This is to certify that the

dissertation entitled

Stability of Symmetrized Probabilities and
Compact Equivariant Compound Decisions

presented by

Mostafa Mashayekhi

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

James Hannan
Major professor

Date August 10, 1990

STABILITY OF SYMMETRIZED PROBABILITIES AND COMPACT EQUIVARIANT DECISIONS

By

Mostafa Mashayekhi

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1990

ABSTRACT

STABILITY OF SYMMETRIZED PROBABILITIES AND COMPACT EQUIVARIANT COMPOUND DECISIONS

By

Mostafa Mashayekhi

Extensions of Hannan and Huang (1972) results on the stability of symmetrization of product probability measures to the compact case and their applications in extensions of some of the results of Gilliland and Hannan (1974), on equivariance in a compound decision problem are obtained.

Let $\mathscr{P}$ be a compact, in the total variation norm, class of pairwise mutually absolutely continuous probability distributions. We show that the total variation norm of the symmetrization of two products of probabilities in $\mathscr{P}$, with differences in one factor, converges to zero uniformly as the number of factors approaches ∞. Rates of convergence are obtained for the case where $\mathscr{P}$ is an exponential family with its parameter space in the interior of the natural parameter space.

The above convergences translate into the convergence to zero of the excess of the simple envelope over the equivariant envelope, for a restricted component risk compound decision problem, as the number of problems approaches ∞.

For compound estimation of continuous functions under squared error loss, and finite action problems with continuous loss functions, the problem of treating the asymptotic excess compound risk of equivariant "delete bootstrap" rules is reduced, under an identifiability condition, to the question of L_1– consistency of certain mixtures. Examples of estimates satisfying the above consistency condition are included.

To my mother

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor Professor James Hannan for suggesting the problem and for the patience he accorded me in the preparation of this thesis. His careful criticism and invaluable suggestions aided greatly in simplifying proofs and improving virtually all of the results in the thesis.

I like to express my thanks to my other committee members, Professors Dennis C. Gilliland Habib Salehi Clifford Weil, and Professor James Stapleton for their suggestions on an earlier draft.

Finally I would like to thank the Department of Statistics and Probability at Michigan State University for their generous support, financial and otherwise, during my stay at Michigan State University.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. The set compound problem.

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. Ordinarily in the component problem, there is a family of probability distributions $\mathscr{P}$ on some common measurable space $(\mathscr{X},\mathscr{B})$, an observable $\mathscr{X}$-valued random element X with distribution P, where $P \in \mathscr{P}$, an action space $\mathscr{A}$, a loss function L: $\mathscr{A}\times\mathscr{P} \to [0,\infty)$, a class $\mathscr{D}$ of (randomized) decision rules t, on $\mathscr{X} \times \mathscr{a}$, where $\mathscr{a}$ is a σ–field of subsets of $\mathscr{A}$, such that for each $x \in \mathscr{X}$ $t(x,\cdot)$ is a probability measure on $\mathscr{a}$ and for each $A \in \mathscr{a}$ $t(\cdot,A)$ is $\mathscr{B}$ measurable. The decision procedure t has risk

$$R(t,P) = \int\int L(a,P)t(x,da)dP(x). \tag{1}$$

In the compound problem, we have the state space $\mathscr{P}^n$, the action space $\mathscr{A}^n$, observations $\underline{X} = (X_1, \ldots, X_n)$ with distribution P , where $P = \times_{\alpha=1}^{n} P_\alpha$, $\underline{P} = (P_1, \ldots, P_n) \in \mathscr{P}^n$, compound rules $\underline{t} = (t_1, \ldots, t_n)$, where for each $1 \le \alpha \le n$, t_α has domain $\mathscr{X}^n \times \mathscr{a}$ with $t_\alpha(\cdot,A)$ $\mathscr{B}^n$ measurable for each A and $t_\alpha(x,\cdot)$ a probability measure on $\mathscr{a}$ for each x. The compound risk is given by

$$\underline{R}(\underline{t},\underline{P}) = \frac{1}{n} \Sigma\int\int L(\alpha_i,P_i)t(\underline{x},da_i)dP(\underline{x}). \tag{2}$$

Sometimes it is preferable (see Hannan and Huang (1972a)) to consider loss functions that may depend on x itself. A more general setting (cf. Gilliland and Hannan (1974–)) is to bypass the consideration of a loss function and identify each decision rule in $\mathscr{D}$ by its risk point in $[0,\infty)^{\mathscr{P}}$. Then a compound rule $\underline{t}$ is identified with $\underline{s} = (s_1, \ldots, s_n)$ where for each i, s_i is a $\mathscr{B}^{n-1}$ measurable mapping into $[0,\infty)^{\mathscr{P}}$. The α-th component risk at P_α is then $s_\alpha(P_\alpha)$ and the compound risk is given by

$$\underline{R}(\underline{t},\underline{P}) = \frac{1}{n} \Sigma\, s_\alpha(P_\alpha).$$

Let $\mathscr{S}$ be the class of all simple procedures (i.e. $\mathscr{S} = \{\underline{t}\colon t_\alpha(\underline{x}) = t(x_\alpha)\ \forall\ 1 \le \alpha \le n$, for some component rule $t\}$), and let $\mathscr{E}$ be the class of compound rules that are equivariant under the permutation group. As functions of $\underline{P}$, $\inf_{\mathscr{S}} \underline{R}(\underline{t},\underline{P})$, $\inf_{\mathscr{E}} \underline{R}(\underline{t},\underline{P})$ are called the simple envelope and the equivariant envelope respectively. It is clear from the definition that the latter is the infimum over a larger class and well known that the former coincides with $R(G_n)$, where $R(w)$ is the component Bayes risk at w, and G_n denotes the empirical distribution of $P_1, \ldots, P_n$.

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

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

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

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

2. History review and a summary of the present work.

Hannan and Robbins (1955) introduced the class of equivariant procedures for the 2×2 $\mathcal{P}\times\mathcal{A}$ compound problem and showed (Theorem 5) that the difference between the simple and equivariant envelopes converges to zero uniformly in $\underline{P}$. Hannan and Huang (1972a) considered the compound problem for finite $\mathcal{P}$ 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–) (Theorems 1 and 2) extended those results to arbitrary bounded risk components for finite $\mathcal{P}$. 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 thesis we consider the compound decision problem in which the set of component distributions $\mathcal{P}$ is compact in the topology induced by the total variation norm and has pairwise mutually absolutely continuous elements. The risk set of the component problem is assumed to be a bounded subset of $[0,\infty)^{\mathcal{P}}$.

In Chapter 2 we consider some extensions of Hannan and Huang (1972 b) results on symmetrization of product measures to the compact case and prove two measure theoretic theorems analogous to Theorem 1 of Hannan and Huang (1972b). Theorem 1 shows convergence to zero , as the number of factors approaches ∞ , of the total variation norm of the symmetrization of the difference of two product probability measures with differences in one factor. Theorem 2 specializes to compact k–dimensional exponential families and obtains rates of convergence for the case where the parameter space is a compact subset of the interior of the natural parameter space.

Chapter 3 considers some extensions of Gilliland and Hannan (1974–) results on equivariance and the compound decision problem. In Remark 4 we observe that the method of proof of their Theorem 1 bounds the difference of the simple and equivariant envelopes by a constant multiple of the norm of two product probability measures considered in our Theorem 1.

Our envelope results strengthen, inter alia, the results of Datta (1988) 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.

Theorem 3 provides sufficient conditions for asymptotic optimality of "delete bootstrap" rules. Examples 3 and 4 show that for squared error loss estimation of continuous functions and for finite $\mathcal{A}$ problems with continuous loss functions Theorem 3 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–), and immediately extends the results of Datta (1990) for the empirical Bayes decision problem, to the corresponding compound decision problem, under appropriate loss functions.

Examples of estimates satisfying the above consistency condition are provided in Section 3 of Chapter 3.

3. Notations and conventions.

Let n be a positive integer. If $P_1, \ldots, P_n$ are probability measures, P denotes the product probability measure $P_1 \times \ldots \times P_n$. We use P^n to denote $\times_{\alpha=1}^{n} P$ and P_1P_2 to denote $P_1 \times P_2$. An n–tuple $(x_1, \ldots, x_n)$ is denoted by $\underline{x}_n$ and $\bar{x}_n$ denotes the average of the components of $\underline{x}_n$ (the subscript n will not be exhibited if it is clear from the context). The empirical distribution of $\underline{P}$, where P_α is a probability measure for each α, is denoted by G_n.

We use $\mu(f)$ or μf to denote the integral of a function f with respect to (wrt hereafter) a signed measure μ. 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.

A function f defined on a set $\mathscr{X}$ into a set $\mathscr{Y}$ is sometimes denoted by $x \in \mathscr{X} \rightsquigarrow f(x)$ or $x \rightsquigarrow f(x)$. Sometimes we abuse notation and denote functions by their values. If $\mathscr{X}$ and $\mathscr{Y}$ are metric spaces with metrics r and ρ respectively, we sometimes denote f by $x \in (\mathscr{X}, r) \rightsquigarrow f(x) \in (\mathscr{Y}, \rho)$. If f is a function of two arguments, $f(\cdot, y)$ denotes the function (section) that is obtained by fixing the second argument at the point y.

If τ is a signed measure, then $|\tau|$ will denote the total variation measure corresponding to τ (i.e. $|\tau| = \tau^+ + \tau^-$) and $\|\tau\|$ will denote the total variation norm of τ. We denote the Euclidian norm and inner product by $| \quad |$ and juxtaposition respectively. The supremum of a real function f is denoted by $\|f\|_\infty$ whatever be its domain.

All incompletely described limits are as $n \longrightarrow \infty$ through positive integers. All sums will be on i from 1 to n unless otherwise indicated. The symbol $\square$ denotes end of proof.

CHAPTER 2

ON SYMMETRIZATION OF PRODUCT MEASURES

1. Introduction

In this chapter we consider some extensions of Hannan and Huang (1972b) results, on the stability of symmetrization of product measures, to the compact case. Our main results (Theorems 1 and 2) are analogous to their Theorem 1.

In Section 1 we reproduce some of the general properties of signed measures and their symmetrization with respect to general groups from their Section 2, with the minor improvement that we consider total, instead of their maximum, variation norm. The substitution of this equivalent norm simplifies some relations and proofs.

Section 2 considers a contraction effect of probability factors in product signed measures that was noted in their Section 3 , and presents an extension of their Lemma 1 with a simpler proof.

In Section 3, specializing to permutation groups, we consider product probability measures with factors in a set which is compact under the topology induced by the total variation norm, and has pairwise mutually absolutely continuous elements. Theorem 1 and Theorem 2 deal with the effect of symmetrization on the difference of two product probability measures with differences in one factor. Theorem 1 shows uniform convergence to zero of the total variation norms, and Theorem 3 specializes to k–dimensional exponential families and obtains rates of convergence.

2. Preliminaries

Let $\mathscr{G}$ be a finite group of measurable transformations g on $(\mathscr{Y}, \mathscr{C})$. For a signed measure τ on $(\mathscr{Y}, \mathscr{C})$, the symmetrization τ^* of τ is defined by

(1) $$\tau^*(C) = N^{-1} \sum_{g \in \mathscr{G}} \tau(g(C)), \quad C \in \mathscr{C},$$

where N is the number of elements in $\mathscr{G}$. Thus symmetrization (*, hereafter) is an expectation operator. We will abbreviate affixes on * by omission.

For any real valued function f on $\mathscr{Y}$, its symmetrization f^* is defined by

(2) $$f^* = N^{-1} \sum_{g \in \mathscr{G}} f \cdot g .$$

τ and f are said to be symmetric if $\tau = \tau^*$ and $f = f^*$, respectively.

The properties, $(\tau . g)^* = \tau^*$ and $(f . g)^* = f^*$ $\forall$ $g \in \mathscr{G}$, will be used later without comment.

The following two facts are taken from Section 2 (Relations (6) and (8)) of Hannan and Huang (1972b).

Let τ be a signed measure, and let μ be a measure such that $d\tau/d\mu$ exists. If $\mu = \mu^*$, then

(3) $$(d\tau/d\mu)^* = (d\tau^*/d\mu).$$

Let μ be a measure. If f is μ^*-integrable, then

(4) $$\mu^*(f) = \mu^*(f^*) = \mu(f^*).$$

The following two relations ((5) and (6)) are simpler analogs of their relations (9), (11) and (12).

If (3) holds,

$$|d\tau^*/d\mu| \le |d\tau/d\mu|^*$$

by subadditivity of $|\ \ |$. Applying this with $\mu = \mu^*$, by (4), integration wrt μ and the isotonicity of μ-integral give

$$\|\tau^*\| \le \|\tau\|. \tag{5}$$

If $\tau\sigma$ is a product signed measure, then

$$|\tau\sigma| = (\tau^+\sigma^+ + \tau^-\sigma^-) + (\tau^+\sigma^- + \tau^-\sigma^+) = |\tau|\ |\sigma|$$

and therefore by the Fubini Theorem,

$$\|\tau\sigma\| = \|\tau\|\ \|\sigma\|. \tag{6}$$

If P and Q are product probability measures, subadditivity of norm and applications of (6) give

$$\|P - Q\| \le \sum_i \|\underset{j<i}{\times} P_j(P_i - Q_i) \underset{j>i}{\times} Q_j\| = \sum_i \|P_i - Q_i\|. \tag{7}$$

3. Contraction effect of probability factors.

Let $(\mathcal{X},\mathcal{B})$ be a measurable space. For each n let $\mathcal{G}_n$ be a measurable group of transformations on $(\mathcal{X},\mathcal{B})^n$ such that $\mathcal{G}_n$ is a subgroup of $\mathcal{G}_{n+1}$. Consider the symmetrization of a measure on $(\mathcal{X},\mathcal{B})^n$ relative to $\mathcal{G}_n$. The following lemma is a strengthened generalization of Lemma 1 of Hannan and Huang (1972b), with a simple proof eliminating the need for developing their (13) and (14). It also serves for the extension of Remark 2 of their

Addendum. Their Lemma 1 was already sufficient for the proof of our theorems in this chapter.

Lemma 1. If τ is a signed measure and P is a probability measure then

(8) $$\| (\tau P)^* \| \le \|\tau^*\|.$$

Proof. Observe that $(\tau P)^* = (\tau^* P)^*$, since they agree on symmetric functions. Therefore (8) follows from the application of (5) and the probability case of (6) to $\tau^* P$. □

Henceforth we specialize $\mathcal{G}_n$ to be the group of transformations on $(\mathcal{X}, \mathcal{B})^n$ induced by the group of permutations on n objects. We also let $\mathcal{G}_n$ denote the permutation group itself. Thus a generic element $g \in \mathcal{G}_n$ will be used both as a permutation and the transformation $g(\underline{x}) = (x_{g1}, \dots x_{gn})$.

4. Two product probability measures with differences in one factor.

Throughout the rest of the thesis

$\mathcal{P}$ is a non-empty norm-compact class of pairwise mutually absolutely continuous probability measures.

Part (i) of Lemma 2, to follow, proves uniform convergence to zero of $\|(\tau Q^n)^*\|$ where $\tau = R - S$ with $(Q, R, S) \in \mathcal{P}^3$. The result is used to prove a stronger assertion in Theorem 1 where Q^n is replaced by a product of n elements of $\mathcal{P}$. Part (ii) of Lemma 2 obtains rates of convergence under an additional assumption and is used in the proof of Theorem 2 in Section 5.

Lemma 2. For $(Q,R,S) \in \mathscr{P}^3$ let t be a density of τ wrt Q and let $T_n = \|(\tau Q^{n-1})^*\|$. Then, with t_i denoting $\underline{x} \leadsto t(x_i)$,

$$T_n = Q^n|\bar{t}|, \tag{9}$$

(i) $\|T_n\|_\infty = o(1)$, and for each $r \in (1,2]$

(ii) $n^{r-1}\|T_n\|_\infty^r \le \alpha_r \|Q|t|^r\|_\infty$ with $\alpha_r = 2^{2-r}$.

Proof. Since t is a density of τ wrt Q, t_1 is a density of τQ^{n-1} wrt Q^n and $\bar{t}$ is a density of $(\tau Q^{n-1})^*$ wrt Q^n. The latter implies (9).

(i) We show that $\langle T_n \rangle$ is a monotone sequence of continuous functions decreasing to zero on compact $\mathscr{P}^3$ and therefore by Dini's Theorem (cf. e.g. Proposition 9.11 of Royden (1968)) it converges to zero uniformly. By Lemma 1 $T_n \ge T_{n+1}$. Since sum and product are, by the norm properties and (6), continuous operations on finite signed measures and * is linear, the composition T_n inherits continuity on $\mathscr{P}^3$. By the L_1 Law of Large Numbers the rhs(9) converges to zero.

(ii) By the independent case of a von Bahr and Esseen inequality (cf. von Bahr and Esseen (1965))

$$Q^n|\Sigma t_i|^r \le \alpha_r \, \Sigma Q^n|t_i|^r = n\alpha_r Q|t|^r. \tag{10}$$

From (9) and the moment inequality, $n^r T_n^r \le$ lhs(10). Weakening this by (10), dividing by n and taking supremum over (Q,R,S), gives the inequality in (ii). □

Remark 1 In the proof of Lemma 2 we used Lemma 1 to obtain the monotonicity of T_n. Only later did we note that Lemma 1 can be used to show the monotonicity of $E|\bar{X}_n|$ when $X_1, \ldots, X_n$ are i.i.d with EX_1 finite, a fact seldom noted in texts, but presumable well known to the authors who discuss the associated reverse martingale.

Any hope that this application of Lemma 1 would be a worthy simpler proof were short lived. My colleague Liu, Zhihui noted non-increasing monotonicity for $\|\bar{X}_n\|_r$ with $r \in [1,\infty]$ and $X_1, \ldots, X_n$ only exchangeable,

$$(n-1)\|\sum_{i=1}^{n} X_i\|_r \leq \sum_{j=1}^{n} \|\sum_{i\neq j} X_i\|_r = n\|\sum_{i=1}^{n-1} X_i\|_r ,$$

as this immediate consequence of the homogeneity and subadditivity properties of the $L_r(P)$ norm appropriately applied.

Remark 2. Pairwise mutual absolute continuity of the elements of $\mathcal{P}$ is a necessary condition for the conclusion of Lemma 2 to hold. For, if B is such that $P(B) = 0$, then

$$\text{(12)} \qquad \|(P^n - QP^{n-1})^*\| \geq (P^n - QP^{n-1})^*((\mathcal{X}-B)^n) = Q(B)$$

so that lhs(12) $= o(1)$ only if $Q(B) = 0$.

Theorem 1. Let $P = \times_{i=1}^{n} P_i$, where for each i $P_i \in \mathcal{P}$, and let $\tau = R - S$ with R and $S \in \mathcal{P}$. Then

$$\text{(13)} \qquad \sup\{\|(\tau P)^*\| : (R,S,\underline{P}) \in \mathcal{P}^{n+2}\} = o(1).$$

Proof. Let Q be a product of N factors of P. By Lemma 1

$$\|(\tau P)^*\| \le \|(\tau Q)^*\|. \tag{14}$$

If Q is a probability measure, by the triangle inequality, (5), (6), and (7),

$$\text{rhs}(14) \le \|\tau\| \; \Sigma\|Q_i - Q\| + \|(\tau Q^N)^*\|. \tag{15}$$

Let $\epsilon > 0$, and use Lemma 2 to choose N such that $\|T_{N+1}\|_\infty < \epsilon/3$. By total boundedness there is a finite covering of $\mathscr{P}$ by m balls of radius $\frac{\epsilon}{3N}$. If $n > (N-1)m$, then some ball contains at least N factors of P. Let Q be the center of such ball and the Q_i be factors of $\underline{P}$ therein. Then by (14) and (15) and the choice of N

$$\|(\tau P)^*\| < \epsilon.$$ □

Example 1: Location family. Let P be a probability measure equivalent to Lebesgue measure on R^k. Let Θ be a compact subset of R^k and, $\forall\ \theta \in \Theta$, let P_θ be the translate of P by θ. Observe that

$$\|P_{\theta+\delta} - P_\theta\| = \|P_\delta - P\| = P|dP_\delta/dP - 1| \longrightarrow 0 \text{ as } \delta \longrightarrow 0$$

by Lebesgue's theorem (cf. Royden pp 90–91 for the proof in the one dimensional case). Compactness of $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ then follows by compactness of Θ, and pairwise mutual absolute continuity of its elements follows by the equivalence of P and the Lebesgue measure together with the translation invariance of the latter.

5. Exponential families.

Let μ be a measure on $\mathscr{X} = R^k$ such that

(16) $$\mathscr{N} = \{\theta \in R^k : \varphi(\theta) = \ln\int e^{\theta x} d\mu(x) < \infty\} \neq \phi.$$

For $\theta \in \mathscr{N}$, let P_θ be the probability measure with μ-density

(17) $$p_\theta(x) = e^{\theta x - \varphi(\theta)}.$$

It has long been known (cf. e.g. Theorem 1.13 of Brown (1986)) that, on $\mathscr{N}$, φ is convex by the Holder inequality and lower semicontinuous by the Fatou Lemma. If $\theta \in \mathscr{N}^\circ$, then (cf. e.g. Theorem 2.2 of Brown (1986)) all derivatives of φ exist at θ and can be obtained by differentiating under the integral sign.

In Example 2 and Remark 3, to follow, $\mathscr{P} = \{P_\theta: \theta \in \Theta\}$ for various $\Theta \subset \mathscr{N}$.

Example 2: Θ polytope. By a Gale–Klee–Rockafellar theorem (Theorem 10.2 of Rockafellar (1970), convex φ is upper semi-continuous on locally simplicial Θ. Since a polytope is a finite union of simplices, it then follows that each x–section of p is continuous . Therefore, by the Scheffé Theorem, P_θ is norm continuous so that $\mathscr{P}$ inherits compactness of Θ.

Remark 3: Compact $\Theta \subset \mathscr{N}^\circ$. Then each x–section of p is continuous so that $\mathscr{P}$ inherits the compactness of Θ as in Example 2. Consider a finite covering of Θ by open cubes with their closures in $\mathscr{N}^\circ$. Then the convex hull of the vertices v of these cubes is a polytope in $\mathscr{N}$ with Θ in its interior. Lemma 2.1 of Brown (1986) then applies and gives a number K_1 such that

(18) $$|e^{\theta\cdot} - e^{\theta'\cdot}| \leq |\theta - \theta'| K_1 \Sigma e^{v\cdot} \quad \forall\ (\theta,\theta') \in \Theta^2.$$

By triangulation about $e^{\theta'\cdot}$ it follows that

$$e^{\varphi(\theta)}|p_\theta - p_{\theta'}| \leq |e^{\theta\cdot} - e^{\theta'\cdot}| + |e^{\varphi(\theta')} - e^{\varphi(\theta)}| p_{\theta'} ,$$

whence properties of integration wrt to μ and (18) bound

$$\|p_\theta - P_{\theta'}\| \text{ by } B|\theta - \theta'| \text{ with } B = \|e^{-\varphi}\|_\infty\ 2\ K_1 \Sigma e^{\varphi(v)}.$$

Theorem 2. Let $\mathscr{P}$ be an exponential family with compact $\Theta \subset \mathscr{N}^\circ$. If $r \in (1,2]$ and, for every θ_0 and θ_1 in Θ, the external convex combination $\theta_r = r\ \theta_1 + (1-r)\theta_0 \in \mathscr{N}^\circ$, then with $\tau = R - S$

(19) $$\sup\{\|(\tau P)^*\| : R, S \in \mathscr{P} \text{ and } \underline{P} \in \mathscr{P}^n\} = O(n^{-\beta})$$

where $\beta = (r-1)/(r+(2r-1)k)$.

Proof. Let $n \in Z^+$, $(R, S) \in \mathscr{P}^2$ and $\underline{P} \in \mathscr{P}^n$. Let $(\theta_1, \ldots, \theta_n) \in \Theta^n$ such that $P_i = P_{\theta_i}$. Choose $N \in Z^+$ such that with $m = ([N^{(2-\frac{1}{r})}]+1)^k$ and $g(N) = (N-1)m + 1$,

$$g(N) \leq n \leq g(N+1).$$

Consider a cube containing Θ and divide it into m equal size subcubes. Since $n \geq g(N)$ there exist N factors of P , say $Q_1, \ldots, Q_N$, with their indices in the intersection of one of the cubes with Θ. By (15)

$$(20) \qquad \|(\tau P)^*\| \le 2 \sum_{i=2}^{N} \|Q_i - Q_1\| + \|(\tau Q_1^N)^*\|.$$

Since the diameter of each cube is less than or equal to a constant multiple of $N^{(\frac{1}{r}-2)}$, where the constant depends on k and the size of the cube containing Θ, by the uniform bound, on $\|P_\theta - P_{\theta'}\| / |\theta - \theta'|$, considered in Remark 3 there is a constant B_1 such that the first term on the rhs(20) is less than or equal to $B_1 N^{(\frac{1}{r}-1)}$. Since

$$P_{\theta_0}(p_{\theta_1}/p_{\theta_0})^r = \exp\{-r\varphi(\theta_1) - (1-r)\varphi(\theta_0) + \varphi(\theta_r)\},$$

it is continuous on compact Θ^2 and therefore is bounded. Thus if t is a density of τ wrt to Q, $\|Q|t|^r\|_\infty \le 2^r$(above bound) by Minkowski's inequality in $L_r(Q)$. Therefore (20) and Lemma 2(ii) imply that

$$(21) \qquad \|(\tau P)^*\| \le BN^{(\frac{1}{r}-1)}$$

where B is a constant independent of R, S, $\underline{P}$. Observe that by definition of g and β

$$(22) \qquad g^\beta (N+1)\mathrm{rhs}(21) = O(1).$$

By choice of N and (21),

$$(23) \qquad n^\beta \mathrm{lhs}(19) \le \mathrm{lhs}(22).$$

The conclusion of the theorem follows by (22) and (23). □

CHAPTER 3

EQUIVARIANCE AND THE COMPOUND DECISION PROBLEM

1. Introduction

Consider a compound problem involving n independent repetitions of a component problem with states $P \in \mathscr{P}$. Let $\mathscr{D}$ be a bounded risk class of decision rules for the component problem and let $M < \infty$ be such that $\forall\ t \in \mathscr{D}$, and $\forall\ P \in \mathscr{P}$, $R(t,P) \leq M$. For an n–tuple $\underline{x} = (x_1, \ldots, x_n)$ let $\underline{x}_{\check{\alpha}}$ denote $\underline{x}$ with the α–th component deleted, and let $P_{\check{\alpha}}$ denote P with the α–th factor deleted. Consider the class $\underline{\mathscr{D}}$ of compound rules $\underline{t} = (t_1, \ldots, t_n)$ where each $\underline{x}_{\check{\alpha}}$–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 $\underline{\mathscr{D}}$ the largest class of compound decision rules. The compound problem with restricted component risk was considered by Gilliland and Hannan (1974–) for finite $\mathscr{P}$, because of the generality it provided for 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 $\underline{\mathscr{D}}$

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

It is well known (see Section 2 of Hannan and Huang (1972a) or Section 1 of Gilliland and Hannan (1974–)) that the compound problem is invariant under the group of n! permutations of coordinates, and that a compound rule $\underline{t}$ is equivariant if and only if there exists a function γ on $\mathscr{X} \times \mathscr{X}^{n-1}$ to $\mathscr{A}$, symmetric on $\mathscr{X}^{n-1}$, such that $t_\alpha(\underline{x}) = \gamma(x_\alpha, \underline{x}_{\check{\alpha}})$ for all α. The latter implies that if $\underline{t}$ 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. For equivariant procedures we will abbreviate $s(\underline{t}, \alpha, P, \cdot)$ by using the affixes on $\underline{t}$ and P. For example $s(\hat{\underline{t}}, 1, P_\alpha, \cdot)$ will be abbreviated to $\hat{s}_\alpha$.

Let $\mathscr{E}$ and $\mathscr{S}$ denote the class of all equivariant rules in $\underline{\mathscr{D}}$ and the class of all simple symmetric rules in $\underline{\mathscr{D}}$ respectively. The equivariant envelope corresponding to $\underline{\mathscr{D}}$ is defined by

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

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

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

In this chapter we use the results of Chapter 2 and properties of equivariant rules to show the asymptotic equivalence of the simple and equivariant envelopes and establish asymptotic optimality of certain equivariant "delete bootstrap" rules.

Section 2 deals with the difference of the two envelopes and asymptotic optimality. In Remark 4 we observe that the method of proof of Theorem 1 of Gilliland and Hannan (1974–) can be applied to translate the results of our Theorems 1 and 2 into convergence to zero of the excess of the simple envelope over the equivariant envelope. Theorem 3 introduces sufficient conditions for asymptotic optimality of equivariant "delete bootstrap" rules. Examples 3 and 4 consider important cases in which, by assuming an identifiability condition, Theorem 3 reduces the problem of treating the asymptotic excess compound risk of "delete bootstrap" rules to the question of L_1– consistency of certain mixtures.

Section 3 provides, as examples, two classes of mixtures that satisfy the required consistency condition. The first example is the class of mixtures based on hyperpriors obtained in Datta (1990), thus showing that his results for empirical Bayes problems are extended to the corresponding compound problems. A mixture in the second class is obtained by minimizing an L_2–distance.

2. Asymptotically optimal "delete bootstrap" rules.

The following remark shows that the excess of the simple envelope over the equivariant envelope has a uniform upper bound for which we have shown convergence to zero in Theorem 1 and obtained rates of convergence, in a case of exponential families, in Theorem 2. The result which is the first proof in the non–finite case, strengthens all the previous results in compound estimation under squared error loss.

Remark 4.

$$(\psi - \tilde{\psi}) \leq M \sup_{\alpha} \sup_{\underline{P}} \|(P_{\check{\alpha}} - P_{\check{n}})^{*}\|. \tag{3}$$

This follows by the method of proof of Theorem 1 of Gilliland and Hannan (1974–) : Let $\underline{t} \in \mathfrak{X}$ By non–negativity and symmetry of s_{α}

$$|(P_{\check{n}} - P_{\check{\alpha}})s_{\alpha}| \leq M \|(P_{\check{\alpha}} - P_{\check{n}})^{*}\| \tag{4}$$

which implies

$$(5) \qquad P_{\check{n}}\, n^{-1}\Sigma\, s_\alpha - \underline{R}(\underline{t},\underline{P}) \leq \text{ rhs}(3)$$

Since $P_{\check{n}}\,(n^{-1}\Sigma\, s_\alpha) - \psi(\underline{P}) = P_{\check{n}}\, G_n(s - \tilde{s})$, where $\tilde{t}$ is Bayes versus G_n in the component problem, by isotonicity of $P_{\check{n}}$

$$P_{\check{n}}\,(n^{-1}\Sigma\, s_\alpha) \geq \psi(\underline{P}).$$

Therefore (5) implies

$$(6) \qquad \psi(\underline{P}) - \underline{R}(\underline{t},\underline{P}) \leq \text{ rhs}(3).$$

Since $\underline{t}$ is arbitrary $\in \mathscr{T}$, we obtain (3).

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

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

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

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

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

Then by symmetry of $\hat{\omega}$, $\hat{\underline{t}}$ is equivariant. The next theorem gives sufficient conditions for asymptotic optimality of $\hat{\underline{t}}$.

Theorem 3. $\hat{\underline{t}}$ is asymptotically optimal if

(i) For each $\epsilon > 0$, $\exists\ \delta_\epsilon > 0$ such that $\forall$ n,

$(\|P_{\hat{\omega}} - P_{G_n}\| < \delta_\epsilon) \Rightarrow ((G_n - \hat{\omega})(\hat{s} - \tilde{s}) < \epsilon)$,

where $\tilde{\underline{t}}$ an equivariant rule with its components Bayes versus G_n.

(ii) $\sup\{P_{\check{n}} \| P_{\hat{\omega}} - P_{G_n}\| : \underline{P} \in \mathcal{P}^n\} = o(1)$.

Proof. By (4)

$$\text{(7)} \qquad \underline{R}(\hat{\underline{t}},\underline{P}) - \psi(\underline{P}) \le P_{\check{n}}\ G_n(\hat{s} - \tilde{s}) + \text{rhs}(3).$$

Weakening (7) by subtracting the non-positive function $\hat{\omega}(\hat{s} - \tilde{s})$ from its right hand side integrand and triangulation about $\tilde{\psi}(\underline{P})$ together with (3) give

$$\text{(8)} \qquad \underline{R}(\hat{\underline{t}},\underline{P}) - \tilde{\psi}(\underline{P}) \le P_{\check{n}}\ (G_n - \hat{\omega})(\hat{s} - \tilde{s}) + 2(\text{rhs}(3)).$$

The rhs(3) converges to zero by Theorem 1. In order to show uniform convergence to zero of the first term on the rhs(8), let $\epsilon > 0$. Choose δ_ϵ with the property assumed in (i). The first term on the rhs(8) is less than or equal to

$$\epsilon + MP_{\check{n}}[\| P_{\hat{\omega}} - P_{G_n}\| \ge \delta_\epsilon]$$

which is less than or equal to

$$\text{(9)} \qquad \epsilon + M\delta_\epsilon^{-1}\ P_{\check{n}}\| P_{\hat{\omega}} - P_{G_n}\|$$

by the Markov inequality. Since ϵ is arbitrary, the conclusion follows by (ii).

Observe that since $\mathscr{P}$ is a compact metric space, by Theorem II 6.4 of Parthasarathy (1967), Ω with the topology of weak convergence is a compact metric space.

Lemma 3 Let ϕ be a continuous function on $\mathscr{P}$. For each ω let ν_ω 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 $\omega \in (\Omega,d) \leadsto \nu_\omega$, with the norm–topology on the range, is uniformly continuous.

Proof. Let ω_n be a sequence in Ω converging to ω. 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 η_n and η on the Lebesgue unit interval with respective distributions ω_n and ω such that η_n converges to η pointwise.

Since ν_ω is the ω-mixture of the $\nu_P = \phi(P)P$,

$$\nu_{\omega_n} - \nu_\omega = (\omega_n - \omega)\nu_{\cdot} = \int_0^1 (\nu_{\eta_n} - \nu_\eta). \tag{10}$$

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

$$\nu_{\eta_n} - \nu_\eta \le \|\phi\|_\infty \, \|\eta_n - \eta\| + |\phi(\eta_n) - \phi(\eta)|. \tag{11}$$

Since variations of a positive mixture are bounded by the mixture of the variations, continuity of ν at ω follows from (10) and (11) by two applications of the Bounded Convergence Theorem for the $\int_0^1$. Uniform continuity follows by compactness of Ω. □

In the rest of this chapter we assume

(12) Ω is identifiable: $\omega \rightsquigarrow P_\omega$ is 1–1,

and let ρ denote the metric on Ω thereby induced by $\| \quad \|$ on the range

(13) $$\rho(\omega,\omega') = \|P_\omega - P_{\omega'}\|.$$

Remark 5. If d metrizes weak convergence in Ω, then d is equivalent to ρ :

By choosing $\phi \equiv 1$ in Lemma 3, it follows that $\omega \in (\Omega, d) \rightsquigarrow P_\omega$ is continuous on Ω. The same conclusion follows directly with d replaced by ρ. By the identifiability assumption and compactness of Ω and metric range, ((cf. Proposition 9.5 of Royden 1968)) both are homeomorphisms. Thus d and ρ are equivalent.

Example 3. Let ϕ be a continuous function on $\mathscr{P}$, and consider the compound decision problem whose component problem is estimation of $\phi(P)$ under squared error loss. Let $\hat{\omega}$ be a symmetric mapping on $\mathscr{X}^{n-1}$ into Ω, and let $\underline{\hat{t}}$ be an equivariant rule which is Bayes versus $\hat{\omega}(X_{\check{\alpha}})$ in the α–th component. Then $\underline{\hat{t}}$ satisfies assumption (i) of Theorem 3.

Proof. Let $B = \|\phi(P)\|_\infty$. Since $(\hat{s} - \tilde{s})(P) = P(\hat{t}^2 - \tilde{t}^2) - 2\phi(P)P(\hat{t} - \tilde{t})$,

(14) $$\begin{aligned}(G_n - \hat{\omega})(\hat{s} - \tilde{s}) &= (P_{G_n} - P_{\hat{\omega}})(\hat{t}^2 - \tilde{t}^2) - 2(\nu_{G_n} - \nu_{\hat{\omega}})(\hat{t} - \tilde{t}) \\ &\le B^2\|P_{G_n} - P_{\hat{\omega}}\| + 4B\|\nu_{G_n} - \nu_{\hat{\omega}}\|\end{aligned}$$

since $\hat{t}$ and $\tilde{t}$ inherit the bound on ϕ. The conclusion now follows by the uniform ρ continuity of Lemma 3 with the choice $d = \rho$ justified by Remark 5. □

Example 4: Finite $\mathcal{A}$ and continuous loss functions Such decision problems satisfy a much stronger property than (i). Note that, for arbitrary $s : P \rightsquigarrow P \sum_a t_a L_a(P)$,

$$(15)\quad (G_n - \hat{\omega})s = \sum_a (G_n - \hat{\omega})(Pt_a)L_a(P) = \sum_a (\nu_{G_n} - \nu_{\hat{\omega}})t_a \le \sum_a \|\nu_{G_n} - \nu_{\hat{\omega}}\|.$$

But by Lemma 3 and Remark 5, $\forall\, \epsilon > 0\ \exists\, \delta > 0$ such that $\rho(G_n,\hat{\omega}) \ge \delta$ or $\|\nu_{G_n} - \nu_{\hat{\omega}}\| \le \epsilon$.

Theorem 3 of Gilliland and Hannan (1974–) reduced the problem of treating the asymptotic excess compound risk of equivariant "delete bootstrap rules" to the question of L_1–consistency of $\|\hat{\omega}_n - G_n\|$ for finite $\mathcal{P}$. Datta (1988) considered the compound estimation problem under squared error loss for real one parameter exponential families with compact parameter space and reduced the problem to the question of L_1–consistency of $\| P_{\hat{\omega}_n} - P_{G_n} \|$, under a domination assumption on translates of μ 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.

3. Examples of L_1–consistent posterior mixtures

In Theorem 3 we listed two conditions under which "delete bootstrap" rules are asymptotically optimal. In Examples 3 and 4 we considered situations where one of the conditions 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 3. Below we consider two classes of estimates of G_n that satisfy that 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 (1990).

More specifically; let μ be a measure and let $\mathcal{P}$ be the class of probability distributions with densities $\{p_\theta : \theta \in \Theta\}$ wrt μ, where Θ is a compact metric space. Suppose p(x) is continuous for each x and, with

$$h_{\theta'} = \sup_{\theta\in\Theta} |\log (p_\theta/p_{\theta'})|, \quad \sup_{\theta\in\Theta} \int(h_\theta - M)^+ p_\theta d\mu \longrightarrow 0 \qquad \text{as} \quad M \longrightarrow \infty.$$

Observe that as pointed out in Remark 3.2 of Datta (1990), the second part of the above assumption forces P_θ's to be pairwise mutually absolutely continuous. By the Scheffé theorem, continuity of p(x) for each x implies the norm–continuity of P_θ. The latter implies that $\mathcal{P}$ inherits the compactness of Θ.

Consider Ω with the topology of weak convergence and let Λ be a probability measure on the Borel subsets of Ω. Let $\hat{\Lambda}$ be the posterior distribution of ω given $\underline{X} = \underline{x}$. Then $\hat{\Lambda}$ is the probability measure on Ω with density proportional to $\prod_{i=1}^{n} p_\omega(x_i)$ with respect to Λ. Let $\hat{\omega}_n$ denote the $\hat{\Lambda}$–mix of ω's. Then Theorem 3.1 of Datta (1990) asserts that if Λ has full support

$$(16) \qquad \sup\{P \, \|P_{\hat{\omega}_n} - P_{G_n}\| : \underline{P} \in \mathscr{P}^n \} = o(1).$$

Since $(n+1)(P_{G_n} - P_{G_{n-1}}) = P_{\theta_n} - P_{\theta_{n+1}}$, its norm does not exceed 2. Thus, by triangulation about $P_{G_{n+1}}$, (16) is equivalent to (16) with G_n replaced by G_{n+1} or, equivalently, with $\hat{\omega}_n$ replaced by $\hat{\omega}_{n-1}$.

Observe that $\hat{\omega}_{n-1}$ is symmetric on $\mathscr{X}^{n-1}$. Therefore $\hat{\omega}_{n-1}$ provides an example of estimates that satisfy assumption (ii) of Theorem 3.

The importance of Datta's estimates is due to the fact that compound Bayes rules against a prior not depending on $\underline{X}$ turn out to be Bayes versus $\hat{\omega}_{n-1}(\underline{X}_{\check{\alpha}})$, in the α–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 $\hat{\omega}_{n-1}(\underline{X}_{\check{\alpha}})$ in the α–th component, will be admissible for each n. The uniqueness of the compound risk of Bayes rules versus a prior ζ in an estimation problem under squared error loss was shown in Section 4 of the appendix in Datta (1988), under the condition that P_θ is dominated by P_ζ for every θ.

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 λ 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 η a probability with support R^k and obtain estimators for the case where $\mathcal{P}$ is a class of distributions on R^k. Theorem 4, to follow, proves L_1– consistency of minimum $L_2(\eta)$–distance estimators of 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 η is a probability measure on R^k, any distribution function H is in $L_2(\eta)$ and d: $\underline{P} \rightsquigarrow \|F_{G_n} - H\|_\eta$ satisfies

$$(17) \qquad |d(\underline{P}) - d(\underline{P}')| \le \|F_{G_n} - F_{G_n'}\|_\eta \le \|P_{G_n} - P_{G_n'}\|$$

so that d is continuous on compact $\mathcal{P}^n$ and therefore attains a minimum.

Lemma 4. Let $\mathcal{X}$ be R^k and let η be a probability measure with support R^k. Let r be the pseudo–metric on Ω induced by the $L_2(\eta)$ norm on the range of $\omega \rightsquigarrow F_\omega$. Then r is a metric equivalent to ρ and $\omega \in (\Omega,\rho) \rightsquigarrow \omega \in (\Omega,r)$ is uniformly bicontinuous.

Proof. If $r(\omega,\omega') = 0$, then $F_\omega = F_{\omega'}$ a.e.(η) and therefore, by continuity from above, everywhere. Thus $P_\omega = P_{\omega'}$ and by identifiability

$\omega = \omega'$. Since $r \leq \rho$, the identity function on (Ω,ρ) to (Ω,r) is continuous. Therefore by compactness, as in Remark (5), it is uniformly bicontinuous.□

Theorem 4. Let $\mathscr{X}$ be R^k and let η be a probability measure with support R^k. Let $\tilde{G}_n$ be the empirical distribution of $\underline{\tilde{P}}$ a measurable minimizer of d_n: $\underline{P} \rightsquigarrow \|F_{G_n} - H_n\|_\eta$ with H_n the empirical distribution of $\underline{X}$. Then

$$\sup\{P\|P_{\tilde{G}_{n-1}} - P_{G_n}\| : \underline{P} \in \mathscr{P}^n\} = o(1). \tag{18}$$

Proof. Since $F_{G_n} = P(H_n)$, H_n as average of P–independent Bernoulli processes, has variance

$$P(F_{G_n} - H_n)^2 = \frac{1}{n} G_n(F(1-F)) \leq \frac{1}{4n}. \tag{19}$$

By the Fubini Theorem $P\|F_{G_n} - H_n\|_\eta^2$ has the same bound. By triangulation about H_n and use of the minimizing property of $\tilde{G}_n$,

$$r^2(\tilde{G}_n,G_n) = \|F_{\tilde{G}_n} - F_{G_n}\|_\eta^2 \leq 2\|F_{G_n} - H_n\|_\eta^2. \tag{20}$$

Let $\epsilon > 0$. By Lemma 4, take $\delta > 0$ such that $\rho \leq \epsilon$ or $r \geq \delta$. Then

$$P\,\|P_{\tilde{G}_n} - P_{G_n}\| \leq \epsilon + P\,\|P_{\tilde{G}_n} - P_{G_n}\|[r(\tilde{G}_n,G_n) \geq \delta]. \tag{21}$$

By the Markov inequality, the last term in (21) is bounded by

(22) $$\frac{1}{2\delta^2} P(\mathrm{lhs}(20)) \leq \frac{1}{\delta^2 n}$$

by (20)and the bound (19) for its P expectation.

The resulting bound for the lhs(21) proves (18) for the equivalent (as for (16) in A.) form with G_n replaced by G_{n-1}. □

Observe that $\tilde{G}_n$ can be taken to depend on $\underline{X}$ only through H_n and therefore is an example of $\hat{\omega}$ of Theorem 3.

BIBLIOGRAPHY

Billingsley, Patrick (1968). Convergence of Probability Measures. John Wiley & Sons.

Billingsley, Patrick (1971). Weak Convergence of Measures : Applications in Probability. SIAM.

Brown, Lawrence D (1986). Fundamentals of Statistical Exponential families with Applications in Statistical Decision Theory. Inst. of Mathematical Statistics, Lecture Notes – Monograph Series, vol. 9.

Datta, Somnath (1988). Asymptotically optimal Bayes compound and empirical Bayes estimators in exponential families with compact parameter space. Ph.D. Thesis, Dept. of Statistics and Probability , Michigan State University.

Datta, Somnath (1990). On the consistency of posterior mixtures and its applications. To appear in Ann. Statist.

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

Fabian, Václav and Hannan, James (1985). Introduction to Probability and Mathematical Statistics. John Willey & Sons.

Gilliland, Dennis C. and Hannan, James (1974). The finite state compound decision problem, equivariance and restricted risk components. RM–317 Dysyidyivd snf Probability, MSU; (1986)Adaptive Statistical Procedures and Related Topics. IMS Lecture Notes–Monograph Series **8** 129–145.

Hannan, James F. (1957). Approximation to Bayes risk in repeated play. Contribution to the Theory of Games **3**, Ann. Math. Studies, No. 39, Princeton University Press, 97–139.

Hannan, James and Huang, J.S. (1972a). Equivariant procedures in the compound decision problem with finite state component problem. Ann. Math. Statis. **43** 102–112.

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

Hannan James F. and Robbins, Herbert (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. Ann. Math. Statist. 36 1743–1752.

Parthasarathy, K. R. (1967). Probability measures on metric spaces. Academic Press.

Robbins, Herbert (1951). Asymptotically sub–minimax solutions of compound decision problems. Proc. Second Berkeley Symp. Math. Statist. Prob., Univ. of California Press, 131–148.

Rockafellar, R. Tyrrell (1970). Convex Analysis. No 28, Princeton Mathematical Series.

von Bahr, Bengt and Esseen, Carl–Gustav (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. Ann. Math. Statist. **36** 299–303.