Supplementary material of
"Sequential multiple testing of multiple composite hypotheses: An asymptotic optimality theory with general information functions"

Yiming Xing

Tongji University

This supplementary material is organized as follows: In Section S1, we illustrate the general theory through three concrete examples. In Section S2, we present the numerical studies of testing the correlation coefficient of autoregressive data. In Section S3, we present all proofs. In Section S4, we discuss about model misspecification.

S1 Examples

In this section we investigate three concrete examples: testing the mean of independent Gaussians with unequal variances, testing the correlation coefficient of AR(1) series, and testing the transition matrix of Markov chains. For each example, we derive the expressions of the test statistics and check all assumptions. The goal is to illustrate the general theory and demonstrate its generality.

S1.1 Testing the mean of independent Gaussians with unequal variances

We first consider an example where, for every $k \in [K]$, $\{X_k(n),\, n \in \mathbb{N}\}$ are independent, Gaussian, with equal but unknown mean $\theta_k \in \mathbb{R}$ and unequal but known variances $\sigma_k(n)^2 > 0$, i.e., $X_k(n) \sim N\left(\theta_k, \sigma_k(n)^2\right)$ for $n \in \mathbb{N}$. For every $k \in [K]$, consider testing

$$\theta_k \in \Theta_k^m \text{ for } m \in [M],$$

where $\Theta_k = [\underline{\theta}_k, \overline{\theta}_k]$, $M \geq 2$,

$$\Theta_k^1 = [\underline{\theta}_k^1, \overline{\theta}_k^1], \quad \ldots, \quad \Theta_k^M = [\underline{\theta}_k^M, \overline{\theta}_k^M],$$

and $\underline{\theta}_k \leq \underline{\theta}_k^1 \leq \overline{\theta}_k^1 < \underline{\theta}_k^2 \leq \cdots \leq \overline{\theta}_k^{M-1} < \underline{\theta}_k^M \leq \overline{\theta}_k^M \leq \overline{\theta}_k$ are real numbers.

We would like to find a sufficient condition on $\{\sigma_k(n)^2,\, n \in \mathbb{N}\}$ so that Assumption 1-3 can be satisfied. Note that when $\sigma_k(n)^2 \equiv \sigma_k^2$ for some constant $\sigma_k^2 > 0$ and all $n \in \mathbb{N}$, this has been checked and extensively studied in, e.g., Song and Fellouris (2019); Deshmukh et al. (2021); Xing and Fellouris (2025). However, the case with unequal variances is underexplored and reveals some interesting findings.

Since the testing problems in all streams are homogeneous, for simplicity of notations, in what follows we suppress the lower index $k \in [K]$.

First of all, for any distinct $\theta, \theta' \in \Theta$, we have

$$\ell(n;\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log\sigma(t)^2 - \frac{1}{2}\sum_{t=1}^{n}\frac{1}{\sigma(t)^2}(X(t)-\theta)^2,$$

$$\ell(n;\theta) - \ell(n;\theta') = (\theta-\theta')\sum_{t=1}^{n}\frac{1}{\sigma(t)}\frac{X(t)-\theta}{\sigma(t)} + \psi(n)\frac{(\theta-\theta')^2}{2},$$

where we denote

$$\psi(n) \equiv \sum_{t=1}^{n}\frac{1}{\sigma(t)^2}.$$

Assuming

$$\lim_{q\to 1}\lim_{n\to\infty}\frac{\psi(qn)}{\psi(n)} = 1 \quad \text{and} \quad \sum_{n=1}^{\infty} e^{-\psi(n)} < \infty, \tag{S1.1}$$

we claim that, for any $m \in [M]$ and $\theta \in \Theta \backslash \Theta^m$, Assumption 1 holds with

$$\begin{aligned} I^m(n;\theta) &\equiv \inf_{\theta'\in\Theta^m} I(n;\theta,\theta') \equiv \psi(n)\inf_{\theta'\in\Theta^m}\frac{(\theta-\theta')^2}{2}, \\ \text{where } I(n;\theta,\theta') &\equiv \mathrm{E}_\theta\left[\ell(n;\theta)-\ell(n;\theta')\right] = \psi(n)\frac{(\theta-\theta')^2}{2}. \end{aligned} \tag{S1.2}$$

Indeed, condition (2.5) follows from the first condition in (S1.1), and condition (2.7) follows from definition. To show condition (2.6), since $|\theta-\theta'| \leq \overline{\theta} - \underline{\theta}$, it suffices to show

$$Y(n) \equiv \frac{1}{\psi(n)}\sum_{t=1}^{n}\frac{1}{\sigma(t)}\frac{X(t)-\theta}{\sigma(t)} \xrightarrow{\textit{completely}} 0 \text{ under } \mathrm{P}_\theta. \tag{S1.3}$$

To see this, note that $Y(n) \sim N(0, 1/\psi(n))$ and, by Mill's inequality, for any $\epsilon > 0$,

$$\mathrm{P}_\theta\left(|Y(n)| \geq \epsilon\right) = \mathrm{P}\left(|N(0,1)| \geq \epsilon\sqrt{\psi(n)}\right) \leq \sqrt{\frac{2}{\pi}}\frac{1}{\epsilon\sqrt{\psi(n)}}e^{-\frac{\epsilon^2}{2}\psi(n)},$$

whose summation over n is finite provided the last condition in (S1.1).

Next, we study the test statistic $\hat{\ell}(n) - \ell^m(n)$ where, naturally, we use the global MLE as the estimator in the adaptive log-likelihood. The global MLE is

$$\hat{\theta}(n) \equiv \left(\left(\frac{1}{\psi(n)} \sum_{t=1}^{n} \frac{X(t)}{\sigma(t)^2} \right) \vee \underline{\theta} \right) \wedge \overline{\theta},$$

and the adaptive log-likelihood in (2.8) is

$$\hat{\ell}(n) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \log \sigma(t)^2 - \frac{1}{2} \sum_{t=1}^{n} \frac{1}{\sigma(t)^2} (X(t) - \hat{\theta}(t-1))^2.$$

For every $m \in [M]$, the local MLE in the m^{th} hypothesis is

$$\hat{\theta}^m(n) \equiv \left(\hat{\theta}(n) \vee \underline{\theta}^m \right) \wedge \overline{\theta}^m,$$

and the corresponding local generalized log-likelihood in (2.9) is

$$\ell^m(n) = \ell(n; \hat{\theta}^m(n)).$$

We claim that, for any $m \in [M]$ and $\theta \in \Theta \backslash \Theta^m$, $\hat{\ell}(n) - \ell^m(n)$ and $I^m(n;\theta)$ in (S1.2) satisfy Assumption 2-3 provided that there exists a function $g(\cdot) : (0, \infty) \to (0, \infty)$ so that for large $n \in \mathbb{N}$ and large $M > 0$,

$$M\psi(n) \leq \psi\big(g(M)n\big), \tag{S1.4}$$

and

$$\sum_{n=0}^{\infty} \sum_{m=n}^{\infty} e^{-\psi(m)} < \infty. \tag{S1.5}$$

It suffices to check condition (2.11).

Before starting, we state and prove the following proposition, which will be an intermediate step in all three examples.

Proposition 1. *Let $\hat{\ell}(n)$ and $\ell^m(n)$ be defined as in (2.8) and (2.9), with lower index k suppressed to ease the notation. Then,*

$$\hat{\ell}(n) - \ell^m(n) \geq \underbrace{\hat{\ell}(n) - \ell(n;\theta) - \sum_{t=1}^{n} \mathrm{E}_\theta \left[\Delta(\hat{\ell}(t) - \ell(t;\theta)) \middle| \mathcal{F}(t-1) \right]}_{A(n)}$$
$$+ \underbrace{\sum_{t=1}^{n} \mathrm{E}_\theta \left[\Delta(\hat{\ell}(t) - \ell(t;\theta)) \middle| \mathcal{F}(n-1) \right]}_{B(n)} + \underbrace{\inf_{\theta' \in \Theta^m} \{\ell(n;\theta) - \ell(n;\theta') - I(n;\theta,\theta')\}}_{C(n)} + I^m(n;\theta), \tag{S1.6}$$

where $\Delta S(t) \equiv S(t) - S(t-1)$. Besides, $\{A(n),\, n \geq 1\}$ is a zero-mean martingale with respect to filtration $\mathcal{F}$ and probability measure P_θ.

Proof of Proposition 1. Indeed,

$$\hat{\ell}(n) - \ell^m(n) = \hat{\ell}(n) + \inf_{\theta' \in \Theta^m} \{-\ell(n;\theta')\}$$
$$= \hat{\ell}(n) - \ell(n;\theta) + \inf_{\theta' \in \Theta^m} \{\ell(n;\theta) - \ell(n;\theta') - I(n;\theta,\theta') + I(n;\theta,\theta')\}$$
$$\geq \hat{\ell}(n) - \ell(n;\theta) + \inf_{\theta' \in \Theta^m} \{\ell(n;\theta) - \ell(n;\theta') - I(n;\theta,\theta')\} + I^m(n;\theta),$$

and the desired form follows after adding and subtracting the same term. The second statement can be checked directly by definition. □

In the example of this subsection, we have

$$\begin{aligned}
&\mathrm{E}_\theta\left[\Delta(\hat{\ell}(n)-\ell(n;\theta))\middle|\mathcal{F}(n-1)\right]\\
=\,&\mathrm{E}_\theta\left[-\frac{1}{2\sigma(n)^2}\left((X(n)-\hat{\theta}(n-1))^2-(X(n)-\theta)^2\right)\middle|\mathcal{F}(n-1)\right]\\
=\,&-\frac{1}{2\sigma(n)^2}(\hat{\theta}(n-1)-\theta)^2.
\end{aligned}$$

It suffices to show $A(n)/\psi(n)$, $B(n)/\psi(n)$ and $C(n)/\psi(n)$ converge to zero completely, respectively.

For the first, it is clear that

$$A(n)=\sum_{t=1}^{n}\frac{1}{\sigma(t)^2}(\hat{\theta}(t-1)-\theta)(X(t)-\theta),$$

so

$$|A(n)|\leq(\overline{\theta}-\underline{\theta})\left|\sum_{t=1}^{n}\frac{1}{\sigma(t)^2}(X(t)-\theta)\right|,$$

which divided by $\psi(n)$ we have shown to converge to zero completely in (S1.3).

For the second, fix $\epsilon>0$ and define random times

$$\begin{aligned}
\tau_1&\equiv\sup\left\{n\in\mathbb{N}:(\hat{\theta}(n)-\theta)^2>\epsilon\right\}+1,\\
\tau_2&\equiv\sup\left\{n\in\mathbb{N}:\frac{1}{\psi(n)}\sum_{t=1}^{n}\frac{(\hat{\theta}(t-1)-\theta)^2}{2\sigma(t)^2}>\epsilon\right\}.
\end{aligned}$$

From, e.g., (Tartakovsky et al., 2014, Chapter 2.4.3), we know that $B(n)/\psi(n)\xrightarrow{completely}0$ is implied by $B(n)/\psi(n)\xrightarrow{quickly}0$, which is defined as: for all $\epsilon>0$, $\mathrm{E}_\theta[\tau_2]<\infty$. Next we show

$$\psi(\tau_2)\leq\frac{(\overline{\theta}-\underline{\theta})^2}{\epsilon}\psi(\tau_1),$$

so that, based on condition (S1.4), we have

$$\psi(\tau_2) \le \psi\left(g\left(\frac{(\bar{\theta}-\underline{\theta})^2}{\epsilon}\right)\tau_1\right) \Rightarrow \tau_2 \le g\left(\frac{(\bar{\theta}-\underline{\theta})^2}{\epsilon}\right)\tau_1,$$

and it suffices to show $\mathrm{E}_\theta[\tau_1] < \infty$. Indeed, for any $n \in \mathbb{N}$ that $\psi(n) \ge \frac{(\bar{\theta}-\underline{\theta})^2}{\epsilon}\psi(\tau_1)$, we have

$$\begin{aligned}
&\frac{1}{\psi(n)}\sum_{t=1}^{n}\frac{(\hat{\theta}(t-1)-\theta)^2}{2\sigma(t)^2}\\
&= \frac{1}{\psi(n)}\left(\sum_{t=1}^{\tau_1}\frac{(\hat{\theta}(t-1)-\theta)^2}{2\sigma(t)^2} + \sum_{t=\tau_1+1}^{n}\frac{(\hat{\theta}(t-1)-\theta)^2}{2\sigma(t)^2}\right)\\
&\le \frac{1}{\psi(n)}\left(\frac{(\bar{\theta}-\underline{\theta})^2}{2}\sum_{t=1}^{\tau_1}\frac{1}{\sigma(t)^2} + \frac{\epsilon}{2}\sum_{t=\tau_1+1}^{n}\frac{1}{\sigma(t)^2}\right)\\
&\le \frac{(\bar{\theta}-\underline{\theta})^2}{2}\frac{\psi(\tau_1)}{\psi(n)} + \frac{\epsilon}{2} \le \epsilon.
\end{aligned}$$

To show $\mathrm{E}_\theta[\tau_1] < \infty$, note that the distance between θ and the restricted MLE $\hat{\theta}(n)$ is less than or equal to the distance between θ and the unrestricted MLE, and

$$\frac{1}{\psi(n)}\sum_{t=1}^{n}\frac{X(t)}{\sigma(t)^2} - \theta = \frac{1}{\psi(n)}\sum_{t=1}^{n}\frac{1}{\sigma(t)}\frac{X(t)-\theta}{\sigma(t)} \equiv Y(n) \sim N\left(0, 1/\psi(n)\right),$$

so

$$\begin{aligned}
\mathrm{E}_\theta[\tau_1] &= \sum_{n=0}^{\infty}\mathrm{P}_\theta(\tau_1 > n) = \sum_{n=0}^{\infty}\mathrm{P}_\theta\left(\exists\, m \ge n,\, (\hat{\theta}(m)-\theta)^2 > \epsilon\right)\\
&\le \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}\mathrm{P}_\theta\left((\hat{\theta}(m)-\theta)^2 > \epsilon\right) \le \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}\mathrm{P}_\theta(|Y(m)| \ge \sqrt{\epsilon})\\
&= \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}\mathrm{P}\left(|N(0,1)| \ge \sqrt{\epsilon\psi(m)}\right) \le \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\epsilon\psi(m)}}e^{-\frac{\epsilon}{2}\psi(m)},
\end{aligned}$$

which is finite provided condition (S1.5).

For the third, it is clear that

$$\ell(n;\theta) - \ell(n;\theta') - I(n;\theta,\theta') = (\theta - \theta') \sum_{t=1}^{n} \frac{1}{\sigma(t)^2}(X(t) - \theta),$$

so

$$|C(n)| \leq (\bar{\theta} - \underline{\theta}) \left| \sum_{t=1}^{n} \frac{1}{\sigma(t)^2}(X(t) - \theta) \right|,$$

which is the same as the upper bound for $|A(n)|$.

Combining (S1.1), (S1.4) and (S1.5), a set of sufficient conditions on $\psi(n) = \sum_{t=1}^{n} 1/\sigma(t)^2$ is

$$\lim_{q\to 1} \lim_{n\to\infty} \frac{\psi(qn)}{\psi(n)} = 1, \quad \sum_{n=0}^{\infty} \sum_{m=n}^{\infty} e^{-\psi(m)} < \infty,$$

and there exists a function $g(\cdot) : (0,\infty) \to (0,\infty)$ so that for large $n \in \mathbb{N}$ and large $M > 0$,

$$M\psi(n) \leq \psi\big(g(M)n\big).$$

Note that this is clearly satisfied by functions with a polynomial rate, i.e., $\psi(n) \sim an^b$ as $n \to \infty$ for some $a, b > 0$.

S1.2 Testing the correlation coefficient of AR(1) series

In the second example, we assume that $X = \{X(n), n \in \mathbb{N}\}$ follows a first-order autoregressive model with Gaussian noises, i.e.,

$$X(n) = \theta X(n-1) + \epsilon(n), \quad n \in \mathbb{N},$$

where $X(0) = 0$, $\{\epsilon(n),\, n \in \mathbb{N}\}$ are i.i.d. standard Gaussian, and $\theta \in (-1,1)$ is unknown. Consider testing $\theta \in \Theta^m$ for $m \in [M]$, where $\Theta = [\underline{\theta}, \overline{\theta}] \subseteq (-1,1)$ and $\{\Theta^m,\, m \in [M]\}$ are M disjoint closed intervals in Θ, with $\Theta^m = [\underline{\theta}^m, \overline{\theta}^m]$.

First of all, for any $\theta \in \Theta$, the conditional distribution of $X(n)$ given $X(n-1)$ is $N(\theta X(n-1), 1)$, so the log-likelihood is

$$\ell(n;\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left(X(t) - \theta X(t-1)\right)^2,$$

and, for any distinct $\theta, \theta' \in \Theta$, the log-likelihood ratio is

$$\ell(n;\theta) - \ell(n;\theta') = (\theta - \theta')\sum_{t=1}^{n}\left(X(t-1)X(t) - \frac{\theta+\theta'}{2}X(t-1)^2\right).$$

It is shown in Bercu et al. (1997) that

$$\frac{1}{n}\sum_{t=1}^{n} X(t-1)X(t) \to \frac{\theta}{1-\theta^2}, \quad \frac{1}{n}\sum_{t=1}^{n} X(t-1)^2 \to \frac{1}{1-\theta^2}, \tag{S1.7}$$

with an exponential rate under P_θ, i.e., there exists a constant $C_\theta > 0$ and positive function $\phi_\theta(\cdot)$, so that

$$\mathrm{P}_\theta(|\,\mathrm{Statistic}(n) - \mathrm{Limit}\,| > \epsilon) \leq C_\theta e^{-n\phi_\theta(\epsilon)} \text{ for all } \epsilon > 0,$$

where Statistics(n) and Limit represent the corresponding quantities in (S1.7) for simplicity. This convergence rate is clearly stronger than complete convergence and quick convergence. so

$$\frac{1}{n}(\ell(n;\theta) - \ell(n;\theta')) \xrightarrow{completely} \frac{(\theta-\theta')^2}{2(1-\theta^2)} \text{ under } \mathrm{P}_\theta.$$

Therefore, for any $m \in [M]$ and $\theta \in \Theta \backslash \Theta^m$, Assumption 1 holds with

$$I^m(n;\theta) = n \inf_{\theta' \in \Theta^m} \frac{(\theta - \theta')^2}{2(1-\theta^2)} = n \min \left\{ \frac{(\theta - \underline{\theta}^m)^2}{2(1-\theta^2)}, \frac{(\theta - \overline{\theta}^m)^2}{2(1-\theta^2)} \right\}. \qquad \text{(S1.8)}$$

Next, we study the test statistic $\hat{\ell}(n) - \ell^m(n)$. We use the global MLE as the adaptive estimator, i.e.,

$$\hat{\theta}(n) \equiv \left(\frac{\sum_{t=1}^n X(t-1)X(t)}{\sum_{t=1}^n X(t-1)^2} \vee \underline{\theta} \right) \wedge \overline{\theta},$$

which is known as the Yule-Walker estimator and is shown in Bercu et al. (1997) to converge to θ with an exponential rate under P_θ. The adaptive log-likelihood in (2.8) is

$$\hat{\ell}(n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^n \Big(X(t) - \hat{\theta}(t-1)X(t-1) \Big)^2.$$

For every $m \in [M]$, the local generalized log-likelihood in (2.9) is $\ell^m(n) = \sup_{\theta' \in \Theta^m} \ell(n;\theta')$. We claim that, for any $m \in [M]$ and $\theta \in \Theta \backslash \Theta^m$, $\hat{\ell}(n) - \ell^m(n)$ and $I^m(n;\theta)$ in (S1.8) satisfy Assumption 2-3.

We start from the decomposition in Proposition 1. Specifically, we have

$$\begin{aligned}
&\mathrm{E}_\theta\left[\Delta(\hat{\ell}(n) - \ell(n;\theta)) \Big| \mathcal{F}(n-1) \right] \\
&= \mathrm{E}_\theta\left[\frac{1}{2}(\hat{\theta}(n-1) - \theta)X(n-1)\Big(2X(n) - (\hat{\theta}(n-1)+\theta)X(n-1)\Big) \Big| \mathcal{F}(n-1) \right] \\
&= -\frac{1}{2}(\hat{\theta}(n-1)-\theta)^2 X(n-1)^2,
\end{aligned}$$

where we used the fact that $X(n)|\mathcal{F}(n-1) \sim N(\theta X(n-1), 1)$ under P_θ.

For the first term in (S1.6), we have

$$A(n)=\sum_{t=1}^{n}(\hat{\theta}(t-1)-\theta)X(t-1)(X(t)-\theta X(t-1)).$$

This cannot be dealt with as simply as in the first example. However, as a martingale, from Stoica (2007) we know that its sequence of averages converge completely to its mean, i.e., $A(n)/n \to 0$ completely, as long as its sequence of differences is uniformly bounded in L^2, i.e.,

$$\sup_{n\geq 1}\mathrm{E}_\theta[(\Delta A(n))^2]<\infty.$$

Indeed,

$$\mathrm{E}_\theta[(\Delta A(n))^2]\leq(\overline{\theta}-\underline{\theta})^2\mathrm{E}_\theta[X(n-1)^2(X(n)-\theta X(n-1))^2]$$

$$=(\overline{\theta}-\underline{\theta})^2\mathrm{E}_\theta\left[X(n-1)^2\mathrm{E}_\theta[(X(n)-\theta X(n-1))^2|\mathcal{F}(n-1)]\right]$$

$$=(\overline{\theta}-\underline{\theta})^2\mathrm{E}_\theta[X(n-1)^2]$$

is uniformly bounded in L^2, so $A(n)/n \to 0$ completely is proved.

For the second term, we have

$$|B(n)|\leq\frac{1}{2}\sum_{t=1}^{n}(\hat{\theta}(t-1)-\theta)^2X(t-1)^2.$$

Although both $\hat{\theta}(n)-\theta$ and $\frac{1}{n}\sum_{t=1}^{n}X(t-1)^2$ converge with an exponential rate, and even $\hat{\theta}(n)-\theta$ is uniformly bounded, $B(n)/n\to 0$ completely is not automatic. A sufficient condition is that $\sup_{n\geq 1}\mathrm{E}_\theta[X(n)^{2p}]<\infty$ for some $p>1$, which is clearly satisfied in this example. To see this is sufficient, fix

$\epsilon > 0$ and define

$$\tau_1 \equiv \sup\left\{n \in \mathbb{N} : (\hat{\theta}(n) - \theta)^2 > \epsilon \text{ or } \frac{1}{n}\sum_{t=1}^{n} X(t-1)^2 > v\right\} + 1,$$

$$\tau_2 \equiv \sup\left\{n \in \mathbb{N} : \frac{1}{n}\sum_{t=1}^{n}(\hat{\theta}(t-1) - \theta)^2 X(t-1)^2 > \epsilon(1+v)\right\},$$

where v represents an arbitrary real number that is greater than $1/(1-\theta^2)$. From the exponential converge rates of $\hat{\theta}(n) - \theta$ and $\frac{1}{n}\sum_{t=1}^{n} X(t-1)^2$, we know that $\mathrm{E}_\theta[\tau_1] < \infty$. Next we show that $\mathrm{E}_\theta[\tau_2] < \infty$, which implies $|B(n)|/n \to 0$ completely. Indeed, for any $n \geq \frac{4}{\epsilon}\sum_{t=1}^{\tau_1} X(t-1)^2$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}(\hat{\theta}(t-1) - \theta)^2 X(t-1)^2 \\
&= \frac{1}{n}\left(\sum_{t=1}^{\tau_1}(\hat{\theta}(t-1) - \theta)^2 X(t-1)^2 + \sum_{t=\tau_1+1}^{n}(\hat{\theta}(t-1) - \theta)^2 X(t-1)^2\right) \\
&\leq \frac{4}{n}\sum_{t=1}^{\tau_1} X(t-1)^2 + \epsilon v \leq \epsilon(1+v),
\end{aligned}$$

so $\tau_2 \leq \frac{4}{\epsilon}\sum_{t=1}^{\tau_1} X(t-1)^2$. It remains to show $\mathrm{E}_\theta[\sum_{t=1}^{\tau_1} X(t-1)^2] < \infty$. Indeed,

$$\begin{aligned}
&\mathrm{E}_\theta\left[\sum_{t=1}^{\tau_1} X(t-1)^2\right] = \mathrm{E}_\theta\left[\sum_{t=1}^{\infty} X(t-1)^2 1\{t \leq \tau_1\}\right] \\
&= \sum_{t=1}^{\infty}\mathrm{E}_\theta\left[X(t-1)^2 1\{t \leq \tau_1\}\right] \leq \sum_{t=1}^{\infty}\mathrm{E}_\theta[X(t-1)^{2p}]^{1/p}\mathrm{P}_\theta(\tau_1 \geq t) \\
&\leq \left(\sup_{n\geq 1}\mathrm{E}_\theta[X(n)^{2p}]\right)^{1/p}\mathrm{E}_\theta[\tau_1] < \infty,
\end{aligned}$$

where in the second equality we could change the order of expectation and summation because of non-negativity, and in the first inequality we used Hölder's inequality.

For the third term, we have, for any $\theta' \in \Theta^m$,

$$\begin{aligned}
&\ell(n;\theta) - \ell(n;\theta') - I(n;\theta,\theta') \\
&= (\theta-\theta')\sum_{t=1}^{n}\left(X(t-1)X(t) - \frac{\theta+\theta'}{2}X(t-1)^2 - \frac{\theta-\theta'}{2(1-\theta^2)}\right) \\
&= (\theta-\theta')\sum_{t=1}^{n}\left(X(t-1)X(t) - \frac{\theta}{2}X(t-1)^2 - \frac{\theta}{2(1-\theta^2)}\right) \\
&\quad + (\theta-\theta')\theta'\sum_{t=1}^{n}\left(-\frac{1}{2}X(t-1)^2 + \frac{1}{2(1-\theta^2)}\right),
\end{aligned}$$

so

$$\begin{aligned}
|C(n)| \leq 2&\left|\sum_{t=1}^{n}\left(X(t-1)X(t) - \frac{\theta}{2}X(t-1)^2 - \frac{\theta}{2(1-\theta^2)}\right)\right| \\
&+ 2\left|\sum_{t=1}^{n}\left(-\frac{1}{2}X(t-1)^2 + \frac{1}{2(1-\theta^2)}\right)\right|.
\end{aligned}$$

Based on (S1.7) it is clear that $C(n)/n \to 0$ completely under P_θ.

S1.3 Testing the transition matrix of a Markov chain

we assume that $X = \{X(n) : n \in \mathbb{N}\}$ follows a Markov process with finite state space $[s] = \{1, \ldots, s\}$, initial state $X(0) = 1$, and transition matrix $\Pi = \{\Pi(i,j) : (i,j) \in [s]^2\}$. We denote by $\mathcal{P}$ a family of irreducible and recurrent transition matrices that satisfies

$$\sup_{\Pi,\Pi'\in\mathcal{P}} \sup_{(i,j)\in[s]^2} \log\frac{\Pi(i,j)}{\Pi'(i,j)} < \infty, \tag{S1.9}$$

so that the log-likelihood ratio between any two transition matrices is finite based on finite samples. Consider testing $\Pi \in \mathcal{P}^m$ for $m \in [M]$ where

$\{\mathcal{P}^m,\, m \in [M]\}$ are M disjoint subsets of $\mathcal{P}$ so that I^m_Π, defined in (S1.11), is positive for all $m \in [M]$ and $\Pi \in \mathcal{P} \backslash \mathcal{P}^m$.

First of all, for any $\Pi \in \mathcal{P}$, the conditional distribution of $X(n)$ given $X(n-1)$ is multinomial with probability $\Pi(X(n-1), i)$ for $i \in [s]$, so the log-likelihood is

$$\ell(n; \Pi) = \sum_{t=1}^{n} \log \Pi(X(t-1), X(t)) = \sum_{(i,j)\in[s]^2} N_n(i,j) \log \Pi(i,j),$$

and, for distinct $\Pi, \Pi' \in \mathcal{P}$, the log-likelihood ratio is

$$\ell(n; \Pi) - \ell(n; \Pi') = \sum_{t=1}^{n} \log \frac{\Pi(X(t-1), X(t))}{\Pi'(X(t-1), X(t))} = \sum_{(i,j)\in[s]^2} N_n(i,j) \log \frac{\Pi(i,j)}{\Pi'(i,j)},$$

where $N_n(i,j) \equiv \sum_{t=1}^{n} 1\{(X(t-1), X(t)) = (i,j)\}$ records the number of transitions from state i to state j up to time n. We denote by $\pi = \{\pi(i),\, i \in [s]\}$ the stationary distribution of X when the transition matrix is Π, and by $\pi_Y = \{\pi_Y(i,j),\, (i,j) \in [s]^2\}$ the stationary distribution of the induced two-step Markov process $Y = \{(X(n), X(n+1)),\, n \geq 0\}$. The existence and uniqueness of the stationary distributions are guaranteed (see,e.g., (Durrett, 2010, Theorem 5.5.9)). It is shown in (Dembo and Zeitouni, 1998, Chapter 3) that

$$\frac{1}{n} N_n(i,j) \to \pi_Y(i,j) \text{ with an exponential rate, for all } (i,j) \in [s]^2, \tag{S1.10}$$

so

$$\frac{1}{n}\left(\ell(n;\Pi)-\ell(n;\Pi')\right) \xrightarrow[under\, \mathrm{P}_\Pi]{completely} \sum_{(i,j)\in[s]^2} \pi_Y(i,j)\log\frac{\Pi(i,j)}{\Pi'(i,j)} \equiv I_{\Pi,\Pi'}.$$

Therefore, for any $m \in [M]$ and $\Pi \in \mathcal{P}\backslash\mathcal{P}^m$, Assumption 1 holds with

$$I^m(n;\Pi) = nI^m_\Pi, \text{ where } I^m_\Pi \equiv \inf_{\Pi'\in\mathcal{P}^m} I_{\Pi,\Pi'}, \tag{S1.11}$$

as long as $\mathcal{P}^m$ is separated from the rest of the hypothesis space in the sense that $I^m_\Pi > 0$ for all $\Pi \in \mathcal{P}\backslash\mathcal{P}^m$. A quick example is, when $s = 2$, testing $\Pi(1,2) < p_0$ v.s. $\Pi(1,2) > p_1$ where $0 \le p_0 < p_1 \le 1$.

Next, we study the test statistic $\hat{\ell}(n) - \ell^m(n)$. For any $n \ge 1$, it is natural to use the following estimator:

$$\hat{\Pi}_n(i,j) \equiv \frac{N_n(i,j)}{\sum_{k\in[s]} N_n(i,k)} \text{ for } (i,j)\in[s]^2,$$

which is shown in Anderson and Goodman (1957) to be the unrestricted global MLE and to converge to $\Pi(i,j)$ with an exponential rate. The adaptive log-likelihood in (2.8) is

$$\begin{aligned}\hat{\ell}(n) &= \sum_{t=1}^n \log\hat{\Pi}_{t-1}(X(t-1),X(t))\\ &= \sum_{(i,j)\in[s]^2}\sum_{t=1}^n 1\{(X(t-1),X(t))=(i,j)\}\log\hat{\Pi}_{t-1}(i,j).\end{aligned}$$

For every $m \in [M]$, the local generalized log-likelihood in (2.9) is $\ell^m = \sup_{\Pi'\in\mathcal{P}^m}\ell(n;\Pi')$. Starting from the decomposition in Proposition 1 and following similar steps as in the second example, it can be shown that, for any $m \in [M]$ and $\Pi \in \mathcal{P} \backslash \mathcal{P}^m$, $\hat{\ell}(n) - \ell^m(n)$ and $I^m(n;\theta)$ in (S1.11) satisfy Assumption 2-3.

S2 Extra numerical studies

In this section, we present a numerical study of testing the correlation coefficient of AR(1) series. Specifically, we consider K first-order autoregressive series,

$$X_k(n) = \theta_k X_k(n-1) + \epsilon_k(n), \quad n \in \mathbb{N},$$

with correlation coefficient θ_k and i.i.d. standard Gaussian noises for $k \in [K]$. For each of them, consider testing the following hypotheses:

$$\Theta_k^1 = [-0.75, -0.5], \quad \Theta_k^2 = [-0.25, 0.25], \quad \Theta_k^3 = [0.5, 0.75],$$
$$\Theta_k^0 = (-0.5, -0.25) \cup (0.25, 0.5),$$

that is, $M = 3$ and $\Theta_k = [-0.75, 0.75]$. All statistics, namely, the log-likelihoods $\ell_k(n;\theta_k)$, the global MLE $\hat{\theta}_k(n)$, the adaptive log-likelihoods $\hat{\ell}(n)$, the local MLEs $\hat{\theta}_k^m(n)$, and the local generalized log-likelihoods have been derived in Section S1.2.

We use $n_0 = 10$ pilot samples and the same importance sampling distribution as the numerical study in the main body. The same set of figures are presented in Figure 1-3. All simulations are based on 10^5 rounds.

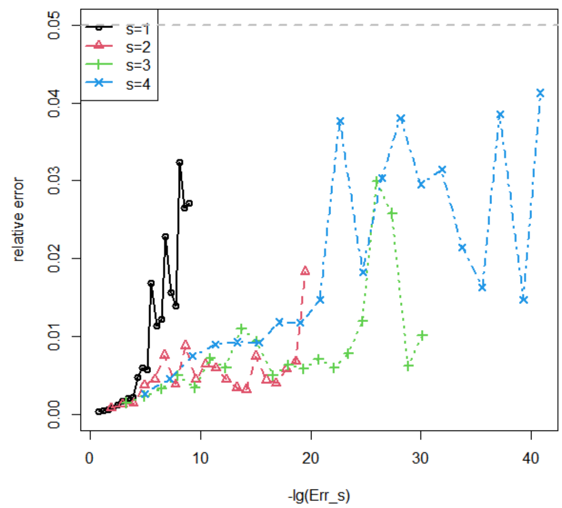

Figure 1: Relative errors of all estimates are below 5% based on 10^5 simulation rounds.

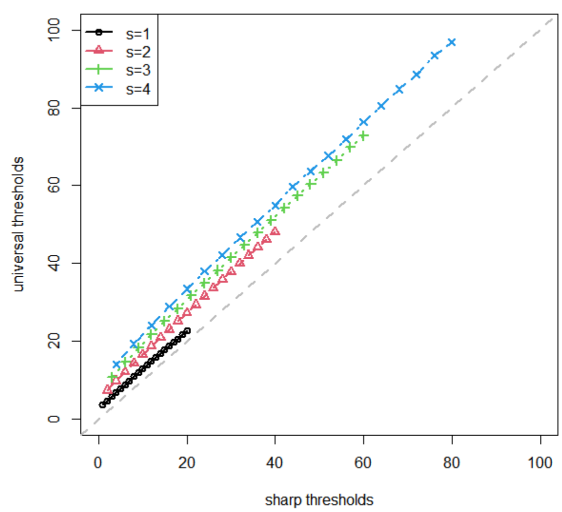

Figure 2: Universal thresholds versus sharp thresholds, showing basically a constant level of conservativeness.

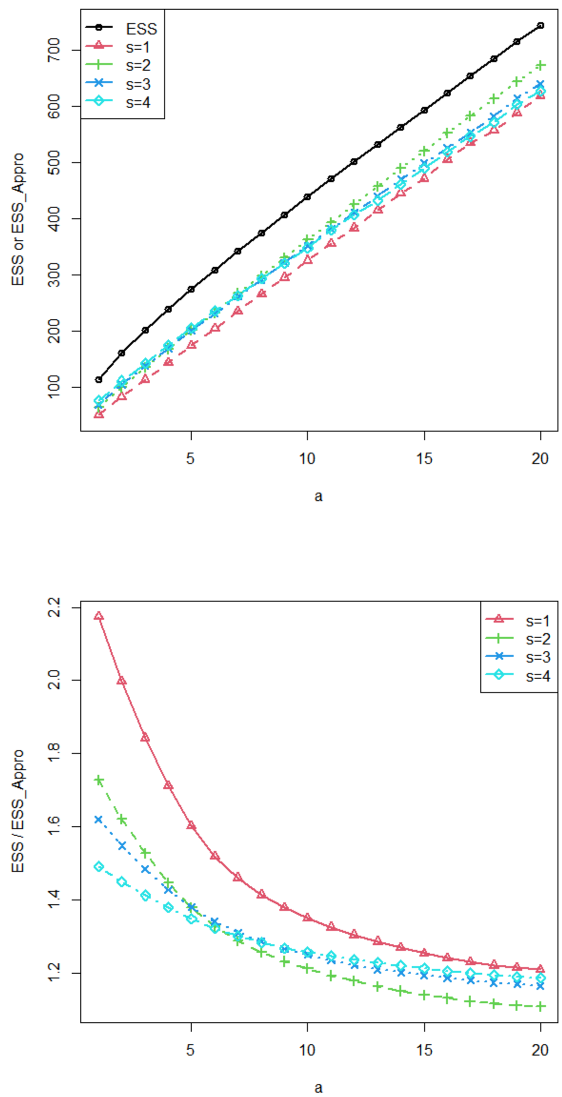

Figure 3: Verifications of the asymptotic theory. All curves at the top are asymptotically equal and all curves at the bottom asymptotically converge to one.

S3 Proofs

This section includes proofs of all theorems, along with some important lemmas and remarks.

S3.1 Related to the asymptotic lower bound

Lemma 1. *Let* $\{A(n),\, n \in \mathbb{N}\}$ *be a sequence of random variables and* $\{a(n),\, n \in \mathbb{N}\}$ *be a sequence of positive real numbers that increase to infinity. If*

$$\mathrm{P}\left(\limsup_{n\to\infty} \frac{A(n)}{a(n)} \leq 1\right) = 1, \tag{S3.12}$$

then, for any $q \in (0,1)$,

$$\lim_{N\to\infty} \mathrm{P}\left(\frac{1}{a(N)} \max_{1\leq n\leq N} A(n) \geq \frac{1}{q}\right) = 0.$$

Proof of Lemma 1. Fix $q \in (0,1)$ and $N \in \mathbb{N}$. By the union bound and the condition that $\{a(n),\, n \in \mathbb{N}\}$ is positive and increasing, for any $1 \leq m \leq N$ we have

$$\begin{aligned}
&\mathrm{P}\left(\frac{1}{a(N)} \max_{1\leq n\leq N} A(n) \geq \frac{1}{q}\right) \\
&\leq \mathrm{P}\left(\frac{1}{a(N)} \max_{1\leq n\leq m} A(n) \geq \frac{1}{q}\right) + \mathrm{P}\left(\frac{1}{a(N)} \max_{m\leq n\leq N} A(n) \geq \frac{1}{q}\right) \\
&\leq \mathrm{P}\left(\frac{1}{a(N)} \max_{1\leq n\leq m} A(n) \geq \frac{1}{q}\right) + \mathrm{P}\left(\sup_{n\geq m} \frac{A(n)}{a(n)} \geq \frac{1}{q}\right).
\end{aligned}$$

Letting $N \to \infty$, we have

$$\limsup_{N\to\infty} \mathrm{P}\left(\frac{1}{a(N)} \max_{1\leq n\leq N} A(n) \geq \frac{1}{q}\right) \leq \mathrm{P}\left(\sup_{n\geq m} \frac{A(n)}{a(n)} \geq \frac{1}{q}\right).$$

Letting $m \to \infty$, the right-hand-side is equal to

$$\mathrm{P}\left(\limsup_{n\to\infty} \frac{A(n)}{a(n)} \geq \frac{1}{q}\right),$$

which is zero based on condition (S3.12). □

Lemma 2. *Let $\{f_k(n), n \in \mathbb{N}\}$, $k \in [K]$ be K sequences of positive real numbers that increase to infinity. Let $f(n) \equiv \sum_{k\in[K]} f_k(n)$. For any $a > 0$, let $f_k(a)^{-1} \equiv \min\{n \in \mathbb{N} : f_k(n) \geq a\}$ and $f(a)^{-1} \equiv \min\{n \in \mathbb{N} : f(n) \geq a\}$. If*

$$\lim_{q\to 1} \lim_{a\to\infty} \frac{f_k(qa)^{-1}}{f_k(a)^{-1}} = 1 \text{ for all } k \in [K], \tag{S3.13}$$

then

$$\lim_{q\in(0,1)\to 1} \lim_{a\to\infty} \frac{f(qa)^{-1}}{f(a)^{-1}} = 1.$$

Proof of Lemma 2. Condition (S3.13) implies that, for any $\epsilon \in (0,1)$, there exist $q_\epsilon \in (0,1)$ and $a_\epsilon > 0$, so that $f_k(qa)^{-1}/f_k(a)^{-1} \geq 1-\epsilon$ for all $q \in (q_\epsilon, 1)$, $a > a_\epsilon$ and $k \in [K]$. Fix a small ϵ and a $q \in (q_\epsilon, 1)$. For any $a > 0$, define

$$a_k \equiv f_k\big(f(qa)^{-1}\big)/q \text{ for } k \in [K],$$

where the dependence on a is suppressed for simplicity of notations, so that

$$\begin{aligned} &\min\Big\{n \in \mathbb{N} : f_k(n) \geq q a_k \text{ for all } k \in [K]\Big\} \\ = &\min\Big\{n \in \mathbb{N} : f_k(n) \geq f_k\big(f(qa)^{-1}\big) \text{ for all } k \in [K]\Big\} = f(qa)^{-1}. \end{aligned}$$

Let a be large enough so that $a_k > a_\epsilon$ for all $k \in [K]$. Now, on the one hand, $f(a)^{-1} \geq f(qa)^{-1}$. On the other hand,

$$\begin{aligned}
f(qa)^{-1} &= \min\Big\{n \in \mathbb{N} : f_k(n) \geq qa_k \text{ for all } k \in [K]\Big\} \\
&= \max_{k\in[K]} \min\Big\{n \in \mathbb{N} : f_k(n) \geq qa_k\Big\} \\
&= \max_{k\in[K]} f_k(qa_k)^{-1} \geq (1-\epsilon) \max_{k\in[K]} f_k(a_k)^{-1} \\
&= (1-\epsilon) \max_{k\in[K]} \min\Big\{n \in \mathbb{N} : f_k(n) \geq a_k\Big\} \\
&= (1-\epsilon) \min\Big\{n \in \mathbb{N} : f_k(n) \geq a_k \text{ for all } k \in [K]\Big\} \\
&\geq (1-\epsilon) \min\Big\{n \in \mathbb{N} : f(n) \geq \sum_{k\in[K]} a_k = \frac{1}{q} \sum_{k\in[K]} f_k\big(f(qa)^{-1}\big) = \frac{1}{q} f\big(f(qa)^{-1}\big)\Big\} \\
&\geq (1-\epsilon) \min\Big\{n \in \mathbb{N} : f(n) \geq a\Big\} = (1-\epsilon) f(a)^{-1}.
\end{aligned}$$

That is to say, we have shown that, for any small ϵ, $f(a)^{-1} \geq f(qa)^{-1} \geq (1-\epsilon)f(a)^{-1}$ for q sufficiently close to 1 and a sufficiently large. Letting $q \in (0,1) \to 1$, $a \to \infty$ and $\epsilon \to 0$ completes the proof. □

Proof of Theorem 1. It suffices to prove (3.13) since (3.14) follows from (3.13) and the fact that $\Delta(\boldsymbol{\alpha}) = \cap_{s\in[K]} \Delta_s(\alpha_s)$.

Fix arbitrary $1 \leq s \leq K$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Denote

$$\mathcal{U}_{s,\boldsymbol{\theta}} \equiv \left\{\boldsymbol{B} \in [M]^K : |\boldsymbol{H}_{\boldsymbol{\theta}} \triangle \boldsymbol{B}| < s\right\},$$

i.e., all identifications of hypotheses that make less than s errors when the truth is $\boldsymbol{H}_{\boldsymbol{\theta}}$. Denote by S^* a minimizer of the minimization problem in

(3.12). Also fix arbitrary $\alpha \in (0,1)$ and $\delta \in \Delta_s(\alpha)$. It is clear that

$$1-\alpha \leq \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{D} \in \mathcal{U}_{s,\boldsymbol{\theta}}) = \sum_{B \in \mathcal{U}_{s,\boldsymbol{\theta}}} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{D} = \boldsymbol{B}).$$

Further fix arbitrary $\boldsymbol{B} \in \mathcal{U}_{s,\boldsymbol{\theta}}$ and $q \in (0,1)$. Let $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ be that, for $k \in [K]$,

$$\tilde{\theta}_k = \begin{cases} \text{an element of } (\Theta_k^{H'_{k,\theta_k}} \cup \Theta_k^{H''_{k,\theta_k}}) \backslash \Theta_k^{B_k} \text{ so that} \\ \qquad \mathrm{P}_{k,\theta_k}\left(\limsup_{n\to\infty} \frac{\ell_k(n;\theta_k) - \ell_{k,\tilde{\theta}_k}(n)}{I_k^{H''_{k,\theta_k}}(n;\theta_k)} \leq \frac{1}{\sqrt{q}} \right) = 1, \quad \text{if } k \in S^*, \\ \theta_k, \qquad \text{otherwise,} \end{cases}$$

whose existence is guaranteed by Assumption 1. Note that $\tilde{\boldsymbol{\theta}}$ satisfies the following two properties: (i) $S^* \subseteq \boldsymbol{H}_{\tilde{\boldsymbol{\theta}}} \triangle \boldsymbol{B}$, so $\mathbf{P}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{D} = \boldsymbol{B}) \leq \alpha$. (ii)

$$\begin{aligned} &\mathbf{P}_{\boldsymbol{\theta}}\left(\limsup_{n\to\infty} \frac{\boldsymbol{\ell}_{\boldsymbol{\theta}}(n) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(n)}{\boldsymbol{I}_s(n;\boldsymbol{\theta})} \leq \frac{1}{\sqrt{q}} \right) \\ &= \mathbf{P}_{\boldsymbol{\theta}}\left(\limsup_{n\to\infty} \frac{\sum_{k\in S^*} \ell_k(n;\theta_k) - \ell_{k,\tilde{\theta}_k}(n)}{\sum_{k\in S^*} I_k^{H''_{k,\theta_k}}(n;\theta_k)} \leq \frac{1}{\sqrt{q}} \right) \\ &\geq 1 - \sum_{k\in S^*} \left(1 - \mathbf{P}_{\boldsymbol{\theta}}\left(\limsup_{n\to\infty} \frac{\ell_k(n;\theta_k) - \ell_{k,\tilde{\theta}_k}(n)}{I_k^{H''_{k,\theta_k}}(n;\theta_k)} \leq \frac{1}{\sqrt{q}} \right) \right) = 1, \end{aligned}$$

where in the inequality we used the fact that for real numbers a_k and positive real numbers b_k,

$$\frac{\sum_k a_k}{\sum_k b_k} > c \text{ implies } \frac{a_k}{b_k} > c \text{ for some } k.$$

Further fix arbitrary $\eta \in (0,1)$. We decompose $\mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{D} = \boldsymbol{B})$ as

$$\mathbf{P}_{\boldsymbol{\theta}}\left(\boldsymbol{\ell}_{\boldsymbol{\theta}}(T) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(T) < \log\frac{\eta}{\alpha},\, \boldsymbol{D} = \boldsymbol{B} \right) + \mathbf{P}_{\boldsymbol{\theta}}\left(\boldsymbol{\ell}_{\boldsymbol{\theta}}(T) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(T) \geq \log\frac{\eta}{\alpha},\, \boldsymbol{D} = \boldsymbol{B} \right) \equiv I + II.$$

By changing measure from $\mathbf{P}_{\boldsymbol{\theta}}$ to $\mathbf{P}_{\tilde{\boldsymbol{\theta}}}$ and applying property (i),

$$I = \mathbf{E}_{\tilde{\boldsymbol{\theta}}}\left[\exp\{\boldsymbol{\ell}_{\boldsymbol{\theta}}(T) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(T)\}; \boldsymbol{\ell}_{\boldsymbol{\theta}}(T) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(T) < \log\frac{\eta}{\alpha}, \, \boldsymbol{D} = \boldsymbol{B}\right]$$
$$\leq \frac{\eta}{\alpha}\,\mathbf{P}_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{D} = \boldsymbol{B}) \leq \eta.$$

Meanwhile,

$$II \leq \mathbf{P}_{\boldsymbol{\theta}}\left(T \leq \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}, \, \boldsymbol{\ell}_{\boldsymbol{\theta}}(T) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(T) \geq \log\frac{\eta}{\alpha}\right)$$
$$+\mathbf{P}_{\boldsymbol{\theta}}\left(T \geq \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}, \, \boldsymbol{D} = \boldsymbol{B}\right).$$

The first term in this equation can be upper bounded by

$$\mathbf{P}_{\boldsymbol{\theta}}\left(\frac{1}{q|\log\alpha|}\max_{1\leq n\leq \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}} \boldsymbol{\ell}_{\boldsymbol{\theta}}(n) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(n) \geq \frac{1}{q}\frac{|\log\alpha| + \log\eta}{|\log\alpha|}\right)$$
$$= \mathbf{P}_{\boldsymbol{\theta}}\left(\frac{1}{\boldsymbol{I}_s(N;\boldsymbol{\theta})}\max_{1\leq n\leq N} \boldsymbol{\ell}_{\boldsymbol{\theta}}(n) - \boldsymbol{\ell}_{\tilde{\boldsymbol{\theta}}}(n) \geq \frac{1}{q}\frac{|\log\alpha| + \log\eta}{|\log\alpha|}\right) \equiv \epsilon_{s,\boldsymbol{\theta}}(\alpha, \boldsymbol{B}, q, \eta),$$

where we denote $N \equiv \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}$ to make the expression cleaner. Based on property (ii) and Lemma 1, we know that $\epsilon_{s,\boldsymbol{\theta}}(\alpha, \boldsymbol{B}, q, \eta) \to 0$ as $\alpha \to 0$.

Combining all above results, we have shown that

$$\mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{D} = \boldsymbol{B}) \leq \eta + \epsilon_{s,\boldsymbol{\theta}}(\alpha, \boldsymbol{B}, q, \eta) + \mathbf{P}_{\boldsymbol{\theta}}\left(T \geq \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}, \, \boldsymbol{D} = \boldsymbol{B}\right).$$

Summing over $\boldsymbol{B} \in \mathcal{U}_{s,\boldsymbol{\theta}}$, we obtain

$$1 - \alpha \leq 2^K\eta + \epsilon_{s,\boldsymbol{\theta}}(\alpha, q, \eta) + \mathbf{P}_{\boldsymbol{\theta}}\left(T \geq \boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}\right),$$

where we denote $\epsilon_{s,\boldsymbol{\theta}}(\alpha, q, \eta) \equiv \sum_{\boldsymbol{B}\in\mathcal{U}_{s,\boldsymbol{\theta}}} \epsilon_{s,\boldsymbol{\theta}}(\alpha, q, \boldsymbol{B}, \eta)$, which also converges to zero as $\alpha \to 0$. Applying Markov's inequality, it follows that

$$\frac{\mathrm{E}_{\boldsymbol{\theta}}[T]}{\boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}} \geq \mathbf{P}_{\boldsymbol{\theta}}\left(T \geq q\boldsymbol{I}_s(|\log\alpha|;\boldsymbol{\theta})^{-1}\right) \geq 1 - \alpha - 2^K\eta - \epsilon_{s,\boldsymbol{\theta}}(\alpha, q, \eta),$$

and, thus,

$$\frac{\mathrm{E}_{\boldsymbol{\theta}}[T]}{\boldsymbol{I}_s(|\log\alpha|;\boldsymbol{\theta})^{-1}} \geq \frac{\boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}}{\boldsymbol{I}_s(|\log\alpha|;\boldsymbol{\theta})^{-1}}\big(1-\alpha-2^K\eta-\epsilon_{s,\boldsymbol{\theta}}(\alpha,q,\eta)\big).$$

Based on condition (2.5) and Lemma 2, we know that $\boldsymbol{I}_s(q|\log\alpha|;\boldsymbol{\theta})^{-1}/\boldsymbol{I}_s(|\log\alpha|;\boldsymbol{\theta})^{-1} \to 1$ as $\alpha \to 0$ and $q \to 1$. Taking infimum over $\delta \in \Delta_s(\alpha)$, first letting $\alpha \to 0$, then letting $q \to 1$ and $\eta \to 0$, the proof is complete. □

S3.2 Related to the error control of the proposed test

Lemma 3. *For any $k \in [K]$ and $\theta \in \Theta_k$,*

$$\exp\left\{\hat{\ell}_k(n) - \ell_k(n;\theta)\right\}, \quad n \in \mathbb{N}$$

is a mean-one martingale with respect to filtration $\{\mathcal{F}_k(n),\, n \in \mathbb{N}\}$ and probability measure P_{k,θ_k}.

Proof of Lemma 3. See Lemma D.2 of Song and Fellouris (2019). □

Remark 1. By this lemma and independence across streams, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $B \subseteq [K]$, we can define a probability measure $\mathbf{Q}_{\boldsymbol{\theta},B}$ so that

$$\frac{d\mathbf{Q}_{\boldsymbol{\theta},B}}{d\mathbf{P}_{\boldsymbol{\theta}}}(\mathcal{F}(n)) = \prod_{k\in B}\exp\left\{\hat{\ell}_k(n) - \ell_k(n;\theta)\right\}.$$

Proof of Theorem 2. Fix $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{a} \in (0,\infty)^K$. To show (4.16), it suffices to note that, for any $k \in [K]$,

$$\xi_k(n) \equiv \hat{\ell}_k(n) - \ell_k^{(M-1)}(n) \geq \hat{\ell}_k(n) - \max_{m\in[M]\setminus H_{k,\theta_k}} \ell_k^m(n), \quad n \in \mathbb{N},$$

which goes to infinity as $n \to \infty$ a.s. under $\mathbf{P}_{\boldsymbol{\theta}}$ provided Assumption 2.

To show (4.17), we fix $s \in [K]$ and note that $\{|\boldsymbol{H}_{\boldsymbol{\theta}} \triangle \hat{\boldsymbol{D}}| \geq s\}$ occurs if and only if there exist (i) $B \subseteq [K]$ that $|B| = s$ and $H_{k,\theta_k} \neq 0$ for all $k \in B$, and (ii) $\{m_k,\, k \in B\}$ that $m_k \in [M] \backslash H_{k,\theta_k}$ for all $k \in B$, so that event

$$\Gamma\big(B,\, \{m_k,\, k \in B\}\big) \equiv \{\hat{D}_k = m_k,\, \forall\, k \in B\}$$

occurs. Since there are at most $\binom{K}{s}(M-1)^s$ such combinations, it suffices to show $\mathbf{P}_{\boldsymbol{\theta}}\big(\Gamma\big(B,\, \{m_k,\, k \in B\}\big)\big) \leq e^{-a_s}$ for each of such combinations.

Fix such a combination. Note that for any $k \in B$, $\hat{D}_k \equiv \underset{m \in [M]}{\operatorname{argmax}}\, \ell_k^m(\hat{T}) \neq H_{k,\theta_k}$ implies $\ell_k^{(M-1)}(\hat{T}) \geq \ell_k^{H_{k,\theta_k}}(\hat{T}) \geq \ell_k(\hat{T}; \theta_k)$, so on the event of $\Gamma\big(B,\, \{m_k,\, k \in B\}\big)$ we have

$$\sum_{k \in B} \Big(\hat{\ell}_k(\hat{T}) - \ell_k(\hat{T}; \theta_k)\Big) \geq \sum_{k \in B} \Big(\hat{\ell}_k(\hat{T}) - \ell_k^{(M-1)}(\hat{T})\Big) = \sum_{k \in B} \xi_k(\hat{T}) \geq a_s.$$

By Remark 1, changing measure from $\mathbf{P}_{\boldsymbol{\theta}}$ to $\mathbf{Q}_{\boldsymbol{\theta},B}$, we have

$$\begin{aligned}
&\mathbf{P}_{\boldsymbol{\theta}}\big(\Gamma\big(B,\, \{m_k,\, k \in B\}\big)\big) \\
&= \mathbf{E}_{\mathbf{Q}_{\boldsymbol{\theta},B}}\left[-\exp\left\{\sum_{k \in B} \Big(\hat{\ell}_k(\hat{T}) - \ell_k(\hat{T}; \theta_k)\Big)\right\}; \Gamma\big(B,\, \{m_k,\, k \in B\}\big)\right] \\
&\leq e^{-a_s}\, \mathbf{Q}_{\boldsymbol{\theta},B}\big(\Gamma\big(B,\, \{m_k,\, k \in B\}\big)\big) \leq e^{-a_s},
\end{aligned}$$

where $\mathbf{E}_{\mathbf{Q}_{\boldsymbol{\theta},B}}$ represents the expectation under $\mathbf{Q}_{\boldsymbol{\theta},B}$. $\square$

S3.3 Related to the asymptotic upper bound of the proposed test

Lemma 4. *Let* $\{A_k(n),\, n \in \mathbb{N}\}$, $k \in [K]$ *be* $K \geq 1$ *sequences of random variables and* $\{a_k(n),\, n \in \mathbb{N}\}$, $k \in [K]$ *be* K *sequences of positive real numbers that increase to infinity. Let* $q \in (0,1)$*. For any* $\boldsymbol{b} = (b_1, \ldots, b_K) \in (0,\infty)^K$, *let*

$$T(\boldsymbol{b}) \equiv \inf\left\{n \in \mathbb{N} : A_k(n) \geq a_k \text{ for all } k \in [K]\right\}.$$

If, for any $k \in [K]$,

$$\sum_{n=1}^{\infty} \mathrm{P}\left(\frac{A_k(n)}{a_k(n)} \leq q\right) < \infty,$$

then as $\max_{k\in[K]} b_k \to \infty$,

$$\mathrm{E}[T(\boldsymbol{b})] \lesssim \max_{k\in[K]} a_k(b_k/q)^{-1}.$$

Proof of Lemma 4. Fix $\boldsymbol{b} \in (0,\infty)^K$. Denote $N(\boldsymbol{b}) \equiv \max_{k\in[K]} a_k(b_k/q)^{-1}$. It is clear that

$$\mathrm{E}[T(\boldsymbol{b})] = \sum_{n=0}^{\infty} \mathrm{P}(T(\boldsymbol{b}) > n) \leq N(\boldsymbol{b}) + \sum_{n=N(\boldsymbol{b})}^{\infty} \mathrm{P}(T(\boldsymbol{b}) > n).$$

For any $n \geq N(\boldsymbol{b})$, which implies $b_k/a_k(n) \leq q$ for all $k \in [K]$, we have

$$\begin{aligned}
\mathrm{P}(T(\boldsymbol{b}) > n) &\leq \mathrm{P}\left(\exists\, k \in [K],\, A_k(n) < b_k\right) \\
&\leq \sum_{k\in[K]} \mathrm{P}(A_k(n) \leq b_k) \leq \sum_{k\in[K]} \mathrm{P}\left(\frac{A_k(n)}{a_k(n)} \leq q\right).
\end{aligned}$$

Thus,

$$\mathrm{E}[T(\boldsymbol{b})] \leq N(\boldsymbol{b}) + \sum_{k\in[K]} \sum_{n=1}^{\infty} \mathrm{P}\left(\frac{A_k(n)}{a_k(n)} \leq q\right).$$

Since the latter term is finite by condition, letting $\max_{k\in[K]} b_k \to \infty$ gives the desired result. □

Proof of Theorem 3. Fix $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Denote by $\tilde{T}$ the following stopping time:

$$\inf\left\{n \in \mathbb{N} : \forall\, s \in [K], \sum_{k\in B}\left(\hat{\ell}_k(n) - \ell_k^{m_k}(n)\right) \geq a_s \text{ for all } B \subseteq [K] \text{ that } |B| = s \right.$$
$$\left. \text{and all } \{m_k,\, k \in B\} \text{ that } m_k \in [M]\backslash H'_{k,\theta_k} \text{ for } k \in B\right\}.$$

Note that since $\ell_k^{(M-1)}(n) \leq \max_{m\in[M]\backslash H'_{k,\theta_k}} \ell_k^m(n)$, we have $\hat{T} \leq \tilde{T}$ for all $\boldsymbol{a}$, so it suffices to show $\mathbf{E}_{\boldsymbol{\theta}}[\tilde{T}] \lesssim \max_{s\in[K]} \boldsymbol{I}_s(a_s/q;\boldsymbol{\theta})^{-1}$ as $\overline{\boldsymbol{a}} \to \infty$.

Assumption 3 implies that, for any such s, B, $\{m_k,\, k \in B\}$, and $q \in (0,1)$,

$$\sum_{n=1}^{\infty} \mathbf{P}_{\boldsymbol{\theta}}\left(\frac{\sum_{k\in B}\hat{\ell}_k(n) - \ell_k^{m_k}(n)}{\sum_{k\in B} I_k^{m_k}(n;\theta_k)} \leq q\right)$$
$$\leq \sum_{k\in B}\sum_{n=1}^{\infty} \mathrm{P}_{k,\theta_k}\left(\frac{\hat{\ell}_k(n) - \ell_k^{m_k}(n)}{I_k^{m_k}(n;\theta_k)} \leq q\right) < \infty,$$

so by Lemma 4 we have, for any $q \in (0,1)$, as $\overline{\boldsymbol{a}} \to \infty$,

$$\mathbf{E}_{\boldsymbol{\theta}}[\tilde{T}] \lesssim \max_{s,\,B,\,\{m_k, k\in B\}} \min\left\{n \in \mathbb{N} : \sum_{k\in B} I_k^{m_k}(n;\theta_k) \geq a_s/q\right\}$$
$$= \max_{s\in[K]} \min\left\{n \in \mathbb{N} : \boldsymbol{I}_s(n;\boldsymbol{\theta}) \geq a_s/q\right\} = \max_{s\in[K]} \boldsymbol{I}_s(a_s/q;\boldsymbol{\theta})^{-1}.$$

Letting $q \to 1$ and applying Lemma 2 gives the desired upper bound. □

S4 Model misspecification

One of our reviewers raised the concern about model misspecification and asked how the proposed test performs under misspecified models. Since the problem caused by model misspecification is basically the same, we focus on the setup of sequential single testing with two hypotheses, in which case the proposed test reduces to the well-known adaptive SPRT (see, e.g., (Tartakovsky et al., 2014, Chapter 5)). We also assume for simplicity that the data are i.i.d., so that the information functions are linear in time.

We first explain what model misspecification means in this setup. Let $\Theta_{\text{sma}} \subsetneqq \Theta_{\text{big}}$ be parameter spaces and Θ^1, Θ^2 be disjoint subsets of Θ_{sma}, and consider the following two models:

Model 1: Assume that $\theta \in \Theta_{\text{sma}}$ and test

$$\theta \in \Theta^1 \quad \text{versus} \quad \theta \in \Theta^2.$$

Model 2: Assume that $\theta \in \Theta_{\text{big}}$ and test the same hypotheses.

Then, if the truth is $\theta \in \Theta_{\text{big}} \backslash \Theta_{\text{sma}}$, Model 1 is misspecified and Model 2 is correctly specified.

We then study how the adaptive SPRT under these two models differs in form and performance. First of all, under model $\square \in \{\text{sma}, \text{big}\}$, the test takes the following form:

$$\hat{T}_{\square} \equiv \inf\left\{n \geq 1 : \hat{\ell}_{\square}(n) - \ell^1_{\square}(n) \wedge \ell^2_{\square}(n) \geq a\right\},$$
$$\hat{D}_{\square} \equiv \operatorname*{argmax}_{m\in\{1,2\}} \ell^m_{\square}(\hat{T}).$$

It is clear that $\ell^m_{\text{sma}} = \ell^m_{\text{big}} \equiv \ell^m$ for $m \in \{1,2\}$, because the hypotheses are the same under the two models, but $\hat{\ell}_{\text{sma}} \leq \hat{\ell}_{\text{big}}$, because the global MLE in the latter model is taken with respect to a bigger parameter space. So a first observation is that, the expected sample size of the test increases under misspecified models.

A second observation is that, as long as Assumption 1 holds under the true parameter, i.e.,

$$\hat{\ell}_{\text{sma}}(n) - \ell^1(n) \wedge \ell^2(n) \to \infty \text{ almost surely under } \mathrm{P}_\theta, \tag{S4.14}$$

the test under the misspecified model still terminates almost surely, i.e., $\mathrm{P}_\theta(\hat{T}_{\text{sma}} < \infty) = 1$. Note that the requirement of error control does not apply, because θ does not belong to any of the hypotheses. Further, if Assumption 3 holds under the true parameter, i.e., if there exists some $J^* > 0$ such that

$$\frac{1}{n}\left(\hat{\ell}_{\text{sma}}(n) - \ell^1(n) \wedge \ell^2(n)\right) \xrightarrow{\text{completely}} J^* \text{ under } \mathrm{P}_\theta, \tag{S4.15}$$

then the test under the misspecified model still admits an asymptotic approximation to its expected sample size, i.e., $\mathrm{E}_\theta[\hat{T}_{\text{sma}}] \sim a/J^*$ as $a \to \infty$.

However, the optimal expected expected sample size under the correctly specified model is asymptotically a/I^*, where constant I^* is characterized by

$$\begin{aligned} &\inf_{\theta' \in \Theta^1 \cup \Theta^2} \mathrm{E}_\theta \left[\ell(1;\theta) - \ell(1;\theta')\right] \equiv I^*, \\ &\frac{1}{n}\left(\hat{\ell}_{\mathrm{big}}(n) - \ell^1(n) \wedge \ell^2(n)\right) \xrightarrow{completely} I^* \text{ under } \mathrm{P}_\theta, \end{aligned} \tag{S4.16}$$

which, generally speaking, satisfies $I^* \geq J^*$. Therefore, the test under the misspecified model terminates almost surely as long as condition (S4.14) holds and admits an asymptotic approximation as long as condition (S4.15) holds, but the asymptotic approximation is, generally speaking, suboptimal. In what follows, we present a concrete example for illustration.

S4.1 Example

We use $(\mathrm{Normal}, \theta)$, $\theta \in \mathbb{R}$ to mean $N(\theta, 1)$, i.e., Normal distribution with mean θ and unit variance, and $(\mathrm{Laplace}, \theta)$, $\theta \in \mathbb{R}$ to mean $L(\theta, 1)$, i.e., Laplace distribution with mean θ and scale parameter 1, and consider the following two models:

Model 1: Assume that the distribution of X belongs to

$$\Theta_{\mathrm{sma}} \equiv \{(\mathrm{Normal}, \theta) : \theta \in \mathbb{R}\},$$

and test

$$\Theta^1 \equiv \{(\mathrm{Normal}, \theta) : \theta \leq -\delta\} \quad \text{versus} \quad \Theta^2 \equiv \{(\mathrm{Normal}, \theta) : \theta \geq \delta\}.$$

Model 2: Assume that the distribution of X belongs to

$$\Theta_{\mathrm{big}} \equiv \{(\mathrm{Normal}, \theta) : \theta \in \mathbb{R}\} \cup \{(\mathrm{Laplace}, \theta) : \theta \in \mathbb{R}\},$$

and test the same hypotheses.

Suppose that the true distribution of X is $(\mathrm{Laplace}, 0)$, so Model 1 is misspecified and Model 2 is correctly specified.

In this case, based on similar methods as in the examples of Section S1, it can be shown that condition (S4.15) holds with $J^* = \delta^2/2$, where

$$\hat{\ell}_{\mathrm{sma}}(n) = \hat{\ell}_{\mathrm{sma}}(n-1) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\left(X(n) - \hat{\theta}_{\mathrm{sma}}(n-1)\right)^2 \text{ and } \hat{\theta}_{\mathrm{sma}}(n) = \bar{X}(n)$$

always adopt the form of log-likelihood and MLE of Normal distributions, although the true data-generating distribution is Laplace. So the test under the misspecified model, i.e., Model 1, terminates almost surely and its expected sample size $\sim a/J^*$ as $a \to \infty$ under $X \sim L(0,1)$.

Besides, we have, for any $\theta \in \mathbb{R}$,

$$I(L(0,1), N(\theta,1)) \equiv \mathrm{E}_{X\sim L(0,1)}\left[\ell(1; L(0,1)) - \ell(1; N(\theta,1))\right] = \frac{\theta^2}{2} + \frac{1}{2}\log\frac{\pi}{2},$$

and it can be shown that condition (S4.16) holds with $I^* = \frac{\delta^2}{2} + \frac{1}{2}\log(2\pi) >$

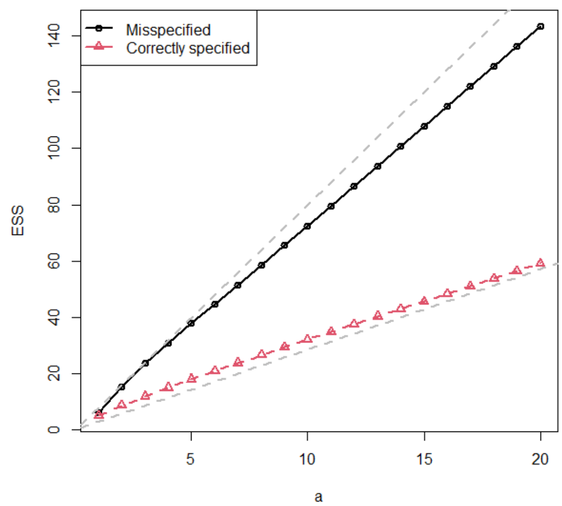

Figure 4: Expected sample size of the adaptive SPRT under misspecified model (Model 1 in Section S4.1) and that under the correctly specified model (Model 2). The gray lines represent asymptotic approximations.

$\delta^2/2 = J^*$, where

$$\hat{\ell}_{\text{big}}(n) = \hat{\ell}_{\text{big}}(n-1) + \begin{cases} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(X(n) - \hat{\theta}_{\text{big}}(n-1)\right)^2, \\ -\log 2 - |X(n) - \hat{\theta}_{\text{big}}(n-1)|, \end{cases} \quad \text{and}$$

$$\hat{\theta}_{\text{big}}(n) = \begin{cases} \bar{X}(n), \text{ if } -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left(X(t) - \bar{X}(n)\right)^2 \geq -n\log 2 - \sum_{t=1}^{n}|X(t) - X_{\text{Median}}(n)|, \\ X_{\text{Median}}(n), \text{ otherwise.} \end{cases}$$

So the test under the misspecified model is suboptimal with asymptotic relative efficiency J^*/I^*. Figure 4 visualizes this phenomenon, where we select $\delta = 0.5$.

Bibliography

Anderson, T. W. and L. A. Goodman (1957). Statistical inference about markov chains. *The Annals of Mathematical Statistics*, 89–110.

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.

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

Deshmukh, A., V. V. Veeravalli, and S. Bhashyam (2021). Sequential controlled sensing for composite multihypothesis testing. *Sequential Analysis 40*(2), 259–289.

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

Song, Y. and G. Fellouris (2019). Sequential multiple testing with generalized error control: An asymptotic optimality theory. *The Annals of Statistics 47*(3), 1776 – 1803.

Stoica, G. (2007). Baum–katz–nagaev type results for martingales. *Journal of Mathematical Analysis and Applications 336*(2), 1489–1492.

Tartakovsky, A., I. Nikiforov, and M. Basseville (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection* (1st ed.). Chapman & Hall/CRC.

Xing, Y. and G. Fellouris (2025). Asymptotically optimal sequential multiple testing with asynchronous decisions. *Bernoulli 31*(1), 271–294.