SOME NONPARAMETRIC BAYESIAN

ESTIMATION PROBLEMS

By

K. M. Lal Saxena

FSU Statistics Report M100

October, 1965
Department of Statistics
Florida State University
Tallahassee, Florida

Work supported by the Office of Naval Research Contract No.
Nonr 988(13).

Reproduction in whole or in part is permitted for any purpose of the United States Government.
1. **INTRODUCTION.**

There is a wealth of literature available dealing with the two sample nonparametric problems for single parameter families of distributions. There has been no attempt to consider these problems from the Bayesian viewpoint, when only the relative magnitudes of the observations are known. For an illustration, consider the following problem: Suppose $X_1$ and $X_2$ are independent random variables with continuous distribution functions $F(x, \theta_1)$ and $F(x, \theta_2)$ respectively. Define $Z = 1$ or $0$ according as $X_1 < X_2$ or $X_1 \geq X_2$. Suppose $Z$ is the only observable random variable. Then, the probability that $Z$ takes the value $1$ is given by

$$\int_{-\infty}^{\infty} F(x, \theta_1) dF(x, \theta_2).$$

Suppose this probability depends on $\theta_1$ and $\theta_2$ through a function $h(\theta_1, \theta_2)$. From the Bayesian point of view, $h(\theta_1, \theta_2)$ is a given value of a random variable $H = H(\theta_1, \theta_2)$. Let $H$ have the prior distribution function $G(h)$. Our interest is in the following two problems: (a) to obtain the Bayes estimate of $H$ with respect to the squared error loss function and to compare its risk with the risk of the Bayes estimate of $H$, with the squared error loss function, if $X_1$ and $X_2$ were observable, (b) with two decisions to choose from, namely, $\delta_1: H > h_0$ and $\delta_2: H \leq h_0$, to obtain the Bayes decision rule with respect to the $(0,1)$ loss function and to compare the risk of the Bayes decision rule with the $(0,1)$ loss function, if $X_1$ and $X_2$ were observable.

Three kinds of data, namely, paired comparison data, rank order data and signed rank order data, are considered in sections
2, 3 and 4 respectively. Both the problems (a) and (b) described above have been considered for the paired comparison data and the rank order data. For the signed rank order data, only problem (a) has been considered. The summaries of the results that have been obtained, for the estimation problem and the two decision problem, for the three kinds of data, are given in the corresponding sections after the problems are formulated. At this point, however, it should be pointed out that a number of results for the estimation problem, with the rank order data, are derived from results of Saxena (1965). The appendix gives some analytic properties of the normal distribution, results about likelihood ratios and probabilities of rank orders. These results are used in sections 2, 3 and 4.

It should be pointed out that the term nonparametric in the title, in the present context, does not imply that the methods adopted in this investigation are distribution free. It only indicates that the methods depend on statistics used in nonparametric methods such as the sign test statistic (for paired comparison data), the rank order statistic and the signed rank order statistic.

In this preliminary investigation of the nonparametric Bayes estimation problem and the nonparametric Bayes two decision problem, a number of results, such as the orderings of the values of the Bayes estimate of \( \theta \) in the estimation problem and the monotone character of the Bayes two decision procedures in the two decision problem, have been obtained under mild restrictions. But results have not been obtained in such generality as to be compatible with reality. The Bayes estimates, the Bayes risk and the efficiency of
the procedures have been discussed for very small sample
sizes when the populations sampled are normal or uniform.

2. PAIRED COMPARISON DATA.

Suppose \((X_1, Y_1), \ldots, (X_n, Y_n)\) are \(2n\) mutually
independent random variables. The \(X_i\)'s and the \(Y_i\)'s have the
continuous distribution functions \(F(x, \theta_1)\) and \(F(x, \theta_2)\) respectively.
Define random variables \(Z_1, \ldots, Z_n\) in the following manner:

\[
Z_i = \begin{cases} 
1 & \text{if } X_i < Y_i \\
0 & \text{if } X_i \geq Y_i, i = 1, \ldots, n.
\end{cases}
\]

The \(Z_i\)'s are independently and identically distributed random variables.
Denote by \(P_{\theta_1, \theta_2}(X < Y)\) the probability that \(Z_i\) takes the value 1. Then,

\[
P_{\theta_1, \theta_2}(X < Y) = \int_{-\infty}^{\infty} F(x, \theta_1) dF(x, \theta_2).
\]

Suppose that \(P_{\theta_1, \theta_2}(X < Y)\) depends on \(\theta_1\) and \(\theta_2\) through the function
\(h(\theta_1, \theta_2)\). Then \(P_{\theta_1, \theta_2}(X < Y)\) is written as \(P(h)\). As a degenerate
example, consider \(F(x, \theta)\) to be the normal distribution function with
variance \(\theta\). Then \(P(h)\) is identically equal to 1/2.

From the Bayesian viewpoint \(h(\theta_1, \theta_2)\) is a given value
of the random variable \(H = H(\theta_1, \theta_2)\). Then \(P(h)\) is the conditional
probability that \(Z_i\) takes the value 1 given \(h\). Let \(H\) have the prior
distribution function \(G(h)\). Whenever \(P_{\theta_1, \theta_2}(X < Y)\) depends on \(\theta_1\) and
\(\theta_2\) through their difference, the prior distribution is considered for
\(H = \theta_2 - \theta_1\). Whenever \(P_{\theta_1, \theta_2}(X < Y)\) depends on \(\theta_1\) and \(\theta_2\) through
their ratio, the prior distribution is considered for \( H = \frac{\theta_2}{\theta_1} \), unless otherwise stated. Define

\[
\Lambda_n = \frac{1}{n} \sum_{i=1}^{n} Z_i .
\]

Notice that \( \Lambda_n \) is a sufficient statistic for \( h(\theta_1, \theta_2) \). The conditional probability that \( \Lambda_n \) takes a value \( \lambda \) given \( h \), denoted by \( f(\lambda|h) \), is given by

\[
f(\lambda|h) = \binom{n}{n\lambda} P^{\lambda}(h)(1-P(h))^n(1-\lambda),
\]

\( \lambda = 0, 1/n, \ldots, 1 \). The random variable \( H \) has the posterior distribution function \( P[H \leq h|\lambda] \) given by

\[
P[H \leq h|\lambda] = \frac{\int_{-\infty}^{h} P^{\lambda}(x)(1-P(x))^n(1-\lambda) dG(x)}{\int_{-\infty}^{\infty} P^{\lambda}(x)(1-P(x))^n(1-\lambda) dG(x)} .
\]

Sections 2.1 through 2.5 deal with the paired comparison Bayes estimation problem with the squared error loss function. Sections 2.6 through 2.9 deal with the paired comparison two decision problem with the \((0,1)\) loss function.

2.1 Statement and summary of results for the estimation problem. Suppose that \( Z_i \)'s are the only observable random variables. All through the discussion of the estimation problem it is assumed that the loss function is squared error and that the prior distribution of \( H \) has finite second moment. Therefore, from lemma 8 of the appendix, it follows that the Bayes estimate of \( H \) has finite risk. The Bayes estimate, denoted by \( \hat{h}^*(\lambda) \), is the mean of the posterior distribution of \( H \). Thus,

\[
\hat{h}^*(\lambda) = \frac{\int_{-\infty}^{\infty} h P^{\lambda}(h)(1-P(h))^n(1-\lambda) dG(h)}{\int_{-\infty}^{\infty} P^{\lambda}(h)(1-P(h))^n(1-\lambda) dG(h)} .
\]
The Bayes risk, denoted by $R(n)$, is given by

$$
R(n) = \sum_{\lambda}^{\infty} \binom{n}{n\lambda} \int_{-\infty}^{\infty} (h^\ast(\lambda) - h)^2 p_n^\lambda(h)(1-P(h))n^{1-\lambda} dG(h) .
$$

Equation (2.4) can be written as

$$
R(n) = \int_{-\infty}^{\infty} h^2 dG(h) - \left[ \sum_{\lambda}^{\infty} (h^\ast(\lambda))^2 \binom{n}{n\lambda} \right] 
\int_{-\infty}^{\infty} p_n^\lambda(h)(1-P(h))n^{1-\lambda} dG(h) .
$$

Denote by $R_0(n)$ the risk of the Bayes estimate of $H$ that would be obtained if the $X_i$'s and the $Y_i$'s were observable random variables. The efficiency of the paired comparison Bayes estimation procedure, denoted by $E(n)$, is defined by

$$
E(n) = \frac{R_0(n)}{R(n)} .
$$

Clearly $0 \leq E(n) \leq 1$ for all $n$. The efficiency is unity when the prior on $H$ is a single point distribution. When the paired comparisons are completely non-informative, that is when $P(h)$ is identically a constant, then $R(n)$ is the variance of the prior distribution of $H$. In such a situation, the efficiency goes to zero with increasing $n$ if $R_0(n)$ goes to zero with increasing $n$.

Section 2.2 gives a number of relations among the values of the Bayes estimate of $H$. In section 2.3 it is assumed that the prior distribution of $H$ has a scale parameter $\eta$. Maclaurin expansions for the Bayes estimate and the Bayes risk are given in a neighborhood of $\eta = 0$. Section 2.4 deals with the case when the populations sampled
belong to a normal family with a translation parameter $\theta$. In this case $H = \theta_2 - \theta_1$. The prior distribution of $H$ is assumed to be a normal distribution. The sample sizes considered are: $n = 1$, 2 and 3. The efficiency is considered for all the three sample sizes. In section 2.5 mild restrictions are put on $F(x, \theta)$ and $G(h)$. Then, it is shown that the Bayes risk $R(n)$ tends to zero as $n$ tends to infinity.

2.2. Relations among the values of the Bayes estimate of $H$. Theorem 2.2.1 gives a simple ordering of the values of $h^*(\lambda)$. Theorem 2.2.2 gives relations between the posterior distribution functions of $H$. Corollaries give applications of the theorems.

**Definition.** A family of density functions $f(x, \theta)$ is said to have an increasing likelihood ratio if for all $x > x'$ and $\theta > \theta'$, the following inequality holds:

$$f(x, \theta)f(x', \theta') - f(x, \theta')f(x', \theta) \geq 0.$$  

**Theorem 2.2.1.** Let $H$ have any prior density. If $P(h)$ is an increasing function of $h$, then $h^*(\lambda) \geq h^*(\lambda')$ whenever $\lambda > \lambda'$.

**Proof.** Consider any $\lambda > \lambda'$. From (2.1), it follows that

$$f(\lambda|h) = C \left[ \frac{P(h)}{1-P(h)} \right]^{n(\lambda'-\lambda)}.$$  

where $C$ does not depend on $h$. Since $P(h)$ is an increasing function of $h$, it follows that the family of density functions $f(\lambda|h)$ has increasing likelihood ratio. As $A_n$ is a sufficient statistic for $h(\theta_1, \theta_2)$, lemma 6 of the appendix applies. Therefore, the family of posterior densities $g(h|\lambda)$ has increasing likelihood ratio. Now an application
of a result of Lehmann (1959), given as lemma 7 of the appendix, completes the proof of the theorem.

It should be noted that in theorem 2.2.1, the inequality is strict if \( P(h) \) is a strictly increasing function of \( h \) and the prior distribution of \( H \) is not a single point distribution.

**Corollary 2.2.1.** Let \( H \) have any prior density. If the \( X_i \)'s and the \( Y_i \)'s have the continuous distribution functions \( F(x/\theta_1) \) and \( F(x/\theta_2) \) respectively, then \( h^*(\lambda) \geq h^*(\lambda') \) whenever \( \lambda > \lambda' \).

**Proof.** Since
\[
P_{\theta_1, \theta_2}(X < Y) = \int_{-\infty}^{\infty} F(x+\theta_2 - \theta_1) dF(x),
\]
it follows that \( H = \theta_2 - \theta_1 \) and
\[
(2.7) \quad P(h) = \int_{-\infty}^{\infty} F(x+h) dF(x).
\]
Clearly \( P(h) \) is an increasing function of \( h \). Hence theorem 2.2.1 applies.

**Corollary 2.2.2.** Let \( H \) have any prior density. Suppose the \( X_i \)'s and the \( Y_i \)'s are positive random variables having continuous distribution functions \( F(x/\theta_1) \) and \( F(x/\theta_2) \) respectively, \( \theta_1, \theta_2 > 0 \). Then, \( h^*(\lambda) \geq h^*(\lambda') \) whenever \( \lambda > \lambda' \).

**Proof.** Since
\[
P_{\theta_1, \theta_2}(X < Y) = \int_{0}^{\infty} F(\theta_2 x/\theta_1) dF(x),
\]
it follows that \( H = \theta_2 / \theta_1 \) and
\[
P(h) = \int_{0}^{\infty} F(hx) dF(x).
\]
Therefore, \( P(h) \) is an increasing function of \( h \) and theorem 2.2.1 applies.

**Theorem 2.2.2.** If

\[
\begin{align*}
(a) & \quad P(h) = 1 - P(-h), \\
(b) & \quad G(h) = 1 - G(-h),
\end{align*}
\]

for all \( h \), then the posterior distribution function \( f \) satisfies:

\[
P[H \leq h | \lambda] = 1 - P[H \leq -h | 1 - \lambda],
\]

for all \( h \).

**Proof.** In the expression for \( P[H \leq h | \lambda] \) given in (2.2), make the transformation \( x = -y \) in both the integrals. Then,

\[
P[H \leq h | \lambda] = \frac{\int_{-\infty}^{h} P^{n\lambda}(-y)(1 - P(-y))^{n(1 - \lambda)} dG(-y)}{\int_{-\infty}^{\infty} P^{n\lambda}(-y)(1 - P(-y))^{n(1 - \lambda)} dG(-y)}
\]

Now using (a) and (b) in the right hand side, the result follows.

**Corollary 2.2.3.** Under the conditions of theorem 2.2.2, \( h^*(\lambda) = -h^*(1 - \lambda) \).

**Corollary 2.2.4.** Suppose the \( X_1 \)'s and the \( Y_1 \)'s have continuous distribution functions \( F(x - \theta_1) \) and \( F(x - \theta_2) \) respectively. If \( G(h) = 1 - G(-h) \) for all \( h \), then \( h^*(\lambda) = -h^*(1 - \lambda) \).

**Proof.** Consider (2.7). Then,

\[
P(h) = \int_{-\infty}^{\infty} F(x + h) dF(x) = \int_{-\infty}^{\infty} F(x) dF(x - h)
\]

\[
= 1 - \int_{-\infty}^{\infty} F(x - h) dF(x) = 1 - P(-h),
\]

for all \( h \). Now Corollary 2.2.3 applies.

**Corollary 2.2.5.** Suppose \( P(h) \) and \( G(h) \) satisfy the conditions of theorems 2.2.1 and 2.2.2. Then \( h^*(\lambda) \geq 0 \) whenever \( \lambda > 1/2 \).
2.3. Maclaurin expansions for the Bayes estimate and its risk. In this subsection it is assumed that H has the prior distribution function \( G(h-h_0)/\eta \). Maclaurin expansion for the Bayes estimate of H and the Bayes risk are obtained in a neighborhood of \( \eta = 0 \) in theorem 2.3.1. It should be pointed out that the case when \( \eta \) is large is more interesting. But expansions in a neighborhood of \( \eta = 0 \) are relatively easy to obtain even though the calculus is involved. The following notation is adopted:

\[
P(n, \lambda, h_0, \eta \pi) = P_n^\lambda (h_0 + \eta \pi)(1-P(h_0 + \eta \pi))^n(1-\lambda),
\]

\[
I(\eta) = \frac{\int_{-\infty}^{\infty} xP(n, \lambda, h_0, \eta \pi)dG(x)}{\int_{-\infty}^{\infty} P(n, \lambda, h_0, \eta \pi)dG(x)},
\]

\[
I_i(0) = \left[ \frac{d^i}{d\eta^i} I(\eta) \right]_{\eta=0}, \text{ } i \text{ a non negative integer},
\]

\[
D_i(0) = \left[ \frac{\partial^i}{\partial(\eta \pi)^i} P(n, \lambda, h_0, \eta \pi) \right]_{\eta=0}, \text{ } i \text{ a positive integer},
\]

\[
D(0) = P(n, \lambda, h_0, 0),
\]

\[
d_i = \left[ \frac{\partial^i}{\partial(\eta \pi)^i} P(h_0 + \eta \pi) \right]_{\eta=0},
\]

\[
d = P(h_0),
\]

\[
r = n\lambda.
\]
The expansions are obtained under the following assumptions:

(a) \[ G(x) = 1 - G(-x) \] for all \( x \),

(b) \[ \frac{\partial^j}{\partial \eta^j} P(n, \lambda, h_0, \eta x) \] exists for \( j = 1, \ldots, 5 \),

(c) \[ \left| \frac{\partial^j}{\partial \eta^j} P(n, \lambda, h_0, \eta x) \right| < Q_j(x) \]

in some neighborhood of \( \eta = 0 \), where

\[
\int_{-\infty}^{\infty} x^i Q_j(x) dG(x) < \infty,
\]

for \( i \) fixed and \( j = 1, \ldots, 5 \). The following relations, needed for the expansion of \( h^*(\lambda) \) in theorem 2.3.1 are obtained after some computations:

\[
D(0) = d^r(1-d)^{n-r},
\]

\[
D_1(0) = d_1 d^{r-1}(1-d)^{n-r-1}(r-nd),
\]

\[
D_2(0) = d_2 d^{r-1}(1-d)^{n-r-1}(r-nd) + d_2^2 d^{r-2}(1-d)^{n-r-2}(r(r-1) - 2r(n-1)d + n(n-1)d^2),
\]

\[
D_3(0) = d_3 d^{r-1}(1-d)^{n-r-1}(r-nd) + d_1 d_2 d^{r-2}(1-d)^{n-r-2}(3r(r-1) - 6r(n-1)d + 3n(n-1)d^2)
\]

\[
+ d_1^3 d^{r-3}(1-d)^{n-r-3}(r(r-1)(r-2) - 3r(r-1)(n-2)d + 3r(n-1)(n-2)d^2) + 3r(n-1)(n-2)d^2 - n(n-1)(n-2)d^3)
\].

**Lemma 2.3.1.** Suppose the assumptions (a), (b) and (c) hold. Denote
\[ J^i_j = \left[ \frac{\partial^j}{\partial \eta^j} \int_{-\infty}^{\infty} x^i \bar{P}(n, \lambda, h_0, \eta x) dG(x) \right]_{\eta = 0} . \]

Then for all \( j \leq 5 \) and \( i \) fixed

\[ J^i_j = \begin{cases} 0 & \text{for } (i+j) \text{ odd} \\ \mu_{i+j} \left[ \frac{\partial^j}{\partial (\eta x)^j} P(n, \lambda, h_0, \eta x) \right]_{\eta = 0} & \text{for } (i+j) \text{ even} \end{cases} , \]

where

\[ \mu_k = \int_{-\infty}^{\infty} x^k dG(x) . \]

**Proof.** From (b) and (c) it follows that

\[ J^i_j = \int_{-\infty}^{\infty} x^{i+j} \left[ \frac{\partial^j}{\partial (\eta x)^j} P(n, \lambda, h_0, \eta x) \right]_{\eta = 0} dG(x) . \]

Now the lemma follows from (a).

Lemma 2.3.2 given below is obtained by repeated application of lemma 2.3.1 to \( I(\eta) \) given in (2.8).

**Lemma 2.3.2.** Suppose the assumptions (a), (b) and (c) hold. Then,

\[ I_0(0) = 0 , \]

\[ I_1(0) = \frac{\mu_2 D_1(0)}{D(0)} , \]

\[ I_2(0) = 0 , \]

\[ I_3(0) = \frac{\mu_4 D_3(0)}{D(0)} - \frac{3 \mu_2^2 D_1(0) D_2(0)}{D^2(0)} , \]

\[ I_4(0) = 0 . \]

**Theorem 2.3.1.** Suppose \( H \) has the prior distribution function \( G((h-h_0)/\eta) \). If the assumptions (a), (b) and (c) hold, then the
Maclaurin expansions of \( h^*(\lambda) \) and \( R(n) \) in a neighborhood of \( \eta = 0 \)
are given by

\[
h^*(\lambda) = h_0 + \frac{\eta^2 \mu_2 d_2(0)}{D(0)} + \frac{4}{3!} \left[ \frac{\mu_4 d_3(0)}{D(0)} + \frac{3 \mu_2^2 d_1(0)d_2(0)}{D^2(0)} \right] + O(\eta^6)
\]

(2.9)

and

\[
R(n) = \eta^2 + \frac{\eta^2 \mu_2 d_1(0)}{d(1-d)} + O(\eta^6).
\]

(2.10)

**Proof.** From (2.3), \( h^*(\lambda) \) can be written as

\[
h^*(\lambda) = h_0 + \eta I(\eta),
\]

where \( I(\eta) \) is given in (2.8). Then, the Maclaurin expansion of \( h^*(\lambda) \)
is given by

\[
h^*(\lambda) = h_0 + \sum_{i=1}^{5} \frac{\eta^i I_{i-1}(0)}{(i-1)!} + O(\eta^6).
\]

Using lemma 2.3.2, the expansion (2.9) for \( h^*(\lambda) \) is obtained. From

(2.5), \( R(n) \) can be written as

\[
R(n) = \eta^2 \mu_2 + h_0^2
\]

(2.11)

\[
- \sum_{\text{all } \lambda} \binom{n}{\lambda} (h^*(\lambda))^2 \int_{-\infty}^{\infty} P(n, \lambda, h_0, \eta x) dG(x).
\]

Use of lemma 2.3.1 gives

\[
\int_{-\infty}^{\infty} P(n, \lambda, h_0, \eta x) dG(x)
\]

(2.12)

\[
= D(0) + \frac{\eta^2 \mu_2 d_2(0)}{2!} + \frac{\eta^4 \mu_4 d_4(0)}{4!} + O(\eta^6).
\]
In (2.11) replace $h^*(\lambda)$ and the integral by (2.9) and (2.12). Then

$$R(n) = \eta \mu_2 + h_0^2 - h_0^2 \sum_{r=0}^{n} \binom{n}{r} D(0) - \frac{1}{2} h_0^2 \eta \mu_2 \sum_{r=0}^{n} \binom{n}{r} D_2(0)$$

$$-2 h_0 \eta \mu_2 \sum_{r=0}^{n} \binom{n}{r} D_1(0) - \frac{1}{4} h_0^2 \eta \mu_4 \sum_{r=0}^{n} \binom{n}{r} D_4(0)$$

$$-\eta \mu_2 \sum_{r=0}^{n} \frac{\binom{n}{r} D_1^2(0)}{D(0)} - \frac{1}{3} h_0 \eta \mu_4 \sum_{r=0}^{n} \binom{n}{r} D_3(0) + O(\eta^6).$$

(2.13)

It should be noted that $D_i(0)$ are functions of $r$. It is found that in (2.13) all summations except $\sum_{r=0}^{n} \binom{n}{r} D(0)$ and $\sum_{r=0}^{n} \binom{n}{r} D_1^2(0)/D(0)$ are zero.

The values of these two summations are given below:

$$\sum_{r=0}^{n} \binom{n}{r} D(0) = 1$$

(2.14)

and

$$\sum_{r=0}^{n} \frac{\binom{n}{r} D_1^2(0)}{D(0)} = \frac{nd^2}{d(1-d)}.$$  

(2.15)

Using (2.14) and (2.15) in (2.13), the expansion (2.10) is obtained.

Notice that the expansions of $h^*(\lambda)$ and $R(n)$ in theorem 2.3.1, though valid for any $\eta$, are useful only for small $\eta$. These expansions are used in section 2.4, dealing with a normal family, to obtain an approximation to the efficiency of the paired comparison Bayes estimation procedure.

2.4. A normal family. Let $F(\cdot)$ and $f(\cdot)$ denote the standard normal distribution function and the density function respectively.
Suppose the \( X_i \)'s and the \( Y_i \)'s have the distribution functions \( F((x-\theta_1)/\sigma) \) and \( F((x-\theta_2)/\sigma) \) respectively, where \( \sigma \) is known. Then

\[
P_{\theta_1, \theta_2}(X < Y) = \int_{-\infty}^{\infty} F\left(x, \frac{h}{\sigma}\right) dF(x),
\]

where \( h = \theta_2 - \theta_1 \). Hence lemma 2 of the appendix gives:

\[
P(h) = F(h/\sigma \sqrt{2}).
\]

Let the prior distribution function of \( H = \theta_2 - \theta_1 \) be \( F((h-h_0)/\eta) \). Corollary 2.2.1 shows that \( h^*(\lambda) > h^*(\lambda') \) whenever \( \lambda > \lambda' \). If in particular, \( h_0 = 0 \), corollaries 2.2.4 and 2.2.5 also apply. Then \( h^*(\lambda) = -h^*(1-\lambda) \) for any \( \lambda \) and \( h^*(\lambda) > 0 \) whenever \( \lambda > 1/2 \).

Suppose the \( X_i \)'s and the \( Y_i \)'s were observable random variables. Then the Bayes estimate of \( H \) would be \( n\eta^2(\bar{y} - \bar{x})/(n\eta^2 + 2\sigma^2) \), where \( \bar{y} \) and \( \bar{x} \) denote the sample means, each based on \( n \) observations. The Bayes risk \( R_0(n) \) would be given by

\[
(2.16) \quad R_0(n) = \frac{2\sigma^2}{n\eta^2 + 2\sigma^2}.
\]

Let \( \eta/\sigma \) be denoted by \( k \). For \( n = 1, 2 \) and \( 3 \), the expressions for the Bayes estimate \( h^*(\lambda) \), obtained from (2.3) and the use of lemma 2 of the appendix, are given below:

(i) \( n = 1 \) and any \( h_0 \):

\[
h^*(0) = h_0 - \frac{n\eta k}{\sqrt{k} + 2} \left( \frac{h_0}{\sqrt{2} + 2} \right) / F\left( \frac{-h_0}{\sqrt{2} + 2} \right),
\]

\[
h^*(1) = h_0 + \frac{n\eta k}{\sqrt{k} + 2} \left( \frac{h_0}{\sqrt{2} + 2} \right) / F\left( \frac{h_0}{\sqrt{2} + 2} \right).
\]
(ii) \( n = 2 \) and any \( h_0 \):

\[
\begin{align*}
h^*(0) &= h_0 - \frac{2\eta k}{\sqrt{k^2+2}} f \left( \frac{h_0}{\sqrt{k^2+2}} \right) F \left( \frac{-h_0}{\sigma \sqrt{k^2+3k^2+2}} \right) \\
&\quad \left[ F^2 \left( \frac{-h_0}{\sigma \sqrt{k^2+2}} \right) + \int_{-\infty}^{\infty} F^2 \left( \frac{kx}{\sigma \sqrt{2}} + \frac{h_0}{\sigma \sqrt{2}} \right) f(x) dx \right]^{-1},
\end{align*}
\]

\[
\begin{align*}
h^* \left( \frac{1}{2} \right) &= h_0 + \frac{\eta k}{\sqrt{k^2+2}} f \left( \frac{h_0}{\sqrt{k^2+2}} \right) \left[ 1 - 2F \left( \frac{h_0}{\sigma \sqrt{k^2+3k^2+2}} \right) \right] \\
&\quad \left[ F \left( \frac{h_0}{\sigma \sqrt{k^2+2}} \right) F \left( \frac{-h_0}{\sigma \sqrt{k^2+2}} \right) - \int_{-\infty}^{\infty} F^2 \left( \frac{kx}{\sigma \sqrt{2}} + \frac{h_0}{\sigma \sqrt{2}} \right) f(x) dx \right]^{-1},
\end{align*}
\]

\[
\begin{align*}
h^*(1) &= h_0 + \frac{2\eta k}{\sqrt{k^2+2}} f \left( \frac{h_0}{\sqrt{k^2+2}} \right) F \left( \frac{h_0}{\sigma \sqrt{k^2+3k^2+2}} \right) \\
&\quad \left[ F^2 \left( \frac{h_0}{\sigma \sqrt{k^2+2}} \right) + \int_{-\infty}^{\infty} F^2 \left( \frac{kx}{\sigma \sqrt{2}} + \frac{h_0}{\sigma \sqrt{2}} \right) f(x) dx \right]^{-1}.
\end{align*}
\]

(iii) \( n = 3 \) and \( h_0 = 0 \):

\[
\begin{align*}
h^*(1) &= \frac{6\eta k}{\sqrt{2\pi(k^2+2)}} \frac{\pi + 2 \arcsin \left( \frac{k^2}{2(k^2+1)} \right)}{\pi + 6 \arcsin \left( k \sqrt{\frac{2}{k^2+1}} \right)} \\
&\quad = -h^*(0),
\end{align*}
\]

\[
\begin{align*}
h^* \left( \frac{2}{3} \right) &= \frac{2\eta k}{\sqrt{2\pi(k^2+2)}} \frac{\pi - 6 \arcsin \left( \frac{k^2}{2(k^2+1)} \right)}{\pi - 2 \arcsin \left( \frac{k^2}{k^2+1} \right)} \\
&\quad = -h^*(1/3).
\end{align*}
\]
For \( n = 1, 2, 3 \) and \( h_0 = 0 \), the expressions for the efficiency of the paired comparison Bayes estimation procedure from (2.5), (2.6), (2.16) and lemma 2 of the appendix are given below:

\[
E(1) = \frac{2\pi}{\pi(k^2+2) - 2k^2}
\]

\[
E(2) = \frac{(k^2+2)(\pi+2 \arcsin \frac{k^2}{(k^2+2)})}{(k^2+1)[(k^2+1)(\pi+2 \arcsin \frac{k^2}{(k^2+2)}) - 4k^2]}
\]

\[
E(3) = \frac{2}{3k^2+2} \left[ 1 - \frac{k^2}{\pi(k^2+1)} \left\{ \frac{9(\pi+2 \arcsin \frac{k^2}{2(k^2+1)})^2}{2(\pi+6 \arcsin \frac{k^2}{(k^2+2)})} \right\} \right]^{-1}
\]

Table 2.4.1 shows that the efficiency decreases, for each \( n \), as \( k \) increases, that is as the prior distribution is more dispersed relative to the dispersion of the distributions of the sampled populations. The efficiency also decreases, for every value of \( k \) considered, as \( n \) increases from 1 to 3. It is not known whether for all \( k \) and \( n \), the efficiency is a decreasing function of \( k \) with \( n \) fixed, and a decreasing function of \( n \) with \( k \) fixed, and if so whether the efficiency decreases to zero or to some other constant.
<table>
<thead>
<tr>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.5917</td>
<td>0.5833</td>
<td>0.5319</td>
<td>0.4760</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9928</td>
<td>0.9792</td>
<td>0.6258</td>
<td>0.6825</td>
<td>0.5833</td>
<td>0.5319</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9718</td>
<td>0.9473</td>
<td>0.8587</td>
<td>0.8587</td>
<td>0.5791</td>
<td>0.4760</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9386</td>
<td>0.8934</td>
<td>1.0</td>
<td>0.8587</td>
<td>0.7865</td>
<td>0.2366</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8953</td>
<td>0.8319</td>
<td>1.0</td>
<td>0.8587</td>
<td>0.7865</td>
<td>0.2366</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8462</td>
<td>0.7679</td>
<td>1.0</td>
<td>0.8587</td>
<td>0.7865</td>
<td>0.2366</td>
</tr>
<tr>
<td>1.2</td>
<td>0.7926</td>
<td>0.7062</td>
<td>1.0</td>
<td>0.8587</td>
<td>0.7865</td>
<td>0.2366</td>
</tr>
</tbody>
</table>

Table 2.1.4.2. The efficiency of the paired comparison Bayes estimation procedure for \( n = 1, 2, 3, h_0 = 0, \) and \( k = 0(2)2(15). \)
For \( n \geq 3 \) and any \( h_0 \), it is found that the computations for the values of the Bayes estimate require the evaluation of the integrals

\[
\int_{-\infty}^{\infty} F^m \left( \frac{kx + h_0\sqrt{2}}{\sqrt{k^2+2}} \right) f(x) dx, \quad \int_{-\infty}^{\infty} F^m \left( \frac{kx + \frac{h_0}{\sqrt{2}}}{\frac{\sqrt{2}}{\sigma\sqrt{2}}} \right) f(x) dx,
\]

for \( m = 0, 1, \ldots, n \). Tables by Milton (1963) give the numerical values of the integral

\[
\int_{-\infty}^{\infty} F^m \left( \frac{x\sqrt{r} + t}{\sqrt{1-r}} \right) f(x) dx
\]

for \( m = 2(1)9(5)24, t = \cdot 00(0.05)5.15 \) and \( r = \cdot 00(0.05)1.00 \). The only value of \( k \) for which the tables can be used without interpolation is \( \sqrt{2/3} \) provided \( h_0 \) is zero.

The Maclaurin expansions considered in section 2.3 can be used to obtain approximate expressions for \( h^*(\lambda) \) and \( R(n) \). Since \( P(h_0 + \eta x) = F((h_0 + \eta x)/\sqrt{2}) \), \( p = f(h_0/\sigma\sqrt{2}) \) and \( p_1 = f(h_0/\sigma\sqrt{2})/\sigma\sqrt{2} \), theorem 2.3.1 gives

\[
R(n) = \eta^2 - \frac{n\eta^4}{2\sigma^2 Q(h_0)} + O(\eta^6),
\]

where

\[
Q(h_0) = \frac{F(h_0/\sigma\sqrt{2})F(-h_0/\sigma\sqrt{2})}{f^2(h_0/\sigma\sqrt{2})}. \tag{2.17}
\]

Using the first two terms in the above expansion of \( R(n) \) and (2.16), an approximation, denoted by \( E_a(n) \), for the efficiency \( E(n) \) is

\[
E_a(n) = \left( \frac{2}{nk^2+2} \right) / \left( 1 - \frac{nk^2}{2Q(h_0)} \right).
\]
As $k$ tends to zero, $E_a(n)$ tends to unity. This is as it should be since $k$ tending to zero means that the prior distribution is tending to a single point distribution. The approximation $E_a(n)$ should not be used if $nk^2 > 2Q(h_0) - 2$, as the efficiency can not be greater than unity. Further, $E_a(n)$ is a decreasing function of $k$ with $n$ fixed and a decreasing function of $n$ with $k$ fixed if and only if

$$nk^2 < Q(h_0) - 1.$$  

(2.18)

Notice that $Q(h_0)$ in (2.17) is symmetric about $h_0 = 0$. Lemma 3 of the appendix shows that $Q(h_0)$ is an increasing function of $h_0$ for $h_0 \geq 0$. Thus, the minimum value of $Q(h_0) - 1$ is $\pi/2 - 1$. Therefore, there exist $n$ and $k$ such that $nk^2 < Q(h_0) - 1$. To be consistent with the numerical evidence of the table 2.4.1, it is recommended that the approximation procedure be used only for those values of $n$ and $k$ which satisfy (2.18). The following table gives a comparison of $E_a(n)$ for $n = 1, 2, 3$ and $h_0 = 0$, with the cases considered earlier in this subsection, over the recommended range of values of $nk^2$. It also gives the values of $E_a(n)$ over the recommended range of values of $nk^2$ for $n = 4, 5$ and 6.
Table 2.4.2. The approximate efficiency of the paired comparison Bayes estimation procedure for $n = 1(1)6$, $h_0 = 0$ and values of $k$ over the recommended range.

<table>
<thead>
<tr>
<th>k</th>
<th>$E_{a(1)}$</th>
<th>$E_{a(2)}$</th>
<th>$E_{a(3)}$</th>
<th>$E_{a(4)}$</th>
<th>$E_{a(5)}$</th>
<th>$E_{a(6)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.1</td>
<td>.9930</td>
<td>.9928</td>
<td>.9866</td>
<td>.9964</td>
<td>.9930</td>
<td>.9900</td>
</tr>
<tr>
<td>0.2</td>
<td>.9982</td>
<td>.9982</td>
<td>.9964</td>
<td>.9964</td>
<td>.9947</td>
<td>.9946</td>
</tr>
<tr>
<td>0.3</td>
<td>.9952</td>
<td>.9839</td>
<td>.9732</td>
<td>.9858</td>
<td>.9946</td>
<td>.9956</td>
</tr>
<tr>
<td>0.4</td>
<td>.9576</td>
<td>.9718</td>
<td>.9598</td>
<td>.9692</td>
<td>.9792</td>
<td>.9571</td>
</tr>
<tr>
<td>0.5</td>
<td>.9567</td>
<td>.9718</td>
<td>.9598</td>
<td>.9692</td>
<td>.9792</td>
<td>.9571</td>
</tr>
<tr>
<td>0.6</td>
<td>.9516</td>
<td>.9514</td>
<td>.9218</td>
<td>.9473</td>
<td>.9519</td>
<td>.9260</td>
</tr>
<tr>
<td>0.7</td>
<td>.9184</td>
<td>.9386</td>
<td>.9218</td>
<td>.9473</td>
<td>.9519</td>
<td>.9260</td>
</tr>
</tbody>
</table>
2.5. The risk of the paired comparison Bayes estimation procedure for large $n$. In this subsection, sufficient conditions are given for the Bayes risk $R(n)$ to go to zero as $n$ goes to infinity.

**Theorem 2.5.1.** Suppose that

(a) $G(h)$ has finite second moment,

(b) $P(h) = 1 - P(-h)$ for all $h$, and $P(h)$ is strictly increasing,

(c) the derivative of $P(h)$, denoted by $P'(h)$, exists for all $h$ and $P'(h) \geq P'(h')$ for $0 \leq h < h'$,

(d) there exists a sequence $\{b_n\}$ of positive real numbers such that

$$\lim_{n \to \infty} [\sqrt{n} P'(2 b_n)]^{-1} = 0$$

and $b_n \to \infty$, then,

$$\lim_{n \to \infty} R(n) = 0.$$

**Remark.** Some of the assumptions in theorem 2.5.1 can easily be relaxed. For instance, the first part of assumption (b) is for convenience in applying the Tchebecheff inequality. Assumption (d) could be modified to accommodate distributions with finite range.

**Proof of theorem 2.5.1.** Consider the following estimate of $H$:

$$s^*(\lambda) = \begin{cases} P^{-1}(\lambda) & \text{for } P(-2 b_n) \leq \lambda \leq P(2 b_n) \\ 0 & \text{otherwise} \end{cases}$$

The risk of $s^*(\lambda)$, denoted by $R_s(n)$, is given by

$$R_s(n) = \sum_{\text{all } \lambda} \left( \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} [s^*(\lambda) - h]^2 P^n(h) P^n(1-\lambda) (-h) \, dG(h) \right).$$
It is clear that \( R(n) \leq R_s(n) \). Therefore, it is sufficient to show that

\[
\lim_{n \to \infty} R_s(n) = 0.
\]

Breaking up the summation and the integral in (2.20), \( R_s(n) \) can be symbolically written as

\[
R_s(n) = \sum_{\text{all } \lambda} \int_{-b_n}^{b_n} + \sum_{\text{all } \lambda} \int_{b_n}^{\infty} + \sum_{\lambda < P(-2b_n)} \int_{-b_n}^{-b_n} + \sum_{\lambda > P(2b_n)} \int_{b_n}^{b_n} + \sum_{P(-2b_n) \leq \lambda \leq P(2b_n)} \int_{-b_n}^{b_n}.
\]

(2.21)

Denote the \( i \)th term on the right hand side of (2.21) by \( I_i, \ i = 1, \ldots, 5 \).

Then,

\[
I_1 = \sum_{\text{all } \lambda} \binom{n}{n\lambda} \int_{-\infty}^{-b_n} \{s^*(\lambda) - h\}^2 p^{n\lambda}(h)p^{n(1-\lambda)}(-h)dG(h)
\]

\[
< \sum_{\text{all } \lambda} \binom{n}{n\lambda} \int_{-\infty}^{-b_n} \{2b_n - h\}^2 p^{n\lambda}(h)p^{n(1-\lambda)}(-h)dG(h)
\]

\[
= \int_{-\infty}^{-b_n} \{2b_n - h\}^2 dG(h)
\]

\[
\leq 4b_n^2 \int_{-\infty}^{-b_n} dG(h) - 4b_n \int_{-\infty}^{-b_n} hdG(h) + \int_{-\infty}^{-b_n} h^2 dG(h)
\]

\[
\leq 9 \int_{-\infty}^{-b_n} h^2 dG(h).
\]

From (a) it follows that

\[
\lim_{n \to \infty} I_1 = 0.
\]
In a similar manner it is found that

$$\lim_{n \to \infty} I_2 = 0 .$$

From (2.19) it follows that

$$I_3 + I_4 = \int_{-b_n}^{b_n} h^2 P_h [\lambda < P(-2b_n) \text{ or } \lambda > P(2b_n)] dG(h).$$

Consider $h \in (-b_n, b_n)$. Then,

$$P_h [\lambda < P(-2b_n) \text{ or } \lambda > P(2b_n)]$$

$$= 1 - P_h [P(-2b_n) - P(h) \leq \lambda - P(h) \leq P(2b_n) - P(h)]$$

$$< 1 - P_h [P(-2b_n) - P(-b_n) < \lambda - P(h) < P(2b_n) - P(b_n)]$$

$$= P_h [|\lambda - P(h)| > P(2b_n) - P(b_n)]$$

$$\leq \frac{P(h)(1-P(h))}{n[P(2b_n) - P(b_n)]^2} ,$$

from the Tchebecheff inequality. Therefore,

$$I_3 + I_4 < \frac{b_n^2}{4n[P(2b_n) - P(b_n)]^2} .$$

Applying the mean value theorem to $P(2b_n) - P(b_n)$, it is found that

$$I_3 + I_4 < \frac{1}{4n[P'(2b_n)]^2} .$$

Hence, from (d) it follows that

$$\lim_{n \to \infty} I_3 + I_4 = 0 .$$
The following Taylor expansion of $P^{-1}(\lambda)$ in a neighborhood of $\lambda = P(h)$ is used for $I_5$:

$$P^{-1}(\lambda) = h + \frac{\lambda - P(h)}{P'[P^{-1}(c\lambda+(1-c)P(h))]} ,$$

where $0 \leq c \leq 1$. Thus,

$$I_5 = \sum_{n=1}^{n} \left( n \right) _{n} \int_{-b_n}^{b_n} \frac{(\lambda-P(h))^{2}P_{n}(h)P_{n}(1-\lambda)(-h)}{P'[P^{-1}(c\lambda+(1-c)P(h))]^{2}} dG(h),$$

where $\Sigma$ denotes summation over $\lambda \in [P(-2b_n), P(2b_n)]$. Hence

$$I_5 < \sum_{n=1}^{n} \left( n \right) _{n} \int_{-b_n}^{b_n} \frac{(\lambda-P(h))^{2}P_{n}(h)P_{n}(1-\lambda)(-h)}{[P'(2b_n)]^{2}} dG(h)$$

$$< \frac{1}{4n[P'(2b_n)]^{2}} .$$

From (d), it follows that

$$\lim_{n \to \infty} I_5 = 0.$$ 

This completes the proof of the theorem.

The conditions (b) through (d) of theorem 2.5.1 are satisfied if the populations sampled are normal. Then in particular, $\{\log \log n\}$ can be taken as the $b_n$ - sequence. It should be noted that theorem 2.5.1 does not give the main term of the expression for the risk of $s^*(\lambda)$.

2.6. Statement and summary of results for the two decision problem.

Suppose the $Z_i$'s are the only observable random variables. Assume that $G(h)$, the prior distribution function of $H$, is continuous (at least at
h_0). One of the following two decisions is to be taken:

\[ \delta_1: \theta > h_0, \quad \delta_2: \theta \leq h_0, \]

where \( h_0 \) is some preassigned constant. All through the discussion of the two decision problem, it is assumed that the loss function, denoted by \( L(\delta, h) \), is

\[
L(\delta, h) = \begin{cases} 
0 & \text{if } H > h_0 \text{ and the decision taken is } \delta_1, \\
1 & \text{if } H > h_0 \text{ and the decision taken is } \delta_2, \\
0 & \text{if } H \leq h_0 \text{ and the decision taken is } \delta_2, \\
1 & \text{if } H \leq h_0 \text{ and the decision taken is } \delta_1.
\end{cases}
\]

The conditional risk of \( \delta_1 \), given \( \lambda \), denoted by \( R(\delta_1, \lambda) \) is given by

\[
(2.22) \quad R(\delta_1, \lambda) = C \int_{-\infty}^{h_0} \binom{n}{n\lambda} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} dG(h)
\]

and the conditional risk of \( \delta_2 \), given \( \lambda \), denoted by \( R(\delta_2, \lambda) \), is given by

\[
(2.23) \quad R(\delta_2, \lambda) = C \int_{h_0}^{\infty} \binom{n}{n\lambda} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} dG(h),
\]

where \( C \) is the reciprocal of the marginal density of \( \Lambda_n \) at \( \lambda \). Then, for a given \( \lambda \), the Bayes decision procedure is the following: Take decision \( \delta_1 \) if \( R(\delta_1, \lambda) < R(\delta_2, \lambda) \) and take decision \( \delta_2 \) if \( R(\delta_1, \lambda) \geq R(\delta_2, \lambda) \). Then, the Bayes risk, denoted by \( R(n) \), is given by

\[
R(n) = \sum_{1}^{\infty} \binom{n}{n\lambda} \int_{-\infty}^{h_0} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} dG(h)
\]

\[
+ \sum_{2}^{\infty} \binom{n}{n\lambda} \int_{h_0}^{\infty} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} dG(h),
\]

\[
(2.24)
\]
where }\sum_1\text{ is the summation over all }\lambda \in [\lambda : R(\delta_1, \lambda) < R(\delta_2, \lambda)] \text{ and } \sum_2 \text{ is the summation over all }\lambda \in [\lambda : R(\delta_1, \lambda) \geq R(\delta_2, \lambda)].

Definition. For the paired comparison two decision problem, the Bayes decision procedure is said to be monotone if there exists a number }\lambda_0', \, 0 \leq \lambda_0' \leq 1, \text{ such that for all }\lambda > \lambda_0' \text{ one specific decision is taken and for all }\lambda < \lambda_0' \text{, the other decision is taken. For }\lambda = \lambda_0' \text{ the decision may be either }\delta_1 \text{ or }\delta_2'.

Remark. In the above definition, it is assumed that for }\lambda = \lambda_0' \text{, the decision may be }\delta_1 \text{ or }\delta_2', \text{ to include the extreme cases in which the Bayes decision is }\delta_1 \text{ for all }\lambda \text{ or the Bayes decision is }\delta_2 \text{ for all }\lambda.

Section 2.7 deals with the monotonic behavior of the Bayes decision procedure. In section 2.8, }h_0\text{ is assumed to be zero. It is shown that, under certain conditions }R(2n-1) = R(2n). \text{ Section 2.9 deals with the Bayes risk }R(n) \text{ for large }n.

2.7. Monotonic behavior of the Bayes two decision procedure. Theorem 2.7.1 gives a sufficient condition for the Bayes two decision procedure to be monotone.

Theorem 2.7.1. Let }H\text{ have any prior density }g(h). \text{ If }P(h)\text{ is an increasing function of }h \text{ and }P(0) = 1/2, \text{ then the paired comparison Bayes two decision procedure is monotone.}

Proof. Suppose }h_0 \leq 0. \text{ Consider }R(\delta_1, \lambda) \text{ and }R(\delta_2, \lambda) \text{ given by (2.22) and (2.23). Make the transformation }y = h - h_0 \text{ in the integral in (2.22) and the transformation }y = -h + h_0 \text{ in the integral in (2.23). Then,}

\begin{equation}
R(\delta_1, \lambda) - R(\delta_2, \lambda) = C_{n\lambda} \int_{-\infty}^{0} [1 - P(y + h_0)]^{B} (y + h_0)^{n\lambda} \left[ \frac{P(y + h_0)}{1 - P(y + h_0)} \right]^{n\lambda}
\end{equation}
\[
\left[ 1 - \left\{ \frac{1 - P(y+h_0)}{1 - P(y)} \right\} \frac{g(y)}{g(y+h_0)} \frac{P(-y|h_0) \left(1 - P(y) \right)}{P(y+h_0) \left(1 - P(-y|h_0) \right)} \right] n^\lambda dy
\]

\[= C_{n^\lambda}^n \lambda I(\lambda),\]

where \(I(\lambda)\) denotes the integral in (2.25). In \(I(\lambda)\), observe that

\[(2.26) \quad \frac{P(y|h_0)}{1 - P(y|h_0)} \leq 1,\]

for all \(y \leq 0\) and \(h_0 \leq 0\) because \(P(h)\) is an increasing function of \(h\) and \(P(0) = 1/2\). Further observe that

\[(2.27) \quad \frac{P(-y|h_0) \left(1 - P(y) \right)}{P(y|h_0) \left(1 - P(-y|h_0) \right)} \geq 1,\]

for all \(y \leq 0\) because \(P(h)\) is an increasing function of \(h\).

Consider any \(\lambda' > \lambda\). Then, using (2.26) and (2.27) in \(I(\lambda)\), it follows that

\[I(\lambda') \leq I(\lambda).\]

Therefore,

\[(2.28) \quad R(\delta_1, \lambda') - R(\delta_2, \lambda') \leq \frac{C_{n^\lambda}^n \lambda'}{C_{n^\lambda}^n \lambda} \left[ R(\delta_1, \lambda) - R(\delta_2, \lambda) \right],\]

where \(C\) and \(C'\) are the evaluations of the marginal density of \(n\) at \(\lambda\) and \(\lambda'\) respectively. The inequality (2.28) shows that, if for \(\lambda\) the Bayes decision is \(\delta_1\), then for all \(\lambda' > \lambda\) the Bayes decision is \(\delta_1\).

Furthermore, (2.28) also shows that, if for \(\lambda'\) the Bayes decision is \(\delta_2\), then for all \(\lambda < \lambda'\) the Bayes decision is \(\delta_2\). This establishes that the Bayes two decision procedure is monotone whenever \(h_0 \leq 0\). In a similar manner, it is found that the Bayes two decision procedure is monotone
whenever \( h_0 \geq 0 \). Thus, the theorem is proved.

It should be noted that even when \( h_0 < 0 \), the Bayes decision may be \( \delta_2 \) for all \( \lambda \). Such a situation can arise if the prior distribution of \( H \) puts almost all the probability mass in a neighborhood of a point far enough to the left of \( h_0 \). Similarly, when \( h_0 > 0 \), the Bayes decision may be \( \delta_1 \) for all \( \lambda \). Such a situation can arise if the prior distribution of \( H \) puts almost all the probability mass in a neighborhood of a point far enough to the right of \( h_0 \).

**Corollary 2.7.1.** If the sampled populations belong to a family of distributions with a translation parameter, then the Bayes two decision procedure is monotone.

**Proof.** It has been shown in corollaries 2.2.1 and 2.2.4 that \( P(h) \) is an increasing function of \( h \) and that \( P(0) = 1/2 \). Therefore, theorem 2.7.1 applies.

**Theorem 2.7.2.** Suppose that \( P(h) \) and \( G(h) \) satisfy the following conditions for all \( h \):

(i) \( P(h) = 1 - P(-h) \),

(ii) \( G(h) = 1 - G(-h) \).

Then, (a) with \( h_0 \geq 0 \), if for some \( \lambda \) the Bayes decision is \( \delta_1 \), then for \( 1 - \lambda \), the Bayes decision is \( \delta_2 \) and (b) with \( h_0 \leq 0 \), if for some \( \lambda \) the Bayes decision is \( \delta_2 \), then for \( 1 - \lambda \), the Bayes decision is \( \delta_1 \). Furthermore, if \( P(h) \) is an increasing function of \( h \), then with \( h_0 \geq 0 \), the Bayes decision is \( \delta_2 \) for all \( \lambda < 1/2 \) and with \( h_0 \leq 0 \), the Bayes decision is \( \delta_1 \) for all \( \lambda > 1/2 \).

**Proof.** Suppose that \( h_0 \geq 0 \) and that the Bayes decision for \( \lambda \) is \( \delta_1 \). Consider \( R(\delta_1, \lambda) \) and \( R(\delta_2, \lambda) \) given by (2.22) and (2.23) and make the
transformation \( y = -h \) in both the integrals. Then,

\[
(2.29) \quad \int_{-h_0}^{\infty} p^{n(1-\lambda)}(y)p^{n\lambda}(-y)dG(y) < \int_{-\infty}^{-h_0} p^{n(1-\lambda)}(y)p^{n\lambda}(-y)dG(y).
\]

Since \( h_0 \geq 0 \), it follows from (2.29) that

\[
\int_{h_0}^{\infty} p^{n(1-\lambda)}(y)p^{n\lambda}(-y)dG(y) < \int_{-\infty}^{h_0} p^{n(1-\lambda)}(y)p^{n\lambda}(-y)dG(y).
\]

Therefore, the Bayes decision for \( 1 - \lambda \) is \( \delta_2 \) and (a) is proved. In a similar manner, it is found that (b) is true.

When \( P(h) \) is an increasing function of \( h \), theorem 2.7.1 applies. Suppose with \( h_0 \geq 0 \), the Bayes decision is \( \delta_1 \) for some \( \lambda < 1/2 \). Then, from (a) it follows that the Bayes decision is \( \delta_2 \) for \( 1 - \lambda( > 1/2) \). This contradicts theorem 2.7.1. Similarly theorem 2.7.1 is contradicted if the Bayes decision is \( \delta_2 \) for some \( \lambda > 1/2 \) when \( h_0 \leq 0 \). This completes the proof of the theorem.

It should be noted that, when the conditions of theorems 2.7.1 and 2.7.2 are satisfied, the Bayes decision can be \( \delta_2 \) for all \( \lambda \) when \( h_0 > 0 \) and the Bayes decision can be \( \delta_1 \) for all \( \lambda \) when \( h_0 < 0 \). This can happen when \( h_0 >> 0 \) or \( h_0 << 0 \).

2.8. The Bayes decision procedure with \( h_0 = 0 \). Suppose the prior density of \( h \) is \( g(h-h') \) satisfying the symmetry condition: \( g(h-h') = g(h'-h) \) for all \( h \). It is assumed all through this section that \( h_0 = 0 \).

Theorem 2.8.1 shows how the Bayes decision depends on \( h' \). Theorem 2.8.2 gives the conditions which give \( R(2n-1) = R(2n) \).

Theorem 2.8.1. Suppose \( g(h-h') \) and \( P(h) \) satisfy the following conditions:
(i) \( g(h-h') = g(h'-h) \),

for all \( h \) and \( g(h-h') \) is a decreasing function of \( h \) for \( h \geq h' \),

(ii) \( P(h) = 1 - P(-h) \), for all \( h \),

(iii) \( P(h) \) is an increasing function of \( h \).

If \( h_0 = 0 \), then (a) with \( h' \geq 0 \), the Bayes decision is \( \delta_1 \) for all \( \lambda \geq 1/2 \)
and (b) with \( h' \leq 0 \), the Bayes decision is \( \delta_2 \) for all \( \lambda \leq 1/2 \).

\textbf{Proof.} Suppose \( h' \geq 0 \). Since \( h_0 = 0 \), it follows from (2.22), (2.23)
and the conditions (i) and (ii) that

\[
R(\delta_1, \lambda) - R(\delta_2, \lambda) = C_{n\lambda} \left[ \int_0^\infty P^n(h)P^n(1-\lambda)(-h)g(h-h') \right. \\
\left. \int_0^{\infty} \left[ \frac{P(h)}{P(-h)} \right]^{n(1-2\lambda)} \frac{g(h+h')}{g(h-h')} \right] dh.
\]

(2.30)

From (i) it follows that \( g(h+h')/g(h-h') \geq 1 \) for \( h' \geq 0 \) and all
\( h \leq 0 \). From (ii) and (iii) it follows that \( [P(h)/P(-h)]^{n(1-2\lambda)} \geq 1 \)
for \( \lambda \geq 1/2 \) and all \( h \leq 0 \). Then (2.30) shows that \( R(\delta_1, \lambda) - R(\delta_2, \lambda) < 0 \)
for all \( \lambda \geq 1/2 \). In a similar manner (b) is found to be true.

\textbf{Corollary 2.8.1.} If the conditions of theorem 2.8.1 are satisfied and
\( h' = 0 \), then the Bayes decision is \( \delta_1 \) for \( \lambda > 1/2 \) and the Bayes
decision is \( \delta_2 \) for \( \lambda \leq 1/2 \).

It should be pointed out that for corollary 2.8.1, it is not
necessary to assume that \( g(h) \) is a decreasing function of \( h \) for \( h \geq 0 \).

\textbf{Theorem 2.8.2.} Suppose \( P(h) \) and \( g(h) \) satisfy the following conditions
for all \( h \): (i) \( P(h) = 1 - P(-h) \), (ii) \( g(h) = g(-h) \). Then, for \( h_0 = 0 \),
\( R(2n-1) = R(2n) \).

\textbf{Proof.} Denote \( n\lambda \) by \( r \). From (2.24) and corollary 2.8.1, it follows
that
\[ R(2n) - R(2n-1) \]
\[ = \int_{-\infty}^{0} 2 \sum_{r=n+1}^{2n} \binom{2n}{r} P_r^x(h) P^{2n-r}(-h) - 2 \sum_{r=n}^{2n-1} \binom{2n-1}{r} P_r^x(h) P^{2n-r-1}(-h) \]
\[ + \binom{2n}{n} P^n(h) P^n(-h) \int_{-\infty}^{\infty} g(h) dh . \]

After some manipulation of the expression inside the big brackets, using condition (i), it is found that the expression is identically equal to zero. Hence the theorem.

2.9. **The risk of the Bayes two decision procedure for large \( n \).** In this subsection sufficient conditions are given for the Bayes risk \( R(n) \) to go to zero as \( n \) goes to infinity.

**Theorem 2.9.1.** Let \( H \) have any prior density which does not assign a positive probability to the point \( h_0 \). If \( P(h) \) is a strictly increasing function of \( h \) and the derivative of \( P(h) \), denoted by \( P'(h) \), exists at \( h_0 \), then

\[ \lim_{n \to \infty} R(n) = 0. \]

**Proof.** Consider the following two decision procedure: take the decision \( \delta_1 \) if \( P^{-1}(\lambda) > h_0 \) and take the decision \( \delta_2 \) if \( P^{-1}(\lambda) \leq h_0 \). The risk of this decision procedure, denoted by \( R^*(n) \), is given by

\[ R^*(n) = \sum_{\lambda \geq P(h_0)} \binom{n}{n\lambda} \int_{-\infty}^{h_0} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} g(h) dh \]
\[ + \sum_{\lambda < P(h_0)} \binom{n}{n\lambda} \int_{h_0}^{\infty} P^{n\lambda}(h)(1-P(h))^{n(1-\lambda)} g(h) dh \]
(2.31) \quad = \int_{-\infty}^{h_0} P_{h}^{*} \left( \lambda > P(h_0) \right) g(h) dh + \int_{h_0}^{\infty} P_{h} \left( \lambda \leq P(h_0) \right) g(h) dh.

Denote the first integral of (2.31) by \( I_1 \) and the second integral by \( I_2 \).

Since \( R(n) \) is the risk of the Bayes two decision procedure, it is clear that \( R(n) \leq R^*(n) \). Therefore, to prove the theorem, it is sufficient to show that

\[
\lim_{n \to \infty} R^*(n) = 0.
\]

Consider \( I_1 \). The Cantelli inequality (see Savage (1961)) shows that

\[
P_{h} \left( \lambda > P(h_0) \right) \leq \frac{P(h)(1-P(h))}{P(h)(1-P(h)) + n(P(h_0) - P(h))^2},
\]

for all \( h \leq h_0 \), because \( P(h) \) is an increasing function of \( h \). Choose a sequence \( \{a_n\} \) of positive real numbers converging to zero such that

(2.32) \quad a_n = 0(\sqrt{n}^{-1})^\delta,

where \( 0 < \delta < 1/2 \). Then,

(2.33) \quad I_1 \leq \frac{1}{4n} \int_{-\infty}^{h_0-a_n} \frac{g(h)}{[P(h_0) - P(h)]^2} dh + \int_{h_0-a_n}^{h_0} g(h) dh.

As the prior density does not prescribe a positive probability to the point \( h_0 \), the second integral in (2.33) goes to zero as \( n \) goes to infinity. Further,

\[
\frac{1}{4n} \int_{-\infty}^{h_0-a_n} \frac{g(h)}{[P(h_0) - P(h)]^2} dh < \frac{1}{4n[P(h_0) - P(h_0-a_n)]^2},
\]
since \( P(h) \) is an increasing function of \( h \). From (2.32) it follows that,

\[
\lim_{n \to \infty} \frac{1}{4n[P(h_0) - P(h_0 - a_n)]^2} = \frac{1}{4[P'(h_0)]^2} \lim_{n \to \infty} \left( \frac{1}{na_n^2} \right),
\]

\[= 0.\]

Therefore, \( \lim_{n \to \infty} I_1 = 0. \)

Consider \( I_2 \). The Cantelli inequality shows that for all \( h \geq h_0 \),

\[
P_h(\lambda \leq P(h_0)) \leq \frac{P(h)(1-P(h))}{P(h)(1-P(h)) + n[P(h_0) - P(h)]^2}.
\]

Then in a similar manner as for \( I_1 \), it is found that \( \lim_{n \to \infty} I_2 = 0. \)

This completes the proof of the theorem.

It should be pointed out that, while theorem 2.9.1 shows that the Bayes risk \( R(n) \) goes to zero as \( n \) goes to infinity, it does not give the main term in the expression of \( R(n) \).

3. THE RANK ORDER DATA.

Suppose \( X_1, \ldots, X_m; Y_1, \ldots, Y_n \) are mutually independent random variables. The \( X_i \)'s and the \( Y_i \)'s have the continuous distribution functions \( F(x, \theta_1), F(x, \theta_2) \) and the density functions \( f(x, \theta_1), f(x, \theta_2) \) respectively. Let the observed values of \( X_1, \ldots, X_m \) be called the first sample and the observed values of \( Y_1, \ldots, Y_n \) be called the second sample. Consider the joint ranking of the two samples in the ascending order. Form the sequence \( z = (z_1, \ldots, z_N) \), where
\[ N = m + n, \] 
by defining \( z_i = 0 \) if the \( i \)th observation in the ranking is from the first sample and \( z_i = 1 \) if the \( i \)th observation in the ranking is from the second sample. The random variable \( Z \), corresponding to \( z \) is called a rank order. So long as there is no confusion, \( z \) will also be called a rank order.

The probability that \( Z = z \), denoted by \( P_{\theta_1', \theta_2'} (z) \), is given by

\[
P_{\theta_1', \theta_2'} (z) = m! n! \int \cdots \int_{i=1}^{N} f(x_1', \theta_1') f(x_i', \theta_2') \, dx_i. 
\]

Suppose \( P_{\theta_1', \theta_2'} (z) \) depends on \( \theta_1 \) and \( \theta_2 \) through the function \( h(\theta_1, \theta_2) \).

Then \( P_{\theta_1', \theta_2'} (z) \) is written as \( P_h (z) \).

The notation \( zRz' \) denotes the following relationship: \( z_i = z'_i \) for all \( i = 1, \ldots, N \) except for some \( j \) and \( k(j < k) \) and \( z_j = z'_j = 0 \), \( z_k = z'_k = 1 \). The notation \( zR^* z' \) denotes either of the following relationships: (a) \( z_i = z'_i \) for \( i = 3, \ldots, N \) and \( z_1 = z'_2 = 0 \), \( z_2 = z'_1 = 1 \), (b) \( z_i = z'_i \) for \( i = 1, \ldots, N - 2 \) and \( z_{N-1} = z'_{N-1} = 0 \), \( z_N = z'_{N-1} = 1 \). The notation \( zR^* z' \) is also used if there exists \( z \) such that \( zR^* z' \). The transpose and complement of \( z \) are defined by \( z^T = (z_1^T, \ldots, z_N^T) \) and \( z^c = (z_1^c, \ldots, z_N^c) \) respectively, where \( z_i^T = z_{N+1-i} \) and \( z_i^c = 1 - z_i, i = 1, \ldots, N \).

From the Bayesian viewpoint, \( h(\theta_1', \theta_2') \) is a given value of a random variable \( H = H(\theta_1, \theta_2) \). Then \( P_h (z) \) is the conditional probability that \( Z = z \) given \( h \). Let \( H \) have the prior distribution function \( G(h) \). Whenever \( P_{\theta_1', \theta_2'} (z) \) depends on \( \theta_1 \) and \( \theta_2 \) through their difference,
the prior distribution is considered for $H = \theta_2 - \theta_1$. Whenever $P_{\theta_1, \theta_2}(z)$ depends on $\theta_1$ and $\theta_2$ through their ratio, the prior distribution is considered for $H = \theta_2/\theta_1$ unless otherwise stated. The random variable $H$ has the posterior distribution function $P[H \leq h | z]$ given by
\begin{equation}
P[H \leq h | z] = \frac{\int_{-\infty}^{h} P_X(z) dG(x)}{\int_{-\infty}^{\infty} P_X(z) dG(x)}
\end{equation}

The posterior density of $H$, given $z$, whenever it is assumed to exist, is denoted by $g(h | z)$.

Sections 3.1 through 3.5 deal with the rank order Bayes estimation of $H$ with the squared error loss function. Section 3.6 deals with the rank order Bayes two decision problem with the $(0,1)$ loss function.

3.1. Statement and summary of results for the estimation problem.

Suppose the vector $Z$ is the only observable random variable. All through the discussion of the rank order Bayes estimation of $H$, the loss function is assumed to be squared error and the prior distribution of $H$ is assumed to have finite second moment. Therefore, from lemma 3 of the appendix, it follows that the Bayes estimate of $H$ has finite risk. The Bayes estimate of $H$, denoted by $h^*(z)$, is the mean of the posterior distribution of $H$. Thus,
\begin{equation}
h^*(z) = \frac{\int_{-\infty}^{\infty} hP_h(z) dG(h)}{\int_{-\infty}^{\infty} P_h(z) dG(h)}
\end{equation}

The Bayes risk, denoted by $R(m, n)$, is given by
(3.4) \[ R(m, n) = \sum_{\text{all } z} \int_{-\infty}^{\infty} \left( h^*(z) - h \right)^2 p_h(z) dG(h). \]

Equation (3.4) can be written as

(3.5) \[ R(m, n) = \int_{-\infty}^{\infty} h^2 dG(h) - \sum_{\text{all } z} \left( h^*(z) \right)^2 \int_{-\infty}^{\infty} p_h(z) dG(h). \]

Denote by \( R_0(m, n) \) the risk of the Bayes estimate of \( H \) that would be obtained if the \( X_i \)'s and the \( Y_i \)'s were observable random variables. The efficiency of the rank order Bayes estimation procedure, denoted by \( E(m, n) \) is defined by

\[ E(m, n) = \frac{R_0(m, n)}{R(m, n)}. \]

Clearly \( 0 \leq E(m, n) \leq 1 \), for all \( m \) and \( n \). If the prior distribution of \( H \) is a single point distribution, then the efficiency is unity.

Section 3.2 gives two analytic properties of the posterior distribution of \( H \). A number of relations among the values of the Bayes estimate are obtained. Section 3.3 deals with the case when the populations sampled belong to a normal family of distributions with a translation parameter. In this case \( H = \theta_2 - \theta_1 \). The prior distribution of \( H \) is assumed to be a normal distribution with zero mean. The sample sizes considered are: (i) \( m = n = 2 \), (ii) \( m = 1, n = 2 \) and (iii) \( m = 2, n = 1 \). The efficiency is obtained for the three cases. Section 3.4 deals with the case when the populations sampled belong to a uniform family of distributions with a scale parameter. In this case \( H = \theta_1/\theta_2 \). The prior distribution of \( H \) is assumed to be the uniform distribution over \([0, 1]\). Simplified expressions for the Bayes estimate of \( H \) and
the Bayes risk are obtained for \( m \geq 3 \) and any \( n \). The efficiency is considered for \( m = 3 \) and \( n = 1(1)10 \). In section 3.5, the behavior of the Bayes risk is discussed for large \( m \) and \( n \).

3.2. **Analytic properties of the posterior distribution of \( H \).** Theorems 3.2.1 and 3.2.2 give two analytic properties of the posterior distribution function of \( H \). Theorem 3.3.3 is about the likelihood ratio of the posterior densities.

**Theorem 3.2.1.** Suppose \( H \) has any prior distribution function \( G(h) \).

If \( f(x, \theta) = f(x-\theta) = f(\theta-x) \), for all \( x \), then the posterior distribution function of \( H \) satisfies:

\[
P[H \leq h | z] = P[H \leq h | z^{tc}],
\]

for all \( h \) and \( z \).

**Proof.** Consider \( P[H \leq h | z] \) given by (3.2). Due to a result by Savage, Sobel and Woodworth (1965), given in lemma 4 of the appendix, \( P_x(z) \) can be replaced by \( P_x(z^{tc}) \) for all \( x \) and the theorem follows.

**Corollary 3.2.1.** If the conditions of theorem 3.2.1 are satisfied then \( h^*(z) = h^*(z^{tc}) \).

**Theorem 3.2.2.** If (a) \( f(x, \theta) = f(x-\theta) = f(\theta-x) \) for all \( x \) and (b) \( G(h) = 1 - G(-h) \) for all \( h \), then

\[
P[H \leq h | z] = 1 - P[H \leq - h | z^c] = 1 - P[H \leq - h | z^t],
\]

for all \( h \) and \( z \).

**Proof.** Consider \( P[H \leq h | z] \) given by (3.2). Make the transformation \( x = -y \) in both the integrals. Then,

\[
\int_{-y}^{-h} P_y(z) dG(-y) = \frac{P[H \leq h | z]}{\int_{-y}^{-h} P_y(z) dG(-y)}.
\]
From a result by Savage, Sobel and Woodworth (1965), given in lemma 4 of the appendix, $P_{-y}(z)$ can be replaced by $P_y(z^c)$ or $P_y(z^t)$ for all $y$. Then from (b), the theorem follows.

**Corollary 3.2.2.** If the conditions of theorem 3.2.2 are satisfied, then $h^*(z) = -h^*(z^t) = -h^*(z^c)$ for all $z$. Further for every $z$ such that $z \equiv z^t$, $h^*(z) = 0$.

**Example 3.2.1.** Denote the condition: $f(x, \theta) = f(x-\theta) = f(\theta-x)$ for all $x$, by ST and if further $G(h) = 1 - G(-h)$ for all $h$, denote the conditions by STS. Then for $m = n = 2$,

\[
\begin{align*}
STS & \quad h^*(0011) = -h^*(1100), \\
STS & \quad h^*(0101) = -h^*(1010), \\
STS & \quad h^*(1001) = h^*(0110) = 0.
\end{align*}
\]

**Theorem 3.2.3.** Suppose $H$ has any prior density $g(h)$. For any $z$ and $z'$ if $P_h(z)/P_h(z')$ is an increasing function of $h$, then $g(h|z)/g(h|z')$ is an increasing function of $h$.

**Proof.** The posterior density of $H$ given $z$ is

\[
g(h|z) = \frac{P_h(z)g(h)}{\int_{-\infty}^{\infty} P_h(z)g(h)dh}.
\]

Therefore, $g(h|z)/g(h|z') = CP_h(z)/P_h(z')$, where $C$ does not depend on $h$. Hence the theorem.

Theorem 3.2.3 is used in corollary 3.2.3 to obtain orderings of the values of the Bayes estimate of $H$. The author (1965) has in-
vestigated in detail the conditions under which different families of distributions have $P_h(z)/P_h(z')$ as an increasing function of $h$. It has been shown that, for a Lehmann family and a uniform family of distributions, $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zRz'$ and also for $z$ and $z'$ in some other special relationships. For a normal family and a logistic family of distributions, each with a translation parameter, it has been shown that $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zR^*z'$. For the normal family, the logistic family and a double exponential family of distributions with a translation parameter and $m = n = 2$, it has been shown that $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zR(z')$. Further, it is shown that $P_{\theta_1,\theta_2}(z)/P_{\theta_1,\theta_2}(z')$ is monotone in each of the two arguments $\theta_1$ and $\theta_2$ when one of the populations sampled has increasing likelihood ratio and minimum $(m,n) = 1$.

**Corollary 3.2.3.** Suppose $H$ has any prior density $g(h)$. For any $z$ and $z'$, if $P_h(z)/P_h(z')$ is an increasing function of $h$, then $h^*(z) \geq h^*(z')$.

**Proof.** Theorem 3.2.3 shows that $g(h|z)/g(h|z')$ is an increasing function of $h$. Then, from a result by Lehmann (1959), given as lemma 7 in the appendix, the corollary follows.

**Corollary 3.2.4.** Suppose $H$ has any prior density $g(h)$. If

(a) $f(x,\theta) = f(x-\theta)$,

(b) $f(x-\theta)$ has an increasing likelihood ratio,

(c) minimum $(m,n) = 1$,

then $h^*(z) \geq h^*(z')$ whenever $zRz'$.

**Proof.** Due to (a), $H = \theta_2 - \theta_1$. From corollaries 5.3 and 5.4 of Saxena (1965) it follows that $P_h(z)/P_h(z')$ is an increasing function of $h$.
whenever $zRz'$. Hence corollary 3.2.3. applies.

**Corollary 3.2.5.** Suppose $H$ has any prior density $g(h)$. If

(a) $f(x, \theta) = \theta f(x), \theta > 0$,

(b) $\theta f(\theta x)$ has an increasing likelihood ratio,

(c) minimum $(m, n) = 1$,

then $h^*(z) \geq h^*(z')$ whenever $zRz'$.

**Proof.** Due to (a), $H = \frac{\theta_2}{\theta_1}$. From corollaries 5.5 and 5.6 of Saxena (1965) it follows that $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zRz'$. Hence corollary 3.2.3 applies.

**Corollary 3.2.6.** Suppose $H$ has any prior density $g(h)$. If

(a) $f(x, \theta) = f(x-\theta) = f(\theta-x)$ for all $x$,

(b) $f(x-\theta)$ has an increasing likelihood ratio,

(c) $f(x-\theta)F(x)/f(x)F(x-\theta)$ is an increasing function of $x$ for $\theta \geq 0$,

then $h^*(z) \geq h^*(z')$ when $zR^*z'$.

**Proof.** Due to (a), $H = \theta_2 - \theta_1$. From theorem 6.2 of Saxena (1965) it is seen that $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zR^*z'$. Therefore, corollary 3.2.3 applies.

It should be pointed out that condition (c) of corollary 3.2.6 is satisfied by a normal family (see Savage, Sobel and Woodworth (1965), lemma 1 of appendix I) and a logistic family (see example 6.2 of Saxena (1965)).

**Corollary 3.2.7.** Suppose $H$ has any prior density $g(h)$. If

(a) $f(x, \theta) = f(x-\theta)$, (b) $z = (0^m, 1^n)$ and $z' = (1^n, 0^m)$ where $0^m$ and $1^n$ denote runs of $m$ zeroes and $n$ ones, then $h^*(z) \geq h^*(z')$.

**Proof.** Due to (a), $H = \theta_2 - \theta_1$. From a remark in section 7 of Saxena (1965) it follows that $P_h(z)/P_h(z')$ is an increasing function of $h$. 
Therefore corollary 3.2.3 applies.

Corollary 3.2.3 is also used for obtaining an ordering of the values of the Bayes estimate in section 3.4.

3.3. A normal family with a translation parameter. Let $F(\cdot)$ and $f(\cdot)$ denote the standard normal distribution function and the density function respectively. Suppose the $X_i$'s and the $Y_i$'s have the distribution functions $F((x-\theta_1)/\sigma)$ and $F((x-\theta_2)/\sigma)$ respectively, where $\sigma$ is known. In this case, $H = \theta_2 - \theta_1$. Assume that $H$ has the prior distribution function $F(x/\eta)$. Let $\eta/\sigma$ be denoted by $k$. The following combinations of the sample sizes are considered: (a) $m = n = 2$, (b) $m = 2$, $n = 1$ and (c) $m = 1$, $n = 2$.

(a) $m = n = 2$: Example 3.2.1 shows that the values of the Bayes estimate of $H$ need to be evaluated only for the rank orders (0011) and (0101). For the rank order (0011), it follows from (3.1) and (3.3) that

$$h^*(0011) = \frac{\eta \int \int h^2 F(kh-x)F(x)df(x)df(h)}{\int \int F^2(kh-x)F(x)df(x)df(h)}.$$ 

After some manipulation both the integrals can be put in the form (d) of lemma 2 of the appendix. Then, it is found that

$$h^*(0011) = \frac{\eta k}{\sqrt{2\pi(k^2+2)}} \pi - 2 \arcsin \left(\frac{1}{(2k^2+3)}\right) \arcsin \left(\frac{1}{(k^2+1)/(k^2+2)}\right)$$

$$= -h^*(1100).$$

Consider the rank order (0101). From a result of Savage, Sobel and
Woodworth (1965) given as lemma 5 in the appendix, it follows that

\[ p_h(0101) = 2p_h^2(01) - 2p_h(0011). \]

From (3.1), (3.3), (3.6) and lemma 2 of the appendix it is found that

\[ h^*(0101) = -h^*(1010) \]

\[ = \frac{4\eta k^{\sqrt{2}}}{\sqrt{\pi(k^2+2)}} \frac{\arcsin \left( \frac{1}{(2k^2+3)} \right)}{\pi + 2 \arcsin \left( k^2/(k^2+2) \right) - 4 \arcsin \left( (k^2+1)/(k^2+2) \right)} \]

From corollaries 3.2.3, 3.2.6 and results in section 7 of Saxena (1965) about a normal family of distributions and \( m = n = 2 \), the following ordering of the values of the Bayes estimate is obtained:

\[ h^*(0011) > h^*(0101) > h^*(1100) = 0 \]

\[ = h^*(1001) > h^*(1101) > h^*(1100). \]

From (3.5) and lemma 2 of the appendix, it is found that

\[ R(2,2) \]

\[ = \eta^2 - \frac{\eta^2 k^2 \left[ \pi - 2 \arcsin \left( \frac{1}{(2k^2+3)} \right) \right]^2}{\pi^2 (k^2+2) \arcsin \left( (k^2+1)/(k^2+2) \right)} - \frac{32\eta^2 k^2 \left[ \arcsin \left( \frac{1}{(2k^2+3)} \right) \right]^2}{\pi^2 (k^2+2) \left[ \pi + 2 \arcsin \left( k^2/(k^2+2) \right) - 4 \arcsin \left( (k^2+1)/(k^2+2) \right) \right]} \]

Suppose the \( X_i \)'s and the \( Y_i \)'s were observable random variables. Then, the Bayes estimate of \( H \) would be \( 2\eta^2 (\bar{y} - \bar{x})/(2\eta^2 + 2\sigma^2) \), where \( \bar{y} \) and \( \bar{x} \) denote the sample means, each based on two observations.
The Bayes risk $R_0(2,2)$ would be given by

$$R_0(2,2) = \frac{n^2}{k^2 + 1}.$$ 

Table 3.3.1. The efficiency of the rank order Bayes estimation procedure with $m = n = 2$ and $k = 0(\cdot 2)2(1)5$.

<table>
<thead>
<tr>
<th>k</th>
<th>0·0</th>
<th>0·2</th>
<th>0·4</th>
<th>0·6</th>
<th>0·8</th>
<th>1·0</th>
<th>1·2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(2,2)$</td>
<td>1·0000</td>
<td>·9908</td>
<td>·9650</td>
<td>·9271</td>
<td>·8802</td>
<td>·8273</td>
<td>·7704</td>
</tr>
<tr>
<td>k</td>
<td>1·4</td>
<td>1·6</td>
<td>1·8</td>
<td>2·0</td>
<td>3·0</td>
<td>4·0</td>
<td>5·0</td>
</tr>
<tr>
<td>$E(2,2)$</td>
<td>°7109</td>
<td>·6525</td>
<td>·5925</td>
<td>·5372</td>
<td>·3188</td>
<td>·1957</td>
<td>·1273</td>
</tr>
</tbody>
</table>

The rank order Bayes estimation procedure with $m = n = 1$ is the same as the paired comparison Bayes estimation procedure with $n = 1$. A comparison of table 3.3.1 with table 2.4.1 shows that for each $k$ the efficiency decreases as the common sample size is increased from 1 to 2. Further, table 3.3.1 shows that the efficiency decreases as $k$ increases. It has not been investigated whether, for all $k$ and $n$ (common sample size), the efficiency of the rank order Bayes estimation procedure is a decreasing function of $n$ with $k$ fixed and a decreasing function of $k$ with $n$ fixed and if so whether the efficiency decreases to zero or to some other constant.

(b) $m = 2$, $n = 1$: The three rank orders are (001), (010) and (100). Corollary 3.2.2 shows that $h^*(010) = 0$ and $h^*(001) = -h^*(100)$. Further, corollary 3.2.4 gives the following ordering of the values of the Bayes estimate: $h^*(001) > 0 = h^*(010) > h^*(100)$. From (3.1), (3.3), (3.5) and lemma 2 of the appendix it is found that
\[ h^*(001) = \frac{2\pi k \sqrt{2\pi}}{\sqrt{k^2 + 2} \left[ \pi + 2 \arcsin \left( (k^2 + 1)/(k^2 + 2) \right) \right] } \]

and

\[ R(2,1) = \eta^2 - \frac{4\eta^2 k^2}{(k^2 + 2) \left[ \pi + 2 \arcsin \left( (k^2 + 1)/(k^2 + 2) \right) \right] } \]

Suppose \( X_1, X_2 \) and \( Y \) were observable random variables. Then the Bayes risk \( R_0(2,1) \) would be \( 3\eta^2/(2k^2 + 3) \).

### Table 3.3.2. The efficiency of the rank order Bayes estimation procedure with \( m = 2, n = 1 \) and \( k = 0(\cdot 2)2(1)5 \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(2,1)</td>
<td>1.0000</td>
<td>0.9925</td>
<td>0.9374</td>
<td>0.8945</td>
<td>0.8448</td>
<td>0.7907</td>
</tr>
<tr>
<td>( k )</td>
<td>1.4</td>
<td>1.6</td>
<td>2.0</td>
<td>3.0</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>E(2,1)</td>
<td>0.7345</td>
<td>0.6783</td>
<td>0.5702</td>
<td>0.3603</td>
<td>0.2338</td>
<td>0.1593</td>
</tr>
</tbody>
</table>

Table 3.3.2, like table 3.3.1, shows that the efficiency decreases as \( k \) increases. Comparing table 3.3.2 with table 3.3.1 it is seen that for each \( k \), \( E(2,1) > E(2,2) \) which indicates that for each \( k \), the efficiency decreases as the sum of the two sample sizes increases.

(c) \( m = 1, n = 2 \): The three rank orders are (011), (101) and (110).

From corollary 3.2.2, it follows that \( h^*(011) = -h^*(100) \), \( h^*(110) = -h^*(001) \) and \( h^*(101) = 0 \). It is found that \( R(1,2) = R(2,1) \).

3.4. The uniform family. Suppose \( X_1, \ldots, X_m; Y_1, \ldots, Y_n \), \( m \geq 3 \), are mutually independent random variables. The \( X_i \)'s and the \( Y_i \)'s have
the distribution functions \( F(x, \theta_1) \) and \( F(x, \theta_2) \) respectively with \( \theta_1 \leq \theta_2 \), where

\[
F(x, \theta) = \begin{cases} 
0 & \text{for } x \leq 0 \\
x/\theta & 0 \leq x \leq \theta
\end{cases}
\]

In this case \( H = \theta_1/\theta_2 \) and

\[
(3.7) \quad P_h(z) = \frac{m!n!}{h^n(r(m+n-r-1)!h^m} \int_0^h (1-x)^r x^{m+n-r-1} dx,
\]

where \( r \) is the number of ones at the right end of \( z \). The random variable \( H \) is assumed to have the prior distribution function \( F(h,1) \). Equation (3.7) shows that \( P_h(z) \) depends on the components of \( z \) through \( r \) alone (for an alternative form of \( P_h(z) \), see Savage (1956)). Therefore, the notation for the Bayes estimate of \( H \) is changed from \( h^\ast(z) \) to \( h^\ast(r) \).

From (3.3) and (3.5), it is found that

\[
h^\ast(r) = \frac{m-1}{(m-2)(m+2)} \left[ \frac{(m+n)!((n-r+1)!(m+n-r-1)!}{(m+n)!((n-r)!(m+n-r-1)! - \frac{1}{(m+1)!(m+n-r-1)!} \right]
\]

and

\[
R(m,n) = \frac{1}{3} - \frac{m}{m-1} \sum_{r=0}^n \left( h^\ast(r) \right)^2 \left[ \frac{1}{n+1} - \frac{n!(m+n-r-1)!}{(n-r)!(m+n)!} \right]
\]

The following theorem gives the ordering of the values of the Bayes estimate of \( H \).

**Theorem 3.4.1.** If

\[
z = (z_1, \ldots, z_{m+n-r-1} 01(r)),
\]

\[
z' = (z'_1, \ldots, z'_{m+n-s-1} 01(s)),
\]
where \( r < s \) then \( h^*(r) > h^*(s) \).

**Proof.** From theorem 4.1 of Saxena (1965), it follows that \( P_h(z)/P_h(z') \) is an increasing function of \( h \). Therefore, corollary 3.2.3 applies.

To obtain the efficiency of the above rank order Bayes estimation procedure suppose that the \( X_i \)'s and the \( Y_i \)'s were observable random variables. Then \( (X_{\text{max}}, Y_{\text{max}}) \) is sufficient for \( (\theta_1, \theta_2) \). Since the prior is not given jointly for \( \theta_1 \) and \( \theta_2 \) but only for \( \theta_1/\theta_2 \), attention is restricted to the Bayes estimators of \( H = \theta_1/\theta_2 \) which are functions of \( X_{\text{max}}/Y_{\text{max}} \). Denote \( X_{\text{max}}/Y_{\text{max}} \) by \( U \) and the observed value of \( U \) by \( u \).

The conditional density of \( U \), given \( h = \theta_1/\theta_2 \), denoted by \( p(u/h) \), is given by

\[
p(u/h) = \begin{cases} \frac{mn u^{m-1}}{(m+n)h^m} & \text{for } 0 \leq u \leq h \\ \frac{mn \cdot n^{r+1}}{(m+n)u^{r+1}} & \text{for } u \leq h. \end{cases}
\]

After some manipulation, it is found that the Bayes estimate of \( H \), denoted by \( h_0^*(u) \), is given by

\[
h_0^*(u) = \begin{cases} \frac{(m-1)(n+1)u[(m+n)-(m+2)u^{m-2}]}{(m-2)(n+2)[m+n-(n+1)u^{m-1}]} & \text{for } u \leq 1 \\ \frac{n+1}{n+2} & \text{for } u \geq 1, \end{cases}
\]

and the Bayes risk \( R_0(m, n) \) is given by

\[
R_0(m, n) = \frac{1}{3} - \frac{mn(n+1)(m-1)}{(n+2)^2(m-2)^2} \int_0^1 \frac{x^2(1-ax^{m-2})}{(1-bx)^{m-1}} \, dx 
\]

\[
- \frac{m(n+1)}{(n+2)^2(m+n)},
\]
where $a = (m+2)/(m+n)$ and $b = (n+1)/(m+n)$. The efficiency is $R_0(m,n)/R(m,n)$. For $m = 3$, $R_0(3,n)$ is simplified to the following form:

$$R_0(3,n) = \frac{1}{3} - \frac{3(n+1)}{(n+2)^2(n+3)} - \frac{6n(n+1)}{(n+2)^2} \left[ \frac{3a-a^2-3}{3b} + \frac{1}{b^2} \left\{ a \log (1-b) - a^2 + \frac{b+a^2}{2\sqrt{b}} \log \frac{1+\sqrt{b}}{1-\sqrt{b}} \right\} \right],$$

where $a = (n+2)/(n+3)$ and $b = (n+1)/(n+3)$.

Table 3.4.1. The efficiency of the rank order Bayes estimation procedure for $m = 3$ and $n = 1(1)10$.

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(3,n)</td>
<td>-6408</td>
<td>.5279</td>
<td>.4854</td>
<td>.4750</td>
<td>.4799</td>
</tr>
<tr>
<td>n</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>E(3,n)</td>
<td>.4730</td>
<td>.4962</td>
<td>.5194</td>
<td>.5266</td>
<td>.5154</td>
</tr>
</tbody>
</table>

In Table 3.4.1, unlike Tables 3.3.1 and 3.3.2, no monotonicity is exhibited.

3.5. $R(m,n)$ for large $m$ and $n$. In theorem 2.5.1 sufficient conditions for the sampled populations are given for the risk of the paired comparison Bayes estimation procedure to go to zero as $n$ goes to infinity.

It is clear that for the rank order data, when the sampled populations satisfy the conditions of theorem 2.5.1, the risk of the Bayes estimation procedure goes to zero when both $m$ and $n$ go to infinity. This is so because, from the rank order data, the paired comparison data can be obtained by making pairs from the two samples by some random mechanism and throwing away the excess number of observations from the sample of bigger size.
3.6. **The rank order Bayes two decision problem.** Suppose the rank order \( Z \) is the only observable random variable. Assume that \( G(h) \), the prior distribution function of \( H \), is continuous. One of the following two decisions is to be taken:

\[
\delta_1: H > h_0, \quad \delta_2: H \leq h_0,
\]

where \( h_0 \) is some preassigned constant. All through the discussion of the two decision problem, it is assumed that the loss function, denoted by \( L(\delta, h) \), is

\[
L(\delta, h) = \begin{cases} 
0 & \text{if } H > h_0 \text{ and the decision taken is } \delta_1, \\
1 & \text{if } H > h_0 \text{ and the decision taken is } \delta_2, \\
0 & \text{if } H \leq h_0 \text{ and the decision taken is } \delta_2, \\
1 & \text{if } H \leq h_0 \text{ and the decision taken is } \delta_1.
\end{cases}
\]

The conditional risk of \( \delta_1 \), given \( z \), denoted by \( R(\delta_1, z) \) is given by

\[
R(\delta_1, z) = C \int_{-\infty}^{h_0} P_h(z) dG(h)
\]

and the conditional risk of \( \delta_2 \), given \( z \), denoted by \( R(\delta_2, z) \) is given by

\[
R(\delta_2, z) = C \int_{h_0}^{\infty} P_h(z) dG(h),
\]

where \( C \) is the marginal density of \( Z \) at \( z \). Then, for a given \( z \), the rank order Bayes two decision procedure is the following: take decision \( \delta_1 \) if \( R(\delta_1, z) < R(\delta_2, z) \) and take decision \( \delta_2 \) if \( R(\delta_1, z) \geq R(\delta_2, z) \).

Then the Bayes risk, denoted by \( R(m, n) \), is given by

\[
R(m, n) = \Sigma_1 \int_{-\infty}^{h_0} P_h(z) dG(h) + \Sigma_2 \int_{h_0}^{\infty} P_h(z) dG(h),
\]
where $\Sigma_1$ is the summation over all $z \in \{ z : R(\delta_1, z) < R(\delta_2, z) \}$ and $\Sigma_2$ is the summation over all $z \in \{ z : R(\delta_1, z) \geq R(\delta_2, z) \}$.

**Definition.** For a rank order $z$, denote $\{ z' : zRz' \}$ by $S_1(z)$ and $\{ z' : z'Rz \}$ by $S_2(z)$. The rank order Bayes two decision procedure is said to be monotone if, whenever the Bayes decision is $\delta_1$ for a rank order $z$, the Bayes decision is $\delta_1$ for all $z' \in S_2(z)$ and whenever the Bayes decision is $\delta_2$ for a rank order $z$, the Bayes decision is $\delta_2$ for all $z' \in S_1(z)$.

Theorem 3.6.1 gives sufficient conditions for the Bayes two decision procedure to be monotone. Theorem 3.6.2 provides tools for obtaining the Bayes decision when the sampled populations belong to a translation parameter family.

**Theorem 3.6.1.** Suppose $H$ has any prior density $g(h)$. If, (a) the density function $f(x, \theta)$ has an increasing likelihood ratio and (b) $P_h(z)/P_h(z')$ is an increasing function of $h$ whenever $zRz'$, then the rank order Bayes two decision procedure is monotone.

**Proof.** Suppose $h_0 \leq 0$. Consider $R(\delta_1, z)$ and $R(\delta_2, z)$ given by (3.8) and (3.9). Make the transformation $y = h - h_0$ in the integral in (3.8) and the transformation $y = -h + h_0$ in the integral in (3.9). Then,

$$R(\delta_1, z) = R(\delta_2, z)$$

(3.10)

$$= C \int_{-\infty}^{0} P_{y+h_0}^{-\gamma}(z)g(y+h_0) \left[ 1 - \frac{P_{y+h_0}^{-\gamma}(z)g(-y+h_0)}{P_{y+h_0}^{-\gamma}(z)g(y+h_0)} \right] dy$$

$$= CI(z),$$

where $I(z)$ denotes the integral in (3.10). Note that $C$ depends on $z$.

Consider any $z'$ such that $z'Rz$. From theorem 6.1 of Savage (1956), it
follows that

\[(3.11) \quad \frac{P_{y+h_0}(z')}{P_{y+h_0}(z)} \leq \frac{P_{y+h_0}(z)}{P_{y+h_0}(z)},\]

for all \(y \leq 0\) and \(h_0 \leq 0\). From (b), it follows that

\[(3.12) \quad \frac{P_{-y+h_0}(z')}{P_{y+h_0}(z')} \geq \frac{P_{-y+h_0}(z)}{P_{y+h_0}(z)},\]

for all \(y \leq 0\). From (3.11) and (3.12), it follows that \(I(z') \leq I(z)\).

Therefore,

\[(3.13) \quad R(\delta_1, z') - R(\delta_2, z') \leq \frac{C'}{C} [R(\delta_1, z) - R(\delta_2, z)].\]

If the Bayes decision is \(\delta_1\) for a rank order \(z\), then (3.13) shows that for all \(z' \in S_2(z)\), the Bayes decision is \(\delta_1\). The inequality (3.13) also shows that if the Bayes decision is \(\delta_2\) for a rank order \(z'\), then for all \(z \in S_1(z')\) the Bayes decision is \(\delta_2\). This proves the theorem with \(h_0 \leq 0\). For \(h_0 \geq 0\), the theorem is proved in a similar manner.

The author (1965) has investigated a number of families of distributions satisfying condition (b) of theorem 3.6.1. A summary of results obtained is to be found following theorem 3.2.3.

For the following theorem define \(T_1\) and \(T_2\) in the following manner:

\[T_1 = \{z: P_h(z) > P_h(z^t), \text{ for all } h \geq 0\},\]
\[T_2 = \{z: P_h(z) \leq P_h(z^t), \text{ for all } h \geq 0\}.

**Theorem 3.6.2.** Suppose

(a) \(h_0 = 0\),

(b) \(f(x, \theta) = f(x-\theta) = f(\theta-x), \text{ for all } x,\)

(c) \(g(h) = g(-h), \text{ for all } h.\)
Then, the Bayes decision is $\delta_1$ for all $z \in T_1$ and the Bayes decision is $\delta_2$ for all $z \in T_2$.

**Proof.** Consider $R(\delta_1, z)$ and $R(\delta_2, z)$ given by (3.8) and (3.9). Make the transformation $y = -h$ in the integral in (3.8). From (a) and (c), it follows that

$$R(\delta_1, z) - R(\delta_2, z) = C \int_0^\infty [P_{-y}(z) - P_y(z)]g(y)dy.$$  

From a result by Savage, Sobel and Woodworth (1965), given in lemma 4 of the appendix, it follows that

$$P_{-y}(z) = P_y(z^c),$$

for all $y$. Using (3.15) in (3.14), the theorem is proved.

It should be noted that if, in theorem 3.6.2 the sets $T_1$ and $T_2$ are redefined by replacing $z^c$ with $z^c$, the theorem is again true. This is so because $P_{-y}(z) = P_y(z^c)$ for all $y$ (see lemma 4 of the appendix).

**Example 3.6.1.** Suppose the conditions of theorem 3.6.2 are satisfied. Then for (a) $m = n = 2$ and (b) $m = 3$, $n = 2$, the rank orders, for which the Bayes decision is $\delta_1$ or $\delta_2$, are listed below.

(a) $m = n = 2$:

Bayes decision $\delta_1$: (0011) and (0101),
Bayes decision $\delta_2$: (0110), (1001), (1010) and (1100).

(b) $m = 3$, $n = 2$:

Bayes decision $\delta_1$: (00011), (00101), (01001) and (00110),
Bayes decision $\delta_2$: (10001), (01010), (01100), (10010),

(10100) and (11000).

Concluding this subsection it should be pointed out that if the conditions of theorem 2.9.1 are satisfied, then the risk of the
rank order Bayes two decision procedure goes to zero as m and n both go to infinity.

4. SIGNED RANK ORDER DATA.

Suppose \((X_1, Y_1), \ldots, (X_n, Y_n)\) are 2n mutually independent random variables. The \(X_i\)'s and the \(Y_i\)'s have continuous distribution functions \(F(x, \theta_1)\) and \(F(x, \theta_2)\) and the density functions \(f(x, \theta_1)\) and \(f(x, \theta_2)\) respectively. Define \(V_i = Y_i - X_i, i = 1, \ldots, n\), and denote the observed values of \(V_i\) by \(v_i\). Consider the ranking of the absolute values of \(v_i\) in the ascending order. Form a sequence \(z = (z_1, \ldots, z_n)\) by defining \(z_1 = 0\) if the \(i\)-th absolute value corresponds to a negative \(v\) and defining \(z_i = 1\) if the \(i\)-th absolute value corresponds to a positive \(v\).

The vector random variable \(Z\) corresponding to \(z\) is called a signed rank order. Whenever it does not cause any confusion, \(z\) will also be called a signed rank order. Denote the density function of \(V_i\)'s by \(p(x, \theta_1, \theta_2)\) and the probability that \(Z = z\) by \(P_{\theta_1, \theta_2}(z)\). Then, from Savage (1959), it follows that

\[
(4.1) \quad P_{\theta_1, \theta_2}(z) = n! \int \cdots \int_{0 < x_1 < \cdots < x_n < \infty} \prod_{i=1}^{n} p_i(x_i, \theta_1, \theta_2) p_{1-z_i}(-x_i, \theta_1, \theta_2) dx_i.
\]

It is assumed that \(p(x_i, \theta_1, \theta_2)\) depends on \(\theta_1\) and \(\theta_2\) through the function \(h(\theta_1, \theta_2)\). Then \(p(x_i, \theta_1, \theta_2)\) and \(P_{\theta_1, \theta_2}(z)\) are denoted by \(p(x_i, h)\) and \(P_h(z)\) respectively.

From the Bayesian viewpoint, \(h(\theta_1, \theta_2)\) is a given value of \(H = H(\theta_1, \theta_2)\) and \(P_h(z)\) is the conditional probability that \(Z = z\), given \(h\). Suppose \(H\) has the prior distribution function \(G(h)\). Then the posterior distribution function of \(H\), denoted by \(P[H \leq h | z]\) is given by
\[
\begin{align*}
    (4.2) \quad P[H \leq h|z] &= \frac{\int_{-\infty}^{x} P(z) dG(x)}{\int_{-\infty}^{\infty} P(z) dG(x)}.
\end{align*}
\]

All through the following discussion of the signed rank order Bayes estimation of \(H\), it is assumed that the loss function is squared error and that \(G(h)\) has finite second moment. Denote the Bayes estimate of \(H\) by \(h^*(z)\) and its risk by \(R(n)\). Then \(R(n)\) is finite. The Bayes estimate is given by

\[
    h^*(z) = \frac{\int_{-\infty}^{\infty} hP_h(z) dG(h)}{\int_{-\infty}^{\infty} P_h(z) dG(h)}
\]

and the Bayes risk is given by

\[
    R(n) = \int_{-\infty}^{\infty} h^2 dG(h) - \sum_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ h^*(z) \right]^2 P_h(z) dG(h).
\]

Section 4.1 gives some analytic properties of the posterior distribution of \(H\) and the Bayes estimate of \(H\). An example is given with \(n = 2\), when the sampled populations are normal. In section 4.2 a modification of the signed rank order Bayes estimation procedure is formulated.

4.1. **Analytic properties of the posterior distribution of \(H\) and the Bayes estimate of \(H\).** An analytic property of the posterior distribution function of \(H\) is obtained in theorem 4.1.1. Two conjectures regarding the monotonicity of the signed rank order likelihood ratio are given.

**Lemma 4.1.1.** If \(p(x,h) = p(x-h) = p(h-x)\) for all \(x\), then \(P_h(z) = P_{-h}(z^C)\).
Proof. From (4.1) it follows that

\[ P_h(z) = n! \int \cdots \int_{0 < x_1 < \cdots < x_n < \infty} \prod_{i=1}^{n} p_i(x_i-h)p_i(-x_i+h) \, dx_i \]

\[ = n! \int \cdots \int_{0 < x_1 < \cdots < x_n < \infty} \prod_{i=1}^{n} p_i(x_i+h)p_i(-x_i-h) \, dx_i \]

\[ = P_{-h}(z^c). \]

The condition of lemma 4.1.1 is satisfied if the populations sampled belong to a normal family with a translation parameter.

Theorem 4.1.1. Suppose that

(a) \( p(x, h) = p(x-h) = p(h-x) \) for all \( x \),

(b) \( G(h) = 1 - G(-h) \) for all \( h \),

then the posterior distribution function of \( H \) satisfies

\[ P[H \leq h \mid z] = 1 - P[H \leq -h \mid z^c]. \]

Proof. Consider \( P[H \leq h \mid z] \) given by (4.2). Make the transformation \( y = -x \) in both the integrals. Then

\[ P[H \leq h \mid z] = \frac{\int_{-y}^{-h} P_{-y}(z) dG(-y)}{\int_{-y}^{\infty} P_{-y}(z) dG(-y)}. \]

Replace \( P_{-y}(z) \) by \( P_{-y}(z) \) by using lemma 4.1.1. Then due to (b) the theorem follows.

Corollary 4.1.1. If the conditions of theorem 4.1.1 are satisfied then

\[ h^*(z) = - h^*(z^c). \]

Example 4.1.1. Let \( F(\cdot) \) and \( f(\cdot) \) denote the standard normal distribution
function and the density function respectively. Let the \( X_i \)'s and the \( Y_i \)'s have the distribution functions \( F((x - \theta_1)/\sigma) \) and \( F((x - \theta_2)/\sigma) \) respectively where \( \sigma \) is known. In this case \( H = \theta_2 - \theta_1 \). Suppose \( H \) has the prior distribution function \( F(h/\eta) \) and denote \( \eta/\sigma \) by \( k \).

Consider \( n = 2 \). Then, after some manipulation it is found that

\[
P_h(11) = F^2\left( \frac{h}{\sigma\sqrt{2}} \right),
\]

\[
P_h(10) = 2F\left( \frac{h}{\sigma\sqrt{2}} \right) - F\left( \frac{h}{\sigma} \right) - F^2\left( \frac{h}{\sigma\sqrt{2}} \right),
\]

\[
P_h(01) = F\left( \frac{h}{\sigma} \right) - F^2\left( \frac{h}{\sigma\sqrt{2}} \right),
\]

\[
P_h(00) = F^2\left( \frac{-h}{\sigma\sqrt{2}} \right).
\]

Whenever \( z \) and \( z' \) are such that \( z_i \geq z'_i \), \( i = 1, 2 \), it is found that \( P_h(z_1,z_2)/P_h(z'_1,z'_2) \) is an increasing function of \( h \). Therefore, the ratio, \( g(h|z)/g(h|z') \), of the posterior densities of \( H \) is an increasing function of \( h \), showing that \( h^*(z) \geq h^*(z') \). It should be pointed out that the inequality holds for any prior density for \( H \). Corollary 4.1.1 and lemma 2 of the appendix give

\[
h^*(11) = -h^*(00) = \frac{2nk\sqrt{2\pi}}{\sqrt{k^2+2}} \frac{1}{\pi + 2 \arcsin \left( \frac{k^2}{(k^2+2)} \right)},
\]

\[
h^*(01) = h^*(10) = \frac{2nk\sqrt{2\pi}[\sqrt{k^2+2} - \sqrt{k^2+1}]}{[\pi - 2 \arcsin \left( \frac{k^2}{(k^2+2)} \right)]\sqrt{(k^2+1)(k^2+2)}}
\]

and

\[
R(2) = \eta^2 - \frac{4\eta^2k^2}{(k^2+2)[\pi + 2 \arcsin \left( \frac{k^2}{(k^2+2)} \right)]}
\]

(\text{cont.})
\[- \frac{4 \pi^2 k^2 \left( \sqrt{k+2} - \sqrt{k+1} \right)^2}{(k+1)(k+2)[\pi-2 \arcsin\left(\frac{k}{k+2}\right)]^2}\]

Notice that the first two terms in $R(2)$ give the Bayes risk of the paired comparison Bayes estimation procedure with $n = 2$. Thus, the third term in $R(2)$ gives the decrease in the risk of the Bayes estimate of $\theta$ from the paired comparison data when the signs of the comparisons are taken into consideration. If the $X_i$'s and the $Y_i$'s were observable random variables, the Bayes risk $R_0(2)$ would be $\eta^2/(k+1)$. The efficiency is given in the following table.

Table 4.1.1. The efficiency of the signed rank order Bayes estimation.

<table>
<thead>
<tr>
<th>$k$</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R(2)$</td>
<td>1.0000</td>
<td>.9899</td>
<td>.9612</td>
<td>.9176</td>
<td>.8630</td>
<td>.8016</td>
<td>.7369</td>
</tr>
<tr>
<td>$k$</td>
<td>1.4</td>
<td>1.6</td>
<td>1.8</td>
<td>2.0</td>
<td>3.0</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>$R(2)$</td>
<td>.6713</td>
<td>.6085</td>
<td>.5486</td>
<td>.4925</td>
<td>.2885</td>
<td>.1783</td>
<td>.1180</td>
</tr>
</tbody>
</table>

As pointed out in example 4.1.1 for $n = 2$, it can be shown that, for any $n$, if $z$ and $z'$ are such that $P_h(z)/P_h(z')$ is an increasing function of $h$, then $h^*(z) \geq h^*(z')$. In analogy to results of Saxena (1965), two unproven conjectures regarding the monotonicity of the likelihood ratio of the signed rank orders are being given. It is doubtful that these conjectures are true. Yet they are being reported since they point to directions in which further study is needed.

**Conjecture I.** Suppose that $f(x, \theta)$ has an increasing likelihood ratio. If $z$ and $z'$ are such that $z_i \geq z'_i$, $i = 1, \ldots, n$, then $P_{\theta_1, \theta_2}(z)/P_{\theta_1, \theta_2}(z')$ is increasing in $\theta_2$ and decreasing in $\theta_1$. 
Conjecture II. Suppose that \( f(x, \theta) \) has an increasing likelihood ratio. Then \( P_{\theta_1, \theta_2}(z)/P_{\theta_1, \theta_2}(z') \) is increasing in \( \theta_2 \) and decreasing in \( \theta_1 \) whenever \( z \leq z' \).

4.2. A modified formulation for the Bayes estimation from the signed rank order data. Consider the formulation of the Bayes estimation problem given in section 4. Suppose \( n = 2 \). Then the four possible signed rank orders are: (00), (01), (10) and (11). The sampling plan is modified in the following manner. Consider a set of two pairs of observations as a sample point and then consider a sample of size \( m \). Suppose in the sample, the signed rank orders (00), (01), (10) and (11) are observed \( r_1, r_2, r_3 \) and \( m - r_1 - r_2 - r_3 \) times respectively.

Denote the random variables corresponding to \( r_i \) by \( R_i \), \( i = 1, 2 \) and 3. The multinomial distribution of \( R_1, R_2 \) and \( R_3 \) is given by

\[
P_h(r_1, r_2, r_3) = \frac{r_1! r_2! r_3!}{m!} \frac{P_1(00) P_2(01) P_3(10) P_4(11)}{r_1! r_2! r_3! r_4!},
\]

where \( r_4 = m - r_1 - r_2 - r_3 \). Then the Bayes estimate of \( H \), denoted by \( h^*(r_1, r_2, r_3) \) is

\[
h^*(r_1, r_2, r_3) = \frac{\int hP_h(r_1, r_2, r_3)dG(h)}{\int P_h(r_1, r_2, r_3)dG(h)}.
\]

The Bayes risk, denoted by \( R(m) \), is

\[
R(m) = \int h^2 dG(h) - \sum \int \left[ h^* (r_1, r_2, r_3) \right]^2 P_h(r_1, r_2, r_3)dG(h),
\]
where \( \Sigma' \) denotes summation over all possible combinations of \( r_1, r_2 \) and \( r_3 \).

The reason for the modified formulation of the signed rank order Bayes estimation problem is that the monotonic properties of the ratios of the posterior densities can be studied by studying the signed rank order likelihood ratios of the form \( \frac{P_h(z_1, z_2)}{P_h(z_1', z_2')} \).

Example 4.1.1 can be treated as an example of the modified formulation of this subsection with \( m = 1 \). As for the rank order data, the Bayes risk of the signed rank order estimation procedure goes to zero as \( n \) goes to infinity if the conditions of theorem 2.5.1 are satisfied.
APPENDIX

The following results are about analytic properties of the normal distribution, likelihood ratios and probabilities of rank orders. Some of these results are known and have been included for the sake of facility as they are used in sections 2, 3 and 4.

Lemma 1 is an extension of a result of David (1953) for normal distributions to distributions symmetric about the origin.

**Lemma 1.** If \( F(x) = 1 - F(-x) \), for all \( x \), then for \( n \) odd and all \( a_i \neq 0 \),

\[
\int_{-\infty}^{\infty} \prod_{i=1}^{n} F(a_i x) dF(x) = \frac{1}{2} \left[ 1 - \sum_{i=1}^{n} \int_{-\infty}^{\infty} F(a_i x) dF(x) \right] + \sum_{1 \leq i < j \leq n} \int_{-\infty}^{\infty} F(a_i x) F(a_j x) dF(x) + \ldots
\]

\[
+ \sum_{j=1}^{n} \int_{-\infty}^{\infty} \prod_{i=1}^{n} F(a_i x) dF(x) \right] .
\]

**Proof.** Write \( F(a_i x) = 1 - F(-a_i x), i = 1, \ldots, n \), in \( \prod_{i=1}^{n} F(a_i x) \) and expand the product. As

\[
\int_{-\infty}^{\infty} \prod_{i=1}^{n} F(-a_i x) dF(x) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} F(a_i x) dF(x),
\]

for all \( n \) and all \( a_i \), the lemma follows.

**Lemma 2.** Let \( F(\cdot) \) and \( f(\cdot) \) denote the standard normal distribution function and the density function respectively. Then,
(a) \( f(ax+b)f(cx+d) \)

\[ = f(\frac{ad - bc}{\sqrt{a^2 + c^2}} f(\frac{\sqrt{a^2 + c^2}}{a + c^2} (x + \frac{a^2 + cd}{a + c^2})), \]

(b) \( \int_{-\infty}^{\infty} f(ax+b)f(cx+d)dx = \frac{1}{\sqrt{2 \pi}} f(\frac{ad - bc}{\sqrt{a^2 + c^2}}), \]

(c) \( \int_{-\infty}^{\infty} F(ax+b)f(x)dx = F(\frac{b}{\sqrt{1+a^2}}), \)

(d) \( \int_{-\infty}^{\infty} F(ax)F(bx)F(cx)f(x)dx \)

\[ = \frac{1}{8} + \frac{1}{4\pi} \left[ \arcsin \frac{ab}{\sqrt{(1+a^2)(1+b^2)}} + \arcsin \frac{ac}{\sqrt{(1+a^2)(1+c^2)}} \right. \]

\[ + \left. \arcsin \frac{bc}{\sqrt{(1+b^2)(1+c^2)}} \right], \]

(e) \( \int_{-\infty}^{\infty} xF(ax+b)F(cx+d)f(x)dx \)

\[ = \frac{a}{\sqrt{1+a^2}} f(\frac{b}{\sqrt{1+a^2}}) F(\frac{d(1+a^2) - abc}{\sqrt{(1+a^2)(1+a^2+c^2)}}) \]

\[ + \frac{c}{\sqrt{1+c^2}} f(\frac{d}{\sqrt{1+c^2}}) F(\frac{b(1+c^2) - acd}{\sqrt{(1+c^2)(1+a^2+c^2)}}), \]

(f) \( \int_{-\infty}^{\infty} xF(ax)F(bx)F(cx)f(x)dx \)

\[ = \frac{a}{\sqrt{2\pi(1+a^2)}} \left[ \frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{bc}{\sqrt{(1+a^2+b^2)(1+a^2+c^2)}} \right] \]

\[ + \frac{b}{\sqrt{2\pi(1+b^2)}} \left[ \frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{ac}{\sqrt{(1+b^2+a^2)(1+b^2+c^2)}} \right]. \]

(cont.)
\[ + \frac{c}{\sqrt{2\pi(1+c^2)}} \left[ \frac{1}{4} \frac{1}{2\pi} \arcsin \frac{ab}{\sqrt{(1+c^2+a^2)(1+c^2+b^2)}} \right]. \]

**Proof.** (a) and (b) are immediate. (c) is a known result and can be obtained by a probability argument. (d) is obtained from lemma 1 and a result due to Sheppard and given in Cramér (1946). (e) and (f) are obtained by integration by parts and using the previous results.

Notice that the integral (d) in lemma 2 can be written as

\[ P[Y_1 < 0, Y_2 < 0, Y_3 < 0], \] where \( Y_1, Y_2 \) and \( Y_3 \) are normally distributed random variables with zero means, variances unity and correlations

\[ r_{12} = \frac{ab}{\sqrt{(1+a^2)(1+b^2)}}, \]
\[ r_{13} = \frac{ac}{\sqrt{(1+a^2)(1+c^2)}} \]
\[ r_{23} = \frac{bc}{\sqrt{(1+b^2)(1+c^2)}}. \]

**Lemma 3.** Let \( F(\cdot) \) and \( f(\cdot) \) denote the standard normal distribution function and the density function respectively. Then \( f^2(x)/F(x)F(-x) \) is a decreasing function of \( x \) for \( x \geq 0 \).

For the proof see Sampford (1953).

**Lemma 4.** Let \( X_1, \ldots, X_m; Y_1, \ldots, Y_n \) be mutually independent random variables. The \( X_i \)'s have the density function \( f(x) \) and the \( Y_i \)'s have the density function \( f(x-\theta) \). If \( f(x-\theta) = f(\theta-x) \) for all \( x \), then the following results hold for all \( \theta \): \( P_{\theta}(z) = P_{-\theta}(z^t) = P_{-\theta}(z^c) \).

For the proof, see theorems 1 and 2 of Savage, Sobel and Woodworth (1965).

**Lemma 5.** Let \( X_1, X_2; Y_1, Y_2 \) be mutually independent random variables. The \( X_i \)'s have the density function \( f(x) \) and the \( Y_i \)'s have the density function \( g(x) \). Then, \( P(0101) = 2P^2(01) - 2P(0011) \).
For the proof see theorem 9 of appendix IV of Savage, Sobel and Woodworth (1965).

Lemma 6. If a family of density functions \( f(x, \theta) \) has increasing likelihood ratio, then for any prior density \( g(\cdot) \) for \( \theta \), the family of posterior densities for \( \theta \) has increasing likelihood ratio.

Proof. The posterior density, denoted by \( g(\theta|x) \) is given by

\[
g(\theta|x) = \frac{f(x, \theta)g(\theta)}{\int_{-\infty}^{\infty} f(x, \theta)g(\theta) d\theta}
\]

\[= c(x)f(x, \theta)g(\theta),\]

where \( c \) is the reciprocal of the integral. Consider any \( x > x' \) and \( \theta > \theta' \). Then,

\[
g(\theta|x)g(\theta'|x') - g(\theta|x')g(\theta'|x)
\]

\[= c(x)c(x')\{f(x, \theta)f(x', \theta') - f(x, \theta')f(x', \theta)\}
\]

\[\geq 0.
\]

Hence the lemma is proved.

If for a random sample of size \( n \), there exists a sufficient statistic \( T(x_1, \ldots, x_n) \) for \( \theta \), then for any prior density for \( \theta \), the posterior densities are indexed by the set of values of the sufficient statistic. Further, if the family of density functions for the sufficient statistic \( T \) has increasing likelihood ratio, then lemma 6 applies.

Lemma 7. Let \( P_0 \) and \( P_1 \) be two distributions with densities \( p_0 \) and \( p_1 \) respectively such that \( p_1(x)/p_0(x) \) is an increasing function of \( x \).

Then, \( E_0 \psi(x) \leq E_1 \psi(x) \) for any increasing function \( \psi \), where \( E_\psi(x) \) denotes the expectation of \( \psi(x) \) with respect to \( P_\psi \).

Lemma 8. If the prior distribution of a parameter $\theta$ has finite second moment, then the Bayes estimate of $\theta$, with respect to the squared error loss function, has finite risk.

The proof is immediate.
ACKNOWLEDGEMENT. The author is greatly indebted to his advisor, Professor I. Richard Savage, who introduced him to the fields of Bayesian inference and nonparametric methods and through whose patient help, inspiration and guidance this work has come to its present form. Gratitude is expressed here to Professor R. A. Bradley for providing facilities for the author's work in the Department of Statistics of the Florida State University. Thanks are due to Professor Fred C. Leone of Case Institute of Technology for providing opportunities to the author to do this work while teaching in the Institute. The author is thankful to Professors B. Lindgren and M. Sobel for acting as advisors during the period the author was not a resident student at the University of Minnesota. The financial help during this work came from the Graduate School of the University of Minnesota and the Office of Naval Research and gratitude for this help is recorded here. Acknowledgement is recorded here for the help of Mr. S. S. Chitgopek in checking some of the computations.
REFERENCES


ment of Statistics, Technical Report No. 29 under the names of the first two authors).

Two sample nonparametric decision problems for single parameter families of distributions are considered from the Bayesian viewpoint when only the relative magnitudes of the observation are known for the paired comparison data, the rank order data and the signed rank order data. The term nonparametric is used as the decision procedures depend on nonparametric statistics. It is assumed that the likelihood functions for the three kinds of data depend on the parameters of the sampled populations through some function $H$. A prior distribution is considered for the random variable $H$.

The problem of the Bayes estimation of $H$, with the squared error loss function, is considered for the three kinds of data. A number of analytic properties of the posterior distribution of $H$ and the orderings of the values of the Bayes estimate of $H$ are obtained. Examples are given from a normal and a uniform family of distributions. Sufficient conditions are given on the sampled populations for the risk of the Bayes estimate of $H$ to go to zero as the sample sizes go to infinity for the three kinds of data.

The Bayes two decision problem, the two decisions being $H > h$ and $H \leq h$, with the $(0,1)$ loss function, is considered for the paired comparison data and the rank order data. Monotonic properties of the Bayes decision procedures are obtained under mild conditions on the sampled populations. Sufficient conditions on the sampled populations are given for the risk of the Bayes two decision procedures to go to zero as the sample sizes go to infinity.
<table>
<thead>
<tr>
<th></th>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Role</td>
<td>Role</td>
<td>Role</td>
</tr>
<tr>
<td></td>
<td>Wt</td>
<td>Wt</td>
<td>Wt</td>
</tr>
<tr>
<td>Mathematical Statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two sample problem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonparametric Bayes estimation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonparametric Bayes decision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired comparison</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank order</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signed rank order</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal distribution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform distribution</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>