \sim \mathbf{P}_g, \quad g \in G, \quad \text{vs.} \quad \mathcal{H}_1 : X^n \sim \mathbf{Q}_g, \quad g \in G. \tag{4}$$

To make notation more succinct, we use $\mathcal{Q} = \{\mathbf{Q}_g\}_{g \in G}$ to denote the alternative hypothesis and $\mathcal{P} = \{\mathbf{P}_g\}_{g \in G}$ for the null. As will follow from our discussion, our results are insensitive to the choices of $\theta_0 \in \delta^{-1}(\delta_0)$ and $\theta_1 \in \delta^{-1}(\delta_1)$.

As mentioned in the Introduction, tests for (4) are classically based on the likelihood ratio T^{M_n} of a maximally invariant statistic $M_n = m_n(X^n)$, as in (2). While the distribution of M_n might be unknown, it is well known that its likelihood ratio can be computed by integration over the group G whenever the following three conditions—which will be explained in brief—hold: (1) the action is continuous and proper, (2) G is a σ-compact locally compact topological group, and (3) for all g, $\mathbf{P}_g$ and $\mathbf{Q}_g$ are dominated by a relatively left invariant measure ν. In (1), an action is proper if the map $G \times \mathcal{X}^n \to \mathcal{X}^n \times \mathcal{X}^n$ defined by $(g, x^n) \mapsto (gx^n, x^n)$ is proper, that is, the inverse of any compact set is compact. In (2), a topological group is a group equipped with a topology, such that the group operation, seen as a function $G \times G \to G$, is continuous. Under (3), we assume the existence of densities p_g and q_g for $\mathbf{P}_g$ and $\mathbf{Q}_g$, respectively, with respect to ν for each $g \in G$. Furthermore, since G is assumed to be locally compact, there exists a measure ρ on G that is right invariant (see Bourbaki (2004), VII, Section 1, number 2). This means that for any $g \in G$ and any $B \subseteq G$ that is measurable, it holds that $\rho\{Bg\} = \rho\{B\}$. The measure ρ, called the right Haar measure, is unique up to a multiplicative factor and is finite if and only if G is compact. Using disintegration-of-measure results from Bourbaki ((2004), VIII.27), Andersson (1982) shows that T^{M_n} can be computed as

$$T^{M_n} = \frac{q^{M_n}(m_n(X^n))}{p^{M_n}(m_n(X^n))} = \frac{\int_G q_g(X^n)\,\mathrm{d}\rho(g)}{\int_G p_g(X^n)\,\mathrm{d}\rho(g)}. \tag{5}$$

This is known as Wijsman's representation theorem (for extended statement and discussion, see Eaton (1989), Theorem 5.9). Note that (5) implies that the likelihood ratio T^{M_n} is independent of the choice of maximal invariant M_n. Remarkably, work by Stein, reported by Hall, Wijsman and Ghosh (1965), shows that it does not even matter whether we consider a maximal invariant of the original data, or whether we first reduce the data through sufficiency and then consider a maximal invariant of the sufficient statistic. In the t-test example, this shows that the likelihood ratio of the t-statistic is equal to that of M_n as in the start of the Introduction. We further discuss this result in Section S1 of the Supplementary Material (Pérez-Ortiz et al. (2024)).

Finally, the classic theorem of Hunt and Stein (Lehmann and Romano (2005), Section 8.5) shows that, under some regularity conditions, when looking for a test that is max-min optimal in the sense of power, it is sufficient to look among invariant tests, that is, tests that can be written as a function of T^{M_n} as in (2). One of the crucial assumptions underlying their result is the *amenability* of G. A group G is amenable if there exists a sequence of almost-right-invariant probability distributions, that is, a sequence $\mathbf{\Pi}_1, \mathbf{\Pi}_2, \dots$ such that, for any measurable set $B \subseteq G$ and $g \in G$

$$\lim_{k\to\infty} \left|\mathbf{\Pi}_k\{H \in B\} - \mathbf{\Pi}_k\{H \in Bg\}\right| = 0.$$

Amenable groups have been thoroughly studied (Paterson (1988)) and include, among others, all finite, compact, commutative, and solvable groups. The easiest example of a nonamenable group is the free group in two elements and any group containing it. Another prominent example of a nonamenable group is that of invertible $d \times d$ matrices with matrix multiplication.

EXAMPLE 1 (t-test under Gaussian assumptions). Consider an i.i.d. sample $X^n = (X_1, \dots, X_n)$ of size $n \in \mathbb{N}$ from an unknown Gaussian distribution $N(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. In the 1-sample t-test, we are interested in testing whether $\mu/\sigma = \delta_0$ or $\mu/\sigma = \delta_1$ for some $\delta_0, \delta_1 \in \mathbb{R}$. For $c \in \mathbb{R}^+$, we have that $cX \sim N(c\mu, c\sigma)$, so it follows that the Gaussian model is invariant under scale transformations. The corresponding group is $G = (\mathbb{R}^+, \cdot)$, which acts on $\mathcal{X}^n$ by component-wise multiplication and on Θ by $(c, (\mu, \sigma)) \mapsto (c\mu, c\sigma)$ for each $c \in G$ and $(\mu, \sigma) \in \Theta$. The parameter of interest, $\delta = \mu/\sigma$, is scale-invariant and indexes the orbits of the action of G on Θ. A maximally invariant statistic is $M_n := (X_1/|X_1|, \dots, X_n/|X_1|)$. The right Haar measure ρ on G is given by $\mathrm{d}\rho(\sigma) = \mathrm{d}\sigma/\sigma$, so that the likelihood ratio of M_n can be expressed, as in (5), by

$$T^{M_n} = \frac{\int_{\sigma>0} \frac{1}{\sigma^n} \exp(-\frac{n}{2}[(\frac{\bar{X}_n}{\sigma} - \delta_1)^2 + \frac{1}{n}\sum_{i=1}^n (\frac{X_i - \bar{X}}{\sigma})^2]) \frac{\mathrm{d}\sigma}{\sigma}}{\int_{\sigma>0} \frac{1}{\sigma^n} \exp(-\frac{n}{2}[(\frac{\bar{X}_n}{\sigma} - \delta_0)^2 + \frac{1}{n}\sum_{i=1}^n (\frac{X_i - \bar{X}}{\sigma})^2]) \frac{\mathrm{d}\sigma}{\sigma}}, \tag{6}$$

where $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. The results by Stein, discussed in Section S1 of the Supplementary Material (Pérez-Ortiz et al. (2024)), show that the likelihood ratio of the t-statistic, that is, $M_{\mathcal{S},n} \propto \hat{\mu}_n/\hat{\sigma}_n$, is equal to the expression obtained in (6).

2.3. *The family of e-statistics, and optional continuation and stopping*. We now define *e*-statistics, our measure of evidence against the null hypothesis. The family of *e*-statistics comprises all nonnegative real statistics whose expected value is bounded by one under all elements of the null, that is, all statistics $T_n = t_n(X^n)$ such that $T_n \geq 0$ and

$$\sup_{g \in G} \mathbf{E}_g^{\mathbf{P}}[T_n] \leq 1. \tag{7}$$

An example of an *e*-statistic is the likelihood ratio statistic in any simple-vs-simple testing problem (see, e.g., GHK, Section 1 or Ramdas et al. (2023)). In particular, (2) is an *e*-statistic for the hypotheses in (4). *e*-statistics are appropriate in optional continuation contexts because of the following two properties that are consequences of (7).

1. The type-I error of the test that rejects the null hypothesis anytime that $T_n \geq 1/\alpha$ is smaller than α, a consequence of (7) and Markov's inequality.
2. Suppose that X^n and X^m are the independent outcomes of two subsequent experiments. Let $T_n = t_n(X^n)$ be an *e*-statistic for X^n and let $\{T_{m,\varphi} : \varphi \in \Phi\}$ be a family of *e*-statistics for X^m indexed by some set Φ. Suppose further that, after observing the first sample X^n, the specific $T_{m,\varphi}$ used to measure evidence for the second sample is chosen as a function of X^n, that is, we use $T_{m,\hat{\varphi}}$ where $\hat{\varphi} = \hat{\varphi}(X^n)$ is some function of X^n. Then $T_{n+m} := T_n T_{m,\hat{\varphi}}$ is also an *e*-statistic, irrespective of the definition of $\hat{\varphi}$. In particular, this includes the scenario where we only continue to the second experiment if a certain outcome is observed in the first one. Indeed, Φ may contain a special value $\mathbf{1}$ so that $t_m(X^m; \mathbf{1}) = 1$ is constant, irrespective of X^m. Then, $T_{n+m} = T_n$ every time that $\hat{\varphi} = \mathbf{1}$.

Together, these two properties imply that the test that rejects the null if $T_{n+m} \geq 1/\alpha$ has type-I error bounded by α, no matter the definition of $\hat{\varphi}$. Such type-I error guarantees are essentially impossible using p-values (GHK, Section 1.3). Some—not all—types of *e*-statistics can additionally be used in two related settings: (a) *optional stopping*, when there is a single sequence of data $X_1, X_2, \dots$ and we want to do a test with type-I error guarantees based on all data seen so far, irrespective of when we stop; and (b) optional continuation as in 2. above, but with individual *e*-statistics whose sample size is itself not fixed but determined by some stopping rule. As is well known, for both (a) and (b) it is sufficient that $(T_n)_{n \in \mathbb{N}}$ is a nonnegative martingale with respect to some filtration $\mathcal{F}$ (see, e.g., Ramdas et al. (2023), or GHK). The first part follows from Ville's inequality for nonnegative martingales: the probability that there will *ever* be a sample size n at which $T_n \geq 1/\alpha$ is bounded by α. We thus have type-I error control under optional stopping, which takes care of (a) above. The optional stopping theorem implies that for every stopping time τ adapted to $\mathcal{F}$, T_τ is also an *e*-statistic, taking care of (b). For completeness, we provide more details in Section S3 of the Supplementary Material (Pérez-Ortiz et al. (2024)), including a subtlety regarding (b): while they seem unlikely to arise in practice, there do exist stopping times τ' relative to the data that are not stopping times relative to $\mathcal{F}$. We show an example where $T_{\tau'}$ is not an *e*-statistic and (b) breaks.

2.4. *Optimality criteria for e-statistics.* The standard optimality criterion for hypothesis tests satisfying a certain type-I error guarantee is worst-case power maximization for a fixed-sample-size or, with classic sequential tests, for a fixed stopping rule. This criterion cannot be used when the stopping rule is unknown because knowledge of the stopping rule is required by the definition of power. Additionally, an e-statistic that optimizes power at fixed stopping time will take the value zero with positive probability, making it useless for optional continuation by multiplication. A more sensible criterion for e-statistics under optional continuation is growth rate optimality in the worst case (GHK). Should it exist, an e-statistic T_n^* is GROW if it maximizes the worst-case expected logarithmic value under the alternative hypothesis, that is, if it maximizes

$$T_n \mapsto \inf_{g\in G} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n] \tag{8}$$

over all e-statistics. The following theorem, stated in our notation for group-invariant problems, shows that in most cases the GROW e-statistic takes the form of a particular Bayes factor.

THEOREM 1 (GHK Theorem 1 in Section 4.3). *Suppose that there exists a statistic $V_n = v_n(X^n)$ such that*

$$\inf_{\mathbf{\Pi}_0,\mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^g\mathbf{Q}_g, \mathbf{\Pi}_0^g\mathbf{P}_g\big) = \min_{\mathbf{\Pi}_0,\mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^g\mathbf{Q}_g^{V_n}, \mathbf{\Pi}_0^g\mathbf{P}_g^{V_n}\big) < \infty, \tag{9}$$

where $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ are probability distributions on G. Let $\mathbf{\Pi}_0^\star$ and $\mathbf{\Pi}_1^\star$ be probability distributions that achieve the minimum on the right-hand side. Then

$$\max_{T_n \text{ e-stat.}} \inf_{g\in G} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n] = \mathrm{KL}\big(\mathbf{\Pi}_1^{\star g}\mathbf{Q}_g^{V_n}, \mathbf{\Pi}_0^{\star g}\mathbf{P}_g^{V_n}\big),$$

and the maximum on the left is achieved, essentially uniquely, by T_n^* as given by

$$T_n^* := \frac{\int q_g^{V_n}(v_n(X^n))\,\mathrm{d}\mathbf{\Pi}_1^\star(g)}{\int p_g^{V_n}(v_n(X^n))\,\mathrm{d}\mathbf{\Pi}_0^\star(g)}.$$

Here 'essentially uniquely' means that any other e-statistic achieving the maximum must coincide with T_n^* almost surely, under all $\mathbf{P}_g$ and $\mathbf{Q}_g$ with $g \in G$. In words, the e-statistic T_n^* is GROW for testing $\{\mathbf{P}_g\}_{g\in G}$ against $\{\mathbf{Q}_g\}_{g\in G}$.

The statistic V_n may be any measurable function taking values in any set $\mathcal{V}_n$ (equipped with its corresponding σ-algebra), but in all our examples we can take $\mathcal{V}_n = \mathbb{R}^m$ for some $m \le n$. By allowing $V_n \ne X^n$, the theorem also covers cases in which the infimum on the left in (9) is not achieved. This might be the case when the group G is not compact, as in the t-test example. Corollary 3 in the next section, which gives the GROW e-statistic when G is possibly noncompact, uses crucially this feature of Theorem 1.

Given their worst-case nature, GROW e-statistic, while appropriate in some scenarios (e.g., testing exponential families with given minimum effect sizes and no nuisance parameters), are too conservative in others (GHK). GHK propose, for those cases, to maximize a relative form of (8), leading to less conservative e-statistics. We say that an e-statistic T_n^* is relatively GROW if it maximizes the gain in expected logarithmic value relative to an oracle that is given the particular distribution in the alternative hypothesis from which data are generated, that is, if T_n^* maximizes, over all e-statistics,

$$T_n \mapsto \inf_{g\in G}\Big\{\mathbf{E}_g^{\mathbf{Q}}[\ln T_n] - \sup_{T_n' \text{ e-stat.}} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n']\Big\}. \tag{10}$$

As we will see and contrary to the general case, in the group-invariant setting, any GROW e-statistic is also relatively GROW. Hence, both criteria coincide and the differences that have been observed between them (raising the sometimes difficult question: which one to choose?) are not a concern for our purposes (Ramdas et al. (2023)).

2.5. *Previous and related work.* Group-invariant problems have a long tradition in statistics. They have been studied both for fixed-sample-size experiments Eaton (1989), Lehmann and Romano (2005) and classical, Wald-type sequential experiments (Cox (1952), Rushton (1950)). For fixed-sample-size tests, our main result can be viewed, to some extent, as an anytime-valid analogue of the Hunt–Stein theorem. The proof techniques that are needed for our result are, however, distinct. At the core of the proof of the Hunt–Stein theorem lies the fact that the power is a linear function of the test under consideration. In its proof, an approximate symmetrization of the test is carried out using almost-right-invariant priors without affecting power guarantees. This line of reasoning cannot be directly translated to our setting because of the nonlinearity of the objective function that characterizes the optimal e-statistics that we consider (see Section 2.4). As for sequential tests with group invariance, most previous work (including the pioneering Cox (1952), Rushton (1950) and in fact, as far as we could ascertain, all work pre-dating Robbins (1970)) dealt, like Wald's original SPRT, with a priori fixed stopping rules and is not directly comparable to our anytime-valid work (see Section S3 of the Supplementary Material (Pérez-Ortiz et al. (2024)) for elaboration of this point). Notable exceptions are the works of Robbins (1970) and Lai (1976), who do consider what we now call anytime validity. Lai (1976) also used the expression in (6) for the t-test, which, in our terminology, is using the fact that it gives an e-statistic. However, our main concern, optimality of e-statistics, has not been explored in this context.

Related ideas can also be found in the Bayesian literature, where group-invariant inference with right Haar priors has been studied (Berger, Pericchi and Varshavsky (1998), Dawid, Stone and Zidek (1973)). It has been shown that, in contrast to some other improper priors, inference based on right Haar priors yields admissible procedures in a decision-theoretical sense (Eaton and Sudderth (1999, 2002)). However, there have also been concerns that the underlying group (and hence the right Haar prior) is not uniquely defined in some situations, and that different choices lead to different conclusions (Berger and Sun (2008), Sun and Berger (2007)). Interestingly, as we briefly discuss in Section 5 and at length in Section S2 of the Supplementary Material (Pérez-Ortiz et al. (2024)), this issue cannot arise in our setting. In the same section of the supplement, we point out similarities and the main difference to the information-theoretic work of Liang and Barron (2004), who provide exact min-max procedures for predictive density estimation for general location and scale families under Kullback–Leibler loss. In a nutshell, despite some similarities, the precise min-max result that they prove is not comparable to the results presented here.

3. Main results. In this section, we state the main results of this article. In Section 3.1, we show that the likelihood ratio T^{M_n} for a maximal invariant M_n is simultaneously GROW and relatively GROW. Next, in Section 3.2, we show that T^{M_n} can be used to build an anytime-valid test. Finally, in Section 3.3 we extend these results to the case that the hypotheses remain composite after reduction by invariance.

3.1. *GROW for simple invariant hypotheses.* In order to build intuition, we first demonstrate our line of reasoning using the very special case of finite groups. So, assume for now that G is a finite group, for instance, a group of permutations. Since the uniform probability distribution $\mathbf{\Pi}_{\mathrm{U}(G)}$ on G is right invariant, the right Haar measure ρ coincides with $\mathbf{\Pi}_{\mathrm{U}(G)}$ up to scaling. By Wijsman's representation theorem (5), the likelihood ratio for any maximal invariant $M_n = m_n(X^n)$ can be written as

$$T^{M_n} = \frac{q^{M_n}(m_n(X^n))}{p^{M_n}(m_n(X^n))} = \frac{\frac{1}{|G|}\sum_{g\in G} q_g(X^n)}{\frac{1}{|G|}\sum_{g\in G} p_g(X^n)}. \tag{11}$$

Furthermore, Theorem 1 above takes a simple form for finite parameter spaces, as is the case here, namely

$$\max_{T_n \text{ } e\text{-stat.}} \min_{g\in G} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n] = \min_{\mathbf{\Pi}_0,\mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^g \mathbf{Q}_g, \mathbf{\Pi}_0^g \mathbf{P}_g\big), \tag{12}$$

where the minimum on the right hand side is taken over all pairs of distributions on G. We now employ the information processing inequality (Cover and Thomas (2006), Section 2.8) which says that KL divergence decreases when taking functions of the data (i.e., if $\mathbf{A}$ and $\mathbf{B}$ are distributions for X and $U = u(X)$, then $\mathrm{KL}(\mathbf{A}\|\mathbf{B}) \geq \mathrm{KL}(\mathbf{A}^U\|\mathbf{B}^U)$). In our setting, the information processing equality implies that for any pair $(\mathbf{\Pi}_0, \mathbf{\Pi}_1)$ of probability distributions on G,

$$\mathrm{KL}\big(\mathbf{\Pi}_1^g \mathbf{Q}_g, \mathbf{\Pi}_0^g \mathbf{P}_g\big) \geq \mathrm{KL}\big(\mathbf{Q}^{M_n}, \mathbf{P}^{M_n}\big). \tag{13}$$

This lower bound can be rewritten as $\mathrm{KL}(\mathbf{Q}^{M_n}, \mathbf{P}^{M_n}) = \mathrm{KL}(\mathbf{\Pi}_{\mathrm{U}(G)}^g \mathbf{Q}_g, \mathbf{\Pi}_{\mathrm{U}(G)}^g \mathbf{P}_g)$ because of the second equality in (11). Therefore, the minimum KL on the right-hand side of (12) is achieved for the particular choice of two uniform priors on G. Finally, we have that $\mathbf{E}_g^{\mathbf{Q}}[\ln T^{M_n}] = \mathrm{KL}(\mathbf{Q}^{M_n}, \mathbf{P}^{M_n})$ for all $g \in G$. Putting everything together

$$\max_{T_n \text{ } e\text{-stat.}} \min_{g\in G} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n] = \mathrm{KL}\big(\mathbf{Q}^{M_n}, \mathbf{P}^{M_n}\big) = \min_{g\in G} \mathbf{E}_g^{\mathbf{Q}}[\ln T^{M_n}];$$

in other words, T^{M_n} is a GROW e-statistic. A natural question is whether this same reasoning can be reproduced for infinite groups. If the Haar measure ρ could always be chosen to be a probability measure, we could replace $\mathbf{\Pi}_{\mathrm{U}(G)}$ by ρ everywhere in the reasoning above and conclude that T^{M_n} is GROW in general. However, ρ is finite if and only if G is compact (see, e.g., Reiter and Stegeman (2000), Proposition 3.3.5). This is a severe limitation; it would not even cover our guiding example, the t-test, because the group $(\mathbb{R}^+, \cdot)$ is not compact (see Example 1). The main technical contribution of this article is the extension of the above optimality result to amenable groups (see Section 2.2). Setting technical details aside, the core of the proof of the main Theorem 2 below is replacing the Haar measure above by a sequence of almost-right-invariant probability measures and showing that the KL converges to its infimum. Our arguments require the following additional assumptions.

ASSUMPTION 1. Let G be a topological group acting on a topological space $\mathcal{X}^n$, both equipped with their Borel σ-algebra. The group G, the observation space $\mathcal{X}^n$, and the probabilistic models under consideration satisfy the following three properties:

1. As topological spaces, G and $\mathcal{X}^n$ are Polish—separable and completely metrizable—and locally compact.
2. The action of G on $\mathcal{X}^n$ is free, continuous and proper.
3. The models $\{\mathbf{P}_g\}_{g\in G}$ and $\{\mathbf{Q}_g\}_{g\in G}$ are invariant and have densities with respect to a common measure μ on $\mathcal{X}^n$ that is relatively left invariant with some multiplier χ—$\mu\{gB\} = \chi(g)\mu\{B\}$ for any measurable set $B \subseteq \mathcal{X}^n$ and $g \in G$. All densities have a single common support.

Assumption 1 holds in most cases of interest for the purpose of parametric inference; some examples where it holds are given in Section 4. The topological assumptions on G and $\mathcal{X}$ have two purposes. The first is to ensure that Wijsman's representation theorem (5) holds. Though (5) requires slightly weaker assumptions than those presented here (see Section 2.2), the strengthened conditions are needed for the second purpose: to ensure that the observation space $\mathcal{X}^n$ can be put in bijective and bimeasurable[1] correspondence with a subset of $G \times \mathcal{X}^n/G$, where the group G acts naturally by multiplication on the first component (Bondar (1976)). This will be used extensively in the proofs given in Section 6. With these assumptions, everything is in place to state the main results of this article.

THEOREM 2. *Let $M_n = m_n(X^n)$ be a maximally invariant statistic under the action of the group G on $\mathcal{X}^n$. Assume that G is amenable, that Assumption 1 holds, and that there is $\varepsilon > 0$ such that*

$$\mathbf{E}_1^{\mathbf{Q}}\left[\left|\ln \frac{q_1(X^n)}{p_1(X^n)}\right|^{1+\varepsilon}\right], \mathbf{E}^{\mathbf{Q}^{M_n}}\left[\left|\ln \frac{q^{M_n}(M_n)}{p^{M_n}(M_n)}\right|^{1+\varepsilon}\right] < \infty, \tag{14}$$

where the subindex 1 refers to the unit element of G. Then

$$\inf_{\mathbf{\Pi}_0, \mathbf{\Pi}_1} \mathrm{KL}(\mathbf{\Pi}_1^g \mathbf{Q}_g, \mathbf{\Pi}_0^g \mathbf{P}_g) = \mathrm{KL}(\mathbf{Q}^{M_n}, \mathbf{P}^{M_n}),$$

where the infimum is over all pairs $(\mathbf{\Pi}_0, \mathbf{\Pi}_1)$ of probability distributions on G.

COROLLARY 3. *Under the assumptions of Theorem 2, a GROW e-statistic for testing $\mathcal{H}_1$ against $\mathcal{H}_0$ as in (4) is given by the likelihood ratio of any maximally invariant statistic $M_n = m_n(X^n)$, that is,*

$$T^{M_n} = \frac{q^{M_n}(m_n(X^n))}{p^{M_n}(m_n(X^n))}.$$

Corollary 3 follows from the combination of Theorem 2 with Theorem 1. The results are stated in terms of the likelihood ratio of any maximal invariant for the original data. However, as mentioned briefly in Section 2.2 and in detail in Section S1 of the Supplementary Material (Pérez-Ortiz et al. (2024)), one can use instead any maximal invariant for a sufficient statistic of the original data, rather than for the data itself. The resulting likelihood ratio is identical and the optimality results therefore remain valid. Next, we show that in the group-invariant setting, any statistic that is GROW is also relatively GROW, meaning that any *e*-statistic that maximizes (8) also maximizes (10). This is not true in general; the result relies crucially on the invariance of the models. For example, for contingency tables, the two *e*-statistics are vastly different (Turner, Ly and Grünwald (2024)).

THEOREM 4. *Suppose that Part 3 of Assumption 1 is satisfied and that, for each $g \in G$, there exists $h \in G$ such that $\mathrm{KL}(\mathbf{Q}_g, \mathbf{P}_h)$ is finite. Then the map defined by*

$$g \mapsto \sup_{T_n \text{ e-stat.}} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n]$$

is constant. Consequently, any maximizer of (8) also maximizes (10), that is, an e-statistic is GROW if and only if it is relatively GROW for the hypothesis testing problem (4).

COROLLARY 5. *T^{M_n} from Corollary 3 is not only GROW, it is also relatively GROW.*

EXAMPLE 1 (continued). It is known that the group $G = (\mathbb{R}^+, \cdot)$ of the t-test is amenable—the sequence of probability distributions $(\mathrm{Uniform}([-n, n]))_{n\in\mathbb{N}}$ is almost right invariant. It is readily verified that Assumption 1 and condition (14) are also satisfied. Hence, Corollary 3 implies that the likelihood ratio for the t-statistic, given in (6), is a GROW *e*-statistic. Moreover, it follows from Corollary 5 that it is also relatively GROW.

[1]We call an invertible map bimeasurable if both the map and its inverse are measurable.

3.2. *Anytime-validity.* As discussed in Section 2.3, any e-statistic can be used in the context of optional continuation with fixed sample sizes, but not all e-statistics are suitable for optional stopping and optional continuation with data-dependent sample sizes. A sufficient condition that allows us to engage in these two additional uses is that the sequence of e-statistics is a nonnegative martingale. We now show that this is the case for the sequence $(T^{M_n})_{n\in\mathbb{N}}$.

PROPOSITION 6. *If $(M_n)_{n\in\mathbb{N}}$ is a sequence of maximally invariant statistics $M_n = m_n(X^n)$ for the action of G on $\mathcal{X}^n$, then the process $(T^{M_n})_{n\in\mathbb{N}}$ is a nonnegative martingale with respect to the filtration $(\sigma(M_1,\dots,M_n))_{n\in\mathbb{N}}$ under any of the elements of the null hypothesis.*

In particular, Proposition 6 implies that under every stopping time τ defined relative to the filtration induced by $(M_n)_{n\in\mathbb{N}}$, T^{M_τ} is itself an e-statistic; see Section S3 of the Supplementary Material (Pérez-Ortiz et al. (2024)) for the (standard) proof. There is an interesting subtlety here however: if τ' is a stopping time relative to the filtration induced by $(X_n)_{n\in\mathbb{N}}$ but not relative to the coarser filtration induced by $(M_n)_{n\in\mathbb{N}}$, then $T^{M_{\tau'}}$ is not necessarily an e-statistic anymore. Thus, with such $T^{M_{\tau'}}$, we cannot engage in optional continuation. This is generally not a problem, since most stopping times encountered in practice are stopping times relative to the filtration induced by $(M_n)_{n\in\mathbb{N}}$. This includes the aggressive stopping time "stop at the smallest n at which $T^{M_n} \geq 1/\alpha$". However, in Section S3.1 of the Supplementary Material (Pérez-Ortiz et al. (2024)) we give an explicit example of a stopping time τ' relative to the filtration induced by $(X_n)_{n\in\mathbb{N}}$ in the t-test such that $T^{M_{\tau'}}$ is not an e-statistic.

3.3. *GROW for composite invariant hypotheses.* Until now, we have considered null and alternative hypotheses that become simple when viewed through the lens of the maximally invariant statistic. As we saw, in the t-test this corresponds to testing simple hypotheses about the effect size δ. In this section, we consider hypotheses that are composite in the maximally invariant parameter. We also consider problems in which a fixed prior is placed on the maximally invariant parameter δ. This implements the method of mixtures, a standard method to combine test martingales (Darling and Robbins (1968), Wald (1945)), which was already used in the context of the anytime-valid t-test (Lai (1976)).

Suppose that the initial hypotheses are not defined by a single value of the maximally invariant parameter $\delta = \delta(\theta)$, as in (1), but are instead given by

$$\mathcal{H}_0 : \delta(\theta) = \delta, \quad \delta \in \Delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \delta(\theta) = \delta, \quad \delta \in \Delta_1, \tag{15}$$

where Δ_0 and Δ_1 are two sets of possible values of $\delta = \delta(\theta)$. In Section 2.2, we reparametrized $\{\mathbf{P}_\theta\}_{\theta\in\Theta:\delta(\theta)=\delta_0}$ and $\{\mathbf{Q}_\theta\}_{\theta\in\Theta:\delta(\theta)=\delta_1}$ in terms of G, and denoted the resulting models as $\{\mathbf{P}_g\}_{g\in G}$ and $\{\mathbf{Q}_g\}_{g\in G}$ respectively. Instead of only considering δ_0 and δ_1, we can do the same for all $\delta \in \Delta_0$ and $\delta \in \Delta_1$. We denote the resulting models as $\{\mathbf{P}_{g,\delta}\}_{g\in G,\delta\in\Delta_0}$ and $\{\mathbf{Q}_{g,\delta}\}_{g\in G,\delta\in\Delta_1}$. As an example, $\mathbf{P}_{g,\delta_0}$ and $\mathbf{Q}_{g,\delta_1}$ correspond to what were previously simply $\mathbf{P}_g$ and $\mathbf{Q}_g$. The problem (15) may now be rewritten as

$$\mathcal{H}_0 : X^n \sim \mathbf{P}_{g,\delta}, \quad \delta \in \Delta_0, g \in G \quad \text{vs.} \quad \mathcal{H}_1 : X^n \sim \mathbf{Q}_{g,\delta}, \quad \delta \in \Delta_1, g \in G. \tag{16}$$

Since the distribution of a maximally invariant function of the data $M_n = m_n(X^n)$ depends on the parameter δ, these hypotheses are not simple when data are reduced through invariance. The main objective of this section is to show that, when searching for a GROW e-statistic for (16), it is enough to do so for the invariance-reduced problem

$$\mathcal{H}_0 : M_n \sim \mathbf{P}^{M_n}_\delta, \quad \delta \in \Delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : M_n \sim \mathbf{Q}^{M_n}_\delta, \quad \delta \in \Delta_1. \tag{17}$$

We follow the same steps that we followed in Section 3.1, and begin by showing that if there exists a minimizer for the KL minimization problem associated to (17), then it has the same value as that associated to (16).

PROPOSITION 7. *Assume that there exists a pair of probability distributions $\mathbf{\Pi}_0^\star$, $\mathbf{\Pi}_1^\star$ on Δ_0 and Δ_1 that satisfy*

$$\mathrm{KL}\big(\mathbf{\Pi}_1^{\star\delta}\mathbf{Q}_\delta^{M_n}, \mathbf{\Pi}_0^{\star\delta}\mathbf{P}_\delta^{M_n}\big) = \min_{\mathbf{\Pi}_0, \mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^{\delta}\mathbf{Q}_\delta^{M_n}, \mathbf{\Pi}_0^{\delta}\mathbf{P}_\delta^{M_n}\big). \tag{18}$$

For each $g \in G$, define the probability distributions $\mathbf{P}_g^\star = \mathbf{\Pi}_0^{\star\delta}\mathbf{P}_{g,\delta}$ and $\mathbf{Q}_g^\star = \mathbf{\Pi}_1^{\star\delta}\mathbf{Q}_{g,\delta}$ on $\mathcal{X}^n$. If the models $\{\mathbf{P}_g^\star\}_{g\in G}$ and $\{\mathbf{Q}_g^\star\}_{g\in G}$ satisfy the assumptions of Theorem 2, then

$$\inf_{\mathbf{\Pi}_0, \mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^{g,\delta}\mathbf{Q}_{g,\delta}, \mathbf{\Pi}_0^{g,\delta}\mathbf{P}_{g,\delta}\big) = \min_{\mathbf{\Pi}_0, \mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^{\delta}\mathbf{Q}_\delta^{M_n}, \mathbf{\Pi}_1^{\delta}\mathbf{P}_\delta^{M_n}\big).$$

From this proposition, using Theorem 1 and the steps used for Corollaries 3 and 5, we can conclude that the ratio of the Bayes marginals for the invariance-reduced data M_n using the optimal priors $\mathbf{\Pi}_0^\star$ and $\mathbf{\Pi}_1^\star$ is both a GROW and a relatively GROW e-statistic for (16). We now state the corollary and apply it to to our running example, the t-test.

COROLLARY 8. *Under the assumptions of Proposition 7, the statistic given by*

$$T^\star = \frac{\int q_\delta^{M_n}(m_n(X^n))\,\mathrm{d}\mathbf{\Pi}_1^\star(\delta)}{\int p_\delta^{M_n}(m_n(X^n))\,\mathrm{d}\mathbf{\Pi}_0^\star(\delta)}$$

is a (*both absolute and relative*) *GROW e-statistic for* (16).

EXAMPLE 1 (continued). Suppose, in the t-test setting, that we are interested in testing

$$\mathcal{H}_0 : \delta \in (-\infty, \delta_0] \quad \text{vs.} \quad \mathcal{H}_1 : \delta \in [\delta_1, \infty)$$

for some $\delta_0 < \delta_1$, where, recall, $\delta = \mu/\sigma$ is the maximally invariant parameter. Corollary 8 shows that no loss is incurred if we only look among e-statistics that are a function of the maximally invariant function M_n, the t-statistic. Since the density of the t-statistic is monotone in δ, we can use Proposition 3 of GHK, Section 3.1. to infer that the minimum in (18) is achieved by the probability distributions $\mathbf{\Pi}_0^\star$ and $\mathbf{\Pi}_1^\star$ that put all of their mass on δ_0 and δ_1, respectively. Corollary 8 yields that $T_n^* = p_{\delta_1}^{M_n}/p_{\delta_0}^{M_n}$ is GROW among all possible e-statistics of the original data (not only the scale-invariant ones). This result can be extended to other families with this type of monotonicity property.

Another approach to deal with the unknown parameter values is to employ proper prior distributions, as is standard practice both within Bayesian statistics and with e-statistics. That is, we may want to use specific priors $\tilde{\mathbf{\Pi}}_0$ and $\tilde{\mathbf{\Pi}}_1$ on Δ_0 and Δ_1 respectively. If we define for each g the probability distributions $\tilde{\mathbf{P}}_g = \tilde{\mathbf{\Pi}}_0^\delta\mathbf{P}_{g,\delta}$ and $\tilde{\mathbf{Q}}_g = \tilde{\mathbf{\Pi}}_1^\delta\mathbf{Q}_{g,\delta}$, and the resulting models $\{\tilde{\mathbf{P}}_g\}_{g\in G}$ and $\{\tilde{\mathbf{Q}}_g\}_{g\in G}$ also satisfy the conditions of Corollary 3, the proof of Proposition 7 also provides the following corollary.

COROLLARY 9. *Let $\tilde{\mathbf{\Pi}}_0$ and $\tilde{\mathbf{\Pi}}_1$ be two probability distributions on Δ_0 and Δ_1, respectively. Let $\{\tilde{\mathbf{P}}_g\}_{g\in G}$ and $\{\tilde{\mathbf{Q}}_g\}_{g\in G}$ be two probability models defined by $\tilde{\mathbf{P}}_g = \tilde{\mathbf{\Pi}}_0^\delta\mathbf{P}_{g,\delta}$ and $\tilde{\mathbf{Q}}_g = \tilde{\mathbf{\Pi}}_1^\delta\mathbf{Q}_{g,\delta}$. If $\{\tilde{\mathbf{P}}_g\}_{g\in G}$ and $\{\tilde{\mathbf{Q}}_g\}_{g\in G}$ satisfy the conditions of Corollary 3, then the e-statistic*

$$\tilde{T}_n = \frac{\int q_\delta(m_n(X^n))\,\mathrm{d}\tilde{\mathbf{\Pi}}_1(\delta)}{\int p_\delta(m_n(X^n))\,\mathrm{d}\tilde{\mathbf{\Pi}}_0(\delta)} \tag{19}$$

is both GROW and relatively GROW for testing $\{\tilde{\mathbf{P}}_g\}_{g\in G}$ against $\{\tilde{\mathbf{Q}}_g\}_{g\in G}$.

EXAMPLE 1 (continued). Jeffreys (1961) proposed a Bayesian version of the t-test based on the Bayes factor (6) with δ_0 to 0 and a Cauchy prior centered at 0 on δ_1. Popularized as the *Bayesian t-test* (Rouder et al. (2009)), it is an instance of (19) with $\tilde{\mathbf{\Pi}}_1$ set to aforementioned Cauchy prior and $\tilde{\mathbf{\Pi}}_0$ putting mass 1 on $\delta_0 = 0$. It is itself an e-statistic (GHK), but condition (14) of Theorem 2 does not hold because the Cauchy distribution does not have any moments. Thus, we cannot verify whether (19) has the relative GROW property. However, as soon as we replace the Cauchy prior by any prior centered at 0 for which, for some $\varepsilon > 0$, the $(2 + \varepsilon)$th moment exists (such as, e.g., a normal distribution centered at 0, as has also been proposed for this problem), we can use Lemma 1 in the next section (applied with $d = 1$) to infer that assumption (14) holds. Finally, Proposition 9 can be applied to conclude that the corresponding Bayes factor is then relatively GROW.

4. Testing multivariate normal distributions under group invariance. We show how the theory developed in the previous sections can be applied to hypothesis testing under normality assumptions. The latter is particularly suited for the group-invariant setting, because the family of normal distributions carries a natural invariance under scale-location transformations, as we have already seen in Example 1. Different subsets of scale-location transformations correspond to different parameters of interest. We develop two examples in detail. The first is an alternative to Hotelling's T^2 for testing whether the (multivariate) mean of the distribution is identically zero. The corresponding group is that of lower triangular matrices with positive entries on the diagonal. This test is in direct relation with the step-down procedure of Roy and Bargmann (1958)[2] (see also Subbaiah and Mudholkar (1978)). The second example that we consider is, in the setting of linear regression, a test for whether or not a specific regression coefficient is identically zero. In this case, the group is a subset of the affine linear group.

4.1. *The lower triangular group.* Consider data $X^n = (X_1, \dots, X_n)$ where $X_i \in \mathcal{X} = \mathbb{R}^d$. We assume each X_i to have a Gaussian distribution $N(\mu, \Sigma)$ with unknown mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ. We consider a test for whether the mean μ of the distribution is zero. To formalize the test, recall that the Cholesky decomposition of a positive definite matrix Σ is $\Sigma = \Lambda\Lambda'$ for a unique $\Lambda \in \mathrm{LT}^+(d)$. Here, $\mathrm{LT}^+(d)$ denotes the group of lower triangular matrices with positive entries on the diagonal, which is amenable. We can therefore parametrize the Gaussians in terms of (μ, Λ), taking the parameter space to be $\Theta = \mathbb{R}^d \times \mathrm{LT}^+(d)$. In this parametrization, consider the following hypothesis testing problem, which generalizes the t-test (Example 1) to dimensions $d \geq 1$:

$$\mathcal{H}_0 : \Lambda^{-1}\mu = \delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \Lambda^{-1}\mu = \delta_1. \tag{20}$$

A test for whether μ is zero can be obtained by setting $\delta_0 = 0$. The group $\mathrm{LT}^+(d)$ acts freely and continuously on $\mathcal{X}^n$ through component-wise matrix multiplication, that is, $(L, X^n) \mapsto (LX_1, \dots, LX_n)$ for any $L \in \mathrm{LT}^+(d)$. This action is continuous and free, and can be shown to be proper on the restriction of $\mathcal{X}^n$ to matrices of rank d if $n \geq d + 1$. If $X_i \sim N(\mu, \Lambda)$, then $LX_i \sim N(L\mu, L\Lambda)$, so that $\mathrm{LT}^+(d)$ acts on Θ by $(L, (\mu, \Lambda)) \mapsto (L\mu, L\Lambda)$ for each $(\mu, \Lambda) \in \Theta$ and $L \in \mathrm{LT}^+(d)$. A maximally invariant parameter under this action is $\delta(\mu, \Lambda) = \Lambda^{-1}\mu$, so that (20) is indeed a test of the form described in Section 2.2. Furthermore, seen as a subset of $\mathbb{R}^{d\times n}$, the restriction of the Lebesgue measure to $\mathcal{X}^n$ is relatively left-invariant with multiplier $\chi(L) = |\det(L)|^n$. It follows that Assumption 1 holds and therefore, the likelihood ratio of any maximally invariant statistic is GROW by Corollary 3.

[2]Even though not explicitly in group-theoretic terms, the test of Roy and Bargmann (1958) test is based on a different maximally invariant function of the data. The fact that the test statistic of Roy and Bargmann (1958) is maximally invariant is shown by Subbaiah and Mudholkar (1978).

By the results of Hall, Wijsman and Ghosh (1965), recapped in Section S1 of the Supplementary Material (Pérez-Ortiz et al. (2024)), this likelihood ratio must coincide with that of an invariantly sufficient statistic for δ. We now proceed to compute one such statistic. Recall that the pair $S_n = s_n(X^n) = (\bar{X}_n, \bar{V}_n)$, consisting of the unbiased estimators $\bar{X}_n$ and $\bar{V}_n$ for the mean and covariance matrix respectively, is a sufficient statistic for (μ, Σ). Analogous to the technique we used for the parameter space, we can perform the Cholesky decomposition $\bar{V}_n = L_n L_n'$. The statistic $M_{\mathcal{S},n} = m_{\mathcal{S},n}(S_n) = \sqrt{n/(n-1)} L_n^{-1} \bar{X}_n$ is maximally invariant under the action of $\mathrm{LT}^+(d)$ on S_n; in other words, $M_{\mathcal{S},n}$ is invariantly sufficient for δ. Hence, the GROW e-statistic can be written as $T^{M_{\mathcal{S},n}} = q^{M_{\mathcal{S},n}}/p^{M_{\mathcal{S},n}}$. Since it was used in Example 1 (underneath Corollary 9), we give an explicit expression for the likelihood ratio $T^{M_{\mathcal{S},n}}$ when $\delta_0 = 0$, from which values for other δ_0 can be computed. It is based on a more general computation in Section S4 of the Supplementary Material (Pérez-Ortiz et al. (2024)).

LEMMA 1. *For the maximally invariant statistic* $M_{\mathcal{S},n} = \sqrt{\frac{n}{n-1}} L_n^{-1} \bar{X}_n$, *we have*

$$\frac{q^{M_{\mathcal{S},n}}(m_{\mathcal{S},n}(S_n))}{p^{M_{\mathcal{S},n}}(m_{\mathcal{S},n}(S_n))} = \mathrm{e}^{-\frac{n}{2}\|\delta_1\|^2} \int \mathrm{e}^{n\langle \delta_1, T A_n^{-1} M_{\mathcal{S},n}\rangle} \, \mathrm{d}\mathbf{P}_{n,I}(T), \tag{21}$$

where A_n *is the lower triangular matrix resulting from the Cholesky decomposition* $I + M_{\mathcal{S},n} M_{\mathcal{S},n}' = A_n A_n'$, *and* $\mathbf{P}_{n,I}^T$ *is the distribution according to which* $nTT' \sim W(n, I)$, *a Wishart distribution.*

PROOF. This follows from Proposition S5 in Section S4 of the Supplementary Material with $\gamma = \sqrt{n}\delta_1$, $X = \sqrt{n}\bar{X}_n$, $m = n-1$, and $S = \bar{V}_n$. $\square$

4.2. *Linear regression.* Consider the problem of testing whether one of the coefficients of a linear regression is zero under Gaussian error assumptions. Assume that the observations are of the form $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$, where $X_i, Y_i \in \mathbb{R}$ and $Z_i \in \mathbb{R}^d$ for each i. We consider the the linear model given by

$$Y_i = \gamma X_i + \beta' Z_i + \sigma \varepsilon_i,$$

where $\gamma \in \mathbb{R}$, $\beta \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^+$ are the parameters, and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. errors with standard Gaussian distribution $N(0,1)$. We are interested in testing

$$\mathcal{H}_0 : \gamma/\sigma = \delta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \gamma/\sigma = \delta_1. \tag{22}$$

A test for whether $\gamma = 0$ is readily obtained by taking $\delta_0 = 0$. This problem is invariant under the action of the group $G = \mathbb{R}^+ \times \mathbb{R}^d$ given by $((c, v), (X, Y, Z)) \mapsto (X, cY + v'Z, Z)$ (Eaton (1989), Kariya (1980)). The corresponding action of G on the parameter space is given by $((c, v), (\gamma, \beta, \sigma)) \mapsto (c\gamma, c\beta + v, c\sigma)$. A maximally invariant parameter is $\delta(\gamma, \beta, \sigma) = \gamma/\sigma$, so that the problem in (22) is of the form described in Section 2.2. Furthermore, it can be shown that the action of G on $\mathcal{X}$ is continuous and proper, and that G is amenable. Since the Lebesgue measure is again relatively left invariant, it follows that Assumption 1 holds. All that remains is to find a maximally invariant function of the data. To this end, define the vectors $Y^n = (Y_1, \dots, Y_n)'$ and $X^n = (X_1, \dots, X_n)'$, and the $n \times d$ matrix $\mathbf{Z}_n = [Z_1, \dots, Z_n]'$ whose rows are the vectors $Z_1, \dots, Z_n$. Assume that $\mathbf{Z}_n$ has full rank. A maximally invariant function of the data is given by $M_n = (\frac{\boldsymbol{A}_n' Y^n}{\|\boldsymbol{A}_n' Y^n\|}, X^n, \mathbf{Z}_n)$, where $\boldsymbol{A}_n$ is an $(n-d) \times n$ matrix whose columns form an orthonormal basis for the orthogonal complement of the column space of $\mathbf{Z}_n$ (Bhowmik and King (2007), Kariya (1980)). In order to compute the likelihood ratio for M_n, we assume that the mechanism that generates X^n and $\mathbf{Z}_n$ is the same under both hypotheses, so that we only need to consider the distribution of $\boldsymbol{U}_n = \frac{\boldsymbol{A}_n' Y^n}{\|\boldsymbol{A}_n' Y^n\|}$ conditionally on X^n and $\boldsymbol{Z}_n$. Bhowmik and King (2007) show that for arbitrary effect size δ, the density of this distribution is given by

$$p_\delta^{\boldsymbol{U}_n}(u|X^n, \boldsymbol{Z}_n) = \frac{1}{2}\Gamma\left(\frac{k}{2}\right)\pi^{-\frac{k}{2}} e^{c(\delta)}\left[{}_1F_1\left(\frac{k}{2}, \frac{1}{2}, \frac{a^2(u,\delta)}{2}\right) + \sqrt{2}a(u,\delta)\frac{\Gamma((1+k)/2)}{\Gamma(k/2)}{}_1F_1\left(\frac{1+k}{2}, \frac{3}{2}, \frac{a^2(u,\delta)}{2}\right)\right],$$

where $k = n - d$, u is a unit vector in $\mathbb{R}^k$, $a(u,\delta) = \delta X^{n\prime}\boldsymbol{A}_n u$, $c(\delta) = -\frac{1}{2}\delta^2 X^{n\prime}\boldsymbol{A}_n\boldsymbol{A}_n' X^n$, and ${}_1F_1$ is the confluent hypergeometric function. This can be used to compute the likelihood ratio for M_n, which is the relatively GROW e-statistic for testing (22). In fact, Bhowmik and King compute in more generality the density of the maximally invariant statistic when X is allowed to have a non-linear effect on Y. This does not impact the group invariance structure of the model, so that our results can also be used in this semilinear setting if the hypotheses are adjusted accordingly.

5. Discussion and future work.

In this concluding section, we bring up an issue that deserves further discussion and may inspire future work. We also use this issue to highlight the differences between our work and related work in a Bayesian context.

5.1. *Amenability is not always necessary.*

We have shown that, if a hypothesis testing problem is invariant under a group G and our assumptions are satisfied, then amenability of G is a sufficient condition for the likelihood ratio of the maximal invariant to be GROW. A natural question is therefore whether amenability is also a necessary condition for the latter to hold. This is not only of theoretical relevance: some groups that are important for statistical practice are not amenable. For instance, the general linear group $\mathrm{GL}(d)$, which is the relevant group in Hotelling's test, is nonamenable. The setup of Hotelling's test is similar to that in Section 4.1, except that the hypotheses are given by

$$\mathcal{H}_0 : \|\Lambda^{-1}\mu\|^2 = 0 \quad \text{vs.} \quad \mathcal{H}_1 : \|\Lambda^{-1}\mu\|^2 = \gamma. \tag{23}$$

A maximally invariant statistic is the T^2-statistic $n\bar{X}_n'\bar{V}_n^{-1}\bar{X}_n$, where, as in Section 4.1, $\bar{X}_n$ and $\bar{V}_n$ are the unbiased estimators of the mean and the covariance matrix, respectively. Notice that this test is equivalent to (20) with the alternative expanded to $\Delta = \{\delta : \|\delta\|^2 = \gamma\}$, but that T^2 is not a maximal invariant under the lower triangular group. However, Giri, Kiefer and Stein (1963) have shown that for $d = 2$ and $n = 3$, the likelihood ratio of the T^2-statistic can be written as an integral over the likelihood ratio in (21) with a proper prior on $\delta \in \Delta$ as defined there. It follows as a result of Proposition 7 that the likelihood ratio of the T^2-statistics is also GROW in the case that $d = 2$ and $n = 3$. These results can be extended to the case that $d = 2$ with arbitrary n by the work of Shalaevskii (1971). An interesting question is whether amenability can be replaced by a weaker condition, and/or whether a counterexample to Theorem 2 for nonamenable groups can be given.

5.2. *Nonuniqueness issues with right Haar priors do not arise.*

As the above example illustrates, it is sometimes possible to represent the same $\mathcal{H}_0$ and $\mathcal{H}_1$ via (at least) two different groups. As we explain in full detail in Section S2 of the Supplementary Material (Pérez-Ortiz et al. (2024)), this is generally unproblematic: as soon as the assumptions of Theorem 2 hold for at least one of the two groups, we can construct the GROW e-statistic, and it is uniquely defined. Superficially, this may seem to contradict Sun and Berger (2007) who point out that in some settings, the underlying group is not uniquely determined and then the right Haar prior for the considered model $\mathcal{P}$ is not uniquely defined. Then, different choices of right Haar prior give different Bayesian posteriors—a fact that has sometimes been taken as a criticism of objective Bayesian approaches. Such nonuniqueness is avoided in our approach. The reason is, essentially, that whereas the GROW e-statistic T_n^* is a ratio between Bayes marginals for different models $\mathcal{H}_0$ and $\mathcal{H}_1$ at the same sample size n, the Bayes predictive distribution based on a single model $\mathcal{P}$ is a ratio between Bayes marginals for the same $\mathcal{P}$ at different sample sizes n and $n-1$. The role of "same" and "different" being interchanged, it turns out that this Bayes predictive distribution *can* depend on the group on which the right Haar prior for $\mathcal{P}$ is based. Since the Bayes predictive distribution can be rewritten as a marginal over the Bayes posterior, which is Sun and Berger (2007)'s quantity of interest, it is then not surprising that this Bayes posterior may also change if the underlying group is changed. Instead, one may quantify uncertainty by the *e-posterior*, an e-statistic-based measure of uncertainty recently put forward by Grünwald (2023): if one replaces the standard Bayes posterior on δ by the e-posterior based on the GROW e-statistic T_n^*, the nonuniqueness issue disappears as well.

6. Proofs.

In this section, we give all the proofs that were omitted earlier. We first provide two remarks that will be useful throughout the proofs.

REMARK 1. Without loss of generality, we may modify 3 in Assumption 1 as follows:

3′ The models $\{\mathbf{P}_g\}_{n\in\mathbb{N}}$ and $\{\mathbf{Q}_g\}_{n\in\mathbb{N}}$ are invariant and have densities with respect to a common measure ν on $\mathcal{X}^n$ that is left invariant.

The reason that there is no loss in generality is that from any relatively left-invariant measure μ with multiplier χ, a left-invariant measure ν can be constructed. Indeed, Bourbaki ((2004), Chapter 7, Section 2 Proposition 7) shows that, under our assumptions, for any multiplier χ there exists a function $\varphi : \mathcal{X}^n \to \mathbb{R}$ with the property that $\varphi(gx) = \chi(g)\varphi(x)$ for any $x \in \mathcal{X}$ and $g \in G$. With this function at hand, one can define the measure $\mathrm{d}\nu(x) = \mathrm{d}\mu(x)/\varphi(x)$, which is left invariant. After multiplication by φ, probability densities with respect to μ are readily transformed into probability densities with respect to ν. The invariance of the models implies that the densities of $\mathbf{P}_g$ and $\mathbf{Q}_g$ with respect to ν take the form $p_g(x^n) = p_1(g^{-1}x^n)$ and $q_g(x^n) = q_1(g^{-1}x^n)$ for any $x^n \in \mathcal{X}^n$, where 1 denotes the unit element of the group G. It follows that for any $g, h \in G$ it holds that $p_g(x^n) = p_h(hg^{-1}x^n)$ for all $x^n \in \mathcal{X}^n$. A similar statement can be made for q_g.

REMARK 2. So far, we have only considered the right Haar measure ρ on G, however on any locally compact group G there also exists a left-invariant measure λ, called the left Haar measure. It can be shown that λ is relatively right invariant with a multiplier Δ, that is, for any measurable $B \subseteq G$ and $g \in G$ it holds that $\lambda\{Bg\} = \Delta(g)\lambda\{B\}$ for any $g \in G$. Moreover, a computation shows that the measure ρ' defined by $\rho'\{B\} = \lambda\{B^{-1}\}$ for each measurable $B \subseteq G$, is right invariant; in other words, ρ' is a right Haar measure. We may therefore choose ρ to be equal to ρ' and in the following, we always refer to right and left Haar measures that are related to each other by that identity. In our proofs, we will use that for any integrable function f defined on G, the identities $\int f(h)\,\mathrm{d}\rho(h) = \int f(h)/\Delta(h)\,\mathrm{d}\lambda(h)$ and $\int f(h^{-1})\,\mathrm{d}\lambda(h) = \int f(h)\,\mathrm{d}\rho(h)$ hold (see Eaton (1989), Section 1.3).

6.1. *Proofs of Theorem 4*, *Proposition 6*, *Proposition 7*.

Here we prove all results in the main text except the main Theorem 2, which is deferred to the next subsection.

PROOF OF THEOREM 4. Let g be a fixed group element of G. Recall from Remark 1 that we may assume that both models are dominated by a left invariant measure ν on $\mathcal{X}$.

Theorem 1 by GHK (its simplest instantiation in their Section 2) implies that

$$\sup_{T_n\ e\text{-stat.}} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n] = \inf_{\mathbf{\Pi}_0} \mathrm{KL}(\mathbf{Q}_g, \mathbf{\Pi}_0^{g'} \mathbf{P}_{g'}), \tag{24}$$

where the infimum is over all distributions $\mathbf{\Pi}_0$ on G. We will show that for any pair $g, h \in G$ and any prior $\mathbf{\Pi}$ on G, there exists a prior $\tilde{\mathbf{\Pi}}$ such that

$$\mathrm{KL}(\mathbf{Q}_g, \mathbf{\Pi}^{g'} \mathbf{P}_{g'}) = \mathrm{KL}(\mathbf{Q}_h, \tilde{\mathbf{\Pi}}^{g'} \mathbf{P}_{g'}). \tag{25}$$

From this, our claim will follow: by symmetry, the previous display implies that $g \mapsto \sup_{T_n\ e\text{-stat.}} \mathbf{E}_g^{\mathbf{Q}}[\ln T_n]$ is constant over G because of its relation to the KL minimization in (24). Let $\bar{p} = \int p_{g'} \, \mathrm{d}\mathbf{\Pi}(g')$, use both the invariance of ν and of $\mathcal{Q}$, and compute

$$\begin{aligned}
\mathrm{KL}(\mathbf{Q}_g, \mathbf{\Pi}^{g'} \mathbf{P}_{g'}) &= \mathbf{E}_g^{\mathbf{Q}}\left[\ln \frac{q_g(X^n)}{\bar{p}(x^n)}\right] = \int q_g(x^n) \ln \frac{q_g(x^n)}{\bar{p}(x^n)} \, \mathrm{d}\nu(x^n) \\
&= \int q_h(hg^{-1}x^n) \ln \frac{q_h(hg^{-1}x^n)}{\bar{p}(x^n)} \, \mathrm{d}\nu(x^n).
\end{aligned}$$

Next, define $\tilde{\mathbf{\Pi}}$ as the probability distribution on G that assigns $\tilde{\mathbf{\Pi}}\{H \in B\} = \mathbf{\Pi}\{H \in gh^{-1}B\}$ for any measurable set $B \subseteq G$. Then

$$\bar{p}(x^n) = \int p_{g'}(x^n) \, \mathrm{d}\mathbf{\Pi}(g') = \int p_{gh^{-1}g'}(x^n) \, \mathrm{d}\tilde{\mathbf{\Pi}}(g') = \int p_{g'}(hg^{-1}x^n) \, \mathrm{d}\tilde{\mathbf{\Pi}}(g').$$

Define $\tilde{p} = \int p_{g'} \, \mathrm{d}\tilde{\mathbf{\Pi}}(g')$. The two last displays together imply that

$$\mathrm{KL}(\mathbf{Q}_g, \mathbf{\Pi}^{g'} \mathbf{P}_{g'}) = \int q_h(hg^{-1}x^n) \ln \frac{q_h(hg^{-1}x^n)}{\tilde{p}(hg^{-1}x^n)} \, \mathrm{d}\nu(x^n).$$

After a change of variable and using the invariance of ν, the right hand side of this equation equals $\mathrm{KL}(\mathbf{Q}_g, \tilde{\mathbf{\Pi}}^{g'} \mathbf{P}_{g'})$. Thus, this last equation is nothing but (25), as was our objective. By our previous discussion, the result follows. $\square$

PROOF OF PROPOSITION 6. Let $g \in G$ be arbitrary but fixed. We start by showing that T^{M_n} equals the likelihood ratio for $M^n = (M_1, \ldots, M_n)$ between $\mathbf{P}_g$ and $\mathbf{Q}_g$. For each $t > 1$, the maximally invariant statistic at $n-1$, $M_{n-1} = m_{n-1}(X^{n-1})$ is invariant if seen as a function of X^n. Hence, by the maximality of m_n, M_{n-1} can be written as a function of M_n. Repeating this reasoning $n-1$ times yields that M_n contains all information about the value of $M^{n-1} = (M_1, \ldots, M_{n-1})$, all the maximally invariant statistics at previous times. Two consequences fall from these observations. First, no additional information about T^{M_n} is gained by knowing the value of $M^{n-1} = (M_1, \ldots, M_{n-1})$ with respect to only knowing M_{n-1}, that is, $\mathbf{E}_g^{\mathbf{P}}[T^{M_n} | M_{n-1}] = \mathbf{E}_g^{\mathbf{P}}[T^{M_n} | M^{n-1}]$. Second, the likelihood ratio between $\mathbf{P}_g$ and $\mathbf{Q}_g$ for the sequence $M_1, \ldots, M_n$ equals the likelihood ratio for M_n alone, that is,

$$T^{M_n} = \frac{q^{M_1,\ldots,M_n}(m_1(X^1), \ldots, m_n(X^n))}{p^{M_1,\ldots,M_n}(m_1(X^1), \ldots, m_n(X^n))}.$$

The previous two consequences, and a computation, together imply that $(T^{M_n})_{n\in\mathbb{N}}$ is an M-martingale under $\mathbf{P}_g$, that is, $\mathbf{E}_g^{\mathbf{P}}[T^{M_n} | M^{n-1}] = T^{M_{n-1}}$. Since $g \in G$ was arbitrary, the result follows. $\square$

PROOF OF PROPOSITION 7. Let $\mathbf{\Pi}_0^{g,\delta}, \mathbf{\Pi}_1^{g,\delta}$ be two probability distributions on $G \times \Delta_0$ and $G \times \Delta_1$, respectively. If we call $\mathbf{\Pi}_0^{\delta}$ and $\mathbf{\Pi}_1^{\delta}$ their respective marginals on Δ_0 and Δ_1, then the information processing inequality implies that

$$\mathrm{KL}(\mathbf{\Pi}_1^{g,\delta} \mathbf{Q}_{g,\delta}, \mathbf{\Pi}_0^{g,\delta} \mathbf{P}_{g,\delta}) \geq \mathrm{KL}(\mathbf{\Pi}_1^{\delta} \mathbf{Q}_{\delta}^{M_n}, \mathbf{\Pi}_0^{\delta} \mathbf{P}_{\delta}^{M_n}) \geq \mathrm{KL}(\mathbf{\Pi}_1^{\star\delta} \mathbf{Q}_{\delta}^{M_n}, \mathbf{\Pi}_0^{\star\delta} \mathbf{P}_{\delta}^{M_n}).$$

This means that the right-most member of the previous display is a lower bound on our target infimum, that is,

$$\inf_{\mathbf{\Pi}_0,\mathbf{\Pi}_1} \mathrm{KL}\big(\mathbf{\Pi}_1^{g,\delta}\mathbf{Q}_{g,\delta}\mathbf{\Pi}_0^{g,\delta}\mathbf{P}_{g,\delta}\big) \geq \mathrm{KL}\big(\mathbf{\Pi}_1^{\star\delta}\mathbf{Q}_\delta^{M_n}, \mathbf{\Pi}_0^{\star\delta}\mathbf{P}_\delta^{M_n}\big). \tag{26}$$

To show that this is indeed an equality, it suffices to prove it when taking the infimum over a smaller subset of probability distributions $\mathbf{\Pi}_0$, $\mathbf{\Pi}_1$. We proceed to build such a subset. Let $\mathcal{P}(\mathbf{\Pi}_0^{\star\delta})$ be the set of probability distributions on $G \times \Delta_0$ with marginal distribution $\mathbf{\Pi}_0^{\star\delta}$. Define analogously the set of probability distributions $\mathcal{P}(\mathbf{\Pi}_1^{\star\delta})$ on $G \times \Delta_1$. By our assumptions, Theorem 2 can be readily used to conclude that

$$\inf_{(\mathbf{\Pi}_0,\mathbf{\Pi}_1)\in\mathcal{P}(\mathbf{\Pi}_0^{\star\delta})\times\mathcal{P}(\mathbf{\Pi}_1^{\star\delta})} \mathrm{KL}\big(\mathbf{\Pi}_1^{g,\delta}\mathbf{Q}_{g,\delta}, \mathbf{\Pi}_0^{g,\delta}\mathbf{P}_{g,\delta}\big) = \mathrm{KL}\big(\mathbf{\Pi}_1^{\star\delta}\mathbf{Q}_\delta^{M_n}, \mathbf{\Pi}_0^{\star\delta}\mathbf{P}_\delta^{M_n}\big) \tag{27}$$

holds; (26) and (27) together imply the result that we were after. □

6.2. *Proof of the main theorem*, *Theorem* 2. For the proof of the main result, we use an equivalent definition of amenability to the one that was already anticipated in Section 2.2. We take the one that suits our purposes best (see Bondar and Milnes (1981), p. 109, Condition A_1). That is, a group G is amenable if there exists an increasing sequence of symmetric compact subsets $C_1 \subseteq C_2 \subseteq \cdots \subset G$ such that, for any compact set $K \subseteq G$,

$$\frac{\rho\{C_i\}}{\rho\{C_i K\}} \to 1 \quad \text{as } i \to \infty.$$

In this formulation, amenability is the existence of *almost invariant* symmetric compact subsets of the group G. We use these sets to build a sequence of *almost invariant* probability measures when G is noncompact.

PROOF OF THEOREM 2. Under our assumptions, Theorem 2 of Bondar (1976) implies the existence of a bimeasurable one-to-one map $r : \mathcal{X}^n \to G \times \mathcal{X}^n/G$ such that $r(x^n) = (h(x^n), m(x^n))$ and $r(gx^n) = (gh(x^n), m(x^n))$ for $h(x^n) \in G$ and $m(x^n) \in \mathcal{X}^n/G$. Hence, by a change of variables, we can take densities with respect to the image measure μ of ν under the map r on $G \times \mathcal{X}^n/G$. Call the random variables $M = m(X^n)$ and $H = h(X^n)$. We can therefore assume, without loss of generality, that the data is of the form (H, M), that the group G acts canonically by multiplication on the first component, and that the measures are with respect to a G-invariant measure $\mu = \lambda \times \beta$ where λ is the left Haar measure on G and β is some measure on $\mathcal{X}^n/G$ (see Remark 1). Note that rewriting the data in this way does not affect our objective because the KL divergence remains unchanged under bijective transformations of the data. For each $g \in G$, write $\mathbf{P}_g^{H|m}$ and $\mathbf{Q}_g^{H|m}$ for the conditional probabilities $\mathbf{P}_g^H\{\,\cdot\,|M = m\}$ and $\mathbf{Q}_g^H\{\,\cdot\,|M = m\}$, which can be obtained through disintegration (see Chang and Pollard (1997)), and write $p_g(\,\cdot\,|m)$ and $q_g(\,\cdot\,|m)$ for their respective conditional densities with respect to the left Haar measure λ.

We turn to our KL minimization objective. The chain rule for the KL divergence implies that, for any probability distribution $\mathbf{\Pi}$ on G,

$$\mathrm{KL}\big(\mathbf{\Pi}^g\mathbf{Q}_g, \mathbf{\Pi}^g\mathbf{P}_g\big) = \mathrm{KL}\big(\mathbf{Q}^M, \mathbf{P}^M\big) + \int \mathrm{KL}\big(\mathbf{\Pi}^g\mathbf{Q}_g^{H|m}, \mathbf{\Pi}^g\mathbf{P}_g^{H|m}\big)\,\mathrm{d}\mathbf{Q}^M(m). \tag{28}$$

In order to prove our claim, we will build a sequence $\{\mathbf{\Pi}_i\}_{i\in\mathbb{N}}$ of probability distributions on G such that the term in (28) pertaining the conditional distributions given M—the second term on the right-hand side—goes to zero, that is, such that

$$\int \mathrm{KL}\big(\mathbf{\Pi}_i^g\mathbf{Q}_g^{H|m}, \mathbf{\Pi}_i^g\mathbf{P}_g^{H|m}\big)\,\mathrm{d}\mathbf{Q}^M(m) \to 0 \quad \text{as } i \to \infty. \tag{29}$$

We define the distributions $\mathbf{\Pi}_i$ as the normalized restriction of the right Haar measure ρ to carefully chosen compact sets $C_i \subset G$, that we describe in brief. In other words, for $B \subseteq G$ measurable, we define $\mathbf{\Pi}_i$ by

$$\mathbf{\Pi}_i\{g \in B\} := \frac{\rho\{B \cap C_i\}}{\rho\{C_i\}}. \tag{30}$$

Next, the choice of compact sets C_i. For technical reasons that will become apparent later, we pick $C_i = J_i K_i L_i$, where J_i, K_i, and L_i are increasing compact symmetric neighborhoods of the unity of G with the growth condition that C_i is not much bigger—measured by ρ—than J_i. More precisely, we choose C_i according to the following lemma.

LEMMA 2. *Under the amenability of G there exist sequences $\{J_i\}_{i\in\mathbb{N}}$, $\{K_i\}_{i\in\mathbb{N}}$ and $\{L_i\}_{i\in\mathbb{N}}$ of compact symmetric neighborhoods of the unity of G, each increasing to cover G, such that*

$$\frac{\rho\{J_i\}}{\rho\{J_i K_i L_i\}} \to 1 \quad \textit{as } i \to \infty.$$

The proof of this lemma is given in Section S4.1 of the Supplementary Material (Pérez-Ortiz et al. (2024)). There is no risk of dividing by ∞ in (30): by the continuity of the group operation each C_i is compact, hence $\rho\{C_i\} < \infty$. Lemma 2 ensures that $\mathbf{\Pi}_i\{g \in J_i\} \to 1$ as $i \to \infty$, a fact that will be useful later in the proof. Write $\mathbf{Q}_i^{H|m} := \mathbf{\Pi}_i^g \mathbf{Q}_g^{H|m}$, and $\mathbf{P}_i^{H|m} := \mathbf{\Pi}_i^g \mathbf{P}_g^{H|m}$, and $q_i(h|m)$ and $p_i(h|m)$ for their respective densities. We use a change of variable and split the integral in our quantity of interest from (29). To this end, notice that for any function $f = f(h, m)$, the expected value $\mathbf{E}_g^{\mathbf{Q}}[f(H, M)] = \mathbf{E}_1^{\mathbf{Q}}[f(gH, M)]$. Indeed,

$$\begin{aligned}\iint f(h, m) q_g(h, m)\, \mathrm{d}\lambda(g)\, \mathrm{d}\beta(m) &= \iint f(h, m) q_1\big(g^{-1}h, m\big)\, \mathrm{d}\lambda(g)\, \mathrm{d}\beta(m) \\ &= \iint f(gh, m) q_1(h, m)\, \mathrm{d}\lambda(g)\, \mathrm{d}\beta(m).\end{aligned}$$

Use this fact to obtain that

$$\begin{aligned}\int \mathrm{KL}\big(\mathbf{\Pi}_i^g \mathbf{Q}_g^{H|m}, \mathbf{\Pi}_i^g \mathbf{P}_g^{H|m}\big)\, \mathrm{d}\mathbf{Q}(m) &= \int \mathbf{E}_1^{\mathbf{Q}}\bigg[\ln \frac{q_i(gH|M)}{p_i(gH|M)}\bigg]\, \mathrm{d}\mathbf{\Pi}_i(g) \\ &= \underbrace{\int \mathbf{E}_1^{\mathbf{Q}}\bigg[\mathbb{1}\{gH \in J_i K_i\} \ln \frac{q_i(gH|M)}{p_i(gH|M)}\bigg]\, \mathrm{d}\mathbf{\Pi}_i(g)}_{\mathrm{A}} \\ &\quad + \underbrace{\int \mathbf{E}_1^{\mathbf{Q}}\bigg[\mathbb{1}\{gH \notin J_i K_i\} \ln \frac{q_i(gH|M)}{p_i(gH|M)}\bigg]\, \mathrm{d}\mathbf{\Pi}_i(g)}_{\mathrm{B}}.\end{aligned} \tag{31}$$

We separate the rest of the proof in two steps, one for bounding each term in (31). These steps use two technical lemmas that we prove in Section S4.1 of the Supplementary Material (Pérez-Ortiz et al. (2024)).

Bound for A in (31): Recall that

$$\ln \frac{q_i(gh|m)}{p_i(gh|m)} = \ln \frac{\int \mathbb{1}\{g' \in J_i K_i L_i\} q_{g'}(gh|m)\, \mathrm{d}\rho(g')}{\int \mathbb{1}\{g' \in J_i K_i L_i\} p_{g'}(gh|m)\, \mathrm{d}\rho(g')}.$$

Use $N = J_i K_i$—not necessarily symmetric—and $L = L_i$ in the following lemma.

LEMMA 3. *Let N and L be compact subsets of G. Assume that L is symmetric. Then for each $m \in \mathcal{X}^n/G$ it holds that*

$$\sup_{h' \in N} \left\{ \ln \frac{\int \mathbb{1}\{g \in NL\}\, q_g(h'|m)\, \mathrm{d}\rho(g)}{\int \mathbb{1}\{g \in NL\}\, p_g(h'|m)\, \mathrm{d}\rho(g)} \right\} \le -\ln \mathbf{P}_1\{H \in L \mid M = m\}.$$

With this lemma at hand, conclude that, for all $gh \in J_i K_i$, and $m \in \mathcal{M}$

$$\ln \frac{q_i(gh|m)}{p_i(gh|m)} \le -\ln \mathbf{P}_1\{H \in L_i \mid M = m\}.$$

At the same time this implies that A in (31) is smaller than

$$-\int \ln \mathbf{P}_1\{H \in L_i \mid M = m\}\, \mathrm{d}\mathbf{Q}(m).$$

Since the sets L_i were chosen to satisfy $L_i \uparrow G$, the probability $\mathbf{P}_1\{H \in L_i \mid M = m\} \to 1$ monotonically for each value of m. Consequently, the quantity in last display tends to 0 by the monotone convergence theorem, and so does A in (31). This ends the first step of the proof. Now, we turn to the second term in (31).

Bound for B in (31): Our strategy at this point is to show that, as $i \to \infty$,

$$\int \mathbf{Q}_1\{gH \notin J_i K_i\}\, \mathrm{d}\mathbf{\Pi}_i(g) \to 0, \tag{32}$$

and to use (14) to show our goal, that B in (31) tends to zero. To show (32), notice that if $g \in J_i$ and $h \in K_i$, then $gh \in J_i K_i$, which implies that

$$\int \mathbf{Q}_1\{gH \in J_i K_i\}\, \mathrm{d}\mathbf{\Pi}_i(g) \ge \mathbf{\Pi}_i\{g \in J_i\} \mathbf{Q}_1\{H \in K_i\}.$$

Since the sets K_i increase to cover G, we have $\mathbf{Q}_1\{H \in K_i\} \to 1$ as $i \to \infty$, and by our initial choice of sets J_i, K_i, L_i, the probability $\mathbf{\Pi}_i\{g \in J_i\} \to 1$, as $i \to \infty$. Hence, (32) holds. To bound the second term, we use the following lemma with $\mathbf{\Pi} = \mathbf{\Pi}_i$.

LEMMA 4. *Let $\mathbf{\Pi}$ be a distribution on G. Then, for each $h \in G$ and $m \in \mathcal{X}^n/G$, setting $\mathrm{d}\mathbf{\Pi}(g|h,m) = \frac{q_g(h|m)\, \mathrm{d}\mathbf{\Pi}(g)}{\int q_g(h|m)\, \mathrm{d}\mathbf{\Pi}(g)}$, it holds that*

$$\ln \frac{\int q_g(h|m)\, \mathrm{d}\mathbf{\Pi}(g)}{\int p_g(h|m)\, \mathrm{d}\mathbf{\Pi}(g)} \le \int \ln \frac{q_g(h|m)}{p_g(h|m)}\, \mathrm{d}\mathbf{\Pi}(g|h,m).$$

After invoking the previous lemma, apply Hölder's and Jensen's inequality consecutively to bound B in (31) by

$$\begin{aligned}
&\iint \left[\mathbb{1}\{gh \notin J_i K_i\} \int \ell\,(gh|m)\, \mathrm{d}\mathbf{\Pi}_i(g'|h,m) \right] \mathrm{d}\mathbf{Q}_1(h,m)\, \mathrm{d}\mathbf{\Pi}_i(g) \\
&\quad \le \underbrace{\left(\int \mathbf{Q}_1\{gH \notin J_i K_i\}\, \mathrm{d}\mathbf{\Pi}_i(g) \right)^{1/q}}_{\to 0 \text{ as } i \to \infty \text{ by } (32)} \\
&\qquad \times \left(\iint \left| \int \ell(gh|m)\, \mathrm{d}\mathbf{\Pi}_i(g'|h,m) \right|^p \mathrm{d}\mathbf{Q}_1(h,m)\, \mathrm{d}\mathbf{\Pi}_i(g) \right)^{1/p},
\end{aligned} \tag{33}$$

where here and in the sequel, $\ell\,(gh|m)$ abbreviates $\ln \frac{q_{g'}(gh|m)}{p_{g'}(gh|m)}$, and $p = 1 + \varepsilon$ and q is p's Hölder conjugate, that is, $1/p + 1/q = 1$. Next, we show that the second factor on the right of (33) remains bounded as $i \to \infty$. By Jensen's inequality, this quantity is smaller than

$$\left(\iiint |\, \ell(gh|m)|^p\, \mathrm{d}\mathbf{\Pi}_i(g'|h,m)\, \mathrm{d}\mathbf{Q}_1(h,m)\, \mathrm{d}\mathbf{\Pi}_i(g) \right)^{1/p}.$$

After a series of rewritings and using our Assumption (14), we will show that this quantity is bounded. First, we deduce that

$$\begin{aligned}
&\iint |\,\ell(gh|m)|^p \,\mathrm{d}\boldsymbol{\Pi}_i(g'|h,m)\,\mathrm{d}\mathbf{Q}_1(h,m)\,\mathrm{d}\boldsymbol{\Pi}_i(g) \\
&\quad = \iint |\,\ell(h|m)|^p \,\mathrm{d}\boldsymbol{\Pi}_i(g'|h,m)\,\mathrm{d}\mathbf{Q}_g(h,m)\,\mathrm{d}\boldsymbol{\Pi}_i(g) \\
&\quad = \iint |\,\ell(h|m)|^p \,\mathrm{d}\boldsymbol{\Pi}_i(g'|h,m)\,\mathrm{d}\mathbf{Q}_i(h,m) = \mathbf{E}_1^{\mathbf{Q}}\left|\ln \frac{q_1(H|M)}{p_1(H|M)}\right|^p,
\end{aligned}$$

where we used again the change of variable that we used to obtain (31)—but now in the opposite direction—and in the final equality, we used Bayes' theorem.

Hence, as

$$\begin{aligned}
\left(\mathbf{E}_1^{\mathbf{Q}}\left[\left|\ln \frac{q_1(H|M)}{p_1(H|M)}\right|^p\right]\right)^{1/p} &\le \left(\mathbf{E}_1^{\mathbf{Q}}\left[\left|\ln \frac{q_1(H,M)}{p_1(H,M)}\right|^p\right]\right)^{1/p} + \left(\mathbf{E}_1^{\mathbf{Q}}\left[\left|\ln \frac{q_1(M)}{p_1(M)}\right|\right]^p\right)^{1/p} \\
&< \infty
\end{aligned}$$

by (14), we have shown that (33) tends to 0 as $i \to \infty$ and that consequently B in (31) tends to 0 in the same limit.

After completing these two steps, we have shown that both A and B in (31) tend to 0 as $i \to \infty$, and that consequently the claim of the theorem follows. All that is left is to prove Lemmas 2, 3, and 4. The proofs being straightforward but tedious, we delegated these to Section S4 of the Supplementary Material (Pérez-Ortiz et al. (2024)). $\square$

Acknowledgements. We thank Wouter Koolen for useful conversations and for inspiring the example in Section S3.1 of the Supplementary Material (Pérez-Ortiz et al. (2024)).

Peter D. Grünwald is also affiliated with the Mathematical Institute of Leiden University.

Funding. This work is part of the research program with project number 617.001.651, which is financed by the Dutch Research Council (NWO).

SUPPLEMENTARY MATERIAL

Online supplement (DOI: 10.1214/24-AOS2394SUPP; .pdf). The online supplement (Pérez-Ortiz et al. (2024)) contains all the proofs that were omitted from the main text, as well as auxiliary results and discussions.

REFERENCES

ANDERSSON, S. (1982). Distributions of maximal invariants using quotient measures. *Ann. Statist.* **10** 955–961. MR0663446

BERGER, J. O., PERICCHI, L. R. and VARSHAVSKY, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhyā Ser. A* **60** 307–321. MR1718789

BERGER, J. O. and SUN, D. (2008). Objective priors for the bivariate normal model. *Ann. Statist.* **36** 963–982. MR2396821 https://doi.org/10.1214/07-AOS501

BHOWMIK, J. L. and KING, M. L. (2007). Maximal invariant likelihood based testing of semi-linear models. *Statist. Papers* **48** 357–383. MR2395928 https://doi.org/10.1007/s00362-006-0342-7

BONDAR, J. V. (1976). Borel cross-sections and maximal invariants. *Ann. Statist.* **4** 866–877. MR0474589

BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. Verw. Gebiete* **57** 103–128. MR0623457 https://doi.org/10.1007/BF00533716

BOURBAKI, N. (2004). *Integration. II. Chapters* 7–9. *Elements of Mathematics* (*Berlin*). Springer, Berlin. Translated from the 1963 and 1969 French originals by Sterling K. Berberian. MR2098271

CHANG, J. T. and POLLARD, D. (1997). Conditioning as disintegration. *Stat. Neerl.* **51** 287–317. MR1484954 https://doi.org/10.1111/1467-9574.00056

COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory. Wiley Series in Telecommunications and Signal Processing*. Wiley, New York.

COX, D. R. (1952). Sequential tests for composite hypotheses. *Proc. Camb. Philos. Soc.* **48** 290–299. MR0047292 https://doi.org/10.1017/s030500410002764x

DARLING, D. A. and ROBBINS, H. (1968). Some nonparametric sequential tests with power one. *Proc. Natl. Acad. Sci. USA* **61** 804–809. MR0238437 https://doi.org/10.1073/pnas.61.3.804

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* **35** 189–233. MR0365805

EATON, M. L. (1989). *Group Invariance Applications in Statistics. NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. IMS, Hayward. MR1089423

EATON, M. L. and SUDDERTH, W. D. (1999). Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli* **5** 833–854. MR1715441 https://doi.org/10.2307/3318446

EATON, M. L. and SUDDERTH, W. D. (2002). Group invariant inference and right Haar measure. *J. Statist. Plann. Inference* **103** 87–99. MR1896985 https://doi.org/10.1016/S0378-3758(01)00199-9

GIRI, N., KIEFER, J. and STEIN, C. (1963). Minimax character of Hotelling's T^2 test in the simplest case. *Ann. Math. Stat.* **34** 1524–1535. MR0156408 https://doi.org/10.1214/aoms/1177703884

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. (2023). Safe testing. Available at: arXiv:1906.07801 [cs, math, stat]. First version on arXiv 2019. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear.

GRÜNWALD, P. D. (2023). The e-posterior. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **381** Paper No. 20220146, 21. MR4590499

HALL, W. J., WIJSMAN, R. A. and GHOSH, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Stat.* **36** 575–614. MR0178552 https://doi.org/10.1214/aoms/1177700169

HENZI, A., PUKE, M., DIMITRIADIS, T. and ZIEGEL, J. (2024). A safe Hosmer–Lemeshow test. *New Engl. J. Statist. Data Sci.* **2** 175–189.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford. MR0187257

KARIYA, T. (1980). Locally robust tests for serial correlation in least squares regression. *Ann. Statist.* **8** 1065–1070. MR0585704

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

LAI, T. L. (1976). On confidence sequences. *Ann. Statist.* **4** 265–280. MR0395103

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LIANG, F. and BARRON, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inf. Theory* **50** 2708–2726. MR2096988 https://doi.org/10.1109/TIT.2004.836922

PATERSON, A. L. T. (1988). *Amenability. Mathematical Surveys and Monographs* **29**. Amer. Math. Soc., Providence, RI. MR0961261 https://doi.org/10.1090/surv/029

PÉREZ-ORTIZ, M. F., LARDY, T., DE HEIDE, R. and GRÜNWALD, P. D. (2024). Supplement to "E-statistics, group invariance and anytime-valid testing." https://doi.org/10.1214/24-AOS2394SUPP

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

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

REITER, H. and STEGEMAN, J. D. (2000). *Classical Harmonic Analysis and Locally Compact Groups*, 2nd ed. *London Mathematical Society Monographs. New Series* **22**. Clarendon, New York. MR1802924

REN, Z. and BARBER, R. F. (2024). Derandomised knockoffs: Leveraging e-values for false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 122–154. MR4754084 https://doi.org/10.1093/jrsssb/qkad085

ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Stat.* **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

ROUDER, J. N., SPECKMAN, P. L., SUN, D., MOREY, R. D. and IVERSON, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16** 225–237. https://doi.org/10.3758/PBR.16.2.225

ROY, S. N. and BARGMANN, R. E. (1958). Tests of multiple independence and the associated confidence bounds. *Ann. Math. Stat.* **29** 491–503. MR0093862 https://doi.org/10.1214/aoms/1177706624

RUSHTON, S. (1950). On a sequential t-test. *Biometrika* **37** 326–333. MR0044080 https://doi.org/10.1093/biomet/37.3-4.326

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

SHALAEVSKII, O. V. (1971). Minimax character of Hotelling's T2 test. I. In *Investigations in Classical Problems of Probability Theory and Mathematical Statistics*: *Part I* (V. M. Kalinin and O. V. Shalaevskii, eds.) 74–101. Springer, Boston. https://doi.org/10.1007/978-1-4684-8211-9_2

SUBBAIAH, P. and MUDHOLKAR, G. S. (1978). A comparison of two tests for the significance of a mean vector. *J. Amer. Statist. Assoc.* **73** 414–418. MR0497422

SUN, D. and BERGER, J. O. (2007). Objective Bayesian analysis for the multivariate normal model. In *Bayesian Statistics* 8. *Oxford Sci. Publ.* 525–562. Oxford Univ. Press, Oxford. MR2433206

TURNER, R. J., LY, A. and GRÜNWALD, P. D. (2024). Generic E-variables for exact sequential k-sample tests that allow for optional stopping. *J. Statist. Plann. Inference* **230** Paper No. 106116, 15. MR4661178 https://doi.org/10.1016/j.jspi.2023.106116

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

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 https://doi.org/10.1214/aoms/1177731118

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

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

ZHANG, Z., RAMDAS, A. and WANG, R. (2023). When do exact and powerful p-values and e-values exist? arXiv preprint. Available at: arXiv:2305.16539.