*The Annals of Statistics*
1993, Vol. 21, No. 2, 736–745

# ON EQUIVARIANCE AND THE COMPOUND DECISION PROBLEM

BY MOSTAFA MASHAYEKHI

*University of Nebraska*

This paper obtains some extensions of Gilliland and Hannan's results on equivariance and the compound decision problem.

Consider a compound decision problem with restricted component risk and component distributions in a norm compact set of mutually absolutely continuous probability measures. Then the method of proof of a theorem of Gilliland and Hannan translates the results of Mashayekhi on symmetrization of product measures into uniform convergence to zero of the excess of the simple envelope over the equivariant envelope.

Our envelope results strengthen, among other things, the results of Datta who obtained admissible asymptotically optimal solutions to the compound estimation problem for a large subclass of the real one parameter exponential family under squared error loss.

Sufficient conditions for asymptotic optimality of "delete bootstrap" rules are given and, for squared error loss estimation of continuous functions and for finite action space problems with continuous loss functions, the problem of treating the asymptotic excess compound risk of Bayes compound rules is reduced to the question of $L_1$-consistency of certain mixtures.

Examples of estimates satisfying the above consistency condition are provided.

**1. Introduction, notation and history.** In the set version of the compound decision problem, pioneered by Robbins (1951), simultaneous decisions are to be made in $n$ problems of the same generic structure, with this structure being possessed by what is called the component problem.

Let $\mathscr{P}$ be a class of probability measures on a measurable space $(\mathscr{X}, \mathscr{B})$ and consider a component problem with action space $\mathscr{A}$ and an observable $\mathscr{X}$-valued random element $X$ with distribution $P \in \mathscr{P}$. Let $\mathscr{D}$ be a bounded risk class of decision rules for the component problem and let $M < \infty$ be such that $R(t, P) \le M \ \forall\ t \in \mathscr{D}$ and $\forall\ P \in \mathscr{P}$, where $R(t, P)$ denotes the risk of $t$ at $P$.

For an $n$-tuple $\mathbf{x} = (x_1, \dots, x_n)$ let $\mathbf{x}_{\breve{\alpha}}$ denote $\mathbf{x}$ with the $\alpha$th component deleted. Consider the class $\mathscr{D}$ of compound rules $\mathbf{t} = (t_1, \dots, t_n)$ where each $\mathbf{x}_{\breve{\alpha}}$-section of $t_\alpha \in \mathscr{D}$. When $\mathscr{D}$ is the largest class of decision rules for the component problem, the above compound problem is the usual compound problem with $\mathscr{D}$ the largest class of compound decision rules. The compound problem with restricted component risk was considered by Gilliland and Hannan (1974, 1986) for finite $\mathscr{P}$ because of the generality it provided for

Received June 1991; revised June 1992.
*AMS* 1991 *subject classifications*. Primary 62C25; secondary 62A05.
*Key words and phrases*. Equivariance, compound decision, equivariant envelope, simple envelope, asymptotically optimal, delete bootstraps rules, consistent mixtures, minimum distance.

their envelope results and the fact that it is the natural setting in which to study "delete bootstrap" procedures. Moreover, as they noted, it allows for choice of $\mathscr{D}$ to control component risk behavior and the construction of asymptotically best equivariant procedures in $\mathscr{D}$.

Let $s$ be the function on $\mathscr{D} \times \{1, \dots, n\} \times \mathscr{P} \times \mathscr{X}^{n-1}$ such that $s(\mathbf{t}, \alpha, P, \mathbf{X}_{\check{\alpha}})$ is the conditional on $\mathbf{X}_{\check{\alpha}}$ risk incurred by $\mathbf{t}$ in the component $\alpha$ when the distribution of $X_\alpha$ is $P$.

It is well known [see Section 2 of Hannan and Huang (1972a) or Section 1 of Gilliland and Hannan (1974, 1986)] that the compound problem is invariant under the group of $n!$ permutations of coordinates, and that a compound rule $\mathbf{t}$ is equivariant if and only if there exists a function $\gamma$ on $\mathscr{X} \times \mathscr{X}^{n-1}$ to $\mathscr{A}$, symmetric on $\mathscr{X}^{n-1}$, such that $t_\alpha(\mathbf{x}) = \gamma(x_\alpha, \mathbf{x}_{\check{\alpha}})$ for all $\alpha$. The latter implies that if $\mathbf{t}$ is equivariant then $s$ is constant in its second argument and symmetric in its fourth argument. The implied property for $s$ will be used as a definition of equivariance when we bypass the consideration of a loss function and identify each decision rule in $\mathscr{D}$ with its risk point in $[0, M]^{\mathscr{P}}$. For equivariant procedures we will abbreviate $s(\mathbf{t}, \alpha, P, \cdot\,)$ by using the affixes on $\mathbf{t}$ and $P$. For example $s(\hat{\mathbf{t}}, 1, P_\alpha, \cdot\,)$ will be abbreviated to $\hat{s}_\alpha$.

Let $\mathscr{E}$ and $\mathscr{S}$ denote the class of all equivariant rules in $\mathscr{D}$ and the class of all simple rules in $\mathscr{D}$, respectively. Let $\mathbf{R}(\mathbf{t}, \mathbf{P})$ denote the compound risk of $\mathbf{t}$ at $\mathbf{P}$. The equivariant envelope corresponding to $\mathscr{D}$ is defined by

$$\tilde{\psi}(\mathbf{P}) = \inf_{\mathbf{t} \in \mathscr{E}} \mathbf{R}(\mathbf{t}, \mathbf{P}) \tag{1}$$

and the simple envelope corresponding to $\mathscr{D}$ is defined by

$$\psi(\mathbf{P}) = \inf_{\mathbf{t} \in \mathscr{S}} \mathbf{R}(\mathbf{t}, \mathbf{P}). \tag{2}$$

Clearly $\tilde{\psi}$ is the infimum over a larger class and it follows from the definition that $\psi(\mathbf{P}) = R(G_n)$, where $R(\omega)$ is the component Bayes restricted risk at $\omega$ and $G_n$ denotes the empirical distribution of $\mathbf{P}$.

Traditionally, a compound rule is called asymptotically optimal if, with the modified regret at $\mathbf{P}$ defined by

$$D_n(\mathbf{t}, \mathbf{P}) = \mathbf{R}(\mathbf{t}, \mathbf{P}) - R(G_n), \tag{3}$$

$\sup_{\mathbf{P}} D_n(\mathbf{t}, \mathbf{P}) \to 0$ as $n \to \infty$.

However, since almost all of the compound rules in the literature are equivariant, the equivariant envelope [cf. Hannan and Huang (1972a), page 104] is considered a more appropriate yardstick of performance than the simple one.

Hannan and Robbins (1955) introduced the class of equivariant procedures for the compound $2 \times 2$ $\mathscr{P} \times \mathscr{A}$ problem and showed (Theorem 5) that the difference between the simple and equivariant envelopes converges to zero uniformly in $\mathbf{P}$. Hannan and Huang (1972a) considered the compound problem for finite $\mathscr{P}$ under a certain class of loss functions and provided an upper bound on the difference of the simple and equivariant envelopes which is $O(n^{-1/2})$. Gilliland and Hannan [(1974, 1986), Theorems 1 and 2] extended those results to arbitrary bounded risk components for finite $\mathscr{P}$. They also showed (Theorems 3 and 4) that for equivariant "delete bootstrap" procedures, the excess compound risk over the simple envelope is bounded in terms of the $L_1$ error of estimation and thus established a large class of asymptotic solutions to the compound decision problem with restricted risk and finite state component. Their proof depended heavily on the Hannan and Huang (1972b) results on the stability of symmetrization of product measures (Theorem 3) which was a strengthened generalization of Theorem II.1 of Hannan (1953).

In this paper we consider the compound decision problem in which the set of component distributions $\mathscr{P}$ is compact in the topology induced by the total variation norm and has pairwise mutually absolutely continuous elements.

Section 3 deals with the difference of the two envelopes and asymptotic optimality. In Proposition 1 we observe that the method of proof of Theorem 1 of Gilliland and Hannan (1974, 1986) provides a setting for a straightforward application of the results of Mashayekhi (1990) to obtain asymptotic equivalence of the simple and equivariant envelopes.

Theorem 1 provides sufficient conditions for asymptotic optimality of "delete bootstrap" rules. Examples 1 and 2 show that for squared error loss estimation of continuous functions and for finite $\mathscr{A}$ problems with continuous loss functions Theorem 1 reduces the problem of treating the asymptotic excess compound risk of Bayes compound rules to the question of $L_1$-consistency of certain mixtures. The reduction is analogous to Theorem 3 of Gilliland and Hannan (1974, 1986) and, combined with the results of Datta (1991a), provides the first solution to the compound problem in the most general compact parameter case.

Section 4 provides, as examples, two classes of mixtures those satisfy the required consistency condition. The first example is the class of mixtures based on hyperpriors obtained in Datta (1991a), thus showing that his results for empirical Bayes problems are extended to the corresponding compound problems under appropriate loss functions. A mixture in the second class is obtained by minimizing an $L_2$-distance.

**2. Notation and conventions.** Let $n$ be a positive integer. An $n$-tuple $(x_1, \dots, x_n)$ is denoted by $\mathbf{x}_n$ (the subscript $n$ will not be exhibited if it is clear from the context). If the components of $\mathbf{P}$ are probability measures, then $\mathbb{P}$ denotes the product probability measure $P_1 \times \cdots \times P_n$ and $\mathbb{P}_{\check{\alpha}}$ denotes $\mathbb{P}$ with the $\alpha$th factor deleted.

We use $\mu(f)$ or $\mu f$ to denote the integral of a function $f$ with respect to (wrt hereafter) a signed measure $\mu$. We sometimes use expressions such as $\int f(x)\, d\mu(x)$ to exhibit dummy variables. The same notation is used for a set and its indicator function when the distinction is clear from the context. In particular [ ] is used to denote an indicator function.

For any signed measure $\mu$ on $\mathscr{B}^n$, $\mu^*$ denotes symmetrization of $\mu$ with respect to the group of transformations on $(\mathscr{X}, \mathscr{B})^n$ induced by the group of permutations on $n$ objects.

If $\tau$ is a signed measure, then $\|\tau\|$ will denote the total variation norm of $\tau$. The absolute supremum of a real function $f$ is denoted by $\|f\|_\infty$ whatever be its domain.

All incompletely described limits are as $n \to \infty$ through positive integers.

**3. Asymptotically optimal "delete bootstrap" rules.** Proposition 1, to follow, shows how Theorems 1 and 2 of Mashayekhi (1990) can be translated into the convergence to zero of the excess of the simple envelope over the equivariant envelope. This new infinite state case strengthens many previous compound results by implying that, for all asymptotically optimal decision rules, the modified regret defined by the more stringent equivariant envelope converges to zero, uniformly on compact cubes.

Theorem 1 of Mashayekhi (1990) which is used in the proof of Theorem 1 of this paper is restated below:

THEOREM [Mashayekhi (1990)]. *Let $\mathscr{P}$ be a norm-compact class of mutually absolutely continuous probability measures and let $\mathbb{P} = \times_{i=1}^{n} P_i$, where each $P_i \in \mathscr{P}$. Let $\tau = R - S$ with $R$ and $S \in \mathscr{P}$. Then*

$$\sup\{\|(\tau \times \mathbb{P})^*\| : (R, S, P_1, \dots, P_n) \in \mathscr{P}^{n+2}\} = o(1). \tag{4}$$

Rates of convergence of the lhs of (4), for the case where $\mathscr{P}$ is an exponential family with its parameter space a compact subset of the interior of the natural parameter space, were obtained in Theorem 2 of Mashayekshi (1990).

PROPOSITION 1.

$$\psi - \tilde{\psi} \le M \sup_{\mathbf{P}} \|(\mathbb{P}_{\check{1}} - \mathbb{P}_{\check{n}})^*\|. \tag{5}$$

The proof is essentially the proof of Theorem 1 of Gilliland and Hannan (1974, 1986). If $\mathbf{t} \in \mathscr{E}$,

$$\mathbb{P}_{\check{n}} s_\alpha - \mathbb{P}_{\check{\alpha}} s_\alpha \le M \|(\mathbb{P}_{\check{\alpha}} - \mathbb{P}_{\check{n}})^*\| \tag{6}$$

because of the boundedness and symmetry of $s_\alpha$. Therefore

$$\mathbb{P}_{\check{n}}\left(n^{-1} \sum_\alpha s_\alpha\right) - n^{-1} \sum_\alpha \mathbb{P}_{\check{\alpha}} s_\alpha \le \text{ the rhs of (5)}.$$

Since $n^{-1}\Sigma_\alpha s_\alpha \ge \psi(\mathbf{P})$,

$$\psi(\mathbf{P}) - \mathbf{R}(\mathbf{t}, \mathbf{P}) \le \text{the rhs of (5)}. \tag{7}$$

We obtain (5), since (7) holds for every $\mathbf{t} \in \mathscr{E}$. □

Consider $\mathscr{P}$ with the topology induced by the total variation norm and let $\Omega$ be the set of all probability measures on Borels of $\mathscr{P}$. For each $\omega \in \Omega$ the mixture $\mathscr{P}_\omega$ is the measure on $\mathscr{X}$ defined by

$$\mathscr{P}_\omega(B) = \omega(P(B)), \qquad B \in \mathscr{B}.$$

For each $\omega \in \Omega$, let $t_\omega$ be a Bayes solution versus $\omega$ in the component problem. Considered as a function on $\Omega$ into $\mathscr{D}$ [cf. Hannan (1957), page 101], $t$ is called a Bayes response.

THEOREM 1. *Let $t$ be a Bayes response, $\widehat{\omega}$ a symmetric mapping on $\mathscr{X}^{n-1}$ into $\Omega$. Let $\hat{\mathbf{t}}$ be the compound rule with*

$$\hat{t}_\alpha(\mathbf{x}) = t_{\widehat{\omega}(\mathbf{x}_{\check{\alpha}})}(x_\alpha)$$

*and let $\tilde{\mathbf{t}}$ be an equivariant rule with its components Bayes versus $G_n$. Then $\hat{\mathbf{t}}$ is asymptotically optimal if*:

(i) *For each $\varepsilon > 0$, $\exists\ \delta > 0$ such that $\forall\ n$*

$$\left(\|\mathscr{P}_{\widehat{\omega}} - \mathscr{P}_{G_n}\| < \delta\right) \Rightarrow \left(G_n(\hat{s} - \tilde{s}) < \varepsilon\right)$$

*and*

(ii) $\sup\{\mathbb{P}_{\check{n}}\|\mathscr{P}_{\widehat{\omega}} - \mathscr{P}_{G_n}\|: \mathbf{P} \in \mathscr{P}^n\} = o(1)$.

PROOF. $\hat{s}_\alpha$ is symmetric by symmetry of $\widehat{\omega}$. Therefore, as in Proposition 1,

$$\mathbb{P}_{\check{\alpha}}\hat{s}_\alpha - \mathbb{P}_{\check{n}}\hat{s}_\alpha \le \text{the rhs of (5)}. \tag{8}$$

Subtracting $\psi(\mathbf{P})$ from both sides of (8) and averaging over the components give

$$\mathbf{R}(\hat{\mathbf{t}}, \mathbf{P}) - \psi(\mathbf{P}) \le \mathbb{P}_{\check{n}} G_n(\hat{s} - \tilde{s}) + \text{the rhs of (5)}. \tag{9}$$

A triangulation about $\tilde{\psi}(\mathbf{P})$ together with (5) give

$$\mathbf{R}(\hat{\mathbf{t}}, \mathbf{P}) - \tilde{\psi}(\mathbf{P}) \le \mathbb{P}_{\check{n}} G_n(\hat{s} - \tilde{s}) + 2[\text{the rhs of (5)}]. \tag{10}$$

The rhs of (5) converges to zero by (4). In order to show uniform convergence to zero of the first term on the rhs of (10), let $\varepsilon > 0$. Choose $\delta$ with the property assumed in (i). Since

$$G_n(\hat{s} - \tilde{s}) \le \varepsilon + M\left[\|\mathscr{P}_{\widehat{\omega}} - \mathscr{P}_{G_n}\| \ge \delta\right],$$

the first term on the rhs of (10) is bounded by $\varepsilon + M$ lhs of (ii)$/\delta$ by the Markov inequality. Since $\varepsilon$ is arbitrary, the conclusion follows by (ii). □

Observe that, by Theorem II6.4 of Parthasarathy (1967), $\Omega$ with the topology of weak convergence inherits the compact metricity of $\mathscr{P}$.

Lemma 1 to follow is used in Examples 1 and 2 to show that condition (i) of Theorem 1 is satisfied for a large class of decision problems.

LEMMA 1. *Let $\phi$ be a real continuous function on $\mathscr{P}$. For each $\omega$ let $\nu_\omega$ be the signed measure defined by*

$$\nu_\omega(B) = \int \phi(P) P(B)\, d\omega(P), \qquad B \in \mathscr{B}.$$

*Let $d$ be a metric of weak convergence. Then $\omega \in (\Omega, d) \rightsquigarrow \nu_\omega$, with the norm-topology on the range, is uniformly continuous.*

PROOF. Let $\omega_n$ be a sequence in $\Omega$ converging to $\omega$. Since $\mathscr{P}$ is a compact metric space, it is complete and separable. By the Skorohod representation theorem [Theorem 3.3 of Billingsley (1971)] there exist $\mathscr{P}$ valued random elements $\eta_n$ and $\eta$ on the Lebesgue unit interval with respective $\lambda$-induced distributions $\omega_n$ and $\omega$ such that $\eta_n$ converges to $\eta$ pointwise.

Since $\nu_\omega$ is the $\omega$-mixture of the $\nu_P = \phi(P)P$,

$$\nu_{\omega_n} - \nu_\omega = (\omega_n - \omega)\nu. = \lambda(\nu_{\eta_n} - \nu_\eta). \tag{11}$$

Triangulation about $\phi(\eta_n)\eta$ and simple norm properties give

$$\|\nu_{\eta_n} - \nu_\eta\| \le \|\phi\|_\infty \|\eta_n - \eta\| + |\phi(\eta_n) - \phi(\eta)| \le 4\|\phi\|_\infty. \tag{12}$$

Since variations of a positive mixture are bounded by the mixture of the variations, continuity of $\nu$ at $\omega$ follows from (11) and (12) by bounded convergence theorem application to the rhs of (11). Uniform continuity follows by compactness of $\Omega$. □

In the rest of this paper we assume

$$\Omega \text{ is identifiable: } \omega \rightsquigarrow \mathscr{P}_\omega \text{ is } 1-1, \tag{13}$$

and let $\rho$ denote the metric on $\Omega$ thereby induced by $\|\ \|$ on the range

$$\rho(\omega, \omega') = \|\mathscr{P}_\omega - \mathscr{P}_{\omega'}\|. \tag{14}$$

REMARK 1. If $d$ metrizes weak convergence in $\Omega$, then $d$ is equivalent to $\rho$:

By the $\phi = 1$ case of Lemma 1, $\omega \in (\Omega, d) \rightsquigarrow \mathscr{P}_\omega$ is uniformly continuous on $\Omega$. By the identifiability assumption and compactness of $\Omega$ and metric range [cf. Proposition 9.5 of Royden (1968)] it is a homeomorphism. So is the isometry with $d$ replaced by $\rho$. Thus $d$ and $\rho$ are equivalent.

EXAMPLE 1. Consider the compound decision problem whose component problem is estimation of $\phi(P)$ under squared error loss for a real continuous $\phi$. Under the hypothesis of Theorem 1, $\hat{\mathbf{t}}$ satisfies condition (i) of Theorem 1.

PROOF. Since $(\hat{s} - \tilde{s})(P) = P(\hat{t}^2 - \tilde{t}^2) - 2\phi(P)P(\hat{t} - \tilde{t})$ and $\hat{t}$ and $\tilde{t}$ inherit the bound on $\phi$,

$$\begin{aligned}(G_n - \hat{\omega})(\hat{s} - \tilde{s}) &= (\mathscr{P}_{G_n} - \mathscr{P}_{\hat{\omega}})(\hat{t}^2 - \tilde{t}^2) - 2(\nu_{G_n} - \nu_{\hat{\omega}})(\hat{t} - \tilde{t}) \\ &\le \|\phi\|_\infty^2 \|\mathscr{P}_{G_n} - \mathscr{P}_{\hat{\omega}}\| + 4\|\phi\|_\infty \|\nu_{G_n} - \nu_{\hat{\omega}}\|.\end{aligned} \tag{15}$$

Observe that $\hat{\omega}(\hat{s} - \tilde{s})$ is nonpositive; equivalently

$$G_n(\hat{s} - \tilde{s}) \le \text{the lhs of } (15).$$

The conclusion now follows by the uniform $\rho$ continuity of Lemma 1 with the choice $d = \rho$ justified by Remark 1. □

EXAMPLE 2 (Finite $\mathscr{A}$ and continuous loss functions). Such decision problems satisfy a much stronger property than (i). Note that, the risk of an arbitrary randomized decision rule $t$ assigning probability $t_a$ to action $a$, is given by

$$s(P) = P\sum_a t_a L_a(P),$$

where $L_a(P)$ is the value of the loss function at $(a, P)$. Then, with $\phi(P) = L_a(P)P$,

$$\begin{aligned}(G_n - \widehat{\omega})s &= \sum_a (G_n - \widehat{\omega})(Pt_a)L_a(P)\\ &= \sum_a (\nu_{G_n} - \nu_{\widehat{\omega}})t_a \le \sum_a \|\nu_{G_n} - \nu_{\widehat{\omega}}\|.\end{aligned} \tag{16}$$

But by Lemma 1 and Remark 1, $\forall\ \varepsilon > 0\ \exists\ \delta > 0$ such that $\rho(G_n, \widehat{\omega}) \ge \delta$ or $\|\nu_{G_n} - \nu_{\widehat{\omega}}\| \le \varepsilon$.

Theorem 3 of Gilliland and Hannan (1974, 1986) reduced the problem of treating the asymptotic excess compound risk of equivariant "delete bootstrap" rules to the question of $L_1$-consistency of $\|\widehat{\omega}_n - G_n\|$ for finite $\mathscr{P}$. Datta (1988) considered the compound estimation problem under squared error loss for real one parameter exponential families with compact parameter space and reduced that problem to the question of $L_1$-consistency of $\|\mathscr{P}_{\widehat{\omega}_n} - \mathscr{P}_{G_n}\|$, under a domination assumption on translates of $\mu$ that implies our identifiability assumption. His proof however, depended heavily on the particular shape of the densities for that family and the functional form of the Bayes estimator under squared error loss.

**4. Examples of $L_1$-consistent posterior mixtures.** In Theorem 1 we listed two conditions under which "delete bootstrap" rules are asymptotically optimal. In Examples 1 and 2 we considered situations where condition (i) was satisfied and the problem of finding asymptotically optimal solutions was reduced to the problem of obtaining estimates of $G_n$ that satisfy the $L_1$-consistency requirement of Theorem 1. Below we consider two classes of estimators of $G_n$ that satisfy the requirement.

*A. Consistent posterior mixtures based on a hyperprior.* Consistent mixtures based on a hyperprior were introduced in Section 1.4 of Datta (1988), for a subclass of one dimensional real exponential families and were extended to a much larger class of probability distributions in Theorem 3.1 of Datta (1991a).

More specifically, let $\mu$ be a measure and let $\mathscr{P}$ be the class of probability distributions with densities $\{p_\theta : \theta \in \Theta\}$ wrt $\mu$, where $\Theta$ is a compact metric space. Suppose $p_\theta(x)$ is continuous for each $x$ and, with $h_{\theta'} = \sup_{\theta \in \Theta} |\log(p_\theta/p_{\theta'})|$ and $x^+ = \max(x, 0)$, $\sup_{\theta \in \Theta} \int (h_\theta - M)^+ p_\theta \, d\mu \to 0$ as $M \to \infty$.

Observe that as pointed out in Remark 3.2 of Datta (1991a), the second part of the above assumption forces $P_\theta$'s to be pairwise mutually absolutely continuous. By the Scheffé theorem, continuity of $p_\theta(x)$ for each $x$ implies the norm-continuity of $P_\theta$. The latter implies that $\mathscr{P}$ inherits the compactness of $\Theta$.

Consider $\Omega$ with the topology of weak convergence and let $\Lambda$ be a probability measure on the Borel subsets of $\Omega$. Let $\hat{\Lambda}$ be the posterior distribution of $\omega$ given $\mathbf{X} = \mathbf{x}$. Then $\hat{\Lambda}$ is the probability measure on $\Omega$ with density proportional to $\prod_{i=1}^n p_\omega(x_i)$ with respect to $\Lambda$. Let $\widehat{\omega}_n$ denote the $\hat{\Lambda}$-mix of $\omega$'s. Now, Theorem 3.1 of Datta (1991a) asserts that if $\Lambda$ has full support then

$$\sup\left\{\mathbb{P}\|\mathscr{P}_{\widehat{\omega}_n} - \mathscr{P}_{G_n}\|: \mathbf{P} \in \mathscr{P}^n\right\} = o(1). \tag{17}$$

Since $(n+1)(\mathscr{P}_{G_n} - \mathscr{P}_{G_{n+1}}) = \mathscr{P}_{G_n} - \mathscr{P}_{\theta_{n+1}}$ its norm does not exceed 2. Thus, by triangulation about $\mathscr{P}_{\widehat{G}_{n+1}}$, (17) is equivalent to (17) with $G_n$ replaced by $G_{n+1}$ or, equivalently, with $\widehat{\omega}_n$ replaced by $\widehat{\omega}_{n-1}$.

Observe that $\widehat{\omega}_{n-1}$ is symmetric on $\mathscr{X}^{n-1}$. Therefore $\widehat{\omega}_{n-1}$ provides an example of estimators that satisfy assumption (ii) of Theorem 1.

The estimators considered by Datta are particularly important because compound Bayes rules against certain hierarchical priors turn out to be Bayes versus $\widehat{\omega}_{n-1}(\mathbf{X}_{\check{\alpha}})$ in the $\alpha$th component [cf. Datta (1988), Section 1.2.1]. Therefore if the Bayes rules versus a given prior have unique risk, the compound rule that is obtained by playing Bayes versus $\widehat{\omega}_{n-1}(\mathbf{X}_{\check{\alpha}})$ in the $\alpha$th component, will be admissible for each $n$. The uniqueness of the compound risk of Bayes rules versus a prior $\zeta$ in an estimation problem under squared error loss was shown in Section 4 of the Appendix in Datta (1988), under the condition that $P_\theta$ is dominated by $P_\zeta$ for every $\theta$.

*B. $L_1$-consistent mixtures based on a minimum distance.* $L_1$-consistent estimators of the mixing distribution for a normal mean were obtained, in Edelman (1988), by minimizing an $L_2(\lambda)$-distance where $\lambda$ denotes Lebesgue measure on $R$. His proof depended heavily on the properties of the normal distribution, especially the functional form of the normal characteristic function.

Instead of $L_2(\lambda)$ we consider minimum distance in $L_2(\eta)$ with $\eta$ a probability with support $R^k$ and obtain estimators for the case where $\mathscr{P}$ is a class of distributions on $R^k$. Theorem 2, to follow, proves $L_1$-consistency of minimum $L_2(\eta)$-distance estimators of $\mathscr{P}_{G_n}$. In what follows we will use $F$, with or without affixes, to denote the distribution function of a probability distribution $P$ and $\|\ \|_\eta$ to denote the norm on $L_2(\eta)$.

Observe that if $\eta$ is a probability measure on $R^k$, any distribution function $H$ is in $L_2(\eta)$ and $d$: $\mathbf{P} \rightsquigarrow \|F_{G_n} - H\|_\eta$ satisfies

$$|d(\mathbf{P}) - d(\mathbf{P}')| \le \|F_{G_n} - F_{G'_n}\|_\eta \le \|\mathscr{P}_{G_n} - \mathscr{P}_{G'_n}\|. \tag{18}$$

Therefore, $d$ is continuous on compact $\mathscr{P}^n$ and hence attains a minimum.

LEMMA 2. *Let $\mathscr{X}$ be $R^k$ and let $\eta$ be a probability measure with support $R^k$. Let $r$ be the pseudo-metric on $\Omega$ induced by the $L_2(\eta)$ norm on the range of $\omega \rightsquigarrow F_\omega$. Then $r$ is a metric equivalent to $\rho$ and $\omega \in (\Omega, \rho) \rightsquigarrow \omega \in (\Omega, r)$ and its inverse are uniformly continuous.*

PROOF. If $r(\omega, \omega') = 0$, then $F_\omega = F_{\omega'}$ a.e. $(\eta)$ and therefore, by continuity from above, everywhere. Thus $\mathscr{P}_\omega = \mathscr{P}_{\omega'}$ and by identifiability $\omega = \omega'$. Since $r \le \rho$, the identity function on $(\Omega, \rho)$ to $(\Omega, r)$ is continuous. Therefore by compactness, as in Remark 1, its inverse is uniformly continuous. □

THEOREM 2. *Let $\mathscr{X}$ be $R^k$ and let $\eta$ be a probability measure with support $R^k$. Let $\tilde{G}_n$ be the empirical distribution of $\tilde{\mathbf{P}}$, a measurable minimizer of $d_n$: $\mathbf{P} \rightsquigarrow \|F_{G_n} - H_n\|_\eta$ with $H_n$ the empirical distribution of $\mathbf{X}$. Then*

$$\sup\{\mathbb{P}\|\mathscr{P}_{\tilde{G}_{n-1}} - \mathscr{P}_{G_n}\| : \mathbf{P} \in \mathscr{P}^n\} = o(1). \tag{19}$$

PROOF. $H_n$ as average of $\mathbb{P}$-independent Bernoulli processes, has mean $F_{G_n}$ and variance

$$\mathbb{P}(F_{G_n} - H_n)^2 = \frac{1}{n} G_n(F(1-F)) \le \frac{1}{4n}. \tag{20}$$

By the Fubini theorem $\mathbb{P}\|F_{G_n} - H_n\|_\eta^2$ has the same bound. By a triangulation about $H_n$ and using the minimizing property of $\tilde{G}_n$ we get

$$r^2(\tilde{G}_n, G_n) = \|F_{\tilde{G}_n} - F_{G_n}\|_\eta^2 \le 4\|F_{G_n} - H_n\|_\eta^2. \tag{21}$$

Let $\varepsilon > 0$. By Lemma 2, take $\delta > 0$ such that $\rho \le \varepsilon$ or $r \ge \delta$. Then

$$\mathbb{P}\|\mathscr{P}_{\tilde{G}_n} - \mathscr{P}_{G_n}\| \le \varepsilon + \mathbb{P}\|\mathscr{P}_{\tilde{G}_n} - \mathscr{P}_{G_n}\|\left[r(\tilde{G}_n, G_n) \ge \delta\right]. \tag{22}$$

By the Markov inequality and that $\rho \le 2$, the last term in (22) is bounded by

$$\frac{2}{\delta^2}\mathbb{P}[\text{the lhs of (21)}] \le \frac{2}{\delta^2 n} \tag{23}$$

by (21) and the bound (20) for its $\mathbb{P}$ expectation.

The resulting bound for the lhs of (22) proves (19) for the equivalent [as for (17) in A] form with $G_n$ replaced by $G_{n+1}$. □

Observe that $\tilde{G}_n$ can be taken to depend on $\mathbf{X}$ only through $H_n$ and therefore is an example of $\widehat{\omega}$ of Theorem 1.

Unlike the mixtures based on a hyperprior, the mixtures based on a minimum distance do not produce admissible rules. However minimum distance procedures provide asymptotically optimal rules for a larger class of compound problems when $\mathscr{X} = R^k$. It also seems to be easier to explore rates of convergence for the modified regret of the minimum distance procedures. In order to establish rates of convergence in our examples it is necessary to have the moduli of continuity of $\omega \rightsquigarrow \nu_\omega$ of Lemma 1 and $\omega \in (\Omega, r) \rightsquigarrow \omega \in (\Omega, \rho)$ of Lemma 2. Our assumptions (compactness of $\mathscr{P}$ and identifiability) are not sufficient for deriving these moduli of continuity. For some discussion of the identifiability condition see Datta (1991a).

**Acknowledgments.** This paper is a part of my Ph.D. dissertation at Michigan State University. I am greatly indebted to my advisor Professor James Hannan for his guidance throughout the research. I would also like to thank the referee for helpful suggestions and remarks.

## REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BILLINGSLEY, P. (1971). *Weak Convergence of Measures: Applications in Probability* **5**. SIAM, Philadelphia.

DATTA, S. (1988). Asymptotically optimal Bayes compound and empirical Bayes estimators in exponential families with compact parameter space. Ph.D. dissertation, Dept. Statistics and Probability, Michigan State Univ.

DATTA, S. (1991a). On the consistency of posterior mixtures and its applications. *Ann. Statist.* **19** 338–353.

DATTA, S. (1991b). Asymptotic optimality of Bayes compound estimators in compact exponential families. *Ann. Statist.* **19** 354–365.

EDELMAN, D. (1988). Estimation of the mixing distribution for a normal mean with applications to the compound decision problem. *Ann. Statist.* **16** 1609–1622.

GILLILAND, D. C. and HANNAN, J. (1974). The finite state compound decision problem, equivariance and restricted risk components. Technical Report RM-317, Dept. Statistics and Probability, Michigan State Univ.

GILLILAND, D. C. and HANNAN, J. (1986). The finite state compound decision problem, equivariance and restricted risk components. In *Adaptive Statistical Procedures and Related Topics* (J. Van Ryzen, ed.) 129–145. IMS, Hayward, CA.

HANNAN, J. F. (1953). Asymptotic solutions of compound decision problems. UNC Inst. Statistics Mimeo Series 68.

HANNAN, J. F. (1957). Approximation to Bayes risk in repeated play. In *Contribution to the Theory of Games 3. Ann. of Math. Stud.* **39** 97–139. Princeton Univ. Press.

HANNAN, J. and HUANG, J. S. (1972a). Equivariant procedures in the compound decision problem with finite state component problem. *Ann. Math. Statist.* **43** 102–112.

HANNAN, J. and HUANG, J. S. (1972b). A stability of symmetrization of product measures with few distinct factors. *Ann. Math. Statist.* **43** 308–319.

HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

MASHAYEKHI, M. (1990). Stability of symmetrized probabilities and compact equivariant compound decisions. Ph.D. dissertation, Dept. Statistics and Probability, Michigan State Univ.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic, New York.

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

ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, New York.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF NEBRASKA
LINCOLN, NEBRASKA 68588-0323