The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 2038–2043

ASYMPTOTIC SHAPES FOR SEQUENTIAL TESTING OF TRUNCATION PARAMETERS[1]

By Gideon Schwarz

University of California, Berkeley and Hebrew University, Jerusalem

1. Introduction. In an earlier paper [2] an asymptotic property of the Bayes sequential testing regions was proved for exponential families. With c, the cost of an observation, tending to zero, the regions, scaled down by a factor of $-\log c$, were shown to approach a limiting region. The limiting region depends on the *a priori* distribution only through its support, and is easily and explicitly described in terms of a modified maximum likelihood statistic. In this paper these results are extended to families with *truncation parameters*, that is, parameters that govern the range of the random variables.

The result in [2] is obtained by: (1) bounding the Bayes regions within and without by *constant a posteriori risk regions*, and (2) studying the asymptotic behaviour of the latter. The first part of the result is easily extended beyond exponential families, and this has been done by Kiefer and Sacks [1]. The second part is extended to truncation parameter families in this paper (Theorem 2). In order to make the paper self-contained, a simple proof of the extended first part is included (Theorem 1). A number of examples conclude the paper.

2. Truncation parameters. In a family $P(\cdot, \eta, \theta)$ of distributions, θ is a *truncation parameter* if it is real-valued, and if there exists a random variable T such that for $\theta_1 > \theta_2$, $P(\cdot, \eta, \theta_2)$ is obtained from $P(\cdot, \eta, \theta_1)$ by conditioning on the event $\{T \leqq \theta_2\}$. The families we are concerned with here depend on a finite number of parameters, some of which are truncation parameters, with the rest, if any, appearing as exponential parameters. Such a family we call an *exponential truncation family* $P(\cdot, \theta_1, \cdots, \theta_t, \eta_1, \cdots, \eta_s)$, characterized as follows: there exist on $(\Omega, \mathcal{B})$ a measure μ and $t + s$ random variables $T_1, \cdots T_t, Y_1, \cdots, Y_s$ such that the density of $P(\cdot, \theta_1, \cdots, \theta_t, \eta_1, \cdots, \eta_s)$ with respect to μ is given by

$$\exp\{\langle \boldsymbol{\eta} \cdot \mathbf{Y}\rangle - b(\boldsymbol{\theta}, \boldsymbol{\eta})\} \quad \text{when} \quad T_k \leqq \theta_k \quad \text{for} \quad k = 1, \cdots, t$$

and

$$0 \qquad\qquad \text{otherwise.}$$

Here $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_t)$, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_s)$, $\mathbf{Y} = (Y_1, \cdots, Y_s)$, $\langle\cdot\rangle$ stands for the dot product, and b is the real-valued function required to normalize the density. The *parameter space* Θ is the Borel set of all $(\boldsymbol{\theta}, \boldsymbol{\eta})$ for which such a b exists.

Received 4 January 1968.

[1] Part of this paper was prepared under contract Nonr-225(52) for the Office of Naval Research (U.S.A.).

For example: if Ω is the real line, μ is Lebesgue measure, $s = t = 2$, $T_1 = -\omega$, $T_2 = \omega$, $Y_1 = \omega$ and $Y_2 = \omega^2$, $P(\cdot, \theta_1, \theta_2, \eta_1, \eta_2)$ is a normal distribution with mean $-\eta_1/2\eta_2$ and variance $-1/2\eta_2$, truncated at $-\theta_1$ on the left and at θ_2 on the right. The parameter space in this example consists of all $(\theta_1, \theta_2, \eta_1, \eta_2)$ such that $\theta_1 < \theta_2$ and $0 < \eta_2$.

The family of normal distributions with unknown mean and variance, truncated at three standard deviations on either side of the mean, is not itself an exponential truncation family; it is, however, a subfamily of the previous example.

For n independent observations $\omega_1, \cdots, \omega_n$ from an exponential truncation family, the $s + t$ dimensional statistic

$$\sum_{i=1}^{n} Y_1(\omega_i), \cdots, \sum_{i=1}^{n} Y_s(\omega_i), \max_i T_1(\omega_i), \cdots, \max_i T_t(\omega_i),$$

or, for short,

$$(\sum_{i=1}^{n} \mathbf{Y}(\omega_i), \max_i \mathbf{T}(\omega_i)),$$

is easily seen to be sufficient. For our purposes, the equivalent statistic $(\sum_{i=1}^{n} \mathbf{Y}(\omega_i), n \max_i \mathbf{T}(\omega_i))$ will be more convenient. We denote it by $(\mathbf{S}, \mathbf{M})$ and proceed to study stopping regions in the $s + t + 1$ dimensional space of the vector $(n, \mathbf{S}, \mathbf{M})$.

3. Constant *a posteriori* risk boundaries. Two disjoint Borel subsets of the parameter space, the *hypotheses*, are denoted by H_0 and H_1. Consider first any one of the hypotheses, say H_0, and let $L_0(\boldsymbol{\theta}, \boldsymbol{\eta})$ be the loss incurred when H_0 is rejected while the "true" parameter point is $(\boldsymbol{\theta}, \boldsymbol{\eta})$. We assume L_0 to be a bounded measurable function on Θ, positive on H_0 and zero on the rest of Θ. By W we denote an *a priori* distribution, a probability measure on the Borel subsets of Θ.

After having observed $\omega_1, \cdots, \omega_n$, the *a posteriori risk of rejecting* H_0 is the conditional expectation of $L_0(\boldsymbol{\theta}, \boldsymbol{\eta})$ given $\omega_1, \cdots, \omega_n$. We denote it by R_0, and though it is a random variable on $\Omega \times \cdots \times \Omega \times \Theta$, it is easily expressed as a function of the sufficient statistic $(n, \mathbf{S}, \mathbf{M})$:

Denoting by $A = A(n, \mathbf{M})$ the set of all $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in Θ such that $n^{-1} \leqslant \theta_k$ (the kth component of $\mathbf{M}$) for $k = 1, 2, \cdots, t$, we obtain

$$R_0(n, \mathbf{S}, \mathbf{M}) = \int_A \exp(\langle \boldsymbol{\eta} \cdot \mathbf{S} \rangle - nb(\boldsymbol{\theta}, \boldsymbol{\eta}))L_0(\boldsymbol{\theta}, \boldsymbol{\eta})\, dW \cdot [\int_A \exp(\langle \boldsymbol{\eta} \cdot \mathbf{S} \rangle - nb(\boldsymbol{\theta}, \boldsymbol{\eta}))\, dW]^{-1}.$$

It is convenient to adopt this as the definition of $R_0(n, \mathbf{S}, \mathbf{M})$ for arbitrary, rather than integer-valued, positive n. Accordingly, we define $\mathcal{R}_0(r)$ for $1 > r > 0$ by $\mathcal{R}_0(r) = \{(n, \mathbf{S}, \mathbf{M}) : R_0(n, \mathbf{S}, \mathbf{M}) \leqq r\}$. Now consider the intersection of $\mathcal{R}_0(r)$ with a fixed ray ρ emanating from the origin of $(n, \mathbf{S}, \mathbf{M})$-space.

Along the ray ρ, the ratios $\mathbf{m} = \mathbf{M}/n$ and $\mathbf{s} = \mathbf{S}/n$ are constant, and since the set A depends only on $\mathbf{M}/n$, it is the same set for all points of ρ. For fixed $\nu > 0$, consider the point on ρ whose first coordinate $n = -\nu \log r$. This point is an element of $\mathfrak{R}_0(r)$ if and only if

$$\int_A \exp\,((\langle \mathbf{n}\cdot \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n}))n)L(\boldsymbol{\theta}, \mathbf{n})\,dW[\int_A \exp\{(\langle \mathbf{n}\cdot \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n}))n\}\,dW]^{-1} \leqq r.$$

Now we take the nth root of both sides. On the right, since $n = -\nu \log r$, we obtain $r^{1/n} = e^{-1/\nu}$.

On the left hand side, the nth roots of the integrals are best expressed as L_p-norms with $p = n$. The root in the numerator is the L_n-norm of $\exp\,(\langle \mathbf{n}, \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n}))$ on A, with respect to the measure $L_0(\boldsymbol{\theta}, \mathbf{n})\,dW$; the root in the denominator is the L_n-norm of the same function on A, with respect to the measure dW.

If we now hold ρ and ν fixed, and send r to zero, $n = -\nu \log r$ tends to infinity, and the L_n-norms tend in the limit to L_∞-norms, that is, to the essential suprema, modulo the measures $L_0\,dW$ and dW, of the integrand on the set $A(n, \mathbf{M})$. Since L_0 is positive on H_0 and zero off H_0, the essential supremum modulo $L_0\,dW$ taken over A, is the essential supremum modulo dW over $A \cap H_0$, and the limiting inequality becomes

$$\sup\,(\mathrm{mod}\ W)_{A\cap H_0} \exp\,(\langle \mathbf{n}\cdot \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n}))/\sup\,(\mathrm{mod}\ W)_A \exp\,(\langle \mathbf{n}\cdot \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n})) \leqq e^{-1/\nu}.$$

Solving for ν, we obtain

$$\nu \geqq [(\sup_A - \sup_{A\cap H_0})(\mathrm{mod}\ W)(\langle \mathbf{n}\cdot \mathbf{s}\rangle - b(\boldsymbol{\theta}, \mathbf{n}))]^{-1}$$

as the necessary and sufficient condition for the point $(\nu, \nu\mathbf{s}, \nu\mathbf{m}) = (\nu, \mathbf{S}, \mathbf{M})$ to be an element of the "asymptotic shape" $\lim_{r\to 0}(-\log r)^{-1}\mathfrak{R}_0(r)$.

Combining this with similar considerations for H_1, and defining $\mathfrak{R}(r) = \mathfrak{R}_0(r) \cup \mathfrak{R}_1(r)$ as the set where at least one of the two available decisions leads to an *a posteriori* risk of at most r, we obtain:

THEOREM 1.

$$\lim_{r\to 0}(-\log r)^{-1}\mathfrak{R}(r)$$
$$= (n, \mathbf{S}, \mathbf{M}):(\sup_A - \min_{i=0,1}(\sup_{A\cap H_i}))(\mathrm{mod}\ W)\ (\langle \mathbf{n}\cdot \mathbf{S}\rangle - nb(\boldsymbol{\theta}, \mathbf{n})) \geqq 1.$$

COROLLARY: Let $\Lambda(n, \mathbf{S}, \mathbf{M})$ be the *"two-sided maximum likelihood statistic"*

$$\Lambda = \sup_A \exp\,(\langle \mathbf{n}\cdot \mathbf{S}\rangle - nb(\boldsymbol{\theta}, \mathbf{n}))/\min_{i=0,1}(\sup_{A\cap H_i} \exp\,(\langle \mathbf{n}\cdot \mathbf{S}\rangle - nb(\boldsymbol{\theta}, \mathbf{n}))).$$

If the support of W contains all of Θ, then

$$\lim_{r\to 0}(-\log r)^{-1}\mathfrak{R}(r) = \{(n, \mathbf{S}, \mathbf{M}):(n, \mathbf{S}, \mathbf{M}) \geqq e\}.$$

4. *d*-testable hypotheses. The hypotheses H_0 and H_1 are *d-testable*, if there exists a fixed-sample-size-test of H_1 against H_0 based on d observations, whose probability of error is bounded on $H_0 \cup H_1$ by a number smaller than $\frac{1}{2}$.

LEMMA 1. *If H_0 and H_1 are d-testable, there exist tests of H_1 against H_0, based on N observations, whose error-bounds decrease geometrically with N.*

PROOF. Let $\tau(\omega_1, \cdots, \omega_d)$ be a test whose probability of error is bounded above by $\alpha < \frac{1}{2}$. Let τ^k be the following test, based on $(2k - 1)d$ observations: Reject H_0 whenever at least k among the tests $\tau(\omega_1, \cdots, \omega_d)$, $\tau(\omega_{d+1}, \cdots, \omega_{2d})$, $\cdots$, $\tau(\omega_{(2k-2)d+1}, \cdots, \omega_{(2k-1)d})$ reject H_0. Since for any parameter point θ in H_0 we have $P_\theta(\tau \text{ rejects } H_0) < \alpha$, we obtain

$$P_\theta(\tau^k \text{ rejects } H_0) < \sum_{j=k}^{2k-1} \binom{2k-1}{j} \alpha^j (1 - \alpha)^{2k-1-j}.$$

Furthermore, since $\alpha < \frac{1}{2}$ implies that the largest term in this sum is the first term,

$$\begin{aligned} P(\tau^k \text{ rejects } H_0) &< k\binom{2k-1}{k}\alpha^k(1 - \alpha)^{k-1} < 2^{2k-1}k\alpha^k(1 - \alpha)^{k-1} \\ &= (4\alpha(1 - \alpha))^k k(2(1 - \alpha))^{-1} \end{aligned}$$

which ultimately decreases geometrically with k. The conclusion of the lemma now follows by putting k equal to the largest integer such that $(2k - 1)d \leqq N$, and repeating the argument with H_0 replaced by H_1.

LEMMA 2. *If the hypotheses are d-testable and the loss function is bounded, there exist, for sufficiently small cost c per observation, fixed sample size procedures whose risk is $O(c \log c^{-1})$.*

PROOF. By Lemma 1, there exists for every N a test based on N observations, such that for some $A > 0$ and $B < 1$ the error probability is at most AB^N. Introducing L for a bound on the loss function, and combining the penalty for error with the observation cost, the bound $ALB^N + cN$ is obtained for the risk. Choosing for N the integer closest to $(\log c^{-1})(\log B^{-1})^{-1}$, the conclusion of the lemma follows easily.

5. Criteria for d-testability. We now prove two sufficient conditions for d-testability, the first being a special case of the second, and two relative conditions, that infer the d-testability of some pairs of hypotheses from the d-testability of others.

(a) *In the purely exponential case, if H_0 and H_1 are compact, disjoint, and contain no boundary points of the parameter space, they are d-testable.*

PROOF. At any interior point $\mathbf{n}$ of the parameter space, $E_{\mathbf{n}}(\mathbf{Y}) = \text{grad}\, b(\mathbf{n})$. The Jacobian matrix of the mapping $\mathbf{n} \to \text{grad}\, b(\mathbf{n})$ is the covariance matrix of $\mathbf{Y}$, and therefore positive definite, and the mapping is one-to-one and continuous. The images of H_0 and H_1 under such a mapping are also compact and disjoint. Let ϵ be the distance between them, and let B be the maximum over $H_0 \cup H_1$ of the eigenvalues of the covariance matrix of $\mathbf{Y}$. Then, with $d > 8B\epsilon^{-2}$, and $\mathbf{s} = d^{-1} \sum_{i=1}^{d} \mathbf{Y}(\omega_i)$, the test that accepts the hypothesis to whose image under grad b the point $\mathbf{s}$ is closer, has error probabilities bounded by $P_{\mathbf{n}}(\|\mathbf{s} - \text{grad}\, b\| \geqq \epsilon/2) \leqq 4B\epsilon^{-2}d^{-1} < \frac{1}{2}$, by the Chebyshev inequality.

In the case of exponential truncation families, different points of the parameter space may correspond to the same distribution, and this redundancy can be eliminated as follows: a parameter point $(\boldsymbol{\theta}, \mathbf{n})$ is *proper* if $\sup\,(\text{mod}\, P_{\boldsymbol{\theta},\mathbf{n}})\, T_k = \theta_k$ for $k = 1, \cdots, t$. The family is uniquely parametrized by the proper parameter points. For interior points of the parameter space, the propriety of $(\boldsymbol{\theta}, \mathbf{n})$ depends only on $\boldsymbol{\theta}$.

(b) *If $P_{\theta,\mathfrak{n}}$ is an exponential truncation family, and H_0 and H_1 are compact disjoint sets of proper interior parameter points, they are d-testable.*

PROOF. Consider the images of H_0 and H_1 under the mapping $(\boldsymbol{\theta}, \mathfrak{n}) \to (\boldsymbol{\theta}, \mathrm{grad}_{\mathfrak{n}}\, b)$, where $\mathrm{grad}_{\mathfrak{n}}$ is the vector of derivatives with respect to the η_i. As in (a), the mapping is one-to-one and continuous, and we can denote the distance between the images by ϵ. For given d, define B and $\mathbf{s}$ as in (a), and let $\mathbf{m} = (\max_{1 \leqq i \leqq d} T_1(\omega_i), \cdots, \max_{1 \leqq i \leqq d} T_t(\omega_i))$. Accept the hypothesis whose image under $(\boldsymbol{\theta}, \mathfrak{n}) \to (\boldsymbol{\theta}, \mathrm{grad}_{\mathfrak{n}}\, b)$ is closer to $(\mathbf{m}, \mathbf{s})$. A wrong decision can occur only if $\|(\mathbf{m}, \mathbf{s}) - (\boldsymbol{\theta}, \mathrm{grad}_{\mathfrak{n}}\, b\| \geqq \epsilon/2$, which implies that one of the events $\|\mathbf{s} - \mathrm{grad}_{\mathfrak{n}}\, b\| \geqq \epsilon/2 \cdot 2^{\frac{1}{2}}$ or $\|\mathbf{m} - \boldsymbol{\theta}\| \geqq \epsilon/2 \cdot 2^{\frac{1}{2}}$ occur. The first has its probability bounded by $8B\epsilon^{-2}d^{-1}$, which is less than $\frac{1}{4}$ when $d > 32B\epsilon^{-2}$. The second event implies $\theta_k - m_k > \epsilon/3t^{\frac{1}{2}}$ for some $1 \leqq k \leqq t$, or equivalently, $\theta_k - T_k(\omega_i) > \epsilon/3t^{\frac{1}{2}} = \delta$ for some $1 \leqq k \leqq t$, for all $1 \leqq i \leqq d$. Denoting the maximum of θ on $H_0 \cup H_1$ by $\boldsymbol{\tau}$, we have

$$P_{\theta,\mathfrak{n}}(\theta_k - T_k \leqq \delta) = P_{\tau,\mathfrak{n}}(\theta_k - T_k \leqq \delta \mid \mathbf{T} \leqq \boldsymbol{\theta})$$
$$\geqq P_{\tau,\mathfrak{n}}(\theta_k - T_k \leqq \delta \quad \text{and} \quad \mathbf{T} \leqq \boldsymbol{\theta}),$$

which is continuous in $\mathfrak{n}$ and lower semi-continuous in $\boldsymbol{\theta}$, and therefore achieves its minimum on $H_0 \cup H_1$. Denoting by Δ_k this minimum, which has to be positive by the assumption of proper parameter points, we have

$$P_{\theta,\mathfrak{n}}(\theta_k - T_k(\omega_i) > \delta) \leqq 1 - \Delta_k,$$
$$P_{\theta,\mathfrak{n}}(\theta_k - \max T_k(\omega_i) > \delta) \leqq (1 - \Delta_k)^d,$$
$$P_{\theta,\mathfrak{n}}(\|\boldsymbol{\theta} - \mathbf{m}\| \geqq \epsilon/2 \cdot 2^{\frac{1}{2}}) \leqq \textstyle\sum_{k=1}^{t} (1 - \Delta_k)^d,$$

which for sufficiently large d is less than $\frac{1}{4}$, yielding $P_{\theta,\mathfrak{n}}(\|(\mathbf{m}, \mathbf{s}) - (\boldsymbol{\theta}, \mathrm{grad}_{\mathfrak{n}}\, b)\| \geqq \epsilon/2) < \frac{1}{2}$.

(c) *If (G_i, H_j) is d_{ij}-testable for $i = 1, \cdots, I$ and $j = 1, \cdots, J$, then $(\cup_i G_i, \cup_j H_j)$ is d-testable.*

PROOF. It is enough to prove the case $I = 2$, $J = 1$. By Lemma 1 of Section 3, there exist tests τ_1 and τ_2 for (G_1, H) and (G_2, H) respectively, with error probability bounded by $\frac{1}{4}$. If we carry out both tests, with a separate batch of observations for each, and accept $G_1 \cup G_2$ whenever τ_1 accepts G_1 and/or τ_2 accepts G_2, we have

$$P(\text{we accept } H) = P(\tau_1 \text{ accepts } H)P(\tau_2 \text{ accepts } H)$$

which is bounded from above by $\frac{1}{4}$ on $G_1 \cup G_2$ and bounded from below by $(1 - \frac{1}{4})^2 = \frac{9}{16}$ on H.

The following is obvious:

(d) *Subsets of d-testable hypotheses are d-testable.*

6. The main theorem. The two are lemmas, together with Theorem 1, the main steps in proving the following:

THEOREM 2. *Let $P(\cdot, \boldsymbol{\theta}, \mathbf{n})$ be a truncation family with parameter space Θ, and let H_0 and H_1 be two d-testable hypotheses. For* a priori *distribution W, bounded loss function L, and cost of an observation c, the Bayes stopping regions $\mathcal{B}_c$ for testing H_1 against H_0 fulfill the asymptotic shape relation*:

$$\lim_{c\to 0}((\log c^{-1})^{-1})\mathcal{B}_c$$
$$= \{(n, \mathbf{S}, \mathbf{M}) : (\sup_A - \min_{i=0,1}(\sup_{A\cap H_i}))(\mathrm{mod}\ W)(\langle \mathbf{n}\cdot\mathbf{S}\rangle - nb(\boldsymbol{\theta}, \mathbf{n})) \geqq 1\}.$$

PROOF. If at some stage of sampling the *a posteriori* risk reaches below c, stopping right there is better than any strategy that takes at least one more observation, therefore $\mathcal{B}_c \supset \mathcal{R}(c)$. On the other hand, Lemma 2 assures the existence of sampling procedures with risk of $Kc \log c^{-1}$ at most, for some fixed K, and c sufficiently small. If at some stage of sampling sequentially the *a posteriori* risk still exceeds $Kc \log c^{-1}$, following one such procedure is better than stopping, and hence

$$\mathcal{R}(Kc \log c^{-1}) \supset \mathcal{B}_c \supset \mathcal{R}(c).$$

The application of Theorem 1, and the observation that $(\log(Kc \log c^{-1}))/\log c \to 1$ finish the proof of the theorem.

7. An example. Let ω_i be uniformly distributed on $(-a, b)$ and let (M_1, M_2) be the sufficient statistic $(-n \min \omega_i,\ n \max \omega_i)$. We test the hypotheses $H_0 : g \leqq g_0$ and $H_1 : g \geqq g_1$ about the midrange $g = (b - a)/2$, but in order to make these hypotheses d-testable, we have to limit the parameter space to the region $\{(-a, b) : 0 \leqq a + b \leqq V\}$ for a fixed V. If we assume that W has this whole region as its support, the region $\lim (\log c^{-1})^{-1}\mathcal{B}_c$ in (n, M_1, M_2) space can be expressed in terms of the *sample range* $t = (M_1 + M_2)/n$ and the *sample midrange* $g = (M_2 - M_1)/2n$ as follows:

$$\lim (\log c^{-1})^{-1}\mathcal{B}_c = \{(n, M_1, M_2) : d_0 + \tfrac{1}{2}\min(V - t, (e^{1/n} - 1)t) \leqq g$$
$$\text{or} \quad g \leqq d_1 - \tfrac{1}{2}\min(V - t, (e^{1/n} - 1)t)\}.$$

Therefore the procedure corresponding to the approximate Bayes region $(\log c^{-1})\lim (\log c^{-1})^{-1}\mathcal{B}_c$ can be described by "continue sampling as long as

$$d_1 - \tfrac{1}{2}\min(V - t, (c^{-1/n} - 1)t) \leqq g \leqq d_0 + \tfrac{1}{2}\min(V - t, (c^{-1/n} - 1)t).$$

As soon as one of the inequalities is violated, stop, and make your final decision according to which of the inequalities made you stop."

REFERENCES

[1] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

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