JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume 3, June 1990

# Optimal Sequential Tests which minimize the Average Sample Size

SUNG LAI KIM

ABSTRACT. For testing a hypothesis $H : \theta = \theta_1$, vs $A : \theta = \theta_2$ $(\theta_1 < \theta_2)$, we obtain a truncated sequential bayes procedure which minimizes the average sample size between $\theta_1$ and $\theta_2$.

## 1. Introduction

Let $X_1, X_2, \dots$ be a sequence of random variables (not necessary idd) with joint distribution depending on a real parameter $\theta \in \Theta$. Let $X = (X_1, X_2, \dots, X_n)$ at $x^n = (x_1, x_2, \dots, x_n)$ have a density $P_{n,\theta}(x^n$ w.r.t. $\mu_n(dx^n)$. Now suppose $\theta = \theta_1$ is to be tested against $\theta = \theta_2$ where $\theta_1 < \theta_2$. Let the error probabilities of any test $\delta$ be $\alpha_i(\delta) = P_{\theta_i}$ ($\delta$ reject $\theta_i$), $i = 1, 2$. Then we can find examples that $E_\theta N$ for SPRT(Sequential Probability Ratio Test) is everywhere less than the sample size for fixed sample size procedure having the same $\alpha_i (i = 1, 2)$ and also there are examples that $\max_{\theta_1 \leq \theta \leq \theta_2} E_\theta N$ for SPRT is large than the sample size for fixed sample size procedure with the same $\alpha_i (i = 1, 2)$. So there naturally arised the problem to find a procedure which minimizes $\max E_\theta N$.

By the optimum properties of the SPRT, given a SPRT $\delta_0$ with stopping bounds $(B, A)$, $(B \leq 1 \leq A)$ and given $\lambda = (\lambda^1, \lambda^2)$ with $0 < \lambda^i < 1$ $(i = 1, 2)$, the SPRT $\delta_0$ is the sequential Bayes procedure $(\lambda_1, \iota_1, \iota_2)$ for some $\iota_1, \iota_2 > 0$. Therefore we can restricted the procedures in the sequential Bayes procedures and furthermore if a sequential Bayes procedure is truncated, then we can easily find the sequential Bayes procedure using the method of backward induction.

Several studies were done trying to minimize $\max_{\theta_1 \leq \theta \leq \theta_2} E_\theta N$. In this paper, we are interested in minimizing the average sample size between $\theta_1$ and $\theta_2$ instead of minimizing $\max_\theta E_\theta N$. Throughout this

Received by the editors on 25 June 1990.
1980 *Mathematics subject classifications*: Primary 62L10.

paper, let $(\mathfrak{X}, \mathfrak{A}, P_\theta : \theta \in \Theta)$ be a probability space where $\Theta \subset R$ and let $\lambda$ be a prior distribution on $(\Theta, \mathfrak{B})$ where $\mathfrak{B}$ is a $\sigma$-field on $\Theta$. We use the following notations;
$\mathfrak{A}$: actions space, $L : \Theta \times \mathfrak{A} \to R$ nonnegative finite loss functionn, $\mathfrak{D} : \mathfrak{X} \to \mathfrak{A}$ the set of decision functions. $R_\delta$ (or $R(\delta, \lambda)$) ; the Bayes Risk of $\delta \in \mathfrak{D}$.

## 2. Sequential Bayes procedures

LEMMA 1. *Let $P = \{P_\theta : \theta \in \Theta\}$ be a dominated family of probability distribution over $(\mathfrak{X}, \mathfrak{A})$. For testing the hypothesis $H : \theta = \theta_1$ vs $A : \theta = \theta_2$ $(\theta_1 < \theta_2)$, put $\alpha_i(\delta) = P_{\theta_i}$ (take wrong decision $|\delta$) $i = 1, 2$ and put $\nu(\delta) = E_{\theta_0}(N|\delta)$ for some $\theta_1 < \theta_0 < \theta_2$. Then for a given $0 < \alpha_i < 1$, $i = 1, 2$, there is a sequential Bayes procedure $\delta_\lambda$ having $\alpha_i(\delta_\lambda) = \alpha_i (i = 1, 2)$ which minimizes $\nu(\delta)$ among all procedures $\delta \in \mathfrak{D}$ with $\alpha_i(\delta) \le \alpha_i$, $i = 1, 2$.*

PROOF: Let $\lambda = (\lambda^1, \lambda^0, \lambda^2)$ be a prior distribution on $\{\theta_1, \theta_0, \theta_2\}$ with $\lambda^i > 0$, $i = 0, 1, 2$. Let

$$L(\theta_i, a_j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } i = j,\ a_j \in \mathfrak{D} \\ 1 & \text{otherwise} \end{cases}$$

and let the cost per observation be constant, say $c > 0$ if $\theta = \theta_0$ and $c = 0$ if $\theta = \theta_1$ or $\theta = \theta_2$.

$$R(\delta, \lambda) = \lambda^1 \alpha_1(\delta) + \lambda^2 \alpha_2(\delta) + \lambda^0 c \nu(\delta) \quad \forall \delta \in \mathfrak{D}.$$

Since $\delta_\lambda$ is Bayes $(\lambda)$, $R(\delta, \lambda) \ge R(\delta_\lambda, \lambda)\, \forall\, \delta \in \mathfrak{D}$. Therefore $\lambda^0 c[\nu(\delta) - \nu(\delta_\lambda)] \ge \sum_{i=1}^{2} \lambda^i [\alpha_i(\delta_\lambda) - \alpha_i(\delta)] \ge 0$. So $\min_{\delta \in \mathfrak{D}} \nu(\delta) = \nu(\delta_\lambda)$.

THEOREM 2. *Let $(X_1, X_2, \ldots, X_n)$ have a p.d.f. $P_{n,\theta_i}(x^n)$ w.r.t $\mu_n(x^n)$ under $\theta_i \in \Theta$ and $P_{n,\theta_i}(x^n) \ll P_{n,\theta_0}(x^n)\ \forall\ n$, $i = 1, 2$. Let $\lambda = (\lambda^1, \lambda^0, \lambda^2)$ be a prior distribution on $\{\theta_1, \theta_0, \theta_2\}$ with $\lambda^i > 0$, $i = 0, 1, 2$. If there exists a converging to zero sequence of constants $\{b_n^0\}$ such that $\sum_{i=1}^{2} \lambda^i L(\theta_i, d_n^0(x^n)) P_{n,\theta_i}(x^n) / P_{n,\theta_0}(x^n) \le b_n^0$ where $d_n^0(x^n)$ is chosen to minimize $\sum_{i=1}^{2} \lambda^i L(\theta_i, d_n(x^n)) P_{n,\theta_i}(x^n)$ among all*

$d_n(x^n)$. *Then the sequential Bayes procedure* $\delta_\lambda$ *to test* $\theta = \theta_1$ vs $\theta_2$ *is truncated.*

PROOF: Let cost per observation be $c(>0)$ if $\theta = \theta_0$ and zero if $\theta = \theta_1$ or $\theta_2$ and let $L(\theta_i, a_j) = 0$ if $i = 0$ or $i = j$ and equals to one otherwise where $a_j$ is the action to accept $\theta_j$. Define the stopping rule as $\{\alpha_0, \alpha_1(x^1), \alpha_2(x^2), \dots\}$ s.t. $\sum_0^\infty \alpha_n(\omega) = 1$ where $\omega = (x_1, x_2, \dots)$ and terminal decision rule $\{d_n(x^n)\}$ values in $\{a_1, a_2\}$. Then for $d_n = d_n(x^n) = d_n^0$,

$$\begin{aligned} R(\delta,\lambda) &= \sum_{n=0}^{\infty} \int \alpha_n(x^n)[\lambda^0 nc \cdot P_{n_0}(x^n) \\ &\quad + \sum_{i=1}^{2} \lambda^i L(\theta_i, d_n^0) P_{n,\theta_i}(x^n)]\mu_n(dx^n) \\ &= \sum_{n=0}^{\infty} \int \alpha_n(x^n)[\lambda^0 nc + b_n(x^n)] P_{n,\theta_0}(x^n)\mu_n(dx^n) \\ &= \sum_{n=0}^{\infty} E_{\theta_0}\alpha_n(\lambda^n)[\lambda^0 nc + b_n(X^n)] \end{aligned}$$

where $b_n(x^n) = \sum_{i=1}^{2} \lambda^i L(\theta_i, d_n^0) P_{n,\theta_i}(x^n)/P_{n,\theta_0}(x^n)$. Put

$$E_{\theta_0}\alpha_n(X^n)b_n(X^n)/E_{\theta_0}\alpha_n(X^n) \equiv \bar{b}_n,$$

then $\bar{b}_n \le b_n^0$ $\forall n$. We have $R(\delta,\lambda) = \sum_{n=0}^{\infty} E_{\theta_0}\alpha_n(X^n)[\lambda^0 nc + \bar{b}_n] = \sum_{n=0}^{\infty} \beta_n(\lambda^0 nc + \bar{b}_n)$ where $\beta_n = E_{\theta_0}\alpha_n(X^n)$. Let $n_0$ be s.t. $b_n^0 < \lambda^0 c$ for all $n \ge n_0$. Then the sequential Bayes procedure $\delta_\lambda$ must have $\beta_n = 0$ for $n > n_0$. So $\alpha_n(x^n) = 0$ a.e. $P_{\theta_0}$, so does a.e. $P_{\theta_i}$.

THEOREM 3. *Let* $\{X_i\}$ *be a sequence of random variables defined on* $(\mathfrak{X}, \mathfrak{A}, P_\theta, \theta \in \Theta \subset R)$ *and let* $X^n = (X_1, X_2, \dots, X_n)$ *have a p.d.f.* $P_{n,\theta}(x^n)$ w.r.t $\mu_n(dx^n)$ at $x^n = (x_1, x_2, \dots, x_n)$. *For testing* $H : \theta = \theta_1$ vs $A : \theta = \theta_2$ $(\theta_1 < \theta_2)$ *with error probabilities* $\alpha_i$ $(0 < \alpha_i < 1)$ $i = 1, 2$.

*If there is a sequence of constants* $\{b_n^0\}$ *with* $b_n^0 \to 0$ *as* $n \to \infty$ *such that* $\min_i P_{n,\theta_i}(x^n)\cdot(\theta_2-\theta_1)/\int_{\theta_1}^{\theta_2} P_{n,\theta}(x^n)\,d\theta \le b_n^0\ \forall n$. *Then there is a truncated sequential Bayes procedure* $\delta_\lambda$ *such that* $\int_{\theta_1}^{\theta_2} E_\theta(N|\delta_\lambda)\,d\theta = \min_{\delta\in\mathfrak{D}} \int_{\theta_1}^{\theta_2} E_\theta(N|\delta)\,d\theta$ *for all* $\alpha_i(\delta) = P_\theta(\delta \text{ reject } \theta_i) \le \alpha_i$, $i = 1,2$.

PROOF: Define a prior distribution on $\Theta$ as follows; $\lambda(\theta_i) = \lambda^i > 0$, $i = 1,2$, $\lambda = 0$ if $\theta < \theta_1$ or $\theta > \theta_2$ and $\lambda^0\,d\theta$ on interval $(\theta_1,\theta_2)$ with $\lambda^1+\lambda^2+\lambda^0(\theta_2-\theta_1) = 1$. Assume that the loss equals to one for wrong terminal decision, and cost per observation equals to one for each $\theta$, $\theta_1 < \theta < \theta_2$ but no cost if $\theta = \theta_1$ or $\theta_2$. Let $\{\alpha_n\}$ be a sequence of stopping rule. Then

$$R(\delta,\lambda) = \sum_{n=0}^{\infty} \alpha_n(x^n) \int \left\{ \sum_{i=1}^{2} \lambda^i L(\theta_i, d_n(x^n)) P_{n,\theta_i}(x^n) \right.$$
$$\left. + n\lambda^0 \int_{\theta_1}^{\theta^2} P_{n,\theta}(x^n)\,d\theta \right\} \mu_n(dx^n).$$

Observe that the best determinal decision $d_n(x^n) = \theta_1$ if $\lambda^1 P_{n,\theta_1}(x^n) > \lambda^2 P_{n,\theta_2}(x^n)$, and equals to $\theta_2$ if $\lambda^1 P_{n,\theta_1}(x^n) < \lambda^2.\ P_{n,\theta_2}(x^n)$. So for best decision rule $R(\delta,\lambda) = \sum_{n=0}^{\infty} \alpha_n(x^n) \int\{\min_i \lambda^i \cdot P_{n,\theta_i}(x^n) + n\lambda^0 \int_{\theta_1}^{\theta_2} P_{n,\theta}(x^n)\,d\theta\}\mu_n(dx^n)$. Let $\overline{P}_n(x^n) = \frac{1}{\theta_2-\theta_1}\int_{\theta_1}^{\theta_2} P_{n,\theta}(x^n)\,d\theta$, then $\overline{P}_n(x^n)$ is a density w.r.t $\mu_n(dx^n)$ and

$$R(\delta,\lambda) = \sum_{n=0}^{\infty} \alpha_n(x^n) \int \overline{P}_n(x^n)\{n(\theta_2-\theta_1)\lambda^0$$
$$+ \min_i \lambda^i P_{n,\theta_i}(x^n)/\overline{P}_n(x^n)\}\mu_n(dx^n).$$

Since $\min_i P_{n,\theta_i}(x^n)/\overline{P}_n(x^n) \le b_n^0$ which converges to zero, there exists a truncated sequential Bayes procedure $\delta_\lambda$ which minimizes the average sample size by the theorem 2.

## REFERENCES

1. Burkholder, D.L., Wijsman, R.A., *Optimum properties and admissibility of sequential tests*, Ann. Math. Statist. **34** (1963), 1–18.

2. Govindarajulu, Z., "The sequential statistical analysis," Amer. Science Press, Inc., 1981.
3. Lehmann, E.L, "Testing statistical hypotheses," Wiley, New-York, 1959.
4. Lai, T.L., *Optimum stopping and sequential tests which minimize the maximum expected sample size*, Ann. Statist. **1** (1973), 659–673.
5. Matthes, T.K., *On the optimality of sequential probability ratio tests*, Ann. Math. Statist **34** (1963), 18–21.
6. Shapiro, C.P., Wardrop, R.L., *Bayes sequential estimation for one parameter exponential families*, J. Amer. Statist. Assoc. **75** (1981), 984–988.
7. Siegmund, D.O., "Sequential Analysis," Springer-Verlag, New-York, 1985.
8. Weiss, L., *On sequential tests which minimize the maximum expected sample size*, J. Amer. Statist. Assoc. **57** (1962), 551–557.

Department of Mathematics
Chungnam National University
Taejeon, 305-764, Korea