Admissible anytime-valid sequential inference must rely on nonnegative martingales

Aaditya Ramdas[1], Johannes Ruf[2], Martin Larsson[3], Wouter M. Koolen[4]

[1] Departments of Statistics and Machine Learning, Carnegie Mellon University
[2] Department of Mathematics, London School of Economics
[3] Department of Mathematics, Carnegie Mellon University
[4] Machine Learning Group, CWI Amsterdam

aramdas@cmu.edu, j.ruf@lse.ac.uk
martinl@andrew.cmu.edu, wmkoolen@cwi.nl

September 22, 2020

Abstract

Wald's anytime-valid p-values and Robbins' confidence sequences enable sequential inference for composite and nonparametric classes of distributions at arbitrary stopping times, as do more recent proposals involving Vovk's 'e-values' or Shafer's 'betting scores'. Examining the literature, one finds that at the heart of all these (quite different) approaches has been the identification of composite nonnegative (super)martingales. Thus, informally, nonnegative (super)martingales are known to be sufficient for *valid* sequential inference. Our central contribution is to show that martingales are also universal—all *admissible* constructions of (composite) anytime p-values, confidence sequences, or e-values must necessarily utilize nonnegative martingales (or so-called max-martingales in the case of p-values). Sufficient conditions for composite admissibility are also provided. Our proofs utilize a plethora of modern mathematical tools for composite testing and estimation problems: max-martingales, Snell envelopes, and new Doob-Lévy martingales make appearances in previously unencountered ways. Informally, if one wishes to perform anytime-valid sequential inference, then any existing approach can be recovered or dominated using martingales. We provide several sophisticated examples, with special focus on the nonparametric problem of testing if a distribution is symmetric, where our new constructions render past methods inadmissible.

Keywords: Admissibility; anytime p-value; composite nonnegative supermartingale; confidence sequence; Doob-Lévy martingale; e-value; max-martingale; optional stopping; sequential inference; Snell envelope; symmetric distribution; Ville's inequality.

Contents

1 Introduction

Our fairly mathematical treatment will be immensely helped by a concrete hypothetical example. Consider a sociology (or your favorite discipline) laboratory that wishes to understand if a particular intervention ('treatment') has any positive effect whatsoever on a prespecified outcome of interest. Without getting too bogged down by the details, suppose the 'average treatment effect' of the intervention over the relevant population is denoted by θ. Suppose they want to test $H_0 : \theta \leq 0$ against $H_1 : \theta > 0$, or to estimate θ using a confidence interval. The lab believes that there is an effect, but has no idea how many subjects to collect data from: a larger sample size means more power, but also more time and money. So they conduct their experiment sequentially: subjects enter the study one at a time and are assigned to treatment or control completely at random; denote the data from subject t as X_t.

After observing X_t, they analyze the data $X_1, \dots, X_t$ they have so far, and decide if they wish to collect more data, or whether what they already have suffices to demonstrate an effect (to themselves, or to a journal, or to the world). Thus, the lab stops their experiment at a data-dependent stopping time τ—maybe time ran out, or they used the money up faster than expected; maybe the effect was sufficiently large, or perhaps they lowered their sights by being satisfied with a smaller effect, or became more optimistic and kept the experiment running longer in the hope for a narrower confidence interval around a (hopefully) large effect. In other words, the stopping criterion used may itself have changed over time with funding coming in or drying up, with initial results being more/less promising than anticipated.

In any case, the experiment was stopped at time τ and not earlier or later, and there could be multiple data-dependent reasons for stopping at τ that were impossible to anticipate in advance.

On adopting the testing approach, they may hope to construct a sequence of p-values $(\mathfrak{p}_t)_{t \in \mathbb{N}}$ that satisfies

$$\text{for any arbitrary stopping time } \tau, \ \Pr_{H_0}(\mathfrak{p}_\tau \leq a) \leq a, \ \text{for all } a \in [0, 1], \tag{1}$$

which is simply the definition of a p-value at stopping time τ. Unfortunately, naively using a t-test, a chi-squared test, or permutations, does not yield a p-value with this property. Indeed, these types of 'standard' non-sequential p-values only satisfy the weaker property

$$\text{for any data-independent time } t, \ \Pr_{H_0}(\mathfrak{p}_t \leq a) \leq a, \ \text{for all } a \in [0, 1].$$

The most straightforward way to construct 'anytime-valid p-values' that satisfy property (1) is to employ bonafide sequential tests like Wald's sequential likelihood ratio test [25, 26], but extensions to nonparametric settings have also been explored recently [10, 27]. These are also called sequentially-adjusted p-values [5], or always-valid p-values [12].

If instead the lab had adopted the confidence interval approach, and specifying an error tolerance $\alpha \in [0, 1]$, they may hope to construct a sequence of confidence sets (often intervals) $(\mathfrak{C}_\tau(\alpha))_{t \in \mathbb{N}}$ which satisfies

$$\text{for any arbitrary stopping time } \tau, \ \Pr(\theta \in \mathfrak{C}_\tau(\alpha)) \geq 1 - \alpha. \tag{2}$$

Once more, unfortunately, a naive confidence interval based on the central limit theorem or the bootstrap does not satisfy the desired property. As before, these 'standard' constructions instead satisfy the above property only at fixed data-independent times t. To satisfy property (2), one could employ 'confidence sequences' proposed by Robbins and collaborators like Darling, Siegmund, and Lai [3, 18, 14], and regaining interest in recent years [17, 11]; this is a central topic of this paper and we return to it later.

Recently, another set of highly interrelated ideas has been put forward under a variety of names by authors such as Shafer, Vovk, Grünwald, and their collaborators: test martingales [20] (since the test statistics are sometimes martingales), e-values or sequential e-values [23, 24] (e for expectation), betting scores [19] (since they have roots in gambling), or safe e-values [9] (safe under optional stopping). Even though these concepts often have origins in parametric settings, the ideas have been extended to complicated nonparametric settings involving composite irregular models [27, 10]. We try to strike a balance of terminologies: we use the terminology of an 'e-value', we denote it by $\mathfrak{e}$ to reinforce the fact that it is an e-value. Importantly, our use of the term 'safe' does not in any way imply that the other two concepts (anytime p-values and confidence sequences) are unsafe; indeed they are also safe against optional stopping and continuation of experiments. For this paper, a 'safe e-value' is a nonnegative sequence $(\mathfrak{e}_t)_{t\in\mathbb{N}}$ that satisfies

$$\text{for any arbitrary stopping time } \tau,\ \mathtt{E}_{H_0}[\mathfrak{e}_\tau] \leq 1. \tag{3}$$

As before, a standard e-value (that is not anytime-valid) exhibits the above property only at fixed times t, which does not suffice for our sequentially motivated applications.

More generally, each of the above modes of inference are often used to perform sequential testing of H_0, but they are not necessarily exhaustive. We may instead want to directly consider a level-α sequential test, which is a decision rule that maps the data (and α) onto $\{0, 1\}$, and stop when the test first outputs one (rejection of the null). Formally, a level-α sequential test is a binary sequence $(\psi_t)_{t\in\mathbb{N}}$ that satisfies

$$\text{for any arbitrary stopping time } \tau,\ \Pr_{H_0}(\psi_\tau = 1) \leq \alpha. \tag{4}$$

Once more, standard nonsequential tests only satisfy such a type-I error guarantee at fixed times t. Instead, anytime-valid p-values, e-values, or confidence sets can each be used to derive a level-α sequential test.

We formally define all of the above concepts for composite nulls in Section 4, but the above semi-formal description suffices for the moment. One common theme amongst all the aforementioned works over the decades is the repeated appearance of various, often sophisticated, nonnegative supermartingales as the central object that enables all four types of anytime-valid inference, no matter what name they go under. In the rest of this paper, we further examine this central role of nonnegative martingales in constructing p-values, confidence sets, and e-values with the desired robustness to optional stopping (and continuation) of experiments. Specifically, we show that all admissible constructions of these objects must employ nonnegative martingales (either explicitly, or implicitly under the hood).

We provide a single example here of condition (2) for the reader to have a concrete instance in mind. In the above setup, suppose X_s is i.i.d. standard Gaussian, then it can be shown [11] that

$$\text{for any finite stopping time } \tau,\ \Pr\left(\theta \in \left[\frac{\sum_{s\leq\tau} X_s}{\tau} \pm \sqrt{\frac{(1+1/\tau)\log((\tau+1)/\alpha^2)}{\tau}}\right]\right) \geq 1-\alpha. \tag{5}$$

Of course, at any fixed time t, with z_q denoting the q-quantile of a standard Gaussian Z, we could have used a width of $z_{1-\alpha/2}/\sqrt{t}$. To approximate $z_{1-\alpha/2}$, note that the Gaussian tail inequality yields that for $t \geq 1$, we have $\Pr(Z > t) \leq (\sqrt{2\pi})^{-1/2}\exp(-t^2/2)$. Setting the right hand side to $\alpha/2$, we get that $z_{1-\alpha/2} \leq \sqrt{\log(2/(\pi\alpha^2))}$, which is known to be reasonably tight for small α. Thus the main difference in the time-uniform bound is the presence of an additional $\approx \sqrt{\log t}$ factor. The above inequality is proved by applying Ville's inequality (see (7a) below) to an exponential Gaussian-mixture martingale. These are tools we encounter later in this paper so we do not elaborate on them further here.

Admissibility. Naturally, the desire for methods satisfying properties like (1), (2), (3), or (4) comes with an implicit wish for efficiency. In other words, setting $\mathfrak{p}_\tau = 1$, $\mathfrak{e}_\tau = 1$, $\psi_\tau = 0$, or $\mathfrak{C}_\tau = \mathbb{R}$ (or $\mathfrak{C}_\tau = \Theta$ for a more general parameter space) trivially satisfies those requirements, but is clearly uninformative. We want $\mathfrak{p}_\tau$ to be as small as possible, $\mathfrak{e}_\tau$ and ψ_τ to be as large as possible, and $\mathfrak{C}_\tau$ to be as narrow as possible, while still being statistically valid measures of uncertainty. We use the term 'dominates' to compare pairs of these objects (in order to avoid using case-by-case adjectives like small/large/narrow)—so, if $\mathfrak{p}' \leq \mathfrak{p}$ then $\mathfrak{p}'$ dominates $\mathfrak{p}$. Similarly, if $\mathfrak{e}' \geq \mathfrak{e}$ then $\mathfrak{e}'$ dominates $\mathfrak{e}$, if $\psi' \geq \psi$ then ψ' dominates ψ, and if $\mathfrak{C}' \subseteq \mathfrak{C}$ then $\mathfrak{C}'$ dominates $\mathfrak{C}$. In this paper, we use the notion of admissibility to capture this idea: informally, a p-value (or e-value, test, confidence set) is inadmissible if it is strictly dominated by another p-value (or e-value, test, confidence set). We define admissibility more formally in Section 4.

Paper outline. Sections 2 and 3 lay out the formal definitions of several of the basic tools—nonnegative (super)martingales, max-martingales, Doob's optional stopping theorem, and Ville's inequality. Section 4 introduces the four central tools of anytime-valid sequential inference: Wald's anytime p-values, safe e-values, sequential tests, and Robbins' confidence sequences. Section 5 provides two simple examples:

Gaussian and symmetric (super)martingales. Then Section 6 and Section 7 summarize this paper's central message about the centrality of nonnegative martingales in constructing the aforementioned tools. Section 6 formalizes the necessary and sufficient conditions for admissibility in the point null setting; it uses a Doob-Lévy max-martingale construction to show the necessary conditions for p-values, a Doob-Lеvy martingale for sequential tests, and uses the Doob decomposition of an appropriate Snell envelope to prove that admissible e-values must also (explicitly or implicitly) employ nonnegative martingales. Section 7 develops several novel reductions of admissibility in the composite null setting to the point null case, and presents extensions to estimation (confidence sequences). Section 8 presents deeper investigations on admissibility, including anti-concentration bounds and the role of randomization. Section 9 utilizes the learnt lessons to produce admissible tests for symmetry. Appendix A recaps certain technical concepts like local domination and essential suprema. Appendix B details all proofs that are not in the main paper. Finally, Appendix C contains examples and counterexamples to support several claims made in the paper.

2 Wald's anytime p-values and Robbins' confidence sequences

2.1 A time-uniform equivalence lemma

The following lemma is quite central to the construction and interpretation of p-values, confidence sets, and sequential tests that are valid at arbitrary stopping times.

Lemma 1 (Equivalence lemma)**.** *Let $(A_t)_{t\in\mathbb{N}}$ be an adapted sequence of events in some filtered probability space and let $A_\infty := \limsup_{t\to\infty} A_t := \bigcap_{t\in\mathbb{N}} \bigcup_{s\ge t} A_s$. The following statements are then equivalent:*

(i) $\Pr(\bigcup_{t\in\mathbb{N}} A_t) \le \alpha$.

(ii) $\sup_T \Pr(A_T) \le \alpha$, *where T ranges over random times, possibly infinite, not necessarily stopping times.*

(iii) $\sup_\tau \Pr(A_\tau) \le \alpha$, *where τ ranges over stopping times, possibly infinite.*

The proof can be found in Appendix B. If the event A_t is associated with making an erroneous claim at time t, we interpret the aforementioned three statements as follows:

(i) The probability of ever making an erroneous claim, from time one to infinity, is at most α.

(ii) The probability of making an erroneous claim at an arbitrary data-dependent time T, perhaps chosen post-hoc as a past time after an experiment is stopped, is at most α.

(iii) When we stop an experiment at an arbitrary stopping time τ, the probability of making an erroneous claim at that time is at most α.

Intuitively, it is clear that (i) implies (ii), which in turn implies (iii), but the aforementioned lemma establishes that all three properties are actually equivalent: if you want one of them, you get all of them for free. This lemma gives the first hint of the centrality of martingales—the third statement is very directly about optional stopping, even though this fact is somewhat masked in the first two ways of framing the desired error control. While (iii) enables inferences at stopping times as initially motivated, property (ii) allows further introspection at previous times, enabling statistically valid answers to questions like 'what was the estimate of the treatment effect at time $\tau/2$?' (where τ was the stopping time).

The above lemma first appeared recently in Howard et al. [11]. While Lemma 1 did not motivate its original definition in 1967, Darling and Robbins [3] first defined a 'confidence sequence' for a parameter θ as an infinite sequence of confidence sets $(\mathfrak{C}_t)_{t\in\mathbb{N}}$ such that

$$\Pr(\exists t \in \mathbb{N} : \theta \notin \mathfrak{C}_t) \le \alpha.$$

In other words the aforementioned confidence sets satisfy property (i) for $A_t := \mathbf{1}_{\theta\notin\mathfrak{C}_t}$.

Since its inception 75 years ago, the field of sequential analysis has devoted much effort to constructing anytime-valid p-values for testing and confidence sequences for estimation. Underlying the construction of these objects in a variety of works, one often finds the repeated use of Ville's inequality for nonnegative supermartingales (NSMs). We will show that this is not a coincidence: we prove that NSMs underlie all admissible constructions for performing anytime-valid sequential inference.

To make these claims more formal, and especially to handle composite hypothesis testing, we need to clarify what the probability Pr means in the above definitions even further, and we do so next.

2.2 Notation and conventions

Let $\mathbb{N}$ represent the natural numbers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We use (B_t) to denote a sequence $(B_t)_{t\in\mathbb{N}}$ or $(B_t)_{t\in\mathbb{N}_0}$ where the indexing of t is either implicitly understood from the context or unimportant, but we use B_t without the brackets to denote a particular element from the sequence. Thus, for example, $\mathcal{F}_t$ will denote a sigma-field at time t but $(\mathcal{F}_t)$ denotes a filtration, which is an increasing sequence of sigma-fields. Unless otherwise mentioned, $\mathcal{F}_0 = \sigma(U)$, where U is a $[0,1]$-uniform random variable that is independent of everything else, signifying an external source of randomness, and $\mathcal{F}_t := \sigma(U, X_1, \ldots, X_t)$ will denote the canonical filtration, where X_t is the data observed at time t. We allow X_t to take values in some general space, which we do not need to specify here, e.g., in $\mathbb{R}^d$, equipped with the Borel sigma algebra.

Earlier, we used Pr to represent the probability taken over all sources of randomness, but in what follows we will use a more explicit notation: we denote the distribution of an infinite sequence of observations by $\mathtt{P}$; this means that $\mathtt{P}$ is a probability measure on $\mathcal{F}_\infty := \sigma(U, (X_t)_{t\in\mathbb{N}}) = \sigma(\bigcup_{t\in\mathbb{N}_0} \mathcal{F}_t)$. Expectations with respect to $\mathtt{P}$ are denoted $\mathtt{E_P}$. A set consisting of distributions over sequences will be denoted $\mathcal{P}$; so $\mathcal{P} = \{\mathtt{P}\}$ is the singleton case, but more generally there may be uncountably many $\mathtt{P} \in \mathcal{P}$. In the case of testing, we denote the null set of distributions by $\mathcal{Q} \subset \mathcal{P}$.

Next, τ will always denote a stopping time, while t denotes a fixed time. A subscript t for $\mathfrak{p}_t$, $\mathfrak{e}_t$, ψ_t, and $\mathfrak{C}_t$ means that these objects were constructed using only the data available up to time t. In other words, $\mathfrak{p}_t$, $\mathfrak{e}_t$, ψ_t, and $\mathfrak{C}_t$ are $\mathcal{F}_t$-measurable, or the sequences $(\mathfrak{p}_t)$, $(\mathfrak{e}_t)$, (ψ_t), and $(\mathfrak{C}_t)$ are adapted to $(\mathcal{F}_t)$. It is also understood that a p-value $\mathfrak{p}_t$ has range $[0,1]$, an e-value $\mathfrak{e}_t$ has range $[0,\infty]$ and a sequential test ψ_t has range $\{0,1\}$; the range of the confidence set $\mathfrak{C}_t$ will be formally specified later.

If $\mathtt{P}$ is a probability measure on $\mathcal{F}_\infty$ and τ is a stopping time, we write $\mathtt{P}_\tau$ for the restriction of $\mathtt{P}$ to $\mathcal{F}_\tau$. This is simply the probability measure on $\mathcal{F}_\tau$ defined by $\mathtt{P}_\tau(A) = \mathtt{P}(A)$, for $A \in \mathcal{F}_\tau$. (Think of this as the 'coarsening' of $\mathtt{P}$ that only operates on events observable up to time τ.)

We sometimes, but not always, assume that $\mathcal{P}$ is 'locally dominated' by (i.e., locally absolutely continuous with respect to) a fixed reference measure $\mathtt{R}$; we review the meaning of this in Appendix A.1. For example, if each observation X_s has a Lebesgue density under all $\mathtt{P} \in \mathcal{P}$, one can choose the reference measure to be the distribution of an i.i.d. sequence of standard Gaussians. Existence of a reference measure is needed to unambiguously interpret conditional expectations like $\mathtt{E_P}[Y \mid \mathcal{F}_t]$ under measures different from $\mathtt{P}$, since *a priori*, such conditional expectations are only defined up to $\mathtt{P}$-nullsets. For completeness, we elaborate on this in Appendix A.1, but this issue will not actually be visible in the proofs of our results.

2.3 A composite time-uniform equivalence lemma

One of this paper's central contributions is to characterize admissibility in composite settings. With that motivation, we present the following extension of Lemma 1 using the notation introduced above.

Lemma 2 (Composite equivalence lemma)**.** *Let $\mathcal{Q}$ be a family of probability measures. Let $(A_t)_{t\in\mathbb{N}}$ be an adapted sequence of events in some filtered probability space and let $A_\infty := \limsup_{t\to\infty} A_t := \bigcap_{t\in\mathbb{N}} \bigcup_{s\geq t} A_s$. The following statements are equivalent:*

(i) $\sup_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}(\bigcup_{t\in\mathbb{N}} A_t) \leq \alpha$.

(ii) $\sup_T \sup_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}(A_T) \leq \alpha$, *where T ranges over all random times, possibly infinite.*

(iii) $\sup_\tau \sup_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}(A_\tau) \leq \alpha$, *where τ ranges over all stopping times, possibly infinite.*

Further, if equality holds for any one, then it holds for the other two.

The proof is exactly the same as in Lemma 1 and is thus omitted. For contrast, we now state a version with expectations instead of probabilities in which the corresponding statements are not equivalent. Such a non-equivalence points to forthcoming differences between p-values and e-values. The following result resembles Lemma 1 but a composite version resembling Lemma 2 can also be easily stated.

Lemma 3 (Non-equivalence lemma). *Let $(N_t)_{t\in\mathbb{N}_0}$ be an adapted sequence of nonnegative integrable random variables in a filtered probability space; let $N_\infty := \limsup_{t\to\infty} N_t$. Consider the following statements:*

(i) $\mathtt{E}[\sup_{t\in\mathbb{N}_0} N_t] \le 1$.

(ii) $\mathtt{E}[N_T] \le 1$ *for all random times T, possibly infinite and not necessarily stopping times.*

(iii) $\mathtt{E}[N_\tau] \le 1$ *for all stopping times τ, possibly infinite.*

(iv) $\mathtt{E}[g(1) \vee \sup_{t\in\mathbb{N}_0} g(N_t)] \le 1$ *for any nondecreasing function g such that $\int_1^\infty g(y)/y^2 \mathrm{d}y = 1$; in particular,* $\mathtt{E}[1 \vee \sup_{t\in\mathbb{N}_0} \sqrt{N_t}] \le 2$.

Then (i) and (ii) are equivalent. Both (i) and (ii) imply (iii), which in turn implies (iv).

The proof can be found in Appendix B. Contrasting the above lemma with Lemma 1 brings out some of the differences between p-values and e-values. To dig deeper at the difference, note that one could have equivalently written Lemma 1 in terms of Bernoulli random variables $B_t := \mathbf{1}_{A_t}$, in which case the formulae above involving $\mathtt{Q}(\dots)$ would be replaced by (i) $\mathtt{E}_\mathtt{Q}[\sup_{t\in\mathbb{N}_0} B_t]$, (ii) $\mathtt{E}_\mathtt{Q}[B_T]$, and (iii) $\mathtt{E}_\mathtt{Q}[B_\tau]$, respectively. Thus, for these specific nonnegative binary random variables (B_t), the relevant statements are all equivalent, but more generally they are not. This difference later manifests itself in the inability to take running suprema for e-values, and overall a rather different underlying structure.

3 Martingales, max-martingales, and Ville's inequality

A martingale is a stochastic process adapted to an underlying filtration, whose value at any time is the conditional expectation of its value at any later time. This is however not the only possible notion of martingale; another interesting notion is obtained by replacing conditional expectations by so-called conditional suprema, leading to *max-martingales*. Both notions play an important role in this paper. In particular, max-martingales turn out to be particularly suitable for dealing with p-values. We briefly review the definitions and basic properties of martingales and max-martingales.

3.1 Martingales (based on conditional expectation)

Given a filtration $(\mathcal{F}_t)$ and a measure $\mathtt{P}$ on $\mathcal{F}_\infty$, a process $(M_t)_{t\in\mathbb{N}_0}$ is called a martingale (with respect to $(\mathcal{F}_t)$) if M_t is $\mathcal{F}_t$-measurable, $\mathtt{P}$-integrable, and

$$\mathtt{E}_\mathtt{P}[M_t|\mathcal{F}_s] = M_s \text{ for any } t \text{ and } s \le t.$$

(Sub- and supermartingales are defined by relaxing the martingale property and allowing for inequality, $\ge$ respectively, $\le$.) Since we had earlier mentioned that $\mathcal{F}_0$ includes an initial source of independent randomness, M_0 is itself allowed to be random. Naturally, we have $\mathtt{E}_\mathtt{P}[M_t] = \mathtt{E}_\mathtt{P}[M_0]$. Often, in this paper, (M_t) will be nonnegative and the latter quantity equals one and so when we say 'a nonnegative martingale with initial value one', we implicitly mean with initial *expected* value one.

Given an $\mathcal{F}_\infty$-measurable integrable random variable Y, the process $M_t := \mathtt{E}_\mathtt{P}[Y|\mathcal{F}_t]$ is known as the Doob (or Doob–Lévy) martingale associated with Y. The fact that this is a martingale follows from the tower rule of the conditional expectation: $\mathtt{E}_\mathtt{P}[M_t|\mathcal{F}_s] = \mathtt{E}_\mathtt{P}[\mathtt{E}_\mathtt{P}[Y|\mathcal{F}_t]\ |\mathcal{F}_s] = M_s$ if $s \le t$.

We now generalize these definitions to hold for an entire set of measures. Given a set $\mathcal{P}$ of measures on $\mathcal{F}_\infty$, a process $(M_t)_{t\in\mathbb{N}_0}$ is called a nonnegative $\mathcal{P}$-supermartingale ($\mathcal{P}$-NSM) (with respect to $(\mathcal{F}_t)$) if M_t is nonnegative, $\mathcal{F}_t$-measurable, and

$$\mathtt{E}_\mathtt{P}[M_t|\mathcal{F}_s] \le M_s \text{ for all } t \in \mathbb{N},\ s \le t \text{ and every } \mathtt{P} \in \mathcal{P}. \tag{6}$$

If (6) holds with equality, (M_t) is called a nonnegative $\mathcal{P}$-martingale ($\mathcal{P}$-NM). If $\mathcal{P} = \{\mathtt{P}\}$ is a singleton, we write '$\mathtt{P}$-NSM' instead of '$\{\mathtt{P}\}$-NSM'. This notational choice is also applied to other objects. We refer to a $\mathcal{P}$-NM (or NSM) as a 'composite' NM (or NSM), while a $\mathtt{P}$-NM is called a 'pointwise' NM (or NSM).

Doob's optional stopping theorem [6] extends the (sub-/super-) martingale property from deterministic times to stopping times. In general, only bounded stopping times are allowed in the optional stopping theorem; however, the nonnegativity of an NSM relieves us of this restriction. In particular, if (N_t) is (upper bounded by) a $\mathcal{P}$-NSM starting in N_0, the optional stopping theorem implies that

$$\mathtt{E}_{\mathtt{P}}[N_\tau] \leq \mathtt{E}_{\mathtt{P}}[N_0] \text{ for all stopping times } \tau \text{, potentially infinite, and every } \mathtt{P} \in \mathcal{P}.$$

In fact, if (M_t) is a $\mathcal{P}$-NSM, then we additionally have $\mathtt{E}_{\mathtt{P}}[M_\tau | \mathcal{F}_\rho] \leq M_\rho$ for all stopping times ρ and τ such that $\rho \leq \tau$, $\mathtt{P}$-almost surely, for each $\mathtt{P} \in \mathcal{P}$. Further, by Doob's supermartingale convergence theorem, we know that if (M_t) is a $\mathcal{P}$-NSM with initial expected value one, then its limit $M_\infty := \lim_{t\to\infty} M_t$ exists $\mathtt{P}$-almost surely and $\mathtt{E}_{\mathtt{P}}[M_\infty] \in [0,1]$, for each $\mathtt{P} \in \mathcal{P}$.

Stemming from his 1939 PhD thesis [21], Ville's inequality is a time-uniform generalization of Markov's inequality; for our purposes, the relevant version states that if (M_t) is (upper-bounded by) a $\mathcal{P}$-NSM with initial expected value one, then the following three equivalent statements hold:

$$\mathtt{P}\left(\exists t \in \mathbb{N} : M_t \geq \frac{1}{\alpha}\right) \leq \alpha \text{ for every } \mathtt{P} \in \mathcal{P} \text{ and } \alpha \in [0,1]; \tag{7a}$$

$$\iff \quad \sup_{\mathtt{P}\in\mathcal{P}} \mathtt{P}\left(\sup_{t\in\mathbb{N}} M_t \geq \frac{1}{\alpha}\right) \leq \alpha \text{ for every } \alpha \in [0,1]; \tag{7b}$$

$$\iff \quad \sup_{\mathtt{P}\in\mathcal{P}, \tau\geq 0} \mathtt{P}\left(M_\tau \geq \frac{1}{\alpha}\right) \leq \alpha \text{ for every } \alpha \in [0,1]. \tag{7c}$$

Note that (7b) and (7c) usually only hold with inequality (for example, for the singleton $\mathcal{P} = \{\mathtt{P}\}$) but it can hold with equality for larger nontrivial nonparametric classes $\mathcal{P}$, as we shall encounter later in Section 8.2. We also remark that a conditional version of Ville's inequality is also true, though we do not utilize it much in this paper. Specifically, if (M_t) is a $\mathcal{P}$-NSM, then

$$\sup_{\mathtt{P}\in\mathcal{P}} \mathtt{P}\left(\exists t \geq s : M_t \geq \frac{M_s}{\alpha} \,\middle|\, \mathcal{F}_s\right) \leq \alpha \text{ for every } \alpha \in [0,1]. \tag{8}$$

The relationship between likelihood ratios and nonnegative martingales. The simplest NM (beyond the constant process $M_t = 1$) that arises rather naturally is the likelihood ratio; indeed this is at the heart of Wald's sequential likelihood ratio test [25, 26]. To be specific, when testing $H_0 : X_s \sim Q$ versus $H_0 : X_s \sim P$, define the likelihood ratio $M_t := \prod_{s\leq t} \frac{\mathrm{d}P}{\mathrm{d}Q}(X_s)$, assuming that the Radon-Nikodym derivative $\mathrm{d}P/\mathrm{d}Q$ exists. If P, Q have densities p, q with respect to a common measure then each term in the product is just $p(X_s)/q(X_s)$. Let Q^∞ now denote the distribution under which the sequence is i.i.d., each element distributed according to Q. Wald effectively proved that (M_t) is a Q^∞-NM and a test that rejects if $M_t \geq 1/\alpha$ controls the Type-I error at level α due to Ville's inequality (Wald proved the result from scratch, but the language of martingales and Ville's thesis was known to Wald [2]). It is also apparent that every nonnegative martingale is a product of nonnegative random variables with conditional mean one, meaning that if (M_t) is a $\mathcal{Q}$-NM, then $M_t = \prod_{s\leq t} Y_s$, where (Y_t) is adapted to $(\mathcal{F}_t)$ and $\mathtt{E}_{\mathtt{Q}}[Y_t|\mathcal{F}_{t-1}] = 1$ for every $\mathtt{Q} \in \mathcal{Q}$; to see this, simply define the multiplicative increment as $Y_t := M_t/M_{t-1}$ with $0/0 := 1$. At first sight, despite having a product form, it may appear like nonnegative martingales are strict generalizations of likelihood ratio processes. However, in fact, a converse statement is also true: not only is every likelihood ratio a martingale (under the null), but every martingale is also implicitly a likelihood ratio; this was discussed by Shafer et al. [20] for point nulls, and we generalize it below to the composite case, borrowing the terminology of 'implied alternative' from Shafer [19].

To make the following result precise we assume that the sequence of observations $(U, (X_t)_{t\in\mathbb{N}})$ is a process on the space $\Omega = \mathbb{R}^{\mathbb{N}}$ of real-valued sequences, and we let $(\mathcal{F}_t)$ be the canonical filtration.

Proposition 4. *Consider any composite null set $\mathcal{Q}$ of probability measures on $\mathcal{F}_\infty$, and let $\mathcal{P}$ consist of all probability measures $\mathtt{P}$ that are locally absolutely continuous with respect to some $\mathtt{Q} \in \mathcal{Q}$. (Thus in particular, $\mathcal{Q} \subset \mathcal{P}$.) If (M_t) is a $\mathcal{Q}$-NM starting at one, then for every $\mathtt{Q} \in \mathcal{Q}$ there exists some 'implied alternative' distribution $\mathtt{P} \in \mathcal{P}$ (depending on $\mathtt{Q}$) that is locally dominated by $\mathtt{Q}$, such that $M_t = \mathrm{d}\mathtt{P}_t/\mathrm{d}\mathtt{Q}_t$. In other words, a composite nonnegative martingale is a 'composite' likelihood ratio (meaning, it takes the form of a likelihood ratio under every element of the null).*

We recognize that the above statement may be known to different researchers in some form, but it does provide useful intuition and we have not seen it stated in the generality above in the statistics literature. The proof is in Section B. The informal takeaway message is that nonnegative martingales are implicitly likelihood ratios, but the former are typically easier to identify (or construct) in composite null settings.

3.2 Max-martingales (based on conditional supremum)

Max-martingales are defined by replacing the conditional expectation by the *conditional supremum*, so we start by reviewing this notion; more information can be found in Barron et al. [1] and Larsson [15]; see also Appendix A.2. For a given probability measure $\mathtt{P}$, random variable Y, and sub-σ-algebra $\mathcal{G}$, the *$\mathcal{G}$-conditional supremum* is defined as the smallest $\mathcal{G}$-measurable almost sure upper bound on Y:

$$\bigvee\nolimits_{\mathtt{P}}[Y \mid \mathcal{G}] := \operatorname{ess\,inf}\{Z \colon Z \text{ is } \mathcal{G}\text{-measurable and } Z \geq Y,\ \mathtt{P}\text{-almost surely}\}.$$

Note that $\bigvee_{\mathtt{P}}[Y \mid \mathcal{G}] \geq Y$ by construction, and is the smallest $\mathcal{G}$-measurable random variable with that property. (Here and in the rest of this subsection, equalities are understood in the $\mathtt{P}$-almost sure sense.) The conditional supremum can be viewed as a nonlinear analog of the conditional expectation, and has similar properties. In particular, one has a 'tower property' which states that for nested sub-σ-algebras $\mathcal{G} \subset \mathcal{H}$, one has

$$\bigvee\nolimits_{\mathtt{P}}\Big[\bigvee\nolimits_{\mathtt{P}}[Y \mid \mathcal{H}]\Big|\mathcal{G}\Big] = \bigvee\nolimits_{\mathtt{P}}[Y \mid \mathcal{G}].$$

Given a filtration $(\mathcal{F}_t)$, a process (Y_t) is called a $\mathtt{P}$-*max-martingale*, or $\mathtt{P}$-MM for short, if

$$Y_s = \bigvee\nolimits_{\mathtt{P}}[Y_t \mid \mathcal{F}_s], \quad s \leq t.$$

Any max-martingale (Y_t) is almost surely decreasing, which ensures that the limit $\lim_{t\to\infty} Y_t$ exists in $[-\infty, \infty]$. We call a max-martingale *closed* if $Y_t = \bigvee_{\mathtt{P}}[\lim_{s\to\infty} Y_s \mid \mathcal{F}_t]$ for all $t \in \mathbb{N}_0$.

As earlier, given a set $\mathcal{P}$ of measures on $\mathcal{F}_\infty$, a process (Y_t) is called a (closed) $\mathcal{P}$-max-martingale ($\mathcal{P}$-MM) if (Y_t) is a (closed) $\mathtt{P}$-MM for each $\mathtt{P} \in \mathcal{P}$.

One can also introduce notions of sub- and supermartingales using the conditional supremum, although we will not use these here. Moreover, one can analogously define *conditional infimum* martingales using $\bigwedge$ instead of $\bigvee$, and all properties stated above and below also hold analogously.

In further analogy with (standard) martingales, max-martingales satisfy an optional stopping theorem [15, Lemma 2.10]. Specifically, consider the process $Y_t := \bigvee_{\mathtt{P}}[Y \mid \mathcal{F}_t]$ for an $\mathcal{F}_\infty$-measurable Y. Thanks to the tower property, (Y_t) is then a max-martingale; we call such a construction a Doob-Lévy MM. We then have

$$Y_\tau = \bigvee\nolimits_{\mathtt{P}}[Y \mid \mathcal{F}_\tau] \qquad \text{and} \qquad Y_\rho = \bigvee\nolimits_{\mathtt{P}}[Y_\tau \mid \mathcal{F}_\rho]$$

for all finite stopping times ρ and τ such that $\rho \leq \tau$.

The connection between $\mathtt{P}$-MMs and $\mathtt{P}$-NMs goes beyond mere analogies. If (M_t) is a $\mathtt{P}$-NM with $M_0 > 0$, the following statement easily follows from [15, Proposition 4.1]:

$$\inf_{s\leq t} \frac{1}{M_s} = \bigvee\nolimits_{\mathtt{P}}\Big[\inf_{s\in\mathbb{N}_0} \frac{1}{M_s}\Big|\mathcal{F}_t\Big], \qquad t \in \mathbb{N}_0. \tag{9}$$

A word of warning: although the conditional supremum and conditional expectation are in some ways similar, they sometimes behave very differently. In particular, the conditional supremum only depends on the underlying measure $\mathtt{P}$ through its zero measure sets. Computing the conditional supremum under a different measure $\mathtt{P}'$ thus gives the same result whenever the two measures are mutually absolutely continuous. This is in stark contrast to the behavior of the conditional expectation. (On the other hand, if $\mathtt{P}$ and $\mathtt{P}'$ are mutually singular, as is often the case in our infinite-horizon situations, then one can make no general statements about the relation between the corresponding conditional suprema.)

Although any nonnegative max-martingale (Y_t) is almost surely decreasing, which ensures that the limit $\lim_{t\to\infty} Y_t$ exists, it is possible that Y_t and $\bigvee_{\mathtt{P}}[\lim_{s\to\infty} Y_s \mid \mathcal{F}_t]$ are not necessarily the same, i.e., that (Y_t) is not closed. Indeed, we have the following non-uniqueness property. If (Y_t) is a max-martingale then we usually can find another max-martingale (Y'_t) with $Y'_t > Y_t$ for each $t \in \mathbb{N}$, but (Y'_t) and (Y_t) have the same limit, almost surely. For example, assume (Z_t) is i.i.d. Bernoulli$(1/2)$, independent of (Y_t), and adapted to $(\mathcal{F}_t)$. Define $Y'_t = Y_t + \prod_{s\le t}(Z_s \vee 1/2)$. Then (Y'_t) is also a max-martingale, with $Y'_t \ge Y_t + 2^{-t} > Y_t$ for all $t \in \mathbb{N}$, but converging almost surely to the same limit as Y'_t as $t \to \infty$.

A similar phenomenon also holds true for martingales: if (M_t) is an NM then $M'_t = M_t + \prod_{s\le t}(Z_s + 1/2)$ is another NM with $M'_t \ge M_t + (1/2)^t > M_t$ for all $t \in \mathbb{N}$, but $\lim_{t\to\infty} M'_t = \lim_{t\to\infty} M_t$. Hence also for nonnegative martingales, we often do not have that M_t and $\mathtt{E}[\lim_{s\to\infty} M_s|\mathcal{F}_t]$ are the same.

We conclude this subsection with an interesting example.

Example 5. *Let V denote a uniformly distributed random variable and assume that (X_t) is, conditionally on V, i.i.d.* Bernoulli(V). *Then $V = \lim_{t\to\infty} \sum_{s\le t} X_s/t$ (the limit frequency of ones), hence V is $\mathcal{F}_\infty$-measurable. Moreover, $Y_t := \bigvee_{\mathtt{P}}[V \mid \mathcal{F}_t]$ yields a max-martingale. It is now relatively easy to check that $Y_t = 1$ for each $t \in \mathbb{N}_0$. This yields an instance where $\lim_{t\to\infty} Y_t \ne V$, but (Y_t) is nonetheless a closed max-martingale since $Y_t = 1 = \bigvee_{\mathtt{P}}[1 \mid \mathcal{F}_t]$.*

4 Tools for anytime sequential inference and their admissibility

We formally introduce the four instruments for anytime-valid sequential inference that play central roles in our paper. We present a rather natural definition of admissibility for each of the instruments below, but recognize that other alternatives may be suitable depending on the context.

4.1 $\mathcal{Q}$-valid p-values, $\mathcal{Q}$-safe e-values, and $(\mathcal{Q}, \alpha)$-sequential tests

Let the (unknown) distribution of the observed data sequence be denoted $\mathtt{P}$. Suppose we wish to test the null $H_0 : \mathtt{P} \in \mathcal{Q}$ for some $\mathcal{Q} \subset \mathcal{P}$, against $H_1 : \mathtt{P} \in \mathcal{P}$ (or against $H_1 : \mathtt{P} \in \mathcal{P}\backslash\mathcal{Q}$). A sequence $(\mathfrak{p}_t)$ is called an anytime valid p-value for H_0 if it satisfies

$$\mathtt{Q}(\mathfrak{p}_\tau \le \alpha) \le \alpha \text{ for arbitrary stopping times } \tau, \text{ every } \mathtt{Q} \in \mathcal{Q}, \text{ and } \alpha \in [0, 1].$$

Above, we have implicitly defined and used

$$\mathfrak{p}_\infty := \liminf_{t\to\infty} \mathfrak{p}_t.$$

For succinctness, we say $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid. By Lemma 2, the above condition is equivalent to requiring that

$$\mathtt{Q}(\exists t \in \mathbb{N} : \mathfrak{p}_t \le \alpha) \le \alpha \text{ for every } \mathtt{Q} \in \mathcal{Q} \text{ and } \alpha \in [0, 1].$$

Note that $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid if and only if the running infimum $(\inf_{s\le t} \mathfrak{p}_s)$ is $\mathcal{Q}$-valid, so it helps to think of $(\mathfrak{p}_t)$ as a nonincreasing sequence; in this case, we have $\mathfrak{p}_\infty = \lim_{t\to\infty} \mathfrak{p}_t$ (the limit exists and equals the lim inf used earlier). When we refer only to the validity of the single random variable $\mathfrak{p}_\infty$, and not the sequence $(\mathfrak{p}_t)$, we say $\mathfrak{p}_\infty$ is 'statically' $\mathcal{Q}$-valid, meaning that its distribution is stochastically larger than uniform under any $\mathtt{Q} \in \mathcal{Q}$. The connection to Ville's inequality (7a), and thus to martingales, should be apparent.

Similarly, a sequence $(\mathfrak{e}_t)$ is called an anytime valid e-value for H_0—or, in short, $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe—if

$$\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_\tau] \le 1 \text{ for arbitrary stopping times } \tau, \text{ and every } \mathtt{Q} \in \mathcal{Q}.$$

Analogous to the p-value case above, we have implicitly defined and used

$$\mathfrak{e}_\infty := \limsup_{t\to\infty} \mathfrak{e}_t.$$

It is worth noting that already in the case of a singleton $\mathcal{Q} := \{\mathtt{Q}\}$, if one additionally desired a conditional safety property to hold for $(\mathfrak{e}_t)$, namely that $\mathtt{E}_\mathtt{Q}[\mathfrak{e}_\tau|\mathcal{F}_s] \le \mathfrak{e}_{s\wedge\tau}$ for arbitrary times s and stopping times τ, then $(\mathfrak{e}_t)$ must necessarily be a $\mathtt{Q}$-NSM; indeed by taking $\tau = t$ for $t \ge s$, we recover the definition of a $\mathtt{Q}$-NSM. In fact, if one would like such a property to hold for every $\mathtt{Q}$ in a composite $\mathcal{Q}$, such a requirement can only be satisfied by a $\mathcal{Q}$-NSM. Despite the fact that we will not require this conditional property, we will see that (composite) $\mathcal{Q}$-NMs or (pointwise) $\mathtt{Q}$-NMs play a central, universal role in constructing e-values that may not themselves be martingales.

A p-value or e-value does not directly yield a decision rule for when to reject the null hypothesis; they are real-valued measures of evidence, and need to be coupled with a decision rule in order to yield a test.

A binary sequence (ψ_t) is called a $(\mathcal{Q},\alpha)$-sequential test for H_0—or, in short, (ψ_t) is a $(\mathcal{Q},\alpha)$-ST—if

$$\mathtt{E}_\mathtt{Q}[\psi_\tau] = \mathtt{Q}(\psi_\tau = 1) \le \alpha \text{ for arbitrary stopping times } \tau, \text{ and every } \mathtt{Q} \in \mathcal{Q}.$$

As with p-values, we may think of (ψ_t) as nondecreasing, and have implicitly defined and used $\psi_\infty := \lim_{t\to\infty} \psi_t$.

Admissibility: We follow the standard convention of using the term 'admissible' as a shorthand for 'not inadmissible', so we only define inadmissibility below.

We say that $(\mathfrak{p}_t)$ is inadmissible if there exists $(\mathfrak{p}'_t)$ that is $\mathcal{Q}$-valid, and is always at least as good but sometimes strictly better; more formally, $\mathtt{Q}(\mathfrak{p}'_t \le \mathfrak{p}_t) = 1$ for all $t \in \mathbb{N}_0$ and all $\mathtt{Q} \in \mathcal{Q}$, and that $\mathtt{Q}(\mathfrak{p}'_t < \mathfrak{p}_t) > 0$ for some $t \in \mathbb{N}_0$, and some $\mathtt{Q} \in \mathcal{Q}$. We say that $(\mathfrak{e}_t)$ is inadmissible if there exists $(\mathfrak{e}'_t)$ that is $\mathcal{Q}$-safe, such that $\mathtt{Q}(\mathfrak{e}'_t \ge \mathfrak{e}_t) = 1$ for all $t \in \mathbb{N}_0$ and all $\mathtt{Q} \in \mathcal{Q}$, and also $\mathtt{Q}(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$ for some $t \in \mathbb{N}_0$, and some $\mathtt{Q} \in \mathcal{Q}$. Finally, (ψ_t) is inadmissible if there exists a $(\mathcal{Q},\alpha)$-sequential test (ψ'_t), such that $\mathtt{Q}(\psi'_t \ge \psi_t) = 1$ for all $t \in \mathbb{N}_0$ and all $\mathtt{Q} \in \mathcal{Q}$, and $\mathtt{Q}(\psi'_t > \psi_t) > 0$ for some $t \in \mathbb{N}_0$, and some $\mathtt{Q} \in \mathcal{Q}$.

Remark 6. *We could have equivalently formulated admissibility in terms of $\mathtt{Q}(\mathfrak{e}'_\tau > \mathfrak{e}_\tau) > 0$ for some stopping time τ, etc.; this is in fact identical to the current definition, because for any discrete-time random sequences (W_t) and (W'_t), the statement '$\mathtt{Q}(W'_t = W_t) = 1$ for all t' is equivalent to '$\mathtt{Q}(W'_t = W_t \text{ for all } t) = 1$', meaning that pointwise equality at fixed times yields simultaneous equality (including stopping times). In case it is useful to the reader, another restatement of inadmissibility is the following:*

$$\inf_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}(\forall t \in \mathbb{N}_0 : \mathfrak{p}'_t \le \mathfrak{p}_t) = 1 \textit{ and } \sup_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}(\exists t \in \mathbb{N}_0 : \mathfrak{p}'_t < \mathfrak{p}_t) > 0.$$

The above definition of admissibility does not specify any alternative. What allows us to do so is that the admissibility conditions are stated 'almost surely'. To elaborate, assume there was an alternative, say $\mathtt{P}^*$ such that there existed some time t and an event $A \in \mathcal{F}_t$ with $\mathtt{P}^*(A) > 0$ but $\mathtt{Q}(A) = 0$ for all $\mathtt{Q} \in \mathcal{Q}$. In this case, on the event A we would always set $\mathfrak{p}_t = 0$, $\mathfrak{e}_t = \infty$, and $\psi_t = 1$. We could always do so because no $\mathtt{Q} \in \mathcal{Q}$ 'is aware of' the event A, hence modifying a p-value on A would not change any of its $\mathtt{Q}$-distributional properties (like validity). For this reason, and to avoid any notational complication arising, from now on we always assume the following:

> If there exists some $t \in \mathbb{N}$, some $A \in \mathcal{F}_t$, and some $\mathtt{P}^* \in \mathcal{P}$ with $\mathtt{P}^*(A) > 0$ then there exists also some $\mathtt{Q} \in \mathcal{Q}$ with $\mathtt{Q}(A) > 0$.

This is a very mild assumption! It is, for example, satisfied if there exists some $\mathtt{Q}^* \in \mathcal{Q}$ such that every $\mathtt{P} \in \mathcal{P}$ is locally absolutely continuous with respect to $\mathtt{Q}^*$ (see Appendix A.1 for a review). This previous condition is satisfied, for example, if all measures in $\mathcal{P}$ are locally equivalent. In interesting situations a locally dominating measure $\mathtt{Q}^*$ may not actually exist—for example when testing for symmetry, as we encounter in this paper—but our standing assumption above is still true. Specifically, the aforementioned assumption also holds if each $\mathtt{P} \in \mathcal{P}$ is locally absolutely continuous with respect to some specific $\mathtt{Q} \in \mathcal{Q}$ (potentially different for each $\mathtt{P}$)—this is often easier to check, and indeed holds in the symmetric example.

Note that validity is subset-proof, meaning that validity for $\mathcal{Q}$ implies validity for $\mathcal{Q}' \subset \mathcal{Q}$; however, admissibility is neither subset-proof nor superset-proof. It is posssible that $(\mathfrak{p}_t)$ may be admissible for $H_0 : \mathtt{Q} \in \mathcal{Q}$, but not for $H_0 : \mathtt{Q} \in \mathcal{Q}'$ for some $\mathcal{Q}' \subset \mathcal{Q}$ or $\mathcal{Q}' \supset \mathcal{Q}$; indeed, it may not even be valid in the latter case. Thus, to avoid confusions, we should say $(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible, but we sometimes drop the additional prefix if it can be inferred from the context. (The same logic applies for $(\mathfrak{e}_t)$ and (ψ_t).)

A family of sequential tests $\{(\psi_t(\alpha))\}_{\alpha\in[0,1]}$ is said to be 'nested in α', if $\psi_t(\alpha') \geq \psi_t(\alpha)$ for any t and $\alpha' \geq \alpha$. This ensures that it is easier to reject the null with a larger Type-I error budget.

We conclude this subsection with the following observation.

Proposition 7. *Assume that $\mathcal{Q}$ is locally dominated. Then any p-value for $\mathcal{Q}$, e-value for $\mathcal{Q}$, or $(\mathcal{Q}, \alpha)$-sequential test can be dominated by an admissible p-value, e-value, or sequential test, respectively.*

It is proven in Appendix B using transfinite induction. Next, we switch from testing to estimation.

4.2 $(\phi, \mathcal{P}, \alpha)$-confidence sequences

Let ϕ be a map from $\mathcal{P}$ to an arbitrary set $\mathcal{Z}$. For every every $\gamma \in \mathcal{Z}$, define

$$\mathcal{P}^\gamma := \{\mathtt{P} \in \mathcal{P} : \phi(\mathtt{P}) = \gamma\},$$

and note that $\{\mathcal{P}^\gamma\}_{\gamma\in\mathcal{Z}}$ is a partition of $\mathcal{P}$. Of special interest is the i.i.d. case where $\mathtt{P} = \mu^\infty$. As examples, consider $\phi(\mathtt{P}) = \phi^{\mathrm{med}}(\mu)$ denoting the median of μ, or $\phi(\mathtt{P}) = \phi^{\mathrm{mean}}(\mu)$ denoting the mean of μ. Another case of interest is where each $\mathtt{P} \in \mathcal{P}$ can be represented as $\mathtt{P}_\theta$ for some unique $\theta \in \Theta$ (the 'parametric' setting) and then $\phi(\mathtt{P}_\theta) = \theta$. One may also care about fully nonparametric functionals, for example ϕ^{cdf}, which maps μ to its cumulative distribution function (or ϕ^{pdf} if μ has a Lebesgue density).

We then define a $(1-\alpha)$-confidence sequence for a functional ϕ as an adapted sequence of confidence sets $(\mathfrak{C}_t)$ such that

$$\sup_{\mathtt{P}\in\mathcal{P}} \mathtt{P}\left(\exists t \in \mathbb{N} : \phi(\mathtt{P}) \notin \mathfrak{C}_t\right) \leq \alpha.$$

This error condition can be phrased equivalently as a coverage criterion: $\mathtt{P}\left(\forall t \in \mathbb{N}_0 : \phi(\mathtt{P}) \in \mathfrak{C}_t\right) \geq 1-\alpha$, for every $\mathtt{P} \in \mathcal{P}$. Above, we have suppressed the dependence of $\mathfrak{C}_t$ on α for notational succinctness. Moreover, we emphasize that the outer probability $\mathtt{P}$ matches with the inner functional $\phi(\mathtt{P})$, as it should, Indeed, the coverage probability for $\phi(\mathtt{P})$ should hold given that the data used to construct $\mathfrak{C}_t$ was drawn according to this $\mathtt{P}$. We then say that $(\mathfrak{C}_t)$ is a $(\phi, \mathcal{P}, \alpha)$-CS. If required, we define $\mathfrak{C}_\infty := \liminf_{t\to\infty} \mathfrak{C}_t$.

Note that even though we defined $(\mathcal{Q}, \alpha)$-sequential tests without explicit reference to any functional ϕ, one could associate $\mathcal{Q}$ to the subset of $\mathcal{P}$ in which ϕ takes on certain values. In other words, we can absorb the role of any such ϕ into the definition of $\mathcal{Q}$ to reduce notational overhead.

We say $(\mathfrak{C}_t)$ is inadmissible if there exists $(\mathfrak{C}'_t)$ that is a $(\phi, \mathcal{P}, \alpha)$-CS such that $\mathtt{P}(\mathfrak{C}'_t \subseteq \mathfrak{C}_t) = 1$ for all t and all $\mathtt{P} \in \mathcal{P}$, and $\mathtt{P}(\mathfrak{C}'_t \subsetneq \mathfrak{C}_t) > 0$ for some t and some $\mathtt{P} \in \mathcal{P}$.

Finally, let us mention a minor technical point. A CS $(\mathfrak{C}_t)$ is by definition an *adapted sequence of random sets*. In this paper, we take 'adapted' to simply mean that all events of the form $\{z \in \mathfrak{C}_t\}$ belong to $\mathcal{F}_t$. This is sufficient for our needs, and lets us avoid dealing with σ-algebras on spaces of sets.

Last, as for sequential tests, a family of CSs $\{(\mathfrak{C}_t(\alpha))\}_{\alpha\in[0,1]}$ is said to be 'nested in α' if $\mathfrak{C}_t(\alpha') \subseteq \mathfrak{C}_t(\alpha)$ for any t and $\alpha' \geq \alpha$.

Remark 8. *An analogue of Proposition 7 also holds for $(\phi, \mathcal{P}, \alpha)$-confidence sequences, where $\mathcal{P}$ is locally dominated. This will later follow from Theorem 28, in conjunction with Proposition 7.*

4.3 Reductions between the four instruments

We now demonstrate how to transform one tool for sequential inference into another; some of these are well-known or 'obvious' but some are new, especially in the composite setting. These transformations are usually not lossless or bidirectional, meaning that something may be lost in going from one to another and back again. All proofs are relegated to Appendix B.

We start with the most straightforward direction: forming $(\mathcal{Q}, \alpha)$-sequential tests from the other three.

Proposition 9. *One can construct $(\mathcal{Q},\alpha)$-sequential tests in the following ways:*

(1) *If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid, then $\{(\mathbf{1}_{\mathfrak{p}_t\le\alpha})\}_{\alpha\in[0,1]}$ is a nested family of $(\mathcal{Q},\alpha)$-STs.*

(2) *If $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe, then $\{(\mathbf{1}_{\mathfrak{e}_t\ge 1/\alpha})\}_{\alpha\in[0,1]}$ is a nested family of $(\mathcal{Q},\alpha)$-STs.*

(3) *If $(\mathfrak{C}_t)$ is a $(\phi,\mathcal{P},\alpha)$-CS, then $(\mathbf{1}_{\phi(\mathcal{Q})\cap\mathfrak{C}_t=\emptyset})$ is a $(\mathcal{Q},\alpha)$-ST.*

Next, we show how to form a composite anytime p-value from the other three.

Proposition 10. *One can construct p-values for $\mathcal{Q}$ in the following ways:*

(1) *If $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe, then $(1\wedge\inf_{s\le t}1/\mathfrak{e}_s)$ is $\mathcal{Q}$-valid.*

(2) *If $\{(\psi_t(\alpha))\}_{\alpha\in[0,1]}$ is a nested family of $\mathcal{Q}$-STs, then $(\inf\{\alpha:\psi_t(\alpha)=1\})$ is $\mathcal{Q}$-valid.*

(3) *If $\{(\mathfrak{C}_t(\alpha))\}_{\alpha\in[0,1]}$ is a nested family of $(\phi,\mathcal{P})$-CSs, then $(\inf\{\alpha:\phi(\mathcal{Q})\cap\mathfrak{C}_t(\alpha)=\emptyset\})$ is $\mathcal{Q}$-valid.*

A confidence sequence can be formed by inverting families of tests, as follows.

Proposition 11. *Recall that $\mathcal{P}^\gamma:=\{\mathtt{P}\in\mathcal{P}:\phi(\mathtt{P})=\gamma\}$. If (ψ_t^γ) is a $(\mathcal{P}^\gamma,\alpha)$-sequential test for each $\gamma\in\mathcal{Z}$ then $\{\gamma\in\mathcal{Z}:\psi_t^\gamma=0\}$ is a $(\phi,\mathcal{P},\alpha)$-confidence sequence. To convert a family of p-values or e-values into confidence sequences, we can first convert them to sequential tests using Proposition 9 and then invert those tests as done in the first part of this proposition.*

Finally, we can form composite e-values by calibrating p-values, following Shafer et al. [20].

Proposition 12. *One can construct e-values for $\mathcal{Q}$ as follows: If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid, then $1/(2\sqrt{\mathfrak{p}_t})$ is $\mathcal{Q}$-safe. In fact, $(f(\mathfrak{p}_t))$ is $\mathcal{Q}$-safe for any nonincreasing 'calibration' function $f:[0,1]\to[0,\infty)$ such that $\int_0^1 f(u)\mathrm{d}u=1$. To convert a nested family of sequential tests or confidence sequences into e-values, we can first convert them to p-values using Proposition 10 and then apply the aforementioned calibration.*

4.4 Some basic closure properties

At several points in this paper, we will see that convexity of the class of distributions plays a central role, the first hint of which is provided below.

Proposition 13. *Let $\mathrm{conv}(\mathcal{Q})$ denote the convex hull of $\mathcal{Q}$. Then the following statements hold.*

(1) *An e-value for $\mathcal{Q}$ is also automatically an e-value for $\mathrm{conv}(\mathcal{Q})$. Such closure under the convex hull also holds for p-values and sequential tests.*

(2) *An admissible e-value for $\mathcal{Q}$ is also automatically an admissible e-value for $\mathrm{conv}(\mathcal{Q})$. Such closure under the convex hull also holds for p-values and sequential tests.*

The proof can be found in Appendix B. Example 42 in Appendix C shows that the statement of Proposition 13 does not extend to confidence sequences.

Remark 14. *Note that stability with respect to taking a running extremum does not hold for e-values, but does hold for the other three. To elaborate, if $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid, then so is $(\min_{s\le t}\mathfrak{p}_s)$; if $(\mathfrak{C}_t)$ is a $(\phi,\mathcal{P},\alpha)$-CS, then so is $(\bigcap_{s\le t}\mathfrak{C}_s)$; if (ψ_t) is $(\mathcal{Q},\alpha)$-ST, then so is $(\max_{s\le t}\psi_s)$. However, a running maximum does not usually preserve safety for e-values! Moreover, if $(\mathfrak{e}_t)$ is $\mathcal{P}$-safe and $(\mathfrak{f}_t)$ is $\mathcal{Q}$-safe, then $\min(\mathfrak{e}_t,\mathfrak{f}_t)$ is $\mathrm{conv}(\mathcal{P}\cup\mathcal{Q})$-safe, while $(\mathfrak{e}_t+\mathfrak{f}_t)/2$ is $\mathrm{conv}(\mathcal{P}\cap\mathcal{Q})$-safe. Similarly, if $(\mathfrak{p}_t)$ is $\mathcal{P}$-valid and $(\mathfrak{q}_t)$ is $\mathcal{Q}$-valid, then $\max(\mathfrak{p}_t,\mathfrak{q}_t)$ is $\mathrm{conv}(\mathcal{P}\cup\mathcal{Q})$-valid, while $2\min(\mathfrak{p}_t,\mathfrak{q}_t)$ is $\mathrm{conv}(\mathcal{P}\cap\mathcal{Q})$-valid. Last, if $(\mathfrak{C}_t)$ is a $(\phi,\mathcal{P},\alpha)$-CS and $(\mathfrak{D}_t)$ is a $(\phi,\mathcal{Q},\alpha)$-CS, then $\mathfrak{C}_t\cup\mathfrak{D}_t$ is a $(\phi,\mathcal{P}\cup\mathcal{Q},\alpha)$-CS, while $\mathfrak{C}_t\cap\mathfrak{D}_t$ is a $(\phi,\mathcal{P}\cap\mathcal{Q},2\alpha)$-CS.*

5 Two instructive examples of exponential (super)martingales

We now illustrate the above concepts with two examples of composite martingales: the Gaussian NM and the symmetric NSM, the former being a simple parametric example, and the latter being a sophisticated nonparametric example (we use the term sophisticated because it includes atomic and nonatomic distributions, and there is no underlying reference measure).

5.1 A Gaussian nonnegative martingale

If (X_t) is a sequence of i.i.d. standard Gaussians, then it is well known that $\exp(\sum_{s\le t} X_s - t/2)$ is a nonnegative martingale. This is one of the simplest nontrivial pointwise NMs that one can construct. Below, we elaborate on how to construct a simple composite NM in the Gaussian case. We generalize the Gaussian case to an arbitrary fixed mean m and a variance process, where each σ_t^2 is revealed to us on demand. Technically, we suppose the process (X_t) consists of two components, $X_t = (Y_t, \sigma_{t+1}^2)$, and let σ_1^2 be a constant. Note that (σ_t^2) is a predictable sequence. We now let $\mathcal{G}^m$ denote the set of distributions such that the outcome Y_t at time t is conditionally Gaussian with deterministic mean m and variance σ_t^2, that is,

$$\mathcal{G}^m := \{\mathtt{P} : Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(m, \sigma_t^2) \text{ for all } t \in \mathbb{N}\}.$$

Since the distribution generating the predictable variances is left unspecified, $\mathcal{G}^m$ is a highly composite set of measures. Next, define the process (G_t^m) by

$$G_t^m := \exp\left(\sum_{s\le t}(Y_s - m) - \frac{1}{2}\sum_{s\le t}\sigma_s^2\right). \tag{10}$$

It is easy to check that (G_t^m) is a $\mathcal{G}^m$-NM for each $m \in \mathbb{R}$, specifically by evaluating the moment generating function of X_t. As a direct consequence of Ville's inequality, a $(1-\alpha)$-CS $(\mathfrak{C}_t^{\text{mean}})$ for an unknown mean is given by

$$\mathfrak{C}_t^{\text{mean}} := \left\{m \in \mathbb{R} : G_t^m < \frac{1}{\alpha}\right\}.$$

Formally, $(\mathfrak{C}_t^{\text{mean}})$ is a $(\phi^{\text{mean}}, \bigcup_{m\in\mathbb{R}} \mathcal{G}^m, \alpha)$-CS for the mean, where we define the mean functional $\phi^{\text{mean}}(\mathtt{P}) = m$ for every $\mathtt{P} \in \mathcal{G}^m$.

Remark 15. *The confidence sequence $(\mathfrak{C}_t^{\text{mean}})$ derived above does not yield interval (5) in the introduction. That CS is obtained by noting that $(G_t^m(\lambda))$, given by*

$$G_t^m(\lambda) := \exp\left(\lambda\sum_{s\le t}(Y_s - m) - \frac{\lambda^2}{2}\sum_{s\le t}\sigma_s^2\right),$$

is a $\mathcal{G}^m$-NM for every $\lambda \in \mathbb{R}$. Fubini's theorem then implies that $(\int G_t^\mu(\lambda)\mathrm{d}\Phi(\lambda))$ is also a $\mathcal{G}^m$-NM for any distribution function Φ. Choosing Φ as a standard Gaussian for example, yields the normal mixture martingale (dating back at least to Darling and Robbins [3]). Applying Ville's inequality then yields (5).

5.2 A nonnegative supermartingale under symmetry

We now move to a nonparametric example that first appears in de la Peña [4] but the core idea can be traced back to Efron [7], who was interested in the robustness of the t-test to heavy tailed (but symmetric) distributions; further, it has recently been extended to the matrix setting by Howard et al. [10]. This example is particularly interesting because of three reasons: (a) it deals with a nonparametric class of distributions that does not have a common dominating measure (it includes atomic and non-atomic measures), (b) there exists a rather elegant well-known 'nonparametric' NSM, meaning that one single process is a composite NSM, (c) the visual form of the NSM below is reminiscent of the aforementioned

Gaussian example, making it intuitively appealing. Unfortunately, we will later demonstrate that this construction is suboptimal, and leads to inadmissible sequential inference.

Consider the convex set of distributions indexed by $m \in \mathbb{R}$, where each increment is conditionally symmetric around m, i.e.,

$$\mathcal{S}^m := \{\mathtt{P} : (X_t - m) \sim -(X_t - m) \mid \mathcal{F}_{t-1}\}. \tag{11}$$

Consider then the following family of processes (S_t^m), indexed by $m \in \mathbb{R}$:

$$S_t^m := \exp\left(\sum_{s\le t}(X_s - m) - \frac{1}{2}\sum_{s\le t}(X_s-m)^2\right). \tag{12}$$

It is known [4] that (S_t^m) is an $\mathcal{S}^m$-NSM for each fixed $m \in \mathbb{R}$; the proof stems from the observation that $\cosh(z) \le \exp(z^2/2)$, and thus for a single symmetric random variable Z, we have

$$\mathtt{E}[e^{Z-Z^2/2}] = \mathtt{E}[e^{-Z-Z^2/2}] = \mathtt{E}\left[\frac{e^{Z-Z^2/2}+e^{-Z-Z^2/2}}{2}\right] = \mathtt{E}[e^{-Z^2/2}\cosh(Z)] \le 1.$$

Note that the symmetric distribution could be different at each time point (e.g., Gaussian with mean m at time one, $(\delta_{m-|X_1|}+\delta_{m+|X_1|})/2$ at time two, etc., where δ_z denotes the Dirac measure at some $z \in \mathbb{R}$).

The process (S_t^m) is visually quite similar to the Gaussian process (G_t^m) from the previous subsection. The relaxation from using the true variance in (G_t^m) to using an empirical variance in (S_t^m), allows the NM property to transform to an NSM property for a much larger class of heavy-tailed distributions (such as t and Cauchy distributions). Even when applied to Gaussians, one no longer needs to know the variance.

One can check that the sets

$$\mathfrak{C}_t^{\text{center}} := \left\{m \in \mathbb{R} : S_t^m < \frac{1}{\alpha}\right\}$$

together form a $(\phi^{\text{center}}, \bigcup_{m\in\mathbb{R}}\mathcal{S}^m, \alpha)$-CS for the functional ϕ^{center}, which maps symmetric distributions to their center of symmetry. Said differently, if $\mathtt{P} \in \mathcal{S}^m$ then $\mathtt{P}(\exists t \in \mathbb{N} : m \notin \mathfrak{C}_t^{\text{center}}) \le \alpha$.

Above, we have described only the confidence sequences, but one could equally well have defined e-values and anytime p-values. For example, to test the null $H_0 : \mathtt{P} \in \mathcal{S}^0$, one can use the fact that (S_t^0) is an NSM under the null, and thus (S_t^0) is an e-value for $\mathcal{S}^0$, and $(\inf_{s\le t} 1/S_s^0)$ is a p-value for $\mathcal{S}^0$. Of course, all of these in turn define $(\mathcal{S}^0, \alpha)$-sequential tests.

We return to these and other examples later in the paper.

6 Admissible inference for point nulls via martingales

We begin our examination of admissibility via the lens of testing (p-values, e-values, and sequential tests) and leave the results on estimation (confidence sequences) for Subsection 7.4.

Theorem 16 (Necessary and sufficient conditions for pointwise admissibility)**.** *Consider a point null $\mathcal{Q} = \{\mathtt{Q}\}$. The following statements describe necessary and sufficient conditions for pointwise admissibility.*

(1) *If $(\mathfrak{p}_t)$ is admissible, then it is a closed $\mathtt{Q}$-MM with $F(\inf_{t\in\mathbb{N}_0}\mathfrak{p}_t) = \inf_{t\in\mathbb{N}_0}\mathfrak{p}_t$, where F is the distribution function of $\inf_{t\in\mathbb{N}_0}\mathfrak{p}_t$. In the other direction, if $(\mathfrak{p}_t)$ is a closed $\mathtt{Q}$-MM and $\inf_{t\in\mathbb{N}_0}\mathfrak{p}_t$ is $\mathtt{Q}$-uniformly distributed, then it is admissible.*

(2) *$(\mathfrak{e}_t)$ is admissible if and only if it is a $\mathtt{Q}$-NM with $\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_0] = 1$.*

(3) *(ψ_t) is an admissible $(\mathtt{Q}, \alpha)$-sequential test if and only if $\psi_t := \mathbf{1}_{\sup_{s\le t} M_s \ge 1/\alpha}$, where (M_t) is a $\mathtt{Q}$-NM with no overshoot at $1/\alpha$ and $M_\infty \in \{0, 1/\alpha\}$.*

The theorem is proven later in this section, with one subsection per statement.

Remark 17. *Similar to max-martingales, introduced in Subsection 3.2, we could have introduced min-martingales by replacing conditional suprema by infima. With such a notion in place it would be easy to see that (ψ_t) is an admissible $(\mathtt{Q}, \alpha)$-sequential test if and only if (ψ_t) is a closed min-martingale such that $\sup_{t\in\mathbb{N}_0} \psi_t$ is $\{0,1\}$-valued and $\mathtt{Q}(\sup_{t\in\mathbb{N}_0} \psi_t = 1) = \alpha$.*

Unfortunately, a version of the first statement in Theorem 16 that is phrased as follows—'if $(\mathfrak{p}_t)$ is admissible then $\mathfrak{p}_t = \inf_{s\le t} 1/M_s$ for all $t \in \mathbb{N}_0$, where (M_t) is a $\mathtt{Q}$-NM'—is incorrect. Below is a simple counterexample.

Example 18. *Let $\mathfrak{p}_0 = U$, and $\mathfrak{p}_0 = \mathfrak{p}_1 = \mathfrak{p}_2 = \mathfrak{p}_3 = \dots$. Then $(\mathfrak{p}_t)$ is admissible since it cannot be improved without violating uniformity. However, since the inverse of a uniform random variable is not integrable, we cannot find a NM (M_t) that yields $(\mathfrak{p}_t)$; indeed, no nonnegative integrable random variable N_0 can yield $\mathfrak{p}_0 = 1/N_0$.*

The above example further demonstrates that max-martingales, not the 'usual' martingales, are the right mathematical construct to deal with p-values. We remark that if we are allowed to construct continuous-time processes, then one can work with usual martingales, see Shafer et al. [20, Theorem 2]. However, we do obtain the following corollary of Theorem 16(1).

Corollary 19. *Let (M_t) denote a $\mathtt{Q}$–martingale with $M_0 > 0$ and let F denote the distribution of $\inf_{t\in\mathbb{N}_0} 1/M_t$. If F is atomless, then $\mathfrak{p}_t := F(\inf_{s\le t} 1/M_s)$ is an admissible p-value.*

Proof. We will make use of the fact that the conditional supremum commutes with continuous nondecreasing functions: $\bigvee[f(Y) \mid \mathcal{G}] = f(\bigvee[Y \mid \mathcal{G}])$ for every continuous nondecreasing function f; we will use this with $f = F$. Combining this with the max-martingale property of the reciprocal of the running supremum of an NM (see (9)) we get,

$$\mathfrak{p}_t = F\left(\inf_{s\le t} \frac{1}{M_s}\right) = F\left(\bigvee\nolimits_{\mathtt{Q}}\Big[\inf_{s\in\mathbb{N}_0} \frac{1}{M_s}\Big|\mathcal{F}_t\Big]\right) = \bigvee\nolimits_{\mathtt{Q}}\Big[F\Big(\inf_{s\in\mathbb{N}_0} \frac{1}{M_s}\Big)\Big|\mathcal{F}_t\Big] = \bigvee\nolimits_{\mathtt{Q}}\Big[\inf_{s\in\mathbb{N}_0} \mathfrak{p}_s\Big|\mathcal{F}_t\Big].$$

The last equality follows because $\inf_{s\in\mathbb{N}_0} \mathfrak{p}_s = F(\inf_{s\in\mathbb{N}_0} 1/M_s)$, which is uniformly distributed since F is atomless. We are now in a position to apply Theorem 16(1) to conclude that $(\mathfrak{p}_t)$ is admissible. $\square$

Let us now provide an example of a closed $\mathtt{Q}$-MM that satisfies $F(\inf_{t\in\mathbb{N}_0} \mathfrak{p}_t) = \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$ in the notation of Theorem 16(1), but is not admissible.

Example 20 (The gap between sufficient and admissible conditions in Theorem 16(1))**.** *This example shows that simply using $F(\inf_{s\le t} 1/M_s)$ for a Q-NM (M_t) and F as in Corollary 19 does not typically yield an admissible p-value. It also provides an example for the gap between the sufficient and admissible conditions in Theorem 16(1). Define the martingale (M_t) by $M_t := 1 + \mathbf{1}_{U\le 1/2}$ for all $t \in \mathbb{N}_0$. Then $F(\inf_{s\le t} 1/M_s) = \inf_{s\le t} 1/M_s$ is an inadmissible p-value. (It is, however, a p-value despite $\mathtt{E}_{\mathtt{Q}}[M_0] > 1$.) To see this, define a p-value $(\mathfrak{p}_t)$ by $\mathfrak{p}_t = U$ for all $t \in \mathbb{N}_0$. Then $(\mathfrak{p}_t)$ strictly dominates $\inf_{s\le t} 1/M_s$ because*

$$U \le \frac{1}{2}\mathbf{1}_{U\le 1/2} + \mathbf{1}_{U>1/2} = \inf_{s\le t} \frac{1}{M_s}.$$

6.1 Necessary and sufficient conditions for p-values (proof)

Proof of Theorem 16(1). We start with the necessary conditions for pointwise admissibility of a p-value. To this end, let $(\mathfrak{p}_t)$ be a $\mathtt{Q}$-admissible p-value, which must necessarily be nonincreasing by Remark 14. First define

$$\bar{\mathfrak{p}} := \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t = \lim_{t\to\infty} \mathfrak{p}_t,$$

and let F be the distribution function of $\bar{\mathfrak{p}}$. Since $(\mathfrak{p}_t)$ is valid, $\bar{\mathfrak{p}}$ is stochastically larger than uniform and so $F(x) \le x$ for all $x \in [0,1]$. For later use, let us observe that F is right-continuous, hence

$$\lim_{t\to\infty} F(\mathfrak{p}_t) = F(\bar{\mathfrak{p}}). \tag{13}$$

We now define

$$\mathfrak{p}'_t := \bigvee_{\mathbb{Q}} \left[F(\bar{\mathfrak{p}}) \middle| \mathcal{F}_t \right].$$

Since F is a nondecreasing function, we have $F(\bar{\mathfrak{p}}) \leq F(\mathfrak{p}_t)$ for all $t \in \mathbb{N}_0$. By definition of conditional supremum, $\mathfrak{p}'_t$ is the *smallest* $\mathcal{F}_t$-measurable random variable with this property; therefore, $\mathfrak{p}'_t \leq F(\mathfrak{p}_t)$ for all $t \in \mathbb{N}_0$. Since we also have $F(x) \leq x$ for all $x \in [0,1]$ we get $\mathfrak{p}'_t \leq \mathfrak{p}_t$ for all $t \in \mathbb{N}_0$. But $(\mathfrak{p}_t)$ is admissible by assumption, so we must in fact have the equality $\mathfrak{p}'_t = \mathfrak{p}_t$ for all $t \in \mathbb{N}_0$.

We have now argued $F(\bar{\mathfrak{p}}) \leq \mathfrak{p}'_t = \mathfrak{p}_t \leq F(\mathfrak{p}_t)$. Taking now limits in t and recalling (13), we get $F(\bar{\mathfrak{p}}) \leq \bar{\mathfrak{p}} \leq \lim_{t\to\infty} F(\mathfrak{p}_t) = F(\bar{\mathfrak{p}})$, thus allowing us to conclude that $F(\bar{\mathfrak{p}}) = \bar{\mathfrak{p}}$ and that $(\mathfrak{p}_t)$ is closed. This yields the necessary conditions of the theorem.

Let us now discuss the sufficient conditions for pointwise admissibility of a p-value. To this end, let $(\mathfrak{p}_t)$ be a closed MM such that $\inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$ is uniformly distributed. It is then clear that $(\mathfrak{p}_t)$ is valid. Consider now an arbitrary p-value $(\mathfrak{p}'_t)$ with $\mathfrak{p}'_t \leq \mathfrak{p}_t$ for all $t \in \mathbb{N}_0$. We must argue that we have equality. We clearly have $\inf_{t\in\mathbb{N}_0} \mathfrak{p}'_t \leq \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$. The validity of $(\mathfrak{p}'_t)$ implies moreover that $\inf_{t\in\mathbb{N}_0} \mathfrak{p}'_t$ stochastically dominates a uniform. This now directly yields that we have indeed $\inf_{t\in\mathbb{N}_0} \mathfrak{p}'_t = \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t =: \bar{\mathfrak{p}}$, which is uniform. By definition of max-martingales, $\mathfrak{p}_t$ is the smallest $\mathcal{F}_t$-measurable upper bound on $\bar{\mathfrak{p}}$; since $\mathfrak{p}'_t$ is another such bound, we must have $\mathfrak{p}'_t \geq \mathfrak{p}_t$ for all $t \in \mathbb{N}_0$. This proves that $(\mathfrak{p}_t)$ is admissible. ☐

6.2 Necessary and sufficient conditions for e-values (proof)

Proof of Theorem 16(2). Again, let us start with the necessary conditions for pointwise admissibility of an e-value. To this end, let us fix an admissible $(\mathfrak{e}_t)$. Next, let us define the 'Snell envelope' of $(\mathfrak{e}_t)$ as the process (L_t) given by

$$L_t := \operatorname*{ess\,sup}_{\tau\geq t} \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_\tau \mid \mathcal{F}_t],$$

where τ ranges over all finite stopping times.

First, observe that $\mathbb{E}_{\mathbb{Q}}[L_0] \leq 1$ because $(\mathfrak{e}_t)$ is $\mathbb{Q}$-safe and $\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_0] \leq 1$. It is self-evident that (L_t) inherits the nonnegativity property directly from $(\mathfrak{e}_t)$. Moreover, it is clear that $L_t \geq \mathfrak{e}_t$ since $\tau = t$ is a valid stopping time, for all $t \in \mathbb{N}_0$. Next, we claim that (L_t) is a supermartingale. This is a well-known result, but for the convenience of the reader we include the short proof. This uses properties of the essential supremum reviewed in Appendix A.2, in particular Proposition 39.

For each fixed $t \in \mathbb{N}_0$, L_t is the essential supremum of the family consisting of all $\mathbb{E}_{\mathbb{Q}}[M_\tau|\mathcal{F}_t]$ where $\tau \geq t$ is a finite stopping time. This family is closed under maxima. To see this, let τ and τ' be given, define $A = \{\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_\tau|\mathcal{F}_t] > \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau'} \mid \mathcal{F}_t]\}$, and set $\tau'' = \tau\mathbf{1}_A + \tau'\mathbf{1}_{A^c}$. Since A lies in $\mathcal{F}_t$ and $\tau, \tau' \geq t$, we have that τ'' is a stopping time and we obtain

$$\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau''} \mid \mathcal{F}_t] = \mathbf{1}_A \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_\tau \mid \mathcal{F}_t] + \mathbf{1}_{A^c}\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau'} \mid \mathcal{F}_t] = \max\left\{\mathbb{E}_{\mathbb{Q}}\left[\mathfrak{e}_\tau \mid \mathcal{F}_t\right], \mathbb{E}_{\mathbb{Q}}\left[\mathfrak{e}_{\tau'} \mid \mathcal{F}_t\right]\right\}.$$

This demonstrates closure under maxima. Consequently we can apply Proposition 39 to obtain finite stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ with $\tau_n \geq t$ such that $\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau_n}|\mathcal{F}_t] \uparrow L_t$ almost surely. Therefore, by the conditional version of the monotone convergence theorem, the tower rule, and the definition of L_{t-1}, we get

$$\mathbb{E}_{\mathbb{Q}}[L_t \mid \mathcal{F}_{t-1}] = \mathbb{E}_{\mathbb{Q}}\left[\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau_n} \mid \mathcal{F}_t] \middle| \mathcal{F}_{t-1}\right] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}[\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau_n} \mid \mathcal{F}_t] \mid \mathcal{F}_{t-1}] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_{\tau_n} \mid \mathcal{F}_{t-1}] \leq L_{t-1}.$$

This shows that (L_t) is a supermartingale.

Since we have established that the Snell envelope (L_t) is a supermartingale we can write down its Doob decomposition as $L_t = M_t - A_t$ for a unique (nonnegative) integrable martingale (M_t) with $M_0 = L_0$, and a unique nondecreasing predictable process (A_t) with $A_0 = 0$. The optional stopping theorem applied to the martingale (M_t) implies that it is an e-value and moreover, $M_t \geq L_t \geq \mathfrak{e}_t$ for all $t \in \mathbb{N}_0$. Since $(\mathfrak{e}_t)$ was assumed admissible we get $\mathfrak{e}_t = M_t$ for all $t \in \mathbb{N}_0$. Finally, we can assume that $\mathbb{E}_{\mathbb{Q}}[M_0] = 1$, else we can replace (M_t) by $(M_t + 1 - \mathbb{E}_{\mathbb{Q}}[M_0])$, which is again an e-value.

Let us now discuss the sufficient conditions for pointwise admissibility of a NM $(\mathfrak{e}_t)$ with $\mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_0] = 1$. First of all, the optional stopping theorem yields that $(\mathfrak{e}_t)$ is an e-value. Consider now some e-value $(\mathfrak{e}'_t)$ for $\mathbb{Q}$ with $\mathfrak{e}'_t \geq \mathfrak{e}_t$. Since $1 = \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}_t] \leq \mathbb{E}_{\mathbb{Q}}[\mathfrak{e}'_t] \leq 1$, we then have $\mathfrak{e}_t = \mathfrak{e}'_t$ for each $t \in \mathbb{N}_0$, yielding the admissibility of $(\mathfrak{e}_t)$, hence the assertion. ☐

6.3 Necessary and sufficient conditions for sequential tests (proof)

Proof of Theorem 16(3). Let us start with the necessary conditions for pointwise admissibility of a sequential test. Recall that by assumption, (ψ_t) satisfies

$$\bar{\alpha} := \mathtt{Q}(\exists t \in \mathbb{N}_0 : \psi_t = 1) \leq \alpha.$$

Define now

$$\psi_t' := \mathbf{1}_{U \leq \alpha - \bar{\alpha}} + \psi_t \mathbf{1}_{U > \alpha - \bar{\alpha}}, \qquad t \in \mathbb{N}_0.$$

Note that $\psi_t' \geq \psi_t$ for all $t \in \mathbb{N}_0$ and (ψ_t') is again a sequential test. If $\bar{\alpha} < \alpha$ then indeed (ψ_t') strictly dominates (ψ_t), in contradiction to the admissibilty of (ψ_t). Hence we may assume that $\bar{\alpha} = \alpha$.

Define now the Doob-Lévy martingale (M_t) by

$$M_t := \frac{\mathtt{Q}(\exists s \in \mathbb{N}_0 : \psi_s = 1 \mid \mathcal{F}_t)}{\mathtt{Q}(\exists s \in \mathbb{N}_0 : \psi_s = 1)} = \frac{\mathtt{Q}(\exists s \in \mathbb{N}_0 : \psi_s = 1 \mid \mathcal{F}_t)}{\alpha}.$$

Note that $\mathtt{E}_{\mathtt{Q}}[M_0] = 1$ and if there exists a time τ at which $\psi_\tau = 1$, then $M_t = 1/\alpha$ for any $t \geq \tau$. So, $M_\infty \in \{0, 1/\alpha\}$. Define next $\widetilde{\psi}_t := \mathbf{1}_{M_t \geq 1/\alpha}$. By Ville's inequality, $(\widetilde{\psi}_t)$ is a sequential test that dominates (ψ_t). Since the latter was assumed admissible, we have established $\psi_t = \widetilde{\psi}_t$ for all $t \in \mathbb{N}_0$.

Consider now a NM (M_t) with no overshoot at $1/\alpha$ and $M_\infty \in \{0, 1/\alpha\}$ and define $\psi_t := \mathbf{1}_{\sup_{s \leq t} M_s \geq 1/\alpha}$. By Ville's inequality, (ψ_t) is a sequential test. Consider next some sequential test (ψ_t') with $\psi_t' \geq \psi_t$ and fix some $t^* \in \mathbb{N}_0$. Since $\mathtt{E}_{\mathtt{Q}}[\psi_\infty] = \alpha$ we know that $\psi_{t^*}' = 1$ only on the event $\{\psi_\infty = 1\} = \{\tau < \infty\}$, where $\tau := \inf\{t \in \mathbb{N}_0 : M_t \geq 1/\alpha\}$. Hence, $\psi_{t^*}' = 1$ implies that $M_t \geq 1/\alpha$; otherwise the martingale property of (M_t) would be contradicted. Thus (ψ_t) is indeed $\mathcal{Q}$-admissible, concluding the proof of the statement. $\square$

7 Reducing admissible composite inference to the pointwise case

To build intuition towards composite admissibility, we begin with a basic question on validity: is there a systematic way to construct tools for valid (potentially inadmissible) inference in composite settings?

7.1 Necessary and sufficient conditions for valid (composite) inference

The following observations are straightforward and arguably well-known in some form or another, but are nevertheless useful to spell out formally in order to lay the path for the admissibility results.

Proposition 21 (Pointwise-to-composite validity)**.** *The following statements lay out necessary and sufficient conditions that connect validity in the 'pointwise' setting to the 'composite' setting.*

(1) $(\mathfrak{p}_t)$ *is* $\mathcal{Q}$*-valid if and only if* $\mathfrak{p}_t \geq \mathfrak{p}_t^{\mathtt{Q}}$, $\mathtt{Q}$*-a.s., for all* t *and* $\mathtt{Q} \in \mathcal{Q}$*, where* $(\mathfrak{p}_t^{\mathtt{Q}})$ *is some p-value for* $\mathtt{Q}$.

(2) $(\mathfrak{e}_t)$ *is* $\mathcal{Q}$*-safe if and only if* $\mathfrak{e}_t \leq \mathfrak{e}_t^{\mathtt{Q}}$, $\mathtt{Q}$*-a.s., for all* t *and* $\mathtt{Q} \in \mathcal{Q}$*, where* $(\mathfrak{e}_t^{\mathtt{Q}})$ *is some e-value for* $\mathtt{Q}$.

(3) (ψ_t) *is a* $(\mathcal{Q}, \alpha)$*-ST if and only if* $\psi_t \leq \psi_t^{\mathtt{Q}}$, $\mathtt{Q}$*-a.s., for all* t *and* $\mathtt{Q} \in \mathcal{Q}$*, where* $(\psi_t^{\mathtt{Q}})$ *is some* $(\mathtt{Q}, \alpha)$*-ST.*

Proof. We only prove (1), the other two assertions are argued analogously. Suppose we are given that $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid. Then, by definition, $\mathfrak{p}_t$ is $\mathtt{Q}$-valid for every $\mathtt{Q} \in \mathcal{Q}$; so choosing $(\mathfrak{p}_t^{\mathtt{Q}}) := (\mathfrak{p}_t)$ itself, we have proved the 'only if' direction. For the other direction, suppose for every $\mathtt{Q} \in \mathcal{Q}$ we are given a p-value $(\mathfrak{p}_t^{\mathtt{Q}})$ for $\mathtt{Q}$, and that $(\mathfrak{p}_t)$ satisfies $\mathfrak{p}_t \geq \mathfrak{p}_t^{\mathtt{Q}}$, $\mathtt{Q}$-almost surely, for all t. Let $\mathtt{Q}^* \in \mathcal{Q}$ be some true (arbitrary) data-generating distribution. We must argue that $(\mathfrak{p}_t)$ is $\mathtt{Q}^*$-valid. Indeed, for any $\alpha \in [0, 1]$, we have

$$\mathtt{Q}^*(\exists t \in \mathbb{N} : \mathfrak{p}_t \leq \alpha) \leq \mathtt{Q}^* \left(\exists t \in \mathbb{N} : \mathfrak{p}_t^{\mathtt{Q}^*} \leq \alpha \right) \leq \alpha,$$

where the first inequality follows because $\mathfrak{p}_t \geq \mathfrak{p}_t^{\mathtt{Q}^*}$, and the second inequality follows because $(\mathfrak{p}_t^{\mathtt{Q}^*})$ is $\mathtt{Q}^*$-valid by assumption. This concludes the proof. $\square$

The proposition provides a generic reduction from the composite setting to the pointwise setting for performing *valid* inference, but we can deduce a similar result for admissible inference, presented later. While Proposition 21 forms a useful building block, it makes no mention of martingales. Nevertheless, we now have the appropriate context in place to summarize some of our central results. Below, we use the notions of *essential supremum* and *essential infimum*, which we review in Appendix A.2, and note that we will need the additional restriction that $\mathcal{P}$ is locally dominated in order for these essential extrema to be well defined.

Corollary 22 (Pointwise supermartingales are sufficient for composite validity)**.** *Let $\mathcal{Q}$ be locally dominated. Then the following statements demonstrate how supermartingales suffice for sequential inference.*

(1) *If $\mathfrak{p}_t = \operatorname{ess\,sup}_{\mathtt{Q}\in\mathcal{Q}} 1 \wedge \inf_{s\le t} 1/N_s^{\mathtt{Q}}$, where $(N_t^{\mathtt{Q}})$ is upper bounded by a $\mathtt{Q}$-NSM, then $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid.*

(2) *If $\mathfrak{e}_t = \operatorname{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} N_t^{\mathtt{Q}}$, where $(N_t^{\mathtt{Q}})$ is upper bounded by a $\mathtt{Q}$-NSM, then $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe.*

(3) *If $\psi_t = \operatorname*{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} \sup_{s\le t} \mathbf{1}_{N_s^{\mathtt{Q}}\ge 1/\alpha}$, where $(N_t^{\mathtt{Q}})$ is upper bounded by a $\mathtt{Q}$-NSM, then (ψ_t) is a $(\mathcal{Q},\alpha)$-ST.*

Above, all $\mathtt{Q}$-NSMs start with initial expected value (at most) one, and $(N_t^{\mathtt{Q}})$ is assumed nonnegative.

This corollary follows directly from Proposition 21 and so its proof is omitted; see also Remark 14.

7.2 Necessary conditions for admissible (composite) inference

Not all constructions based on martingales are admissible. We next provide an analog to Proposition 21, now for admissibility.

Proposition 23 (Admissible composite tests must aggregate admissible pointwise tests)**.** *Let $\mathcal{Q}$ be locally dominated. The following statements show how composite admissible instruments must aggregate (some) pointwise admissible instruments.*

(1) *If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible, then $\mathfrak{p}_t = \operatorname{ess\,sup}_{\mathtt{Q}\in\mathcal{Q}} \mathfrak{p}_t^{\mathtt{Q}}$ for all t, where $(\mathfrak{p}_t^{\mathtt{Q}})$ is $\mathtt{Q}$-admissible.*

(2) *If $(\mathfrak{e}_t)$ is $\mathcal{Q}$-admissible, then $\mathfrak{e}_t = \operatorname{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} \mathfrak{e}_t^{\mathtt{Q}}$ for all t, where $(\mathfrak{e}_t^{\mathtt{Q}})$ is $\mathtt{Q}$-admissible.*

(3) *If (ψ_t) is $\mathcal{Q}$-admissible, then $\psi_t = \operatorname{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} \psi_t^{\mathtt{Q}}$ for all t, where $(\psi_t^{\mathtt{Q}})$ is $\mathtt{Q}$-admissible.*

Proof. Let $(\mathfrak{p}_t)$ denote a $\mathcal{Q}$-admissible p-value. For each $\mathtt{Q}\in\mathcal{Q}$, let $(\mathfrak{p}_t^{\mathtt{Q}})$ be a $\mathtt{Q}$-admissible p-value that dominates $(\mathfrak{p}_t)$. Such $(\mathfrak{p}_t^{\mathtt{Q}})$ exists thanks to Proposition 7. Let us now define $\mathfrak{p}_t' := \operatorname{ess\,sup}_{\mathtt{Q}\in\mathcal{Q}} \widetilde{\mathfrak{p}}_t^{\mathtt{Q}}$, which is $\mathcal{Q}$-valid thanks to Proposition 21(1). Clearly, we have $\mathfrak{p}_t' \ge \mathfrak{p}_t$ for all $t\in\mathbb{N}_0$. Since $(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible, we indeed have $\mathfrak{p}_t' = \mathfrak{p}_t$ for all $t\in\mathbb{N}_0$, yielding the assertion for p-values. The assertions for e-values and sequential tests are shown in the same manner. $\square$

The following corollary describes the restrictions that every admissible construction necessarily satisfies.

Corollary 24 (Pointwise martingales are necessary for composite admissibility)**.** *Let $\mathcal{Q}$ be locally dominated. Then the following statements demonstrate how martingales underpin all admissible constructions.*

(1) *If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible, then $(\mathfrak{p}_t)$ is nonincreasing and $\mathfrak{p}_t = \operatorname{ess\,sup}_{\mathtt{Q}\in\mathcal{Q}} \mathfrak{p}_t^{\mathtt{Q}}$ for all t, where $(\mathfrak{p}_t^{\mathtt{Q}})$ is a closed $\mathtt{Q}$-MM.*

(2) *If $(\mathfrak{e}_t)$ is $\mathcal{Q}$-admissible, then $\mathfrak{e}_t = \operatorname{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} M_t^{\mathtt{Q}}$ for all t, where $(M_t^{\mathtt{Q}})$ is a $\mathtt{Q}$-NM with $\mathtt{E}_{\mathtt{Q}}[M_0^{\mathtt{Q}}]=1$.*

(3) *If (ψ_t) is $\mathcal{Q}$-admissible, then $\psi_t = \operatorname{ess\,inf}_{\mathtt{Q}\in\mathcal{Q}} \sup_{s\le t} \mathbf{1}_{M_s^{\mathtt{Q}}\ge 1/\alpha}$, where $(M_t^{\mathtt{Q}})$ is a $\mathtt{Q}$-NM with $\mathtt{E}_{\mathtt{Q}}[M_0^{\mathtt{Q}}]=1$, $M_\infty^{\mathtt{Q}}\in\{0,1/\alpha\}$ and $(M_t^{\mathtt{Q}})$ has no overshoot at level $1/\alpha$.*

The aforementioned three statements are a direct consequence of Proposition 23 and Theorem 16.

7.3 Sufficient conditions for admissible (composite) inference

The next proposition argues that it suffices to consider only a subset of $\mathcal{Q}$ when constructing $\mathcal{Q}$-admissible p-values, e-values, or sequential tests. Note that we do not require $\mathcal{Q}$ to be locally dominated below.

Proposition 25 (Pointwise-to-composite admissibility). *Assume there exists a 'reference family' $(\mathtt{Q}_i)_{i\in I} \subset \mathcal{Q}$ such that for each t and $A \in \mathcal{F}_t$,*

if there exists $\mathtt{Q} \in \mathcal{Q}$ with $\mathtt{Q}(A) > 0$, then there exists $i \in I$ with $\mathtt{Q}_i(A) > 0$.

Then we have the following.

(1) *If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid and $(\mathfrak{p}_t)$ is $\mathtt{Q}_i$-admissible for each $i \in I$, then $(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible.*

(2) *If $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe and $(\mathfrak{e}_t)$ is $\mathtt{Q}_i$-admissible for each $i \in I$, then $(\mathfrak{e}_t)$ is $\mathcal{Q}$-admissible.*

(3) *If (ψ_t) is a $(\mathcal{Q}, \alpha)$-ST and (ψ_t) is $\mathtt{Q}_i$-admissible for each $i \in I$, then (ψ_t) is $\mathcal{Q}$-admissible.*

Proof. Let us only argue here the case of e-values. The other cases follow in exactly the same manner. Assume that there exists an e-value $(\mathfrak{e}'_t)$ for $\mathcal{Q}$ such that $\mathtt{Q}(\mathfrak{e}'_t \geq \mathfrak{e}_t) = 1$ for all $t \in \mathbb{N}_0$ and all $\mathtt{Q} \in \mathcal{Q}$, and that $\mathtt{Q}^*(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$ for some $t \in \mathbb{N}_0$, and some $\mathtt{Q}^* \in \mathcal{Q}$. By assumption, there exists some $i \in I$ such that $\mathtt{Q}_i(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$. Since $(\mathfrak{e}_t)$ is assumed to be $\mathtt{Q}_i$-admissible, we get a contradiction. □

Of course, two special cases are found at the extremes: when the reference family is a singleton, it means there is a common reference measure $\mathtt{R}$, and when the reference family is $\mathcal{Q}$ itself, the proposition is vacuous. The proposition is particularly useful in the first case; then to get an admissible e-value, for example, it suffices to construct a $\mathcal{Q}$-NSM (M_t) that is also a $\mathtt{Q}^*$-NM, thanks to Proposition 25(2) and Theorem 16(2). The following example demonstrates one such setting.

Example 26. *Recalling notation from Section 5.1, let $\mu^m \in \mathcal{G}^m$ denote the measure under which (X_t) is i.i.d. Gaussian with unit variance and mean m, and consider $\mathcal{Q} := \{\mu^m : m \leq 0\}$. Then $G_t := \exp(\sum_{s\leq t} X_s - t/2)$ is not a $\mathcal{Q}$-NM—it is a μ^0-NM when X_t is standard Gaussian, but is a μ^m-NSM for $m < 0$. Nevertheless, (G_t) is a $\mathcal{Q}$-admissible e-value. The reason is that (G_t), being a μ^0-NM, is immediately μ^0-admissible, and the singleton reference family $\{\mu^0\}$ satisfies the local absolute continuity condition required to invoke Proposition 25(2).*

Corollary 27 (Composite martingales are sufficient for composite admissibility). *Consider a general composite family $\mathcal{Q}$.*

(1) *$(\mathfrak{p}_t)$ is $\mathcal{Q}$-admissible if is a closed $\mathcal{Q}$-MM and $\inf_{t\in\mathbb{N}} \mathfrak{p}_t$ is $\mathtt{Q}$-uniformly distributed for every $\mathtt{Q} \in \mathcal{Q}$.*

(2) *$(\mathfrak{e}_t)$ is is $\mathcal{Q}$-admissible it is a $\mathcal{Q}$-NM with $\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_0] = 1$ for all $\mathtt{Q} \in \mathcal{Q}$.*

(3) *(ψ_t) is $\mathcal{Q}$-admissible if $\psi_t = \sup_{s\leq t} \mathbf{1}_{M_s \geq 1/\alpha}$, where (M_t) is a $\mathcal{Q}$-NM with $M_0 = 1$, $M_\infty \in \{0, 1/\alpha\}$, $\mathtt{Q}$-almost-surely, for every $\mathtt{Q} \in \mathcal{Q}$, and no overshoot at $1/\alpha$.*

This corollary is again a direct consequence of Proposition 25 and Theorem 16. Recall that as in Corollary 19, a sufficient condition for $(\mathfrak{p}_t)$ to be a closed $\mathtt{Q}$-MM and $\inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$ be $\mathtt{Q}$-uniformly distributed, for some fixed $\mathtt{Q} \in \mathcal{Q}$, is that the p-value has the representation $\mathfrak{p}_t = F^{\mathtt{Q}}(\inf_{s\leq t} 1/M_s^{\mathtt{Q}})$, where $(M_t^{\mathtt{Q}})$ is a $\mathtt{Q}$-NM and $\inf_{s\in\mathbb{N}_0} 1/M_s$ has an atomless distribution function $F^{\mathtt{Q}}$ under $\mathtt{Q}$.

As an immediate consequence of Corollary 27(2), recall the Gaussian example in Section 5.1, and consider testing if the underlying mean is zero. Since (G_t^0) is a $\mathcal{G}^0$-NM, it is also $\mathcal{G}^0$-safe, and hence a $\mathcal{G}^0$-admissible e-value when testing against, for example, $\mathcal{P} = \bigcup_{m\in\mathbb{R}} \mathcal{G}^m$.

7.4 Necessary and sufficient conditions for confidence sequences

Proposition 11 already shows that we can construct a confidene sequence by inverting a family of sequential tests. We now show that their admissibility is also tightly linked to those of the underlying tests.

Theorem 28. *Recall that $\mathcal{P}^\gamma := \{\mathtt{P} \in \mathcal{P} : \phi(\mathtt{P}) = \gamma\}$. If (ψ_t^γ) is an admissible $(\mathcal{P}^\gamma, \alpha)$-sequential test for each $\gamma \in \mathcal{Z}$ then $\{\gamma \in \mathcal{Z} : \psi_t^\gamma = 0\}$ is an admissible $(\phi, \mathcal{P}, \alpha)$-confidence sequence. Similarly, if $(\mathfrak{C}_t)$ is an admissible $(\phi, \mathcal{P}, \alpha)$-confidence sequence, then $\psi_t^\gamma := \mathbf{1}_{\gamma \notin \mathfrak{C}_t}$ yields an admissible $(\mathcal{P}^\gamma, \alpha)$-sequential test for each $\gamma \in \mathcal{Z}$, so that $\mathfrak{C}_t = \bigcup_{\gamma \in \mathcal{Z}} \{\gamma \in \mathcal{Z} : \psi_t^\gamma = 0\}$. As a result, we can infer the following.*

(1) *(Validity) If the process $(N_t^{\mathtt{P}})$ is upper bounded by a $\mathtt{P}$-NSM with initial expected value one, then $\mathfrak{C}_t := \bigcup_{\mathtt{P} \in \mathcal{P}} \{\phi(\mathtt{P}) : \sup_{s \leq t} N_s^{\mathtt{P}} < 1/\alpha\}$, is a $(\phi, \mathcal{P}, \alpha)$-CS.*

(2) *(Admissibility) If (M_t^γ) is a $\mathcal{P}^\gamma$-NM with $M_0^\gamma = 1$, $M_\infty^\gamma \in \{0, 1/\alpha\}$, $\mathtt{P}$-almost surely, for every $\mathtt{P} \in \mathcal{P}^\gamma$, and has no overshoot at $1/\alpha$, then $\mathfrak{C}_t := \{\gamma \in \mathcal{Z} : \sup_{s \leq t} M_s^\gamma < 1/\alpha = 0\}$ is $\mathcal{P}$-admissible.*

Assume now that $\mathcal{P}^\gamma$ is locally dominated for each γ, and that $(\mathfrak{C}_t)$ is $\mathcal{P}$-admissible. Then, for all $t \in \mathbb{N}_0$, we have $\mathfrak{C}_t = \bigcup_{\mathtt{P} \in \mathcal{P}} \{\phi(\mathtt{P}) : \psi_t^{\mathtt{P}} = 0\}$ where $(\psi_t^{\mathtt{P}})$ is $\mathtt{P}$-admissible. Moreover, we can write

$$\mathfrak{C}_t = \bigcup_{\mathtt{P} \in \mathcal{P}} \left\{ \phi(\mathtt{P}) : \sup_{s \leq t} M_s^{\mathtt{P}} < \frac{1}{\alpha} \right\},$$

where $(M_t^{\mathtt{P}})$ is a $\mathtt{P}$-NM that has no overshoot at level $1/\alpha$, with $M_0^{\mathtt{P}} = 1$ and $M_\infty^{\mathtt{P}} \in \{0, 1/\alpha\}$.

Proof. For the first statement of the theorem, note that $\mathfrak{C}_t = \{\gamma \in \mathcal{Z} : \psi_t^\gamma = 0\}$ yields a $(\phi, \mathcal{P}, \alpha)$-confidence sequence by Proposition 11. Suppose for contradiction that $(\mathfrak{C}_t')$ is another $(\phi, \mathcal{P}, \alpha)$-confidence sequence that witnesses the inadmissibility of $(\mathfrak{C}_t)$. Proposition 9(3) then yields a corresponding family $\{(\eta_t^\gamma)\}_{\gamma \in \mathcal{Z}}$ of sequential tests. The inadmissibility of $(\mathfrak{C}_t)$ then yields some $\gamma \in \mathcal{Z}$ such that (η_t^γ) strictly dominates (ψ_t^γ), a contradiction to the assumption that (ψ_t^γ) is admissible. The second statement follows in exactly the same way, again by an application of Propositions 11 and 9(3). Statements (1) and (2) are direct corollaries of combining the first part of the theorem with Corollary 22(3) and Corollary 27(3). Assume now that $\mathtt{P}^\gamma$ is locally dominated, for each γ, and that $(\mathfrak{C}_t)$ is $\mathcal{P}$-admissible. The statement then follows from Proposition 23(3) and Corollary 24(3). □

This and the previous section argued in detail that restricting our attention to constructions based on NMs (not NSMs!) does not hurt us: these are *universal* constructions. Indeed, if one is presented with a p-value, e-value, sequential test, or confidence sequence constructed in some arbitrary fashion, we show that one can always uncover a 'hidden' underlying NM, such that applying Ville's inequality or the optional stopping theorem yields an instrument that is at least as good as the original one.

8 A deeper investigation on admissibility

8.1 The gap between necessary and sufficient conditions (composite)

Recall that we were able to crisply summarize the necessary conditions for admissibility using martingales in Corollary 24 and sufficient conditions in Corollary 27. The following discussion probes at the gap between necessary and sufficient conditions for composite admissibility, in order to demonstrate that the gap is real. We begin with two instructive examples that demonstrate that the necessary conditions of Corollary 24 are not actually sufficient for admissibility.

Example 29 (Necessary conditions for Corollary 24(1) are not sufficient)**.** *Assume that $\mathcal{Q}$ is the set of probability measures under which X_1 is Bernoulli and $X_2 = X_3 = \ldots = 0$. Consider now*

$$\mathfrak{p}_1 := \mathbf{1}_{X_1=0} U + \mathbf{1}_{X_1=1} \sqrt{U}$$

and $\mathfrak{p}_2 = \mathfrak{p}_3 = \ldots = \mathfrak{p}_1$*. Note that* $\mathfrak{p}_1 \geq U$ *by construction and hence is valid. Then* $(\mathfrak{p}_t)$ *is an anytime valid p-value and satisfies Corollary 24(1) and Proposition 23(1). Here*

$$\mathfrak{p}_1^{\mathtt{Q}} = \mathfrak{p}_2^{\mathtt{Q}} = \mathfrak{p}_3^{\mathtt{Q}} = \ldots = F_{\mathtt{Q}}(\mathfrak{p}_1) = (1-q)\mathfrak{p}_1 + q\mathfrak{p}_1^2,$$

where $q := \mathtt{Q}(X_1 = 1) \in [0,1]$*, for each* $\mathtt{Q} \in \mathcal{Q}$ *and* $F_{\mathtt{Q}}$ *is the* $\mathtt{Q}$*-distribution function of* $\mathfrak{p}_1$*. Indeed, by definition* $\mathfrak{p}_1^{\mathtt{Q}}$ *is* $\mathtt{Q}$*-uniform and considering small* q*'s it is easy to see that* $\mathfrak{p}_1 = \operatorname{ess\,sup}_{\mathtt{Q}\in\mathcal{Q}} \mathfrak{p}_1^{\mathtt{Q}}$*. However,* $(\mathfrak{p}_t)$ *is indeed inadmissible as it is dominated by* (U)*.*

Example 30 (Necessary conditions for Corollary 24(2) are not sufficient)**.** *Assume that* $X_t = (Y_t, Z_t)$*, where* Z_t *is some 'nuisance parameter', and let* $\mathcal{Q} := \{\mathtt{Q}_+, \mathtt{Q}_-\}$ *consist of two measures: under both* $\mathtt{Q}_+$ *and* $\mathtt{Q}_-$*, the first components* Y_t *of the observation process are i.i.d. standard Gaussian. The two measures may be different on the 'nuisance parameter'* Z_t*. Following the Gaussian example in Section 5.1, define*

$$M_t^+ := \exp\left(\sum_{s\leq t} Y_s - \frac{t}{2}\right) \text{ and } M_t^- := \exp\left(-\sum_{s\leq t} Y_s - \frac{t}{2}\right).$$

Then (M_t^+) *is a* $\mathtt{Q}_+$*-NM, and* (M_t^-) *is a* $\mathtt{Q}_-$*-NM. Indeed, both processes are martingales under both measures. The corresponding e-value for* $\mathcal{Q}$ *corresponding to Proposition 23(2) is*

$$\mathfrak{e}_t = \min\{M_t^+, M_t^-\} = \exp\Big(-\Big|\sum_{s\leq t} Y_s\Big| - \frac{t}{2}\Big).$$

Then $(\mathfrak{e}_t)$ *is the minimum of two martingales, both under* $\mathtt{Q}_+$ *and* $\mathtt{Q}_-$*, hence a supermartingale under both measures. Moreover, the martingale part of* $(\mathfrak{e}_t)$ *is the same under both measures (as both measures agree on the filtration generated by the process* (Y_t)*) and strictly dominates* $(\mathfrak{e}_t)$ *under both measures. This proves that* $(\mathfrak{e}_t)$ *is inadmissible since it is strictly dominated by a* $\mathcal{Q}$*-NM.*

8.2 Composite Ville-like anti-concentration bounds

Next, we discuss anti-concentration results, which are somewhat surprisingly insufficient for admissibility. Since Ville's inequality has been often used in this paper to demonstrate validity, one would hope that if Ville's inequality is 'essentially' tight (it holds with 'almost' equality) then the corresponding inferential instruments may be close to admissible. We examine this angle next. Below, we derive an anti-concentration (lower) bound to complement the upper bound of Ville's inequality.

Lemma 31 (Anti-concentration for pointwise NMs)**.** *Let* (M_t) *be a* $\mathtt{Q}$*–NM with* $\mathtt{E}_{\mathtt{Q}}[M_0] = 1$*, so that the 'multiplicative increment'* $Y_t := M_t/M_{t-1}$ *(with* $0/0 := 1$*) has unit mean. Assume that the aggregate empirical variance of* (Y_t) *is* $\mathtt{Q}$*-almost surely infinite, i.e.,*

$$\mathtt{Q}\left(\sum_{t\in\mathbb{N}} (Y_t - 1)^2 = \infty\right) = 1. \tag{14}$$

Then $M_\infty = 0$*,* $\mathtt{Q}$*-almost surely. Fix now some* $\varepsilon > 0$*. Assume that for each* $t \in \mathbb{N}_0$ *the multiplicative increment* Y_t *with* $Y_0 := M_0$ *satisfies a tail condition, namely for each* $\mathcal{F}_{t-1}$*-measurable random variables* $\beta > 1$*, where* $\mathcal{F}_{-1} := \{\emptyset, \Omega\}$*, we have*

$$\mathtt{E}_{\mathtt{Q}}\left[Y_t \,|\mathcal{F}_{t-1}, Y_t \geq \beta\right] \leq \beta(1+\varepsilon). \tag{15}$$

Then, for any $\alpha \in (0,1]$ *we have*

$$\alpha \;\geq\; \mathtt{Q}\left(\sup_{t\in\mathbb{N}_0} M_t \geq \frac{1}{\alpha}\right) \;\geq\; \frac{\alpha}{1+\varepsilon}.$$

The proof is in Appendix B. We observe that (15) is satisfied, for example, when $\mathtt{Q}(Y_t \leq 1+\varepsilon) = 1$ for all $t \in \mathbb{N}_0$. We note that (14) is easily satisfied when $\mathtt{P}$ is a product measure and thus (Y_t) is a sequence of i.i.d. random variables, as long as $\mathtt{Q}(Y_t \neq 1) > 0$. We can now extend the above pointwise result to the composite setting, and once more this result is of independent interest.

Corollary 32 (Anti-concentration for bounded composite NMs)**.** *Consider a family $\mathcal{Q}$ of probability measures and a $\mathcal{Q}$-NM (M_t) with $M_0 = 1$. Define $Y_t := M_t/M_{t-1}$ (with $0/0 := 1$) and assume that for each $\varepsilon > 0$ there exists some $\mathtt{Q} \in \mathcal{Q}$ such that conditions (14) and (15) hold for each $t \in \mathbb{N}$ and $\mathcal{F}_{t-1}$-measurable random variable $\beta > 1$. Then*

$$\sup_{\mathtt{Q}\in\mathcal{Q}} \mathtt{Q}\left(\sup_{t\in\mathbb{N}_0} M_t \geq \frac{1}{\alpha}\right) = \alpha, \quad \textit{for any } \alpha \in (0,1].$$

In words, the above result establishes rather simple and interpretable sufficient conditions under which $\mathfrak{p}_t := \inf_{s\leq t} 1/M_s$ uses up all its type-I error budget, meaning that at least in a worst-case sense, Ville's inequality did not lead to a conservative test.

Unfortunately, Example 44 shows, in the context of conditionally symmetric distributions (see Section 9), that even under the assumptions of the previous corollary, such $(\mathfrak{p}_t)$ does not need to be admissible. This further demonstrates the subtleties of establishing sufficient conditions for admissibility. Nevertheless, Corollary 32 is of independent interest; but the fairly intuitive condition it yields does usually not suffice for admissibility.

8.3 The role of initial randomization in determining admissibility

Throughout this subsection, let us consider $\mathcal{Q} = \{\mathtt{Q}\}$.

The admissibility of p-values is subtle and randomization appears to play a key role in enabling admissible constructions. The key difficulty is in dealing with atomic limiting distributions, and we delve more into this topic here with several examples.

In Corollary 19, $\inf_{s\in\mathbb{N}_0} 1/M_s$ is assumed to have an atomless distribution function F. Initial randomization turns out to be necessary for this to hold. To formalize this claim, consider a martingale (M_t). We now argue the following fact:

$$\text{If } M_0 = 1,\ \mathtt{Q}\text{-almost surely, then } \sup_{t\in\mathbb{N}_0} M_t \text{ has an atom at one under } \mathtt{Q}.$$

The proof is simple, so we present it immediately. Define $Y_t := M_t/M_{t-1}$ with $0/0 := 1$ Without loss of generality, we may assume that $\mathtt{Q}(Y_1 \neq 1) > 0$. Since $\mathtt{E}_{\mathtt{Q}}[Y_1] = 1$, there exists some $\eta > 0$ such that $\mathtt{Q}(Y_1 \leq 1-\eta) > \eta$. On the event $\{Y_1 \leq 1-\eta\}$, the conditional version of Ville's inequality (8) yields that $\mathtt{Q}(\sup_{t\in\mathbb{N}} M_t \geq 1|\mathcal{F}_1) \leq M_1 \leq 1-\eta$. Hence on this event we have $\mathtt{Q}(\sup_{t\in\mathbb{N}} M_t < 1|\mathcal{F}_1) \geq \eta$, yielding the unconditional bound $\mathtt{Q}(\sup_{t\in\mathbb{N}} M_t < 1) \geq \eta^2$. This then gives $\mathtt{Q}(\sup_{t\in\mathbb{N}_0} M_t = 1) \geq \eta^2$. Hence $\sup_{t\in\mathbb{N}_0} M_t$ has an atom at one, and so does the induced p-value, completing the proof of the aforementioned fact.

In contrast, if we consider the martingale (M'_t) with randomized initial value $M'_t := M_t + \varepsilon U$, where $\varepsilon > 0$, and recall that U is the (independent) $\mathcal{F}_0$-measurable $[0,1]$-uniformly distributed random variable, then $\sup_{t\in\mathbb{N}_0} M'_t = \sup_{t\in\mathbb{N}_0} M_t + \varepsilon U$ has a density since it is the convolution of $\sup_{t\in\mathbb{N}_0} M_t$ with a random variable that has a density.

Let us consider for the moment a p-value constructed as $\mathfrak{p}_t := F(\inf_{s\leq t} 1/M_s)$, where (M_t) is a $\mathtt{Q}$-martingale with $M_0 = 1$ and F is the distribution function of $\inf_{s\in\mathbb{N}_0} 1/M_s$. (Note that such a p-value always dominates $(\inf_{s\leq t} 1/M_s)$.) Then $(\mathfrak{p}_t)$ is *always* inadmissible. To see this, define $\mathfrak{p}_\infty := \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$ and $\delta := \mathtt{Q}(\mathfrak{p}_\infty = 1) > 0$, where the inequality follows from the fact argued above. Moreover, define the conditional distribution function G by $[0,1] \ni u \mapsto \mathtt{Q}(U \leq u|\mathfrak{p}_\infty = 1)$. Let us then define $\mathfrak{p}'_t := \mathfrak{p}_t \wedge (1-\delta+\delta G(U))$. Then clearly $(\mathfrak{p}'_t)$ strictly dominates $(\mathfrak{p}_t)$. Moreover, $(\mathfrak{p}'_t)$ is a p-value since $\mathfrak{p}'_\infty := \inf_{t\in\mathbb{N}_0} \mathfrak{p}'_t = \mathfrak{p}_\infty \wedge (1-\delta+\delta G(U))$ stochastically dominates a uniform. Indeed, for $\alpha \in (0,1-\delta)$ we have $\mathtt{Q}(\mathfrak{p}'_\infty \le \alpha) = \mathtt{Q}(\mathfrak{p}_\infty \le \alpha) \le \alpha$ and for $\alpha \in [1-\delta, 1]$ we get

$$\begin{aligned}\mathtt{Q}(\mathfrak{p}'_\infty \le \alpha) &= \mathtt{Q}(\mathfrak{p}_\infty \le 1-\delta) + \mathtt{Q}(\mathfrak{p}_\infty = 1, 1-\delta+\delta G(U) \le \alpha) \\ &= 1-\delta+\delta\mathtt{Q}(\delta G(U) \le \alpha-(1-\delta) \mid \mathfrak{p}_\infty = 1) \\ &= 1-\delta+\delta\frac{\alpha-(1-\delta)}{\delta} = \alpha,\end{aligned}$$

where we used that $G(U)$ is uniformly distributed under the conditional measure $\mathtt{Q}(\cdot|\mathfrak{p}_\infty = 1)$.

We note above that atoms at one are 'obviously' undesirable for p-values. Quite surprisingly, *there do exist admissible anytime p-values with atomic limiting distributions* (where the atoms are not at one); see Example 43. In that example, we have $\mathtt{Q}(\inf_{t\in\mathbb{N}_0} \mathfrak{p}_\infty < 1/2) = 0$, and $(\mathfrak{p}_t)$ is independent of the randomization device U; nevertheless it is *impossible* to 'randomize' the atom.

To end the discussion about atoms in the context of p-values, we remark that atomic limiting distributions occur more often in discrete time than in continuous time. For example, if $(B_t)_{t\in[0,\infty)}$ is a standard Brownian motion, then $(\exp(B_t - t/2))_{t\in[0,\infty)}$ is a martingale, and $\inf_{t\ge 0} 1/\exp(B_t - t/2)$ is exactly $[0,1]$-uniformly distributed. However, the corresponding standard Gaussian NM from (10) has that $\inf_{t\in\mathbb{N}_0} 1/G^0_t$ is atomic when (X_t) under $\mathtt{Q}$ follow the law of i.i.d. standard Gaussians.

In sharp contrast, initial randomization causes sequential tests based on e-values to become inadmissible. Indeed, if the jumps of $(\mathfrak{e}_t)$ are continuous with positive probability then the corresponding sequential test is not admissible for *any* $\alpha \in (0,1)$ due to overshoot. Only e-values that have atomic jumps can possibly lead to admissible tests; however, any such e-value cannot lead to an admissible test for every α (it will overshoot for some and not for others). Example 37 in the next section derives an admissible sequential test for (composite) symmetric distributions.

9 Admissible inference for symmetric distributions

Considering the results of this paper presented thus far, we demonstrate how they may inform practice. At a high level, this section constructs (using different NMs) admissible version of all four instruments studied in this paper for the class of conditionally symmetric distributions.

We return to the example from Section 5.2 where we had presented a rather elegant, and intuitive, exponential NSM for distributions that yield conditionally symmetric observations. However, the results following the example showed that the tests or confidence sequences stemming from it are inadmissible, since all admissible constructions must use NMs. We will construct such NMs, which appear to be new to the best of our knowledge (but we would not be surprised if they have been discovered before). Recalling the notation from Section 5.2, let $\mathcal{S} := \mathcal{S}^0$ be the set of laws such that X_t conditional on $\mathcal{F}_{t-1}$ is symmetric around zero. We will demonstrate here that inference based on (S^0_t) defined in (12) is inadmissible by explicitly constructing procedures that dominate it.

We note that $\mathcal{S}$ is *not* locally dominated. Indeed, just consider $\mathtt{P} \in \mathcal{S}$ of the form $\mathtt{P} = \mathtt{U} \times \mu^\infty$ (where $\mathtt{U}$ denotes the uniform measure), and note that there are uncountably many mutually singular choices for μ; take for instance $\mu = (\delta_x + \delta_{-x})/2$ for $x \in \mathbb{R}$. Here δ_x denotes the Dirac measure at x. Nevertheless, despite the lack of a reference measure, it is still possible to construct a family of $\mathcal{S}$-NMs, and thus admissible e-values for $\mathcal{S}$. Indeed, we have the following proposition.

Proposition 33. *An adapted process (M_t) with M_0 bounded and nonnegative is an $\mathcal{S}$-NM if and only if $Y_t := M_t/M_{t-1}$ (with $0/0 := 1$) is of the form $Y_t = f_t(X_t)$, where (f_t) is a nonnegative predictable function such that $f_t - 1$ is odd, or equivalently, $f_t(x) + f_t(-x) = 2$ for all $x \in \mathbb{R}$. Moreover, if $M_0 = 1$ then (M_t) is an $\mathcal{S}$-admissible e-value by Corollary 27(2).*

The proof is in Section B. A similar characterization as above also holds for any $\mathcal{S}$-NSM; in the notation of Proposition 33, (M_t) is an $\mathcal{S}$-NSM if and only if $Y_t = f_t(X_t)$ where (f_t) is a nonnegative predictable function such that

$$f_t(x) + f_t(-x) \le 2, \qquad t \in \mathbb{N}. \tag{16}$$

Moreover, an $\mathcal{S}$-NSM (M_t) can be converted to an $\mathcal{S}$-NM $(\widetilde{M_t})$ with $\widetilde{M}_t = \prod_{s\le t} \widetilde{f}_s(X_s)$ by the following mirroring operation:

$$\widetilde{f}_t(x) = \begin{cases} f_t(x), & f_t(x) \ge f_t(-x); \\ 2 - f_t(-x), & f_t(x) < f_t(-x). \end{cases}$$

Indeed, we get that $\widetilde{f}_t \ge f_t$ and that equality holds in (16) with f_t replaced by $\widetilde{f}_t$.

Proposition 33 demonstrates how to construct admissible e-values for symmetry, and we give two instantiations that we have found (subjectively) elegant. Let h be an odd function and consider

$$f(x) = 1 + \arctan h(x) \qquad \text{or} \qquad f(x) = 1 + \cos h(x).$$

Then, $(\prod_{s\le t} f(X_s))$ is an $\mathcal{S}$–NM and thus an e-value for $\mathcal{S}$ which is admissible by Proposition 33.

Finally, we return to the exponential $\mathcal{S}$-NSM from [4] from Section 5.2, showing that it is leads to an inadmissible e-value for $\mathcal{S}$, and improving it to an admissible one by converting the NSM to an NM.

Example 34. *Let $\mathcal{P} \supset \mathcal{S}$ and recall from Section 5.2 that the process $(S_t) := (S_t^0)$ given by*

$$S_t = \prod_{s\le t} g(X_s), \qquad \textit{where} \qquad g(x) := \exp\left(x - \frac{x^2}{2}\right),$$

is an $\mathcal{S}$-NSM. Further, (S_t) is not a martingale unless $X_t = 0$ for all $t \in \mathbb{N}$, and the corresponding $\mathcal{S}$-safe e-value is inadmissible due to the following argument. Define

$$f(x) = \begin{cases} g(x), & x \ge 0; \\ 2 - g(-x), & x < 0. \end{cases}$$

Then $f \ge g$ with equality only if $x = 0$. Further, $f(-x) - 1 = 1 - f(x)$ and finally f is nonnegative since $g \le e^{1/2} \approx 1.65$; hence $(\prod_{s\le t} f(X_s))$ is an $\mathcal{S}$-NM by Proposition 33. This also yields that the corresponding e-value is admissible. Even though the original $\mathcal{S}$-NSM is inadmissible, we recognize the aesthetic and analytical advantage in having a simple exponential formula.

Let us now illustrate that the p-values corresponding to the $\mathcal{S}$-NMs fully utilize the available Type-I error budget in the sense of the next proposition.

Proposition 35. *Consider a nonnegative function f, continuous and strictly monotone at zero and such that $f - 1$ is odd. Then $M_t := \prod_{s\le t} f(X_s)$ is an $\mathcal{S}$-NM by Proposition 33 and we have*

$$\sup_{\mathtt{Q}\in\mathcal{S}} \mathtt{Q}\left(\sup_{t\in\mathbb{N}_0} M_t \ge \frac{1}{\alpha}\right) = \alpha, \quad \textit{for any } \alpha \in (0,1]. \tag{17}$$

Thus, defining $\mathfrak{p}_t := \inf_{s\le t} 1/M_s$, we have that $(\mathfrak{p}_t)$ is a p-value for $\mathcal{S}$ that satisfies

$$\sup_{\mathtt{Q}\in\mathcal{S},\tau\ge 0} \mathtt{Q}\left(\mathfrak{p}_\tau \le \alpha\right) = \alpha, \quad \textit{for any } \alpha \in (0,1]. \tag{18}$$

The proof is in Section B. An analogous result to the above proposition is known for the class of subGaussian distributions [11, Proposition 4], but had only been conjectured for other nonparametric classes like $\mathcal{S}$. Unfortunately, not every p-value for $\mathcal{S}$ constructed as (18) above is admissible; see Example 44 in Appendix C.

Finally, let us construct an $\mathcal{S}$-admissible p-value in the next example.

Example 36. *Define the following subset of symmetric distributions:*

$$\widetilde{\mathcal{S}} := \left\{ \mathtt{P} \in \mathcal{S} : \mathtt{P}\left(\sum_{t\in\mathbb{N}} \mathbf{1}_{X_t\neq 0} = \infty\right) = 1 \right\}. \tag{19}$$

Next, we define the process $(\mathfrak{p}_t)$ *as* $\mathfrak{p}_0 := 1$ *and*

$$\mathfrak{p}_t := 1 - \sum_{s \leq t} \frac{1}{2^{N_s}} \mathbf{1}_{X_s > 0} = \mathfrak{p}_{t-1} - \frac{1}{2^{N_t}} \mathbf{1}_{X_t > 0}, \qquad \textit{where} \qquad N_s := \sum_{i \leq s} \mathbf{1}_{X_i \neq 0}.$$

Then it can be checked that $(\mathfrak{p}_t)$ *is a closed* $\mathcal{S}$*-MM. Moreover,* $\inf_{t \in \mathbb{N}_0} \mathfrak{p}_t$ *is* $\mathtt{Q}$*-uniform for each* $\mathtt{Q} \in \widetilde{\mathcal{S}}$*. Assume for the moment that* $\mathtt{Q} \in \mathcal{S} \setminus \widetilde{\mathcal{S}}$ *can be locally dominated by some* $\mathtt{Q}' \in \widetilde{\mathcal{S}}$*. Proposition 25(1) and Corollary 27(1) then yield that* $(\mathfrak{p}_t)$ *is* $\mathcal{S}$*-admissible.*

Let us now fix some $\mathtt{Q} \in \mathcal{S} \setminus \widetilde{\mathcal{S}}$ *and argue that indeed it can be locally dominated by some* $\mathtt{Q}' \in \widetilde{\mathcal{S}}$*. To do so, define the measure* $\mu := 1/2(\delta_1 + \delta_{-1})$ *with* δ_x *denoting again the Dirac measure at* $x \in \mathbb{R}$*. Moreover, let* $\mathtt{H}$ *denote the law of a Poisson random variable with expectation one. On the appropriate canonical space, consider the measure* $\mathtt{Q} \times \mu^\infty \times \mathtt{H}$*, and write* $U, (X_t), (Y_t), H$ *for the canonical random variables. To summarize,* U *is uniform,* (X_t) *is our original conditionally symmetric sequence,* (Y_t) *is an independent sequence of Rademacher random variables, and* H *is Poisson. Define a new sequence* (X'_t) *by* $X'_t := X_t \mathbf{1}_{H > t} + Y_t \mathbf{1}_{H \leq t}$ *for all* $t \in \mathbb{N}$ *and let* $\mathtt{Q}'$ *denote the measure induced by* U *and* (X'_t)*. Then* $\mathtt{Q}' \in \widetilde{\mathcal{S}}$ *and it can be checked that* $\mathtt{Q}'$ *locally dominates* $\mathtt{Q}$*. This completes the proof of our initial claim.*

As a final observation, observe that if we replace $\mathcal{S}$ *by the superset of probability measures for which the conditional laws of* X_t *have median zero, all statements still hold.*

This section has now developed admissible e-values and p-values for testing for symmetry, and we move next to admissible sequential tests (and thus confidence sequences).

Example 37. *Consider the null* $H_0 : \mathtt{P} \in \mathcal{S}$ *from definition (11), and let* $\alpha = 0.05$ *so that* $1/\alpha = 20$*. Consider the process* (M_t) *defined as follows. Let* $M_0 = 1$*, and zero is an absorbing state, meaning that if* $M_t = 0$*, then the process stays at zero from then on. If* M_t *is nonzero, then define* $M_{t+1} := M_t + \mathrm{sign}(\mathrm{X_t}) \mathbf{1}_{\mathrm{X_t} \neq 0}$*. It is easy to check that* (M_t) *is an* $\mathcal{S}$*-NM. Define* τ *as the first time* M_t *reaches* 20*. Then,* $M_\tau = 20$ *on* $\{\tau < \infty\}$*, and also* $M_\infty = 0$*,* $\mathtt{Q}$*-almost surely, for each* $\mathtt{Q} \in \widetilde{\mathcal{S}}$*, as in (19). Invoking Proposition 25(3) and Corollary 27(3) as in the previous example yields that* $(\mathbf{1}_{M_t \geq 20})$ *is* $(\mathcal{S}, 0.05)$*-admissible. More generally,* $(\mathbf{1}_{M_t \geq 1/\alpha})$ *is* $(\mathcal{S}, \alpha)$*-admissible whenever* $1/\alpha \in \mathbb{N}$*.*

Of course, there is nothing special about 0.05 and 20; for any other α, the process (M_t) can be altered accordingly to yield an admissible test for that α. The above process (M_t) also delivers admissible tests for several subsets of $\mathcal{S}$, for example for if we restricted to only Gaussian distributions with any variance. This is interesting because admissibility is generally not subset-proof or superset-proof, but above we have a single process (M_t) that yields admissible e-values and sequential tests for a variety of subsets of $\mathcal{S}$.

Continuing from Example 37 and using Theorem 28, we can construct an admissible level α test for any $(\mathcal{S}^m)_{m \in \mathbb{R}}$, so the above construction yields an admissible confidence sequence for the center of symmetry.

Thus, we have accomplished our goal of constructing admissible versions of all four instruments for sequential inference, for a composite nonparametric class of distributions.

10 Summary

The central contribution of this work is to identify the central role of nonnegative martingales in anytime-valid sequential inference. As a by-product, we have added several modern mathematical techniques to the toolkit of the methodologist who wishes to design statistically efficient methods for inference at arbitrary stopping times. We end with a few comments.

It is apparent to us that some of our analysis may have been simpler in continuous time. Indeed, some of the difficulty in constructing admissible sequential tests using e-values arises from the 'overshoot', while the difficulty in designing admissible p-values arises because we do not observe the process 'in between' the fixed times and thus the running infimum is not *exactly* uniformly distributed in the limit. Several of these problems go away with continuous time/path martingales. However, continuous path martingales only represent large-scale approximations of most actual experimental setups, which typically involve discrete events. The accuracy of these approximations would have to be assessed, especially outside very high-frequency settings like finance, and it may not be clear how to do so. We believe that the additional effort to understand admissibility in the discrete time setup was fruitful.

Following the literature, our sequential inference tools were only required to have marginal guarantees, and not conditional ones. To pick one example, we required that for each $\mathtt{Q} \in \mathcal{Q}$, an e-value must satisfy $\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_\tau] \leq 1$ at arbitrary stopping times τ, but it need not satisfy $\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_t|\mathcal{F}_s] \leq \mathfrak{e}_s$. This gap between conditional and marginal guarantees is paramount: it allows for the construction of e-values that are not simply $\mathcal{Q}$-NMs, because in several settings of interest, one can show that the only nontrivial $\mathcal{Q}$-NM is the constant process that equals one at all times, but nontrivial e-values with power to detect deviations from $\mathcal{Q}$ can still be constructed. We explore these connections further by using a structural notion called 'fork-convexity' in a separate work.

Finally, the paper provides a rather general treatment of the inferential tools and problem settings. However, perhaps additional insights could be gained when $\mathcal{P}$ or $\mathcal{Q}$ have special structure, or when we pay attention to particular classes of stopping times, or restrict ourselves to a bounded horizon; these may all be promising directions to explore. Finally, while we take a step forward in relating the various concepts used for sequential inference, and present a thorough analysis of their validity and admissibility, the question of optimality is unaddressed by our work. Of course, this usually needs to be studied by specifying appropriate alternatives, and introducing metrics by which to judge optimality (such as the GROW criterion of Grünwald et al. [9]), and so we leave such considerations for future work.

Acknowledgments

The authors are thankful to the organizers of the International Seminar on Selective Inference, which stimulated conversations that led to this paper. AR acknowledges NSF DMS grant 1916320.

11 References

[1] E. N. Barron, P. Cardaliaguet, and R. Jensen. Conditional essential suprema with applications. *Appl. Math. Optim.*, 48:229–253, 2003. 9

[2] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. On the history of martingales in the study of randomness. *Electronic Journal for History of Probability and Statistics*, 5, 2009. 8

[3] D. A. Darling and Herbert Robbins. Confidence Sequences for Mean, Variance, and Median. *Proceedings of the National Academy of Sciences*, 58(1):66–68, July 1967. ISSN 0027-8424, 1091-6490. 3, 5, 14

[4] Victor H. de la Peña. A General Class of Exponential Inequalities for Martingales and Ratios. *The Annals of Probability*, 27(1):537–564, January 1999. ISSN 0091-1798, 2168-894X. 14, 15, 25

[5] William D Dupont. Sequential stopping rules and sequentially adjusted p-values: Does one require the other? *Controlled Clinical Trials*, 4(1-2):3–10, 1983. 3

[6] Rick Durrett. *Probability: Theory and Examples.* Cambridge University Press, 5a edition, 2017. 8

[7] Bradley Efron. Student's t-Test Under Symmetry Conditions. *Journal of the American Statistical Association*, 64(328):1278–1302, 1969. ISSN 0162-1459. 14

[8] Hans Föllmer and Alexander Schied. *Stochastic Finance*, volume 27 of *De Gruyter Studies in Mathematics.* Walter de Gruyter & Co., Berlin, extended edition, 2004. 29

[9] Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe testing. *arXiv:1906.07801*, June 2019. 3, 27

[10] Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020. 3, 14

[11] Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, forthcoming, 2020. 3, 4, 5, 25

[12] Ramesh Johari, Leo Pekelis, and David J. Walsh. Always valid inference: Bringing sequential analysis to A/B testing. *arXiv preprint arXiv:1512.04922*, 2015. 3

[13] Alisa Kirichenko and Peter Grünwald. Minimax rates without the fixed sample size assumption. *arXiv preprint arXiv:2006.11170*, 2020. 31

[14] Tze Leung Lai. On Confidence Sequences. *The Annals of Statistics*, 4(2):265–280, March 1976. ISSN 0090-5364, 2168-8966. 3

[15] Martin Larsson. Conditional infimum and recovery of monotone processes. *arXiv preprint arXiv:1802.08628*, 2018. 9

[16] Martin Larsson and Johannes Ruf. Convergence of local supermartingales. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, forthcoming, 2020. 33

[17] Luigi Pace and Alessandra Salvan. Likelihood, replicability and Robbins' confidence sequences. *International Statistical Review*, 2019. 3

[18] Herbert Robbins and David Siegmund. The Expected Sample Size of Some Tests of Power One. *The Annals of Statistics*, 2(3):415–436, May 1974. ISSN 0090-5364, 2168-8966. 3

[19] Glenn Shafer. The language of betting as a strategy for statistical and scientific communication. *Journal of the Royal Statistical Society, Series A*, forthcoming discussion paper, 2020. 3, 8

[20] Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk. Test Martingales, Bayes Factors and p-Values. *Statistical Science*, 26(1):84–101, February 2011. ISSN 0883-4237, 2168-8745. 3, 8, 13, 16, 31

[21] J Ville. *Étude Critique de la Notion de Collectif.* Gauthier-Villars, Paris, 1939. 8

[22] Vladimir Vovk. Testing randomness. *arXiv:1906.09256*, 2019. 30

[23] Vladimir Vovk. Non-algorithmic theory of randomness. *Fields of Logic and Computation III. Lecture Notes in Computer Science*, 12180, 2020. 3

[24] Vladimir Vovk and Ruodu Wang. Combining e-values and p-values. *arXiv preprint arXiv:1912.06116*, 2019. 3

[25] Abraham Wald. Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*, 16(2): 117–186, 1945. 3, 8

[26] Abraham Wald. *Sequential Analysis*. John Wiley & Sons, New York, 1947. 3, 8

[27] Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 2020. ISSN 0027-8424. 3

A Additional technical concepts and definitions

A.1 Reference measures and local absolute continuity

Consider a probability space with a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$. Let $\mathtt{R}$ be a particular probability measure on $\mathcal{F}_\infty$; we think of $\mathtt{R}$ as a *reference measure*. We now explain the concept of local domination and how it allows us to unambiguously define conditional expectations.

- $\mathtt{P}$ is called *locally absolutely continuous with respect to* $\mathtt{R}$ (or *locally dominated by* $\mathtt{R}$), if $\mathtt{P}_t \ll \mathtt{R}_t$ for all $t \in \mathbb{N}$. We write this $\mathtt{P} \ll_{\mathrm{loc}} \mathtt{R}$. More explicitly, this means that

$$\mathtt{R}(A) = 0 \quad \Rightarrow \quad \mathtt{P}(A) = 0, \qquad \text{for any } A \in \mathcal{F}_t \text{ and } t \in \mathbb{N}.$$

  Local absolute continuity does *not* imply that $\mathtt{P} \ll \mathtt{R}$. However, it does imply that $\mathtt{P}_\tau \ll \mathtt{R}_\tau$ for any finite (but possibly unbounded) stopping time τ. Indeed, if $A \in \mathcal{F}_\tau$ and $\mathtt{R}(A) = 0$, then $A \cap \{\tau \le t\} \in \mathcal{F}_t$ for all t, and hence $\mathtt{P}(A) = \lim_{t\to\infty} \mathtt{P}(A \cap \{\tau \le t\}) = 0$.

- A set $\mathcal{P}$ of probability measures on $\mathcal{F}_\infty$ is called locally dominated by $\mathtt{R}$ if every element of $\mathcal{P}$ is locally dominated by $\mathtt{R}$.

- Any $\mathtt{P} \ll_{\mathrm{loc}} \mathtt{R}$ has an associated *density process*, namely the $\mathtt{R}$-martingale (Z_t) given by $Z_t := \mathrm{d}\mathtt{P}_t/\mathrm{d}\mathtt{R}_t$. Being a nonnegative martingale, once Z_t reaches zero it stays there. Thus with the convention $0/0 := 1$, ratios Z_τ/Z_t are well-defined for any $t \in \mathbb{N}$ and any finite stopping time $\tau \ge t$. Note that each Z_t is defined up to $\mathtt{R}$-nullsets, and therefore also up to $\mathtt{P}$-nullsets.

- If $\mathtt{P} \ll_{\mathrm{loc}} \mathtt{R}$ has density process (Z_t), the following 'Bayes formula' holds: for any $t \in \mathbb{N}$, any finite stopping times τ, and any nonnegative $\mathcal{F}_\tau$-measurable random variable Y, one has

$$\mathtt{E}_{\mathtt{P}}[Y \mid \mathcal{F}_t] = \mathtt{E}_{\mathtt{R}}\left[\frac{Z_\tau}{Z_t} Y \,\middle|\, \mathcal{F}_t\right], \ \mathtt{P}\text{-almost surely.}$$

  The right-hand side is uniquely defined $\mathtt{R}$-almost surely (not just $\mathtt{P}$-almost surely), and therefore provides a 'canonical' version of $\mathtt{E}_{\mathtt{P}}[Y \mid \mathcal{F}_t]$. We always use this version. This allows us to view such conditional expectations under $\mathtt{P}$ as being well-defined up to $\mathtt{R}$-nullsets.

One might ask why we work with *local* domination, rather a 'global' condition like $\mathtt{P} \ll \mathtt{R}$ for all $\mathtt{P}$ of interest. The answer is that such a condition would be far too restrictive, as we now illustrate. Let $(X_t)_{t\in\mathbb{N}}$ be a sequence of random variables. For each $\eta \in \mathbb{R}$, let $\mathtt{P}^\eta$ be the distribution such that the X_t become i.i.d. Gaussian with mean η and unit variance. By the strong law of large numbers, $\mathtt{P}^\eta$ assigns probability one to the event $A^\eta := \{\lim_{t\to\infty} t^{-1}\sum_{s=1}^t X_s = \eta\}$. Moreover, the events A^η are mutually disjoint: $A^\eta \cap A^\nu = \emptyset$ whenever $\eta \ne \nu$. Therefore, by definition, the measures $\{\mathtt{P}^\eta\}_{\eta\in\mathbb{R}}$ are all mutually singular. Since there is an uncountable number of them, there cannot exist a measure $\mathtt{R}$ such that $\mathtt{P}^\eta \ll \mathtt{R}$ for all $\eta \in \mathbb{R}$. On the other hand, if $\mathtt{P}^\eta_t$ denotes the law of the partial sequence $X_1, \ldots, X_t$, then the measures $\{\mathtt{P}^\eta\}_{\eta\in\mathbb{R}}$, are all mutually absolutely continuous. In particular, we could (for instance) use $\mathtt{R} = \mathtt{P}^0$ as reference measure and obtain $\mathtt{P}^\eta_t \ll_{\mathrm{loc}} \mathtt{R}_t$ for all $\eta \in \mathbb{R}$.

A.2 Essential supremum and infimum

We briefly review the notions of essential supremum and infimum. For more information, as well as proofs of the results below, we refer to Section A.5 in [8].

On some probability space, consider a collection $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$ of random variables, where $\mathcal{A}$ is an arbitrary index set. If $\mathcal{A}$ is uncountable, the pointwise supremum $\sup_{\alpha\in\mathcal{A}} Y_\alpha$ might not be measurable (not a random variable). Alternatively, it might happen that $Y_\alpha = 0$ almost surely for every $\alpha \in \mathcal{A}$, but $\sup_{\alpha\in\mathcal{A}} Y_\alpha = 1$. For this reason, the pointwise supremum is often not useful. Instead, one can use the *essential supremum.*

Proposition 38. *There exists a* $[-\infty, \infty]$*-valued random variable* Y*, called the* essential supremum *and denoted by* $\operatorname{ess\,sup}_{\alpha\in\mathcal{A}} Y_\alpha$*, such that*

1. *$Y \geq Y_\alpha$, almost surely, for every $\alpha \in \mathcal{A}$,*

2. *if Y' is a random variable that satisfies $Y' \geq Y_\alpha$, almost surely, for every $\alpha \in \mathcal{A}$, then $Y' \geq Y$, almost surely.*

The essential supremum is almost surely unique.

In words, the essential supremum is the smallest almost sure upper bound on $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$. The proposition guarantees that it always exists. In some cases, more can be said: the essential supremum can be obtained as the limit of an increasing sequence.

Proposition 39. *Suppose $\{Y_\alpha\}$ is closed under maxima, meaning that for any $\alpha, \beta \in \mathcal{A}$ there is some $\gamma \in \mathcal{A}$ such that $Y_\gamma = Y_\alpha \vee Y_\beta$. Then there is a sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ such that $\{Y_{\alpha_n}\}_{n\in\mathbb{N}}$ is an increasing sequence and $\operatorname{ess\,sup}_{\alpha\in\mathcal{A}} Y_\alpha = \lim_{n\to\infty} Y_{\alpha_n}$ almost surely.*

One can also define the *essential infimum* by setting

$$\operatorname*{ess\,inf}_{\alpha\in\mathcal{A}} Y_\alpha := -\operatorname*{ess\,sup}_{\alpha\in\mathcal{A}}(-Y_\alpha).$$

This is the largest almost sure lower bound on $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$. It satisfies properties analogous to those in the propositions above.

A.3 On the choice of filtration

In the paper, we assume that the filtration $(\mathcal{F}_t)$ in use is by default the canonical filtration $\mathcal{F}_t := \sigma(U, X_1, \ldots, X_t)$. However, there are examples of hypothesis tests for $H_0 : \mathtt{Q} \in \mathcal{Q}$ where the only $\mathcal{Q}$-NMs with respect to $(\mathcal{F}_t)$ are almost surely constants. For the purpose of designing more powerful tests, it may make sense to coarsen the filtration.

As a first example, consider the problem of testing if a sequence is exchangeable:

$$H_0 : X_1, X_2, \ldots \text{ form an exchangeable sequence.}$$

Vovk [22] demonstrates that all martingales with respect to $(\mathcal{F}_t)$ (under the null) are constants, and are hence all derived tests are powerless to reject the null. Nevertheless, Vovk demonstrates that one can derive interesting and nontrivial 'conformal' martingales (M_t) with respect to the restricted filtration $\mathcal{G}_t := \sigma(M_1, \ldots, M_t) \subset \mathcal{F}_t$, that do indeed have power to reject the null (for appropriate deviations from the null). In short, coarsening the filtration is a design tool that could aid in the construction of more powerful sequential tests, and p-values and e-values.

In the following example, we show how the choice of including external randomization U into $\mathcal{F}_0$ also helps design better p-values. (However, it is not always possible to randomize atoms as Example 43 illustrates.)

Example 40. *Assume that $\mathcal{Q} = \{\mathtt{Q}\}$, where under $\mathtt{Q}$ we have that X_1 is Bernoulli($1/2$) and $X_2, X_3, \ldots = 0$. Consider the canonical filtration $(\mathcal{G}_t)$, so that $\mathcal{G}_\infty = \sigma(X_1)$. Then (M_t) with $M_0 = 1$, $M_t = 2X_1$ for all $t \in \mathbb{N}$ is a $\mathtt{Q}$-NM and the corresponding p-value $(\inf_{s\leq t} 1/M_s)$ is admissible. (Indeed, any p-value $(\mathfrak{p}_t)$ has to satisfy $\mathtt{Q}(\mathfrak{p}_1 \leq \alpha) \in \{0, 1/2, 1\}$ for each $\alpha \in [0, 1]$). However, by expanding the filtration using external randomization, one can easily derive a strictly smaller p-value $(\mathfrak{p}'_t)$ such that $\mathfrak{p}'_1$ is uniform. In other words, the original p-value is only admissible under the filtration generated solely by the observations, but is inadmissible under an expanded filtration that includes a randomization device (which is the filtration $(\mathcal{F}_t)$ in this paper).*

The main idea of the following randomization device is somewhat folklore, but we find the following succinct lemma useful.

Lemma 41 (Randomization device)**.** *If Y is a random variable with distribution function F and U is an independent uniformly distributed random variable, then*

$$Y' := UF(Y) + (1 - U)F(Y-)$$

is uniformly distributed and satisfies $Y' \leq F(Y)$.

Proof. Fix $a \in [0,1]$ and define $y := \inf\{x \in \mathbb{R} : F(x) \geq a\}$. Note that

$$\Pr(Y' \leq a|Y) = \Pr\left(U \leq \frac{a - F(y-)}{F(y) - F(y-)} \middle| Y\right) \mathbf{1}_{Y=y} + \Pr(UF(Y) + (1-U)F(Y-) \leq a \mid Y)\mathbf{1}_{Y \neq y}.$$

(If $\Pr(Y = y) = 0$, the first term should be understood as zero.) On $\{Y > y\}$ we have $a < F(Y-)$, so that the second term equals zero. On $\{Y < y\}$ we have $F(Y) \leq a$, so that the second term instead equals one. Since also U is uniform and independent of Y, we get

$$\Pr(Y' \leq a|Y) = \frac{a - F(y-)}{F(y) - F(y-)} \mathbf{1}_{Y=y} + \mathbf{1}_{Y<y}.$$

Taking expectations and simplifying gives $\Pr(Y' \leq a) = a$, showing that Y' is uniformly distributed. Finally, it is clear from the definition of Y' that $Y' \leq F(Y)$. $\square$

B Omitted proofs

Proof of Lemma 1. It is clear that (ii) $\implies$ (iii). The implication (i) $\implies$ (ii) follows from

$$A_T = \left(\bigcup_{t\in\mathbb{N}} (A_t \cap \{T = t\})\right) \cup (A_\infty \cap \{T = \infty\}) \subseteq \bigcup_{t\in\mathbb{N}} A_t.$$

For (iii) $\implies$ (i), take $\tau := \inf\{t \in \mathbb{N} : A_t \text{ occurs}\}$, so that $A_\tau = \bigcup_{t\in\mathbb{N}} A_t$. $\square$

Proof of Lemma 3. First, (i) implies (ii) since $N_T \leq \sup_{t\in\mathbb{N}_0} N_t$, hence $\mathtt{E}[N_T] \leq \mathtt{E}[\sup_{t\in\mathbb{N}_0} N_t]$, for all random times T. Conversely, for any $\varepsilon > 0$ there exists some random time T such that $N_T \geq \sup_{t\in\mathbb{N}_0} N_t - \varepsilon$. Thus if (ii) holds, then $\mathtt{E}[\sup_{t\in\mathbb{N}_0} N_t] \leq \mathtt{E}[N_T] + \varepsilon \leq 1 + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we find that (ii) implies (i). It is clear that (ii) implies (iii).

The fact that (iii) implies (iv) is however not obvious, and is essentially a consequence of a result by Shafer et al. [20, Theorem 3], as also noted recently by Kirichenko and Grünwald [13, Lemma 5.1]. First, we note that if (N_t) satisfies (iii), then $\mathfrak{p}_t := 1 \wedge \inf_{s\leq t} 1/N_s$ is a p-value (see also Proposition 10(1)). In particular, $\mathfrak{p}_\infty := \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t$ stochastically dominates a uniform. Therefore, for any nonnegative, nonincreasing function $f(u)$ such that $\int_0^1 f(u)\mathrm{d}u = 1$, we have $\mathtt{E}[f(\mathfrak{p}_\infty)] \leq \int_0^1 f(u)\mathrm{d}u = 1$ (see also the proof of Proposition 12). The function $f(u) := g(1/u)$, with g as in the lemma, satisfies this condition. Consequently, $\mathtt{E}[g(1 \vee \sup_{s\in\mathbb{N}_0} N_s)] = \mathtt{E}[f(\mathfrak{p}_\infty)] \leq 1$, as required. $\square$

Proof of Proposition 4. For each $t \in \mathbb{N}_0$, define a probability measure $\mathtt{P}'_t$ on the Borel sets of $\mathbb{R}^t$ by $\mathtt{P}'_t(A) := \mathtt{E}_\mathtt{Q}[M_t \mathbf{1}_A]$. Because (M_t) is a martingale under $\mathtt{Q}$, the sequence $(\mathtt{P}'_t)_{t\in\mathbb{N}}$ forms a consistent system of finite-dimensional distributions. Therefore, by Kolmogorov's extension theorem, there exists a single probability measure $\mathtt{P}$ on the Borel sets of $\Omega = \mathbb{R}^\mathbb{N}$ whose projection onto $\mathbb{R}^t$ is exactly $\mathtt{P}'_t$ for each $t \in \mathbb{N}$. Put differently, $\mathtt{P}$ satisfies $\mathtt{P}_t = \mathtt{P}'_t$ for all $t \in \mathbb{N}$, as desired. $\square$

Proof of Proposition 7. We prove the statement for p-values; the same argument holds for e-values and sequential tests. The proof is based on transfinite induction. Fix some p-value $(\mathfrak{p}_t)$. For all countable ordinals β, we now recursively define p-values $(\mathfrak{p}_t^\beta)$ as follows. For $\beta = 1$, we set $(\mathfrak{p}_t^\beta) := (\mathfrak{p}_t)$. For any successor ordinal $\gamma := \beta + 1$, if $(\mathfrak{p}_t^\beta)$ is $\mathcal{Q}$-admissible we set $\mathfrak{p}_t^\gamma := \mathfrak{p}_t^\beta$, and otherwise we let $(\mathfrak{p}_t^\gamma)$ be any p-value that strictly dominates $(\mathfrak{p}_t^\beta)$. For any limit ordinal $\gamma := \lim_{n\to\infty} \beta_n$, we define $(\mathfrak{p}_t^\gamma) := (\lim_{n\to\infty} \mathfrak{p}_t^{\beta_n})$.

Let us now use the induction assumption that $(\mathfrak{p}_t^\beta)$ is $\mathcal{Q}$-valid for all $\beta < \gamma$, for this limit ordinal γ. Since $(\lim_{n\to\infty} \mathfrak{p}_t^{\beta_n})$ is a decreasing limit, we have for every $\varepsilon > 0$, $\alpha \in [0,1]$, and $\mathtt{Q} \in \mathcal{Q}$ that

$$\mathtt{Q}\left(\inf_{t\in\mathbb{N}_0} \mathfrak{p}_t^\gamma \le \alpha\right) \le \mathtt{Q}\left(\lim_{n\to\infty} \inf_{t\in\mathbb{N}_0} \mathfrak{p}_t^{\beta_n} < \alpha + \varepsilon\right) = \lim_{n\to\infty} \mathtt{Q}\left(\inf_{t\in\mathbb{N}_0} \mathfrak{p}_t^{\beta_n} < \alpha + \varepsilon\right) \le \alpha + \varepsilon.$$

It follows that $(\mathfrak{p}_t^\gamma)$ is $\mathcal{Q}$-valid. By transfinite induction, this holds for all countable ordinals β.

Writing $\mathtt{R}$ for the reference probability measure, $\{\mathtt{E}_\mathtt{R}[\sum_{t\in\mathbb{N}_0} 2^{-t}\mathfrak{p}_t^\beta]\}_\beta$ defines a decreasing $[0,2]$-valued transfinite sequence. This sequence must eventually become stationary, that is, it becomes constant for all β beyond some countable ordinal β_0. Thus $\mathfrak{p}_t^\beta = \mathfrak{p}_t^{\beta_0}$ for all $\beta \ge \beta_0$ and all $t \in \mathbb{N}_0$. By construction, $(\mathfrak{p}_t^{\beta_0})$ must then be admissible and dominate $(\mathfrak{p}_t)$. This shows that any p-value for $\mathcal{Q}$ can be dominated by a $\mathcal{Q}$-admissible p-value.

Let us also remark that in the case of $\mathcal{Q}$ being a singleton the statement for e-values and sequential tests could be proved in a more constructive manner as in Subsections 6.2 and 6.3. $\square$

Proof of Proposition 9. We prove the three statements in order. Let (ψ_t) denote the constructed binary sequence, which we will now show is a $(\mathcal{Q},\alpha)$-sequential tets. Let τ denote an arbitrary stopping time, potentially infinite, and fix $\mathtt{Q} \in \mathcal{Q}$.

(1) $\mathtt{Q}(\psi_\tau = 1) = \mathtt{Q}(\mathfrak{p}_\tau \le \alpha) \le \alpha$ since $(\mathfrak{p}_t)$ is $\mathtt{Q}$-valid.

(2) $\mathtt{Q}(\psi_\tau = 1) = \mathtt{Q}(\mathfrak{e}_\tau \ge 1/\alpha) \le \alpha\mathtt{E}_\mathtt{Q}[\mathfrak{e}_\tau] \le \alpha$, where we used Markov's inequality and the fact that $(\mathfrak{e}_t)$ is $\mathtt{Q}$-safe. In short, e-values satisfy Ville's inequality.

(3) $\mathtt{Q}(\psi_\tau = 1) = \mathtt{Q}(\phi(\mathcal{Q}) \cap \mathfrak{C}_\tau = \emptyset) \le \mathtt{Q}(\phi(\mathtt{Q}) \notin \mathfrak{C}_\tau) \le \alpha$, where the first inequality follows because the event $\{\phi(\mathcal{Q}) \cap \mathfrak{C}_\tau = \emptyset\}$ implies that $\mathfrak{C}_\tau$ does not contain $\phi(\mathtt{Q})$, which is improbable under $\mathtt{Q}$.

The fact that the $(\mathcal{Q},\alpha)$-sequential tests in (1) and (2) are nested is obvious. This completes the proof. $\square$

Proof of Proposition 10. We prove the three statements in order. Let $(\mathfrak{p}_t)$ denote the constructed sequence of random variables, which we will now show in each case is a p-value. Let τ denote an arbitrary stopping time, potentially infinite, and fix $\mathtt{Q} \in \mathcal{Q}$.

(1) $\mathtt{Q}(1/\mathfrak{e}_\tau \le \alpha) = \mathtt{Q}(\mathfrak{e}_\tau \ge 1/\alpha) \le \mathtt{E}_\mathtt{Q}[\mathfrak{e}_\tau] \cdot \alpha \le \alpha$, where we used Markov's inequality and the fact that $(\mathfrak{e}_t)$ is $\mathtt{Q}$-safe. Since a p-value remains valid after taking the running infimum, we obtain $(\mathfrak{p}_t)$ is valid.

(2) $\mathtt{Q}(\mathfrak{p}_\tau > \alpha) = \mathtt{Q}(\psi_\tau(\alpha) = 0) \ge 1-\alpha$, where the equality follows since the sequential tests are nested and the inequality because $(\psi_t(\alpha))$ is a $(\mathcal{Q},\alpha)$-sequential test.

(3) $\mathtt{Q}(\mathfrak{p}_\tau > \alpha) = \mathtt{Q}(\phi(\mathcal{Q}) \cap \mathfrak{C}_\tau(\alpha) = \emptyset) \ge \mathtt{Q}(\phi(\mathtt{Q}) \notin \mathfrak{C}_\tau(\alpha)) \ge 1-\alpha$ as in (2).

This completes the proof. $\square$

Proof of Proposition 11. Let $(\mathfrak{C}_t)$ denote the constructed sequence of sets, which we will now show is a $(\phi, \mathcal{P}, \alpha)$-confidence sequence. To this end, let τ denote an arbitrary stopping time, potentially infinite, and fix $\mathtt{P} \in \mathcal{P}$; note that $\mathtt{P} \in \mathcal{P}^\gamma$ for some $\gamma \in \mathcal{Z}$. Then, we have $\mathtt{P}(\phi(\mathtt{P}) \notin \mathfrak{C}_\tau) = \mathtt{P}(\gamma \notin \mathfrak{C}_\tau) = \mathtt{P}(\psi_\tau^\gamma = 1) \le \alpha$ since (ψ_t^γ) is a $(\mathcal{P}^\gamma, \alpha)$-sequential test. This completes the proof. $\square$

Proof of Proposition 12. Define $\mathfrak{e}_t := f(\mathfrak{p}_t)$; we must show that $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe. If $(\mathfrak{p}_t)$ is $\mathcal{Q}$-valid then for any stopping time τ and $\mathtt{Q} \in \mathcal{Q}$, the distribution of $\mathfrak{p}_\tau$ is stochastically larger than a uniform random variable (denoted V). Thus for any calibrator f, we have $\mathtt{E}_\mathtt{Q}[\mathfrak{e}_\tau] = \mathtt{E}_\mathtt{Q}[f(\mathfrak{p}_\tau)] \le \mathtt{E}[f(V)] = \int_0^1 f(v)\mathrm{d}v = 1$. Since this result holds for any τ and $\mathtt{Q} \in \mathcal{Q}$, the result follows. $\square$

Proof of Proposition 13. To see (1), fix a probability measure $\mathtt{Q} \in \operatorname{conv}(\mathcal{Q})$. Then there exist $\mathtt{Q}_1, \mathtt{Q}_2 \in \mathcal{Q}$ and $\lambda \in [0,1]$ such that $\mathtt{Q} = \lambda \mathtt{Q}_1 + (1-\lambda)\mathtt{Q}_2$. Let now $(\mathfrak{e}_t)$ denote an e-value for $\mathcal{Q}$. Consider some stopping time τ and note that

$$\mathtt{E}_{\mathtt{Q}}[\mathfrak{e}_\tau] = \lambda \mathtt{E}_{\mathtt{Q}_1}[\mathfrak{e}_\tau] + (1-\lambda)\mathtt{E}_{\mathtt{Q}_2}[\mathfrak{e}_\tau] \le \lambda + 1 - \lambda = 1,$$

since $(\mathfrak{e}_t)$ is $\mathcal{Q}$-safe. This yields that $(\mathfrak{e}_t)$ is also $\operatorname{conv}(\mathcal{Q})$-safe. The same argument also applies for valid p-values and sequential tests.

Next, (2) follows in a similar way. Assume that $(\mathfrak{e}_t)$ is $\mathcal{Q}$-admissible and consider some $\operatorname{conv}(\mathcal{Q})$-valid e-value $(\mathfrak{e}'_t)$ that satisfies $\mathtt{Q}(\mathfrak{e}'_t \ge \mathfrak{e}_t) = 1$ for all $t \in \mathbb{N}$ and $\mathtt{Q} \in \operatorname{conv}(\mathcal{Q})$ and there exists some $\mathtt{Q}^* \in \operatorname{conv}(\mathcal{Q})$ and some $t \in \mathbb{N}$ such that $\mathtt{Q}^*(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$. Since we always have $\mathtt{Q}^* = \lambda \mathtt{Q}_1 + (1-\lambda)\mathtt{Q}_2$ for some $\mathtt{Q}_1, \mathtt{Q}_2 \in \mathcal{Q}$ and $\lambda \in [0,1]$ we also have $\mathtt{Q}_1(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$ or $\mathtt{Q}_2(\mathfrak{e}'_t > \mathfrak{e}_t) > 0$, leading to a contradiction. Again, the same argument also applies for admissibile p-values and sequential tests. $\square$

Proof of Lemma 31. Fix some $\alpha \in (0,1]$, let τ denote the first time t that $M_t \ge 1/\alpha$, and let

$$q := \mathtt{Q}(\tau < \infty) = \mathtt{Q}\left(\sup_{t \in \mathbb{N}_0} M_t \ge \frac{1}{\alpha}\right).$$

Next, (14) yields $\mathtt{Q}(M_\infty = 0) = 1$, for example, by [16, Theorem 4.2]. Note that the stopped process M^τ is a uniformly integrable martingale, yielding $\mathtt{E}_{\mathtt{Q}}[M^\tau_\infty] = 1$. On the event $\{\tau = \infty\}$, we have $M^\tau_\infty = 0$. With $M_{-1} := 1$, $Y_0 := 0$, and $\mathcal{F}_{-1} := \{\emptyset, \Omega\}$, these observations then yield

$$\begin{aligned}
1 = \mathtt{E}_{\mathtt{Q}}\left[M^\tau_\infty\right] &= \sum_{t \in \mathbb{N}_0} \mathtt{E}_{\mathtt{Q}}\left[M_t \mathbf{1}_{\tau = t}\right] = \sum_{t \in \mathbb{N}_0} \mathtt{E}_{\mathtt{Q}}\left[M_{t-1} Y_t \mathbf{1}_{\tau = t}\right] = \sum_{t \in \mathbb{N}_0} \mathtt{E}_{\mathtt{Q}}\left[\mathtt{E}_{\mathtt{Q}}\left[Y_t \middle| \mathcal{F}_{t-1}, Y_t \ge \frac{1}{\alpha M_{t-1}}\right] M_{t-1}\mathbf{1}_{\tau = t}\right] \\
&\le \frac{1}{\alpha}(1+\varepsilon)\sum_{t \in \mathbb{N}_0} \mathtt{E}_{\mathtt{Q}}\left[\mathbf{1}_{\tau = t}\right] = q\frac{1+\varepsilon}{\alpha}.
\end{aligned}$$

This then gives $q \ge \alpha/(1+\varepsilon)$, yielding the claim. $\square$

Proof of Proposition 33. First, assume (M_t) is an $\mathcal{S}$-NM and fix a time t. Since (Y_t) is adapted, Y_t is a function of $U, X_1, \ldots, X_t$. Hence we may write $Y_t = f_t(X_t)$ for some nonnegative predictable function $f_t(\cdot)$. More explicitly, $Y_t = f_t(U, X_1, \ldots, X_{t-1}; X_t)$. Now pick any real numbers $x_1, \ldots, x_t$. Consider the two-point measures $\mu_s := (\delta_{-x_s} + \delta_{x_s})/2$ for all $s \le t$ and let $\mathtt{P} := \mathtt{U} \times \prod_{s \in \mathbb{N}} \mu_{s \wedge t}$ be the distribution that makes the data independent with $X_s \sim \mu_{s \wedge t}$. Then $\mathtt{P} \in \mathcal{S}$. Moreover,

$$1 = \mathtt{E}_{\mathtt{P}_x}[Y_t | \mathcal{F}_{t-1}] = \frac{1}{2}\left(f_t(U, X_1, \ldots, X_{t-1}; x_t) + f_t(U, X_1, \ldots, X_{t-1}; -x_t)\right).$$

Since the event $\{X_i = x_i,\ i = 1, \ldots, t-1\}$ has positive probability, we get

$$\frac{1}{2}\left(f_t(U, x_1, \ldots, x_{t-1}; x_t) + f_t(U, x_1, \ldots, x_{t-1}; -x_t)\right) = 1.$$

But the numbers $x_1, \ldots, x_t$ were arbitrary, so it follows that the function $x \mapsto f_t(U, x_1, \ldots, x_{t-1}; x) - 1$ is odd for all $x_1, \ldots, x_{t-1}$.

For the reverse direction fix some $\mathtt{P} \in \mathcal{S}$ and some $t \in \mathbb{N}$. Then

$$\begin{aligned}
\mathtt{E}_{\mathtt{P}}[Y_t | \mathcal{F}_{t-1}] &= \mathtt{E}_{\mathtt{P}}[f_t(X_t) | \mathcal{F}_{t-1}] \\
&\overset{(i)}{=} \frac{1}{2}\left(\mathtt{E}_{\mathtt{P}}[f_t(X_t) | \mathcal{F}_{t-1}] + \mathtt{E}_{\mathtt{P}}[f_t(-X_t) | \mathcal{F}_{t-1}]\right) \\
&= 1 + \frac{1}{2}\left(\mathtt{E}_{\mathtt{P}}[f_t(X_t) - 1 | \mathcal{F}_{t-1}] + \mathtt{E}_{\mathtt{P}}[f_t(-X_t) - 1 | \mathcal{F}_{t-1}]\right) \\
&\overset{(ii)}{=} 1,
\end{aligned}$$

yielding the statement. Above, equality (i) follows by symmetry of $\mathtt{P}$, and equality (ii) follows because $f_t - 1$ is odd. $\square$

Proof of Proposition 35. This follows from an application of Corollary 32. Fix some $\varepsilon > 0$. Since f is continuous at zero and $f(0) = 1$ there exists some $\eta > 0$ such that $f(x) \leq 1+\varepsilon$ for all $x \in (-2\eta, 2\eta)$. This implies (15). Moreover, since f is strictly monotone at zero we may assume that $f(\eta) \neq 1$.

Consider now the measure $\mu_\eta := (\delta_{-\eta} + \delta_\eta)/2$ and note that $\mathtt{Q}_\eta := \mathtt{U} \times \mu_\eta^\infty \in \mathcal{S}$. Hence $\mathtt{Q}_\eta(f(X_t) \leq 1+\varepsilon) = 1$ for all $t \in \mathbb{N}$. Moreover,

$$\mathtt{Q}_\eta\left(\sum_{t\in\mathbb{N}}(f(X_t)-1)^2 = \infty\right) = \mathtt{Q}_\eta\left(\sum_{t\in\mathbb{N}}(f(\eta)-1)^2 = \infty\right) = 1,$$

since $f-1$ is odd and $f(\eta) \neq 1$. This shows that (14) holds. Hence Corollary 32 can indeed be applied and the statement follows. $\square$

C Auxiliary examples

The following example shows that Proposition 13 cannot be extended to confidence sequences.

Example 42 (Confidence sequences do not mesh with convex closures)**.** *Let μ_+ (μ_-) denote the law of a Gaussian random variable with unit variance and mean 1 (-1). Moreover, let $\mathcal{P} = \{\mu_+^\infty, \mu_-^\infty\}$ be the family of i.i.d. laws of such distributions. Consider ϕ^{mean}, which satisfies $\phi^{\text{mean}}(\mu_-^\infty) = -1$ and $\phi^{\text{mean}}(\mu_+^\infty) = 1$. Then $(\mathfrak{C}_t)$ given by $\mathfrak{C}_t = \{-1, +1\}$ is a (trivial) $(\phi^{\text{mean}}, \mathcal{P}, \alpha)$-valid confidence sequence for $\alpha \in [0,1]$. Now consider the measure $\mathtt{P} = (\mu_+^\infty + \mu_-^\infty)/2 \in \operatorname{conv}(\mathcal{P})$, which satisfies $\phi^{\text{mean}}(\mathtt{P}) = 0 \notin \mathfrak{C}_t$. It is clear that $(\mathfrak{C}_t)$ is not a $(\phi^{\text{mean}}, \operatorname{conv}(\mathcal{P}), \alpha)$-valid confidence sequence for any $\alpha \in [0,1]$.*

The next example also elaborates further on the discussion in Subsection 8.3 by providing an admissible p-value that has an atomic limiting distribution.

Example 43 (Atomic admissible p-values exist even in the presence of an independent $\mathcal{F}_0$-measurable random device)**.** *Consider $\mathtt{Q}$ under which (X_t) are i.i.d. uniformly distributed. Then an adapted process $(\mathfrak{p}_t)$ with the following properties can be constructed.*

- *$\mathfrak{p}_t$ is supported on $\{1/2 + k/2^{t+1}\}_{k=1,\dots 2^t}$;*
- *$\mathtt{Q}(\mathfrak{p}_t = 1/2 + 1/2^{t+1}) = 1/2 + 1/2^{t+1}$ and $\mathtt{Q}(\mathfrak{p}_t = 1/2 + k/2^{t+1}) = 1/2^{t+1}$ for all $k = 2, \dots 2^t$;*
- *$(\mathfrak{p}_t)$ is a $\mathtt{Q}$-MM with*

$$\mathtt{Q}\left(\mathfrak{p}_{t+1} - \frac{1}{2} \in \left\{\frac{2k-1}{2^{t+2}}, \frac{k}{2^{t+1}}\right\} \middle| \, \mathfrak{p}_t - \frac{1}{2} = \frac{k}{2^{t+1}}\right) = 1$$

  for all $k = 1, \dots 2^t$ and $t \in \mathbb{N}$;
- *$(\mathfrak{p}_t)$ is independent of U.*

Note that $\mathfrak{p}_\infty := \inf_{t\in\mathbb{N}} \mathfrak{p}_t$ satisfies $\mathtt{Q}(\mathfrak{p}_\infty \leq \alpha) = \alpha \mathbf{1}_{\alpha \geq 1/2} \leq \alpha$ for all $\alpha \in [0,1]$; in particular $(\mathfrak{p}_t)$ is an anytime p-value and its limit $\mathfrak{p}_\infty$ has an atom at $1/2$.

We claim that $(\mathfrak{p}_t)$ is $\mathtt{Q}$-admissible. Indeed, assume there exists an anytime p-value $(\mathfrak{p}'_t)$ that dominates $(\mathfrak{p}_t)$ (we explicitly allow $(\mathfrak{p}'_t)$ to depend on the randomization device U). Then there exists some $t \in \mathbb{N}$ such that $\mathtt{Q}(\mathfrak{p}'_t < \mathfrak{p}_t) > 0$. Let us first assume that $\mathtt{Q}(\mathfrak{p}'_t < 1/2 + 1/2^{t+1}) > 0$. In combination with the fact that $(\mathfrak{p}_t)$ is a $\mathtt{Q}$-MM there exists some $n > t+2$ such that

$$\mathtt{Q}\left(\left\{\mathfrak{p}'_t \leq \frac{1}{2} + \frac{1}{2^{t+1}}\right\} \cap \left\{\mathfrak{p}_\infty \geq \frac{1}{2} + \frac{1}{2^{t+2}}\right\}\right) > 0.$$

Since $\mathfrak{p}'_\infty \leq \mathfrak{p}_\infty$ and since $\mathtt{Q}(\mathfrak{p}_\infty \geq 1/2 + 1/2^{t+2}) = 1/2 - 1/2^{t+2}$, we hence obtain $\mathtt{Q}(\mathfrak{p}'_\infty \geq 1/2 + 1/2^{t+2}) < 1/2 - 1/2^{t+2}$, a contradiction to the fact that $(\mathfrak{p}'_t)$ is an anytime p-value. We obtain similar contradictions if we assume $\mathtt{Q}(\{\mathfrak{p}'_t < \mathfrak{p}_t\} \cap \{\mathfrak{p}_t = 1/2 + k/2^{t+1}\}) > 0$ for some $k = 2, \dots, 2^t$. This shows that $(\mathfrak{p}_t)$ is indeed $\mathtt{Q}$-admissible, despite having an atom and being independent of the randomization device.

The next example illustrates how anti-concentration bounds can be satisfied by NMs that lead to inadmissible p-values.

Example 44 (A p-value for $\mathcal{S}$ that satisfies Proposition 35 need not be admissible)**.** *Fix the function* $f : x \mapsto ((1+x) \wedge 2)^+$. *This function satisfies the criteria of Proposition 35. Hence the process* $M_t := \prod_{s \le t} f(X_s)$ *is an* $\mathcal{S}$*-martingale and we also have* (17)*. Consider the* p*-value* $(\mathfrak{p}_t)$ *given by* $\mathfrak{p}_t := \inf_{s \le t} 1/M_s$*. This is a* p*-value by* (17)*.*

Define next $\mathfrak{p}'_t := \mathfrak{p}_t$ *for* $t = 0, 1$ *and* $\mathfrak{p}'_t := \mathfrak{p}_t - 1/4\, \mathbf{1}_{X_1 \vee X_2 \le -1}$ *for* $t \ge 2$*. Clearly we have* $\mathtt{Q}(\mathfrak{p}'_t \le \mathfrak{p}_t) = 1$ *for all* $t \in \mathbb{N}_0$ *and* $\mathtt{Q} \in \mathcal{S}$ *and there exists* $\mathtt{Q}^* \in \mathcal{S}$ *such that* $\mathtt{Q}^*(\mathfrak{p}'_2 < \mathfrak{p}_2) > 0$*. We now claim that* $(\mathfrak{p}'_t)$ *is also* $\mathcal{S}$*-valid. We will prove this claim below. This assertion then yields that* (17) *is* not *sufficient for the admissibility of the corresponding* p*-value in general.*

We now prove the claim that $(\mathfrak{p}'_t)$ *is* $\mathcal{S}$*-valid. To do so we consider a subset* $\widetilde{\mathcal{S}} \subset \mathcal{S}$*, namely those measures* $\mathtt{Q} \in \mathcal{S}$ *that satisfy* $\mathtt{Q}(X_1 \le -1) \in \{0, 1/2\}$*. Note that* $\mathcal{S} = conv(\widetilde{\mathcal{S}})$*, the convex hull of* $\widetilde{\mathcal{S}}$*. Thanks to Proposition 13(1) it suffices to argue that* $(\mathfrak{p}'_t)$ *is* $\widetilde{\mathcal{S}}$*-valid. To this end, note that on the event* $\{X_1 > -1\}$ *we have* $\mathfrak{p}_t = \mathfrak{p}'_t$ *for all* $t \in \mathbb{N}_0$*. On the other hand, on the event* $\{X_1 \le -1\}$ *we have* $\mathfrak{p}_t = 1$ *for all* $t \in \mathbb{N}_0$*.*

Fix now some $\alpha \in (0,1)$ *and* $\mathtt{Q} \in \widetilde{\mathcal{S}}$*. Without loss of generality we can assume that* $\mathtt{Q}(X_1 \le -1) = 1/2$*, otherwise there is nothing to be argued. We now need to show that*

$$\mathtt{Q}(\mathfrak{p}'_\infty \le \alpha) \le \alpha. \tag{20}$$

To make headway, note that

$$\begin{aligned}
\{\mathfrak{p}'_\infty \le \alpha\} &= (\{\mathfrak{p}'_\infty \le \alpha\} \cap \{X_1 > -1\}) \cup (\{\mathfrak{p}'_\infty \le \alpha\} \cap \{X_1 \le -1\}) \\
&\subset \{\mathfrak{p}_\infty \le \alpha\} \cup \left\{ \mathfrak{p}_\infty - \frac{1}{4} \mathbf{1}_{X_1 \vee X_2 \le -1} \le \alpha \right\} \\
&= \{\mathfrak{p}_\infty \le \alpha\} \cup \left\{ \frac{1}{4} \mathbf{1}_{X_1 \vee X_2 \le -1} \ge 1 - \alpha \right\}.
\end{aligned}$$

Thus, if $\alpha < 3/4$ *then* $\{\mathfrak{p}'_\infty \le \alpha\} \subset \{\mathfrak{p}_\infty \le \alpha\}$ *and we have* (20)*. Let us now assume that* $\alpha \ge 3/4$*. Note that since* $\mathtt{Q}(\mathfrak{p}_\infty = 1) \ge \mathtt{Q}(X_1 \le -1) = 1/2$ *and hence* $\mathtt{Q}(\mathfrak{p}_\infty \le \alpha) < 1/2$*. This then yields*

$$\mathtt{Q}(\mathfrak{p}'_\infty \le \alpha) \le \mathtt{Q}(\mathfrak{p}_\infty \le \alpha) + \mathtt{Q}(X_1 \vee X_2 \le -1) < 1/2 + 1/2^2 = 3/4 \le \alpha,$$

yielding the $\widetilde{S}$*-validity of* $(\mathfrak{p}'_t)$*, hence also the* S*-validity of* $(\mathfrak{p}'_t)$*.*