SIAM J. CONTROL OPTIM.
Vol. 51, No. 4, pp. 2922–2948

REWARD-RATE MAXIMIZATION IN SEQUENTIAL IDENTIFICATION UNDER A STOCHASTIC DEADLINE*

SAVAS DAYANIK† AND ANGELA J. YU‡

Abstract. Any intelligent system performing evidence-based decision making under time pressure must negotiate a speed-accuracy trade-off. In computer science and engineering, this is typically modeled as minimizing a Bayes-risk functional that is a linear combination of expected decision delay and expected terminal decision loss. In neuroscience and psychology, however, it is often modeled as maximizing the long-term reward rate, or the ratio of expected terminal reward and expected decision delay. The two approaches have opposing advantages and disadvantages. While Bayes-risk minimization can be solved with powerful dynamic programming techniques unlike reward-rate maximization, it also requires the explicit specification of the relative costs of decision delay and error, which is obviated by reward-rate maximization. Here, we demonstrate that, for a large class of sequential multihypothesis identification problems under a stochastic deadline, the reward-rate maximization is equivalent to a special case of Bayes-risk minimization, in which the optimal policy that attains the minimal risk when the unit sampling cost is exactly the maximal reward rate is also the policy that attains maximal reward rate. We show that the maximum reward rate is the unique unit sampling cost for which the expected total observation cost and expected terminal reward break even under every Bayes-risk optimal decision rule. This interplay between reward-rate maximization and Bayes-risk minimization formulations allows us to show that maximum reward rate is always attained. We can compute the policy that maximizes reward rate by solving an inverse Bayes-risk minimization problem, whereby we know the Bayes risk of the optimal policy and need to find the associated unit sampling cost parameter. Leveraging this equivalence, we derive an iterative dynamic programming procedure for solving the reward-rate maximization problem exponentially fast, thus incorporating the advantages of both the reward-rate maximization and Bayes-risk minimization formulations. As an illustration, we will apply the procedure to a two-hypothesis identification example.

Key words. reward-rate maximization, Bayes-risk minimization, sequential multihypothesis testing, dynamic programming, speed-accuracy trade off

AMS subject classifications. 62L15, 62C10, 60G40

DOI. 10.1137/100818005

1. Introduction.

Evidence-based decision-making under conditions of uncertainty is a fundamental problem facing any intelligent, interactive system. The brain excels in making such decisions under changing and competing objectives, a feat particularly impressive given its noisy sensors, fallible communication channels, and imperfect controllers. Similar challenges riddle artificial systems, for many applications in computer science and engineering. Understanding the computational basis of decision making within an optimality framework, therefore, would not only shed light on a critical problem in natural intelligence, but may also inspire new designs for artificial systems.

One major challenge of evidence-based decision-making is negotiating the trade-off between speed and accuracy: longer deliberation duration tends to improve the quality of the decision, but incur a concomitant opportunity cost in time. In neuro-

*Received by the editors December 13, 2010; accepted for publication (in revised form) May 21, 2013; published electronically July 16, 2013.
http://www.siam.org/journals/sicon/51-4/81800.html
†Bilkent University, Departments of Industrial Engineering and Mathematics, Bilkent 06800, Ankara, Turkey (sdayanik@bilkent.edu.tr). This author's work was partially supported by the TÜBİTAK Research grant 110M610.
‡Department of Cognitive Science, University of California San Diego, La Jolla, CA 92093 (ajyu@ucsd.edu).

science and psychology, humans [4] and animals [14] are often modeled as maximizing the long-run average reward rate, or the ratio of accuracy to expected temporal delay. In computer science and engineering modeling, the speed-accuracy trade-off is typically formalized in terms of Bayes-risk minimization, which minimizes a linear combination of expected temporal delay and response errors [18, 16, 10, 11, 15, 9, 8, 12]. The advantage of the risk minimization formulation is that the linear speed-accuracy trade-off makes it amenable to a substantial body of tools for solving or characterizing the optimal solution, including Wald's sequential statistical decision formulation [17] and Bellman's dynamic programming principle [1]. The disadvantage is the need for a free parameter specifying the relative importance of time and error, which may not be easily determined or uniquely constrained in a given application. The reward rate formulation has just the converse properties: it obviates the need for that extra speed-accuracy parameter, but also does not lend itself easily to theoretical or computational analysis. In practice, when maximizing reward-rate in neuroscience modeling, a particular parametrized class of policies is typically assumed for computational ease [14, 6, 4, 19], but which may contain neither the optimal policy nor the actual policy effectively implemented by the brain. Relatedly, when experimental subjects' behavior deviates from the conditionally optimal policy within the assumed policy space, it cannot be known whether the brain is suboptimal or the policy space itself is unsuitable.

The goal in this paper is to investigate the formal relationship between reward-rate maximization and Bayes-risk minimization, in a setting where a subject repeatedly performs statistically independent and identical experiments to identify an unknown distribution from which a stream of noisy data is being observed, while there are costs associated with misidentification, number of samples (amount of time) taken, and exceeding a stochastically distributed decision deadline. In a typical experiment, the subject samples, as long as she wants, independently and identically distributed random variables $X_1, X_2, \ldots$ with some unknown common probability density function f, which is selected by nature or the experimenter according to some known prior probability distribution from a set of m distinct alternative probability density functions $f_1, \ldots, f_m$. The subject eventually stops sampling to identify the unknown density function (chooses one of the m hypotheses), with her choice registering after an additional $T_0 > 0$ units of time that captures any fixed and known nondecision time such as motor delay. Independently of the the subject's observation and decision process, a random deadline Θ, selected by nature or the experimenter, may prematurely terminate the experiment without allowing the subject to register her choice. The subject earns a positive reward r_j for some $1 \le j \le m$ if (i) f_j is the true density and the subject correctly identifies it, and (ii) if the subject's decision is registered before the deadline Θ. At every moment in time, the subject faces the trade-off between taking longer samples to increase the probability of getting positive reward and acting fast enough to register an answer before the deadline arrives. We are interested in finding a decision rule (τ, μ) that maximizes the reward rate per unit time in the long run, whereby τ is the decision time or the number of samples observed, and $\mu \in \{1, \ldots, m\}$ is the terminal decision (choice) of one of the m hypotheses.

If M identifies the unknown true density function of the observations, then the reward in a typical experiment equals $R = 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^{m} r_j 1_{\{\mu=j, M=j\}}$, where $1_{\{\cdot\}}$ is the indicator function evaluating to 1 only when its argument is satisfied. The experiment is terminated at time $T = (\tau + T_0) \wedge \Theta$ by the deadline Θ, or by the successful registry of the subject's decision, whichever occurs earlier—"$\wedge$" denotes the minimum of the two arguments on either side. Then by the strong law of large num-

bers the long-run average reward per unit time equals $\mathbb{E}R/\mathbb{E}T$ with probability one. Therefore, the maximum reward-rate problem is equivalent to solving the stochastic optimization problem

$$V := \sup_{(\tau,\mu)} \frac{\mathbb{E}\left[1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right]}{\mathbb{E}\left[(\tau+T_0)\wedge\Theta\right]}$$

for which we will show that an optimal solution always exists and describe how to calculate the supremum and an admissible decision rule (τ,μ) which attains the supremum.

An important theoretical question is whether and how Bayes-risk minimization and reward-rate maximization are related to each other. In this work, we assume that a known prior distribution of m hypotheses is initially available and that random deadline Θ has a known geometric distribution. We demonstrate that reward-rate maximization for this class of problems is formally equivalent to solving the family $(W(c))_{c>0}$ of Bayes-risk minimization problems,

$$W(c) := \inf_{(\tau,\mu)} \mathbb{E}\left[c\big((\tau+T_0)\wedge\Theta\big)+1_{\{\tau+T_0<\Theta\}}\sum_{i\neq j} r_j 1_{\{\mu=i,M=j\}}+1_{\{\tau+T_0\geq\Theta\}}\sum_{j=1}^m r_j 1_{\{M=j\}}\right],$$

indexed by the unit sampling (observation or time) cost $c>0$, thus rendering the reward-rate maximization problem amenable to a large array of existing analytical and computational tools in stochastic control theory. In particular, we show that the maximum reward rate V is the unique unit sampling cost $c>0$ which makes the minimum Bayes risk $W(c)$ equal to the maximal expected reward $\sum_{j=1}^m r_j\mathbb{P}(M=j)$ under the prior distribution. Using the identity

$$W(c) = \sum_{j=1}^m r_j\mathbb{P}(M=j) + \inf_{(\tau,\mu)} \mathbb{E}\left[c\big((\tau+T_0)\wedge\Theta\big) - 1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right],$$

we also derive the striking relationship

$$c \gtreqless V \quad \text{if and only if} \quad \inf_{(\tau,\mu)} \mathbb{E}\left[c\big((\tau+T_0)\wedge\Theta\big) - 1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right] \gtreqless 0;$$

namely, that *the maximum reward rate V is the unique unit sampling cost c for which expected total observation cost $\mathbb{E}[c((\tau^*+T_0)\wedge\Theta)]$ and expected terminal reward $\mathbb{E}[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu^*=j,M=j\}}]$ break even under any optimal decision rule* (τ^*,μ^*). Intuitively, it also makes sense that the unit sampling cost that strikes an optimal balance between speed and accuracy in the above sense should be the maximum expected reward that can be gained per unit time.

Unlike the standard Bayes-risk minimization problem in which the unit sampling cost is a fixed known constant and the minimum Bayes risk is sought, in the Bayes-risk minimization problem dictated by the reward-rate maximization problem the minimum Bayes risk is known and the unknown unit sampling cost is sought. In other words, solving the reward-rate maximization problem is equivalent to solving

an *inverse* Bayes-risk minimization problem. The unit sampling cost in the inverse Bayes-risk minimization problem determines the optimal trade-off between speed and accuracy if and only if it coincides with the maximum reward rate of the reward-rate maximization problem.

In section 2, we characterize the Bayes-risk minimization solution to the multihypothesis sequential identification problems $W(c)$, $c > 0$ under a stochastic deadline. This treatment extends our previous work on Bayes-risk minimization in sequential testing of multiple hypotheses [7] and of binary hypotheses under a stochastic deadline [13], in which there are penalties associated with breaching a stochastic deadline in addition to typical observation and misidentification costs. In section 3, we characterize the formal relationship between reward-rate maximization and Bayes-risk minimization, and leverage it to obtain a numerical procedure for optimizing reward rate. Significantly, we will show that the optimal policy for reward-rate maximization depends on the initial belief state, unlike for Bayes-risk minimization—this is because the former identifies with a different setting of the latter depending on the initial state. This dependence on initial belief state shows explicitly that the reward-rate maximizing policy cannot satisfy any iterative, Markovian form of Bellman's dynamic programming equation [1]. Finally, in section 4, we demonstrate how the procedure can be applied to solve a numerical example involving binary hypotheses.

2. Multihypothesis sequential testing: Bayes-risk minimization. In the Bayes-risk minimization, the objective is to minimize a linear combination of sampling (observation or time) cost and response errors. In our problem, the response errors are of two types, misidentification and exceeding the deadline. In the following, we characterize properties of the Bayes-risk minimization problem:

- it reduces to an optimal stopping problem (section 2.1);
- value iteration yields successive approximations that converge to the optimal solution exponentially fast (section 2.2);
- the optimal stopping region, before the deadline, is a union of m convex regions containing the m respective cases of perfect identification certainty (section 2.3); the associated optimal policy is stationary and a random-walk process with absorbing boundaries

2.1. Bayes-risk minimization as optimal stopping. Assume we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common but unknown probability density function $f(\cdot)$. We know that $f(\cdot)$ is one of m known densities $f_1(\cdot), \ldots, f_m(\cdot)$, and the index M of the true density function is a random variable with the discrete prior probability distribution $\pi = (\pi_1, \ldots, \pi_m)$, where

$$\pi_j = \mathbb{P}\{M = j\}, \quad j = 1, \ldots m.$$

The problem is to identify the unknown density $f(\cdot)$ before a random deadline Θ, which is unknown but observable and has geometric distribution

$$\mathbb{P}\{\Theta = n\} = (1-p)^{n-1}p, \quad n = 1, 2 \ldots$$

for some known constant $0 < p < 1$ independent of $X_1, X_2, \ldots$. In addition, we assume that the observer's choice is registered $T_0 > 0$ units of "nondecision time" after the decision is made, so that the deadline may occur during that extra time interval even if

it had not appeared before the decision time. In a real application, this may represent motor delay or any other nontrivial delay in registering the choice after the decision has been made.

Let us denote any decision rule by a pair $\delta = (\tau, \mu)$ consisting of a stopping time τ of observation filtration

$$\begin{aligned}&\mathcal{F}_0 = \{\varnothing, \Omega\},\\ &\mathcal{F}_n = \sigma\{X_1 1_{\{\Theta \geq 1\}}, X_2 1_{\{\Theta \geq 2\}}, \ldots, X_n 1_{\{\Theta \geq n\}}, \Theta 1_{\{\Theta \leq n\}}, 1_{\{\Theta > n\}}\},\ n \geq 0,\end{aligned}$$

and $\{1, \ldots, m\}$-valued $\mathcal{F}_\tau$-measurable random variable μ that indicates the terminal choice. Observe that Θ is a stopping time of $(\mathcal{F}_n)_{n\geq 0}$. Let us also define the $(\mathcal{F}_n)_{n\geq 0}$-adapted process

$$S_n = 1_{\{\Theta \leq n\}}, \quad n \geq 0,$$

indicating whether the deadline Θ has already been observed. Suppose that initially $S_0 = s \in \{0, 1\}$.

For each $(\pi, s) \in \mathcal{S}_{m-1} \times \{0, 1\}$, $\mathcal{S}_{m-1} = \{(\pi_1, \ldots, \pi_m);\ \pi_j \geq 0, 1 \leq j \leq m, \text{and } \pi_1 + \cdots + \pi_m = 1\}$ being the $(m$–$1)$-dimensional simplex, we define $R_{\tau,\mu}(\pi, s) \equiv R_{\tau,\mu}(\pi, s; c, T_0)$ as the expected total cost associated with admissible rule (τ, μ),

$$\begin{aligned}(1)\quad R_{\tau,\mu}(\pi, s) := \mathbb{E}_{\pi,s}\Bigg[c\big((\tau + T_0) \wedge \Theta\big) + \sum_{j=1}^{m} \sum_{i:i\neq j} c_{ij} 1_{\{\tau + T_0 < \Theta,\, \mu = i, M = j\}}\\ + \sum_{j=1}^{m} d_j 1_{\{\tau + T_0 \geq \Theta, M = j\}}\Bigg],\end{aligned}$$

where c is the observation cost, c_{ij} is the cost of misidentification of j with i for every $1 \leq i \neq j \leq m$, and d_j is the cost of missing the deadline when $f_j(\cdot)$ is the true common probability density function for every $1 \leq j \leq m$. If the deadline has not yet passed (i.e., $\Theta > 0$), then we say $s = 0$; otherwise (i.e., $\Theta \leq 0$), we have $s = 1$.

Consider now the Bayes-risk minimization problem

$$(2)\qquad W(\pi, s) \equiv W(\pi, s; c, T_0) := \inf_{(\tau,\mu)} R_{\tau,\mu}(\pi, s; c, T_0), \quad (\pi, s) \in \mathcal{S}_{m-1} \times \{0, 1\}\ .$$

We first write down the Bayesian belief update equations and then show that it is a Markov process. Let $\Pi_n^{(j)} := \mathbb{P}\{M = j \mid \mathcal{F}_n\}$, $1 \leq j \leq m$, and recall that $S_n = 1_{\{\Theta \leq n\}}$ for every $n \geq 0$. Then the posterior distribution is

$$\Pi_{n+1}^{(j)} = S_{n+1}\Pi_n^{(j)} + (1 - S_{n+1})\frac{\Pi_n^{(j)} f_j(X_{n+1})}{\sum_{k=1}^{m} \Pi_n^{(k)} f_k(X_{n+1})}\ , \quad 1 \leq j \leq m,\ n \geq 0,$$

and the predictive distribution is

$$\mathbb{P}\{X_{n+1} \in \mathrm{d}x, S_{n+1} = 0 \mid \mathcal{F}_n\} = (1 - S_n)(1 - p)\sum_{j=1}^{m} \Pi_n^{(j)} f_j(x)\mathrm{d}x, \quad n \geq 0\ .$$

The sequence $(\Pi_n, S_n)_{n\geq 1}^{\infty}$ is a Markov process, because for every $n \geq 0$ we have

$$\begin{aligned}
\Pi_{n+1} &= S_{n+1} + (1 - S_{n+1})D(\Pi_n, X_{n+1}), \quad \text{where} \\
D(\pi, x) &= \left(\frac{\pi_1 f_1(x)}{\sum_{j=1}^m \pi_j f_j(x)}, \ldots, \frac{\pi_m f_m(x)}{\sum_{j=1}^m \pi_j f_j(x)} \right), \\
\mathbb{P}\{S_{n+1} = 1 \mid \mathcal{F}_n\} &= 1 - (1 - S_n)(1 - p) = p + S_n - pS_n,
\end{aligned}$$

which imply for every $n \geq 0$ and bounded function $f : \mathcal{S}_{m-1} \times \{0, 1\} \mapsto \mathbb{R}$, that

$$\begin{aligned}
\mathbb{E}[f(\Pi_{n+1}, S_{n+1}) \mid \mathcal{F}_n] &= \mathbb{E}\big[S_{n+1} f(\Pi_n, 1) + (1 - S_{n+1}) f\big(D(\Pi_n, X_{n+1})\big) \big| \mathcal{F}_n\big] \\
&= (p + S_n - pS_n) f(\Pi_n, 1) \\
&\quad + (1 - S_n)(1 - p) \int f\big(D(\Pi_n, x), 0\big) \sum_{j=1}^m \Pi_n^{(j)} f_j(x) \mathrm{d}x \ ,
\end{aligned}$$

which is (Π_n, S_n)-measurable.

Following Shiryaev [16, p. 167], we first reduce the Bayes-risk minimization problem to a pure optimal stopping problem of a suitable Markov process. Shiryaev showed that the posterior probability process $(\Pi_n)_{n=0}^{\infty}$ is a sufficient Markov statistic for the classical Bayes-risk minimization problem. In our new Bayes-risk minimization problem motivated by the setup of the neuroscience experiments, however, both running and terminal costs account for the extra cost incurred during the registration of terminal decision T_0 time units after stopping and depend in the first place on whether the decision is successfully registered before the random deadline. Therefore, the costs are more complex, and the sufficient Markov process now becomes the pair $(\Pi_n, S_n)_{n=0}^{\infty}$, consisting of posterior probability and survival processes, which together may be thought of as the *killed* posterior probability process. Proposition 1 describes precisely the new equivalent optimal stopping problem by carefully taking care of the technical differences between old and new formulations of Bayes-risk minimization problems.

PROPOSITION 1. *The original problem in* (2) *can be reduced to an optimal stopping problem*

$$W(\pi, s) = \inf_{\tau} R_{\tau, \mu(\tau)} = \inf_{\tau} \mathbb{E}_{\pi, s}\left[\sum_{k=0}^{\tau - 1} c(1 - S_k) + h(\Pi_\tau, S_\tau) \right] \tag{3}$$

of the Markov process $(\Pi_n, S_n)_{n=0}^{\infty}$, *where* $\mu(\tau)$ *is the optimal terminal decision rule for any stopping time* τ*:*

$$\mu(n) := \arg\min_{1 \leq i \leq m} \sum_{i=1}^m c_{ij} \Pi_n^{(j)} \quad \text{for every } n = 0, 1, \ldots, \tag{4}$$

$\sum_{k=0}^{\tau-1} c(1 - S_k)$ *is the observation cost, and* $h(\pi, s) \equiv h(\pi, s; c, T_0)$ *is the terminal decision cost function incorporating both misidentifications and the deadline; for each* $(\pi, s) \in \mathcal{S}_{m-1} \times \{0, 1\}$

$$\begin{aligned}
h(\pi, s) = (1 - p)^{T_0}(1 - s) \min_{1 \leq i \leq m} \Bigg[& \sum_{j : j \neq i} c_{ij} \pi_j + \big((1 - (1 - p)^{T_0})(1 - s) + s\big) \sum_{j=1}^m d_j \pi_j \\
& + \frac{c}{p}(1 - (1 - p)^{T_0})(1 - s) \Bigg].
\end{aligned}$$

Proof. We derive expressions for each of the three terms on the right-hand side of (1).

(a) We first note

$$(\tau+T_0)\wedge\Theta=\sum_{k=0}^{\infty}1_{\{(\tau+T_0)\wedge\Theta>k\}}=\sum_{k=0}^{\infty}1_{\{\tau+T_0>k\}}1_{\{\Theta>k\}}=\sum_{k=0}^{\tau+T_0-1}1_{\{\Theta>k\}}$$
$$=\sum_{k=0}^{\tau-1}(1-S_k)+\sum_{k=\tau}^{\tau+T_0-1}(1-S_k)=\sum_{k=0}^{\tau-1}(1-S_k)+\sum_{k=0}^{T_0-1}(1-S_{\tau+k})\ .$$

Because $\mathbb{E}[1-S_{\tau+k}]=\mathbb{E}[\mathbb{E}(1-S_{\tau+k}\mid\mathcal{F}_\tau)]=\mathbb{E}[(1-S_\tau)\,\mathbb{P}\{S_{\tau+k}=0\mid\mathcal{F}_\tau\}]=\mathbb{E}[(1-S_\tau)\,\mathbb{P}\{S_{\tau+k}=0\mid\tau,S_\tau=0\}]=\mathbb{E}[(1-S_\tau)(1-p)^k]$ for every $k\geq 0$, the expected decision delay is

$$\mathbb{E}[(\tau+T_0)\wedge\Theta]=\mathbb{E}\Big[\sum_{k=0}^{\tau-1}(1-S_k)\Big]+\sum_{k=0}^{T_0-1}\mathbb{E}(1-S_{\tau+k})=\mathbb{E}\Big[\sum_{k=0}^{\tau-1}(1-S_k)\Big]$$
$$+\mathbb{E}\Big[(1-S_\tau)\sum_{k=0}^{T_0-1}(1-p)^k\Big]$$
$$=\mathbb{E}\Big[\sum_{k=0}^{\tau-1}(1-S_k)\Big]+\frac{1-(1-p)^{T_0}}{p}\mathbb{E}(1-S_\tau).$$

(b) The misidentification probability is

$$\begin{aligned}
&\mathbb{E}[1_{\{\tau+T_0<\Theta\,\mu=i,M=j\}}]\\
&\quad=\mathbb{P}\{\tau+T_0<\Theta,\mu=i,M=j\}\\
&\quad=\sum_{n=0}^{\infty}\mathbb{E}\big[1_{\{\tau=n,\mu=i\}}\mathbb{P}\{n+T_0<\Theta,M=j\mid\mathcal{F}_n\}\big]\\
&\quad=\sum_{n=0}^{\infty}\mathbb{E}\big[1_{\{\tau=n,\mu=i\}}(1-S_n)\mathbb{P}\{S_{n+T_0}=0,M=j\mid\mathcal{F}_n\}\big]\\
&\quad=\sum_{n=0}^{\infty}\mathbb{E}\big[1_{\{\tau=n,\mu=i\}}(1-S_n)\mathbb{P}\{S_{n+T_0}=0\mid S_n=0\}\mathbb{P}\{M=j\mid X_1,\dots,X_n\}\big]\\
&\quad=\sum_{n=0}^{\infty}\mathbb{E}\big[1_{\{\tau=n,\mu=i\}}(1-S_n)(1-p)^{T_0}\Pi_n^{(j)}\big]\\
&\quad=(1-p)^{T_0}\mathbb{E}\big[1_{\{\tau<\infty,\mu=i\}}(1-S_\tau)\Pi_\tau^{(j)}\big]\\
&\quad=(1-p)^{T_0}\mathbb{E}\big[1_{\{\mu=i\}}(1-S_\tau)\Pi_\tau^{(j)}\big]\quad\text{for every } 1\leq i,j\leq m,
\end{aligned}$$

since $S_\infty=\lim_{n\to\infty}S_n=1$ a.s. and $(1-S_\tau)\Pi_\tau=(1-S_\infty)\Pi_\infty=0\cdot\Pi_\Theta=0$ a.s. on $\{\tau=\infty\}$. This is because $S_\Theta=1$ a.s., and $\Pi_\Theta=S_\Theta\Pi_{\Theta-1}+(1-S_\Theta)D(\Pi_{\Theta-1},X_\Theta)=\Pi_{\Theta-1}$. Thus $\Pi_{\Theta-1}=\Pi_\Theta=\cdots$ a.s.; consequently, $\Pi_\infty:=\lim_{n\to\infty}\Pi_n=\Pi_\Theta$ and $\Pi_n 1_{\{n\geq\Theta\}}=\Pi_\Theta 1_{\{n\geq\Theta\}}$ a.s. for every $n\geq 0$.

(c) The probability of breaching the deadline is

$$\mathbb{P}\{\tau+T_0\geq\Theta,M=j\}=\mathbb{P}\{\tau<\Theta,\tau+T_0\geq\Theta,M=j\}+\mathbb{P}\{\tau\geq\Theta,M=j\}$$
$$=\mathbb{E}\Big[\big((1-(1-p)^{T_0})(1-S_\tau)+S_\tau\big)\Pi_\tau^{(j)}\Big],$$

because $\tau \wedge \Theta$ is an $(\mathcal{F}_n)_{n\geq 0}$ stopping time and $\mathcal{F}_\Theta \equiv \mathcal{F}_\tau$ on $\{\tau \geq \Theta\}$ imply

$$\begin{aligned}\mathbb{P}\{\tau \geq \Theta, M = j\} &= \mathbb{E}[1_{\{\tau\geq\Theta\}}\mathbb{P}\{M = j \mid \mathcal{F}_{\tau\wedge\Theta}\}] = \mathbb{E}[1_{\{\tau\geq\Theta\}}\mathbb{P}\{M = j \mid \mathcal{F}_\Theta\}] \\ &= \mathbb{E}[1_{\{\tau\geq\Theta\}}\mathbb{P}\{M = j \mid \mathcal{F}_\tau\}] = \mathbb{E}[1_{\{\tau\geq\Theta\}}\Pi_\tau^{(j)}] = \mathbb{E}[S_\tau\Pi_\tau^{(j)}],\end{aligned}$$

and $(1 - S_\tau)\Pi_\tau = 0$ a.s. on $\{\tau = \infty\}$ implies

$$\begin{aligned}&\mathbb{P}\{\tau < \Theta, \tau + T_0 \geq \Theta, M = j\} \\ &= \sum_{n=0}^{\infty} \mathbb{E}[1_{\{\tau=n\}}\mathbb{P}\{n < \Theta \leq n + T_0, M = j \mid \mathcal{F}_n\}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[1_{\{\tau=n\}}(1 - S_n)\mathbb{P}\{n < \Theta \leq n + T_0 \mid \Theta > n\}\mathbb{P}\{M = j \mid X_1, \ldots, X_n\}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[1_{\{\tau=n\}}(1 - S_n)(1 - (1 - p)^{T_0})\Pi_n^{(j)}] = (1 - (1 - p)^{T_0})\,\mathbb{E}[1_{\{\tau<\infty\}}(1 - S_\tau)\Pi_\tau^{(j)}] \\ &= (1 - (1 - p)^{T_0})\,\mathbb{E}[(1 - S_\tau)\Pi_\tau^{(j)}].\end{aligned}$$

Combining (a), (b), and (c), we can now rewrite $R_{\tau,\mu}(\pi, s)$ of (1) as follows:

$$\begin{aligned}&R_{\tau,\mu}(\pi, s) \\ &= \mathbb{E}_{\pi,s}\Big[c\big((\tau+T_0)\wedge\Theta\big) + \sum_{j=1}^{m}\sum_{i:i\neq j} c_{ij}1_{\{\tau+T_0<\Theta\,\mu=i,M=j\}} + \sum_{j=1}^{m} d_j 1_{\{\tau+T_0\geq\Theta,M=j\}}\Big] \\ &= \mathbb{E}_{\pi,s}\Big[\sum_{k=0}^{\tau-1} c(1 - S_k)\Big] + \frac{c}{p}\big(1 - (1 - p)^{T_0}\big)\mathbb{E}_{\pi,s}(1 - S_\tau) \\ &\quad + (1 - p)^{T_0}\sum_{j=1}^{m}\sum_{i:i\neq j} c_{ij}\mathbb{E}_{\pi,s}\big[1_{\{\mu=i\}}(1 - S_\tau)\Pi_\tau^{(j)}\big] \\ &\quad + \sum_{j=1}^{m} d_j\mathbb{E}_{\pi,s}\Big[\big((1 - (1 - p)^{T_0})(1 - S_\tau) + S_\tau\big)\Pi_\tau^{(j)}\Big] \\ &= \mathbb{E}_{\pi,s}\Big[\sum_{k=0}^{\tau-1} c(1 - S_k) + (1 - p)^{T_0}(1 - S_\tau)\sum_{i=1}^{m} 1_{\{\mu=i\}}\sum_{j:j\neq i} c_{ij}\Pi_\tau^{(j)} \\ &\quad + \big((1 - (1 - p)^{T_0})(1 - S_\tau) + S_\tau\big)\sum_{j=1}^{m} d_j\Pi_\tau^{(j)} + \frac{c}{p}\big(1 - (1 - p)^{T_0}\big)(1 - S_\tau)\Big] \\ &\geq \mathbb{E}_{\pi,s}\Big[\sum_{k=0}^{\tau-1} c(1 - S_k) + (1 - p)^{T_0}(1 - S_\tau)\min_{1\leq i\leq m}\sum_{j:j\neq i} c_{ij}\Pi_\tau^{(j)} \\ &\quad + \big((1 - (1 - p)^{T_0})(1 - S_\tau) + S_\tau\big)\sum_{j=1}^{m} d_j\Pi_\tau^{(j)} + \frac{c}{p}\big(1 - (1 - p)^{T_0}\big)(1 - S_\tau)\Big].\end{aligned}$$

Combined with (2), this proves (3). □

Remark 2. For every admissible rule (τ, μ), the rule $(\tau\wedge\Theta, \mu(\tau\wedge\Theta))$ is admissible and has expected total cost less than or equal to that of (τ, μ) because

$$S_{\tau\wedge\Theta} = S_\tau,\ \Pi_{\tau\wedge\Theta} = \Pi_\tau, \quad \text{and} \quad \sum_{k=0}^{\tau\wedge\Theta-1} c(1 - S_k) = \sum_{k=0}^{\tau-1} c(1 - S_k) \tag{5}$$

imply that

$$\begin{aligned}
R_{\tau,\mu} &\geq R_{\tau,\mu(\tau)} \\
&= \mathbb{E}\Bigg[\sum_{k=0}^{\tau-1} c(1-S_k) + (1-p)^{T_0}(1-S_\tau) \min_{1\leq i\leq m} \sum_{j:j\neq i} c_{ij}\Pi_\tau^{(j)} \\
&\qquad + \big((1-(1-p)^{T_0})(1-S_\tau) + S_\tau\big) \sum_{j=1}^{m} d_j \Pi_\tau^{(j)} + \frac{c}{p}(1-(1-p)^{T_0})(1-S_\tau)\Bigg] \\
&= \mathbb{E}\Bigg[\sum_{k=0}^{\tau\wedge\Theta-1} c(1-S_k) + (1-p)^{T_0}(1-S_{\tau\wedge\Theta}) \min_{1\leq i\leq m} \sum_{j:j\neq i} c_{ij}\Pi_{\tau\wedge\Theta}^{(j)} \\
&\qquad + \big((1-(1-p)^{T_0}) \times (1-S_{\tau\wedge\Theta}) + S_{\tau\wedge\Theta}\big) \sum_{j=1}^{m} d_j \Pi_{\tau\wedge\Theta}^{(j)} \\
&\qquad + \frac{c}{p}(1-(1-p)^{T_0})(1-S_{\tau\wedge\Theta})\Bigg] = R_{\tau\wedge\Theta,\mu(\tau\wedge\Theta)}.
\end{aligned}$$

Finally, the identities in (5) follow from

$$\begin{aligned}
S_{\tau\wedge\Theta} = 0 \iff \Theta > \tau\wedge\Theta \iff \Theta > \tau \iff S_\tau = 0, \\
\Pi_{\tau\wedge\Theta} = \Pi_\tau 1_{\{\tau<\Theta\}} + \Pi_\Theta 1_{\{\tau\geq\Theta\}} = \Pi_\tau 1_{\{\tau<\Theta\}} + \Pi_\tau 1_{\{\tau\geq\Theta\}} = \Pi_\tau, \\
\sum_{k=0}^{\tau-1} c(1-S_k) = \sum_{k=0}^{\tau\wedge\Theta-1} c(1-S_k) + 1_{\{\tau>\Theta\}} \cancel{\sum_{k=\Theta}^{\tau-1} c(1-S_k)} = \sum_{k=0}^{\tau\wedge\Theta-1} c(1-S_k),
\end{aligned}$$

because $S_k = 1$ for every $k \geq \Theta$ a.s.

2.2. Successive approximation of value function. The dynamic programming principle implies that

$$W(\pi,s) = \min\Big\{h(\pi,s), c(1-s) + \mathbb{E}[W(\Pi_1,S_1) \mid (\Pi_0,S_0) = (\pi,s)]\Big\}, \tag{6}$$

where the expectation $\mathbb{E}[W(\Pi_1,S_1) \mid (\Pi_0,S_0) = (\pi,s)]$ becomes

$$sW(\pi,s) + (1-s)\mathbb{E}\Big[W\big(S_1\Pi_0 + (1-S_1)D(\Pi_0,X_1),0\big)\Big|(\Pi_0,S_0) = (\pi,s)\Big].$$

More precisely, we have $\mathbb{E}[W(\Pi_1,S_1) \mid (\Pi_0,S_0) = (\pi,1)] = W(\pi,1)$ and

$$\begin{aligned}
\mathbb{E}[W(\Pi_1,S_1) \mid (\Pi_0,S_0) &= (\pi,0)] \\
&= pW(\pi,1) + (1-p)\,\mathbb{E}[W(D(\Pi_0,X_1),0) \mid (\Pi_0,S_0) = (\pi,0)] \\
&= pW(\pi,1) + (1-p)\int W(D(\pi,x),0) \sum_{j=1}^{m} \pi_j f_j(x)\mathrm{d}x.
\end{aligned}$$

On the collection of bounded functions $w : \mathcal{S}_{m-1} \times \{0,1\} \mapsto \mathbb{R}$, let us define operators

$$(Tw)(\pi,s) = s\,w(\pi,1) + (1-s)\Big[p\,w(\pi,1) + (1-p)\int w(D(\pi,x),0) \sum_{j=1}^{m} \pi_j f_j(x)\mathrm{d}x\Big], \tag{7}$$

$$(Mw)(\pi,s) = \min\{h(\pi,s), c(1-s) + (Tw)(\pi,s)\}.$$

The value function $W(\pi, s)$ is a fixed point of operator M. If $S_0 \equiv s = 1$ in (3), then $S_0 = S_1 = \cdots = 1$ and

$$(8)\quad W(\pi,1) = \inf_{\tau} \mathbb{E}_{\pi,1}\left[\sum_{j=1}^{m} d_j \Pi_\tau^{(j)}\right] = \inf_{\tau} \sum_{j=1}^{m} d_j \pi_j = \sum_{j=1}^{m} d_j \pi_j \quad \text{for every } \pi \in \mathcal{S}_{m-1},$$

because $\Pi_n^{(j)} = \mathbb{P}\{M = j \mid \mathcal{F}_n\}$, $n \geq 0$ is a bounded martingale. Therefore, it is uniformly integrable, and the optional sampling theorem implies that $\mathbb{E}_{\pi,1}\Pi_\tau^{(j)} = \Pi_0^{(j)} = \pi_j$ for every $(\mathcal{F}_n)_{n\geq 0}$ stopping time τ.

The optimality equation in (6) turns out to have a unique solution, which can be found as the pointwise limit of successive approximations; see, for example, Shiryaev [16, pp. 168–169] for similar results for the classical Bayesian binary hypothesis testing problem. Here we follow the general theory of stochastic dynamic programming as, for example, described by Bertsekas and Shreve [2, Chapter 4], and show that the dynamic programming operator M in (7) is a contraction by Proposition 3 and that the value function $W(\cdot)$ is its unique fixed point by Corollary 4. The successive approximations of the fixed point of a contraction therefore lead naturally to the successive approximations of the value function as described by Proposition 5 and Corollary 6. Here, the optimal stopping problem is not a discounted optimal control problem with bounded costs and the contraction property of the dynamic programming operator is not automatic. We establish this property by taking advantage of the exponential decay in the excess life distribution of the random deadline.

PROPOSITION 3. *The operator M is a contraction mapping on the collection of bounded functions $w : \mathcal{S}_{m-1} \times \{0,1\} \mapsto \mathbb{R}$ with $w(\pi,1) = h(\pi,1) = \sum_{j=1}^{m} d_j\pi_j$ for every $\pi \in \mathcal{S}_{m-1}$.*

Proof. Let $w_1, w_2 : \mathcal{S}_{m-1} \times \{0,1\} \mapsto \mathbb{R}$ be two bounded functions such that $w_i(\pi,1) = h(\pi,1)$ for every $\pi \in \mathcal{S}_{m-1}$ and $i = 1,2$. Then $|(Mw_1)(\pi,s) - (Mw_2)(\pi,s)|$ equals

$$\begin{aligned}
&|\min\{h(\pi,s), c(1-s) + (Tw_1)(\pi,s)\} - \min\{h(\pi,s), c(1-s) + (Tw_2)(\pi,s)\}| \\
&\quad \leq |(c(1-s) + (Tw_1)(\pi,s)) - (c(1-s) + (Tw_2)(\pi,s))| \\
&\quad \leq \Big|\cancel{w_1(\pi,1)} + (1-s)\Big[\cancel{p\, w_1(\pi,1)} + (1-p)\int w_1(D(\pi,x),0)\sum_{j=1}^{m}\pi_j f_j(x)\mathrm{d}x\Big] \\
&\qquad - \Big(\cancel{w_2(\pi,1)} + (1-s)\Big[\cancel{p\, w_2(\pi,1)} + (1-p)\int w_2(D(\pi,x),0)\sum_{j=1}^{m}\pi_j f_j(x)\mathrm{d}x\Big]\Big)\Big| \\
&\quad = \Big|(1-s)(1-p)\int (w_1 - w_2)(D(\pi,x),0)\sum_{j=1}^{m}\pi_j f_j(x)\mathrm{d}x\Big| \\
&\quad \leq (1-p)\sup_{\pi\in\mathcal{S}_{m-1}} |w_1(\pi,0) - w_2(\pi,0)| \leq (1-p)\|w_1 - w_2\|
\end{aligned}$$

for every $(\pi,s) \in \mathcal{S}_{m-1} \times \{0,1\}$. Therefore, $\|Mw_1 - Mw_2\| \leq (1-p)\|w_1 - w_2\|$. ∎

COROLLARY 4. *The value function $W(\cdot,\cdot)$ of (2) is the unique fixed point of operator M in the class of bounded functions $w : \mathcal{S}_{m-1} \times \{0,1\} \mapsto \mathbb{R}$ such that $w(\pi,1) = h(\pi,1)$ for every $\pi \in \mathcal{S}_{m-1}$.*

Proof. If $V : \mathcal{S}_{m-1} \times \{0,1\} \mapsto \mathbb{R}$ is another fixed point of M such that $V(\pi,1) = h(\pi,1)$ for every $\pi \in \mathcal{S}_{m-1}$, then by Proposition 3 we have $\|V - W\| = \|MV - MW\| \leq (1-p)\|V - W\|$, which holds if and only if $\|V - W\| = 0$. ∎

To numerically calculate $W(\cdot,\cdot)$, let us introduce the successive approximations

$$\begin{aligned} w_0(\pi,s) &= h(\pi,s) = sh(\pi,1)+(1-s)h(\pi,0), && (\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\},\\ w_{n+1}(\pi,s) &= (Mw_n)(\pi,s), && (\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}. \end{aligned} \tag{9}$$

We can show by induction on $n\geq 0$ that

$$w_n(\pi,1) = h(\pi,1) \quad \text{for every } \pi\in\mathcal{S}_{m-1}. \tag{10}$$

By definition, $w_0(\pi,1)=h(\pi,1)$ for every $\pi\in\mathcal{S}_{m-1}$. Suppose that for some $n\geq 0$ we have $w_n(\pi,1)=h(\pi,1)$ for every $\pi\in\mathcal{S}_{m-1}$. Then (7) implies that

$$\begin{aligned} w_{n+1}(\pi,1) &= (Mw_n)(\pi,1) = \min\{h(\pi,1),(Tw)(\pi,1)\} = \min\{h(\pi,1),w_n(\pi,1)\}\\ &= \min\{h(\pi,1),h(\pi,1)\} = h(\pi,1) \quad \text{for every } \pi\in\mathcal{S}_{m-1}. \end{aligned}$$

Using (10) we can write

$$\begin{aligned} w_{n+1}(\pi,s) &= (Mw_n)(\pi,s) = sh(\pi,1)+(1-s)(Mw_n)(\pi,0)\\ &= sh(\pi,1)+(1-s)\min\Bigg\{h(\pi,0),c+ph(\pi,1)\\ &\qquad +(1-p)\int w_n(D(\pi,x),0)\sum_{j=1}^{m}\pi_j f_j(x)\mathrm{d}x\Bigg\}. \end{aligned} \tag{11}$$

PROPOSITION 5. *For every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$, the sequence $(w_n(\pi,s))_{n\geq 0}$ is decreasing and $w_\infty(\pi,s):=\lim_{n\to\infty}w_n(\pi,s)$ exists.*

Proof. From (11), we notice that $0\leq w_1(\pi,s)\leq sh(\pi,1)+(1-s)h(\pi,0)=w_0(\pi,s)$ for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$. Suppose that $0\leq w_n(\pi,s)\leq w_{n-1}(\pi,s)$ for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$ for some $n\geq 1$. Then

$$\begin{aligned} 0\leq w_{n+1}(\pi,s) &= (Mw_n)(\pi,s) = \min\{h(\pi,s),c(1-s)+(Tw_n)(\pi,s)\}\\ &\leq \min\{h(\pi,s),c(1-s)+(Tw_{n-1})(\pi,s)\} = (Mw_{n-1})(\pi,s) = w_n(\pi,s) \end{aligned}$$

for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$. Therefore, $(w_n(\pi,s))_{n\geq 0}$ is decreasing and $w_\infty(\pi,s):=\lim_{n\to\infty}w_n(\pi,s)$ exists for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$. ∎

COROLLARY 6. *The value function W and the limit w_∞ of successive approximations coincide; namely, $W(\pi,s)=w_\infty(\pi,s)$ for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$. Moreover, $\|W-w_n\|\leq(1-p)^n\|h\|$ for every $n\geq 0$.*

Proof. Because $0\leq w_n\leq w_0$, taking the limit as $n\to\infty$ in (11) and the bounded convergence theorem imply that

$$\begin{aligned} w_\infty(\pi,s) &= sh(\pi,1)+(1-s)\min\Bigg\{h(\pi,0),c+ph(\pi,1)\\ &\qquad +(1-p)\int w_\infty(D(\pi,x),0)\sum_{j=1}^{m}\pi_j f_j(x)\mathrm{d}x\Bigg\}\\ &= (Mw_\infty)(\pi,s) \end{aligned}$$

for every $(\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}$. Therefore, w_∞ is a fixed point of operator M. Because $w_\infty(\pi,1)=\lim_{n\to\infty}w_n(\pi,1)=\lim_{n\to\infty}h(\pi,1)=h(\pi,1)$ for every $\pi\in\mathcal{S}_{m-1}$, Corollary 6 implies that $W(\cdot,\cdot)=w_\infty(\cdot,\cdot)$. Finally, $\|W-w_n\|=\|MW-Mw_{n-1}\|\leq(1-p)\|W-w_{n-1}\|\leq\cdots\leq(1-p)^n\|W-w_0\|\leq(1-p)^n\|w_0\|=(1-p)^n\|h\|$ for every $n\geq 0$. ∎

2.3. Structure of optimal policy. The optimal stopping region is

$$\Gamma(c,T_0) := \{(\pi,s) \in \mathcal{S}_{m-1} \times \{0,1\};\ W(\pi,s;c,T_0) = h(\pi,s;c,T_0)\}, \quad c>0,\ T_0 \geq 1,$$

and an optimal (stationary) decision rule is $(\tau(c,T_0), \mu(\tau(c,T_0)))$, where $\mu(\cdot)$ is defined by (4) and

$$\tau(c,T_0) := \inf\{n \geq 0;\ (\Pi_n, S_n) \in \Gamma(c,T_0)\} \quad \text{for every } c>0 \text{ and } T_0 \geq 1. \tag{12}$$

Because $h(\pi,s;c,T_0) = \min_{1\leq i\leq m} h_i(\pi,s;c,T_0)$ in terms of

$$\begin{aligned} h_i(\pi,s;c,T_0) = (1-s)\Bigg[(1-p)^{T_0} \sum_{j:j\neq i} c_{ij}\pi_j + \left(1-(1-p)^{T_0}\right)\left(\frac{c}{p} + \sum_{j=1}^m d_j\pi_j\right)\Bigg] \\ + s\sum_{j=1}^m d_j\pi_j, \\ (\pi,s) \in \mathcal{S}_{m-1}\times\{0,1\},\ 1\leq i\leq m, \end{aligned}$$

and $W(\pi,1;c,T_0) = h(\pi,1;c,T_0)$ for every $\pi \in \mathcal{S}_{m-1}$, we have

$$\begin{aligned} \Gamma(c,T_0) &= \Gamma_0(c,T_0) \cup \Gamma_1(c,T_0), \\ \Gamma_1(c,T_0) &= \{(\pi,1);\ \pi\in\mathcal{S}_{m-1}, W(\pi,1;c,T_0) = h(\pi,1;c,T_0)\} = \mathcal{S}_{m-1}\times\{1\}, \\ \Gamma_0(c,T_0) &= \{(\pi,0);\ \pi\in\mathcal{S}_{m-1}, W(\pi,0;c,T_0) = h(\pi,0;c,T_0)\} \\ &= \Gamma_0^{(1)}(c,T_0) \cup \cdots \cup \Gamma_0^{(m)}(c,T_0), \end{aligned}$$

where

$$\Gamma_0^{(i)}(c) = \{(\pi,0);\ \pi\in\mathcal{S}_{m-1}, W(\pi,0;c) = h_i(\pi,0)\}, \quad 1\leq i\leq m.$$

Next, we show that the stopping region, before the deadline, is the union of m convex regions containing the m respective cases of the perfect identification certainty. This result is similar to the findings of Shiryaev [16, p. 169] in the simple classical case of the Bayesian sequential binary hypothesis testing problem and those of Blackwell and Girshick [3, Theorem 9.4.3] for more general Bayesian sequential procedures. Here, the new and more complex form of the transition function T in (7) of the two-dimensional Markov sufficient statistic $(\Pi_n, S_n)_{n\geq 0}^{\infty}$ demands extra care. To establish the convexity of stopping regions by Proposition 7, we first show that the transition function is concave by means of the general convexity-preserving property of perspective functions; see, for example, Boyd and Vandenberghe [5, section 3.2.6].

PROPOSITION 7. *Let $e_1,\ldots,e_m$ be the unit vectors in $\mathbb{R}_m$. Then $e_i \in \Gamma_0^{(i)}(c,T_0)$ and $\Gamma_0^{(i)}(c,T_0)$ is convex for every $i=1,\ldots,m$.*

We first show that $\pi \mapsto W(\pi,0) \equiv W(\pi,0;c,T_0)$ is concave. Let us prove that

$$\begin{aligned} &\text{for every bounded function } w : \mathcal{S}_{m-1}\times\{0,1\} \mapsto \mathbb{R} \text{ such} \\ &\text{that } w(\pi,1) = h(\pi,1) \text{ for every } \pi\in\mathcal{S}_{m-1} \text{ and } \pi\mapsto w(\pi,0) \\ &\text{is concave, the mapping } \pi\mapsto (Mw)(\pi,0) \text{ is concave.} \end{aligned} \tag{13}$$

Recall that $(Mw)(\pi,0) = \min\{h(\pi,0), c + (Tw)(\pi,0)\}$. Because the minimum of two concave functions is concave and $\pi\mapsto h(\pi,0)$ is concave, it is sufficient to show that

$$\pi \mapsto (Tw)(\pi,0) = ph(\pi,1) + (1-p)\int w(D(\pi,x),0)\sum_{j=1}^m \pi_j f_j(x)\mathrm{d}x$$

is concave. Because $\pi \mapsto h(\pi,1) = \sum_{j=1}^m d_j \pi_j$ is concave, it suffices to show for every $x \in \mathbb{R}$

$$\pi \mapsto w\left(\left(\frac{\pi_1 f_1(x)}{\sum_{k=1}^m \pi_k f_k(x)}, \dots, \frac{\pi_m f_m(x)}{\sum_{k=1}^m \pi_k f_k(x)}\right), 0\right) \sum_{j=1}^m \pi_j f_j(x) \quad \text{is concave.} \tag{14}$$

Take any $a, b \in \mathcal{S}_{m-1}$, $0 < \alpha < 1$, and let $\beta = 1 - \alpha$. The concavity of $\pi \mapsto w(\pi, 0)$ implies

$$\begin{aligned}
&w\left(\left(\frac{(\alpha a_1 + \beta b_1) f_1(x)}{\sum_{k=1}^m (\alpha a_k + \beta b_k) f_k(x)}, \dots, \frac{(\alpha a_m + \beta b_m) f_m(x)}{\sum_{k=1}^m (\alpha a_k + \beta b_k) f_k(x)}\right), 0\right) \left(\sum_{j=1}^m (\alpha a_j + \beta b_j) f_j(x)\right) \\
&\quad = w\Bigg(\Bigg(\frac{\alpha \sum_{k=1}^m a_k f_k(x) \frac{a_1 f_1(x)}{\sum_{k=1}^m a_k f_k(x)} + \beta \sum_{k=1}^m b_k f_k(x) \frac{b_1 f_1(x)}{\sum_{k=1}^m b_k f_k(x)}}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x))}, \dots, \\
&\qquad \frac{\alpha \sum_{k=1}^m a_k f_k(x) \frac{a_m f_1(x)}{\sum_{k=1}^m a_k f_k(x)} + \beta \sum_{k=1}^m b_k f_k(x) \frac{b_m f_1(x)}{\sum_{k=1}^m b_k f_k(x)}}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x))}\Bigg), 0\Bigg) \\
&\qquad \times \left(\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x))\right) \\
&\quad = w\Bigg(\Bigg(\frac{\alpha \sum_{k=1}^m a_k f_k(x)}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x)} \left[\frac{a_1 f_1(x)}{\sum_{k=1}^m a_k f_k(x)}, \dots, \frac{a_m f_m(x)}{\sum_{k=1}^m a_k f_k(x)}\right] \\
&\qquad + \frac{\beta \sum_{k=1}^m b_k f_k(x)}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x)} \left[\frac{b_1 f_1(x)}{\sum_{k=1}^m b_k f_k(x)}, \dots, \frac{b_m f_m(x)}{\sum_{k=1}^m b_k f_k(x)}\right]\Bigg), 0\Bigg) \\
&\qquad \times \left(\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x))\right) \\
&\quad \geq \Bigg\{\frac{\alpha \sum_{k=1}^m a_k f_k(x)}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x)} w\left(\left[\frac{a_1 f_1(x)}{\sum_{k=1}^m a_k f_k(x)}, \dots, \frac{a_m f_m(x)}{\sum_{k=1}^m a_k f_k(x)}\right], 0\right) \\
&\qquad + \frac{\beta \sum_{k=1}^m b_k f_k(x)}{\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x)} w\left(\left[\frac{b_1 f_1(x)}{\sum_{k=1}^m b_k f_k(x)}, \dots, \frac{b_m f_m(x)}{\sum_{k=1}^m b_k f_k(x)}\right], 0\right)\Bigg\} \\
&\qquad \times \left(\alpha \sum_{k=1}^m a_k f_k(x) + \beta \sum_{k=1}^m b_k f_k(x))\right) \\
&\quad = \alpha\, w\left(\left(\frac{a_1 f_1(x)}{\sum_{k=1}^m a_k f_k(x)}, \dots, \frac{a_m f_m(x)}{\sum_{k=1}^m a_k f_k(x)}\right), 0\right) \sum_{k=1}^m a_k f_k(x) \\
&\qquad + \beta\, w\left(\left(\frac{b_1 f_1(x)}{\sum_{k=1}^m b_k f_k(x)}, \dots, \frac{b_m f_m(x)}{\sum_{k=1}^m b_k f_k(x)}\right), 0\right) \sum_{k=1}^m b_k f_k(x),
\end{aligned}$$

which implies (14) and completes the proof of (13). Recall now that $W(\pi, s) = \lim_{n\to\infty} w_n(\pi, s)$ is the pointwise limit of the successive approximations in (9). Because the mapping $w(\cdot,\cdot) = w_0(\cdot,\cdot) = h(\cdot,\cdot)$ satisfies the hypothesis of (13), an induc-

tion on n shows that every $w(\cdot,\cdot) = w_n(\cdot,\cdot)$ satisfies the hypothesis of (13). Therefore, $\pi \mapsto w_n(\pi, 0)$ is concave for every $n \geq 0$. Because the pointwise limit of a sequence of concave functions is concave, the mapping $\pi \mapsto W(\pi, 0) = \lim_{n\to\infty} w_n(\pi, 0)$ is also concave.

Proof of Proposition 7. Let us first prove that $e_i \in \Gamma_0^{(i)}(c, T_0)$ for every $i = 1, \ldots, m$. We will suppress c and T_0 and write $\Gamma_0^{(i)}$, $W(\pi, s)$, $h(\pi, s)$, $h_i(\pi, s)$ instead of $\Gamma_0^{(i)}(c, T_0)$, $W(\pi, s; c, T_0)$, $h(\pi, s; c, T_0)$, $h_i(\pi, s; c, T_0)$. Because for every $1 \leq i \leq m$

$$h_i(e_i, 0) = \big(1 - (1-p)^{T_0}\big)\left(\frac{c}{p} + d_i\right), \quad h(e_i, 1) = d_i, \quad h(e_i, s) = h_i(e_i, s) \text{ for } s = 0, 1,$$

$$W(e_i, 1) = h(e_i, 1), \quad D(e_i, x) = e_i, \text{ and } W(D(e_i, x), 0) = W(e_i, 0) \text{ for } x \in \mathbb{R},$$

we have

$$\begin{aligned}(TW)(e_i, 0) &= pW(e_i, 1) + (1-p)\int W(D(e_i, x), 0) f_i(x)\mathrm{d}x \\ &= ph(e_i, 1) + (1-p)\int W(e_i, 0) f_i(x)\mathrm{d}x = p\, d_i + (1-p)W(e_i, 0), \\ W(e_i, 0) &= \min\{h(e_i, 0),\ c + (TW)(e_i, 0)\} \\ &= \min\{h_i(e_i, 0),\ c + p\, d_i + (1-p)W(e_i, 0)\}.\end{aligned}$$

Let us assume on the contrary that $e_i \notin \Gamma_0^{(i)}$. Then

$$\big(1 - (1-p)^{T_0}\big)\left(\frac{c}{p} + d_i\right) = h_i(e_i, 0) > W(e_i, 0) = c + p\, d_i + (1-p)W(e_i, 0).$$

Because the last equality implies that $W(e_i, 0) = (c/p) + d_i$, the strict inequality gives $(1-(1-p)^{T_0})((c/p) + d_i) > W(e_i, 0) = (c/p) + d_i$, which contradicts $1 - (1-p)^{T_0} < 1$. Therefore, $e_i \in \Gamma_0^{(i)}$ for every $i = 1, \ldots, m$.

To show that $\Gamma_0^{(i)}$ is convex, let us take any two fixed points $a, b \in \Gamma_0^{(i)}$ and $0 < \alpha < 1$. Because $\pi \mapsto h_i(\pi, 0)$ is affine and $\pi \mapsto W(\pi, 0)$ is concave,

$$\begin{aligned}h_i(\alpha a + (1-\alpha)b, 0) &= \alpha h_i(a, 0) + (1-\alpha)h_i(b, 0) = \alpha W(a, 0) + (1-\alpha)W(b, 0) \\ &\leq W(\alpha a + (1-\alpha)b, 0) \leq h(\alpha a + (1-\alpha)b, 0) \\ &\leq h_i(\alpha a + (1-\alpha)b, 0)\end{aligned}$$

implies that $h_i(\alpha a + (1-\alpha)b, 0) = W(\alpha a + (1-\alpha)b, 0)$ and $\alpha a + (1-\alpha)b \in \Gamma_0^{(i)}$. Therefore, $\Gamma_0^{(i)}$ is convex for every $i = 1, \ldots, m$. □

3. Multihypothesis sequential testing: Reward rate maximization. In this section, we study the same deadlined sequential identification problem as in section 2, but optimize a different objective function, the average reward rate. We show that an optimal policy, which depends on the initial belief state, exists, and we describe a numerical procedure for solving it. We show the following in turn:

- the reward-rate maximizing policy is equivalent to the solution of a special case of the Bayes-risk minimization problem in (2), whose value function $W(\pi, s; c^*, T_0)$ we know but whose observation cost c^* is unknown; c^* turns out to be the maximal reward rate (section 3.1);

- the Bayes-risk value function is strictly increasing, concave, and continuous in the observation cost c, before the deadline arrives, implying c^* is the unique solution that yields $W(\pi,0;c^*,T_0)=\sum_{j=1}^m r_j\pi_j$ (section 3.2);
- a bisection procedure, in the c values explored, can solve the reward-rate problem exponentially fast (section 3.3).

3.1. Reward-rate maximization versus Bayes-risk minimization. Suppose we earn $r_j \geq 0$ on $\{M=j\}$, $0 \leq j \leq m$ for correctly identifying M, and receive no rewards otherwise. The experiment takes a random $T=T(\tau,\Theta)=(\tau+T_0)\wedge\Theta$ units of time, depending on whether it terminates with an identification decision or with the deadline. The reward received is $R=R(\tau,\mu,\Theta,M)=1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}$. By the strong law of large numbers, the long-run average reward per unit time, when the experiment is repeated *ad infinitum*, equals

$$\frac{\mathbb{E}R}{\mathbb{E}T}=\frac{\mathbb{E}\left[1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right]}{\mathbb{E}\left[(\tau+T_0)\wedge\Theta\right]} \quad \text{with probability one.}$$

Our goal is to find the maximum reward rate

(15)
$$V(\pi,s):=\sup_{(\tau,\mu)}\frac{\mathbb{E}_{\pi,s}\left[1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right]}{\mathbb{E}_{\pi,s}[(\tau+T_0)\wedge\Theta]},\quad (\pi,s)\in\mathcal{S}_{m-1}\times\{0,1\}\ .$$

We first note that $V(\pi,1)$ is undefined and uninteresting, because both the numerator and denominator in (15) evaluate to 0. In the remainder, we will work on how to characterize and calculate $V(\pi,0)$ and find an admissible decision rule (τ,μ) whenever the supremum in (15) is attained for $s=0$. Note also that the assumption of $T_0>0$ precludes the optimal policy from being the trivial one of choosing $\tau=0$ a.s., which makes the denominator in (15) evaluate to 0.

Our first key insight is that the reward-rate maximizing policy is equivalent to the solution of a special case of the Bayes-risk minimization problem in (2).

PROPOSITION 8. *For every $\pi\in\mathcal{S}_{m-1}$,*

$$\sum_{j=1}^m r_j\pi_j=\inf_{(\tau,\mu)}\mathbb{E}_{\pi,0}\Big[V(\pi,0)\big((\tau+T_0)\wedge\Theta\big)+1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^m\sum_{i:i\neq j}r_j 1_{\{\mu=i,M=j\}}+1_{\{\tau+T_0\geq\Theta\}}\sum_{j=1}^m r_j 1_{\{M=j\}}\Big],$$

which is the value function $W(\pi,0;V(\pi,0),T_0)$ of the Bayes-risk minimization problem in (2)*, whereby $c=V(\pi,0)$, $c_{ij}=r_j 1_{\{i\neq j\}}$, $d_j=r_j$, for every $1\leq i,j\leq m$, and any reaction time $T_0>0$.*

Proof. We prove the equality in two steps:

(a) $W(\pi,0;V(\pi,0),T_0)\geq\sum_{j=1}^m r_j\pi_j$;

(b) $W(\pi,0;V(\pi,0),T_0)\leq\sum_{j=1}^m r_j\pi_j$.

(a) Let us fix any $\pi \in \mathcal{S}_{m-1}$. For every admissible (τ, μ), we have

$$V(\pi,0) \geq \frac{\mathbb{E}_{\pi,0}\left[1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right]}{\mathbb{E}_{\pi,0}[(\tau+T_0)\wedge\Theta]},$$

$$\begin{aligned}
V(\pi,0)\,\mathbb{E}_{\pi,0}[(\tau+T_0)\wedge\Theta] &\geq \mathbb{E}_{\pi,0}\left[1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right] \\
&= \mathbb{E}_{\pi,0}\left[1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j \left(1_{\{M=j\}} - \sum_{i:i\neq j} 1_{\{\mu=i,M=j\}}\right)\right] \\
&= \mathbb{E}_{\pi,0}\left[\left(1-1_{\{\tau+T_0\geq\Theta\}}\right)\sum_{j=1}^m r_j 1_{\{M=j\}} - 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j \sum_{i:i\neq j} 1_{\{\mu=i,M=j\}}\right] \\
&= \sum_{j=1}^m r_j \pi_j - \mathbb{E}_{\pi,0}\Bigg[1_{\{\tau+T_0\geq\Theta\}} \sum_{j=1}^m r_j 1_{\{M=j\}} \\
&\qquad + 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m \sum_{i:i\neq j} r_j 1_{\{\mu=i,M=j\}}\Bigg],
\end{aligned}$$

which leads to

$$\begin{aligned}
W(\pi,0;V(\pi,0),T_0) = \inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\Bigg[&V(\pi,0)\big((\tau+T_0)\wedge\Theta\big) \\
&+ 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m \sum_{i:i\neq j} r_j 1_{\{\mu=i,M=j\}} \\
&+ 1_{\{\tau+T_0\geq\Theta\}} \sum_{j=1}^m r_j 1_{\{M=j\}}\Bigg] \geq \sum_{j=1}^m r_j \pi_j.
\end{aligned}$$

(b) Because

$$\mathbb{E}_{\pi,0}[T_0\wedge\Theta] = \mathbb{E}_{\pi,0}\Big[\sum_{k=0}^{T_0-1} 1_{\{\Theta>k\}}\Big] = \sum_{k=0}^{T_0-1}(1-p)^k = \frac{1-(1-p)^{T_0}}{p}, \tag{16}$$

it is clear from (15) that

$$0 \leq V(\pi,0) \leq \frac{\max_{1\leq j\leq m} r_j}{\mathbb{E}[T_0\wedge\Theta]} = \frac{p\,\max_{1\leq j\leq m} r_j}{1-(1-p)^{T_0}} < \infty.$$

Therefore, for every $\varepsilon > 0$ there exists some $(\tau^*,\mu^*) \equiv (\tau^*(\pi,\varepsilon),\mu^*(\pi,\varepsilon))$ such that

$$V(\pi,0) - \varepsilon \leq \frac{\mathbb{E}_{\pi,0}\left[1_{\{\tau^*+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{\mu^*=j,M=j\}}\right]}{\mathbb{E}_{\pi,0}[(\tau^*+T_0)\wedge\Theta]},$$

which can be rearranged as

$$
\begin{aligned}
&(V(\pi,0)-\varepsilon)\,\mathbb{E}_{\pi,0}[(\tau^*+T_0)\wedge\Theta]\\
&\quad\le \mathbb{E}_{\pi,0}\left[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m r_j 1_{\{\mu^*=j,M=j\}}\right]\\
&\quad= \mathbb{E}_{\pi,0}\left[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m r_j\left(1_{\{M=j\}}-\sum_{i:i\neq j}1_{\{\mu^*=i,M=j\}}\right)\right]\\
&\quad= \mathbb{E}_{\pi,0}\left[\left(1-1_{\{\tau^*+T_0\ge\Theta\}}\right)\sum_{j=1}^m r_j 1_{\{M=j\}}-1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m r_j\sum_{i:i\neq j}1_{\{\mu^*=i,M=j\}}\right]\\
&\quad= \sum_{j=1}^m r_j\pi_j-\mathbb{E}_{\pi,0}\left[1_{\{\tau^*+T_0\ge\Theta\}}\sum_{j=1}^m r_j 1_{\{M=j\}}+1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m\sum_{i:i\neq j} r_j 1_{\{\mu^*=i,M=j\}}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{j=1}^m r_j\pi_j &\ge \mathbb{E}_{\pi,0}\Big[(V(\pi,0)-\varepsilon)\big((\tau^*+T_0)\wedge\Theta\big)+1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m\sum_{i:i\neq j} r_j 1_{\{\mu^*=i,M=j\}}\\
&\qquad +1_{\{\tau^*+T_0\ge\Theta\}}\sum_{j=1}^m r_j 1_{\{M=j\}}\Big]\\
&\ge \mathbb{E}_{\pi,0}\Big[V(\pi,0)\big((\tau^*+T_0)\wedge\Theta\big)+1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^m\sum_{i:i\neq j} r_j 1_{\{\mu^*=i,M=j\}}\\
&\qquad +1_{\{\tau^*+T_0\ge\Theta\}}\sum_{j=1}^m r_j 1_{\{M=j\}}\Big]-\varepsilon\mathbb{E}_{\pi,0}\Theta\ge W(\pi,0;V(\pi,0),T_0)-\varepsilon\mathbb{E}_{\pi,0}\Theta,
\end{aligned}
$$

and letting $\varepsilon\downarrow 0$ gives $\sum_{j=1}^m r_j\pi_j\ge W(\pi,0;V(\pi,0),T_0)$. □

Proposition 8 tells us that we can compute the maximal reward rate $V(\pi,0)$ by solving an inverse case of the Bayes-risk minimization problem, whereby we know the minimal Bayes risk $W(\pi,0;V(\pi,0),T_0)$ and need to find the appropriate sampling cost $c^*:=V(\pi,0)$ associated with that minimal risk. Intuitively, it makes sense that the sampling cost, which determines the trade-off between speed and accuracy, should be the maximal expected reward that can be gained per unit time.

3.2. Uniqueness of c^*. Finding the appropriate $c^*=V(\pi,0)$ would be greatly facilitated if we knew c^* was the unique value of c that satisfies $W(\pi,0;c,T_0)=\sum_{j=1}^m r_j\pi_j$, and if $W(\pi,0;c,T_0)$ is continuous and monotonic in c. The following proposition gives us the desiderata.

PROPOSITION 9. *For every $\pi\in\mathcal{S}_{m-1}, T_0\ge 0$, the mapping $c\mapsto W(\pi,0;c,T_0):(0,\infty)\mapsto\mathbb{R}$ is increasing, concave, and continuous. Moreover,*

(17)
$$
c\,\frac{1-(1-p)^{T_0}}{p}\le W(\pi,0;c,T_0)\le c\,\frac{1-(1-p)^{T_0}}{p}+\sum_{j=1}^m r_j\pi_j-(1-p)^{T_0}\max_{1\le i\le m} r_i\pi_i,
$$

so that $W(\pi,0;c,T_0) > \sum_{j=1}^m r_j\pi_j$ *if* $c > u_0$, $W(\pi,0;c,T_0) < \sum_{j=1}^m r_j\pi_j$ *if* $0 < c < l_0$, *where*

$$l_0 := \frac{p(1-p)^{T_0}}{(1-(1-p)^{T_0})} \max_{1\le j\le m} r_j\pi_j \quad < \quad u_0 := \frac{p}{(1-(1-p)^{T_0})} \sum_{j=1}^m r_j\pi_j.$$

Taken together, there exists unique $c^* \ge 0$ *such that* $W(\pi,0;c^*,T_0) = \sum_{j=1}^m r_j\pi_j$. *Moreover,* $c^* \in [l_0,u_0]$ *and* $c^* = V(\pi,0)$ *in light of Proposition* 8.

Proof. Note that $W(\pi,0;c,T_0)$ is the infimum of a family of nondecreasing affine functions of c. Therefore, the mapping $c \mapsto W(\pi,0;c,T_0) : (0,\infty) \mapsto \mathbb{R}$ is nondecreasing and concave, and also continuous. Thus, $c \mapsto (T(W(\cdot,\cdot;c,T_0)))(\pi,0)$ is nondecreasing, and $c \mapsto c + (T(W(\cdot,\cdot;c,T_0)))(\pi,0)$ is strictly increasing. Moreover, for every $\pi \in \mathcal{S}_{m-1}$, we have

$$h(\pi,0;c,T_0) = (1-p)^{T_0} \min_{1\le i\le m} \sum_{j:j\ne i} r_j\pi_j + \big(1-(1-p)^{T_0}\big)\left(\frac{c}{p} + \sum_{j=1}^m r_j\pi_j\right), \tag{18}$$

implying that $c \mapsto h(\pi,0;c,T_0)$ is strictly increasing. Therefore, the minimum of strictly increasing functions,

$$c \mapsto W(\pi,0;c,T_0) = \min\{h(\pi,0;c,T_0), c + (T(W(\cdot,\cdot;c,T_0)))(\pi,0)\},$$

is also strictly increasing. The first inequality in (17) follows from (16) and

$$W(\pi,0;c,T_0) \ge \mathbb{E}_{\pi,0}[c(T_0 \wedge \Theta)] = c\,\frac{1-(1-p)^{T_0}}{p},$$

and the second inequality follows from $W(\pi,0;c,T_0) \le h(\pi,0;c,T_0)$ after rearranging the right-hand side of (18). ∎

Because $W(\pi,0;c,T_0) - \sum_{j=1}^m r_j\pi_j$ equals

$$\begin{aligned}
&\inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\Bigg[c\big((\tau+T_0)\wedge\Theta\big) + 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m \sum_{i:i\ne j} r_j 1_{\{\mu=i,M=j\}} \\
&\qquad\qquad -1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{M=j\}}\Bigg] \qquad (19)\\
&= \inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\left[c\big((\tau+T_0)\wedge\Theta\big) - 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j \left(1_{\{M=j\}} - \sum_{i:i\ne j} 1_{\{\mu=i,M=j\}}\right)\right] \\
&= \inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\left[c\big((\tau+T_0)\wedge\Theta\big) - 1_{\{\tau+T_0<\Theta\}} \sum_{j=1}^m r_j 1_{\{\mu=j,M=j\}}\right],
\end{aligned}$$

Proposition 9 implies that

$$c \gtreqless V(\pi,0) \text{ if and only if } \inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\left[c\big((\tau+T_0)\wedge\Theta\big) - 1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^{m} r_j 1_{\{\mu=j,M=j\}}\right] \gtreqless 0. \tag{20}$$

COROLLARY 10. *The maximum reward rate $V(\pi,0)$ is the unique unit sampling cost c in the Bayes-risk minimization problem*

$$W(\pi,0;c,T_0) = \inf_{(\tau,\mu)} \mathbb{E}_{\pi,0}\Bigg[c\big((\tau+T_0)\wedge\Theta\big) + 1_{\{\tau+T_0<\Theta\}}\sum_{j=1}^{m}\sum_{i:i\neq j} r_j 1_{\{\mu=i,M=j\}} + 1_{\{\tau+T_0\geq\Theta\}}\sum_{j=1}^{m} r_j 1_{\{M=j\}}\Bigg], \tag{21}$$

for which the expected total observation cost $\mathbb{E}_{\pi,0}[c\big((\tau^*+T_0)\wedge\Theta\big)]$ and expected terminal reward $\mathbb{E}_{\pi,0}[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^{m} r_j 1_{\{\mu^*=j,M=j\}}]$ break even under any optimal decision rule (τ^*,μ^*), which attains the infimum in (21) *or, equivalently, in* (20).

Finally, Proposition 11 below shows that the reward-rate maximization problem always admits an optimal decision rule. Note that, unlike the optimal decision rules for the Bayes-risk minimization problem, *optimal decision rules for the reward-rate maximization problem depend on the initial belief states.*

PROPOSITION 11. *For every $\pi\in\mathcal{S}_{m-1}$, an optimal decision rule for the reward-rate maximization problem in* (15) *with $s=0$ is given by*

$$(\tau^*,\mu^*) \equiv \big(\tau^*(\pi,T_0),\mu^*(\pi,T_0)\big) := \big(\tau(V(\pi,0),T_0),\mu(\tau(V(\pi,0),T_0))\big), \tag{22}$$

where $(\tau(c,T_0),\mu(\tau(c,T_0)))$ is the optimal decision rule given by (12) *and* (4) *for the Bayes-risk minimization problem $W(\cdot,0;V(\pi,0),T_0)$ in* (2) *with unit sampling cost $c=V(\pi,0)$ and misidentification and deadline cost parameters $c_{ij}=d_j\equiv r_j$ for every $1\leq i\neq j\leq m$.*

Proof. For any fixed $\pi\in\mathcal{S}_{m-1}$ and (τ^*,μ^*) as in (22), Proposition 8 and (19) imply that $0 = W(\pi,0;V(\pi,0),T_0) - \sum_{j=1}^{m} r_j\pi_j = \mathbb{E}_{\pi,0}[V(\pi,0)\big((\tau^*+T_0)\wedge\Theta\big) - 1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^{m} r_j 1_{\{\mu^*=j,M=j\}}]$ which is equivalent to $V(\pi,0)\,\mathbb{E}_{\pi,0}[(\tau^*+T_0)\wedge\Theta] = \mathbb{E}_{\pi,0}[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^{m} r_j 1_{\{\mu^*=j,M=j\}}]$ or

$$V(\pi,0) = \frac{\mathbb{E}_{\pi,0}\Big[1_{\{\tau^*+T_0<\Theta\}}\sum_{j=1}^{m} r_j 1_{\{\mu^*=j,M=j\}}\Big]}{\mathbb{E}_{\pi,0}[(\tau^*+T_0)\wedge\Theta]}$$

and this proves the optimality of (τ^*,μ^*) for the reward-rate maximization problem. ∎

3.3. Numerical procedure for maximizing reward rate. Thanks to Proposition 9, the maximum reward rate always lies in $[l_0,u_0]$ and can be found by a binary search on $[l_0,u_0]$ as described in Figure 1. The procedure is schematically illustrated in Figure 2. Proposition 11 implies that, unlike the optimal strategies for the Bayes-risk minimization problem, the optimal strategy for maximizing reward rate depends on the initial belief state. In other words, depending on the prior distribution over M, the stopping regions will take on different shapes. This is because different π results in different $V(\pi,0)$, equivalent to minimizing Bayes risk with different $c^*=V(\pi,0)$.

Step 0 Fix any $\pi \in \mathcal{S}_{m-1}$ and tolerance limit $\varepsilon > 0$ to check convergence. Set $n = 0$,

$$l_0 := \frac{p(1-p)^{T_0}}{1-(1-p)^{T_0}} \max_{1\le j\le m} r_j\pi_j, \quad \text{and} \quad u_0 := \frac{p}{1-(1-p)^{T_0}} \sum_{j=1}^{m} r_j\pi_j.$$

Step 1 If $\left| \sum_{j=1}^{m} r_j\pi_j - W(\pi, s; \frac{l_n+u_n}{2}, T_0) \right| < \varepsilon$, then stop and set

$$V(\pi, 0) = \frac{l_n + u_n}{2}.$$

Otherwise, set n to $n+1$. If $\sum_{j=1}^{m} r_j\pi_j > W(\pi, s; \frac{l_n+u_n}{2}, T_0)$ then set l_n to $\frac{l_{n-1}+u_{n-1}}{2}$ and u_n to u_{n-1}; otherwise, set l_n to l_{n-1} and u_n to $\frac{l_{n-1}+u_{n-1}}{2}$, and repeat to Step 1.

FIG. 1. *The algorithm to find $V(\pi, 0)$ for every fixed $\pi \in \mathcal{S}_{m-1}$.*

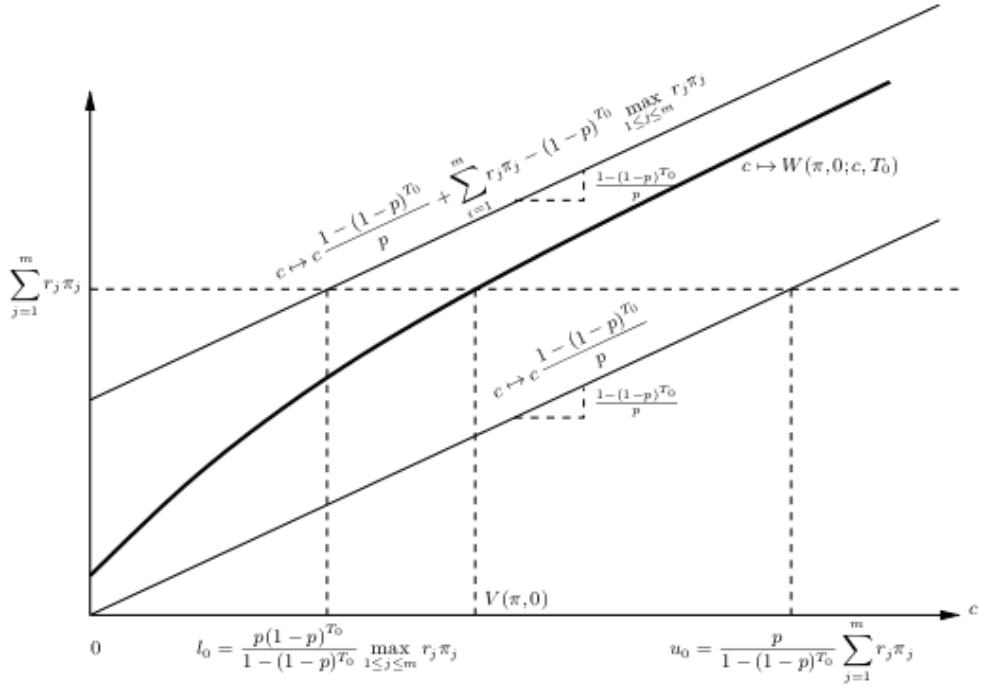

FIG. 2. *Finding $V(\pi, 0)$ for every fixed $\pi \in \mathcal{S}_{m-1}$. The strictly increasing concave continuous mapping $c \mapsto W(\pi, 0; c, T_0)$ is sandwiched between two increasing straight lines both of which intersect the vertical axis below $\sum_{j=1}^{m} r_j\pi_j$. Therefore, $c \mapsto W(\pi, 0; c, T_0)$ crosses the level $\sum_{j=1}^{m} r_j\pi_j$ at some unique $c > 0$, which coincides with $V(\pi, 0)$ by Proposition 8 and lies in the bounded interval $[l_0, u_0]$. One can find $V(\pi, 0)$ with a bisection search in $[l_0, u_0]$.*

4. Numerical examples. For illustration, we shall describe in detail the solution of the maximum reward-rate problem for sequential testing of $m = 2$ hypotheses; namely, there are two alternatives to choose from after stopping. Shiryaev [16, Chapter 4] solves the Bayes-risk minimization problem for sequential testing of two

hypotheses. Recall that there are a few fundamental differences between the two formulations and their solution methods. Let us summarize the fundamental differences between Shiryaev's *Bayes-risk minimization problem* (BRm) and our *reward-rate maximization problem* (RRM).

(i) In BRm, the unit sampling cost is a known fixed constant, and the minimum Bayes risk is sought. In RRM, the sampling costs are not considered at all, but to solve RRM we formulate an *inverse Bayes-risk minimization problem* (invBRm), in which—contrary to BRm—the minimum Bayes risk is known, and the unit sampling cost (= maximum reward rate in the original RRM) is sought. Hence, to solve RRM, one has to solve an inverse BRm problem.

(ii) Shiryaev [16] shows that BRm admits an optimal decision rule independently of the initial prior probability distribution of the hypotheses. We show that RRM also admits an optimal decision rule, but it depends on the initial prior probability distribution of the hypotheses.

(iii) Finally, BRm penalizes the decision time and misidentification, while invBRm penalizes the decision time plus time to register the decision capped by the unknown random deadline, misidentification, and late registered decisions after deadline even if they are correct.

The one-dimensional posterior probability process $\Pi_n = \mathbb{P}\{M = 1 \mid \mathcal{F}_n\}$, $n \geq 0$ and $S_n = 1_{\{\Theta \leq n\}}$, $n \geq 0$ together form a Markov sufficient statistic $(\Pi_n, S_n)_{n=1}^{\infty}$ with the dynamics

$$\mathbb{P}\{X_{n+1} \in \mathrm{d}x, S_{n+1} = 0 \mid \mathcal{F}_n\} = (1 - S_n)(1 - p)[\Pi_n f_1(x) + (1 - \Pi_n) f_2(x)]\mathrm{d}x,$$

$$\Pi_{n+1} = S_{n+1}\Pi_n + (1 - S_{n+1})\frac{\Pi_n f_1(X_{n+1})}{\Pi_n f_1(X_{n+1}) + (1 - \Pi_n) f_2(X_{n+1})}$$

for every $n \geq 0$. The maximum reward-rate and minimum Bayes-risk problems become

$$V(\pi, 0) = \sup_{(\tau,\mu)} \frac{\mathbb{E}_{\pi,0}[1_{\{\tau+T_0<\Theta\}}(r_1 1_{\{\mu=1,M=1\}} + r_2 1_{\{\mu=2,M=2\}})]}{\mathbb{E}_{\pi,0}[(\tau + T_0) \wedge \Theta]}, \quad \pi \in [0,1],$$

$$\begin{aligned} W(\pi, s; c, T_0) = \inf_{(\tau,\mu)} \mathbb{E}_{\pi,s}\big[& c\big((\tau + T_0) \wedge \Theta\big) \\ & + 1_{\{\tau+T_0<\Theta\}}\big(r_1 1_{\{\mu=2,M=1\}} + r_2 1_{\{\mu=1,M=2\}}\big) \\ & + 1_{\{\tau+T_0\geq\Theta\}}\big(r_1 1_{\{M=1\}} + r_2 1_{\{M=2\}}\big)\big], \\ (\pi, s) \in [0,1] \times \{0,1\}, & \end{aligned}$$

respectively, where supremum and infimum are taken over the pairs (τ, μ) of a stopping time τ of observation filtration $(\mathcal{F}_n)_{n\geq 0}$ and an $\mathcal{F}_\tau$-measurable $\{1,2\}$-valued random variable μ. The latter problem can be rewritten as

$$W(\pi, s; c, T_0) = \inf_{\tau} \mathbb{E}_{\pi,s}\left[\sum_{k=0}^{\tau-1} c(1 - S_k) + h(\Pi_\tau, S_\tau; c, T_0)\right]$$

for every $(\pi, s) \in [0,1] \times \{0,1\}$, where

$$\begin{aligned} &h(\pi, s; c, T_0) \\ &\quad = (1-s)\Big\{(1-p)^{T_0} \min\{r_1\pi, r_2(1-\pi)\} + \big(1-(1-p)^{T_0}\big)\Big(\frac{c}{p} + r_1\pi + r_2(1-\pi)\Big)\Big\} \\ &\qquad + s(r_1\pi + r_2(1-\pi)), \quad (\pi, s) \in [0,1] \times \{0,1\}. \end{aligned}$$

The function $W(\pi, s) \equiv W(\pi, s; c, T_0)$ is the unique bounded fixed point of operator M defined by

$$(Mw)(\pi, s) = \min\{h(\pi, s), c(1-s) + (Tw)(\pi, s)\}, \quad (\pi, s) \in [0,1] \times \{0,1\}$$

for all bounded functions $w : [0,1] \times \{0,1\} \mapsto \mathbb{R}$ such that $w(\pi, 1) = h(\pi, 1)$ for every $\pi \in [0,1]$, where

$$\begin{aligned} &(Tw)(\pi, s) \\ &\quad = s\, w(\pi, 1) + (1-s)\Big[pw(\pi, 1) + (1-p) \\ &\qquad\qquad \times \int w\Big(\frac{\pi f_1(x)}{\pi f_1(x) + (1-\pi) f_2(x)}, 0\Big)(\pi f_1(x) + (1-\pi) f_2(x))\mathrm{d}x\Big]. \end{aligned}$$

For every fixed observation cost $c > 0$ and reaction time $T_0 \geq 1$, the value function $W(\cdot, \cdot; c, T_0)$ is the pointwise limit of a decreasing sequence of successive approximations

$$w_0(\pi, s) = h(\pi, s) \quad \text{and} \quad w_{n+1}(\pi, s) = (Mw_n)(\pi, s) \quad \text{for every } (\pi, s) \in [0,1] \times \{0,1\}.$$

Finally, for every $\pi \in [0,1]$, the maximum reward rate $c = V(\pi, 0)$ is the unique solution of

$$r_1\pi + r_2(1-\pi) = W(\pi, 0; c, T_0), \tag{23}$$

which can be found by running the following algorithm of a bisection search on $[l_0, u_0]$ with

$$l_0 = \frac{p(1-p)^{T_0}}{1-(1-p)^{T_0}} \max\{r_1\pi, r_2(1-\pi)\} \quad \text{and} \quad u_0 = \frac{p}{1-(1-p)^{T_0}} \big(r_1\pi + r_2(1-\pi)\big) :$$

Step 0 Fix any $\pi \in [0,1]$ and any $\varepsilon > 0$. Set $n = 0$.
Step 1 If $\big|r_1\pi + r_2(1-\pi) - W(\pi, 0; \frac{l_n+u_n}{2}, T_0)\big| < \varepsilon$, then stop and set

$$V(\pi, 0) = \frac{l_n + u_n}{2}.$$

Otherwise, set n to $n+1$. If $r_1\pi + r_2(1-\pi) > W(\pi, 0; \frac{l_n+u_n}{2}, T_0)$ then set l_n to $\frac{l_{n-1}+u_{n-1}}{2}$ and u_n to u_{n-1}; otherwise, set l_n to l_{n-1} and u_n to $\frac{l_{n-1}+u_{n-1}}{2}$, and repeat Step 1.

For every $c > 0$ and $T_0 \geq 1$, the optimal stopping region before deadline

$$\Gamma_0(c, T_0) = \{(\pi, 0);\ \pi \in [0,1],\ W(\pi, 0; c, T_0) = h(\pi, 0; c, T_0)\}$$

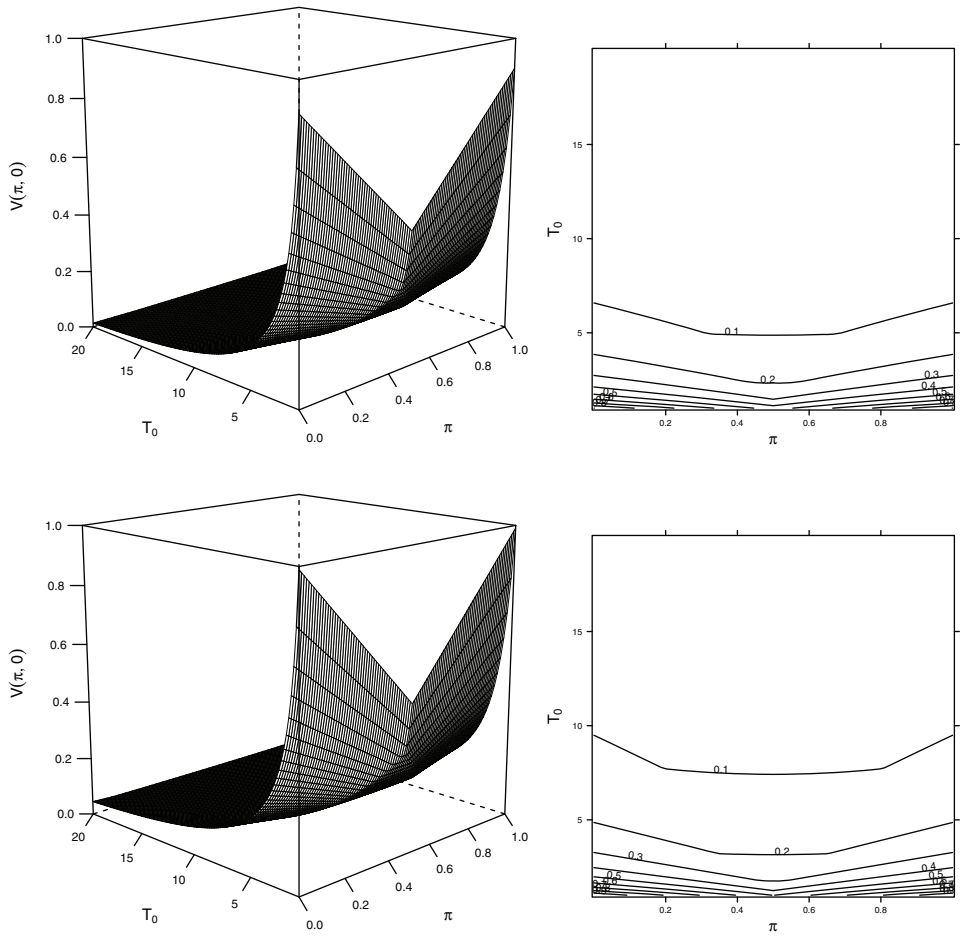

FIG. 3. *Value function* $V(\pi,0)$, $\pi\in[0,1]$ *of the reward-rate maximization problem for different* $T_0\in[1,20]$ *values (*$p=0.1$ *above and* $p=0.01$ *below,* $r_1=r_2=1$*).*

of the Bayes-risk minimization problem is in the form of $\Gamma_0(c,T_0)=\big([0,l(c,T_0)]\cup[u(c,T_0),1]\big)\times\{0\}$ for some optimal lower and upper control bounds $0<l(c,T_0)\le u(c,T_0)<1$. Therefore, starting at $(\pi,0)$ for any $\pi\in[0,1]$, an optimal decision rule for the maximum reward-rate problem is $\big(\tau(V(\pi,0),T_0),\mu(\tau(V(\pi,0),T_0))\big)$, where

$$\tau(V(\pi,0),T_0)=\inf\big\{n\ge 0;\ \Pi_n\notin\big(l(V(\pi,0),T_0),u(V(\pi,0),T_0)\big)\big\},$$
$$\mu(n)=\begin{cases}1, & \pi r_1>(1-\pi)r_2,\\ 2, & \pi r_1\le(1-\pi)r_2,\end{cases}\tag{24}$$

which depends on the initial value $\Pi_0=\pi$ of the $\Pi=(\Pi_n)_{n=0}^{\infty}$ process.

In the first numerical example, we take

$$p=0.1 \text{ and } 0.01,\quad r_1=r_2=1,\quad T_0\in[1,20].$$

Before the deadline, the value function $V(\pi,0)$, $\pi\in[0,1]$ of the reward-rate Bayesian maximization problem is plotted in Figure 3 as T_0 changes. For every fixed $\pi\in[0,1]$,

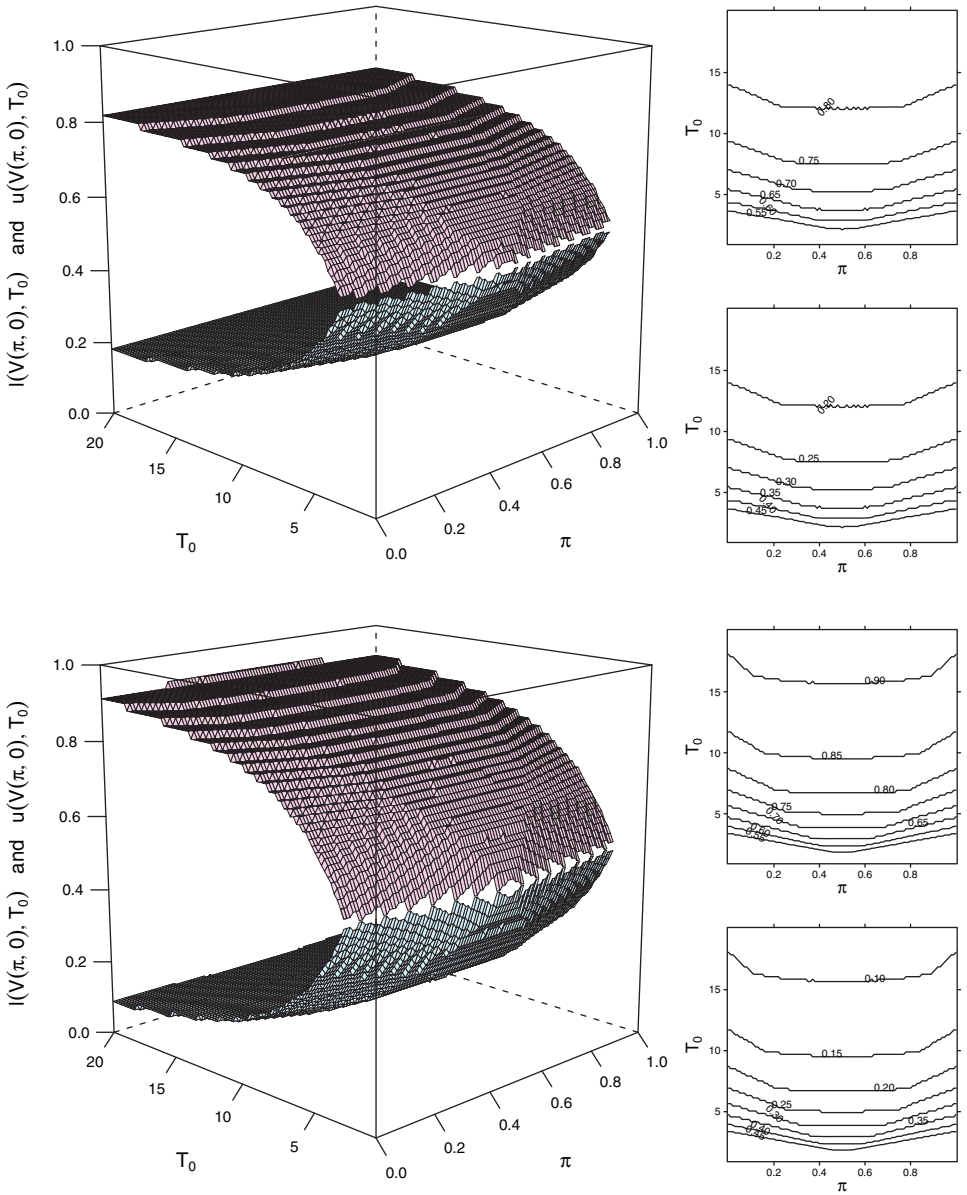

FIG. 4. *Optimal lower and upper control boundaries* $l(V(\pi,0),T_0)$ *and* $u(V(\pi,0),T_0)$ *for the reward-rate maximization problem for different* $T_0 \in [1,20]$ *values (*$p = 0.1$ *above and* $p = 0.01$ *below,* $r_1 = r_2 = 1$*).*

the value function $V(\pi,0)$ decreases as T_0 increases. For every fixed T_0, as π goes farther away from either endpoint of $[0,1]$, the uncertainty about the true hypothesis increases and the maximum reward rate $V(\pi,0)$ decreases. Because the cost of a wrong terminal decision is the same under both hypotheses, $V(\pi,0) = V(1-\pi,0)$ which is reflected by the symmetry in the figures about $\pi = 1/2$. As p decreases, the deadline Θ gets stochastically longer, the maximum reward rate increases, and the optimal continuation regions widen.

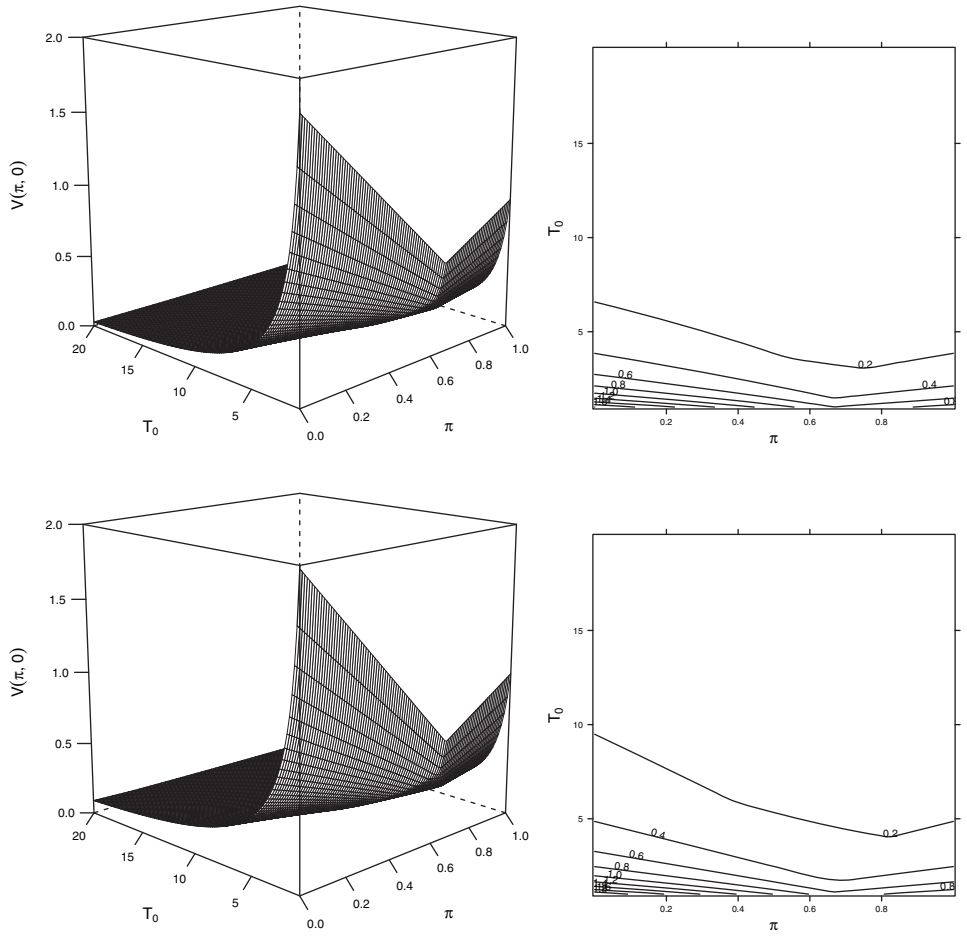

FIG. 5. *Value function* $V(\pi,0)$, $\pi\in[0,1]$ *of the reward-rate maximization problem for different* $T_0\in[1,20]$ *values (*$p=0.1$ *above and* $p=0.01$ *below,* $r_1=1$, $r_2=2$*).*

Figure 4 displays optimal lower and upper control bounds $l(V(\pi,0),T_0)$ and $u(V(\pi,0),T_0)$ in (24) for every initial $\pi\in[0,1]$ as T_0 changes. For every $\pi\in[0,1]$, the optimal continuation region $[0,1]\setminus\Gamma(V(\pi,0),T_0)=\big(l(V(\pi,0),T_0),u(V(\pi,0),T_0)\big)$ is symmetric about the center of the unit interval; namely, $l(V(\pi,0),T_0)=1-u(V(\pi,0),T_0))$, because we have equal wrong terminal decision costs $r_1=r_2$. The continuation regions $\big(l(V(\pi,0),T_0),u(V(\pi,0),T_0)\big)$, $\pi\in[0,1]$ enlarge as π approaches $1/2$ and/or as T_0 increases. Decreasing p has the same effect on maximum reward rate and optimal continuation regions as before.

In the second example, we only double the wrong terminal decision cost $r_2=2$; see Figure 5 for the maximum reward rates and Figure 6 for the optimal lower and upper control bounds. The symmetries disappear as expected, but the general properties of maximum reward rate and optimal boundaries do not change.

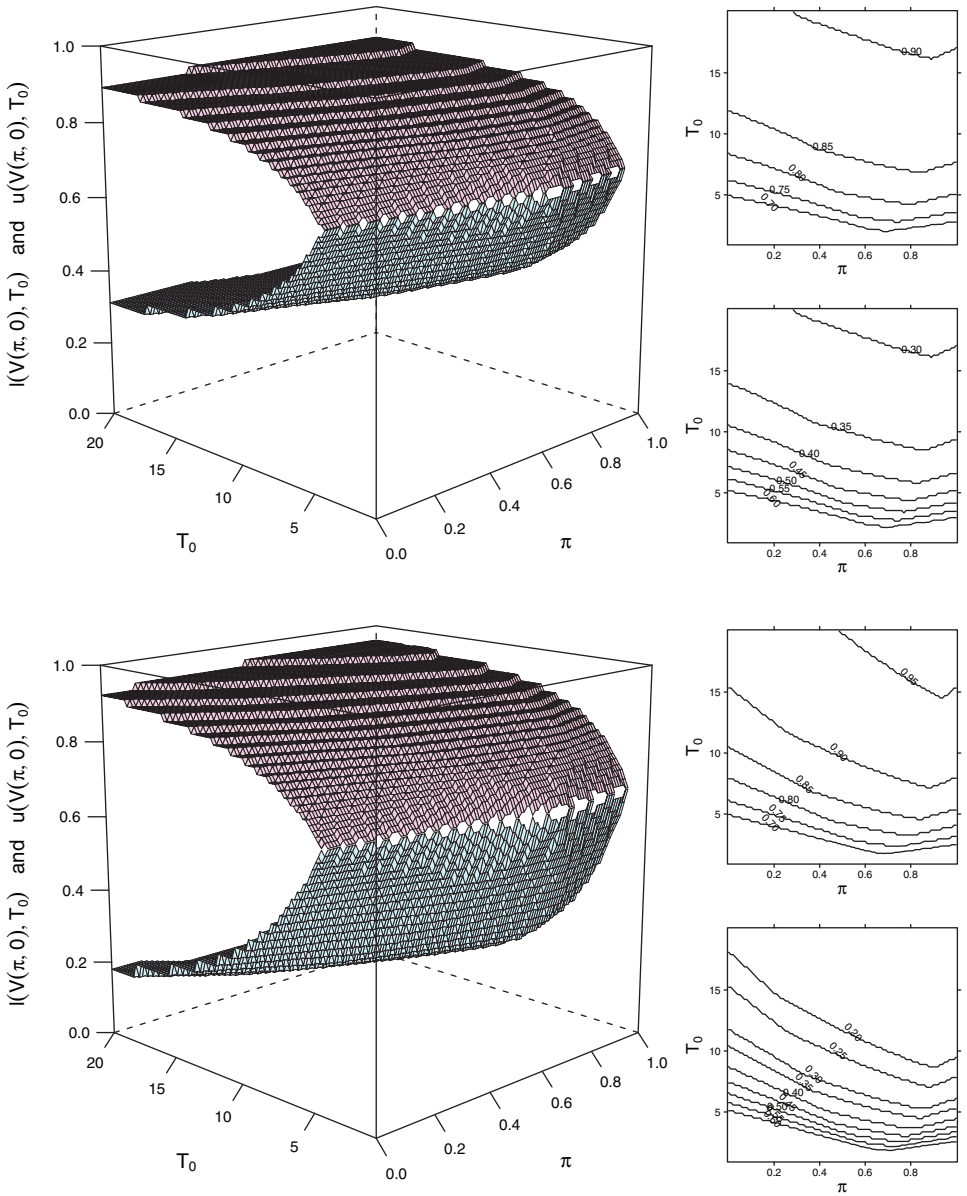

FIG. 6. *Optimal lower and upper control boundaries* $l(V(\pi,0),T_0)$ *and* $u(V(\pi,0),T_0)$ *for the reward-rate Bayesian maximization problem for different* $T_0 \in [1,20]$ *values (*$p = 0.1$ *above and* $p = 0.01$ *below,* $r_1 = 1$, $r_2 = 2$*).*

Acknowledgments. We thank William Bialek and Peter Frazier for stimulating and fruitful discussions.

REFERENCES

[1] R. BELLMAN, *On the theory of dynamic programming*, Proc. Natl. Acad. Sci. USA, 38 (1952), pp. 716–719.

[2] D. P. BERTSEKAS AND S. E. SHREVE, *Stochastic optimal control: The discrete time case*, Math. Sci. Eng. 139, Academic Press, New York, 1978.

[3] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, Dover, New York, 1979.
[4] R. Bogacz, E. Brown, J. Moehlis, P. Hu, P. Holmes, and J. D. Cohen, *The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks*, Psychol. Rev., 113 (2006), pp. 700–765.
[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
[6] A. K. Churchland, R. Kiani, and M. N. Shadlen, *Decision-making with multiple alternatives*, Nature Neurosci., 11 (2008), pp. 693–702.
[7] S. Dayanik, C. Goulding, and H. V. Poor, *Bayesian sequential change diagnosis*, Math. Oper. Res., 33 (2008), pp. 475–496.
[8] S. Dayanik, H. V. Poor, and S. O. Sezer, *Sequential multi-hypothesis testing for compound Poisson processes*, Stochastics, 80 (2008), pp. 19–50.
[9] S. Dayanik and S. O. Sezer, *Sequential testing of simple hypotheses about compound Poisson processes*, Stochastic Process. Appl., 116 (2006), pp. 1892–1919.
[10] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, *Multihypothesis sequential probability ratio test* (i): *Asymptotic optimality*, IEEE Trans. Inform. Theory, 45 (1999), pp. 2448–2461.
[11] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, *Multihypothesis sequential probability ratio test* (ii): *Accurate asymptotitc expansions for the expected sample size*, IEEE Trans. Inform. Theory, 46 (2000), pp. 1366–1383.
[12] G. Fellouris and G. V. Moustakides, *Decentralized sequential hypothesis testing using asynchronous communication*, IEEE Trans. Inform. Theory, 57 (2011), pp. 534–548.
[13] P. Frazier and A. J. Yu, *Sequential hypothesis testing under stochastic deadlines*, Adv. Neural Inform. Process. Systems, 20 (2008).
[14] J. I. Gold and M. N. Shadlen, *Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward*, Neuron, 36 (2002), pp. 299–308.
[15] Y. Mei, *Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks*, IEEE Trans. Inform. Theory, 54 (2008), pp. 2072–2089.
[16] A. N. Shiryaev, *Optimal Stopping Rules*, Springer-Verlag, New York, 1978.
[17] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.
[18] A. Wald and J. Wolfowitz, *Optimal character of the sequential probability ratio test*, Ann. Math. Statist., 19 (1948), pp. 326–339.
[19] K. F. Wong-Lin, P. Eckhoff, P. Holmes, and J. D. Cohen, *Optimal performance in a countermanding saccade task*, Brain Res., 1318 (2010), pp. 178–187.