A CLASS OF STOPPING RULES FOR TESTING PARAMETRIC HYPOTHESES

HERBERT ROBBINS and DAVID SIEGMUND
COLUMBIA UNIVERSITY and HEBREW UNIVERSITY

Let $f_\theta(x)$, $\theta \in \Omega$, be a one parameter family of probability densities with respect to some σ-finite measure μ on the Borel sets of the line. Denote by P_θ the probability measure under which random variables $x_1, x_2, \cdots$ are independent with the common probability density $f_\theta(x)$. Let θ_0 be an arbitrary fixed element of Ω and ε any constant between 0 and 1. We are interested in finding stopping rules N for the sequence $x_1, x_2, \cdots$ such that

$$P_\theta(N < \infty) \leqq \varepsilon \qquad \text{for every } \theta \leqq \theta_0, \tag{1}$$

and

$$P_\theta(N < \infty) = 1 \qquad \text{for every } \theta > \theta_0. \tag{2}$$

Among such rules, we wish to find those which in some sense minimize $E_\theta(N)$ for all $\theta > \theta_0$.

A method of constructing rules which satisfy (1) and (2) by using mixtures of likelihood ratios was given in [3]. Here we sketch an alternative method.

Let $\theta_{n+1} = \theta_{n+1}(x_1, \cdots, x_n)$ for $n = 0, 1, 2, \cdots$, be any sequence of Borel measurable functions of the indicated variables such that

$$\theta_{n+1} \geqq \theta_0. \tag{3}$$

In particular, θ_1 is some constant $\geqq \theta_0$. Define

$$z_n = \prod_1^n \frac{f_{\theta_i}(x_i)}{f_{\theta_0}(x_i)}, \qquad n = 1, 2, \cdots, \tag{4}$$

and for any constant $b > 0$, let

$$N = \begin{cases} \text{first } n \geqq 1 \text{ such that } z_n \geqq b, \\ \infty \text{ if no such } n \text{ occurs.} \end{cases} \tag{5}$$

We shall show that under a certain very general assumption on the structure of the family $f_\theta(x)$, the inequality (1) holds at least for all $b \geqq 1/\varepsilon$.

ASSUMPTION. *For every triple $\alpha \leqq \gamma \leqq \beta$ in Ω,*

$$\int \frac{f_\alpha(x) f_\beta(x)}{f_\gamma(x)}\, d\mu(x) \leqq 1. \tag{6}$$

Research supported by Public Health Service Grant No. 1-R01-GM-16895-03.

We remark without proof that this holds for the general one parameter Koopman-Darmois-Pitman exponential family and many others.

Denote by $\mathfrak{F}_n$ the Borel field generated by $x_1, \cdots, x_n$. *Then for each fixed* $\theta \leqq \theta_0$, $\{z_n, \mathfrak{F}_n, P_\theta; n \geqq 1\}$ *is a nonnegative supermartingale sequence.* For, given any $n \geqq 1$,

$$(7) \qquad E_\theta(z_{n+1}|\mathfrak{F}_n) = z_n E_\theta\left(\frac{f_{\theta_{n+1}}(x_{n+1})}{f_{\theta_0}(x_{n+1})}\bigg|\mathfrak{F}_n\right) = z_n \int \frac{f_\theta(x) f_{\theta_{n+1}}(x)}{f_{\theta_0}(x)}\, d\mu(x) \leqq z_n,$$

since by hypothesis $\theta \leqq \theta_0 \leqq \theta_{n+1}$. We can therefore apply the following.

LEMMA. *Let* $\{z_n, \mathfrak{F}_n, P; n \geqq 1\}$ *be any nonnegative supermartingale. Then for any constant* $b > 0$,

$$(8) \qquad P(z_n \geqq b \text{ for some } n \geqq 1) \leqq P(z_1 \geqq b) + \frac{1}{b}\int_{(z_1 < b)} z_2\, dP \leqq \frac{E(z_1)}{b}.$$

PROOF. Defining N by (5), we have

$$(9) \qquad P(z_n \geqq b \text{ for some } n \geqq 1) = P(z_1 \geqq b) + P(1 < N < \infty).$$

Since z_n is a nonnegative supermartingale,

$$(10) \qquad \int_{(N>1)} z_1\, dP \geqq \int_{(N>1)} z_2\, dP = \int_{(N=2)} z_2\, dP + \int_{(N>2)} z_2\, dP \geqq \cdots \geqq \sum_{i=2}^{n} \int_{(N=i)} z_i\, dP + \int_{(N>n)} z_n\, dP \geqq bP(1 < N \leqq n) + 0,$$

because $z_i \geqq b$ on $(N = i)$ and $z_n \geqq 0$. Since n is arbitrary,

$$(11) \qquad P(1 < N < \infty) \leqq \frac{1}{b}\int_{(z_1 < b)} z_2\, dP,$$

and hence from (9)

$$(12) \qquad P(z_n \geqq b \text{ for some } n \geqq 1) \leqq P(z_1 \geqq b) + \frac{1}{b}\int_{(z_1 < b)} z_2\, dP \leqq \frac{1}{b}\int_{(z_1 \geqq b)} z_1\, dP + \frac{1}{b}\int_{(z_1 < b)} z_1\, dP = \frac{E(z_1)}{b},$$

which proves (8).

Applying this lemma to (4) and (5), we see that for each fixed $\theta \leqq \theta_0$,

$$(13) \qquad P_\theta(N < \infty) \leqq P_\theta(z_1 \geqq b) + \frac{1}{b}\int_{(z_1 < b)} z_2\, dP_\theta \leqq \frac{E_\theta(z_1)}{b} = \frac{1}{b}\int \frac{f_\theta(x) f_{\theta_1}(x)}{f_{\theta_0}(x)}\, d\mu(x) \leqq \frac{1}{b},$$

and hence, as claimed above, (1) holds at least for $b \geqq 1/\varepsilon$.

As an example, suppose that under P_θ the x are $N(\theta, 1)$, so that $f_\theta(x) = \varphi(x - \theta)$, where $\varphi(x)$ is the standard normal density, and that $\theta_0 = 0$. It is easily seen that if $\theta_1 > 0$ then

$$(14)\qquad z_n = \prod_1^n \exp\left\{\theta_i x_i - \frac{\theta_i^2}{2}\right\}, \qquad E_\theta(z_1) = \exp\{\theta\theta_1\},$$
$$P_\theta(z_1 \geqq b) = \Phi\left(\theta - \frac{\log b}{\theta_1} - \frac{\theta_1}{2}\right),$$

$$(15)\qquad \int_{(z_1<b)} z_2\, dP_\theta = \int_{-\infty}^{\log b/\theta_1+\theta_1/2} \int_{-\infty}^{\infty} z_2\varphi(x_2 - \theta)\varphi(x_1 - \theta)\, dx_2\, dx_1$$
$$\leqq \exp\{\theta\theta_1\}\, \Phi\left(\frac{\log b}{\theta_1} - \frac{\theta_1}{2} - \theta\right),$$

where $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt$. Hence, (13) gives for any $\theta \leqq 0$, the inequality

$$(16)\qquad P_\theta\left(\prod_1^n \exp\left\{\theta_i x_i - \frac{\theta_i^2}{2}\right\} \geqq b \text{ for some } n \geqq 1\right)$$
$$\leqq \Phi\left(\theta - \frac{\log b}{\theta_1} - \frac{\theta_1}{2}\right) + \frac{1}{b}\exp\{\theta\theta_1\}\, \Phi\left(\frac{\log b}{\theta_1} - \frac{\theta_1}{2} - \theta\right)$$
$$\leqq \frac{1}{b}\exp\{\theta\theta_1\}.$$

The middle term of (16) is increasing in θ, so

$$(17)\qquad P_\theta\left(\prod_1^n \exp\left\{\theta_i x_i - \frac{\theta_i^2}{2}\right\} \geqq b \text{ for some } n \geqq 1\right)$$
$$\leqq \Phi\left(-\frac{\log b}{\theta_1} - \frac{\theta_1}{2}\right) + \frac{1}{b}\Phi\left(\frac{\log b}{\theta_1} - \frac{\theta_1}{2}\right) \leqq \frac{1}{b}$$

for every $\theta \leqq 0$.

We shall now suppose that in addition to the requirement that $\theta_{n+1} = \theta_{n+1}(x_1, \cdots, x_n) \geqq 0$, *the sequence* θ_n *converges to* θ *with probability* 1 *under* P_θ *for each* $\theta > 0$. For example, both

$$(18)\qquad \theta_{n+1} = \frac{\max(0, s_n)}{n}$$

and

$$(19)\qquad \theta_{n+1} = \frac{s_n}{n} + \frac{\varphi(s_n/\sqrt{n})}{\sqrt{n}\Phi(s_n/\sqrt{n})},$$

where $s_n = x_1 + \cdots + x_n$, have this desired property (equation (19) is the posterior expected value of θ given $x_1, \cdots, x_n$ when the prior distribution of θ is flat for $\theta > 0$). Thus, for large n,

$$(20)\qquad z_n = \prod_1^n \exp\left\{\theta_i x_i - \frac{\theta_i^2}{2}\right\} \approx \prod_1^n \exp\left\{\theta x_i - \frac{\theta^2}{2}\right\} = \exp\left\{\theta s_n - \frac{n\theta^2}{2}\right\} = z_n(\theta),$$

say. Now it has been remarked elsewhere [2], and a proof based on [1], pp. 107–108, is easily given, that for any fixed $\theta > 0$,

$$(21)\qquad N_{\theta,b} = \begin{cases} \text{first } n \geqq 1 \text{ such that } z_n(\theta) \geqq b, \\ \infty \text{ if no such } n \text{ occurs,} \end{cases}$$

is optimal in the sense that if T is any stopping rule of $x_1, x_2, \cdots$ such that

$$P_0(T < \infty) \leqq P_0(N_{\theta,b} < \infty), \tag{22}$$

then $E_\theta(N_{\theta,b}) < \infty$ and $E_\theta(T) \geqq E_\theta(N_{\theta,b})$. Thus, the N using (18) or (19) may be expected to be "almost optimal" simultaneously for all values $\theta > 0$. Monte Carlo methods will be needed to get accurate estimates of $P_0(N < \infty)$ and $E_\theta(N)$ for $\theta > 0$. We have, however, been able to find the asymptotic nature of $E_\theta(N)$ as $\theta \to 0$ or $b \to \infty$ in the normal and other cases for various choices of the θ_n sequence, and the results will be published elsewhere. For example, using (18), we can show that, for $\theta > 0$,

$$E_\theta(N) \sim P_0(N = \infty) \left(\log \frac{1}{\theta} \Big/ \theta^2\right) \qquad \text{as } \theta \to 0, \tag{23}$$

and

$$E_\theta(N) = \frac{2 \log b + \log_2 b}{\theta^2} + o(\log_2 b) \qquad \text{as } b \to \infty. \tag{24}$$

By putting

$$\theta_{n+1} = \begin{cases} \dfrac{s_n}{n} & \text{if } s_n \geqq [n(2 \log_2^+ n + 3 \log_3^+ n)]^{1/2}, \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

where $\log_2 n = \log(\log n)$, and so on, equation (23) is replaced by

$$E_\theta(N) \sim 2P_0(N = \infty) \log_2 \frac{1}{\theta} \Big/ \theta^2 \qquad \text{as } \theta \to 0, \tag{26}$$

which is optimal for $\theta \to 0$.

In evaluating $P_\theta(N < \infty)$ for $\theta \leqq 0$ with an arbitrary sequence $\theta_{n+1} = \theta_{n+1}(x_1, \cdots, x_n) \geqq 0$, $n = 0, 1, 2, \cdots$, and $b > 1$, we see that this probability is equal to

(27)

$$P_\theta \left(\prod_1^n \exp\left\{\theta_i x_i - \frac{\theta_i^2}{2}\right\} \geqq b \text{ for some } n \geqq 1\right) = \sum_{n=1}^{\infty} \int_{(N=n)} \exp\left\{\theta s_n - \frac{n\theta^2}{2}\right\} dP_0.$$

For any fixed x and n the function $f(\theta) = \exp\{\theta x - n\theta^2/2\}$ is increasing for $-\infty < \theta < x/n$. Hence if the condition

$$s_n > 0 \quad \text{whenever} \quad N = n, \qquad n = 1, 2, \cdots, \tag{28}$$

is satisfied, then $P_\theta(N < \infty)$ will be an increasing function of $\theta \leqq 0$ (as is the middle term of (16)). Recalling that

$$N = \begin{cases} \text{first } n \geqq 1 \text{ such that } \sum_1^n \left(\theta_i x_i - \dfrac{\theta_i^2}{2}\right) \geqq \log b, \\ \infty \text{ if no such } n \text{ occurs,} \end{cases} \tag{29}$$

we see that if $N = 1$, then $\theta_1 x_1 \geqq \log b + \theta_1^2/2$ so $s_1 = x_1 > 0$, while if $N = n > 1$, then

$$\sum_1^{n-1} \theta_i x_i < \log b + \frac{1}{2}\sum_1^{n-1} \theta_i^2,$$
$$\sum_1^{n} \theta_i x_i \geqq \log b + \frac{1}{2}\sum_1^{n} \theta_i^2, \tag{30}$$

so $\theta_n x_n > 0$, and hence $\theta_n > 0$ and $x_n > 0$. In cases (18) and (25), it follows that $s_{n-1} \geqq 0$, and hence $s_n = s_{n-1} + x_n > 0$. Thus, $P_\theta(N < \infty)$ is an increasing function of $\theta \leqq 0$ in these cases. Whether this is true for the choice (19) we do not know. Likewise, we do not know whether $P_\theta(N \leqq n)$ is an increasing function of θ for each fixed $n = 1, 2, \cdots$, even for (18) or (25). For $\theta > 0$, $P_\theta(N < \infty) = 1$ and $E_\theta(N) < \infty$ in all three cases.

In the case of a general parametric family $f_\theta(x)$, we can try to make $E_\theta(N)$ small for $\theta > \theta_0$ by choosing θ_n to converge properly to θ under P_θ for $\theta > \theta_0$, but a comparison with the methods of [3] remains to be made. The present method of sequentially estimating the true value of θ when it is $>\theta_0$ appears somewhat more natural in statistical problems.

If we do not wish to take advantage of the property (6), we can use, instead of (4),

$$z_n' = \prod_1^n \frac{f_{\theta_i}(x_i)}{h_n}, \tag{31}$$

where $h_n = h_n(x_1, \cdots, x_n) = \sup_{\theta \leqq \theta_0} \{\prod_1^n f_\theta(x_i)\}$. The use of (31) has been independently suggested by Edward Paulson. For $\theta \leqq \theta_0$, we then have

$$P_\theta(z_n' \geqq b \text{ for some } n \geqq 1) \leqq P_\theta\left(\prod_1^n \frac{f_{\theta_i}(x_i)}{f_\theta(x_i)} \geqq b \text{ for some } n \geqq 1\right) \leqq \frac{1}{b}, \tag{32}$$

by the lemma above. It would seem, however, that (31) should be less efficient than (4) when the assumption (6) holds.

REFERENCES

[1] Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Boston, Houghton-Mifflin, 1971.

[2] D. A. Darling and H. Robbins, "Some further remarks on inequalities for sample sums," *Proc. Nat. Acad. Sci.*, Vol. 60 (1968), pp. 1175–1182 (see p. 1181).

[3] H. Robbins, "Statistical methods related to the law of the iterated logarithm," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 1397–1409.