On the behavior of the posterior for an extreme observation in a multivariate setting

Dale Edward Umbach

Iowa State University
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
UMBACH, Dale Edward, 1950-
ON THE BEHAVIOR OF THE POSTERIOR FOR AN
EXTREME OBSERVATION IN A MULTIVARIATE
SETTING.

Iowa State University, Ph.D., 1976
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106
On the behavior of the posterior for an extreme observation in a multivariate setting

by

Dale Edward Umbach

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1976
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. EXPONENTIAL FAMILY</td>
<td>8</td>
</tr>
<tr>
<td>A. Main Results</td>
<td>8</td>
</tr>
<tr>
<td>B. Applications</td>
<td>17</td>
</tr>
<tr>
<td>III. LOCATION VECTOR</td>
<td>21</td>
</tr>
<tr>
<td>A. Main Results</td>
<td>21</td>
</tr>
<tr>
<td>B. Discussion</td>
<td>40</td>
</tr>
<tr>
<td>IV. VARYING THE LOSS FUNCTION</td>
<td>47</td>
</tr>
<tr>
<td>A. Main Results</td>
<td>47</td>
</tr>
<tr>
<td>B. Discussion</td>
<td>59</td>
</tr>
<tr>
<td>V. BIBLIOGRAPHY</td>
<td>63</td>
</tr>
<tr>
<td>VI. ACKNOWLEDGEMENTS</td>
<td>64</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

If a single observation is taken from a particular distribution, we in general expect this observation to be close to the mean of the distribution. Indeed, for a normal population the probability that the observation is not within three standard deviations of the mean is approximately .003. Thus, if the population mean is unknown, it is reasonable to suppose that it is close to the observation. This is certainly the basic idea behind all statistical inference. Thus, if we are estimating the mean of a population, we might reasonably wish our estimator to be close to the observation.

Consider this as it relates to Bayesian estimation of a population mean $\theta$. For convenience consider squared error loss. The Bayesian, in estimating $\theta$, combines his prior distribution and the distribution for the observation, $x$. The estimator that is used for $\theta$ is the mean of the posterior distribution. Suppose, for example, that we wish to estimate the mean of a normal population with variance equal to 1. If the prior is a normal distribution with mean equal to zero and variance equal to $\sigma^2 > 0$, then the posterior mean is $(\sigma^2/\sigma^2 + 1)x$, where $x$ is the observation. But note, as $x$ gets large this estimator grows far from $x$.

In the discussion section of (Lindley, 1968), Beale and Lindley make note of this. They note that the example is even more disconcerting in that the posterior mean grows further from the observation, $x$, as $x$ gets larger. And a very large observation may tend to make us
think that the prior was ill-chosen. This is so, again, since we expect $x$ to be close to the mean of the distribution. However, for the example given we have $x - \text{E}(\theta|X = x) \rightarrow \infty$ as $x \rightarrow \infty$, where $\text{E}(\theta|X = x)$ is the posterior mean. Thus, Beale and Lindley raise the question: What sort of priors lead to estimators that avoid this undesirable behavior? Lindley shows that for a normal distribution if the prior is a student's $t$ distribution then $x - \text{E}(\theta|X = x) \rightarrow 0$ as $x \rightarrow \infty$.

These ideas were taken up next in (Dawid, 1973). He considers a single observation, $X$, which has the location parameter, $\theta$. For convenience, we consider the case when $\theta$ is the mean of the distribution of $X|\theta = \theta$. Dawid considers the behavior of the posterior mean as $x$, the observation, tends to infinity. He gives sufficient conditions on the distribution of the observation and the prior to insure that $x - \text{E}(\theta|X = x) \rightarrow 0$ as $x \rightarrow \infty$. Dawid notes that the case Lindley refers to (normal distribution and student's $t$ prior) is covered by his results.

Dawid also notes that if the properties on the distribution of the observation and the prior are interchanged then we have $\text{E}(\theta|X = x) + \text{E}(\theta)$, the prior mean, as $x \rightarrow \infty$! This is more disconcerting than the behavior noted earlier. The extremely large observation may at times be thought of as discrediting the prior; yet in the limit (as $x \rightarrow \infty$) we base our inference solely on the prior.

In (Meeden and Isaacson, 1976) these ideas are considered as they
relate to the one dimensional exponential family. They consider estimation of the natural parameter. Sufficient conditions are given for the posterior to be asymptotically normally distributed about the posterior mode as the observed \( x \) tends to infinity. These same conditions are also sufficient for \( \mathbb{E}(\theta|X = x) - \theta^*_x \to 0 \) as \( x \to \infty \), where \( \theta^*_x \) is the mode of the distribution of \( \theta|X = x \).

This work proposes to extend these results to a multivariate situation. To this end the following notation will be used throughout. All vectors are written as column vectors in \( \mathbb{R}^n \), \( n \) dimensional Euclidean space. The transpose is denoted by \( ^T \). The prime, ' , is used to represent a derivative. Thus if \( f: \mathbb{R}^n \to \mathbb{R} \) has all of its first partial derivatives existing, we write

\[
\begin{pmatrix}
  f_1(x) \\
  f_2(x) \\
  \vdots \\
  f_n(x)
\end{pmatrix} = \begin{pmatrix}
  f_1(x) \\
  f_2(x) \\
  \vdots \\
  f_n(x)
\end{pmatrix}^T,
\]

where \( f_i(x) = \frac{\partial f}{\partial x_i}(x) \) for \( i = 1, 2, \ldots, n \). Somewhat less conventionally, if \( f: \mathbb{R}^n \to \mathbb{R} \) has all of its second partials and mixed second partial derivatives existing, then we write,

\[
f''(x) = \begin{pmatrix} f_{ij}(x) \end{pmatrix} \quad \text{for } i, j = 1, 2, \ldots, n,
\]
where \( f_{ij}(x) = \frac{2^2 x}{2x_i 2x_j} (x) \) for \( i,j = 1, 2, \ldots, n \).

In chapter II we extend some of Meeden and Isaacson's results to estimation of the natural parameter of the multivariate exponential family. The situation considered is the case when the posterior of \( \Theta | X = x \) has a density which can be written as

\[
\int_{\mathbb{R}^n} \frac{\exp(\theta^T X - \lambda(\theta))}{\int_{\mathbb{R}^n} \exp(\theta^T X - \lambda(\theta))} d\theta.
\]

Here \( \Theta \) and \( x \) are vectors in \( \mathbb{R}^n \). Suppose that \( \lambda''(y) = \Sigma \), a positive definite matrix, as \( ||y|| \to \infty \). Then under suitable conditions it is shown that \( \Theta | X = x \) is asymptotically normal with mean \( (\lambda')^{-1}(x) \) and variance-covariance matrix \( \Sigma^{-1} \) as \( ||x|| \to \infty \).

The same conditions yield \( \mathbb{E}(\Theta | X = x) = (\lambda')^{-1}(x) \to 0 \), the zero vector, as \( ||x|| \to \infty \).

In chapter III we generalize the one dimensional results of Dawid. His results basically cover the case when the prior density, \( f \), satisfies \( f(x - y)/f(x) \to 1 \) as \( x \to \infty \). We allow \( f(x - y)/f(x) \to h(y) \) as \( x \to \infty \). We also extend the results to the multivariate situation. Here for a single observation \( X = (X_1, X_2, \ldots, X_n)^T \) from a distribution with the location parameter \( \Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)^T \), we investigate the posterior distribution of \( \Theta | X = x \) as \( x \) gets "large". The key ingredient to the chapter is the term \( f(x - y)/f(x) \), where \( f \)
is the prior density. Consider a path in $\mathbb{R}^n$ traced by $\gamma(t)$ for $t \in \mathbb{R}^+$. If it happens that $f(\gamma(t) - y)/f(\gamma(t)) \to h(y)$ as $t \to \infty$, then generally the distribution of $\gamma(t) - (\theta|X = \gamma(t))$ as $t \to \infty$ is given by

\begin{equation}
(1.1) \quad g(y) h(\gamma(y)) \int_{\mathbb{R}^n} g(y) h(y) dy.
\end{equation}

Here $g$ is the density for $X|\theta = 0$. Slightly stronger conditions insure convergence of $\gamma(t) - E(\theta|X = \gamma(t))$ to the expected value of the distribution with density (1.1).

Some particular paths are of special interest. We may require the same limit (1.1) for all paths which have $\gamma_i(t) \to \infty$ as $t \to \infty$ for $i = 1, 2, \ldots, n$. These are considered. Another case of interest is as one coordinate of the observed vector $x$ goes off to infinity.

Suppose we let $\tilde{x}_i \in \mathbb{R}^{n-1}$ be the vector $x$ with the $i$th coordinate, $x_i$, removed. If $f(x - y)/f(z) \to h(y)$ as $x_i \to \infty$, where $\gamma_i(\tilde{x}_i)$ is the marginal distribution of the prior for $\gamma_i$, then sufficient conditions are given for the density of $x - (\theta|X = x)$ to converge to the density

\begin{equation}
(1.1) \quad h(y) \frac{\gamma_i(\tilde{x}_i)}{\gamma_i(\tilde{y}_i)} g(y) / \int_{\mathbb{R}^n} h(y) \frac{\gamma_i(\tilde{x}_i)}{\gamma_i(\tilde{y}_i)} g(y) dy.
\end{equation}

One may also wish to consider the behavior of the posterior as simply $||x|| \to \infty$. If we require the same limit (1.1) for all paths
with $||y(t)|| \to \infty$ as $t \to \infty$, then much the same can be done.

Here we desire $f(x - y)/f(x) \to h(y)$ as $||x|| \to \infty$. In (Hill, 1974) it is shown that generally, if a limiting distribution exists simply as $||x|| \to \infty$, then that distribution is the same as the distribution of $X|\theta = 0$. This corresponds to the case $h(y) \equiv 1$.

We give conditions for behavior similar to that mentioned in the previous two paragraphs with Hill's result in mind.

Examples are given which show that a small change in the prior, especially in the tails, can lead to a quite different estimator. We advocate having somewhat fat tails for the prior. It is shown that this leads to posterior means which are close to the observed $x$ when $x$ is "large". In this discussion it is shown that if the distribution for $X|\theta = 0$ and the prior distribution for $\theta$ are the same, then the posterior mean is simply $\frac{1}{2}x$. This estimator gives us a case between the extremes of $x - E(\theta|X = x) \to 0$ and $E(\theta|X = x) \to 0$ as $x \to \infty$.

In chapters II and III, the basic concern is with the posterior mean, the Bayes estimate under squared error loss. With other loss functions, one uses different estimators. For example, if the loss is absolute error, the Bayes estimate is the posterior median. For loss function $L$, let $\delta_L(x)$ be the Bayes estimate for $\theta$. In chapter IV a wide class of loss functions is presented which have the property that $\delta_L(x) - E(\theta|X = x) \to 0$ as $x \to \infty$. This result basically holds when the density of $(\theta|X = x) - E(\theta|X = x)$ converges to a
symmetric density. Thus even though the posterior mean is considered in chapters II and III, many of the results remain valid with $\delta_L(x)$ replacing $E(\theta|X = x)$. 
II. EXPONENTIAL FAMILY

A. Main Results

Consider a σ-finite measure, μ, defined on the Borel sets of \( \mathbb{R}^n \). Define

\[
\mathcal{P}_\mu = \{ \theta \in \mathbb{R}^n : \int_{\mathbb{R}^n} \exp(\theta^T x) d\mu(x) < \infty \}.
\]

Consideration is given here only to measures, μ, with \( \mathcal{P}_\mu = \mathbb{R}^n \). Thus define \( \beta(\theta) \) by

\[
\exp(\beta(\theta)) = \int_{\mathbb{R}^n} \exp(\theta^T x) d\mu(x) \quad \text{for} \quad \theta \in \mathbb{R}^n.
\]

Now let \( X = (X_1, X_2, \ldots, X_n)^T \) be a random vector taking values in \( \mathbb{R}^n \), whose family of possible densities with respect to μ is given by

\[
\{ \exp(\theta^T x - \beta(\theta)), \theta \in \mathbb{R}^n \}.
\]

For this family, belonging to the multiparameter exponential family, consider estimation of the parameter \( \theta \). Let \( g(\theta) \) denote a prior density for \( \theta \) with respect to Lebesgue measure over \( \mathbb{R}^n \). If \( g(\theta) > 0 \) for all \( \theta \in \mathbb{R}^n \), then write \( g(\theta) = \exp(-\gamma(\theta)) \). So with \( \lambda = \beta + \gamma \) the posterior density for \( \theta | X = x \) is seen to be
Conditions are given on \( \lambda \) for the posterior density to converge to a normal density as \( ||x|| \rightarrow \infty \). In fact, under appropriate conditions, it is shown that \( \theta|x = x \) is asymptotically normal with mean vector \( (\lambda')^{-1}(x) \) and variance-covariance matrix given by \( \Sigma^{-1} \), where \( \Sigma = \lim_{||x|| \rightarrow \infty} \lambda''(x) \). The following discussion and lemmas will greatly simplify the proof of the main theorem of this chapter.

The notion of a positive definite (p.d.) matrix will be often used. The square matrix, \( A \), is said to be positive definite if, for all \( x \in \mathbb{R}^n \), \( x^T A x > 0 \) with equality only for \( x \) the zero vector. \( A \) is said to be nonnegative definite (n.n.d.) if only \( x^T A x \geq 0 \) for all \( x \in \mathbb{R}^n \).

If \( \lambda: \mathbb{R}^n \rightarrow \mathbb{R} \) has all of its second partials existing on a convex set \( B \) in \( \mathbb{R}^n \) and \( \lambda''(x) \) is p.d. for each \( x \in B \) then \( \lambda \) is a strictly convex function over \( B \). If \( \lambda''(x) \) is n.n.d. for each \( x \in B \) then \( \lambda \) is a convex function over \( B \). Thus, consider the following lemmas.

Lemma 2.1: Suppose \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) has all second partials existing for each \( x \in B \), a convex, open set in \( \mathbb{R}^n \). If \( f''(x) \) is p.d. for all \( x \in B \), then there exists at most one point, \( x_0 \), such that \( f'(x_0) = 0 \), the zero vector.
Proof. Assume $x_0 \in B$ such that $f'(x_0) = 0$. Then the Taylor expansion of $f$ about $x_0$ can be written as

$$f(x) = f(x_0) + (x - x_0)^T f'(x_0) + \frac{1}{2}(x - x_0)^T f''(\tau)(x - x_0),$$

where $\tau = \tau(x_0, x)$ is a point on the line segment from $x_0$ to $x$. See (Widder, 1961) for more details. But since $f'(x_0) = 0$, (2.2) can be written as

$$f(x_0) + \frac{1}{2}(x - x_0)^T f''(\tau)(x - x_0).$$

But $f''(\tau)$ is p.d. since $\tau \in B$. Thus

$$\frac{1}{2}(x - x_0)^T f''(\tau)(x - x_0) > 0 \text{ for } x \neq x_0.$$

Thus, by (2.3) it is seen that $f$ has a minimum over $B$ at $x_0$.

Assume $x_1 \in B$ with $x_1 \neq x_0$ and $f'(x_1) = 0$. The same argument can be used to deduce $f$ has a minimum over $B$ at $x_1$, so $f(x_0) = f(x_1)$. But by (2.3) one can write

$$f(x_1) = f(x_0) + \frac{1}{2}(x_1 - x_0)^T f''(\tau)(x_1 - x_0).$$

Thus

$$0 = f(x_1) - f(x_0)$$

$$= \frac{1}{2}(x_1 - x_0)^T f''(\tau)(x_1 - x_0) > 0.$$
Thus, a contradiction arises with the assumption that two points, \( x_0, x_1 \), exist such that \( f'(x_1) = f'(x_0) = 0 \).

Lemma 2.2: Let \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) have all its second partials existing. If there exists \( K > 0 \) such that \( f''(y) - KI \) is n.n.d. for each \( y \in \mathbb{R}^n \), then \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is one-to-one and onto.

Proof: First consider the one-to-one-ness of \( f' \). Fix \( x \in \mathbb{R}^n \). Define \( f'(y) = \lambda(y) - x^T y \). Then \( f''(y) = \lambda''(y) - x \) and \( f''(y) = \lambda''(y) \).

Thus \( f''(y) \) is p.d. for \( y \in \mathbb{R}^n \). Thus by lemma 2.1, there exists at most one point \( y_o \in \mathbb{R}^n \) such that \( f'(y_o) = 0 \). But \( f'(y_o) = \lambda'(y_o) - x \).

So \( f'(y_o) = 0 \) if and only if \( \lambda'(y_o) = x \). So there exists at most one \( y_o \in \mathbb{R}^n \) such that \( \lambda'(y_o) = x \). Thus \( \lambda' \) is one-to-one.

Next consider onto-ness. Again fix \( x \in \mathbb{R}^n \) and define \( f''(y) = \lambda(y) - x^T y \). Again \( f''(y) = \lambda''(y) \). First note there exists \( y_o \in \mathbb{R}^n \) such that \( \lambda'(y_o) = 0 \). This is so since it is assumed \( \lambda' \) is continuous and it will be shown that \( \lambda \) has a minimum. Thus the point \( y_o \) which minimizes \( \lambda(y) \) has \( \lambda'(y_o) = 0 \). To see that \( \lambda \) has a minimum it is shown that \( \lambda(x) \rightarrow \infty \) as \( ||x|| \rightarrow \infty \). Expand \( \lambda \) in a Taylor series about zero. Then

\[
\lambda(y) = \lambda(0) + y^T \lambda'(0) + \frac{1}{2} y^T \lambda''(\tau)y
\]

where \( \tau \) is some point between 0 and y. Expand \( \lambda''(\tau) \) in a Taylor series about zero. Then

\[
\lambda''(\tau) = \lambda''(0) + \frac{1}{2} y^T \lambda'''(\tau) - kI y + \frac{k T}{2} y
\]
\[ 2 \lambda(0) + y^T \lambda'(0) + \frac{k}{2} y^T y \]

\[ 2 \lambda(0) - |y||\lambda'(0)|| + \frac{k}{2} y^T y \]

(2.5) \[ = \lambda(0) + |y||\frac{k}{2}|y|| - |\lambda'(0)||. \]

But by (2.5) it is obvious that \( \lim_{|y| \to \infty} \lambda(y) = \infty \) since \( k > 0 \).

Now if the minimum for \( \lambda \) occurs at \( y_o \), then \( \lambda'(y_o) = 0 \). Now expand \( f_x(y) \) in a Taylor series about \( y_o \). Thus

(2.6) \[ f_x(y) = f_x(y_o) + (y - y_o)^T f'_x(y_o) + \frac{1}{2}(y - y_o)^T f''_x(\tau)(y - y_o), \]

where \( \tau \) is a point on the line segment from \( y \) to \( y_o \). But one can rewrite (2.6) as

\[ f_x(y) = \lambda(y_o) - x^T y_o + (y - y_o)^T \{\lambda'(y_o) - x\} + \frac{1}{2}(y - y_o)^T \lambda''(\tau)(y - y_o) \]

\[ = \lambda(y_o) - x^T y_o + (y - y_o)^T (-x) + \frac{1}{2}(y - y_o)^T \lambda''(\tau)(y - y_o) \]

\[ = \lambda(y_o) - x^T y_o - (y - y_o)^T x + \frac{1}{2}(y - y_o)^T \{kI + \lambda''(\tau) - kI\} (y - y_o) \]

\[ = \lambda(y_o) - x^T y_o - (y - y_o)^T x + \frac{1}{2}(y - y_o)^T kI(y - y_o) \]

\[ + \frac{1}{2}(y - y_o)^T \{\lambda''(\tau) - kI\} (y - y_o) \]
\[ z \lambda(y_o) - x^T y_o - |(y - y_o)^T x| + \frac{k}{2} (y - y_o)^T (y - y_o) \]

\[ \lambda(y_o) - x^T y_o - |y - y_o| + \frac{k}{2} |y - y_o|^2 \]

(2.7) = \lambda(y_o) - x^T y_o + \frac{k}{2} |y - y_o|^2

But since \( \frac{k}{2} > 0 \), it is seen that (2.7) converges to infinity as \(|y - y_o| \rightarrow \infty\), or as \(|y| \rightarrow \infty\) since \( y_o \) is fixed. Thus

\[ \lim f_x(y) = \infty. \]

Thus \( f_x \) has a unique minimum for each \( x \in \mathbb{R}^n \) and the minimum occurs at the point \( y \) such that \( f'_x(y) = 0 \) or \( \lambda'(y) = x \). So for each \( x \in \mathbb{R}^n \) there exists \( y \in \mathbb{R}^n \) such that \( \lambda'(y) = x \), or \( \lambda' \) is onto.

Remark: Since \( \lambda \) is convex and \( \lambda' \) is one-to-one and onto, it follows that \( \lim_{|y| \rightarrow \infty} |\lambda'(y)| = \infty. \)

Thus, conversely \( \lim_{|x| \rightarrow \infty} |(\lambda')^{-1}(x)| = \infty. \) Thus in taking limits with \(|x| \rightarrow \infty \) going to infinity one could equivalently take limits as \(|y| \rightarrow \infty \).

Recall that the posterior for \( \theta | \mathbf{X} = \mathbf{x} \) is given by (2.1).

**Theorem 2.1:** Suppose there exists \( k > 0 \) such that \( \lambda''(\theta) = kI \) is n.n.d. for all \( \theta \in \mathbb{R}^n \). Suppose \( \lambda''(\theta) + \Sigma \) p.d. as \(|\theta| \rightarrow \infty. \)

Then \( \theta | \mathbf{X} = \mathbf{x} \) is asymptotically normal with mean \( (\lambda')^{-1}(\mathbf{x}) \) and variance-covariance matrix \( \Sigma^{-1} \) as \(|x| \rightarrow \infty. \)
Proof: Let \( y = (\lambda')^{-1}(x) \). Define \( Z_y = (0, Y = y) - y \). Thus the density of \( Z_y \) is given by

\[
(2.8) \quad f_y(z) = \frac{\exp \left( \left( (z + y)^T \lambda'(y) - \lambda(z + y) \right) \right)}{\int_{\mathbb{R}^n} \exp \left( \left( (r + y)^T \lambda'(y) - \lambda(r + y) \right) \right) \, dr}
\]

Now the Taylor expansion of \( \lambda(z + y) \) about \( z = 0 \) is given by

\[
(2.9) \quad \lambda(z + y) = \lambda(y) + z^T \lambda'(y) + \frac{1}{2} z^T \lambda''(\xi_y + y)z,
\]

where \( \xi_y \) is on the line segment from \( 0 \) to \( z \). Thus the numerator of (2.8) can be written as

\[
\exp \left\{ z^T \lambda'(y) + y^T \lambda'(y) - \lambda(y) - z^T \lambda'(y) - \frac{1}{2} z^T \lambda''(\xi_y + y)z \right\}
\]

\[
= \exp \left\{ y^T \lambda'(y) - \lambda(y) - \frac{1}{2} z^T \lambda''(\xi_y + y)z \right\}.
\]

Thus (2.8) can be rewritten as

\[
(2.10) \quad f_y(z) = \frac{\exp \left\{ y^T \lambda'(y) - \lambda(y) - \frac{1}{2} z^T \lambda''(\xi_y + y)z \right\}}{\int_{\mathbb{R}^n} \exp \left\{ y^T \lambda'(y) - \lambda(y) - \frac{1}{2} r^T \lambda''(\xi_y + y)r \right\} \, dr}
\]

Now for each fixed \( z \in \mathbb{R}^n \) the numerator of (2.10) converges to \( \exp \left\{ -\frac{1}{2} z^T \lambda'' \right\} \) as \( ||y|| \to \infty \), since \( \xi_y \) is on the line segment from
0 to $z$. Thus the theorem is established once it is shown that the denominator of (2.10) converges to $\int_\mathbb{R} \exp \left\{ -\frac{1}{2} z^T \Sigma z \right\} \, dz$.

In the denominator consider the term

$$-\frac{1}{2} x^T \lambda''(\xi_y + y)r = -\frac{1}{2} x^T \{ kI + \lambda''(\xi_y + y) - kI \} r$$

$$= -\frac{1}{2} x^T kIr - \frac{1}{2} r^T \{ \lambda''(\xi_y + y) - kI \} r$$

$$\leq -\frac{1}{2} r^T kIr$$

$$= -\frac{k}{2} r^T r.$$

But $\int_\mathbb{R} \exp \left\{ -\frac{k}{2} r^T r \right\} \, dr = (2k\|r\|^2)^{n/2} < \infty$.

So by the Lebesgue dominated convergence theorem, the limit may be moved inside the integral. Thus

$$\lim_{||y|| \to \infty} \int_\mathbb{R} \exp \left\{ -\frac{1}{2} x^T \lambda''(\xi_y + y) r \right\} \, dr$$

$$= \int_\mathbb{R} \exp \left\{ -\frac{1}{2} r^T \Sigma r \right\} \, dr.$$

**Theorem 2.2**: Suppose the conditions of theorem 2.1 hold. Let $Q: \mathbb{R}^n \to \mathbb{R}$ with $E(\|Q(W)\|) < \infty$, where $W$ is a normal random vector with mean $0$ and variance-covariance matrix $kI$. Then $E(Q(Z_y)) \to$
E(Q(U)) as ||y|| → ∞. Here U is a normal random vector with mean 0 and variance-covariance matrix $U^{-1}$.

Proof: $E(|Q(W)|)$ is given by

\[ (2.11) \quad \int_{R^n} |Q(x)| \exp \left\{ -\frac{k}{2} r^T r \right\} dr. \]

Now (2.11) is finite if and only if

\[ (2.12) \quad \int_{R^n} |Q(x)| \exp \left\{ -\frac{k}{2} r^T r \right\} dr < \infty. \]

Now $E(Q(Z_y))$ can be written as

\[ (2.13) \quad \frac{\int_{R^n} Q(z) \exp \left\{ -\frac{1}{2} z^T \lambda''(\xi_y + y) z \right\} dz}{\int_{R^n} \exp \left\{ -\frac{1}{2} z^T \lambda''(\xi_y + y) z \right\} dz} \]

Now the integrand in the numerator is bounded by $|Q(z)| \exp \left\{ -\frac{k}{2} z^T z \right\}$, as in theorem 2.1. But the integral of the expression is (2.12) which is finite. Thus by the Lebesgue dominated convergence theorem, one can move the limit inside the integral in the numerator of (2.13). The same operation is justified in the denominator as in theorem 2.1. Thus the limit of (2.13) as ||y|| → ∞ is
\[
\frac{\int_{\mathbb{R}^n} Q(z) \exp \left\{ -\frac{1}{2} z^T \Sigma z \right\} \, dz}{\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} z^T \Sigma z \right\} \, dz} = E(Q(u))
\]

Since a normal random vector has all of its moments, the conditions of theorem 2.1 imply that \( E(Z_y) \to 0 \), the zero vector, as \( ||y|| \to \infty \). But \( Z_y = (\theta | Y = y) - y = (\theta | (\lambda')^{-1}(x)) - (\lambda')^{-1}(x) \).

Thus the conditions of theorem 2.1 imply \( E(\theta | X = x) - (\lambda')^{-1}(x) \to 0 \) as \( ||x|| \to \infty \). Thus with norm-squared error loss the Bayes estimate for \( \theta \) is asymptotically \( (\lambda')^{-1}(x) \) as \( ||x|| \to \infty \).

B. Applications

Consider estimation of a normal mean vector \( \theta \). If the variance-covariance matrix is the identity, \( I \), then \( \theta \) is the natural parameter. For this case it is well known that if the prior is normal mean \( 0 \) and variance-covariance matrix \( \Gamma \), p.d., then the posterior for \( \theta \) is exactly normal with mean \( (I + \Gamma^{-1})^{-1} x \) and variance-covariance matrix \( (I + \Gamma^{-1})^{-1} \). But the results of this chapter agree with this, since

\[
\lambda(y) = \frac{1}{2} y^T (I + \Gamma^{-1}) y,
\]

\[
\lambda'(y) = (I + \Gamma^{-1}) y,
\]
\[ \lambda''(y) = I + \Gamma^{-1} \text{ p.d.} \]

Here \((\lambda')^{-1}(x) = \{I + \Gamma^{-1}\}^{-1} x\).

At the other extreme, if the prior is the improper uniform prior over \(R^n\), then it is well established that the posterior is exactly normal with mean \(x\) and variance-covariance matrix \(I\). Again, our results agree. Here

\[ \lambda(y) = \frac{1}{2} y^T y, \]

\[ \lambda'(y) = y \]

\[ \lambda''(y) = I \text{ p.d.} \]

Here \((\lambda')^{-1}(x) = x\).

The cases between these are of interest. For example, again suppose we desire to estimate a normal mean with variance-covariance matrix \(I\). Suppose the prior for \(\theta\) is proportional to \(\exp(-||\theta||)\) for \(||\theta|| \geq 1\) and \(\exp\left(-\frac{3}{8} + \frac{3}{2}\left(\frac{||\theta||^2}{2} - \frac{||\theta||^4}{12}\right)\right)\) for \(||\theta|| \leq 1\). This is done to make \(\lambda\) differentiable at \(0\). Now for \(||y|| \geq 1\),

\[ \lambda(y) = \frac{1}{2} ||y||^2 + ||y|| + K_1, \]
\[
\lambda'(y) = \frac{||y|| + 1}{||y||} y,
\]

\[
\lambda''(y) = I + \frac{1}{||y||} I - \frac{1}{||y||^3} H_y.
\]

Here \( H_y = (y_i y_j) \) for \( i, j = 1, 2, \ldots, n \). But if \( ||y|| \geq 1 \), then \( \lambda''(y) - kI \) is n.n.d. for \( 0 < k \leq 1 \) since for \( x \in \mathbb{R}^n \) we have

\[
x^T(\lambda''(y) - kI)x
\]

\[
= x^T x + \frac{x^T H y x}{||y||^3} - k x^T x
\]

\[
= (1 - k) ||x||^2 + \frac{||x||^2}{||y||^3} - \frac{\left( \sum_{i=1}^{n} x_i y_i \right)^2}{||y||^3}.
\]

However

\[
\frac{||x||^2}{||y||^3} - \frac{\left( \sum_{i=1}^{n} x_i y_i \right)^2}{||y||^3} \geq 0 \quad \text{by the Cauchy inequality.}
\]

Similarly \( \lambda''(y) - kI, 0 < k \leq 1 \), is n.n.d. for \( ||y|| \leq 1 \). In this case

\[
\lambda''(y) = \frac{5}{2} I - \frac{||y||^2}{2} I - H_y.
\]

However, \( \lim_{||y|| \to \infty} \lambda''(y) = I \) and \( (\lambda')^{-1}(x) = \frac{||x|| - 1}{||x||} x \) for \( ||x|| \geq 2 \). Thus by theorem 2.2, then,
\[ \lim_{||x|| \to \infty} \{E(\theta | X = x) - \frac{||x|| - 1}{||x||} x\} = 0.\]

In particular if we consider the \( j \)th coordinate of \( E(\theta | X = x) \) and take the limit as \( x_i \to \infty \), then we get

\[ \lim_{x_i \to \infty} E(\theta_j | X = x) - \left( \frac{||x|| - 1}{||x||} \right) x_j = 0 \]

for \( j = 1, 2, \ldots, n \). However \( \lim_{x_i \to \infty} \left( \frac{||x|| - 1}{||x||} \right) x_j = x_j \) for \( i \neq j \).

Thus \( \lim_{x_i \to \infty} E(\theta_j | X = x) = x_j \) for \( i \neq j \).

In general then, as the prior gets sharper the Bayes estimate gets shrunk more toward the prior's center. The asymptotic normality will hold in general if \( \lambda(y) \) is bounded by a quadratic form, assuming of course that \( \lambda''(y) \to \Gamma \), p.d., as \( ||y|| \to \infty \).
III. LOCATION VECTOR

A. Main Results

Consider a random vector \( X = (X_1, X_2, \ldots, X_n)^T \) which has a location vector \( \theta = (\theta_1, \theta_2, \ldots, \theta_n)^T \). \( \theta \) being a location vector implies that the distribution of \( (X|\theta = \theta) - \theta \) is independent of \( \theta \).

We define the random vector \( D = (X|\theta = \theta) - \theta = (X|\theta = 0) \).

We assume throughout this chapter that \( D \) has a density, \( g \), with respect to Lebesgue measure over \( \mathbb{R}^n \). This distribution is called the error distribution. Thus we write \( X = \theta + D \), where \( D \) has the error density, \( g \), and \( \theta \) has a prior density, \( f \), and \( \theta \) and \( D \) are independent. The parameter space is taken to be all of \( \mathbb{R}^n \). The only priors to be considered are those which have densities, \( f \), which are everywhere positive with respect to Lebesgue measure over \( \mathbb{R}^n \). With these assumptions we consider the behavior of the posterior distribution when \( x \) gets large in various ways.

Let \( \gamma: \mathbb{R}^+ \to \mathbb{R}^n \). Thus \( \gamma(t) \), for \( t \in \mathbb{R}^+ \), maps out a path in \( \mathbb{R}^n \). Suppose \( \lim_{t \to \infty} ||\gamma(t)|| = \infty \). We will consider \( x \) getting large along the path \( \gamma \). For the next theorem, however, it is not required that \( ||\gamma(t)|| \to \infty \) as \( t \to \infty \).

**Theorem 3.1:** Define \( k_y(y) = \sup_{t \in \mathbb{R}^+} \frac{f(\gamma(t) - y)}{f(\gamma(t))} \). Suppose the following two conditions hold,
\( (1) \quad \frac{f(\gamma(t) - y)}{f(\gamma(t))} + h_\gamma(y) > 0 \) a.e. as \( t \to \infty \), and
\( (2) \quad \int_{\mathbb{R}^n} k_\gamma(y) g(y) \, dy < \infty. \)

Then \( D|X = \gamma(t) \xrightarrow{\text{law}} Z_\gamma \) as \( t \to \infty \). Here \( Z \) has a density given by
\( (3.1) \quad g(y) h_\gamma(y) \frac{1}{\int_{\mathbb{R}^n} g(y) h_\gamma(y) \, dy}. \)

Proof: The density for \( D|X = \gamma(t) \) is
\( (3.2) \quad g(y) f(\gamma(t) - y) \frac{1}{\int_{\mathbb{R}^n} g(y) f(\gamma(t) - y) \, dy}. \)

But since \( f(x) > 0 \) for each \( x \), we can rewrite (3.2) as
\( (3.3) \quad g(y) \frac{f(\gamma(t) - y)}{f(\gamma(t))} \frac{1}{\int_{\mathbb{R}^n} g(y) f(\gamma(t) - y) / f(\gamma(t)) \, dy}. \)

Now consider the limit of (3.3) as \( t \to \infty \). The numerator converges to \( g(y) h_\gamma(y) \) by assumption (1). Consider the denominator. Since
\( g(y)f(\gamma(t) - y) / f(\gamma(t)) \leq g(y)k_\gamma(y) \) and by (2) \( \int_{\mathbb{R}^n} g(y)k_\gamma(y) \, dy < \infty \), the Lebesgue dominated convergence theorem may be used to move the limit inside the integral. Thus the denominator of (3.3) converges to \( \int_{\mathbb{R}^n} g(y) h_\gamma(y) \, dy \) as \( t \to \infty \). Thus the limit of (3.2) as \( t \to \infty \) is (3.1). Thus the posterior density converges to the proper
density (3.1), so by Scheffe's theorem the convergence is complete. Note that (3.1) is proper since

$$\int_{\mathbb{R}^n} g(y) h_{\gamma}(y) \, dy \leq \int_{\mathbb{R}^n} g(y) k_{\gamma}(y) \, dy < \infty,$$

and

$$\int_{\mathbb{R}^n} g(y) h_{\gamma}(y) \, dy > 0 \text{ since } h_{\gamma}(y) > 0 \text{ a.e.}$$

Slightly stronger conditions are needed to insure convergence of expected values.

Corollary 3.1.1: Let $m: \mathbb{R}^n \to \mathbb{R}$. Suppose conditions (1) and (2) of theorem 3.1 hold. Then

$$\int_{\mathbb{R}^n} |m(y)| g(y) k_{\gamma}(y) \, dy < \infty$$

implies $E(m(D)|X = \gamma(t)) \to E(Z_{\gamma})$ as $t \to \infty$.

Proof: Again the density for $D|X = \gamma(t)$ is given by (3.3). Thus,

$$E(m(D)|X = \gamma(t))$$

can be written as

$$(3.4) \quad \frac{\int_{\mathbb{R}^n} m(y) g(y) f(\gamma(t) - y)/f(\gamma(t)) \, dy}{\int_{\mathbb{R}^n} g(y) f(\gamma(t) - y)/f(\gamma(t)) \, dy}.$$
converges to \( \int_{\mathbb{R}^n} g(y) h(y) \, dy \) as \( t \to \infty \). Thus all that remains is to show the numerator converges to \( \int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy \) as \( t \to \infty \).

But since

\[
m(y) g(y) \frac{r(x(t) - y)}{r(y(t))} \leq |m(y)| g(y) k(t)
\]

and

\[
\int_{\mathbb{R}^n} |m(y)| g(y) k(y) \, dy < \infty,
\]

the Lebesgue dominated convergence theorem can be used to move the limit inside the integral. Thus the numerator of (3.4) converges to

\[
\int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy
\]

as \( t \to \infty \). So \( E(m(D)|X = \gamma(t)) \) converges to

\[
\frac{\int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy}{\int_{\mathbb{R}^n} g(y) h(y) \, dy} = E(m(Z))
\]

Note here that \( E(|m(Z)|) < \infty \) since

\[
E(|m(Z)|) = \frac{\int_{\mathbb{R}^n} |m(y)| g(y) h(y) \, dy}{\int_{\mathbb{R}^n} g(y) h(y) \, dy}
\]
\[
\int_{R^n} |m(y)| g(y) k_\gamma(y) \, dy < \frac{\int_{R^n} g(y) h_\gamma(y) \, dy}{\int_{R^n} s(y) \, dy} < \infty.
\]

Corollary 3.1.2: Suppose conditions (1) and (2) of theorem 3.1 hold. Suppose also that

\[
\int_{R^n} |y_i| k_\gamma(y) g(y) \, dy < \infty \quad \text{for } i = 1, 2, \ldots, n.
\]

Then \( \gamma(t) - E(\Theta|X = \gamma(t)) \to E(\lambda) \) as \( t \to \infty \).

Proof: Define \( m:R^n \to R \) by \( m_i(d) = d_i \) for \( i = 1, 2, \ldots, n \).

Then

\[
\int_{R^n} |m(y)| k_\gamma(y) g(y) \, dy
= \int_{R^n} |y_i| k_\gamma(y) g(y) \, dy < \infty \quad \text{for } i = 1, 2, \ldots, n.
\]

Thus by corollary 3.1.1, \( E(m_i(D)|X = \gamma(t)) \to E(m_i(\lambda)) \) as \( t \to \infty \),

for \( i = 1, 2, \ldots, n \). But

\[
E(m_i(D)|X = \gamma(t)) = E(D_i|X = \gamma(t))
= E(X_i - \theta_i|X = \gamma(t))
\]
where \( \gamma_i(t) \) is the \( i \)th coordinate of the vector \( \gamma(t) \). Also

\[
E(m(Z_y)) = E(Z_{i,y}),
\]

where \( Z_{i,y} \) is the \( i \)th coordinate of the random vector \( Z_y \). Thus \( \gamma(t) - E(\theta | X = \gamma(t)) \rightarrow E(Z_{i,y}) \) coordinatewise as \( t \rightarrow \infty \). Again note that the conditions assumed imply \( E(|Z_{i,y}|) < \infty \) for \( i = 1, 2, \ldots, n \).

Some particular paths and families of paths are of special interest. In some case we may be interested in the limit as only one coordinate of \( x \), say \( x_i \), goes off to infinity as the other coordinates, \( x^j \), remain fixed. Or we may wish to investigate cases where a limit exists for all paths with \( ||\gamma(t)|| \rightarrow \infty \) as \( t \rightarrow \infty \). First though, we consider paths with \( \gamma_i(t) \rightarrow \infty \) as \( t \rightarrow \infty \) for \( i = 1, 2, \ldots, n \). In fact, in the following we require the same limiting distribution for all paths with \( \gamma_i(t) \rightarrow \infty \) for \( i = 1, 2, \ldots, n \).

**Theorem 3.2:** Define \( k(y) = \sup_x f(x - y)/f(x) \). Suppose the following two conditions hold,

(1) \( f(x - y)/f(x) \rightarrow h(y) > 0 \) a.e. as \( \min x_i \rightarrow \infty \),

and

(2) \( \int_{R^n} k(y) g(y) \, dy < \infty \).

Then \( D|X = x \overset{Law}{\rightarrow} Z \) as \( \min x_i \rightarrow \infty \).
Here $Z$ is a random vector which has a density given by

\begin{equation}
(3.5) \quad g(y) \frac{h(y)}{\int_{\mathbb{R}^n} g(y) h(y) \, dy}, \quad \text{for } y \in \mathbb{R}^n.
\end{equation}

**Proof:** The density of $D|X = x$ is

\begin{equation}
(3.6) \quad \frac{g(y) f(x - y)}{\int_{\mathbb{R}^n} g(y) f(x - y) f(x) \, dy} \frac{f(x - y)}{f(x)} = g(y) \frac{f(x - y)}{f(x)} k(y) \, dy
\end{equation}

Now consider the limit of (3.6) as $\min x_i \to \infty$. The numerator converges to $g(y) h(y)$ since it is assumed that $\lim_{\min x_i \to \infty} \frac{f(x - y)}{f(x)} = h(y)$. Now consider the denominator. Since

\[ g(y) \frac{f(x - y)}{f(x)} k(y) \leq g(y) k(y) \]

and

\[ \int_{\mathbb{R}^n} g(y) k(y) \, dy < \infty, \]

the Lebesgue dominated convergence theorem may be used to move the limit inside the integral. Thus the denominator of (3.6) converges to $\int_{\mathbb{R}^n} g(y) h(y) \, dy$ as $\min x_i \to \infty$. Thus the densities converge to
the proper density (3.5), so by Scheffe's theorem the convergence is complete.

Corollary 3.2.1: Let \( m: \mathbb{R}^n \rightarrow \mathbb{R} \). Suppose conditions (1) and (2) of theorem 3.2 hold. Then

\[
\int_{\mathbb{R}^n} |m(y)| g(y) k(y) \, dy < \infty
\]

implies \( E(m(D)|X = x) \rightarrow E(Z) \) as \( \min x_i \rightarrow \infty \).

Proof: Again the density for \( D|X = x \) is given by (3.6). Thus \( E(m(D)|X = x) \) can be written as

\[
E\left( \frac{\int_{\mathbb{R}^n} m(y) g(y) f(x - y)/f(x) \, dy}{\int_{\mathbb{R}^n} g(y) f(x - y)/f(x) \, dy} \right)
\]

Again, the denominator of (3.7) converges to \( \int_{\mathbb{R}^n} g(y) h(y) \, dy \).

Thus all that remains to show is that the numerator converges to \( \int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy \). But since

\[
m(y) g(y) f(x - y)/f(x) \leq |m(y)| g(y) k(y)
\]

and

\[
\int_{\mathbb{R}^n} |m(y)| g(y) k(y) \, dy < \infty,
\]
the Lebesgue dominated convergence theorem can be used to move the limit inside the integral. Thus the numerator of (3.7) converges to \( \int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy \) as \( \min x_i \to \infty \). So \( E(m(D) | X = x) \) converges to

\[
\frac{\int_{\mathbb{R}^n} m(y) g(y) h(y) \, dy}{\int_{\mathbb{R}^n} g(y) h(y) \, dy} = E(m(Z)) .
\]

Again note \( E(|m(Z)|) < \infty \).

Corollary 3.2.2. Suppose conditions (1) and (2) of theorem 3.2 hold. Suppose also that

\[
\int_{\mathbb{R}^n} |y_i| g(y) k(y) \, dy < \infty \quad \text{for } i = 1, 2, \ldots, n.
\]

Then \( x - E(\theta | X = x) \to E(Z) \) as \( \min x_i \to \infty \).

Proof: Define \( m_i : \mathbb{R}^n \to \mathbb{R} \) by \( m_i(d) = d_i \) for \( i = 1, 2, \ldots, n \). Then

\[
\int_{\mathbb{R}^n} |m_i(y)| k(y) g(y) \, dy
\]

\[
= \int_{\mathbb{R}^n} |y_i| k(y) g(y) \, dy < \infty \quad \text{for } i = 1, 2, \ldots, n.
\]

Thus by corollary 3.2.1, \( E(m_i(D) | X = x) \to E(m_i(Z)) \) as \( \min x_i \to \infty \). But
\[ E(m(D) \mid X = x) = E(D \mid X = x) \]

\[ = E(X_i - \theta_i \mid X = x) \]

\[ = \theta_i - E(\theta_i \mid X = x) \text{ for } i = 1, 2, \ldots, n. \]

But since \( E(m(Z)) = E(Z_i) \) for \( i = 1, 2, \ldots, n \), we get
\[ x - E(\theta \mid X = x) \rightarrow E(Z) \text{ coordinatewise as } \min x_i \rightarrow \infty. \]
Again note that \( E(|Z_i|) < \infty \) for \( i = 1, 2, \ldots, n \).

The next theorem and corresponding corollaries consider the behavior of the posterior as one coordinate of \( x \), say \( x_i \), goes off to infinity. The results involve the marginal distributions of the prior density, \( f \). Thus to avoid some difficulties that arise in this area when \( f \) is improper, the assumption is made that \( f \) is a proper prior.

The following notation is used. Let \( f(\theta) \) be a prior density over \( \mathbb{R}^n \). Let \( f_i(\theta_i) \) represent the marginal density of \( \theta_i \). The conditional density of \( \theta_i \mid \theta_i = \hat{\theta}_i \) is written as \( f_i(\theta_i \mid \hat{\theta}_i) \).

(Recall, \( \hat{\theta}_i \in \mathbb{R}^{n-1} \) is the vector \( \theta \) with the \( i \)th coordinate removed.) Let \( \hat{f}_i(\hat{\theta}_i) \) represent the marginal density of \( \hat{\theta}_i \).

**Theorem 3.3:** Define \( k(x_i, y) = \sup_{x_i} f(x - y) / f(x) \). Suppose \( x_i \in \mathbb{R}^{n-1} \) is such that the following conditions hold,

\[ \frac{f_i(x_i - y_i \mid \hat{x}_i - \hat{y}_i)}{\hat{f}_i(x_i \mid \hat{x}_i)} \rightarrow h(y) > 0 \text{ a.e. as } x_i \rightarrow \infty \]
and

\[ \int_{\mathbb{R}^n} k(x_i, y) g(y) \, dy < \infty. \]

Then \( D|X = x \overset{Law}{\rightarrow} Z(x_i) \) as \( x_i \to \infty \). Here \( Z(x_i) \) is a random vector with a density given by

\[ \frac{\gamma_i(x_i - z) h(z) g(z)}{\int_{\mathbb{R}^n} \gamma_i(x_i - z) h(z) g(z) \, dz} \] for \( z \in \mathbb{R}^n \).

Proof: The theorem is actually a corollary of theorem 3.1. Here \( y(t) \) has \( t \) for the \( i \)th coordinate and \( x_i \) for the other coordinate. Also

\[ \frac{f(y(t) - y)}{f(y(t))} + \frac{\gamma_i(x_i - y_i)}{\gamma_i(x_i)} h(y) = h_Y(y), \]

and \( Z(x_i) \overset{d}{=} Z \) since \( Z \) has a density

\[ \frac{g(y) h_Y(y)}{\int_{\mathbb{R}^n} g(y) h_Y(y) \, dy} \cdot \frac{\gamma_i(x_i - y_i)}{\gamma_i(x_i)} h(y) \]

\[ = \frac{\int_{\mathbb{R}^n} g(y) \frac{\gamma_i(x_i - y_i)}{\gamma_i(x_i)} h(y) \, dy}{\int_{\mathbb{R}^n} g(y) \frac{\gamma_i(x_i - y_i)}{\gamma_i(x_i)} h(y) \, dy} \]
The same calculations show that the next two corollaries are special cases of corollaries 3.1.1 and 3.1.2.

Corollary 3.3.1: Suppose conditions (1) and (2) of theorem 3.3 hold. If

\[ \int_\mathbb{R}^n |m(y)| g(y) k(x, y) \, dy < \infty \]

then \( E(m(D)|X = x) \rightarrow E(m(Z(x))) \) as \( x \rightarrow \infty \).

Corollary 3.3.2: Suppose that conditions (1) and (2) of theorem 3.3 hold. Suppose also that

\[ \int_\mathbb{R}^n |y_j| g(y) k(x, y) \, dy < \infty \] for \( j = 1, 2, \ldots, n \).

Then \( x \rightarrow E(\theta|X = x) \rightarrow E(Z(x)) \) as \( x \rightarrow \infty \).

Some densities have the property that \( f(x - y)/f(x) \rightarrow h(y) \) as \( ||x|| \rightarrow \infty \). These are considered in the next theorem and corollaries. However, in (Hill, 1974) it is shown that quite generally if a limiting distribution for \( D|X = x \) exists as simply \( ||x|| \rightarrow \infty \), then that
limiting distribution has the distribution of the error vector $D$.

In this work, we give general conditions for a limiting distribution to exist, whereas Hill works with the assumption that a limiting distribution does exist, making some remarks as to when one will exist. Now $D'X = x + D$ corresponds to the case where $f(x - y)/f(x) \to h(y) = 1$ as $||x|| \to \infty$. One assumption that Hill uses is that

$$\lim_{|C| \to \infty} f(C\theta) = 0 \text{ for almost all } \theta.$$ 

This is certainly not a strong assumption; however, it is not a necessary assumption in the following theorems. With this in mind, we write the next theorems with $f(x - y)/f(x) \to h(y)$ as $||x|| \to \infty$, even though we have not shown the existence of a density, $f$, with $f(x - y)/f(x)$ converging to anything other than the constant 1 as $||x|| \to \infty$.

Also, the proofs of the next theorem and corollaries are omitted. The proofs follow exactly like theorem 3.2 and corollaries 3.2.1 and 3.2.2 with the limits taken as $||x|| \to \infty$ instead of as $\min x_i \to \infty$.

Also, the assumption that $f$ be proper is relaxed; $f$ may be proper or improper for the remainder of the chapter.

**Theorem 3.4:** Again, let $k(y) = \sup_x f(x - y)/f(x)$. Suppose the following two conditions hold,

1. $f(x - y)/f(x) \to h(y) > 0 \text{ a.e. as } ||x|| \to \infty$,

and

2. $\int_{\mathbb{R}^n} g(y) k(y) \, dy < \infty$,
Then $D|X = x \xrightarrow{\text{Law}} Z$ as $||x|| \to \infty$. Here again $Z$ has the density (3.5).

Corollary 3.4.1: Let $m: \mathbb{R}^n \to \mathbb{R}$. Suppose conditions (1) and (2) of theorem 3.4 hold. Then

$$\int_{\mathbb{R}^n} |m(y)||g(y)|k(y)\,dy < \infty$$

implies $E(m(D)|X = x) \to E(Z)$ as $||x|| \to \infty$.

Corollary 3.4.2: Suppose conditions (1) and (2) of theorem 3.4 hold. Suppose also that

$$\int_{\mathbb{R}^n} |y_i||g(y)|k(y)\,dy < \infty \quad \text{for} \quad i = 1, 2, \ldots, n.$$

Then $x - E(g|X = x) \to E(Z)$ as $||x|| \to \infty$.

An important density with the property that $f(x - y)/f(x) \to 1$ as $||x|| \to \infty$ is the multivariate t distribution as in (DeGroot, 1970). Here for $\mu \in \mathbb{R}^n$, and $k$ a positive integer, and $L$ a positive definite matrix, the density $f(x)$ is proportional to

$$[1 + \frac{1}{k} (x - \mu)^T L(x - \mu)]^{-\frac{k+m}{2}}$$

for $x \in \mathbb{R}^n$.

Densities with $f_i(x_i - y_i)/f_i(x_i)$ converging to $h(y)$ are also available. If $f(x) = f_i(x_i) \frac{h_i(y_i)}{f_i(x_i)}$, then all that is required is that $f_i(x_i - y_i)/f_i(x_i) \to h(y_i)$. For example,
if \( f_1(x_1) \alpha \exp\{-\lambda_1|x_1|\} \), then \( f_1(x_1 - y_1)/f_1(x_1) \) converges to \( \exp\{\lambda_1y_1\} \) as \( x_1 \to \infty \). Another, somewhat more artificial, family of densities in this class can be defined conditionally. For example, suppose \( n = 2 \) and

\[
f(x) \alpha \left(1 + \frac{1}{k_1} (x_1 - x_2)^2\right)^{\frac{k_1+1}{2}} \left(1 + \frac{1}{k_2} x_2^2\right)^{\frac{k_2+1}{2}}
\]

for \( k_1 \) and \( k_2 \) positive integers. Then

\[
f_1(x_1 - y_1|x_2 - y_2)/f_1(x_1|x_2) \to 1
\]

as \( x_1 \to \infty \). Here \( f_2(x_2) \alpha \left(1 + \frac{1}{k_2} x_2^2\right)^{-\frac{k_2+1}{2}} \).

It is noted in the introduction that interchanging the properties assumed for \( f \) and \( g \) can produce drastically different behavior of the posterior. Theorems to cover these cases are presented here. The proof of the first theorem is included; the rest are omitted, their proofs following clearly established patterns.

**Theorem 3.5:** Define \( \ell_\gamma(y) = \sup_t g(\gamma(t) - y)/g(\gamma(t)) \). Suppose the following conditions hold.

1. \( g(\gamma(t) - y)/g(\gamma(t)) \to \Gamma_\gamma(y) > 0 \) a.e. as \( t \to \infty \)

and
Then $\theta \mid X = \gamma(t) \xrightarrow{\text{Law}} W_\gamma$ as $t \to \infty$. Here $W_\gamma$ is a random vector with a density given by

\[(3.8) \quad f(y) \frac{\Gamma_Y(y)}{\int_{\mathbb{R}^n} f(y) \Gamma_Y(y) \, dy}.\]

**Proof:** The density for $\theta \mid X = \gamma(t)$ can be written as

\[(3.9) \quad f(y) \frac{g(\gamma(t) - y)}{\int_{\mathbb{R}^n} f(y) \frac{g(\gamma(t) - y)}{g(\gamma(t))} \, dy}.\]

Now since (1) holds there exists $N < \infty$ such that $t > N$ implies $g(\gamma(t)) > 0$. So for $t > N$ (3.9) can be rewritten as

\[(3.10) \quad \frac{f(y) \frac{g(\gamma(t) - y)}{g(\gamma(t))}}{\int_{\mathbb{R}^n} f(y) \frac{g(\gamma(t) - y)}{g(\gamma(t))} \, dy}.\]

Now as $t \to \infty$ the numerator converges to $f(y) \frac{\Gamma_Y(y)}{\Gamma_Y(y)}$ by assumption (1). Now consider the denominator. Since

\[f(y) \frac{g(\gamma(t) - y)}{g(\gamma(t))} \leq f(y) \frac{\Gamma_Y(y)}{\Gamma_Y(y)}\]

and

\[\int_{\mathbb{R}^n} f(y) \frac{\Gamma_Y(y)}{\Gamma_Y(y)} \, dy < \infty,\]
the Lebesgue dominated convergence theorem can be used to move the
limit inside the integral. Thus the denominator of (3.10) converges
to \( \int_{\mathbb{R}^n} f(y) \gamma(y) \, dy \). But again \( \gamma(y) > 0 \) a.e. implies
\( \int_{\mathbb{R}^n} f(y) \gamma(y) \, dy > 0 \), and \( \gamma(y) \leq \lambda(y) \) implies
\[
\int_{\mathbb{R}^n} f(y) \gamma(y) \, dy < \infty
\]
by assumption (2). Thus (3.10) converges to the proper density
(3.8) as \( t \to \infty \), so by Scheffe's theorem the convergence is complete.

Corollary 3.5.1: Let \( m: \mathbb{R}^n \to \mathbb{R} \). Suppose conditions (1) and (2)
of theorem 3.5 hold. Then
\[
\int_{\mathbb{R}^n} |m(y)| \lambda(y) \, dy < \infty
\]
implies \( E(m(W)) = \gamma(t) \to E(m(W)) \) as \( t \to \infty \).

Corollary 3.5.2: Suppose conditions (1) and (2) of theorem 3.5
hold. Suppose also that
\[
\int_{\mathbb{R}^n} |y_i| f(y) \lambda(y) \, dy < \infty \quad \text{for } i = 1, 2, \ldots, n.
\]
Then \( E(W) = \gamma(t) \to E(W) \) as \( t \to \infty \).

**Theorem 3.6:** Define \( \lambda(y) = \sup_{x} g(x - y)/g(x) \). Suppose the following
two conditions hold,

\[(1) \quad \frac{g(x - y)}{g(x)} \rightarrow r(y) > 0 \text{ a.e. as } \min x_i \rightarrow \infty,\]

and

\[(2) \quad \int_{\mathbb{R}^n} f(y) g(y) \, dy < \infty.\]

Then \( g|X = x \rightarrow W \) as \( \min x_i \rightarrow \infty \). Here \( W \) is a random vector with a density given by

\[(3.11) \quad f(y) g(y) = \int_{\mathbb{R}^n} f(y) g(y) \, dy.\]

**Corollary 3.6.1:** Let \( m: \mathbb{R}^n \rightarrow \mathbb{R} \). Suppose conditions (1) and (2) of theorem 3.6 hold. Then

\[\int_{\mathbb{R}^n} |m(y)| f(y) g(y) \, dy < \infty\]

implies \( E(m(\Theta|X = x)) \rightarrow E(m(W)) \) as \( \min x_i \rightarrow \infty \).

**Corollary 3.6.2:** Suppose conditions (1) and (2) of theorem 3.6 hold. Suppose also that

\[\int_{\mathbb{R}^n} |y_i| f(y) g(y) \, dy < \infty \quad \text{for} \quad i = 1, 2, \ldots, n.\]

Then \( E(\Theta|X = x) \rightarrow E(W) \) as \( \min x_i \rightarrow \infty \).
Theorem 3.7: Define \( \lambda(x_1, y) = \sup_{x_1} g(x - y)/g(x) \). Suppose \( x_1 \in \mathbb{R}^{n-1} \) is such that the following conditions hold,

\[
(1) \quad \frac{g_i(x_1 - y_1 | x_1 - y_1)}{g_i(x_1 | x_1)} + r(y) > 0 \text{ a.e. as } x_1 \to \infty,
\]

and

\[
(2) \quad \int_{\mathbb{R}^n} \lambda(x_1, y) f(y) \, dy < \infty.
\]

Then \( x \xrightarrow{\text{Law}} W(x_1) \) as \( x_1 \to \infty \). Here \( W(x_1) \) is a random vector with density

\[
\frac{g_i(x_1 - y_1) \Gamma(y) f(y)}{\int_{\mathbb{R}^n} g_i(x_1 - y_1) \Gamma(y) f(y) \, dy}.
\]

Corollary 3.7.1: Let \( m: \mathbb{R}^2 \to \mathbb{R} \). Suppose conditions (1) and (2) of theorem 3.7 hold. Then \( \int_{\mathbb{R}^n} |m(y)| f(y) \lambda(x_1, y) \, dy < \infty \) implies \( E(m(\theta) | X = x) \to E(m(W(x_1))) \) as \( x_1 \to \infty \).

Corollary 3.7.2: Suppose conditions (1) and (2) of theorem 3.7 hold. Suppose also that \( \int_{\mathbb{R}^n} |y_j| f(y) \Gamma(x_1, y) \, dy < \infty \) for \( j = 1, 2, \ldots, n \). Then \( E(\theta | X = x) \to E(W(x_1)) \) as \( x_1 \to \infty \).

Theorem 3.8: Again let \( \lambda(y) = \sup_{x} g(x - y)/g(x) \). Suppose the following two conditions hold,
(1) \[ g(x - y)/g(x) + f(y) > 0 \text{ a.e. as } ||x|| \to \infty, \]

and

(2) \[ \int_{\mathbb{R}^n} f(y) \ell(y) \, dy < \infty. \]

Then \( 0|X = x \stackrel{\text{Law}}{\Rightarrow} W \) as \( ||x|| \to \infty \). Here again \( W \) has the density (3.11).

**Corollary 3.8.1:** Let \( m: \mathbb{R}^n \to \mathbb{R} \). Suppose conditions (1) and (2) of theorem 3.8 hold. Then \( \int_{\mathbb{R}^n} |m(y)|f(y) \ell(y) \, dy < \infty \) implies \( E(m(\theta)|X = x) \to E(m(W)) \) as \( ||x|| \to \infty \).

**Corollary 3.8.2:** Suppose conditions (1) and (2) of theorem 3.8 hold. Suppose also that \( \int_{\mathbb{R}^n} |y_i|f(y) \ell(y) \, dy < \infty \) for \( i = 1, 2, \ldots, n \). Then \( E(\theta|X = x) \to E(W) \) as \( ||x|| \to \infty \).

**B. Discussion**

Densities, \( f \), with the property that \( f(x - y)/f(x) \to 1 \) are generally flat in the tails. If \( f(x - y)/f(x) + h(y), h(y) \not\equiv 1 \), then \( f \) is a little sharper. For example, the student's t distribution has a density, \( f \), such that \( f(x - y)/f(x) \to 1 \) as \( x \to \infty \).

If a sharper density is considered, say \( f(x) \propto \exp \{-\lambda |x|\} \), then \( f(x - y)/f(x) \to \exp \{\lambda y\} \). Sharper densities than \( \exp \{-\lambda |x|\} \) lead to difficulties. For example, a normal density with mean 0 and variance 1 has
\[
f(x - y)/f(x) \rightarrow \begin{cases} 
0 & \text{for } y < 0 \\
1 & \text{for } y = 0 \\
\infty & \text{for } y > 0 
\end{cases}
as x \to \infty
\]

Functions of regular variation, as in (Feller, 1966), are closely related to densities with \( f(x - y)/f(x) \to h(y) \). The possible limits can be determined by regular variation techniques. The following lemma is a combination of a lemma and a footnote on pages 268-269 of (Feller, 1966).

Lemma 3.1: Let \( U \) be a positive function on \((0, \infty)\) such that \( U(tx) = U(t) \) for each \( x \) as \( t \to \infty \). If \( 0 < \psi(x) < \infty \) then
\[
\psi(x) = x^f \text{ for some } f \text{ with } \|f\| < \infty.
\]

Now suppose \( f(x - y)/f(x) \to h(y) \) as \( x \to \infty \). Suppose also that \( 0 < h(y) \to 0 \) as \( y \to 0 \). This implies that
\[
f \log(e^x) = h(y)
f \log(e^{e^{-y}}) \to h(y)
\]
as \( x \to \infty \), or as \( e^x \to \infty \). Let \( t = e^x \) and \( w = e^{-y} \), then
\[
\frac{f \log(tw)}{f \log(t)} \to h(y) = h(-\log w).
\]

But \( 0 < h(-\log w) < \infty \), so by lemma 3.1, \( h(-\log w) = e^f \) for \( -\infty < f < \infty \).

But this in turn implies that \( h(y) = (e^{-y})^f = e^{-f y} \). Thus the possible
limit functions \( h(y) \) are given by \( e^{f(y)} \) for \(-\infty < f < \infty\).

An important case to study is when the prior, \( f \), has the property \( f(x - y)/f(x) \rightarrow 1 \) as \( x \) gets "large". This corresponds to the case where \( f(x - y)/f(x) \rightarrow 1 \) as \( x \) gets "large". For these densities the conclusions of all the theorems and corollaries can be presented in a much simpler manner. For example, the conclusion of theorem 3.1, for the case \( h(y) \equiv 1 \), is actually \( D|X = \gamma(t) \xrightarrow{\text{Law}} D \) as \( t \rightarrow \infty \). In the theorem the random variable \( Z_\gamma \) is introduced as the limit. \( Z_\gamma \) has a density, for \( h(y) \equiv 1 \), given by \( g(y)/f_\gamma (x) \xrightarrow{\text{R^n}} g(y) \ dy = g(y) \).

But \( g \) is the density for \( D \). In a similar fashion the error random vector \( D \) can be substituted for the random vector \( Z_\gamma \) in corollaries 3.1.1 and 3.1.2, and for the random vector \( Z \) in theorems 3.2 and 3.4 and corollaries 3.2.1, 3.2.2, 3.4.1, and 3.4.2. The random vector \( D|\hat{x}_1 = \hat{\gamma}_1 \) can be substituted for the random vector \( Z|\hat{x}_1 \) in corollaries 3.3.1 and 3.3.2, and \( Z \) in theorems 3.3.6 and 3.3.2. Similarly \( \Theta \) can be substituted for \( W \) in theorem 3.5 and corollaries 3.5.1 and 3.5.2, and \( W \) for \( V \) in theorems 3.6 and 3.8 and corollaries 3.6.1, 3.6.2, 3.8.1, and 3.8.2 for the case where \( g(x - y)/g(x) \rightarrow 1 \). Also \( \Theta|\hat{x}_1 = \hat{x}_1 \) can be substituted for \( W(\hat{x}_1) \) in theorem 3.7 and corollaries 3.7.1 and 3.7.2 when \( g(x - y)/g(x) \rightarrow 1 \).

Consider now estimation of \( \Theta \) with norm squared error loss. As was mentioned in the introduction, an unusually "large" observation can, in some cases, be thought of as discrediting the prior. If the prior is uniform enough so that \( f(x - y)/f(x) \rightarrow 1 \) as \( x \) gets "large".
then \( E(\theta | X = x) \), the Bayes estimate, will be close to the observed \( x \) under appropriate conditions.

In particular, consider estimation of a normal mean vector with variance-covariance matrix \( I \). Suppose the prior is a multivariate t with \( L = I \) and \( \mu = 0 \). Here \( f(x - y)/(x) \rightarrow 1 \) as \( ||x|| \rightarrow \infty \) and \( \sup_x f(x - y)/(x) \) behaves essentially like \( ||y||^P \), for some \( p > 0 \), for \( ||y|| \) large. Thus since \( g \) is a normal density \( \int_\mathbb{R}^n ||y||^P g(y) \, dy < \infty \) and \( \int_\mathbb{R}^i ||y||^P g(y) \, dy < \infty \) for \( i = 1, 2, \ldots, n \). Thus \( x - E(\theta | X = x) \rightarrow 0 \), the zero vector, as \( ||x|| \rightarrow \infty \). The prior is gentle enough in the tails to allow the posterior to take drastic revision with the large \( ||x|| \). Contrast this to the situation when a normal prior is used. Then

\[
\left| \left| x - E(\theta | X = x) \right| \right| \rightarrow \infty \quad \text{as} \quad \left| \left| x \right| \right| \rightarrow \infty .
\]

The situation in corollary 3.3.2 is more involved. Suppose

\[
f_i(x_i - y_i, x_i - y_i) \rightarrow 1 \quad \text{as} \quad x_i \rightarrow \infty .
\]

Suppose the hypotheses on \( f \) and \( g \) are satisfied. The conclusion, then, is \( x - E(\theta | X = x) \rightarrow E(D|X = x) \) as \( x_i \rightarrow \infty \). Consider estimation of \( \theta_i \). Here

\[
x_i - E(\theta_i | X = x) \rightarrow E(D_i | X_i = x_i) \quad \text{as} \quad x_i \rightarrow \infty .
\]

Consider estimation of \( \theta_i \). Again, if corollary 3.3.2 is satisfied, then \( \tilde{x}_i - E(\tilde{\theta}_i | X = x) \rightarrow E(D_i | X_i = \tilde{x}_i) = \tilde{x}_i - E(\tilde{\theta}_i | X_i = \tilde{x}_i) \) as \( x_i \rightarrow \infty \). Thus \( E(\tilde{\theta}_i | X = x) \rightarrow E(\tilde{\theta}_i | X_i = \tilde{x}_i) \) as \( x_i \rightarrow \infty \), so that the affect of the prior on \( E(\tilde{\theta}_i | X = x) \) remains. But the prior's affect
on $\mathbb{E}(\theta_1 | X = x)$ does not.

If $\mathbb{E}(D) = 0$ then $\theta = \mathbb{E}(X|\theta = \theta)$. However, the results also present desirable behavior when $\mathbb{E}(D) \neq 0$. In this case $\theta + \mathbb{E}(D) = \mathbb{E}(X|\theta = \theta)$. So in estimating $\theta$, one is estimating the mean of $(X|\theta = \theta) \sim D$. So that in the limiting case one might want the estimator for $\theta$, $\mathbb{E}(\theta | X = x)$ say, to look like $x - \mathbb{E}(D)$. Thus the property given in chapter I as desired by Beale and Lindley is actually that $x - \mathbb{E}(\theta | X = x) \rightarrow \mathbb{E}(D)$ as $x$ gets "large".

In summary, if the prior is gentle enough to satisfy theorems 3.1, 3.2, 3.3, or 3.4 with $h(y) \equiv 1$, then with suitable error densities, the affect of the "large" observation is to wash out the prior's affect on the Bayes estimate. The posterior becomes asymptotically fiducial.

If $h(y) \neq 1$ then $\sup_x f(x - y)/f(x) = k(y)$ will be fairly large in general. Thus the condition on $g$ that $\int g(y) k(y) \, dy < \infty$ will be fairly strong. Indeed, for the one dimensional case, since $h(y) = e^{\epsilon y}$ for some $\epsilon (-\infty, \infty)$ and $k(y) > h(y)$, a necessary condition for $\int k(y) g(y) \, dy < \infty$ is that $g$ have a moment generating function at $f$. However, some interesting cases can be addressed.

For convenience, consider the one dimensional case. Let $g(x) = \frac{1}{2} \exp \left\{ - \frac{1}{2} |x| \right\}$ for $x \in \mathbb{R}$ and $\lambda_1 > 0$. Suppose $f(x) = \frac{2}{\lambda_2} \exp \left\{ - \frac{1}{2} |x| \right\}$ for $x \in \mathbb{R}$ and $\lambda_2 > 0$. Now if $\lambda_1 = \lambda_2$, it is easy to show that $\mathbb{E}(\theta | X = x) = x/2$.

Indeed, this is the case whenever the prior and error distributions are identical. Provided, of course, that $\mathbb{E}(\theta | X = x)$ exists. To
see this consider

\[ x = \mathbb{E}(X | X = x) \]

\[ = \mathbb{E}(\theta + D | X = x) \]

\[ = \mathbb{E}(\theta | X = x) + \mathbb{E}(D | X = x). \]

But \( f = g \) implies the distribution of \( \theta | X = x \) is the same as the distribution of \( D | X = x \). Thus \( \mathbb{E}(\theta | X = x) = \mathbb{E}(D | X = x) \). Thus \( x = 2\mathbb{E}(\theta | X = x) \), or \( \mathbb{E}(\theta | X = x) = \frac{x}{2} \).

If \( \lambda_1 > \lambda_2 \) then the error density is sharper. Here

\[ \sup_x f(x - y)/f(x) = \exp \{\lambda_2 |y|\} \quad \text{and} \]

\[ \int_{\mathbb{R}} e^{\lambda_2 |y|} g(y) \, dy = \int_{\mathbb{R}} \exp \{\lambda_2 |y| - \lambda_1 |y|\} \frac{\lambda_1}{2} \, dy \]

\[ = \frac{\lambda_1}{2} \int_{\mathbb{R}} \exp \{-\lambda_1 - \lambda_2\} |y| \, dy < \infty. \]

Also

\[ \int_{\mathbb{R}} |y| e^{\lambda_2 |y|} g(y) \, dy = \frac{\lambda_1}{2} \int_{\mathbb{R}} |y| \exp \{-\lambda_1 - \lambda_2\} |y| \, dy < \infty. \]

So by corollary 3.2.2 \( x - \mathbb{E}(\theta | X = x) \rightarrow \mathbb{E}(Z) \) where \( Z \) has a density proportional to \( \exp \{-\lambda_1 |z| + \lambda_2 |z|\} \). But \( \mathbb{E}(Z) = \frac{2\lambda_1}{\lambda_1 - \lambda_2} \).
Thus \[ \lim_{x \to \infty} \{ x - E(\theta | X = x) \} = \frac{2\lambda_2}{\lambda_1 - \lambda_2}. \]

However, if the prior is sharper, \( \lambda_2 > \lambda_1 \), then corollary 3.6.2 can be used to deduce \( E(\theta | X = x) \to E(W) \), where \( W \) has a density proportional to \( \exp \{ -\lambda_2 |w| \} \), and thus \( E(W) = \frac{2\lambda_1}{\lambda_2 - \lambda_1} \). On page 331 of (Lehmann, 1959) it is shown that densities proportional to \( \exp \{ -\lambda |x - \theta| \} \) have the monotone likelihood ratio property in \( x \) for each fixed \( \lambda \). Thus by (Karlin and Rubin, 1955), \( E(\theta | X = x) \) is monotonically increasing. Thus, by symmetry, \( E(\theta | X = x) \) goes monotonically from \( \frac{-2\lambda_1}{\lambda_2 - \lambda_1} \) at \( x = -\infty \) to \( \frac{2\lambda_1}{\lambda_2 - \lambda_1} \) at \( x = \infty \).

The behavior of the posterior was drastically altered by a rather slight change in the prior.

This indicates that a Bayesian should choose his prior with care, since slight changes in the prior can lead to drastic changes in the posterior for a large observation. Also, it seems that one's prior knowledge is usually sharpest near the "center" of the prior and more vague in the tails. But it is the tails of the prior and error distributions which determine the limiting behavior of the posterior. Thus one should take care so that the tails of the prior are not unduly sharp. The resulting estimator will then have what might be considered more desirable limiting properties.
IV. VARYING THE LOSS FUNCTION

A. Main Results

The main emphasis up until now has been on the posterior mean, the Bayes estimate under norm-squared error loss. The theorems presented in this chapter are designed to extend the results obtained under norm-squared error to some other loss functions. In many cases the limiting behavior is seen to be the same. In general, we consider loss functions which are convex functions of absolute error.

Throughout this chapter it is assumed that \( W : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) is a strictly increasing convex function. Also, \( W \) is assumed differentiable on \((0, \infty)\). Thus we can define \( W'(0) = \lim_{x \to 0} W'(x) \) since \( W' \) is non-decreasing (by convexity) and \( W'(x) > 0 \) for \( x > 0 \) since \( W \) is strictly increasing.

Let \( h(\theta | x) \) represent the posterior density for \((\theta | x)\).

If the loss function is given by \( L(\theta, a) = W(|\theta - a|) \), for one dimension, then any Bayes estimate for \( \theta, \delta^*(x) \), satisfies the following equation:

\[
\delta^*(x) = \frac{\int_{-\infty}^{\infty} W'(\delta^*(x) - \theta) h(\theta | x) d\theta}{\int_{-\infty}^{\infty} W'(\delta^*(x) - \theta) h(\theta | x) d\theta} - \delta(x).
\]

See (DeGroot and Rao, 1963) for details.

Quite generally, it is true that if \( \delta(x) \) is a function such that
\[
\lim_{x \to \infty} \{ \int_{-\infty}^{\infty} w'(\delta(x) - \theta) h(\theta|x) \, d\theta - \int_{\delta(x)}^{\infty} w'(\theta - \delta(x)) h(\theta|x) \, d\theta \} = 0,
\]
then
\[
\lim_{x \to \infty} \{ \delta(x) - \delta^*(x) \} = 0.
\]

This is the key idea behind the following results. To extend the results of chapters II and III one may take \( \delta(x) = E(\theta|X = x) \) or \( \delta(x) = x \).

**Theorem 4.1:** Let \( a \in \mathbb{R} \). Define

\[
\gamma_a(x) = \int_{-\infty}^{\infty} w'(\theta - a) h(\theta|x) \, d\theta - \int_{-\infty}^{a} w'(a - \theta) h(\theta|x) \, d\theta.
\]

If \( \gamma_a(x) \to 0 \) as \( x \to \infty \) and given \( \varepsilon > 0 \), there exists \( N < \infty \) and \( \delta > 0 \) such that

\[
\int_{a}^{a+\varepsilon} w'(\theta - a) h(\theta|x) \, d\theta \geq \delta
\]
and

\[
\int_{a-\varepsilon}^{a} w'(a - \theta) h(\theta|x) \, d\theta \geq \delta
\]
for \( x > N \), then \( \lim_{x \to \infty} \delta^*(x) = a \).

**Proof:** Assume \( \delta^*(x) \) does not converge to \( a \). Then there exists \( \{x_n\}_{n=1}^{\infty} \) with \( \lim_{n \to \infty} x_n = \infty \) such that either

\[
\sum_{n=1}^{\infty} x_n = \infty
\]
(a) $\delta^*(x_n) \geq a + \epsilon, \epsilon > 0$, or

(b) $\delta^*(x_n) \leq a - \epsilon, \epsilon > 0$.

Assume (a). By assumption $\lim_{n \to \infty} \gamma_{a(x_n)} = 0$. However,

$$
\gamma_{a(x_n)} = \int_{-\infty}^{\infty} w'(\theta - a) h(\theta|x_n) \, d\theta - \int_{-\infty}^{a} w'(a - \theta) h(\theta|x_n) \, d\theta
$$

$$
= \int_{-\infty}^{\infty} w'(\theta - a) h(\theta|x_n) \, d\theta - \int_{-\infty}^{a} w'(a - \theta) h(\theta|x_n) \, d\theta
$$

$$
- \int_{\delta(x_n)}^{\infty} w'(*) - \delta^*(x_n) h(\theta|x_n) \, d\theta
$$

$$
\delta^*(x_n) = \int_{-\infty}^{\infty} w'(\theta - a) h(\theta|x_n) \, d\theta
$$

$$
+ \int_{-\infty}^{\infty} w'(** - \delta^*(x_n) \quad h(\theta|x_n) \, d\theta
$$

$$
\delta^*(x_n) = \int_{-\infty}^{a} w'(\theta - a) h(\theta|x_n) \, d\theta
$$

$$
+ \int_{a}^{\infty} w'(\theta - a) h(\theta|x_n) \, d\theta
$$

$$
- \int_{-\infty}^{-\infty} w'(a - \theta) h(\theta|x_n) \, d\theta
$$
Consider (4.3). Since \( a < \delta(x_n) < \theta \) and \( w' \) is non-decreasing, 
\[ w'(\theta - a) \geq w'(\theta - \delta(x_n)) \] 
so (4.3) \( \geq 0 \). Similarly, for \( \theta < a \) 
\[ w'(\delta(x_n) - \theta) \geq w'(a - \theta) \] 
so (4.4) \( \geq 0 \). Also for 
\( a < \delta(x_n) \), 
\[ w'(\delta(x_n) - a) \geq 0 \] 
since \( w' \) is non-decreasing. Thus (4.6) \( \geq 0 \). Now since \( \delta(x_n) \geq a + \varepsilon \), (4.5) is greater than or equal to
\[(4.7) \quad \int_{a}^{a+\epsilon} w'(\theta - a) h(\theta|\mathbf{x}_n) \, d\theta \]

But for \(x_n \geq N\), \((4.7) \geq \delta > 0\). So the limit as \(n \to \infty\) of \((4.7)\) is \(\geq \delta\). Thus \(\lim_{n \to \infty} \gamma_{a|n} \geq \delta > 0\). Thus the contradiction. So assumption (a) cannot hold. Assumption (b) is handled similarly, and thus theorem is proved.

Corollaries 4.1.1 and 4.1.2 present some conditions under which theorem 4.1 will hold.

Corollary 4.1.1: Suppose \(h(\theta|x) \to h^\#(\theta)\), where \(h^\#\) is a proper density, as \(x \to \infty\). Also assume that \(h^\#\) is symmetric about \(a\) with \(\int_{a}^{a+\epsilon} h^\#(\theta) \, d\theta > 0\) and \(\int_{a-E}^{a} h^\#(\theta) \, d\theta > 0\) for each \(\epsilon > 0\). Then if there exists \(p(\theta)\) and \(N < \infty\) such that \(h(\theta|x) \leq p(\theta)\) for \(x > N\) and \(\int_{-\infty}^{\infty} w'(|\theta - a|) p(\theta) \, d\theta < \infty\) then \(\lim_{x \to \infty} \delta(x) = a\).

Proof: \(\gamma_a(x) = \int_{a}^{\infty} w'(\theta - a) h(\theta|x) \, d\theta \)

\[- \int_{-\infty}^{a} w'(a - \theta) h(\theta|x) \, d\theta \, .\]

However,

\[\lim_{x \to \infty} \gamma_a(x) = \int_{a}^{\infty} w'(\theta - a) h^\#(\theta) \, d\theta \, - \int_{-\infty}^{a} w'(a - \theta) h^\#(\theta) \, d\theta \, .\]
Moving the limit inside the integral is justified by the Lebesgue dominated convergence theorem, since for $x > N$ we have

\[ w'(\theta - a) \frac{h(x)}{\theta} \leq w'(\theta - a) p(\theta), \text{ for } \theta > a \]

and

\[ w'(a - \theta) \frac{h(x)}{\theta} \leq w'(a - \theta) p(\theta), \text{ for } \theta < a \]

and

\[ \int_a^\infty w'(\theta - a) p(\theta) \, d\theta < \infty \]

and

\[ \int_{-\infty}^a w'(a - \theta) p(\theta) \, d\theta < \infty. \]

But since $h'$ is symmetric about $a$, we see

\[ \int_{-\infty}^\infty w'(\theta - a) h'(\theta) \, d\theta = \int_a^a w'(a - \theta) h'(\theta) \, d\theta \]

Thus $\lim_{x \to \infty} \gamma_a(x) = 0$. So all that remains is to show that given $\varepsilon > 0$, there exists $N < \infty$ and $\delta > 0$ such that for $x > N$

\[ \int_a^{a+\varepsilon} w'(\theta - a) \frac{h(x)}{\theta} \, d\theta > \delta \]

and

\[ \int_{a-\varepsilon}^a w'(a - \theta) \frac{h(x)}{\theta} \, d\theta > \delta, \]

since then we can apply theorem 4.1. But we know
\[
\lim_{x \to a^+} \int w'(\theta - a) \, h(\theta | x) \, d\theta = \int w'(\theta - a) \, h'(\theta) \, d\theta
\]

by the Lebesgue dominated convergence theorem. But for \( a < \theta < a^+ \), we see \( w'(\theta - a) > 0 \) and by assumption \( \int h'(\theta) \, d\theta > 0 \), so

\[
\int_{a}^{a+\epsilon} w'(\theta - a) \, h'(\theta) \, d\theta = \delta_2 > 0.
\]

Thus there exists \( N < \infty \) such that for \( x \geq N \)

\[
\int_{a}^{a+\epsilon} w'(\theta - a) \, h(\theta | x) \, d\theta \geq \frac{\delta_2}{2} = \delta > 0.
\]

One can handle

\[
\int_{a-\epsilon}^{a} w'(a - \theta) \, h(\theta | x) \, d\theta
\]

in exactly the same manner. Thus we can apply theorem 4.1 and conclude

\[
\lim_{x \to \infty} \lambda(x) = a.
\]

Corollary 4.1.2: Let \( a \in \mathbb{R} \). Suppose there exists \( \lambda(x) \) with

\[
\lim_{x \to \infty} \lambda(x) = \infty,
\]

such that the density, \( q_x(y) \), for

\[
Y_x = \lambda(x) \{(\theta - a) | X = x\},
\]

converges to a proper density, \( q(y) \), as \( x \to \infty \). Suppose there exists \( p(y), C > 0 \), and \( N < \infty \) such that \( q_x(y) \leq p(y) \) for
\[ x \geq N \quad \text{and} \]

\[ \int_{-\infty}^{\infty} w'(C|y|) p(y) \, dy < \infty. \]

Then

(I) if \( w'(0) \neq 0 \) and \( q \) has a median at 0 then

\[ \lim_{x \to \infty} \delta^*(x) = a \]

and

(II) if \( w'(0) = 0 \) and \( \int_{-\infty}^{0} q(y) \, dy > 0 \) and \( \int_{-\infty}^{0} q(y) \, dy > 0 \)

then \( \lim_{x \to \infty} \delta^*(x) = a. \)

Proof: \( \gamma_a(x) = \int_{-\infty}^{\infty} w'(a - \theta) h(\theta|x) \, d\theta - \int_{a}^{\infty} w'(a - \theta) h(\theta|x) \, d\theta \)

\[ = \int_{0}^{\infty} w'(y/\lambda(x)) \frac{1}{\lambda(x)} h(y/\lambda(x) + a|x) \, dy \]

\[ - \int_{-\infty}^{0} w'(-y/\lambda(x)) \frac{1}{\lambda(x)} h(y/\lambda(x) + a|x) \, dy \]

\[ = \int_{0}^{\infty} w'(y/\lambda(x)) q_x(y) \, dy - \int_{-\infty}^{0} w'(-y/\lambda(x)) q_x(y) \, dy. \]

But since there exists \( N < \infty \) such that \( x \geq N \) implies \( q_x(y) \leq p(y) \) and \( 1/\lambda(x) \to 0 \) as \( x \to \infty \), for each \( C > 0 \) there exists \( N_1 < \infty \) such that \( x \geq N_1 \) implies \( w'(y/\lambda(x))q_x(y) \leq w'(Cy)p(y) \) for \( y \geq 0. \)
But
\[ \int_0^\infty w'(Cy) p(y) \, dy < \infty \]
so that the Lebesgue dominated convergence theorem can be applied to deduce that (4.8) converges to

\[ (4.9) \int_0^\infty w'(0) q(y) \, dy - \int_{-\infty}^0 w'(0) q(y) \, dy. \]

(We can handle \( \int w'(-y/\lambda(x)) q(x) \, dy \) similarly.) Thus \( \lim_{x \to \infty} \gamma_a(x) \) is equal to (4.9).

Now consider (I). Since \( q \) has a median at 0, we have

\[ \int_0^\infty q(y) \, dy - \int_{-\infty}^0 q(y) \, dy = 0. \]

Thus \( \gamma_a(x) \to 0 \) as \( x \to \infty \). Now let \( \varepsilon > 0 \) be given. For each \( K \) with \( 0 < K < \infty \), there exists \( N_2 < \infty \) such that \( \varepsilon \lambda(x) > K \) for \( x > N_2 \), since \( \lambda(x) \to \infty \) as \( x \to \infty \). Now pick \( K < \infty \) such that

\[ \int q(y) \, dy = \delta_1 > 0. \]

Thus

\[ \int_{a+\varepsilon}^a w'(\theta - a) h(\theta \mid x) \, d\theta \]

\[ = \int_0^{\varepsilon \lambda(x)} w'(y/\lambda(x)) q_x(y) \, dy \]
(4.10) \[ \int_{0}^{K} w'(y/x(x)) q_x(y) \, dy \]

for \( x \geq N \). But the Lebesgue dominated convergence theorem implies that (4.10) converges to

\[ \int_{0}^{K} w'(0) q(y) \, dy = w'(0) \delta_1 > 0. \]

Thus there exists \( N < \infty \) such that \( x \geq N \) implies

\[ \int_{a-\epsilon}^{a+\epsilon} w'(a - \theta) h(\theta \mid x) \, d\theta > \frac{w'(0)\delta_1}{2} = \delta > 0. \]

In a similar fashion one can show that there exists \( N < \infty \) such that

\[ \int_{a-\epsilon}^{a-\epsilon} w'(\theta - a) h(\theta \mid x) \, d\theta > \delta > 0 \]

for \( x \geq N \). Thus by theorem 4.1 we get \( \lim_{x \to \infty} \delta^*(x) = a \).

Consider (II). Since \( w'(0) = 0 \), we get by (4.9) that \( Y_n(x) \to 0 \) as \( x \to \infty \). To prove this proposition we will go back to theorem 4.1 and show that a different term of the sum (4.3) to (4.6) is bounded away from zero. In particular we will show that (4.6) \( \geq \delta > 0 \) for \( x \geq N \). Recall (4.6) is

\[ \int_{a}^{\delta^*(x_n)} w'(\delta^*(x_n) - \theta) h(\theta \mid x_n) \, d\theta. \]

The assumption has been made that \( \delta^*(x_n) \geq a + \epsilon \) for some
\( \varepsilon > 0 \) and all \( n \geq N \). However,

\[
\int_{\delta^*(x_n)}^{a+\varepsilon} w'(\delta^*(x_n) - \theta) h(\theta|x_n) \, d\theta
\]

\[
\geq \int_{a+\varepsilon}^{a+\varepsilon} w'(\delta^*(x_n) - \theta) h(\theta|x_n) \, d\theta
\]

\[ (4.11) \]

Now for each \( 0 < K < \alpha \), there exists \( N_3 \) such that \( \varepsilon\lambda(x_n) \geq K \) for \( n \geq N_3 \), since \( \lambda(x_n) \to \infty \) as \( n \to \infty \). Now pick \( K \) such that

\[
\int_0^K w'(\varepsilon) q(y) \, dy = \delta_3 > 0.
\]

Then we get (4.11) equals

\[
\int_0^\varepsilon \frac{\varepsilon\lambda(x_n)}{w'(\varepsilon - y/\lambda(x_n))} q_x(y) \, dy
\]

\[ (4.12) \]

for \( n \geq \) some \( N_3 < \infty \). But again the Lebesgue dominated convergence theorem implies that (4.12) converges to

\[
\int_0^K w'(\varepsilon) q(y) \, dy = \delta_3 > 0.
\]

Similarly one can handle

\[
\int_{\delta^*(x_n)}^{a} w'(\theta - \delta^*(x_n)) h(\theta|x_n) \, d\theta
\]

\[ \delta^*(x_n) \]
with the assumption that $\delta^*(x_n) < a - \varepsilon$. Thus as in theorem 4.1, the conclusion is that $\lim_{x \to \infty} \delta^*(x_n) = a$.

These results apply directly when the posterior is not drifting off to infinity. The more common case considered in this work is when $h(\theta|x)$ is a density centered at $b(x)$ and $b(x) \to \infty$ as $x \to \infty$. These cases are handled by theorem 4.2. Again $h(\theta|x)$ is the posterior density for $\Theta|X = x$. Suppose $(\Theta|X = x) \to b(x)$ converges in law as $x \to \infty$ to a random variable, say $Z$. Examples of this are throughout this work, as when $D|X = x \overset{\text{Law}}{\to} D$ as in chapter III. Recall the loss function can be written $L(\theta, a) = w(|\theta - a|)$, where $W$ is strictly increasing and convex. The Bayes estimate, $\delta^*(x)$, for $\theta$ satisfies (4.1). Define the random variable $Y = \theta - b(x)$. Then $Y|b(X) = b(x) = (\Theta|b(x) = b(x)) - b(x) = (\Theta|X = x) - b(x)$, with the assumption that $b(x)$ is strictly increasing. Thus the density for $Y|X = x$, say $q(y|x)$, is given by $q(y|x) = h(b(x) + y|x)$. Thus define $\delta^*(x)$ by the equation

$$
\delta^*(x) = \int_{-\infty}^{\infty} w'(\delta(x) - y) q(y|x) dy = \int_{\delta^*(x)}^{\infty} w'(y - \delta^*(x)) q(y|x) dy.
$$

With these assumptions, consider the following.

**Theorem 4.2:** $\delta^*(x) = \delta(x) - b(x)$.

**Proof:**

$$
\int_{\delta^*(x)}^{\infty} w'(y - \delta^*(x)) q(y|x) dy
$$
(4.13) \[ \int_{\delta(x)+b(x)}^{\infty} w'(\theta - (\delta(x) - b(x))) q(\theta - b(x)|x) \, d\theta \]

by the change of variable \( \theta = y + b(x) \). But \( q(\theta - b(x)|x) = h(b(x) + \theta - b(x)|x) = h(\theta|x) \). So that (4.13) equals

\[ \int_{\delta(x)+b(x)}^{\infty} w'(\theta - (\delta(x) + b(x))) h(\theta|x) \, d\theta. \]

Similarly,

\[ \int_{-\infty}^{\delta(x)+b(x)} w'(\delta(x) - y) q(y|x) \, dy \]

\[ = \int_{\delta(x)+b(x)}^{\infty} w'(\delta(x) - (\theta - b(x))) h(\theta|x) \, d\theta \]

\[ = \int_{\delta(x)+b(x)}^{\infty} w'((\delta(x) + b(x)) - \theta) h(\theta|x) \, d\theta \]

Thus \( \delta(x) \) satisfies

\[ \int_{\delta(x)+b(x)}^{\infty} w'(\theta - (\delta(x) + b(x))) h(\theta|x) \, d\theta \]

\[ = \int_{\delta(x)+b(x)}^{\infty} w'(\delta(x) + b(x), - \theta) h(\theta|x) \, d\theta. \]

Thus as in (DeGroot and Rao, 1963) we can choose \( \delta(x) \) such that

\[ \delta(x) + b(x) = \delta(x). \]

Thus \( \delta(x) + b(x) = \delta(x) \). Thus \( \delta(x) + b(x) = \delta(x) \) is equivalent to \( \delta(x) - b(x) = 0 \),

where \( \delta(x) \) is the Bayes estimate for \( \theta \).
B. Discussion

As an example, consider a situation where \( h(\theta|x) \) is asymptotically normal with mean \( b(x) \) and variance \( c > 0 \), as \( x \to \infty \). Conditions for behavior of this sort are given in (Meeden and Isaacson, 1976). The setting is the natural parameter space of the one-dimensional exponential family. Here \( \theta|X = x \) has a density given by

\[
\frac{\exp \{aX - \lambda(a)\}}{\int \exp \{tx - \lambda(t)\} dt},
\]

where \( \lambda \) is convex and strictly increasing on \([K, \infty)\). Also \( \lambda'(y) \to \infty \) as \( y \to \infty \). Assume \( \lambda''(y) + c > 0 \). Then \( \theta|X = x \) is asymptotically normal with mean \((\lambda')^{-1}(x)\) and variance \(1/c\) as \( x \to \infty \). It is also shown that \( E(\theta|X = x) - (\lambda')^{-1}(x) \to 0 \) as \( x \to \infty \). Let the density of \( \theta|X = x - (\lambda')^{-1}(x) \) be given by \( p_x(z) \) where \( z = \theta - (\lambda')^{-1}(x) \). By the properties assumed on \( \lambda \) it can be shown that

\[
\lim_{x \to \infty} \int_{-\infty}^{\infty} f(z) p_x(z) \, dz = \int_{-\infty}^{\infty} f(z) \frac{e^{-z^2/2}}{\sqrt{2\pi/c}} \, dz \quad \text{for } 0 \leq f \leq \infty
\]

Thus if \( W \) is at all reasonably well behaved \((W'(x) \geq x^\Gamma \) for some \( \Gamma < \infty \) and \( x \geq \) some \( N < \infty \)) we can move the limit inside the integral by corollary 4.1.1 to obtain \( \lim_{x \to \infty} \delta^{**}(x) = 0 \). (Here \( \delta^{**}(x) \) is the...
"Bayes estimate" for $z = \theta - (\lambda')^{-1}(x))$. But by theorem 4.2 we see that the Bayes estimate for $\theta$, $\delta^*(x)$, is asymptotically $(\lambda')^{-1}(x)$ as $x \to \infty$ for this wide class of loss functions. More specifically $\delta^*(x) - (\lambda')^{-1}(x) \to 0$ as $x \to \infty$.

Corollary 4.1.2 can be used for the case that $\lambda''(y) \to \infty$ as $y \to \infty$ in the above example. Here Meeden and Isaacson show (under appropriate conditions) that

$$\frac{1}{\sqrt{\lambda''((\lambda')^{-1}(x))}} (\theta | x = x - (\lambda')^{-1}(x))$$

is asymptotically normal with mean 0 and variance 1. Again $\delta^*(x) - (\lambda')^{-1}(x) \to 0$ as $x \to \infty$.

These theorems can be extended directly to $n$ dimensions when

$$L(\theta, \varepsilon) = \sum_{i=1}^{n} W_i(|\theta_i - \varepsilon_i|),$$

where each $W_i$ is strictly increasing and convex. Again write $h(\theta | x)$ for the posterior density of $\theta | X = x$. Then to minimize

$$\int_{R} L(\theta, \delta(x)) h(\theta | x) d\theta$$

one can write,

$$(4.14) \quad \int_{R} L(\theta, \delta(x)) h(\theta | x) d\theta$$

$$= \int_{R} \sum_{i=1}^{n} W_i(|\theta_i - \delta_i(x)|) h(\theta | x) d\theta$$
\[ n \sum_{i=1}^{n} \int_{\mathbb{R}} w_i(|\theta_i - \delta_i(x)|) h(\theta | x) \, d\theta \]

\[ = n \sum_{i=1}^{n} \{ \int_{\mathbb{R}} w_i(|\theta_i - \delta_i(x)|) \int_{\mathbb{R}^{n-1}} h(\theta | x) \, d\delta_i \, d\theta_i \} \]

\[ = n \sum_{i=1}^{n} \int_{\mathbb{R}} w_i(|\theta_i - \delta_i(x)|) h_i(\theta_i | x) \, d\theta_i , \]

where \( h_i(\theta_i | x) \) is the density for \( \theta_i | X = x \). So to minimize (4.14) we can choose \( \delta(x) = (\delta_1(x), \delta_2(x), \ldots, \delta_n(x))^T \) such that \( \delta_i(x) \) is a function which satisfies

\[ \delta_i(x) \]

\[ \int_{-\infty}^{\infty} w_i'(\delta_i(x) - \theta_i) h_i(\theta_i | x) \, d\theta_i \]

\[ = \int_{\delta_i(x)}^{\infty} w_i'(\theta_i - \delta_i(x)) h_i(\theta_i | x) \, d\theta_i . \]

Thus to minimize (4.14) the problem can be reduced to the one dimensional case via marginal distributions, \( h_i(\theta_i | x) \). Then as \( x \) gets large in some appropriate fashion one may get a limiting distribution for \( h_i(\theta_i | x) \) as considered in the earlier chapters.

The conditions given in the theorems are not very strong. In general they apply when the limiting distributions are symmetric and put some mass about the point of symmetry. The theorems are valuable
in that they cover a reasonably wide class of loss functions. One can then usually concentrate on the more easily handled case of squared error loss and get the limiting behavior of $E(\theta|X = x)$. This can then, under appropriate conditions, be extended to the particular loss function under consideration.

The results of this chapter apply when the loss is a convex, increasing function of absolute error. An interesting question to consider is what happens if we drop the assumption of convexity or of strictly increasing loss. For example, one may wish to consider zero-one loss, where the loss is zero if estimate and parameter are equal and is one otherwise. The resulting Bayes estimate is the posterior mode, which is actually $(\lambda')^{-1}(x)$ of chapter II and of the results noted in this chapter from (Meeden and Isaacson, 1976). Thus a partial answer is already provided.
V. BIBLIOGRAPHY


VI. ACKNOWLEDGEMENTS

I am, of course, indebted to my major professor, Glen Meeden, for his contribution to this work. He introduced me to the problem and sat through many discussions of it with me. I would also like to thank H. T. David for the suggestion that functions of regular variation be used in this work. Lastly, I would like to thank my wife, Nancy, for her financial and emotional assistance in getting me to and through graduate school.