Evaluating Sequential Tests for A Class of Stochastic Processes *

Xiaoping Xiong and Ming Tan

St. Jude Children's Research Hospital

ABSTRACT

We propose computational methods for sequential tests for a class of stochastic processes for which the probability density of test statistic can be factorized into a product of a known likelihood function that is independent of the stopping rule and a conditional probability that is independent of parameters of interest using the sufficiency principle. The proposed methods improve accuracy and efficiency of computation and enable us to evaluate properties of special interests in sequential tests such as the probability of discordance between a sequential test and the nonsequential test at the last stage of the sequential test. We give examples for evaluating sequential tests standardized in terms of Brownian motion in an interval of information time.

Key Words: Sequential analysis; computation method; discordance probability

Contact-X Xiong, 332 N. Lauderdale, Memphis, TN 38105 (xiaoping.xiong@stjude.org)

1 Introduction

A major difficulty in sequential analysis is to evaluate various properties of sequential tests and post-sequential estimation (Siegmund 1985). Analytical solutions are not generally available, especially for tests with nonlinear and discrete sequential boundaries. Asymptotic approximations are usually difficult and not generally applicable because the requirement of large sample size contradicts the very goal of early stopping for sequential tests. Numeric methods can be used to evaluate basic characteristics in the design and analysis for sequential tests but are difficult in dealing with characteristics with probabilistic complexity. Additionally, the insights from numerical solutions are limited compared with that from analytical solutions. In this paper, we introduce a general approach for evaluating sequential tests and estimation. This approach is analytic except where a function needs to be numerically calculated. Since many sequential tests have the same type I and II errors, other properties of sequential tests have to be examined to select the most appropriate sequential design. To do so, we need to evaluate properties of the tests and the post-testing parameter estimation efficiently. For example, it is of interest to know the *probability of discordance* which measures the chance that the sequential test does not agree with the non-sequential test at the planned end of the very sequential test, or does not agree with the non-sequential test that has the same significance level and power. This approach was first proposed in Xiong (1991, 1996) to improve the computational methods by Aroian (1968) and Armitage et al. (1969), and is generalized for samples of random variables in Xiong et al. (2000). The purpose of this paper is to extend these results to a class of stochastic processes.

In Section 2, we introduce equations for computation in sequential analysis for a class of stochastic processes based on sufficiency principle. In Section 3, we derive methods for calculating parameter-free function $l(t, s)$ which plays an important role in the proposed approach. In Section 4, we evaluate sequential tests and post-test estimation and give an example of sequential test normalized on information time.

*This research is in part supported by NIH grant R01HL61681, CA 21765, and by the American Lebanese Syrian Associated Charities.

2 Equations

Let Y_t, $0 \le t < \infty$ be a stochastic process, e.g., a Brownian process or a Poisson process. Let θ be a parameter for distribution of Y_t and assume S_t is a sufficient statistic for θ such that for any $t \le t'$, the conditional distribution of S_t given $S_{t'}$ does not depend on θ. Let T be a stopping time, i.e. for any t, $(T \le t)$ is determined only by $S_\tau, \tau < t$. The probability mass or density function of (T, S_T) is $p_\theta(t,s) \equiv \lim_{\delta\to 0} P_\theta(T = t, s < S_T < s+\delta)/\delta$. Then for any (t,s),

$$p_\theta(t,s) = f_\theta(t,s) l(t,s), \tag{1}$$

where $f_\theta(t,s)$ is the density of S_t and $l(t,s)$ is the conditional probability

$$l(t,s) = P(T = t | S_t = s), \tag{2}$$

which does not depend on θ. The importance of equation (1) is that the $f_\theta(t,s)$ depends on the parameter θ but not on the stopping rule, and that the conditional probability $l(t,s)$ depends on the stopping rule but not on θ, if S_t is a sufficient statistic for θ. The $l(t,s)$ in (1) functions similarly as the eigenvalue in linear algebra. We hence shall call $l(t,s)$ the *eigenvalue function*. From (1), a general equation useful for computation in sequential procedures is as follows. Assume $P_\theta(T < \infty) = 1$ for any θ, and for a function $H(t,s)$, $E_\theta[H(t,S_t)]$ exists for any t. Assume T takes only discrete values $t_1, t_2, \cdots$, then

$$E_\theta\{H(T,S_T)\} = \sum_{t_k} E_\theta\{H(t_k, S_{t_k}) l(t_k, S_{t_k})\}. \tag{3}$$

Define function $h(\theta) = E_\theta\{H(T,S_T)\}$, then $h'(\theta) = E_\theta\left\{H(T,S_T)\frac{\partial}{\partial\theta}\{\log f_\theta(T,S_T)\}\right\}$, where $f_\theta(t,s) \equiv f_\theta^{(t)}(s)$ is the density function of S_t. $h'(\theta)$ can be evaluated as for $h(\theta)$ in (3). $h'(\theta)$ is very useful for calculating θ solved from equation $h(\theta) = constant$. One application is to obtain bias-adjusted estimator.

3 Eigenvalue Function

Let $\mathcal{B}$ be the stopping region and $\mathcal{C}$ be the continuation region for (T, S_T). In general, $\mathcal{B}$ may overlap $\mathcal{C}$ because the stopping rule depends on the history of the path (t, S_t), $t > 0$. In this paper, we only consider the stopping rules for which $\mathcal{B}$ and $\mathcal{C}$ are not overlapping, i.e., $\mathcal{B} \cap \mathcal{C} = \emptyset$. Assume for any event $A \in \sigma(S_{t_1}, \cdots, S_{t_k})$, $P(A|S_{t_k}, S_{t_{k+i}}) = P(A|S_{t_k})$ for any $k, i \ge 1$. This condition holds if S_{t_k}s is a Markov sequence; as a special case, if S_{t_k}s is a sequence with independent increments. For any (t,s), define function $l^*(t,s) \equiv P(T > t | S_t = s)$. Then for any $(t,s) \in \mathcal{B}$, $l^*(t_k, s) = 0$. Assume (t, S_t) starts from $t = 0$ and t_1 is the first possible time for stopping. Then for any $(t_k, s) \in \mathcal{C}$, $l^*(t_1, s) \equiv 1$ and $l^*(t_k, s) = E\{l^*(t_{k-1}, S_{t_{k-1}}) | S_{t_k} = s\}$ for $k \ge 2$. For any $(t_k, s) \in \mathcal{C}$, $l(t_k, s) \equiv 0$. For any $(t_k, s) \in \mathcal{B}$, $l(t_1, s) \equiv 1$ and $l(t_k, s) = E\{l^*(t_{k-1}, S_{t_{k-1}}) | S_{t_k} = s\}$ for $k \ge 2$. Denote conditional density function for S_{t_j} given S_{t_k} for $t_j < t_k$ as

$$p_{t_j|t_k}(t|s) \equiv P(S_{t_j} = t | S_{t_k} = s). \tag{4}$$

$p_{t_j|t_k}(t|s)$ does not depend on θ if S_t is a sufficient statistic for θ. For example, for Brownian process, assume $S_t \sim N(t\mu, t\sigma^2)$ for any t. Then for $t_j < t_k$ and given S_{t_k}, the S_{t_j} has a normal distribution of $N\left(S_{t_k}\frac{t_j}{t_k}, t_j(1-\frac{t_j}{t_k})\sigma^2\right)$,

$$p_{t_j|t_k}(t|s) = \phi_{t_j|t_k}(t|s) \equiv \quad \phi\left((t - \tfrac{t_j}{t_k}s)/\sqrt{t_j(1-\tfrac{t_j}{t_k})}\sigma\right)/\sqrt{t_j(1-\tfrac{t_j}{t_k})}\sigma, \tag{5}$$

where $\phi(x)$ is the density function of standard normal distribution. For Poisson process, assume $S_t \sim \mathcal{P}(t\lambda)$ for any t. For $t_j < t_k$ and given S_{t_k}, the S_{t_j} has a binomial distribution $B\left(S_{t_k}, \frac{t_j}{t_k}\right)$,

$$p_{t_j|t_k}(t|s) = b_{t_j|t_k}(t|s) \equiv \binom{t}{s}\left(\frac{t_j}{t_k}\right)^t\left(1-\frac{t_j}{t_k}\right)^{s-t}. \tag{6}$$

Let T be the first time the process exits from $\mathcal{C}$ to $\mathcal{B}$ where $\mathcal{B} = \{(t_k, s): s \leq b_{t_k} \text{ or } s \geq a_{t_k} \text{ for } k = 1, \cdots\}$ and $\mathcal{C} = \{(t_k, s): b_{t_k} < s < a_{t_k} \text{ for } k = 1, \cdots\}$. Then for $s \in (b_{t_k}, a_{t_k})$, $l(t_k, s) = 0$; and for $s \in (b_{t_k}, a_{t_k})^c$,

$$l(t_k, s) = \int_{b_{t_{k-1}}}^{a_{t_{k-1}}} l^*(t_{k-1}, x) p_{t_{k-1}|t_k}(x|s)dx. \tag{7}$$

The $l^*(t_{k-1}, s)$ for $s \in (b_{t_{k-1}}, a_{t_{k-1}})$ is calculated by

$$l^*(t_{k-1}, s) = \int_{b_{t_{k-2}}}^{a_{t_{k-2}}} l^*(t_{k-2}, x) p_{t_{k-2}|t_{k-1}}(x|t)dx. \tag{8}$$

Summation is used instead of integration in above equations if $p_{t_j|t_k}(t|s)$ is discrete as that in (6) for Poisson processes.

4 Evaluating Sequential Tests and Estimation

4.1 Operating Characteristics and Post-test Estimation

Let $H(t, s) = 1_{(s > a_t)}$, the power function of sequential test is

$$\beta(\theta) = P_\theta(S_T \geq a_T) = E_\theta\left\{1_{(S_T \geq a_T)}\right\} = \sum_{t_k} E_\theta\left\{1_{(S_{t_k} > a_{t_k})} l(t_k, S_{t_k})\right\}. \tag{9}$$

Let $H(t, s) = t$, then the expected stopping time is $E_\theta(T) = \sum_{t_k} E_\theta\{t_k l(t_k, S_{t_k})\}$, which is analogous to the expected sample size for samples of random variables. It is well known that if S_t has density $f_\theta^{(t)}(s) = exp(\theta s + \eta(\theta)t)b(t, s)$, then $E_\theta(S_t) = \theta t$. The maximum likelihood estimator $\hat{\theta} = S_T/T$ is biased. A bias-adjusted estimator $\tilde{\theta}$ is the solution of θ from $E_\theta(S_T/T) = (S_T/T)_{observed}$ (Whitehead 1986). To solve this equation, we note $h(\theta) = E_\theta(S_T/T) = \sum_{t_k} E_\theta\left\{\frac{S_{t_k}}{t_k} l(t_k, S_{t_k})\right\}$ and $h'(\theta) = \frac{1}{Var(S_1)}\sum_{t_k} E_\theta\left\{S_{t_k}\left(\frac{S_{t_k}}{t_k} - \theta\right) l(t_k, S_{t_k})\right\}$ and then use the Newton-Ralphson method.

4.2 Probabilities of Discordance

Let D be the event that a sequential test and a non-sequential test lead to a different rejection/acceptance decision when both tests were used on the same sequence of observations. We say a sequential procedure and a non-sequential test are *comparable* if the probability of discordance $P(D)$ is negligible. Let $\mathcal{B}^a$ and $\mathcal{B}^r$ be the acceptance region and the rejection region for (T, S_T) for testing $H_0: \theta \in \Theta_0$ vs. $H_a: \theta \in \Theta_a$. Let $\mathcal{R}^a$ and $\mathcal{R}^r$ be the acceptance region and the rejection region for S_{t^*} for testing the same hypotheses, where t^* is a given time. Define events

$$D^a = \{(T, S_T) \in \mathcal{B}^r, S_{t^*} \in \mathcal{R}^a\} \text{ and } D^r = \{(T, S_T) \in \mathcal{B}^a, S_{t^*} \in \mathcal{R}^r\}. \tag{10}$$

$D = D^a \cup D^r$ includes all possible events of discordance between the two testing statistics, (T, S_T) and S_{t^*}. The probability of discordance is $\rho(\theta) \equiv P_\theta(D) = P_\theta(D^a) + P_\theta(D^r)$. Let $\mathcal{B}^a_{t_k} = \{s : (t_k, s) \in \mathcal{B}^a\}$ and $\mathcal{B}^r_{t_k} = \{s : (t_k, s) \in \mathcal{B}^r\}$. Then

$$P_\theta(D \cap (T \le t^*)) = E_\theta\{P(D \cap (T \le t^*)|S_{t^*})\} \tag{11}$$

where $P(D \cap (T \le t^*)|S_{t^*}) = \sum_{t_k \le t^*} E\{1_{(S_{t_k} \in \mathcal{B}^a_{t_k})} l(t_k, S_{t_k})|S_{t^*}\}$ if $S_{t^*} \in \mathcal{R}^r$, and $P(D \cap (T \le t^*)|S_{t^*}) = \sum_{t_k \le t^*} E\{1_{(S_{t_k} \in \mathcal{B}^r_{t_k})} l(t_k, S_{t_k})|S_{t^*}\}$ if $S_{t^*} \in \mathcal{R}^a$.

Let $\rho_s = P(D|S_{t^*} = s)$ be the conditional probability of discordance given $S_{t^*} = s$. If S_{t^*} is unknown, then ρ_s is unknown, however, is bounded by the *maximum conditional probability of discordance* $\rho = \max_s \rho_s$. $\rho(\theta)$ is the probability of discordance given true θ. The probabilities of discordance ρ_s and $\rho(\theta)$ measure the chances that the sequential test contradicts the non-sequential test: for ρ_s, the probability is conditioning on S_{t^*} and is not dependent of θ; for $\rho(\theta)$, the probability is unconditional on S_{t^*} and is dependent of θ. The ρ provides an upper-bound for both probabilities: $\rho_s < \rho$ for any s and $\rho(\theta) < \rho$ for any θ.

If the non-sequential test is at the last stage of the sequential test, its sampling time is t^*, the maximum time of the sequential test. Hence we have $P_\theta(T \le t^*) \equiv 1$ and $P_\theta(D) = P_\theta\{D \cap (T \le t^*)\}$. Then $\rho_s = P(D \cap (T \le t^*)|S_{t^*} = s)$ and $\rho(\theta)$ can be obtained by (11). If the non-sequential test is conducted before the last stage of the sequential test, evaluations of discordant probabilities ρ_s and $\rho(\theta)$ are more difficult and not given here. These evaluations are similar to those in Xiong et al. (2000) for samples of random variables.

4.3 Examples

Assume $X_i, i = 1, 2, \cdots, m$ are observations from a normal distribution with mean μ and variance σ^2. To test hypothesis $H_0 : \mu \le \mu_0$ vs. $H_a : \mu > \mu_0$, with significance level α and power $1-\beta$ for alternative $\mu = \mu_a$, the sample size for a nonsequential test is $m = \sigma^2(\mu_a - \mu_0)^2/(z_\alpha + z_\beta)^2$, where z_α is the upper α quantile of the standard normal distribution. Let $S_t = \frac{1}{\sqrt{m}\sigma}\sum_{i=1}^n X_i$, where $t = n/m$. Then S_t emulates a Brownian motion for $0 \le t \le 1$. For a sequential conditional probability ratio test (Xiong 1995, Tan et al 1998, Xiong et al. 2001), the upper boundary is $a_{t_k} = z_\alpha + \sqrt{2at_k(1-t_k)}$ and the lower boundary is $b_{t_k} = z_\alpha - \sqrt{2at_k(1-t_k)}$. The stopping region for (T, S_T) is $\mathcal{B} = \{(t_k, s) : \ s \le b_{t_k}$ or $s \ge a_{t_k}$ for $k = 1, 2, \cdots, K\}$, and the continuation region is $\mathcal{C} = \{(t, s) : t = t_k, \ b_{t_k} < s < a_{t_k}, k = 1, 2, \cdots, K\}$, with $t_K = 1$ and $a_K = b_K$. The test statistic for the nonsequential test is S_{t^*} where $t^* = 1$ and the cutoff value for S_{t^*} is z_α. Assume $\alpha = 0.05$ and $1-\beta = 0.8$ and $K = 4$ and $a = 2.953$ and $t_1, \cdots, t_4 = 0.25, 0.5, 0.75, 1$. Then upper boundaries are $a_{t_1}, \cdots, a_{t_4} = 1.461, 2.035, 2.284, 1.645$ and the lower boundaries are $b_{t_1}, \cdots, b_{t_4} = -0.639, -0.390, 0.184, 1.645$. The maximum conditional discordance probability is $\rho = 0.02$; the maximum discordance probability is $\rho_{max} = 0.0054$. The actual significance level is $\alpha = 0.051$ and actual power is $1-\beta = 0.80$. The expected sample size is $0.746m$ when $\mu = \mu_0$, and is $0.856m$ when $\mu = \mu_a$.

REFERENCES

Aroian, L.A. (1968) Sequential analysis, direct method. *Technometrics*, **10**, 125-132.

Armitage, P., McPherson, C. K., and Rowe, B. C. (1969) Repeated significance tests on accumulating data, *Journal of the Royal Statistical Society*, Ser. **A, 132**, 235-244.

Siegmund, D. (1985). *Sequential Analysis.* Springer-Verlag, New York.

Tan, M., Xiong, X., and Kutner, M.H. (1998). Clinical trial designs based on sequential conditional probability ratio tests and Reverse Stochastic Curtailing. *Biometrics*, **54**, 682-695.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, **73**, 573-581.

Xiong, X. (1991). Sequential Tests for Hypergeometric Distributions. *Ph.D. Dissertation, Purdue University, West Lafayette, Indiana.*

Xiong, X. (1995). A class of sequential conditional probability ratio tests. *Journal of American Statistical Association*, **90**, 1463-1473.

Xiong, X. (1996). Absorption probability distributions of random paths from finite populations. *Sequential Analysis*, **15**, 1-19.

Xiong, X., Tan, M., and Kutner, M.H. (2000). Computational methods for evaluating sequential tests and post-test estimation via sufficiency principle. *Statistica Sinica* (In revision).

Xiong, X., Tan, M., and Boyett, J. (2001). Sequential conditional probability ratio tests for normalized test statistic on information time. *Biometrics* (Submitted).