ADMISSIBILITY RESULTS IN LOSS ESTIMATION

BY

CHITRA LELE

TECHNICAL REPORT NO. 352
August 1990

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89–05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Admissibility Results in Loss Estimation

By

Chitra Lele

Technical Report No. 352

August 1990

Prepared Under the Auspices

Of

National Science Foundation Grant DMS89-05874

Department of Statistics

Stanford University

Stanford, California
ADMISSIBILITY RESULTS IN LOSS ESTIMATION

By
Chitra Lele

Abstract

Some theoretical development in the area of loss estimation is sought, especially with respect to admissibility of loss estimators. It's often important to report the discrepancy between an unknown parameter $\theta \in \mathbb{R}^p$ and its point estimator $\delta(X)$. This discrepancy (or 'loss') $L(\delta(X), \theta)$, being a function of both parameter and data, is itself unobservable, but a data dependent estimator $\gamma(X)$ of the loss would be a desirable measure of the performance of $\delta(X)$. Throughout the present work, squared error is used as the distance measure to study how well $\gamma(X)$ estimates $L(\delta(X), \theta)$. The risk incurred by $\gamma(X)$ is

$$R(\gamma, \theta) = E_\theta [\gamma(X) - L(\delta(X), \theta)]^2.$$ 

Here we consider the point estimator $\delta(X)$ as fixed. This problem has both conditional and frequentist aspects: the conditional one is that the actual loss (which is a function of both the observed data as well as the unknown parameter) is estimated instead of the risk, and the frequentist one is that the loss estimators are compared by averaging their performance over the sampling distribution of $X|\theta$. An unbiased estimator of this loss is provided by Stein's unbiased estimator of risk (Annals of Statistics, 1981). Johnstone (Statistical Decision Theory and Related Topics, 1988) proved that in the Gaussian case, the unbiased loss estimators corresponding to the mle (maximum likelihood estimator) and to the James-Stein estimator as point estimators, are inadmissible for $p \geq 5$ and he constructed 'improvements' over these unbiased estimators. An attempt is made herein to extend these kind of results to other cases. Related work has been done by Berger and Lu (Annals of Statistics, 1989), and more recently, by Brown and Hwang, in a different setting.
In the Poisson setup, with the mle as point estimator and the information normalized loss function, i.e. \(X_i, i = 1, \ldots, p\), are independent \(\text{Poisson}(\lambda_i)\) variables, 
\[ X = (X_1, \ldots, X_p)^T, \lambda = (\lambda_1, \ldots, \lambda_p)^T, \delta(X) = X \text{ and} \]
\[ L_{-1}(\delta, \lambda) = \sum_{i=1}^{p} \frac{(\delta_i - \lambda_i)^2}{\lambda_i}, \]
the unbiased loss estimator 'p' is proved admissible for \(p \leq 2\) and for \(p \geq 3\), an 'improved' estimator
\[ p = \frac{(p - 2)}{\sum X_i + 1} \]
is constructed and is also given a Bayesian interpretation. Improvements over unbiased loss estimators are also derived for distributions belonging to Hudson's subclass of continuous exponential families; this subclass contains the gamma(shape parameter) and the inverse Chi-square distributions, in addition to the Gaussian.

Next, a unified admissibility theory is developed for loss estimation and is applied to various more general situations. In the Gaussian case, the admissibility of the posterior loss estimator associated with the Strawderman prior and its point estimator, is proved. Similarly, in the Poisson case with \(L_{-1}\) loss, the posterior loss estimator associated with the Clevenson-Zidek prior is proved to be admissible. Thus, in both these cases, the priors yielding admissible point estimators are seen to yield admissible loss estimators as well. For the Poisson case, with the mle \(X\) as the point estimator and squared error loss, i.e. \(L(X, \lambda) = \sum (X_i - \lambda_i)^2\), the unbiased loss estimator \(\sum X_i\) is proved admissible for \(p \leq 2\). The relation between the Gaussian and the Poisson cases is explained via the polydisc transform. Further, it is proved that the phenomenon 'if an estimator is admissible, then it is generalized Bayes' also holds for loss estimates.
Contents

Abstract iv

Acknowledgements vi

1 Introduction 1
   1.1 Description of the problem 1
   1.2 History and Review 3

2 The James-Stein positive part estimator 6

3 Poisson Distribution 10
   3.1 Improvement over the unbiased estimator of loss 10
      3.1.1 Data Examples 15
   3.2 Admissibility for $p \leq 2$ 21
   3.3 Clevenson-Zidek point estimator 27

4 Hudson's Subclass Of Exponential Families 37

5 Unified Admissibility Theory 41
   5.1 Normal Distribution 44
      5.1.1 Verification of existing result 50
      5.1.2 The Strawderman prior case 50
   5.2 General exponential family distributions 52
   5.3 Poisson with information-normalized loss 55
      5.3.1 Verification of earlier result 61
5.3.2 The Clevenson-Zidek estimator ....................... 61
5.4 Poisson with squared error loss .................... 62
5.5 Role of the polydisc transform .................... 66
  5.5.1 Point Estimation ................................. 67
  5.5.2 Loss Estimation ................................. 68

6 Necessary Condition for Admissibility ............... 70

Bibliography ........................................... 75
Chapter 1

Introduction

In the present work, we shall endeavour to achieve further theoretical development in the area of loss estimation, especially with respect to admissibility of loss estimators. It is often important to report the discrepancy between an unknown (vector) parameter $\theta \in \mathbb{R}^p$ and its point estimator $\delta(X)$. This discrepancy (or loss), being a function of both parameter and data, is itself unknown, but a data-dependent estimator $\gamma(X)$ of this loss would be a desirable measure of the performance of $\delta(X)$.

1.1 Description of the problem

Let us first take a closer look at the problem and the set-up that we shall be working with; and later on we shall take a quick look at the other approaches that have been tried. Formulating the problem rigorously, let $X \sim p_\theta(x)$, $p_\theta(x)$ is the density of $X$ relative to a dominating measure $\nu(dx)$, $\theta \in \Theta \subset \mathbb{R}^p$. Let $\delta(X) \in \mathbb{R}^p$ be an estimator of $\theta$. A loss $L(\delta(X), \theta)$ is incurred. We now need another distance measure to study how well $\gamma(X)$ behaves as an estimator of the loss $L(\delta(X), \theta)$; we shall use squared error as this distance measure. Note that $L$ doesn't have to be a squared error loss function. Hence the risk incurred by $\gamma(X)$ is

$$R(\gamma, \theta) = E_\theta [\gamma(X) - L(\delta(X), \theta)]^2.$$ 

This is really $R(\gamma, \theta, \delta(X))$, but we assume $\delta(X)$ is fixed and given.
CHAPTER 1. INTRODUCTION

This approach is both conditional and frequentist: conditional because the actual loss (which is a function of both the observed data and the unknown parameter) is being estimated instead of the risk, and frequentist because the loss estimators are compared by averaging their performance over the sampling distribution of $X|\theta$. Although, in this sense, the present problem is different from the conventional ones, all the conventional definitions and methods can be used in studying the admissibility of loss estimators. We say that $\gamma(X)$ is inadmissible if there exists another loss estimator

$$\gamma(X) \text{ such that } R(\gamma, \theta) \leq R(\gamma, \theta) \text{ for all } \theta$$

with strict inequality for some $\theta$. The Bayes rule $\gamma_G$ corresponding to the prior $G(d\theta)$ is the minimizer of

$$r(\gamma, G) = \int R(\gamma, \theta) G(d\theta)$$

and

$$\min_{\gamma} r(\gamma, G) = r(\gamma_G, G) = r(G).$$

If $r(G) < \infty$, then the Bayes rule $\gamma_G$ also minimizes the posterior expected distance

$$E_G[(L(\delta(X), \theta) - \gamma)^2 | X = x]$$

and

$$\gamma_G = E_G[L(\delta(X), \theta)|X]. \quad (1.1)$$

If $r(G) = \infty$, $\gamma_G$ defined as above, is called a formal, or a generalized Bayes rule. Note that $\delta(X)$ can be arbitrary.

Blyth's lemma (cf. Berger, 1985) is the most commonly used tool in admissibility proofs. There is a version of this lemma applicable to the present situation; we shall state it later.
1.2 History and Review

The problem of confidence interval estimation has been long studied, and a collection of references appears in Rukhin (1988). There has also been recent interest in the study of loss estimation itself. Rukhin (1988) considers loss functions for the simultaneous estimation of $\theta$ and $L(\delta(X), \theta)$; but we shall focus on estimating $L(\delta(X), \theta)$, once $\delta(X)$ is fixed. Berger and Lu (1989) consider a related problem and we will take a closer look at it later. More recently, Brown and Hwang (1989) have considered a variant of our problem, i.e. they consider admissibility of the unbiased estimator of the coverage function $I(\theta \in C(X))$, in the Gaussian setting, and prove admissibility for $p \leq 4$. A point to note here is that Berger and Lu (1989) and Rukhin (1988) require that a loss estimator be conservative, i.e.,

$$E_\theta(\gamma(X)) \geq E_\theta[L(\delta(X), \theta)].$$

(1.2)

But we don't impose this restriction.

Let us first note that an unbiased estimator of risk is also unbiased for estimating loss. Stein (1981) gives the unbiased estimator of risk in the Gaussian setting, and the same method can be used to determine the unbiased estimators in other distributions as well. Our aim here is to see if we can estimate $L$ better than just by using the unbiased estimator of risk, and also to see what admissibility properties, or methods, carry over from point estimation to loss estimation. The approach here is modeled after the work of Johnstone (1988). Johnstone considered admissibility of unbiased loss estimates in the Gaussian case. The Gaussian setting that he considers is $X \sim N_p(\theta, I)$, $\delta(X)$ is the estimator of $\theta$, and the loss incurred is $L(\delta(X), \theta) = ||\delta(X) - \theta||^2$. For the case $\delta(X) = X$ (the mle), $\gamma_{unb}(X) = p$ is the unbiased estimator of loss, and Johnstone proved that 'p' is admissible as an estimator of $L$ for $p \leq 4$ and that

$$\gamma(X) = p - \frac{2(p-4)}{\|X\|^2}$$

(1.3)

improves over 'p' for $p \geq 5$. 
CHAPTER 1. INTRODUCTION

For the case of the James-Stein estimator i.e. for

\[ \delta(X) = \left(1 - \frac{(p-2)}{\|X\|^2}\right)X \]

Johnstone proved that

\[ \gamma(X) = p - \frac{(p-2)^2}{\|X\|^2} + \frac{2p}{\|X\|^2} \tag{1.4} \]

improves over the unbiased estimator

\[ \gamma_{unb}(X) = p - \frac{(p-2)^2}{\|X\|^2} \quad \forall \quad p \geq 5. \]

He also gave an empirical Bayesian justification for this ‘correction factor’ and extended the results to normal theory linear models. We note here that both estimators (1.3) and (1.4) can be further improved by considering their positive-part versions i.e. by truncating them at zero.

Since the practitioners of statistics often deal with many other distributions besides the Gaussian, it would be useful to derive (in)admissibility results for loss estimators in other distributions and find ‘better’ estimators wherever possible. At the same time, it would be beneficial for the theoretical development of the subject, to try and extend as much of the methodology and results as possible from point estimation to loss estimation. This is the motivation behind this present work.

In Chapter 2 we shall derive improvements over the unbiased loss estimator in the case of the James-Stein positive-part point estimator. In Chapter 3, we consider an important discrete distribution, the Poisson, with the mle as the point estimator, and the information normalized loss function. Thus we have \( X_i \sim^{\text{ind}} \text{Poisson}(\lambda_i) \) for \( i = 1, \cdots, p \), \( X = (X_1, \cdots, X_p)^T \), \( \lambda = (\lambda_1, \cdots, \lambda_p)^T \), \( \delta(X) = X \), and

\[ L_{-1}(\delta, \lambda) = \sum_{i=1}^{p} \lambda_i^{-1}(\delta_i - \lambda_i)^2. \]

We prove admissibility of the unbiased estimator ‘p’ for \( p \leq 2 \), and for \( p \geq 3 \), we construct an improved estimator

\[ p - \frac{(p-2)}{\sum X_i + 1} \]
CHAPTER 1. INTRODUCTION

and give it a Bayesian interpretation. We present computational results and data examples to demonstrate the behavior of this estimator. Chapter 4 deals with construction of improved loss estimators for distributions belonging to Hudson's subclass of continuous exponential families; this subclass contains the gamma (with shape parameter) and the inverse Chi-square distributions besides the Gaussian. Some simulation results will be presented for the Gamma distribution.

Instead of having to prove the results individually, it would certainly be helpful to have one methodology which will help us prove admissibility of loss estimators in a variety of situations. So, in Chapter 5 we shall develop a unified admissibility theory for loss estimation. The basis for this is the theory developed by Brown and Hwang (1982) for point estimation. With this unified theory, in the Gaussian case, the admissibility of the posterior loss estimator associated with the Strawderman prior and its point estimator, is proved. Similarly, in the Poisson case with $L_1$ loss, the posterior loss estimator associated with the Clevenson-Zidek prior is proved to be admissible. Thus, in both these cases, priors yielding admissible point estimators are seen to yield admissible loss estimators as well. For the Poisson case with the mle (i.e., $X$) as the point estimator, and with the squared error loss $L(X, \lambda) = \sum_{i=1}^{p} (\delta_i - \lambda_i)^2$, we prove that the unbiased loss estimator $\sum_{i=1}^{p} X_i$ is admissible for $p \leq 2$. The relation between the Gaussian and Poisson cases is clarified via the polydisc transform. In the end, we prove that the phenomenon 'if an estimator is admissible, then it is generalized Bayes' also holds for loss estimators.
Chapter 2

The James-Stein positive part estimator

Johnstone (1988) derived improvements over the unbiased loss estimator corresponding to the James-Stein point estimator, but didn’t look at the positive part point estimator. However, since the James-Stein positive part estimator dominates the James-Stein estimator, the problem of estimating the loss of the former is worth looking into. Let $X \sim N_p(\theta, I)$,

$$
\delta(X) = \left(1 - \left(\frac{p-2}{\|X\|^2}\right)\right)_+ X,
$$

$$
L(\delta(X), \theta) = \|\delta(X) - \theta\|^2.
$$

Then the unbiased estimator of loss is given by

$$
\gamma_{unb}(X) = \|X\|^2 - p \quad \text{on} \quad \|X\|^2 < p - 2 \\
= p - \left(\frac{(p-2)^2}{\|X\|^2}\right) \quad \text{on} \quad \|X\|^2 \geq p - 2.
$$

The unbiased loss estimator over $\{\|X\|^2 \geq p - 2\}$ is the same as the unbiased loss estimator for the James-Stein estimator. So the correction used in the James-Stein case (see Johnstone(1988)), can be used on this region. Johnstone proves that using the correction factor gives lower risk than that of the unbiased estimator, so use of the correction factor over a smaller region will also result in improvement.
in risk. On \( \{ \|X\|^2 < p - 2 \} \), \( -p \leq \gamma_{unb} \leq -2 \), so anything sensible (i.e. non-negative) can be used here; preferably something that makes the estimator continuous at \( \|X\|^2 = p - 2 \). Using the positive part of the unbiased estimator over this region, i.e. using the constant zero, would certainly give better risk (see Anderson, 1984, p.91). Use of the constant '2' instead of '0' also gives lower risk as shown by Berger and Lu (1989). We, in addition, try using functions linear and quadratic in \( \|X\|^2 \) over this region near zero. The linear piece is chosen so as to make the overall estimator continuous at \( \{ \|X\|^2 = p - 2 \} \) and the quadratic piece is such that the estimator also has continuity of the first derivative. The fact that both these estimators with the linear and quadratic pieces improve over the unbiased estimator of loss will be verified only computationally.

**Proposition 2.1** The following estimators improve over \( \gamma_{unb} \) for all \( p > 4 \):

\[
\gamma^{(i)} = \begin{cases} 
\gamma^{(i)} & \text{if } \|X\|^2 < p - 2 \\
p - \frac{(p - 2)^2}{\|X\|^2} + \frac{2p}{\|X\|^2} & \text{if } \|X\|^2 \geq p - 2
\end{cases}
\]

where for \( i = 1, ..., 4 \), the \( \gamma^{(i)} \) are

\[
\begin{align*}
\gamma^{(1)} &\equiv 0 \\
\gamma^{(2)} &\equiv 2 \\
\gamma^{(3)} &= 0.5 + \left[ \frac{4(p - 1)}{(p - 2)} - 0.5 \right] \frac{1}{(p - 2)} \|X\|^2 \\
\gamma^{(4)} &= 0.5 - \frac{(p^2 - 13p + 10)}{(p - 2)^2} \|X\|^2 + \frac{(2p^2 - 19p + 14)}{2(p - 2)^3} \|X\|^4.
\end{align*}
\]

The percentage improvement in risk obtained by \( \gamma^{(i)} \) over \( \gamma_{unb} \) i.e.,

\[
\frac{R(\gamma_{unb}) - R(\gamma^{(i)})}{R(\gamma_{unb})} \times 100 \%
\]

for different values of \( p \) is plotted in figure (2.1). These computational results were obtained by numerical integration (using IMSL) and not by simulation. There is hardly any difference in the performance of \( \gamma^{(3)} \) and \( \gamma^{(4)} \), i.e. whether we use the linear or the quadratic piece near zero doesn't matter. So we have plotted the
percentage improvement in risk of $\gamma^{(i)}$ for $i = 1, 2, 3$. It is seen that the percentage improvements are comparable in magnitude to the percentage improvements obtained by Johnstone in the James-Stein case, they are not clearly either smaller or bigger. The choice between the estimators $\gamma^{(1)}$, $\gamma^{(2)}$ and $\gamma^{(3)}$ is certainly easier if there is some clue as to the magnitude of $\|\theta\|^2$. For example, $\gamma^{(3)}$ (or equivalently $\gamma^{(4)}$) is clearly better when $\|\theta\|^2$ is bigger than $p$. But in absence of any knowledge, $\gamma^{(2)}$ seems to be a safe bet.

Also plotted are results for another estimator $\gamma$. The percentage improvement in risk of this estimator over $\gamma_{unb}$ is plotted. Percentage improvement obtained by $\gamma^{(2)}$ over $\gamma$ is also shown. This is the estimator used by Berger and Lu(1989).

$$
\gamma = 2 \quad \text{if} \quad \|X\|^2 < p - 2 \\
= p - \frac{(p - 2)^2}{\|X\|^2} \quad \text{if} \quad \|X\|^2 \geq p - 2.
$$

(2.1)

This estimator is non-negative, and continuous at $\|X\|^2 = p - 2$, but the correction factor is not used over the region $\{\|X\|^2 \geq p - 2\}$. Comparing the risk reductions of this estimator with our estimator $\gamma^{(2)}$ gives an idea of how much the correction factor contributes to the overall improvement. As one would expect, the contribution of the correction factor is as small as 5% near the origin and as large as 80% for large $\|\theta\|^2$. The plot of the percentage improvement of $\gamma^{(2)}$ over $\gamma$ indicates that the percentage improvements obtained by Johnstone in the James-Stein point estimator case were not only due to the undesirable behavior (i.e. negativity) of $\gamma_{unb}$ near the origin, but that the correction factor really had a non-trivial contribution to make.

Remark: All the estimators $\gamma^{(i)}$, $i = 1, \cdots, 4$, and $\gamma$ satisfy condition (1.2) and hence are conservative estimators.
Figure 2.1: Percentage improvement of (a) $\gamma^{(1)}$, (b) $\gamma^{(2)}$, (c) $\gamma^{(3)}$ and (d) $\gamma$ over $\gamma_{unb}$ (e): Percentage improvement of $\gamma^{(2)}$ over $\gamma$

The X-axis is $\|\theta\|^2/p$
Chapter 3

Poisson Distribution

The Poisson is the most widely used model after the Gaussian. Hence, investigating properties of loss estimators in this case would be very illuminating. It would enable us to measure the performance of point estimators in this extensively used setting.

We let $X_i \overset{ind}{\sim} \text{Poisson}(\lambda_i), \ i = 1, \ldots, p$. We first consider $\delta(X) = X$ and the information-normalized loss function

$$L_{-1}(X, \lambda) = \sum_{i=1}^{p} \lambda_i^{-1} (X_i - \lambda_i)^2 .$$

The justification for using this loss function is the need to penalize more for small values of the $\lambda_i$s (cf. Clevenson and Zidek, 1975). Besides, the risk of the maximum likelihood point estimator $X$ is constant ($= p$) for this loss function. The weighting in this weighted loss function can be interpreted as scaling by Fisher information (cf. Lehmann, 1983, p.120); hence the name information-normalized loss function.

3.1 Improvement over the unbiased estimator of loss

Stein’s identities for the Gaussian case (1981) were generalized to some other distributions by Hudson (1978). Application of the same methodology yields the following lemma.
CHAPTER 3. POISSON DISTRIBUTION

Lemma 3.1 Let $h : \mathbb{R}_+^p \rightarrow \mathbb{R}^p$ and $h(X)$ is defined to be zero if any component of $X$ is negative. The following Poisson identities hold true if $h$ is such that all the necessary expectations exist:

$$E[\lambda_i h_i(X)] = E[X_i h_i(X - e_i)]$$  \hspace{0.5cm} (3.1)

$$E[\lambda_i^{-1} h_i(X)] = E \left[ \frac{h_i(X + e_i)}{X_i + 1} \right] + e^{-\lambda_i} \lambda_i^{-1} h_i(0)$$ \hspace{0.5cm} (3.2)

$$E[\lambda_i^{-2} h_i(X)] = E \left[ \frac{h(X + 2e_i)}{(X_i + 1)(X_i + 2)} \right] + e^{-\lambda_i} \lambda_i^{-2} h_i(0) + e^{-\lambda_i} \lambda_i^{-2} h_i(1)$$ \hspace{0.5cm} (3.3)

where $e_i$ is a column vector with '1' in the $i$th place and '0's elsewhere. $h_i(0) = h_i(X)|_{X_i=0}$ and $h_i(1)$ is similarly defined.

Proof: We shall only prove (3.2); the proofs of (3.1) and (3.3) follow similarly.

$$E[\lambda_i^{-1} h_i(X)] = \sum_{x_1, \ldots, x_p \geq 0} h_i(x) \lambda_i^{-1} \frac{\exp\left(-\sum_j \lambda_j \right) \prod_j \lambda_j^{x_j}}{x_i!} \prod_j x_j!$$

$$= \sum_{x_1, \ldots, x_p \geq 0} h_i(x) \frac{\lambda_i^{x_i-1} e^{-\lambda_i} \exp\left(-\sum_{j \neq i} \lambda_j \right) \prod_{j \neq i} \lambda_j^{x_j}}{x_i!} \prod_{j \neq i} x_j!$$

$$= \frac{h_i(0) e^{-\lambda_i}}{\lambda_i} + \sum_{y_1, \ldots, y_p \geq 0} \frac{h_i(y + e_i) \exp\left(-\sum_j \lambda_j \right) \prod_j \lambda_j^{y_j}}{y_i + 1} \prod_j y_j!$$

$$= \frac{h_i(0) e^{-\lambda_i}}{\lambda_i} + E \left[ \frac{h_i(X + e_i)}{X_i + 1} \right]$$

where, in the third step we use the change of variables $Y_1 = X_1, \ldots, Y_{i-1} = X_{i-1}, Y_i = X_i - 1, Y_{i+1} = X_{i+1}, \ldots, Y_p = X_p$ and write the expectation back in terms of $X$ for the sake of notational consistency.

This proves (3.2).

Let us assume that $\delta(X) = X + g(X)$, where $X_i = 0 \Rightarrow g_i(X) = 0$, and we also assume that $E_\theta ||g(X)||^2 < \infty \ \forall \ \theta$. Using the above identities,

$$E \left[ \sum_{i=1}^p \lambda_i^{-1} (X_i + g_i(X) - \lambda_i)^2 \right]$$
is
\[ p + \sum_{i=1}^{p} E \left\{ \frac{g_i^2(X + e_i)}{X_i + 1} + 2[g_i(X + e_i) - g_i(X)] \right\}. \]
(cf. Tsui and Press, 1982.)

For \( \delta(X) = X \),
\[ E \left[ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 \right] = p. \]
So \( \gamma_{n_{ub}}(X) = p \), and
\[ R(\gamma_{n_{ub}}, \lambda) = E \left[ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 - p \right]^2 \]
\[ = 2p + \sum \lambda_i^{-1}. \]

We shall henceforth drop the notational dependency of the risk function on \( \lambda \) and just write \( R(\gamma_{n_{ub}}) \) for the above.

For \( \gamma = p - \phi(\sum X_j) = p - \phi(p\bar{X}) \),
\[ R(\gamma) = E \left[ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 - p + \phi(p\bar{X}) \right]^2 \]
\[ = R(\gamma_{n_{ub}}) + E \left[ \phi^2(p\bar{X}) + 2\phi(p\bar{X}) \{ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 - p \} \right], \]
and
\[ E \left[ \phi(p\bar{X}) \{ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 - p \} \right] = E \left[ \sum \phi(p\bar{X}) \{ \lambda_i^{-1}X_i^2 + \lambda_i - 2X_i - 1 \} \right] \]
\[ = E \{ \sum \phi(p\bar{X} + 1)(X_i + 1) + X_i\phi(p\bar{X} - 1) \]
\[ - 2X_i\phi(p\bar{X}) - \phi(p\bar{X}) \} \]
\[ = E[p\bar{X}\{ \phi(p\bar{X} + 1) - 2\phi(p\bar{X}) + \phi(p\bar{X} - 1) \} \]
\[ + p\{ \phi(p\bar{X} + 1) - \phi(p\bar{X}) \}]. \]

The second equality follows from using the first two identities from lemma (3.1) and because \( X_i = 0 \Rightarrow \phi(p\bar{X})X_i^2 = 0 \).

Hence we have the following expression for the difference in risks:
\[ R(\gamma) - R(\gamma_{n_{ub}}) = E \{ \phi^2(p\bar{X}) + 2p\bar{X}[\phi(p\bar{X} + 1) - 2\phi(p\bar{X}) + \phi(p\bar{X} - 1)] \]
\[ + 2p[\phi(p\bar{X} + 1) - \phi(p\bar{X})] \}. \quad (3.4) \]
CHAPTER 3. POISSON DISTRIBUTION

We shall now use the variance stabilizing square root transformation and use the Gaussian results to guess the form of \( \phi \) that makes \( R(\gamma) - R(\gamma_{\text{unb}}) \leq 0 \) everywhere, with strict inequality somewhere. This idea was suggested by Brown (1979). The transformation we make is \( Z^2 = \sum X_i \). Let \( h \) be defined as \( \phi(\sum X_i) = h(Z) \).

If \( \sum X_i = x \), \( z_x = z \),

\[
  z_{x+1} = (x + 1)^\frac{1}{2} \\
  = x^{1/2} \left(1 + \frac{1}{x}\right)^{1/2} \\
  = z + \frac{1}{2z} - \frac{1}{8z^3} + O\left(\frac{1}{z^5}\right),
\]

\[
  z_{x-1} = (x - 1)^\frac{1}{2} \\
  = x^{1/2} \left(1 - \frac{1}{x}\right)^{1/2} \\
  = z - \frac{1}{2z} - \frac{1}{8z^3} + O\left(\frac{1}{z^5}\right).
\]

Since \( \phi(x) = h(z) \), \( \phi(x+1) = h(z_{x+1}) \) and

\[
  h \left( z + \frac{1}{2z} - \frac{1}{8z^3} \right) = h(z) + \left( \frac{1}{2z^4} - \frac{1}{8z^4} \right) h'(z) + \left( \frac{1}{4z^5} - \frac{1}{8z^5} \right) h''(z) + O\left(\frac{1}{z^5}\right),
\]

\[
  h \left( z - \frac{1}{2z} - \frac{1}{8z^3} \right) = h(z) - \left( \frac{1}{2z^4} + \frac{1}{8z^4} \right) h'(z) + \left( \frac{1}{4z^5} + \frac{1}{8z^5} \right) h''(z) + O\left(\frac{1}{z^5}\right),
\]

Therefore the difference in risks, ignoring terms that are \( O(z^{-5}) \), is

\[
  R(\gamma) - R(\gamma_{\text{unb}}) \approx E\{h^2(Z) + 2Z^2 \left[\frac{1}{4Z^3} h'(Z) + \frac{1}{4Z^2} h''(Z)\right] + 2p \left[\frac{1}{2Z^3} - \frac{1}{8Z^3} h'(Z) + \left(\frac{1}{4Z^3} - \frac{1}{8Z^4}\right) \frac{1}{2} h''(Z)\right]\}.
\]

Since in the Gaussian case, the correction factor is proportional to \( (\|X\|^2)^{-1} \), here we let \( h(Z) = cZ^{-2} \), where \( c \) is any constant that may depend on \( p \).

\[
  R(\gamma) - R(\gamma_{\text{unb}}) \approx (c^2 + 4c - 2pc) E\left(\frac{1}{Z^4}\right),
\]

\[c^2 - (2p-4)c \leq 0 \text{ for all } c \leq 2p-4 \text{ and is the least at } c = p-2.\]
CHAPTER 3. POISSON DISTRIBUTION

This suggests that
\[ \phi(p\bar{X}) = (p - 2) \frac{1}{\sum X_j} \]
would be the desired improved estimator. But this is only a heuristic. Since
\[ E(\sum X_j) = \sum \lambda_j \], if \( \sum \lambda_j \) is very small, then the correction factor would blow up,
and this could cause problems. Computational results confirm that this is indeed
the case. So, our heuristic estimator doesn’t improve over the unbiased estimator
when \( \sum \lambda_j \) is very small. Using \( \sum X_j + 1 \) instead of \( \sum X_j \) suffices; a proof that
this estimator actually dominates the unbiased loss estimator follows.

The first and second differences of \( \phi \) are
\[ \phi(p\bar{X} + 1) - \phi(p\bar{X}) = -\frac{(p - 2)}{(p\bar{X} + 1)(p\bar{X} + 2)} \]
and
\[ \phi(p\bar{X} + 1) - 2\phi(p\bar{X}) + \phi(p\bar{X} - 1) = \frac{2(p - 2)}{p\bar{X}(p\bar{X} + 1)(p\bar{X} + 2)} \].
Hence, the difference in risks, from equation (3.4), is
\[ R(\gamma) - R(\gamma_{\text{unb}}) = \mathbb{E} \left[ \frac{(p - 2)^2}{(p\bar{X} + 1)^2} + \frac{4(p - 2)}{(p\bar{X} + 1)(p\bar{X} + 2)} - \frac{2p(p - 2)}{(p\bar{X} + 1)(p\bar{X} + 2)} \right] \]
\[ = -(p - 2)^2 \mathbb{E} \left[ \frac{p\bar{X}}{(p\bar{X} + 1)^2(p\bar{X} + 2)} \right]. \]
Thus the loss estimator \( \gamma(X) \) has lower risk than the unbiased estimator, and
hence it dominates \( \gamma_{\text{unb}}(X) \).

Remark: The \( \gamma \) above is always non-negative, so unlike in the Gaussian
situation, there is no need for a positive-part modification.

We thus have the following proposition:

**Proposition 3.1**

\[ \gamma(X) = p - \frac{(p - 2)}{\sum X_j + 1} \quad (3.5) \]

improves over \( \gamma_{\text{unb}}(X) \equiv p \) as an estimator of \( \sum \lambda_i^{-1}(X_i - \lambda_i)^2 \) for all \( p \geq 3 \).
CHAPTER 3. POISSON DISTRIBUTION

The percentage improvement of the estimator

\[ \gamma(X) = p - \frac{(p - 2)}{\sum X_j + 1} \]

over \( \gamma_{unb} \) is

\[ \frac{R(\gamma_{unb}) - R(\gamma)}{R(\gamma_{unb})} \times 100\% \]

This is given in the following table.

<table>
<thead>
<tr>
<th>p</th>
<th>0.01</th>
<th>0.1</th>
<th>0.2</th>
<th>0.5</th>
<th>0.8</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.63</td>
<td>4.17</td>
<td>5.22</td>
<td>4.02</td>
<td>2.34</td>
<td>1.56</td>
<td>0.63</td>
<td>0.33</td>
</tr>
<tr>
<td>4</td>
<td>0.96</td>
<td>5.80</td>
<td>6.96</td>
<td>4.56</td>
<td>2.48</td>
<td>1.72</td>
<td>0.84</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
<td>6.50</td>
<td>7.50</td>
<td>4.50</td>
<td>2.50</td>
<td>1.70</td>
<td>1.00</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>1.26</td>
<td>6.84</td>
<td>7.50</td>
<td>4.38</td>
<td>2.52</td>
<td>1.92</td>
<td>1.08</td>
<td>0.66</td>
</tr>
<tr>
<td>8</td>
<td>1.36</td>
<td>7.20</td>
<td>7.52</td>
<td>4.32</td>
<td>2.64</td>
<td>1.92</td>
<td>1.04</td>
<td>0.64</td>
</tr>
<tr>
<td>10</td>
<td>1.50</td>
<td>7.20</td>
<td>7.50</td>
<td>4.40</td>
<td>2.50</td>
<td>1.80</td>
<td>1.00</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 3.1: Percentage improvement in risk of \( \gamma \) over ‘\( p \)’

Remark: It can be seen from the table and from the figure (3.1) that the improvements here are not as remarkable as the improvements in the Gaussian case. We have found no real justification for why this should be the case.

3.1.1 Data Examples

We shall illustrate the improvement of our loss estimator using two sets of Poisson data.

(1) Let us first consider ‘The number of deaths by horsekicks in the Prussian army’ data which are classically used to illustrate a Poisson model and any relevant analysis (cf. Andrews and Herzberg, 1985). These data consist of the number of deaths by horsekicks per year recorded during the period 1875 to 1894, for each of the fourteen corps of the army. Let us look at the year 1879.
CHAPTER 3. POISSON DISTRIBUTION

Let \( X_i \) = the number of deaths by horsekicks in the \( i^{th} \) corp for the year 1879. The \( \lambda \)'s are treated as being known, although we really estimate them from the data by averaging over the twenty years. Table (3.2) contains the data.

\[
p = 14 = \gamma_{unb} ,
\]

\[
L_{-1}(X,\lambda) = 12.836 ,
\]

\[
\gamma(X) = p - \frac{(p-2)}{\sum X_i + 1} = 12.909 .
\]

The loss incurred by estimating \( L_{-1} \) by \( \gamma_{unb} \) is 1.355 and that incurred by \( \gamma \) is 0.0054. The risks of the estimators are 51.36, and 49.66 respectively.

Although our improved estimator does much better with respect to this particular realization, the overall improvement is not great, i.e., the loss incurred is very little, but there isn't that much reduction in risk. To illustrate this point, let us look at the whole ensemble of results, i.e., instead of looking only at the year 1879, we shall look at each of the 20 years. So we have 19 more problems which are the same as the one above. In all we now have 20 independently and identically distributed 14-dimensional Poisson variables with mean \( \lambda = (\lambda_1, \cdots, \lambda_{14})^T \). Thus we have 20 different \( L_{-1}(X,\lambda) \)s and \( \gamma(X) \)'s, but \( \gamma_{unb}(X) \) is the same for all the problems. Figure (3.2) plots the loss, and the unbiased and modified loss estimators for each of the 20 years. It can be seen that our estimator does better than the unbiased one in 12 of the 20 cases. The average loss (squared error) incurred by the unbiased loss estimator is 22.54 and that incurred by our modified estimator is 18.59. Here, the average is taken over the 20 years. When the modified estimator is plotted against the loss, a positive correlation is evident, and this is what gives our estimator a lower mean squared error; see figure (3.3).

(2) The second data set we consider is the 'Admissions to intensive care unit' data from Cox and Snell (1981,p.53-54). These data give the arrival times of patients at an intensive care unit. The data were collected by Dr.A.Barr of the Oxford Regional Hospital Board. The data are from February 1963 to March 1964, but we shall exclude the last month because the data for that month are incomplete. We will also neglect what hour of the day the arrivals occurred. We thus have just the frequency of arrivals per day for all of these thirteen months.
CHAPTER 3. POISSON DISTRIBUTION

Let $X_i$ = the number admitted on the first day of the $i^{th}$ month. We assume that the Poisson model holds true i.e. that $X_i \sim \text{Poisson}(\lambda_i)$. Once again, we treat the $\lambda_i$ s as being known, but calculate them from the data; $\lambda_i$ is the average over all days of the $i^{th}$ month. Table (3.3) gives the corresponding reduction of the data.

$$p = 13 = \gamma_{unb},$$

$$L_{-1}(X, \lambda) = 8.533,$$

$$\gamma(X) = p - \frac{(p - 2)}{\sum X_i + 1} = 11.778.$$

The loss incurred by $\gamma_{unb}$, $(\gamma_{unb} - L_{-1})^2$ is 19.957 and that incurred by $\gamma$ is 10.53. The risks of the unbiased and the corrected estimators, treating the $\lambda_i$ s as being known, are 48.149 and 46.258 respectively.

Thus, although the loss incurred by our estimator is quite a bit smaller than that of the unbiased loss estimator, comparison of the risks indicates that the overall improvement attained by our estimator is not huge. We shall again consider all the cases together. Instead of only considering the number admitted on the first day, we shall consider each of the first, second, ..., twenty-eighth days, thus giving rise to 27 more problems like the one above. The actual loss, unbiased estimator of loss and our modified estimator of loss are all plotted in figure (3.4), and the correlation picture is in figure (3.5). Here, in 18 of the 28 cases, our estimator does better than the unbiased one. The average loss incurred by our estimator is 48.04 versus the 51.47 incurred by the unbiased estimator. There is also a positive correlation between the actual loss and our estimator, but it is not as marked as in the previous case.
### Table 3.2: The Prussian Army Data

<table>
<thead>
<tr>
<th>i</th>
<th>$X_i$</th>
<th>$\lambda_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.55</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0.85</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.65</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1.25</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

### Table 3.3: The Intensive Care Data

<table>
<thead>
<tr>
<th>i</th>
<th>$X_i$</th>
<th>$\lambda_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>13/28</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>16/31</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>12/30</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>18/31</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>23/30</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>16/31</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>15/31</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>17/30</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>17/31</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>28/30</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>32/31</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>26/31</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>17/29</td>
</tr>
</tbody>
</table>
CHAPTER 3. POISSON DISTRIBUTION

This improved estimator has a formal Bayes interpretation in the sense of being the limit of a sequence of proper Bayes estimators. The improper prior which is the limit of the corresponding sequence of proper priors is:

$$
\pi_u(\lambda) : \lambda_i|u \overset{ind}{\sim} \frac{u}{1-u} e^{-\frac{u}{1-u}\lambda_i}, \quad 0 < u < 1
$$

(3.6)

and

$$
g(u) \propto \frac{1}{u}, \quad 0 < u < 1.
$$

These are the kind of two-stage priors used by Ghosh (1983).

To establish the formal Bayes interpretation, we construct a sequence of estimators that improve over $\gamma_{unb}$, using the relative savings loss method (cf. Efron-Morris, 1973). This sequence of estimators uses the same $\pi_u$ as in (3.6), but will correspond to a sequence of Beta hyperpriors (on $u$). We then prove that the estimator given by (3.5) can be obtained as a limit of this sequence of Bayes estimators.

Given $u$, $\sum X_i \sim NB(p,u)$ and $S \Delta \sum X_i$ is sufficient for $u$. Let $\gamma \Delta \gamma_{unb} - c(S)$ be a candidate for an improved estimator, where $c(S)$ is any function of $S$. Let us denote the posterior loss estimator of $L_{-1}$, given $u$, by $\gamma_u$.

$$
L_{-1} = L_{-1}(X, \lambda) = \sum_i \lambda_i^{-1}(X_i - \lambda_i)^2,
$$

$$
\gamma_u = E_{\pi_u}(L_{-1}|X)
= p(1-u) + u^2(1-u)^{-1} \sum X_i.
$$

Note that $\gamma_u \rightarrow p$ as $u \rightarrow 0$. Let

$$
\eta(u, S) \Delta \gamma_{unb} - \gamma_u
= pu - u^2(1-u)^{-1} \sum X_i.
$$

The integrated risk of $\gamma$ with respect to the prior $\pi_u$ is

$$
r(\gamma, \pi_u) = E_u \left[ E_{\pi_u} \{(\gamma - L_{-1})^2}\right].
$$

It can be shown that the difference in the integrated risks $r(\gamma, \pi_u) - r(\gamma_u, \pi_u)$ can be written as

$$
r(\gamma, \pi_u) - r(\gamma_u, \pi_u) = E_u \left[ (\gamma - \gamma_u)^2 \right].$$
(See the argument following equation (3.11).) The relative savings loss is defined as

\[ RSL(u, \gamma) = \frac{r(\gamma, \pi_u) - r(\gamma_u, \pi_u)}{r(\gamma_{unb}, \pi_u) - r(\gamma_u, \pi_u)} = \frac{E_u[c(S) - \eta(u, S)]^2}{E_u[\eta(u, S)]^2}. \]

It can be easily seen that \( E_u[\eta(u, S)]^2 = p u^2 (1 - u)^{-1} \). So we now consider the problem of estimating \( c(S) \) with respect to the loss function

\[ \frac{[c - \eta(u, S)]^2}{u^2(1 - u)^{-1}}. \]

We use the \( Beta(a, b) \) prior on \( u \) with \( b = 1 \).

\[ Beta(a, b) : \frac{u^{a-1}(1 - u)^{b-1}}{\beta(a, b)}. \]

The Bayes estimator of \( \eta(u, S) \) with respect to this prior, and with respect to the above loss, is

\[
c_a(S) = \frac{E[(1 - u)u^{-2}\eta(u, S)|S]}{E[(1 - u)u^{-2}|S]}
= \frac{\int_0^1 u^{p+a-3}(1 - u)^{S+1}[pu - u^2(1 - u)^{-1}S]du}{\int_0^1 u^{p+a-3}(1 - u)^{S+1}}
= \frac{p\beta(p + a - 1, S + 2)}{\beta(p + a - 2, S + 2)} - \frac{\beta(p + a, S + 1)}{\beta(p + a - 2, S + 2)}
= \frac{p(p + a - 2)}{p + a + S} - \frac{S(p + a - 2)(p + a - 1)}{(S + 1)(p + a + S)}.
\]

The equality in the second step is due to the fact that \( S|u \sim NB(p, u) \) and that \( u \sim Beta(a, 1) \). The third step follows from carrying out the integration, and the fourth step is a result of algebraic manipulation.

So, as \( a \to 0 \),

\[
c_a(S) \to c_*(S) = \frac{(p - 2)}{(S + 1)}.
\]

As \( a \to 0 \), the Beta prior becomes approximately proportional to \( u^{-1} \). So our improved estimator (3.5) is the limit of the sequence of estimators \( \gamma_{unb} - c_a(S) \), and (3.6) is the limit of the corresponding priors.
CHAPTER 3. POISSON DISTRIBUTION

3.2 Admissibility for $p \leq 2$

Having found improvements for $p > 2$, we suspect that $\gamma_{unb}$ is admissible for $p \leq 2$ and try to prove it. We shall use Blyth’s method. Here is a version of Blyth’s lemma that we can use (see e.g. Johnstone(1988), p.364, Berger(1985), p.547, Brown-Hwang(1982), p.209).

**Lemma 3.2** Assume the densities $p_\theta$ to be mutually absolutely continuous. Then, $\gamma(X)$ is admissible for estimating $L(\delta(X), \theta)$ under squared error distance if there exists $\theta_o \in \Theta$ and a sequence $G_n(d\theta)$ of non-negative measures satisfying

$$G_n\{\theta : |\theta - \theta_o| \leq 1\} \geq c > 0, \ r(G_n) < \infty, \ and$$

$$r(\gamma, G_n) - r(G_n) \to 0 \ as \ n \to \infty \quad (3.7)$$

Let $g_n(x) = \int p_\theta(x) G_n(d\theta)$ be the marginal distribution of $X$. Then the above difference in integrated risks can also be written as

$$r(\gamma, G_n) - r(G_n) = \iint \left[ (\gamma(x) - L)^2 - (\gamma G_n(x) - L)^2 \right] p_\theta(x) G_n(d\theta) \nu(dx)$$

$$= \int [\gamma(x) - \gamma G_n(x)]^2 g_n(x) d\nu(x) \quad (3.8)$$

(Here, $L(\delta(x), \theta)$ is abbreviated by $L$.)

Let $\lambda_i | u \overset{ind}{\sim} \pi_u(\lambda_i)$, where $\pi_u(\lambda_i)$ is the same exponential prior of equation (3.6). Then, marginally, given $u$,

$$\sum X_i \sim NB(p, u),$$

$$\lambda_i | X \overset{ind}{\sim} \text{Gamma}\left((1 - u)^{-1}, X_i + 1\right)$$

and

$$\hat{L} = E(L|X) = \sum E\left[\lambda_i^{-1}(X_i - \lambda_i)^2|X_i\right]$$

$$= \sum\left[\frac{X_i^2}{X_i(1-u)} - 2X_i + (1 + X_i)(1 - u)\right]$$

$$= (1 - u)^{-1}\sum X_i - 2\sum X_i + (1 - u)(\sum X_i + p).$$
Let \( \phi(\sum X_i) = p - E(L|X) \)
\[ = p - (1 - u)^{-1}p\bar{X} + 2p\bar{X} - (1 - u)(p\bar{X} + p). \]

Since the second difference of a linear function is zero,
\[ \phi(p\bar{X} + 1) - 2\phi(p\bar{X}) + \phi(p\bar{X} - 1) = 0. \]
Hence from equation (3.4), we have
\[ R(\hat{L}) = R(\gamma_{unb}) + E \left[ \phi^2(p\bar{X}) + 2p(1 - u)^{-1}p_1 - (1 - u)^{-1}p_2 - (1 - u)^{-1}p_3 \right] \]
and from the remark in Lemma (3.2), the difference in integrated risks is given as
\[ r(\hat{L}) - r(\gamma_{unb}) = E \sum X \left[ \phi^2(p\bar{X}) \right], \]
where the expectation on the right is the expectation of \( \phi^2(p\bar{X}) \) under the Negative Binomial distribution \( NB(p, u) \) of \( p\bar{X} = \sum X_i. \) In order to apply Blyth's lemma, \( \pi_u(\cdot) \) should satisfy Blyth's condition stated in Lemma (3.2). Hence we let
\[ g_u(\lambda_i) = \frac{1 - u}{u} \frac{u}{1 - u} e^{-\frac{u}{1 - u} \lambda_i}, \quad \lambda_i \geq 0, \]
so that \( G_u\{\lambda_i : \lambda_i \leq 1\} > 0 \) for all \( i. \) Then, the marginal distribution of \( \sum X_i \) given \( u \) is
\[ \sum X_i \sim \left( \frac{1 - u}{u} \right)^p NB(p, u). \]
In order to prove admissibility of \( \gamma_{unb} = p \), we need to prove that
\[ u^{-p}(1 - u)^p E_Z[\phi^2(Z)] \to 0 \text{ as } u \to 0 \]
where \( Z \sim NB(p, u). \) Now,
\[ u^{-p}(1 - u)^p E_Z[\phi^2(z)] = u^{-p}(1 - u)^p E \left[ p - \frac{z}{1 - u} + 2z - (1 - u)(p + z) \right]^2 \]
\[ = \frac{u^{-p}(1 - u)^p}{p^{-1}(1 - u)^u}. \]
This goes to 0 only when \( p = 1 \), and tends to the constant 2 for \( p = 2 \).
CHAPTER 3. POISSON DISTRIBUTION

Hence this sequence of priors proves admissibility only for \( p = 1 \). A different, more sophisticated prior would be needed to prove the admissibility in two dimensions.

**Admissibility for \( p = 2 \):**

Let \( \pi(\lambda) \) be the prior on \( \lambda \). Here, we shall first try to bound the difference in integrated risks involved in Lemma (3.2), by an expression involving the derivatives of \( \pi \). Our task will then be reduced to finding a proper sequence of priors \( \{\pi_n\} \) such that in the limit as \( n \to \infty \), \( \pi_n \) tends to a constant.

**Lemma 3.3**

\[
r(p, \pi) - r(\pi) \leq \int \frac{[\lambda \cdot \nabla^2 \pi(\lambda) + \nabla \cdot \pi(\lambda)]^2}{\pi(\lambda)} d\lambda
\]

where

\[
\nabla \cdot \pi(\lambda) = \sum \frac{\partial}{\partial \lambda_i} \pi(\lambda) \quad \text{and} \quad \nabla^2 \pi(\lambda) = \left( \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) \right)_{p \times 1}.
\]

**Proof:** Let

\[
P_\lambda(X) = \frac{e^{-\sum \lambda_i X_i}}{\prod X_i! \lambda_i^{X_i}}
\]

and let the posterior expected loss be defined as

\[
I_\pi \triangleq E_\pi \left[ \sum \lambda_i^{-1} (X_i - \lambda_i)^2 | X \right] = \frac{\int \sum \lambda_i^{-1} (X_i - \lambda_i)^2 P_\lambda(X) \pi(\lambda) d\lambda}{\int P_\lambda(X) \pi(\lambda) d\lambda}.
\]

The function \( P_\lambda(X) \) can also be treated as a \( \text{Gamma}(1, X_i + 1) \) density of \( \lambda_i \). We can use the following identities that hold for the Gamma density (cf. Hudson, 1978):

\[
E[(\lambda_i - X_i - 1) \pi(\lambda)] = E[\lambda_i \pi'_i(\lambda)],
\]

\[
E[(\lambda_i - X_i - 1)^2 \pi(\lambda)] = E[\lambda_i \pi(\lambda) + \lambda_i \pi'_i(\lambda_i) + \lambda_i^2 \pi''_i(\lambda_i)]
\]

(3.9)
for \( i = 1, \cdots, p \), provided \( \pi : \mathbb{R}^p \rightarrow \mathbb{R} \) is continuously differentiable and all the above expectations exist. In the above,

\[
\pi_i'(\lambda) = \frac{\partial}{\partial \lambda_i} \pi(\lambda) = \nabla_i \pi(\lambda) \quad \text{and} \\
\pi_i''(\lambda) = \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) = \nabla_i^2 \pi(\lambda).
\]

So, assuming that \( \pi \) satisfies the above regularity conditions, the posterior expected loss is

\[
I_\pi = p + \frac{\int \sum_i [\lambda_i \nabla_i^2 \pi(\lambda) + \nabla_i \pi(\lambda)] P_\lambda(X)d\lambda}{\int P_\lambda(X) \pi(\lambda)d\lambda}
= p + E\left[ \frac{\sum \{\lambda_i \nabla_i^2 \pi(\lambda) + \nabla_i \pi(\lambda)\}}{\pi(\lambda)} \right]_X.
\]

Therefore, for a sequence of priors \( \Pi_n(d\lambda) = \pi_n(\lambda)d\lambda \) such that \( \pi_n \) satisfies the above regularity conditions for all \( n \),

\[
I_{\pi_n} = p + E_{\pi_n} \left[ \frac{\sum \lambda_i \nabla_i^2 \pi_n + \nabla_i \pi_n}{\pi_n} \right]_X
= p + \mu_n(X), \quad \text{say}.
\tag{3.10}
\]

If \( \gamma_\pi = E_\pi(L|X) \), and \( \gamma \) is any other estimator,

\[
r(\gamma, \pi) - r(\gamma_\pi) = E(\gamma - \gamma_\pi)^2.
\tag{3.11}
\]

This can be proved as follows using the fact that \( E_{\lambda|X}(\gamma_\pi - L) = 0 \).

\[
r(\gamma, \pi) - r(\gamma_\pi) = E_X E_{\lambda|X} \left[ (\gamma - L)^2 - (\gamma_\pi - L)^2 \right]
= E_X E_{\lambda|X} \left[ (\gamma - \gamma_\pi)^2 + 2(\gamma_\pi - L)(\gamma - \gamma_\pi) \right]
= E_X E_{\lambda|X} \left[ (\gamma - \gamma_\pi)^2 \right]
= E(\gamma - \gamma_\pi)^2.
\]

For the particular case \( \gamma \equiv p \),

\[
r(p, \pi_n) - r(\pi_n) = E[\mu_n^2(X)]
\leq \int \frac{[\lambda \cdot \nabla^2 \pi_n(\lambda) + \nabla \cdot \pi_n(\lambda)]^2}{\pi_n(\lambda)}d\lambda.
\tag{3.12}
\]
CHAPTER 3. POISSON DISTRIBUTION

Now we have to find an appropriate sequence of priors \( \{\pi_n\} \) such that the integrand in the inequality (3.12) vanishes as \( n \to \infty \) and the integral is finite. We shall adapt to our case the prior used by Johnstone in proving admissibility of \( \gamma_{unb} (=p) \) in the Gaussian situation for \( p = 4 \). This case of \( p = 4 \) is a borderline case for the Gaussian distribution, in the sense that it separates admissibility from inadmissibility. The corresponding Poisson borderline is \( p = 2 \). This phenomenon of dimension-doubling when going from the Poisson to the Gaussian distribution will be further illustrated and explained in Chapter 5. Also see the remark at the end of this section.

Let

\[
\pi_n(\lambda) = \begin{cases} 
2^{k-1} \eta_n^k & 0 \leq \eta_n \leq 1/2 \\
1 - 2^{k-1}(1 - \eta_n)^k & 1/2 \leq \eta_n \leq 1 
\end{cases}
\]  (3.13)

Here \( \eta_n \) is defined as

\[
\eta_n(\lambda) = \begin{cases} 
1 & \Lambda < 1 \\
1 - \frac{\ln \Lambda}{\ln n} & 1 \leq \Lambda < n \\
0 & \Lambda > n 
\end{cases}
\]  (3.14)

for \( n = 2, 3, \ldots \), and \( \Lambda = \sum_i^p \lambda_i \).

Let us first derive an upper bound on the integrand of (3.12). We start with the two cases \( \eta_n \in [0, 1/2) \) and \( \eta_n \in [1/2, 1] \) separately, but it turns out that there exists a common upper bound. This common bound tends to 0 as \( n \to \infty \), and so does the integrand. Therefore, proving that the integral in (3.12) is bounded and invoking the Dominated Convergence Theorem, we can conclude that the difference in integrated risks in (3.12) tends to 0 with increasing \( n \). The transformation that we apply to carry out the integration is

\[
\theta_i = \lambda_i \Lambda^{-1}, \ i = 1, \cdots, p - 1, \ \text{and} \ \theta_p = \Lambda. \]  (3.15)

For \( 0 \leq \eta_n < 1/2 \),

\[
\nabla_i \pi_n = 2^{k-1} k \eta_n^{k-1} \left( -\frac{1}{\Lambda \ln n} \right) I\{1 < \Lambda \leq n\},
\]

\[
\nabla_i^2 \pi_n = \left[ 2^{k-1} k(k - 1) \eta_n^{k-2} \frac{1}{\Lambda^2 \ln^2 n} + \frac{2^{k-1} k \eta_n^{k-1}}{\Lambda^2 \ln n} \right] I\{1 < \Lambda \leq n\}.
\]
And for \( 1/2 \leq \eta_n \leq 1 \),
\[
\nabla_i \pi_n = 2^{k-1} k (1 - \eta_n)^{k-1} \left( -\frac{1}{\Lambda \ln n} \right) I\{1 < \Lambda \leq n\},
\]
\[
\nabla_i^2 \pi_n = \left[ 2^{k-1} k (k - 1)(1 - \eta_n^{k-2}) \left( -\frac{1}{\Lambda^2 \ln^2 n} \right) + \frac{2^{k-1} k (1 - \eta_n^{k-1})}{\Lambda^2 \ln n} \right] I\{1 < \Lambda \leq n\}.
\]

We shall consider \( k \geq 4 \). After some algebraic manipulation, we can bound the integrand in (3.12) as follows, for \( 0 \leq \eta_n \leq 1 \) and for all \( n \) :-
\[
\frac{(\lambda \cdot \nabla^2 \pi_n + \nabla \cdot \pi_n)^2}{\pi_n} \leq \frac{B_k}{\Lambda^2 \ln^4 n} [2(k - 1) + (p - 1) \ln n]^2 I\{1 < \Lambda \leq n\}
\]
\[
\leq \frac{B_k}{\Lambda^2 \ln^4 (\Lambda \lor 2)} [2(k - 1) + (p - 1) \ln (\Lambda \lor 2)]^2 I\{1 < \Lambda\}
\]
where \( B_k \) is a constant depending on \( k \). The second inequality above holds because \( n \geq (\Lambda \lor 2) \), and because \( I\{1 < \Lambda \leq n\} \leq I\{1 < \Lambda\} \).

Making the transformation (3.15) and integrating out \( \theta_1, \ldots, \theta_{p-1} \),
\[
\int_{\Lambda \geq 1} \frac{B_k [2(k - 1) + (p - 1) \ln (\Lambda \lor 2)]^2}{\Lambda^2 \ln^4 (\Lambda \lor 2)} d\lambda
\]
\[
= B_k \int_{\theta_p=1}^{\infty} \int_{\theta_{p-1}=1}^{\theta_p} \cdots \int_{\theta_1=0}^{1} \frac{[2(k - 1) + (p - 1) \ln (\theta_p \lor 2)]^2 \theta_{p-1}^{p-1}}{\theta_p^2 \ln^4 (\theta_p \lor 2)} d\theta_1 d\theta_2 \cdots d\theta_p
\]
\[
= \int_{\theta_p=1}^{\infty} \frac{[2(k - 1) + (p - 1) \ln (\theta_p \lor 2)]^2 \theta_{p-3}^{p-3}}{\ln^4 (\theta_p \lor 2)} d\theta_p
\]
The above integral is finite for \( p < 3 \), and in particular for \( p = 2 \).

Hence by the Dominated Convergence Theorem, and from inequality (3.12), it follows that
\[
r(p, \pi_n) \rightarrow 0 \text{ for } p = 2.
\]
In fact the above also holds for \( p = 1 \).

Thus we have proved the following theorem -

**Theorem 3.1** \( \gamma_{um} = p \) is admissible as an estimator of the loss

\[
L_{-1}(X, \lambda) = \sum \lambda_i^{-1} (X_i - \lambda_i)^2
\]

for \( p \leq 2 \), and is inadmissible for \( p \geq 3 \).
CHAPTER 3. POISSON DISTRIBUTION

Remarks: (1) We shall elaborate on this later in Chapter 5, nonetheless this is an observation worth making here. Using the polydisc transform as in Johnstone and McGibbon (1989), we can transform the above admissibility problem to the Gaussian case and then use Johnstone's result (1988) almost directly. The integral in the inequality (3.12) transforms to the corresponding Gaussian integral in $2p$ dimensions, and hence, admissibility for $p = 4$ in the Gaussian case implies admissibility for $p = 2$ in this Poisson setting.

(2) The admissibility part of the above theorem will be derived as a special case of a more general theorem in Chapter 5. The method of proof there will naturally be more general. The above admissibility proof for $p = 2$ is presented here at the risk of repetition because the procedure is easy to follow in this specific case. Having explained the procedure in detail here, we will not spell out all the details in the general setting. Here it was also interesting to see how the Gaussian methods could be adapted to this case.

3.3 Cleveenson-Zidek point estimator

Let us now consider the admissible Cleveenson-Zidek point estimator in place of the maximum likelihood point estimator. We shall consider the problem of estimating the loss incurred by this point estimator and see if we can extend any of the previously obtained results to this situation. Cleveenson and Zidek (1975) proved that for the information-normalized loss, the estimator

$$
\delta(X) = \left(1 - \frac{p + \beta - 1}{\sum X_i + p + \beta - 1}\right) X
$$

improves over $X$ as an estimator of $\lambda$ for $p > 1$, and is admissible for all $\beta \geq 0$, $p \geq 2$. For simplicity, we shall consider the case $\beta = 0$.

The unbiased estimator of the loss

$$
L = \sum \lambda_i^{-1}(\delta_i(X) - \lambda_i)^2
$$

is

$$
\gamma_{unb}(X) = p - \frac{(p-1)^2}{\sum X_i + p - 1} \left[ 1 + \frac{1}{\sum X_i + p} \right].
$$
CHAPTER 3. POISSON DISTRIBUTION

Note that, unlike the unbiased loss estimator corresponding to the James-Stein estimator in the Gaussian case (see Johnstone, 1988), this estimator doesn't exhibit undesirable behavior on any part of the sample space; it is positive everywhere.

The Clevenson-Zidek estimator is Bayes for the prior

$$\lambda_i | u \sim \left( \frac{u}{1-u} \right) e^{-\frac{1}{\alpha u^\alpha}}, \quad u \sim u^{\alpha - 1}, \quad \alpha = -1,$$

i.e.

$$\pi(\lambda) = \frac{\Gamma(p-1)}{(\sum \lambda_i)^{p-1}}.$$

The posterior estimator of \( L \) corresponding to this prior is

$$\gamma_{post} = p - \frac{(p-1)^2}{\sum X_i + p - 1}.$$

We shall later prove the admissibility of \( \gamma_{post} \) in Chapter 5. It has not been possible to obtain any (in)admissibility results for \( \gamma_{unb} \), but the computational results presented in figures (3.6) and (3.7) suggest that even if \( \gamma_{unb} \) is inadmissible, the inadmissibility is not at all serious. The admissible estimator \( \gamma_{post} \) doesn't improve over \( \gamma_{unb} \) for small values of \( \sum \lambda_i \), and the improvement for large values of \( \sum \lambda_i \) is quite small. So, the inadmissibility, if at all, would be only pathological. In fact, if we, a priori, expect the \( \lambda_i \)'s to be small, it would be better to use \( \gamma_{unb} \) although \( \gamma_{post} \) is admissible. This serves to illustrate the inadequacy of admissibility as an optimality criterion.

Remark: Note that in the normal distribution, Johnstone's (1988) improved loss estimator corresponding to the maximum likelihood point estimator is not conservative, i.e. it doesn't satisfy the inequality (1.2), but the improved estimator for the James-Stein point estimator is conservative. Here in case of the Poisson distribution too, the improved estimator of the loss incurred by the mle fails to be conservative. But the (admissible) posterior loss estimator corresponding to the Clevenson-Zidek point estimator satisfies the condition (1.2) and is conservative.

Plotted in figures (3.6) and (3.7) is the percentage change in risk with respect to the unbiased loss estimator, of two estimators \( \gamma_{post} \) and \( \gamma \). The estimator \( \gamma \)
is such that it behaves like the unbiased estimator near zero and like the posterior estimator away from zero, i.e. $\gamma(0) = 1/p = \gamma_{unb}(0)$, and $\gamma(p-1) = (p+1)/2 = \gamma_{post}(p-1)$. The $\gamma$ used here is

$$\gamma(\sum X_i) = p - \frac{(p-1)(p-2)}{\sum X_i + p - 1} - \frac{(p-1)(2p-1)}{(\sum X_i + p)(\sum X_i + p - 1)}$$

The percentage change in risk of an estimator $\gamma$ with respect to $\gamma_{unb}$ is

$$\frac{R(\gamma) - R(\gamma_{unb})}{R(\gamma_{unb})} \times 100 \%$$

The changes are plotted for values of $p$ from 3 to 10.
Figure 3.1: Percentage improvement of $\gamma$ over 'p' for $p = 3, \ldots, 10$. The X-axis is $(\sum \lambda_i)/p$. 
Figure 3.2: Actual loss (*), Modified loss estimator (o) and the Unbiased loss estimator (-) for each of the 20 years for the Prussian army data
Figure 3.3: Estimated vs Actual loss for the Prussian army data. X axis: Actual loss.
Figure 3.4: Actual loss (*), Modified loss estimator (o) and the Unbiased loss estimator (−) for each of the first 28 days for the Intensive Care data
Figure 3.5: Estimated vs Actual loss for the Intensive Care data. X axis: Actual loss
Figure 3.6: Percentage change in risk of (a) $\gamma_{post}$ and (b) $\gamma$ over $\gamma_{unb}$ for the Clevenson-Zidek point estimator, $p = 3, 4, 5, 6$. X-axis: $(\sum \lambda_i)/p$. 
% change in risk for $p = 7$  
% change in risk for $p = 8$

% change in risk for $p = 9$  
% change in risk for $p = 10$

Figure 3.7: Percentage change in risk of (a) $\gamma_{post}$ and (b) $\gamma$ over $\gamma_{unb}$ for the Clevenson-Zidek point estimator, $p = 7, 8, 9, 10$. X-axis: $(\sum \lambda_i)/p$. 
Chapter 4

Hudson's Subclass Of
Exponential Families

Let us now see if we can further broaden the scope of the previous results of loss estimation to other distributions in order to be able to construct good estimators in a wide variety of settings. We shall first consider Hudson's subclass (cf. Hudson, 1978) for its structure. It consists of distributions for which the (Gaussian) Stein-like identities hold, and we only need these identities in finding improvements over unbiased loss estimators. The following description will make this clear.

Consider $X = (X_1, \ldots, X_p)^T$, $X_i$ are independent, and have probability densities of the form

$$f_{\theta_i}(x) = \exp\{\theta_i \tau(x) - \psi(\theta_i)\} k_\theta(x)$$ (4.1)

for $i = 1, \ldots, p$. Hudson considers a subclass with densities $f_{\theta}$ for which

$$E_\theta\{(X - \mu)g(X)\} = E_\theta\{a(X)g'(X)\}$$ (4.2)

holds for some $a : \mathbb{R} \rightarrow \mathbb{R}$ and for all absolutely continuous functions $g$ such that $E|ag'| < \infty$, $E_\theta(X) = \mu$. So, the density $f_{\theta}$ is

$$f_{\theta}(x) = \exp\left\{\mu \int [a(x)]^{-1} dx - \chi(\theta)\right\} [a(x)]^{-1} \exp\left\{- \int x[a(x)]^{-1} dx\right\}.$$ (4.3)

If $b(x) = \int [a(x)]^{-1} dx$, $B = b(X)$ has the density

$$f_{\theta}(b) = \exp\{\mu b - \psi(\mu)\} k(b).$$ (4.4)
Hence from identity (4.2), for a function $h$ satisfying the conditions therein,

$$
E\{(t(B) - \mu)h(B)\} = E\{h'(B)\}, \quad t(B) = -\frac{k'(B)}{k(B)}.
$$

But $t(B) = X$, and $X$ is unbiased for $\mu$ and also sufficient. So, from equation (4.2), setting $g(X) = X$ implies

$$
E\{a(X)\} = E\{(X - \mu)^2\}.
$$

Distributions belonging to this class are the gamma (with shape parameter) and the inverse Chi-square, besides the Gaussian. Let $X_i$ have densities as in equation (4.3), so that equation (4.2) holds. Let $S = \sum_{i=1}^{p} b^2(X_i)$. In the same paper Hudson proved that if $g_i(X) = -(p - 2)S^{-1}b(X_i)$, then under squared error loss, $X + g(X)$ improves upon $X$, as an estimator of $\mu$. We shall consider the general estimator $g_i(X) = -cS^{-1}b(X_i)$, where $c$ can be 0 or $p - 2$.

Let us first note that

$$
\frac{\partial}{\partial(x_i)} \left( \frac{1}{S} \right) = -\frac{1}{S^2} \frac{2b(x_i)}{a(x_i)},
$$

and

$$
\frac{\partial^2}{\partial x_i \partial S} \left( \frac{1}{S} \right) = \frac{3b^2(x_i)}{S^3a^2(x_i)} - \frac{2}{S^2} \left[ \frac{1}{a^2(x_i)} - \frac{a'(x_i)b(x_i)}{a^2(x_i)} \right].
$$

Now using (4.2), (4.6), and (4.7), the unbiased estimator of the loss

$$
L(X + g(X), \mu) = \sum_{i=1}^{p} \left( X_i - cS^{-1}b(X_i) - \mu \right)^2
$$

is

$$
\gamma_{unb} = \sum_{i=1}^{p} a(X_i) + \frac{[c^2 - 2c(p - 2)]}{S}.
$$

Applying (4.2) twice gives the identity

$$
E_{\mu}\{(X_i - \mu_i)^2h(X)\} = E\{a(X_i)[a'(X_i)h'_i(X) + a(X_i)h''_i(X) + h(X)]\}
$$

for a function $h : \mathbb{R}^p \rightarrow \mathbb{R}$, such that $h$ is continuously differentiable, and all the relevant expectations exist. In the above equation,

$$
h'_i(X) = \partial h/\partial X_i \text{, and } h''_i(X) = \partial^2 h/\partial X_i^2.
$$
We now want to find an estimator that improves over $\gamma_{\text{unb}}(X)$ as an estimator of $L$. Let us consider $\gamma(X) = \gamma_{\text{unb}}(X) - d / S$, where the constant $d$ is to be determined so that $\gamma(X)$ gives the maximum risk improvement over $\gamma_{\text{unb}}(X)$. Using the identities, we shall obtain the unbiased estimator of the difference in risks of $\gamma_{\text{unb}}$ and $\gamma$, and then determine the $d$ that makes this unbiased estimator of the difference most negative. To begin with, let us note that

$$R(\gamma) = R(\gamma_{\text{unb}}) + E\left\{ \frac{d^2}{S^2} + \frac{2d}{S} [L - \gamma_{\text{unb}}] \right\}.$$  \hspace{1cm} (4.11)

Substituting (4.9) into (4.11), and using (4.10) with the assistance of (4.7) and (4.8), the difference in risks can be written as

$$R(\gamma) - R(\gamma_{\text{unb}}) = \left[ d^2 + 4d(4 - p + 2c) \right] E\left( \frac{1}{S^2} \right)$$  \hspace{1cm} (4.12)

If $c = 0$, maximum improvement is attained for $d = 2(p - 4)$, and the improvement in risk is $4(p - 4)^2 E(S^{-2})$. For the shrinkage estimator, $c = p - 2$, and then the optimal value of $d$ is $-2p$. The improvement in risk in this case is $4p^2 E(S^{-2})$. In both these cases improvements are obtained only for $p > 4$. We have thus proved the following proposition:

**Proposition 4.1**

$$\gamma(X) = \sum a(X_i) - \frac{2(p - 4)}{S}$$

improves over $\gamma_{\text{unb}} = \sum a(X_i)$, as an estimator of $L(X, \mu)$ for $p > 4$ and the improvement in risk is

$$4(p - 4)^2 E\left( \frac{1}{S^2} \right).$$

For the estimator $X + g(X)$ with $g_i(X) = -(p - 2)S^{-1}b(X_i)$,

$$\gamma(X) = \gamma_{\text{unb}}(X) + \frac{2p}{S}$$

improves over

$$\gamma_{\text{unb}}(X) = \sum a(X_i) - \frac{(p - 2)^2}{S}$$

as an estimator of $L(X + g(X), \mu)$ for all $p > 4$. The improvement in risk is $4p^2 E(S^{-2})$. 


For example, suppose $X_i \overset{ind}{\sim} \text{Gamma}(\mu_i, 1)$, i.e., the density of $X_i$ is

$$f_{\mu_i}(x) = \frac{e^{-\frac{x}{\mu_i}}}{\Gamma(\mu_i)}.$$ 

Here, $a(X_i) = X_i$ for all $i$, and $b(X_i) = \log X_i$. When the point estimator used is the maximum likelihood estimator $X$, the unbiased estimator of squared error loss is $\gamma_{\text{unb}} = \sum X_i$ and

$$\gamma(X) = \sum X_i - \frac{2(p-4)}{\sum \log^2 X_i}$$

improves over $\gamma_{\text{unb}}$. For the shrinkage point estimator

$$\delta(X) = X_i - \frac{p-2}{\sum \log^2 X_i},$$

$$\gamma_{\text{unb}} = \sum X_i - \frac{(p-2)^2}{\sum \log^2 X_i}$$

and the improved estimator is

$$\gamma(X) = \sum X_i - \frac{(p-2)^2}{\sum \log^2 X_i} + \frac{2p}{\sum \log^2 X_i}.$$ 

Improvements in both these cases are for all $p > 4$. 
Chapter 5

Unified Admissibility Theory

It would be very convenient to have one methodology that can be used to prove admissibility of loss estimators in a variety of situations. This is what we shall try to achieve in this section. It would lead to a result similar to that of Brown and Hwang (1982) who gave an admissibility proof for a variety of generalized Bayes estimators in the context of estimating the natural mean vector of an exponential family under a quadratic-form loss. We shall now try to achieve this for loss estimation.

The principal tools that we use here are Blyth's lemma, integration by parts, and the Cauchy-Schwarz inequality. Our aim is to propose a general purpose method of proving admissibility of loss estimators in exponential families, where the loss is squared error. To illustrate the procedure, we shall first establish admissibility results in the Gaussian setting. Then we shall develop proofs independently in case of the Poisson, for both the squared error and the information-normalized loss functions using the same method. We will also comment about the feasibility of generalizing the methodology even further.

Consider an exponential family

\[ f_\theta(x) = \exp\{\theta \cdot x - \psi(\theta)\} \]

relative to a \( \sigma \)-finite Borel measure \( \nu \) on \( \mathcal{X} \subset \mathbb{R}^p \), and \( \Theta \subset \mathbb{R}^p \) is the natural
parameter space,
\[ \Theta = \{ \theta : \int e^{\beta x} \nu(dx) < \infty \} . \]
Assume \( \Theta \) is open in \( \mathbb{R}^p \). The mean can be expressed as
\[ E_\theta(X) = \nabla \psi(\theta) . \]

First, using exponential identities derived from integration by parts (i.e., Green's theorem), we shall express the difference between posterior loss estimators corresponding to different priors in terms of the derivatives of the priors.

**Lemma 5.1** Assume \( \Theta = \mathbb{R}^p \). We shall henceforth make this assumption unless stated otherwise. From Hudson (1978) we have the identities:
\[
\int (\psi'_i(\theta) - X_i)g(\theta)f_\theta(X)d\theta = \int g'_i(\theta)f_\theta(X)d\theta \quad \text{and} \\
\int (\psi'_i(\theta) - X_i)^2g(\theta)f_\theta(X)d\theta = \int [g''_i(\theta) + g(\theta)\psi''_i(\theta)]f_\theta(X)d\theta
\]
for all \( i = 1, \ldots, p \), for a continuously differentiable \( g : \mathbb{R}^p \to \mathbb{R} \), for which all the integrals involved in the above identities exist. Using these identities we have the following relation that can be easily verified.
\[
\int \|X - \nabla \psi(\theta)\|^2 g(\theta)f_\theta(X)d\theta = \int [\Delta g(\theta) + g(\theta)\Delta \psi(\theta)]f_\theta(X)d\theta . \tag{5.1}
\]

In particular, if \( g(\theta) \equiv 1 \), the above reduces to
\[
\int \|X - \nabla \psi(\theta)\|^2 f_\theta(X)d\theta = \int \Delta \psi(\theta)f_\theta(X)d\theta .
\]

Hence, for \( L(\theta, X) = \|X - \nabla \psi(\theta)\|^2 \), the posterior loss estimate corresponding to the prior \( g \) is
\[
\gamma_g(X) \triangleq E_g(L(\theta, X)|X) = \frac{\int (\Delta g(\theta))(g(\theta))^{-1} f_\theta(X) g(\theta) d\theta}{\int f_\theta(X) g(\theta) d\theta} + \frac{\int \Delta \psi(\theta) f_\theta(X) g(\theta) d\theta}{\int f_\theta(X) g(\theta) d\theta} .
\]

To prove admissibility using Blyth's lemma (Lemma 3.2), we need to prove that
\[ B(g_n, \gamma_g) - B(g_n, \gamma_{g_n}) \to 0 \quad \text{as} \quad n \to \infty . \]
for a sequence of priors \( g_n \rightarrow g \). Using the representation in (3.8),

\[
B(g_n, \gamma_g) - B(g_n, \gamma_{g_n}) = \int (\gamma_g - \gamma_{g_n})^2 (I_x g_n) \nu(dx)
\]

where \( I_x h = \int h(\theta)f_0(x)d\theta \). Note that we shall use \( x \) instead of \( X \) when used as a variable of integration.

Instead of the point estimator \( X \), let us consider the generalized Bayes point estimator

\[
\delta_g(X) = E_g(\theta|X) = X + \frac{I_x \nabla g}{I_x g}
\]

(cf. Stein(1981), Brown-Hwang(1982)). Then the posterior estimator, with respect to the prior \( g \), of the loss incurred by this point estimator can be written as

\[
\gamma_g = E_g \left( \|X + \frac{I_x \nabla g}{I_x g} - \nabla \psi(\theta)\|^2 |X \right)
\]

\[
= E_g \left( \|X - \nabla \psi(\theta)\|^2 |X \right)
\]

\[
+ \frac{1}{I_x g} \int \frac{I_x \nabla g}{I_x g} \|g(\theta)f_0(X)d\theta
\]

\[
+ \frac{2}{I_x g} \int \left( \frac{I_x \nabla g}{I_x g} \right) \cdot (\nabla \psi(\theta) - X)g(\theta)f_0(X)d\theta
\]

\[
= \frac{I_x \Delta g}{I_x g} + \frac{I_x (g\Delta \psi)}{I_x g} - \| \frac{I_x \nabla g}{I_x g} \|^2.
\]

In the first step above, we just expand the square and separate the terms. Then we apply Lemma (5.1) to the first term, and use the first exponential identity stated in the same Lemma for the third term.

Define

\[
\gamma_{g_n} = E_{g_n} \left( \|X + \frac{I_x \nabla g_n}{I_x g_n} - \nabla \psi(\theta)\|^2 |X \right).
\]

Then the difference between \( \gamma_{g_n} \) and \( \gamma_g \) is

\[
\gamma_{g_n} - \gamma_g = \frac{I_x \Delta g_n}{I_x g_n} + \frac{I_x (g_n\Delta \psi)}{I_x g_n} - \| \frac{I_x \nabla g_n}{I_x g_n} \|^2
\]

\[
- \frac{I_x \Delta g}{I_x g} - \frac{I_x (g\Delta \psi)}{I_x g} + \| \frac{I_x \nabla g}{I_x g} \|^2.
\] (5.2)
The presence of the terms involving $\psi(\theta)$ could cause difficulties. So we shall first look at the simplest case, instead of directly considering the general case.

### 5.1 Normal Distribution

We have chosen the Gaussian case first because $\Delta\psi = \text{constant}$, and the difference in equation (5.2) above depends only on terms involving derivatives of the priors. So,

$$
\Delta_n = B(g_n, \gamma_n) - B(g_n, \gamma_g)
= \int \left( \frac{I_x \Delta g_n}{I_x g_n} - \frac{I_x \Delta g}{I_x g} - \frac{\| I_x \nabla g_n \|^2}{\| I_x \nabla g \|^2} + \frac{\| I_x \nabla g \|^2}{\| I_x \nabla g \|^2} \right)^2 I_x g_n \nu(dx). \tag{5.3}
$$

We assume that conditions of Lemma (5.1) also hold for $g_n$. Since $g_n \to g$ as $n \to \infty$, and due to the regularity conditions on $g_n$ and $g$, the integrand in (5.3) converges to zero for each $x$, as $n$ tends to infinity. Hence proving the boundedness of this integral, or deriving conditions on the prior that will bound the integral, will enable us to apply the Dominated Convergence Theorem and to conclude that the integral vanishes as $n \to \infty$. For this, we shall initially bound the integrand by four different terms, and then find conditions on the prior in order for these four terms to be bounded.

We consider the sequence of priors $g_n(\theta) = \xi_n(\theta)g(\theta)$, with $\xi_n(\theta) \leq 1$ for all $n$. We shall first substitute $\xi_n g$ for $g_n$ and separate out the terms that involve only the derivatives of $g$. Then for the remaining terms, we shall define $\xi_n$ to be of a specific form and obtain bounds on the terms separately.

Now, 

$$
\Delta g_n = \xi_n \Delta g + g \Delta \xi_n + 2 \nabla \xi_n \cdot \nabla g
$$
and the inequality
\[(a_1 + \cdots + a_m)^2 \leq m (a_1^2 + \cdots + a_m^2)\] (5.4)

used with \(m = 2\), together imply that
\[
\Delta_n \leq 2 \int \left( \frac{I_x (g \Delta \xi_n) + 2I_x (\nabla \xi_n \cdot \nabla g)}{I_x g_n} - \left\| \frac{I_x \nabla g_n}{I_x g} \right\|^2 \right)^2 I_x g_n \nu(dx)
+ 2 \int \left( \frac{I_x (\nabla \xi_n g)}{I_x g_n} - \frac{I_x \nabla g}{I_x g} + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 \right)^2 I_x g_n \nu(dx)
\Delta \leq 2A_n + 2B_n. \tag{5.5}
\]

Here \(B_n\) involves only the derivatives of \(g\), and we shall first deal with this term.

\[
B_n = \int \frac{1}{I_x g_n} \left( \int \left[ g_n \left( \frac{\Delta g}{g} - \frac{I_x \Delta g}{I_x g} + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 \right) \right]^2 \nu(dx)
\leq \int \frac{1}{I_x g} \left[ g \left( \frac{\Delta g}{g} - \frac{I_x \Delta g}{I_x g} + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 \right) \right]^2 \nu(dx)
\leq \int \frac{1}{I_x g} \left[ g \left( \frac{\Delta g}{g} - \frac{I_x \Delta g}{I_x g} + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 \right) \right]^2 \nu(dx). \tag{5.6}
\]

In the first step we multiply and divide the integrand by \((I_x g_n)^2\) and do some algebraic manipulation. The second step follows from the Cauchy-Schwarz inequality and the third step uses the fact that \(g_n \leq g\).

The boundedness of (5.6) implies that of \(B_n\). But we can find a condition which is much simpler and easily verifiable, although less general. This is shown in the following lemma.

Lemma 5.2
\[
\int \frac{(\Delta g)^2}{g} d\theta + \int \frac{||\nabla g||^4}{g^3} d\theta < \infty \Rightarrow (5.6) < \infty.
\]

Proof: All that needs to be done to prove the lemma is to prove that the sum of the integrals on the left is an upper bound for the integral in (5.6).
Expanding the square and simplifying,
\[
I_x \left[ g \left( \frac{\Delta g}{g} - \frac{I_x \Delta g}{I_x g} + \| \frac{I_x \nabla g}{I_x g} \|^2 \right) \right]^2 \leq I_x \frac{(\Delta g)^2}{g} - \frac{(I_x \Delta g)^2}{I_x g} + \frac{\| I_x \nabla g \|^4}{(I_x g)^3} + 2 \frac{I_x \nabla g}{I_x g} \| I_x \left[ g \left( \frac{\Delta g}{g} - \frac{I_x \Delta g}{I_x g} \right) \right] \|^2 \leq I_x \frac{(\Delta g)^2}{g} + \frac{\| I_x \nabla g \|^4}{(I_x g)^3}.
\]

Applying Hölder's inequality \( \int uv \leq \|u\|_p^{1/p} \|v\|_q^{1/q} \) with \( u = \nabla g / g^{3/4} \), \( v = g^{3/4} \), \( p = 4 \), \( q = 4/3 \) to the second term of the above expression,
\[
\frac{\| I_x \nabla g \|^4}{(I_x g)^3} \leq \frac{1}{I_x g} \left[ I_x \left( \frac{\| \nabla g \|^2}{g} \right) \right]^2 \leq I_x \left( \frac{\| \nabla g \|^4}{g^3} \right)
\]
and then noticing that
\[
\int [I_x h(\theta)] \nu(dx) = \int h(\theta)d\theta
\]
completes the proof.

So, the condition sufficient for the boundedness of \( B_n \) is
\[
\int \frac{(\Delta g)^2}{g} d\theta + \int \frac{\| \nabla g \|^4}{g^3} d\theta < \infty.
\]

Let us now look at \( A_n \). From (5.5) and using (5.4) for \( m = 3 \) along with repeated application of the Cauchy-Schwarz inequality, we first bound \( A_n \) by three integrals as follows.
\[
A_n \leq 3 \int \frac{[I_x(g(\theta) \Delta \xi_n(\theta))]^2}{I_x g_n(\theta)} \nu(dx) + 12 \int \frac{[I_x(\nabla \xi_n(\theta) \cdot \nabla g(\theta))]^2}{I_x g_n(\theta)} \nu(dx)
+ 3 \int \frac{\| I_x(\nabla g_n(\theta)) \|^4}{(I_x g_n(\theta))^5} \nu(dx)
\leq 3 \int \frac{(\Delta \xi_n(\theta))^2}{\xi_n(\theta)} g(\theta) d\theta + 12 \int \frac{\| \nabla \xi_n(\theta) \|^2}{\xi_n(\theta)} \frac{\| \nabla g(\theta) \|^2}{g(\theta)} d\theta
+ 3 \int \frac{\| \nabla (g(\theta) \xi_n(\theta)) \|^4}{g(\theta) \xi_n(\theta))^3} d\theta
\equiv 3C_n + 12D_n + 3E_n.
\] (5.7)
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

We now have the task of choosing an appropriate sequence \( \{ \xi_n \} \). Let

\[
\eta_n(\theta) = \begin{cases} 
1 & ||\theta|| \leq 1 \\
1 - \frac{\ln ||\theta||}{\ln n} & 1 \leq ||\theta|| \leq n \\
0 & ||\theta|| \geq n
\end{cases}
\] (5.8)

for \( n = 2, 3, ... \)

Using \( \xi_n = \eta_n^2 \) as in Brown and Hwang (1982) doesn't work here. It leads to smoothness difficulties in using Lemma 5.1 and since \( \eta_n \) vanishes at \( ||\theta|| = n \), the integrals in \( C_n, D_n \) and \( E_n \) also diverge. Note that this same difficulty was encountered by Johnstone (1988). Hence we shall also resort to the spline used by Johnstone. We also used it in Chapter 3 (prior (3.13)). We define

\[
\xi(\eta) = \begin{cases} 
2^{k-1} \eta^k & 0 \leq \eta \leq 1/2 \\
1 - 2^{k-1}(1-\eta)^k & 1/2 \leq \eta \leq 1
\end{cases}
\] (5.9)

for some integer \( k > 1 \). Define \( \xi_n(\theta) = \xi(\eta_n(\theta)) \). We note that the Brown-Hwang prior corresponds closely to \( k = 2 \).

We shall now compute the first two derivatives of this \( \xi \), and later on substitute these into the integrands of \( C_n, D_n \) and \( E_n \).

\[
2^{1-k}\xi'(\eta) = \begin{cases} 
k\eta^{k-1} & 0 \leq \eta \leq 1/2 \\
k(1-\eta)^{k-1} & 1/2 \leq \eta \leq 1
\end{cases}
\]

and

\[
2^{1-k}\xi''(\eta) = \begin{cases} 
k(k-1)\eta^{k-2} & 0 \leq \eta < 1/2 \\
-k(k-1)(1-\eta)^{k-2} & 1/2 \leq \eta \leq 1.
\end{cases}
\]

It can be easily seen that

\[
\nabla_\theta \xi = \xi'(\eta) \nabla_\eta \psi, \ \Delta_\theta \xi = \xi''(\eta) \|\nabla_\eta \|^2 + \xi'(\eta) \Delta(\eta)
\]

and

\[
\nabla_\eta \eta_n(\theta) = -\frac{1}{\ln n} \frac{\theta}{\|\theta\|^2} I\{1 < \|\theta\| \leq n\}, \ \Delta_\eta \eta_n = -\frac{1}{\ln n} \frac{(p-2)}{\|\theta\|^2} I\{1 < \|\theta\| \leq n\}.
\]

Here \( I \) denotes the indicator function. Note that \( \eta_n \) does not have a continuous first derivative and the second derivative of \( \xi \) is discontinuous at \( \eta = 1/2 \) for \( k \geq 2 \). We shall require \( k \geq 4 \) here.
Let us first derive a condition that assures boundedness of $C_n$. From the above computation of derivatives, we get

$$\frac{(\Delta \xi_n)^2}{\xi_n} = \frac{k^2}{\|\theta\|^4 \ln^4 n} \begin{cases} 2^{-1} \eta^{k-4} \left[(k - 1) - 2 \eta \ln n\right]^2 & 0 \leq \eta(\theta) < 1/2 \\ 2^{k-2} (1 - \eta)^{2k-2} \frac{(k - 1) - 2(1 - \eta) \ln n}{1 - 2^{k-2} (1 - \eta)^k} & 1/2 \leq \eta(\theta) \leq 1 \end{cases}$$

If $k \geq 4$, it follows that

$$\frac{(\Delta \xi_n)^2}{\xi_n} \leq \frac{M_k (1 + \ln^2 n)}{\|\theta\|^4 \ln^4 n} I\{1 < \|\theta\| \leq n\} \leq \frac{M'_k}{\|\theta\|^4 \ln^2 (\|\theta\| \vee 2)} I\{1 < \|\theta\|\}$$

for some constants $M_k$ and $M'_k$.

So the growth condition for finiteness of $C_n$ is

$$\int_{R^p - S} \frac{g(\theta)}{\|\theta\|^4 \ln^2 (\|\theta\| \vee 2)} d\theta < \infty \quad (5.10)$$

where $S = \{\theta : \|\theta\| \leq 1\}$.

We shall now look at the integrand involved in $D_n$. We have

$$\frac{\|\nabla \xi_n\|^2}{\xi_n} = \frac{k^2}{\|\theta\|^2 \ln^3 n} \begin{cases} 2^{-1} \eta^{k-2} \|\theta\| & 0 \leq \eta(\theta) < 1/2 \\ 2^{k-2} (1 - \eta)^{2k-2} \frac{1 - 2^{k-2} (1 - \eta)^k}{1 - 2^{k-2} (1 - \eta)^k} & 1/2 \leq \eta(\theta) \leq 1 \end{cases}$$

Hence it follows that

$$\frac{\|\nabla \xi_n\|^2}{\xi_n} \leq \frac{P_k}{\|\theta\|^2 \ln^2 n} I\{1 < \|\theta\| \leq n\} \leq \frac{P_k}{\|\theta\|^2 \ln^2 (\|\theta\| \vee 2)} I\{1 < \|\theta\|\}$$

for some constant $P_k$ depending on $k$.

Thus the condition needed in order that $D_n$ be bounded is

$$\int_{R^p - S} \frac{\|\nabla g(\theta)\|^2}{g(\theta) \|\theta\|^4 \ln^2 (\|\theta\| \vee 2)} d\theta < \infty \quad (5.11)$$

Let us note here that the condition (5.10), in conjunction with the condition $\int \|\nabla g\|^4 / g^3 d\theta < \infty$ included in bounding $B_n$ implies (5.11). Applying Cauchy-Schwarz inequality to the integrals, if the conditions for $B_n$ and $C_n$ hold, it follows that

$$\left\{ \int_{R^p - S} \frac{\|\nabla g(\theta)\|^2}{g(\theta) \|\theta\|^4 \ln (\|\theta\| \vee 2)} d\theta \right\}^2 \leq \infty$$
and this implies that (5.11) holds true.

Lastly, we shall consider $E_n$. Since

$$
\nabla (g_\xi \xi_n) = \xi_n \nabla g + g \nabla \xi_n ,
$$

$$
|| \nabla (g_\xi \xi_n) ||^4 \leq 8 || g \nabla \xi_n ||^4 + 8 || \xi_n \nabla g ||^4
$$

and hence

$$
\frac{|| \nabla (g_\xi \xi_n) ||^4}{(g_\xi \xi_n)^3} \leq 8 \frac{|| \nabla \xi_n ||^4}{\xi_n^3} g + 8 \frac{|| \nabla g ||^4}{g^3} .
$$

The integrability of $|| \nabla g ||^4 / g^3$ is already taken care of in the condition for boundedness of (5.6). So it remains to tackle the term involving $\xi_n$. From the computations done for the term $D_n$, it follows that

$$
\frac{|| \nabla \xi_n ||^4}{\xi_n^3} = \frac{k^4}{|| \theta ||^4 \ln^4 n} \left\{ \begin{array}{ll}
2^{k-1} \eta^{k-4} & 0 \leq \eta(\theta) < 1/2 \\
2^{k-4} (1-\eta)^{k-4} & 1/2 \leq \eta(\theta) \leq 1
\end{array} \right.
$$

And then we have

$$
\frac{|| \nabla \xi_n ||^4}{\xi_n^3} \leq \frac{Q_k}{|| \theta ||^4 \ln^4 n} I\{1 < || \theta || \leq n\} \leq \frac{Q_k (1 + \ln^2 n)}{|| \theta ||^4 \ln^4 n} I\{1 < || \theta || \leq n\} .
$$

The bound here being the same as the one for $C_n$ upto a constant, the same condition as for the boundedness of $C_n$ is also sufficient for $E_n$.

The above two conditions along with the Dominated Convergence Theorem imply that $\Delta_n \to 0$. Hence the following theorem:

**Theorem 5.1** In case of the Gaussian distribution, for

$$
L = || X + \frac{I_z \nabla g}{I_z g} - \theta ||^2 ,
$$

the posterior loss estimator

$$
\gamma_g = E_g(L | X)
$$
is admissible if the prior \( g \) is such that the following two conditions are satisfied –

\[
\int \frac{(\Delta g)^2}{g} d\theta + \int \frac{\| \nabla g \|^4}{g^3} d\theta < \infty \tag{5.12}
\]

\[
\int_{B^p - S} \frac{g(\theta)}{\| \theta \|^4 \ln^2(\| \theta \| \vee 2)} d\theta < \infty \tag{5.13}
\]

where \( S = \{ \theta : \| \theta \| \leq 1 \} \).

### 5.1.1 Verification of existing result

We can verify Johnstone’s result using the above theorem. The prior with respect to which the maximum likelihood point estimator is Bayes is \( g(\theta) \equiv 1 \) and the posterior loss estimator for \( \| X - \theta \|^2 \) is \( p \). The condition (5.12) of the theorem holds trivially for this \( g \) because \( \nabla g \) and \( \Delta g \) are identically zero. The integral in condition (5.13), after transformation to polar co-ordinates, becomes –

\[
\int_{1}^{\infty} \frac{r^{p-1}}{r^4 \ln^2(r \vee 2)} dr
\]

which is finite for \( p - 5 \leq -1 \), i.e. for \( p \leq 4 \). So, the estimator \( \gamma(X) \equiv p \) is admissible when \( p \leq 4 \).

### 5.1.2 The Strawderman prior case

Let us now apply the above theorem to the point and loss estimators derived from the Strawderman prior (cf. Berger and Lu, 1989). We consider here the loss associated with the generalized Bayes point estimator, and examine admissibility of the posterior estimator of this loss.

The Strawderman prior is

\[
g_m(\theta) = \int_{0}^{1} \left\{ \det[D(\lambda)] \right\}^{\frac{1}{2}} \exp \left\{ -\frac{(\theta - \alpha)^T D(\lambda)(\theta - \alpha)}{2} \right\} \lambda^{m-\frac{1}{2}} d\lambda
\]
where
\[ D(\lambda) = \left[ I - \lambda \frac{(I - B)}{\tau^2} \right]^{-1} - I. \]

The point estimator which is generalized Bayes with respect to this prior is
\[ \delta_m(X) = X - \frac{r_m(v(X))}{\tau^2 v(X)} (X - \mu(X)) \]
where
\[ \mu(X) = \alpha + B(X - \alpha), \]
\[ v(X) = (X - \alpha)^T (I - B)(X - \alpha)/\tau^2 \]
and
\[ r_m(v) = \frac{v \int_0^1 \lambda^m e^{-\frac{\lambda v}{2}} d\lambda}{\int_0^1 \lambda^{m-1} e^{-\frac{\lambda v}{2}} d\lambda}. \]
Here \( I \) is the identity matrix and \( B \) is an idempotent matrix of rank \( \eta \). The estimator \( \delta_m \) is admissible for \( m \geq (p - 2)/2 \) (cf. Berger, 1980); in particular, \( \delta_m \) is admissible when \( \eta = 0 \) and \( m = (p - 2)/2 \) for \( p \geq 3 \).

We consider the simplest case with \( \alpha = \mu(X) = 0 \), \( B = 0 \), \( \tau^2 = 1 \) and \( m = (p - 2)/2 \). This implies that \( v = \|X\|^2 \) and
\[ \delta(X) = X - \frac{r_{(p-2)/2}(v)}{v} X. \]

The prior with respect to which this \( \delta \) is Bayes is
\[
g(\theta) = \int_0^1 \left( \frac{\lambda}{1 - \lambda} \right)^{\frac{p}{2}} \exp \left\{ -\frac{\lambda}{1 - \lambda} \frac{\theta^T \theta}{2} \right\} \frac{1}{\lambda^2} d\lambda
= \Gamma\left(\frac{p}{2} - 1\right) \left(\frac{\theta^T \theta}{2}\right)^{-\frac{p}{2}+1}. \tag{5.14} \]

We shall now examine whether this prior satisfies the conditions of theorem (5.1). It can be seen that
\[ \nabla g(\theta) \sim O\left(\theta(\|\theta\|^2)^{-\frac{p}{2}}\right) \]
and that
\[ \Delta g(\theta) \equiv 0. \]
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

Thus, both the conditions can be verified to hold for this prior $g(\theta)$. Let

$$u(v) = \frac{r(p-2)/2(v)}{v}.$$ 

The loss estimator (estimating $L = \|\delta(X) - \theta\|^2$) that is posterior with respect to the prior $g(\theta)$ is

$$\gamma_{\text{post}} = 2 + vu(v) - vu^2(v).$$

We now have the following as a corollary of Theorem 5.1.

**Corollary 5.1** $\gamma_{\text{post}}$ is admissible as an estimator of $L$ for all $p \geq 3$.

Even for a general $m = (p - 2) / 2 + \beta$, $\beta > 0$,

$$g(\theta) = O\left(\|\theta\|^{-\frac{p}{2} + 1 - \beta}\right)$$

and the conditions can be verified to hold. So the corollary holds true in this general case too.

**Remarks:** (1) The same integral as in (5.14), but with the integration limits 0 and $\infty$, gives the prior whose posterior is asymptotically equivalent to the James-Stein point estimator. The James-Stein estimator is, of course, not generalized Bayes for any prior. We also note that

$$\gamma_{\text{post}} \sim \gamma_{\text{unb}}^{JS} \text{ as } \|X\|^2 \to \infty.$$

(2) An interesting observation here is that the prior yielding an admissible point estimator also yields an admissible loss estimator.

### 5.2 General exponential family distributions

If we consider distributions other than the Gaussian, $\Delta \psi \neq \text{constant}$. Then,

$$\Delta_n \leq \int \left[ \frac{I_x(\Delta g_n + g_n \Delta \psi)}{I_x g_n} - \frac{I_x(\Delta g + g \Delta \psi)}{I_x g} + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 - \left\| \frac{I_x \nabla g_n}{I_x g_n} \right\|^2 \right]^2 I_x g_n \nu(dx)$$
\[ \leq 2 \int \left[ \frac{I_x \Delta g_n}{I_x g_n} - \frac{I_x \Delta g}{I_x g} - \left\| \frac{I_x \nabla g_n}{I_x g_n} \right\|^2 + \left\| \frac{I_x \nabla g}{I_x g} \right\|^2 \right]^2 (I_x g_n) \nu(dx) \\
+ 2 \int \left[ \frac{I_x (g_n \Delta \psi)}{I_x g_n} - \frac{I_x (g \Delta \psi)}{I_x g} \right]^2 (I_x g_n) \nu(dx) \]
\[\triangleq 2U_n + 2V_n.\]

Even here, the original integrand in \( \Delta_n \) tends to zero, and so, bounding the terms and using the Dominated Convergence Theorem is again the method to follow. In the Gaussian situation we had \( U_n \), but \( V_n \) was zero. So, the same conditions as in the Gaussian case should be sufficient to bound \( U_n \) and we will need additional conditions for \( V_n \) to be bounded. Using the Cauchy-Schwarz inequality, the fact that \( g_n \leq g \) and the inequality \( \text{Var}(\Delta \psi|X) \leq E((\Delta \psi)^2|X) \) gives us the following:

\[ V_n \leq \int I_x \left( g \left[ \Delta \psi - \frac{I_x (g \Delta \psi)}{I_x g} \right] \right)^2 \nu(dx) \]
\[ \leq \int g(\Delta \psi)^2 d\theta. \]

Hence with the additional condition

\[ \int (\Delta \psi(\theta))^2 g(\theta) d\theta < \infty, \tag{5.15} \]

we can extend the previous results to other distributions belonging to exponential families in situations with squared error loss.

But this is not of much practical use because the condition (5.15) is apparently too strong. For example, the Poisson density, with the appropriate transformation to the canonical form, doesn't satisfy this condition for relevant priors \( g(\theta) \). More explicitly, let \( X_i \sim \text{Poisson}(\lambda_i) \) independently, with \( X = (X_1, \cdots, X_p)^T \), \( \lambda = (\lambda_1, \cdots, \lambda_p)^T \). The appropriate canonical transformation is \( \theta_i = \ln \lambda_i \), \( i = 1, \cdots, p \). Then,

\[ \psi(\theta) = \sum_{i} e^{\theta_i} \quad \text{and} \quad \Delta \psi(\theta) = \sum_{i} e^{\theta_i}. \]
The appropriate prior to examine the admissibility of \( \gamma_{unk}(X) \equiv p \) is \( \pi(\lambda) \equiv 1 \), and on transformation to \( \theta \), (5.15) doesn’t hold.

As a result, it seems that the previous results can’t be extended directly from the Gaussian distribution to other exponential family distributions. But it turns out that we can carry out a similar procedure, separately, for some other distributions like the Poisson. We will do that in the next section, for the squared error as well as the information-normalized loss functions.

For the remainder of this section we shall consider the Gamma distribution. We have assumed so far that \( \Theta = \mathbb{R}^p \), but for the Gamma distribution with the scale parameter unknown, \( \Theta = (\mathbb{R}^+)^p \), and so we need an extension of the previous theorem for this case if we are to apply our methodology to the Gamma distribution.

Consider the following Gamma set-up:

\[
X_i \overset{\text{ind}}{\sim} \alpha_i^\beta \exp\{-\alpha_i x_i\} x_i^{(\beta-1)} \frac{1}{\Gamma(\beta)}
\]

where \( \beta > 1 \) is known and \( \alpha_i > 0 \), \( X_i > 0 \), \( X = (X_1, \cdots, X_p)^T \), \( \alpha = (\alpha_1, \cdots, \alpha_p)^T \).

The best equivariant point estimator of the mean

\[
\beta \left( \frac{1}{\alpha_1}, \cdots, \frac{1}{\alpha_p} \right)^T
\]

is

\[
\delta(X) = \frac{\beta X}{\beta + 1}.
\]

Consider the problem of estimating the accuracy of this point estimator. We use the invariant loss function

\[
L = \sum \alpha_i^2 \left( \delta_i(X) - \frac{\beta}{\alpha_i} \right)^2
\]

and the problem is that of estimating \( L \).

\( \delta(X) \) is Bayes for the prior

\[
\pi(\alpha) \propto (\prod \alpha_i)^{-1}.
\]
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

The posterior mean of $L$ with respect to this same prior $\pi$ also happens to be the unbiased estimator of $L$.

$$
\gamma_{post} = E_\pi(L|X) = \frac{p \beta^2}{(\beta + 1)} = \gamma_{unb}.
$$

Application of Blyth's lemma to the sequence of priors

$$
\pi_a(\alpha) \propto (\prod \alpha_i)^{-a},
$$
as $a \rightarrow 1$ proves admissibility of $\gamma_{unb}$ for $p = 1$.

An extension of the unified admissibility proof idea from the Gaussian to the Gamma case also yields admissibility of $\gamma_{post}$ only for $p = 1$. This is why we don't discuss this extension in detail. The case $p \geq 2$ needs further investigation; there are as yet no inadmissibility results either.

5.3 Poisson with information-normalized loss

As said earlier, this situation corresponds very closely to the Gaussian one, and hence the methodology is very similar. First we express the Bayes point estimator with respect to a prior $\pi$ in terms of the derivatives of $\pi$. Then we do the same for the posterior expectation of the loss incurred by this Bayes estimator. Then considering a sequence of priors $\{\pi_n\}$ tending to $\pi$, we have an expression for the difference in the integrated risks in terms of derivatives of the priors. This difference has to be proved to approach zero as $n$ increases.

More formally, let $X_i \text{ i.i.d. Poisson}(\lambda_i), i = 1, \cdots, p, X = (X_1, \cdots, X_p)^T$ and $\lambda = (\lambda_1, \cdots, \lambda_p)^T$. Let $\delta_\pi(X)$ denote the Bayes estimator corresponding to the prior $\pi(\lambda)$. As in Chapter 3, let $P_\lambda(X)$ denote the joint Poisson density of $X$. Then, provided the prior $\pi$ is absolutely continuous and the relevant expectations exist (cf. Hudson, 1978), we have

$$
\delta_{\pi,i}(X) = \left[ E(\lambda_i^{-1}|X) \right]^{-1}
$$
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

\[ = X_i + \frac{\int (\lambda_i - X_i)\lambda_i^{-1}P\lambda(X)\pi(\lambda)d\lambda}{\int \lambda_i^{-1}P\lambda(X)\pi(\lambda)d\lambda} \]

\[ = X_i + \frac{I_{X-e_i}(\lambda_i\pi'_i(\lambda))}{I_{X-e_i}(\pi)} \]

\[ = X_i + \frac{I_X(\pi'_i)}{I_X(\lambda_i^{-1}\pi)} \]

\[ = X_i + \phi_i(X) \text{, say,} \]

where

\[ \pi'_i = \frac{\partial}{\partial \lambda_i} \pi(\lambda) , \]

\[ I_X h = \int h(\lambda)P\lambda(X)d\lambda , \]

and \( e_i \) is a \( p \)-dimensional column vector with '1' in the \( i^{th} \) place and '0' elsewhere.

The posterior loss estimator corresponding to the same prior is

\[ \gamma_p(X) \overset{\Delta}{=} E \left[ \sum \lambda_i^{-1}(X_i + \phi_i(X) - \lambda_i)^2 | X \right] \]

\[ = E \left[ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 | X \right] + E \left[ \sum \frac{\phi^2_i(X)}{\lambda_i} | X \right] \]

\[ - 2E \left[ \sum \frac{(\lambda_i - X_i)}{\lambda_i} \phi_i(X) | X \right] . \]

From the result obtained while proving Lemma (3.3),

\[ E \left[ \sum \lambda_i^{-1}(X_i - \lambda_i)^2 | X \right] = p + \frac{I_X(\lambda \cdot \nabla^2 \pi + \nabla \cdot \pi)}{I_X \pi} , \]

where

\[ \nabla^2 \pi = \left( \frac{\partial^2}{\partial \lambda_i^2} \pi \right)_{px1} , \]

\[ \nabla \cdot \pi = \sum \frac{\partial}{\partial \lambda_i} \pi(\lambda) . \]

Using gamma identities (3.9) for the last two terms in the above expression for \( \gamma_p(X) \),

\[ \gamma_p = p + \frac{I_X(\lambda \cdot \nabla^2 \pi + \nabla \cdot \pi)}{I_X \pi} - \sum_i \frac{[I_{X-e_i}(\lambda_i\pi'_i)]^2}{I_{X-e_i}(\pi)I_X(\pi)} \]
and hence the difference in the posterior estimators corresponding to the priors \( \pi_n \) and \( \pi \) is

\[
\gamma_{\pi_n} - \gamma_{\pi} = \frac{I_X(\lambda \cdot \nabla^2 \pi_n) - I_X(\lambda \cdot \nabla^2 \pi)}{I_X \pi_n} - \frac{I_X(\nabla \cdot \pi_n) - I_X \nabla \cdot \pi}{I_X \pi} \\
+ \frac{I_X(\nabla \cdot \pi_n) - I_X \nabla \cdot \pi}{I_X \pi} \\
- \sum_i \frac{[I_{X-e_i}(\lambda_i \pi_{n,i})]^2}{I_{X-e_i}(\pi_n)I_X(\pi_n)} + \sum_i \frac{[I_{X-e_i}(\lambda_i \pi_i')]^2}{I_{X-e_i}(\pi)I_X(\pi)}. \]

Therefore, the difference in the risks in Blyth’s condition is

\[
\Delta_n = B(\pi_n, \gamma_{\pi_n}) - B(\pi_n, \gamma_{\pi}) \\
= \sum_{x_1, \ldots, x_n \geq 0} (\gamma_{\pi_n} - \gamma_{\pi})^2 (I_x \pi_n) \\
= \sum_x \left[ \left( \frac{I_x(\lambda \cdot \nabla^2 \pi_n) - I_x(\lambda \cdot \nabla^2 \pi)}{I_x \pi_n} \right)^2 \\
+ \left( \frac{I_x(\nabla \cdot \pi_n) - I_x(\nabla \cdot \pi)}{I_x \pi_n} \right)^2 \\
- \sum_i \frac{[I_{x-e_i}(\lambda_i \pi_{n,i})]^2}{I_{x-e_i}(\pi_n)I_x(\pi_n)} + \sum_i \frac{[I_{x-e_i}(\lambda_i \pi_i')]^2}{I_{x-e_i}(\pi)I_x(\pi)} \right] I_x(\pi_n). \tag{5.16} \]

First of all, we note that for a sequence \( \{\pi_n\} \) converging to \( \pi \) and satisfying all the regularity conditions of \( \pi \), the summand in (5.16) converges to zero. So again, by proving boundedness of the entire sum, we can apply the discrete version of the Dominated Convergence Theorem to conclude that \( \Delta_n \to 0 \). The method that we follow to prove boundedness is similar to what we did in the Gaussian context, and hence we will not provide any details here. We let the prior be of the form used in the Gaussian case, i.e.,

\[
\pi_n = \xi_n \pi , \quad \xi_n(\lambda) \leq 1. \]

We bound the terms involving derivatives of \( \pi \) separately, and then employ a similar form for \( \xi_n \) to that used in the Gaussian setting to bound the other terms.
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

The condition for the terms with derivatives of $\pi$ to be bounded is

$$\sum_x I_x \left\{ \pi \left[ \frac{\lambda \cdot \nabla^2 \pi}{\pi} - \frac{I_x(\lambda \cdot \nabla^2 \pi)}{I_x(\pi)} + \frac{\nabla \cdot \pi}{\pi} - \frac{I_x(\nabla \cdot \pi)}{I_x(\pi)} + \sum_i \frac{[I_x(\pi_i)]^2}{I_x - \pi} \right] \right\}^2 < \infty$$

Even here, we can replace this condition with a simpler and a less general (i.e. stronger) condition:

$$\int \frac{(\lambda \cdot \nabla \pi)^2}{\pi^2} d\lambda + \int \frac{(\lambda \cdot (\nabla \pi)^2)^2}{\pi^3} d\lambda < \infty , \quad (5.17)$$

where

$$\lambda \cdot (\nabla \pi)^2 = \sum \lambda_i \left[ \frac{\partial}{\partial \lambda_i} \pi(\lambda) \right]$$ and

$$\nabla^2 \pi = \left( \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) \right)_{p \times 1} .$$

Thus $\nabla^2 \pi$ is a $p$-dimensional vector consisting of the diagonal elements of the Hessian matrix of $\pi$.

Now we shall find conditions for the term involving derivatives of $\xi_n$ to be bounded. The term is

$$\sum_x \left\{ I_x \left[ \lambda \cdot (\pi \nabla^2 \xi_n) + 2 \sum \lambda_i \nabla_i \xi_n \nabla_i \pi + \pi \nabla \cdot \xi_n \right] - \sum_i \frac{[I_x - e_i(\lambda_i \pi_n)]^2}{I_x(\pi_n)} \right\} I_x(\pi_n)$$

$$\leq 4 \int \frac{(\lambda \cdot \nabla^2 \xi_n)^2}{\xi_n} \pi(\lambda) d\lambda + 16 \int \frac{(\sum \lambda_i \nabla_i \xi_n \nabla_i \pi)^2}{\xi_n \pi} d\lambda$$

$$+ 4 \int \frac{(\nabla \cdot \xi_n)^2}{\xi_n} \pi(\lambda) d\lambda + 4 \int \frac{(\sum \lambda_i (\pi_n)^2)^2}{\pi_n^3} d\lambda$$

$$\Delta 4C_n + 16D_n + 4E_n + 4F_n .$$

The above step follows from the Cauchy-Schwarz inequality, the inequality (5.4) with $m = 4$ and due to the fact that

$$\sum_x I_x[h(\lambda)] = \int h(\lambda) d\lambda .$$
We use the same prior as defined by (3.13) and (3.14); we repeat it here for the sake of completeness. The prior is

\[
\xi_n(\lambda) = \begin{cases} 
2^{k-1} \eta_n^k & 0 \leq \eta_n \leq 1/2 \\
1 - 2^{k-1}(1 - \eta_n)^k & 1/2 \leq \eta_n \leq 1 
\end{cases}
\]

where \( \eta_n \) is defined as

\[
\eta_n(\lambda) = \begin{cases} 
1 & \Lambda < 1 \\
1 - \frac{\ln A}{\ln n} & 1 \leq \Lambda < n \\
0 & \Lambda > n 
\end{cases}
\]

for \( n = 2, 3, \ldots \), and \( \Lambda = \sum_1^p \lambda_i \). The derivatives are

\[
2^{1-k} \xi'(\eta) = \begin{cases} 
k \eta^{k-1} & 0 \leq \eta \leq 1/2 \\
k(1 - \eta)^{k-1} & 1/2 \leq \eta \leq 1 
\end{cases}
\]

and

\[
2^{1-k} \xi''(\eta) = \begin{cases} 
k(k-1) \eta^{k-2} & 0 \leq \eta < 1/2 \\
-k(k-1)(1-\eta)^{k-2} & 1/2 \leq \eta \leq 1 
\end{cases}
\]

and

\[
\nabla_i \xi = \xi'(\eta) \nabla_i \eta \, , \, \nabla_i^2 \xi = \xi''(\eta) \nabla_i^2 \eta + \xi'(\eta) \nabla_i^2 \eta \, ,
\]

where

\[
\nabla_i \eta_n(\theta) = \eta_{n,i} = -\frac{1}{\Lambda \ln n} I\{1 < \Lambda \leq n\} \, ,
\]

\[
\nabla_i^2 \eta_n = \eta_{n,i} = \frac{1}{\Lambda^2 \ln n} I\{1 < \Lambda \leq n\}.
\]

From these derivatives we can compute and bound the integrands of \( C_n, D_n, E_n \) and \( F_n \). All the arguments are identical to those in the Gaussian context, so we will not present any of the details here. We shall require \( k \geq 4 \).

It can be verified that the condition

\[
\int_{R_+^k - S} \frac{\pi(\lambda)}{\Lambda^2 \ln^2 (\Lambda + 2)} d\lambda < \infty ,
\]

where

\[
S = \{\lambda : \Lambda \leq 1\}
\]
is sufficient for the boundedness of $C_n$. This condition also ensures that $E_n$ is bounded, and along with condition (5.17) it implies that $F_n$ is bounded.

It is also easy to see that
\[
\int_{R^2_+ - S} \frac{(\lambda \cdot \nabla \pi)^2}{\pi^2} \, d\lambda < \infty \tag{5.19}
\]
guarantees boundedness of $D_n$.

Again, similar to the Gaussian case, we can prove that the conditions (5.17) and (5.18), if true, together imply that the condition (5.19) holds. By applying the Cauchy-Schwarz inequality to
\[
\int \left( \frac{\lambda \cdot (\nabla \pi)^2}{\pi^3} \right) \, d\lambda < \infty
\]
and the integral in (5.18), we can conclude that
\[
\left\{ \int_{R^2_+ - S} \frac{(\lambda \cdot \nabla \pi^2)}{\pi^2} \, d\lambda \right\}^2 < \infty.
\]
But by the Cauchy-Schwarz inequality for summations, we have
\[
(\lambda_i \pi_i^2)^2 \leq \Lambda \sum \lambda_i (\pi_i^2)^2
\]
and it is also true that $\ln^2(\Lambda \lor 2) > \ln(\Lambda \lor 2)$. Hence, once we have conditions (5.17) and (5.18), the condition (5.19) is redundant.

We have thus proved the following theorem:

**Theorem 5.2** For the Poisson distribution, $\gamma_\pi(X) = E_{\pi}[L_{-1}|X]$ is admissible as an estimator of

\[
L_{-1} = \sum \lambda_i^{-1}(\delta_{\pi_i}(X) - \lambda_i)^2
\]

if $\pi$ satisfies the following two conditions:

\[
\int_{R^2_+} \frac{(\lambda \cdot \nabla^2 \pi + \nabla \cdot \pi)^2}{\pi^2} \, d\lambda + \int_{R^2_+} \frac{(\lambda \cdot (\nabla \pi)^2)^2}{\pi^3} \, d\lambda < \infty \tag{5.20}
\]

\[
\int_{R^2_+ - S} \frac{\pi(\lambda)}{\Lambda^2 \ln^2(\Lambda \lor 2)} \, d\lambda < \infty \tag{5.21}
\]

where $S = \{ \lambda : \Lambda \leq 1 \}$.
5.3.1 Verification of earlier result

For \( \pi(\lambda) \equiv 1 , \phi(X) \equiv 0 \), and
\[
E \left[ \sum \lambda_i^{-1} (X_i - \lambda_i)^2 \mid X \right] = p ,
\]
the integrands in the first condition are obviously identically zero. The second condition can be seen to hold only for \( p \leq 2 \). Thus, \( \gamma(X) \equiv p \) is admissible as an estimator of \( \sum \lambda_i^{-1} (X_i - \lambda_i)^2 \), for \( p \leq 2 \). This result was obtained in Chapter 3.

5.3.2 The Clevenson-Zidek estimator

The Clevenson-Zidek point estimator of \( \lambda \), under \( L_1 \) loss, is
\[
\delta_{CZ} = \delta_{CZ}(X) = \left( 1 - \frac{p - 1 + \beta}{\sum X_i + p - 1 + \beta} \right) X .
\]

This estimator is admissible for \( p > 1 \) and \( \beta \geq 0 \) (cf. Clevenson and Zidek, 1975).
Let us consider the simplest case of \( \beta = 0 \). In this case, the prior corresponding to which the Clevenson-Zidek estimator is Bayes, is
\[
\pi_{CZ}(\lambda) = \frac{\Gamma(p - 1)}{(\sum \lambda_i)^{p-1}} .
\] (5.22)

Then,
\[
\nabla \pi = -\Gamma(p) \Lambda^{-(p)} \mathbf{e} \text{ and } \nabla^2 \pi = \Gamma(p + 1) \Lambda^{-(p+1)} \mathbf{e} ,
\]
where \( \mathbf{e} \) is a \( p \)-column vector with '1' everywhere. The integrand in condition (5.21) is
\[
O \left( \Lambda^{-(p+1)} \ln^{-2} (\Lambda + 2) \right)
\]
and the two integrands in the condition (5.20) are
\[
0 \text{ and } \Lambda^{-(p+1)} \text{ respectively .}
\]

So, both conditions are satisfied, yielding the following corollary.
Corollary 5.2 The posterior estimator

\[ p = \frac{(p-1)^2}{\sum X_i + p - 1} \]

of the loss

\[ \sum \lambda_i^{-1}(\delta_{GZ,i} - \lambda_i)^2 \]

is admissible for all \( p > 1 \).

(Note: The C-Z point estimator reduces to the mle when \( p = 1 \).)

Remark: As in the Gaussian case, the prior giving rise to admissible point estimator also yields admissible loss estimator in the Poisson situation.

5.4 Poisson with squared error loss

The setting here is the same as in the previous section, but we consider the loss

\[ L(\delta(X), \lambda) = \sum_i (\delta_i(X) - \lambda_i)^2. \]

The maximum likelihood point estimator \( \delta(X) = X \) is formal Bayes with respect to the prior

\[ \pi_0(\lambda) \propto (\prod \lambda_j)^{-1} \]

and the corresponding posterior expectation of the loss \( L \) is \( \gamma_{post} = \sum X_i \). Let

\[ \pi^*(\lambda) \propto \left( \prod \lambda_j \right)^{-1} \pi(\lambda) \]

for any prior \( \pi(\lambda) \). Let us denote by \( \delta^*_p \) the (generalized) Bayes estimator corresponding to the prior \( \pi^* \). Again, \( P_\lambda(X) \) denotes the joint Poisson density of the \( X_i \)'s. Then,

\[ \delta^*_{p,i} = E_{\pi^*}(\lambda_i|X) \]

\[ = X_i + \frac{\int (\lambda_i - X_i)P_\lambda(X)(\prod \lambda_j)^{-1} \pi(\lambda)d\lambda}{\int P_\lambda(X)(\prod \lambda_j)^{-1} \pi(\lambda)d\lambda} \]

\[ = X_i + \frac{I^*_X(\lambda_i \pi_i^*)}{I^*_X(\pi)}. \]
Here $I^*_X$ is defined as

$$I^*_X h = \int h(\lambda) \frac{\exp\{-\sum_j \lambda_j\} \prod_j \lambda_j^{X_j-1}}{\prod_j (X_j-1)!} d\lambda$$

$$= \int h(\lambda) P_\lambda(X - e) d\lambda$$

for any function $h$ and where $e$ is a unit vector of length $p$. Let us now note that

$$E_{\pi^*}[\sum (X_i - \lambda_i)^2 | X] = \frac{\int \sum (X_i - \lambda_i)^2 P_\lambda(X) \pi^*(\lambda) d\lambda}{\int P_\lambda(X) \pi^*(\lambda) d\lambda}$$

$$= \frac{I^*_X \sum \lambda_i \pi + \sum \lambda_i \pi' + \sum \lambda_i^2 \pi''}{I^*_X \pi}$$

$$= \sum X_i + \frac{I^*_X (2 \sum \lambda_i \pi' + \sum \lambda_i^2 \pi'')}{I^*_X \pi}.$$ 

The second step follows from using the gamma identities (3.9) with $X_i$ instead of $X_i + 1$ and in the third step we use the expression for $\delta^*_{\pi,i}$ as derived above. Let $\gamma^*_\pi$ denote the estimator of $L$ corresponding to the prior $\pi^*$. Then, from the preceding computation, an expression for this loss estimator in terms of the derivatives of $\pi$ is:

$$\gamma^*_\pi(X) \equiv E_{\pi^*}[\sum (\delta^*_{\pi,i}(X) - \lambda_i)^2 | X]$$

$$= \sum X_i + \frac{I^*_X (2 \sum \lambda_i \pi' + \sum \lambda_i^2 \pi'')}{I^*_X \pi} - \sum \frac{[I^*_X (\lambda_i \pi')]^2}{[I^*_X \pi]^2}.$$ 

Considering a sequence of priors $\pi_n$ (note that there is a corresponding sequence $\pi_n^* = (\prod_j \lambda_j)^{-1} \pi$), the difference in the posterior loss estimators corresponding to the priors $\pi_n^*$ and $\pi^*$ can be written as

$$\gamma^*_\pi_n - \gamma^*_\pi = \left[ \frac{I^*_X (2 \lambda \cdot \nabla \pi_n)}{I^*_X \pi_n} - \frac{I^*_X (2 \lambda \cdot \nabla \pi)}{I^*_X \pi} \right]$$

$$+ \left[ \frac{I^*_X (\lambda^2 \cdot \nabla^2 \pi_n)}{I^*_X \pi_n} - \frac{I^*_X (\lambda^2 \cdot \nabla^2 \pi)}{I^*_X \pi} \right]$$

$$- \frac{ \sum \{I^*_X (\lambda_i \pi'_{n,i})\}^2}{\{I^*_X \pi_n\}^2} - \sum \frac{\{I^*_X (\lambda_i \pi'_i)\}^2}{\{I^*_X \pi\}^2}.$$ (5.23)

where

$$\lambda^2 \cdot \nabla^2 \pi_n = \sum_i \lambda_i^2 \pi_i''.$$
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

Note that for the above to hold, \( \pi_n^* \) and \( \pi^* \) should satisfy the conditions of Hudson (1978, p.474).

We shall now make the canonical transformation at this stage and see that this enables us to circumvent the difficulties we encountered in the previous section. So,

\[
\theta_i = \ln \lambda_i , \quad i = 1, \cdots, p .
\]

Let \( \varphi_n(\theta) = \pi_n(\lambda) , \quad \varphi(\theta) = \pi(\lambda) \). Then,

\[
\nabla \cdot \varphi = \lambda \cdot \nabla \pi \quad \text{and} \quad \Delta \varphi = \lambda^2 \cdot \nabla^2 \pi + \lambda \cdot \nabla \pi .
\]

Also, let \( P_\lambda (X - e) d\lambda = f_\theta (X) d\theta \), and let

\[
I_X g = \int g(\theta) f_\theta (X) d\theta .
\]

Now, in terms of the transformed parameter, the difference (5.23) becomes

\[
\gamma_{n \cdot \varphi_n}^* - \gamma_\varphi^* = \left[ \frac{I_X (\Delta \varphi_n)}{I_X \varphi_n} - \frac{I_X (\Delta \varphi)}{I_X \varphi} \right] + \left[ \frac{I_X (\nabla \cdot \varphi_n)}{I_X \varphi_n} - \frac{I_X (\nabla \cdot \varphi)}{I_X \varphi} \right] - \left[ \frac{\|I_X (\nabla \varphi_n)\|^2}{(I_X \varphi_n)^2} - \frac{\|I_X (\nabla \varphi)\|^2}{(I_X \varphi)^2} \right] \quad (5.24)
\]

and as before, using (3.8),

\[
\Delta_n = B(\varphi_n, \gamma_{n \cdot \varphi_n}^*) - B(\varphi_n, \gamma_\varphi^*)
\]

\[
= \sum_x (\gamma_{n \cdot \varphi_n}^*(x) - \gamma_\varphi^*(x))^2 (I_x \varphi_n) .
\]

It can be seen from (5.24) that the summand in the above expression goes to zero because \( \varphi_n(\theta) \to \varphi(\theta) \) as \( n \to \infty \). So, once again we only need to establish conditions for \( \Delta_n \) to be bounded, and then use the discrete version of the Dominated Convergence Theorem. The first and the third terms in (5.24) correspond to the terms in equation (5.2) that we had for the Gaussian situation. Hence the same conditions as there can be put on \( \varphi(\theta) \), to ensure boundedness. In addition, we need conditions for the following expression to be bounded:

\[
\sum_x \left[ \frac{I_x \nabla \cdot \varphi_n}{I_x \varphi_n} - \frac{I_x \nabla \cdot \varphi}{I_x \varphi} \right]^2 (I_x \varphi_n) .
\]

(5.25)
CHAPTER 5. UNIFIED ADMISSION THEORY

We again use the priors \( \varphi_n = \xi_n \varphi \), where \( \varphi \) corresponds to the \( g \) in the Gaussian context, and \( \xi_n \) will be defined exactly as in (5.9), with \( \eta_n \) being the same as in (5.8). From the term in (5.25) we first separate out the terms containing only the derivatives of \( \varphi \). Applying the Cauchy-Shwarz inequality and using the fact that

\[
\text{Var} \left( \frac{\nabla \cdot \varphi}{\varphi} \vert X \right) \leq E \left( \frac{(\nabla \cdot \varphi)^2}{\varphi} \right),
\]

the condition which implies that these terms will be bounded is

\[
\int \frac{(\nabla \cdot \varphi(\theta))^2}{\varphi(\theta)} d\theta < \infty.
\]

The remaining term of (5.25), i.e., the term involving the derivative of \( \xi_n \) is

\[
\int \frac{(\nabla \cdot \xi_n(\theta))^2}{\xi(\theta)} \varphi(\theta) d\theta < \infty.
\]

After some computation it can be seen that a sufficient condition for guaranteeing the finiteness of this term is:

\[
\int_{R^p - S} \frac{\varphi(\theta)}{\|\theta\|^2 \ln^2(\|\theta\| \vee 2)} d\theta < \infty.
\]

But this is already one of the conditions that we directly adapt from the Gaussian situation.

We thus have the following theorem:

**Theorem 5.3** Let \( \varphi(\theta) \) be a prior satisfying the regularity conditions of Hudson (1978). Then the posterior expected loss \( \gamma_\varphi(X) \) is admissible for estimating the squared error loss \( L \) if \( \varphi(\theta) \) satisfies the following conditions:

\[
\int_{R^p} \frac{(\Delta \varphi(\theta))^2}{\varphi(\theta)} d\theta + \int \frac{\|\nabla \varphi(\theta)\|^4}{\varphi^3(\theta)} d\theta < \infty
\]

(5.26)

\[
\int_{R^p} \frac{(\nabla \cdot \varphi(\theta))^2}{\varphi(\theta)} d\theta < \infty
\]

(5.27)

\[
\int_{R^p - S} \frac{\varphi(\theta)}{\|\theta\|^2 \ln^2(\|\theta\| \vee 2)} d\theta < \infty
\]

(5.28)

Here \( S = \{\theta : \|\theta\| \leq 1\} \).
The unbiased loss estimator $\sum X_i$ corresponds to the prior $\varphi(\theta) \equiv 1$. So, the first two conditions above hold trivially, and the last condition holds only for $p \leq 2$. Thus follows the corollary –

**Corollary 5.3** $\sum X_i$ is admissible as an estimator of $L = \sum (X_i - \lambda_i)^2$ if $p \leq 2$.

The problem of finding improvements or proving admissibility for $p > 2$ is still unresolved.

Having established all these unified admissibility theory extensions, we will now attempt to explore the relation between the Gaussian and Poisson cases which we had observed in a remark towards the end of Chapter 3.

### 5.5 Role of the polydisc transform

Let us consider the polydisc transform introduced by Johnstone and McGibbon (1989). We now reproduce from their paper some fundamental aspects of this transform. Polydisc transform is defined as a many-to-one mapping $\tau : \mathbb{R}^{2p} \to \mathbb{R}^p_+$,

$$\tau : (\omega_1, \ldots, \omega_{2p-1}, \omega_{2p}) \to (\omega_1^2 + \omega_2^2, \ldots, \omega_{2p-1}^2 + \omega_{2p}^2). \quad (5.29)$$

The set $\Omega = \tau^{-1}(T)$ is called the polydisc transform of $T$. The transform of a rectangle $[0, a] \subset \mathbb{R}^p_+$, namely, $\{\omega : \omega_{2i-1}^2 + \omega_{2i}^2 \leq a_i, \ i = 1, \ldots, p\}$, is known as a polydisc in functional analysis, and hence the name of the transform. The inverse mapping $\tau^{-1}$ is the ‘dimension-doubling’ version of the traditional square-root variance stabilizing transformation for Poisson data. A function $v(\tau)$ defined on $T$ induces a function $u(\omega) = v(\tau(\omega))$ on $\Omega$. Associating polar co-ordinates $(r_i, \theta_i), \ i = 1, \ldots, p$ with $\omega$, we have $\tilde{u}(r)$ such that $v(\tau) = u(\omega) = \tilde{u}(r)$. We note that $d\omega_{2i-1}d\omega_{2i} = r_idr_id\theta_i = 2^{-1}d\tau_id\theta_i$ and

$$\int_{\Omega} u(\omega)d\omega = (2\pi)^p \int \tilde{u}(r) \prod_i(r_idr_i) = \pi^p \int_T v(\tau)d\tau. \quad (5.30)$$
Thus the polydisc transform is volume-preserving.

Since our problem here is that of deriving conditions on the prior for admissibility, we will be applying the polydisc transform to the parameter space. In our context, \( \lambda \) will replace \( \tau \). Also we define \( \beta(\omega) = \pi(\lambda(\omega)) \). Then we have the relations

\[
\frac{\partial}{\partial \omega_{2i-1}} \beta(\omega) = 2\omega_{2i-1} \frac{\partial}{\partial \lambda_i} \pi(\lambda),
\]

\[
\frac{\partial}{\partial \omega_{2i}} \beta(\omega) = 2\omega_{2i} \frac{\partial}{\partial \lambda_i} \pi(\lambda),
\]

\[
\frac{\partial^2}{\partial \omega_{2i-1}^2} \beta(\omega) = 4\omega_{2i-1}^2 \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) + 2 \frac{\partial}{\partial \lambda_i} \pi(\lambda)
\]

and

\[
\frac{\partial^2}{\partial \omega_{2i}^2} \beta(\omega) = 4\omega_{2i}^2 \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) + 2 \frac{\partial}{\partial \lambda_i} \pi(\lambda).
\]

From the above, we have the following identities:

\[
\|
abla \beta(\omega)\|^2 = 4 \sum \lambda_i (\pi_i'(\lambda))^2 = 4\lambda \cdot (\nabla \pi)^2,
\]

\[
\Delta \beta(\omega) = 4[\nabla \cdot \pi(\lambda) + \lambda \cdot \nabla^2 \pi],
\]

where

\[
(\nabla \pi)^2 = \left( \frac{\partial}{\partial \lambda_i} \pi(\lambda) \right)_{p \times 1}
\]

and

\[
\nabla^2 \pi(\lambda) = \left( \frac{\partial^2}{\partial \lambda_i^2} \pi(\lambda) \right)_{p \times 1}.
\]

We will now see how a Poisson problem in \( p \) dimensions can be transformed to a corresponding Gaussian problem in \( 2p \) dimensions with this transformation, both in the point estimation and the loss estimation contexts.

### 5.5.1 Point Estimation

Let \( X_i \sim \text{Poisson}(\lambda_i) \), \( i = 1, \ldots, p \), and we consider the \( L_{-1} \) loss. Then, the posterior estimator of \( \lambda \) with respect to the prior \( \pi \) is \( \delta_\pi(X) \), with

\[
\delta_{\pi,i}(X) = \frac{1}{E_\pi[\lambda_i^{-1}|X]}.
\]
CHAPTER 5. UNIFIED ADMISSIBILITY THEORY

Using Blyth's lemma and the method of Brown and Hwang (1982), conditions on the prior \( \pi \) in order that \( \delta_\pi(X) \) be an admissible estimator of \( \lambda \) with respect to the \( L_\infty \) loss are:

\[
\int_{\Lambda > 1} \frac{\pi(\lambda)}{\Lambda \ln^2(\Lambda + 1)} d\lambda < \infty \tag{5.33}
\]

\[
\int_{R^p_+} \frac{\sum \lambda_i (\pi_i(\lambda))^2}{\pi(\lambda)} d\lambda < \infty \tag{5.34}
\]

Here we have considered the sequence of priors \( \{\pi_n\} \), with \( \pi_n = h_n^2 \pi \), where \( h_n(\lambda) \) has the same definition as \( \eta_n \) in (3.14).

Using the polydisc transform as in (5.29), and with the assistance of (5.30) and the relations following it, the above conditions reduce to

\[
\int_{||\omega||^2 > 1} \frac{\beta(\omega)}{||\omega||^2 \ln^2(||\omega||^2 + 1)} d\omega < \infty \tag{5.35}
\]

and

\[
\int_{R^p} \frac{||\nabla \beta(\omega)||^2}{\beta(\omega)} d\omega < \infty \tag{5.36}
\]

respectively. The condition (5.36) is identical to condition (3.3) of Brown and Hwang (1982) in \( 2p \) dimensions. Noticing that \( \{\omega : ||\omega||^2 > 1\} = \{\omega : ||\omega|| > 1\} \) and that \( \ln^2(||\omega||^2) = 4 \ln^2(||\omega||) \), the condition (5.35) is also equivalent to condition (3.1) of Brown and Hwang in \( 2p \) dimensions. Conditions (3.1) and (3.3) of Brown-Hwang are conditions on the prior for admissibility of the Bayes point estimator in the Gaussian case, with respect to squared error loss.

Thus the Poisson problem of determining admissibility in \( p \) dimensions of a generalized Bayes point estimator can be transformed to the corresponding Gaussian problem in \( 2p \) dimensions.

5.5.2 Loss Estimation

First let us consider the problem from Chapter 3 of proving the admissibility of \( \gamma_{unk} = p \) for \( p = 2 \). With the polydisc transformation, and letting \( \pi(\lambda) = \beta(\omega) \),

\[
\lambda \cdot \nabla^2 \pi(\lambda) + \nabla \cdot \pi(\lambda) = \frac{1}{4} \Delta \beta(\omega)
\]
and so, the upper bound on the difference in risks of Lemma (3.2) is proportional to

\[
\int_{\mathbb{R}^{2p}} \frac{\|\Delta \beta(\omega)\|^2}{\beta(\omega)} d\omega.
\]

This is exactly the same bound as in Lemma (3.4) of Johnstone (1988) for the Gaussian case in 2p dimensions. Hence Johnstone’s admissibility result for \( p = 4 \) in the Gaussian case directly yields admissibility for \( p = 2 \) for this Poisson case (cf. concluding remark of section 3.2).

Let us now look at the conditions on the prior for admissibility of the posterior loss estimator in the Gaussian setting (as derived in section 1 of this chapter) and the corresponding ones for the Poisson with the information-normalized loss function (as in section 3 here). So we shall be comparing conditions (5.12) and (5.13) with conditions (5.20) and (5.21). It is not hard to verify that after application of the polydisc transformation, the left hand side expressions of (5.20) and (5.21) transform to the \( 2p \) dimensional versions of the left hand side expressions of (5.12) and (5.13) respectively. Once again, as in the point estimation context, the integrals in (5.13) and in the transformed version of (5.21), are identical up to a constant. Nonetheless, we still have equivalence of the conditions in the two situations.

Moreover, the Clevenson-Zidek prior (5.22) is equivalent to the Strawderman prior (5.14) after the transformation and the same holds true even if we don’t restrict to the special values of the parameters.

Brown (1979) first suggested that the Poisson distribution in \( p \) dimensions is equivalent to the Gaussian distribution in \( 2p \) dimensions in the context of admissibility of point estimators. The polydisc transform provides some insight into how this phenomenon occurs, and we find that the equivalence also carries over to the loss estimation setting.
Chapter 6

Necessary Condition for Admissibility

For situations where we don’t have sufficient conditions for admissibility as in the previous chapter, having a necessary condition for admissibility will help us in deducing inadmissibility.

We shall consider the general exponential family set-up and prove that a loss estimator is admissible (with respect to squared error) only if it is generalized Bayes for some prior. A similar result is, of course, known to hold for point estimators (cf. Theorem 4.16, Brown, 1986).

Let \( \nu \) be a \( \sigma \)-finite measure on Borel subsets of \( \mathbb{R}^p \). \( X \in \mathcal{X} \subset \mathbb{R}^p \), \( \theta \in \Theta \subset \mathbb{R}^p \). The natural parameter space is

\[
N = \left\{ \theta : \int e^{\theta \cdot X} \nu(dx) < \infty \right\}
\]

and

\[
\psi(\theta) = \ln \int e^{\theta \cdot X} \nu(dx)
\]

is the cumulant generating function. Let

\[
f_\theta(x) = \exp\{\theta \cdot x - \psi(\theta)\}.
\]
CHAPTER 6. NECESSARY CONDITION FOR ADMISSIONIBILITY

Then, for $\Theta \subset N$, the family of probability densities $\{f_\theta : \theta \in \Theta \}$ is the standard exponential family of probability densities with respect to the measure $\nu$. Let $S$ be the support of $\nu$, i.e., the minimal closed set such that $\nu(S^c) = 0$. Let $K$ be the convex support of $\nu$, i.e., the closure of the convex hull of $S$. We denote the boundary of $\Theta$ by $\partial \Theta$ and the interior of $K$ by $K^o$.

Let $L(\delta(X), \theta)$ denote the loss incurred in estimating $\theta$ by $\delta(X)$. We consider the problem of estimating $L$ by $\gamma(X)$, under squared error discrepancy. It can be shown that the following version of a well-known result (cf. Theorem 4.14, Brown, 1986) holds true. We shall not present the proof here.

Theorem 6.1 In the standard exponential family setup and with squared error loss, every admissible procedure is a limit of Bayes estimators for prior distributions with finite support, i.e., if $\gamma_0(X)$ is admissible, there corresponds a sequence $\{\Pi_n\}$ of distributions supported on a finite set and having finite Bayes risk such that

$$\gamma_{\Pi_n}(x) \to \gamma_0(x) \quad \text{a.e. } (\nu). \quad (6.1)$$

So now, if we prove that the limit of a sequence of Bayes (loss) estimators is itself Bayes or generalized Bayes, we will have the following theorem.

Theorem 6.2 Consider the standard exponential family of probability densities defined earlier. Then $\gamma(X)$ is admissible under squared error loss as an estimator of $L = \|\delta(X) - \theta\|^2$ only if there exists a measure $H$ on $\tilde{\Theta} \subset N$ such that

$$\gamma(x) = \frac{\int \|\delta(x) - \theta\|^2 e^{\theta x} H(d\theta)}{\int e^{\theta x} H(d\theta)} \quad \text{a.e. } (\nu), \quad x \in K^o. \quad (6.2)$$

[The conditions on $H$ that are implicitly included are the existence of the numerator and the denominator in (6.2).]

If $\Theta = \tilde{\Theta}$, or if $H(\partial \Theta) = 0$, then with $G(d\theta) \triangleq \exp\{\psi(\theta)\} H(d\theta)$,

$$\gamma(x) = \frac{\int \|\delta(x) - \theta\|^2 f_\theta(x) G(d\theta)}{\int f_\theta(x) G(d\theta)} \quad \text{a.e. } (\nu), \quad x \in K^o. \quad (6.3)$$

Thus $\gamma(X)$ is generalized Bayes on $K^o$ relative to $G(d\theta)$. 
CHAPTER 6. NECESSARY CONDITION FOR ADMISSIBILITY

Proof: The proof is an extension of the one given by Brown (1986, Theorem 4.16) in the point estimation context. Hence the details, being the same as there, will be omitted from this proof; we only sketch an outline of the argument.

Let $\gamma(X)$ be admissible as an estimator of $L$, under squared error loss. Then, from theorem 6.1, there exists a sequence of distribution functions $\{\Pi_m\}$ such that each member of the sequence has a finite support and a finite Bayes risk, and

$$\gamma_{\Pi_m}(x) \rightarrow \gamma(x) \text{ a.e. } (\nu).$$

We assume that $K^o \neq \emptyset$ and that $f_\theta(x) > 0$ a.e. $(\nu)$, $\theta \in \Theta$. Let $x' \in K^o$ and $x'' \in K^o$ such that (6.1) holds at both these points. Then there exists a finite set $T \subset K^o$ such that (6.1) holds on $T$ and such that $x' \in B^o$ and $x'' \in B^o$, where $B$ is the convex hull of $T$ and $B^o$ is the interior of $B$. (If $T = S$, then $B^o = K^o$.) For $x \in T$, since $\Pi_m$ has a finite support, $\int \exp\{\theta \cdot x - \psi(\theta)\} \Pi_m(d\theta) < \infty$. We can define

$$\tilde{H}_m(d\theta) = \frac{e^{-\psi(\theta)} \Pi_m(d\theta)}{\int \exp\{\xi \cdot x - \psi(\xi)\} \Pi_m(d\theta)}.$$

It can be shown that $\gamma_{\Pi_m}$ satisfies (6.2) with $\tilde{H}_m$ as the distribution function.

We will use the result (Prop.1.6 of Brown) that in the standard exponential family setting, an arbitrary affine transformation can be applied either to $\Theta$ or to $X$.

Since $\int \exp\{x \cdot \theta\} \tilde{H}_m(d\theta) = 1$, using Lemma 2.1 of Brown (1986), we can prove the tightness of the sequence $\{\tilde{H}_m\}$ and conclude that there exists a subsequence $\{\tilde{H}_n\}$ and a limiting measure $H$ such that $\{\tilde{H}_n\} \rightarrow H$ (weak *). We have dropped the subsequence notation here for convenience of representation.

Similarly, it can also be proved that the sequence $\{e^{\theta \cdot x} \tilde{H}_n\}$ is tight. Therefore, for $x \in B^o$,

$$\int e^{\theta \cdot x} H_n(d\theta) \rightarrow \int e^{\theta \cdot x} H(d\theta).$$

This result when used in conjunction with Lemma 2.1 of Brown, yields the following two convergence relations (see Theorem 2.17 of Brown for details):

$$\int \theta e^{\theta \cdot x} H_n(d\theta) \rightarrow \int \theta e^{\theta \cdot x} H(d\theta).$$
and
\[ \int \|\theta\|^2 e^{\theta \cdot x} H_n(d\theta) \rightarrow \int \|\theta\|^2 e^{\theta \cdot x} H(d\theta). \tag{6.6} \]

The above three results together imply that for \( x \in B^0 \),
\[ \int \|\delta(x) - \theta\|^2 e^{\theta \cdot x} H_n(d\theta) \rightarrow \int \|\delta(x) - \theta\|^2 e^{\theta \cdot x} H(d\theta). \]

So, (6.2) holds for \( x' \). But since \( x' \) is arbitrarily chosen in \( K^0 \) and (6.1) is satisfied a.e. \((\nu)\), it follows that (6.2) is satisfied a.e. \((\nu)\). This proves the theorem.

We now state the following as corollaries of the above theorem.

**Corollary 6.1** If \( \Theta \) is closed in \( \mathbb{R}^p \) and \( \nu(\partial K) = 0 \), then (6.3) holds a.e. \((\nu)\), i.e., generalized Bayes procedures form a complete class. When \( \Theta \neq \mathbb{R}^p \), all admissible procedures need not be generalized Bayes, but (6.2) is still valid.

See Corollary 4.17 of Brown for proof.

**Corollary 6.2** The above theorem can be applied to the Poisson distribution when it is viewed as a canonical exponential family. That is, on transformation to the canonical parameter, under squared error discrepancy, an admissible estimator of the information-normalized loss or of the squared error loss can be represented in the form (6.2).

This is easy to verify.

We now suggest how the result of this chapter might be used to check for inadmissibility, though we shall not complete the computation. Let us consider the Gaussian case, the point estimator \( X + g(X) \), and \( L = \|X + g(X) - \theta\|^2 \). The posterior loss estimator of \( L \) with respect to the prior \( \pi(\theta) \) and the corresponding marginal density \( f_\pi(X) \) of \( X \), has the form
\[ p + \frac{\Delta f_\pi(X)}{f_\pi(X)} + \|g(X)\|^2 - 2g(X) \cdot \frac{\nabla f_\pi(X)}{f_\pi(X)} \tag{6.7} \]
CHAPTER 6. NECESSARY CONDITION FOR ADMISSIBILITY

Theorem 6.2 says that a necessary condition for $\gamma(X)$ to be an admissible estimator of $L$ is that there exist a distribution function $\Pi(d\theta)$ and a corresponding marginal $f_\pi(X)$ such that

$$\gamma(X) = (6.7) \quad \text{a.e.} \ (\nu).$$

From Stein (1981), the unbiased estimator of $L$ is given by

$$\gamma_{\text{unb}}(X) = p + \|g(X)\|^2 + 2\nabla \cdot g(X).$$

So a necessary condition for $\gamma_{\text{unb}}$ to be admissible is that the following differential equation have a solution $f_\pi(X)$ that corresponds to some prior $\pi$. This means that insolubility of the differential equation would imply inadmissibility of $\gamma_{\text{unb}}$. The differential equation is

$$2\nabla \cdot g(X) = \frac{\Delta f_\pi(X)}{f_\pi(X)} - 2g(X) \cdot \frac{\nabla f_\pi(X)}{f_\pi(X)},$$

that is,

$$\Delta u - 2g \cdot \nabla u - 2u\nabla \cdot g = 0$$

is the differential equation to be solved for $u(X)$. 
Bibliography


