A COMPROMISE BETWEEN THE BAYES AND MINIMAX APPROACHES 
TO ESTIMATION

BY

STANLEY H. SHAPIRO

TECHNICAL REPORT NO. 31
MARCH 16, 1972

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A COMPROMISE BETWEEN THE BAYES AND MINIMAX APPROACHES TO ESTIMATION

BY

STANLEY H. SHAPIRO

TECHNICAL REPORT NO. 31
MARCH 16, 1972

PREPARED UNDER THE AUSPICES
OF

PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>THE LOCATION PARAMETER PROBLEM</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Background</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Limited Translation Rules - Development</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Limited Translation Rules - Properties</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>The Exponential Shift Problem</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td>Multivariate Considerations</td>
<td>31</td>
</tr>
<tr>
<td>2.6</td>
<td>The Multivariate Normal Problem</td>
<td>35</td>
</tr>
<tr>
<td>2.7</td>
<td>Optimality Considerations</td>
<td>40</td>
</tr>
<tr>
<td>III</td>
<td>THE POISSON PROBLEM</td>
<td>44</td>
</tr>
<tr>
<td>3.1</td>
<td>Formulation</td>
<td>44</td>
</tr>
<tr>
<td>3.2</td>
<td>Limited Translation Rule and its Properties</td>
<td>49</td>
</tr>
<tr>
<td>3.3</td>
<td>Optimality Considerations</td>
<td>55</td>
</tr>
<tr>
<td>IV</td>
<td>THE SCALE PARAMETER PROBLEM</td>
<td>57</td>
</tr>
<tr>
<td>4.1</td>
<td>Formulation</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Limited Scale Rules - Development</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>Limited Scale Rules - Properties</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
<td>66</td>
</tr>
</tbody>
</table>
Chapter I

Introduction

The various philosophies of statistical estimation, as well as the wide variety of contexts in which statistical problems arise, have contributed to the development of a host of differing methodologies. The main thrust of estimation theory has been devoted to the development of estimators which are optimal in the light of some specific criterion, with vigorous debates sometimes arising between proponents of the differing schools of thought.

In this paper our efforts will be focused on compromising between the Bayes and minimax approaches to estimation theory. We shall not attempt to grapple with the essential philosophical differences between them, but instead attempt to combine the desirable features of both approaches in a cohesive manner.

The basic fixed sample size estimation problem involves making a more or less highly sophisticated guess as to the value of some unknown parameter, $\theta = (\theta_1, \ldots, \theta_k)$, based on a random sample of fixed size $\mathbf{X} = (X_1, \ldots, X_n)$, where the distribution of the observed random variables, $F_\theta(x)$, is conditionally dependent upon the parameter of interest. Rigorously the estimation problem consists of a sample space with an attendant $\sigma$-algebra, $(\mathcal{X}, \mathcal{F})$, a parameter space with its $\sigma$-algebra, $(\Theta, \mathcal{G})$, a measurable function $P$ defined on $\mathcal{G}$,
which maps elements of \( \Theta \) into probability measures, \( P_\theta \), on \( \mathcal{B}_\mathcal{X} \), an 
action space and its \( \sigma \)-algebra \( (\mathcal{B}_\mathcal{X}, \mathcal{B}_\mathcal{A}) \), and a loss function \( L \) defined 
on the cross product \( \Theta \times \mathcal{A} \), and taking values in the non-negative real 
numbers. In fuller generality the loss function could be defined on 
\( \Theta \times \mathcal{X} \times \mathcal{A} \), but this extension is not necessary for what follows.

Each member of the family of probability measures \( \{ P_\theta \}_{\theta \in \Theta} \) gives rise 
to an associated member of a family of distribution functions 
\( \{ F_\theta \}_{\theta \in \Theta} \). We assume the existence of a dominating measure on \( \mathcal{B}_\mathcal{X} \) and 
thus each distribution function \( F_\theta \) possesses an associated density 
function \( f_\theta \).

We shall call a formula for estimating \( \theta \) an estimation rule, or 
estimator, and we require that for each \( x \) it be a probability measure 
on \( \mathcal{B}_\mathcal{A} \). We shall denote the set of all possible estimators by \( \mathcal{A} \). For 
the specific problems that we shall be dealing with \( \Theta \) and \( \mathcal{A} \) will be 
subsets of \( \mathbb{E}^K \) where \( \mathbb{E}^K \) is \( K \)-dimensional Euclidean space, and the 
loss functions shall be, for fixed \( \theta \), convex functions of \( \mathcal{A} \). Under 
these conditions we can restrict our attention to nonrandomized 
estimators \( \delta(x) \), that is probability measures, which for each \( x \) 
concentrate all their mass at one point of \( \mathcal{A} \).

The risk function, \( R(\theta, \delta) \), for a fixed nonrandomized estimator 
\( \delta \) is a function from \( \Theta \) into the extended non negative reals:

\[
R(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta(x)) f_\theta(x) d\mu(x)
\]  

(1.1)
R(θ, δ) represents the average loss to the statistician when the true parameter value is θ, and the rule used to estimate it is δ.

The keystone of the Bayesian philosophy is the treatment of the unknown parameter as a random variable with some specified distribution function, $\Xi(\theta)$, referred to as the prior distribution. We assume that $\Xi$ is dominated by a $\sigma$-finite measure on $(\Theta, \mathcal{B}_\Theta)$. Although much controversy exists between Bayesians and non-Bayesians as to the validity of regarding the parameter as a random variable, the techniques of Bayesian estimation are widely used.

The goal of the Bayesian approach is the minimization of the integrated risk function $R(\pi, \delta)$, where the integration is performed over 'the' prior distribution $\pi$. The Bayes rule which we denote by $\delta_{\pi}$, is defined to be that estimator which accomplishes this minimization and the minimum value is referred to as the Bayes risk, $R(\pi, \delta_{\pi})$:

$$R(\pi, \delta_{\pi}) = \inf_{\delta \in \mathcal{D}} \int_{\Theta} R(\theta, \delta) d\pi(\theta).$$

(1.2)

Alternatively, the Bayes rule is often defined as the rule which minimizes the posterior risk

$$R^*(\pi, \delta_{\pi}) = \inf_{\delta} \int L(\theta, \delta) d\pi(\theta|\mathbf{x})$$

(1.3)

where $\pi(\theta|\mathbf{x})$ is the conditional distribution of the parameter given the observations. If $R^*(\pi, \delta_{\pi})$ is integrable with respect to the
marginal distribution of $X$, which we denote by $F(x)$ then the two minimization problems are identical.

Heuristically, if we have many independent repetitions of the same estimation problem, then the best estimator, as defined above, over the entire population of problems is the Bayes estimator. Moreover in this setting the Bayesian definition of 'best' seems quite reasonable.

One of the major drawbacks of Bayesian methods, however, is that they often give rise to estimators whose risk functions are unbounded. Experimentally there usually exists many subpopulations that are of interest, but distributionally get subsumed under one prior distribution. The fact that the Bayes estimator will at times do poorly is a reflection of this fact (a good discussion of this may be found in [5]). What we seek is a means of giving less weight to the 'composite' prior when we have reason to believe that it is leading us astray. If we are in a situation where 'large' risks cannot be tolerated we might be very apprehensive about blindly using the Bayes estimate.

Minimax estimation, on the other hand, focuses its efforts not on an averaged risk function, but on the risk function itself. Its goal is an estimator, which we shall denote by $\delta_0(x)$, such that the supremum of the estimator's risk function is less than or equal to the supremum of the risk function of any other estimator:
\[(1.4) \quad \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta) = \sup_{\theta \in \Theta} \inf_{\delta \in \mathcal{D}} R(\theta, \delta) \quad \forall \delta \in \mathcal{D}, \theta \in \Theta.\]

If we denote the supremum of any risk function as 'the worst', then the minimax rule guarantee's that 'its worst is the best of all possible worsts'.

Minimax estimators often behave poorly when looked at from an averaged risk point of view, the averaging being done with respect to the particular prior distribution of the problem. It is well known that if a least favorable distribution exists, i.e. a distribution \(\pi_0 \Theta\):

\[(1.5) \quad \inf_{\delta \in \mathcal{D}} R(\pi_0, \delta) = \sup_{\pi \in \mathcal{P}} \inf_{\delta \in \mathcal{D}} R(\pi, \delta),\]

where \(\mathcal{P}\) is the set of all prior distributions on \(\Theta\), then the minimax rule is Bayes with respect to it. Often the least favorable distribution is neither a mathematically convenient nor pragmatically relevant prior, and it is not surprising that a rule which is Bayes for one prior does not do particularly well for another. If we demand good performance with respect to our prior information about the parameter, then use of the minimax estimate would not be wise.

There are many situations for which both the local focus of minimax estimation and the global focus of Bayesian estimation are both very relevant. For example, medical clinics often draw blood samples from
their patients with the aim of implementing particular regimens based on the subsequent analysis. The analysis may involve the estimation of various parameters of the blood. Clearly it is desirable to use an estimator which will not do poorly for any one individual and which will also do well over the entire population of patients.

As a means of accomplishing this objective for any particular problem we propose a class of rules which we shall call limited deviation rules.

We impose the restraint that the rules shall not deviate by more than a specified amount from the minimax estimator and subject to this restraint we keep the rules as close to the Bayes rule as possible. Such vague statements as these are obviously not meant to satiate your thirst, but only to quicken it. I hope they do not kill it.
Chapter II

Translation Parameter Problem

2.1 Background

In this chapter we look at the problem of estimating a translation parameter. First we state some well known results dealing with minimax and Bayes estimates for the translation parameter problem, and then proceed to study the properties of the limited translation rules in this framework.

Definition 2.1 A parameter \( \theta \) is a location parameter for the distribution of a random variable \( X \) if the distribution function \( F_\theta(x) \) is a function solely of \( x-\theta \) i.e. \( F_\theta(x) = F_0(x-\theta) \)

We shall be dealing with location parameter families which are absolutely continuous with respect to some dominating measure, and in terms of the resulting density function \( f_\theta(x) \), the condition is \( f_\theta(x) = f_0(x-\theta) \).

As our choice of loss function we shall take squared error loss, \( L(\theta, a) = (a-\theta)^2 \). It's use is not essential to the following results, and we could substitute any loss function which is continuous and increasing in \( |a-\theta| \). Squared error loss is chosen here mainly for its mathematical convenience and its general usage in the location
parameter problem.

For completeness we state two famous results which we shall often use implicitly. Their proofs may be found in Girshick and Savage [7].

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with density function of the form $F_\theta(x) = F_\theta(x - \theta)$ and $E_\theta X_1^2 < \infty \forall \theta$. We shall let $\mathbf{x} = (x_1, \ldots, x_n)'$ denote the vector of observations, and as our loss function we take $L(\theta, a) = (a - \theta)^2$.

If $\eta(x)$ is any translation invariant estimator, i.e. $\eta(x + ce) = \eta(x) + c$, where $e = (1, 1, \ldots, 1)'$, then we have the following:

Theorem 2.1 Under the above conditions the rule

$$
\delta_0(x) = \eta(x) - E_\theta \left[ \frac{\eta(x)}{\mathbf{x}} - \eta(x)e \right]
$$

is minimax.

The rule $\delta_0(x)$ was first proposed by E. G. Pitman [Biometrika 39] and is referred to as the Pitman estimator. It is invariant, unbiased, has constant risk, and under certain conditions, Stein [16], is admissible.

The following theorem deals with the form of the Bayes rule. We let $\pi(\theta)$ be an arbitrary prior density on $(\theta, \mathcal{C}_0)$. Again we assume we have a random sample from a location parameter family with finite second moment, and squared error loss.
Theorem 2.2. If for each \( x \) actions \( a_x, b_x \in \Theta \)

\[
E_x \left[ \frac{(\theta - a_x)^2}{x} \right] < \infty, \quad E_x \left[ \frac{(\theta - b_x)^2}{x} \right] < \infty
\]

then

\( (2.2) \quad \delta_n(x) = E_x \theta \quad \text{and} \quad R(\pi, \delta_n) < \infty \)

i.e. the Bayes rule is the expected value of \( \theta \) under the posterior distribution and the Bayes risk is finite.

The assumptions of the above theorems avoid uninteresting cases such as every rule being minimax or Bayes. To avoid annoying repetition we note that the hypotheses of the above theorems will be assumed throughout this chapter unless otherwise noted.

We shall assume throughout that the only unknown parameter is the parameter of interest \( \theta \) which may be either scalar or vector valued. For the present we take \( \theta \) to be 1-dimensional and treat the problem of \( \theta \) k-dimensional in section 2.5. Also we shall use the terms "Fitman estimator" and "minimax estimator" interchangeably, again unless we note otherwise.

2.2 Limited Translation Rules - Development

Our aim is to estimate the unknown parameter \( \theta \) by means of a rule which performs well with respect to the prior distribution of
the problem and yet whose risk function is bounded by some constant at our control. One means of achieving this goal is to restrict the minimum deviation of our estimate from the minimax estimate while simultaneously keeping as close to the Bayes rule as possible. Since the variance of the underlying family of distributions is constant and we are dealing with problems of location, it seems intuitively reasonable to look at translation deviations, and, to limit the translation by a fixed quantity independent of the observed sample.

Letting the maximum deviation be $d$ then if

\[(2.3) \quad \delta_\pi(x) \in \left[\delta_0(x) - d, \delta_0(x) + d\right],\]

our estimate would be $\delta_\pi(x)$. If $\delta_\pi(x)$ is not in the above interval, than the estimate which is closest to $\delta_\pi(x)$ and yet does lie in the interval is either $\delta_0(x) - d$, if $\delta_\pi(x) < \delta_0(x)$, or $\delta_0(x) + d$, if $\delta_\pi(x) > \delta_0(x)$.

Now since we are concerned with translations from $\delta_0(x)$, a reasonable choice for $d$ would be to fix some constant $m > 0$ and take $d = m\sigma_0$, where $\sigma_0$ is the standard deviation of the Pitman estimator. What this does is limit the deviation from $\delta_0$ in terms of its own standard deviation.

We observe the sample $x_1, x_2, \ldots, x_n$ generated from the distribution $F_\theta$ and proceed to calculate the Bayes and minimax estimators. The constraint on the estimator $\delta(x)$ is:
$$|\delta(x) - \delta_0(x)| \leq m\sigma_0$$

If (1.21) is satisfied with $\delta(x) = \delta_\pi(x)$ then our estimate would be the Bayes estimate. If it is violated then our estimate would be the minimax estimate shifted $m\sigma_0$ units in the direction of the Bayes estimate, i.e. $\delta_0(x) + \text{sgn}(\delta_\pi(x) - \delta_0(x))m\sigma_0$.

Letting $A^* = \{x: \delta_0(x) - \delta_\pi(x) < -m\sigma_0\}$, $B^* = \{x: |\delta_0(x) - \delta_\pi(x)| \leq m\sigma_0\}$,

$C^* = \{x: \delta_0(x) - \delta_\pi(x) > m\sigma_0\}$ we have the following

**Definition 2.2.** The limited translation rule for estimating the location parameter $\theta$, with prior $\pi$ and maximum translation $m\sigma_0$ is $\delta_{\pi, m}(x)$ where:

$$\delta_{\pi, m}(x) = \begin{cases} \delta_0(x) + m\sigma_0 & x \in A^- \\ \delta_\pi(x) & x \in B^- \\ \delta_0(x) - m\sigma_0 & x \in C^- \end{cases}$$

In cases where the minimax rule is also the maximum likelihood estimator, the procedure has an appealing intuitive interpretation. If $|\delta_\pi(x) - \delta_0(x)|$ is 'large' then we have a warning that a discrepancy exists between the data we have observed and our assumed prior. The sample has already exerted its formal influence on the prior through
means of the posterior distribution, and at this point if we
were anything other than a 'come Hell or high water - my priors are
the right priors' Bayesian, we would probably have doubts about the
appropriateness of the specified prior to the problem at hand. In
such a situation it would seem reasonable to de-emphasize the role
of the prior.

An obvious drawback of the above procedure is that in practice
the determination of the sets $A^*, B^*, C^*$ may be quite involved. However,
for a wide variety of families we are guaranteed that this complica-
tion can be avoided, specifically, those families of distributions
which independent of the sample size possess a one dimensional
sufficient statistic. This embraces both the exponential family as
well as those families of distributions which would be included in
the former except for the fact that their support $\{x: F_\theta(x) > 0\}$
depends on $\theta$. In the literature such distribtuions are often referred
to as 'irregular exponential distributions'. By sufficiency arguments
we may restrict our attention to the sufficient statistic, call it $t$.
In this situation the induced sets:

$$A = \{t: \delta_0(t) - \delta_\pi(t) < -m\sigma_0\}$$
$$B = \{t: |\delta_0(t) - \delta_\pi(t)| < m\sigma_0\}$$
$$C = \{t: \delta_0(t) - \delta_\pi(t) > m\sigma_0\}$$

will be unions of intervals, possibly degenerate, on the real line.
Moreover if the Bayes and minimax rules are both linear functions of \( t \) then the sets themselves will be intervals. We present an example below.

Let

\[
X_i \sim \text{iid } e^{-(x-\theta)} \quad 0 < \theta < x_i < \infty \quad i=1, \ldots, n
\]

(2.6)

\[
\pi(\theta) = \frac{\theta^{\alpha-1} e^{-n\theta}}{\Gamma(\alpha)} \quad 0 < \theta < \infty, \quad \alpha > 0
\]

\[
L(\theta, a) = n^2 (a-\theta)^2
\]

Zacks [17] presents an appealing argument for deducing Pitman's estimator without resort to computation. Note that the minimal sufficient statistic is \( t(x) = x_{(1)} \). If \( \delta \) is any translation invariant estimate based on \( x_{(1)} \) then \( \delta(x_{(1)} + c) - \delta(x_{(1)}) = c \) \( \forall c \) and thus every translation invariant rule is of the form \( \delta_b(x_{(1)}) = x_{(1)} + b \)

where \( b \) is some constant. We want the rule, or equivalently the constant \( b_0 \) which minimizes \( n^2 \mathbb{E}_\theta(x_{(1)} + b - \theta)^2 \). Since

\[
-x_{(1)} - \theta \sim \text{ne} \quad x_{(1)} > \theta \quad \text{we have}
\]

\[
n^2 \mathbb{E}_\theta(x_{(1)} - \theta - \frac{1}{n} + \frac{1}{n} + b)^2 = 1 + (1 + nb)^2
\]

From the above it is obvious that \( b_0 = -\frac{1}{n} \) and the Pitman estimator is

(2.7)

\[
\delta_0(x_{(1)}) = x_{(1)} - \frac{1}{n}.
\]
From direct calculation we get that

\[(2.8) \quad \pi(\theta | x(1)) = \frac{\alpha^\theta x^{\alpha-1}}{x(1)^\alpha} \quad 0 < \theta < x(1)\]

and thus

\[(2.9) \quad \delta_\pi(x) = \frac{\alpha}{\alpha+1} x(1)\]

and so

\[(2.10) \quad \delta_{\pi,m}(x) = \begin{cases} 
\frac{x(1) + (m-1)}{n} & x(1) < \frac{(\alpha+1)(1-m)}{n} \\
\frac{\alpha}{\alpha+1} x(1) & \frac{(\alpha+1)(1-m)}{n} \leq x(1) < \frac{(\alpha+1)(1+m)}{n} \\
x(1) - \frac{(1+m)}{n} & x(1) > \frac{(\alpha+1)(1+m)}{n}
\end{cases}\]

Note that for \(m < 1\) and \(x(1) < \frac{1-m}{n}\), \(\delta_{\pi,m}\) will produce a negative estimate for \(\theta\), although a priori we know \(\theta\) to be positive. This is a reflection of the fact that for \(x(1) < \frac{1}{n}\), \(\delta^0(x) < 0\). There are several alternatives which we might choose to follow. One would be to truncate \(\delta^0\) at 0 giving rise to the new rule \(\delta^1_0 = \max(\delta^0_0, 0)\). Intuitively we would expect \(\delta^1_0\) to do better for small values of \(x(1)\) and just as well for larger values. This is reflected in the fact that \(R(\theta, \delta^1_0) \leq R(\theta, \delta^0_0)\), with strict inequality holding for \(\theta < \frac{1}{n}\). \(\delta^1_0\) however, does not possess several of the appealing features which \(\delta^0_0\) does, specifically invariance, unbiasedness, and linearity in \(x(1)\). Another option would be to truncate
\( \delta_{\pi, m} \) at 0 giving rise to the estimator \( \hat{\delta}_{\pi, m} \). This choice however, broadens the scope of the limited translation rules to a point which precludes the development of the relevance function approach of later sections. A third alternative would be to restrict ourselves to values of \( m \geq 1 \). It does not seem unreasonable to assume that in practice we would tolerate deviations of one standard deviation. We shall deal more fully with this problem in section 2.4.

2.3 Limited Translation Rules - Properties

The properties of the limited translation rules will obviously depend on the quantity we choose to limit deviation from \( \delta_0 \). It is intuitively clear that \( R(\pi, \delta_{\pi, m}) \) is a monotonically decreasing function of \( m \), as the larger the value of \( m \), the closer we will be able to follow the Bayes rule. A rigorous proof of this is presented in corollary 2.6. This is bought at a price however, for the larger \( m \) is the greater the deviation from the minimax rule and thus we may incur larger individual risks. The following theorem shows however, that for any value of \( m \) the risk function is bounded by the line \( R = \sigma_0^2 (1+m)^2 \). The theorem holds in general and not for just the exponential family.

Theorem 2.3. Let \( \theta \) be a location parameter, \( L(\theta, a) = (a - \theta)^2 \), and \( \delta_{\pi, m} \) be as defined in (2.5) then
\((2.11) \quad R(\theta, \delta_{\pi, m}) \leq \sigma_0^2 (1 + m)^2 \quad \forall \theta \)

**Proof**

\[
R(\theta, \delta_{\pi, m}) = E_\theta (\delta_{\pi, m} - \theta)^2 \\
= E_\theta (\delta_{\pi, m} - \delta_0 + \delta_0 - \theta)^2 \\
= E_\theta (\delta_{\pi, m} - \delta_0)^2 + 2E_\theta (\delta_{\pi, m} - \delta_0)(\delta_0 - \theta) + E_\theta (\delta_0 - \theta)^2 \\
\leq m \sigma_0^2 + 2m \sigma_0^2 + \sigma_0^2 \\
= \sigma_0^2 (1 + m)^2
\]

\((2.12)\)

(2.12) follows from the fact that \(|\delta_{\pi, m}(x) - \delta_0(x)| \leq m \sigma_0 \forall x\), and an application of the Cauchy-Schwarz Inequality. 

In certain cases we can provide a better bound for \(R(\theta, \delta_{\pi, m})\).

In particular if \(\delta_0(x)\) and \(\delta_{\pi}(x)\) are both linear functions of the sufficient statistic, then for some \(c_1, c_2\) (possibly \(\pm \infty\)) respectively we have \(A = (-\infty, c_1)\) \(B = [c_1, c_2]\) \(C = (c_2, \infty)\) which gives rise to the following theorem.

**Theorem 2.4** If \(\delta(t) \equiv \delta_{\pi, m}(t) - \delta_0(t)\) is a monotonically non-increasing function of \(t\) then

\[
R(\theta, \delta_{\pi, m}) \leq \sigma_0^2 (1 + m^2)
\]
Proof

From the proof of the preceding theorem we have that

\[ R(\theta, \delta, \pi, m) = \sigma^2 + 2E_\theta \delta(\dot{\delta}_0 - \theta) + m^2 \sigma_0^2 \]

Note that \( \delta_0(t) - \theta \) is monotone increasing in \( t \). We let

\[
\begin{align*}
(\delta_0(t) - \theta)^- &= \max(- (\delta_0(t) - \theta), 0) \\
(\delta_0(t) - \theta)^+ &= \max(\delta_0(t) - \theta, 0)
\end{align*}
\]

Let \( t_0 \) be the point where \( \delta_0(t) = \theta \) and let \( V \) denote the value of \( \delta(t) \) at \( t_0 \), \( V = \delta(t_0) \).

From the unbiasedness of \( \delta_0(t) \) we have

\[ VE_\theta(\delta_0(t) - \theta)^- = VE_\theta(\delta_0(t) - \theta)^+ \]

(2.13): \( E_\theta \delta(t) (\delta_0(t) - \theta)^- \geq VE_\theta(\delta_0(t) - \theta)^- = VE_\theta(\delta_0(t) - \theta)^+ \geq E_\theta \delta(t)(\delta_0(t) - \theta)^+ \)

\[ 0 \geq E_\theta \delta(t) [(\delta_0(t) - \theta)^+ - (\delta_0(t) - \theta)^-] = E_\theta \delta(t)(\delta_0(t) - \theta) \]

(2.13) follows from the monotonicity of \( \delta(t) \) and the last line from the definitions. \( \Box \)

17
The proof of the theorem is essentially a proof of the fact that the covariance between a monotone increasing and monotone decreasing function is negative. Note that the unbiasedness of \( \delta_0(t) \) is not crucial in that we could subtract its bias without altering the covariance structure.

Turning our attention to the other side of the problem, we seek a means of quantifying how well the limited translation rules perform from a Bayesian analysis. Following the example of Efron and Morris, we shall call the difference between the integrated risk of a rule \( \delta \), the integration being performed with respect to the particular prior, \( \pi \), of the problem, and the integrated risk of the minimax estimator \( \delta_0 \), as the "savings of \( \delta \) vs \( \pi \)". Note that for any particular problem the maximum savings is achieved by the Bayes rule for that problem. In the exponential example of section 2.4, we see that \( R(\pi, \delta_0) = 1 \), \( R(\pi, \delta_\pi) = \frac{\alpha}{\alpha + 1} \) and so the maximum savings is \( \frac{1}{\alpha + 1} \). Normalizing the savings of \( \delta_{\pi, m} \) by the maximum savings for the problem provides a relative measure of how well \( \delta_{\pi, m} \) performs. Efron and Morris call this quantity the 'relative savings of \( \delta_{\pi, m} \) vs \( \pi \)' and denote it by \( s \)

\[
2.14 \quad s = \frac{R(\pi, \delta_0) - R(\pi, \delta_{\pi, m})}{R(\pi, \delta_0) - R(\pi, \delta_\pi)}
\]

We shall choose to work with
(2.15) \[ 1 - s = \frac{R(\pi, \delta_{\pi, m}) - R(\pi, \delta_{\pi})}{R(\pi, \delta_{\sigma}) - R(\pi, \delta_{\pi})} \]

the 'relative savings loss', which provide a measure of how well \( \delta_{\pi, m} \) does vs \( \delta_{\pi} \) on the latter's home ground. By construction \( 0 \leq 1 - s \leq 1 \), the bounds being achieved for the degenerate cases \( m = 0 \), i.e. \( \delta_0 \), and \( m = \infty \), i.e. \( \delta_{\pi} \). Note that small values of \( 1 - s \) are desirable from the Bayesian view.

Central to the study of the relative savings loss of the limited translation rules, as well as the theory of limited deviation rules in general, is the notion of relevance functions.

Let

(2.16) \[ u(x) = \frac{|\delta_{\pi}(x) - \delta_0(x)|^2}{r}, \quad B = \sqrt{r m \sigma_0}, \quad r = R(\pi, \delta_0) - R(\pi, \delta_{\pi}) \]

and define the function \( \rho(u) = \min (1, \frac{B}{\sqrt{u}}) \) i.e.

(2.17) \[ \rho(u) = \begin{cases} 1 & u \leq B^2 \\ \frac{B}{\sqrt{u}} & u > B^2 \end{cases} \]

We associate with \( \rho \) the decision rule \( \delta_{\rho}(x) \)

(2.18) \[ \delta_{\rho}(x) = [1 - \rho(u)] \delta_0(x) + \rho(u) \delta_{\pi}(x) = \delta_0(x) + \rho(u) [\delta_{\pi}(x) - \delta_0(x)] \]
A moment's reflection should convince us that with \( \rho(u) \) defined in the above way, the rule \( \delta_{\rho} \) is merely our old friend \( \delta_{\pi,m} \). To help that reflection note that

\[
\begin{align*}
\text{if } u < B^2 &\Rightarrow \delta_{\rho}(x) = \delta_{\pi}(x) \\
\text{if } u > B^2 &\Rightarrow \delta_{\rho}(x) = \delta_{0}(x) \pm m\sigma_0
\end{align*}
\]

the appropriate sign depending upon whether \( \delta_{\pi}(x) - \delta_{0}(x) \) is greater than or less than 0. It should not be surprising that \( \delta_{\rho}(x) = \delta_{\pi,m}(x) \) since we have defined \( \rho(u) \) so that this would follow. Of course there are an infinite number of ways we could choose to define the function \( \rho \) and each one would give rise to an associated decision rule.

**Definition 2.3** Letting \( \rho(u) \) be an arbitrary function taking values in the unit interval and \( \delta_{\rho}(x) \) be as defined in (2.16). We shall call \( \rho(u) \) the relevance function of the rule \( \delta_{\rho} \).

The definition is intuitively appealing in that \( \rho \) represents the weight or relevance we assign to the Bayes rule for the problem of estimating \( \theta \), and \( (1-\rho) \) represents the relevance of the minimax estimate.

With the above formulation we can derive an appealing representation of the Bayes risk of the class of rules \( \delta_{\rho} \) and consequently their relative savings loss. Recall that the limited translation rules are a sub-class of these rules. The following theorem exhibits the Bayes risk of \( \delta_{\rho} \) as a linear combination of the Bayes risks of
the minimax rule and the Bayes rule.

Theorem 2.5

For any \( \rho \) let

\[
1 - s = \frac{E_{\pi}(\delta_{\pi} - \delta_{\theta})^2(1-\rho)^2 = \mathbb{E}U(1-\rho(U))^2}
\]

where the expectation is over the marginal distribution of \( x \).

Then:

\[
R(\pi, \delta_{\rho}) = (1-s)R(\pi, \delta_{\theta}) + sR(\pi, \delta_{\pi})
\]

Proof

\[
R(\pi, \delta_{\rho}) = \mathbb{E}_{\theta} \left[ \theta - \{ (1-\rho) \delta_{\theta} + \rho \delta_{\pi} \} \right]^2
\]

\[
= \mathbb{E}_{\theta} \left[ \theta - \delta_{\pi} + (\delta_{\pi} - \delta_{\theta})(1-\rho) \right]^2
\]

\[
= R(\pi, \delta_{\pi}) + \mathbb{E}(\delta_{\pi} - \delta_{\theta})^2(1-\rho)^2
\]

(2.23)

\[
= R(\pi, \delta_{\pi}) + (1-s)[R(\pi, \delta_{\theta}) - R(\pi, \delta_{\pi})]
\]

\[
= (1-s)R(\pi, \delta_{\theta}) + sR(\pi, \delta_{\pi})
\]

As the following corollary shows the choice of \( 1 - s \) to
to denote $\frac{E(\delta_{\pi} - \delta_{\pi_0})^2(1-\nu)^2}{r}$ is consistent with our previous notation in that $1 - s$ is in fact the relative savings loss of the rule $\delta_{\rho}$.

Corollary 2.5.

(2.24) \[ l - s = \frac{R(\pi, \delta_{\rho}) - R(\pi, \delta_\mu)}{R(\pi, \delta_{\mu}) - R(\pi, \delta_\mu)} \]

Proof

The proof is immediate from (2.21):

\[
R(\pi, \delta_{\rho}) = R(\pi, \delta_{\pi}) + (1-s)[R(\pi, \delta_{\mu}) - R(\pi, \delta_{\pi})]
\]

\[
\therefore l - s = \frac{R(\pi, \delta_{\rho}) - R(\pi, \delta_{\mu})}{R(\pi, \delta_{\mu}) - R(\pi, \delta_{\pi})}. \quad ||
\]

Recalling that $\rho(u)$ as defined in (2.17) is the relevance function of $\delta_{\rho} = \delta_{\pi, m}$ we can easily prove the following.

Corollary 2.6 The Bayes risk of the limited translation rule $R(\pi, \delta_{\pi, m})$ is a monotonically decreasing function of $m$.

Proof

Let $m' > m$ and denote the relevance functions corresponding to the rules $\delta_{\pi, m'}$, $\delta_{\pi, m}$ as $\rho_{m'}$ and $\rho_{m}$. From their definition we have

\[ 0 \leq \rho_{m}(u) \leq \rho_{m'}(u) \leq 1 \quad \forall u \]
and thus

\[ \text{EU}[(1-p_m^-(U))^2] \leq \text{EU}[1-p_m(U)]^2. \]

Therefore, from Theorem 2.5 we have

\[ (2.25) \quad R(\pi, \delta, \nu, m) \leq R(\pi, \delta, \nu, m) . \]

### 2.4 The Exponential Shift Problem

We return now to the problem of estimating the shift parameter of the exponential distribution. Recalling we have:

\[ X_1, X_2, \ldots, X_n \sim \text{iid } f_\theta(x) \text{ with} \]

\[ (2.26) \quad f_\theta(x) = e^{(x-\theta)} \quad 0 < \theta < x < \infty \]

\[ (2.27) \quad \pi(\theta) = \frac{n^\alpha e^{-\alpha \theta}}{\Gamma(\alpha)} \quad 0 < \theta < \infty , \quad \alpha > 0 \]

The loss function is normalized squared error loss:

\[ L(\theta, \alpha) = n^2(\alpha - \theta)^2 \]

The minimal sufficient statistic is \( t(x) = x_{(1)} \), the smallest order statistic of the sample. Note that its density function is:

\[ f_\theta(t) = e^{-\theta (t-\theta)} \quad t > \theta \]
By making the transformation \( t = nt \quad \theta = n\theta \)
we can reduce the problem to canonical form:

\[
(2.28) \quad f_\theta(t) = e^{-(t-\theta)} \quad 0 < t
\]

\[
(2.29) \quad \pi(\theta) = \frac{\theta^{a-1} e^{-\theta}}{\Gamma(a)} \quad 0 < \theta < \infty
\]

\[L(\theta,a) = (a-\theta)^2\]

From the above we have that the conditional distribution of \( \theta \) given \( t \) is:

\[
(2.30) \quad \pi(\theta/t) = \frac{\theta^{a-1}}{t^a} \cdot 0 < \theta < t,
\]

and thus,

\[
(2.31) \quad E_t \theta = \delta_{\pi}(t) = \frac{\alpha}{\alpha+1} t
\]

Note that:

\[
R(\theta,\delta_\pi) = \int_0^\infty \left(\frac{\alpha}{\alpha+1} t-\theta\right)^2 e^{-t} dt
\]

\[
= \left(\frac{\alpha}{\alpha+1}\right)^2 \left[\theta^2+2\theta+2\right] - \frac{2\theta\alpha}{\alpha+1} [\theta+1] + \theta^2
\]

\[
(2.32) \quad = [\theta - \frac{\alpha}{\alpha+1}(\theta+1)]^2 + \frac{\alpha^2}{(\alpha+1)^2} \quad \theta^2 + \left(\frac{\alpha}{\alpha+1}\right)^2
\]

24
A convex function of $\theta$ with minimum \(\frac{-\alpha}{\alpha+1}^2\) achieved at $\theta = \alpha$, which goes off to $+\infty$ as $\theta$ grows large. Since $\alpha = E\theta = V(\theta)$, from (2.29) we see that

\[
R(\pi,\delta_\pi) = \frac{\alpha + \alpha^2}{(\alpha+1)^2} = \frac{\alpha}{\alpha+1}
\]

(2.33)

From (2.7) we know that $\delta_0(t) = t-1$ is minimax with constant risk $R(\theta,\delta_0) = 1$. For future use we state the obvious fact that

(2.34)

\[R(\pi,\delta_0) = 1.\]

We note again that the rule $\delta_0^1(t) = \max (\delta_0(t), 0)$ outperforms $\delta_0(t) = t-1$, a consequence of the fact that $\theta e(0, \infty) \triangleright (\delta_0(t)-\theta)^2 \geq (\delta_0^1(t)-\theta)^2$ with strict inequality holding for $t < 1$. From direct computation we have:

\[
R(\theta, \delta_0^1) = \int_0^1 \theta^2 e^{-(t-\theta)} I_{\{t > \theta\}} dt + \int_1^\infty (t-1-\theta)^2 e^{-(t-\theta)} I_{\{t > \theta\}} dt
\]

\[
= \theta^2 (1-e^{-\theta}) I_{\{\theta < 1\}} + \theta^{-1} (2-2\theta + \theta^2) I_{\{\theta < 1\}} + I_{\{\theta \geq 1\}}
\]

where the last two terms come from the second integral

(2.35)

\[
R(\theta, \delta_0^1) = (\theta^2 - 2\theta e^{\theta-1} + 2e^{\theta-1}) I_{\{\theta < 1\}} + I_{\{\theta \geq 1\}}
\]

25
Integrating this with respect to (2.50) and noting that
\[ g(\alpha+1, x) = g(\alpha, x) - \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} \]
(where \( g(\alpha, x) \) is the incomplete gamma function \( \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \)) we have

\[ R(\Pi, \delta^1_0) = \int_0^1 2(1-e^{-\theta}) \frac{\theta^{\alpha-1} e^{-\theta}}{\Gamma(\alpha)} d\theta + \int_1^\infty \frac{\theta^{\alpha-1} e^{-\theta}}{\Gamma(\alpha)} d\theta \]

\[ = \alpha(\alpha+1) g(\alpha+2, 1) - \frac{2e^{-1}}{(\alpha+1) \Gamma(\alpha)} + \frac{2e^{-1}}{\Gamma(\alpha+1)} + 1 - g(\alpha, 1) \]

(2.36)

\[ = 1 + g(\alpha, 1) (\alpha^2 + \alpha - 1) - \frac{e^{-1} (\alpha^3 + 3\alpha^2 + 2\alpha - 2)}{\Gamma(\alpha) (\alpha + 1)} \]

The marginal (unconditional) density of \( t \) is:

(2.37)

\[ f(t) = \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha+1)} \quad t > 0 \]

and thus for 'small' values of \( \alpha \) we would expect to observe small values of \( t \). In this situation we would be more inclined to limit deviations from \( \delta^1_0 \) than \( \delta_0 \).

Basing the limited translation rule on \( \delta^1_0 \) and \( \delta_0 \) it is not immediately obvious by what quantity we would want to limit deviations from \( \delta^1_0 \). The problem arising from the fact that \( \delta^1_0 \) does not have constant risk. One possibility would be \( m^1 \sigma^0 \)

\[ \sigma^1_0 = \sqrt{R(\Pi, \delta^1_0)} \]. Since \( \sigma^1_0 < \sigma_0 \) we would be inclined to choose values of \( m^1 \) larger than the corresponding values of \( m \).

For the particular problem we are dealing with now, the
condition \( |\delta_0^{1}(t) - \delta_{-1}(t)| \leq m_0^{1} \) becomes

\[(2.38) \quad 0 - \frac{\alpha}{\alpha + 1} t \leq m_0^{1}, \quad t \leq l \quad \text{or} \quad t - l - \frac{\alpha}{\alpha + 1} t \leq m_0^{1}, \quad t > l \]

or

\[(2.39) \quad t \leq m_0^{1} \left( \frac{\alpha + 1}{\alpha} \right), \quad t \leq l \quad (\alpha + 1)(1 - m_0^{1}) \leq t \leq (\alpha + 1)(1 + m_0^{1}), \quad t > l \]

and the limited translation rule is:

\[(2.40) \]

\[
\delta^{1}_{\pi, m} = \begin{cases} 
\frac{\alpha}{\alpha + 1} t, & 0 < t \leq m_0^{1} \left( \frac{\alpha + 1}{\alpha} \right) \cup \left( (\alpha + 1)(1 - m_0^{1}) \leq t \leq (\alpha + 1)(1 + m_0^{1}) \right) \\
\frac{m_0^{1}}{\alpha} \quad & \left( \frac{\alpha + 1}{\alpha} \right) < t \leq l \\
t - l + m_0^{1}, & l < t < (\alpha + 1)(1 - m_0^{1}) \\
t - l - m_0^{1}, & t > (\alpha + 1)(1 + m_0^{1}) 
\end{cases}
\]

The limited translation rule based on \( \delta_{0} \) and restricting deviations to be less than \( m \), where \( m \geq l \), is

\[(2.41) \]

\[
\delta^{1}_{\pi, m} = \begin{cases} 
\frac{\alpha}{\alpha + 1} t, & 0 < t \leq (\alpha + 1)(1 + m) \\
t - l, & t > (\alpha + 1)(1 + m) 
\end{cases}
\]

27
The suggestion of taking \( m > l \) corresponds to taking \( m^l > \frac{1}{\sigma_0^l} \). Since \( \sigma_0^l < 1 \) we know, a priori, that the lower bound on \( m^l \) is greater than 1. Fortunately or unfortunately, \( \sigma_0^l \) cannot claim any more validity to the problem than can \( \sigma_0 \), both are arbitrary. We would like to have \( m^l < \frac{\alpha}{(\alpha+1)\sigma_0^l} \), for it is the set \( \{ t : m^l \sigma_0^l \frac{(\alpha+1)}{\alpha} < t < 1 \} \) which primarily motivates the use of \( \delta_{\pi,m^l} \). In Table 1 below we present the value of this upper bound for various values of \( \alpha \). From the table we see that requiring that \( m^l < \frac{\alpha}{(\alpha+1)\sigma_0^l} \) places a greater restriction on \( m^l \) then would seem prudent.

In addition the fact that for increasing values of \( \alpha, \delta_0 \) and \( \delta_0^l \) tend to better agreement and the fact that the investigation of \( \delta_{\pi,m^l} \) although in principle no more involved, is computationally more difficult than that of \( \delta_{\pi,m} \), provide motivation for our studying the behavior of \( \delta_{\pi,m} \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>.1</th>
<th>.25</th>
<th>.50</th>
<th>1</th>
<th>2.5</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\alpha}{(\alpha+1)\sigma_0^l} )</td>
<td>.10</td>
<td>.23</td>
<td>.37</td>
<td>.53</td>
<td>.72</td>
<td>.83</td>
</tr>
</tbody>
</table>

**TABLE 1** Values of the bound \( \frac{\alpha}{(\alpha+1)\sigma_0^l} \) as a function of \( \alpha \)
Conditional upon a specific value of \( \theta \), \( f_\theta(t) \) concentrates its mass close to that value of \( \theta \), while \( \pi(\theta) \) puts most of its weight about \( \alpha \). \( \delta_{\pi,m} \) reflects the fact that for large observations the appropriateness of \( \pi(\theta) \) is in question. From a Bayesian viewpoint when \( t > (\alpha+1)(l+m) \), \( \delta_{\pi,m} \) de-emphasizes the role of \( \pi \).

Letting \( c = (\alpha+1)(l+m) \), \( k = \frac{\alpha}{\alpha+1} \) the risk function of \( \delta_{\pi,m} \) for \( \theta < c \) is

\[
R(\theta, \delta_{\pi,m}) = \int_0^c (kt-\theta)^2 e^{-(t-\theta)} \, dt + \int_c^{\infty} (t-l+m-\theta)^2 e^{-(t-\theta)} \, dt
\]

\[
= e^{\theta-c} \left[ (\theta-l(c+1))^2 + k^2 \right] + \left[ \frac{\theta}{\alpha+1} - k \right]^2 + k^2 + e^{\theta-c} \left[ (\theta-c)^2 + 1 \right]
\]

(2.42)

\[
R(\theta, \delta_{\pi,m}) = 2e^{\theta-c} \left[ \frac{\alpha(m+1)}{\alpha+1} + 1 - \left( \frac{\alpha}{\alpha+1} \right)^2 - \frac{\theta}{\alpha+1} \right] + \frac{1}{(\alpha+1)^2} \left[ (\theta-\alpha)^2 + \alpha^2 \right]
\]

For \( \theta \geq c \) we have

\[
R(\theta, \delta_{\pi,m}) = \int_0^{\infty} (t-l+m-\theta)^2 e^{-(t-\theta)} \, dt
\]

(2.43)

\[
R(\theta, \delta_{\pi,m}) = l + m^2
\]

Combining the above we have

(2.44)

\[
R(\theta, \delta_{\pi,m}) = \left\{ 2e^{\theta-c} \left[ \frac{\alpha(m+1)}{\alpha+1} + 1 - \left( \frac{\alpha}{\alpha+1} \right)^2 - \frac{\theta}{\alpha+1} \right] + \frac{1}{(\alpha+1)^2} \left[ (\theta-\alpha)^2 + \alpha^2 \right] \right\} I_{\{\theta < c\}}
\]

\[+ (l+m^2) I_{\{\theta > c\}} \]
In Figure 1 below we present a plot of the risk functions $R(\theta, \delta_0)$, $R(\theta, \delta_\pi)$, and $R(\theta, \delta_{\pi,m})$ for the case $\alpha = 1$, $m = 1$. In general $R(\theta, \delta_{\pi,m})$ will be above $R(\theta, \delta_\pi)$ for 'small' values of $\theta$ and below $R(\theta, \delta_\pi)$ for moderate and large values of $\theta$.

In Table 2 we have the relative savings loss of $\delta_{\pi,m}$ for various values of $\alpha$ and $m$. For example if $\alpha = .5$ by choosing $m = 1.4$ we can insure that that $\sup R(\theta, \delta_{\pi,m}) \leq 2.96$ while at the same time losing only about 10% of the Bayes savings.

The computed entries of Table 2 were calculated from the relations

\[(2.45) \quad R(\pi, \delta_{\pi,m}) = 2e^{-c} \left\{ \frac{-e^{\alpha + 1}}{(\alpha + 1)^2 \Gamma(\alpha)} + \frac{e^{\alpha}}{\Gamma(\alpha + 1)} \left[ \frac{\alpha}{\alpha + 1} (\frac{\alpha}{\alpha + 1}) + 1 - \left( \frac{\alpha}{\alpha + 1} \right)^2 \right] \right\} \]

\[\quad + \frac{1}{(\alpha + 1)^3} \{ \alpha (\alpha + 1) g(\alpha + 2, c) - 2\alpha^2 g(\alpha + 1, c) + 2\alpha^2 g(\alpha, c) \} \]

\[\quad + (1 + m^2) \left[ 1 - g(\alpha, c) \right] \]

\[(2.46) \quad 1 - s = (\alpha + 1)R(\pi, \delta_{\pi,m}) - \alpha \]

where (2.45) is found directly by integrating (2.44) with respect to $\pi$.  

30
TABLE 2  Relative Savings loss of $\delta_{\pi,m}$
tabled for various values of $\alpha$ and $m$

<table>
<thead>
<tr>
<th>$\alpha/m$</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>.249</td>
<td>.200</td>
<td>.162</td>
<td>.130</td>
<td>.105</td>
<td>.084</td>
</tr>
<tr>
<td>.25</td>
<td>.220</td>
<td>.173</td>
<td>.137</td>
<td>.108</td>
<td>.085</td>
<td>.067</td>
</tr>
<tr>
<td>.50</td>
<td>.182</td>
<td>.138</td>
<td>.105</td>
<td>.079</td>
<td>.060</td>
<td>.045</td>
</tr>
<tr>
<td>1.00</td>
<td>.128</td>
<td>.091</td>
<td>.064</td>
<td>.045</td>
<td>.032</td>
<td>.022</td>
</tr>
<tr>
<td>2.50</td>
<td>.052</td>
<td>.031</td>
<td>.018</td>
<td>.010</td>
<td>.006</td>
<td>.003</td>
</tr>
<tr>
<td>5.00</td>
<td>.015</td>
<td>.006</td>
<td>.003</td>
<td>.001</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Figure 2 presents a plot of the three integrated risk function $R(\pi,\delta_0)$, $R(\pi,\delta_\pi)$ and $R(\pi,\delta_{\pi,1})$ versus $\alpha$.

2.5 Multivariate Considerations

In this section we discuss the extension to the problem of simultaneously estimating the parameter $\theta = (\theta_1, \ldots, \theta_k)^\top$, $k > 1$, a location parameter for the distribution of $X = (X_1, \ldots, X_k)^\top$. We take $\gamma \sigma \theta \Theta \gamma 
\sigma \theta \Theta \gamma$. We shall assume that by the various reduction arguments the problem has been transformed to canonical form wherein the covariance matrix of $X, \Sigma_{\theta}(X-\theta)^\top(X-\theta)$, is equal to the $k \times k$ identity matrix $I$. Moreover we shall make the assumption that the individual components $\theta_i$ are independent of one another, and the loss function is the sum of the squared error.

31
of the component

\[ \mathcal{L}(\theta) = |x - \theta| = \sum_{i=1}^{k} (\theta_i - \hat{\theta}_i)^2 \]

Stein's famous result [15] showed that if

\[(2.47) \quad X \sim \mathcal{N}(\theta, I) \quad k \geq 3 \]

then the uniformly most wonderful rule in lower dimensions, \( \delta_0(x) = x \), is not so wonderful for \( k > 2 \). In particular the estimator \( \frac{1}{k} \delta_0(x) = (1 - \frac{k-2}{||x||^2}) x \) dominates \( \delta_0(x) \) and for large \( k \) the improvement can be considerable. It has been shown by Brown [2], that for \( k \geq 3 \) it is the rule rather than the exception that there exist a Stein type estimator

\[(2.48) \quad \delta_0^1 = (1 - \frac{c}{||x||^2}) x \]

which does considerably better than the Pitman estimator \( \delta_0 \) in the sense that \( R(\theta, \delta_0^1) \leq R(\theta, \delta_0) \forall \theta \).

The situation is by no means clearcut however, for while in higher dimensions the above is generally true, it is not true that \( R(\theta, \delta_0^1) \leq R(\theta, \delta_0, i) \forall \theta \), where we adopt the notation \( \delta = (\delta_1, \ldots, \delta_k) \). Note that by the assumption of independence \( \delta_i \) is the component-wise estimator of \( \theta_i \).

In the multivariate problem, we are faced not only with the tradeoff between Bayes risk and the total, or ensemble risk, but also
with the interrelationship of the above to component risk. We do not propose a definitive answer to the problem of whether δ₀ or δ¹₀ would make a 'better' choice on which to base the limited translation, it is a question best answered by the motivation and underlying structure of the particular problem being dealt with. In keeping with the spirit of this paper, however, we do feel a strong preference toward the use of δ₀. The main reason for this is that playing off δ₀ against δᵢ we can compensate for the fact that δᵢ may have large component risk by the fixed component risk of δ₀. If we play δ¹₀ vs δᵢ we do not have this advantage working for us. Both δ¹₀ and δᵢ may have large component risk and compromising between them will not help the situation. The above ideas will be demonstrated in the multivariate normal problem presented later in the section.

We now give the following

**Definition 2.4** The limited translation rule of maximum translation mσ₀ based on δ₀ and prior π is denoted by \( δᵢₘ,ᵢ = (δᵢₘ,ᵢ, \cdots, δᵢₘ m, k) \) where

\[
\begin{align*}
δ₀,i + mσ₀ & \quad Xᵢ \mathcal{A} \\
δᵢᵢ & \quad Xᵢ \mathcal{B} \\
δ₀,i - mσ₀ & \quad Xᵢ \mathcal{C}
\end{align*}
\]

(2.49)
and \( \sigma_0 = \sqrt{R(\pi, \delta_{0,i})} \), \( \mathcal{A} = \{ y : \delta_{0,i}(y) - \delta_{\pi,i}(y) < -m\sigma_0 \} \)

\[ \mathcal{B} = \{ y : |\delta_{0,i}(y) - \delta_{\pi,i}(y)| \leq m\sigma_0 \} , \mathcal{C} = \{ y : \delta_{0,i}(y) - \delta_{\pi,i}(y) > m\sigma_0 \} \]

The companion theorems to Theorems 2.3 and 2.4 follow directly from the fact that \( R(\theta, \delta_{\pi,m}) = \sum_{i=1}^{k} R(\theta, \delta_{\pi,m,i}) \) and we state them below:

**Theorem 2.6** \( R(\theta, \delta_{\pi,m}) \leq k\sigma_0^2(1+m)^2 \) \( \forall \theta \)

**Theorem 2.7** If \( \delta_{\pi,m,i}(x) - \delta_{0,i}(x) \) is a monotonically non-increasing function of \( x_i \) then

\[ R(\theta, \delta_{\pi,m}) \leq k\sigma_0^2(1+m^2) \]

Letting

\[ u_i(x) = r|\delta_{\pi,i}(x) - \delta_{0,i}(x)|^2 , \quad B = \sqrt{m\sigma_0} \quad r = R(\pi, \delta_{0,i}) - R(\pi, \delta_{\pi,i}) \]

and \( \rho \) and \( \delta \rho,i \) be as defined in (2.15) and (2.16) we have

\[ l-s_i = EU_i(l-\rho(U_i))^2 = l-s_j \quad i,j = 1, \ldots k \]

We shall denote the common value of \( l-s_i \) as "L-s" and we state below the companion to Theorem 2.5

**Theorem 2.8**

\[ R(\pi, \delta \rho) = (l-s)R(\pi, \delta_0) + sR(\pi, \delta_{\pi}) \]
2.6 The Multivariate Normal Problem

We now turn to the problem of estimating the mean vector of the multivariate normal distribution. In canonical form we have:

\[(2.52) \quad \mathbf{X} \sim \mathcal{N}_k(\mathbf{\theta}, \mathbf{I})\]

\[(2.53) \quad \mathbf{\theta} \sim \mathcal{N}_k(\mathbf{\theta}, \mathbf{\Sigma}) \text{ where } \mathbf{\Sigma} \text{ is known}\]

\[(2.54) \quad L(\mathbf{\theta}, \mathbf{a}) = || \mathbf{a} - \mathbf{\theta} ||^2 = \sum_{i=1}^{k} (a_i - \theta_i)^2\]

Both the case with \(k=1\), the univariate problem, and the case with \(\sigma^2\) unknown, the empirical Bayes problem have been extensively treated in Efron and Morris [5] and [6].

Stein has shown [15] that for \(k=2\) the Pitman estimator \(\delta_0(x) = (x_1, x_2)^T\) is the admissible minimax estimate of \(\theta\) and the problem of a dominating rule \(\delta_0^1\) need not concern us. In the case of \(k \geq 3\) we shall first deal with \(\delta_{\Pi, m}\) as defined above and then comment on the use of \(\delta_0^0\).

For completeness we note that \(\delta_0(x) = (x_1, \ldots, x_k)\) with

\[(2.55) \quad R(\mathbf{\theta}, \delta_0) = k, \quad R(\mathbf{\theta}, \delta_{\Pi, i}) = 1, \quad R(\Pi, \delta_0) = k, \quad R(\Pi, \delta_{\Pi, i}) = 1\]

The fact that

\[(2.56) \quad \delta_{\Pi}(x) = \left( \frac{\sigma^2}{\sigma^2 + 1} x_1, \ldots, \frac{\sigma^2}{\sigma^2 + 1} x_k \right)\]

35
follows directly from \( \pi_{\theta/x} = \eta_k \left( \frac{\sigma^2}{\sigma^2 + 1} x, \frac{\sigma^2}{\sigma^2 + 1} \frac{1}{x} \right) \).

Note that:

\[
(2.57) \quad R(\theta, \delta_{\pi}) = k \left( \frac{\sigma^2 + |\theta|^2}{(\sigma^2 + 1)^2} \right) \quad R(\theta, \delta_{\pi, i}) = \frac{\sigma^2 + \theta_i^2}{(\sigma^2 + 1)^2}
\]

\[
(2.58) \quad R(\pi, \delta_{\pi}) = k \frac{\sigma^2}{\sigma^2 + 1} \quad R(\pi, \delta_{\pi, i}) = \frac{\sigma^2}{\sigma^2 + 1}
\]

For \( k \geq 2 \) the analysis of \( \delta_{\pi, m} \) follows directly from the univariate case. The condition \( |\delta_{0, i} - \delta_{\pi, i}| \leq m \sigma_0 \) translates to

\[
(2.59) \quad \left| x_i - \frac{\sigma^2}{\sigma^2 + 1} x_i \right| \leq m \implies |x_i| \leq m(\sigma^2 + 1)
\]

Letting \( C = m(\sigma^2 + 1) \) we have

\[
(2.60) \quad \delta_{\pi, m, i} = \begin{cases} 
  x_i + m & x_i \in \left( -\infty, -c \right) \\
  \frac{\sigma^2}{\sigma^2 + 1} x_i & x_i \in [-c, c] \\
  x_i - m & x_i \in \left( c, \infty \right)
\end{cases}
\]

From theorem 3.2 of [5] which gives the expression for \( R(\theta, \delta_{\pi, m}) \) in the univariate case and from the fact that
\[ (2.61) \quad R(\theta, \delta_{\pi, m}) = \frac{k}{\sum_{c=1}^{C} R(\theta, \delta_{\pi, m,c})} \]

we have

\[ (2.62) \quad R(\theta, \delta_{\pi, m}) = \frac{k}{\sum_{i=1}^{C} \left( \frac{1}{\sigma_i^2 + 1} \right)^2} \left( \theta_i - \frac{1}{\sigma_i^2 + 1} \right) - \frac{1}{\sigma_i^2 + 1} \left( \delta_i(x) \phi(\theta_i) \right) \]

where \( \phi \) and \( \phi \) are the standard normal distribution and density function respectively.

Writing \( \delta_{\pi, m} \) in its relevance function form we have

\[ (2.63) \quad \delta_{\pi} (x) = \left( \delta_{\pi, 1}, \ldots, \delta_{\pi, m} \right) \quad \text{where} \quad \rho_i = \min \left( 1, \frac{\sqrt{x_i^2 + 1}}{\bar{x}_i^2} \right) \]

From (2.28) and (2.29) we have that the relative savings loss for component \( i \) is just

\[ (2.64) \quad 1 - s_i = \frac{x_i^2}{\sigma_i^2 + 1} \left[ 1 - \rho \left( \frac{x_i^2}{\sigma_i^2 + 1} \right) \right]^2 = \frac{1}{\sigma_i^2 + 1} \left( 1 - \rho \left( x_3^2 \right) \right)^2 \]

This follows from the fact that \( U_i = \frac{x_i^2}{\sigma_i^2 + 1} \sim x_i^2 \) and for any function \( g \), \( E x_i^2 g(x_i^2) = Eg(x_3^2) \). This is essentially Theorem 6.1 of [5]. Now from Theorem 2.8 we have

37
(2.65) \[ R(\pi, \delta_0) = k \frac{\sigma^2}{\sigma^2 + 1} + k \frac{1}{\sigma^2 + 1} (1-s) \]

From (2.61) and (2.64) we then have

(2.66) \[ 1-s = 2[(D^2 + 1)(1-\Phi(D)) - D\Phi(D)] \quad \text{where} \quad D = \frac{m}{\sqrt{\sigma^2 + 1}} \]

the same result as for the univariate problem. The relationship between \( D \) and \( 1-s \) is available from Figure 1 and Tables 1 and 2 of [5].

As mentioned, for the case \( k \geq 3 \), James and Stein [9] showed that the estimator

(2.67) \[ \delta_0^1(x) = (1-\frac{k-2}{\|x\|^2}) x \]

although itself inadmissible, dominates \( \delta_0(x) = x \) by showing that

(2.68) \[ R(\delta, \delta_0^1) = k - (k-2) E \| \delta \|^2 \frac{k-2}{k-2+2j} < k \]

where the expectation is taken with respect to the distribution of \( J \) a Poisson random variable with parameter \( \frac{\| \delta \|^2}{2} \). The component risk of \( \delta_0^1 \) has been found to be:
\[ R(\theta_{i,1}, \delta_{0,1}) = 1 + 2 (k - 4) \frac{\theta_{i}^2}{||\theta||^2} \sum_{j=1}^{(k+2J)(k-2+2J)} \frac{1}{(k+2J)(k-2+2J)} \]

\[ - (k-2)E \frac{4J + k - 2}{(k+2J)(k-2+2J)} \]

Morris [11] shows that for large \( k \) the maximum of \( R(\theta_{i,1}, \delta_{0,1}) \) is approximately \( k/4 \) whereas we know that \( \forall \theta R(\theta_{i,1}, \delta_{0,1}) = 1 \). Since \( R(\theta_{i,1}, \delta_{0,1}) \) is maximized for fixed \( ||\theta||^2 \) at \( \theta_{i}^2 = ||\theta||^2 \), a situation in which we had a strong belief that the components of \( \theta_{i} \) were not too different would lend support to the use of \( \delta_{1,0} \). Similarly since (2.69) decreases from its maximum value near \( ||\theta||^2 = k \) to a minimum of \( 2/k \) as \( ||\theta||^2 \) shrinks to \( 0 \), the case where we believed \( ||\theta||^2 \) to be small would also lend justification to the use of \( \delta_{1,0} \).

However part of the motivating force behind this paper is a concern with the appropriateness of our prior beliefs to the specific problem we encounter, and the above arguments are not persuasive in this setting.

We note again that if in a particular problem one chooses to base the limited translation rule on \( \delta_{1,0} \) and \( \delta_{\pi} \), it is not immediately obvious by what quantity we would want to limit deviation from \( \delta_{1,0} \).

We shall adopt the convention of limiting deviations by \( m \sigma_{1,0} \)

where \( m \) is arbitrary and

\[ \sigma_{1,0} = \sqrt{R(\pi, \delta_{0,1})} = \frac{1}{\sigma_{+1}} \sqrt{\left( \frac{\sigma^2 + 2}{k} \right) (\sigma^2 + 1)} \]

39
Imposing the condition:

\[
(2.71) \quad \left| \delta_{0, i}^{l} - \delta_{\pi, i}^{l} \right| \leq \lambda_{0}^{1}
\]

\[
(2.72) \quad \implies \quad \left| \frac{1}{\sigma^{2} + 1} - \frac{k-2}{||x||^2} \right| x_{1} \leq \lambda_{0}^{1}
\]

Letting \( C = \lambda_{0}^{1} \left( \left| \frac{1}{\sigma^{2} + 1} - \frac{k-2}{||x||^2} \right| \right) \) we have

\[
\delta_{0, i}^{l} + \lambda_{0}^{1} x_{i} \epsilon(-\infty, -c)
\]

\[
\delta_{\pi, m, i}^{l} = \begin{cases} 
\delta_{\pi, i}^{l} & x_{i} \epsilon[-c, c] \\
\delta_{0, i}^{l} - \lambda_{0}^{1} x_{i} \epsilon(c, \infty)
\end{cases}
\]

2.7 Optimality Considerations

We now turn our attention to investigating the limited translation rule (2.60) in the context of an optimality criterion put forth by Hodges and Lehman [9]. Their approach to the Bayes minimax compromise was to specify an upper bound for the risk functions and then, among the class of rules whose risk functions lay below this bound, to find that rule which minimized the Bayes risk. They call such a rule a 'restricted Bayes' rule and present a sufficient condition for a procedure to be restricted Bayes. Unfortunately it is extremely difficult to obtain explicit results using their approach, although their framework does provide a good touchstone for judging the limited
translation rules.

Suppose we decide to look only at procedures $\delta_\pi$.

\begin{equation}
(2.74) \quad \sup_{\theta} R(\delta, \theta) \leq a
\end{equation}

where $a$ is an arbitrary constant which we specify (to insure that the class of such rules is non empty we choose $a$ to be not less than the minimax risk). Then if we let $\delta^a_\pi$ be the associated restricted Bayes rule, it is easy to prove, and Hodges and Lehman do so, that $R(\pi, \delta^a_\pi)$ is a convex, continuous, and strictly decreasing function of $a$. From this we also have that \( \sup_{\theta} R(\theta, \delta_\pi) = a \).

Although it is extremely difficult to obtain the precise curve of $R(\pi, \delta^a_\pi)$ as a function of $C$, Efron and Morris propose a method for bounding the curve from below.

Let

\begin{equation}
(2.75) \quad a(\delta) = \sup_{\theta} R(\delta, \delta) \quad B(\delta) = R(\pi, \delta)
\end{equation}

\begin{equation}
(2.76) \quad B_a = \inf_{\{\delta: a(\delta) \leq a\}} B(\delta) = \inf_{\{\delta: a(\delta) = a\}} B(\delta) = R(\pi, \delta^a_\pi)
\end{equation}

If we let $\mu$ be an arbitrary probability measure on $\Theta$, $\mu(0,1)$, we can form the mixture $\lambda = (1-w)\pi + w\mu$. We denote the Bayes rule for this prior as $\delta^a_\lambda$. Now for any rule $\delta$:
\[(2.77) \quad (1-w)B(\delta) + w\alpha(\delta) \geq (1-w)B(\delta) + wR(\mu, \delta) \]
\[= R(\lambda, \delta) \]
\[\geq R(\lambda, \delta_\lambda) \]
\[(2.78) \quad = (1-w)B(\delta_\lambda) + wR(\mu, \delta_\lambda) \]
\[(2.79) \quad \Rightarrow B(\delta) \geq B(\delta_\lambda) + \frac{w}{(1-w)} [R(\mu, \delta_\lambda) - \alpha(\delta)] \]

Table 2 of [4] presents lower bounds for the Hodges Lehman figure of the univariate normal problem at several values of \(a = l+m^2\) and for several values of \(\sigma^2\). Table 3 below follows straightforwardly from (2.45) and section 4 of the above paper. Letting \(\mu\) be the mixture of a shifted multivariate exponential density

\[
\sum_{i=1}^{k} B_{x_i - b} \quad \theta_i \geq b \quad i = 1, \ldots, k \quad \text{and a distribution which concentrates its mass at the vertices of a } k\text{-dimensional cube, we have the lower bounds of Table 3 for various values of } a = k(l+m^2) \text{ and } \sigma^2.
\]
<table>
<thead>
<tr>
<th>Relative Savings Loss</th>
<th>$\sigma^2 = .5$</th>
<th>$\sigma^2 = 1$</th>
<th>$\sigma^2 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l-s$ km $m$ m $R(\pi, \delta m)$ Lower Bound</td>
<td>$m$ m $R(\pi, \delta m)$ Lower Bound</td>
<td>$m$ m $R(\pi, \delta m)$ Lower Bound</td>
<td></td>
</tr>
<tr>
<td>.1 .963 .400k .388k .834 .550k .545k .681 .700k .698k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.25 .623 .500k .483k .540 .625k .618k .441 .750k .750k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.40 .429 .600k .583k .371 .700k .696k .525 .800k .800k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.50 .331 .667k .654k .286 .750k .750k .234 .833k .833k</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Risk of Limited Translation Rules Compared to Computed Lower Bound
Chapter III

The Poisson Problem

3.1 Formulation

We turn now to a family of discrete random variables, the Poisson family. It does not fit into either of the two main classifications of this paper, location and scale parameter families, but yet is interesting in its own right. It poses a challenge to the estimation procedures put forth in this paper in that $\sigma_0$ is a function of the unknown parameter we seek to estimate, as we shall see shortly aside from all of the above, it was the first example we tackled and it has a fondness in our hearts.

Suppose $X$ is a random variable, whose distribution conditional upon the random variable $\theta$, is Poisson:

$$X \sim P_0(\theta) \text{ with mass function } f_\theta(x) = \frac{e^{-\theta} \theta^x}{x!} \quad x = 0, 1, 2, \ldots$$

We are interested in estimating the parameter of the process $\theta$, based on a random sample of size $n$ from $P_0(\theta)$. For our loss function we take

$$L^*(a, \theta) = n \frac{(a - \theta)^2}{\theta}.$$ 

Since $\sigma^2(x) = \theta$ (3.12) can be thought of as a normalized
relative error loss function for the problem.

Since the Poisson distribution is closed under convolutions we have that the sample sum, $n\bar{x}$, is distributed as a Poisson random variable with parameter $n\theta$, i.e. $n\bar{x} \sim P_0(n\theta)$

\[
(3.3) \quad f_{n\theta}(n\bar{x}) = \frac{e^{-n\theta}(n\theta)^{n\bar{x}}}{n^{n\bar{x}}} \quad n\bar{x} = 0, 1, 2, ...
\]

With simplicity in mind, and without loss of generality, we can reduce the problem to canonical form by the following transformations

\[
(3.4) \quad t = n\bar{x} \\
\theta = n\theta \\
a_\theta = na
\]

In this context we have:

\[
(3.5) \quad t \sim e^{-\theta} \frac{e^{\theta t}}{t!} \quad t = 0, 1, 2, ...
\]

and

\[
\text{(3.6)} \quad L^*(\theta, a) = \frac{n(a-\theta)^2}{\theta} = \frac{(na-n\theta)^2}{n\theta} = \frac{(a-\theta)^2}{\theta} = L(\theta, a)
\]

The prior distribution most commonly assumed for $\theta$, and the one we shall work with, is the gamma prior:

\[
(3.7) \quad \pi(\theta) = \frac{\theta^{a-1} e^{-\theta/B}}{\Gamma(a)B^{a-1}} \quad \text{where} \quad \theta > 0, \ a > 1, B > 0
\]

For computational convenience (as will be seen below), and without
any essential loss of generality, we restrict ourselves to values of 
a greater than or equal to one.

The joint distribution of \( t \) and \( \theta \) is:

\[
(3.8) \quad h(t, \theta) = \frac{\theta^{t+\alpha-1} e^{-\theta(1+B)/B}}{\Gamma(t+1) \Gamma(\alpha)B^\alpha}
\]

and the marginal distribution of \( t \) is negative binomial with 
parameters \( \alpha, \frac{B}{1+B} \):

\[
(3.9) \quad f(t) = \int h(t, \theta) \, d\theta = \frac{\Gamma(t+\alpha)}{\Gamma(t+1) \Gamma(\alpha) \left( \frac{1}{1+B} \right)^\alpha \left( \frac{B}{1+B} \right)^t} \quad t = 0, 1, 2, \ldots
\]

Combining the above we have the conditional distribution of the 
parameter given the sample is:

\[
(3.10) \quad \pi_{\theta|t}(\theta) = \frac{\theta^{t+\alpha-1} e^{-\theta(1+B)/B}}{\Gamma(t+\alpha) \frac{B}{1+B}^t} \quad t + \alpha - 1, \frac{B}{1+B}
\]
a gamma distribution with parameters \( t + \alpha - 1, \frac{B}{1+B} \)

Now with loss function of the form \((3.6)\) the Bayes rule, which 
we denote by \( \delta_\pi(t) \) is merely:

\[
(3.11) \quad \delta_\pi(t) = \frac{1}{E_{\theta|t}(\theta)}
\]
provided the expectation exists. From the above we see that for the present case

\[(3.12) \quad \delta_n(t) = (t+\alpha-1) \frac{B}{1+B} \quad t = 0,1,2,\ldots\]

Note that

\[(3.13) \quad (t+\alpha-1) \frac{B}{1+B} = t \frac{B}{1+B} + (\alpha-1)B \frac{1}{1+B}\]

that is, the Bayes rule is a linear combination of the sample sum and the mode of the a priori distribution.

If \( \alpha < 1 \), then \( E_{\theta/t=0} \frac{1}{\theta} \) is infinite and the Bayes rule is:

\[(3.14) \quad \delta_n(t) = (1-\delta_{0t}) (t+\alpha-1) \frac{B}{1+B} \quad t = 0,1,2,\ldots\]

\[(3.15) \quad \text{where} \quad \delta_{0t} = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}\]

This introduces an unessential complication in the subsequent calculations and as mentioned above we restrict ourselves to priors \( \alpha > 1 \).

From 3.1.5, 3.1.6, and 3.1.12 we find that
(3.16) \( R(\theta, \delta, \pi) = \frac{1}{\theta} \sum_{t=0}^{\infty} \frac{(t+\alpha-1)B - \theta)^2}{1+B} e^{-\theta t} \frac{e^{-\theta t}}{t!} \) when \( t = 0, 1, \ldots \)

\[ = \frac{1}{(1+B)^2} \sum_{t=0}^{\infty} \frac{(B(t-\theta)+(\alpha-1)B-\theta)^2}{\theta} e^{-\theta t} \frac{e^{-\theta t}}{t!} \]

(3.17) \[ = \frac{1}{(1+B)^2} \left[ B^2 + \frac{(\alpha-1)B-\theta)^2}{\theta} \right] \]

Exploring the behavior of \( R(\theta, \delta, \pi) \) we see that for \( 0 < \theta \leq (\alpha-1)B \)
\( R(\theta, \delta, \pi) \) is a decreasing function of \( \theta \) with minimum value \( (\frac{B}{1+B})^2 \) achieved at \( \theta = (\alpha-1)B \), while for \( (\alpha-1)B < \theta < \infty \) \( R(\theta, \delta, \pi) \)
increases with \( \theta \) and tends to \( \infty \) as \( \theta \to \infty \).

\( R(\pi, \delta, \pi) \) is found in a straightforward manner.

(3.18) \( R(\pi, \delta, \pi) = \int_0^\infty R(\theta, \delta, \pi) \pi(\theta) d\theta = \frac{1}{(1+B)^2} \left[ B^2 + \frac{(\alpha-1)B-\theta)^2}{\theta} \right] \frac{\alpha}{\Gamma(\alpha)B^\alpha} \]

\[ = \frac{1}{(1+B)^2} \left[ B^2 + \frac{(\alpha-1)B^2}{(\alpha-1)B} \right] \]

\[ = \frac{B}{1+B} \]

Now if we take a sequence of gamma priors

\( \{\pi_n\}_{n=1}^\infty \) where \( \pi_n = g(\alpha, B_n) \) and \( B_n \to \infty \)
the sequence of corresponding Bayes rules $\delta_n$ have by (3.18) Bayes risks tending to 1. The maximum likelihood estimator of $\theta, \delta_0(t) = t$, has constant risk equal to 1. The standard argument, presented below, shows that $\delta_0(t)$ is minimax.

Letting $\delta(t)$ be an estimate of $\theta$ we have

$$\sup_{\theta} R(\theta, \delta) \geq \int_{0}^{\infty} R(\theta, \delta)d\pi_n(\theta)$$

$$\geq R(\pi_n, \delta_n) \quad \forall n$$

where $\delta_n$ is the Bayes rule for $\pi_n$. Since $R(\pi_n, \delta_n) \to 1$ we have that

$$(3.19) \quad \sup_{\theta} R(\theta, \delta) \geq 1 = \sup_{\theta} R(\theta, \delta_0)$$

For the sake of future reference we state the obvious fact that:

$$(3.20) \quad R(\pi, \delta_0) = 1$$

3.2 Limited Translation Rule and its Properties

Attempting to compromise between the Bayes and minimax rules we turn to the limited deviation rules. The Poisson case presents an interesting situation in that the expected value and variance of $T$ both equal $\theta$. The essence of our scheme in the previous chapter was to limit the deviation of our estimator from $\delta_0$ by some
fixed number of standard deviations, \( m\sigma_0 \), of the minimax rule. In
the situation at hand we should be willing to tolerate greater
deviations from \( \delta_0(t) \) for \( t \) large than for \( t \) small, since the
larger \( t \) is the larger we believe \( \theta \) to be; hence the larger the
standard deviation.

As an estimate of \( \sigma_0 = \theta^{1/2} \) we propose using the maximum
likelihood estimate \( \hat{s}^{1/2} \). The fact that it is a biased estimator
of \( \theta^{1/2} \) is not in itself damaging since there does not exist any unbiased
estimator of \( \theta^{1/2} \). Moreover it somehow seems appealing to use the maximum
likelihood estimator as an arbiter in a case we are playing minimax and Bayes
off against each other. With \( L(\theta, a) = \frac{(a-\theta)^2}{\theta} \) the situation is not neutral however
in that \( \delta_0(t) = t \) aside from being the minimax rule, is also the
maximum likelihood estimator.

As before, we shall estimate \( \theta \) to be \( \delta_\pi(t) \) when the following
condition holds:

\[
(3.21) \quad |\delta_0(t) - \delta_\pi(t)| \leq m\sigma_0
\]

With the above setting this is just:

\[
|t - (t+a-1) \frac{B}{1+B}| \leq mt^{1/2}
\]

(3.22) \quad \quad \quad \quad |t - (a-1)B| \leq m(1+B)t^{1/2}
Letting

\[ (3.23) \quad h = (\alpha - 1)B \quad g = m(1+B) \quad \text{we have} \]

\[ (3.24) \quad h - gt \frac{h}{4} \leq t \leq h + gt \frac{h}{4} \]

and the right hand side reduces to

\[ (3.25) \quad c_1 \leq t \leq c_2 \]

\[ (3.26) \quad c_1 = h + \frac{h^2}{2} - \sqrt{h + \frac{h^2}{4} g} \quad c_2 = h + \frac{h^2}{2} + \sqrt{h + \frac{h^2}{4} g} \]

Note that \( c_1 > 0 \). From the above we have that:

\[ (3.27) \quad \delta_{\pi, m}(t) = \begin{cases} 
  t + mt \frac{h}{4} & 0 \leq t < c_1 \\
  (t + \alpha - 1) \frac{B}{1+B} & c_1 \leq t \leq c_2 \\
  t - mt \frac{h}{4} & t > c_2 
\end{cases} \]

The risk function of the limited translation rule is:

\[ (3.28) \quad R(\theta, \delta_{\pi, m}) = \frac{1}{\theta} \left[ \sum_{t=0}^{\infty} (t + mt \frac{h}{4}) f_\theta(t) \right] + \left[ c_2 \right] \left( (t + \alpha - 1) \frac{B}{1+B} - \theta \right)^2 f_\theta(t) \]

\[ + \sum_{t=\lceil c_2 \rceil + 1}^{\infty} (t - mt \frac{h}{4} - \theta)^2 f_\theta(t) \]
In the following theorem we provide a sharp bound for the risk functions of $\delta_{\pi,m}$.

**Theorem 3.1**

With (3.5), (3.6), (3.7) and $\delta_{\pi,m}$ as defined in (3.27) we have:

\[(3.29) \quad R(\theta, \delta_{\pi,m}) \leq (1+m)^2\]

We have

\[(3.30) \quad R(\theta, \delta_{\pi,m}) = E_{\theta} \left( \frac{\delta_{\pi,m} - \delta_0}{\delta} \right)^2 = E_{\theta} \left( \frac{\delta_{\pi,m} - \delta_0 + \delta - \delta_0}{\delta} \right)^2\]

\[= E_{\theta} \left( \frac{\delta_{\pi,m} - \delta_0}{\delta} \right)^2 + 2E_{\theta} \left( \frac{\delta_{\pi,m} - \delta_0}{\delta} \right) \left( \frac{\delta - \delta_0}{\delta} \right) + E_{\theta} \left( \frac{\delta - \delta_0}{\delta} \right)^2\]

\[(3.31) \quad \therefore \quad R(\theta, \delta_{\pi,m}) \leq (1+m)^2\]

(3.31) following from the fact that $|\delta_{\pi,m} - \delta_0| \leq mt^{1/2}$ and Hoelder's Inequality. \|

We now show that when $c_l > 1$, $(1+m)^2$ is a sharp bound. We have

\[(3.32) \quad R(\theta, \delta_{\pi,m}) = \frac{1}{\theta} \{ (\theta^2 e^{-\theta} + (1+m-\theta)^2 e^{-\theta} + \theta(\theta) \}

where the first two terms follow from the first term of (3.28)
Rewriting (3.30):

\[ R(\theta, \delta_{\pi, m}) = \frac{1}{\theta} \{(1+m)^2\theta e^{-\theta} + \theta(\theta)\} \]

\[ \lim_{\theta \to 0} R(\theta, \delta_{\pi, m}) = (1+m)^2 \]

In Figure 3 below we present plots of the risk functions of the minimax, limited translation, and Bayes rules for the case \( \alpha = 2, B = 1, m = 1 \).

Examining \( \delta_{\pi, m} \) from the relevance function approach the limited translation rule \( \delta_{\pi, m} \) can be expressed as:

\[ (3.33) \quad \delta p(t) = [1 - \rho(U)] \delta_0(t) + \rho(U) \delta_\pi(t) \]

where

\[ U = \frac{(t-h)^2}{1+B} \]

\[ (3.34) \]

\[ (3.35) \quad \rho(U) = \begin{cases} 
1 & \text{if} \quad U \leq (1+B)M^2t \\
\frac{m(1+B)t}{|t-h|} & \text{if} \quad U > (1+B)M^2t 
\end{cases} \]

The argument \( \frac{(t-h)^2}{1+B} \) is the distance between the observed value \( t \) and the mode of the prior distribution \( h = (\alpha-1)B \) in units of \( (1+B) \).

Note that since \( L(\theta, a) = \frac{(a-\theta)^2}{\theta} \) Theorem 2.3 is not directly applicable, but the same ideas as used to prove it can be incorporated in the proof of the following
Theorem 3.2

(3.36) \[ R(\pi, \delta_\rho) = (1-s)R(\pi, \delta_0) + sR(\pi, \delta_\pi) \]

(3.37) \[ 1 - s = \mathbb{E} \left\{ \frac{\left( \frac{t-h}{t+\alpha-1} \right)_B (1-\rho(\frac{t-h}{1+B}))}{(t+\alpha-1)_B} \right\}^2 \]

and the expectation is taken with respect to the marginal distribution of \( t \)

Proof

(3.38) \[ R(\Pi, \delta_\rho) = \mathbb{E} \frac{(\theta - \delta_\rho)^2}{\theta} \]

\[ = \mathbb{E} \frac{1}{\theta} \left\{ \theta - [(1-\rho(.))t + \rho(.) \left( \frac{tB+h}{1+B} \right)] \right\}^2 \]

\[ = \mathbb{E} \frac{1}{\theta} \left\{ \theta - \frac{t+h}{1+B} \right\}^2 \]

\[ = \mathbb{E} \frac{1}{\theta} \left\{ \theta - \left( \frac{t+h}{1+B} \right) (1-\rho(.)) \right\}^2 \]

\[ = \mathbb{E} \frac{1}{\theta} \left\{ \theta - \left( \frac{t+\alpha-1}B \right) \left( \frac{t-h}{1+B} \right) \right\}^2 \]

\[ = \frac{B}{1+B} + \frac{1}{1+B} \mathbb{E} \frac{1}{\theta} \left\{ \frac{t-h}{1+B} (1-\rho(.)) \right\}^2 \]

(3.39) \[ = \frac{B}{1+B} + \frac{1}{1+B} \mathbb{E} \frac{(t-h)}{(t+\alpha-1)_B} (1-\rho(.))^2 \]

(3.40) \[ \mathbb{E} \frac{(t-h)}{(t+\alpha-1)_B} (1-\rho(.))^2 + \frac{B}{1+B} [1-\mathbb{E} \frac{(t-h)}{(t+\alpha-1)_B} (1-\rho(.))^2] \]
3.3 Optimality Considerations

We now turn our attention to how well the limited translation rule performs with respect to the computed lower bounds for the corresponding Hodges-Lehman figure. Table 4 below presents, for two corresponding Bayes risk of the limited translation rule, \( R(\pi, \delta_{\pi,m}^{L}) \), compared with lower bounds obtained for \( R(\pi, \delta_{\pi}^{B}) \). For the cases presented \( \delta_{\pi,m}^{L}(t) - \delta_{0}(t) \) is nonincreasing in \( t \), and from (3.30) and the proof of Theorem 2.4 we have \( \sup_{\theta} R(\theta, \delta_{\pi,m}^{L}) = 1 + m^2 \), and we take \( a = 1 + m^2 \).

Initially in generating the lower bounds \( m(\theta) \) was chosen to be of the form:

\[
(3.41) \quad \mu(\theta) = r\delta(\theta) + (1-r)\gamma(\theta) \quad 0 < r < 1
\]

where

\[
(3.42) \quad \delta(\theta) = Ke^{-2m}\theta \quad \theta > \theta_0
\]

\[
(3.43) \quad \gamma(\theta) = \frac{1}{\epsilon} \quad 0 < \theta \leq \epsilon
\]

The choice of (3.42) was based on choosing a prior such that asymptotically (for large \( t \))

\[
(3.44) \quad \delta_{\pi}^{L}(t) \approx \delta_{\pi,m}^{L}(t) = t - mt^{1/2}.
\]

A discussion with Charles Stein resulted in the suggestion of solving
\[(3.45) \quad \delta_{\psi}(t) = \frac{\int_{\theta}^{\infty} e^{-(\theta + \psi(\theta))} d\theta}{\int_{\theta}^{\infty} e^{-(\theta + \psi(\theta))} d\theta} \approx t - mt^2 \]

by Laplace's method. This gives rise to \((3.44)\).

The particular form of \(\gamma(\theta)\) is motivated by the desire to have \(\delta_{\lambda}(0) = 0\), since if this is not the case \(R(\lambda, \delta_{\lambda}) \to \infty\) as \(\theta \to 0\). By choosing \(\varepsilon\) sufficiently small the role of \(\gamma(\theta)\), except for the situation \(t = 0\), becomes negligible.

**TABLE 4**

\[
\begin{array}{cccccccc}
\alpha = 2, \ B = 1 & & \alpha = 1, \ B = 2 \\
\hline
m & 1-s & R(\pi, \delta_{\pi,m}) & \text{Lower Bound} & m & 1-s & R(\pi, \delta_{\pi,m}) & \text{Lower Bound} \\
1.00 & .136 & .568 & .568 & 1.00 & .003 & .668 & .665 \\
.75 & .170 & .585 & .585 & .75 & .018 & .673 & .666 \\
\end{array}
\]
Chapter IV
The Scale Parameter Problem

4.1 Formulation

In this chapter we examine the problem of estimating a scale parameter. We have the following.

Definition 4.1 Let $X$ be a random variable with distribution function $F_\theta(x)$ is solely a function of $x/\theta$ where $\theta$ is positive, then we shall call $\theta$ a scale parameter.

Note that the definition implies $F_\theta(x) = F_{1/\theta}(x/\theta)$. If a density exists then the appropriate condition is $f_\theta(x) = \frac{1}{\theta} f_{1/\theta}(x/\theta)$.

The general scale parameter problem can be thought of as arising in the following manner. We have observations on some random variable $Z$ generated from an underlying distribution $F_{1}(z)$ which is completely specified. What we observe however is not realizations of $Z$, but rather of the random variable $X$ where $X = \theta Z$ and $\theta$ is some unobservable quantity. Based on our observations of $X$ we wish to estimate the unknown parameter $\theta$.

We shall make the usual assumption of requiring that the loss function be a function of $a/\theta$ only, i.e. $L(\theta,a) = L(a/\theta)$. Although the specific choice of loss function will obviously depend upon the underlying structure and motivation of the particular problem, the two most widely used loss functions for the scale parameter problem seem to be relative error loss, $L(\theta,a) = (a/\theta - 1)^2 = \frac{(a-\theta)^2}{\theta^2}$ and log
squared error loss, \( L(\theta, a) = (\log (a/\theta))^2 = (\log \theta - \log a)^2 \)

Note that \( a/\theta - 1 \) is the first term of the Taylor series expansion of \( \log a/\theta \) about 1 so for relatively small errors, the loss functions are in reasonable agreement.

For simplicity we will assume that \( X \) assumes only positive values. The assumption is not restrictive in that we can either focus our attention on a sufficient statistic which is positive valued, or if that fails, deal with the induced distribution of \( Y = |X| \) for which \( \theta \) will still be a scale parameter. If we make a log transformation both on the data and the parameter, i.e.

\[
(4.1) \quad X' = \log X \quad \theta' = \log \theta
\]

then

\[
(4.2) \quad f_{\theta}(x) = \frac{1}{\theta} f(x/\theta) \Rightarrow f_{\theta'}(x') = e^{x'-\theta'} f(e^{x'-\theta'}) \quad \text{and}
\]

we transform the problem of estimating a scale parameter into that of estimating a location parameter. Now since \( L(\theta, a) = L(a/\theta) \) we have

\[
(4.3) \quad R(\theta, \delta(x)) = \int_{X} L(\delta(x)/\theta) \frac{1}{\Pi} \frac{1}{\theta} f(x/\theta) dx_i
\]

\[
(4.4) \quad = \int_{X} \frac{d(e^{x_i}/\theta)}{e^{-\theta'}} \prod_{i=1}^{\eta} x_i' - \theta' f(e^{x_i'-\theta'}) dx_i
\]
If, for the problem of estimating the translation parameter $\theta^-$ of the distribution of $X'$, we let the loss function be of the form

$$(4.5) \quad L'(\theta^-, \hat{a}^-) = L(e^{\hat{a}^- - \theta^-})$$

we have

$$(4.6) \quad R'(\theta, \delta^- (x^-)) = \int_{\mathbb{R}} L(e^{\delta^- (x^-) - \theta^-}) e^{-\theta^-} f(e^{x^- - \theta^-}) \, dx^-$$

On choosing $\delta^- (x^-) = \log \delta (x) = \log \delta (e^{x^-})$ we see that the two estimation problems are identical, i.e. $R(\theta, \delta (x)) = R'(\theta, \delta^- (x^-))$. For example if $\log \delta_0 (x)$ is the minimax estimator of $\log \theta$ then $\delta_0 (x)$ is the minimax estimator of $\theta$.

4.2 Limited Scale Rules - Development

In the transformed setting the problem is the same as that studied in Chapter 1. For convenience we shall take the loss function to be transformed loss function of Chapter 1, log squared error loss,

$L(\theta, \hat{a}) = (\log \hat{a} - \log \theta)^2$. Following the scheme proposed for the estimation of a location parameter, we will constrain our estimate of $\log \theta$ to be within $\pm m \sigma_0$ standard deviations of the minimax estimate of $\log \theta$, that is

$$(4.7) \quad \log \delta (x) \in [\log \delta_0 (x) - m \sigma_0, \log \delta_0 (x) + m \sigma_0]$$

where $\sigma_0 = R'(\log \theta, \log \delta_0 (x)) = R(\theta, \delta_0 (x))$. Reinterpreting this for the original problem of estimating $\theta$ we have that:

59
\[(4.8)\]
\[\delta(x) \in [\delta_0(x)e^{-m\sigma_0}, \delta_0(x)e^{m\sigma_0}] \iff e^{-m\sigma_0} \leq \delta_0(x) \leq e^{m\sigma_0}\]

and thus the appropriate deviations to be considered here are scale deviations. With this in mind we define the limited scale rule for the estimation of \(\theta\) with prior \(\pi\) and maximum allowable scaling \(M_\sigma\) to be

\[(4.9)\]
\[\delta_{\pi,m} = \begin{cases} 
\delta_0(x)e^{m\sigma_0} & x \in A^- \\
\delta_0(x) & x \in B^- \\
\delta_0(x)e^{-m\sigma_0} & x \in C^- 
\end{cases}\]

where \(A^- = \{x: \frac{\delta_0(x)}{\delta_\pi(x)} < e^{-m\sigma_0}\}\), \(B^- = \{x: e^{-m\sigma_0} \leq \frac{\delta_0(x)}{\delta_\pi(x)} \leq e^{m\sigma_0}\}\), and \(C^- = \{x: \frac{\delta_0(x)}{\delta_\pi(x)} > e^{m\sigma_0}\}\).

We need not be concerned about negative estimates, since the Bayes rule for a positive parameter will, for any same choice of prior, be positive and since we can always choose the truncated form of a minimax estimator which gives rise to negative estimates. Comments concerning the form of the sets \(A^-, B^-, C^-\), are similar to those of section 2.2. In particular if \(\delta_0(t), \delta_\pi(t)\) are of the form \(\delta_0(t) = at^a, \delta_\pi(t) = bt^B\) where \(t\) is the sufficient statistic, then the induced sets \(A, B, C\) will be intervals of the positive real line.

We now turn to a familiar problem in an unfamiliar setting.
Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta(x)$, with

\begin{equation}
(4.10) \quad f_\theta(x) = \frac{\theta}{x^2} \quad 1 < \theta < x < \infty
\end{equation}

\begin{equation}
(4.11) \quad \pi(\theta) = \frac{(\log \theta)^{\alpha-1}}{r(\alpha)\theta^2} \quad 1 < \theta < \infty
\end{equation}

\begin{equation}
(4.12) \quad L(\theta, \alpha) = n^{2}(\log \alpha - \log \theta)^2
\end{equation}

From the definition it is clear that $\theta$ is a scale parameter for $F_\theta(x)$. If we make the transformation $\theta' = \log \theta$, $X' = \log X$, we recognize an old friend. Performing the transformation we have

$X_1', \ldots, X_n' \overset{iid}{\sim} F_\theta'(x')$

\begin{equation}
(4.13) \quad f_\theta'(x') = e^{-(x' - \theta')} \quad 0 < \theta' < x' < \infty
\end{equation}

\begin{equation}
(4.14) \quad \pi(\theta') = \frac{\theta^{-\alpha-1}e^{-\theta'}}{r(\alpha)} \quad 0 < \theta' < \infty
\end{equation}

\begin{equation}
(4.15) \quad L(\theta', \alpha') = n^2(\alpha' - \theta')^2
\end{equation}

From our previous results we know that by the transformations $t' = nX'(1)$, $\theta' = n\theta$, $\alpha' = n\alpha'$, we can reduce the problem to the form
\[(4.16) \quad f_{\theta'}(t^-) = e^{-(t^- - \theta^-)} \quad 0 < \theta < t\]

\[(4.17) \quad \Pi(\theta^-) = \frac{\theta^- x - 1}{\pi(\alpha)} e^{-\theta^-} \quad 0 < \theta\]

\[(4.18) \quad L(\theta^- a^-) = (a^- - \theta^-)^2\]

From section (1.2) we have that \(\delta_0(t^-) = t^- - 1\)

\(\delta_n(t^-) = \frac{\alpha}{\alpha + 1} t^-\) so the minimax and Bayes estimators for the scale parameter problem are

\[(4.19) \quad \delta_0(t) = t e^{-1}, \quad \delta_n(t) = t \frac{\alpha}{\alpha + 1}\]

Since \(\sigma_0 = 1\), we have

\[(4.20) \quad \delta_{\pi,m}(t) = \begin{cases} 
  t e^{-(l-m)} & t < e^{(l-m)(\alpha + 1)} \\
  t \frac{\alpha}{\alpha + 1} & e^{(l-m)(\alpha + 1)} \leq t \leq e^{(l+m)(\alpha + 1)} \\
  t e^{-(l+m)} & t > e^{(l+m)(\alpha + 1)} 
\end{cases}\]

Again note that if we take \(m \geq 1\) the set \(\{t : t < e^{(l-m)(\alpha + 1)}\}\) is a set of probability 0 since \(P[t \geq l] = 1\) \(\forall \theta \in \Theta\). The previous arguments pertaining to the use of the invariant minimax estimator \(\delta_0(t^-) = t^- - 1\) vs the truncated form of the estimator also apply.
to the choice of \( \delta_0^-(t) = \frac{1}{te^{-1}} \) vs \( \delta_0^+(t) = \max (te^{-1}, 1) \).

### 4.3 Limited Scale Rules - Properties

With little additional effort we can exploit the relationship of the scale and location parameter problems to prove corresponding theorems to those of Chapter 1. In particular we can show that the risk function of the limited scale rule is bounded by \( \sigma_0^2 (1+m)^2 \).

**Theorem 4.1** Let \( \delta_{\pi, m} \) be as defined in (4.9), \( L(\theta, \alpha) = (\log \alpha - \log \theta)^2 \) then:

\[
R(\theta, \delta_{\pi, m}) \leq \sigma_0^2 (1+m)^2
\]

(4.21) \( R(\theta, \delta_{\pi, m}) = E_0 (\log \delta_{\pi, m}(x) - \log \theta)^2 \)

(4.22) \[
= E_0 (\log \delta_{\pi, m}(x) \delta_0(x) - \log \theta)^2
\]

\[
= E_0 (\log \delta_{\pi, m}(x) - \log \delta_0(x) + \log \delta_0(x) - \log \theta)^2
\]

(4.23) \[
= E_0 (\delta_{\pi, m}(x) - \delta_0(x') + \delta_0(x) - \theta')^2
\]

and the proof follows exactly as that of Theorem 2.3.//

Similarly, again using the fact that the correlation of an increasing and decreasing function is negative, we shall state without proof Corollary 4.1. If \( \hat{\delta}(t) \equiv \frac{\delta_{\pi, m}(t)}{\delta_0(t)} \) is a non-increasing
of the sufficient statistic $t$, then:

\[(4.24) \quad R(\theta, \delta_{\pi,m}) \leq \sigma_0^2 (1+m^2)\]

Suppose we let $\rho$ be some function which we shall presently specify and $\delta_0$ the estimation rule.

\[(4.25) \quad \delta_0 = \frac{\delta_{\pi}(x)}{\delta_{0}(x)} \delta_0(x)\]

Substituting $\delta_{\pi,m}$ for $\delta_0$ and taking the logarithm of both sides we have

\[(4.26) \quad \rho(u) = \min \left(1, \frac{B}{\sqrt{u}}\right)\]

where $u(x) = \left| \frac{\log \delta_{\pi}(x) - \log \delta_{0}(x)}{r}\right|^2, \quad r = R(\pi, \delta_{\pi}) - R(\pi, \delta_{0})$

and $B = \sqrt{r} M\sigma_0$.

In general letting $u$ be as defined above, $\rho(u)$ a function taking values in $[0,1]$, and

\[(4.27) \quad \delta_{\rho}(x) = [\delta_{\pi}(x)]^\rho [\delta_{0}(x)]^{1-\rho}\]

we define $\rho(u)$ as the relevance function of the rule $\delta_{\rho}$.

Since it shall be expedient below we note that with $\rho(u)$ as specified in (4.26) we have $\rho(u) \equiv \rho(u(x)) = \rho'(u(x')) \equiv \rho'(u)$ where
\begin{align}
(4.28) \quad \rho^-(u) &= \min \left( 1, \frac{\text{rm} \sigma_0}{\delta_0^\pi(x^-) - \delta_0^\upsilon(x^-)} \right)
\end{align}

Now the companion to Theorem 2.5 follows by the same principle as have the other corresponding theorems.

\textbf{Theorem 4.2}

Let
\begin{align}
(4.29) \quad 1 - s &= E\left[ \frac{(\delta_0^\pi - \delta_0^\upsilon)}{r} (1 - \rho^-) \right]^2 = E \left( U - \rho^-(U) \right)^2
\end{align}

where the expectation is taken over the marginal distribution of \( X' \).

Then
\begin{align}
(4.30) \quad R(\pi, \delta_\rho) &= (1 - s) R(\pi, \delta_0^\upsilon) + s R(\pi, \delta_\pi^\upsilon)
\end{align}

[Note that \( 1 - s = \frac{R(\pi, \delta_0^\upsilon) - R(\pi, \delta_\pi^\upsilon)}{R(\pi, \delta_0^\upsilon) - R(\pi, \delta_\pi^\upsilon)} \)]

\textbf{Proof}

\begin{align}
R(\pi, \delta_\rho) &= E_{\theta /X} [\log \theta - \{(1 - \rho) \log \delta_0^\pi(x^-) + \rho(x) \log \delta_\pi^\pi(x^-)\}]^2 \\
&= E_{\theta^- /X^-} [\theta^- - \{(1 - \rho^-) \delta_0^\upsilon + \rho^- \delta_\pi^\upsilon\}]^2
\end{align}

and the proof follows from that of Theorem 2.5. //
FIGURE 1. RISK FUNCTIONS OF MINIMAX, BAYES, AND LIMITED TRANSLATION RULE (\(\alpha=1, m=1\))

FIGURE 2. BAYES RISK OF MINIMAX, BAYES, AND LIMITED TRANSLATION RULE AS A FUNCTION OF \(\alpha\) (\(m=1\))
Figure 3. Risk functions of minimax, Bayes, and limited translation rules ($\alpha=2$, $\beta=1$, $m=1$)
REFERENCES


