This is to certify that the

dissertation entitled

Asymptotically optimal and admissible estimators in compound compact Gaussian shift experiments

presented by

Suman Majumdar

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

James Hannan
Major professor

Date August 5, 1992

# ASYMPTOTICALLY OPTIMAL AND ADMISSIBLE ESTIMATORS IN COMPOUND COMPACT GAUSSIAN SHIFT EXPERIMENTS

By

Suman Majumdar

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

1992

# ABSTRACT

## ASYMPTOTICALLY OPTIMAL AND ADMISSIBLE ESTIMATORS IN COMPOUND COMPACT GAUSSIAN SHIFT EXPERIMENTS

By

Suman Majumdar

The problem of finding admissible and asymptotically optimal compound and empirical Bayes rules is investigated in the context of decision about an infinite dimensional parameter.

The component experiment considered is a homogeneous experiment $\{P_\theta : \theta \in H\}$ on some measurable space $(\mathscr{X},\mathscr{F})$, where $H$ is a real separable Hilbert space, such that the map

$$\theta \mapsto \langle \theta, . \rangle_0 := \ln p_\theta(.) + \|\theta\|^2/2$$

is linear from $H$ into the real-valued measurable functions on $(\mathscr{X},\mathscr{F})$, where $p_\theta$ is a density of $P_\theta$ wrt $\mu = P_0$. This experiment is a *Gaussian shift* experiment in the sense of LeCam (1986) and $\{\langle \theta, . \rangle_0 : \theta \in H\}$ is the *isonormal* process on $(\mathscr{X},\mathscr{F},\mu)$ in the sense of Dudley (1967). The component problem estimates the shift parameter $\theta$ restricted to a compact subset of $H$ under squared error loss.

We consider the compound and empirical Bayes formulations of the above component problem and show that all Bayes estimators in the various formulations are *admissible*. Our main result : Any Bayes compound estimator versus a mixture of iid priors on the compound parameter is *asymptotically optimal* if the mixing hyperprior has full support. Analogously any Bayes empirical Bayes estimator is *asymptotically optimal* if the empirical Bayes prior has full support.

Using the (weak) conditional expectation representation of the Bayes estimator in the component problem and weak compactness of the unit ball, along with the fact that $\{<\theta,.>_0 : \theta \in H\}$ is the isonormal process and consequences thereof, we reduce the question of asymptotic optimality to that of an $L_1$ consistency of posterior mixtures. We prove the consistency result, which complements Datta (1991a), by assembling some previously known results and repeatedly using the Gaussian shift structure.

The dissertation also characterizes the support of a Dirichlet hyperprior on the set of all probability measures on a separable metric space to be those probability measures whose supports are contained in that of the parameter measure (of the Dirichlet hyperprior), proving a result stated in Ferguson (1973) for the line and providing examples of a full support hyperprior.

To my father and to the memory of my late mother

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Professor James F. Hannan whose erudition, vision of and dedication to Statistics always guided and inspired me. His caring personality provided strong support in some rather trying circumstances.

I would also like to thank Professors Dennis Gilliland, Hira L. Koul, V. Mandrekar and Habib Salehi for serving on my guidance committee, Professor Gilliland for some very useful comments on an earlier draft, Professor Mandrekar for alerting me to the (now realized) possibility of generalizations of the results initially presented and Professor R.V. Ramamoorthi for an important reference.

The opportunity provided by Professors Habib Salehi and James Stapleton on behalf of the Department of Statistics during the last two years to enhance my professional experience by teaching a wide variety of independent courses and working in the Statistical Consulting Service is gratefully acknowledged. Professors Gilliland and Stapleton very kindly taught me different aspects of the role of a Statistical Consultant.

The support provided by my sisters during my difficult student days in India and their encouragement to pursue graduate studies, as well as the support and encouragement provided by my wife during the preparation of this dissertation deserve special mention.

TABLE OF CONTENTS


Chapter | Page

0. INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . 1

0.1. The component and the compound problem . . . . . . 1

0.2. Literature review and a summary . . . . . . . . . 4

0.3. Notational Conventions . . . . . . . . . . . . . . 7

1. THE COMPOUND ESTIMATION . . . . . . . . . . . . . . . . . 9

1.1. The Gaussian shift component . . . . . . . . . . . 9

1.2. Estimators induced by hyperpriors . . . . . . . 14

1.2.1. Bayes versus mixture of iid priors . . . . 15

1.2.2. A useful inequality on the modified regret 16

1.3. A bound on the $L_1(P_\theta)$ distance between two component Bayes estimators . . . . . . . . . . . . . . . . . 17

1.4. Asymptotic optimality . . . . . . . . . . . . . . 23

1.5. Admissibility . . . . . . . . . . . . . . . . . . 27

2. CONSISTENCY OF THE POSTERIOR MIXTURES . . . . . . . . . . 29

3. THE EMPIRICAL BAYES ESTIMATION . . . . . . . . . . . . . 42

3.1. Bayes empirical Bayes . . . . . . . . . . . . . . 42

3.2. Asymptotic optimality . . . . . . . . . . . . . . 43

APPENDIX . . . . . . . . . . . . . . . . . . . . . . . . . 46

A.1. On measurability . . . . . . . . . . . . . . . . . 46

A.2. On topological support of Dirichlet prior . . . . 47

BIBLIOGRAPHY . . . . . . . . . . . . . . . . . . . . . . . 53

# CHAPTER 0
# INTRODUCTION

In Section 1 we describe the idea of compounding a decision problem (called the component problem) first espoused by Robbins (1951). In Section 2 we review that part of the existing literature on compound decision theory which can be considered to be a forerunner to our work and present a summary of it. In Section 3 we state the notational conventions to be followed throughout the dissertation (some of these conventions will be used informally in Sections 1 and 2).

## 1. The component and the compound problem.

The component problem is a usual decision theory problem, consisting of a parameter set $\Theta$, a family of probability measures $\{P_\theta : \theta \in \Theta\}$ on some measurable space $(\mathfrak{X},\mathfrak{F})$, an observable $\mathfrak{X}$-valued random element $X \sim P_\theta$ under $\theta$, an action space $\mathcal{A}$, a loss function $L:\mathcal{A}\times\Theta\mapsto[0,\infty)$ and decision rules t, $t:\mathfrak{X}\mapsto\mathcal{A}$ such that $L(t,\theta)$ is measurable $\forall\ \theta$ with risk $R(t,\theta) := P_\theta L(t,\theta)$.

For consideration of Bayes solutions, we fix a $\sigma$-algebra of subsets of $\Theta$ such that each of the maps $(x,\theta)\mapsto L(t(x),\theta)$ is jointly measurable. Let $\Omega=\{\omega : \omega \text{ is a probability on } \Theta\}$. For $\omega\in\Omega$, let $r(\omega)$ and $\tau_\omega$ respectively denote the minimum Bayes risk and a Bayes rule versus $\omega$ in the component problem (we assume existence of $\tau_\omega$ for every $\omega$). That is,

$$r(\omega) = \bigwedge_t \int_\Theta R(t,\theta)d\omega(\theta) = \int_\Theta R(\tau_\omega,\theta)d\omega(\theta).$$

The compound problem simultaneously considers a number, say n, of independent decision problems, each of which is structurally identical to the above component problem. The compound loss is taken to be the average of the component losses. In the set compound version a decision about each component parameter is reached by using data from all the component problems, while in the sequence compound version only $\underline{X}_\alpha$, data up to stage $\alpha$, is used in making the $\alpha$-th decision. Thus for each $n \geq 1$, the compound problem is also a decision problem, with parameter set $\Theta^n$, family of probability measures $\{P_{\underline{\theta}} := \times_{\alpha=1}^{n} P_{\theta_\alpha} : \underline{\theta} := (\theta_1,\ldots,\theta_n) \in \Theta^n\}$ on the measurable space $\mathfrak{X}^n$, observations $\underline{X} = (X_1,\ldots,X_n) \sim P_{\underline{\theta}}$ under $\underline{\theta}$, action space $\mathcal{A}^n$, decision rules $\underline{t}:\mathfrak{X}^n \mapsto \mathcal{A}^n$ such that each $L(t_\alpha,\theta_\alpha)$ is measurable, loss

$$L_n(\underline{t},\underline{\theta}) := n^{-1} \sum_{\alpha=1}^{n} L(t_\alpha,\theta_\alpha)$$

and corresponding risk

$$R_n(\underline{t},\underline{\theta}) := P_{\underline{\theta}} L_n(\underline{t},\underline{\theta}). \tag{1.1}$$

If we were going to use only the data from a particular component problem to decide about that component parameter, then the component problems being structurally identical, there is an intuitive reason to use the same procedure (with different data) in the different problems. Formally, that amounts to using a compound procedure $\underline{t}$, for which $t_\alpha(\underline{x}) = t(x_\alpha)$ $\forall$ $\alpha = 1,\ldots,n$, where t is a component procedure; such a compound procedure is called *simple symmetric*.

Let $G_n$ denote the empirical distribution of $(\theta_1,\ldots,\theta_n)$. The

compound risk at $\underline{\theta}$ of a simple symmetric $\underline{t}$ reduces to the component Bayes risk of t versus $G_n$, where $t_\alpha(\underline{x}) = t(x_\alpha)\ \forall\ \alpha = 1,\ldots,n$; as such it is at least $r(G_n)$, the minimum Bayes risk versus $G_n$, which is referred to as the *simple envelope* at $\underline{\theta}$. For a compound rule $\underline{t}$, the difference $D_n(\underline{t},\underline{\theta}) = R_n(\underline{t},\underline{\theta}) - r(G_n)$ is called the *modified regret* of $\underline{t}$ at $\underline{\theta}$ and a sequence of compound rules $\{\underline{t} : n \geq 1\}$ is said to be *asymptotically optimal* (a.o.) if

$$(1.2)\qquad \bigvee_{\underline{\theta}} D_n(\underline{t},\underline{\theta}) \to 0 \text{ as } n\to\infty.$$

However, it has long been recognized (Hannan and Robbins (1955)) that the compound problem is invariant under the group of n! permutations of coordinates; also, almost all the compound rules in the literature are equivariant under the permutation group. Hence a more appropriate yardstick to judge the performance of a compound rule should be the *equivariant envelope*, the minimum compound risk of equivariant rules (see Gilliland and Hannan (1986) for a discussion of equivariance in compound decision problems).

Mashayekhi (1990) has shown that if the component problem involves a compact (in total variation norm) class of mutually absolutely continuous probability measures, then the excess of the simple envelope over the equivariant envelope goes to zero uniformly in the measures. We shall use that result to extend our optimality result against the simple envelope to that against the equivariant envelope.

A sequence of compound rules $\{\underline{t} : n \geq 1\}$ is said to be *admissible* if for every n it is admissible in the usual sense.

## 2. Literature review and a summary.

The problem of exhibiting compound rules which are a.o. as well as admissible has been an interesting and challenging question ever since it was put forward by Robbins (1951) in his pioneering paper of compound decision theory. He considered the problem of decision between $\mathcal{N}(-1,1)$ and $\mathcal{N}(1,1)$, exhibited an a.o. compound procedure and conjectured that the Bayes compound rule versus the symmetric prior uniform on proportions might have better risk behavior, exactly or asymptotically, than his a.o. rule. [That it will not be exactly superior to the bootstrap rule of Robbins was shown by Huang (1972).]

A.o. compound rules whose components are typically Bayes versus some estimates of the unknown $G_n$ or direct estimates of the Bayes rule versus $G_n$ have been worked out for many different component problems. In particular, when the component problem is an estimation problem under squared error loss, Gilliland (1968) and Singh (1974) obtained a.o. sequence compound rules with rates (we say $\underline{t}$ is a.o. with rate $\alpha_n$ if $\bigvee_{\underline{\theta}} D_n(\underline{t},\underline{\theta}) = O(\alpha_n)$) for discrete and Lebesgue exponential components respectively. But these rules are inadmissible in the sense of the previous section.

Making use of results from Gilliland and Hannan (1974), which was later published in 1986, Gilliland, Hannan and Huang (1976) obtained admissible and a.o. rules with rate $n^{-1/2}$ where the component problem was a two-state restricted risk problem. [They did not specify admissibility of their rules. But they considered full support hyperprior mixing of independent identically distributed (iid hereafter) priors on the compound state space to generate full support priors on it and looked at the resulting Bayes rules. Since the risk in their problem is trivially continuous (the state space is discrete), the resulting Bayes rules are admissible.]

The first solution to the problem of exhibiting compound rules which are a.o. as well as admissible when the component problem involves decision among infinitely many probability measures, has been provided by Datta (1988/91b). The component problem there is the squared error loss estimation of an arbitrary continuous transform of the natural parameter of a large compact subclass of a one parameter exponential family.

Since then, Mashayekhi (1990) proposed a class of admissible and a.o. procedures in the restricted risk compact component compound decision problem. This was extended by Zhu (1992), who successfully exploited Datta's (1991a) result about consistency of the posterior mixtures to obtain admissible and a.o. rules when the component problem involves equi(in actions) continuous loss functions in a multiparameter exponential family with parameter set restricted to a polytope inside the natural parameter set.

The present work seems to be the first to accomplish asymptotic optimality when the component problem is the estimation of an infinite dimensional parameter. In fact, it accomplishes admissibility and asymptotic optimality simultaneously. Our component distributions, indexed by a real separable Hilbert space, form a *Gaussian shift* experiment. We consider the component problem of squared error loss estimation of the Hilbert-valued shift parameter restricted to a compact subset of the Hilbert space. We note that all Bayes estimators in our compound problem are admissible. Our main result is that a Bayes compound estimator versus a mixture of iid priors on the compound parameter is a.o. if the mixing hyperprior has full support.

The dissertation is organized as follows.

Chapter 1 treats the compound estimation problem. Section 1 formally introduces the component distributions as satisfying an assumption (A). That assumption immediately identifies the experiment to be a Gaussian shift experiment. Section 2 describes the Bayes estimator versus the above mentioned mixture of iid priors and establishes a bound on the modified regret of such an estimator. Section 3 establishes an upper bound on the distance between two component Bayes estimators in terms of the $L_1$ distance between the corresponding mixtures. Section 4 combines the results in Sections 2 and 3 to establish asymptotic optimality, first against the simple envelope and then against the equivariant envelope, assuming posterior mixtures are $L_1$ consistent for the empirical mixture. In this section we provide a closed form expression of our estimator and examples of a full support hyperprior. In Section 5, we show that every Bayes estimator in our compound problem, in particular a Bayes estimator versus a mixture of iid priors, is admissible.

In Chapter 2 we establish the consistency of the posterior mixtures assumed in proving asymptotic optimality in Chapter 1. In the process we get that the very general sufficient conditions given by Datta (1991a) for this kind of consistency of the posterior mixtures are by no means necessary.

Chapter 3 looks at the empirical Bayes problem of Robbins (1951, 1956) with the component problem described above. Admissibility and asymptotic optimality (defined in that chapter) follow from the compound results.

Finally, in Section 1 of the Appendix we prove two measurability lemmas that are used in the main body of the dissertation; in Section 2, we characterize the topological support of a Dirichlet prior on a separable metric space, which is used in Section 4, Chapter 1 to give examples of a full support hyperprior.

**3. Notational conventions.**

Given any n-tuple $\underline{x} = (x_1, \ldots, x_n)$ of elements from a set, for each $1 \leq \alpha \leq n$, $\underline{x}_\alpha$ denotes the $\alpha$-tuple $(x_1, \ldots, x_\alpha)$. For probabilities $P_1, \ldots, P_n$, $\times_{i=1}^{n} P_i$ denotes their measure theoretic product; when $P_i = P$ $\forall$ i, $\times_{i=1}^{n} P_i$ is denoted by $P^n$. For sets $\{A_i : 1 \leq i \leq n\}$, $\times_{i=1}^{n} A_i$ denotes their set theoretic product; when $A_i = A$ $\forall$ i, $\times_{i=1}^{n} A_i$ is denoted by $A^n$. To denote the integral of a function f with respect to (wrt hereafter) a measure $\mu$, we will interchangeably use the standard integral notation $\int f d\mu$ and the left operator notation $\mu(f)$, or even $\mu f$; depending on typographical convenience and the emphasis to be conveyed, the dummy variable of integration in the integral notation will be sometimes displayed, sometimes only partially displayed and sometimes hidden altogether. Sets are always identified with their indicator functions. The same is true for probabilities and their induced expectations. $\Re$ stands for the real line. If X is a random element on a probability space (.,.,P), then $PX^{-1}$ denotes the P-induced distribution of X on the range space. The notation a := b will mean that a is defined to be b. The set theoretic complement of a set A will be denoted by $\tilde{A}$, except in Section 2 of the Appendix, where the more traditional $A^c$ will be used. The following numbering convention will be used throughout : All numberings of displays and statements are local within a chapter. For chapters with multiple sections, (2.1) will refer to the first display in the second section; for chapters with a single section, (3) will refer to the third display. On occasions when we have to refer to numberings in other chapters, the reference will be explicit, e.g. Theorem 1 of Chapter 2 or Lemma 1.1 of Chapter 1.

# CHAPTER 1
# THE COMPOUND ESTIMATION

In this chapter we consider the compound problem as described in Chapter 0 corresponding to the *Gaussian shift* component problem to be introduced below. We prove asymptotic optimality of Bayes rules versus (full support hyperprior) mixture of iid priors [Theorem 4.1], which is the main result of the dissertation, using the consistency of the posterior (distribution of $\theta_n$ under the mixed compound prior given $\underline{x}_{n-1}$) mixtures, a result of independent interest stated and proved in Chapter 2. Section 1 describes the component problem to be investigated and assembles some pertinent facts about it. In Section 2 we calculate a Bayes estimator in the compound problem versus a mixture of iid priors on the compound parameter and obtain a useful upper bound on its absolute modified regret. In Section 3, we obtain an upper bound on the distance between two component Bayes rules in terms of the $L_1$ distance between the corresponding mixtures, which is used in Section 4 in conjunction with the bound on the absolute modified regret obtained in Section 2 to prove the main result. In Section 5 we show that every Bayes estimator in the compound problem, in particular a Bayes estimator versus a mixture of iid priors, is admissible.

**1. The Gaussian shift component.**

We consider the squared error loss estimation problem in a Hilbert indexed Gaussian shift experiment. Let H be a real separable

Hilbert space (with $\|f\|$ denoting the norm of an element f in H and $<.,.>$ the inner product) and $\{P_\theta : \theta \in H\}$ be a family of probabilities on a measurable space $(\mathfrak{X},\mathfrak{F})$ specified by (strictly positive) densities $\{p_\theta : \theta \in H\}$ wrt $\mu = P_0$, such that

(A) the map $\theta \mapsto <\theta,.>_0 := \ln p_\theta(.) + \|\theta\|^2/2$ is linear from H into the linear space of all real-valued measurable functions on $(\mathfrak{X},\mathfrak{F})$.

We consider the component problem with $\Theta$ a compact subset of $H \supset \mathcal{A} \supset \Theta$ and $L(a,\theta) = \|a-\theta\|^2$.

The contents of the remainder of this section are as described below : We show that $\{<\theta,.>_0 : \theta \in H\}$ is the *isonormal* process on $(\mathfrak{X},\mathfrak{F},\mu)$ in the sense of Dudley (1967) [Remark 1.1], which in turn identifies the experiment under investigation to be a ***Gaussian shift*** experiment in the sense of LeCam (1986) [Remark 1.2]. We show that $(\theta,x) \mapsto p_\theta(x)$ and $(\omega,x) \mapsto p_\omega(x) := \int p_\theta(x)d\omega(\theta)$ are jointly measurable when $\Omega$ (the set of all probabilities on $\Theta$) is endowed with the topology of weak convergence and the corresponding Borel $\sigma$-field [Remark 1.3]. We then show that a Bayes estimator in the component problem must be the (weak) posterior expectation [Lemma 1.1]. We close the section by proving two lemmas [Lemma 1.2 and Lemma 1.3] describing certain features of the component problem that are used in the sequel.

Remark 1.1 (The *isonormal* process). Since by (A) $p_\theta = \exp(-\|\theta\|^2/2 + <\theta,.>_0)$ $\forall$ $\theta \in H$, by representing the lhs below as a $\mu$ integral, using the linearity of the map in (A) to treat the integrand and representing the resulting integral as a $P_{\eta+t\theta}$

integral, we get $\forall\ t \in \Re$ and $\theta,\eta \in H$,

$$P_\eta[\exp\{t < \theta,. >_0\}] = \exp\{t < \theta,\eta > + t^2\|\theta\|^2/2\},$$

which by uniqueness of moment generating function proves

$$\text{(1.1a)} \qquad P_\eta < \theta,. >_0{}^{-1} = \mathcal{N}( < \theta,\eta > ,\ \|\theta\|^2) \qquad \forall\ \theta,\eta \in H.$$

The linearity assumption in (A) then shows

(1.1b) $\{ < \theta,. >_0\ :\ \theta \in H\}$ is a centered Gaussian process on $(\mathfrak{X},\mathfrak{F},\mu)$.

By the (polar) representation of the product of two numbers in terms of the square of their sum and the individual squares, using the linearity of the map in (A) and (1.1a) with $\eta = 0$, we get

$$\text{(1.2)} \qquad \mu( < \theta,. >_0 < \eta,. >_0) = < \theta,\eta > \qquad \forall\ \theta,\eta \in H.$$

Now, the assertions in (1.1b) and (1.2) show that the process $\{ < \theta,. >_0\ :\ \theta \in H\}$ is *isonormal* in the sense of Dudley (1967). //

<u>Remark</u> <u>1.2</u> (*Gaussian shift* experiment). Note that by (1.1b), the experiment under investigation is a *Gaussian shift* experiment in the sense of Definition 2 of Chapter 9 of LeCam (1986). Even though the definition in LeCam does not require the indexing set to be a Hilbert space, discussions following it show that it suffices to restrict attention to that case. //

<u>Remark</u> <u>1.3</u> (Joint measurability of densities). Let $\{e_j : j \geq 1\}$ be an orthonormal basis of $\mathbf{H}$. By (1.1b) and (1.2), we get that $\{<e_j, . >_0 : j \geq 1\}$ are independent random variables on $(\mathfrak{X}, \mathfrak{F}, \mu)$. Let $\theta_n := \sum_{j=1}^{n} <\theta, e_j> e_j$. By linearity of the map in (A), $<\theta_n, . >_0 = \sum_{j=1}^{n} <\theta, e_j> <e_j, . >_0$. Since $\theta_n \to \theta$ in $\mathbf{H}$, $<\theta_n, . >_0$ converges to $<\theta, . >_0$ in $L_2(\mu)$ by (1.2); since the $<\theta_n, . >_0$ are the partial sums of a sequence of independent random variables, by Levy's Theorem [Theorem 3.3.1, Chow and Teicher (1988)], the convergence is $\mu$-a.s. as well. Since $<\theta_n, . >_0$ is continuous in $\theta$ and a measurable function on $\mathfrak{X}$, it is jointly (in $\theta$ and x) measurable by Doob's Theorem. That implies the joint measurability of its $\mu$-a.s. limit and hence that of $p_\theta$.

Let $\Omega$, the set of all probabilities on the Borel $\sigma$-field of $\Theta$, be endowed with the topology of weak convergence and the corresponding Borel $\sigma$-field. For $\omega \in \Omega$, let $p_\omega(x) := \int p_\theta(x) d\omega$; this is clearly a density of the mixture $P_\omega := \int P_\theta d\omega$ wrt $\mu$. The map $(\omega, x) \mapsto p_\omega(x)$ is jointly measurable by the joint measurability of $(\theta, x) \mapsto p_\theta(x)$ and Lemma 1.2 of the Appendix. //

The next lemma characterizes a Bayes estimator in our component problem. Specializing the notation introduced in Chapter 0 we shall denote a Bayes estimator (versus $\omega$) in the component problem by $\tau_\omega$.

Throughout the remainder of the dissertation, let

$$(1.3) \qquad M = \sup\{\|\theta\| : \theta \in \Theta\}.$$

**Lemma** 1.1. On the common support of $\{P_\nu : \nu \in \Omega\}$, $\tau_\omega$ is the unique mapping into H satisfying

$$<\tau_\omega, h> \; = \int_\Theta <\eta, h> (p_\eta/p_\omega) d\omega(\eta) \qquad \forall\ h \in H.$$

**Proof.** We first show that for any probability measure $\pi$ on $\Theta$, $\exists$ an unique element $v(\pi)$ in H satisfying

$$(1.4) \qquad <v(\pi), h> \; = \int_\Theta <\eta, h> d\pi(\eta) \qquad \forall\ h \in H.$$

Since the map $h \mapsto \int_\Theta <\eta, h> d\pi(\eta)$ is a linear functional on H whose norm is bounded by M, the assertion of (1.4) follows from the Riesz-Frechet Theorem [Theorem 5.5.1, Dudley (1989)].

Note that if $p_\omega(x)$ is positive, the map $\theta \mapsto p_\theta(x)/p_\omega(x)$ is a density (wrt $\omega$) of a probability measure $\tilde{\omega}_x$ on $\Theta$. By (1.4), it is enough to show that $\tau_\omega = v(\tilde{\omega})$ on the common support of $\{P_\nu : \nu \in \Omega\}$. Now, by Fubini's Theorem, the Bayes risk (versus $\omega$) of an estimator t is equal to

$$(1.5) \qquad \int_{\mathfrak{X}} [ \int_\Theta \| t(x) - \theta \|^2 d\tilde{\omega}_x ] dp_\omega(x) d\mu(x).$$

Triangulating around $v(\tilde{\omega}_x)$ and expanding the norm square of the sum, the inner integral in (1.5) is

$$\| t(x) - v(\tilde{\omega}_x) \|^2 + \int_\Theta \| v(\tilde{\omega}_x) - \theta \|^2 d\tilde{\omega}_x,$$

which is minimized iff $t(x) = v(\tilde{\omega}_x)$, completing the proof. //

**Lemma 1.2.** For every finite sequence $\{\theta_i : 1 \le i \le k\} \subset H$ and $\{a_i : 1 \le i \le k\} \subset \Re$,

$$2\log\left( \int \prod_{i=1}^k p_{\theta_i}^{a_i} d\mu \right) = \left\| \sum_{i=1}^k a_i \theta_i \right\|^2 - \sum_{i=1}^k a_i \| \theta_i \|^2.$$

**Proof.** Starting with the functional form of $p_{\theta_i}$ implicit in (A), the assertion follows by using the linearity of the map in (A) and the functional form of $p_\xi$, where $\xi = \sum_{i=1}^{k} a_i\theta_i$. //

Throughout the remainder of the dissertation, let $\|f\|_q$ denote the $L_q(\mu)$ norm of a function f in $L_q(\mu)$.

**Lemma 1.3.** For every $\omega \in \Omega$ and every integer $q \geq 1$,

$$p_\omega \in L_q(\mu) \text{ and } \|p_\omega\|_q \leq e^{(q-1)M^2/2}.$$

**Proof.** Writing $p_\omega{}^q$ as a q-fold iterated integral, interchanging the order of integration on $\mathfrak{X}$ and $\Theta^q$, applying Lemma 1.2 (with k=q and $a_i = 1 \ \forall\ i$), and using (1.3), we get

$$\mu(p_\omega{}^q) \leq \exp\{q(q-1)M^2/2\}, \tag{1.6}$$

completing the proof. //

## 2. Estimators induced by hyperpriors.

In Subsection 2.1 we show that the $\alpha$-th component of a Bayes estimator in the compound problem versus a mixture of iid priors on the compound parameter is the Bayes estimator in the component problem versus the posterior mean under the mixing hyperprior given the data from the other problems; in Subsection 2.2 we obtain an upper bound on the absolute modified regret of such an estimator in terms of the distance between its $\alpha$-th component and a component Bayes rule versus the empirical state distribution.

*2.1. Bayes versus mixture of iid priors.*

Since $\Theta$ is a compact metric space, by Theorem II.6.4 of Parthasarathy (1967), $\Omega$ with the topology of weak convergence is also a compact metric space; let $\mathfrak{B}(\Omega)$ denote its Borel $\sigma$-field. Let $\Lambda$ be a probability measure on $(\Omega,\mathfrak{B}(\Omega))$. We take $\Lambda$-mixture of iid priors on $\Theta^n$ (for each n) and denote that prior by $\bar{\omega}_{\Lambda,n}$. [The measure $\bar{\omega}_{\Lambda,n}$ is defined on the class of measurable rectangles by

$$(2.1) \qquad \bar{\omega}_{\Lambda,n}(B_1 \times B_2 \times \dots \times B_n) = \int_{\Omega} \prod_{i=1}^{n} \omega(B_i) d\Lambda,$$

and then extended to the product $\sigma$-field. Note that by Lemma 1.1 of the Appendix the above integrand is measurable.]

Let $\underline{t} = (t_1,\dots,t_n)$, where $t_\alpha : \mathfrak{X}^n \mapsto \mathcal{A}$ is a measurable function, be an estimator in the set compound problem. The $\alpha$-th component Bayes risk of $\underline{t}$ versus $\bar{\omega}_{\Lambda,n}$ is

$$(2.2) \qquad R(t_\alpha,\bar{\omega}_{\Lambda,n}) = \int_{\Omega} \int_{\mathfrak{X}^{n-1}} [\int_{\Theta} \int_{\mathfrak{X}} \| t_\alpha - \theta_\alpha \|^2 \, dP_{\theta_\alpha} d\omega] dP_\omega^{\,n-1} d\Lambda.$$

Disintegrating the joint probability on $\mathfrak{X}^{n-1} \times \Omega$ determined by $(dP_\omega^{\,n-1} d\Lambda)$ as $(d\Lambda_{\alpha,n} dP_{\bar{\omega}_{\Lambda,n-1}})$, where $\Lambda_{\alpha,n}$ is the posterior distribution of $\omega$ (under $\Lambda$) given $(x_1,\dots,x_{\alpha-1},x_{\alpha+1},\dots,x_n)$ [since $\Omega$ is a Polish (in fact compact metric) space, by Theorem 10.2.2 of Dudley (1989), such a disintegration exists], we get

$$(2.3) \qquad \text{lhs}(2.2) = \int_{\mathfrak{X}^{n-1}} [\int_{\Theta} \int_{\mathfrak{X}} \| t_\alpha - \theta_\alpha \|^2 \, dP_{\theta_\alpha} d\omega_{\alpha,n}] dP_{\bar{\omega}_{\Lambda,n-1}},$$

where $\omega_{\alpha,n}$ denotes the $\Lambda_{\alpha,n}$ mix of $\omega$'s. Clearly, rhs(2.3) is minimized by choosing $t_\alpha(\underline{x}) = \tau_{\omega_{\alpha,n}}(x_\alpha)$. Since the compound risk is the average of the component risks, the Bayes estimator in the set compound problem versus the prior $\bar{\omega}_{\Lambda,n}$ is given by $\hat{\underline{t}}$, where

(2.4a) $$\hat{t}_\alpha(\underline{x}_n) = \tau_{\omega_{\alpha,n}}(x_\alpha).$$

A similar argument shows that the Bayes estimator in the sequence compound problem versus the prior $\bar{\omega}_{\Lambda,n}$ is given by $\underline{t}'$, where

(2.4b) $$t'_\alpha(\underline{x}_n) = \tau_{\omega_{\alpha,\alpha}}(x_\alpha).$$

*2.2. A useful inequality on the modified regret.*

Recall from Chapter 0 that $G_n$ stands for the empirical distribution of $\theta_1,\ldots,\theta_n$. For every $\underline{\theta} \in \Theta^n$, by definition,

$$D_n(\hat{\underline{t}},\underline{\theta}) = n^{-1}\sum_{\alpha=1}^{n} P_{\underline{\theta}}[\|\hat{t}_\alpha - \theta_\alpha\|^2 - \|\tilde{t}_\alpha - \theta_\alpha\|^2],$$

where $\tilde{t}_\alpha(\underline{x}) = \tau_{G_n}(x_\alpha)$. Using Cauchy-Schwartz inequality to bound the absolute difference between $\|d\|^2$ and $\|b\|^2$ by $\|d+b\|$ times $\|d-b\|$, triangle inequality in H and (1.3), we get

(2.5) $$|D_n(\hat{\underline{t}},\underline{\theta})| \le 4Mn^{-1}\sum_{\alpha=1}^{n} P_{\underline{\theta}}\|\hat{t}_\alpha - \tilde{t}_\alpha\|.$$

Since $\hat{t}_\alpha(\underline{x}) = \tau_{\omega_{\alpha,n}}(x_\alpha)$,

(2.6) $$P_{\underline{\theta}}\|\hat{t}_\alpha - \tilde{t}_\alpha\| = P_{\underline{\theta}}P_{\theta_\alpha}\|\tau_{\omega_{\alpha,n}} - \tau_{G_n}\|;$$

to investigate the bound on the absolute modified regret given by (2.5), we therefore consider $P_\theta\|\tau_\omega - \tau_\pi\|$ , where $\theta \in \Theta$ and $\omega,\pi \in \Omega$.

## 3. A bound on the $L_1(P_\theta)$ distance between two component Bayes rules.

In Proposition 3.1 we derive a bound on $P_\theta\|\tau_\omega - \tau_\pi\|$ essentially in terms of the total variation distance between the corresponding mixtures. Abusing notation we shall use $\|\sigma\|$ to denote the total variation norm of a signed measure $\sigma$ on $(\mathfrak{X},\mathfrak{F})$ as well.

The next three lemmas are used to prove Proposition 3.1.

**Lemma 3.1.** Let $<\omega,x>_0 := \int <\theta,x>_0 d\omega(\theta)$. Then

$$\mu <\omega,.>_0^{-1} = \mathcal{N}(0,\ \int\int <\eta,\xi> d\omega(\eta)d\omega(\xi)).$$

**Proof.** By (1.1b), $<\omega,.>_0$ is normally distributed if $\omega$ is finitely supported. Since $(\theta,\eta)\mapsto <\theta,\eta>$ is continuous and bounded (on compact $\Theta^2$), the map taking $(\omega,\pi)$ to the $L_2(\mu)$ inner product of $<\omega,.>_0$ and $<\pi,.>_0$ (which by interchanging the order of integration and using (1.2) is seen to be $(\omega\times\pi)<.,.>$) is continuous. Continuity of $\omega\mapsto <\omega,.>_0$ in $L_2(\mu)$ follows. Since $\Omega$ has a dense subset consisting of finitely supported measures [Theorem II.6.3, Parthasarathy (1967)], and a family of normally distributed random variables is closed under $L_2$ convergence, we get that $<\omega,.>_0$ is normally distributed. The expression for the mean and the variance follows by using Fubini's Theorem. //

The following lemma is Lemma A.1 of Datta (1988).

**Lemma (Datta-Singh):** For $(y,z,Y,Z,L) \in \Re^5$ such that $z \neq 0$ and $L \geq 0$,

$$|z|\{|\tfrac{y}{z} - \tfrac{Y}{Z}| \wedge L\} \leq |y - Y| + (|\tfrac{y}{z}| + L)|z - Z|.$$

**Lemma 3.2.** Given $\delta > 0$, $\exists\ \{h_1,\ldots,h_I\} \subset \mathcal{W} := \{h \in H : \|h\| \leq 1\}$ such that, for all real numbers a and b,

$$\exp(-M^2/2 - a + b)\mu(p_\theta \|\tau_\omega - \tau_\pi\| [<\theta,.>_0 \leq a][<\omega,.>_0 > b]) \tag{3.1}$$
$$\leq 2\delta + \sum_{i=1}^{I} \mu|\int <\theta,h_i> p_\theta d(\omega - \pi)| + 3M\|P_\omega - P_\pi\|.$$

**Proof.** Starting with the definition of $p_\omega (= \int p_\theta d\omega)$, recalling the functional form of $p_\theta$ implicit in (A), using (1.3) to bound $p_\theta$ below, applying Jensen's inequality to the exponential function, and noting that $p_\theta[<\theta,.>_0 \leq a]e^{-a} \leq 1$ and $e^{<\omega,.>_0} \geq e^b[<\omega,.>_0 > b]$, we get

$$\exp(-M^2/2 - a + b)p_\theta[<\theta,.>_0 \leq a][<\omega,.>_0 > b] \leq p_\omega.$$

In view of the above it suffices to show that $\mu(p_\omega \|\tau_\omega - \tau_\pi\|)$ can be bounded by rhs(3.1).

By Lemma 1.1,

$$\|\tau_\omega - \tau_\pi\| = \bigvee_{\mathcal{W}}\{|\int_\Theta <.,h> d\tilde{\omega} - \int_\Theta <.,h> d\tilde{\pi}|\}, \tag{3.2}$$

where $\tilde{\omega}$ and $\tilde{\pi}$ are as in the proof of Lemma 1.1.

Applying Datta-Singh Lemma with $z = p_\omega$, $y = \int_\Theta <\eta,h> p_\eta d\omega(\eta)$, $Z = p_\pi$, $Y = \int_\Theta <\eta,h> p_\eta d\pi(\eta)$ and $L = 2M$,

$$(3.3) \qquad p_\omega \left| \left( \int_\Theta <\eta,h> p_\eta d\omega(\eta)/p_\omega \right) - \int_\Theta <\eta,h> p_\eta d\pi(\eta)/p_\pi \right|$$

$$\leq \left| \int_\Theta <\eta,h> p_\eta d(\omega-\pi)(\eta) \right| + 3M|p_\omega - p_\pi|.$$

Since $\Theta$ is compact by assumption and $\mathcal{W}$ is weakly compact by the Banach-Alaoglu Theorem, $\Theta\times\mathcal{W}_w$ is compact. Since H is separable, $\mathcal{W}_w$ and hence $\Theta\times\mathcal{W}_w$ is metrizable. Since $(\theta,h)\mapsto<\theta,h>$ is a continuous function on $\Theta\times\mathcal{W}_w$, it is uniformly continuous. That implies $\{h\mapsto<\theta,h> \ : \ \theta\in\Theta\}$ is an equi(in $\theta$) uniformly continuous family of functions on $\mathcal{W}_w$, so that for every $\omega\in\Omega$

$$(3.4) \qquad \mu\left|\int(<\theta,h> - <\theta,h'>)p_\theta d\omega\right|$$
$$\leq \bigvee_\Theta |<\theta,h> - <\theta,h'>|$$
$$\leq \delta,$$

if the distance between h and h', in a metric metrizing $\mathcal{W}_w$, is less than $\epsilon = \epsilon(\delta)$. If weak-balls of radius $\epsilon$ around $\{h_1,\ldots,h_I\}$ cover $\mathcal{W}$, then triangulating around appropriate $h_i$, using (3.4) and dominating the maximum of I non-negative terms by their sum, we get

$$(3.5) \qquad \mu\left(\bigvee_{\mathcal{W}} \left|\int<\theta,h>p_\theta d(\omega-\pi)\right|\right) \leq 2\delta + \sum_{i=1}^{I}\mu\left|\int<\theta,h_i>p_\theta d(\omega-\pi)\right|.$$

The lemma follows from (3.2),(3.3) and (3.5). //

**Proposition 3.1.** Let $\gamma>0$ be fixed arbitrarily. Then, $\exists$ a number $\mathcal{K}$ such that

$$P_\theta\|\tau_\omega - \tau_\pi\| \le 5\gamma + 9G\|P_\omega - P_\pi\|.$$

**Proof.** For arbitrary real numbers a and b, partitioning H into the sets $[<\theta,.>_0 \le a][<\omega,.>_0 > b]$, $[<\theta,.>_0 \le a][<\omega,.>_0 \le b]$ and $[<\theta,.>_0 > a]$, using the bound $\|\tau_\omega - \tau_\pi\| \le 2M$ on the last two sets and Cauchy-Schwartz inequality in $L_2(\mu)$ on the remaining factors, and bounding $\|p_\theta\|_2$ by $e^{M^2/2}$ (see Lemma 1.3), we get

$$\begin{aligned} &P_\theta\|\tau_\omega - \tau_\pi\| \\ (3.6)\qquad &\le 2Me^{M^2/2}\{(\mu[<\theta,.>_0 > a])^{1/2} + (\mu[<\omega,.>_0 \le b])^{1/2}\} \\ &+ \mu(p_\theta\|\tau_\omega - \tau_\pi\|[<\theta,.>_0 \le a][<\omega,.>_0 > b]). \end{aligned}$$

By (1.1a), using the familiar bound on the upper tail of a normal distribution and (1.3), we get, for $a > 0$

$$(3.7)\qquad \mu[<\theta,.>_0 > a] \le (2\pi)^{-1/2} M a^{-1} \exp(-a^2/2M^2).$$

Similarly, using Lemma 3.1, for $b < 0$

$$(3.8)\qquad \mu[<\omega,.>_0 \le b] \le (2\pi)^{-1/2} M(-b)^{-1} \exp(-b^2/2M^2).$$

In view of (3.7) and (3.8), the first term in rhs(3.6) can be made arbitrarily small by appropriate choice (to be made later) of a and b. To treat the second term, we shall use Lemma 3.2 and concentrate on the term $\mu|\int <\theta,h_i> p_\theta d(\omega-\pi)|$ in the bound (3.1).

Expanding the function $\lambda \mapsto e^{\lambda<\theta,h>}$ in a Taylor series around

$\lambda = 0$ up to 2nd order, collecting the terms in lhs(3.9) on one side of the equality, and using Cauchy-Schwartz inequality in $\mathbb{H}$ and (1.3) to bound the other side, we get for $\lambda > 0$ and $h \in \mathcal{W}$,

$$(3.9) \qquad | < \theta, h > - \tfrac{1}{\lambda}(e^{\lambda < \theta, h >} - 1) | \leq \lambda M^2 e^{\lambda M}/2.$$

By (3.9) and the triangle inequality, with $\sigma$ abbreviating $\omega - \pi$,

$$(3.10) \qquad \mu | \int < \theta, h > p_\theta d\sigma | \leq \lambda M^2 e^{\lambda M} + \tfrac{1}{\lambda}[\| P_\omega - P_\pi \| + \mu | \int e^{\lambda < \theta, h >} p_\theta d\sigma | ].$$

We now show

$$(3.11) \qquad \mu | \int e^{\lambda < \theta, h >} p_\theta d(\omega - \pi) | = P_{\lambda h} | p_\omega - p_\pi |$$

as a consequence of

$$(3.11a) \qquad \mu( \int e^{\lambda < \theta, h >} p_\theta d\omega, \int e^{\lambda < \theta, h >} p_\theta d\pi)^{-1} = P_{\lambda h}(p_\omega, p_\pi)^{-1}.$$

By (1.1a), linearity of inner product and the map in (A), we get, $\forall$ $m \geq 1$ and $\forall$ $(\theta_1, \ldots, \theta_m) \in \Theta^m$,

$$\mu(\{\lambda < \theta_i, h > + < \theta_i, . >_0\}_{i=1}^{i=m})^{-1} = P_{\lambda h}(\{ < \theta_i, . >_0\}_{i=1}^{i=m})^{-1},$$

or equivalently,

$$\mu(\{e^{\lambda < \theta_i, h >} p_{\theta_i}\}_{i=1}^{i=m})^{-1} = P_{\lambda h}(\{p_{\theta_i}\}_{i=1}^{i=m})^{-1}.$$

Hence, if $\omega$ and $\pi$ are finitely supported, (3.11a) holds. Since by Theorem II.6.3 of Parthasarathy (1967) $\Omega$ has a dense subset consisting of finitely supported measures, to prove (3.11a) for general $\omega$ and $\pi$ it will suffice to show that for every $\nu$ in $\Omega$, as $\nu_k \to \nu$, $\int e^{\lambda < \theta, h >} p_\theta d\nu_k(\theta)$ $[p_{\nu_k}]$ goes to $\int e^{\lambda < \theta, h >} p_\theta d\nu(\theta)$ $[p_\nu]$ along a subsequence $\mu$ $[P_{\lambda h}]$ a.s.. Actually we shall show the continuity of the map taking $(\nu,\nu')$ to the $L_2(\mu)$ $[L_2(P_{\lambda h})]$ inner product of $\int e^{\lambda<\theta,h>}p_\theta d\nu(\theta)$ and $\int e^{\lambda<\theta,h>}p_\theta d\nu'(\theta)$ $[p_\nu$ and $p_{\nu'}]$. We do that by interchanging the order of integration on $\mathfrak{X}$ and $\Theta^2$, using Lemma 1.2 (with $k=2$, $a_1=a_2=1$) to evaluate the $\mu$ integral (which is continuous on $\Theta^2$, by continuities of vector addition and inner product and the exponential function, and bounded on $\Theta^2$ by (1.3)) and Lemma III.1.1 of Parthasarathy (1967). The bracket alternative is shown by representing the $L_2(P_{\lambda h})$ inner product as a $\mu$ integral, again interchanging the order of integration on $\mathfrak{X}$ and $\Theta^2$, using Lemma 1.2 (this time with $k=3$, $a_i=1$ $\forall$ i) to evaluate the integral (which is bounded continuous on $\Theta^2$ by the same reasons as above) and Lemma III.1.1 of Parthasarathy (1967) again.

Combining (3.10) and (3.11), we get

$$(3.12)\qquad \text{lhs}(3.10)\le\lambda M^2e^{\lambda M}+\tfrac{1}{\lambda}\|P_\omega-P_\pi\|+\tfrac{1}{\lambda}\mu(|p_\omega-p_\pi|p_{\lambda h}).$$

By partitioning $\mathfrak{X}$ into $[p_{\lambda h}>c]$ and $[p_{\lambda h}\le c]$, and applying Cauchy-Schwartz inequality in $L_2(\mu)$, we get

$$(3.13)\qquad \mu(|p_\omega-p_\pi|p_{\lambda h})\le c\|P_\omega-P_\pi\|+\|p_\omega-p_\pi\|_2\{\mu p_{\lambda h}{}^2[p_{\lambda h}>c]\}^{1/2}.$$

Since the family $\{p_{\lambda h}{}^2 : \lambda\in[0,K],\ h\in\mathcal{W}\}$ is uniformly $\mu$-integrable (it has uniformly bounded higher moments) for every $K>0$, $\{\mu p_{\lambda h}{}^2[p_{\lambda h}>c]\}^{1/2}$ can be made arbitrarily small, uniformly in $\lambda$ and h, by choosing c large enough.

Now choose a in (3.7) and b in (3.8) so that, uniformly in $\omega$ and $\theta$, $2Me^{M^2/2}\{(\mu[<\theta,.>_0>a])^{1/2}+(\mu[<\omega,.>_0\le b])^{1/2}\}<\gamma$. Then choose $\delta$ small enough so that $\exp(M^2/2+a-b)<\gamma/\delta$. Let I correspond to this $\delta$ as in Lemma 3.2. Now choose $\lambda$ small enough so that $\lambda M^2e^{\lambda M}<\delta/I$. Then choose c large enough so that, uniformly in $\omega$ and $\pi$ as well as in $h\in\mathcal{W}$, $(1/\lambda)\|p_\omega-p_\pi\|_2\{\mu p_{\lambda h}{}^2[p_{\lambda h}>c]\}^{1/2}\leq\delta/I$ (possible since by Lemma 1.3 and the triangle inequality in $L_2(\mu)$, $\|p_\omega-p_\pi\|_2\leq 2e^{M^2/2}$).

With these choices, by (3.12) and (3.13),

$$(3.14)\qquad \mathrm{lhs}(3.10)\leq 2\delta/I+(c+1)\|P_\omega-P_\pi\|/\lambda.$$

The proof of the proposition is now completed [with $\mathcal{K}=\{3M+\lambda^{-1}I(c+1)\}\exp(\frac{M^2}{2}+a-b)$] by (3.6), choice of a and b, use of Lemma 3.2 with the above mentioned choice of $\delta$ and substitution of the bound from (3.14) in Lemma 3.2. //

## 4. Asymptotic optimality.

In view of the bound obtained in Proposition 3.1, (2.5) and (2.6), the question of convergence of the modified regret to 0 reduces to the question, loosely speaking, whether $P_{\omega_{\alpha,n}}$ is $L_1$ consistent for $P_{G_n}$. More specifically, it suffices to show

$$(4.1)\qquad \bigvee_{\alpha=1}^{n} P_{\underline{\theta}}\left\|P_{\omega_{\alpha,n}}-P_{G_n}\right\|\to 0,\ \text{uniformly in } \underline{\theta},\ \text{as } n\to\infty.$$

In Theorem 1 in Chapter 2 we establish such a consistency result for the non-delete version for sufficiently diffuse $\Lambda$. The result involving the delete versions will follow as a corollary (i.e. Corollary 1 in Chapter 2).

Now we are in a position to prove our main result. For a finite measure $m$ on the Borel $\sigma$-field of a second countable topological space $\mathcal{Y}$, let $S_m$ denote the topological support of $m$. [For the definition of the topological support of a finite measure on a second countable topological space see Section 2 of the Appendix.]

**Theorem 4.1** (Main Result). If $S_\Lambda = \Omega$ and $\hat{\underline{t}}$ is the Bayes estimator in the set compound problem given in (2.4a), then

$$(4.2) \qquad \bigvee_{\alpha=1}^{n} P_{\underline{\theta}} \| \hat{t}_\alpha - \tilde{t}_\alpha \| \to 0, \text{ uniformly in } \underline{\theta}, \text{ as } n \to \infty.$$

Consequently, $\hat{\underline{t}}$ is a.o.

**Proof.** The second part of the assertion follows from the first part and the bound (2.5).

For the first part recall from (2.6) the representation $P_{\underline{\theta}} \| \hat{t}_\alpha - \tilde{t}_\alpha \| = P_{\underline{\theta}} P_{\theta_\alpha} \| \tau_{\omega_{\alpha,n}} - \tau_{G_n} \|$; since $\gamma$ in the statement of Proposition 3.1 is arbitrary, the assertion follows from that proposition and the $L_1$ consistency (4.1). //

Remark 4.1 (Asymptotic optimality against the equivariant envelope). As indicated in the introduction we now extend our optimality result against the simple envelope to that against the equivariant envelope. If the component problem involves a compact (in total variation norm) class of mutually absolutely continuous probability measures, then the excess of the simple envelope over the equivariant envelope goes to zero uniformly in the measures (Remark 4 in Mashayekhi (1990)). Recall that by assumption the measures $\{P_\theta : \theta \in \Theta\}$ are mutually absolutely continuous. Since $\Theta$ is topologically embedded in $\Omega$ by Lemma 2 of Chapter 2 the map $\theta \mapsto p_\theta$ is continuous in $L_4(\mu)$. That implies continuity of $\theta \mapsto P_\theta$ in total variation norm by the moment inequality. Since $\Theta$ is compact, $\{P_\theta : \theta \in \Theta\}$ is compact in the total variation norm. By triangulation around the simple envelope, the asymptotic optimality against the equivariant envelope follows from Theorem 4.1. //

<u>Remark</u> <u>4.2</u> (Asymptotic optimality of Bayes sequence compound estimators). We now prove the asymptotic optimality of the Bayes sequence compound estimator $\underline{t}'$ given in (2.4b). For $1 \le \alpha \le n < \infty$, let $\tilde{t}_{\alpha n}(\underline{x}_n) = \tau_{G_n}(x_\alpha)$, $\underline{\tilde{t}} = (\tilde{t}_{1n}, \ldots, \tilde{t}_{nn})$ and $\underline{\tilde{t}}' = (\tilde{t}_{11}, \ldots, \tilde{t}_{nn})$. Now note that (with $P_j$ and $\tau_j$ abbreviating $P_{\theta_j}$ and $\tau_{G_j}$ respectively), by the definition of $G_{k-1}$,

$$\sum_{j=1}^{k-1} P_j \|\tau_k - \theta_j\|^2 \ge \sum_{j=1}^{k-1} P_j \|\tau_{k-1} - \theta_j\|^2 \quad \forall\ k = n, n-1, \ldots, 2.$$

Applying the above iteratively with $k = n, n-1, \ldots, 2$

$$\sum_{j=1}^{n} P_j \|\tau_j - \theta_j\|^2 \le \sum_{j=1}^{n} P_j \|\tau_n - \theta_j\|^2 .$$

That is,

$$R_n(\underline{\tilde{t}}', \underline{\theta}) \le R_n(\underline{\tilde{t}}, \underline{\theta}) \qquad \forall\ n \ge 1,$$

which implies

$$D_n(\underline{t}', \underline{\theta}) \le R_n(\underline{t}', \underline{\theta}) - R_n(\underline{\tilde{t}}', \underline{\theta}). \tag{4.3}$$

It should be noted that the display immediately preceding (4.3) is essentially inequality (8.8) of Hannan (1957).

From the definition of $R_n$, $\underline{t}'$ and $\underline{\tilde{t}}'$, following the steps involved in showing (2.5),

$$|\mathrm{rhs}(4.3)| \le 4Mn^{-1} \sum_{\alpha=1}^{n} P_{\underline{\theta}} \| t'_\alpha - \tilde{t}_{\alpha\alpha} \|. \tag{4.4}$$

From (2.4b) and the definition of $\tilde{t}_{\alpha\alpha}$, using an analog of (2.6), Proposition 3.1 and (4.1), it follows that

$$\bigvee_{\underline{\theta}} P_{\underline{\theta}} \| t'_n - \tilde{t}_{nn} \| \to 0 \text{ as } n \to \infty.$$

Using subadditivity of supremum and the fact that the limit of a convergent sequence equals its Ce'saro limit, we get that rhs(4.4)$\to$0 uniformly in $\underline{\theta}$. If we can show that $\bigvee_{\underline{\theta}} D_n(\underline{t}', \underline{\theta})$ is positive, the asymptotic optimality of $\underline{t}'$ will follow by (4.3) and convergence (uniform in $\underline{\theta}$) of rhs(4.4) to 0.

We shall show that $\bigvee_{\underline{\theta}} D_n(\underline{t}, \underline{\theta})$ is positive for every compound procedure $\underline{t}$. Since $\int_{\Theta^n} R_n(\underline{t}, \underline{\theta}) d\omega^n \geq r(\omega)$ for every $\omega$, in particular for $G_n$, we get that $\bigvee_{\underline{\theta}} R_n(\underline{t}, \underline{\theta}) \geq \bigvee_{\underline{\theta}} r(G_n)$. That, by definition of $\bigvee_{\underline{\theta}} D_n(\underline{t}, \underline{\theta})$ and subadditivity of supremum, implies the positivity of $\bigvee_{\underline{\theta}} D_n(\underline{t}, \underline{\theta})$.//

<u>Remark</u> <u>4.3</u> (Calculation of the a.o. Bayes compound estimator). From (2.4a), Lemma 1.1 and the definition of $\omega_{\alpha,n}$, it follows by a successive deconditioning argument that

$$(4.5) \qquad <\hat{t}_\alpha(\underline{x}), h> = \frac{[\int \cdots \int <\theta_\alpha, h> \prod_{i=1}^{n} p_{\theta_i}(x_i) \prod_{i=1}^{n} d\omega^{\underline{\theta}_{i-1}}(\theta_i)]}{[\int \cdots \int \prod_{i=1}^{n} p_{\theta_i}(x_i) \prod_{i=1}^{n} d\omega^{\underline{\theta}_{i-1}}(\theta_i)]},$$

where $\omega^{\underline{\theta}_i}$ is the posterior mean of $\omega$ given $\underline{\theta}_i$ and $\omega^{\underline{\theta}_0} = \int \omega d\Lambda$; for details see Section 3 in Chapter 4 of Datta (1988).

To use (4.5) to calculate our Bayes compound estimator, we need to choose a hyperprior $\Lambda$ such that the posterior mean $\omega^{\underline{\theta}_i}$ has a nice form for all i. With that end in mind, we settle for the Dirichlet priors described below.

Let $\alpha$ be a non-null finite Borel measure on $\Theta$, where $\Theta$ is an arbitrary separable metric space. In Section 2 of the Appendix we show (compiling some results from Section 4 in Ferguson (1973)) that there exists a probability measure $\mathfrak{D}(\alpha)$ on $(\Omega,\mathfrak{B}(\Omega))$ with the following property : for every finite measurable partition $\{B_1,\ldots B_m\}$ of $\Theta$, the distribution of $(\omega(B_1),\ldots,\omega(B_m))$ under $\mathfrak{D}(\alpha)$ is Dirichlet with parameters $(\alpha(B_1),\ldots,\alpha(B_m))$. We call $\mathfrak{D}(\alpha)$ the Dirichlet prior with parameter $\alpha$. By Theorem 2.1 of the Appendix, the topological support of $\mathfrak{D}(\alpha)$ is $\Omega$ if that of $\alpha$ is $\Theta$. An example of a finite Borel measure $\alpha$ on $\Theta$ with full support is obtained by choosing a countable dense subset $\{\theta_n : n \geq 1\}$ of $\Theta$ and selecting $\alpha = \sum_{n=1}^{\infty} c_n \delta_{\theta_n}$, where $c_n \geq 0\ \forall\ n$ and $\sum_{n=1}^{\infty} c_n \in (0,\infty)$. By Theorem 1 in Ferguson (1973),

$$\omega^{\underline{\theta}_n} = (\alpha(\Theta)+n)^{-1}(\alpha + \sum_{i=1}^{n} \delta_{\theta_i}), \qquad n \geq 0.$$

When $\Theta$ is a subset of the line, a Monte Carlo method for calculation of rhs(4.5) has been given by Kuo (1986). The problem of numerical evaluation of our estimator remains and is worth investigating. //

## 5. Admissibility.

The argument we use to prove admissibility of Bayes compound estimators is fairly standard in decision theory : A unique Bayes rule is admissible (see Theorem 1 in Section 2.3 of Ferguson (1967) for a precise statement).

Let $\xi$ be a prior on the compound parameter $\underline{\theta}$. $\mathbb{Q}$ will denote the joint distribution $\xi \circ \mathbf{P}_{\underline{\theta}}$ on $(\underline{x},\underline{\theta})$. Note that $n^{-1}\sum_{\alpha=1}^{n} \mathbb{Q}\| t_\alpha - \theta_\alpha \|^2$, the

Bayes (versus $\xi$) compound risk of an estimator $\underline{t}$, is minimal iff $Q\|t_\alpha - \theta_\alpha\|^2$ is minimal for every $\alpha$. Now $Q\|t_\alpha - \theta_\alpha\|^2$ can be represented as $\int \mathbf{P}_{\underline{\theta}}(\int\int\|t_\alpha - \theta_\alpha\|^2 dP_{\theta_\alpha} d\xi_\alpha)d\xi$, where $\xi_\alpha = \xi\theta_\alpha^{-1}$. Since the expression inside parenthesis in the previous line has, by Lemma 1.1, a unique minimizer, there exists a unique Bayes compound estimator versus every prior $\xi$. That implies the admissibility of every Bayes compound estimator.

# CHAPTER 2
# CONSISTENCY OF THE POSTERIOR MIXTURES

In this chapter we show [Theorem 1] that $P_{\omega_{n,n}}$ (the non-delete version of the discussion at the beginning of Section 4 in Chapter 1) is $L_1$ consistent for $P_{G_{n-1}}$ in the sense of (4.1) of Chapter 1. We actually prove the result with n replaced by $(n+1)$ and obtain (4.1) of Chapter 1 as a corollary [Corollary 1]. For the rest of the chapter let $\hat{\omega}$ and $\hat{\Lambda}$ abbreviate $\omega_{n+1,n+1}$ and $\Lambda_{n+1,n+1}$ respectively. Before proceeding further we note that $\hat{\omega}$ can be interpreted as the posterior distribution of $\theta_{n+1}$ given $(X_1,\ldots,X_n) = (x_1,\ldots,x_n)$ in the Bayes compound model with $(n+1)$ components.

Consider the following Bayes model on $\Omega \times \Theta^n \times \mathfrak{X}^n$:

(i) Bayes model: $\omega$ is distributed as $\Lambda$ and given $\omega$, $\underline{\theta}$ is distributed as $\omega^n = \mathop{\times}\limits_{\alpha=1}^{n} \omega$ and given $\theta$ and $\omega$, $\underline{X}$ is distributed as $\mathbf{P}_{\underline{\theta}} = \mathop{\times}\limits_{\alpha=1}^{n} P_{\theta_\alpha}$.

The above model gives rise to the following marginal model:

(ii) Bayes compound model: $\underline{\theta} = (\theta_1,\ldots,\theta_n)$ is distributed as $\bar{\omega}_{\Lambda,n}$ and given $\underline{\theta}$, $\underline{X} = (X_1,\ldots,X_n)$ is distributed as $\mathbf{P}_{\underline{\theta}}$, where $\bar{\omega}_{\Lambda,n}$ is the $\Lambda$ mixture of $\omega^n$.

Since $\Theta$ and hence $\Omega$ (with the weak convergence topology) is a Polish (in fact compact metric) space, all conditional distributions are regular by Theorem 10.2.2 of Dudley (1989). Datta (1991a) shows [see his Proposition 2.1] that under model (ii), with n replaced by $n+1$, $\hat{\omega}$ is the posterior distribution of $\theta_{n+1}$ given $(X_1,\ldots,X_n) = (x_1,\ldots,x_n)$.

We now develop the machinery needed to prove Theorem 1. There are four propositions leading to the proof of Theorem 1. Four auxiliary lemmas are needed to prove the propositions.

The key to the proof of Theorem 1 is the inequality (17) proved in Proposition 3. The force of Proposition 1 is used in part in the proof of Proposition 3 and later in full in the proof of Theorem 1 to treat the denominator of the second term in the bound (17); it is the only link in the proof where the assumption $S_\Lambda = \Omega$ is used. Proposition 2 is used to treat the numerator of the second term in the bound (17). Proposition 4 disposes of the third term in the bound (17).

**Lemma 1.** For every $\{\omega,\pi\} \subset \Omega$, $\log(p_\omega/p_\pi) \in L_2(\mu)$ and

$$\|\log(p_\omega/p_\pi)\|_2 \leq e^{5M^2/2}\|p_\omega - p_\pi\|_4.$$

**Proof.** Since the reciprocal function is convex on $(0,\infty)$, the area under the reciprocal curve between a and b, where $0 < a \leq b < \infty$, is smaller than the area under the straightline joining the points $(a,a^{-1})$ to $(b,b^{-1})$, which is equal to $(b-a)(a^{-1}+b^{-1})/2$. That gives

$$|\log(p_\omega/p_\pi)| \leq |p_\omega - p_\pi|(p_\omega{}^{-1} + p_\pi{}^{-1})/2 \quad \text{a.s. } (\mu),$$

which implies, via Cauchy-Schwartz inequality,

$$2\|\log(p_\omega/p_\pi)\|_2 \leq \|p_\omega - p_\pi\|_4[\mu(p_\omega{}^{-1} + p_\pi{}^{-1})^4]^{1/4}. \tag{1}$$

Applying Jensen's inequality to the function $x \mapsto x^{-j}$, which is convex on $(0,\infty)$ $\forall$ $j \geq 1$ and (trivially) for $j = 0$,

(2) $$p_\omega^{-j} \le \int p_\theta^{-j} d\omega \ \forall\ \omega \in \Omega \text{ and } \forall\ j \text{ described above.}$$

Applying (2) with $j = i$ on $p_\omega$ and $j = 4 - i$ on $p_\pi$, interchanging the order of integration on $\Theta^2$ and $\mathfrak{S}$, and using Lemma 1.2 of Chapter 1 (with $k = 2$, $a_1 = -i$, $a_2 = i - 4$), we get

(3) $$\mu(p_\omega^{-i} p_\pi^{i-4}) = \exp\{2^{-1}[\| -i\theta + (i-4)\eta \|^2 + i\|\theta\|^2 + (4-i)\|\eta\|^2]\}.$$

The exponent in rhs(3) simplifies to

$$2^{-1}[(i^2 + i)\|\theta\|^2 + (i^2 - 9i + 20)\|\eta\|^2 + 2i < \theta,\eta > (4-i)].$$

For all $i = 0,1,\ldots,4$, the coefficients of $\|\theta\|^2$, $\|\eta\|^2$ and $<\theta,\eta>$ are all non-negative and hence, by (1.3) of Chapter 1, the exponent in rhs(3) is bounded by $10M^2$. Since $\sum_{i=1}^{4} \binom{4}{i} = 2^4$, using (3) with $10M^2$ bounding the exponent, we get

(4) $$\text{second factor in rhs}(1) \le 2e^{5M^2/2}.$$

The lemma follows from (1) and (4). //

In what follows, any reference to a topology of $\Omega$ will be to the topology of weak convergence.

**Lemma 2.** $\omega \mapsto p_\omega$ is uniformly continuous from $\Omega$ into $L_4(\mu)$.

**Proof.** For $j = 0,1,2,3,4$, writing $p_\omega^j$ (and $p_\pi^{4-j}$) as a $j$ (and $4-j$) fold iterated integral, interchanging the order of integration on $\mathfrak{S}$ and $\Theta^4$ and using Lemma 1.2 of Chapter 1 (with $k = 4$, $a_1 = \ldots = a_4 = 1$),

(5) $$\int p_\omega^j\ p_\pi^{4-j} d\mu = \omega^j \times \pi^{4-j}(\exp[2^{-1}\{\| \sum_{i=1}^{4} \theta_i \|^2 - \sum_{i=1}^{4} \|\theta_i\|^2\}]).$$

By repeated application of Lemma III.1.1 of Parthasarathy (1967), if $\omega_n \to \omega$ then $\omega_n{}^j \times \omega^{4-j} \to \omega^4$ weakly on $\Theta^4$. Since the integrand in rhs(5) is a bounded continuous function on $\Theta^4$, using (5) twice we get

$$\int p_{\omega_n}{}^j p_\omega{}^{4-j} d\mu \to \int p_\omega{}^4 d\mu \text{ as } \omega_n \to \omega.$$

Expanding $(p_{\omega_n} - p_\omega)^4$ and applying the above to the integral of each term,

$$\int (p_{\omega_n} - p_\omega)^4 d\mu \to 0 \text{ as } \omega_n \to \omega;$$

that establishes the asserted uniform continuity because the weak topology of $\Omega$ is metrizable as a compact metric space. //

Let

(I) $$\Delta_\pi(\omega) := \int \log(p_\pi/p_\omega) dP_\pi.$$

**Lemma 3.** $\pi \mapsto \Delta_\pi(\omega)$ is equi(in $\omega$) continuous.

**Proof.** For $\pi$ and $\nu$ in $\Omega$, triangulating around $\int \log(p_\pi/p_\omega) dP_\nu$,

(6) $$|\Delta_\pi(\omega) - \Delta_\nu(\omega)| \leq |\int \log(p_\pi/p_\omega) d(P_\pi - P_\nu)| + |\Delta_\nu(\pi)|.$$

By Cauchy-Schwartz inequality in $L_2(\mu)$,

(7) $$\text{1st term in rhs(6)} \leq \| \log(p_\pi/p_\omega) \|_2 \, \| p_\pi - p_\nu \|_2;$$

by Lemma 1, the triangle inequality in $L_4(\mu)$ and Lemma 1.3 of Chapter 1 with $q = 4$,

(8) $$\text{rhs(7)} \leq 2e^{4M^2} \| p_\pi - p_\nu \|_2.$$

By Cauchy-Schwartz inequality in $L_2(\mu)$ again,

(9) $$\text{2nd term in rhs(6)} \leq \| \log(p_\nu/p_\pi) \|_2 \, \| p_\nu \|_2 \;;$$

by Lemma 1.3 of Chapter 1 with $q = 2$ and Lemma 1,

$$\text{rhs}(9) \leq e^{3M^2} \| p_\pi - p_\nu \|_4. \tag{10}$$

Since $\| p_\pi - p_\nu \|_2 \leq \| p_\pi - p_\nu \|_4$ the proof is completed by combining (6)-(10) and applying Lemma 2. //

**Lemma 4.** $\omega \mapsto \Delta_\pi(\omega)$ is equi(in $\pi$) continuous.

**Proof.** For $\omega$ and $\nu$ in $\Omega$, by Cauchy-Schwartz inequality in $L_2(\mu)$,

$$|\Delta_\pi(\omega) - \Delta_\pi(\nu)| \leq \| \log(p_\nu/p_\omega) \|_2 \| p_\pi \|_2$$

$$\leq e^{3M^2} \| p_\omega - p_\nu \|_4, \text{ by Lemma 1.3 of Chapter 1 with } q = 2 \text{ and Lemma 1;}$$

the lemma follows by Lemma 2. //

**Proposition 1.** If $S_\Lambda = \Omega$, then

$$\bigwedge_\pi \Lambda\{\Delta_\pi < \delta\} > 0 \;\; \forall \; \delta > 0.$$

**Proof.** [Taken from Lemma 6.6 of Datta (1991a)] By Lemma 4 $\{\Delta_\pi < \delta\}$ is open; since it is non-empty (it contains $\pi$) and $S_\Lambda = \Omega$, $\Lambda\{\Delta_\pi < \delta\} > 0$. By Lemma 3, if $\pi_n \to \pi$ then $\Delta_{\pi_n} \to \Delta_\pi$ pointwise on $\Omega$, hence in $\Lambda$-distribution. Therefore, by Theorem II.6.1(d) of Parthasarathy (1967),

$$\liminf \Lambda\{\Delta_{\pi_n} < \delta\} \geq \Lambda\{\Delta_\pi < \delta\} \text{ as } \pi_n \to \pi;$$

in other words, $\pi \mapsto \Lambda\{\Delta_\pi < \delta\}$ is lower semi-continuous. Hence the infimum is attained over compact $\Omega$ and is positive. //

Let

$$\Upsilon_n(\omega) := n^{-1} \sum_{\alpha=1}^{n} \log p_\omega(x_\alpha) - \int \log p_\omega dP_{G_n}. \tag{II}$$

**Proposition 2.** Let $\rho$ be a metric on $\Omega$ for the topology of weak convergence. For every $\delta > 0$, $\exists$ an $\epsilon > 0$, such that $\rho(\omega,\pi) < \epsilon$ implies

$$(11) \qquad \mathbf{P}_{\underline{\theta}}(\exp\{2n[\mathcal{V}_n(\omega) - \mathcal{V}_n(\pi)]\}) \le e^{n\delta}.$$

**Proof.** Using Cauchy-Schwartz inequality in $L_2(\mu)$, Lemma 1.3 of Chapter 1 with $q = 2$ to bound $\|p_{\theta_\alpha}\|^2$ and Lemma 1,

$$(12) \qquad |\int \log(p_\omega/p_\pi)dP_{\theta_\alpha}| \le e^{3M^2}\|p_\omega - p_\pi\|_4.$$

Since

$$2n[\mathcal{V}_n(\omega) - \mathcal{V}_n(\pi)] = 2\sum_{\alpha=1}^{n} \log(p_\omega/p_\pi)(x_\alpha) - 2\sum_{\alpha=1}^{n} \int \log(p_\omega/p_\pi)dP_{\theta_\alpha},$$

by isotonicity of the exponential function and the bound in (12),

$$(13) \qquad \mathrm{lhs}(11) \le [\mathbf{P}_{\underline{\theta}}(\prod_{\alpha=1}^{n}(p_\omega/p_\pi)^2(x_\alpha))]\exp\{2n\|p_\omega - p_\pi\|_4 e^{3M^2}\}.$$

We shall now show that

$$(14) \qquad \mathbf{P}_{\underline{\theta}}(\prod_{\alpha=1}^{n}(p_\omega/p_\pi)^2(x_\alpha)) \le \exp\{2n\|p_\omega - p_\pi\|_2 e^{16M^2}\}.$$

Since the $L_2$ norm of a random variable is less than or equal to its $L_4$ norm, (13) and (14), in view of Lemma 2, will complete the proof of the proposition.

Using independence of the factors in the integrand under $\mathbf{P}_{\underline{\theta}}$,

$$\mathrm{lhs}(14) = \prod_{\alpha=1}^{n} P_{\theta_\alpha}(p_\omega/p_\pi)^2,$$

which, by bounding each of the factors using the inequality $v \le e^{v-1}$, is bounded by

$$(15) \qquad \exp\{\sum_{\alpha=1}^{n} P_{\theta_\alpha}[(p_\omega{}^2/p_\pi{}^2) - 1]\}.$$

Converting the integrand into an expression with common denominator, factoring the numerator and applying Cauchy-Schwartz inequality in $L_2(\mu)$,

(16) $$P_{\theta_\alpha}[(p_\omega^2/p_\pi^2)-1] \le \|p_\omega - p_\pi\|_2 \left\| P_{\theta_\alpha} p_{\omega+\pi}/p_\pi^2 \right\|_2 .$$

It now suffices to show that the square of the second factor in rhs(16) can be bounded by $4e^{32M^2}$. We do that by considering it as a $\mu$ integral, using the bound on the inverse fourth power of a mixed density obtained in (2), writing $(p_{\omega+\pi})^2$ as an iterated integral, interchanging the order of integration on $\mathfrak{X}$ and $\Theta^3$, applying Lemma 1.2 of Chapter 1 (with $k=4$, $a_1=2$, $a_2=a_3=1$, $a_4=-4$), simplifying the resulting exponent by expanding the squared norm of the sum of four terms and making possible cancellations, and using Cauchy-Schwartz inequality in H and (1.3) of Chapter 1 to bound the remaining terms in the exponent, in that order. //

The basic idea in the following proposition can be traced back to (iii)′ of the Addendum of Gilliland, Hannan and Huang (1976). For a similar exploitation of that basic idea see Lemma 6.1 of Datta (1991a).

Let

(III) $$\mathfrak{U}_\delta := \{\omega : \Delta_{G_n}(\omega) < \delta\}.$$

**Proposition 3.** Let $\rho$ be as in Proposition 2. Fix a $\delta > 0$. Let

$$A_\delta = \bigcap_{i=1}^{r} \{|\mathfrak{r}_n(\omega_i)| < \delta/2\},$$

where $\rho$-balls of radius $\epsilon$ (corresponding to $\delta$ as in Proposition 2) around $\{\omega_1,\ldots,\omega_r\}$ cover $\Omega$. Then

(17) $$\tfrac{1}{2}\|P_{\hat\omega} - P_{G_n}\| \le 2\sqrt{\delta} + [\,\frac{e^{-3n\delta}}{\Lambda^2(\mathfrak{U}_\delta)} \int e^{-n\mathfrak{r}_n} d\Lambda \int e^{n\mathfrak{r}_n} d\Lambda\,]A_\delta + \tilde{A}_\delta .$$

**Proof.** Since $\frac{1}{2}\|P_\omega - P_\pi\| \le 1$ for all $\{\omega,\pi\} \subset \Omega$

$$\text{lhs}(17) \le \tfrac{1}{2}\left\|P_{\hat\omega} - P_{G_n}\right\|A_\delta + \tilde A_\delta. \tag{18}$$

By definition of $\hat\omega$, using the inequality $|\int f| \le \int |f|$ with $f = p_\omega - p_{G_n}$ and interchanging the order of $\mu$ and $\hat\Lambda$ integration, we get

$$\left\|P_{\hat\omega} - P_{G_n}\right\| \le \int \left\|P_\omega - P_{G_n}\right\| d\hat\Lambda. \tag{19}$$

Since by inequality (3.6) of Hannan (1960)

$$\tfrac{1}{2}\|P_\omega - P_\pi\| \le \sqrt{\Delta_\pi(\omega)},$$

bounding the integrand in rhs(19) by $4\sqrt{\delta}$ on $\mathfrak{U}_{4\delta}$ and by 2 on the complement, and $\hat\Lambda(\mathfrak{U}_{4\delta})A_\delta$ by 1,

$$\tfrac{1}{2}\left\|P_{\hat\omega} - P_{G_n}\right\|A_\delta \le 2\sqrt{\delta} + \hat\Lambda(\tilde{\mathfrak{U}}_{4\delta})A_\delta. \tag{20}$$

In view of (18) and (20), it remains to show that the second term in rhs(20) can be bounded by the second term in rhs(17).

By definition of $\hat\Lambda$ [it is the posterior distribution of $\omega$ given $(X_1,\ldots,X_n) = (x_1,\ldots,x_n)$, when given $\omega$, $X_1,\ldots,X_n$ are iid $\sim P_\omega$ and $\omega \sim \Lambda$],

$$\hat\Lambda(\tilde{\mathfrak{U}}_{4\delta}) \le \frac{\int_{\tilde{\mathfrak{U}}_{4\delta}} e^g \, d\Lambda}{\int_{\mathfrak{U}_\delta} e^g \, d\Lambda}, \tag{21}$$

where $g(\omega) = \sum_{\alpha=1}^{n} \log\, p_\omega(x_\alpha)$.

Using the identity

$$g(\omega) = n\mathcal{V}_n(\omega) - n\Delta_{G_n}(\omega) + n\int \log\, p_{G_n} dP_{G_n},$$

bounding $\Delta_{G_n}$ below by $4\delta$ on $\tilde{\mathfrak{U}}_{4\delta}$ in the numerator of rhs(21) and above by $\delta$ on $\mathfrak{U}_\delta$ in the denominator of rhs(21), we get

$$\text{rhs}(21) \le e^{-3n\delta} \frac{\int_{\tilde{\mathfrak{U}}_{4\delta}} e^{n\Upsilon_n} d\Lambda}{\int_{\mathfrak{U}_\delta} e^{n\Upsilon_n} d\Lambda} . \tag{22}$$

Normalizing $\Lambda$ on $\mathfrak{U}_\delta$ ($\Lambda(\mathfrak{U}_\delta) > 0$ by Proposition 1) and applying Jensen's inequality to the reciprocal function, which is convex on $(0,\infty)$,

$$\frac{1}{\int_{\mathfrak{U}_\delta} e^{n\Upsilon_n} d\Lambda} \le \frac{1}{\Lambda^2(\mathfrak{U}_\delta)} \int_{\mathfrak{U}_\delta} e^{-n\Upsilon_n} d\Lambda. \tag{23}$$

Substituting (23) in rhs(22) and weakening the resulting bound by enlarging the ranges of integration, the proposition follows from the remark following (20). //

**Proposition 4.** Fix a $\delta > 0$. Let $A_\delta$ be as in Proposition 3. Then

$$\bigvee_{\underline{\theta}} \mathbf{P}_{\underline{\theta}} \tilde{A}_\delta = O(n^{-1}) \text{ as } n \to \infty.$$

**Proof.** By the definition of $A_\delta$ and subadditivity of measures,

$$\mathbf{P}_{\underline{\theta}} \tilde{A}_\delta \le \sum_{i=1}^{r} \mathbf{P}_{\underline{\theta}}[\,|\Upsilon_n(\omega_i)| \ge \delta/2]$$

which, by applying Chebychev's inequality to each of the terms and bounding the sum of r nonnegative terms by the maximum of the terms times r, is bounded by

$$(4r/\delta^2) \bigvee_{\omega} \mathbf{P}_{\underline{\theta}}(\Upsilon_n(\omega))^2.$$

Since $\Upsilon_n(\omega)$ is the centered average of n independent random variables under $\mathbf{P}_{\underline{\theta}}$,

$$P_{\underline{\theta}}(\Upsilon_n(\omega))^2 \le n^{-1} \bigvee_{\theta} \mathrm{var}_\theta(\log p_\omega).$$

Since the variance is smaller than the second moment, it suffices to show that $\bigvee_{\omega} \bigvee_{\theta} P_\theta(\log p_\omega)^2 < \infty$.

By the elementary log inequality used in the proof of Lemma 1,

(24) $$4P_\theta(\log p_\omega)^2 \le P_\theta(p_\omega - 1)^2(1 + p_\omega{}^{-1})^2.$$

By Cauchy-Schwartz inequality in $L_2(\mu)$,

(25) $$\mathrm{rhs}(24) \le \| p_\theta(1 + p_\omega{}^{-1})^2 \|_2 \| p_\omega - 1 \|_4{}^2.$$

By the triangle inequality in $L_4(\mu)$ and Lemma 1.3 of Chapter 1 with $q = 4$,

(26) $$\text{2nd factor in rhs}(25) \le (e^{3M^2/2} + 1)^2 .$$

By Cauchy-Schwartz inequality in $L_2(\mu)$ and Lemma 1.3 of Chapter 1 with $q = 4$,

(27) $$\text{1st factor in rhs}(25) \le e^{3M^2/2} \| 1 + p_\omega{}^{-1} \|_8{}^2.$$

By the triangle inequality in $L_8(\mu)$,

(28) $$\text{2nd factor in rhs}(27) \le 2(1 + \| p_\omega{}^{-1} \|_8{}^2).$$

Applying the bound on the inverse eighth power of a mixed density obtained in (2), interchanging the order of $\mu$ and $\omega$ integration, using Lemma 1.2 of Chapter 1 (with $k = 1$, $a_1 = -8$), and using (1.3) of Chapter 1 to bound the resulting exponent, we get that $\| p_\omega{}^{-1} \|_8{}^2$ is bounded by $e^{18M^2}$; substituting that bound in rhs(28) and combining the result with (27) and (26), we complete the proof using

(25) and the remark preceding (24). //

**Theorem** 1 [Consistency of the posterior mixtures]. If $S_\Lambda = \Omega$, then

$$\mathrm{P}_{\underline{\theta}}\|\mathrm{P}_{\hat{\omega}} - \mathrm{P}_{\mathrm{G}_n}\| \to 0, \text{ uniformly in } \underline{\theta}, \text{ as } n \to \infty.$$

**Proof.** We shall show that for every $\delta > 0$,

$$\mathrm{P}_{\underline{\theta}}\left(\int e^{-n\Upsilon_n} d\Lambda \int e^{n\Upsilon_n} d\Lambda\right) A_\delta \le e^{2n\delta}, \tag{29}$$

where $A_\delta$ is as in Proposition 3.

Taking $\mathrm{P}_{\underline{\theta}}$ expectation of both sides of (17) and using (29),

$$\mathrm{P}_{\underline{\theta}}\|\mathrm{P}_{\hat{\omega}} - \mathrm{P}_{\mathrm{G}_n}\| \le 4\sqrt{\delta} + 2\,\frac{e^{-n\delta}}{\Lambda^2(\mathcal{U}_\delta)} + 2\mathrm{P}_{\underline{\theta}}\tilde{A}_\delta. \tag{30}$$

Since the above inequality holds for every $\delta > 0$, we complete the proof of Theorem 1 by taking supremum over $\underline{\theta}$ and lim sup as $n \to \infty$ (in that order) of both sides of (30), using subadditivity of these operations, and applying Propositions 1 and 4.

Applying Cauchy-Schwartz inequality in $L_2(\mathrm{P}_{\underline{\theta}})$ and then the moment inequality to the $\Lambda$ integrals in both the factors in the Cauchy-Schwartz bound,

$$\mathrm{lhs}(29) \le \left[\mathrm{P}_{\underline{\theta}}\left(\int e^{-2n\Upsilon_n} d\Lambda\right) A_\delta\right]^{1/2} \left[\mathrm{P}_{\underline{\theta}}\left(\int e^{2n\Upsilon_n} d\Lambda\right) A_\delta\right]^{1/2}. \tag{31}$$

Consider the finite cover of $\Omega$ described in Proposition 3. Clearly, for every $\omega \in \Omega$, choosing an $\omega_i$ such that $\rho(\omega, \omega_i) < \epsilon$ and using Proposition 2 with $\omega$ and $\omega_i$, by the definition of $A_\delta$,

$$\mathrm{P}_{\underline{\theta}}\left(e^{-2n\Upsilon_n(\omega)} A_\delta\right) \le e^{2n\delta} \quad \text{and} \quad \mathrm{P}_{\underline{\theta}}\left(e^{2n\Upsilon_n(\omega)} A_\delta\right) \le e^{2n\delta}. \tag{32}$$

Interchanging the order of $\Lambda$ and $\mathrm{P}_{\underline{\theta}}$ integration in rhs(31) and using

(32), we establish (29). //

Remark 1 (Comparison of Theorem 1 and Theorem 3.1 of Datta (1991a)). In his Theorem 3.1 Datta (1991a) proves the assertion of Theorem 1 for compact metric $\Theta$ and a class of probability densities $\{p_\theta : \theta \in \Theta\}$ on an arbitrary measurable space, under two regularity assumptions including the one (A1) that $p_\theta(x)$ is continuous in $\theta$ for every x. Unless $\Theta$ is what Dudley (1967) called a GC set (GC is an abbreviation for *Gaussian Continuity*; a subset of a Hilbert space is defined to be a GC set if the isonormal process indexed by that subset has a sample continuous version) (A1) is not satisfied for the Gaussian shift experiment. For an example (among the ellipsoids) of a compact $\Theta$ which is not GC, see the introduction to Section 6 and Proposition 6.3 of Dudley (1967).

Datta (1991a) obtains a bound on $\frac{1}{2}\|P_{\hat{\omega}} - P_{G_n}\|$ in his Lemma 6.1 similar to our bound in Proposition 3. His bound consists of an arbitrarily small term, a term involving a measure of diffuseness of $\Lambda$ (which we dispose of by Proposition 1, which is essentially his Lemma 6.6) and another term involving the probability of the tail of $\bigvee_{\omega} |\Upsilon_n(\omega)|$. By his assumption (A1), the quantity $\bigvee_{\omega} |\Upsilon_n(\omega)|$ is the Banach norm of a $\mathcal{C}(\Omega)$ valued random element, where $\mathcal{C}(S)$ is the Banach space (with sup norm) of all real-valued continuous functions on compact metric S. He develops an uniform $L_1$ law of large numbers for $\mathcal{C}(S)$-valued random elements to dispose of that term. In our context $\Upsilon_n$ is not a $\mathcal{C}(\Omega)$ valued random element, making Datta's method of proof inapplicable. //

Corollary 1. If $S_\Lambda = \Omega$, then (4.1) of Chapter 1 holds; i.e.,

$$\bigvee_{\alpha=1}^{n} \mathbf{P}_{\underline{\theta}} \left\| P_{\omega_{\alpha,n}} - P_{G_n} \right\| \to 0, \text{ uniformly in } \underline{\theta}, \text{ as } n \to \infty.$$

Proof. As in the proof of Lemma 4.3 of Datta (1991b), we observe that

$$(33) \qquad \bigvee_{\underline{\theta} \in \Theta^n} \bigvee_{\alpha \le n} \mathbf{P}_{\underline{\theta}} \left\| P_{\omega_{\alpha,n}} - P_{G_{n\alpha}} \right\| \le \bigvee_{\underline{\theta} \in \Theta^{n-1}} \mathbf{P}_{\underline{\theta}} \left\| P_{\omega_{n,n}} - P_{G_{n-1}} \right\|,$$

where $G_{n\alpha}$ is the empirical distribution based on $(\theta_1, \ldots, \theta_{\alpha-1}, \theta_{\alpha+1}, \ldots \theta_n)$. Since $G_n - G_{n\alpha} = n^{-1}(\delta_{\theta_\alpha} - G_{n\alpha})$ with $\delta_{\theta_\alpha}$ the unit mass at $\theta_\alpha$, the definition of $p_\omega$ gives

$$(34) \qquad \left\| P_{G_n} - P_{G_{n\alpha}} \right\| = \mu(|p_{G_n} - p_{G_{n\alpha}}|) = n^{-1} \mu(|p_{\theta_\alpha} - p_{G_{n\alpha}}|) \le 2n^{-1}.$$

By the triangle inequality, (33) and (34), the corollary follows if rhs(33) goes to 0. But that, with n replacing $n-1$, is the assertion (with some notational changes) of Theorem 1. //

# CHAPTER 3
# THE EMPIRICAL BAYES ESTIMATION

In this chapter we look at the empirical Bayes [Robbins (1951, 1956)] formulation of our component problem. Consider a Bayes decision problem involving $\{P_\theta : \theta \in \Theta\}$ and a Bayes prior $\omega$, where $\omega$ is unknown. Suppose we have iid pairs $(\theta_1, X_1), \ldots, (\theta_n, X_n), \ldots$, where $\theta_1$ is distributed as $\omega$ and given $\theta_1$, $X_1$ is distributed as $P_{\theta_1}$. At stage n, a decision $t_n = t_n(\underline{X}_n)$ about $\theta_n$ is taken incurring loss $\| t_n - \theta_n \|^2$ and risk $\int \int \| t_n - \theta_n \|^2 dP_{\underline{\theta}} d\omega^n$. The sequence $\{t_n : n \geq 1\}$ is called an empirical Bayes rule. An empirical Bayes rule $\{t_n\}$ is called *asymptotically optimal* (a.o.) if

$$\int \int \| t_n - \theta_n \|^2 dP_{\underline{\theta}} d\omega^n \to r(\omega), \text{ for each } \omega \in \Omega, \text{ as } n \to \infty.$$

The notion of *admissibility* in the class of empirical Bayes rules is the same as the corresponding notion in the case of compound rules, with the understanding that the risk now is a function of $\omega$.

Let $\Lambda$ be a hyperprior on $\Omega$. We will prove that any sequence of Bayes (versus $\Lambda$ at each stage) empirical Bayes rules is admissible; if $S_\Lambda = \Omega$, a sequence of Bayes empirical Bayes rules versus $\Lambda$ is asymptotically optimal as well.

## 1. Bayes empirical Bayes.

For any given n, the stage n Bayes risk versus $\Lambda$ in the empirical Bayes problem is

$$\int_{\Omega}\int_{\Theta^n}\int_{\mathfrak{X}^n} \| t_n - \theta_n \|^2 dP_{\underline{\theta}} d\omega^n d\Lambda(\omega) = \int_{\Theta^n}\int_{\mathfrak{X}^n} \| t_n - \theta_n \|^2 dP_{\underline{\theta}} d\bar{\omega}_{\Lambda,n},$$

which is the n-th component Bayes risk versus the prior $\bar{\omega}_{\Lambda,n}$ on the compound parameter $\underline{\theta}$ in the set compound problem with n components. Hence a Bayes empirical Bayes estimator is $\hat{t}_n$ given in (2.4a), with $\alpha$ replaced by n.

Admissibility. Since a Bayes empirical Bayes estimator is $\hat{t}_n$ given in (2.4a), and as observed in Section 2.5 of Chapter 1 every Bayes compound estimator is unique up to $\mu^n$ equivalence, the admissibility again follows from the uniqueness of Bayes rule argument.

## 2. Asymptotic optimality.

**Theorem** 2.1. If $S_\Lambda = \Omega$ then the Bayes empirical Bayes estimator $\{\hat{t}_n : n \geq 1\}$ is asymptotically optimal.

**Proof.** Let $\tau_{\omega,n}$ be a component Bayes estimator versus $\omega$ based on $X_n$. Then, as in (2.5) of Chapter 1,

(2.1) $$\left| \int\int \| \hat{t}_n - \theta_n \|^2 dP_{\underline{\theta}} d\omega^n - r(\omega) \right| \leq 4M\ P_\omega^{\ n} \left\| \tau_{\omega_{n,n}} - \tau_{\omega,n} \right\|;$$

by Proposition 3.1 of Chapter 1 it is enough to show that

(2.2) $$P_\omega^{\ n-1} \left\| P_{\omega_{n,n}} - P_\omega \right\| \to 0 \text{ as } n \to \infty.$$

The uniform (in $\omega$) version of (2.2), with n replacing $n-1$, is the assertion (with some changes in notation) of the following corollary [Corollary 2.1] to Theorem 1 in Chapter 2; applying that corollary we complete the proof. //

Corollary 2.1. Let $\hat{\omega}$ be as in Chapter 2. Under the assumption $S_\Lambda = \Omega$,

$$P_\omega{}^n \| P_{\hat{\omega}} - P_\omega \| \to 0, \text{ uniformly in } \omega, \text{ as } n \to \infty.$$

Proof. Noting that $P_\omega{}^n$ is the marginal on $\mathfrak{X}^n$ of the joint distribution on $\mathfrak{X}^n \times \Theta^n$ obtained from $\mathbf{P}_{\underline{\theta}}$ and $\omega^n$, and triangulating around $P_{G_n}$,

$$(2.3) \qquad P_\omega{}^n \| P_{\hat{\omega}} - P_\omega \| \le \int \mathbf{P}_{\underline{\theta}} \| P_{\hat{\omega}} - P_{G_n} \| d\omega^n + \omega^n(\| P_{G_n} - P_\omega \|) \; .$$

Now the first term in rhs(2.3) goes to zero uniformly in $\omega$ by Theorem 1 in Chapter 2.

By the moment inequality, applied first to the $\omega^n$ integral and then to the $\mu$ integral,

$$(2.4) \qquad (\omega^n \| P_{G_n} - P_\omega \|)^2 \le \omega^n(\| p_{G_n} - p_\omega \|_2{}^2) \; .$$

Now by interchanging the order of $\mu$ and $\omega^n$ integration, noting that $\omega^n(p_{G_n} - p_\omega)^2$ is the variance of the average of n iid random variables and bounding the variance of a random variable by its second moment, we get

$$(2.5) \qquad \text{rhs}(2.4) \le n^{-1} \int \int p_\theta{}^2 d\omega(\theta) d\mu.$$

Interchanging the order of $\mu$ and $\omega$ integration, and applying Lemma 1.3 of Chapter 1 with $q = 2$ to bound $\mu(p_\theta{}^2)$,

$$(2.6) \qquad \text{rhs}(2.5) \le n^{-1} e^{M^2}.$$

The corollary follows by (2.4)-(2.6). //

<u>Remark</u> <u>2.1.</u> By (2.1), Proposition 3.1 of Chapter 1 and Corollary 2.1, we get that lhs(2.1) goes to 0 uniformly in $\omega$, which is a stronger form of asymptotic optimality.

# APPENDIX

# APPENDIX

**1. On measurability.**

In this section, we prove two lemmas concerning measurability of two maps which have been used in the main body of the dissertation.

**Lemma 1.1.** Let $\Theta$ be a separable metric space endowed with its Borel $\sigma$-field. Let $\Omega$ be the set of all probabilities on $\Theta$, endowed with the topology of weak convergence and the corresponding Borel $\sigma$-field. Then, for every bounded real-valued measurable function h on $\Theta$, the map $\omega \mapsto \omega(h)$ is measurable.

**Proof.** We shall use the following theorem (TI.20) from Meyer (1966):

*Let $\mathcal{H}$ be a vector space of bounded real-valued functions defined on $\Gamma$, which contains the constant 1, is closed under uniform convergence, and is such that for every increasing, uniformly bounded sequence of non-negative functions $g_n \in \mathcal{H}$, the function $g = \lim_{n \to \infty} g_n$ belongs to $\mathcal{H}$. Let $\mathcal{C}$ be a subset of $\mathcal{H}$, closed under multiplication. Then the space $\mathcal{H}$ contains all the bounded functions measurable with respect to the $\sigma$-field $\mathcal{T}$ generated by the elements of $\mathcal{C}$.*

Let $\Gamma = \Theta$, $\mathcal{H} = \{h : h$ is a bounded real-valued function on $\Theta$ and $\omega \mapsto \omega(h)$ is measurable$\}$; clearly $\mathcal{H}$ satisfies all the conditions of TI.20, Meyer (1966). Let $\mathcal{C} = \{B \subseteq \Theta : B$ is closed$\}$. Clearly $\mathcal{C}$ is closed under multiplication. Since, by the portmanteau Theorem [Theorem II.6.1(c), Parthasarathy (1967)], for every $B \in \mathcal{C}$ and every $k \in \Re$ the set $\{\omega : \omega(B) \geq k\}$ is closed in the topology of weak convergence, we get that $\mathcal{C}$ is a subset of $\mathcal{H}$. Therefore $\mathcal{H}$ contains all the bounded real-valued measurable functions on $\Theta$. //

**Lemma 1.2.** Let $(\mathfrak{X},\mathfrak{F})$ be a measurable space. Let $\Theta$ be a separable metric space endowed with its Borel $\sigma$-field. Let $f:\Theta\times\mathfrak{X}\mapsto[0,\infty)$ be a measurable function. Let $\Omega$ be the set of all probabilities on $\Theta$, endowed with the topology of weak convergence and its Borel $\sigma$-field. For $\omega\in\Omega$, let $f(\omega,x) := \int f(.,x)d\omega$; then $f:\Omega\times\mathfrak{X}\mapsto[0,\infty)$ is measurable.

**Proof.** We shall again use TI.20, Meyer (1966). Let $\Gamma=\Theta\times\mathfrak{X}$, $\mathcal{H}=\{h$ : $h$ is a bounded real-valued function on $\Theta\times\mathfrak{X}$ and $(\omega,x)\mapsto\int h(.,x)d\omega$ is measurable$\}$; clearly $\mathcal{H}$ satisfies all the conditions of TI.20, Meyer (1966). Let $\mathcal{C}=\{A\times B$ : $A$ is a measurable subset of $\Theta$, $B$ is a measurable subset of $\mathfrak{X}\}$. $\mathcal{C}$ is clearly closed under multiplication. That $\mathcal{C}$ is a subset of $\mathcal{H}$ follows from Lemma 1.1. Therefore $\mathcal{H}$ contains all the bounded real-valued measurable functions on $\Theta\times\mathfrak{X}$, in particular, $f\wedge M$ for every integer $M$. Since $\{f\wedge M$ : $M$ is an integer$\}$ is an increasing, uniformly bounded sequence of non-negative functions, its pointwise limit $f$ also belongs to $\mathcal{H}$. That completes the proof of the lemma. //

## 2. On topological support of Dirichlet prior.

In this section we present a result of independent interest characterizing the topological support of a Dirichlet prior on an arbitrary separable metric space, which is used in Remark 4.3 of Chapter 1 to give examples of $\Lambda$ with full support.

Ferguson (1973) states that the topological support of a Dirichlet prior on the Borel $\sigma$-field (corresponding to the weak convergence topology) of the set of all probability measures on the line is the set of all probability measures with their topological supports contained in that of the parameter measure of the Dirichlet prior. We prove that statement with the line replaced by an arbitrary separable metric space.

Let $\mathfrak{X}$ be a separable metric space and $\mathcal{A}$ be the Borel $\sigma$-field of $\mathfrak{X}$. Let $\Omega$ be the set of all probability measures on $(\mathfrak{X},\mathcal{A})$. The topology of weak convergence on $\Omega$ is metrizable as a separable metric space [Theorem II.6.2, Parthasarathy (1967)]; let $\mathfrak{B}(\Omega)$ denote its Borel $\sigma$-field.

We consider the random probability measure $P$ defined in (4.7) of Ferguson (1973). Let $\{V_n : n \geq 1\}$ be a sequence of iid random elements taking values in $(\mathfrak{X},\mathcal{A})$ with common distribution $\mathbb{Q}$, where $\mathbb{Q}(A) = \alpha(A)/\alpha(\mathfrak{X})$ and $\alpha$ is a finite non-null measure on $(\mathfrak{X},\mathcal{A})$. Let $\{J_n : n \geq 1\}$ be a sequence of non-negative random variables independent of $\{V_n : n \geq 1\}$. For $j \geq 2$, let the conditional distribution of $J_j$ given $J_{j-1}, \ldots, J_1$ be equal to the distribution of $J_1$ truncated above at $J_{j-1}$; let the distribution function of $J_1$ be $\exp(N(.))$, where $N(x) = -\alpha(\mathfrak{X})\int_x^\infty e^{-y}y^{-1}dy$ for $x > 0$. In Theorem 4.1 of Ferguson (1973) it is proved that $\sum_1^\infty J_n$ converges w.p. 1. For $A \in \mathcal{A}$, define

$$P(A) = \sum_1^\infty P_j \chi_{V_j}(A),$$

where $P_n = \dfrac{J_n}{\sum_1^\infty J_n}$ and $\chi_v(A) = 1$ if $v \in A$ $= 0$ otherwise.

Clearly, for every point in the set (in the probability space underlying the random sequences $\{P_n\}$ and $\{V_n\}$) on which $\sum_1^\infty J_n$

converges, $P$ is a probability measure on $\mathcal{A}$. Therefore without loss of generality we can assume $P$ to be $\Omega$ valued. Let $\phi_A$ be the map on $\Omega$ taking $\omega$ to $\omega(A)$. Since the real-valued map $P(A)$ is Borel measurable, $P$ is measurable with respect to $\sigma\{\phi\}$, the $\sigma$-field generated by the family $\{\phi_A : A \in \mathcal{A}\}$. Note that by Lemma 1.1, $\sigma\{\phi\}$ is a sub $\sigma$-field of $\mathfrak{B}(\Omega)$. We shall show that $\mathfrak{B}(\Omega)$ is a sub $\sigma$-field of $\sigma\{\phi\}$. We shall denote the induced distribution of $P$ on $(\Omega, \mathfrak{B}(\Omega))$ by $\mathfrak{P}$ and refer to it as the Dirichlet prior on $\mathfrak{B}(\Omega)$ with parameter $\alpha$. By Theorem 4.2 of Ferguson (1973), for every $k = 1,2,\ldots,$ and measurable partition $(B_1,\ldots,B_k)$ of $\mathfrak{X}$, the distribution of $(P(B_1),\ldots,P(B_k))$ is Dirichlet with parameters $(\alpha(B_1),\ldots,\alpha(B_k))$. Note that in the sense of Ferguson (1973) if the j-th parameter of a Dirichlet distribution is equal to 0, then the j-th coordinate is degenerate at 0.

To prove $\mathfrak{B}(\Omega) \subset \sigma\{\phi\}$, recall [Theorem 3, Appendix III, Billingsley (1968)] that

$$\begin{aligned}\mathfrak{U} := \{&N(\mu\ ;\ A_1,\ldots,A_k\ ;\ \epsilon_1,\ldots,\epsilon_k) : \mu \in \Omega,\ \epsilon_i > 0,\\ &A_i\ \mu\text{-continuity subset of } \mathfrak{X},\ i = 1,2,\ldots,k,\ k = 1,2,\ldots\}\end{aligned}$$

is an open base for the topology of weak convergence on $\Omega$, where

$$N(\mu\ ;\ A_1,\ldots,A_k\ ;\ \epsilon_1,\ldots,\epsilon_k) := \bigcap_{i=1}^{k} \{\nu \in \Omega\ :\ |\nu(A_i) - \mu(A_i)| < \epsilon_i\}.$$

Obviously, every set in $\mathfrak{U}$ is in $\sigma\{\phi\}$. Using separability and metrizability of $\Omega$, which together imply second countability, we conclude by Lindelöf's Theorem that every open subset of $\Omega$ is in $\sigma\{\phi\}$; hence $\mathfrak{B}(\Omega) \subset \sigma\{\phi\}$.

For a finite Borel measure $m$ on a second countable topological space $\mathcal{S}$, the topological support of $m$ is defined to be the set

$$S_m = \cap\{F : F \text{ is closed and } m(F^c) = 0\}.$$

Note that $s \in S_m$ iff for any open set $O$ containing $s$, we have $m(O) > 0$. Since $\mathcal{S}$ is second countable, by Lindelöf's Theorem $S_m{}^c$ can be expressed as a countable union of $F^c$ sets. Therefore

$$(2.1) \qquad m(S_m{}^c) = 0;$$

hence

$$(2.2) \qquad \text{if B is closed and } m(B) = m(S_m), \text{ then } S_m \subset B.$$

In the sequel, the set

$$\{(x_1, \ldots, x_m) \in \Re^m : 0 \le x_i \ \forall \ i = 1, 2, \ldots, m \text{ and } \sum_{i=1}^{m} x_i \le 1\}$$

will be referred to as the *sub-simplex* in m-dimension.

We now state the main result of this section characterizing the topological support of a Dirichlet prior $\mathcal{P}$ with parameter $\alpha$.

**Theorem 2.1.** $S_{\mathcal{P}} = \{\mu \in \Omega : S_\mu \subset S_\alpha\}$.

**Proof.** We first show $S_{\mathcal{P}} \supset \{\mu \in \Omega : S_\mu \subset S_\alpha\}$. Let $S_\mu \subset S_\alpha$ ; we shall show that every basic open set in $\mathcal{U}$ containing $\mu$ has positive $\mathcal{P}$-probability to conclude $\mu \in S_{\mathcal{P}}$. Now for arbitrary positive integer k, $\mu$-continuity subsets $A_1, \ldots, A_k$ of $\mathcal{X}$ and positive numbers $\epsilon_1, \ldots, \epsilon_k$,

$$(2.3) \qquad \mathcal{P}(\bigcap_{i=1}^{k} \{\nu \in \Omega : |\nu(A_i) - \mu(A_i)| < \epsilon_i\}) = \Pr(\bigcap_{i=1}^{k} \{|P(A_i) - \mu(A_i)| < \epsilon_i\}),$$

where Pr is the probability measure on the domain of $P$.

Let $\{F_{\nu_1 \ldots \nu_k} : \nu_i = 0 \text{ or } 1 \ \forall \ i = 1, \ldots, k\}$ denote the measurable partition generated by $A_1,\ldots,A_k$ ; i.e.

$$F_{\nu_1\ldots\nu_k} = \bigcap_{j=1}^{k} A_j^{\nu_j} \quad \text{where } A^{\nu_j} = A \text{ if } \nu_j = 1$$
$$= A^c \text{ otherwise.}$$

Then, noting that $A_i = \bigcup_{\nu_i=1} F_{\nu_1\ldots\nu_k}$ and using subadditivity of distance, we get

$$\text{(2.4)} \qquad \bigcap_{i=1}^{k}\{|(P-\mu)(A_i)| < \epsilon_i\} \supset \bigcap_{i=1}^{k}\{\sum_{\nu_i=1}|(P-\mu)(F_{\nu_1\ldots\nu_k})| < \epsilon_i\}.$$

Since the class of $\mu$-continuity sets form a field [Lemma II.6.4, Parthasarathy (1967)], $F_{\nu_1\ldots\nu_k}$ is a $\mu$-continuity set; i.e.,

$$\text{(2.5)} \qquad \mu(\delta F_{\nu_1\ldots\nu_k}) = 0.$$

Note that,

$$\text{(2.6)} \qquad \text{if } \alpha(F_{\nu_1\ldots\nu_k}) = 0, \text{ then } \mu(\text{int}F_{\nu_1\ldots\nu_k}) = 0.$$

Because $\alpha(F_{\nu_1\ldots\nu_k}) = 0$ implies $\alpha((\text{int}F_{\nu_1\ldots\nu_k})^c \cap S_\alpha) = \alpha(S_\alpha)$, and since $(\text{int}F_{\nu_1\ldots\nu_k})^c \cap S_\alpha$ is closed, by the observation in (2.2), $(\text{int}F_{\nu_1\ldots\nu_k})^c \supset S_\alpha$. Since $S_\mu \subset S_\alpha$, the claim follows by (2.1).

Since $(P(F_{\nu_1\ldots\nu_k}), P((F_{\nu_1\ldots\nu_k})^c))$ has a Dirichlet distribution with parameters $(\alpha(F_{\nu_1\ldots\nu_k}), \alpha((F_{\nu_1\ldots\nu_k})^c))$, $P(F_{\nu_1\ldots\nu_k})$ is degenerate at 0 if $\alpha(F_{\nu_1\ldots\nu_k}) = 0$; hence, by (2.5) and (2.6),

$$\text{(2.7)} \qquad \text{rhs}(2.4) \supset \bigcap\{[|(P-\mu)(F_{\nu_1\ldots\nu_k})| < 2^{-k}\epsilon] \ : \ \alpha(F_{\nu_1\ldots\nu_k}) > 0\},$$

where $\epsilon = \bigwedge_{i=1}^{k} \epsilon_i$.

Now $\{P(F_{\nu_1\ldots\nu_k}) \ : \ \alpha(F_{\nu_1\ldots\nu_k}) > 0\}$ has a Dirichlet distribution with all parameters positive; since, by (2.5) and (2.6),

$$\sum\{\mu(F_{\nu_1\ldots\nu_k}) : \alpha(F_{\nu_1\ldots\nu_k}) > 0\} = 1,$$

temporarily abbreviating $(\nu_1,\ldots,\nu_k)$ to $\underline{\nu}$ and fixing a $\tilde{\underline{\nu}}$ for which $\alpha(F_{\tilde{\underline{\nu}}}) > 0$, we get

$$\text{(2.8)} \qquad \text{rhs}(2.7) \supset \bigcap\{[\,|P(F_{\underline{\nu}}) - \mu(F_{\underline{\nu}})| < 2^{-2k}\epsilon] : \underline{\nu} \neq \tilde{\underline{\nu}} \text{ and } \alpha(F_{\underline{\nu}}) > 0\}.$$

Since $\{P(F_{\nu_1\ldots\nu_k}) : \alpha(F_{\nu_1\ldots\nu_k}) > 0\}$ has a Dirichlet distribution with all parameters positive, the induced distribution (over the sub-simplex in appropriate dimension) of a one-component-deleted subvector of $\{P(F_{\nu_1\ldots\nu_k}) : \alpha(F_{\nu_1\ldots\nu_k}) > 0\}$ puts positive mass on every subset of the sub-simplex with non-empty interior. By (2.4), (2.7) and (2.8), lhs(2.3) is positive. That completes the proof of $\mu \in S_{\mathcal{P}}$.

Conversely, suppose $\mu \in S_{\mathcal{P}}$; to show $S_\mu \subset S_\alpha$ it is enough (by the observation in (2.2)) to show $\mu(S_\alpha{}^c) = 0$. Since by Theorem II.6.1(d) of Parthasarathy (1967) $\underline{\lim}\ \omega_n(A) \geq \omega(A)$ whenever $\omega_n \to \omega$ and $A \subset \mathfrak{X}$ is open, the set $\{\nu \in \Omega : \mu(A) \geq \nu(A) + \epsilon\}$ is closed (in the topology of weak convergence) for every open set $A \subset \mathfrak{X}$ and every $\epsilon > 0$. Since $S_\alpha{}^c$ is open, $\{\nu \in \Omega : \mu(S_\alpha{}^c) < \nu(S_\alpha{}^c) + \epsilon\}$ is an open set containing $\mu$ for every $\epsilon > 0$; since $\mu \in S_{\mathcal{P}}$,

$$\mathcal{P}\{\nu \in \Omega : \mu(S_\alpha{}^c) < \nu(S_\alpha{}^c) + \epsilon\} > 0 \text{ for every } \epsilon > 0.$$

Now $\nu(S_\alpha{}^c) = 0$ a.s. $(\mathcal{P})$, because $(P(S_\alpha{}^c), P(S_\alpha))$ has a Dirichlet distribution with parameters $(\alpha(S_\alpha{}^c), \alpha(S_\alpha))$ and $\alpha(S_\alpha{}^c) = 0$ by (2.1). Therefore, $\mu(S_\alpha{}^c) < \epsilon$ for every $\epsilon > 0$; that is, $\mu(S_\alpha{}^c) = 0$. That completes the proof. //

# BIBLIOGRAPHY

# BIBLIOGRAPHY


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

CHOW, Y.S. and TEICHER, H. (1988). *Probability Theory, Independence, Interchangeability, Martingales.* Springer Verlag, New York.

DATTA, S. (1988). Asymptotically optimal Bayes compound and empirical Bayes estimators in exponential families with compact parameter space. Ph.D. dissertation, Dept. Statist and Probab., 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.

DUDLEY, R.M. (1967). The sizes of compact subsets of Hilbert Space and continuity of Gaussian processes. *J. Functional Analysis* **1** 290-330.

DUDLEY, R.M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, California.

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

FERGUSON, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209-230.

GILLILAND, D.C. (1968). Sequential compound estimation. *Ann. Math. Statist.* **39** 1890-1904.

GILLILAND, D.C. and HANNAN, J. (1974/86). The finite state compound decision problem, equivariance and restricted risk components. In *Adaptive Statistical Procedures and Related Topics* (J.Van Ryzin, ed.) *IMS Lecture Notes-Monograph Series* **8** 129-145. IMS, Haywood, Calif.

GILLILAND, D. C., HANNAN, J. and HUANG, J. S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4** 1101-1112.

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.

HANNAN, J.F. (1957). Approximation to Bayes risk in repeated play. *The Theory of Games 3. Ann. Math. Studies.* **39** 97-139. Princeton Univ. Press.

HANNAN, J. (1960). Consistency of maximum likelihood estimation of discrete distributions. In *Contributions to Probability and Statistics : Essays in Honor of Harold Hotelling* (I. Olkin et. al., eds.) 244-257. Stanford University Press.

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

KUO, L. (1986). A note on Bayes empirical Bayes estimation by means of Dirichlet processes. *Statist. Probab. Lett.* **4** 145-150.

LECAM, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

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

MEYER, P.A. (1966). *Probability and Potential.* Blaisdell Publishing Company, Waltham, Massachusetts.

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

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

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 1 157-164. Univ. California Press, Berkeley.

SINGH, R.S. (1974). Estimation of derivatives of average $\mu$-densities and sequence-compound estimation in exponential families. Ph.D. dissertation, Dept. Statist. and Probab., Michigan State Univ.

ZHU, J. (1992). Asymptotic behavior of compound rules in compact regular and nonregular families. Ph.D. dissertation, Dept. Statist. and Probab., Michigan State Univ.