This is to certify that the

dissertation entitled

Set Compound Decision Estimation
Under Entropy Loss In Exponential Families

presented by

Zhihui Liu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor

Date June 17, 1997

Major Professor

SET COMPOUND DECISION ESTIMATION UNDER ENTROPY LOSS IN EXPONENTIAL FAMILIES

By

Zhihui Liu

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

1997

ABSTRACT

SET COMPOUND DECISION ESTIMATION UNDER ENTROPY LOSS IN EXPONENTIAL FAMILIES

By

Zhihui Liu

Set compound decision estimation has been studied for half a century, starting with finite state examples.

Set compound estimation under the entropy (Kullback-Leibler, same through the thesis) loss for a k-dimensional standard exponential family with a compact parameter space is discussed here. The entropy loss with the exponential family and related properties are investigated in detail. Asymptotically optimal set compound estimators with rates $O(n^{-\frac{1}{2}})$ under this loss are established for one dimensional discrete exponential families including Poisson and negative binomial by being able to view the Bayes estimators as a ratio of two power series and representing them in form of mixture density and using Singh-Datta Lemma. Some remarks related to squared error loss and noncompact state case are also given. Generalization to a higher dimensional discrete exponential family is illustrated with a two dimensional example. Secondly going from cumulant generating functions and using kernel density estimation, continuous exponential families are studied under the same setting. Normal and Gamma distribution families are examples.

To my husband, my son and my daughter.

ACKNOWLEDGMENTS

First the author would like to take this opportunity to thank Professors Dennis Gilliland and James Hannan for suggesting the problems, for valuable recommendations and for their criticism through which the author has gained research abilities. The author has benefited from their knowledge both in statistics and language.

The author likes to express her thanks to the department for the financial support and consideration during her Ph. D. study here. She wants to thank Professor Salehi, Professor Stapleton, Professor LePage and other committee members for their help.

At the end, the author thanks her parents for their encouragement and her husband for his understanding and support in many ways.

TABLE OF CONTENTS

Chapter 0

INTRODUCTION

Compound decision theory was introduced by Robbins (1951) through an example of decision between N(-1,1) and N(1,1). In his paper he proposed a compound procedure that was optimal in the sense of asymptotic subminimaxity and showed the necessity of this kind of procedure. The theory was greatly developed and in much general finite state situation by Hannan (1956) and (1957). He used a randomization technique to overcome the difficulty caused by discontinuity of a Bayes response with respect to priors in his (1957) paper. Since then many works have been done in this area. Gilliland (1968) further demonstrated the necessity of compound procedures by showing that the supremum of regrets over both a parameter space and stages for any simple procedure is positive. Vardeman (1982) applied Hannan (1957)'s randomization technique to extended sequence compound problems for finite state case. Gilliland and Hannan (1986) considered set compound problems in a setting of restricted risk components that avoided action space and loss function. In their paper a large class of asymptotic solutions to the set compound decision problems for finite state case were established. Singh (1974) considered sequence compound problems for exponential families with compact parameter spaces for parameter functions in three cases: θ, e^{θ} and

θ^{-1}, and yielded some rates. Datta (1991a) generalized and strengthened Gilliland, Hannan and Huang (1976)'s work from finite state case to compact state case and established an asymptotically optimal Bayes compound estimator based on a hyperprior for a continuous parameter function of an exponential family with a compact parameter space under squared error loss. Mashayekhi (1991 and 1995) strengthened Datta (1991a)'s result to the equivariant envelope. He also successfully extended Hannan and Huang (1972)'s results on the stability of symmetrization of product measures to the compact state case for exponential families. Zhu (1992) developed Datta's work (1991) to multidimensional case. In addition he considered a nonregular family - two dimensional truncation family, and got some rate for this family. Majumdar (1993) developed Datta (1991a) and (1991b)'s work to Hilbert spaces.

The present work considers the problem of set compound estimation of the natural parameter in an exponential family with entropy loss. The families considered here had been the focus of sequence compound estimation of θ, e^{θ} and θ^{-1} with squared error loss [see Samuel (1965), Gilliland (1968), Swain (1965), Yu (1971) and Singh (1974)]. Estimation with entropy loss was also studied extensively. [see M. Ghosh and M. C. Yang (1988), D. K. Dey, M. Ghosh and C. Srinivasan (1987)]. The work draws on the earlier work, but the results are not immediate extensions of that work. Compound estimation in the Hannan's sense with entropy loss is first attempted apparently. Terms not previously considered in the compound literature have been analyzed.

Chapter 1 conducts a general discussion about component and set compound problems for exponential families with compact parameter spaces under entropy loss.

Chapter 2 obtains asymptotically optimal set compound estimators with rates $O(n^{-1/2})$ for some discrete exponential families with compact parameter spaces by representing the Bayes estimators in terms of mixture density under entropy loss. Chapter 3 discusses set compound estimation with continuous exponential families under entropy loss through using the kernel density estimation considered in Singh (1974).

Chapter 1

SET COMPOUND DECISION PROBLEM UNDER ENTROPY LOSS IN EXPONENTIAL FAMILIES

1.1 The Component Decision Problem

The component statistical decision problem considered has a standard exponential family $\{P_\theta : \theta \in \Theta\}$, where P_θ has density

$$(1.1) \qquad \tilde{p}_\theta(x) = e^{\theta' x - \psi(\theta)}$$

with respect to a measure μ on R^k and $\theta' x$ denotes the inner product of θ and x in R^k. Θ is a subset of the natural parameter space

$$(1.2) \qquad N = \{\theta: \int e^{\theta' x} d\mu < \infty\}$$

and

$$(1.3) \qquad \psi(\theta) = \ln(\int e^{\theta' x} d\mu),\ \theta \in N$$

is the cumulant generating function (see Brown (1986, page 1)). Of course, in this thesis, we consider only the case where Θ, and therefore, N are nonempty.

Define

$$(1.4) \qquad N^{\sim} = \{\theta \in N:\ E_\theta |X| < \infty\}$$

where $|\cdot|$ is the Euclidean norm and

(1.5) $$\eta(\theta) = E_\theta X \ , \ \theta \in N^{\sim} .$$

For future reference, note that the interior N^0 of the natural parameter space is a subset of $N^{\sim}$ and that ψ has derivatives of all orders on N^0 with

$$\dot{\psi}(\theta) = E_\theta(X) = \eta(\theta) \, , \ \ddot{\psi}(\theta) = \mathrm{Var}_\theta(X) \, ,$$

the variance-covariance matrix of $X \sim P_\theta$ (see Brown (1986, Theorem 2.2)).

Throughout this thesis we assume that P_θ is identified by θ. Note that P_θ is equivalent to μ for all $\theta \in N$, so that for every distribution G on Θ, the mixture P_G is equivalent to μ.

The action space A is a nonempty subset of N and the loss function is the entropy loss function:

(1.6) $$L(\theta, a) = \int (\ln \frac{\tilde{p}_\theta}{\tilde{p}_a}) \tilde{p}_\theta d\mu \, , \theta \in N, a \in N.$$

Note that $L(\theta,a) = K(P_\theta, P_a)$, the Kullback-Leibler information of P_a at P_θ and that $L(\theta,a) \geq 0$ and $= 0$ if and only if $P_\theta = P_a$ (see Brown (1986, page 174)). Because of our identifiability assumption, $L(\theta,a) = 0$ if and only if $\theta = a$. It follows from (1.1) and (1.6) that

(1.7) $$L(\theta, a) = (\theta - a)' \eta(\theta) - \psi(\theta) + \psi(a) , \theta \in N^{\sim} , a \in N.$$

A non-randomized decision rule t is a mapping from the underlying measure space X into A such that L(θ, t) is a measurable function of X for each $\theta \in \Theta$. The risk of t at θ is

$$R(\theta, t) = \int L(\theta, t) dP_\theta \, .$$

If G is a probability distribution on Θ and $R(\cdot,t)$ is measurable, then the Bayes risk of t at G is $R(G, t) = \int R(\cdot,t)dG$. The infimum Bayes risk over all decision rules is denoted by R(G). A decision rule τ_G such that $R(G,\tau_G) = \inf_t R(G,t)$ is called a Bayes decision rule versus G or simply Bayes versus G. The terms non-randomized decision rule and estimator are synonymous in this thesis.

The following notations are used in this thesis. Let f^{-1} stand for the inverse of a function f if it exists. Let $f[\Theta]$ be the image of f on a domain Θ. ":=" is used as "denote" or "is denoted". $P_G(B) := \int P_\theta(B)dG(\theta)$ for prior probability G and measurable subsets B. We denote the vector and matrix of the first and second derivatives by dot and double dot notations. For example, $\dot{\psi}$ and $\ddot{\psi}$. Let E be the joint expectation operator of x and θ in the rest of the chapter unless otherwise noted.

Bayes estimators under the entropy loss (1.6) are different from that under squared error loss. The following propositions demonstrate the unique (a.e. μ) Bayes estimator under certain conditions and entropy loss. We will use the following lemma in the proof of Proposition 1.2.

Lemma 1.1 η is a 1-1 function on $N^{\sim}$.

Proof. Let $\theta_1, \theta_2 \in N^{\sim}$ and note that by (1.7),

$$L(\theta_1,\theta_2) + L(\theta_2,\theta_1) = (\theta_1 - \theta_2)'(\eta(\theta_1) - \eta(\theta_2)). \tag{1.8}$$

Thus, $\eta(\theta_1) = \eta(\theta_2)$ implies that the nonnegative quantities $L(\theta_1,\theta_2)$, $L(\theta_2,\theta_1)$ are both zero from which $\theta_1 = \theta_2$. ◊

Proposition 1.1 Suppose that Θ and A are measurable subsets of $N^{\sim}$ with $\Theta \subseteq A$, A is compact and η is continuous on A. Let G be a distribution on Θ. Suppose that τ is an estimator satisfying

(1.9) $$\eta(\tau(x)) = E(\eta(\theta)|x) \qquad \text{a.e. } \mu .$$

Then for any estimator t

$$R(G,t) \geq R(G,\tau)$$

with equality iff $t = \tau$ a.e. μ. Thus, τ is the unique Bayes estimator.

Proof. Using (1.7),

(1.10) $$R(G,t) - R(G,\tau) = EL(\theta,t) - EL(\theta,\tau) =$$

$$E[(\tau - t)'\eta(\theta) + \psi(t) - \psi(\tau)] = EL(\tau,t) + E\{(\tau - t)'[\eta(\theta) - \eta(\tau)]\}.$$

It follows by taking conditional expectation given x that the second term on the right hand side (1.10) = 0 by the hypothesis (1.9). The first term on RHS(1.10) is ≥ 0 since $L \geq 0$ and furthermore it equals to 0 if and only if $\tau = t$ a.e. P_G. As noted previously, P_G is equivalent to μ, which completes the proof. ◊

Proposition 1.2 Suppose that Θ and A are measurable subsets of $N^{\sim}$ with $\Theta \subseteq A$, A is compact and η is continuous on A. Let G be a prior on Θ. Suppose that

(1.11) $$\text{convex hull } \{\eta[\Theta]\} \subseteq \eta[A] .$$

Then there is an estimator $\tau = \tau_G$ satisfying (1.9) (a.e. μ) and

(1.12) $$\tau_G(x) = \eta^{-1}[E(\eta(\theta)|x)] \qquad \text{a.e. } \mu$$

Proof. In the beginning we show that RHS (1.12) is well defined. Since η is continuous and A is compact, $\eta[A]$ is compact. Thus by (1.11) the closure of the convex hull of $\eta[\Theta]$ is compact. It follows from Theorem 3.27 of Rudin (1973) that there is a version of the conditional expectation $E(\eta(\theta)|x)$ that takes values in $\eta[A]$. η is 1-1 by Lemma 1.1 so that with this version $\tau_G := \eta^{-1}[E(\eta(\theta)|x)]$ takes values in A and is seen to satisfy (1.9). ◊

Remark 1.1 Suppose that $\eta(\cdot)$ is continuous on Θ which is compact. Then $\eta[\Theta]$ is compact, Therefore the convex hull of Θ is a compact subset of $N^{\sim}$ by Theorem 3.25(b) of Rudin (1973).

Remark 1.2 Propositions 1.1 and 1.2 develop the unique a.e. μ Bayes response τ_G with respect to a prior distribution G on Θ when the loss function is entropy loss for estimation of θ in an exponential family. The assumptions are $\Theta \subseteq A \subseteq N^{\sim}$ where A is compact, η is continuous on A and (1.11) is valid. These assumptions are satisfied if $\Theta \subseteq A \subseteq N^0$ where A is compact and (1.11) holds. This is the case since $N^0 \subseteq N^{\sim}$ and ψ is infinitely differentiable on N^0 with $\dot{\psi} = \eta$.

Remark 1.3 The condition (1.11) holds in the one-dimensional case if $\Theta \subseteq A \subseteq N^{\sim}$, A = [a, b] and η is continuous on A. When k = 1, η is seen to be strictly increasing by (1.8) so that $\eta[A]=[\eta(a), \eta(b)]$. Hence $\eta[\Theta] \subseteq \eta[A]$ which implies (1.11).

Remark 1.4 The condition (1.11) holds in the one-dimensional case if $\Theta \subseteq A \subseteq N^{\sim}$, Θ = [a, b] and η is continuous on Θ. Here convex hull$\{\eta[\Theta]\} = \eta[\Theta] \subseteq \eta[A]$.

Remark 1.5 If $\Theta = A$, (1.11) is equivalent to $\eta[\Theta]$ convex.

1.2 The Corresponding Set Compound Problem

The set compound decision problem with n components is defined for each n = 1, 2,$\cdots$ as follows: for given n it consists of n independent repetitions of the component decision problem with compound loss equal to the average loss across the component decisions. The observation $\underline{X} = (X_1, \cdots, X_n)$ has distribution $P_{\underline{\theta}} := \prod_{j=1}^{n} P_{\theta_j}$ with $\underline{\theta} = (\theta_1, \cdots, \theta_n) \in \Theta^n$. A compound estimator $\underline{t} := (t_1(\underline{X}), \cdots, t_n(\underline{X}))$ has compound risk

(1.13) $$R_n(\underline{\theta}, \underline{t}) = n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} L(\theta_i, t_i(\underline{X}))$$

and regret

(1.14) $$D_n(\underline{\theta}, \underline{t}) = R_n(\underline{\theta}, \underline{t}) - R(G_n)$$

where R is the Bayes envelope in the component decision problem and G_n denotes the empirical distribution of $\theta_1, \cdots, \theta_n$. (Note that the notation hides the dependence of the component decision rules $t_i(\underline{X})$ on n, i = 1, ... , n). If a sequence of compound estimators $\{\underline{t}\}$ satisfies

(1.15) $$\sup_{\underline{\theta}} D_n(\underline{\theta}, \underline{t}) = o(1)(O(\alpha_n)),$$

we say it is asymptotically optimal (with rate α_n). ((1.15) is equivalent to $\sup_{\underline{\theta}} D_n(\underline{\theta}, \underline{t})_+ = o(1)(O(\alpha_n))$ used by Datta (1988, page 3)). In this thesis our main concern is the construction of compound estimators $\{\underline{t}\}$ satisfying (1.15) with demonstrated rate.

Remark 1.6 The property (1.15) is uniform in parameter sequences $\{\underline{\theta}\}$ so that results for given Θ with $\Theta \subseteq A$ are a corollary to (1.15) for the case $\Theta = A$. Since the construction of rules $\{\underline{t}\}$ satisfying (1.15) is the focus of this thesis, we will henceforth take $\Theta = A$.

1.3 Bounds for the Regret $D_n(\underline{\theta}, \underline{t})$

Proposition 1.3 Suppose $\Theta = A \subseteq N^0$ where Θ is compact. Suppose that $\eta[\Theta]$ is convex. Then there exists a constant B_0 (not depending on $\underline{\theta}$ and n) such that

$$(1.16) \qquad |D_n(\underline{\theta}, \underline{t})| \leq B_0\, n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} |t_i - \tau_{G_n}(X_i)| \quad \text{for all } \underline{\theta} \text{ and n.}$$

Proof. By Remark 1.5, (1.11) is satisfied. By Remark 1.2 τ_{G_n} exists where G_n is the empirical distribution of $\theta_1, \cdots, \theta_n$. Note that

$$R(G_n) = \int E_\theta L(\theta, \tau_{G_n}) dG_n(\theta) = n^{-1} E_{\underline{\theta}} \sum_{i=1}^{n} L(\theta_i, \tau_{G_n}(X_i)).$$

Substituting this along with (1.13) into (1.14) and using (1.7) results in

$$(1.17) \quad D_n(\underline{\theta}, \underline{t}) = \frac{1}{n} \sum_{i=1}^{n} E_{\underline{\theta}} [\eta(\theta_i)'(\tau_{G_n}(X_i) - t_i(\underline{X})) + \psi(t_i(\underline{X})) - \psi(\tau_{G_n}(X_i))].$$

Thus, (1.16) follows from (1.17) and the mean value theorem with $B_0 = \sup\{\,|\eta(\theta_1) - \eta(\theta_2)| : \theta_1, \theta_2 \in \text{convex hull}\{\Theta\}\}$. ◊

The form of the Bayes response (1.12) and the bound RHS (1.16) motivate compound rules of the form

$$(1.18) \qquad t_i(\underline{X}) = \eta^{-1}[s_i(\underline{X})]\ ,\ i = 1, 2, \ldots, n,$$

where s_i is an estimator for the conditional expectation

$$(1.19) \qquad \tau^{*}_{G_n}(X_i) := E_n(\eta(\theta) | X_i),$$

where E_n denotes expectation on (θ, X_i) when $\theta \sim G_n$ and given θ, $X_i \sim P_\theta$. Note that $\tau^{*}_{G_n}(X)$ is a component Bayes response for estimating $\eta(\theta)$ with squared error loss and prior G_n.

Proposition 1.4 Suppose $\Theta = A \subseteq N^0$ where Θ is compact. Suppose that $\eta[\Theta]$ is convex and η^{-1} is a Lipschitz function on $\eta[\Theta]$ with constant C_0. Then $\underline{t}$ defined by (1.18) and (1.19) satisfies

(1.20) $$|D_n(\underline{\theta}, \underline{t})| \le B_0 C_0 n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} |s_i - \tau^{*}_{G_n}(X_i)|, \text{ for all } \underline{\theta} \text{ and n.}$$

Proof. The inequality (1.20) follows from (1.16), (1.18), (1.12), (1.19) and the assumption $|\eta^{-1}(a) - \eta^{-1}(b)| \le C_0 |a - b|$ for all a, b $\in \eta[A]$. ◊

Remark 1.7 It is easy to show that RHS(1.20) with $C_0 = 1$ is a bound for the absolute regret of the set compound estimator $\underline{s}$ in the estimation of $\eta(\theta)$ with squared error loss. In subsequent chapters we show that RHS(1.20) converges to zero uniformly in $\underline{\theta}$ with rates under various conditions. Hence, all of the results for the set compound estimator $\underline{t}$ for $\underline{\theta}$ under entropy loss transfer to corresponding results for the set compound estimator $\underline{s}$ of $\eta(\theta)$ under squared error loss. (For examples of estimators $\underline{t}$ and $\underline{s}$ see (2.14) and (2.15)).

Proposition 1.4 serves as the starting point for finding asymptotically optimal set compound estimators. For the exponential families considered in Chapters 2 and 3, the conditional expectations $\tau^{*}_{G_n}(X_i)$ are expressed in terms of ratios that are estimated. For bounding the error in ratio estimates, we will use Singh-Datta Lemma (see page 40 of Datta (1988)).

Singh-Datta Lemma For real numbers a, b, A, B, and D, with $b \neq 0 < D$, 0/0:=0,

$$|b|\{|\frac{a}{b}-\frac{A}{B}|\wedge D\} \le |b|^{-1}|aB-bA|+D(|b|-|B|)_+ \tag{1.21}$$
$$\le |a-A|+(|\frac{a}{b}|+D)|b-B|.$$

1.4 Summary

In the previous section, as is typical of the analysis of compound risk, the excess compound risk over the simple envelope is bounded by a Cesaro mean estimation error. In Proposition 1.3 the estimation is of the Bayes estimator $\tau_{G_n}(X_i)$ of θ under entropy loss and the Cesaro mean is RHS(1.16). In Proposition 1.4, the estimation is of the Bayes estimator $\tau^{\bullet}_{G_n}(X_i)$ for $\eta(\theta)$ under squared error loss and the Cesaro mean is RHS(1.20). In subsequent chapters, the estimation of $\tau^{\bullet}_{G_n}$ is accomplished in both discrete and continuous cases under assumptions on ψ and with compact parameter sets.

In Theorem 2.1 of Chapter 2, for certain one-dimensional discrete exponential families, set compound estimators $\{\underline{t}\}$ are established for which RHS (1.20) is $O(n^{-1/2})$ uniform in $\underline{\theta}$. The proof starts by expressing η as a power series in e^{θ}, $\eta(\theta)=\sum q_j(e^{\theta})^j$. Then interchanging the order of summation and conditional expectation in $\tau^{\bullet}_{G_n}(X_i)$ (see (1.19)), it is found that the terms can be consistently estimated. Substituting the estimates into the form of $\tau^{\bullet}_{G_n}(X_i)$ (see (1.19)) results in a compound estimator that is consistent in the Cesaro mean with rate. Poisson and negative binomial families with compact parameter spaces interior to their natural parameter spaces are covered. The proof can be extended to higher dimensional families, which is illustrated by working out the details for a particular 2-dimensional family.

In Chapter 3, for certain one-dimensional continuous exponential families, set compound rules $\{\underline{t}\}$ are established for which RHS (1.20) is $O(n^{-\gamma})$ uniform in $\underline{\theta}$ where γ is smaller than ½ (γ can be made arbitrarily close to 1/2 by choice of kernels in the estimation process).

In Theorem 3.1, the proof starts by expressing ψ as a finite power series in θ, $\psi(\theta) = \sum a_q \theta^q$. Then interchanging the order of summation and the conditional expectation of $\dot{\psi}(\theta) = \eta(\theta)$, one finds that the terms can be consistently estimated using kernel estimates for densities and derivatives of densities. (These kernel estimates have been studied by other authors and applied in sequence compound decision theory by Yu (1971) and Singh (1974)). Substituting the estimates into the form of $\tau^{*}_{G_n}(X_i)$ results in a compound estimator that is consistent in the Cesaro mean with rate. A family of normal distributions with fixed variance is an example since $\psi(\theta) = \theta^2 / 2$ (when σ^2=1). In addition, cumulant generating functions of the form $\psi(\theta) = a\theta^j$ where j $\geq$ 2 are investigated. It is shown that the integer part of j/2 is necessarily odd.

For Theorem 3.2, the proof starts for ψ that can be expressed as a finite sum $\psi(\theta) = \sum a_q \theta^{-q} + b \ln(-\theta)$. Then interchanging the order of summation and the conditional expectation of $\dot{\psi}(\theta) = \eta(\theta)$ and using a representation of $E[\theta^{-q} | X_i]$ in a form that can be estimated consistently, the same result as in the first case is essentially derived. A family of Gamma distributions serves as an example. In addition, cumulant generating functions of the form $\psi(\theta) = \sum a_q \theta^{-q} + b \ln(-\theta)$ are investigated and some sufficient conditions for such ψ to be a cumulant generating function for some exponential family are given.

Chapter 2

SET COMPOUND DECISION ESTIMATION UNDER ENTROPY LOSS IN DISCRETE EXPONENTIAL FAMILIES

We discussed the set compound decision problem under entropy loss for exponential families in Chapter 1. In this chapter, we specialize this set compound decision problem to a one dimensional discrete exponential family. Under certain conditions, a sequence of set compound decision rules $\underline{t}$ is shown to be asymptotically optimal with rate $n^{-1/2}$. Poisson and negative binomial families serve as examples. A two dimensional example ends the chapter to illustrate how the ideas extend to multidimensional cases.

2.1 Introduction

Consider the exponential family (1.1) with respect to a discrete measure μ on the nonnegative integers such that $g(x) := \mu(\{x\}) > 0, x = 0,1,2,\cdots$. It follows that P_θ has the probability mass function

$$(2.1) \qquad p_\theta(x) = g(x)e^{\theta x - \psi(\theta)}, \qquad x = 0,1,2,\cdots,\theta \in \Theta.$$

Clearly, $\psi(\theta)=\ln\left[\sum_{x=0}^{\infty}e^{\theta x}g(x)\right]$ is strictly increasing in θ. Here the natural parameter space is an interval (-∞, v}, open or closed on the right with $-\infty < v \le +\infty$. (From display (9) of Rainville (1967) on page 111, $v = -\ln[\limsup_{k\to+\infty} \sqrt[k]{g(k)}\,]$.) Consider

(2.2) $$\Theta = A = [v_1, v_2] \subset N^0.$$

Remark 2.1 If (2.2) holds, then the hypotheses of Propositions 1.1 - 1.4 are satisfied so that for any probability distribution G on Θ, the unique Bayes response is given by (1.12), that is , by

(2.3) $$\tau_G(x) = \eta^{-1}[E(\eta(\theta)|\, x)], \;\; x = 0,1,2,\cdots,$$

Moreover, in Proposition 1.4, the Lipschitz constant C_0 can be taken to be $1/\dot{\eta}(v_1)$, that is,

(2.4) $$|\eta^{-1}(a) - \eta^{-1}(b)| \le |a-b|/\dot{\eta}(v_1) \;\text{ for all a, b} \in A.$$

Proof. Since $\Theta=A=[v_1, v_2]$ is a compact subset of $N^0 \subseteq N^{\sim}$ and ψ is differentiable of all orders on N^0, $\eta = \dot{\psi}$ is continuous on Θ. Therefore, $E(\eta(\theta)|\,x)$ exists and is integrable. Also $\eta[\Theta] = [\eta(v_1), \eta(v_2)]$ is a convex set so that (1.11) is satisfied. Also note that $\frac{d}{dy}\eta^{-1}(y) = 1/\dot{\eta}(\eta^{-1}(y))$. Since $\dot{\eta}(\theta) = \ddot{\psi}(\theta) = \mathrm{Var}_\theta(X) > 0$ on N^0, $\dot{\eta}$ is strictly increasing. Thus, $\frac{d}{dy}\eta^{-1}(y) \le 1/\dot{\eta}(\eta^{-1}(\eta(v_1))) = 1/\dot{\eta}(v_1)$. (2.4) follows from the mean value theorem. ◊

We will use the bounds in (2.5) in proofs that follow.

Remark 2.2 If (2.2) holds, then

$$c^{-1}p_{v_1}(x) \le p_\theta(x) \le c p_{v_2}(x),\ x = 0, 1, 2, \ldots,\ \theta \in [v_1, v_2] \tag{2.5}$$

where $c = e^{\psi(v_2)-\psi(v_1)}$. (This is an analog of the inequality on page 1894 of Gilliland (1968).)

Proof. (2.5) follows directly from (2.1) and (2.2). ◊

Suppose that η has the power series representation

$$\eta(\theta) = \sum_{j=1}^{\infty} q_j e^{j\theta},\quad \theta \in \Theta = [v_1, v_2] \tag{2.6}$$

and that the series converges uniformly on Θ. It follows from (2.1) and (2.6) that for any probability distribution G on Θ,

$$\tau_G^*(x) := E(\eta(\theta)|\, x) = \sum_j q_j E(e^{j\theta}|\, x) = \sum_{j=1}^{\infty} q_j \frac{p_G(x+j)}{p_G(x)} \frac{g(x)}{g(x+j)},\ x = 0, 1, 2, \cdots \tag{2.7}$$

where $p_G(x) = \int p_\theta(x)dG(\theta)$. (2.7) expresses the conditional expectation in terms of ratios. In view of the bound (1.20), we turn our attention to the estimation of $\tau_{G_n}^*(X_i) := E_n(\eta(\theta)|\, X_i)$, i=1, 2, $\cdots$, n through estimation of the probabilities in the ratios, namely, the $p_{G_n}(x)$, $x = 0, 1, 2, \cdots$. For this consider

$$\hat{p}(x) = n^{-1}\sum_{j=1}^{n} [X_j = x],\ x = 0, 1, 2, \cdots \tag{2.8}$$

and, for $i = 1, 2, \cdots, n$, $n \ge 2$

$$\hat{p}_i(x) = (n-1)^{-1}\sum_{\substack{j \ne i \\ j=1}}^{n} [X_j = x],\ x = 0, 1, 2, \cdots; \tag{2.9}$$

where square brackets denote indicator functions.

Lemma 2.1 With $c = e^{\psi(v_2)-\psi(v_1)}$,

(2.10) $$E_{\underline{\theta}}\left|\hat{p}(x) - p_{G_n}(x)\right| \le [cp_{v_2}(x)/n]^{1/2},\ n \ge 1$$

(2.11) $$E_{\underline{\theta}}\left|\hat{p}_i(x) - p_{G_n}(x)\right| \le 2[cp_{v_2}(x)/(n-1)]^{1/2}\ ,\ n\ge 2.$$

Proof. Since $E_{\underline{\theta}}(\hat{p}(x)) = p_{G_n}(x)$, LHS (2.10) is no greater than the standard deviation of $\hat{p}(x)$, that is,

(2.12) $$\text{LHS (2.10)} \le n^{-1}\left[\sum_{j=1}^{n} p_{\theta_j}(x)\left(1 - p_{\theta_j}(x)\right)\right]^{1/2}.$$

(2.10) follows from weakening (2.12) and applying the second inequality of (2.5).

To prove (2.11) we triangulate in LHS (2.11) about $E_{\underline{\theta}}(\hat{p}_i(x))$ to get

(2.13) $$\text{LHS (2.11)} \le E_{\underline{\theta}}|\hat{p}_i(x) - E_{\underline{\theta}}(\hat{p}_i(x))| + |E_{\underline{\theta}}(\hat{p}_i(x)) - p_{G_n}(x)|.$$

The first term in RHS (2.13) is bounded by $[cp_{v_2}(x)/(n-1)]^{1/2}$ by applying (2.10) to the set with θ_i deleted. The second term in RHS (2.13) is equal to $|\overline{w} - w_i|/(n-1)$ where $w_j := p_{\theta_j}(x)$, $j = 1, 2, \cdots, n$. Since the $w_j \ge 0$, $j = 1,2,\cdots,n$, $|\overline{w} - w_j|/(n-1) \le \max_j w_j/(n-1)$, so that the second term in RHS (2.13) is bounded by $\max_j p_{\theta_j}(x)/(n-1)$ for $n \ge 2$. Bounding this by its square root and applying the second inequality of (2.5) to the w_j completes the proof of (2.11). ◊

2.2 An Asymptotically Optimal Set Compound Estimator for a One Dimensional Family

With the results of Chapter 1 and the previous section, we are prepared to demonstrate a compound estimator that satisfies (1.15) with rate $O(n^{-1/2})$ uniform in $\underline{\theta}$. Let $(\cdot)_{[a,b]}$ denote retraction to the interval [a, b], that is $(\cdot)_{[a,b]} = (a\vee\cdot)\wedge b$.

Consider the compound estimator $\{\underline{t}\}$ defined by

$$t_i(\underline{X}) = \eta^{-1}(s_i(\underline{X})) \tag{2.14}$$

where

$$s_i(\underline{X}) = \left(\sum_{j=1}^{\infty} q_j \frac{\hat{p}_i(X_i + j)}{\hat{p}_i(X_i)} \frac{g(X_i)}{g(X_i + j)}\right)_{[\eta(v_1),\eta(v_2)]}, \quad i = 1, 2, \ldots, n \tag{2.15}$$

where the q_j, $\hat{p}_i$ are defined by (2.6) and (2.9) and 0/0 :=0. $s_i(\underline{X})$ is an estimator of $\tau^{\bullet}_{G_n}(X_i)$ defined by (2.7) with $G = G_n$.

Theorem 2.1 Let $\{P_\theta : \theta \in \Theta\}$ be the discrete exponential family defined by (2.1) and suppose (2.2) holds. Also suppose that (2.6) holds with uniform convergence on $\Theta=[v_1, v_2]$ and that

$$\sum_{x=0}^{\infty} \sum_{j=1}^{\infty} |q_j| \left(e^{v_2 j} p_{v_2}(x) \frac{g(x)}{g(x+j)}\right)^{1/2} < \infty . \tag{2.16}$$

Let $\{\underline{t}\}$ be defined by (2.14) and (2.15). Then

$$\sup_{\underline{\theta}} |D_n(\underline{\theta}, \underline{t})| = O(n^{-1/2}). \tag{2.17}$$

Proof. By Remark 2.1 and Proposition 1.4,

$$\left|D_n(\underline{\theta}, \underline{t})\right| \le (B_0 / \dot{\eta}(v_1))n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} \left|s_i - \tau^{\bullet}_{G_n}(X_i)\right|, \text{ for all } \underline{\theta} \text{ and n.} \tag{2.18}$$

The conditional expectation $\tau^{\bullet}_{G_n}(X_i) = a / b$, where

$a := \sum_{j=1}^{\infty} q_j p_{G_n}(X_i + j)[g(X_i)/g(X_i + j)]$ and $b := p_{G_n}(X_i)$, takes values in $[\eta(v_1), \eta(v_2)]$.

Letting $A := \sum_{j=1}^{\infty} q_j \hat{p}_i(X_i + j)[g(X_i)/g(X_i + j)]$ and $B := \hat{p}_i(X_i)$ we see that

$$(2.19) \qquad |s_i - \tau^{*}_{G_n}(X_i)| \le |\frac{A}{B} - \frac{a}{b}| \wedge D$$

where $D = \eta(v_2) - \eta(v_1)$. Applying Singh-Datta Lemma (1.21) and weakening the result we have

$$(2.20) \qquad |s_i - \tau^{*}_{G_n}(X_i)|$$

$$\le \frac{1}{p_{G_n}(X_i)} \{ \sum_{j=1}^{\infty} |q_j| |\hat{p}_i(X_i + j) - p_{G_n}(X_i + j)| \frac{g(X_i)}{g(X_i + j)}$$

$$+ (2\eta(v_2) - \eta(v_1)) |\hat{p}_i(X_i) - p_{G_n}(X_i)| \}.$$

We first consider the Cesaro mean expectation of the last term in RHS(2.20). Applying (2.11) to each summand shows that for $n \ge 2$,

$$(2.21) \qquad n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} \frac{|\hat{p}_i(X_i) - p_{G_n}(X_i)|}{p_{G_n}(X_i)} \le \frac{2c^{1/2}}{\sqrt{n-1}} \frac{1}{n} \sum_{i=1}^{n} E_{\theta_i} \left(\frac{p^{1/2}_{v_2}(X_i)}{p_{G_n}(X_i)} \right).$$

Since $\frac{1}{n} \sum_{i=1}^{n} E_{\theta_i} = E_{G_n}$, RHS(2.21) $\le 2c^{1/2} \sum_{x=0}^{\infty} p^{1/2}_{v_2}(x) / \sqrt{n-1}$. Since v_2 is in the interior of the natural parameter space, $p^{1/2}_{v_2}(x)$ is summable with respect to x as indicated in Gilliland (1966, page 24). (This is easily seen to be the case because there is a number b larger than v_2 such that $\sum_{x=0}^{\infty} e^{bx} g(x) < \infty$. Thus, $\sup_x [e^{bx} g(x)] < \infty$, so that $e^{\psi(v_2)/2} p^{1/2}_{v_2}(x) =$

$e^{(v_2-b)x/2}[e^{bx}g(x)]^{1/2} \le \{\sup_x[e^{bx}g(x)]\}^{1/2}e^{(v_2-b)x/2}$ which is the summable.) Therefore we have RHS(2.21) $=O(n^{-1/2})$ uniform in $\underline{\theta}$.

We now consider the Cesaro mean expectation of the first term in RHS(2.20). Applying (2.11) to each summand, using an argument similar to that used above, and using $p_{v_2}(x+j)=e^{v_2 j}p_{v_2}(x)g(x+j)/g(x)$ shows that for $n\ge 2$,

$$\text{(2.22)}\quad n^{-1}\sum_{i=1}^{n}E_{\underline{\theta}}\sum_{j=1}^{\infty}|q_j|\frac{|\hat{p}_i(X_i+j)-p_{G_n}(X_i+j)|}{p_{G_n}(X_i)}\frac{g(X_i)}{g(X_i+j)}\le 2c^{1/2}\text{LHS}(2.16)/\sqrt{n-1}$$

Using (2.16) we see that RHS(2.22) is $O(n^{-1/2})$ uniform in $\underline{\theta}$ completing the proof. ◊

Remark 2.3 (2.16) does not follow from (2.2). Consider the example with $g(2k)=e^{2k}$, $g(2k+1)=(2k+1)^{-3}$, $k=0,1,2,\cdots$. Then, N=(-∞, -1). Take $v_2=-\frac{3}{2}$ which is smaller than -1, but

$$\sum_{k=0}^{\infty}p^{\frac{1}{2}}_{-\frac{3}{2}}(2k)[\frac{g(2k)}{g(2k+1)}]^{\frac{1}{2}}=\sum_{k=0}^{\infty}e^{-\frac{3}{4}(2k)}(2k+1)^{\frac{3}{2}}e^{2k}=\infty.$$

Remark 2.4 If $\psi(\cdot)$ can be represented as a series of powers of exponential function $e^{\cdot}$, i.e., $\psi(\theta)=\sum_{j=1}^{\infty}q_je^{j\theta}$ with $q_j\ge 0$, then the dominating measure μ indicated in (1.1) has to be discrete. In fact $g(0)=e^{\psi(-\infty)}$, $g(1)=e^{\psi(-\infty)}q_1$, ... , $g(k)=k^{-1}[\sum_{j=0}^{k-1}g(j)q_{(k-j)}]$, by the argument in Theorem 51 of Rainville (1967) (see page 129). But if $q_{j_0}<0$ for some integer $j_0>0$, the conclusion is not true. For example, $\psi(\theta)=\frac{\theta^2}{2}$ in N(θ, 1),

$\dot{\psi}(\theta) = \theta = \sum_{j=1}^{\infty}(-1)^{j+1}\frac{(e^{\theta}-1)^j}{j}$, for $-\infty<\theta\leq\ln 2$, μ is not discrete, actually μ is absolutely continuous with respect to the Lebesgue measure.

2.3 Poisson and Negative Binomial Distributions

Poisson Case. Consider the Poisson family with the mean λ in a closed interval $[\lambda_1, \lambda_2] \subseteq (0, \infty)$. It is transformed to standard form by letting $\theta = \ln\lambda$ resulting in

$$p_\theta(x) = \frac{e^{\theta x - \psi(\theta)}}{x!} \quad x=0,\ 1,\ 2,\cdots;\ \theta \in [\ln\lambda_1, \ln\lambda_2] \tag{2.23}$$

where $\psi(\theta) = e^\theta$. In this case $\dot{\psi}(\theta) = \eta(\theta) = e^\theta$ and the loss function (1.7) is

$$L(\theta, a) = (\theta - a)e^\theta - e^\theta + e^a.$$

In the power series representation (2.6), $q_1 = 1$ and $q_j = 0$ for $j \geq 2$. Thus,

LHS(2.16) $=$ $\lambda_2^{1/2}\sum_{x=0}^{\infty}[(x+1)p_{v_2}(x)]^{1/2}$ which is easily shown to be finite. Thus, this Poisson case is covered by Theorem 2.1.

Negative Binomial Case. Consider the family with probability mass functions

$$(1-p)^\alpha p^x \frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)},\ x\in 0,1,2,\cdots\ ;\ 0 < p_1 \leq p \leq p_2 < 1$$

where $\alpha>0$ is fixed. It is transformed into standard form by letting $\theta = \ln p$ resulting in

$$p_\theta(x) = e^{\theta x - \psi(\theta)}\frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)},\ x=0,1,2,\cdots;\ \ln p_1 \leq \theta \leq \ln p_2. \tag{2.24}$$

where $\psi(\theta) = -\alpha\ln(1-e^\theta)$. In this case $\dot{\psi}(\theta) = \eta(\theta) = \alpha e^\theta/(1-e^\theta)$ and the loss function (1.7) is

$$L(\theta,a)=(\theta-a)\frac{\alpha e^{\theta}}{1-e^{\theta}}+\alpha\ln(1-e^{\theta})-\alpha\ln(1-e^{a}).$$

In the power series representation (2.6), $q_j=\alpha$, $j=1,2,\cdots$. Thus

$$\text{LHS}(2.16)=\alpha\sum_{x=0}^{\infty}\sum_{j=1}^{\infty}e^{\frac{v_2}{2}j}p_{v_2}^{\frac{1}{2}}(x)=\frac{\alpha\sqrt{p_2}}{1-\sqrt{p_2}}\sum_{x=0}^{\infty}(1-p_2)^{\frac{\alpha}{2}}p_2^{\frac{x}{2}}\sqrt{\frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)}}$$

which is finite since $\frac{\Gamma(x+\alpha)}{x^{\alpha}\Gamma(x+1)}\to 0$ as $x\to\infty$. Hence this negative binomial case is covered by Theorem 2.1.

An interesting observation is that the Bayes estimator with geometric distribution (α=1) has the simple form

$$\tau_G(x)=\ln\left(\frac{P_G(X\geq x+1)}{P_G(X\geq x)}\right) \tag{2.25}$$

by Remark 1.4, Proposition 1.2 and the following fact:

$$\int e^{\theta(x+1)}dG=P_G(X\geq x+1).$$

It is not difficult to prove the above fact. First note that

$$p_G(x)=\int e^{\theta x+\ln(1-e^{\theta})}dG=\int e^{\theta x}dG-\int e^{\theta(x+1)}dG$$

so that by telescoping,

$$\int e^{\theta(x+1)}dG=1-\sum_{j=0}^{x}p_G(j)=P_G(X\geq x+1).$$

One can estimate the ratio in $\tau_{G_n}(x)$ using natural estimates of the tail probabilities in constructing asymptotically optimal compound estimators in the geometric case.

Remark 2.5 (i) No one has done the set compound problem with discrete exponential families and continuous parameter spaces. The achievement in this chapter is that instead of the usual squared error loss, the entropy loss is used and with this loss and a different proof, set compound asymptotic estimators with rate $n^{-1/2}$ are established for some discrete exponential families.

(ii) For Poisson exponential family (2.23), squared error loss $(x-e^{\theta})^2$, $\Theta=(-\infty, \ln\lambda_1]$ (which is unbounded) and compound estimator $\underline{t}$ with

$$t_i(\underline{X}) = [(X_i+1)\frac{\hat{p}_i(X_i+1)}{\hat{p}_i(X_i)}]\wedge\lambda_2,$$

we can show

$$\sup_{\underline{\theta}\in\Theta^n}|D_n(\underline{\theta},\underline{t})| = \sup_{\underline{\theta}\in\Theta^n}|\frac{1}{n}\sum_{i=1}^{n}E_{\underline{\theta}}[(t_i(\underline{X})-e^{\theta_i})^2-(t^*-e^{\theta_i})^2]| = O(n^{-\frac{1}{2}}), \tag{2.26}$$

where t^* is the Bayes estimator $E(e^{\theta}|x)$ corresponding to squared error loss and prior G_n. This is a stronger result than what follows from Remark 1.7 since Θ is unbounded.

(iii) For geometric exponential family, squared error loss $(x-e^{\theta})^2$, noncompact $\Theta=(-\infty, \ln p_1]$ and compound estimator $\underline{t}$ with

$$t_i(\underline{X}) = [\frac{\hat{p}_i(X_i+1)}{\hat{p}_i(X_i)}]\wedge p_2,$$

(2.26) can be shown. This is a stronger result than what follows from Remark 1.7 since Θ is unbounded.

2.4 A Two Dimensional Example

The following is a two dimensional example.

Let X be distributed as Poisson (λ) and conditional on $X=x$, let Y be distributed as x+Z, where Z has the negative binomial distribution (x+1, p). Then (X, Y) has a 2-dimensional exponential distribution

(2.27) $$\binom{y}{x}(1-p)^{x+1}p^{y-x}\frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{for } y \geq x,$$

$$0 < p_1 \leq p \leq p_2 < 1, \qquad 0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty.$$

Let $\theta_1 = \ln[\frac{(1-p)\lambda}{p}]$ and $\theta_2 = \ln p$. Then the distribution above can be transformed into the standard form:

$$p_{\theta_1,\theta_2}(x,y) = \exp[\theta_1 x + \theta_2 y - \frac{e^{\theta_1+\theta_2}}{1-e^{\theta_2}} + \ln(1-e^{\theta_2})]\binom{y}{x} / x!, \ y \geq x.$$

The natural parameter space N is the open set $R\times(-\infty,0)$. Let $g(x,y)= \binom{y}{x} / x!$ for $y \geq x$ throughout the following.

Define T: $(0,1)\times R_+ \to R^2$ by $T(p,\lambda) = (\ln\frac{(1-p)\lambda}{p}, \ln p)$. Then T is continuous, which yields the compactness of $\Theta := T\{[p_1,p_2]x[\lambda_1,\lambda_2]\}$.

In this example let the action space A=Θ. Now

$$\eta_1(\theta) = \dot{\psi}_1(\theta) = E_\theta(X) = \lambda = \frac{e^{\theta_1+\theta_2}}{1-e^{\theta_2}} = \sum_{j=1}^{\infty} e^{\theta_1+j\theta_2}$$

and

$$\eta_2(\theta) = \dot{\psi}_2(\theta) = E_\theta(Y) = \frac{p+\lambda}{1-p} = \sum_{j=1}^{\infty}[je^{\theta_1+j\theta_2} + e^{j\theta_2}].$$

$\eta[\Theta] = [(\zeta_1,\zeta_2):\lambda_1 \leq \zeta_1 \leq \lambda_2, \frac{p_1+\zeta_1}{1-p_1} \leq \zeta_2 \leq \frac{p_2+\zeta_1}{1-p_2}]$, which is a polytope in R^2. Thus $\eta[\Theta]$ is convex. By using Remark 1.5 we see that the hypotheses of Propositions 1.1 and

1.2 are satisfied. By Proposition 1.2 the unique Bayes estimator of θ with the entropy loss is

$$\tau_G(w) := (\tau_{1G}, \tau_{2G}) = \eta^{-1}[E(\eta(\theta)|w)],$$

where w := (x, y). Interchange the summation and conditional expectation, obtain

$$E(\eta_1(\theta)|w) = \sum_{j=1}^{\infty} \frac{p_G(x+1, y+j)}{p_G(x,y)} \frac{g(x,y)}{g(x+1, y+j)},$$

$$E(\eta_2(\theta)|w) = \sum_{j=1}^{\infty} [j \frac{p_G(x+1, y+j)}{p_G(x,y)} \frac{g(x,y)}{g(x+1,y+j)} + \frac{p_G(x,y+j)}{p_G(x,y)} \frac{g(x,y)}{g(x+1,y+j)}], \text{ for } y \geq x.$$

Let

$$\hat{p}(x,y) = n^{-1}\sum_{j=1}^{n}[X_j = x, Y_j = y] \text{ and } \hat{p}_i(x,y) = (n-1)^{-1}\sum_{j \neq i}[X_j = x, Y_j = y]$$

for $i = 1, \cdots, n$, where as before square brackets denote the indicator function.

More widely, let $\{v_k = T(\lambda_{0k}, p_{0k}) : k = 1, \cdots, m\} \subset N$ and $\Theta \subset$ convex hull$\{v_k\} \subset N^0$. Similar to one dimensional case the following inequalities can be verified:

$$\text{(2.28)} \qquad E_{\underline{\theta}}|(\hat{p} - p_{G_n})(x,y)| \leq n^{-\frac{1}{2}} e^{\frac{\psi^* - \psi_*}{2}} \sum_{k=1}^{m} p^{\frac{1}{2}}{}_{v_k}(x,y),$$

$$E_{\underline{\theta}_i}|(\hat{p}_i - p_{G_n})(x,y)| \leq 2n^{-\frac{1}{2}} e^{\frac{\psi^* - \psi_*}{2}} \sum_{k=1}^{m} p^{\frac{1}{2}}{}_{v_k}(x,y)$$

by using (17) on page 13 of Zhu (1992) and $(a+b)^{\frac{1}{2}} \leq a^{\frac{1}{2}} + b^{\frac{1}{2}}$ for all a, b ≥0. Establish a set compound estimator based on observations of $\underline{W} := [(X_1, Y_1), \cdots, (X_n, Y_n)]$ for i= 1 , $\cdots$, n as follows:

$$t_i(\underline{W}) := (t_{1i}(\underline{W}), t_{2i}(\underline{W})) = \eta^{-1}\{(\sum_{j=1}^{\infty} \frac{\hat{p}_i(X_i+1, Y_i+j)}{\hat{p}_i(X_i, Y_i)} \frac{g(X_i, Y_i)}{g(X_i+1, Y_i+j)},$$

$$\sum_{j=1}^{\infty}[j \frac{\hat{p}_i(X_i+1, Y_i+j)}{\hat{p}_i(X_i, Y_i)} \frac{g(X_i, Y_i)}{g(X_i+1, Y_i+j)} + \frac{\hat{p}_i(X_i, Y_i+j)}{\hat{p}_i(X_i, Y_i)} \frac{g(X_i, Y_i)}{g(X_i, Y_i+j)}])[\eta[\Theta]]\}$$

and $\underline{t} := (t_1, t_2, \cdots, t_n)$.

Then $\{\underline{t}\}$ is an asymptotically optimal set compound estimator with (2.8) satisfied under the condition that

$$\vee\lambda_{0k} < \frac{1 - \sqrt{\vee p_{0k}}}{1 + \sqrt{\vee p_{0k}}}. \tag{2.29}$$

The proof follows.

From (1.16)

$$\begin{aligned} |D_n(\underline{\theta}, \underline{t})| &\le B_0 n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} |t_i(\underline{W}) - \tau_{G_n}(X_i, Y_i)| \\ &\le B_0 n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} \{|t_{1i}(\underline{W}) - \tau_{1G_n}(X_i, Y_i)| + |t_{2i}(\underline{W}) - \tau_{2G_n}(X_i, Y_i)|\}. \end{aligned} \tag{2.30}$$

Let C be a constant in the following. Use (2.28), Mean Value Theorem to the functions η^{-1} and η^{-1} and Singh-Datta Lemma (see Section 1.3),

(2.31)

LHS(1.16)

$$\le Cn^{-1}\sum_{i=1}^{n} E_{\underline{\theta}} p_{G_n}^{-1}(X_i,Y_i)[|\sum_{j=1}^{\infty}\hat{p}_i(X_i+1,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}$$

$$-\sum_{j=1}^{\infty} p_{G_n}(X_i+1,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}|+C|\hat{p}_i(X_i,Y_i)-p_{G_n}(X_i,Y_i)|$$

$$+|\sum_{j=1}^{\infty} j\hat{p}_i(X_i+1,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}-\sum_{j=1}^{\infty} jp_{G_n}(X_i+1,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}|$$

$$+|\sum_{j=1}^{\infty}\hat{p}_i(X_i,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}-\sum_{j=1}^{\infty} p_{G_n}(X_i,Y_i+j)\frac{g(X_i,Y_i)}{g(X_i+1,Y_i+j)}|]$$

$$\le Cn^{-\frac{1}{2}}\sum_{y\ge x}[\sum_{j=1}^{\infty} j\sum_{k=1}^{m} p^{\frac{1}{2}}{}_{v_k}(x+1,y+j)\frac{g(x,y)}{g(x+1,y+j)}+\sum_{j=1}^{\infty}\sum_{k=1}^{m} p^{\frac{1}{2}}{}_{v_k}(x,y+j)\frac{g(x,y)}{g(x,y+j)}].$$

Furthermore by (2.27) and $\frac{(y+j-x-1)!}{(y+j)!}\le 1$,

$$\sum_{y\ge x}\sum_{j=1}^{\infty} j\sum_{k=1}^{m} p^{\frac{1}{2}}{}_{v_k}(x+1,y+j)\frac{g(x,y)}{g(x+1,y+j)}$$

$$=\sum_{y\ge x}\sum_{j=1}^{\infty} j\sum_{k=1}^{m}[\binom{y+j}{x+1}(1-p_{0k})^{x+2}p_{0k}{}^{y+j-x-1}\frac{e^{-\lambda_{0k}}\lambda_{0k}{}^{x+1}}{(x+1)!}]^{1/2}\frac{\binom{y}{x}\Big/x!}{\binom{y+j}{x+1}\Big/(x+1)!}$$

$$\le C\sum_{k=1}^{m}\sum_{j=1}^{\infty} jp_{0k}{}^{1/2}\sum_{y=0}^{\infty}(1+y)p_{0k}{}^{y/2}\sum_{x=0}^{y}\left(\frac{(1-p_{0k})\lambda_{0k}}{p_{0k}}\right)^{(x+1)/2}\binom{y}{x}\left(\frac{(y+j-x-1)!}{(y+j)!}\right)^{1/2}$$

$$\le C\sum_{k=1}^{m}\sum_{j=1}^{\infty} jp_{0k}{}^{1/2}\sum_{y=0}^{\infty}(1+y)p_{0k}{}^{y/2}[\sqrt{(1-p_{0k})\lambda_{0k}/p_{0k}}+1]^{y}$$

$$\le C\sum_{k=1}^{m}\sum_{j=1}^{\infty} jp_{0k}{}^{1/2}\sum_{y=0}^{\infty}(1+y)[\sqrt{(1-p_{0k})\lambda_{0k}}+\sqrt{p_{0k}}]^{y},$$

which is finite by the fact that $\sqrt{(1-p_{0k})\lambda_{0k}}+\sqrt{p_{0k}}<1$ for $k=1,\cdots,m$ from the condition (2.29).

In the same way, the second sum in RHS(2.31) is finite. Combined with above, the proof is finished.

At last take a look at the condition (2.29). If $\vee p_{0k} = r^{-2}$(r an integer), then $\vee\lambda_{0k} \le (r-1)/(r+1)$.

Remark 2.6 If $\Theta = T\{[p_1, p_2] \times [\lambda_1, \lambda_2]\}$, then we can take $v_1 = T(\lambda_1, p_2)$, $v_2 = (\ln\frac{(1-p_2)\lambda_1}{p_2}, \ln p_1), v_3 = T(\lambda_2, p_1)$ and $v_4 = (\ln\frac{(1-p_1)\lambda_2}{p_1}, \ln p_2)$ and $\Theta \subset$ convex hull$\{v_1, v_2, v_3, v_4\}$. Actually $v_2 = T(\lambda_1^\bullet, p_1)$ and $v_4 = T(\lambda_2^\bullet, p_2)$ with $\lambda_1^\bullet = \frac{(1-p_2)p_1}{p_2(1-p_1)}\lambda_1$ and $\lambda_2^\bullet = \frac{(1-p_1)p_2}{p_1(1-p_2)}\lambda_2$. Condition (2.31) becomes

$$\lambda_2^\bullet < \frac{1-\sqrt{p_2}}{1+\sqrt{p_2}}.$$

Chapter 3

SET COMPOUND DECISION ESTIMATION UNDER ENTROPY LOSS IN CONTINUOUS ONE-DIMENSIONAL EXPONENTIAL FAMILIES

Set compound estimation under entropy loss with continuous one-dimensional exponential families is pursued in this chapter. Before we get into details, let us look at (1.12). If we can represent $E(\eta(\theta)|x)$ in some form of $p_G(x)$, as we did in Chapter 2, then we will be able to give a compound estimator. In continuous case, we use the kernel density estimation considered in Singh (1974) to construct compound estimators. We start our work with some notations that will be used in the chapter. Sections 3.2 and 3.3 discuss continuous exponential families with cumulant generating functions of the form $\sum_{q=0}^{k} a_q\theta^q$ and $\sum_{q=1}^{k} a_q\theta^{-q} + b\ln(-\theta)$ respectively. Normal and Gamma distributions are covered by these cases.

Recall that our distribution has density (1.1) with k=1 in the one-dimensional case. For simplicity we will drop the ~ and let

$$p_\theta(x) = e^{\theta x - \psi(\theta)} \tag{3.1}$$

denote the density of P_θ with respect to μ. In the chapter we assume that μ is dominated by Lebesgue measure λ with density denoted by u. Furthermore, we assume that u is positive iff x is larger than c for some $c \geq -\infty$. Without loss of generality we assume that the variable x and the random variables $X_1, X_2, \cdots$ take values in (c, ∞) in what follows. In addition we assume throughout that

(3.2) $$\Theta = [\alpha, \beta].$$

We continue the convention 0/0 :=0.

Remark 3.1

(3.3) $$p_\theta \leq e^{\psi^* - \psi_*}(p_\alpha + p_\beta)$$

where $\psi^* = \sup\{\psi(\theta) | \theta \in \Theta\}$ and $\psi_* = \inf\{\psi(\theta) | \theta \in \Theta\}$.

Proof. (3.3) follows directly from the inequality: $e^{\theta x} \leq e^{\alpha x} + e^{\beta x}$ for all x when $\theta \in [\alpha, \beta]$.

3.1 Kernel Density Estimation

The kernel functions K satisfy K(s) =0 if s ∉ [0,1] and $\int K^2(s) ds < \infty$. We will consider K that satisfy

(3.4) $$(i!)^{-1} \int s^i K(s) ds = [i = \upsilon] \quad , \quad i = 0, \cdots, r-1$$

for given r and $\upsilon < r$. The kernels satisfying (3.4) for $\upsilon = 0$ are used in the estimation of the density and those for $\upsilon > 0$ are used in the estimation of its υth derivative.

Such K(s) exists since there is a linear functional on $L^2[0,1]$ which satisfies (3.4) (see Rudin (1973), Theorem 3.5). Let

$$\hat{p}^{(\upsilon)}(x) := \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h^{\upsilon+1}} K\left(\frac{X_j - x}{h}\right) \frac{1}{u(X_j)},$$

$x > c$ for $\upsilon = 0, 1, \ldots, r-1$, $n \geq 1$ and $0 < h < 1$. We let $\hat{p}_i^{(\upsilon)}(x)$ denote the (n-1) term average with the ith term deleted, i = 1, 2, …, n. The upper index (υ) will be omitted in case $\upsilon = 0$.

In the following let $T = e^{\beta h}\int K^2(s)ds$, $T_r = [(r-1)!]^{-1} s_0^{r-1}\int |K(s)||s^r| ds$ and $s_0 = |\alpha| \vee |\beta|$. Also let $p_i^{(k)}(x) = \frac{1}{n-1}\sum_{j\neq i} p_{\theta_j}^{(k)}(x)$, $p_{\theta_j}^{(0)}(x) = p_{\theta_j}(x)$, $p_i(x) = p_i^{(0)}(x)$,

$D_i(n) = \sqrt{\frac{Tp_i(x)}{(n-1)h^{2k+1}u(x,h)}} + T_r(e^{s_0}-1)h^{r-k}p_i(x)$ and $D_0(n) = \sqrt{\frac{Tp_{G_n}(x)}{(n-1)h^{2k+1}u(x,h)}}$ $+ T_r(e^{s_0}-1)h^{r-k}p_{G_n}(x)$. The upper index (k) indicates the kth derivative with respect to x there. The next lemma concerns the variance and bias of $\hat{p}^{(\upsilon)}$ as an estimator for the kth derivative $p_{G_n}^{(\upsilon)}$ of the empirical mixture $p_{G_n} = \frac{1}{n}\sum_{i=1}^{n} p_{\theta_i}$.

Lemma 3.1 Let u(x,h) = essinf{u(t): $x < t \leq x+h$ } with respect to λ. Then

(i) $\mathrm{Var}_{\underline{\theta}}(\hat{p}^{(\upsilon)}(x)) \leq \frac{Tp_{G_n}(x)}{nh^{2\upsilon+1}u(x,h)}$,

(ii) $|E_{\underline{\theta}}\hat{p}^{(\upsilon)}(x) - p_{G_n}^{(\upsilon)}(x)| \leq T_r(e^{s_0}-1)h^{r-\upsilon}p_{G_n}(x)$,

(3.6) (iii) $E_{\underline{\theta}}|\hat{p}^{(\upsilon)}(x) - p_{G_n}^{(\upsilon)}(x)| \leq D_0(n)$.

Proof. Note that $E|W| \leq [\mathrm{Var}(W)]^{\frac{1}{2}} + |EW|$ by triangulation at E(W) and moment inequality for E|W-EW| . Thus, (iii) follows from (i) and (ii). Singh (1974, page 59) proved (i) and (ii). The bound in (ii) is slightly different from the Singh's bound because of use of the inequality $|e^x - 1| \leq |e^a - 1||x/a|$ if $|x| \leq a$ instead of the Mean Value Theorem. ◊

Lemma 3.2 For $n \geq 2$,

$$(3.7) \qquad E_{\underline{\theta}}|\hat{p}_i^{(\upsilon)}(x) - p_{G_n}^{(\upsilon)}(x)| \leq D_i(n) + 2s_0^{\upsilon}e^{\psi^{*}-\psi_{*}}(p_{\alpha}(x) + p_{\beta}(x))/(n-1).$$

Proof. Triangulate LHS(3.7) at $p_i^{(\upsilon)}(x)$. Use the fact $|\overline{w} - w_i|/(n-1) \leq 2\max_j |w_j|/(n-1)$

for $w_j := p_{\theta_j}^{(\upsilon)}(x)$ and (3.6) for the following second inequality and (3.3) for the third

inequality below

$$\begin{aligned}
\text{LHS}(3.7) &\leq E_{\underline{\theta}}|\hat{p}_i^{(\upsilon)}(x) - p_i^{(\upsilon)}(x)| + |p_i^{(\upsilon)}(x) - p_{G_n}^{(\upsilon)}(x)| \\
&\leq D_i(n) + 2\max_{j \neq i} p_{\theta_j}^{(\upsilon)}(x)/(n-1) \\
(3.8) \qquad &\leq D_i(n) + 2s_0^{\upsilon}e^{\psi^{*}-\psi_{*}}(p_{\alpha}(x) + p_{\beta}(x))/(n-1) \quad .
\end{aligned}$$

The proof is finished. ◊

Lemma 3.3 Assume for a positive number h_0 with $h_0 < 1$,

$$(3.9) \qquad C := \sup_{\theta \in [\alpha,\beta]} \left\{ \int \frac{p_{\theta}^{1/2}(x)}{u^{1/2}(x,h_0)} d\mu(x) \right\} < \infty.$$

Let $U_{\upsilon}(x)$ be RHS(3.7) with $\upsilon < r$. Then

$$(3.10) \qquad \begin{aligned} \int U_{\upsilon}(x)d\mu(x) \leq \frac{4s_0^{\upsilon}e^{\psi^{*}-\psi_{*}}}{(n-1)} &+ 2C\sqrt{T}e^{\frac{1}{2}(\psi^{*}-\psi_{*})}(n-1)^{-\frac{1}{2}}h^{-(\upsilon+\frac{1}{2})} \\ &+ T_r(e^{s_0} - 1)h^{r-\upsilon} \end{aligned}$$

for $0 < h \leq h_0$ and $n \geq 2$.

Proof. The first and last terms in RHS(3.10) follow directly from integration with respect to μ. The middle term in RHS(3.10) follows from integration with respect to μ applied to the first term in $D_i(n)$ after noting that

$$\sqrt{\frac{p_i(x)}{u(x,h)}} \leq e^{1/2(\psi^{*}-\psi_{*})}\frac{(p_{\alpha}^{1/2}(x) + p_{\beta}^{1/2}(x))}{u^{1/2}(x,h_0)},$$

which follows from (3.3), the fact that u(x,h) is monotone in h, and the inequality

$(a+b)^{1/2} \le a^{1/2} + b^{1/2}$ for a, b ≥ 0. ◊

3.2 Case $\psi(\theta)=\sum_{q=0}^{k} a_q \theta^q$ (k ≥2 and $\{a_q\}$ are constants)

Example 3.1 For the normal distribution family $N(\theta,1)$, $\psi(\theta)=\dfrac{\theta^2}{2}$.

We first establish a set compound estimator. We assume

$$\Theta = [\alpha, \beta] \subseteq N^{0}. \tag{3.11}$$

By Proposition 1.4, for any compound estimator

$$t_i(\underline{X}) = \eta^{-1}[s_i(\underline{X})],\ i=1, \ldots, n. \tag{3.12}$$

$$|D_n(\underline{\theta}, \underline{t})| \le B_0 C_0 n^{-1} \sum_{i=1}^{n} E_{\underline{\theta}} |s_i - \tau^{*}_{G_n}(X_i)|, \tag{3.13}$$

Because $\psi(\theta)=\sum_{q=0}^{n} a_q \theta^q$ and $\eta = \dot{\psi}$, $\tau^{*}_{G}(x) := E(\eta(\theta)|x) = \sum_{q=1}^{k} q a_q E[\theta^{q-1}|x]$. With density

(3.1) it is easy to show that $E[\theta^{q-1}|x] = p_G^{(q-1)}(x)/p_G(x)$ so that

$$\tau^{*}_{G_n}(x) = a_1 + \sum_{q=2}^{k} q a_q \frac{p_{G_n}^{(q-1)}(x)}{p_{G_n}(x)}. \tag{3.14}$$

We consider

$$s_i(\underline{X}) := [a_1 + \sum_{q=2}^{k} q a_q \frac{\hat{p}_i^{(q-1)}(X_i)}{\hat{p}_i(X_i)}]_{[\eta(\alpha),\eta(\beta)]},\ i = 1,2,\cdots,n. \tag{3.15}$$

Theorem 3.1 Suppose that $\psi(\theta)=\sum_{q=0}^{k} a_q \theta^q$ and that (3.11) and (3.9) hold. Then $\{\underline{t}\}$ defined by (3.12) and (3.15) with choice $h = n^{-\delta/(2r+1)}$ is asymptotically optimal with rate $n^{-\gamma}$ where $\gamma = \dfrac{r-k+1}{2r+1}$.

Proof. From (3.13) and the Singh-Datta inequality (Section 1.3) with

$a := \sum_{q=1}^{k} q a_q p_{G_n}^{(q-1)}(X_i)$, $b := p_{G_n}(X_i)$, $A := \sum_{q=1}^{k} q a_q \hat{p}_i^{(q-1)}(X_i)$ and $B := \hat{p}_i(X_i)$

$$|s_i - \tau_{G_n}^{*}(X_i)| \le \left|\frac{A}{B} - \frac{a}{b}\right| \wedge D \le \frac{1}{b}\{|A-a| + (|\frac{a}{b}| + D)|B-b|\} \tag{3.16}$$

where $D=\eta(\beta)-\eta(\alpha)$. Here $|a/b| \in [\eta(\beta), \eta(\alpha)]$,

$$|A-a| \le \sum_{q=2}^{k} q|a_q||\hat{p}_i^{(q-1)}(X_i) - p_{G_n}^{(q-1)}(X_i)| \tag{3.17}$$

and

$$|B-b| = |\hat{p}_i(X_i) - p_{G_n}(X_i)|. \tag{3.18}$$

By applying (3.13), (3.16)-(3.18), Lemmas 3.2 and 3.3,

$$\begin{aligned} \sup_{\underline{\theta}} |D_n(\underline{\theta}, \underline{t})| \le B_0 C_0 \{ & \sum_{q=2}^{k} \{q|a_q|[\frac{4 s_0^{q-1} e^{\psi^{*}-\psi_{*}}}{(n-1)} + 2C\sqrt{T} e^{\frac{1}{2}(\psi^{*}-\psi_{*})}(n-1)^{-\frac{1}{2}} h^{\frac{1}{2}-q} \\ & + T_r(e^{s_0}-1)h^{r-q+1}]\} + [|\eta(\alpha)| \vee |\eta(\beta)| + \eta(\beta) - \eta(\alpha)][\frac{4e^{\psi^{*}-\psi_{*}}}{(n-1)} \\ & + 2C\sqrt{T} e^{\frac{1}{2}(\psi^{*}-\psi_{*})}(n-1)^{-\frac{1}{2}} h^{-\frac{1}{2}} + T_r(e^{s_0}-1)h^{r}]\}. \end{aligned} \tag{3.19}$$

For the choice of $h = n^{-\delta}$ with $\delta = \frac{1}{2r+1}$,

$$\mathrm{RHS}(3.7) \le C_1 n^{-\gamma} \quad ,$$

where C_1 is a constant and $\gamma = \frac{r-k+1}{2r+1}$. ◊

Remark 3.2 (i) If $\psi(\theta) = \sum_{q=0}^{k} a_q \theta^q$ ($k \ge 2$ and $\{a_q\}$ are constants), then N is closed because $\int e^{\theta x} d\mu = e^{\sum a_q \theta^q}$. If $\theta_n \to \alpha$, a finite boundary point, $\int \underline{\lim}_n e^{\theta_n x} d\mu \le$

$\underline{\lim}_n e^{\Sigma a_q \theta_n^q} = e^{\Sigma a_q \alpha^q} < \infty$ by Fatou Lemma. Note that $E_\theta(X^m) = (e^{\Sigma a_q \theta^q})^{(m)}$, $E_\theta(X^m)$ exists and is continuous of θ on N (m> 0) by the same argument before. And also $0 \notin N$ if $a_2 = 0$. If $\psi(\theta) = a\theta^k$ (k > 1), then a > 0 for all even integers k >1. If k is odd, aθ is nonnegative for all θ belonging to N if N^0 is not empty.

(ii) In the above case, *i.p.*(k/2) (integer part) can not be even. The reason is as follows: assume μ is the corresponding measure to the cumulant generating function $\psi(\theta) = a\theta^k$. Take $\theta_0 \in N_\mu^0$, the interior of the natural parameter space for μ and define $d\mu' = e^{\theta_0 x} d\mu$. Then μ' is a finite measure. $a(\theta + \theta_0)^k$ will be the corresponding cumulant generating function. First let k be even and k/2 be even, for any $t_1, t_2 \in R$,

$$\det \begin{vmatrix} e^{a\theta_0^k} & e^{a[i(t_1-t_2)+\theta_0]^k} \\ e^{a[i(t_2-t_1)+\theta_0]^k} & e^{a\theta_0^k} \end{vmatrix} = e^{2a\theta_0^k} - e^{2a\theta_0^k - k(k-1)a\theta_0^{k-2}(t_1-t_2)^2 + \cdots + 2a(t_1-t_2)^k},$$

which is smaller than 0 for large $(t_1 - t_2)$. This means $e^{a(it+\theta_0)^k}$ is not a Laplace-Stieljes transform, contradicting to the argument of Brown (1986) (see page 42).

Now let k be odd and (k-1)/2 be even. In this case, $a\theta_0$ is nonnegative. For any $t_1, t_2 \in R$,

$$\det \begin{vmatrix} e^{a\theta_0^k} & e^{a[i(t_1-t_2)+\theta_0]^k} \\ e^{a[i(t_2-t_1)+\theta_0]^k} & e^{a\theta_0^k} \end{vmatrix} = e^{2a\theta_0^k} - e^{a\theta_0[2\theta_0^{k-1} - k(k-1)\theta_0^{k-3}(t_1-t_2)^2 + \cdots + 2k(t_1-t_2)^{k-1}]},$$

which can be negative for large $(t_1 - t_2)$ if $\theta_0 \neq 0$.

In case $\theta_0 = 0$, take $t_1, t_2, t_3 \in R$ such that

$$a[(t_1-t_2)^k + (t_3-t_1)^k + (t_2-t_3)^k] \neq 2m\pi$$

for any integer m,

$$\begin{vmatrix} 1 & e^{ai(t_1-t_2)^k} & e^{ai(t_1-t_3)^k} \\ e^{ai(t_2-t_1)^k} & 1 & e^{ai(t_2-t_3)^k} \\ e^{ai(t_3-t_1)^k} & e^{ai(t_3-t_2)^k} & 1 \end{vmatrix}$$

$$= -2+2\cos\{a[(t_1-t_2)^k + (t_3-t_1)^k + (t_2-t_3)^k]\} < 0.$$

Therefore $e^{a(it+\theta_0)^k}$ is not positive definite, and is not a Laplace-Stieljes transform. So $a(\theta+\theta_0)^k$ is not a cumulant generating function, neither is $a\theta^k$ in case that k is odd and (k-1)/2 is even.

(iii) If k is odd, the dominating measure for $\psi(\theta) = a\theta^k$ $(k > 1)$ can not be finite from the latest argument above.

(iv) In Theorem 2 on page 62 of Singh (1974), we can reduce the condition (A2.1) there to

$$\int \frac{q_\gamma(x)}{u^{\frac{\gamma}{2}}(x,h_0)} d\mu < \infty$$

for some $0 < h_0 < 1$ because of the fact that LHS(A2.1) is increasing with respect to h, where

$$q_\gamma = \frac{\bigvee_\theta p_\theta}{(p_\alpha \wedge p_\beta)^{\frac{\gamma}{2}}}.$$

An interesting thing is that we can get a better rate with this weaker and simpler condition.

3.3 Case

(3.20) $$\psi(\theta) = \sum_{q=1}^{k} a_q \theta^{-q} + b\ln(-\theta),$$

($-\infty < \alpha \le \theta \le \beta < 0$, $b \le 0$ and $\text{sign}\{a_q\} = (-1)^q$ for $1 \le q \le k$)

Some common distribution families have this kind of cumulant generating functions. We will see through the following examples.

Example 3.2 If $a_q = 0$ for all q in (3.9), ψ is actually the cumulant generating function of a Gamma distribution family $\Gamma(-\theta,-b)$ which has density:

$$(-\theta)^{-b} e^{\theta x - b\ln(-\theta)}$$

with respect to μ, where μ has density $x^{-b-1} I_{(0,\infty)}(x)$ with respect to the Lebesgue measure.

Example 3.3 Consider another example (see page 3 of Brown (1986) for further reference). Let $Y_1, \cdots, Y_n$ be i.i.d. coming from normal distribution family :

$$\phi_{\mu,\sigma^2}(y) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-(y-\mu)^2 / 2\sigma^2).$$

Let

$$X_1 = \overline{Y} := n^{-1} \sum_{i=1}^{n} Y_i,\ S^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \text{ and } X_2 = S^2 + \overline{Y}^2 = n^{-1} \sum_{i=1}^{n} Y_i^2.$$

Then the joint density of $Y = (Y_1, \cdots, Y_n)$ is

$$(3.21)\quad f_{\mu,\sigma^2}(y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp((n\mu / \sigma^2)x_1 + (-n / 2\sigma^2)x_2) \exp(-n\mu^2 / 2\sigma^2).$$

Let $\theta_1 = \dfrac{n\mu}{\sigma^2}$ and $\theta_2 = -\dfrac{n}{2\sigma^2}$. (3.12) becomes

$$f_{\theta_1,\theta_2}(x_1, x_2) = \exp(\theta_1 x_1 + \theta_2 x_2 - \psi(\theta))$$

with $\psi(\theta) = -\dfrac{\theta_1^2}{4\theta_2} - \dfrac{n}{2}\log(-\dfrac{2\theta_2}{n})$, and the natural parameter space

$$N = \{(\theta_1, \theta_2) : \theta_1 \in R, \quad \theta_2 < 0\}.$$

For fixed $\theta_1 = \theta_{10}$, define a measure $\tau(A) = \int_{R\times A} e^{\theta_{10}x_1} d\mu$ for any measurable set A in R. Then the marginal distribution of (3.12) becomes an exponential family with

$$\psi(\theta_2) = -\frac{\theta_{10}^2}{4\theta_2} - \frac{n}{2}\log(-\frac{2\theta_2}{n})$$

as the cumulant generating function and τ as the dominating measure.

Proposition 3.1 ψ in (3.9) is a cumulant generating function of some exponential family and now the natural parameter space N = (-∞, v}, open or closed for some v ≤ ∞.

Proof. Let

$$f(w) = e^{\psi(-w)} = [\prod_{q=1}^{k} e^{|a_q|w^{-q}}]w^b, \quad w \in (0,\infty).$$

Then $f(w) \leq \prod_{q=1}^{k} e^{|a_q|}$ for $w \in [1,\infty)$.

Let $c_1 = 1$, $\qquad c_j = \dfrac{1}{j!(j-2)!}$ $\qquad$ (j=2, 3, $\cdots$), and let

$$Q_j(t,w) = c_j \frac{\partial^j}{\partial w^j}[w^{2j-1}e^{-tw}] \quad , \text{ for } j = 1, 2, \cdots.$$

According to Theorem 7.3 of Boas and Widder (1940), if we can verify the following inequality:

(3.22) $$A_j(c,t) := e^{ct}t^{j-1}\int_0^c Q_j(t,w)f(w+c)dw \geq 0$$

for $c > 1$, $j = 1, 2, \ldots$, and $0 < t < \infty$, (which is (7.6) there), then the proposition will follow.

First, let us show that

(3.23) $$f^{(j)}(w) = (-1)^j f(w) P_j(\frac{1}{w}),$$

where

$$P_j(\frac{1}{w}) = \sum_{i=1}^{j(k+1)} C_{ji} (\frac{1}{w})^i$$

with $C_{ji} \geq 0$ for $1 \leq i \leq j(k+1)$.

$$\begin{aligned} f^{(1)}(w) &= f(w)[\sum_{m=1}^{k} (-|a_m| m w^{-m-1}) + b/w] \\ &= (-1) f(w)[\sum_{m=1}^{k} (|a_m| m w^{-m-1}) + (-b)/w] \\ &= (-1) f(w) P_1(\frac{1}{w}), \end{aligned}$$

where $P_1(\frac{1}{w}) = \sum_{m=1}^{k} (|a_m| m w^{-m-1}) + (-b)/w$.

Assume (3.23) is true for j.

$$\begin{aligned} f^{(j+1)}(w) &= (-1)^{j+1} f(w) P_1(\frac{1}{w}) P_j(\frac{1}{w}) + (-1)^{j+1} f(w) (\sum_{i=1}^{j(k+1)} C_{ji} \frac{i}{w^{i+1}}) \\ &= (-1)^{j+1} f(w) [P_1(\frac{1}{w}) P_j(\frac{1}{w}) + \sum_{i=1}^{j(k+1)} C_{ji} \frac{i}{w^{i+1}}] \\ &:= (-1)^{j+1} f(w) P_{j+1}(\frac{1}{w}) \quad . \end{aligned}$$

(3.23) is verified.

Using the argument on page 4 of Boas and Widder (1940),

$$A_j(c,t) = (-1)^j c_j t^{j-1} e^{ct} \int_0^\infty w^{2j-1} f^{(j)}(w+c) dw \geq 0,$$

for any $c > 1, j \geq 1$ and $0 < t < \infty$.

The proof is complete. ◊

The derivative of $\psi(\theta)$ is

$$\dot{\psi}(\theta) = -\sum_{q=1}^{k} q a_q \theta^{-q-1} + b\theta^{-1}.$$

According to Proposition 1.2,

$$\tau_G(x) = \dot{\psi}^{-1}[E(\dot{\psi}(\theta)|x)] = \dot{\psi}^{-1}[-\sum_{q=1}^{k} q a_q E(\theta^{-q-1}|x) + bE(\theta^{-1}|x)].$$

Let $p_G^{[-q]}(x) = \int_\alpha^\beta \frac{1}{\theta^q} p_\theta(x)dG(\theta)$ and $p_G(x) = \int_\alpha^\beta p_\theta(x)dG(\theta)$. Then with these notations $\tau_G(x) = \dot{\psi}^{-1}[-\sum_{q=1}^{k} q a_q \frac{p_G^{[-(q+1)]}(x)}{p_G(x)} + b\frac{p_G^{[-1]}(x)}{p_G(x)}]$.

The following lemma generalizes a result of Singh (1974) for k = 1 (see page 69) to k > 1 case.

Lemma 3.4 Let $x \in R$, G a prior on $[\alpha,\beta]$. Then for any integer $k > 0$,

$$p_G^{[-k]}(x) = (-1)^k \int_x^{+\infty} ds_{k-1} \int_{s_{k-1}}^{+\infty} \cdots \int_{s_1}^{+\infty} p_G(s)ds. \tag{3.24}$$

Proof. Induction on $k \geq 1$ will achieve the proof. Observe that the integral of $p_G(s)$ with respect to the Lebesgue measure ds on $[x, \infty)$ is the integral of $-\frac{1}{\theta}p_\theta(x)$ with respect to $G(\cdot)$ on $[\alpha,\beta]$. Assume (3.24) is true for k = n-1. Now consider k = n case.

$$\begin{aligned} RHS(3.24) &= (-1) \int_x^\infty ds_{n-1} \int_\alpha^\beta \theta^{-(n-1)} p_\theta(s_{n-1})dG(\theta) \\ &= \int_\alpha^\beta (-1)\ \theta^{-(n-1)} e^{-\psi(\theta)} dG(\theta) \int_x^\infty e^{\theta s_{n-1}} ds_{n-1} = LHS(3.24). \end{aligned}$$

We have used the assumption for the first equation above. The lemma follows by induction. ◊

Let $L = \log(h^r)/\beta$ and

$$\hat{p}_i^{[-k]}(x) = (-1)^k \int_x^{x+L} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} \hat{p}_i(s)ds \quad .$$

Lemma 3.5 In the context above,

$$E_{\underline{\theta}_i} |\hat{p}_i^{[-k]} - p_{G_n}^{[-k]}| \le |\beta|^{-k} \{kh^r p_{G_n}(x) + e^{\psi^* - \psi_*} \frac{2}{n-1}(p_\alpha(x) + p_\beta(x)) \tag{3.25}$$
$$+ T_r(e^{-\alpha} - 1)h^r p_i(x)\} + \sqrt{\frac{L^k T p_i(x)}{(n-1)hu(x, kL+h)|\beta|^k}} \quad .$$

Proof. Triangulate the LHS(3.25) about k iterated integrals of $p_{G_n}(s)$ changing each iterated integral's interval at a time and apply the subadditivity of the absolute value,

$$\begin{aligned}
&\text{LHS(3.25)} \\
&\le \int_x^{x+L} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} E_{\underline{\theta}_i} |\hat{p}_i(s) - p_{G_n}(s)| ds \\
&\quad + \int_{x+L}^{+\infty} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} p_{G_n}(s)ds \\
&\quad + \int_x^{+\infty} ds_{k-1} \int_{s_{k-1}+L}^{+\infty} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} p_{G_n}(s)ds + \cdots \\
&\quad + \int_x^{+\infty} ds_{k-1} \int_{s_{k-1}}^{+\infty} ds_{k-2} \cdots \int_{s_2}^{+\infty} ds_1 \int_{s_1+L}^{+\infty} p_{G_n}(s)ds \quad .
\end{aligned}$$

Consider the second one of k+1 iterated integrals above, denoted as I_2. Replace the upper limits of k-1 inner integrals of I_2 by $+\infty$ and use Lemma 3.4 in the first inequality below,

$$\begin{aligned}
I_2 &\le (-1)^k \int_\alpha^\beta \theta^{-k} p_\theta(x+L) dG_n(\theta) \le |\beta|^{-k} p_{G_n}(x+L) \\
&\le |\beta|^{-k} p_{G_n}(x) e^{\beta L} = |\beta|^{-k} h^r p_{G_n}(x) \quad .
\end{aligned}$$

Similarly, we have $|\beta|^{-k} h^r p_{G_n}(x)$ as the common bound for each iterated integral after the second one. For the first one, denoted as I_1, applying Lemma 3.2 in the first inequality and Lemma 3.4 in the second inequality below,

$$I_1 = \int_x^{x+L} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} [\frac{2}{n-1} e^{\psi^*-\psi_*}(p_\alpha(s)+p_\beta(s))$$

$$+ D_i(n)]ds$$

$$\leq |\beta|^{-k} [\frac{2e^{\psi^*-\psi_*}}{n-1}(p_\alpha(x)+p_\beta(x)) + T_r(e^{-\alpha}-1)p_i(x)]$$

$$+ \int_x^{x+L} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} \sqrt{\frac{Tp_i(s)}{(n-1)hu(s,h)}} ds$$

$$\leq |\beta|^{-k} [\frac{2e^{\psi^*-\psi_*}}{n-1}(p_\alpha(x)+p_\beta(x)) + T_r(e^{-\alpha}-1)p_i(x)]$$

$$+ \sqrt{\frac{T}{(n-1)hu(x,kL+h)}} \int_x^{x+L} ds_{k-1} \int_{s_{k-1}}^{s_{k-1}+L} ds_{k-2} \cdots \int_{s_2}^{s_2+L} ds_1 \int_{s_1}^{s_1+L} \sqrt{p_i(s)} ds \quad .$$

Because by the concavity of $x^{\frac{1}{2}}$ and Jensen inequality,

$$\int_{s_1}^{s_1+L} \sqrt{p_i(s)} ds \leq L(\frac{1}{L}\int_{s_1}^{s_1+L} p_i(s) ds)^{\frac{1}{2}} \leq L^{\frac{1}{2}} |\beta|^{-\frac{1}{2}} p_i^{\frac{1}{2}}(s_1) \quad .$$

Therefore,

$$I_1 \leq |\beta|^{-k} [\frac{2e^{\psi^*-\psi_*}}{n-1}(p_\alpha(x)+p_\beta(x)) + T_r(e^{-\alpha}-1)h^r p_i(x)] + \sqrt{\frac{L^k T p_i(x)}{(n-1)hu(x,kL+h)|\beta|^k}} \, .$$

Combined with the first part of this proof, we have our lemma. ◊

Lemma 3.6 Let $\{p_\theta(\cdot):\theta \in \Theta\}$ be the one dimensional exponential family (1.1) mentioned in the beginning of Section 3.1. Assume that for numbers C_1, C_2 and ℓ with $C_1, C_2 \geq 0$ and $\ell \leq 0$,

$$(3.26) \qquad \bigvee_{v \in \{\alpha,\beta\}} \int (\frac{p_v(x)}{u(x,(k+1)L+h)})^{\frac{1}{2}} d\mu(x) \leq h^{\ell}(C_1 + C_2 |\log h|^m)$$

holds for $0 < h < 1$ and some real number m. Let $V_k(x)$ be the RHS (3.25). Here $V_0(x)$ equals to $U_0(x)$ in Lemma 3.3. Then

(3.27)

$$\int V_q(x)d\mu(x)$$

$$\leq |\beta|^{-q}[qh^r + \frac{4}{n-1}e^{\psi^*-\psi_*} + T_r(e^{-\alpha}-1)h^r] + 2C\sqrt{\frac{L^qT}{|\beta|^q}}(n-1)^{-\frac{1}{2}}h^{-\frac{1}{2}+\ell}(C_1 + C_2|\log h|^m)$$

for $0 \leq q \leq k+1$ and $0 < h < 1$.

Proof. Through (3.25),

$$\text{LHS}(3.27) \leq |\beta|^{-q}[qh^r + \frac{4}{n-1}e^{\psi^*-\psi_*} + T_r(e^{-\alpha}-1)h^r]$$
$$+\sqrt{\frac{L^qT}{(n-1)h|\beta|^q}}e^{\frac{1}{2}(\psi^*-\psi_*)}\int\frac{p_\alpha^{\frac{1}{2}}(x)+p_\beta^{\frac{1}{2}}(x)}{u^{\frac{1}{2}}(x,qL+h)}d\mu(x) \quad .$$

(3.27) is obtained by the fact that $u(x,qL+h) \geq u(x,(k+1)L+h)$ for $0 \leq q \leq k+1$ and our assumption. ◊

Example 3.4 Let μ be a measure that has density $e^{-m_0x}x^{-m_1}1_{[0,\infty)}$ ($m_1 < 1$ and $m_0 \geq 0$) with respect to the Lebesgue measure. For the exponential family (1.1) with respect to μ, $N = (-\infty, m_0)$. For $0 < h < 1$,

$$\bigvee_{v\in\{\alpha,\beta\}}\int\frac{p_v^{\frac{1}{2}}(x)}{u^{\frac{1}{2}}(x,kL+h)}d\mu(x) \leq \bigvee_{v\in\{\alpha,\beta\}}\int_{[0,\infty)}p_v^{\frac{1}{2}}(x)e^{\frac{m_0}{2}(kL+h)}(x+kL+h)^{\frac{m_1}{2}}d\mu(x)$$
$$\leq h^{\frac{m_0k}{2\beta}r}(C_1 + C_2|\log h|^{\frac{m_1}{2}}) \quad .$$

Condition (3.26) is satisfied with $\ell = \frac{m_0k}{2\beta}r$ and $m = \frac{m_1}{2}$.

Theorem 3.2 Let $\{p_\theta(\cdot):\theta \in \Theta\}$ be the exponential family with (3.9) as the cumulant generating function. Assume that the condition (3.26) holds. Let

$$\hat{t}_i(\underline{X}) = \dot{\psi}^{-1}\{\dot{\psi}(\alpha) \vee [-\sum_{q=1}^{k} q a_q \frac{\hat{p}_i^{[-(q+1)]}(X_i)}{\hat{p}_i(X_i)} + b\frac{\hat{p}_i^{[-1]}(X_i)}{\hat{p}_i(X_i)}] \wedge \dot{\psi}(\beta)\}.$$

Then $\{\underline{\hat{t}} = (\hat{t}_1(\underline{X}), \cdots, \hat{t}_n(\underline{X}))\}$ is an asymptotically optimal set compound estimator and

$$\text{(3.28)} \qquad \sup_{\underline{\theta}} |D_n(\underline{\theta}, \underline{\hat{t}})| = O\{n^{-\gamma}[\sum_{q=0}^{k+1}(M_q'(\log n)^{\frac{q}{2}} + M_q(\log n)^{m+\frac{q}{2}})]\} \quad ,$$

where $\gamma = \dfrac{r}{2(r-\ell)+1}$, $\{M_q'\}$ and $\{M_q\}$ are nonnegative constants.

Proof. Using the fact that $|a \vee f \wedge b - a \vee g \wedge b| \le |f-g| \wedge |a-b|$ for any real numbers a, b and any functions f, g and the Mean Value Theorem to the function $\dot{\psi}^{-1}(\cdot)$ for the first inequality, Singh-Datta Lemma for the second inequality and Lemma 3.6 for the third inequality below,

$$\text{RHS}(3.13) \le$$

$$B_0 C_0 n^{-1}\sum_{i=1}^{n} E_{\underline{\theta}}\{|\frac{\sum_{q=1}^{k} q a_q \hat{p}_i^{[-(q+1)]}(X_i) - b\hat{p}_i^{[-1]}(X_i)}{\hat{p}_i(X_i)} - \frac{\sum_{q=1}^{k} q a_q p_{G_n}^{[-(q+1)]}(X_i) - b p_{G_n}^{[-1]}(X_i)}{p_{G_n}(X_i)}| \wedge (\dot{\psi}(\beta) - \dot{\psi}(\alpha))\}$$

$$\le B_0 C_0 n^{-1}\sum_{i=1}^{n} E_{\underline{\theta}} p_{G_n}^{-1}(X_i)\{\sum_{q=1}^{k}(q|a_q||\hat{p}_i^{[-(q+1)]}(X_i) - p_{G_n}^{[-(q+1)]}(X_i)|)$$

$$+ (-b)|\hat{p}_i^{[-1]}(X_i) - p_{G_n}^{[-1]}(X_i)| + [|\dot{\psi}(\alpha)| \vee |\dot{\psi}(\beta)| + \dot{\psi}(\beta) - \dot{\psi}(\alpha)]|\hat{p}_i(X_i) - p_{G_n}(X_i)|\}$$

$$\le B_0 C_0 \{\sum_{q=1}^{k}[q|a_q|(|\beta|^{-(q+1)}[(q+1)h^r + \frac{4}{n-1}e^{\psi^* - \psi_*} + T_r(e^{-\alpha}-1)h^r]$$

$$+ 2C\sqrt{\frac{L^{q+1}T}{|\beta|^{q+1}}}(n-1)^{-\frac{1}{2}}h^{-\frac{1}{2}+\ell}(C_1 + C_2|\log h|^m))] + (-b)(|\beta|^{-1}[h^r + \frac{4}{n-1}e^{\psi^* - \psi_*} + T_r(e^{-\alpha}-1)h^r]$$

$$+ 2C\sqrt{\frac{L\,T}{|\beta|}}(n-1)^{-\frac{1}{2}}h^{-\frac{1}{2}+\ell}(C_1 + C_2|\log h|^m)) + [|\dot{\psi}(\alpha)| \vee |\dot{\psi}(\beta)| + \dot{\psi}(\beta) - \dot{\psi}(\alpha)]$$

$$(\frac{4}{n-1}e^{\psi^* - \psi_*} + T_r(e^{-\alpha}-1)h^r + 2C\sqrt{T}e^{\frac{1}{2}(\psi^* - \psi_*)}(n-1)^{-\frac{1}{2}}h^{-\frac{1}{2}+\ell}(C_1 + C_2|\log h|^m))\} \quad ,$$

for $0 < h < 1$.

At the choice of $h = n^{-\frac{1}{2(r-\ell)+1}}$, (3.28) is obtained. ◊

Remark 3.2 Let $\psi(\theta) = \sum_{q=1}^{k} a_q \theta^{-q} + b \ln(-\theta)$, with $0< \alpha \le \theta \le \beta < -\infty$, $b \le 0$, $\{a_q\}$ being all nonnegative. Then it is a cumulant generating function with support$\{\mu\}$ being a subset of $(-\infty, 0]$ and $N = \{v, +\infty)$, open or closed for some $v \ge -\infty$.

BIBLIOGRAPHY

Boas, R.P. and Widder, D.V. (1940). An Inversion Formula for the Laplace Integral. Duke Mathematical Journal, 6, 1 - 26.

Brown, L.D. (1986). Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. IMS Lecture Notes -- Monograph Series 9.

Datta, S. (1991a). Asymptotic Optimality of Bayes Compound Estimators in Compact Exponential Families. Ann. Statist. 19, 354 - 365.

Datta, S. (1991b). On the Consistency of Posterior Mixtures and its Applications. Ann. Statist. 19, 338 - 353.

D. K. Dey, M. Ghosh and C. Srinivasan (1987). Simultaneous Estimation of Parameter under Entropy Loss. Journal of Statistical Planning and Inference 15 (1987) 347-363.

M. Ghosh and M. C. Yang (1988). Simultaneous Estimation of Poisson Means under Entropy Loss. Ann. Statist. Vol. 16, No.1, 278-291.

Gilliland, Dennis C. (1968). Sequential Compound Estimation. Ann. Math. Statist. 39, No.6, 1890 - 1904.

Gilliland, Dennis C., Hannan, James F. and Huang, J.S. (1976). Asymptotic Solutions to the Two State Component Decision Problem, Bayes Versus Diffuse Priors on Proportions. Ann. Statist. 4, 1101 - 1112.

Gilliland, Dennis C. and Hannan, James F. (1986). The Finite State Compound Decision Problem, Equivariance and Restricted Risk Components. Adaptive Statistical Procedures and Related Topics, IMS Lecture Notes - Monograph Series 8, 129 - 145.

Hannan, James F. (1956). The Dynamical Statistical Decision Problem When the Component Problem Involves a Finite Number, m of Distributions. (Abstract) Ann. Math. Statist. 27, 212.

Hannan, James F. (1957). Approximation to Bayes Risk in Repeated Play, Contribution to the Theory of Games 3. Ann. Math. Studies, No.39, Princeton University Press, 97 - 139.

Hannan, James and Huang, J.S. (1972a). Equivariant Procedures in the Compound Decision Problem. Ann. Math. Statist. 43, 102-112.

Hannan, James and Huang, J.S. (1972b). A Stability of Symmetrization of Product Measures with Few Distinct Factors. Ann. Math. Statist. 43, No.1, 308 - 319.

Majumdar, S. (1993). Asymptotically Optimal and Admissible Estimators in Compound Compact Gaussian Shift Experiments. Ph.D. Thesis, Department of Statistics and Probability, Michigan State University.

Mashayekhi, M. (1990). Stability of Symmetrized Probabilities and Compact Equivariant Compound Decisions. Ph.D. Thesis, Department of Statistics and Probability, Michigan State University.

Mashayekhi, M. (1995). On Symmetrization of Product Measures. Mathematical Methods of Statistics 4, No.3, 334-343.

Rainville, E. D. (1967). Infinite Series, Macmillan, New York.

Robbins, H. (1951). Asymptotically Subminimax Solutions of Compound Statistical Decision Problems. Proc. Second Berkeley Symp. Math. Statist. Prob., 131 - 148, University of California Press.

Rudin, W. (1973). Functional Analysis, McGraw-Hill.

Samuel, E. (1965). Sequential Compound estimators. Ann. Math. Statist. 36, 879-889

Singh, R.S. (1974). Estimation of Derivatives of Average of μ-Densities and Sequence - Compound Estimation in Exponential Families. Ph.D. Thesis, Department of Statistics and Probability, Michigan State University.

Swain, D. D. (1965). Bounds and Rates of Convergence for the Extended Compound Estimation Problem in the Sequence Case. TR-81, Department of Statistics, Stanford.

Vardeman, S.B. (1982). Approximation to Minimum K - Extended Bayes Risk in Sequences of Finite State Decision Problems and Games. Bulletin of the Institute of Mathematics Academica Sinica 10, No. 1, 35 - 52.

Yu, B. (1971). Rates of Convergence in Empirical Bayes Two-action and Estimation Problems and in Extended Sequence - Compound Estimation Problems. RM-279, Statistics and Probability MSU.

Zhu, J. (1992). Asymptotic Behavior of Compound Rules in Compact Regular and Nonregular Families. Ph.D. Thesis, Department of Statistics and Probability, Michigan State University.