TP12 - 3:30

87-02

Proceedings of the 26th Conference on Decision and Control
Los Angeles, CA • December 1987

Sequential Testing on the Rate of a Counting Process

John S. Baras * and David C. Mac Enany **

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, MD 20742

Abstract

In this paper, optimality results are given for the problems of Bayesian and Wald sequential (simple, binary) hypothesis testing on the rate of a counting process. An explicit formula is given for the Bayes risk, and the system to solve for the exact optimal thresholds is also given.

Introduction

Space limitations do not permit a gentle introduction into the nature and importance of sequential testing problems. Fortunately, these notions have been extensively documented since Wald's original contribution [6] until the present [5]. For a discussion of their importance in applications see [2]. In an earlier report [4], sufficient conditions are given to prove Bayesian optimality results for general semimartingales, along with complete proofs of the results herein. These results will appear elsewhere with recent extensions and improvements.

I. Problem Framework

The Bayesian sequential, simple (binary) hypothesis testing problem is outlined as follows. One is given a probability triple $(\Theta,\sigma(\Theta),P)$ and some $\sigma(\Theta)$-measurable, binary-valued random variable (r.v.), θ, where Θ defines a set of hypotheses consisting of two elements, $\sigma(\Theta)$ is the trivial σ-algebra on Θ, and P assigns θ mass according to $P\{\theta=0\}=1-\pi$; $P\{\theta=1\}=\pi$, for some arbitrary, fixed real number $\pi\in[0,1]$. One is also given a measurable space $(\Omega,\mathcal{F})$ upon which there are defined two $\mathcal{F}$-completed probability measures P_0 and P_1, independent of P. The hypothesis testing problem takes place on the product space defined next. Construct the complete probability triple $(\Omega^\pi,\mathcal{F}^\pi,P^\pi)$ where $\Omega^\pi=\Theta\times\Omega$, $\mathcal{F}^\pi=\sigma(\Theta)\otimes\mathcal{F}$, and with the probability measure P^π defined via,

$$P^\pi\{\{\theta=i\}\cap F\}=P\{\theta=i\}\,P_i\{F\}\qquad \forall F\in\mathcal{F};\ i=0,1. \quad (1.1)$$

Hence,

$$P^\pi\{F\}=\pi\,P_1\{F\}+(1-\pi)\,P_0\{F\}\qquad \forall F\in\mathcal{F}. \quad (1.2)$$

It is assumed that the random variable θ is unobservable, but that one can observe a counting process whose statistics under each of the hypotheses—$\theta=0$, $\theta=1$—are governed by the probability measures P_0 and P_1, respectively. To be precise, for each $\omega^\pi=(\theta,\omega)\in\Omega^\pi$, one observes a counting process, $\{n_t:t\ge 0\}$, with semimartingale representation,

$$n_t(\omega^\pi)=\int_0^t\left[(1-\theta)\lambda_s^0(\omega)+\theta\lambda_s^1(\omega)\right]ds+m_t(\omega^\pi)\quad \forall t\ge 0, \quad (1.3)$$

where $\{m_t:t\ge 0\}$ is a $(P^\pi,\mathcal{F}_t^\pi)$-martingale, and the intensity, $\{\lambda_t^i:t\ge 0\}$, is a nonnegative $\mathcal{F}_t$-predictable process satisfying $E_i\int_0^t\lambda_s^i\,ds<\infty$, $\forall t<\infty$, and for each $i=0,1$. Denote by $\{\mathcal{O}_t:t\ge 0\}$ the family of σ-algebras generated by the observation process. Based on this family, the manner in which the true value for θ is chosen has two parts. First, the decision to terminate the observation procedure is made according to a $(P^\pi,\mathcal{O}_t)$-stopping time, say τ, and second, a judgment as to the true value of θ is made according to a $(P^\pi,\mathcal{O}_\tau)$ binary-valued random variable, say δ. Any such pair, (τ,δ) is called an **admissible policy** and one seeks an admissible policy which minimizes the P^π-average **Bayes cost**, $\rho(\pi,\tau,\delta)$, here defined as,

$$\rho(\pi,\tau,\delta)=E^\pi\Big[\int_0^\tau c_s\,ds+w(\theta,\delta)\Big], \quad (1.4)$$

where if for $c^0,c^1>0$ one defines,

$$w(\theta,\delta)=\begin{cases}c^0, & \text{if } \theta=1 \text{ and } \delta=0;\\ c^1, & \text{if } \theta=0 \text{ and } \delta=1;\\ 0, & \text{if } \theta=\delta,\end{cases}$$

then $E^\pi[w(\theta,\delta)]$ yields the P^π-average cost of (incorrect) decisions, and where $E^\pi\int_0^\tau c_s\,ds$ gives the P^π-average **waiting cost** when $\{c_t:t\ge 0\}$ is some nonnegative $\mathcal{O}_t$-adapted process which models a desired costing behavior and satisfies,

$$E^\pi\int_0^t c_s\,ds<\infty\quad \forall t<\infty. \quad (1.5)$$

Note that it is without loss of generality that no cost is levied for correct decisions in 1.4. Using 1.4 one can now define the Bayesian optimality criterion.

Definition 1 *An admissible policy,* (τ_*,δ_*) *is said to be* **Bayesian**, *or* **Bayesian optimal**, *if*

$$\rho(\pi,\tau_*,\delta_*)=\inf_{(\tau,\delta)}\rho(\pi,\tau,\delta),\qquad \forall\pi\in[0,1], \quad (1.6)$$

where the infimum is over all admissible policies, and then $\rho(\pi)=\rho(\pi,\tau_*,\delta_*)$ *is called the* **Bayes cost.**

The minimization indicated in 1.6 can be greatly simplified, as is pointed out next. Define the *a posteriori* probability of the "$\theta=1$" hypothesis as,

$$\pi_t=P^\pi\{\theta=1|\mathcal{O}_t\}\qquad t\ge 0. \quad (1.7)$$

With only reasonable assumptions on the nature of $\{c_t:t\ge 0\}$, one can prove the following lemma [4,5]:

Lemma: *Define,*

$$\rho(\pi,\tau)=E^\pi\Big[\int_0^\tau c_s\,ds+e(\pi_\tau)\Big], \quad (1.8)$$

where, $e(\pi)=\min\{c^0\pi,c^1(1-\pi)\}$. *Let* $\mathcal{T}$ *denote the class of* $(P^\pi,\mathcal{O}_t)$*-a.s. finite stopping times, and suppose* τ_* *satisfies,* $\rho(\pi,\tau_*)=\inf_{\tau\in\mathcal{T}}\rho(\pi,\tau)$. *Also, let* $\pi_e=\frac{c^1}{c^0+c^1}$ *and define,*

$$\delta_*=\delta(\tau_*)=\begin{cases}1 & \pi_{\tau_*}\ge\pi_e;\\ 0 & \pi_{\tau_*}<\pi_e.\end{cases} \quad (1.9)$$

Then (τ_*,δ_*) *satisfies 1.6;* $e(\pi_\tau)$ *is called the* **terminal cost.**

Thus, *the search for a Bayesian optimal policy can be reduced to a search for an optimal stopping time.* In the light of this and other facts [4,5], it becomes clear that an important subclass of admissible policies are the threshold policies.

Definition 2 **A threshold policy** *based on the a posteriori probability process, is an admissible policy pair,* $(\bar\tau,\bar\delta)$, *with*

$$\bar\tau=\inf\{t\ge 0:\ \pi_t\notin I\}, \quad (1.10)$$

* Research supported in part by NSF Grant CDR-85-00108, ONR grant N00014-83-K-0731, and by a grant from UNISYS Corporation.

** Research supported by an ONR Fellowship.

and,

$$\bar{\delta} = \begin{cases} 1 & \pi_{\bar{\tau}} \geq b; \\ 0 & \pi_{\bar{\tau}} \leq a, \end{cases} \tag{1.11}$$

where $0 < a \leq \pi_e \leq b < 1$ are the **thresholds**, *and $I = (a, b)$ is called the* **threshold**, *or* **continuation interval**. *If $a = \pi_e = b$, the threshold policy is said to be* **degenerate** *and one defines $\bar{\tau} \equiv 0$ and $\bar{\delta} = 1$ if $\pi_0 = \pi_e$. From 1.9, note that $\bar{\delta} \equiv \delta_*$, i.e., both yield exactly the same decisions.*

In the next section we indicate that a Bayesian threshold policy exists in the problem of testing two simple hypotheses on the rate of a counting process, and show how to compute the optimal thresholds.

II. Testing on the Rate of a Counting Process

With the set-up given in the previous section it remains only to specify the nature of the intensities under each hypothesis, and to choose a reasonable form for the waiting cost. For simplicity, consider the important special case where the count-rate intensities are given by

$$\lambda_t^i = \lambda^i \cdot \lambda_t \qquad \forall t \geq 0 \quad i = 0, 1, \tag{2.1}$$

with $\lambda^1 > \lambda^0$ strictly positive constants and $\{\lambda_t : t \geq 0\}$ some nonnegative $(P_i, \mathcal{F}_t)$-predictable process. A reasonable choice for the average waiting cost for this problem is,

$$\int_0^\tau c_s \, ds = \int_0^\tau c \cdot (\hat{\lambda}_s^1 - \hat{\lambda}_s^0) \, ds, = \int_0^\tau c \cdot (\lambda^1 - \lambda^0) \hat{\lambda}_s \, ds \tag{2.2}$$

with $c > 0$ a known constant, and where $\hat{\lambda}_t = E_i[\lambda_t | \mathcal{O}_{t-}]$, $\forall t > 0$. This choice of cost avoids analytical intractability and is the one used in the theorem to follow. For technical reasons, it is required that the common factor intensity satisfy,

$$P_i\{\int_0^\infty \hat{\lambda}_s \, ds = \infty\} = 1. \tag{2.3}$$

With this set-up, one can state the following theorem [4].

Theorem *Assume 2.3 holds. In the problem of Bayesian sequential hypothesis testing, based on the partial observation process $\{n_t : t \geq 0\}$ (see 1.3 and 2.1), with running cost, $E^\pi[\int_0^\tau c_s \, ds] = E^\pi[\int_0^\tau c\,(\hat{\lambda}_s^1 - \hat{\lambda}_s^0)\, ds]$, (see 2.2), and terminal cost $e(\pi_\tau)$, the threshold policy, (τ_*, δ_*), based on $\{\pi_t : t \geq 0\}$ and with threshold interval $I_* = (a_*, b_*)$ is Bayesian optimal. The Bayes cost is defined as,*

$$\rho(\pi) = \begin{cases} e(\pi) & \pi \notin (a_*, b_*); \\ r_*(\pi) & \pi \in (a_*, b_*), \end{cases}$$

and the **subcost** [4], *$r_*(\pi) = r(\pi; a_*, b_*)$, is given by,*

$$r(\pi; a_*, b_*) = \mathcal{D}(\pi; b_*) + \mathcal{H}(\pi; a_*, b_*) \quad \forall \pi \in (0, 1),$$

where,

$$\mathcal{H}(\pi; a, b) = c^1(1 - \pi) + K(a, b)\,\bar{H}(\pi; b);$$

$$\mathcal{D}(\pi; b) = c\,\frac{\lambda^1 - \lambda^0}{\lambda^0 \lambda^1}\,[d(\pi; b) + \bar{D}(\pi; b)];$$

$$d(\pi; b) = (\lambda^1(1 - \pi) + \lambda^0 \pi)\,(1 + N_b(\pi));$$

$$N_b(\pi) = \lfloor x(b) - x(\pi) - 1 \rfloor,$$

with, $x(\pi) = \log[\frac{\pi}{1-\pi}] / \log \frac{\lambda^1}{\lambda^0}$, and with $\lfloor \cdot \rfloor$ denoting the floor function. The subcost definition is completed by specifying,

$$\bar{D}(\pi; b) = -\lambda^1(1 - \pi)\,e^{\nu_0 X_b(\pi)} D_0(\pi; b) - \lambda^0 \pi\, e^{\nu_1 X_b(\pi)} D_1(\pi; b);$$

$$\bar{H}(\pi; b) = \lambda^1(1 - \pi)\,e^{-\nu_0 x(\pi)} H_0(\pi; b) + \lambda^0 \pi\, e^{-\nu_1 x(\pi)} H_1(\pi; b),$$

with H_i, D_i, $i = 0, 1$ defined via,

$$H_i(\pi; b) = \sum_{n=0}^{N_b(\pi)} \frac{(-1)^n}{n!} [(X_b(\pi) - n)\nu_i e^{-\nu_i}]^n;$$

$$D_i(\pi; b) = e^{-\nu_i} \sum_{n=0}^{N_b(\pi)-1} e^{-\nu_i n} \sum_{m=0}^{n} \frac{(-1)^m}{m!} [(X_b(\pi) - n - 1)\nu_i]^m,$$

where,

$$X_b(\pi) = x(b) - x(\pi); \quad \nu_i = \frac{\lambda^i \log \frac{\lambda^1}{\lambda^0}}{\lambda^1 - \lambda^0}, \; i = 0, 1.$$

The constant $K(a, b)$ is given by,

$$K(a, b) = \frac{c^0 a - [c^1(1 - b) + \mathcal{D}(a; b)]}{\bar{H}(a; b)}.$$

The interval $I_* = (a_*, b_*)$ is the unique solution to

$$r'(a_*; a_*, b_*) = c^0;$$

$$r'(b_*^-; a_*, b_*) = \frac{c}{b_*(1 - b_*)} - c^1.$$

Note, that the 'empty sum equals zero', and '$0^0 = 1$' conventions are used.

There is no *easy* extension of this result to the more general case where the ratio of the intensities of the counting process under each hypothesis is not deterministic. In general when the ratio of the intensities is stochastic, the state-space of the process must be enlarged to include the ratio. This can be shown to lead to a partial functional differential equation for the cost in two variables. In addition, since one is forced to consider a 2-dimensional process, this has the immediate consequence that in general, one must extend the notion of threshold policies from intervals to include open sets.

In the Wald problem of sequential testing [6], θ is viewed as deterministic and unknown, and one uses a threshold policy with threshold chosen to minimize the P_i-average waiting cost 1.5 for both $i = 0, 1$, with fixed probabilities of error of both types. Using the theorem above and the lemma due to Le Cam [3], it follows that a unique, Wald-optimal threshold policy exists. See [1,2] for the system to solve for the optimal thresholds. It is clear that solving this system and the one in 2.4 is quite difficult in general. However, their availability in closed form is certainly necessary to judge the efficacy of approximation schemes.

Acknowledgments

The authors thank A. Makowski and J. Gubner for their helpful discussions during the course of this research.

References

[1] A. Dvoretzky, J. Kiefer & J. Wolfowitz, "Sequential Decision Problems for Processes with Continuous Time Parameter. Testing Hypotheses", *Ann. Math. Stat.*, vol. 24, 254-264, 1953.

[2] D. Kazakos, P. Papantoni-Kazakos[1980] "Sequential Detection Between Poisson Processes", *IEEE Trans. on Info. Theory*, Vol. IT-26, No 1, Jan 1980.

[3] E.L. Lehmann, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1958.

[4] D.C. Mac Enany, "Bayesian Sequential Hypothesis Testing", SRC-TR87-107, Systems Research Center, Univ. of Maryland (*Master's Thesis* 1986).

[5] A.N. Shiryayev, *Optimal Stopping Rules* (English Translation), Springer-Verlag, New York-Heidelberg-Berlin 1977.

[6] A. Wald, *Sequential Analysis*, Dover Publications, Inc., New York, N.Y. 1947.