

BAYESIAN SEQUENTIAL COMPOSITE HYPOTHESIS TESTING IN DISCRETE TIME*

ERIK EKSTRÖM** AND YUQIONG WANG

Abstract. We study the sequential testing problem of two alternative hypotheses regarding an unknown parameter in an exponential family when observations are costly. In a Bayesian setting, the problem can be embedded in a Markovian framework. Using the conditional probability of one of the hypotheses as the underlying spatial variable, we show that the cost function is concave and that the posterior distribution becomes more concentrated as time goes on. Moreover, we study time monotonicity of the value function. For a large class of model specifications, the cost function is non-decreasing in time, and the optimal stopping boundaries are thus monotone.

Mathematics Subject Classification. 62L10, 60G40, 62C10.

Received October 14, 2021. Accepted April 11, 2022.

1. INTRODUCTION

Assume that a sequence of random variables X_1, X_2, \dots is observed sequentially, and that the sequence is drawn from a one-parameter family of distributions depending on a real-valued random variable Θ in such a way that X_1, X_2, \dots are independent (conditional on Θ). Consider a tester who wants to test the two alternative hypotheses

$$\begin{aligned} H_0 : \quad \Theta &\leq \theta_0, \\ H_1 : \quad \Theta &> \theta_0, \end{aligned}$$

where $\theta_0 \in \mathbb{R}$ is a given constant (the 'threshold'). In the presence of an observation cost, a tradeoff between statistical precision and costly observation arises.

In a Bayesian formulation of the problem, the tester's initial belief is described by a prior distribution μ for the unknown parameter Θ . Denote by \mathcal{T} the set of \mathbb{F}^X -stopping times with values in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, where $\mathbb{F}^X = \{\mathcal{F}_n^X\}_{n=1}^\infty$ is the filtration generated by the observation process $X = \{X_n, n \geq 0\}$. Given a stopping time $\tau \in \mathcal{T}$, let \mathcal{D}^τ be the set of \mathcal{F}_τ^X -measurable random variables D with values in $\{0, 1\}$. The random variable D here represents the decision of the tester, with ' $D = d$ ' representing that hypothesis H_d is accepted, $d = 0, 1$.

*The first author gratefully acknowledges support from Vetenskapsrådet (The Swedish Research Council).

Keywords and phrases: Sequential analysis, hypothesis testing, exponential family, optimal stopping.

Department of Mathematics, Uppsala University, Box 256, 75105 Uppsala, Sweden.

** Corresponding author: ekstrom@math.uu.se

We define the cost

$$V := \inf_{\tau \in \mathcal{T}} \inf_{D \in \mathcal{D}^\tau} \{ \mathbb{P}(D = 1, \Theta \leq \theta_0) + \mathbb{P}(D = 0, \Theta > \theta_0) + c\mathbb{E}[\tau] \}, \quad (1.1)$$

where $c > 0$ is a given and fixed cost of each observation.

The case when μ is a two-point distribution with

$$\begin{aligned} \mathbb{P}(\Theta = \theta_2) &= \pi, \\ \mathbb{P}(\Theta = \theta_1) &= 1 - \pi, \end{aligned}$$

where $\pi \in (0, 1)$ and $\theta_1 \leq \theta_0 < \theta_2$ was studied in the classical reference [24], see also [21, Chapter 4.1]. It turns out that the statistical problem (1.1) can be reduced to an optimal stopping problem in terms of the posterior probability process $\Pi_n := \mathbb{P}(\Theta = \theta_2 | \mathcal{F}_n)$, and since Π in this case is a (time-homogeneous) Markov process, the stopping problem can be embedded in a Markovian framework. It is shown in [21] that the cost function is concave in the prior belief $\Pi_0 = \pi$; as a consequence, the continuation region is an interval, and the optimal stopping time is the first exit time from this interval (the latter property was also obtained in [24]).

In the current article we relax the assumption about a two-point prior distribution and study the sequential analysis problem (1.1) in a Bayesian set-up for general prior distributions μ . To do that, we impose a one-dimensional exponential structure on the distribution of X_k . As in [21], the conditional probability process Π is then still Markovian; however, Π is in general time-inhomogeneous, which leads to time-dependence in the cost function, and the study of optimal strategies is more involved. In the absence of explicit solutions for the cost and the optimal strategy, we focus on structural properties of the solution. In particular, we prove that spatial concavity of the cost function holds regardless of the prior distribution. We also show a concentration result for the posterior distribution, which combined with the concavity result has implications for the monotonicity of the cost with respect to the time parameter.

While the analysis in the classical studies [21, 24] is restricted to the case with a two-point distribution for the unknown parameter, it allows for an arbitrary distribution of the observation sequence. In contrast, our set-up allows for an arbitrary distribution of the unknown parameter, which then forces us to restrict our attention to the exponential class of distributions for the observations. We also note that the key complication when abandoning the two-point distribution is the time-dependencies that are intrinsic in the problem; by exploiting the fact of the exponential class that the possible posterior distributions at each time point belongs to a one-parameter family of distributions, we manage to overcome irregularities with respect to time. In fact, we use a concentration result for the posterior distribution to prove that a large class of models exhibits monotonicity with respect to time.

1.1. Literature review

The problem of sequential testing of an unknown parameter has attracted much attention in the statistical literature, with [24] as an early reference covering the case of two simple hypotheses and independent and identically distributed observations. Sequential testing of composite hypotheses in a discrete time setting with Bernoulli distributed observations is studied in [17, 18], with linear penalty for wrong decisions and relying on a conjugate prior for the unknown parameter. In [23], Sobel studies sequential testing of composite hypotheses for an arbitrary class of distributions in the exponential family and with a general prior distribution of the unknown parameter. In a key result, he establishes the existence of two stopping boundaries beyond which it is optimal to stop. Related literature in discrete time, but more focused on the case of sequential *estimation*, includes [1, 6].

Another strand of literature has focused on continuous time approximations of sequential testing problems and their connections with free boundary problems. For the sequential testing of two simple hypotheses, [22] solved the problem of determining the unknown drift of a Brownian motion, and [19] solved the corresponding sequential testing problem of determining an unknown intensity of a Poisson process. In [2], a problem with

composite hypotheses was studied in continuous time and for a normal prior distribution, with a ‘0-1’ loss function for wrong decisions (as in (1.1)), and in a series of papers (see [7] and the references therein), Chernoff studied the same problem but with linear penalty functions. In the case of sequential composite hypothesis testing, explicit solutions are rare, and a main focus in this literature is on deriving asymptotics of the problem as the cost of observation tends to zero, as well as asymptotically optimal solutions (*e.g.* [3, 16, 20]) and deriving bounds for the stopping boundaries.

More recent literature has focused on different variants of these continuous-time problems. To mention a few, [12] studies a version with finite horizon, [8] studies a setting with combined learning from several Brownian motions and compound Poisson processes, and [11] studies Wiener sequential testing in a multi-dimensional set-up. All these papers study simple hypotheses, *i.e.* set-ups where the unknown parameters can take only two possible values. In [25], a hypothesis testing problem for a case with three possible drifts is examined, and in [10] a composite hypothesis problem for the drift of a Wiener process is studied with a general prior distribution. Moreover, [9] study a sequential estimation problem for a Wiener process in the same set-up. Key to the analysis in [9, 10] is the choice of appropriate variables. In fact, in [10] it is shown that if instead of the observation process one uses the conditional probability Π as state variable, then the corresponding continuation region is shrinking in time; a similar result holds for sequential least-square estimation if one uses the conditional expectation as state variable.

1.2. Our contribution

Even though the sequential composite hypothesis testing problem is very natural, it seems that qualitative studies of properties of the stopping boundaries are missing in the literature, even in the fundamental case of the exponential family. In the current article, we study the sequential composite hypothesis testing problem (1.1) using a Markovian approach. Our analysis is general in the sense that we treat the whole one-parameter exponential family with arbitrary prior distribution, and we thus do not rely on conjugate priors. Following [10], we use the conditional probability process as the underlying state variable, and we show that a concavity result holds in these coordinates. An immediate consequence of concavity is that each time section of the continuation region is an interval. Another reason to study concavity is the close connection between spatial concavity and monotonicity results with respect to time; this connection is well-known in continuous time models of option pricing (*cf.* [14, 15]), but still unexplored in the current setting. Our analysis shows that spatial concavity implies monotonicity results in time also in the current setting, provided that an assumption of stochastic domination holds. Consequently, under this assumption, the continuation region (in terms of the conditional probability) shrinks in time; in addition to being a fundamental property in itself, time monotonicity also is useful for the derivation of upper bounds on the growth of the stopping boundaries when translating back to the observation coordinates. Moreover, we provide a concentration result for the posterior distribution, showing that the posterior becomes more concentrated along level curves of the conditional probability; this can then be used to identify large classes of problem specifications for which the assumption of stochastic dominance is fulfilled, thereby guaranteeing time monotonicity.

The paper is organised as follows. In Section 2 we recall some basic properties of statistical inference in the exponential family, and we introduce the notion of π -level curves along which the value of the conditional probability Π is constant. In Section 3, we provide a Markovian embedding of (1.1), and we prove that the embedded cost function is spatially concave. In Section 4 we prove that the posterior distribution becomes more concentrated about the threshold θ_0 along level curves. Sections 5 and 6 deal with the question whether the value function is monotone with respect to the time parameter.

2. PRELIMINARIES ON THE EXPONENTIAL FAMILY

In this article, we will consider the case of a one-dimensional exponential family of distributions for X_k , $k \geq 1$. More precisely, let ν be a σ -finite measure ν on \mathbb{R} , and define

$$B(u) := \log \left(\int_{\mathbb{R}} \exp\{ux\} \nu(dx) \right)$$

and

$$N = \left\{ u \in \mathbb{R} : \int_{\mathbb{R}} \exp\{ux\} \nu(dx) < \infty \right\}$$

so that

$$B(u) < \infty$$

for $u \in N$. For $u \in N$, let

$$p_u(x) := \exp\{ux - B(u)\} \tag{2.1}$$

so that $\int_{\mathbb{R}} p_u \nu(dx) = 1$. We assume that the distribution of X_k , conditional on $\Theta = u$, is

$$\mathbb{P}(X_k \in A | \Theta = u) = \int_A p_u(x) \nu(dx). \tag{2.2}$$

Remark 2.1. In some literature, the notion of an exponential family allows for densities on the form $p_u(x) = \exp\{\eta(u)T(x) - B(u)\}$, and the case (2.2) in which $\eta(u) = u$ and $T(x) = x$ is then referred to as a *natural* exponential family. Using the transformed variables $\eta = \eta(u)$ and $T = T(x)$, an exponential form can be transformed into a natural form, so we may consider the natural form (as above) without loss of generality.

We start with some well-known results.

Lemma 2.2. *We have that*

(i) *B is convex, and N is an interval.*

Denote by N° the interior of N . Then

(ii) *all derivatives of B exist on N° , and they are given by the expressions obtained by formally differentiating inside the integral. In particular,*

$$B'(u) = \frac{\int_{\mathbb{R}} x \exp\{ux\} \nu(dx)}{\int_{\mathbb{R}} \exp\{ux\} \nu(dx)} = \mathbb{E}[X_1 | \Theta = u];$$

(iii) *the function $u \mapsto \mathbb{E}[G(X_1) | \Theta = u]$ is non-decreasing for any non-decreasing function $G : \mathbb{R} \rightarrow \mathbb{R}$.*

Proof. For (i) and (ii) we refer to ([5], Thm. 1.13) and ([5], Thm. 2.2), respectively. For (iii), we have

$$\begin{aligned} \frac{\partial}{\partial u} \mathbb{E}[G(X_1) | \Theta = u] &= \frac{\partial}{\partial u} \int_{\mathbb{R}} G(x) p_u(x) \nu(dx) = \int_{\mathbb{R}} G(x) (x - B'(u)) p_u(x) \nu(dx) \\ &= \mathbb{E}[G(X_1) X_1 | \Theta = u] - \mathbb{E}[G(X_1) | \Theta = u] \mathbb{E}[X_1 | \Theta = u] \geq 0, \end{aligned}$$

where the final inequality is due to the fact the covariance of two non-decreasing functions evaluated at the same random variable is non-negative. \square

We use a Bayesian set-up in which the unknown parameter Θ has a given prior distribution μ ; we assume that μ is a measure on N° , and we denote the support of μ by S . Moreover, denote

$$S^+ = S \cap (\theta_0, \infty) \quad \& \quad S^- = S \cap (-\infty, \theta_0] = S \setminus S^+.$$

Naturally, to avoid degenerate cases we assume that $0 < \mu(S^+) < 1$.

Next, by standard means, the optimization problem (1.1) can be reduced to an optimal stopping problem, *i.e.* a problem in which only one optimization (namely over τ) takes place. In fact, given a stopping time $\tau \in \mathcal{T}$, an optimal decision rule $D \in \mathcal{D}^\tau$ is given by

$$D = \begin{cases} 0 & \text{if } \Pi_\tau \leq 1/2 \\ 1 & \text{if } \Pi_\tau > 1/2, \end{cases}$$

where the posterior probability process Π is given by

$$\Pi_n := \mathbb{P}(\Theta > \theta_0 | \mathcal{F}_n^X).$$

Consequently,

$$V = \inf_{\tau \in \mathcal{T}} \mathbb{E} [\Pi_\tau \wedge (1 - \Pi_\tau) + c\tau],$$

where $a \wedge b = \min\{a, b\}$. To derive an expression for Π_1 , note that

$$\mathbb{P}(\Theta > \theta_0 | X_1 = x_1) = \frac{\int_{S^+} p_u(x_1) \mu(du)}{\int_S p_u(x_1) \mu(du)},$$

so

$$\Pi_1 = \frac{\int_{S^+} p_u(X_1) \mu(du)}{\int_S p_u(X_1) \mu(du)}.$$

More generally, at time n , given observations $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, we have by independence

$$\begin{aligned} \mathbb{P}(\Theta > \theta_0 | X_1 = x_1, \dots, X_n = x_n) &= \frac{\int_{S^+} \prod_{i=1}^n p_u(x_i) \mu(du)}{\int_S \prod_{i=1}^n p_u(x_i) \mu(du)} \\ &= \frac{\int_{S^+} \exp\{u \sum_{i=1}^n x_i - nB(u)\} \mu(du)}{\int_S \exp\{u \sum_{i=1}^n x_i - nB(u)\} \mu(du)}. \end{aligned}$$

Thus, denoting

$$Y_n := \sum_{i=1}^n X_i$$

we have

$$\Pi_n = q(n, Y_n),$$

where

$$q(n, y) := \frac{\int_{S^+} e^{uy-nB(u)} \mu(du)}{\int_S e^{uy-nB(u)} \mu(du)}.$$

Remark 2.3. The fact that Y_n is a sufficient statistic in any exponential family is well-known. Moreover, also a converse holds: under some mild conditions it holds that any family of distributions that admits a real-valued sufficient statistic for sample size larger than one is a one-parameter exponential family, see *e.g.* [4, 13].

We denote by

$$\mu_{n,y}(du) := \frac{e^{uy-nB(u)} \mu(du)}{\int_S e^{uy-nB(u)} \mu(du)}$$

the posterior distribution of Θ at time n conditional on $Y_n = y$. Note that the prior distribution satisfies $\mu = \mu_{0,0}$; however, for reasons of Markovian embedding, below we will consider simultaneously the whole family $\{\mu_{0,y}, y \in \mathbb{R}\}$ of alternative prior distributions.

Lemma 2.4. *The function $y \mapsto q(n, y) : \mathbb{R} \rightarrow (0, 1)$ is an increasing bijection for each fixed n .*

Proof. We have

$$\begin{aligned} \frac{\partial q(n, y)}{\partial y} &= \frac{\int_{S^+} u e^{uy-nB(u)} \mu(du)}{\int_S e^{uy-nB(u)} \mu(du)} - \frac{\int_S u e^{uy-nB(u)} \mu(du) \int_{S^+} e^{uy-nB(u)} \mu(du)}{(\int_S e^{uy-nB(u)} \mu(du))^2} \\ &= \mathbb{E}[\Theta \mathbb{1}_{\{\Theta > \theta_0\}} | Y_n = y] - \mathbb{P}(\Theta > \theta_0 | Y_n = y) \mathbb{E}[\Theta | Y_n = y]. \end{aligned} \quad (2.3)$$

Since $\mu_{n,y}$ assigns positive mass on each side of the threshold θ_0 , the above covariance is strictly positive. Thus $\frac{\partial q(n, y)}{\partial y} > 0$, so $q(n, \cdot)$ is strictly increasing. Moreover,

$$\frac{\int_{S^+} e^{uy-nB(u)} \mu(du)}{\int_{S^-} e^{uy-nB(u)} \mu(du)} \geq \frac{\int_{S^+} e^{(u-\theta_0)y-nB(u)} \mu(du)}{\int_{S^-} e^{-nB(u)} \mu(du)} \rightarrow \infty$$

as $y \rightarrow \infty$, so $q(n, y) \rightarrow 1$ as $y \rightarrow \infty$. A similar argument shows that $q(n, y) \rightarrow 0$ as $y \rightarrow -\infty$, so $q(n, \cdot)$ is surjective. \square

For each fixed value $\pi \in (0, 1)$, denote by $y(n, \pi)$ the unique value such that $q(n, y(n, \pi)) = \pi$. The set $\{(n, y(n, \pi)), n \geq 0\}$ consists of all points (n, y) with $q(n, y) = \pi$, and is referred to as the π -level curve. Since the function $y \mapsto q(n, y)$ is a bijection, two π -level curves with different π -values never intersect. Furthermore, they are ordered so that if $\pi_1 < \pi_2$, then $y(n, \pi_1) < y(n, \pi_2)$.

3. MARKOVIAN EMBEDDING

It follows from Lemma 2.4 that the process Π is a (time-inhomogeneous) Markov process, and we can write the Π -process in terms of Y as

$$\Pi_n = \int_{S^+} \mu_{n, Y_n}(du) = \frac{\int_{S^+} p_u(X_n) \mu_{n-1, Y_{n-1}}(du)}{\int_S p_u(X_n) \mu_{n-1, Y_{n-1}}(du)}.$$

Furthermore, this allows us to embed the optimal stopping problem (1.1) as a time-dependent problem in terms of the Markov process Π as

$$V(n, \pi) = \inf_{\tau \in \mathcal{T}} \mathbb{E}_{n, \pi} [\Pi_{\tau+n} \wedge (1 - \Pi_{\tau+n}) + c\tau]. \tag{3.1}$$

Here $\mathbb{P}_{n, \pi}(\cdot) := \mathbb{P}(\cdot | \Pi_n = \pi)$ is the probability measure under which Θ has distribution $\mu_{n, y(n, \pi)}$. We emphasize that $V : \mathbb{N}_0 \times (0, 1) \rightarrow [0, \infty)$, i.e. π can take any value in $(0, 1)$.

Lemma 3.1. *The value function $V(n, \pi)$ satisfies*

$$V(n - 1, \pi) = \min\{\pi \wedge (1 - \pi), c + \mathbb{E}_{n-1, \pi}[V(n, \Pi_n)]\}.$$

Proof. This follows directly from the Markovian structure of the process Π . □

Lemma 3.2. *Let $f : [0, 1] \rightarrow [0, \infty)$ be a concave function. Then $\pi \mapsto \mathbb{E}_{n, \pi}[f(\Pi_{n+1})]$ is concave on $(0, 1)$.*

Proof. To simplify the notation, we prove the statement for $n = 0$. Moreover, we will assume that f is twice continuously differentiable; the general case follows readily by approximation.

First note that

$$\mathbb{E}_{0, \pi}[f(\Pi_1)] = \int_{\mathbb{R}} f\left(\frac{\alpha(x, \pi)}{\beta(x, \pi)}\right) \beta(x, \pi) \, dx,$$

where

$$\begin{aligned} \alpha(x, \pi) &= \int_{S^+} p_u(x) \mu_{0, y(0, \pi)}(du), \\ \beta(x, \pi) &= \int_S p_u(x) \mu_{0, y(0, \pi)}(du). \end{aligned}$$

Define

$$H_1(z) := f(z) + (1 - z)f'(z)$$

and

$$H_2(z) := f(z) - zf'(z).$$

Straightforward differentiation yields

$$\begin{aligned} \frac{\partial^2 \mathbb{E}_{0, \pi}[f(\Pi_1)]}{\partial \pi^2} &= \int_{\mathbb{R}} \left(f\left(\frac{\alpha}{\beta}\right) \beta_{\pi\pi} + f'\left(\frac{\alpha}{\beta}\right) \frac{(\beta\alpha_{\pi\pi} - \alpha\beta_{\pi\pi})^2}{\beta} + f''\left(\frac{\alpha}{\beta}\right) \frac{(\beta\alpha_{\pi} - \alpha\beta_{\pi})^2}{\beta^3} \right) dx \\ &\leq \int_{\mathbb{R}} \left(f\left(\frac{\alpha}{\beta}\right) \beta_{\pi\pi} + f'\left(\frac{\alpha}{\beta}\right) \frac{(\beta\alpha_{\pi\pi} - \alpha\beta_{\pi\pi})^2}{\beta} \right) dx \\ &= \int_{\mathbb{R}} \left(\alpha_{\pi\pi} H_1\left(\frac{\alpha}{\beta}\right) + (\beta - \alpha)_{\pi\pi} H_2\left(\frac{\alpha}{\beta}\right) \right) dx \\ &= I_1 + I_2, \end{aligned}$$

where

$$I_1 := \int_{\mathbb{R}} \alpha_{\pi\pi} H_1 \left(\frac{\alpha}{\beta} \right) dx \quad \& \quad I_2 := \int_{\mathbb{R}} (\beta - \alpha)_{\pi\pi} H_2 \left(\frac{\alpha}{\beta} \right) dx$$

Note that H_1 is decreasing on $(0, 1)$, and H_2 is increasing. Furthermore, by Lemma 2.4, $\frac{\alpha(x,\pi)}{\beta(x,\pi)}$ increases in x . We will show that

$$I_1 \leq 0 \quad \& \quad I_2 \leq 0.$$

To do that, first note that

$$\alpha(x, \pi) = \int_{S^+} p_u(x) \frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \mu(du),$$

so

$$\begin{aligned} I_1 &= \int_{S^+} \left(\frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \right)_{\pi\pi} \int_S p_u(x) H_1 \left(\frac{\alpha(x, \pi)}{\beta(x, \pi)} \right) dx \mu(du) \\ &= \int_{S^+} \left(\frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \right)_{\pi\pi} \mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \mid \Theta = u \right] \mu(du). \end{aligned} \tag{3.2}$$

By Lemma 2.2, the function

$$u \mapsto \mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \mid \Theta = u \right] \tag{3.3}$$

is non-increasing.

To study the first factor of the integrand in (3.2), denote $g(y) := \int_{\mathbb{R}} e^{uy} \mu(du)$ and note that

$$\frac{\partial}{\partial \pi} \left(\frac{e^{uy(0,\pi)}}{\int_S e^{uy(0,\pi)} \mu(du)} \right) = \frac{\partial}{\partial y} \left(\frac{e^{uy}}{g(y)} \right) \Big|_{y=y(0,\pi)},$$

where

$$\pi(y) := \frac{\int_{S^+} e^{uy} \mu(du)}{\int_S e^{uy} \mu(du)}.$$

Consequently,

$$\frac{\partial^2}{\partial \pi^2} \left(\frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \right) = \frac{\pi'(y) \frac{\partial^2}{\partial y^2} \left(\frac{e^{uy}}{g(y)} \right) - \pi''(y) \frac{\partial}{\partial y} \left(\frac{e^{uy}}{g(y)} \right)}{\pi'(y)^3} \Big|_{y=y(0,\pi)}$$

Using

$$\frac{\partial}{\partial y} \frac{e^{uy}}{g} = \frac{e^{uy}}{g^2} (ug - g')$$

and

$$\pi' = \frac{g \int_{S^+} u e^{uy} \mu(du) - g' \int_{S^+} e^{uy} \mu(du)}{g^2},$$

straightforward calculations show that

$$\frac{\partial^2}{\partial \pi^2} \left(\frac{e^{uy(0,\pi)}}{\int_S e^{uy(0,\pi)} \mu(du)} \right) = \frac{e^{uy(0,\pi)} F(u)}{(\pi')^3 g^3},$$

where

$$\begin{aligned} F(u) &= u^2 \left(g \int_{S^+} u e^{uy} \mu(du) - g' \int_{S^+} e^{uy} \mu(du) \right) \\ &\quad + u \left(g'' \int_{S^+} e^{uy} \mu(du) - g \int_{S^+} u^2 e^{uy} \mu(du) \right) \\ &\quad + g' \int_{S^+} u^2 e^{uy} \mu(du) - g'' \int_{S^+} u e^{uy} \mu(du). \end{aligned}$$

Note that F is a quadratic function in u , and that the coefficient of u^2 is positive since

$$g \int_{S^+} u e^{uy} \mu(du) - g' \int_{S^+} e^{uy} \mu(du) = g^2 \text{Cov}_{0,\pi}(\Theta, \mathbb{1}_{\{\Theta > \theta_0\}}) > 0.$$

Consequently, the set $\{F < 0\}$ is a bounded interval (possibly empty). Moreover, since

$$\pi = \frac{\int_{S^+} e^{uy(0,\pi)} \mu(du)}{g(y(0,\pi))} = 1 - \frac{\int_{S^-} e^{uy(0,\pi)} \mu(du)}{g(y(0,\pi))},$$

we have

$$\int_{S^+} \frac{\partial^2}{\partial \pi^2} \left(\frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \right) \mu(du) = \int_{S^-} \frac{\partial^2}{\partial \pi^2} \left(\frac{e^{uy(0,\pi)}}{\int_{\mathbb{R}} e^{uy(0,\pi)} \mu(du)} \right) \mu(du) = 0. \tag{3.4}$$

Therefore we must have

$$F(\theta_0) < 0,$$

so the interval $\{F < 0\} \neq \emptyset$. Denote the end-points of this interval by u_0 and u_1 , respectively, so that $\{F < 0\} = (u_0, u_1)$, with $\theta_0 \in (u_0, u_1)$. Then, using (3.3) we find that

$$\begin{aligned} I_1 &= \int_{S \cap (-\infty, u_1)} \frac{e^{uy(0,\pi)} F(u)}{(\pi'(y(0,\pi)))^3 g^3(y(0,\pi))} \mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u \right] \mu(du) \\ &\quad + \int_{S \cap [u_1, \infty)} \frac{e^{uy(0,\pi)} F(u)}{(\pi'(y(0,\pi)))^3 g^3(y(0,\pi))} \mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u \right] \mu(du) \\ &\leq \mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u_1 \right] \int_{(\theta_0, u_1)} \frac{e^{uy(0,\pi)} F(u)}{(\pi'(y(0,\pi)))^3 g^3(y(0,\pi))} \mu(du) \end{aligned}$$

$$\begin{aligned}
 & +\mathbb{E} \left[H_1 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u_1 \right] \int_{[u_1, \infty)} \frac{e^{uy(0, \pi)} F(u)}{(\pi'(y(0, \pi)))^3 g^3(y(0, \pi))} \mu(du) \\
 & = 0,
 \end{aligned}$$

where we used (3.4) in the last equality.

Similarly, $\mathbb{E} \left[H_2 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u \right]$ increases in u , so

$$\begin{aligned}
 I_2 & = \int_{S^+ \cap (-\infty, u_0]} \frac{e^{uy(0, \pi)} F(u)}{(\pi'(y(0, \pi)))^3 g^3(y(0, \pi))} \mathbb{E} \left[H_2 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u \right] \mu(du) \\
 & \quad + \int_{S^- \cap (u_0, \infty)} \frac{e^{uy(0, \pi)} F(u)}{(\pi'(y(0, \pi)))^3 g^3(y(0, \pi))} \mathbb{E} \left[H_2 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u \right] \mu(du) \\
 & \leq \mathbb{E} \left[H_2 \left(\frac{\alpha(X_1, \pi)}{\beta(X_1, \pi)} \right) \middle| \Theta = u_0 \right] \int_{S^-} \frac{e^{uy(0, \pi)} F(u)}{(\pi'(y(0, \pi)))^3 g^3(y(0, \pi))} \mu(du) \\
 & = 0.
 \end{aligned}$$

Thus $\pi \mapsto \mathbb{E}_{0, \pi}[f(\Pi_1)]$ is concave. □

Theorem 3.3. *The function $\pi \mapsto V(n, \pi)$ is concave for each fixed $n \geq 0$.*

Proof. Define the cost function $V^N(n, \pi)$ as in (3.1), but with the infimum being taken over stopping times $\tau \leq N - n$ (V^N is then the value function in a problem with a finite horizon). By an iterated use of Lemma 3.1 and Lemma 3.2 and the fact that the minimum of two concave functions is concave, $\pi \mapsto V^N(n, \pi)$ is concave. Moreover, it is straightforward to check that $V^N(n, \pi) \rightarrow V(n, \pi)$ as $N \rightarrow \infty$, and since the pointwise limit of concave functions is concave, the result follows. □

So far we have been working under the assumption that $\pi \in (0, 1)$. One can further extend the value function V to the boundary points $\pi \in \{0, 1\}$ by setting $V(n, 0) = V(n, 1) = 0$ for all n . In this way, V is defined for every $\pi \in [0, 1]$ and the concavity is preserved.

In accordance with standard stopping theory, we introduce the continuation region \mathcal{C} by

$$\mathcal{C} := \{(n, \pi) \in \mathbb{N}_0 \times [0, 1] : V(n, \pi) < \pi \wedge (1 - \pi)\},$$

and the stopping region \mathcal{D} by

$$\mathcal{D} := \{(n, \pi) \in \mathbb{N}_0 \times [0, 1] : V(n, \pi) = \pi \wedge (1 - \pi)\}.$$

The stopping time

$$\tau^* := \inf\{k \geq 0 : (n + k, \Pi_{n+k}) \in \mathcal{D}\}$$

is an optimal strategy for our testing problem.

The concavity of the value function has important implications for the structure of the continuation region.

Corollary 3.4. *There exist functions $b_1 : \mathbb{N}_0 \rightarrow [0, \frac{1}{2}]$ and $b_2 : \mathbb{N}_0 \rightarrow [\frac{1}{2}, 1]$ such that*

$$\mathcal{C} = \{(n, \pi) \in \mathbb{N}_0 \times [0, 1] : b_1(n) < \pi < b_2(n)\}.$$

Proof. Since $V(n, 0) = V(n, 1) = 0$, we have $\{(n, 0)\} \cup \{(n, 1)\} \subseteq \mathcal{D}$. The result then follows from concavity of $\pi \mapsto V(n, \pi)$ and the piecewise linearity of $\pi \mapsto \pi \wedge (1 - \pi)$. □

Remark 3.5. In view of the bijection in Lemma 2.4, the fact that time sections of the continuation region are intervals in the (n, π) -coordinates implies that also time sections of the continuation region expressed in (n, y) -coordinates are intervals. This is a well-known result, see [23] (under somewhat different assumptions).

4. CONCENTRATION OF THE POSTERIOR DISTRIBUTION

Recall that the mass above θ_0 of the posterior distribution remains constantly equal to π along a π -level curve. In this section we show that the posterior distribution becomes more concentrated around θ_0 along a level curve. This result, however natural it appears, seems to be new in the literature; for related results showing that the conditional variance of the mean-square estimate is a supermartingale, see [9].

Theorem 4.1. *If $a < \theta_0 < b$, then*

$$n \mapsto \mathbb{P}_{n,\pi}(\Theta \leq a) \quad \& \quad n \mapsto \mathbb{P}_{n,\pi}(\Theta > b)$$

are decreasing.

Proof. For the first claim, it suffices to show that

$$\mathbb{P}_{0,\pi}(\Theta \leq a) \geq \mathbb{P}_{1,\pi}(\Theta \leq a) \tag{4.1}$$

where $a < \theta_0$. Moreover, without loss of generality, we may assume that $y(0, \pi) = 0$ so that $\mu_{0,y(0,\pi)} = \mu$. Let

$$f(u) := e^{uy(1,\pi) - B(u)},$$

and let $S^a := S \cap (-\infty, a]$. Note that

$$\mathbb{P}_{0,\pi}(\Theta \leq a) = \int_{S^a} \mu(du)$$

and

$$\mathbb{P}_{1,\pi}(\Theta \leq a) = \frac{\int_{S^a} f(u)\mu(du)}{\int_S f(u)\mu(du)}.$$

Also note that

$$\frac{\partial f(u)}{\partial u} = f(u)(y(1, \pi) - B'(u)).$$

Therefore, since B is convex, we have that f changes its monotonicity (from increasing to decreasing) at most once. Now we consider two separate cases:

(i) $f(a) \leq f(\theta)$

and

(ii) $f(a) > f(\theta)$.

If (i) holds, then $(f(u) - f(a))(u - a) \geq 0$ for $u \leq \theta_0$ (since f changes its monotonicity at most once). Consequently, if $\mu(S^- \setminus S^a) \neq 0$, then

$$\frac{\int_{S^a} f(u)\mu(du)}{\int_{S^- \setminus S^a} f(u)\mu(du)} \leq \frac{f(a) \int_{S^a} \mu(du)}{f(a) \int_{S^- \setminus S^a} \mu(du)} = \frac{\int_{S^a} \mu(du)}{\int_{S^- \setminus S^a} \mu(du)}. \tag{4.2}$$

(if $\mu(S^- \setminus S^a) = 0$, then (4.1) holds trivially with equality). Since

$$\frac{\int_{S^-} f(u)\mu(du)}{\int_S f(u)\mu(du)} = 1 - \pi = \int_{S^-} \mu(du),$$

(4.2) implies that

$$\begin{aligned} \mathbb{P}_{1,\pi}(\Theta \leq a) &= \frac{\int_{S^a} f(u)\mu(du)}{\int_S f(u)\mu(du)} = \frac{\int_{S^a} f(u)\mu(du) \int_{S^a} \mu(du)}{\int_{S^-} f(u)\mu(du)} \\ &= \frac{\int_{S^a} f(u)\mu(du) \int_{S^-} \mu(du)}{\int_{S^- \setminus S^a} f(u)\mu(du) \int_{S^-} \mu(du)} \leq \int_{S^a} \mu(du) = \mathbb{P}_{0,\pi}(\Theta \leq a), \end{aligned}$$

so (4.1) holds.

On the other hand, if (ii) holds, then the fact that f changes its monotonicity at most once gives that $(f(u) - f(\theta_0))(u - \theta_0) \leq 0$ for all $u \geq a$. Consequently,

$$\frac{\int_{S^- \setminus S^a} f(u)\mu(du)}{\int_{S^+} f(u)\mu(du)} \geq \frac{f(\theta_0) \int_{S^- \setminus S^a} \mu(du)}{f(\theta_0) \int_{S^+} \mu(du)} = \frac{\int_{S^- \setminus S^a} \mu(du)}{\int_{S^+} \mu(du)}. \quad (4.3)$$

Since

$$\frac{\int_{S^+} f(u)\mu(du)}{\int_S f(u)\mu(du)} = \pi = \int_{S^+} \mu(du),$$

the inequality (4.3) yields

$$\begin{aligned} \mathbb{P}_{1,\pi}(\Theta > a) &= \frac{\int_{S \setminus S^a} f(u)\mu(du)}{\int_S f(u)\mu(du)} = \frac{\int_{S \setminus S^a} f(u)\mu(du) \int_{S^+} \mu(du)}{\int_{S^+} f(u)\mu(du)} \\ &= \left(1 + \frac{\int_{S^- \setminus S^a} f(u)\mu(du)}{\int_{S^+} f(u)\mu(du)} \right) \int_{S^+} \mu(du) \geq \int_{S \setminus S^a} \mu(du) = \mathbb{P}_{0,\pi}(\Theta > a), \end{aligned}$$

from which (4.1) follows. Finally, the second inequality (for $b > \theta_0$) follows by a similar argument. \square

As a consequence, we can show that the level curves are spreading out along the time axis.

Corollary 4.2. *Let $0 < \pi_1 < \pi_2 < 1$. Then $n \mapsto y(n, \pi_2) - y(n, \pi_1)$ is non-decreasing.*

Proof. Recall from (2.3) that

$$\begin{aligned} \frac{\partial q}{\partial y}(n, y(n, \pi)) &= \mathbb{E}_{n,\pi}[\Theta \mathbf{1}_{\{\Theta > \theta_0\}}] - \mathbb{P}_{n,\pi}(\Theta > \theta_0) \mathbb{E}_{n,\pi}[\Theta] \\ &= (1 - \pi) \mathbb{E}_{n,\pi}[\Theta \mathbf{1}_{\{\Theta > \theta_0\}}] - \pi \mathbb{E}_{n,\pi}[\Theta \mathbf{1}_{\{\Theta \leq \theta_0\}}] \end{aligned}$$

By Theorem 4.1, this covariance (with respect to the posterior distribution) is non-increasing in n . By this, the level curves are spreading out. \square

5. CONDITIONS FOR MONOTONICITY IN TIME

In this section we investigate whether $n \mapsto V(n, \pi)$ is non-decreasing. If this monotonicity holds, then the stopping boundaries b_1 and b_2 will be non-decreasing and non-increasing, respectively. To prove the monotonicity of V , we will use the following assumption.

Assumption 5.1. For any $\pi \in (0, 1)$ and $n \geq m \geq 0$, the random variable $\Pi_{m+1}|\{\Pi_m = \pi\}$ dominates $\Pi_{n+1}|\{\Pi_n = \pi\}$ in convex order.

Theorem 5.2. Assume that Assumption 5.1 holds. Then $V(n, \pi)$ is non-decreasing in n , and the boundaries b_1 and b_2 are thus non-decreasing and non-increasing, respectively.

Proof. For any concave function f , we have $\mathbb{E}_{m,\pi}[f(\Pi_{m+1})] \leq \mathbb{E}_{n,\pi}[f(\Pi_{n+1})]$. It thus follows from Lemma 3.1 that $V(n, \pi)$ is non-decreasing in n ; the monotonicity of the boundaries b_1 and b_2 is a direct consequence. \square

Since $\Pi_{m+1}|\{\Pi_m = \pi\}$ and $\Pi_{n+1}|\{\Pi_n = \pi\}$ have the same expected value π , a sufficient condition for stochastic domination in convex order is that there exists a point π_0 around which the distribution of $\Pi_{n+1}|\{\Pi_n = \pi\}$ is more concentrated compared to the distribution of $\Pi_{m+1}|\{\Pi_m = \pi\}$ in the sense that

$$\mathbb{P}_{m,\pi}(\Pi_{m+1} \leq \alpha) \geq \mathbb{P}_{n,\pi}(\Pi_{n+1} \leq \alpha)$$

and

$$\mathbb{P}_{m,\pi}(\Pi_{m+1} > \beta) \geq \mathbb{P}_{n,\pi}(\Pi_{n+1} > \beta)$$

for $\alpha < \pi_0 < \beta$. Using Theorem 4.1, we now state a sufficient condition under which the above concentration property holds.

Theorem 5.3. Assume that the observations are continuously distributed with density of the form

$$h(x)p_u(x),$$

with $\nu(dx) = h(x)dx$ for some nonnegative continuous function h such that $I := \{h > 0\}$ is an interval. Assume that either

(i) $h(x)p_u(x)$ is increasing in x on I , and S^+ is a singleton

or

(ii) $h(x)p_u(x)$ is decreasing in x on I , and S^- is a singleton

holds. Then Assumption 5.1 holds, so $n \mapsto V(n, \pi)$ is non-decreasing.

Proof. We will verify Assumption 5.1 for $m = 0$ and $\pi = \mu(S \cap [0, \infty))$; the general case follows by translation. Thus we consider two distributions μ and $\mu' := \mu_{n,y(n,\pi)}$. Let $F(a) = \mathbb{P}_{0,\pi}(\Pi_1 \leq a)$ and $G(a) = \mathbb{P}_{n,\pi}(\Pi_{n+1} \leq a)$. Since F and G are continuous distribution functions with the same expected value π , there exists $\pi_0 \in (0, 1)$ with $F(\pi_0) = G(\pi_0)$. We claim that

$$F'(\pi_0) < G'(\pi_0) \tag{5.1}$$

at such a point (unless μ is a two-point distribution), which implies that there is a single intersection point π_0 and that G is more concentrated about π_0 in comparison with F .

To prove the claim, let π_0 be such that $F(\pi_0) = G(\pi_0)$, and denote by x_1 and x_{n+1} the unique values such that $q(1, x_1) = \pi_0 = q(n + 1, y(n, \pi) + x_{n+1})$, so that $(1, x_1)$ and $(n + 1, y(n, \pi) + x_{n+1})$ are both on the π_0 -level curve. We then have

$$F(\pi_0) = \int_{-\infty}^{x_1} \int_S h(x)p_u(x)\mu(du)dx$$

and

$$G(\pi_0) = \int_{-\infty}^{x_{n+1}} \int_S h(x)p_u(x)\mu'(du)dx,$$

so

$$F'(\pi_0) = \frac{\int_S h(x_1)p_u(x_1)\mu(du)}{\frac{\partial q}{\partial y}(1, x_1)}$$

and

$$G'(\pi_0) = \frac{\int_S h(x_{n+1})p_u(x_{n+1})\mu'(du)}{\frac{\partial q}{\partial y}(n + 1, y(n, \pi) + x_{n+1})}.$$

Note that

$$\frac{\partial q}{\partial y}(1, x_1) = Cov_{1, \pi_0}(\Theta, 1_{\{\Theta \leq \theta_0\}}) \geq Cov_{n+1, \pi_0}(\Theta, 1_{\{\Theta \leq \theta_0\}}) = \frac{\partial q}{\partial y}(n + 1, y(n, \pi) + x_{n+1})$$

since the covariance decreases along the π_0 -level curve. Furthermore, the covariance is strictly decreasing along the level curve (unless μ has support on only two points). Therefore, it suffices to show that $\int_S h(x_1)p_u(x_1)\mu(du) \leq \int_S h(x_{n+1})p_u(x_{n+1})\mu'(du)$.

To do that, assume that (i) holds so that $\text{supp } \mu \cap S^+ = \{\theta_1\}$ for some $\theta_1 > \theta_0$. Then μ and both μ' are identical on S^+ , so it follows from Theorem 4.1 that μ' stochastically dominates μ . Consequently,

$$\int_{-\infty}^x \int_S h(y)p_u(y)\mu(du)dy \geq \int_{-\infty}^x \int_S h(y)p_u(y)\mu'(du)dy$$

for any x , so $F(\pi_0) = G(\pi_0)$ implies $x_1 \leq x_{n+1}$. Moreover, the relation $p(1, x_1) = p(n + 1, y(n, \pi) + x_{n+1})$ implies that

$$\frac{\int_{S^-} h(x_1)p_u(x_1)\mu(du)}{h(x_1)p_{\theta_1}(x_1)} = \frac{\int_{S^-} h(x_{n+1})p_u(x_{n+1})\mu'(du)}{h(x_{n+1})p_{\theta_1}(x_{n+1})}. \tag{5.2}$$

Since the density function is increasing in x , we have $h(x_1)p_{\theta_1}(x_1) \leq h(x_{n+1})p_{\theta_1}(x_{n+1})$. Equation (5.2) thus implies

$$\int_S h(x_1)p_u(x_1)\mu(du) \leq \int_S h(x_{n+1})p_u(x_{n+1})\mu'(du). \tag{5.3}$$

If instead (ii) holds, then $x_1 \geq x_{n+1}$, and (5.3) is derived in a similar way.

It follows that F and G have a unique intersection point π_0 (unless the support of μ has only two points, in which $F \equiv G$), and that G is more concentrated about π_0 compared to F . Consequently, the random variable $\Pi_1|\{\Pi_0 = \pi\}$ dominates $\Pi_{n+1}|\{\Pi_n = \pi\}$ in convex order. \square

Remark 5.4. At first it may seem counterintuitive that the cost $V(\pi, n)$ is increasing in n , despite the fact that more observations should enhance statistical precision. The explanation is that when moving along a level curve, the posterior distribution concentrates about the level curve, compare Theorem 4.1, and a concentrated distribution is disadvantageous from the hypothesis testing perspective.

Example (Exponential observations.) Assume that $\{X_k, k \geq 1\}$ are exponentially distributed with unknown intensity Θ and independent (conditional on Θ) so that

$$\mathbb{P}(X_1 \leq x | \Theta = u) = 1 - e^{-ux}.$$

This is not of the exponential form (2.2), but it is straightforward to check that if one instead considers $X'_k := -X_k$, then X'_k has density

$$h(x)p_u(x) = \begin{cases} \exp\{ux + \log u\} & x < 0 \\ 0 & x \geq 0 \end{cases}$$

with respect to Lebesgue measure. For $u > 0$, this density is increasing in x on $I = (-\infty, 0)$. Consequently, if S^+ is a singleton, then (i) in Theorem 5.3 gives that $V(n, \pi)$ is increasing in n , which leads to the monotonicity of the stopping boundaries.

Example (Gaussian observations with unknown variance.) If $\{X_k, k \geq 1\}$ are normally distributed with mean 0 and unknown standard deviation Θ and independent (conditional on Θ), then the random variables $X'_k = -\frac{1}{2}(X_k)^2$ are of the exponential form (2.2) with respect to the unknown variable $\Theta' := \Theta^{-2}$, with density

$$h(x)p_u(x) = \begin{cases} \frac{2}{\sqrt{-\pi x}} \exp\{ux + \frac{1}{2} \log u\} & x < 0 \\ 0 & x \geq 0 \end{cases}$$

with respect to Lebesgue measure. Note that this density is increasing in x . Also note that H_0 holds precisely when $\Theta' \geq \theta_0^{-2}$. Therefore, the value function is decreasing in time for any prior distribution μ such that S^- is a singleton.

Remark 5.5. Theorem 5.3 considers random variables that are continuously distributed on an interval. However, a closer inspection of the proof reveals that one can relax this assumption and instead assume that $H := \text{supp}(h)$ is the union of disjoint intervals. Moreover, when these intervals become small, using

approximation arguments one would expect a similar result for discrete distributions; we leave the details, as well as the precise formulation, of such a result.

6. FURTHER DISCUSSION ON TIME MONOTONICITY

While the conditions in Theorem 5.3 may appear somewhat restrictive, we have failed to remove the conditions. On the other hand, we have also been unable to produce examples within the exponential family for which the asserted time monotonicity fail. In this final section, we show that time monotonicity holds for arbitrary prior distributions μ in a few particular examples, see Sections 6.1-3 below. Based on these findings, we formulate the following conjecture.

Conjecture 6.1. *The function $V(n, \pi)$ in (3.1) is non-decreasing in n for any prior distribution μ and any ν .*

6.1. Gaussian observations with unknown mean

Assume that the Gaussian sequence $\{X_k, k \geq 1\}$ has unknown mean Θ and known standard deviation 1 so that the density (conditional on $\Theta = u$) is

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \exp\{ux - B(u)\},$$

where $B(u) = u^2/2$. In this case, the discrete time Π -process can be embedded in the corresponding continuous time process, as studied in [10]. Moreover, the discrete time problem then corresponds to the continuous time problem in [10] but with the restriction that stopping is only allowed at integer times. For such a problem, the techniques used in [10] (in particular, preservation of concavity for martingale diffusions coupled with time-decay of the diffusion coefficient of Π) show that $V(n, \pi)$ is non-decreasing in n .

6.2. Bernoulli observations

Consider a sequence $\{X_k, k \geq 1\}$ which is Bernoulli distributed with parameter Θ so that

$$\mathbb{P}(X_k = 1 | \Theta = u) = 1 - \mathbb{P}(X_k = 0 | \Theta = u) = u.$$

(This is on the exponential form if one instead uses $\Theta' := \log \frac{\Theta}{1-\Theta}$ as the unknown parameter, because then

$$\mathbb{P}(X_k = x | \Theta' = u) = \mathbb{P}(X_k = x | \Theta = \frac{e^u}{1 + e^u}) = e^{ux - \log(1 + e^u)}$$

for $x \in \{0, 1\}$.) Also assume that the distribution of Θ is μ , which is a given measure on $[0, 1]$.

Note that since observations are binary, Π_{n+1} can only take two different values if started at a given point (n, π) . For $m \leq n$ we have

$$\begin{aligned} \mathbb{P}_{m,\pi}(\Theta > \theta_0 | X_{m+1} = 1) &= \frac{\int_{S^+} u \mu_{m,y(m,\pi)}(du)}{\int_S u \mu_{m,y(m,\pi)}(du)} = \frac{1}{1 + \frac{\mathbb{E}_{m,\pi}[\theta \mathbb{1}_{\{\theta \leq \theta_0\}}]}{\mathbb{E}_{m,\pi}[\theta \mathbb{1}_{\{\theta > \theta_0\}}]}} \\ &\geq \frac{1}{1 + \frac{\mathbb{E}_{n,\pi}[\theta \mathbb{1}_{\{\theta \leq \theta_0\}}]}{\mathbb{E}_{n,\pi}[\theta \mathbb{1}_{\{\theta > \theta_0\}}]}} = \mathbb{P}_{n,\pi}(\Theta > \theta_0 | X_{n+1} = 1), \end{aligned}$$

where the inequality is a consequence of Theorem 4.1. Similarly,

$$\begin{aligned} \mathbb{P}_{m,\pi}(\Theta > \theta_0 | X_{m+1} = 0) &= \frac{1}{1 + \frac{\mathbb{E}_{m,\pi}[(1-\theta)\mathbb{1}_{\{\theta \leq \theta_0\}}]}{\mathbb{E}_{m,\pi}[(1-\theta)\mathbb{1}_{\{\theta > \theta_0\}}]}} \\ &\leq \frac{1}{1 + \frac{\mathbb{E}_{n,\pi}[(1-\theta)\mathbb{1}_{\{\theta \leq \theta_0\}}]}{\mathbb{E}_{n,\pi}[(1-\theta)\mathbb{1}_{\{\theta > \theta_0\}}]}} = \mathbb{P}_n(\Theta > \theta_0 | X_{n+1} = 0). \end{aligned}$$

Since two two-point distributions with the same mean and with mass on $\{a, b\}$ and $\{a', b'\}$, where $a \leq a' < b' \leq b$, are ordered in convex order, it follows that the distribution of Π_{m+1} under $\mathbb{P}_{m,\pi}$ dominates the distribution of Π_{n+1} under $\mathbb{P}_{n,\pi}$ in convex order. Thus Assumption 5.1 is satisfied, so time monotonicity for arbitrary priors μ holds by Theorem 5.2.

6.3. Binomial observations

Consider a general prior distribution μ on $[0, 1]$ for Θ , and observations $\{X_k, k \geq 1\}$ that are $\text{Bin}(N, \Theta)$. As in Section 6.2, this is on the exponential form if one uses $\Theta' = \log \frac{\Theta}{1-\Theta}$ as the unknown parameter.

Now consider a sequential testing problem for Bernoulli observations with the same unknown parameter Θ , but where the cost c is only imposed on the $(Nk + 1)$ -th observation, for all $k \in \mathbb{N}_0$. Denoting the value function of that sequential problem by $V^{\text{Ber}}(n, \pi)$, arguments similar to those in Section 6.2 imply that the function $k \mapsto V^{\text{Ber}}(kN, \pi)$ is non-decreasing. However, the value function $V(n, \pi)$ for Binomial observations clearly coincide with $V^{\text{Ber}}(nN, \pi)$, so time monotonicity holds also for $V(n, \pi)$.

REFERENCES

- [1] M. Alvo, Bayesian sequential estimation. *Ann. Statist.* **5** (1977) 955–968.
- [2] J.A. Bather, Bayes procedures for deciding the sign of a normal mean. *Proc. Cambridge Philos. Soc.* **58** (1962) 599–620.
- [3] P.J. Bickel, On the asymptotic shape of Bayesian sequential tests of $\theta \leq 0$ versus $\theta > 0$ for exponential families. *Ann. Statist.* **1** (1973) 231–240.
- [4] L. Brown, Sufficient statistics in the case of independent random variables. *Ann. Math. Statist.* **35** (1964) 1456–1474.
- [5] L. Brown, Fundamentals of statistical exponential families with applications in statistical decision theory. Institute of Mathematical Statistics, Hayward, CA (1986).
- [6] P. Cabilio, Sequential estimation in Bernoulli trials. *Ann. Statist.* **5** (1977) 342–356.
- [7] H. Chernoff, Sequential tests for the mean of a normal distribution. IV. (Discrete case). *Ann. Math. Statist.* **36** (1965) 55–68.
- [8] S. Dayanik and S. Sezer, Multisource Bayesian sequential binary hypothesis testing problem. *Ann. Oper. Res.* **201** (2012) 99–130.
- [9] E. Ekström, I. Karatzas and J. Vaicenavicius, Bayesian sequential least-squares estimation for the drift of a Wiener process. *Stoch. Process. Appl.* **145** (2022) 335–352.
- [10] E. Ekström and J. Vaicenavicius, Bayesian sequential testing of the drift of a Brownian motion. *ESAIM: PS.* **19** (2015) 626–648.
- [11] E. Ekström and Y. Wang, Multi-dimensional sequential testing and detection. To appear *Stochastics* (2022).
- [12] P.V. Gapeev and G. Peskir, The Wiener sequential testing problem with finite horizon. *Stoch. Stoch. Rep.* **76** (2004) 59–75.
- [13] C. Hipp, Sufficient statistics and exponential families. *Ann. Statist.* **2** (1974) 1283–1292.
- [14] D. Hobson, Volatility misspecification, option pricing and superreplication via coupling. *Ann. Appl. Probab.* **8** (1998) 193–205.
- [15] S. Janson and J. Tysk, Volatility time and properties of option prices. *Ann. Appl. Probab.* **13** (2003) 890–913.
- [16] T.L. Lai, Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16** (1988) 856–886.
- [17] D. Lindley and B. Barnett, Sequential sampling: Two decision problems with linear losses for binomial and normal random variables. *Biometrika* **52** (1965) 507–532.
- [18] S. Moriguti and H. Robbins, A Bayes test of “ $p \leq 1/2$ ” versus “ $p > 1/2$ ”. *Rep. Statist. Appl. Res. Un. Jpn. Sci. Eng.* **9** (1962) 39–60.
- [19] G. Peskir and A. Shiryaev, Sequential testing problems for Poisson processes. *Ann. Statist.* **28** (2000) 837–859.
- [20] G. Schwarz, Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** (1962) 224–236.
- [21] A. Shiryaev, *Optimal stopping rules*. Applications of Mathematics, Vol. 8. Springer-Verlag, New York-Heidelberg (1978).
- [22] A. Shiryaev, Two problems of sequential analysis. *Cybernetics* **3** (1967) 63–69 (1969).
- [23] M. Sobel, An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Stat.* **24** (1953) 319–337.

- [24] A. Wald and J. Wolfowitz, Bayes solutions of sequential decision problems. *Ann. Math. Stat.* **21** (1950) 82–99.
- [25] M. Zhitlukhin and A.N. Shiryayev, A Bayesian sequential testing problem of three hypotheses for Brownian motion. *Stat. Risk Model.* **28** (2011) 227–249.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/math-s2o-programme>