Statistics and Probability Letters 229 (2026) 110574

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Supermartingales for one-sided tests: Sufficient monotone likelihood ratios are sufficient ☆

Peter Grünwald [a,b], Wouter M. Koolen [a,c,*]

[a] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[b] Leiden University, Leiden, The Netherlands
[c] Twente University, Enschede, The Netherlands

ARTICLE INFO

Keywords:
Sequential t-test
Supermartingales
e-variables
Group invariance
Monotone likelihood ratio

ABSTRACT

The t-statistic is a widely-used scale-invariant statistic for testing the null hypothesis that the mean is zero. Martingale methods enable sequential testing with the t-statistic at every sample size, while controlling the probability of falsely rejecting the null. For one-sided sequential tests, which reject when the t-statistic is too positive, a natural question is whether they also control false rejection when the true mean is negative. We prove that this is the case using monotone likelihood ratios and sufficient statistics. We develop applications to the scale-invariant t-test, the location-invariant χ^2-test and sequential linear regression with nuisance covariates.

1. The problem

Over the last few years, it has become abundantly clear that e-variables are of fundamental importance for multiple testing (Goeman et al., 2025; Ignatiadis et al., 2024). The latter paper employed the standard *t-test e-variable* (Pérez-Ortiz et al., 2024) in a multiple testing context. Moreover, recently researchers have started to explore e-processes and their special cases, test martingales, for anytime-valid testing in combination with multiple testing as well (Fischer et al., 2024, 2025), and we expect that the *t-test martingale*, an extension of the t-test variable that is a standard tool in anytime-valid testing, will be used in such anytime-valid-multiple testing combinations as well. Surprisingly, it has not been known whether the t-test martingale for testing a zero mean is still suitable for one-sided testing when the mean can be negative (Wang and Ramdas, 2025). In this paper we develop a generic argument based on sufficiency showing that it is suitable. The argument turns out to apply to many other cases of interest as well, including the location-invariant χ^2-test and sequential linear regression with nuisance covariates. From a practical stance (see Section 3), this means that in all such settings, we may employ test martingales that have strong (growth-) optimality properties.

The set-up. We start by reviewing the setting of the one-sided sequential testing problem. Let $U_1, U_2, \ldots$ be a stochastic process. Each U_i is a random variable with support $\mathcal{U}_i \subset \mathbb{R}$ and we abbreviate the sequence of outcomes by $U^n = (U_1, \ldots, U_n)$ and its domain by $\mathcal{U}^{(n)} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_n$. The underlying filtration is $(\sigma(U^n))_{n \in \mathbb{N}}$. Let $\{P_\delta : \delta \in \Delta\}$ be a 1-parameter statistical model for such a process, where the parameter domain $\Delta \subseteq \mathbb{R}$ is an interval. We restrict attention to models in which we are allowed to condition just as in elementary probability. While we make this assumption precise later, for now we note that it implies the following: for all i, $\delta \in \Delta$, $u^{i-1} \in \mathcal{U}^{i-1}$, the conditional distributions $P_\delta(U_i = \cdot \mid U^{i-1} = u^{i-1})$ have a density $f_\delta^{U_i}(U_i \mid u^{i-1})$ relative to a common underlying measure ν_i on $\mathcal{U}_i$; in all our examples this will be either counting or Lebesgue measure. This further allows us to fix any element $\delta_0 \in \Delta$ and re-represent the conditional distributions $P_\delta^{U_i} \mid U^{i-1} := P_\delta(U_i = \cdot \mid U^{i-1})$ by their conditional densities $p_\delta^{U_i}(U_i \mid U^{i-1})$

☆ This article is part of a Special issue entitled: 'E-values and Multiple Testing' published in Statistics and Probability Letters.

* Corresponding author at: Centrum Wiskunde & Informatica, Amsterdam, The Netherlands.
E-mail address: wmkoolen@cwi.nl (W.M. Koolen).

https://doi.org/10.1016/j.spl.2025.110574

Available online 24 October 2025

relative to $P_{\delta_0}^{U_i} \mid U^{i-1}$. Here, following Pérez-Ortiz et al. (2024), we denote the distribution and density of a measurable random quantity V under P_δ by P_δ^V and p_δ^V, respectively. We shall reserve the symbol p_δ for the density of P_δ relative to P_{δ_0} — these may also be thought of as likelihood ratios. When using densities relative to the underlying measures ν_i we use the symbol f_δ instead.

Simple null. We first consider testing a point null $\mathcal{H}_{\delta=\delta_0} = \{P_{\delta_0}\}$ against an alternative P_W, where W is an arbitrary prior measure on Δ and P_W has density $p_W^{U^n}(U^n) := \int p_\delta^{U^n}(U^n) dW(\delta)$. We allow W to be degenerate, i.e. put all its mass on a single point. We may think of $(p_W^{U^n})_{n\in\mathbb{N}}$ as a *likelihood ratio process*. As is well known, this likelihood ratio process is a *test martingale* (non-negative and starting from 1) relative to null $\mathcal{H}_{\delta=\delta_0} = \{P_{\delta_0}\}$ (Ramdas et al., 2023). Therefore, it handles optional stopping and anytime-validity. Specifically, under any stopping time τ, the test which rejects $\mathcal{H}_{\delta=\delta_0}$ if $p_W^{U^\tau}(U^\tau) \geq \alpha^{-1}$ has Type-I error bounded by α. The (super)-martingale property requires that for all n, all $u^{n-1} \in \mathcal{U}^{(n-1)}$, the conditional one-step likelihood ratios $p_W^{U_n}(U_n \mid u^{n-1})$ are *past-conditional e-variables*, i.e.

$$\mathbf{E}_{\delta_0}\left[p_W^{U_n}(U_n \mid u^{n-1}) \middle| u^{n-1}\right] \leq 1. \tag{1}$$

This, in turn, readily follows from the standard cancellation argument:

$$\mathbf{E}_{\delta_0}\left[p_W^{U_n}(U_n \mid u^{n-1}) \middle| u^{n-1}\right] = \int_u \frac{dP_W^{U_n}|u^{n-1}}{dP_{\delta_0}^{U_n}|u^{n-1}}(u) d\left(P_{\delta_0}^{U_n}|u^{n-1}\right) = 1. \tag{2}$$

One-sided tests. Now fix $\delta^+ > \delta_0$ and consider the one-sided null hypothesis

$$\mathcal{H}_{\delta\leq\delta_0} := \{P_\delta : \delta \in \Delta, \delta \leq \delta_0\}.$$

As alternative we may take either the Bayesian point alternative $\mathcal{H}_1 = \{P_W\}$ with W a prior on $\{\delta \in \Delta : \delta \geq \delta^+\}$ or a composite alternative $\mathcal{H}_1 \subseteq \{P_\delta : \delta \in \Delta, \delta \geq \delta^+\}$. For simplicity we concentrate on the case with $\mathcal{H}_1 = \{P_{\delta^+}\}$ for now, returning to the general case in Section 3.

The question we aim to answer is: *does the likelihood ratio process $(p_{\delta^+}^{U^n})_{n\in\mathbb{N}}$ still constitute a test supermartingale for the enlarged null hypothesis $\mathcal{H}_{\delta\leq\delta_0}$?* In other words, does (1) still hold with the expectation taken over P_δ instead of P_{δ_0} (while the δ_0 hidden by the density notation p_W remains fixed)?

In applications, the parameter δ often represents a notion of effect size and one would then expect that, at least under some further conditions, (1) still holds in this case — indeed, for some special cases such as the test supermartingales appearing in Turner et al. (2024) ($k \times 2$ tables) and ter Schure et al. (2024) (logrank test), the test supermartingale property (1) was shown to hold. *But how general is this phenomenon?* Below we first indicate why existing results do not directly tell us either way; Our main result, Lemma 3, provides a condition under which the supermartingale property does hold. To state it and earlier results, we need the following property.

Definition 1 (*Monotone Likelihood Ratio Property*). Let T be a random variable with common support $\mathcal{T}$ under all P_δ with $\delta \in \Delta$. We say that T satisfies the *Monotone Likelihood Ratio (MLR) Property* if for all $\delta_0, \delta^+ \in \Delta$ with $\delta_0 \leq \delta^+$, the likelihood ratio $p_{\delta^+}^T(t)$ is increasing[1] as a function of $t \in \mathcal{T}$.

(Grünwald et al., 2024, Proposition 3) implies the following:

Proposition 2. *Suppose that T satisfies the MLR Property and fix $\delta \leq \delta_0 \leq \delta^+$ with $\delta, \delta_0, \delta^+ \in \Delta$. Then $\mathbf{E}_\delta[p_{\delta^+}^T(T)] \leq 1$, i.e. the likelihood ratio $p_{\delta^+}^T(T)$ is an e-variable for $\mathcal{H}_{\delta\leq\delta_0}$.*

Appendix A includes a direct proof, building on Brown et al. (1981), Lehmann and Romano (1986). This proposition can and has been used to show that some likelihood ratios $p_{\delta^+}^{T_n}(T_n)$, at fixed sample size n, of statistic $T_n = t_n(U^n)$ set to some fixed function of U^n, provide e-*variables*. Yet, it tells us nothing about whether the sequence $(p_{\delta^+}^{T_n}(T_n))_{n\in\mathbb{N}}$ is a supermartingale.

Example 1 (*t-test*). Let us illustrate this using the standard anytime-valid t-test as presented by both (Grünwald et al., 2024; Pérez-Ortiz et al., 2024). Here $\Delta = \mathbb{R}$ and $P_\delta^{Y^\infty}$ expresses that the data $Y_1, Y_2, \ldots$ are i.i.d. $\sim N(\mu, \sigma)$ for some (μ, σ) such that $\delta = \mu/\sigma$. Thus δ is not sufficient to determine the distribution of the data $Y_1, Y_2 \ldots$ Yet the sequence of *maximal invariants* (Pérez-Ortiz et al., 2024) $U_1, U_2, \ldots$ with $U_i := \frac{Y_i}{|Y_1|}$, has the same distribution $P_\delta^{U^\infty}$ under all $N(\mu, \sigma)$ with given effect size $\mu/\sigma = \delta$. It is invariant to rescaling, e.g. changing the unit of measurement of the Y_i (note that the support $\mathcal{U}_1 = \{\pm 1\}$ is discrete whereas $\mathcal{U}_i = \mathbb{R}$ for $i > 1$). The standard anytime-valid t-test is defined in terms of the likelihood ratio process $(p_{\delta^+}^{U^n})_{n\in\mathbb{N}}$ with P_δ as above. By the argument (2) above, it provides a test-martingale in the filtration $(\sigma(U^n))_{n\in\mathbb{N}}$ relative to the null $\mathcal{H}_{\delta=\delta_0}$. Yet, as Wang and Ramdas (2025, Section 4.5) write for the case $\delta_0 = 0$ (notation adapted):

> .... Pérez-Ortiz et al. (2024, Corollary 8) show that at any fixed sample size n the scale- invariant t-likelihood ratio $p_{\delta^+}^{U^n}$ is actually an e-value for the larger one-sided null $\mathcal{H}_{\delta\leq 0}$ if $\delta^+ \geq 0$ [...] meaning that one can non-sequentially test that null using this statistic. However, it is unclear if this t-likelihood ratio process is an e-process for $\mathcal{H}_{\delta\leq 0}$. *We leave this question open for future investigation.*

[1] throughout, increasing will be used in the non-strict sense.

...an observation which prompted nervousness among some members of the e-community, who had simply assumed the t-likelihood ratio to be a super-martingale for $\mathcal{H}_{\delta \leq \delta_0}$, and now discovered that this was not straightforward to prove.

The difficulty. At first sight, the problem might seem easy to solve: if we could show that for every sample size n and past $u^{n-1} \in \mathcal{U}^{(n-1)}$, the *conditional likelihood ratio*

$$p_{\delta^+}^{U_n}(u \mid u^{n-1}) \tag{3}$$

is monotone in the next outcome u, then we could still apply Proposition 2 above pointwise, for each n and conditional on each u^{n-1}, and we would be done. This is how (Turner et al., 2024; ter Schure et al., 2024) proceed: in their settings, the U_i are (in essence) i.i.d. making it easy to employ Proposition 2 pointwise.

The t-test setting is more complicated, as the maximal invariants U_i are not i.i.d. One might be tempted to exploit that the original, uncoarsened data Y_i are i.i.d., but unfortunately the t-likelihood ratio simply is not an e-process relative to the corresponding filtration $(\sigma(Y^n))_{n \in \mathbb{N}}$, leading to trouble when optional stopping is desired. This subtle point is discussed in detail by Pérez-Ortiz et al. (2024). It reflects the more general fact that in group-invariant testing, of which the t-test is merely a very special case, *even when the raw data are i.i.d. we often need to deal with U_i that are not i.i.d.*. This is further illustrated by all examples in Appendix B. One might then directly set to prove that the conditional likelihood ratio (3) is monotone in U_n. However, Wang (2024) showed via numerical experiments that for the t-test it is *not*.

So what to do? In Theorem 4 below we show the following: if for all n, there exists a *sufficient statistic* $T_n = t_n(U^n)$ for the model $P_\delta^{U^n}$ such that the MLR Property holds for T_n, then $(p_{\delta^+}^{U^n})_{n \in \mathbb{N}}$ is a test supermartingale relative to $\mathcal{H}_{\delta \leq \delta_0}$ after all. We will see that within the t-test setting, the t-statistic provides just such a sufficient statistic. Thus, our result resolves the issue for the t-test likelihood ratio but also, as we will show in Appendix B, for several other cases of interest.

2. The solution

Let T_n be a sufficient statistic for model $\{P_\delta : \delta \in \Delta\}$ and data U^n. Recall that this means that, with $f_\delta^{T_n}$ and $f_\delta^{U^n}$ the densities of $P_\delta^{T_n}$ and $P_\delta^{U^n}$ relative to some background measures ν^{T_n} and ν^{U^n}, there exist functions $t_n : \mathcal{U}^{(n)} \to \mathbb{R}$ and $q_n : \mathcal{U}^{(n)} \to \mathbb{R}^+$ such that $T_n = t_n(U^n)$ and for all $\delta \in \Delta$, all $u^n \in \mathcal{U}^{(n)}$, we have

$$f_\delta^{U^n}(u^n) \; = \; f_\delta^{T_n}(t_n(u^n)) \cdot q_n(u^n). \tag{4}$$

To state our main results we first need to make precise the assumption that we ‘can condition as in elementary probability’: formally, we call measurable random quantities (G_1, G_2), *elementary* if there exist sets $\mathcal{G}_1, \mathcal{G}_2$ and measures ν_1 on $\mathcal{G}_1$ and ν_2 on $\mathcal{G}_2$ respectively, such that, for all $\delta \in \Delta$, there exist joint densities $f_\delta^{(G_1,G_2)}$ relative to $\nu_1 \times \nu_2$ and versions of the conditional distributions $P_\delta^{G_1} \mid G_2$ and $P_\delta^{G_2} \mid G_1$ with densities $f_\delta^{G_1|G_2}(g_1 \mid g_2)$ (relative to ν_1) and $f_\delta^{G_2|G_1}(g_2 \mid g_1)$ (relative to ν_2) such that for all $g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2$, we have

$$f_\delta^{G_1|G_2}(g_1 \mid g_2) \cdot f_\delta^{G_2}(g_2) = f_\delta^{G_2|G_1}(g_2 \mid g_1) \cdot f_\delta^{G_1}(g_1) = f_\delta^{(G_1,G_2)}(g_1, g_2).$$

This assumption will automatically hold if for all δ, it holds that $f_\delta^{G_1,G_2}$ exists relative to a $\nu_1 \times \nu_2$ such that ν_1 is either Lebesgue or counting measure; ν_2 is either Lebesgue or counting measure; and $f_\delta^{G_1,G_2}(g_1, g_2) > 0$ for all $(g_1, g_2) \in \mathcal{G}_1 \times \mathcal{G}_2$ — under these conditions, we can take the $f^{G_1|G_2}, f^{G_2|G_1}$ to be ‘elementary’ conditional densities (Williams, 1991, Section 9.6) implying that (G_1, G_2) are elementary in the above sense. This will be the case in all our examples, for all G_1, G_2 for which we will require the assumption. We henceforth write f^{G_i} instead of $f^{G_i|G_j}$, the conditioning random variable always being clear from context. With that, we are ready to state our main results, both proved further below:

Lemma 3. *Assume that for all n, the pair (U_n, U^{n-1}) is elementary and let T_n be a sufficient statistic such that (T_n, U^{n-1}) is elementary and such that the MLR Property (Definition 1) holds with T set to T_n. Then for any $\delta_0, \delta^+ \in \Delta$ with $\delta_0 < \delta^+$, and all $n \geq 1$, for all $u^n \in \mathcal{U}^{(n)}$, we have*

$$p_{\delta^+}^{U_n}(u_n | u^{n-1}) \; = \; p_{\delta^+}^{T_n}(t_n(u^n) | u^{n-1}).$$

and the conditional likelihood ratio $p_{\delta^+}^{T_n}(t | u^{n-1})$ is increasing in t.

Importantly, the likelihood ratio $p_{\delta^+}^{U_n}(u \mid u^{n-1})$ need *not* be monotone in u and indeed it is not in the t-test setting; this caused the perceived difficulty of the problem. Combining the lemma with Proposition 2 gives:

Theorem 4. *Under the assumptions of Lemma 3, the process $(\prod_{i=1}^n p_\delta^{T_i}(T_i \mid U^{i-1}))_{n \in \mathbb{N}}$ is identical to the likelihood ratio process $(p_\delta^{U^n})_{n \in \mathbb{N}}$ and both are test supermartingales relative to $\mathcal{H}_{\delta \leq \delta_0}$.*

Example 2 (*t-test, Continued*)**.** Here are three well-known facts: (a) if data $Y_1, Y_2, \ldots$ are sampled i.i.d. from a normal $N(\mu, \sigma)$ with effect size $\delta = \mu/\sigma$, then the t-statistic

$$T_n \; = \; t_n(U^n) \; = \; \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i}{\sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n U_i^2 - \frac{1}{n} (\sum_{i=1}^n U_i)^2 \right)}} \qquad \text{with} \qquad U_i \; = \; \frac{Y_i}{|Y_1|}. \tag{5}$$

at sample size n has a noncentral t-distribution with $\nu := n - 1$ degrees of freedom and noncentrality parameter $\lambda := \sqrt{n}\delta$. Let us denote the (Lebesgue) density of this distribution by $f_{\mathrm{T}(\nu,\lambda)}$. (b) Also, as easily established, for each n, the t-statistic T_n is sufficient for the coarsening $(U_1, \ldots, U_n)$, see also Pérez-Ortiz et al. (2024, Example 1). That is, for each n, as an instance of (4), we can write (with $\lambda_0 = \sqrt{n}\delta_0$ and $\lambda^+ = \sqrt{n}\delta^+$):

$$p_{\delta^+}^{U^n}(u^n) \;=\; \frac{f_{\mathrm{T}(\nu,\lambda^+)}(t_n(u^n))}{f_{\mathrm{T}(\nu,\lambda_0)}(t_n(u^n))}. \tag{6}$$

Finally, (c) for fixed ν, the family of noncentral t-distributions has the MLR property (Kruskal, 1954). That is, the ratio $f_{\mathrm{T}(\nu,\lambda^+)}(t)/f_{\mathrm{T}(\nu,\lambda_0)}(t)$ is increasing in t when $\delta^+ \geq \delta_0$. Taken together, facts (a)–(c) imply via Lemma 3 and Theorem 4 that the likelihood ratio process $(p_{\delta^+}^{U^n}(U^n))_{n\in\mathbb{N}}$ is a test supermartingale relative to the one-sided composite null $\mathcal{H}_{\delta\leq\delta_0}$, answering (Wang and Ramdas, 2025)'s question in the affirmative.

A collection of examples, including the χ^2-test, linear regression and label agnostic Bernoulli are presented in Appendix B. Our linear regression treatment extends Bhowmik and King (2007), Lindon et al. (2024). We now proceed with the proofs of Lemma 3 and Theorem 4. The proofs are embarrassingly simple once one realizes that to establish the supermartingale property one should consider the conditional density of T_n rather than U_n — *that* realization is the real contribution of this note.

Proof of Lemma 3. Let us fix $u^{n-1} \in \mathcal{U}^{(n-1)}$ throughout the argument. By sufficiency (4), the likelihood ratio of $u_n \in \mathcal{U}_n$ conditioned on u^{n-1} is

$$p_{\delta_+}^{U^n}(u_n|u^{n-1}) \;=\; \frac{p_{\delta_+}^{T_n}(t_n(u^n))}{p_{\delta_+}^{U^{n-1}}(u^{n-1})}, \tag{7}$$

and given u^{n-1} this is a function of u_n only through $t_n(u^n)$. For any possible $t \in \{t_n(u^{n-1}, u_n) \mid u_n \in \mathcal{U}_n\}$ the conditional likelihood ratio, using (7), equals

$$p_{\delta_+}^{T_n}(t|u^{n-1}) \;=\; \frac{p_{\delta_+}^{(T_n,U^{n-1})}(t, u^{n-1})}{p_{\delta_+}^{U^{n-1}}(u^{n-1})} \;=\; \frac{p_{\delta_+}^{U^{n-1}}(u^{n-1} \mid t)\cdot p_{\delta_+}^{T_n}(t)}{p_{\delta_+}^{U^{n-1}}(u^{n-1})} \;\overset{(*)}{=}\; \frac{p_{\delta_+}^{T_n}(t)}{p_{\delta_+}^{U^{n-1}}(u^{n-1})}. \tag{8}$$

Here we used the elementary probability assumption for (T_n, U^{n-1}) in the first two equations. $(*)$ follows because, as a consequence of sufficiency (4), for any random quantity G that is determined by U^n, i.e. $G = \gamma(U^n)$ for some function γ, there exist versions of the conditional distributions $P_\delta^G \mid T_n$ such that for all $t \in \{t_n(u^{n-1}, u_n) \mid u_n \in \mathcal{U}_n\}$, it holds that the density $f_\delta^G \mid T_n = t$ is the same for all $\delta \in \Delta$ so that $p_{\delta^+}^G(g \mid T_n = t) = 1$; in particular this then holds for $G = U^{n-1}$.

Now note that the right-hand side of (8) is increasing in t for fixed u^{n-1} by assumption, showing the second claim. The first claim follows from combining (8) for $t = t_n(u^n)$ with (7), giving

$$p_{\delta_+}^{T_n}(t_n(u^n)|u^{n-1}) \;=\; \frac{p_{\delta_+}^{T_n}(t_n(u^n))}{p_{\delta_+}^{U^{n-1}}(u^{n-1})} \;=\; p_{\delta_+}^{U^n}(u_n|u^{n-1}). \quad \square$$

Proof of Theorem 4. We apply Proposition 2 for each n and each $u^{n-1} \in \mathcal{U}^{(n-1)}$, with T replaced by T_n and $p_{\delta^+}^T$ replaced by $p_{\delta^+}^{T_n} \mid u^{n-1}$. By Lemma 3, the MLR property holds for these *conditional* densities, so the proposition can be applied and gives that for each n and u^{n-1}, it holds that $p_{\delta^+}^{T_n}(T_n \mid u^{n-1})$ is an e-variable on domain $\mathcal{U}^{(n)}$ conditional on past data $U^{n-1} = u^{n-1}$, i.e. for all $\delta \leq \delta_0$, we have:

$$\mathbf{E}_\delta\left[p_{\delta^+}^{T_n}(T_n \mid u^{n-1})\Big|u^{n-1}\right] \leq 1 \tag{9}$$

and hence the products $(\prod_{i=1}^n p_\delta^{T_i}(T_i \mid U^{i-1}))_{n\in\mathbb{N}}$ of these past-conditional e-variables constitute a supermartingale (see Proposition 2 of Grünwald et al. (2024)). But from the first claim of Lemma 3 we know that for all n and $u^{n-1} \in \mathcal{U}^{(n-1)}$, we have for each factor in the product: $p_{\delta^+}^{T_n}(T_n \mid u^{n-1}) = p_{\delta^+}^{U_n}(U_n \mid u^{n-1})$. Since also for each n, it holds that $p_\delta^{U^n}(U^n) = \prod_{i=1}^n p_\delta^{U_i}(U_i \mid U^{i-1})$, the processes $(\prod_{i=1}^n p_\delta^{T_i}(T_i \mid U^{i-1}))_{n\in\mathbb{N}}$ and $(p_\delta^{U^n}(U_n))_{n\in\mathbb{N}}$ must coincide. $\square$

3. Further discussion and future work

Priors on alternative and growth-rate optimality considerations. What if we put a general prior W on $\{\delta \in \Delta : \delta \geq \delta_0\}$ instead of the point-mass on $\delta^+ \geq \delta_0$ that we implicitly used in Lemma 3 and Theorem 4 above? Following exactly the same steps as above we see that both results still hold with $p_{\delta^+}^{\cdot}$ replaced everywhere by $p_W^{\cdot}$ for an arbitrary such prior W. What can we say about the quality of the general supermartingale $(P_W^{U_n})_{n\in\mathbb{N}}$, measured in the standard sense of *e-power* or, equivalently, *log-optimality* (Ramdas et al., 2023)? First, note that for every stopping time τ, we have that $p_W^{U^\tau}$ is an e-variable. Following standard definitions, we call this e-variable GRO (growth-rate optimal) or equivalently *log-optimal* or *numéraire* (Larsson et al., 2025) for alternative $\{P_W\}$ on sample space $\mathcal{U}^{(\tau)}$ against null $\mathcal{H}_0$ if it achieves

$$\max_{S \in \mathcal{E}(\mathcal{H}_0; U^\tau)} \mathbf{E}_{P_W} \left[\ln S(U^\tau)\right] \tag{10}$$

where $\mathcal{E}(\mathcal{H}_0; U^\tau)$ is the set of all e-variables relative to null hypothesis $\mathcal{H}_0$, that can be written as a function of U^τ. We first note that the likelihood ratio $p_W^{U^\tau}$ is clearly GRO (growth-rate-optimal) for $\{P_W\}$ against null $\mathcal{H}_0 := \mathcal{H}_{\delta=\delta_0} = \{P_{\delta_0}\}$, i.e. it achieves (10) with $\mathcal{H}_0$ replaced by $\mathcal{H}_{\delta=\delta_0}$. This follows from the general fact that likelihood ratios are growth-rate optimal for simple nulls (Grünwald et al., 2024; Ramdas et al., 2023). But it turns out that $p_W^{U^\tau}$ is also GRO (growth-rate-optimal) for $\{P_W\}$ against one-sided null $\mathcal{H}_0 := \mathcal{H}_{\delta \le \delta_0}$. To see this, note that, for any e-variable S, if $S \in \mathcal{E}(\mathcal{H}_a; U^\tau)$ and $S \in \mathcal{E}(\mathcal{H}_b; U^\tau)$ with $\mathcal{H}_a \subset \mathcal{H}_b$, and S is growth-optimal relative to $\mathcal{H}_a$, then it must also be growth-optimal relative to $\mathcal{H}_b$ with the same alternative (the maximum in (10) is taken over a smaller set if we set $\mathcal{H}_0 = \mathcal{H}_b$ rather than $\mathcal{H}_0 = \mathcal{H}_a$). Since this holds for *every* stopping time τ, we can in fact infer the much stronger conclusion that the process $(p_W^{U^n})_{n \in \mathbb{N}}$ has the LOAVEV (log-optimal anytime-valid e-value) property for alternative P_W against null $\mathcal{H}_{\delta \le \delta_0}$, meaning that it is growth-optimal in the strongest possible sense (Koolen and Grünwald, 2022). In a similar manner one can show that Theorem 4 also implies that, for any stopping time τ, we have that $p_{\delta^+}^{U^\tau}$ is GROW (growth-rate optimal in the worst case) for the one-sided composite alternative $\mathcal{H}_{\delta \ge \delta^+}$ against $\mathcal{H}_{\delta \le \delta_0}$, i.e. it achieves (10) with $\mathbf{E}_{U^\tau \sim P_W}$ replaced by $\inf_{\delta \ge \delta^+} \mathbf{E}_{U^\tau \sim P_\delta}$.

What does this imply for practice? As explained by Wang and Ramdas (2025), in the one-sided t-test setting, one may obtain an e-process by the method of universal inference (UI); their reasoning extends to all settings we cover in Appendix B. Prior to the results we just presented, UI would have been the method of choice to obtain tests that remain valid under optional stopping. Our result shows that one may replace UI by a test supermartingale that has the log-optimality property. Indeed, in terms of e-power, log-optimal e-values tend to beat UI by an amount that is logarithmic (yet nonnegligible in practice) in the sample size: Hao and Grünwald (2024) provide many examples of this phenomenon. While their results do not cover the t-test, they do (see the remarks underneath their Theorem 2) include the χ^2-test treated in Appendix B.

Future work. In several classical testing scenarios that are closely related to the ones we present here, the minimal sufficient statistic is multivariate — a case in point is the setting of Hotelling's T^2 test, or the 2×2 tables considered by Turner et al. (2024). It would be interesting (but certainly challenging) to see if our arguments can be extended to such cases.

Acknowledgment

We thank Aaditya Ramdas and Hongjiang Wang for noticing the issue that was adressed in this paper. Peter Grünwald has been supported by ERC ADG project No 101142168 (FLEX).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2025.110574.

Data availability

No data was used for the research described in the article.

References

Bhowmik, J., King, M., 2007. Maximal invariant likelihood based testing of semi-linear models. Statist. Papers 48, 357–383.
Brown, L.D., Johnstone, I.M., MacGibbon, K.B., 1981. Variation diminishing transformations: a direct approach to total positivity and its statistical applications. J. Amer. Statist. Assoc. 76 (376), 824–832.
Fischer, L., Barry, T., Ramdas, A., 2024. Multiple testing with anytime-valid Monte-Carlo p-values. URL https://arxiv.org/abs/2404.15586.
Fischer, L., Xu, Z., Ramdas, A., 2025. An online generalization of the (e-) Benjamini-Hochberg procedure. URL https://arxiv.org/abs/2407.20683.
Goeman, J., de Heide, R., Solari, A., 2025. The e-Partitioning principle of false discovery rate control. URL https://arxiv.org/abs/2504.15946.
Grünwald, P.D., de Heide, R., Koolen, W.M., 2024. Safe testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 86 (5), 1091–1128, With Discussion.
Hao, Y., Grünwald, P., 2024. E-Values for exponential families: the general case. URL https://arxiv.org/abs/2409.11134.
Ignatiadis, N., Wang, R., Ramdas, A., 2024. Asymptotic and compound e-values: multiple testing and empirical Bayes. URL https://arxiv.org/abs/2409.1981.
Koolen, W.M., Grünwald, P., 2022. Log-Optimal anytime-valid E-values. Internat. J. Approx. Reason. 141, 69–82, Festschrift for G. Shafer's 75th Birthday.
Kruskal, W., 1954. The Monotonicity of the Ratio of Two Noncentral t-Density Functions. Ann. Math. Stat. 25 (1), 162–165.
Larsson, M., Ramdas, A., Ruf, J., 2025. The numeraire e-variable and reverse information projection. Ann. Stat. 53 (3), 1015–1–43.
Lehmann, E., Romano, J.P., 1986. Testing statistical hypotheses, vol. 3, Springer.
Lindon, M., Ham, D.W., Tingley, M., Bojinov, I., 2024. Anytime-Valid linear models and regression adjusted causal inference in randomized experiments. URL https://arxiv.org/abs/2210.08589.
Pérez-Ortiz, M.F., Lardy, T., de Heide, R., Grünwald, P., 2024. E-Statistics, group invariance and anytime valid testing. Ann. Statist. 52 (4), 1410–1432.
Ramdas, A., Grünwald, P., Vovk, V., Shafer, G., 2023. Game-Theoretic statistics and safe Anytime-Valid inference. Statist. Sci. (ISSN: 0883-4237) 38 (4), 576–601. http://dx.doi.org/10.1214/23-STS894.
ter Schure, J., Pérez-Ortiz, M., Ly, A., Grünwald, P., 2024. The Anytime-Valid logrank test: Error control under continuous monitoring with unlimited horizon. New Engl. J. Stat. Data Sci. 2 (2), 190–214.
Turner, R., Ly, A., Grünwald, P., 2024. Generic E-Variables for exact sequential k-Sample tests that allow for optional stopping. Stat. Plan. Inference 230, 106116.
Wang, H., 2024. Personal communication.
Wang, H., Ramdas, A., 2025. Anytime-valid t-tests and confidence sequences for Gaussian means with unknown variance. Sequential Anal. 44 (1), 56–110. http://dx.doi.org/10.1080/07474946.2024.2428245.
Williams, D., 1991. Probability with Martingales. In: Cambridge Mathematical Textbooks.