The Annals of Statistics
1976, Vol. 4, No. 6, 1101–1112

ASYMPTOTIC SOLUTIONS TO THE TWO STATE COMPONENT COMPOUND DECISION PROBLEM, BAYES VERSUS DIFFUSE PRIORS ON PROPORTIONS[1]

By Dennis C. Gilliland, James Hannan and J. S. Huang

Michigan State University and University of Guelph

Gilliland and Hannan (1974, Section 3) consider a general finite state compact risk component and reduce the problem of treating the asymptotic excess compound risk of Bayes compound rules to the question of L_1 consistency of certain induced estimators. This present paper considers the two state case and for several classes of diffuse symmetric priors on proportions establishes the L_1 consistency with rate. The rate $O(n^{-\frac{1}{2}})$ uniform in state sequences is shown for the uniform prior giving strong affirmation to the asymptotic form of a conjecture by Robbins (1951). The same or logarithmically weakened rate is shown for symmetric priors which are Λ-mixtures for several classes of Λ. A corollary shows a nonnull consistency, without regularity conditions, of a maximum likelihood estimator.

1. Introduction, notations and summary. Robbins (1951, page 140) introduced the compound decision problem and, in a featured example witn component problem discrimination between $N(-1, 1)$ and $N(1, 1)$, suggested that the Bayes compound rule versus the symmetric prior uniform on proportions might be superior, exactly or asymptotically, to his bootstrap rule. Gilliland and Hannan ((1974), Section 3) consider a general finite state compact risk component and reduce the problem of treating the asymptotic excess compound risk of Bayes compound rules to the question of consistency of certain induced estimators.

The present paper treats the consistency question in the two state case and, for several classes of symmetric priors including the Robbins prior, establishes L_1 consistency with rate.

In the last example of Section 3, Gilliland and Hannan (1974) show that a Bayes compound rule versus a symmetric prior $\boldsymbol{\beta}$ is provided by an equivariant delete bootstrap rule $\mathbf{s}^{\mathbf{w}}$ with $\mathbf{w}$ an estimator induced by $\boldsymbol{\beta}$. In their Theorems 3 and 4, the N-component compound risk excess over the simple envelope, $\mathbf{R}(\mathbf{N}, \mathbf{s}^{\mathbf{w}}) - \psi(\mathbf{N})$, is bounded by finite sums of L_1 estimation errors of $k\mathbf{w}$ (where k is a normalizing factor) plus terms $O(N^{\frac{1}{2}})$ uniformly in empirical state distributions $\mathbf{N}$. Thus, $\mathbf{s}^{\mathbf{w}}$ is an asymptotic solution to the compound problem.

$$(1) \qquad \mathbf{R}(\mathbf{N}, \mathbf{s}^{\mathbf{w}}) - \psi(\mathbf{N}) = o(N) \qquad \text{uniformly in} \quad \mathbf{N}\,,$$

Received June 1974.

[1] Research supported by NSF Grant No. GP-33677X1.

AMS 1970 *subject classifications.* Primary 62C25; Secondary 62F10.

Key words and phrases. Two state compound decision problem, Bayes compound procedures, consistent estimator of proportion, consistency of maximum likelihood estimator, consistency of posterior mean.

provided the L_1 estimation errors are no larger. We treat estimation on the (usual) relative frequency scale and obtain rates which carry over to near $O(N^{\frac{1}{2}})$ in (1).

The following notations will be used throughout the paper. For states $\mathscr{P} = \{F_0, F_1\}$, where $F_0 \neq F_1$ are probability measures on $(\mathscr{X}, \mathscr{B})$, let $\mu = F_0 + F_1$ and $z = dF_1/d\mu$ with $0 \leqq z \leqq 1$. Let $\mathbf{P} = \times P_i \in \mathscr{P}^\infty$ and $\|\ \|$ denote the norm in $L_1(\mathbf{P})$. For each n, let

$$(2) \qquad p = n^{-1}\#[i \mid 1 \leqq i \leqq n, P_i = F_1]$$

and let $\boldsymbol{\beta}$ be a probability distribution on $\mathscr{P}^{n+1}$ symmetric under permutation of coordinates. $\boldsymbol{\beta}$ is identified henceforth with the distribution on orbits $\beta_0, \beta_1, \cdots, \beta_{n+1}$. For each $i = 1, 2, \cdots, n$, let $z_i = z(x_i)$ and for each $K = 0, 1, \cdots, n$ let s_K denote the symmetrization of the measure $F_0^{n-K} \times F_1^K$ on $(\mathscr{X}, \mathscr{B})^n$.

With the induced estimator (40) of Gilliland and Hannan (1974) with $N = n + 1$ and $m = 1$, we will use the normalizing factor $k = n(\mathbf{w}_0 + \mathbf{w}_1)^{-1}$ and let

$$(3) \qquad p(\boldsymbol{\beta}) \equiv \mathbf{w}_1/(\mathbf{w}_0 + \mathbf{w}_1)$$

denote the induced estimator of the proportion p of states F_1. To show the L_1 consistency of the estimators $p(\boldsymbol{\beta})$ and $1 - p(\boldsymbol{\beta})$, it suffices to treat the common norm $\|p(\boldsymbol{\beta}) - p\|$.

Theorems 1–4 establish rates of uniform L_1 consistency for the estimator (3) for various classes of $\boldsymbol{\beta}$. Theorem 1 (Section 2) establishes $O(n^{-\frac{1}{2}})$ for Robbins' $\boldsymbol{\beta}$ and Theorem 2 (Section 3) establishes $O((n/\log n)^{-\frac{1}{2}})$ for $\boldsymbol{\beta}$ which are Beta-mixtures of Bernoulli distributions. Section 4 considers Λ-mixtures of Bernoulli distributions and, in Theorems 3 and 4, obtains consistency results for two classes of Λ. The proofs are carried by analyses (Lemmas 2 and 3) of the concentration of a posterior distribution about a maximum likelihood estimator and triangulation about the estimator of Theorem 1. A nonnull consistency of that MLE, without regularity conditions, is a corollary to Lemma 2 and Theorem 1. Lemmas A and B of the appendix establish that log concave discrete and Lebesgue densities possess increasing hazard rates from which follow tail probability bounds used in the proofs of Lemmas 1 and 2. Lemma C and its corollary are results from Hoeffding (1963), reshaped for direct application in the proofs of Lemmas 1 and 2.

In his example, Robbins (1951) demonstrated a bootstrap rule satisfying (1) and suggested that the Bayes compound rule versus the symmetric prior uniform on proportions (Example 2, page 137) might have lower compound risk across N components. Huang (1972) has shown otherwise for $N = 2$ components. However, Theorems 1–4 of this paper together with Theorems 3 and 4 of Gilliland and Hannan (1974) establish classes of Bayes rules in a more general compound problem which satisfy (1) with strengthened rates near $O(N^{\frac{1}{2}})$. One very special case has been treated earlier: with the component problem discrimination between a true coin and a 2-headed one, Samuel ((1967), Section 4) has indicated an inductive proof of (1) for the Bayes compound rule versus the Robbins' prior.

2. Consistency of $p(\boldsymbol{\beta})$ when $\boldsymbol{\beta}$ is the uniform prior. For $n \geqq 1$ and $K = 0, 1, \cdots, n$, let

$$L_K = (n!)^{-1} \sum_g z_{g(1)} \cdots z_{g(K)}(1 - z_{g(K+1)}) \cdots (1 - z_{g(n)}) \tag{4}$$

where the sum is over all $n!$ permutations g and let $L_K = 0$ otherwise. Note that L_K is a density of s_K with respect to μ^n and $\binom{n}{K}L_K$ is a generalized binomial probability of K successes in n independent trials. (The L_K notation was used by Hannan and Robbins (1955) following their (6.5).) The induced estimators, (40) of Gilliland and Hannan (1974) with $N = n + 1$, are

$$\mathbf{w}_0 = \sum (n - K + 1)\beta_K L_K, \qquad \mathbf{w}_1 = \sum K\beta_K L_{K-1}. \tag{5}$$

Before treating $p(\boldsymbol{\beta})$ defined in (3), we find it convenient to treat a somewhat different symmetric estimator. For $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \cdots, \gamma_n), \gamma_K > 0, K = 0, 1, \cdots, n$, let

$$\rho(\boldsymbol{\gamma}) = \frac{\sum K\gamma_K L_K}{n \sum \gamma_K L_K}. \tag{6}$$

Note that

$$\Delta \equiv \int z\,dF_1 - \int z\,dF_0 = 2 \int z^2\,d\mu - 1 > 0 \tag{7}$$

by the Schwarz inequality and the linear independence in $L_2(\mu)$ of 1 and z necessitated by $F_0 \neq F_1$.

LEMMA 1. *With* $l = \log(\bigvee \gamma_K / \bigwedge \gamma_K)$,

$$\tfrac{1}{2}\Delta n^{\frac{1}{2}} \|\rho(\boldsymbol{\gamma}) - p\| \leqq l^{\frac{1}{2}} + (8\pi)^{\frac{1}{2}}. \tag{8}$$

PROOF. From (6) it follows that for integer M, $0 \leqq M \leqq n$, $(n\rho(\boldsymbol{\gamma}) - M)_+$ is bounded by

$$\frac{\sum (K - M)_+ \gamma_K L_K}{\sum \gamma_K L_K} = \frac{\sum_{M+1}^n \sum_K^n \gamma_J L_J}{\sum \gamma_K L_K} \leqq \sum_{M+1}^n \left(e^l \frac{\sum_K^n L_J}{\sum L_J} \wedge 1\right) \tag{9}$$

where the last inequality follows from bounding out the ratio of γ's. For each $\mathbf{z}$, the sequence $\{L_K\}$ is known to be log concave (see Definition A of the appendix and, e.g., Samuels (1965), (5)) so that by Lemma A of the appendix,

$$\sum_K^n L_J / \sum_M^n L_J \leqq L_K / L_M \qquad \text{for all} \quad K \geqq M \quad \text{and} \quad L_M > 0.$$

Since $L_M = ds_M/d\mu^n$, it follows that RHS (9) is a.e. s_M bounded by

$$\sum_{M+1}^n \left(e^l \frac{L_K}{L_M} \wedge 1\right). \tag{10}$$

Since the minimum is less than (or equal to) every convex combination, for every test function T

$$\int \left(e^l \frac{L_K}{L_M} \wedge 1\right) ds_M \leqq e^l \int T\,ds_K + \int (1 - T)\,ds_M. \tag{11}$$

Consider the test $T \equiv [\sum z_i < c_K]$ and let

$$\mu_K = \int \sum z_i \, ds_K \,(= K \int z \, dF_1 + (n-K) \int z \, dF_0) \,. \tag{12}$$

With $\mu_M \leqq c_K \leqq \mu_K$, applications of the corollary in the Appendix give

$$\int T \, ds_K \leqq \exp - 2n^{-1}(\mu_K - c_K)^2 \tag{13}$$

$$\int (1 - T) \, ds_M \leqq \exp - 2n^{-1}(c_K - \mu_M)^2 \,. \tag{14}$$

When K is so large that

$$e^l \exp - n^{-1}(\mu_K - \mu_M)^2 \leqq 1 \,, \tag{15}$$

there is a c_K between μ_M and μ_K such that e^l RHS (13) = RHS (14), namely

$$c_K = \frac{\mu_K + \mu_M}{2} - \frac{nl}{4(\mu_K - \mu_M)} \,. \tag{16}$$

For this choice,

$$c_K - \mu_M = \frac{\mu_K - \mu_M}{2}\left(1 - \frac{nl}{2(\mu_K - \mu_M)^2}\right) \geqq \frac{\mu_K - \mu_M}{4} = \frac{\Delta(K-M)}{4} \tag{17}$$

and

$$\text{RHS}\,(11) \leqq 2 \exp - (8n)^{-1}\Delta^2(K-M)^2 \,. \tag{18}$$

For K not satisfying (15), namely $K < M + \Delta^{-1}(nl)^{\frac{1}{2}}$, we use the bound 1 for LHS (11) and, henceforth taking $M = np$, obtain

$$\Delta n^{\frac{1}{2}} \int (\rho(\boldsymbol{\gamma}) - p)_+ \, ds_M - l^{\frac{1}{2}} \leqq 2\Delta n^{-\frac{1}{2}} \textstyle\sum_1^{n-M} \exp - (8n)^{-1}\Delta^2 J^2 < (8\pi)^{\frac{1}{2}} \,. \tag{19}$$

Since, with $'$ denoting the $F_0 \leftrightarrow F_1$ interchange and the changes thereby induced, $s_M = s'_{n-M}$, $L_{n-K} = L_K'$, $\Delta = \Delta'$, $\gamma_K = \gamma'_{n-K}$, $l = l'$ and

$$M - n\rho(\boldsymbol{\gamma}) = -(n-M) + \frac{\sum (n-K)\gamma_K L_K}{\sum \gamma_K L_K} = \frac{\sum K\gamma_K' L_K'}{\sum \gamma_K' L_K'} - (n-M) \,,$$

we see that

$$\Delta n^{\frac{1}{2}} \int (\rho(\boldsymbol{\gamma}) - p)_- \, ds_M \leqq \text{RHS}\,(8) \tag{20}$$

completing the proof of Lemma 1. □

Lemma 1 will later be used for the proof of L_1 consistency when $\boldsymbol{\beta}$ is a Beta mixture of Bernoulli in Section 3. The following theorem is a rather immediate corollary to it.

THEOREM 1. *For the uniform prior* $\beta_K = (n+2)^{-1}$, $K = 0, 1, \cdots, n+1$,

$$\|p(\boldsymbol{\beta}) - p\| \leqq (4\Delta^{-1}(2\pi n)^{\frac{1}{2}} + 1)/(n+2) \,. \tag{21}$$

PROOF. It follows from (3), (5) and (6) that, for uniform $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$,

$$p(\boldsymbol{\beta}) = \frac{n}{n+2}\rho(\boldsymbol{\gamma}) + \frac{1}{n+2} \,. \tag{22}$$

An application of the triangle inequality and the $l = 0$ case of Lemma 1 complete the proof of (21). □

As remarked in the introduction, the rate $O(n^{-\frac{1}{2}})$ in Theorem 1 carries over to $O(N^{\frac{1}{2}})$ in (1), providing strong affirmation of Robbins' conjecture concerning his Bayes compound procedure. In subsequent sections we establish bounds for $\|p(\boldsymbol{\beta}) - p\|$ for several classes of symmetric priors. To close this section, we remark on the L_1 error of finite mixtures of symmetric priors.

REMARK 1. If $\boldsymbol{\beta}$ is a convex combination $\sum c_i \boldsymbol{\beta}_i$, then

$$\|p(\boldsymbol{\beta}) - p\| \leqq \bigvee \|p(\boldsymbol{\beta}_i) - p\| \, . \tag{23}$$

PROOF. Let $p_i = p(\boldsymbol{\beta}_i)$ and $p_i S_i = \sum_K K\beta_{iK} L_{K-1}$. Then $p(\boldsymbol{\beta}) = \sum c_i p_i S_i / \sum c_i S_i$ and therefore

$$p(\boldsymbol{\beta}) - p = \sum (p_i - p) c_i S_i / \sum c_i S_i$$

from which (23) is immediate. □

3. Consistency of $p(\boldsymbol{\beta})$ when $\boldsymbol{\beta}$ is a Beta mixture of Bernoulli. The Bayesian approach of Shapiro (1972) (cf. also (1974)) to a two state classification component empirical Bayes problem first prompted the authors to examine $\boldsymbol{\beta} = \boldsymbol{\beta}(\Lambda)$ which are mixtures of Bernoulli distributions with respect to a probability measure Λ on $[0, 1]$. As seen in Remark 2 of Gilliland and Hannan ((1974), Section 4) equivariant Bayes rules versus Λ in the empirical Bayes problem are Bayes rules versus $\boldsymbol{\beta} = \boldsymbol{\beta}(\Lambda)$ in the corresponding compound problem. Section 4 treats rather general Λ-mixtures whereas this section treats the Beta mixture with parameters $a, b > 0$,

$$\beta_K = \int \binom{n+1}{K} \omega^K (1-\omega)^{n+1-K} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \omega^{a-1} (1-\omega)^{b-1} \, d\omega \tag{24}$$

for $K = 0, 1, \cdots, n+1$. This mixture is sufficiently smooth in K to permit the use of Lemma 1 in treating $\|p(\boldsymbol{\beta}) - p\|$ and specializes to Robbins' prior when $a = b = 1$.

THEOREM 2. *For $\boldsymbol{\beta}$ defined by* (24)

$$\|p(\boldsymbol{\beta}) - p\| \leqq 2\Delta^{-1}\{\eta^{\frac{1}{2}} + (8\pi)^{\frac{1}{2}}\} n^{-\frac{1}{2}} + (a \vee b) n^{-1} \tag{25}$$

when

$$\eta = \delta\left(1 + \frac{\delta}{n}\right)\left(1 + \frac{1}{a \wedge b} + \log n\right) \qquad \textit{with} \quad \delta = |a-1| \vee |b-1| \, . \tag{26}$$

PROOF. Substituting the evaluation of (24) into (3) and (5) shows that

$$p(\boldsymbol{\beta}) = \frac{1}{n+a+b} (n\rho(\boldsymbol{\gamma}) + a) \, , \tag{27}$$

where ρ is defined by (6) and

$$\gamma_K = \frac{\Gamma(K+a)}{K!} \frac{\Gamma(n-K+b)}{(n-K)!} \, , \qquad K = 0, 1, \cdots, n \, . \tag{28}$$

Weakening the result of the triangle inequality applied to (27),

$$n\|p(\boldsymbol{\beta}) - p\| \leqq n\|\rho(\boldsymbol{\gamma}) - p\| + a \vee b\,. \tag{29}$$

The bound (25) will then follow from the application of Lemma 1 upon showing that

$$\log\left(\bigvee \gamma_K / \bigwedge \gamma_K\right) \leqq \eta\,. \tag{30}$$

Let $\gamma_t = \bigvee \gamma_K$ and $\gamma_s = \bigwedge \gamma_K$ and suppose $t \geqq s$. With $r_K = \gamma_{K+1}\gamma_K^{-1}$, $K = 0, 1, \cdots, n-1$, $\gamma_t\gamma_s^{-1} = \prod_s^{t-1} r_K \leqq \prod_0^{n-1}(r_K \vee 1)$. Since $\log_+ r_K \leqq |r_K - 1|$ and

$$r_K - 1 = \frac{(a-1)(n+1)}{(n+b)(K+1)} - \frac{(b-1)(n+a+b-1)}{(n+b)(n-K+b-1)}\,,$$

appropriate use of the telescoping bounds, $(z+1)^{-1} < \int_z^{z+1} x^{-1}\,dx$, shows LHS (30) is bounded by

$$|a-1|\frac{n+1}{n+b}(1+\log n) + |b-1|\frac{n+a+b-1}{n+b}\left(\frac{1}{b} + \frac{1}{b+1} + \log(n-1)\right). \tag{31}$$

Interchanging K with $n - K$ and a with b in (28) shows that, if $t \leqq s$, LHS (30) is bounded by (31) with a and b interchanged. Thus LHS (30) is always bounded by η. □

REMARK 2. If $a = b = 1$, then $\eta = 0$ and the bound RHS (25) is only slightly weaker than the bound RHS (21) developed directly in the Robbins' prior case.

4. Consistency of $p(\boldsymbol{\beta})$ in the mixture case. When $\boldsymbol{\beta}$ is a mixture of Bernoulli distributions, $p(\boldsymbol{\beta})$ is the mean of a posterior distribution. This distribution is shown to concentrate about a maximum likelihood estimator which is itself shown to be consistent for p.

For given $\mathbf{z}$, let

$$g(\omega) = \sum \log[\omega z_i + (1-\omega)(1-z_i)]\,, \qquad 0 \leqq \omega \leqq 1 \tag{33}$$

where here, and henceforth, all sums on i are from 1 to n. Since L_K defined in (4) is given by

$$L_K = \binom{n}{K}^{-1} \sum \prod_1^K z_{i_j} \prod_{K+1}^n (1 - z_{i_j}) \tag{34}$$

where the sum is over all $\binom{n}{K}$ partitions of $\{1, 2, \cdots, n\}$ into subsets $\{i_1, \cdots, i_K\}$, $\{i_{K+1}, \cdots, i_n\}$, it follows that

$$e^{g(\omega)} = \sum \binom{n}{K}\omega^K(1-\omega)^{n-K}L_K\,. \tag{35}$$

When $\boldsymbol{\beta}$ is a mixture,

$$\beta_K = \int \binom{n+1}{K}\omega^K(1-\omega)^{n-K+1}\,d\Lambda\,, \qquad K = 0, 1, \cdots, n+1\,, \tag{36}$$

we write $p(\boldsymbol{\beta}) = p(\Lambda)$ and note that

$$\sum K\beta_K L_{K-1} = \int \omega e^{g(\omega)}\,d\Lambda\,, \qquad \sum (n-K+1)\beta_K L_K = \int (1-\omega)e^{g(\omega)}\,d\Lambda\,. \tag{37}$$

Hence, by (3) and (5),

$$p(\Lambda) = \int \omega \, dQ \,, \tag{38}$$

where Q is the (posterior) probability measure on [0, 1] with density

$$e^g / \int e^g \, d\Lambda \qquad \text{with respect to} \quad \Lambda \,. \tag{39}$$

Let $\hat{\omega} = \hat{\omega}(\mathbf{z})$ be a maximizer with respect to ω of the concave g defined in (33) ($g'' < 0$ except for $z_1 = \cdots = z_n = \frac{1}{2}$) and record for later use the bound and Fubini representation

$$|p(\Lambda) - \hat{\omega}| \leqq \int |\omega - \hat{\omega}| \, dQ = \int_0^1 Q[|\omega - \hat{\omega}| > t] \, dt \,. \tag{40}$$

To bound the integrand in RHS (40), we will use the Taylor expansion about $\hat{\omega}$ for $\omega \in [0, 1]$ with $g(\omega) > -\infty$,

$$g(\omega) = g(\hat{\omega}) + (\omega - \hat{\omega})g'(\hat{\omega}) - \tfrac{1}{2}(\omega - \hat{\omega})^2 \sum \frac{(2z_i - 1)^2}{[\omega^* z_i + (1 - \omega^*)(1 - z_i)]^2} \tag{41}$$

for some ω^* between ω and $\hat{\omega}$. Since $(\omega - \hat{\omega})g'(\hat{\omega}) \leqq 0$ and $0 < \omega^* z_i + (1 - \omega^*)(1 - z_i) \leqq 1$,

$$g(\omega) - g(\hat{\omega}) \leqq -\tfrac{1}{2}(\omega - \hat{\omega})^2 S \qquad \text{for all} \quad \omega \in [0, 1] \,, \tag{42}$$

where

$$S = \sum (2z_i - 1)^2 \,. \tag{43}$$

Note that Δ defined by (7) is given by $\int (2z - 1) \, dF_1 = \int (1 - 2z) \, dF_0$ so that

$$\int (2z - 1)^2 \, dF_0 \wedge \int (2z - 1)^2 \, dF_1 \geqq \Delta^2 > 0 \,, \tag{44}$$

and, therefore,

$$||S|| \geqq n\Delta^2 \,. \tag{45}$$

LEMMA 2. *For $\Lambda = U$, the uniform measure on* [0, 1], *and Q defined by* (39),

$$||\int |\omega - \hat{\omega}| \, dQ|| \leqq \pi^{\frac{1}{2}} \Delta^{-1} n^{-\frac{1}{2}} + e^{-2n\Delta^4} \,. \tag{46}$$

PROOF. Lemma B of the appendix shows that for $0 < t < 1 - \hat{\omega}$,

$$Q[\omega > \hat{\omega} + t] \leqq Q[\omega > \hat{\omega}] \exp\{g(\hat{\omega} + t) - g(\hat{\omega})\}$$

and, applied to $1 - \omega$, shows that for $0 < t < \hat{\omega}$,

$$Q[\omega < \hat{\omega} - t] \leqq Q[\omega < \hat{\omega}] \exp\{g(\hat{\omega} - t) - g(\hat{\omega})\} \,.$$

Using (42) it follows that

$$Q[|\omega - \hat{\omega}| > t] \leqq e^{-\frac{1}{2}t^2 S} \qquad \text{for} \quad 0 < t < 1 \,. \tag{47}$$

Applying Lemma C of the Appendix to centered $-\frac{1}{2}t^2(2z_i - 1)^2$ and using (45) gives

$$\log ||e^{-\frac{1}{2}t^2 S}|| \leqq -\tfrac{1}{2} n\Delta^2 t^2 + nt^4/32 \,. \tag{48}$$

For $t^2 \leqq 8\Delta^2$, RHS (48) $\leqq -n\Delta^2 t^2/4$. Since LHS (48) decreases with respect to t^2, RHS (48) evaluated at $t^2 = 8\Delta^2$, namely $-2n\Delta^4$, $\geqq$ LHS (48) for $t^2 > 8\Delta^2$. Therefore, partitioning the integration indicated in RHS (40) at $8^{\frac{1}{2}}\Delta \wedge 1$ and applying (47), the Fubini theorem, and the above bounds, we obtain

$$(49) \qquad ||\textstyle\int |\omega - \hat{\omega}|\, dQ|| \leqq \int_0^{8^{\frac{1}{2}}\Delta\wedge 1} e^{-n\Delta^2 t^2/4}\, dt + (1 - 8^{\frac{1}{2}}\Delta)_+ e^{-2n\Delta^4},$$

completing the proof. □

COROLLARY 1. *The* MLE $\hat{\omega}$ *is* L_1 *consistent for* p *at a rate* $O(n^{-\frac{1}{2}})$ *uniformly in* **P**. *Specifically*,

$$(50) \qquad ||\hat{\omega} - p|| \leqq \text{RHS}\,(46) + \text{RHS}\,(21)\,.$$

PROOF. By Lemma 2 and (40), $||p(U) - \hat{\omega}|| \leqq$ RHS (46). Since $p(\boldsymbol{\beta})$ of Theorem 1 is $p(U)$, (50) follows from the triangle inequality about $p(U)$. □

THEOREM 3. *If* Λ *has a density* λ *with respect to* U *with* (ess sup λ)/(ess inf λ) $=$ $c < \infty$, *then* $p(\Lambda)$ *is* L_1 *consistent for* p *at a rate* $O(n^{-\frac{1}{2}})$ *uniformly in* **P**. *Specifically*,

$$(51) \qquad ||p(\Lambda) - p|| \leqq (c + 1)\,\text{RHS}\,(46) + \text{RHS}\,(21)\,.$$

PROOF. Displaying the dependence of Q of (39) on Λ, we see that $Q_\Lambda \leqq cQ_U$. Hence, by (40) and Lemma 2, it follows that $||p(\Lambda) - \hat{\omega}|| \leqq c\,$RHS (46) and the proof follows from Corollary 1 and the triangle inequality about $\hat{\omega}$. □

The key to the proof of Lemma 2 is the bound (47) for the posterior probability $Q_U[|\omega - \hat{\omega}| > t]$. An analysis of Q_Λ leads to a bound which can be used to establish the L_1 consistency of $p(\Lambda)$ for a larger class of Λ than covered by Theorem 3 but with some loss of rate. For $0 < t < 1 - \hat{\omega}$, the line $g(\hat{\omega}) + s(\omega - \hat{\omega})$ where

$$(52) \qquad st = g(\hat{\omega} + t) - g(\hat{\omega})\,,$$

is below the concave g on $(\hat{\omega}, \hat{\omega} + t]$ and above it on $(\hat{\omega} + t, 1]$ so that by (39)

$$(53) \qquad \frac{Q[\omega > \hat{\omega} + t]}{Q[\hat{\omega} < \omega \leqq \hat{\omega} + t]} \leqq \frac{\int [\omega > \hat{\omega} + t] e^{s(\omega - \hat{\omega})}\, d\Lambda}{\int [\hat{\omega} < \omega < \hat{\omega} + t/2] e^{s(\omega - \hat{\omega})}\, d\Lambda} \leqq \frac{e^{st}\Lambda(\hat{\omega} + t, 1]}{e^{\frac{1}{2}st}\Lambda(\hat{\omega}, \hat{\omega} + t/2)}\,.$$

With

$$(54) \qquad L(2v) \equiv \inf\{\Lambda(a, b)\,|\,0 < a,\, a + v = b < 1\} \qquad \text{for} \quad 0 < v < 1\,,$$

(52) and (42) show that RHS (53) $\leqq (L(t))^{-1} \exp -\frac{1}{4}t^2 S$ and, since $(1 + x^{-1})^{-1}$ is increasing in $x > 0$,

$$(55) \qquad Q[\omega > \hat{\omega} + t] \leqq Q[\omega > \hat{\omega}](1 + L(t)\exp \tfrac{1}{4}t^2 S)^{-1}\,.$$

Treating the left tail by symmetry and combining the result with (55) gives

$$(56) \qquad Q[|\omega - \hat{\omega}| > t] < (1 + L(t)\exp \tfrac{1}{4}t^2 S)^{-1} \qquad \text{for} \quad 0 < t < 1\,.$$

Partitioning the integration indicated in RHS (40) at ε and using weakened (56) on each part,

$$\int |\omega - \hat{\omega}|\, dQ \leqq \varepsilon + (L(\varepsilon))^{-1} \exp -\tfrac{1}{4}\varepsilon^2 S \qquad \text{for} \quad 0 < \varepsilon < 1\,. \tag{57}$$

LEMMA 3. *If* Λ *is such that* $L(\varepsilon) \geqq c\varepsilon^r$ *for* $0 < \varepsilon < 1$ *for some* $c, r > 0$, *then, with* A *abbreviating* $2(r+1)\Delta^{-2}$,

$$\|\int |\omega - \hat{\omega}|\, dQ\| \leqq n^{-\frac{1}{2}}\{(A \log n)^{\frac{1}{2}} + c^{-1}(A \log n)^{-r/2} \exp A^2 e^{-2}/32\}\,. \tag{58}$$

PROOF. Weakening (57) by the hypothesis, the triangle inequality and the bound (48) with $t^2 = \frac{1}{2}\varepsilon^2$ give

$$\|\int |\omega - \hat{\omega}|\, dQ\| \leqq \varepsilon + c^{-1}\varepsilon^{-r} \exp\{-n\Delta^2\varepsilon^2/4 + n\varepsilon^4/128\}\,. \tag{59}$$

Specifying $\varepsilon^2 = An^{-1} \log n$ and using $n^{-1}(\log n)^2 \leqq 4e^{-2}$ completes the proof. □

For sufficiently diffuse prior, we obtain the L_1 consistency with a logarithmically weakened rate as an immediate consequence of Lemma 3 together with Corollary 1.

THEOREM 4. *If* Λ *satisfies the hypothesis of Lemma* 3, *then* $p(\Lambda)$ *is* L_1 *consistent for* p *at the rate* $O((n/\log n)^{-\frac{1}{2}})$ *uniformly in* **P**. *Specifically*,

$$\|p(\Lambda) - p\| \leqq \text{RHS (58)} + \text{RHS (50)}\,. \tag{60}$$

PROOF. The proof follows from (40), Lemma 3, Corollary 1 and the triangle inequality about $\hat{\omega}$. □

We conclude with the following remarks about the function L.

REMARK 3. If Λ satisfies the hypothesis of Lemma 3, then $c \leqq 1 \leqq r$. The left inequality follows from the fact $c \leqq L(1-) \leqq 1$. The right inequality follows from the fact that, for every integer J, $1 \geqq L(1) \geqq 2^J L(2^{-J}) \geqq 2^J c 2^{-rJ}$.

REMARK 4. If Λ is the Beta distribution with parameters $a, b > 0$, then $L(\varepsilon) \geqq c\varepsilon^r$ for $c > 0$ and $r = a \vee b \vee 1$. In case either $a \neq 1$ or $b \neq 1$, then RHS (26) is ratewise the same as RHS (60) whereas if $a = b = 1$, RHS (26) is $O(n^{-\frac{1}{2}})$.

APPENDIX

Lemmas A and B to follow establish that log concave discrete and Lebesgue densities have increasing hazard rate (IHR). Bounds on tail probabilities follow from IHR and are used in the proofs of Lemma 1 (discrete case) and Lemma 2 (Lebesgue case). Barlow and Prochan (1965) have indicated following their Definition 3, page 24, the equivalence of log concavity and Pólya frequency function of order 2 for Lebesgue densities and, in their Theorem 1 (Appendix, page 229) have a result implying that PF_2 densities have IHR. Karlin (1968) in his Proposition 1.2, page 332, establishes that PF_2 implies log concavity but doesn't include the converse.

Lemma C and its corollary are results from Hoeffding (1963) in forms which readily apply in the proofs of Lemma 1 and Lemma 2.

DEFINITION A. A sequence $L_i \geqq 0$ is *log concave* if $[i \mid L_i > 0]$ is an interval of integers and $L_{i+1}/L_i \downarrow$ on that interval.

LEMMA A. *Let* $L_i \geqq 0$ *be log concave with* $[i \mid L_i > 0] = I$. *Then* $L_m/(\sum_m^\infty L_i) \uparrow$ *with respect to* $m \in I$.

PROOF. If $k \geqq m$ then $L_{k+j}L_m \leqq L_k L_{m+j}$ for all $j \geqq 0$ by the log-concavity. Adding these inequalities gives $L_m \sum_k^\infty L_i \leqq L_k \sum_m^\infty L_i$. □

DEFINITION B. A function $f \geqq 0$ on the reals is *log concave* if $I = \{x \mid f(x) > 0\}$ is an interval and $\log f$ is concave on I.

LEMMA B. *Let* f *be log concave and integrable and let* $J(v) = \int_v^\infty f(x)\,dx$ *for* $v \in I$. *Then* $f/J \uparrow$ *on* I.

PROOF. Let $g = \log f$ and let $'$ denote the right derivative. On the right interior of I, $(\log(e^g/J))' \equiv g' + e^g/J \geqq 0$ since $g(x) \leqq g(v) + (x - v)g'(v)$ for $v < x \in I$ so that, if $g'(v) < 0$, $e^{-g(v)}J(v) \leqq \int_v^\infty e^{(x-v)g'(v)}\,dx = |g'(v)|^{-1}$. □

The following lemma is a useful combination of Lemma 1 of Hoeffding (1963) and the proof of his Theorem 2.

LEMMA C. *If* Z *is a random variable with* $EZ = 0$ *and* $R = \operatorname{ess\,sup} Z - \operatorname{ess\,inf} Z$, *then*

$$Ee^Z \leqq pe^{-qR} + qe^{pR} \leqq e^{R^2/8}$$

with $p = 1 - q = R^{-1} \operatorname{ess\,sup} Z$.

PROOF. By convexity of exp, $\mathrm{R}e^Z \leqq (pR - Z)e^{-qR} + (Z + qR)e^{pR}$ and expectation then gives the first inequality. Letting $L = -qR + \log(p + qe^R)$ denote the log of the middle term and $'$ denote d/dR,

$$L' = -q + \frac{qe^R}{p + qe^R} \qquad L'' = \frac{pqe^R}{(p + qe^R)^2} \leqq \frac{1}{4}$$

and thus $L \leqq 0 + 0 \cdot R + \frac{1}{4}R^2/2!$ by the Taylor theorem. □

COROLLARY (Theorem 2, Hoeffding (1963)). *If* $X_1, \cdots, X_n$ *are independent random variables with zero means and finite ranges, then* $\forall\, T \geqq 0$

$$P[\sum X_i \geqq T] \leqq \exp\{-2T^2/\sum R_i^2\}.$$

PROOF. For $0 < h$, LHS $\leqq E \exp h(\sum X_i - T)$. By independence and the application of Lemma C to each hX_i, these bounds do not exceed $\exp\{h^2 \sum R_i^2/8 - hT\}$. Minimization with respect to h completes the proof. □

Addendum. The results (Theorems 2 and 5) of Inglis (1973) on uniform a.s. convergence to 0 of $p(\Lambda) - p$ came to our attention after our results were complete. (Our proof of Theorem 1 was done in 1971; some aspects of its natural generalization were the objects of abortive interim efforts.) In the following remarks, we (I) record a simplified version of his proof, (II) indicate how such results follow from our work and (III) reinterpret our simplified version for nonfinite states.

(I) With $F_\omega \equiv \omega F_1 + (1-\omega)F_0$ and $Z_\omega \equiv \omega z + (1-\omega)(1-z)$ for $\omega \in [0, 1]$, let

(i) $$\mathscr{U}_\delta = \{\omega \mid \int \log (Z_p/Z_\omega)\, dF_p < \delta\}\,.$$

From well-known properties of Kullback–Liebler information (cf. Lemma 1 of Hannan (1960))

(ii) $$(pe^{-\delta}, 1-(1-p)e^{-\delta}) \subset \mathscr{U}_\delta \subset \mathscr{U}_{2\delta} \subset (p-\eta, p+\eta)$$

for $\eta > 0$ and $8\delta = \eta^2(\int d|F_1 - F_0|)^2$. Abbreviating $\bigvee_\omega |n^{-1}g(\omega) - \int \log Z_w\, dF_p|$ to $\mathscr{V}$, crude bounds based on the above inclusions give

(iii) $$\frac{Q(p-\eta, p+\eta)^c}{Q(p-\eta, p+\eta)} \leqq \frac{e^{-2n\delta + n\mathscr{V}}}{L(2(1-e^{-\delta}))e^{-n\delta - n\mathscr{V}}}\,.$$

It is easily shown that L is positive if Λ is positive on every nondegenerate subinterval of [0, 1]. The above bound then shows that, as $n \to \infty$, Q concentrates near p provided $\mathscr{V}$ converges to 0.

The uniform a.s. form of the latter is furnished by the Wald–Le Cam SLLN, slightly upgraded to treat this independent nonidentically distributed case and to establish the uniformity in **P**. The upgrading to finitely many possible distributions and establishment of uniformity are standard in the compound problem. They result from application to norms of sums of i.i.d. random variables of the trivial fact that, if $'$ denotes a map of the positive integers into the nonnegative integers with $n' \leqq n$ and $0 = S_0, S_1, \cdots$ is a sequence of numbers, then uniformly over the class of such maps $S_{n'}$ is $o(n)$ if S_n is.

(II) The uniform a.s. convergence to 0 of $p(\Lambda) - p$ is a roundabout corollary to our work. It follows from adding the bounds of the corollary of the appendix (and (45)) that

(iv) $$\sum_{n>m} \mathbf{P}\left[\frac{n}{S} \geqq 2\Delta^{-2}\right] \leqq 2\Delta^{-4} \exp - \tfrac{1}{2} m\Delta^4\,.$$

Thus, $S \to \infty$ a.s. uniformly in **P** and, from (57), $p(\Lambda) - \hat{\omega} \to 0$ a.s. uniformly in **P** provided L is positive. The difference $\hat{\omega} - p$ is treated by first noting that Corollary 1 gives a constant C such that for every r,

(v) $$\mathbf{P}[\bigvee \{|\hat{\omega} - p| \mid n \in \{r^3, (r+1)^3, \cdots\}\} \geqq \varepsilon] \leqq C\varepsilon^{-1}r^{-\frac{1}{2}}\,.$$

By analyzing g it can be shown that for all integers $1 \leqq m \leqq o$,

(vi) $$\bigvee_{m \leqq n \leqq o} \{|\hat{\omega}_n - \hat{\omega}_m| \vee |\hat{\omega}_n - \hat{\omega}_o|\} \leqq 3\left(\frac{o}{m} - 1\right) \bigvee_{n \geqq m} \frac{n}{S}\,.$$

This together with (iv) gives the uniform stability of $\hat{\omega}$ between cubes and since p is uniformly stable between cubes, it follows that $\hat{\omega} - p \to 0$ a.s. uniformly in **P**.

(III) Some of (I) applies to compact dominated $\mathscr{P} = \{F_\theta \mid \theta \in \Theta\}$ with $z(\theta) = dF_\theta/d\mu$ replacing z, $1-z$. Reinterpreting ω as a probability on Θ belonging to a class Ω including all those with finite support, Z_ω as $\int z\,d\omega$ and p as the empiric distribution of $\theta_1, \cdots, \theta_n$, the leftmost inclusion of (ii) is not available but a more basic form of (iii) is immediate:

$$\text{(iii)}' \qquad \frac{Q(\mathscr{U}_{2\delta}^c)}{Q(\mathscr{U}_\delta)} \leqq \frac{e^{-2n\delta+n\mathscr{V}}}{\Lambda(\mathscr{U}_\delta)e^{-n\delta-n\mathscr{V}}}\,.$$

If $\forall\,\delta > 0$, $\Lambda(\mathscr{U}_\delta)$ is bounded away from 0 uniformly in p, (iii)′ then shows that, as $n \to \infty$, Q concentrates "near" p again provided $\mathscr{V}$ converges to 0.

REFERENCES

BARLOW, RICHARD E. and PROSCHAN, FRANK (1965). *Mathematical Theory of Reliability.* Wiley, New York.

GILLILAND, DENNIS C. and HANNAN, JAMES (1974). The finite state compound decision problem, equivariance and restricted risk components. To be published *Ann. Statist.*

HANNAN, JAMES (1960). Consistency of maximum likelihood estimation of discrete distributions. In *Contributions to Probability and Statistics, Hotelling Volume.* Stanford Univ. Press.

HANNAN, JAMES F. and ROBBINS, HERBERT (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37-51.

HOEFFDING, WASSILY (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

HUANG, J. S. (1972). A note on Robbins' compound decision problem. *Ann. Math. Statist.* **43** 348-350.

INGLIS, JAMES (1973). Admissible decision rules for the compound problem. TR 37, Dept. of Statistics, Stanford Univ.

KARLIN, SAMUEL (1968). *Total Positivity* **1**. Stanford Univ. Press.

ROBBINS, HERBERT (1951). Asymptotically sub-minimax solutions of compound decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131-148, Univ. of California Press.

SAMUEL, ESTER (1967). The compound statistical decision problem. *Sankhyā* **29** 123-140.

SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36** 1272-1278.

SHAPIRO, C. P. (1972). Bayesian classification. Doctoral thesis, Univ. of Michigan.

SHAPIRO, C. P. (1974). Bayesian classification: asymptotic results. *Ann. Statist.* **2** 763-774.

DENNIS C. GILLILAND AND JAMES HANNAN
DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824

J. S. HUANG
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO N1G 2W1 CANADA