The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 313–315

A SECOND-ORDER APPROXIMATION TO OPTIMAL SAMPLING REGIONS[1]

BY GIDEON SCHWARZ

University of California, Berkeley, and Hebrew University

1. Introduction. In an earlier paper [3] an asymptotic description of the optimal sequential testing regions for separated hypotheses was given. It involved an asymptotic formula for curves of constant posterior risk. The need for a second-order correction term for this formula was demonstrated by Fushimi [1], who found such a term for normal and binomial sampling distributions, with their conjugate *a priori* distributions, and a truncated linear loss function. In this paper we find the general correction term. While it does not depend on the sampling distributions as long as the latter form an exponential family, it does vary with the loss function, and depends also on some properties of the *a priori* distribution which carry over into the *a posteriori* distributions: the locations of its atoms and the zeros of its density.

2. Preliminaries and statement. We assume a sampling distribution with density $f(x, \theta) = e^{\theta x - b(\theta)}$ with respect to some measure. The parameter θ ranges over the interior Θ of the natural parameter space. There $b(\theta)$ has all derivatives, and since $b'(\theta)$ and $b''(\theta)$ are the expectation and variance of the sampling variable, $b(\theta)$ is a convex function. For expository convenience we assume "one sided" hypotheses H_0: $\theta \leqq M'$, H_1: $\theta \geqq M > M'$ and an *a priori* distribution W that dominates Lebesgue measure on Θ. The local behaviour of W at θ is described by a number $-1 \leqq \tau(\theta) < \infty$, as follows:

(a) if $W(\{\theta_0\}) > 0$, $\tau(\theta_0) = -1$;

(b) if, in a neighbourhood of θ_0, W has a density of the form $|\theta - \theta_0|^\alpha g(\theta)$ with $g(\theta)$ bounded away from 0 and ∞, $\tau(\theta_0) = \alpha$.

We assume the τ is defined at every $\theta \,\varepsilon\, \Theta$ either by (a) or by (b). This restricts the generality of the *a priori* distributions somewhat, but it seems general enough for any conceivable application.

The loss for deciding "H_0" when the true parameter value is θ, is given on H_1 by $l(\theta) = (\theta - M)^\eta d(\theta)$ with d bounded away from 0 and ∞. If $\tau(M) > -1$, we assume $\eta + \tau(M) > -1$, to avoid infinite Bayes risks; if $\tau(M) = -1$ any bounded loss function would do, but we assume in this case that it is positive at M, and define $\eta = 0$ independently of the behaviour of $l(\theta)$ near M.

The posterior risk of deciding "H_0" after having made n observations with sum $S_n = kn$ is

$$(*) \qquad R_0(n, S_n) = \int_{\theta \geqq M} l(\theta)\, e^{(k\theta - b(\theta))n}\, dW(\theta) \Big/ \int_\Theta e^{(k\theta - b(\theta))n}\, dW(\theta).$$

Received 18 July 1968.

[1] This paper was prepared with the partial support of National Science Foundation Grant GP-7454.

We shall study the asymptotic behaviour of R_0 as n and S_n tend to infinity and $k = n^{-1}S_n$ is fixed and less then $b'(M)$, and obtain:

THEOREM. *For $k < b'(M)$ and $r \to 0$, the solution of the equation $R_0(n, kn) = r$ is given by*

$$n = (\log r^{-1} - (\tfrac{1}{2} + \eta + \tau(M) - \tfrac{1}{2}\tau(\theta(k))) \log\log r^{-1}) \log \lambda_0(k) + O(1),$$

where $\theta(k)$ is defined by $b'(\theta(k)) = k$, and $\lambda_0(k)$ is the generalized likelihood ratio statistic

$$(\sup_\Theta / \sup_{H_1})e^{k\theta - b(\theta)}.$$

3. Proof of the theorem. We first state and prove some lemmas, essentially about the asymptotic behaviour of L_p norms as $p \to \infty$.

LEMMA 1. *Let h be a bounded integrable function on a measure space. If K is the essential supremum of h, and $L < K$, then, as $n \to \infty$*

$$\int h^n / \int_{h>L} h^n \to 1.$$

PROOF. As is well known ([2], p. 160), when $n \to \infty$, $(\int_{h>L} h^n)^{1/n} \to K$ and $(\int_{h\leqq L} h^n)^{1/n} \to L' \leqq L$; hence $(\int_{h\leqq L} h^n / \int_{h>L} h^n)^{1/n} \to L'/K < 1$ and $\int_{h\leqq L} h^n / \int_{h>L} h^n \to 0$. Now $\int h^n / \int_{h>L} h^n = 1 + \int_{h\leqq L} h^n / \int_{h>L} h^n$ implies the statement of the lemma.

LEMMA 2. *If $f'(\theta)$ is negative for $M \leqq \theta \in \Theta$, and $\rho > -1$, then, as $n \to \infty$,* $\log \int_{\theta \geqq M} (\theta - M)^\rho e^{nf(\theta)}\, d\theta = nf(M) - (\rho + 1) \log n + O(1)$.

PROOF. By the definition of derivative, if $A_1 > -f'(M) > A_2 > 0$, then $-A_1 < (f(\theta) - f(M))/(\theta - M) < -A_2$, and consequently

$$(\theta - M)^\rho e^{n(f(M) - A_1(\theta - M))} < (\theta - M)^\rho e^{nf(\theta)} < (\theta - M)^\rho e^{n(f(M) - A_2(\theta - M))},$$

for θ in some interval $(M, M + \epsilon]$. Integrating the left and right sides of the last inequality from M to ∞, we obtain $(A_i n)^{-(\rho+1)}\Gamma(\rho + 1)e^{nf(M)}$, $i = 1, 2$, which has the form stated by the lemma. According to Lemma 1, applied to the function e^f and the measure $(\theta - M)^\rho\, d\theta$, the behaviour of f outside $(M, M + \epsilon]$, where $f(\theta) < f(M + \epsilon) < f(M)$ by $f'(\theta) < 0$, does not affect the conclusion.

LEMMA 3. *If $f''(\theta) < 0$ on Θ, and f attains its maximum at $\theta = m \in \Theta$, then for $\rho > -1$,*

$$\log \int_\Theta |\theta - m|^\rho e^{nf(\theta)}\, d\theta = nf(m) - \tfrac{1}{2}(\rho + 1)\log n + O(1).$$

PROOF. We must have $f'(m) = 0$, and therefore, when $A_1 > -f''(m) > A_2 > 0$, Taylor's theorem implies $-A_1 < 2(f(\theta) - f(m))/(\theta - M)^{-2} < -B_1$, and consequently $|\theta - m|^\rho e^{n(f(m) - \frac{1}{2}A_1(\theta - m)^2)} < |\theta - m|^\rho e^{nf(\theta)} < |\theta - m|^\rho e^{n(f(m) - \frac{1}{2}A_2(\theta - m)^2)}$ for θ in some interval $[m - \epsilon, m + \epsilon]$. Integration of the left and right terms over the real line yields $(nA_i/2)^{-\frac{1}{2}(\rho+1)}\Gamma((\rho + 1)/2)e^{nf(m)}$, $i = 1, 2$, and since f decreases as θ moves away from m in either direction, Lemma 1 can be applied to obtain the statement of Lemma 3.

LEMMA 4. *If f achieves its maximum at m, and $W(\{m\}) > 0$, then, as $n \to \infty$*

$$\log \int e^{nf(\theta)}\, dW = nf(m) + O(1).$$

PROOF. From $f(\theta) \leqq f(m)$, we obtain $nf(\theta) \leqq nf(m) + (f(\theta) - f(m))$, and therefore

$$W(\{m\})e^{nf(m)} \leqq \int e^{nf(\theta)}\, dW \leqq e^{nf(m)} \int e^{f(\theta)-f(m)}\, dW,$$

which implies the statement of the lemma.

To prove the theorem, we note that $k\theta - b(\theta)$ is strictly convex, and attains its maximum at $\theta = \theta(k)$. For $k < b'(M)$, $\theta(k) < M$ holds, and $k\theta - b(\theta)$ has a negative derivative for $M < \theta \varepsilon \Theta$. We now apply Lemma 2 with $f(\theta) = k\theta - b(\theta)$ and $\rho = \eta + \tau(M)$ to the numerator of $(*)$, and Lemma 3 with the same f and $\rho = \tau(\theta(k))$ to the denominator. If $\tau(\theta(k)) = -1$ we apply Lemma 4 instead of 3. This way we obtain

$$\log R_0(n, kn) = -n \log \lambda(k) - (\tfrac{1}{2} + \eta + \tau(M) - \tfrac{1}{2}\tau(\theta(k)))\log n + O(1),$$

and the equation $R_0 = r$ yields, after some manipulation, the statement of the theorem.

4. Remarks and conclusion. (a) All calculations were made for the lower boundary. A formula for the upper boundary, valid for $k > b'(M')$, is completely analogous. In the overlap region $b'(M) < k < b'(M')$ the boundary closer to the origin is the one used.

(b) For $\tau \equiv 0$, $\eta \equiv 1$ and $b(\theta) = 1/2\theta^2$ or $b(\theta) = \log(1 + e^\theta)$ the theorem reproduces Fushimi's correction term [1].

(c) In the first-order approximation [3], all the continuation regions were convex. For the second order approximation this is only true for $\tau \equiv 0$. In the directions k corresponding to atoms or "zeros" of the *a priori* distributions, the continuation regions have "cusps" of length $O(\log \log r^{-1})$, pointing in or out, respectively.

(d) The first case for which the asymptotic shape was found is that of a three-point parameter space, with each hypothesis containing one point; it turned out to be a pentagon. In this, as in any case where W is discrete, Lemma 4 shows that the log log-term is zero: the pentagon is unaffected by the correction!

(e) According to Theorem I in [3], the Bayes regions are enclosed between constant-posterior-risk regions with risks c and $\sigma c \log c^{-1}$, where c is the cost of an observation, and σ is a positive constant. Now $\log (\sigma c \log c^{-1})^{-1} = \log c^{-1} - \log\log c^{-1} - \log \sigma$. Consequently, the bottle-neck of the search for approximations to the Bayes regions is now shifted from the problem of finding asymptotic formulas for the constant-risk regions, to the problem of improving Theorem I of [3], by finding closer constant-risk approximations to the Bayes regions.

REFERENCES

[1] FUSHIMI, M. (1967). On the rate of convergence of asymptotically optimal Bayes tests. *Rep. Stat. Appl. Res. JUSE* **14** 1–7.

[2] LOÈVE, M. (1963). *Probability Theory* (3rd Edition). Van Nostrand, Princeton.

[3] SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.