ASYMPTOTICALLY OPTIMAL MULTISTAGE TESTS FOR NON-I.I.D. DATA

Yiming Xing and Georgios Fellouris

University of Illinois at Urbana-Champaign

Supplementary Material

The supplement is organized as follows. In Section S1 we state sufficient conditions for the asymptotic analysis in Section 4. In Section S2 we develop an importance sampling approach for the efficient implementation of the proposed tests when the error probabilities are small. In Section S3 we illustrate the general theory in three specific testing problems. In Section S4 we present the results of numerical studies. All proofs are presented in Section S5.

Before starting, we introduce some extra notations. For an event A, we denote by $1\{A\}$ its indicator function. For a set $A \subseteq \mathbb{R}$, we denote by A^o its interior. For a function $f : \mathbb{R} \to (-\infty, \infty]$, we call $\{x \in \mathbb{R} : f(x) < \infty\}$ the effective domain of f, and denote by $f(x+)$ (resp. $f(x-)$) the right (resp. left) limit of f at $x \in \mathbb{R}$ when it exists.

S1 Sufficient conditions

In this section we state sufficient conditions for the existence of functions ψ_0, ψ_1 that satisfy (4.9)–(4.10), which we also specify. To this end, we rely on the *Gärtner-Ellis theorem* from large deviation theory. We start by stating a version of this theorem that focuses on events of form (κ, ∞) or $(-\infty, \kappa]$, where $\kappa \in \mathbb{R}$, and requires somewhat weaker conditions compared to standard formulations in the literature, such as (Dembo and Zeitouni, 1998, Theorem 2.3.6) or (Bucklew, 2010, Theorem 3.2.1).

S1.1 The Gärtner-Ellis theorem

In this subsection we consider an arbitrary $\mathsf{P} \in \mathcal{P}$ and for every $\theta \in \mathbb{R}$ we set

$$\phi_n(\theta) \equiv \frac{1}{n} \log \mathsf{E} \left[\exp\{n\, \theta\, T_n\} \right], \quad n \in \mathbb{N},$$

and assume that

$$\phi(\theta) \equiv \lim_n \phi_n(\theta) \quad \text{exists in} \quad (-\infty, \infty]. \tag{S1.1}$$

We denote by Θ the effective domain of ϕ, i.e., $\Theta \equiv \{\theta \in \mathbb{R} : \ \phi(\theta) < \infty\}$, and by ϕ^* its Legendre-Fenchel transform:

$$\phi^*(\kappa) \equiv \sup_{\theta \in \mathbb{R}} \{\theta \kappa - \phi(\theta)\}, \quad \kappa \in \mathbb{R}.$$

We further assume that $\Theta^o \neq \emptyset$, and that

$$\phi \text{ is } \textit{strictly} \text{ convex and continuous in } \Theta \text{ and differentiable in } \Theta^o. \quad \text{(S1.2)}$$

This assumption implies that ϕ' is strictly increasing in Θ^o, that $\phi'(\Theta^o)$ is a non-trivial open interval, and as a result that

$$\phi^*(\kappa) = \vartheta(\kappa)\,\kappa - \phi(\vartheta(\kappa)), \quad \text{for any } \kappa \in \phi'(\Theta^o),$$

where ϑ is the inverse of ϕ' in Θ^o.

Finally, we assume that for every $\theta \in \Theta$ there exists a distribution of X, Q_θ, such that

$$\frac{d\mathrm{Q}_\theta}{d\mathsf{P}}(\mathcal{F}_n) = \exp\{n\,(\theta\, T_n - \phi_n(\theta))\}, \quad \text{for any } n \in \mathbb{N}. \tag{S1.3}$$

This is known as an *exponential tilting* of P, and for its existence it suffices, for example, that $\mathbb{S}$ be Polish (see, e.g., (Parthasarathy, 2005, p. 144, Theorem 5.1)).

Theorem S1. *Suppose* (S1.1) (S1.2) (S1.3) *hold.*

(i) If $\Theta^o \cap (0,\infty) \neq \emptyset$, then $\phi^*(\phi'(0+)) = 0$, ϕ^* is strictly increasing in $\phi'(\Theta^o \cap (0,\infty))$ and, for every $\kappa \in \phi'(\Theta^o \cap (0,\infty))$,

$$\lim_n \frac{1}{n} \log \mathsf{P}\,(T_n > \kappa) = -\phi^*(\kappa). \tag{S1.4}$$

(ii) If $\Theta^o \cap (-\infty, 0) \neq \emptyset$*, then* $\phi^*(\phi'(0-)) = 0$*,* ϕ^* *is strictly decreasing in* $\phi'(\Theta^o \cap (-\infty, 0))$ *and, for every* $\kappa \in \phi'(\Theta^o \cap (-\infty, 0))$*,*

$$\lim_n \frac{1}{n} \log \mathsf{P}\left(T_n \leq \kappa\right) = -\phi^*(\kappa). \tag{S1.5}$$

(iii) For every $\theta \in \Theta^o$*,* $\mathrm{Q}_\theta\left(T_n \to \phi'(\theta)\right) = 1$.

Proof. Section S5.4. □

Remark S1. 1) Theorem S1 implies that, for any $\epsilon > 0$, $\mathsf{P}\left(T_n > \phi'(0+) + \epsilon\right)$ decays exponentially fast in n if Θ^o intersects $(0, \infty)$, and $\mathsf{P}(T_n \leq \phi'(0-) - \epsilon)$ decays exponentially fast in n if Θ^o intersects $(-\infty, 0)$.

2) In standard formulations of the Gärtner-Ellis theorem, such as (Dembo and Zeitouni, 1998, Theorem 2.3.6) or (Bucklew, 2010, Theorem 3.2.1), it is additionally assumed that $0 \in \Theta^o$, in which case the conditions in both (i) and (ii) of Theorem S1 hold, $\phi'(0)$ exists, and thus $\mathsf{P}(|T_n - \phi'(0)| > \epsilon)$ decays exponentially fast in n for any $\epsilon > 0$, and $\mathsf{P}(T_n \to \phi'(0)) = 1$. It is also assumed that ϕ is *steep*, i.e., $\phi'(\Theta^o) = \mathbb{R}$, (see, e.g., (Dembo and Zeitouni, 1998, Definition 2.3.5)), in which case

$$\phi'(\Theta^o \cap (0, \infty)) = (\phi'(0), \infty) \quad \text{and} \quad \phi'(\Theta^o \cap (-\infty, 0)) = (-\infty, \phi'(0)).$$

S1.2 Sufficient conditions for the asymptotic theory of Section 4

We next apply Theorem S1 to establish sufficient conditions for the asymptotic theory of Section 4. To this end, when the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_i$, where $i \in \{0, 1\}$, we write $\phi_{i,n}, \phi_i, \Theta_i, \phi_i^*, \vartheta_i$ instead of $\phi_n, \phi, \Theta, \phi^*, \vartheta$ and, for each $\theta \in \Theta_i^o$, we denote by $\mathrm{Q}_{i,\theta}$ the exponential tilting of P_i, i.e.,

$$\frac{d\mathrm{Q}_{i,\theta}}{d\mathsf{P}_i}(\mathcal{F}_n) = \exp\left\{n\left(\theta\, T_n - \phi_{i,n}(\theta)\right)\right\}, \quad \text{for any } n \in \mathbb{N}. \tag{S1.6}$$

Corollary S1. *Suppose* (4.8) *holds for some* $J_0, J_1 \in \mathbb{R}$ *so that* $J_0 < J_1$.

(i) If the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_0$ *and*

$$\begin{aligned} \Theta_0^o \cap (0, \infty) \neq \emptyset, \quad \phi_0'(0+) = J_0, \\ \textit{and there exists } \theta_0 \in \Theta_0 \cap (0, \infty): \ \phi_0'(\theta_0-) = J_1, \end{aligned} \tag{S1.7}$$

then (4.9) *holds, with* $\psi_0 = \phi_0^*$, *for every* $\kappa \in (J_0, J_1)$. *If also* $\theta_0 \in \Theta_0^o$, *then* (4.9) *holds, with* $\psi_0 = \phi_0^*$, *in a neighborhood of* J_1.

(ii) If the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_1$ *and*

$$\begin{aligned} \Theta_1^o \cap (-\infty, 0) \neq \emptyset, \quad \phi_1'(0-) = J_1, \\ \textit{and there exists } \theta_1 \in \Theta_1 \cap (-\infty, 0): \ \phi_1'(\theta_1+) = J_0, \end{aligned} \tag{S1.8}$$

then (4.10) *holds, with* $\psi_1 = \phi_1^*$, *for every* $\kappa \in (J_0, J_1)$. *If also* $\theta_1 \in \Theta_1^o$, *then* (4.10) *holds, with* $\psi_1 = \phi_1^*$, *in a neighborhood of* J_0.

(iii) For $i \in \{0, 1\}$*, if the assumptions of Subsection S1.1 hold for* $\mathsf{P} = \mathsf{P}_i$*, then*

$$\mathrm{Q}_{i,\theta}\left(T_n \to \phi_i'(\theta)\right) = 1 \quad \text{for any } \theta \in \Theta_i^o.$$

Proof. We only prove (i), as the proof of (ii) is similar, whereas that of (iii) follows directly from Theorem S1.(iii). Since $\phi_0'(\Theta_0^o)$ is, by assumption, an open interval, (S1.7) implies that

$$(J_0, J_1) \subseteq \phi_0'\left(\Theta_0^o \cap (0, \infty)\right), \tag{S1.9}$$

and the first claim in (i) follows by an application of Theorem S1.(i).

If also $\theta_0 \in \Theta_0^o$, (S1.7) implies that

$$(J_0, J_1] \subseteq \phi_0'(\Theta_0^o \cap (0, \infty)), \tag{S1.10}$$

and the second claim in (i) follows again by an application of Theorem S1.(i). □

Corollary S2. *Suppose that the assumptions of Subsection S1.1 hold for both* $\mathsf{P} = \mathsf{P}_0$ *and* $\mathsf{P} = \mathsf{P}_1$*.*

(i) If $0 \in \Theta_0^o \cap \Theta_1^o$ *and* $\phi_0'(0) < \phi_1'(0)$*, then* (4.8) *holds with* $J_i = \phi_i'(0)$*,* $i \in \{0, 1\}$*.*

(ii) If, also, both ϕ_0 *and* ϕ_1 *are steep, then* (4.9) *holds, with* $\psi_0 = \phi_0^*$*, for every* $\kappa > J_0$*, and* (4.10) *holds, with* $\psi_1 = \phi_1^*$*, for every* $\kappa < J_1$*.*

Proof. This is a direct consequence of Remark S1. □

Remark S2. The conditions in Corollary S1 are weaker than the conditions in Corollary S2. In what follows, we consider an example in which the former are satisfied while the latter are not, in the sense that 0 *is not in the interior of either* Θ_0 *or* Θ_1.

S1.3 The likelihood ratio case

In what follows, we focus on the case where $T = \bar{\Lambda}$ and the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_0$. Then, in view of the fact that

$$\mathsf{E}_1[\exp\{\theta\Lambda_n\}] = \mathsf{E}_0[\exp\{(\theta+1)\Lambda_n\}], \quad \text{for any } n \in \mathbb{N}, \quad \theta \in \mathbb{R},$$

the assumptions of Subsection S1.1 also hold for $\mathsf{P} = \mathsf{P}_1$, with

$$\phi_1(\theta) = \phi_0(\theta+1), \quad \theta \in \mathbb{R}, \tag{S1.11}$$

$$\Theta_1 = \Theta_0 - 1, \tag{S1.12}$$

$$\phi_1^*(\kappa) = \phi_0^*(\kappa) - \kappa, \quad \kappa \in \mathbb{R}. \tag{S1.13}$$

From (S1.11) it follows that 1 is the non-zero root of ϕ_0 and, as a result, $[0,1] \subseteq \Theta_0$ since Θ_0 is an interval. Since also ϕ_0 is strictly convex and continuous in $[0,1]$, and differentiable in $(0,1)$, we conclude that

$$-\infty < \phi_0'(0+) < 0 < \phi_0'(1-) < \infty.$$

From (S1.11) and (S1.12) it similarly follows that -1 is the non-zero root of ϕ_1, $[-1,0] \subseteq \Theta_1$, and

$$-\infty < \phi_1'(-1+) < 0 < \phi_1'(0-) < \infty.$$

Based on these observations, we can see that the conditions of Corollary S1 simplify considerably.

Corollary S3. *If $T = \bar{\Lambda}$, the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_0$, and* (4.2) *holds with $I_0 = -\phi_0'(0+)$ and $I_1 = \phi_0'(1-)$, then* (4.9) *and* (4.10) *hold for every $\kappa \in (-I_0, I_1)$ with $\psi_0 = \phi_0^*$ and $\psi_1 = \phi_1^*$, respectively. Moreover,*

$$C = \phi_0^*(0) = \phi_1^*(0) = -\inf_{\theta \in \mathbb{R}} \phi_0(\theta), \tag{S1.14}$$

where C is defined in (4.13)*.*

Proof. From the discussion prior to statement of Corollary S3 it follows that the conditions of Corollary S1 are satisfied, by applying which (4.9) and (4.10) follow. To show (S1.14), we note that the supremum in the definition of C in (4.13) is attained when $\psi_0 = \psi_1$, or equivalently when $\phi_0^* = \phi_1^*$. Comparing with (S1.13) completes the proof.

□

Remark S3. Suppose that $T = \bar{\Lambda}$ and that the assumptions of Subsection S1.1 hold for $\mathsf{P} = \mathsf{P}_0$. Then, from Corollary S2 it follows that a sufficient condition for (4.2), with $I_0 = -\phi_0'(0)$ and $I_1 = \phi_0'(1)$, is that $\{0,1\} \subset \Theta_0^o$. However, as we mentioned earlier, the assumptions of Corollary S3 may hold even when $\Theta_0 = [0,1]$, in which case $\Theta_1 = [-1,0]$ and $(-I_0, I_1) = \phi_i'(\Theta_i^o)$, $i \in \{0,1\}$. The importance of this observation becomes clear in the iid setup, on which we focus next.

The iid setup

We show that the conditions of Corollary S3 are satisfied in the iid setup of Subsection 4.1.1 as long as the Kullback-Leibler divergences defined in (4.6) are positive and finite, or equivalently, the expectation of $\Lambda_1 = \log\left(f_1(X_1)/f_0(X_1)\right)$ is non-zero and finite under both P_0 and P_1.

Indeed, in this case, (4.2) holds with $I_0 = D(f_0 \| f_1)$ and $I_1 = D(f_1 \| f_0)$ by Kolmogorov's Strong Law of Large Numbers, and clearly

$$\phi_0(\theta) = \log \mathsf{E}_0[\exp\{\theta \Lambda_1\}], \quad \theta \in \mathbb{R}.$$

Since ϕ_0 is the cumulant generating function of a non-degenerate distribution and $[0,1] \subseteq \Theta_0$, ϕ_0 is strictly convex in Θ_0, differentiable in Θ_0^o, continuous at 0 and 1, and satisfies

$$\phi_0'(0+) = \mathsf{E}_0[\Lambda_1] = -I_0 \quad \text{and} \quad \phi'(1-) = \frac{\mathsf{E}_0[\Lambda_1 \exp\{\Lambda_1\}]}{\mathsf{E}_0[\exp\{\Lambda_1\}]} = I_1$$

(see, e.g. (Dembo and Zeitouni, 1998, Excercise 2.2.24)).

Remark S4. As a quick summary of this section, for the asymptotic analysis in Section 4 to apply,

- when $T \neq \bar{\Lambda}$, it suffices to check the conditions in Corollary S1 or S2, where the latter is stronger than the former,
- when $T = \bar{\Lambda}$, it suffices to check the conditions in Corollary S3; when, in particular, X is an iid sequence, it suffices to check that the Kullback-Leibler divergences are positive and finite.

S2 Implementation via importance sampling

The proposed designs for the multistage tests in Section 2 require knowledge of the functions n^* and κ^*, defined in (2.4). These do not admit, in general, closed-form expressions and need to be approximated. For any given α and β in $(0,1)$, $n^*(\alpha,\beta)$ and $\kappa^*(\alpha,\beta)$ can be approximated by estimating $\mathsf{P}_0(T_n > \kappa)$ and $\mathsf{P}_1(T_n \leq \kappa)$ for different n and κ, and finding the minimum n for which there exists a κ so that the first probability does not exceed α and the second does not exceed β.

If it is convenient to simulate the sequence X under P_0 and P_1, a simple method for the estimation of $\mathsf{P}_0(T_n > \kappa)$ and $\mathsf{P}_1(T_n \leq \kappa)$ is plain Monte-Carlo simulation. However, when these probabilities are very small, this approach may not be efficient, or even feasible. Indeed, if the probability of interest is 10^{-a} for some $a > 0$, the minimum number of simulation runs needed for the *relative error* of the Monte-Carlo estimator to be at most 1% is 10^{a+4}. Therefore, when the probability of interest is very small, a different method is needed for its estimation, such as *importance sampling* Bucklew (2010).

To illustrate this method, we focus on the estimation of $\mathsf{P}_0(T_n > \kappa)$, as a completely analogous discussion applies to the estimation of $\mathsf{P}_1(T_n \leq \kappa)$. We observe that if Q is a distribution of X that is mutually absolutely continuous with P_0 on $\mathcal{F}_n$ for every $n \in \mathbb{N}$, then $\mathsf{P}_0(T_n > \kappa) = \mathsf{E}_{\mathrm{Q}}\left[Z_{n,\kappa}(\mathrm{Q})\right]$, where

$$Z_{n,\kappa}(\mathrm{Q}) \equiv \frac{d\mathsf{P}_0}{d\mathrm{Q}}(\mathcal{F}_n) \cdot 1\{T_n > \kappa\} \tag{S2.15}$$

and E_{Q} denotes expectation under Q. Thus, if it is possible to simulate X under Q, $\mathsf{P}_0(T_n > \kappa)$ can be estimated by averaging $Z_{n,\kappa}(\mathrm{Q})$ over a large number of independent realizations of X in which it is distributed according to Q.

The question then is how to select the *importance sampling distribution* Q, so that the relative error of the induced estimator is small even when $\mathsf{P}_0(T_n > \kappa)$ is small. To answer it, we assume that the assumptions of Corollary S1.(i) (resp. Corollary S3) hold when $T \neq \bar{\Lambda}$ (resp. $T = \bar{\Lambda}$)

and fix κ in (J_0, J_1) (resp. $(-I_0, I_1)$), in which case $\mathsf{P}_0(T_n > \kappa)$ decays exponentially fast in n. Then, squaring both sides in (S2.15), applying the Cauchy-Schwarz inequality, taking logarithms on both sides, dividing by n, letting $n \to \infty$, and applying (4.9), we obtain

$$\varliminf_n \frac{1}{n} \log \mathsf{E}_{\mathrm{Q}} \left[Z^2_{n,\kappa}(\mathrm{Q}) \right] \geq -2\, \psi_0(\kappa). \tag{S2.16}$$

The latter is essentially a universal asymptotic lower bound on the variance of any importance sampling estimator. As it is common in the relevant literature (see, e.g., (Bucklew, 2010, Chapter 5)), we refer to Q as *logarithmically efficient* for the estimation of $\mathsf{P}_0(T_n > \kappa)$ if it attains this asymptotic lower bound, i.e., if

$$\varlimsup_n \frac{1}{n} \log \mathsf{E}_{\mathrm{Q}} \left[Z^2_{n,\kappa}(\mathrm{Q}) \right] \leq -2\, \psi_0(\kappa). \tag{S2.17}$$

Recalling the definition of the exponential tilting $\mathrm{Q}_{0,\theta}$ in (S1.3), for every $n \in \mathbb{N}$ and $\theta \in \Theta_0^o$ we have

$$\begin{aligned} \mathsf{E}_{\mathrm{Q}_{0,\theta}} \left[Z^2_{n,\kappa}(\mathrm{Q}_{0,\theta}) \right] &= \mathsf{E}_{\mathrm{Q}_{0,\theta}} \left[\exp\{-2n\,(\theta\, T_n - \phi_{0,n}(\theta))\};\, T_n > \kappa \right] \\ &\leq \exp\{-2n(\theta\, \kappa - \phi_{0,n}(\theta))\}. \end{aligned}$$

Taking logarithms, dividing by n and letting $n \to \infty$ we obtain

$$\varlimsup_n \frac{1}{n} \log \mathsf{E}_{\mathrm{Q}_{0,\theta}} \left[Z^2_{n,\kappa}(\mathrm{Q}_{0,\theta}) \right] \leq -2(\theta\, \kappa - \phi_0(\theta)).$$

Therefore, when $\theta = \vartheta_0(\kappa)$, where ϑ_0 is the inverse function of ϕ_0', the right-hand-side is equal to $-2\psi_0(\kappa)$, which proves that $\mathrm{Q}_{0,\vartheta_0(\kappa)}$ is logarithmically efficient for the estimation of $\mathsf{P}_0(T_n > \kappa)$.

Working similarly, we can see that if the assumptions of Corollary S1.(ii) (resp. Corollary S3) hold when $T \neq \bar{\Lambda}$ (resp. $T = \bar{\Lambda}$), a logarithmically efficient importance sampling distribution for the estimation of $\mathsf{P}_1(T_n \leq \kappa)$ is $\mathrm{Q}_{1,\vartheta_1(\kappa)}$, where ϑ_1 is the inverse function of ϕ_1'. In Subsection S3.1, we show that in the case of testing in a one-parameter exponential family, $\mathrm{Q}_{0,\vartheta_0(\kappa)}$ and $\mathrm{Q}_{1,\vartheta_1(\kappa)}$ coincide.

Finally, we observe that by Corollary S1.(iii) it follows that $\mathrm{Q}_{i,\vartheta_i(\kappa)}(T_n \to \kappa) = 1$, $i \in \{0, 1\}$. This suggests that if it is not convenient to simulate X under the logarithmically efficient importance sampling distributions, a potential strategy for estimating $\mathsf{P}_0(T_n > \kappa)$ and $\mathsf{P}_1(T_n \leq \kappa)$, simultaneously, is to apply importance sampling using a distribution under which *it is convenient to simulate X and T_n converges almost surely to κ as $n \to \infty$*. We apply this strategy successfully in two non–iid testing problems in Section S4.

S3 Examples

In this section we focus on three concrete testing problems, with which we illustrate the general results of the previous sections. Specifically, for each of these testing problems we show that the conditions of Subsection 4.1 hold, and also that the conditions of Subsection 4.2 hold for $T = \bar{\Lambda}$, as well as for an alternative test statistic. In view of Theorem 3 and 4, we can see that the asymptotic relative efficiency of the proposed multistage tests when $T \neq \bar{\Lambda}$ and (4.11) holds, against when $T = \bar{\Lambda}$, under P_0 and P_1, as $\alpha, \beta \to 0$ satisfying the required relative decay rates, can be written as

$$\mathrm{ARE}_0 \equiv \frac{\psi_1(J_0)}{I_0} \quad \text{and} \quad \mathrm{ARE}_1 \equiv \frac{\psi_0(J_1)}{I_1}, \tag{S3.18}$$

which we also compute in these examples.

S3.1 Testing in a one-parameter exponential family

In the first example of this section we let h be a density with respect to a σ-finite measure ν on $\mathbb{S}$ such that $M \neq \emptyset$, where

$$M \equiv \{\mu \in \mathbb{R} : \varphi(\mu) < \infty\}^o, \qquad \varphi(\mu) \equiv \log \int_{\mathbb{S}} e^{\mu x}\, h(x)\, \nu(dx), \tag{S3.19}$$

and, for each $\mu \in M$, we set

$$h_\mu(x) \equiv h(x)\; e^{\mu x - \varphi(\mu)}, \quad x \in \mathbb{S}.$$

Noting that h_μ is also a density with respect to ν, with the same support as h, we denote by $\mathbb{P}_\mu$ the distribution of X, and by $\mathbb{E}_\mu$ the corresponding expectation, when X is a sequence of independent random elements with common density h_μ, and consider the testing setup of Subsection 2.1. In this context, the log-likelihood ratio statistic in (4.7) becomes

$$\Lambda_n = (\mu_1 - \mu_0) \sum_{i=1}^{n} X_i - n\,(\varphi(\mu_1) - \varphi(\mu_0)), \quad n \in \mathbb{N}, \tag{S3.20}$$

and, for any $\mu \in M$, it is a random walk under $\mathbb{P}_\mu$ with drift

$$\mathbb{E}_\mu[\Lambda_1] = (\mu_1 - \mu_0)\,\varphi'(\mu) - (\varphi(\mu_1) - \varphi(\mu_0)). \tag{S3.21}$$

Thus, setting μ equal to μ_0 and μ_1, we obtain the following expressions for the Kullback-Leibler divergences in (4.6):

$$\begin{aligned} D(f_0||f_1) &= -\big((\mu_1 - \mu_0)\,\varphi'(\mu_0) - (\varphi(\mu_1) - \varphi(\mu_0))\big) \\ D(f_1||f_0) &= (\mu_1 - \mu_0)\,\varphi'(\mu_1) - (\varphi(\mu_1) - \varphi(\mu_0))\,. \end{aligned}$$

Since these quantities are positive and finite, by the discussion in Subsection S1.3 it follows that all assumptions in Subsections 4.1-4.2 hold with

$$\begin{aligned} I_0 =& D(f_0\|f_1), \quad I_1 = D(f_1\|f_0), \quad C = \psi_0(0), \\ \psi_0(\kappa) =& \vartheta_0(\kappa)\kappa - \phi_0(\vartheta_0(\kappa)), \quad \text{for any } \kappa \in (-I_0, I_1) \\ \psi_1(\kappa) =& \vartheta_1(\kappa)\kappa - \phi_1(\vartheta_1(\kappa)), \quad \text{for any } \kappa \in (-I_0, I_1), \end{aligned} \tag{S3.22}$$

where ϑ_i is the inverse of ϕ_i', $i \in \{0,1\}$, and

$$\phi_0(\theta) = \varphi(\mu_0 + \theta(\mu_1 - \mu_0)) - (\varphi(\mu_0) + \theta(\varphi(\mu_1) - \varphi(\mu_0)))\,,\ \theta \in [0,1]$$
$$\phi_1(\theta) = \varphi(\mu_1 + \theta(\mu_1 - \mu_0)) - (\varphi(\mu_1) + \theta(\varphi(\mu_1) - \varphi(\mu_0)))\,,\ \theta \in [-1,0]. \tag{S3.23}$$

As a result, in this context, the asymptotic optimality of the proposed multistage tests holds when $T = \bar{\Lambda}$. In fact, it also holds when

$$T = \bar{X} \equiv \{\bar{X}_n, n \in \mathbb{N}\}, \quad \text{where} \quad \bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n \in \mathbb{N}, \tag{S3.24}$$

regardless of $\mu_0, \mu_1 \in M^o$. This is because there is a bijection between $\bar{\Lambda}_n$ and $\bar{X}_n$:

$$\bar{\Lambda}_n = (\mu_1 - \mu_0)\,\bar{X}_n - (\varphi(\mu_1) - \varphi(\mu_0)), \quad n \in \mathbb{N}, \tag{S3.25}$$

which implies that the values of $n^*(\alpha,\beta)$ and $\kappa^*(\alpha,\beta)$, which in general depend on the choice of the test statistic T, coincide when $T = \bar{X}$ and when $T = \bar{\Lambda}$.

Importance sampling distributions

In this setup, it is convenient to obtain an explicit form for the logarithmically efficient importance sampling distributions for the estimation of both $\mathsf{P}_0(\bar{\Lambda}_n > \kappa)$ and $\mathsf{P}_1(\bar{\Lambda}_n \leq \kappa)$ for any $\kappa \in (-I_0, I_1)$. Indeed, for any $\kappa \in (-I_0, I_1)$ we have:

$$\mathrm{Q}_{0,\vartheta_0(\kappa)} = \mathrm{Q}_{1,\vartheta_1(\kappa)} = \mathbb{P}_\mu,$$

where $\mu \in (\mu_0, \mu_1)$ is such that $\mathbb{E}_\mu[\Lambda_1] = \kappa$. To prove this statement, we first note that for any $n \in \mathbb{N}$ and $\theta \in (0,1)$, by (S3.20) we have

$$\begin{aligned}\Lambda_n\left(\mathbb{P}_{\mu_0+\theta(\mu_1-\mu_0)}, \mathsf{P}_0\right) &= \Lambda_n\left(\mathbb{P}_{\mu_0+\theta(\mu_1-\mu_0)}, \mathbb{P}_{\mu_0}\right) \\ &= \theta(\mu_1-\mu_0)\sum_{i=1}^{n} X_i - n\left(\varphi(\mu_0+\theta(\mu_1-\mu_0)) - \varphi(\mu_0)\right) \\ &= n\left(\theta\bar{\Lambda}_n - \phi_0(\theta)\right),\end{aligned}$$

and similarly, for any $n \in \mathbb{N}$ and $\theta \in (-1,0)$,

$$\Lambda_n\left(\mathbb{P}_{\mu_1+\theta(\mu_1-\mu_0)}, \mathsf{P}_1\right) = n\left(\theta\bar{\Lambda}_n - \phi_1(\theta)\right).$$

Thus, the exponential tiltings of P_0 and P_1, defined in (S1.6), are given by

$$\begin{aligned}\mathrm{Q}_{0,\theta} &= \mathbb{P}_{\mu_0+\theta(\mu_1-\mu_0)}, \quad \theta \in (0,1), \\ \mathrm{Q}_{1,\theta} &= \mathbb{P}_{\mu_1+\theta(\mu_1-\mu_0)}, \quad \theta \in (-1,0).\end{aligned}$$

Differentiating the identities in (S3.23) and comparing with (S3.21) we obtain

$$\begin{aligned}\mathbb{E}_{\mu_0+\theta(\mu_1-\mu_0)}[\Lambda_1] &= \phi_0'(\theta), \quad \theta \in (0,1), \\ \mathbb{E}_{\mu_1+\theta(\mu_1-\mu_0)}[\Lambda_1] &= \phi_1'(\theta), \quad \theta \in (-1,0).\end{aligned} \tag{S3.26}$$

The statement now follows by the definition of ϑ_i as the inverse of ϕ_i', where $i \in \{0,1\}$.

A binary statistic

An approach to the testing problem of this subsection, which can be motivated by practical constraints or robustness considerations, is to binarize the data, recording only whether each observation is larger, or not, than some user-specified value in the interior of the support of h, say x_*. Then, the test statistic can be written as

$$\begin{gathered} T = \bar{Z} \equiv \{\bar{Z}_n, n \in \mathbb{N}\}, \\ Z_n \equiv 1\{X_n > x_*\}, \quad \bar{Z}_n \equiv \frac{1}{n}\sum_{i=1}^{n} Z_i, \quad n \in \mathbb{N}, \end{gathered} \tag{S3.27}$$

and all assumptions in Subsection 4.2, including (4.11), are satisfied with

$$\begin{gathered} J_i = \mathsf{P}_i(X_1 > x_*), \quad C = -\log\sqrt{4\, J_0\, J_1}, \\ \phi_i(\theta) = \log\left(J_i\, e^{\theta} + (1 - J_i)\right), \quad \theta \in \mathbb{R}, \\ \psi_i(\kappa) = \mathrm{Ber}(\kappa || J_i), \quad \kappa \in (0,1), \end{gathered}$$

where $i \in \{0,1\}$, and $\mathrm{Ber}(x||y)$ is the Kullback-Leibler divergence between two Bernoulli distributions with success probability x and y respectively, i.e.,

$$\mathrm{Ber}(x||y) \equiv x\log(x/y) + (1-x)\log((1-x)/(1-y)), \quad x, y \in (0,1). \tag{S3.28}$$

Testing the Gaussian mean

We next specialize the above results to the special case of testing the mean of a Gaussian distribution with unit variance, i.e., when $M = \mathbb{R}$ and $\varphi(\mu) = \mu^2/2$ for every $\mu \in \mathbb{R}$ in (S3.19). For simplicity, we assume that the two parameter values under which we control the two error probabilities, μ_0 and μ_1, are opposite, i.e., $\mu_1 = -\mu_0 = \eta$ for some $\eta > 0$.

In this case, $n^*(\alpha, \beta)$ and $\kappa^*(\alpha, \beta)$ in (2.4) can be computed explicitly when $T = \bar{\Lambda}$, for any $\alpha, \beta \in (0, 1)$. Specifically, we have

$$I_0 = I_1 = 2\eta^2 \equiv I, \quad C = 4I$$

$$\phi_0(\theta) = \theta(\theta - 1)\, I, \quad \phi_1(\theta) = \theta(\theta + 1)\, I, \quad \theta \in \mathbb{R}$$

$$\psi_0(\kappa) = (I + \kappa)^2/(4I), \quad \psi_1(\kappa) = (I - \kappa)^2/(4I), \quad \kappa \in \mathbb{R},$$

and, for any $\alpha, \beta \in (0, 1)$,

$$n^*(\alpha, \beta) = \left\lceil \frac{(z_\alpha + z_\beta)^2}{2I} \right\rceil, \qquad \kappa^*(\alpha, \beta) = \frac{z_\alpha - z_\beta}{2\sqrt{n^*(\alpha, \beta)}}, \tag{S3.29}$$

where z_p is the upper p-quantile of the standard Gaussian distribution. In Figure 1a we plot the functions ψ_0, ψ_1, for $T = \bar{\Lambda}$ and $T = \bar{Z}$ in (S3.27) with $x_* = 0$, when $\eta = 0.5$.

Finally, we note that in this case the asymptotic relative efficiencies in

(S3.18) coincide when $T = \bar{Z}$, since

$$\mathrm{ARE}_0 = \frac{\mathrm{Ber}(\Phi(-\eta)||\Phi(\eta))}{2\eta^2} = \frac{\mathrm{Ber}(\Phi(\eta)||\Phi(-\eta))}{2\eta^2} = \mathrm{ARE}_1, \tag{S3.30}$$

where Φ denotes the cumulative distribution function of the standard Gaussian distribution and the function $\mathrm{Ber}(x||y)$ is defined in (S3.28). We note also that this quantity converges to 0.25 as $\eta \to \infty$ and to $2/\pi$ as $\eta \to 0$. In Figure 1b we plot the asymptotic relative efficiency in (S3.30) as a function of $\eta \in (0, 5)$.

S3.2 Testing the coefficient of a first-order autoregressive model

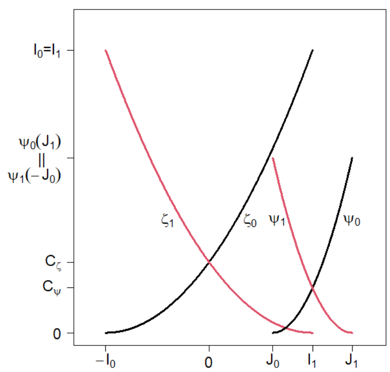
In the second example of this section we assume that X follows a Gaussian first-order autoregressive model, i.e.,

$$X_n = \mu X_{n-1} + \epsilon_n, \quad n \in \mathbb{N},$$

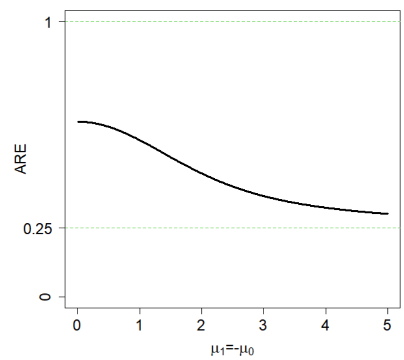
where $X_0 = 0$, $\{\epsilon_n, n \in \mathbb{N}\}$ are iid standard Gaussian, and μ is an unknown parameter taking values in $M = (-1, 1)$. We denote by $\mathbb{P}_\mu$ the distribution and by $\mathbb{E}_\mu$ the corresponding expectation when the true parameter is μ, and consider the testing problem of Subsection 2.1.

In this setup, the log-likelihood ratio statistic in (4.1) becomes

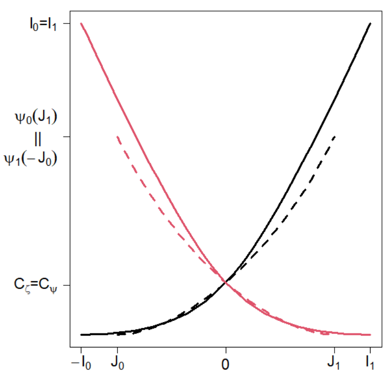
$$\Lambda_n = (\mu_1 - \mu_0) \left(\sum_{i=1}^{n} X_{i-1} X_i - \frac{\mu_1 + \mu_0}{2} \sum_{i=1}^{n} X_{i-1}^2 \right), \quad n \in \mathbb{N}. \tag{S3.31}$$



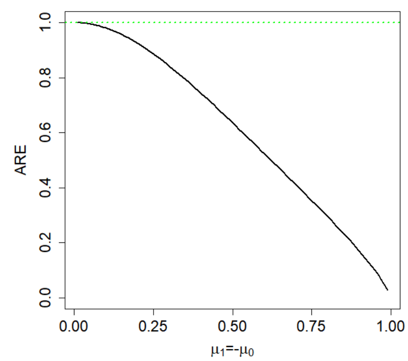
(a)



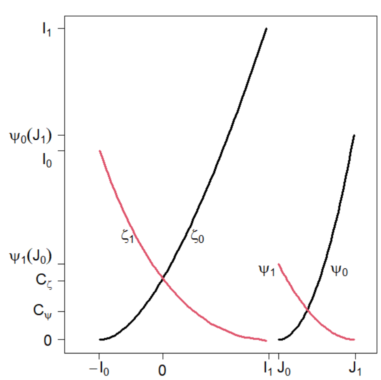
(b)



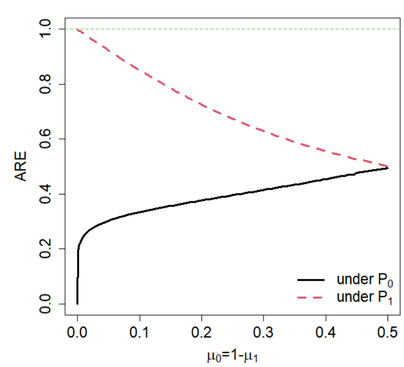
(c)



(d)



(e)



(f)

Table 1: Each row represents an example considered in Section S3. The left column draws the function images of ψ_0 and ψ_1 when $T = \bar{\Lambda}$ and when T is the alternative test statistic considered, where to distinguish them, they are denoted by ζ_0 and ζ_1 when $T = \bar{\Lambda}$. The right column draws the corresponding asymptotic relative efficiencies, defined in (S3.18).

For every $\mu \in M$, from (Brockwell and Davis, 1986, Chapter 3) it follows that

$$\frac{1}{n}\sum_{i=1}^{n} X_{i-1}^2 \to \frac{1}{1-\mu^2} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_{i-1}X_i \to \frac{\mu}{1-\mu^2} \quad \mathbb{P}_\mu - \text{a.s.}, \tag{S3.32}$$

and consequently

$$\bar{\Lambda}_n \to \frac{\mu_1-\mu_0}{1-\mu^2}\left(\mu - \frac{\mu_1+\mu_0}{2}\right) \qquad \mathbb{P}_\mu - \text{a.s.} \tag{S3.33}$$

Moreover, from Bercu et al. (1997) it follows that, for every $\mu \in M$,

$$\frac{1}{n}\log \mathbb{E}_\mu\left[e^{\theta\Lambda_n}\right] \to -\frac{1}{2}\log\left(\frac{1}{2}\,p_\mu(\theta) + \frac{1}{2}\sqrt{p_\mu^2(\theta) - 4q_\mu^2(\theta)}\right) \equiv \phi(\theta;\mu), \tag{S3.34}$$

where $\theta \in \mathcal{D}_\mu \equiv \mathcal{D}_{\mu,1} \cup \mathcal{D}_{\mu,2} \cup \mathcal{D}_{\mu,3}$,

$$\mathcal{D}_{\mu,1} \equiv \left\{\theta \in \mathbb{R} : \, \mu^2 < p_\mu(\theta) \le 2\mu^2, \; q_\mu^2(\theta) \le \mu^2(p_\mu(\theta) - \mu^2)\right\},$$

$$\mathcal{D}_{\mu,2} \equiv \left\{\theta \in \mathbb{R} : \, 2\mu^2 < p_\mu(\theta) < 2, \; p_\mu(\theta) > 2|q_\mu(\theta)|\right\},$$

$$\mathcal{D}_{\mu,3} \equiv \left\{\theta \in \mathbb{R} : \, p_\mu(\theta) \ge 2, \; q_\mu^2(\theta) \le p_\mu(\theta) - 1\right\},$$

$$\text{and} \quad p_\mu(\theta) \equiv 1 + \mu^2 + (\mu_1 - \mu_0)(\mu_1 + \mu_0)\,\theta,$$

$$q_\mu(\theta) \equiv -\mu - (\mu_1 - \mu_0)\theta.$$

The function $\phi(\cdot;\mu)$ in (S3.34) is differentiable in $\mathcal{D}_\mu^o$, and

$$0 \in \mathcal{D}_{\mu_0,2}^o, \qquad 1 \in (\mathcal{D}_{\mu_0,1} \cup \mathcal{D}_{\mu_0,2})^o.$$

Thus, setting μ equal to μ_0 and μ_1 in (S3.33)-(S3.34), we conclude that all assumptions in Corollary S3 are satisfied with

$$I_0 = \frac{(\mu_1 - \mu_0)^2}{2(1-\mu_0^2)}, \qquad I_1 = \frac{\mu_1^2 - \mu_0^2}{2(1-\mu_1^2)}, \qquad \phi_i = \phi(\cdot;\mu_i), \quad i \in \{0,1\}.$$

Moreover, from (S1.14) it follows, by minimizing $\phi(\cdot;\mu_0)$, that

$$C = \log \sqrt{\frac{1-\mu_0\,\mu_1}{1-(\mu_0+\mu_1)^2/4}}. \tag{S3.35}$$

The functions ψ_0 and ψ_1 in this context are computed numerically and are plotted in Figure 1c when $\mu_1 = -\mu_0 = 0.5$. We note that, in this case, they are symmetric about the y-axis, a property that does not hold, in general, when $\mu_1 \neq -\mu_0$.

The Yule-Walker estimator

An alternative test statistic for this testing problem is the Yule-Walker estimator, i.e., $T = \hat{\mu} \equiv \{\hat{\mu}_n, n \in \mathbb{N}\}$, where

$$\hat{\mu}_n \equiv \frac{\sum_{i=1}^n X_{i-1}X_i}{\sum_{i=1}^n X_i^2}, \quad n \in \mathbb{N}. \tag{S3.36}$$

From (S3.32) it follows that $\hat{\mu}_n$ is a strongly consistent estimator of μ, i.e., for every $\mu \in M$,

$$\mathbb{P}_\mu(\hat{\mu}_n \to \mu) = 1. \tag{S3.37}$$

Moreover, from Bercu et al. (1997) it follows that, for any $\mu \in M$,

$$\begin{aligned} -\frac{1}{n}\log \mathbb{P}_{\mu}(\hat{\mu}_n > \kappa) &\to \psi(\kappa;\mu), \quad \text{for any } \kappa \in (\mu, 1) \\ -\frac{1}{n}\log \mathbb{P}_{\mu}(\hat{\mu}_n \leq \kappa) &\to \psi(\kappa;\mu), \quad \text{for any } \kappa \in (-1, \mu), \end{aligned} \tag{S3.38}$$

where the function

$$\psi(\kappa;\mu) \equiv \log\sqrt{\frac{1+\mu^2-2\mu\kappa}{1-\kappa^2}}, \quad \kappa \in (-1,1)$$

is strictly convex, has a unique root at μ, and goes to ∞ as κ goes to -1 or 1. Therefore, setting μ equal to μ_0 and μ_1 in (S3.37)-(S3.38), we conclude that assumptions (4.8)-(4.11) hold with

$$J_i = \mu_i, \quad \psi_i = \psi(\cdot;\mu_i), \quad i \in \{0,1\}.$$

Interestingly, equating ψ_0 and ψ_1 we obtain the same value for C as in (S3.35). In view of (S5.52), this implies that using $\hat{\mu}$, instead of $\bar{\Lambda}$, as the test statistic, does not reduce the asymptotic relative efficiency of the *fixed-sample-size* test as $\alpha, \beta \to 0$ so that $|\log\alpha| \sim |\log\beta|$. This is not the case for the proposed multistage tests, as can be seen in Figure 1d, where we plot ARE_0 and ARE_1 when $\mu_0 = -\mu_1$, in which case they coincide, for different values of μ_1 in $(0,1)$.

S3.3 Testing the transition matrix of a Markov chain

In the third example of this section we assume that X is an irreducible and recurrent Markov chain with state space $[I] = \{0,1,\ldots,I\}$ where $I \in \mathbb{N}$, initial value $X_0 = 0$, transition matrix Π, and stationary distribution π. Moreover, we note that (see, e.g., (Durrett, 2010, Theorem 5.5.9))

$$Y \equiv \{Y_n \equiv (X_{n-1}, X_n),\, n \in \mathbb{N}\}$$

is also an irreducible and recurrent Markov chain, with state space $[I]^2$, transition matrix $\Pi^{\mathcal{Y}}$ whose $((i_1, i_2), (i_3, i_4))$-th element is

$$\Pi(i_3, i_4)\, 1\{i_2 = i_3\}, \quad (i_1, i_2), (i_3, i_4) \in [I]^2,$$

and stationary distribution $\pi^{\mathcal{Y}}$ whose (i, j)-th element is

$$\pi(i)\, \Pi(i, j), \quad i, j \in [I].$$

For simplicity, we identify the family of all possible distributions of X, $\mathcal{P}$, with the class of all irreducible and recurrent transition matrices of dimension $I + 1$. For each $\Pi \in \mathcal{P}$, we denote by $\mathbb{P}_\Pi$ the distribution of X, and by $\mathbb{E}_\Pi$ the corresponding expectation, when the transition matrix of X is Π. We consider the general testing setup of Section 2, where $\mathcal{P}_0$ and $\mathcal{P}_1$ are two arbitrary subclasses of $\mathcal{P}$, and

$$\mathsf{P}_i \equiv \mathbb{P}_{\Pi_i}, \quad i \in \{0, 1\}$$

for some arbitrary $\Pi_i \in \mathcal{P}_i$, $i \in \{0, 1\}$. In this setup, the log-likelihood ratio statistic in (4.1) takes the form:

$$\Lambda_n = \sum_{(i,j) \in [I]^2} r(i, j)\, N_n(i, j) = \sum_{m=1}^{n} \mathcal{U}(Y_m), \quad n \in \mathbb{N},$$

where, for each (i,j), $y \in [I]^2$,

$$r(i,j) \equiv \log\left(\frac{\Pi_1(i,j)}{\Pi_0(i,j)}\right), \quad N_n(i,j) \equiv \sum_{m=1}^{n} 1\{Y_m = (i,j)\},$$

$$\mathcal{U}(y) \equiv \sum_{(i,j)\in[I]^2} r(i,j) \cdot 1\{y = (i,j)\}.$$

For any $\Pi \in \mathcal{P}$, from (Durrett, 2010, Example 6.2.4) it follows that, for every $(i,j) \in [I]^2$,

$$\frac{1}{n} N_n(i,j) \to \pi^{\mathcal{Y}}(i,j) \qquad \mathbb{P}_{\Pi} - \text{a.s.}$$

and, as a result,

$$\bar{\Lambda}_n \to \sum_{(i,j)\in[I]^2} r(i,j)\, \pi^{\mathcal{Y}}(i,j) \qquad \mathbb{P}_{\Pi} - \text{a.s.} \tag{S3.39}$$

Moreover, by (Dembo and Zeitouni, 1998, Theorem 3.1.1 & 3.1.2), it follows that, for any $\Pi \in \mathcal{P}$,

$$\frac{1}{n} \log \mathsf{E}_{\Pi}[\exp\{\theta\Lambda_n\}] \to \log \xi\left(\Pi^{\mathcal{Y}}_{\theta,\mathcal{U}}\right) \equiv \phi(\theta;\Pi), \quad \text{for every} \quad \theta \in \mathbb{R}, \tag{S3.40}$$

where ξ is the functional that maps a matrix to its greatest eigenvalue, $\Pi^{\mathcal{Y}}_{\theta,\mathcal{U}}$ is a matrix of the same dimension as $\Pi^{\mathcal{Y}}$ whose $((i_1,i_2),\,(i_3,i_4))$-th element is

$$\Pi^{\mathcal{Y}}((i_1,i_2),(i_3,i_4))\, \exp\{\theta\,\mathcal{U}((i_3,i_4))\}, \quad (i_1,i_2),\,(i_3,i_4) \in [I]^2,$$

and the limit in (S3.40) is a finite and differentiable function of θ. Therefore, setting Π equal to Π_0 and Π_1 in (S3.39)-(S3.40) we conclude that all assumptions in Corollary S3 are satisfied, and I_i, ψ_i, $i \in \{0, 1\}$ can be computed accordingly.

The two-state case

We next specialize the previous setup to the case that $I = 1$, where the transition matrix and stationary distribution of X are of the form

$$\Pi = \begin{pmatrix} p & 1-p \\ 1-\mu & \mu \end{pmatrix}, \quad \pi = \left(\frac{1-\mu}{2-p-\mu}, \frac{1-p}{2-p-\mu}\right), \text{ where } \; p, \mu \in (0, 1).$$

We fix $p \in (0, 1)$, so that the only unknown parameter is μ, which takes values in $M = (0, 1)$. Thus, we now denote by $\mathbb{P}_\mu$ the distribution, and by $\mathbb{E}_\mu$ the corresponding expectation, of X when the unknown parameter is μ, and consider the testing setup of Subsection 2.1. In this case, (S3.39) reduces to

$$\bar{\Lambda}_n \to \frac{1-p}{2-p-\mu}\big(\operatorname{Ber}(\mu||\mu_0) - \operatorname{Ber}(\mu||\mu_1)\big) \quad \mathbb{P}_\mu - \text{a.s.}, \tag{S3.41}$$

where $\operatorname{Ber}(x||y)$ is defined in (S3.28), and I_0, I_1 become

$$I_0 = \frac{1-p}{2-p-\mu_0} \operatorname{Ber}(\mu_0\|\mu_1), \quad I_1 = \frac{1-p}{2-p-\mu_1} \operatorname{Ber}(\mu_1\|\mu_0).$$

An alternative test statistic in this setup is the sample average in (S3.24), or equivalently,

$$T_n = \bar{X}_n \equiv \frac{1}{n}\sum_{m=1}^{n} \mathcal{V}(x_m), \quad \text{where} \quad \mathcal{V}(x) = x.$$

Unlike the first example of this section, however, this test statistic does not lead to asymptotic optimality, as it does not admit a bijection with the log-likelihood ratio, as in (S3.25). To compute the resulting asymptotic relative efficiency, (S3.18), we note that, by (Durrett, 2010, Example 6.2.4), for any $\mu \in M$,

$$\bar{X}_n \to \sum_{i \in [I]} i\, \pi(i) = \frac{1-p}{2-p-\mu} \qquad \mathbb{P}_\mu - \text{a.s.} \tag{S3.42}$$

Moreover, by (Dembo and Zeitouni, 1998, Theorem 3.1.1 & 3.1.2) it follows that, for any $\mu \in M$,

$$\frac{1}{n} \log \mathsf{E}_\mu[\exp\{\theta\, n\, \bar{X}_n\}] \to \log \xi(\Pi_{\theta,\mathcal{V}}) \equiv \phi(\theta; \Pi), \quad \text{for any } \theta \in \mathbb{R}, \tag{S3.43}$$

where $\Pi_{\theta,\mathcal{V}}$ is a matrix of the same dimension as Π, whose (i,j)-th element is

$$\Pi(i,j)\, e^{\theta\, \mathcal{V}(j)}, \quad i,j \in [I],$$

and the limit is finite, differentiable and steep in $\mathbb{R}$ as a function of θ. Therefore, setting μ equal to μ_0 and μ_1 in (S3.42)-(S3.43) we conclude that all assumptions in Corollary S2 are satisfied, with

$$J_i = \frac{1-p}{2-p-\mu_i} \quad \text{and} \quad \phi_i(\theta) = \phi(\theta; \Pi_i), \quad \text{for any } \theta \in \mathbb{R}, \quad i \in \{0,1\}.$$

In Figure 1e we plot the functions ψ_0, ψ_1 for $T = \bar{\Lambda}$ and $T = \bar{X}$ when $\mu_0 = 1 - \mu_1 = 0.25$. In Figure 1f we plot the asymptotic relative efficiencies in (S3.18) against $\mu_0 = 1 - \mu_1$ for different values of $\mu_0 \in (0, 0.5)$.

S4 Numerical studies

In this section we present the results of two numerical studies in which we compare the 3-stage test, $\tilde{\chi}$, the 4-stage test, $\hat{\chi}$, both with $T = \bar{\Lambda}$, against the SPRT, χ', when

- testing the mean of an iid Gaussian sequence with unit variance (Subsection S3.1), with $\mu_1 = -\mu_0 = 0.5$,
- testing the correlation coefficient of a first-order autoregression (Subsection S3.2), with $\mu_1 = -\mu_0 = 0.5$,
- testing an entry in the transition matrix of a two-state Markov chain (Subsection S3.3), with $p = 0.5$ and $\mu_1 = 1 - \mu_0 = 0.75$.

Before we describe the two studies and present the main findings, we discuss how the tests are designed and how their average sample sizes are computed.

S4.1 Design of tests

In all cases, the SPRT in (4.4) is designed with $B = |\log \alpha|$ and $A = |\log \beta|$, whereas the multistage tests are designed according to Theorems 1 and 2, with free parameters selected according to (3.14) and (3.29). The functions n^* and κ^*, defined in (2.4), are evaluated using the closed-form expressions in (S3.29) in the first testing problem and using the importance sampling method of Section S2 in the other two.

Specifically, in the latter two examples, for n from 1 to a sufficiently large integer (we used 150) and for κ in a sufficiently fine grid over (J_0, J_1) (we used 100 equidistant points), we estimate, via importance sampling, $\mathsf{P}_0(T_n > \kappa)$ and $\mathsf{P}_1(T_n \leq \kappa)$. The importance sampling distribution we employed in the second (resp. third) example is the distribution $\mathbb{P}_\mu$ under which the limit in (S3.33) (resp. (S3.41)) is equal to κ. For each pair of n and κ, we implemented 10^4 simulation runs, which leads to relative errors below 0.5% in all cases. Moreover, grid search with 10^4 grid points is used for the determination of the free parameters.

S4.2 Computation of the expected sample sizes

The expected sample sizes of the multistage tests are computed using the formulas (3.8)-(3.9) and (3.20)-(3.22) in the first testing problem, as it is possible to compute the multivariate Gaussian probabilities in these expressions, and plain Monte Carlo in the other two. The expected sample size of the SPRT is estimated with plain Monte Carlo in all cases. In each Monte Carlo application, 10^4 replications are run, leading to *relative* errors below 0.5% in all cases.

S4.3 The first study

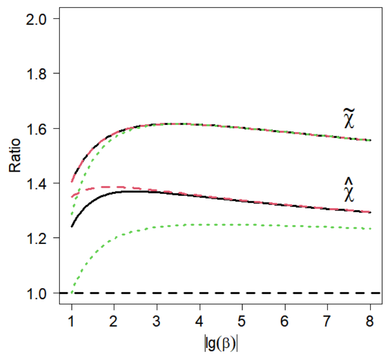
In the first study we compare the expected sample sizes of $\tilde{\chi}$, $\hat{\chi}$ and χ' under P_0, with the understanding that analogous results can be obtained when comparing $\tilde{\chi}$, $\check{\chi}$ and χ' under P_1. Specifically, we evaluate $\mathsf{E}_0[\tilde{\tau}]/\mathsf{E}_0[\tau']$ and $\mathsf{E}_0[\hat{\tau}]/\mathsf{E}_0[\tau']$, i.e., the ratio of the expected sample sizes under H_0 of $\tilde{\chi}$ and $\hat{\chi}$ over that of χ', in the context of the first testing problem, for different values of β, when α is given by one of the following relationships:

$$\alpha = \beta, \quad \alpha = \beta^4, \quad |\log \alpha| = |\log \beta|^{1.5}, \quad |\log \alpha| = |\log \beta|/\beta^{0.08}. \quad \text{(S4.44)}$$

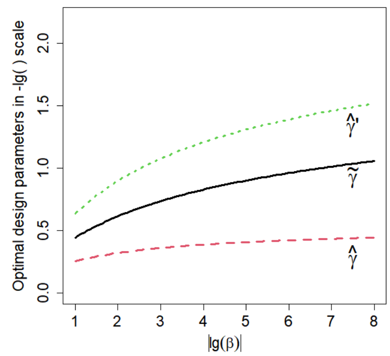
In the left column of Figure 2 and 3 we present these ratios, together with the non-asymptotic bounds implied by (3.12)-(3.13) and (3.25)-(3.26). In these graphs we observe a slow, downward trend, as α and β decrease, in all ratios but the one that corresponds to $\tilde{\chi}$ *in the last asymptotic regime*. This is consistent with Theorem 3, in which $\hat{\chi}$ is shown to achieve asymptotic optimality under P_0 in all asymptotic regimes in (S4.44), whereas $\tilde{\chi}$ only in the first three.

From these graphs we also see that, under P_0, the average sample of the 4-stage test, $\hat{\chi}$, is substantially smaller than that of the 3-stage test, $\tilde{\chi}$, in all cases, and exceeds that of the SPRT by about $20\% \sim 40\%$

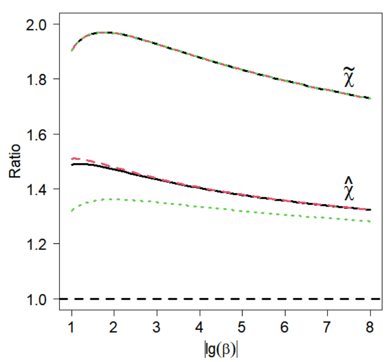
Finally, we see that the upper bounds are very accurate approximations



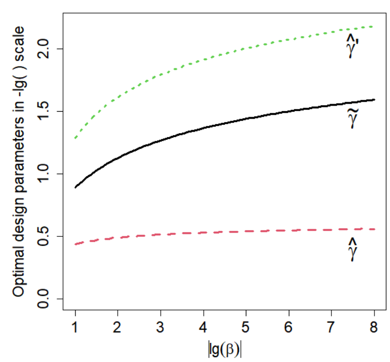
(a) $\alpha = \beta$, ratios of expected sample sizes



(b) Optimally selected free parameters

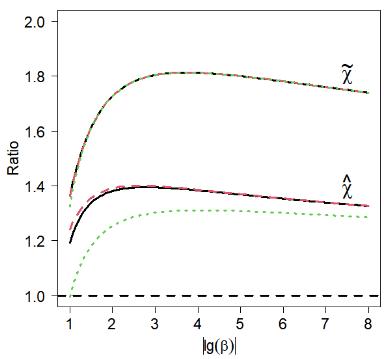


(c) $|\log \alpha| = 4\,|\log \beta|$

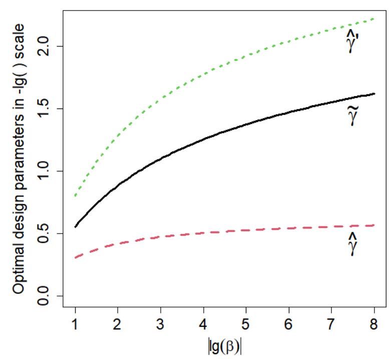


(d)

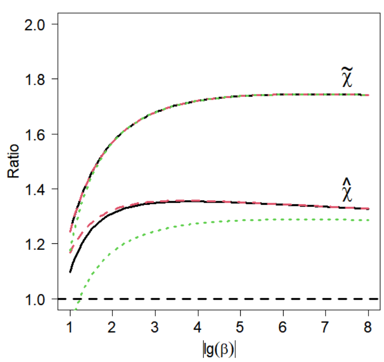
Table 2: In this and the next tables, each row represents an asymptotic relationship between α and β in (S4.44), the left column draws $\mathsf{E}_0[\tilde{\tau}]/\mathsf{E}_0[\tau']$ and $\mathsf{E}_0[\hat{\tau}]/\mathsf{E}_0[\tau']$, along with the corresponding bounds from Section 2, against $|\lg \beta|$, in testing the mean of an iid Gaussian sequence with unit variance, and the right column draws the corresponding optimally selected $|\lg \tilde{\gamma}|$ in $\tilde{\chi}$ and $|\lg \hat{\gamma}|, |\lg \hat{\gamma'}|$ in $\hat{\chi}$, against $|\lg \beta|$.



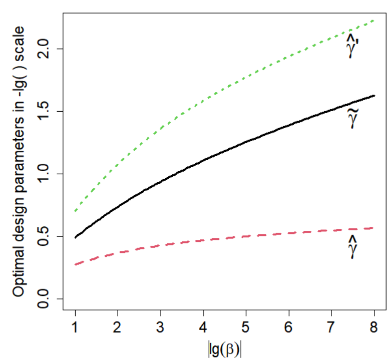
(a) $|\log \alpha| = |\log \beta|^{1.5}$



(b)



(c) $|\log \alpha| = |\log \beta|/\beta^{0.08}$



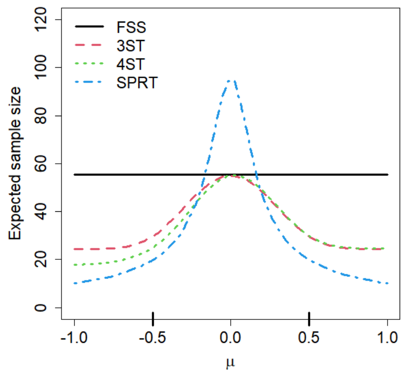
(d)

Table 3

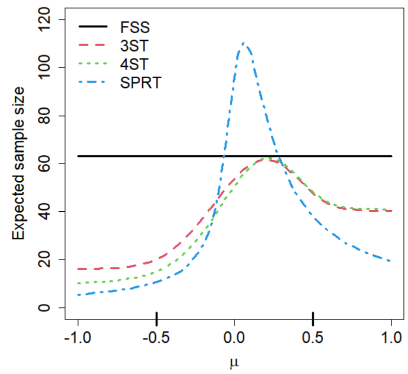
to the expected sample sizes in all cases, even for large values of α and β. On the other hand, the lower bounds are similarly accurate for $\tilde{\chi}$, but relatively conservative for $\hat{\chi}$. To illustrate the selection of the free parameters of the two multistage tests, in the right column of Figure 2 and 3 we plot $\tilde{\gamma}$ in $\tilde{\chi}$ and $\hat{\gamma}, \hat{\gamma}'$ in $\hat{\chi}$, against β, all of them in the $|\lg(\cdot)|$ scale.

S4.4 The second study

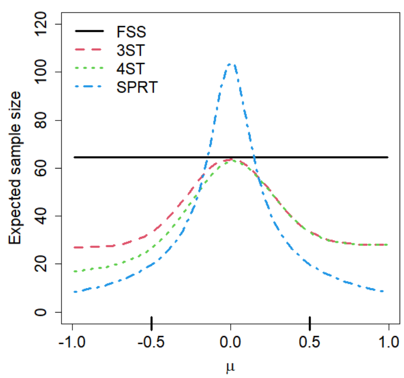
In the second study we compare the expected sample sizes of the four tests when the true distribution is not necessarily P_0 or P_1. Specifically, we compute $\mathbb{E}_\mu[\tau]$ for different values of μ, in each of the three testing problems considered in Section S3, when $\alpha = \beta = 10^{-4}$ and when $\alpha = 10^{-8}$, $\beta = 10^{-2}$. The results are presented in Figure 4, where each row represents a testing problem, each column represents a pair of error probabilities, and the values of the parameter at which we control the error probabilities are highlighted on the x-axis. Consistent with our discussion in Subsection 4.3.1, we can see that when the true parameter is around the middle of μ_0 and μ_1, the expected sample sizes of the multistage tests are much smaller than that of the SPRT. On the other hand, the expected sample size of the SPRT further decreases to the left of μ_0 and to the right of μ_1, while that of the multistage tests is lower bounded, at least by $n_0 \wedge n_1$.



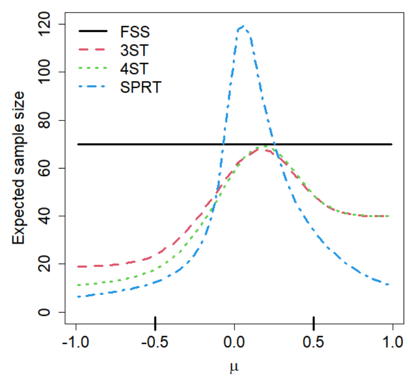
(a) IID Gaussian, $\alpha = \beta = 10^{-4}$



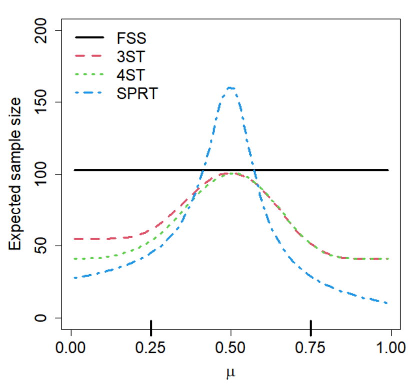
(b) $\alpha = 10^{-8}, \beta = 10^{-2}$



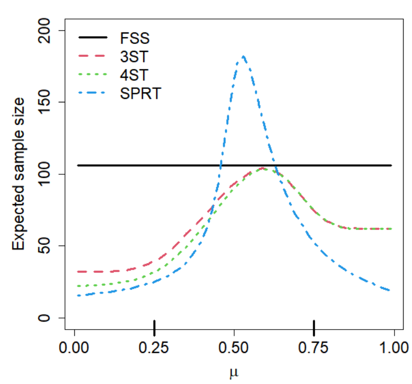
(c) AR(1), $\alpha = \beta = 10^{-4}$



(d) $\alpha = 10^{-8}, \beta = 10^{-2}$



(e) Two-state Markov, $\alpha = \beta = 10^{-4}$



(f) $\alpha = 10^{-8}, \beta = 10^{-2}$

Table 4

S5 Proofs

This section is devoted to presenting all the proofs that are omitted in the main content and in Section S1 of the supplement. It is organized as follows. In Section S5.1, we list some preliminary results about asymptotic bounds and approximations to the fixed-sample-size test, which, as mentioned at the beginning of Subsection 4.3, are the building blocks for the asymptotic analysis of the proposed multistage tests. In Section S5.2 we prove these preliminary results. In Subsection S5.3 we prove Lemma 1, Theorem 3 and Theorem 4, which are the main theoretical results of this work. In Subsection S5.4, we prove Theorem S1 in Section S1.

S5.1 Asymptotic analysis for the fixed-sample-size test

The asymptotic results in this subsection are based on the assumptions of Subsection 4.1 and 4.2, and, unless specifically mentioned, hold *as at least one of* α *and* β *goes to 0, while the other one either goes to 0 as well or remains fixed*. When any of these asymptotic regimes holds, we simply write $\alpha \wedge \beta \to 0$.

Asymptotic bounds

Theorem S2. *As* $\alpha \wedge \beta \to 0$,

$$\min\left\{\frac{|\log\beta|}{\psi_1(\kappa)}, \frac{|\log\alpha|}{\psi_0(\kappa)}\right\} \lesssim n^*(\alpha,\beta) \lesssim \max\left\{\frac{|\log\beta|}{\psi_1(\kappa)}, \frac{|\log\alpha|}{\psi_0(\kappa)}\right\} \tag{S5.45}$$

for every $\kappa \in (J_0, J_1)$, *and consequently*

$$n^*(\alpha,\beta) \lesssim \frac{|\log(\alpha\wedge\beta)|}{C}, \tag{S5.46}$$

where C *is defined as in* (4.13).

We present the following asymptotic lower bounds separately when $T = \bar{\Lambda}$ and when $T \neq \bar{\Lambda}$, as in the latter case we also need assumption (4.11).

Theorem S3. *(i) If* $T = \bar{\Lambda}$, *then*

$$n^*(\alpha,\beta) \gtrsim \max\left\{\frac{|\log\beta|}{I_0}, \frac{|\log\alpha|}{I_1}\right\} \quad as \quad \alpha\wedge\beta \to 0. \tag{S5.47}$$

(ii) If $T \neq \bar{\Lambda}$ *and* (4.11) *holds, then*

$$n^*(\alpha,\beta) \gtrsim \max\left\{\frac{|\log\beta|}{\psi_1(J_0)}, \frac{|\log\alpha|}{\psi_0(J_1)}\right\} \quad as \quad \alpha\wedge\beta \to 0. \tag{S5.48}$$

Asymptotic approximations

Unlike the preceding bounds, asymptotic approximations to $n^*(\alpha,\beta)$ depend on the relative decay rate of α and β. To describe them, we introduce the following function:

$$g(\kappa) \equiv \frac{\psi_0(\kappa)}{\psi_1(\kappa)}, \quad \kappa \in (J_0, J_1), \tag{S5.49}$$

which, by assumption, is continuous and strictly increasing in (J_0, J_1) with $g(J_0+) = 0$ and $g(J_1-) = \infty$. As a result, its inverse, g^{-1}, is well-defined in $(0, \infty)$ and satisfies $g^{-1}(0, \infty) = (J_0, J_1)$.

We start with the asymptotic regime where $\alpha, \beta \to 0$ so that

$$|\log \alpha| \sim r\, |\log \beta| \quad \text{for some} \quad r \in (0, \infty), \tag{S5.50}$$

in which case the approximation is expressed in terms of the function g.

Corollary S4. *As $\alpha, \beta \to 0$ so that (S5.50) holds,*

$$n^*(\alpha, \beta) \sim \frac{|\log \alpha|}{\psi_0(g^{-1}(r))} \sim \frac{|\log \beta|}{\psi_1(g^{-1}(r))}. \tag{S5.51}$$

When in particular, $r = 1$,

$$n^*(\alpha, \beta) \sim \frac{|\log \alpha|}{C} \sim \frac{|\log \beta|}{C}. \tag{S5.52}$$

Remark S5. From the previous corollary and the asymptotically optimal performance in (4.5) we obtain the asymptotic relative efficiency of the fixed-sample-size test as $\alpha, \beta \to 0$ so that (S5.50) holds. Specifically,

$$n^*(\alpha, \beta) \sim \frac{I_1}{\psi_0(g^{-1}(r))} \mathcal{L}_1(\alpha, \beta) \sim \frac{I_0}{\psi_1(g^{-1}(r))} \mathcal{L}_0(\alpha, \beta), \tag{S5.53}$$

and when in particular $r = 1$,

$$n^*(\alpha, \beta) \sim \frac{I_1}{C} \mathcal{L}_1(\alpha, \beta) \sim \frac{I_0}{C} \mathcal{L}_0(\alpha, \beta). \tag{S5.54}$$

When $\alpha \wedge \beta \to 0$ so that $|\log \alpha|/|\log \beta|$ either goes to zero or diverges, the asymptotic lower bounds in Theorem S3 turn out to be sharp.

Corollary S5. *Let* $T = \bar{\Lambda}$.

(i) If $\alpha \wedge \beta \to 0$ *so that* $|\log \alpha| \ll |\log \beta|$, *then* $n^*(\alpha, \beta) \sim |\log \beta|/I_0$.

(ii) If $\alpha \wedge \beta \to 0$ *so that* $|\log \alpha| \gg |\log \beta|$, *then* $n^*(\alpha, \beta) \sim |\log \alpha|/I_1$.

Corollary S6. *Let* $T \neq \bar{\Lambda}$ *and assume that* (4.11) *holds.*

(i) If $\alpha \wedge \beta \to 0$ *so that* $|\log \alpha| \ll |\log \beta|$, *then* $n^*(\alpha, \beta) \sim |\log \beta|/\psi_1(J_0)$.

(ii) If $\alpha \wedge \beta \to 0$ *so that* $|\log \alpha| \gg |\log \beta|$, *then* $n^*(\alpha, \beta) \sim |\log \alpha|/\psi_0(J_1)$.

Remark S6. When $T = \bar{\Lambda}$ and one of α and β is fixed, Corollary S5 is known as *Stein's lemma* (see, e.g., (Dembo and Zeitouni, 1998, Lemma 3.4.7)). We stress, however, that both α and β are allowed to go to 0 in the previous corollaries.

When both α and β go to 0, Corollary S5, in conjunction with (4.5), implies that the fixed-sample-size test is asymptotically optimal under one of the two hypotheses, while being of a larger order of magnitude compared to the optimal under the other hypothesis. This is formalized in the following corollary.

Corollary S7. *Let* $T = \bar{\Lambda}$.

(i) If $\alpha, \beta \to 0$ *so that* $|\log \alpha| \ll |\log \beta|$, *then*

$$\mathcal{L}_1(\alpha, \beta) \ll n^*(\alpha, \beta) \sim \mathcal{L}_0(\alpha, \beta).$$

(ii) If $\alpha, \beta \to 0$ *so that* $|\log \alpha| \gg |\log \beta|$*, then*

$$\mathcal{L}_0(\alpha, \beta) \ll n^*(\alpha, \beta) \sim \mathcal{L}_1(\alpha, \beta).$$

We end this subsection with the corresponding result when $T \neq \bar{\Lambda}$.

Corollary S8. *Let* $T \neq \bar{\Lambda}$ *and assume that* (4.11) *holds.*

(i) If $\alpha, \beta \to 0$ *so that* $|\log \alpha| \ll |\log \beta|$*, then*

$$\mathcal{L}_1(\alpha, \beta) \ll n^*(\alpha, \beta) \sim \frac{I_0}{\psi_1(J_0)} \mathcal{L}_0(\alpha, \beta).$$

(ii) If $\alpha, \beta \to 0$ *so that* $|\log \alpha| \gg |\log \beta|$*, then*

$$\mathcal{L}_0(\alpha, \beta) \ll n^*(\alpha, \beta) \sim \frac{I_1}{\psi_0(J_1)} \mathcal{L}_1(\alpha, \beta).$$

S5.2 Proof of results in Section S5.1

We start with a preliminary lemma, which holds under only some of the assumptions of Section 4.

Lemma S1. *(i) If, for every* $n \in \mathbb{N}$, P_1 *and* P_0 *are mutually absolutely continuous when restricted to* $\mathcal{F}_n$*, then*

$$n^*(\alpha, \beta) \to \infty \quad \textit{as} \quad \alpha \wedge \beta \to 0.$$

(ii) If also (4.8) *holds, then*

$$J_0 \leq \varliminf \kappa^*(\alpha, \beta) \quad \textit{and} \quad \varlimsup \kappa^*(\alpha, \beta) \leq J_1 \quad \textit{as} \quad \alpha \wedge \beta \to 0.$$

Proof. (i) Since $n^*(\cdot,\cdot)$ is decreasing in both its arguments, it suffices to show $n^*(\alpha,\beta)$ goes to infinity when one of α and β goes to 0, while the other one is fixed. Without loss of generality, we assume that α is fixed and $\beta \to 0$. We argue by contradiction and suppose that $n^*(\alpha,\beta) \not\to \infty$ as $\beta \to 0$. From this assumption and the fact that n^* is decreasing in both its arguments we conclude that there exists an $m \in \mathbb{N}$ and a sequence (β_n) with $\beta_n \to 0$ such that $n^*(\alpha,\beta_n) = m$ for any $n \in \mathbb{N}$. Then, for every $n \in \mathbb{N}$ we have $\kappa^*(\alpha,\beta_n) \geq z_\alpha$, where

$$z_\alpha \equiv \inf\{z \in \mathbb{R} : \mathsf{P}_0(T_m > z) \leq \alpha\} > -\infty,$$

and subsequently

$$\beta_n \geq \mathsf{P}_1(T_{n^*(\alpha,\beta_n)} \leq \kappa^*(\alpha,\beta_n)) = \mathsf{P}_1(T_m \leq \kappa^*(\alpha,\beta_n)) \geq \mathsf{P}_1(T_m \leq z_\alpha).$$

Letting $n \to \infty$ we obtain $\mathsf{P}_1(T_m \leq z_\alpha) = 0$. By the definition of z_α we also have $\mathsf{P}_0(T_m \leq z_\alpha) \geq 1-\alpha > 0$. This violates the assumption that P_0 is absolutely continuous to P_1 when restricted to $\mathcal{F}_m$, thus, we have reached a contradiction.

(ii) We only prove the first inequality, as the proof of the second is similar. We argue by contradiction and suppose that $\underline{\lim}\, \kappa^*(\alpha,\beta) < J_0$ as $\alpha \wedge \beta \to 0$. Then there exists an $\epsilon > 0$ and a sequence $(\alpha_n,\beta_n)_{n\in\mathbb{N}}$ such that at least one of $(\alpha_n)_{n\in\mathbb{N}}$ or $(\beta_n)_{n\in\mathbb{N}}$ converges to zero, both are bounded away from one, and $\kappa^*(\alpha_n, \beta_n) \leq J_0 - \epsilon$ for every $n \in \mathbb{N}$. Then for every $n \in \mathbb{N}$,

$$\alpha_n \geq \mathsf{P}_0(T_{n^*(\alpha_n,\beta_n)} > \kappa^*(\alpha_n, \beta_n)) \geq \mathsf{P}_0(T_{n^*(\alpha_n,\beta_n)} > J_0 - \epsilon).$$

In view of (i) and assumption (4.8), the lower bound goes to 1 as $n \to \infty$, which contradicts the fact that the sequence $(\alpha_n)_{n\in\mathbb{N}}$ is bounded away from one.

□

Proof of Theorem S2. The upper bound in (S5.45) implies that

$$n^*(\alpha, \beta) \lesssim \frac{|\log \alpha| \vee |\log \beta|}{\psi_1(\kappa) \wedge \psi_0(\kappa)} = \frac{|\log(\alpha \wedge \beta)|}{\psi_1(\kappa) \wedge \psi_0(\kappa)} \quad \text{for every } \kappa \in (J_0, J_1),$$

and optimizing with respect to κ we obtain (S5.46). Therefore, it suffices to show (S5.45).

We first fix arbitrary $\alpha, \beta \in (0, 1)$. To lighten the notation, we write $n^*(\alpha, \beta)$ and $\kappa^*(\alpha, \beta)$ as n^* and κ^*. By the definitions of these quantities we have

$$\mathsf{P}_0(T_{n^*} > \kappa^*) \leq \alpha \quad \text{and} \quad \mathsf{P}_1(T_{n^*} \leq \kappa^*) \leq \beta = \alpha^{\frac{|\log \beta|}{|\log \alpha|}},$$

and as a result

$$\max\left\{\mathsf{P}_0(T_{n^*} > \kappa^*),\ \mathsf{P}_1(T_{n^*} \leq \kappa^*)^{\frac{|\log \alpha|}{|\log \beta|}}\right\} \leq \alpha.$$

Since for any $n \in \mathbb{N}$ and $\kappa_1, \kappa_2 \in \mathbb{R}$,

$$\text{either} \quad \mathsf{P}_0(T_n > \kappa_1) \geq \mathsf{P}_0(T_n > \kappa_2) \quad \text{or} \quad \mathsf{P}_1(T_n \leq \kappa_1) \geq \mathsf{P}_1(T_n \leq \kappa_2),$$

we have, for any $\kappa \in (J_0, J_1)$,

$$\min\left\{\mathsf{P}_0(T_{n^*} > \kappa),\ \mathsf{P}_1(T_{n^*} \leq \kappa)^{\frac{|\log\alpha|}{|\log\beta|}}\right\} \leq \alpha,$$

and consequently

$$\frac{\min\left\{\frac{1}{n^*}\log \mathsf{P}_0(T_{n^*} > \kappa),\ \frac{1}{n^*}\log \mathsf{P}_1(T_{n^*} \leq \kappa) \cdot \frac{|\log\alpha|}{|\log\beta|}\right\}}{\frac{1}{n^*}\log\alpha} \geq 1. \tag{S5.55}$$

On the other hand, the definition of n^* also implies

$$\text{either} \quad \alpha < \mathsf{P}_0(T_{n^*-1} > \kappa^*) \quad \text{or} \quad \beta < \mathsf{P}_1(T_{n^*-1} \leq \kappa^*),$$

and consequently

$$\alpha < \max\left\{\mathsf{P}_0(T_{n^*-1} > \kappa^*),\ \mathsf{P}_1(T_{n^*-1} \leq \kappa^*)^{\frac{|\log\alpha|}{|\log\beta|}}\right\}.$$

Similarly we conclude that, for any $\kappa \in (J_0, J_1)$,

$$\frac{\max\left\{\frac{1}{n^*-1}\log \mathsf{P}_0(T_{n^*-1} > \kappa),\ \frac{1}{n^*-1}\log \mathsf{P}_1(T_{n^*-1} \leq \kappa) \cdot \frac{|\log\alpha|}{|\log\beta|}\right\}}{\frac{1}{n^*-1}\log\alpha} < 1. \tag{S5.56}$$

Now, we let $\alpha \wedge \beta \to 0$ in (S5.55) and (S5.56). Based on Lemma S1 and (4.9)-(4.10) we conclude that for any $\kappa \in (J_0, J_1)$,

$$\varliminf \frac{n^*(\alpha,\beta)}{\log\alpha} \min\left\{-\psi_0(\kappa),\ -\psi_1(\kappa) \cdot \frac{|\log\alpha|}{|\log\beta|}\right\} \geq 1,$$

and

$$\overline{\lim} \frac{n^*(\alpha,\beta)}{\log\alpha} \max\left\{-\psi_0(\kappa),\ -\psi_1(\kappa)\cdot\frac{|\log\alpha|}{|\log\beta|}\right\} \leq 1.$$

which complete the proof of (S5.45).

□

Proof of Theorem S3. (i) When both α and β go to 0, this follows from the universal asymptotic lower bound in (4.5). Therefore, it suffices to consider the case that only one of them goes to 0, while the other one is fixed. Without loss of generality, we assume that $\beta \to 0$, while α is fixed, in which case it suffices to show that, for every $\epsilon > 0$,

$$\underline{\lim} \frac{\log\beta}{n^*(\alpha,\beta)} \geq -I_0 - \epsilon.$$

To this end, we fix $\epsilon > 0$ and observe that, by Lemma S1.(ii), for β small enough we have $\kappa^*(\alpha,\beta) > -I_0 - \epsilon$ and consequently

$$\begin{aligned}
\beta &\geq \mathsf{P}_1(\bar{\Lambda}_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)) \\
&\geq \mathsf{P}_1(-I_0 - \epsilon < \bar{\Lambda}_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)) \\
&= \mathsf{E}_0\left[\exp\{\Lambda_{n^*(\alpha,\beta)}\};\ -I_0 - \epsilon < \bar{\Lambda}_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)\right] \\
&\geq \exp\{-n^*(\alpha,\beta)\,(I_0+\epsilon)\}\cdot \mathsf{P}_0(-I_0 - \epsilon < \bar{\Lambda}_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)).
\end{aligned} \tag{S5.57}$$

Moreover, for any $\alpha, \beta \in (0,1)$ we have

$$\begin{aligned} &\mathsf{P}_0(-I_0 - \epsilon < \bar{\Lambda}_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)) \\ &= 1 - \mathsf{P}_0(\bar{\Lambda}_{n^*(\alpha,\beta)} \leq -I_0 - \epsilon) - \mathsf{P}_0(\bar{\Lambda}_{n^*(\alpha,\beta)} > \kappa^*(\alpha,\beta)) \qquad \text{(S5.58)} \\ &\geq \; 1 - \mathsf{P}_0(\bar{\Lambda}_{n^*(\alpha,\beta)} \leq -I_0 - \epsilon) - \alpha, \end{aligned}$$

and the probability in the lower bound of (S5.58) goes to zero as $\beta \to 0$ because of Lemma S1.(i) and assumption (4.2). Therefore, taking logarithms on both sides of (S5.57), dividing by $n^*(\alpha,\beta)$ and letting $\beta \to 0$ complete the proof.

(ii) We only prove that, as $\alpha \wedge \beta \to 0$,

$$n^*(\alpha,\beta) \gtrsim \frac{|\log \beta|}{\psi_1(J_0)},$$

as the proof that $n^*(\alpha,\beta) \gtrsim |\log \alpha|/\psi_0(J_1)$ is similar. By assumption (4.11), there is an $\epsilon > 0$ so that ψ_1 is finite and (4.10) holds in $(J_0 - 2\epsilon, J_1)$. From Lemma S1.(ii) it follows that, when at least one of α and β is small enough, $\kappa^*(\alpha,\beta) > J_0 - \epsilon$ and consequently

$$\beta \geq \mathsf{P}_1(T_{n^*(\alpha,\beta)} \leq \kappa^*(\alpha,\beta)) \geq \mathsf{P}_1(T_{n^*(\alpha,\beta)} \leq J_0 - \epsilon).$$

Thus, taking logarithms, dividing by $n^*(\alpha,\beta)$ and letting $\alpha \wedge \beta \to 0$ we obtain

$$\underline{\lim}\, \frac{\log \beta}{n^*(\alpha,\beta)} \geq \underline{\lim}\, \frac{1}{n^*(\alpha,\beta)} \log \mathsf{P}_1(T_{n^*(\alpha,\beta)} \leq J_0 - \epsilon) = -\psi_1(J_0 - \epsilon),$$

where the equality follows from Lemma S1.(i) and assumption (4.10). Since ψ_1 is convex, it is continuous in the interior of its effective domain. Thus, letting $\epsilon \downarrow 0$ completes the proof.

□

Proof of Corollary S4. The first asymptotic approximation in (S5.51) follows by setting $\kappa = g^{-1}(r)$ in (S5.45), whereas the second by the definition of g which implies

$$\psi_0\left(g^{-1}(r)\right) = r\,\psi_1\left(g^{-1}(r)\right) \quad \text{for any } r \in (0,\infty).$$

To prove (S5.52) it suffices to show that

$$C = \psi_0(g^{-1}(1)) = \psi_1(g^{-1}(1)).$$

Indeed, the strict monotonicity of ψ_0 and ψ_1 in (J_0, J_1) implies the supremum in (4.13) is attained when $\psi_0 = \psi_1$, or equivalently, when $g = 1$.

□

Proofs of Corollaries S5 and S6. In view of the asymptotic lower bounds in Theorem S3, it satisfies to establish only the corresponding upper bounds. We only prove part (i) of each Corollary, as the proof of (ii) is similar.

We first show that, for any test statistic T, *even if* (4.11) *does not hold*,

$$n^*(\alpha,\beta) \lesssim \frac{|\log\beta|}{\psi_1(J_0)}, \quad \text{or equivalently,} \quad \psi_1(J_0) \lesssim \frac{|\log\beta|}{n^*(\alpha,\beta)},$$

as $\alpha \wedge \beta \to 0$ so that $|\log \alpha| \ll |\log \beta|$.

By assumption, ψ_1 is continuous in $[J_0, J_1]$. Therefore, to prove the above claim it suffices to show that, as $\alpha \wedge \beta \to 0$ so that $|\log \alpha| \ll |\log \beta|$,

$$\psi_1(\kappa) \lesssim \frac{|\log \beta|}{n^*(\alpha, \beta)}, \quad \text{for any } \kappa \in (J_0, J_1),$$

which follows directly by Theorem S2. When $T \neq \bar{\Lambda}$, the proof is complete. When $T = \bar{\Lambda}$, it remains to show that $\psi_1(-I_0) \geq I_0$. Since ψ_1 is continuous in $[-I_0, I_1]$, it suffices to show that $\psi_1(\kappa) \geq -\kappa$ for every $\kappa \in (-I_0, 0)$. Indeed, for any $\kappa < 0$, by Markov's inequality we have

$$\mathsf{P}_1(\bar{\Lambda}_n \leq \kappa) \leq e^{n\kappa}\, \mathsf{E}_0[\exp\{\Lambda_n\}] = e^{n\kappa} \quad \text{for all} \quad n \in \mathbb{N}.$$

Taking logarithms, dividing by n, letting $n \to \infty$, and applying (4.10) for κ in $(-I_0, 0)$ complete the proof.

□

S5.3 Proof of results in Section 4.3

Proof of Lemma 1. We only prove the asymptotic lower bounds under P_0, as those under P_1 are similar. We first prove the result for $\tilde{\chi}$, in which case it suffices to show that, for any $\epsilon \in (0, 1)$,

$$\inf_{\gamma \in \left[\frac{\alpha \vee \beta}{2}, 1\right)} \mathsf{E}_0[\tilde{\tau}] \gtrsim (1 - \epsilon)\, \frac{|\log \beta|}{\psi_1(J_0)} \quad \text{as} \quad \alpha, \beta \to 0. \tag{S5.59}$$

Fix $\epsilon \in (0,1)$. By the non-asymptotic lower bound in (3.12) it follows that, for any $\alpha, \beta \in (0,1)$ and $\gamma \in [(\alpha \vee \beta)/2, 1)$,

$$\mathsf{E}_0[\tilde{\tau}] \geq \max\left\{n^*(\gamma, \beta/2) \cdot (1-\alpha/2),\ n^*(\alpha/2, \beta/2) \cdot (\gamma - \alpha/2)\right\}.$$

When, in particular, $\gamma \leq 1-\epsilon$,

$$\mathsf{E}_0[\tilde{\tau}] \geq n^*(\gamma, \beta/2) \cdot (1-\alpha/2) \geq n^*(1-\epsilon,\ \beta/2) \cdot (1-\alpha/2).$$

and when $\gamma \geq 1-\epsilon$,

$$\mathsf{E}_0[\tilde{\tau}] \geq n^*(\alpha/2,\ \beta/2) \cdot (\gamma-\alpha/2) \geq n^*(\alpha/2,\ \beta/2) \cdot (1-\epsilon-\alpha/2).$$

By Theorem S3.(ii) it then follows that, as $\alpha, \beta \to 0$,

$$\inf_{\gamma \in \left[\frac{\alpha \vee \beta}{2}, 1-\epsilon\right]} \mathsf{E}_0[\tilde{\tau}] \gtrsim n^*(1-\epsilon,\ \beta/2) \gtrsim \frac{|\log \beta|}{\psi_1(J_0)},$$

$$\inf_{\gamma \in [1-\epsilon, 1)} \mathsf{E}_0[\tilde{\tau}] \gtrsim (1-\epsilon) \cdot n^*(\alpha/2, \beta/2) \gtrsim (1-\epsilon)\, \frac{|\log \beta|}{\psi_1(J_0)},$$

and they imply (S5.59).

The proof for $\check{\chi}$ is similar and omitted.

To prove the result for $\hat{\chi}$, it suffices to show that, for any $\epsilon \in (0,1)$,

$$\inf_{\left(\frac{\alpha}{2} \vee \frac{\beta}{3}\right) \leq \gamma' \leq \gamma < 1} \mathsf{E}_0[\hat{\tau}] \gtrsim (1-2\epsilon)\, \frac{|\log \beta|}{\psi_1(J_0)} \quad \text{as} \quad \alpha, \beta \to 0. \tag{S5.60}$$

Fix $\epsilon \in (0,1)$. By the non-asymptotic lower bound in (3.25) it follows that, for any $\alpha, \beta \in (0,1)$ and $(\alpha/2) \vee (\beta/3) \leq \gamma' \leq \gamma < 1$,

$$\mathsf{E}_0[\hat{\tau}] \geq \max\left\{n^*(\gamma,\ \beta/3) \cdot (1-\alpha/2),\ n^*(\gamma', \beta/3) \cdot (\gamma - \alpha/2),\right.$$
$$\left. n^*(\alpha/2, \beta/3) \cdot ((1-\alpha/2) - (1-\gamma) - (1-\gamma'))\right\}.$$

When, in particular, $\gamma \le 1-\epsilon$,

$$\mathsf{E}_0[\hat{\tau}] \ge n^*(\gamma,\, \beta/3) \cdot (1-\alpha/2) \ge n^*(1-\epsilon,\, \beta/3) \cdot (1-\alpha/2),$$

when $\gamma' \le 1-\epsilon \le \gamma$,

$$\mathsf{E}_0[\hat{\tau}] \ge n^*(\gamma',\, \beta/3) \cdot (\gamma-\alpha/2) \ge n^*(1-\epsilon,\, \beta/3) \cdot (1-\epsilon-\alpha/2),$$

and when $\gamma' \ge 1-\epsilon$,

$$\begin{aligned}\mathsf{E}_0[\hat{\tau}] &\ge n^*(\alpha/2,\, \beta/3) \cdot ((1-\alpha/2)-(1-\gamma)-(1-\gamma')) \\ &\ge n^*(\alpha/2,\, \beta/3) \cdot (1-2\epsilon-\alpha/2).\end{aligned}$$

By Theorem S3.(ii) it then follows that, as $\alpha, \beta \to 0$,

$$\begin{aligned}\inf_{(\frac{\alpha}{2}\vee\frac{\beta}{3})\le\gamma'\le\gamma\le 1-\epsilon} \mathsf{E}_0[\hat{\tau}] &\gtrsim n^*(1-\epsilon,\, \beta/3) \gtrsim \frac{|\log\beta|}{\psi_1(J_0)}, \\ \inf_{(\frac{\alpha}{2}\vee\frac{\beta}{3})\le\gamma'\le 1-\epsilon\le\gamma} \mathsf{E}_0[\hat{\tau}] &\gtrsim n^*(1-\epsilon,\, \beta/3)\cdot(1-\epsilon) \gtrsim (1-\epsilon)\,\frac{|\log\beta|}{\psi_1(J_0)}, \\ \inf_{1-\epsilon\le\gamma'\le\gamma<1} \mathsf{E}_0[\hat{\tau}] &\gtrsim n^*(\alpha/2,\, \beta/3)\cdot(1-2\epsilon) \gtrsim (1-2\epsilon)\,\frac{|\log\beta|}{\psi_1(J_0)},\end{aligned}$$

which together imply (S5.60).

□

The proof of Theorem 3 is omitted, as it is almost identical to that of Theorem 4. For the proof of the latter, in view of Lemma 1, it suffices to prove in each case the corresponding asymptotic upper bounds.

Our technique is to show that the asymptotic upper bounds hold for the proposed tests as long as the free parameters are selected to meet certain very mild asymptotic relationships with α and β, these asymptotic relationships must be satisfied by a sequence of points on the grids, and thus, the asymptotic upper bounds must hold if the free parameters are selected as the minimizers over the grids.

Proof of Theorem 4. (i) We first prove the result for $\tilde{\chi}$. First, suppose that there is a $\delta_{\alpha,\beta}$ such that $\delta_{\alpha,\beta} \in [(\alpha \vee \beta)/2, 1)$ for every $\alpha, \beta \in (0,1)$, and

$$|\log \delta_{\alpha,\beta}| \ll |\log \alpha|, \quad |\log(\alpha \wedge \beta)|\, \delta_{\alpha,\beta} \ll |\log \alpha| \tag{S5.61}$$

as $\alpha, \beta \to 0$. Then, if we select δ as $\delta_{\alpha,\beta}$, by (3.13), Corollary S5.(ii) and Theorem S2 we have, as $\alpha, \beta \to 0$,

$$\begin{aligned}
\mathsf{E}_1[\tilde{\tau}] &\leq n^*(\alpha/2, \delta_{\alpha,\beta}) + \big(n^*(\alpha/2, \beta/2) - n^*(\alpha/2, \delta_{\alpha,\beta})\big) \cdot \delta_{\alpha,\beta} \\
&\leq n^*(\alpha/2, \delta_{\alpha,\beta}) + n^*(\alpha/2, \beta/2) \cdot \delta_{\alpha,\beta} \\
&\lesssim \frac{|\log \alpha|}{\psi_0(J_1)} + \frac{|\log(\alpha \wedge \beta)|}{C}\, \delta_{\alpha,\beta} \sim \frac{|\log \alpha|}{\psi_0(J_1)}.
\end{aligned}$$

Second, for any $\alpha, \beta \in (0,1)$ let $\delta^{\tilde{L}}_{\alpha,\beta}$ denote the grid point on $\tilde{L}_{\alpha,\beta}$ that is the closest to $\delta_{\alpha,\beta}$. If, also, the grid-length goes to 0 so that

$$|\delta^{\tilde{L}}_{\alpha,\beta} - \delta_{\alpha,\beta}| \leq \tilde{l}_{\alpha,\beta} \ll \delta_{\alpha,\beta} \tag{S5.62}$$

as $\alpha, \beta \to 0$, then $\delta^{\tilde{L}}_{\alpha,\beta}$ satisfies (S5.61) as well and, consequently, the same asymptotic upper bound on $\mathsf{E}_1[\tilde{\tau}]$ holds for $\delta \equiv \delta^{\tilde{L}}_{\alpha,\beta}$ and, furthermore, for δ that is selected as the minimizer over $\tilde{L}_{\alpha,\beta}$.

Thus, it remains to show that it is possible to find such $\delta_{\alpha,\beta}$ and $\tilde{l}_{\alpha,\beta}$ that satisfy (S5.61)-(S5.62). This is not always possible if α and β go to 0 at arbitrary rates. However, when $\alpha, \beta \to 0$ so that $|\log \alpha| \gtrsim |\log \beta|$, an example of such a selection is

$$\begin{aligned} &\tilde{l}_{\alpha,\beta} = |\log(\alpha \wedge \beta)|^{-1} \quad \text{and} \\ &\delta_{\alpha,\beta} = |\log(\alpha \wedge \beta)|^{-\epsilon} \vee ((\alpha \vee \beta)/2) \end{aligned} \tag{S5.63}$$

for some $\epsilon \in (0, 1)$. This completes the proof for $\tilde{\chi}$.

The proof for $\hat{\chi}$ is similar and omitted.

To prove the result for $\check{\chi}$, we observe that by (3.28) it follows that

$$\begin{aligned} &n^*(\alpha/3, \delta) + (n^*(\alpha/3, \delta') - n^*(\alpha/3, \delta)) \cdot \delta + (n^*(\alpha/3, \beta/2) - n^*(\alpha/3, \delta') \cdot \delta' \\ &\leq n^*(\alpha/3, \delta) + n^*(\alpha/3, \beta/2) \cdot \delta \end{aligned}$$

for any $\alpha, \beta \in (0, 1)$ and δ, δ' such that $(\alpha/3) \vee (\beta/2) \leq \delta' \leq \delta < 1$. The proof then continues in exactly the same way as for $\tilde{\chi}$.

(ii) If $\gamma \equiv \gamma_{\alpha,\beta},\ \gamma' \equiv \gamma'_{\alpha,\beta}$ satisfy

$$(\alpha/2) \vee (\beta/3) \leq \gamma'_{\alpha,\beta} \leq \gamma_{\alpha,\beta} < 1$$

for every $\alpha, \beta \in (0,1)$, and

$$\begin{aligned}
|\log \gamma_{\alpha,\beta}| &\ll |\log \beta| \\
|\log(\gamma'_{\alpha,\beta} \wedge \beta)|\, \gamma_{\alpha,\beta} &\ll |\log \beta| \qquad \text{(S5.64)} \\
|\log(\alpha \wedge \beta)|\, \gamma'_{\alpha,\beta} &\ll |\log \beta|
\end{aligned}$$

as $\alpha, \beta \to 0$, then, by (3.25), Corollary S5.(i) and Theorem S2 we conclude that, as $\alpha, \beta \to 0$,

$$\begin{aligned}
\mathsf{E}_0[\hat{\tau}] &\leq n^*(\gamma_{\alpha,\beta}, \beta/3) + \big(n^*(\gamma'_{\alpha,\beta}, \beta/3) - n^*(\gamma_{\alpha,\beta}, \beta/3)\big) \cdot \gamma_{\alpha,\beta} \\
&\qquad + \big(n^*(\alpha/2, \beta/3) - n^*(\gamma'_{\alpha,\beta}, \beta/3)\big) \cdot \gamma'_{\alpha,\beta} \\
&\leq n^*(\gamma_{\alpha,\beta}, \beta/3) + n^*(\gamma'_{\alpha,\beta}, \beta/3) \cdot \gamma_{\alpha,\beta} + n^*(\alpha/2, \beta/3) \cdot \gamma'_{\alpha,\beta} \\
&\lesssim \frac{|\log \beta|}{\psi_1(J_0)} + \frac{|\log(\gamma'_{\alpha,\beta} \wedge \beta)|}{C}\, \gamma_{\alpha,\beta} + \frac{|\log(\alpha \wedge \beta)|}{C}\, \gamma'_{\alpha,\beta} \sim \frac{|\log \beta|}{\psi_1(J_0)}.
\end{aligned}$$

For any $\alpha, \beta \in (0,1)$, let $\gamma^{\hat{L}}_{\alpha,\beta}$ and $\gamma'^{\hat{L}}_{\alpha,\beta}$ denote the grid points on $\hat{L}_{\alpha,\beta}$ that are the closest to $\gamma_{\alpha,\beta}$ and $\gamma'_{\alpha,\beta}$ respectively. If the grid lengths satisfy

$$|\gamma^{\hat{L}}_{\alpha,\beta} - \gamma_{\alpha,\beta}|,\ |\gamma'^{\hat{L}}_{\alpha,\beta} - \gamma'_{\alpha,\beta}| \leq \hat{l}_{\alpha,\beta} \ll \gamma'_{\alpha,\beta},$$

as $\alpha, \beta \to 0$, then $\gamma^{\hat{L}}_{\alpha,\beta}$ and $\gamma'^{\hat{L}}_{\alpha,\beta}$ satisfy (S5.64) as well and, consequently, the same asymptotic upper bound on $\mathsf{E}_0[\hat{\tau}]$ holds for $\gamma \equiv \gamma^{\hat{L}}_{\alpha,\beta}$, $\gamma' \equiv \gamma'^{\hat{L}}_{\alpha,\beta}$ and, furthermore, for γ, γ' that are selected as the minimizers over $\hat{L}_{\alpha,\beta}$.

Now, it remains to find such $\gamma_{\alpha,\beta}, \gamma'_{\alpha,\beta}$ and $\hat{l}_{\alpha,\beta}$, when $\alpha, \beta \to 0$ so that

$|\log \beta| \lesssim |\log \alpha| \lesssim |\log \beta|/\beta^r$ for some $r > 0$. The following is an example:

$$\begin{aligned}
\hat{l}_{\alpha,\beta} &= |\log(\alpha \wedge \beta)|^{-1} \\
\gamma_{\alpha,\beta} &= |\log \beta|^{-\epsilon} \vee \frac{\alpha}{2} \vee \frac{\beta}{3}, \\
\gamma'_{\alpha,\beta} &= \left(\frac{|\log \beta|^{\epsilon'}}{|\log \alpha|} \wedge \gamma_{\alpha,\beta} \right) \vee \frac{\alpha}{2} \vee \frac{\beta}{3}
\end{aligned} \tag{S5.65}$$

for some $\epsilon, \epsilon' \in (0, 1)$.

(iii) To prove the result for $\tilde{\chi}$, note that if $\gamma \equiv \gamma_{\alpha,\beta}$ satisfies $\gamma_{\alpha,\beta} \in [(\alpha \vee \beta)/2, 1)$ for every $\alpha, \beta \in (0, 1)$, and

$$|\log \gamma_{\alpha,\beta}| \ll |\log \beta|, \quad |\log(\alpha \wedge \beta)|\, \gamma_{\alpha,\beta} \ll |\log \beta|, \tag{S5.66}$$

as $\alpha, \beta \to 0$, then, by (3.12), Corollary S5.(i) and Theorem S2 we conclude that, as $\alpha, \beta \to 0$,

$$\begin{aligned}
\mathsf{E}_0[\tilde{\tau}] &\leq n^*(\gamma_{\alpha,\beta}, \beta/2) + \big(n^*(\alpha/2, \beta/2) - n^*(\gamma_{\alpha,\beta}, \beta/2)\big) \cdot \gamma_{\alpha,\beta} \\
&\leq n^*(\gamma_{\alpha,\beta}, \beta/2) + n^*(\alpha/2, \beta/2) \cdot \gamma_{\alpha,\beta} \\
&\lesssim \frac{|\log \beta|}{\psi_1(J_0)} + \frac{|\log(\alpha \wedge \beta)|}{C}\, \gamma_{\alpha,\beta} \sim \frac{|\log \beta|}{\psi_1(J_0)}.
\end{aligned}$$

The rest then follows by a similar argument as in the previous cases. An example of $\gamma_{\alpha,\beta}$ and $\check{l}_{\alpha,\beta}$ that satisfy (S5.66) and $\check{l}_{\alpha,\beta} \ll \gamma_{\alpha,\beta}$ when $\alpha, \beta \to 0$

so that $|\log\beta| \lesssim |\log\alpha| \lesssim |\log\beta|^r$ for some $r \geq 1$ is:

$$\begin{aligned}\check{l}_{\alpha,\beta} &= |\log(\alpha\wedge\beta)|^{-1} \\ \gamma_{\alpha,\beta} &= \frac{|\log\beta|^{\epsilon}}{|\log\alpha|} \vee \frac{\alpha}{2} \vee \frac{\beta}{2} \quad \text{for some } \epsilon \in (0,1).\end{aligned} \tag{S5.67}$$

The proof for $\check{\chi}$ is essentially the same with $\alpha/2$ replaced by $\alpha/3$.

□

S5.4 Proof of Theorem S1

This is a version of the Gärtner-Ellis Theorem. Its proof follows the same technique as in (Dembo and Zeitouni, 1998, Theorem 2.3.6) or (Bucklew, 2010, Theorem 3.2.1), and is presented only for completeness. Specifically, we establish first the asymptotic upper bounds in (i) and (ii). Using these, we establish (iii). Finally, using (iii), we establish the asymptotic lower bounds in (i) and (ii).

Proof of Theorem S1. We establish the asymptotic upper bound only for (i), as the corresponding proof for (ii) is similar. Thus, we assume that $\Theta^o \cap (0,\infty) \neq \emptyset$. For any $\kappa_1, \kappa_2 \in \phi'(\Theta^o \cap (0,\infty))$ such that $\kappa_1 < \kappa_2$,

$$\phi^*(\kappa_1) = \vartheta(\kappa_1)\kappa_1 - \phi(\vartheta(\kappa_1)) < \vartheta(\kappa_1)\kappa_2 - \phi(\vartheta(\kappa_1)) \leq \phi^*(\kappa_2),$$

which proves that ϕ^* is strictly increasing in $\phi'(\Theta^o \cap (0,\infty))$. From (Dembo and Zeitouni, 1998, Lemma 2.2.5)) it follows that the Legendre-Fenchel transform of ϕ, ϕ^*, is non-negative and lower-semicontinuous, and these properties imply that

$$0 \leq \phi^* \left(\phi'(0+)\right) \leq \lim_{\theta \downarrow 0} \phi^* \left(\phi'(\theta)\right) = \lim_{\theta \downarrow 0} \left\{\theta \phi'(\theta) - \phi(\theta)\right\} = 0.$$

Since $\phi'(\Theta^o \cap (0,\infty))$ is an open interval (whose right endpoint may be infinity), to show that (S1.4) holds for every $\kappa \in \phi'(\Theta^o \cap (0,\infty))$ it suffices to show that it holds for every $\kappa \in \phi'((0,\theta_*))$, where $\theta_* \in \Theta^o \cap (0,\infty)$. Thus, we fix $\theta_* \in \Theta^o \cap (0,\infty)$ and denote $\phi'((0,\theta_*)) \equiv (a,b)$, where $a \equiv \phi'(0+)$ and $b \equiv \phi'(\theta_*)$.

For any $\kappa \in (a,b)$ and $\theta \in (0,\theta_*)$, we have

$$\mathsf{P}(T_n > \kappa) \leq \exp\{-n\,\theta\,\kappa\}\,\mathsf{E}\left[\exp\{n\,\theta\,T_n\}\right] = \exp\{-n(\theta\kappa - \phi_n(\theta))\},$$

which, after taking logarithm, dividing by n and letting $n \to \infty$, gives

$$\overline{\lim_{n}}\, \frac{1}{n} \log \mathsf{P}(T_n > \kappa) \leq -\left(\theta\kappa - \phi(\theta)\right).$$

Optimizing the right-hand-side with respect to $\theta \in (0,\theta_*)$, we obtain $-\phi^*(\kappa)$. The proof for the asymptotic upper bounds is complete.

Note that this asymptotic upper bound is non-trivial for every $\kappa \in (a,\infty)$ since $\phi^*(a) = 0$ and ϕ^* is strictly increasing in (a,b). Therefore, it implies that $\mathsf{P}\left(T_n - \phi'(0+) > \epsilon\right)$ is exponentially decaying for every $\epsilon > 0$. Similarly it follows that if $\Theta^o \cap (-\infty,0) \neq \emptyset$, then $\mathsf{P}\left(T_n - \phi'(0-) \leq -\epsilon\right)$ is exponentially decaying for every $\epsilon > 0$. From these observations we conclude that if $0 \in \Theta^o$, then $\mathsf{P}\left(|T_n - \phi'(0)| > \epsilon\right)$ is exponentially decaying for every $\epsilon > 0$, and as a result $\mathsf{P}\left(T_n \to \phi'(0)\right) = 1$. Therefore, (iii) follows using exactly the same argument as long as the sequence of functions

$$\lambda \in \mathbb{R} \to \frac{1}{n} \log \mathsf{E}_{\mathrm{Q}_\theta}\left[\exp\{n\lambda T_n\}\right], \quad n \in \mathbb{N}$$

satisfies the assumptions of the theorem, 0 belongs to the interior of the effective domain of its limit, and the derivative of its limit at 0 is $\phi'(\theta)$. To show this, we fix $\theta \in (0, \theta_*)$. Then, for any $\lambda \in \mathbb{R}$,

$$\begin{aligned}\mathsf{E}_{\mathrm{Q}_\theta}\left[\exp\{n\lambda T_n\}\right] &= \mathsf{E}\left[\exp\{n((\lambda+\theta)T_n - \phi_n(\theta))\}\right] \\ &= \exp\{n(\phi_n(\lambda+\theta) - \phi_n(\theta))\},\end{aligned}$$

and consequently

$$\lim_n \frac{1}{n} \log \mathsf{E}_{\mathrm{Q}_\theta}\left[\exp\{n\lambda T_n\}\right] = \phi(\lambda+\theta) - \phi(\theta).$$

The limit is finite for $\lambda \in (-\theta, \theta_* - \theta)$, which contains 0 in its interior, inherits all the smoothness properties of ϕ, and its derivative at $\lambda = 0$ is $\phi'(\theta)$. The proof for (iii) is complete.

It remains to prove the asymptotic lower bounds in (i) and (ii). Again, we only do so for (i), as the proof for (ii) is similar. Fix $\kappa \in (a, b)$. For any

$n \in \mathbb{N}$, $\theta \in (0, \theta_*)$ and $\epsilon \in (0, b - \kappa)$,

$$\begin{aligned}\mathsf{P}(T_n > \kappa) &= \mathsf{E}_{\mathrm{Q}_\theta}\left[\exp\{-n(\theta T_n - \phi_n(\theta))\};\, T_n > \kappa\right] \\ &\geq \mathsf{E}_{\mathrm{Q}_\theta}\left[\exp\{-n(\theta T_n - \phi_n(\theta))\};\, \kappa < T_n \leq \kappa + \epsilon\right] \\ &\geq \exp\{-n(\theta(\kappa+\epsilon) - \phi_n(\theta))\}\, \mathrm{Q}_\theta(\kappa < T_n \leq \kappa + \epsilon).\end{aligned}$$

If we now set $\theta = \vartheta(\kappa + \epsilon/2)$, take logarithms, divide by n and let $n \to \infty$, by (iii) we obtain

$$\varliminf_n \frac{1}{n} \log \mathsf{P}(T_n > \kappa) \geq -\vartheta(\kappa + \epsilon/2)(\kappa + \epsilon) + \phi(\vartheta(\kappa + \epsilon/2)).$$

To complete the proof, we let $\epsilon \downarrow 0$ and observe that the right-hand-side converges to $-\big(\vartheta(\kappa)\kappa - \phi(\vartheta(\kappa))\big) = -\phi^*(\kappa)$, since ϑ and ϕ are both continuous in the corresponding neighborhoods.

$\square$

Bibliography

Bercu, B., F. Gamboa, and A. Rouault (1997). Large deviations for quadratic forms of stationary gaussian processes. *Stochastic Processes and their Applications 71*, 75–90.

Brockwell, P. J. and R. A. Davis (1986). *Time Series: Theory and Methods.* Berlin, Heidelberg: Springer-Verlag.

Bucklew, J. (2010). *Introduction to Rare Event Simulation* (1st ed.). Springer Publishing Company, Incorporated.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications*. Springer.

Durrett, R. (2010). *Probability: Theory and Examples* (4th ed.). USA: Cambridge University Press.

Parthasarathy, K. R. (2005). *Probability measures on metric spaces*. American Mathematical Soc.

Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright St. Champaign, IL 61822, USA

E-mail: yimingx4@illinois.edu, fellouri@illinois.edu