The Annals of Statistics
1979, Vol. 7, No. 6, 1339–1344

ON BAYES SEQUENTIAL TESTS REGARDING TWO KOOPMAN-DARMOIS DISTRIBUTIONS: BOUNDS ON THE MAXIMUM SAMPLE SIZE

BY J. E. MANN AND T. L. BRATCHER

Virginia Polytechnic Institute and State University and The University of Southwestern Louisiana

A result due to Ray is extended to cover hypothesis testing problems in which samples are taken sequentially in pairs from two independent distributions of a one-dimensional univariate exponential family. In particular, for certain reasonable restrictions on losses and costs, an upper bound to the exact stage of truncation is obtained for a truncated Bayes procedure.

1. Introduction. S. N. Ray (1965) gave conditions for the existence of a truncated Bayes sequential procedure and developed a criterion for determining upper bounds to the point of truncation of a sequential test regarding the parameter of a one-dimensional exponential family of distributions. These results are contained in Theorems 1.1 and 1.2 below.

In this paper Ray's result, Theorem 1.2, is extended to cover hypothesis testing problems in which sampling is from two independent distributions belonging to the same one-dimensional exponential family. This result is given in Theorem 2.2 below. It is also shown that the assumptions for Theorem 1.2, as stated by Ray (1965), should be modified.

Since the notation to be used is standard, we shall simply identify the symbols used and refer the reader to Ray (1965) for precise definitions. Thus, let A be the action space; Γ the parameter space; $L(\theta, a)$, $\theta \in \Gamma$, $a \in A$, the loss; and $c(\theta)$ the cost of a single observation. For any prior distribution, ξ, of Θ, let $\rho(\xi)$ be the Bayes risk over all measurable randomized sequential procedures; let $\rho_n(\xi)$ be the Bayes risk over the subclass of such procedures which are truncated at n; and let $\lambda(\xi)$ be the expected reduction in the stopping risk by taking one more observation. If $c(\theta)$ is the cost of an observation then $c(\xi)$ represents the expected cost with respect to θ.

The motivation for Theorem 4.1 of Ray (1965) and Theorem 2.2 of this paper is the following well-known proposition which, for the sake of completeness, is stated here.

THEOREM 1.1. (See Ray (1965, page 861)). *If* $\lim_{n\to\infty}\rho_n(\xi) = \rho(\xi)$, *there is a Bayes sequential procedure with respect to* ξ *that is truncated provided there exists an integer* n' *such that*

$$\text{essup}_{\mathbf{x}^n}\left[\lambda(\xi_{\mathbf{x}^n})/c(\xi_{\mathbf{x}^n})\right] \leqslant 1 \tag{1.1}$$

for every $n \geqslant n'$.

Received July 1977; revised May 1978.

AMS 1970 *subject classifications.* Primary 62L10; secondary 62F05, 62F15.

Key words and phrases. Tests of hypotheses, sequential analysis, exponential family, truncated Bayes sequential procedures.

Here ξ is the prior distribution of Θ, and $\xi_{\mathbf{x}^n}$ is the posterior distribution of Θ after observing the vector $\mathbf{x}^n$.

Thus, if it is possible to determine the left-hand side of (1.1) as a function of n, then (1.1) provides a test of truncation as well as a means of obtaining an upper bound to the exact stage of truncation.

In the remainder of this paper, our attention will be focused upon the hypothesis testing problem for which the action space is $A = \{a_1, a_2\}$. Before considering our extension of Theorem 4.1 of Ray (1965), we shall briefly discuss Ray's assumptions.

Let X have the generalized conditional density, given $\Theta = \theta$,

$$f(x|\theta) = \Psi(\theta)\exp[\theta P(x)] \tag{1.2}$$

with respect to a σ-finite measure μ on the real line where $\theta \in \Gamma$, an interval of the real line, and P is monotone in x.

The prior (marginal) distribution, ξ, of Θ is assumed to admit the density

$$g(\theta) = K(a, b)\exp[a\theta + bS(\theta)] \tag{1.3}$$

with respect to τ, the measure induced on Γ by Lebesgue measure on the real line, where $\exp[S(\theta)] = \Psi(\theta)$ as given in (1.2). Note that (1.2) and (1.3) define naturally conjugate families.

The loss characteristic associated with A is defined by

$$L(\theta) = L(\theta, a_1) - L(\theta, a_2). \tag{1.4}$$

Now let $X_1, \cdots, X_n$ be i.i.d. as (1.2). Since the statistic $T_n = \Sigma_{i=1}^n P(X_i)$ is sufficient for Θ, the posterior distribution $\xi_{\mathbf{x}}n$ of Θ may be denoted by (n, t), and the posterior expectation of any function, say g, of Θ may be denoted by $g(n, t)$, where $t = T_n(\mathbf{x}^n)$. Remarks (1)–(3) below pertain to the hypotheses of Theorem 4.1, Ray (1965, page 865), which is stated here as Theorem 1.2.

THEOREM 1.2. (Ray (1965, page 865)). *Let X and Θ have distributions* (1.2) *and* (1.3), *respectively. If μ is absolutely continuous with respect to Lebesgue measure and the loss characteristic is monotone in θ taking both positive and negative values with positive ξ probability, then $\lambda(n, t)$ attains its supremum on the set, $\{(n, t_n^o) : L(n, t_n^o) = 0, n = 1, 2, \cdots\}$, called the neutral boundary. Moreover, this supremum is given by*

$$\lambda(n, t_n^o) = |\textstyle\int_\Gamma L(\theta)F_\theta(t_{n+1}^o - t_n^o)\, d_{(n, t_n^o)}(\theta)| \tag{1.5}$$

where F_θ is the conditional distribution function of $P(X)$ given $\Theta = \theta$.

(1) The hypothesis regarding the change of sign of $L(\theta)$ is insufficient to insure a change of sign of $L(n, t)$ as may be seen by setting $f(x|\theta) = \theta\exp(-\theta x)$, $x > 0$ ($\theta > 0$), taking the distribution of Θ to be gamma (2, 10), and letting $L(\theta) = \theta - \frac{1}{2}$ and $n = 2$. Then, $L(n, t) = 4/(10 + t) - 2^{-1}$ and $P[L(n, T_n) > 0] = 0$, where $T_n = \Sigma_{i=1}^n X_i$. Thus, it should be assumed that $L(n, t)$ changes sign.

(2) The assumption that μ is absolutely continuous with respect to Lebesgue measure is made in order to assure the continuity of $L(n, t)$ in t for all n. However,

since $L(n, t)$ depends on the posterior distribution of Θ, this hypothesis should be replaced by an appropriate assumption regarding (n, t). It will be sufficient to assume that (n, t) is absolutely continuous with respect to τ and that there exist closed intervals J_1 and J_2 of the real line such that $[d(n, t)/d\tau](\theta)$ is continuous in (θ, t) over $J_1 \times J_2$. If J_1 is compact, assume $L(\theta)$ is Riemann integrable over J_1; if J_1 is an infinite interval assume $E_{(n, t)}L(\Theta)$ converges uniformly. See Bartle (1964, pages 306, 312, 355) for a discussion of these conditions.

(3) Most importantly, Theorem 1.2—as stated—does not apply to a discrete X. The revisions suggested in (1) and (2) broaden Ray's important result to include the discrete case. His proof remains valid without change. It should be noted, however, that $\operatorname{essup}_{\mathbf{x}^n}\lambda(\xi_{\mathbf{x}^n})$ may not coincide with $\sup_t \lambda(n, t)$, but will never exceed it.

2. Comparing two distributions. Let $\Theta = (\Theta_1, \Theta_2)$, and let $\mathbf{X}^n = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ where $\mathbf{X}_i = (X_{1i}, X_{2i})$, be a random sample from the joint distribution of X_1 and X_2. It is assumed that X_1 and X_2 are independently distributed with generalized densities, with respect to μ, of the form

$$f(x|\theta_i) = \Psi(\theta_i)\exp[\theta_i P(x)], \qquad i = 1, 2; \tag{2.1}$$

and Θ_1 and Θ_2 are independently distributed with generalized densities, with respect to τ, of the form

$$g(\theta) = K(a_i, b_i)\exp[a_i\theta + b_i S(\theta)], \qquad i = 1, 2, \tag{2.2}$$

as in (1.2) and (1.3), respectively.

For $i = 1, 2$, define $T_{i,n} = \Sigma_{j=1}^n P(X_{ij})$, $U_n = T_{1,n} + T_{2,n}$, and $d = \theta_1 - \theta_2$. It is obvious from (1.2) and (1.3) that $T_{1,n}$ and $T_{2,n}$ are jointly sufficient for Θ_1 and Θ_2, and the posterior distribution of Θ, given $(n, T_{1,n} = t_1, T_{2,n} = t_2)$, has density

$$g(\theta_1, \theta_2|t_1, t_2) = C' \exp\{\Sigma_{i=1}^2[(a_i + t_i)\theta_i + (b_i + n)S(\theta_i)]\}.$$

Thus, the posterior distribution of Θ may be denoted by (n, t_1, t_2). If $U_n = u$ is fixed, then $(n, t_1, t_2) = (n, t_1, u - t_1)$. In this case, we may denote (n, t_1, t_2) by (n, t_1, u).

The following lemmas are needed for the proof of Theorem 2.1.

LEMMA 2.1. *The conditional distribution of $T_{1,n}$, given $U_n = u$, belongs to a one-dimensional univariate family and depends on $\theta = (\theta_1, \theta_2)$ only through the difference d.*

PROOF. Definitions (2.1) and (2.2) imply that the conditional joint density of $(T_{1,n}, U_n)$ given (θ_1, θ_2) is $K'(d, \theta_2)\exp(dt_1 + \theta_2 u)$ with respect to a suitable σ-finite measure ν on the product space. The conclusion now follows immediately from Lemma 8, page 52, of Lehmann (1959).

LEMMA 2.2. *For each fixed value u of U_n, the conditional posterior distribution of Θ, given (t_1, u), has monotone likelihood ratio in d when t_1 is regarded as the parameter.*

PROOF. The posterior density of Θ is $C_1(t_1, u)C_2(d, \theta_2) \exp(dt_1 + \theta_2 u)$.

THEOREM 2.1. *If the loss characteristic, $L(\Theta)$, is monotone in d and if U_n has the fixed value u, then*

$$L(n, t_1, u) = E_{(n, t_1, u)}L(\Theta) \tag{2.3}$$

is a monotone function of t_1. If for each positive integer n, $L(n, t_1, u)$ assumes both positive and negative values as t_1 varies over the essential range of $T_{1, n}$, then the equation $L(n, t_1, u) = 0$ has a unique root, $t_1 = t'$.

PROOF. The result of Lemma 2.2, along with Lemma 2(i), page 74, of Lehmann (1959) and the assumptions made here imply that $L(n, t_1, u)$ is monotone in t_1. According to Lemma 3, page 129, of Ferguson (1967), $L(n, t_1, u)$ is an analytic function of t_1 and u separately. Thus, if $L(n, t_1, u)$ assumes both positive and negative values, it follows from the monotonicity of $L(n, t_1, u)$ in t_1 and from the properties of analytic functions (see Bartle (1964, page 412)) that the equation $L(n, t_1, u) = 0$ has exactly one solution in t_1 for each n.

As a consequence of the preceding, whenever the loss characteristic satisfies the assumptions of Theorem 2.1, the following result applies. We now state our generalization of Theorem 1.2 due to Ray (1965).

THEOREM 2.2. *It is assumed that an hypothesis is to be tested regarding the parameters Θ_1 and Θ_2 of two independent distributions having generalized densities of the form* (2.1) *while Θ_1 and Θ_2 are independent with prior distributions* (2.2). *Let samples be taken in pairs from the distributions* (2.1). *If $L(n, t_1, t_2) = E_{(n, t_1, t_2)}L(\Theta)$ is nondecreasing in t_1 and nonincreasing in t_2, and for each n and u, $L(n, t_1, u)$ assumes both positive and negative values, then the equation $L(n, t_1, u) = 0$ has the unique solution $t_1 = t'$ and $\lambda(n, t_1, u)$ assumes its supremum on the neutral boundary $\{(n, t', u)\colon L(n, t', u) = 0, u \in R, n = 1, 2, \cdots\}$, that is*

$$\operatorname{essup}_{(n, t_1, u)}\lambda(n, t_1, u) \leqslant \sup_{(n, t_1, u)}\lambda(n, t_1, u) = \lambda(n, t', u). \tag{2.4}$$

Furthermore, if $t_1 = t''$, is the unique solution, given $U_n = u + y$, of $L(n + 1, t_1, u + y) = 0$, then

$$\begin{aligned}\lambda(n, t', u) &= \left|\int_{T^2} L(\theta) \int_{-\infty}^{\infty} \left\{\int_{-\infty}^{t''-t'} f_\theta(x|y)\, d\mu(x)\right\} g_\theta(y)\, d\mu(y)\, d\xi_n^0(\theta)\right| \\ &= \left|E_{\xi_n^0} L(\Theta) E_{Y|\Theta} F_{X|\Theta}(T'' - t'|Y\right|.\end{aligned} \tag{2.5}$$

In (2.5), *$\xi_n^0 = (n, t', u)$. Also, $X = P(X_{1, n+1})$, $F_{X|\theta}$ is the conditional distribution function of $P(X_1)$ given $\Theta = \theta$, and T'' depends on $u + Y$ where $Y = P(X_{1, n+1}) + P(X_{2, n+1})$. Finally, $E_{Y|\theta}$ and $E_{\xi_n^0}$ are the expectation operators with respect to the conditional distribution of Y given $\Theta = \theta$ and ξ_n^0, respectively.*

PROOF. The proof is a simple adaptation of the proof of Theorem 1.2 (Ray (1965)) and is not given here. A detailed proof may be found in Mann (1973).

Although the expectation (2.3) is continuous in both t_1 and u, it should be observed that if μ is absolutely continuous with respect to counting measure, there may exist a neighborhood N of (t', u) such that $(T_{1,n}, U_n)$ assumes values in N with zero probability. In such cases it is convenient to introduce the concept of an extended neutral boundary, Ray (1965, page 868). If the expected posterior sampling cost attains its minimum on the neutral boundary at the point (n, t', u^0) where u^0 is the value of u which maximizes $\lambda(n, t', u)$, then an upper bound n' to the exact stage of truncation is determined by choosing n' to be the least nonnegative integer n such that $\lambda(n, t', u^0) \leqslant c(n, t', u^0)$ a.s. for every $n \geqslant n'$.

3. Applications. In each of the following examples, the hypotheses to be tested are $H_0 : \theta_1 \leqslant \theta_2$ vs $H_1 : \theta_1 > \theta_2$ with symmetric linear loss given by

$$\begin{aligned} L(\theta, a_1) &= 0, && \theta_1 \leqslant \theta_2, \\ &= k(\theta_1 - \theta_2), && \theta_2 < \theta_1; \\ L(\theta, a_2) &= 0, && \theta_2 < \theta_1; \\ &= k(\theta_2 - \theta_1), && \theta_1 \leqslant \theta_2, \end{aligned}$$

where k is a positive constant. Also, $P(X_i) = X_i$; so that $T_{i,n} = \Sigma_{j=1}^{n} X_{ij}$, $Y = X_{1,n+1} + X_{2,n+1}$ and $U_n = T_{1,n} + T_{2,n}$. In addition, we shall define X and t_i by $X = X_{1,n+1}$ and t_i is a value of $T_{i,n}$.

EXAMPLE 1. *Normal.* Suppose $X_i \sim N(\theta_i, 1)$, $\Theta_i \sim N(\mu_i, \tau_i^{-1})$, $i = 1, 2$, where τ_i is the precision. Applying Theorem 2.2, we obtain

$$\text{(3.1)} \quad \sup_{t_1} \lambda(n, t_1, u) = k | \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta_1 - \theta_2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{t''-t'} f_\theta(x|y)\, dx g_\theta(y)\, dy\, h_1(\theta_1) h_2(\theta_2)\, d\theta_1\, d\theta_2 |$$

where

$$\begin{aligned} f_\theta(x|y) &= \Pi^{-\frac{1}{2}} \exp\{-[x - (\theta_1 - \theta_2 + y)/2]^2\}, \\ g_\theta(y) &= (\Pi^{-\frac{1}{2}}/2) \exp\{-4^{-1}[y - (\theta_1 + \theta_2)]^2\}, \\ h_i(\theta_i) &= [(\tau_i + n)/(2\pi)]^{\frac{1}{2}} \exp\{-[(\tau_i + n)/2][\theta_i - \mu']^2\} \end{aligned}$$

where

$$\mu' = (\tau_1 \mu_1 + \tau_2 \mu_2 + u)/(\tau_1 + \tau_2 + 2n).$$

Also,

$$\begin{aligned} t'' - t' &= [(\alpha + 1)/(\alpha + \beta + 2)]y \\ &\quad + [(\tau_2 - \tau_1)(\tau_1 \mu_1 + \tau_2 \mu_2 + u)]/[(\alpha + \beta)(\alpha + \beta + 2)] \end{aligned}$$

where $\alpha = \tau_1 + n$ and $\beta = \tau_2 + n$. We have evaluated (3.1) exactly if $\tau_1 = \tau_2 = \tau$. In this case the right side of (3.1) becomes $k[\pi(\tau + n)(\tau + n + 1)]^{-\frac{1}{2}}$, and n' is chosen to be the smallest nonnegative integer n such that $n \geqslant k[\pi c^2 + 0.25]^{-\frac{1}{2}} - \tau - 0.5$. For unequal values of τ_1 and τ_2, a 16-point Gaussian quadrature was used to approximate (3.1) which appears to be independent of u. The accuracy was checked for several cases of equal τ_1 and τ_2. In each case, we obtained at least three place accuracy, and the error diminishes with increasing sample size. Thus, it appears that such a numerical approximation is sufficiently accurate for applications.

EXAMPLE 2. *Bernoulli.* Suppose $X_i \sim \text{bin}(1, \theta_i)$, $\Theta_i \sim \text{beta}(\alpha_i, \beta_i)$, $\alpha_i > 0$, $\beta_i > 0$, $i = 1, 2$. In this case,

$$\sup_{(t_1, u)} \lambda(n, t_1, u) = k/[4(m + n + 1)]$$

where $m = \min\{\alpha_1 + \beta_1, \alpha_2 + \beta_2\}$. An upper bound n' to the exact stage of truncation is obtained by letting n' be the smallest nonnegative integer n such that $n \geqslant k/(4c) - m - 1$. This result is consistent with that given by Bratcher (1971).

For a more detailed analysis of both examples, the reader is referred to Mann (1973).

REFERENCES

[1] BARTLE, R. G. (1964). *The Elements of Real Analysis*. Wiley, New York.

[2] BRATCHER, T. L. (1971). An upper bound for the sequential comparison of two binomial probabilities. Presented at the Joint Statistical Meetings, Fort Collins. Abstract *Ann. Math. Statist.* **42** 1791.

[3] FERGUSON, T. S. (1967). *Mathematical Statistics—A Decision Theoretic Approach*. Academic Press, New York.

[4] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[5] MANN, J. E. (1973). On a Bayesian sequential comparison of two Koopman-Darmois distributions. Unpublished thesis, Univ. Southwestern Louisiana.

[6] RAY, S. N. (1965). Bounds on the maximum sample size of a Bayes sequential procedure. *Ann. Math. Statist.* **36** 859–878.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA 24061

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHWESTERN LOUISIANA
LAFAYETTE, LOUISIANA 70504