THE MINIMAX PRINCIPLE AND THE UNBIASEDNESS PRINCIPLE
CONSIDERED IN THE LIGHT OF A 2-STEP MULTIPLE DECISION
PROBLEM

By

Harald Goldstein
Lehmann [2], [3] has developed a technique for constructing decision procedures which uniformly minimize the risque among the unbiased procedures for certain types of multiple decision problems. He also gives examples where this technique is not applicable. Such an example is treated in Sverdrup [1] (p 24-30) from the point of view of the performance function. Here the same example shall be treated from the point of view of various loss functions. The main task of this paper will be to throw some light on some aspects of the minimax principle and the unbiasedness principle by means of this example.
l. The Decision Problem and Loss Functions.

(1.1) The problem:

Let \( X_{ijv} \) be independent, normally distributed variables where \( \text{var} X_{ijv} = \sigma^2 \) (unknown) and \( \text{EX}_{ijv} = \left\{ \alpha_i + \beta_j + \gamma_{ij} \right\} \), \( i, j \in \{1, 2\} \); \( v \in \{1, 2, \ldots, m\} \).

The number of variables are \( n = 4m \). The side conditions are

\[
\begin{align*}
\alpha_1 &= -\alpha_2 = \alpha \\
\beta_1 &= -\beta_2 = \beta \\
\gamma_{11} &= -\gamma_{12} = -\gamma_{21} = \gamma_{22} = \gamma
\end{align*}
\]

We want to investigate whether there is some interaction or not; that is, we have a 3-decision problem where we want to decide between \( \gamma < 0 \), \( \gamma > 0 \) and not saying anything about \( \gamma \).

If we do not find support for \( \gamma < 0 \) or \( \gamma > 0 \) we want to test the hypothesis that \( \alpha \leq 0 \) against the alternative that \( \alpha > 0 \).

On the other hand if we find support for \( \gamma < 0 \) or \( \gamma > 0 \) we are not interested in testing any hypothesis about \( \alpha \).

This leads to the following decision space

\[
D = \{ d_1, d_2, d_3, d_4 \} = \{ "\Theta", "\alpha > 0", "\gamma < 0", "\gamma > 0" \}
\]

"\( \Theta \)" means that we are not making any statement about the parameters, \( \Theta \subseteq \mathbb{R}^5 \) being the parameter space.

A decision problem like this where it is possible to make more or less specific conclusions is called free.

It is natural from the situation described above to call the problem a 2-step decision problem where the first step consists of a 3-decision problem and the second one of an ordinary problem of hypothesis testing. This will mainly influence the loss function or the requirements one wants to put on the performance function.

Suppose too that we want to be as "reasonable" as possible in our conclusions on step 2 even if we made a mistake on step 1.
This wish will also influence the loss function or the requirements on the performance function.

**Notation:** Since $D$ is finite any decision function can in this case be thought of as a function:

$$\delta(x) = (\varphi_1(x), \varphi_2(x), \varphi_3(x), \varphi_4(x))$$

$$(\delta = \delta(\varphi_1, \ldots, \varphi_4))$$

where $\varphi_i : \mathcal{X} \rightarrow [0, 1], i = 1, \ldots, 4$ are measurable functions defined on the sample space $\mathcal{X}$, fulfilling

$$\sum_{i=1}^{4} \varphi_i(x) = 1.$$ 

$\varphi_i(x)$ is then interpreted as the conditional probability of making the decision $d_i$ given $X = x$, $X$ being the stochastic vector observed.

The performance function $(\mathcal{C}_\delta, \Theta)$ is for all $\Theta \in \Theta$ and $\delta$ a probability distribution over $D$ given by

$$E_{\Theta} \delta(X) = (E_{\Theta} \varphi_1(X), \ldots, E_{\Theta} \varphi_4(X))$$

Thus $(\mathcal{C}_\delta, \Theta)(d_i) = \Pr(d_i | \delta, \Theta) = E_{\Theta} \varphi_i(X); i = 1, \ldots, 4$

A loss function is a function

$$W : \Theta \times D \rightarrow [0, \infty).$$

We are here only going to consider simple loss functions, eg. $W(\Theta, d_i) = w_{ij}$ for all $\Theta \in \Theta_j; i = 1, \ldots, 4, j = 1, \ldots, r$ where $\{\Theta_1, \ldots, \Theta_r\}$ is a finite partition of $\Theta$.

The risque function becomes

$$r(\Theta, \delta) = \int W(\Theta, t) \mathcal{C}_\delta, \Theta(\mathrm{d}t) =$$

$$= \sum_{i=1}^{4} W(\Theta, d_i) E_{\Theta} \varphi_i (X)$$

The generalized risque (see for instance Sverdrup [1]) is given by:

$$r(\Theta, \Theta', \delta) = \sum_{i=1}^{4} w(\Theta', d_i) E_{\Theta} \varphi_i (X).$$
We are now going to construct some loss functions which should accord with the description of the problem given above.

It is common in a free decision problem to consider two kinds of errors, e.g., errors of the first kind which are wrong statements and errors of the second kind which are not wrong statements but owe to the fact that one could have been more specific in the conclusions.

When constructing and judging different decision rules there are two types of general principles that are usually used: (a) in terms of the performance function and (b) in terms of the loss function. In the case of (a) it is customary to try to maximize the probabilities of correct statements in some sense or other under certain side conditions on the probability of errors of the first kind. In Sverdrup [1] the example above is treated from the aspect in (a). We will try to discuss the example with the main weight on (b).

Referring to the loss function, the fact that the decision problem is free, may be expressed by making the losses due to errors of the first kind "essentially" greater than those due to errors of the second kind. The meaning of "essentially greater" will usually depend on the circumstances of the investigation. It may also depend on a discussion corresponding to the one given in this paper.

The wish to be reasonable at step two in case of a false statement at step one may be expressed by adding the corresponding losses from the two steps (see Lehmann [3], p 558). Thus, let the loss function at step one \((W'(\theta, d))\) be the following (Sverdrup [1], p 70):
(W' is made slightly less general here than in [1] for the sake of simplicity. This restriction, however, is not essential to the discussion below.)

The loss function at step two \((W''(\theta, d))\):

\[
\begin{array}{c|c|c}
\text{dec.} & "-\infty<\gamma<\infty" & "\gamma>0" \\
\hline
\text{nature} & b_1 & a_1+b_1 \\
\gamma = 0 & b_1 & a_1+b_1 \\
\gamma > 0 & b_1 & 0 \\
\end{array}
\]

\((b_1 < a_1)\)

Following Lehmann [3], p 558, the total loss function \((W_1(\theta, d))\) becomes:

\[
\begin{array}{c|c|c|c|c|c|c}
\text{dec.} & d_1 & d_2 & d_3 & d_4 \\
\hline
\text{nature} & "H" & "\alpha>0" & "\gamma<0" & "\gamma>0" \\
\hline
H_1 & \gamma=0, \alpha \leq 0 & 0 & a_2 & a_1 & a_1 \\
H_2 & \gamma=0, \alpha > 0 & b_2 & 0 & a_1 & a_1 \\
H_3 & \gamma<0, \alpha \leq 0 & b_1 & a_2+b_1 & 0 & a_1+b_1 \\
H_4 & \gamma>0, \alpha \leq 0 & b_1 & a_2+b_1 & a_1+b_1 & 0 \\
H_5 & \gamma>0, \alpha > 0 & b_1+b_2 & b_1 & a_1+b_1 & 0 \\
H_6 & \gamma<0, \alpha > 0 & b_1+b_2 & b_1 & 0 & a_1+b_1 \\
\end{array}
\]
The loss matrix we write:

\[
W_1 = \begin{pmatrix}
    w_{11} & \cdots & w_{14} \\
    \vdots \\
    w_{61} & \cdots & w_{64}
\end{pmatrix}
\]

As an alternative loss function we propose the following modification: \( W_2 (\theta, d) \) is the same as \( W_1 (\theta, d) \) except that \( w_{23} = a_1 \) and \( w_{24} = a_1 \) are replaced by \( w_{23} = a_1 + b_2 \) and \( w_{24} = a_1 + b_2 \). The reason for this might be: If we make the decision "\( \gamma < 0 \)" ("\( \gamma > 0 \)") whereas \( \gamma = 0 \) and \( \alpha > 0 \) we are not only saying something false about \( \gamma \), but the consequence of this is that we are losing the possibility of making a true statement about \( \alpha \). Therefore the loss ought to be greater than just \( a_1 \).

This is not the case if for example \( \gamma > 0, \alpha > 0 \) and we state that \( \gamma < 0 \), for here \( \gamma \neq 0 \) and we are then not interested in \( \alpha \) a priori.

In order to be able to discuss \( W_1 \) and \( W_2 \) simultaneously and also possibly to see the importance of the addition itself of the losses from the two steps we are introducing a more general loss function \( W_3 (\theta, d) \):

<table>
<thead>
<tr>
<th>dec.</th>
<th>nat.</th>
<th>&quot;( \gamma = 0 )</th>
<th>&quot;( \alpha &gt; 0 )&quot;</th>
<th>&quot;( \gamma &lt; 0 )&quot;</th>
<th>&quot;( \gamma &gt; 0 )&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = 0 )</td>
<td>( \alpha \leq 0 )</td>
<td>0</td>
<td>( t_7 )</td>
<td>( t_4 )</td>
<td>( t_4 )</td>
</tr>
<tr>
<td>( \gamma &gt; 0 )</td>
<td>( \alpha &gt; 0 )</td>
<td>( t_2 )</td>
<td>0</td>
<td>( t_5 )</td>
<td>( t_5 )</td>
</tr>
<tr>
<td>( \gamma &lt; 0 )</td>
<td>( \alpha \leq 0 )</td>
<td>( t_1 )</td>
<td>( t_8 )</td>
<td>0</td>
<td>( t_6 )</td>
</tr>
<tr>
<td>( \gamma &gt; 0 )</td>
<td>( \alpha &lt; 0 )</td>
<td>( t_1 )</td>
<td>( t_8 )</td>
<td>( t_6 )</td>
<td>0</td>
</tr>
<tr>
<td>( \gamma &lt; 0 )</td>
<td>( \alpha &gt; 0 )</td>
<td>( t_3 )</td>
<td>( t_1 )</td>
<td>( t_6 )</td>
<td>0</td>
</tr>
</tbody>
</table>
The losses due to errors of the first kind are $t_4$, $t_5$, $t_6$, $t_7$, $t_8$ and to errors of the second kind: $t_1$, $t_2$, $t_3$. Let $b = \max \{ t_1, t_2, t_3 \}$ and $a = \min \{ t_4, t_5, t_6, t_7, t_8 \}$. Since the decision problem is free it is natural to suppose that $b < a$.

We call $W_1(\theta, d)$ Lehmann's loss function, $W_2(\theta, d)$ the modified one and $W_3(\theta, d)$ the general one. Let $W_1, W_2, W_3$ denote the corresponding loss matrixes.

We are in the next sections going to study $W_3$ under the following restrictions:

$$b < a, \quad t_7 \leq t_8, \quad t_3 > t_2 - \frac{t_1}{t_8} (t_2 - t_1),$$

$$b = t_3 \quad \text{and} \quad t_6 > 2t_3$$

If $a_1 > 2b_2 + b_1$ both $W_1$ and $W_2$ fulfill (1.5).

2. The Minimax Principle.

The problem concerns two different parameters $(\alpha, \gamma)$. Part of the motivation for proposing the decision space $D$ in section 1 is supposed to be an interest in both of the parameters for their own sake. (See Sverdrup [1] p 29-30 for corresponding remarks.) As for $\alpha$ this is expressed through the wish to be reasonable at step two even if a mistake was made at step one. On the other hand, the purpose of testing hypotheses concerning $\gamma$ is supposed to be (a) an interest in $\gamma$ for its own sake and (b) to motivate an interest in $\alpha$ (the interest depends on the result at step one).

This interest in both $\alpha$ and $\gamma$ might be the reason for not wanting to use a decision rule which with probability 1 does not say anything about one or more of these parameters:

1. A rule which with probability 1 does not say anything about $\alpha$ would for example never state that $\alpha > 0$ if actually $\alpha > 0$ and $\gamma = 0$, and this is incompatible with the interest in $\alpha$. 
(ii) A rule which with probability 1 does not say anything about \( \gamma \), would for example never discover it if \( \gamma \) actually is greater than 0, and this goes against (a) above.

If on the other hand the purpose of testing hypotheses about \( \gamma \) only was (b), there is no reason why one should object against a decision rule of type (ii). Such a rule would then have to compete on equal terms with other decision rules for example with respect to the risk function.

We have now motivated the following definition: (Only \( d_2 = "\alpha > 0" \) says something about \( \alpha \) and only \( d_3 = "\gamma < 0" \) and \( d_4 = "\gamma > 0" \) about \( \gamma \).)

**Definition 1**
A decision function
\[
\delta = \delta(\varphi_1, ..., \varphi_4)
\]
is called tolerant if both
\[
\Pr (d_2 | \delta, \vartheta) > 0 \quad \text{and} \quad \Pr (d_3 | \delta, \vartheta) + \Pr (d_4 | \delta, \vartheta) > 0
\]
(It is not possible in our model that \( \Pr (d_i | \delta, \vartheta) > 0 \) for some \( \vartheta \) and = 0 for others.)

It will be shown in this section that the minimax principle leads to intolerant decision rules.

Our example is a special case of the following situation:

\begin{equation}
(2.1) \text{Consider the experiment } (\mathcal{X}, A, P_\vartheta; \vartheta \in \mathcal{H}) \text{ where } \mathcal{X} \text{ is the sample space, } A \text{ a } \sigma\text{-field in } \mathcal{X} \text{ and } P_\vartheta \text{ a probability measure over } (\mathcal{X}, A). \text{ } \mathcal{F} = \{P_\vartheta; \vartheta \in \mathcal{H}\} \text{ is assumed to be equivalent with a } \sigma\text{-finite measure } \mu \text{ (e.g. } \mathcal{F} \ll \mu \text{ and } \mu \ll \mathcal{F}). \text{ Let the parameter space } \mathcal{H} \subseteq \mathbb{R}^r \text{ (the } r\text{-dimensional euclidian space). A } \vartheta \in \mathcal{H} \text{ we write } \vartheta' = (\gamma, \alpha, \theta_3, ..., \theta_r) \text{ where } \gamma \in I_1, \alpha \in I_2; \text{ } I_1, I_2 \text{ being intervals with zero as an inner point. Assume that } E_\vartheta \varphi(X) \text{ is a continuous function of } \vartheta \text{ for any test function } \varphi.\end{equation}
Let the decision space be as before \( D \) and let the loss function be given by \( W_3(\theta, d) \):

\[
W_3 = \begin{pmatrix}
0 & t_7 & t_4 & t_4 \\
t_2 & 0 & t_5 & t_5 \\
t_1 & t_8 & 0 & t_6 \\
t_1 & t_8 & t_6 & 0 \\
t_3 & t_1 & t_6 & 0 \\
t_3 & t_1 & 0 & t_6
\end{pmatrix} = (a_1, a_2, a_3, a_4) = \begin{pmatrix}
\xi_1' \\
\xi_2' \\
\vdots \\
\xi_6'
\end{pmatrix}
\]

(2.2) Let \( G \) be the set of strategies for player I and \( H \) for player II in the game with loss matrix \( W_3 \). Let \( G^0 \) be the set of optimal strategies for I (maximin strategies) and \( H^0 \) the corresponding one for II (minimax strategies).

For future reference we state some results and notation from Schaafsma [4]: Let the decision space be \( D' = \{d_1, \ldots, d_n\} \) and \( \overline{H} = \overline{H}_1 \cup \ldots \cup \overline{H}_m \) (disjoint union). The loss function \( W : H \times D' \rightarrow [0, \infty) \) is simple, and a decision function is described by \( \delta = \delta(\varphi_1, \ldots, \varphi_n) \) where the \( \varphi_i : \mathcal{X} \rightarrow [0, 1] \) are measurable and \( \sum_{i=1}^{n} \varphi_i(x) = 1 \) \( \forall x \in \mathcal{X} \). Then the following definition (Schaafsma [4]) covers our situation:

**Definition 2** A multiple decision problem is said to be of type I if

1) \( E_\Theta \varphi(X) \) is a continuous function of \( \Theta \) for every test function \( \varphi \) and

2) \( \overline{H}_o = \overline{H}_1 \cap \overline{H}_2 \cap \ldots \cap \overline{H}_m \neq \emptyset \),

(\( \overline{H}_i \) denotes the closure of \( H_i \))

The loss matrix is \( W = \begin{pmatrix}
w_{11} & \cdots & w_{1n} \\
\vdots & \ddots & \vdots \\
w_{ml} & \cdots & w_{mn}
\end{pmatrix} = (w_1, \ldots, w_n) \)
and the risque \( r(\theta, \delta) = \sum_{j=1}^{n} w_{ij} E_{\theta} \phi_{j}(X) \)
for \( \theta \in \Omega_i \); \( i = 1, 2, \ldots, m \).

\( \delta^* \) has minimax risque \( (\delta^* \text{ is MR}) \)
if \( \sup_{\theta \in \Omega} r(\theta, \delta^*) = \inf_{\delta} \sup_{\theta \in \Omega} r(\theta, \delta) \).

(2.3) Put \( S(W) = \{ w \mid w \in \mathbb{R}^m; \ w = \sum_{j=1}^{n} p_j w_j, \ p_j \geq 0, \ \sum p_j = 1 \} \).

\( S(W) \) is then a convex set in \( \mathbb{R}^m \).

\[
\begin{pmatrix}
  w_1^* \\
  \vdots \\
  w_m^*
\end{pmatrix} \in S(W) \text{ is called a }
\]

(2.4) minimax point in \( S(W) \) if

\[ \forall \ w = \begin{pmatrix}
  w_1 \\
  \vdots \\
  w_m
\end{pmatrix} \in S(W); \ \max_{i} w_i^* \leq \max_{i} w_i. \]

(We have that \( w^* = \sum_{j=1}^{n} p_j^* w_j \) is a minimax point in \( S(W) \)),

\[ (p_1^*, \ldots, p_n^*) \in \mathcal{H}_0. \]

\[
\begin{pmatrix}
  w_1^* \\
  \vdots \\
  w_m^*
\end{pmatrix} \in S(W) \text{ is called a maximin point }
\]

(2.5) \[ \forall \ w \in S(W); \ \min_{i} w_i^* > \min_{i} w_i \]

The following theorem (Schaafsma [4]) shows a certain connection between optimal strategies for II and MR rules.

**Theorem 2.1** For m.d.p.'s of type I the following holds:

(1) \((p_1^*, \ldots, p_n^*) \in \mathcal{H}_0 \Rightarrow \delta_1 = \delta(\varphi_1^{(1)}, \ldots, \varphi_n^{(1)}), \)

defined by \( \varphi_j^{(1)}(x) = p_j^* \) for \( j = 1, \ldots, n, \)

is a MR-rule.
(ii) If $\delta^* = \delta(f_1^*, \ldots, f_n^*)$ is MR then
\[ \forall \theta_o \in \Theta_o = \Theta_1 \cap \ldots \cap \Theta_m ; (p_1^*(\theta_o), \ldots, p_n^*(\theta_o)) \in H^o, \text{where we have put } p_j^*(\theta_o) = E_{\Theta_o} f_j^*(X); j=1,\ldots, n. \]

When determining $H^o, G^o$, the following lemma can be useful (Schaafsma [4], lemma 3.1):

**Lemma 2.2** \( w^* = \sum_{j=1}^{n} p_j^* w_j \) is a minimax point in $S(W)$ if and only if \( \exists g' = (g_1^*, \ldots, g_m^*) \in G \) such that (i) $g_j = 0$ for all $i$ with $w_i^* < \max_k w_k^* = v$ and (ii) $g' w_j > v$ for $j = 1,2,\ldots,n$. (In that case $g' \in G^o$ and $v$ is the value of the game.)

Returning to the experiment described in (2.1), we will determine $H^o$:

**Lemma 2.3** If $W_3$ satisfies (1.5) the following holds:

The game with matrix $W_3$ has value
\[ v = \frac{t_3 - t_1}{d_{13} + d_{18}}, \text{where we have put } \]
\[ d_{ij} = t_j - t_i. \text{ Moreover: } \]
\[ p^* = (\frac{d_{18}}{d_{13} + d_{18}}, \frac{d_{13}}{d_{13} + d_{18}}, 0, 0) \in H^o \text{ and } \]
\[ p^* \text{ is the only strategy in } H^o. \]
\[ g_0^* = (0, 0, q_1, q_2, q_3, q_4) \in G^o, \text{ where } \]
\[ q_1 = q_2 = \frac{d_{13}}{2 (d_{13} + d_{18})}, \quad q_3 = q_4 = \frac{d_{18}}{2 (d_{13} + d_{18})} \]
\[ \text{and } g_0^* a_3 > v, \quad g_0^* a_4 > v. \]

**Proof:** From (2.2) we see that if (1.5) is satisfied then
\[ \frac{1}{2} a_3 + \frac{1}{2} a_4 > a_1 \left( \begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right) \left( \begin{array}{c} y_1 \\ \vdots \\ y_n \end{array} \right) \] means that $x_i > y_i$ for $i = 1, \ldots, n.$
Then $a_3$ or $a_4$ can be deleted from $W_3$ without changing $H^0$ (see Karlin [6]). Because of symmetry it is reasonable to believe that both can be deleted, thus expecting that there exists a minimax strategy of the form $h' = (p, 1-p, 0, 0)$.

$$W_3h = w^* = \begin{pmatrix} w_1^* \\ \vdots \\ w_6^* \end{pmatrix} = pa_1 + (1-p)a_2 = \begin{pmatrix} (1-p)t_7 \\ pt_2 \\ c \\ c \\ d \\ d \end{pmatrix}$$

where $c = pt_1 + (1-p)t_8$, $d = pt_3 + (1-p)t_1$

For $0 < p < 1$ we have $w_1^* < c$ and $w_2^* < d$. Referring to lemma 1 we have here three possibilities to look for a $g' \in G^0$: for a $p$ such that $c < d$, or such that $c > d$ or $c = d$. We try the last one:

$$c = d \iff pt_1 + (1-p)t_8 = pt_3 + (1-p)t_1$$

$$\iff p = \frac{d_{18}}{d_{13} + d_{18}}.$$ 

Hence $v = \max_i w_i^* = \frac{t_1d_{18} + t_8d_{13}}{d_{13} + d_{18}} = \frac{t_3d_{18} + t_1d_{13}}{d_{13} + d_{18}}$

or $v = \frac{t_3t_8 - t_1^2}{d_{13} + d_{18}}$.

According to lemma 2.2 we try to find a $g' \in G$ of the form $g' = (0, 0, q_1, \ldots, q_4)$ satisfying $\sum q_4 = 1$ and $g'_{a_j} \geq v$ for $j = 1, \ldots, 4$.

For $0 < p < 1$ we must have $g'_{a_j} = v$; $j = 1, 2$ (see Karlin [6]),

$$g'_{a_1} = t_1(q_1 + q_2) + t_3(q_3 + q_4)$$
$$g'_{a_2} = t_8(q_1 + q_2) + t_1(q_3 + q_4)$$
$$g'_{a_3} = t_6(q_2 + q_3)$$
$$g'_{a_4} = t_6(q_1 + q_4)$$
Letting $q_1 + q_2 = \frac{d_{13}}{d_{13} + d_{18}}$, $q_3 + q_4 = \frac{d_{18}}{d_{13} + d_{18}}$, we have $g'a_1 = g'a_2 = v$. Letting $q_1 = q_2$ and $q_3 = q_4$ we have:

\[(2.6) \quad g'a_3 = g'a_4 = \frac{1}{2} \frac{t_6 d_{13} + t_6 d_{18}}{d_{13} + d_{18}} > v = \frac{t_1 d_{18} + t_3 d_{13}}{d_{13} + d_{18}}\]

since by (1.5), $(t_6 - 2t_3)d_{13} + (t_6 - 2t_1)d_{18} > 0$.

Hence by lemma 2.2, $g' \in G^0$ and

\[p^* = (\frac{d_{18}}{d_{13} + d_{18}}, \frac{d_{13}}{d_{13} + d_{18}}, 0, 0) \in H^0.\]

It now remains to show that $p^*$ is the only minimax strategy:

Let $h' = (p_1, p_2, p_3, p_4)$ be any strategy for II such that $p_1 + p_2 < 1$. We then have:

\[g'W_3'h = g' \sum_{j=1}^{4} p_j a_j = \frac{4}{4} \sum_{j=1}^{4} p_j g'a_j = (p_1 + p_2)v + p_3 g'a_3 + p_4 g'a_4 > v = g'W_2p^*\]

and consequently $h' \notin H^0$.

Thus $h' \in H^0 \Rightarrow p_3 = p_4 = 0$.

Let $h' = (p, (1-p), 0, 0) \in H^0$ be arbitrary. If $t_1 < t_3$ then all $x_3', x_4', x_5', x_6'$ (2.2) must be relevant for $h'$ since $g' \in G^0$, and $t_1 < t_3 \Rightarrow q_1, q_2 > 0$ (see Karlin [6]).

Hence $x_i'h = v$; $i = 3, 4, 5, 6$ and we have:

\[\begin{align*}
t_1p + (1-p)t_8 &= v \\
t_2p + (1-p)t_1 &= v
d \end{align*}\]

\[p = \frac{d_{18}}{d_{13} + d_{18}} \Rightarrow h' = p^*.\]

If $t_1 = t_3$ then $v = t_1$. Let $g_1' = (0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

We then have: $g_1'a_3 = t_1 = v$, $g_1'a_2 = \frac{1}{2}(t_1 + t_8) > t_1$, $g_1'a_3 = \frac{1}{2} t_6 > t_3 = t_1$, $g_1'a_4 = \frac{1}{2} t_6 > v$.

Thus (lemma 2.2): $(1, 0, 0, 0) \in H^0$ and $g_1' \in G^0$.

Let $h' = (p_1, \ldots, p_4)$ be any strategy such that $p_1 < 1$. Then

\[g_1'W_3'h = p_1v + \frac{4}{j=2} p_j g_1'a_j > v\]

so that $h' \notin H^0$, and the proof is completed.
For a decision procedure \( \delta = \delta(q_1, \ldots, q_4) \) put \( p_j(\theta) = E_{\theta} \phi_j(\chi) \); \( j = 1, \ldots, 4 \).

Let \( \mathcal{M} \) denote the class of minimax decision procedures (MR-rules).

Proposition 2.4 below may now represent a warning against the use of a decision space like \( D \) if the interest in \( \gamma \) for its own sake is not strong enough (see the introduction to this section).

**Proposition 2.4** If \( w_3 \) satisfy (1.5) and \( t_3 > t_1 \) the following holds:

\[
\delta^* = \delta(q_1^*, \ldots, q_4^*) \in \mathcal{M} \iff
\]

(i) \( q_3^*(x) = q_4^*(x) = 0 \) a.e. \( (\mu) \) and

(ii) \( p_2(\theta) \geq \frac{d_{13}}{d_{13}^* + d_{18}} \quad \forall \theta; \quad \alpha > 0 \) and

\[
p_2(\theta) \leq \frac{d_{13}}{d_{13}^* + d_{18}} \quad \forall \theta; \quad \alpha \leq 0
\]

(Thus: \( \delta^* \in \mathcal{M} \Rightarrow P_0(d_3 \text{ or } d_4 | \delta^*) = 0 \forall \theta \in \overline{\Omega} \);

e.g. every MR-rule is intolerant w.r.t. \( \gamma \))

**Proof:** Let \( \delta^* \in \mathcal{M} \). By theorem 2.1 (ii) and lemma 2.3.,

\[
E_{\theta_0} \phi_3^* (\chi) = E_{\theta_0} \phi_4^* (\chi) = 0 \text{ for } \theta_0 \in \overline{\Omega}.
\]

Then, since all \( \phi_j > 0, \phi_3^* = \phi_4^* = 0 \) a.e. \( (P_{\theta_0}) \)

Hence \( \phi_3^* = \phi_4^* = 0 \) a.e. \( (\mu) \) since \( \mu \) is equivalent to \( \overline{P} \).

Since our decision problem is of type I, it follows from the proof of theorem 2.1 (Schaafsma [4], p 1687) that

\[
(2.7) \quad \delta^* \in \mathcal{M} \iff \sup_{\theta \in \overline{\Omega}} r(\theta, \delta^*) = v
\]

We have:

\[
t_1 + d_{18} \frac{d_{13}}{d_{13} + d_{18}} = \frac{t_1 d_{13} + t_1 d_{18} + d_{18} d_{13}}{d_{13} + d_{18}} = \frac{t_3 + d_{18} - t_1}{d_{13} + d_{18}}. \quad \text{Thus:}
\]

\[
(2.8) \quad v = t_1 + d_{18} \frac{d_{13}}{d_{13} + d_{18}}
\]
Now $\delta^* = \delta(\varphi_1^*, \ldots, \varphi_4^*) \in \mathcal{M} \Rightarrow p_1(\theta) + p_2(\theta) = 1$ for all $\theta \in \Theta$, where $p_j(\theta) = E_{\theta} \varphi_j^*(X)$.

The risk then becomes:

$$r(\theta, \delta^*) = \begin{cases} t_7 p_2(\theta) & \alpha \leq 0, \gamma = 0 \\ t_2 p_1(\theta) & \alpha > 0, \gamma = 0 \\ t_1 p_1(\theta) + t_8 p_2(\theta) & \alpha \leq 0, \gamma \neq 0 \\ t_3 p_1(\theta) + t_1 p_2(\theta) & t_1 + d_{18} p_2(\theta) \leq t_2 p_1(\theta) \\ t_3 p_1(\theta) + t_1 p_2(\theta) & t_1 + d_{13} p_1(\theta) \leq t_2 p_1(\theta) \end{cases}$$

(ii) now follows from (2.7), (2.8) and (2.9).

In order to prove the converse, assume (i) and (ii). Since the $p_j(\theta)$'s are continuous and for all $\theta \in \Theta$,

$$t_1 p_1(\theta) + t_8 p_2(\theta) \geq t_7 p_2(\theta) \text{ and } t_3 p_1(\theta) + t_1 p_2(\theta) \geq t_2 p_1(\theta),$$

we must have that

$$\sup_{\theta \in \Theta} r(\theta, \delta^*) = \frac{t_1 + d_{13} + d_{18}}{t_1 + d_{13} + d_{18}} = v,$$

implying that $\delta^* \in \mathcal{M}$.

Q.E.D.

Thus the minimax principle recommends an unbiased test for the hypothesis $\alpha \leq 0$ against $\alpha > 0$, while ignoring $\gamma$.

The following corollary gives a limit case:

**Corollary 2.5** If $W_3$ satisfy $t_1 = t_3$ in addition to (1.5) then $\mathcal{M}$ consists only of the rule which with probability 1 does not say anything ($\varphi_1(x) = 1$ a.e. $(\mu)$).

**Proof:** $t_1 = t_3 \Rightarrow d_{13} = 0 \Rightarrow p^* = (1, 0, 0, 0)$ (lemma 2.3).

Since $p^*$ is the only minimax strategy, it follows from theorem 2.1 (ii):

$$\delta^* = \delta(\varphi_1^*, \ldots, \varphi_4^*) \in \mathcal{M} \Rightarrow E_{\theta_0} \varphi_1^*(X) = 1$$

for $\theta_0 \in \Theta_0$, or $E_{\theta_0} (1 - \varphi_1^*(X)) = 0$.

Since $\varphi_1^* \leq 1$, $\varphi_1^*(x) = 1$ a.e. $(\mathbb{P}_{\theta_0})$ and so $\varphi_1^*(x) = 1$ a.e. $(\mu)$.

Q.E.D.
It is difficult to decide to what extent corollary 2.5 says something about the problem and the loss function or to what extent it says something about the minimax principle. On the other hand, the problem may be somewhat uninteresting since the condition $t_1 = t_3$ seems to harmonize poorly with the wish to be reasonable at step two even after a mistake at step one which is an essential part of the original problem.

We see that $W_1$ and $W_2$ both satisfy (1.5) if and only if

$$\min \{a_1, a_2\} \geq b_1 + b_2 \text{ and } a_1 > b_1 + 2b_2.$$  

If $a_1 = b_1 + 2b_2$ ($t_6 = 2t_3$), lemma 2.3 still holds because $b_1 + b_2 > b_1$ ($t_3 > t_1$) implies $g'_{a_3} > v$ and $g'_{a_4} > v$ (2.6). Hence:

**Corollary 2.6** If $a_1 \geq b_1 + 2b_2$ and $a_2 \geq b_1 + b_2$, the minimax principle leads to decision rules which are intolerant w.r.t. $Y$ for both $W_1$ and $W_2$.

(2.10) An essential feature about the loss functions used by Lehmann in [2] and [3] is:

The decision problem is decomposed into component problems and the loss functions from these are added in order to obtain the total loss function of the original problem. We see that the results in this section do not depend on the restriction resulting from this addition.

Thus the minimax principle is of little help when we want tolerant decision rules with good properties. (With somewhat different conditions on $W_1$ than stated in corollary 2.6 this situation will change considerably in section 6.) It is therefore natural to turn the attention towards other principles.
3. The Principle of Unbiasedness in the Risque.

A possible mathematical expression of the requirement that a
decision rule should not favour any set of parameter values at the
expense of the others, is given by the principle of unbiasedness
(in the risque) which is due to Lehmann (see for example Sverdrup
[1], p 69). A decision rule \( \delta = \delta(\varphi_1, \ldots, \varphi_n) \) is said to be un-
biased if for all \( \theta, \theta' \in \mathbb{H} \):

\[
r(\theta, \theta', \delta) = E_{\theta} W(\theta', \mathcal{T}) \geq E_{\theta} W(\theta, \mathcal{T}) = r(\theta, \delta).
\]

For a m.d.p with simple loss function we immediately have
(Schaafsma [4]):

**Proposition 3.1** \( \delta = \delta(\varphi_1, \ldots, \varphi_n) \) is unbiased if and only if
for \( i = 1, \ldots, m \) the following holds:

\[
\forall \theta \in \mathbb{H}_i : \sum_{j=1}^n w_{hj} E_{\theta} \varphi_j(X) \geq \sum_{j=1}^n w_{ij} E_{\theta} \varphi_j(X) \quad \text{for} \quad h = 1, \ldots, m
\]

where \( W = (w_{ij}) \) is the loss matrix.

**Corollary 3.2** For a m.d.p of type I, the unbiasedness of
\( \delta = \delta(\varphi_1, \ldots, \varphi_n) \) implies that

\[
\sum_{j=1}^n w_{hj} E_{\theta} \varphi_j(X) = \sum_{j=1}^n w_{ij} E_{\theta} \varphi_j(X) \quad \text{for all} \quad \theta \in \mathbb{H}_1 \cap \ldots \cap \mathbb{H}_m \quad (i, h \in \{1, \ldots, m\})
\]

We see that corollary 3.2 implies a strong condition on the
loss matrix \( W = (w_{ij}) \) for m.d.p's of type I: \( S(W), (2.3) \), must
contain at least one point \( w^* = (w^*, \ldots, w^*)' \) with all coordinates
equal (Schaafsma [4]).

**Proposition 3.3** below, shows that the unbiasedness principle
can only be applied to \( W_2(\theta, d) \) in special cases.

Assume the experiment to be as described in (2.1):
Proposition 3.3 Let $w_3$ satisfy (1.5). Then:

(i) If $2t_4 < t_6$, or $2t_4 = t_6$ and $t_4 \neq t_5$, no decision rules are unbiased.

(ii) If $2t_4 = t_6$ and $t_4 = t_5$ then every unbiased decision rule is intolerant w.r.t. $\alpha$.

(iii) If $2t_4 > t_6$ then there exist unbiased decision rules only if $S(W_3)$ contains a point $w^* = (w^*, \ldots, w^*)'$ with all coordinates equal, and this is true if and only if

$$(3.1) \quad (v - \frac{d_{18}}{d_{13} + d_{18}} t_2) (2t_4 - t_6) = (v - \frac{d_{18}}{d_{13} + d_{18}} t_7) (2t_5 - t_6)$$

where $v = \frac{t_3 t_8 - t_1^2}{d_{13} + d_{18}}$ (lemma 2.3).

In this case $w^* = \sum_{j=1}^{4} p_j^* a_j$ is uniquely determined; the strategy $p^* = (p_1^*, \ldots, p_4^*) = (d_{18}, d_{13}, d_{18}, d_{13}, d_{18})$ as well, where $d = (d_{13} + d_{18}) (2t_4 - t_6)$, $c = t_1 d_{18} + d_{18} d_{13} + t_6 d_{13}$.

Proof: $w_3 = (a_1, \ldots, a_4) = \left( \begin{array}{c} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{array} \right)$

Let $p' = (p_1, \ldots, p_4) \in H$ and $u_i = u_i(p) = p'_i p$, $i = 1, \ldots, 6$:

$u_1 = t_7 p_2 + t_4 p_3 + t_4 p_4$
$u_2 = t_2 p_1 + t_5 p_3 + t_5 p_4$
$u_3 = t_1 p_1 + t_6 p_2 + t_6 p_4$
$u_4 = t_1 p_1 + t_8 p_2 + t_6 p_3$
$u_5 = t_3 p_1 + t_1 p_2 + t_6 p_3$
$u_6 = t_3 p_1 + t_1 p_2 + t_6 p_4$
We will try to determine $p$ such that

(1) $u_i = u_j$ for all $i, j \in \{1, \ldots, 6\}$, and $\sum p_i = 1$.

Since $u_3 = u_4 \iff p_3 = p_4$, we have (1) $\iff$ (2) where

\[
\begin{align*}
(a) & \quad p_1 + p_2 + 2p_3 = 1 \quad (p_3 \leq \frac{1}{2}) \\
& \quad u_1 = t_7 p_2 + t_4 2p_3 = \\
& \quad u_2 = t_2 p_1 + t_5 2p_3 = \\
& \quad u_3 = t_1 p_1 + t_8 p_2 + t_6 p_3 = \\
& \quad u_5 = t_3 p_1 + t_1 p_2 + t_6 p_3
\end{align*}
\]

We find:

\[
\begin{align*}
& u_3 = u_5 \iff p_2 = \frac{d_{13}}{d_{18}} \frac{d_{13}}{d_{18}} (1 - 2p_3) \\
& \text{Hence by (a)}: \quad u_3 = u_5 \iff p_2 = \frac{d_{13}}{d_{18}} (1 - 2p_3)
\end{align*}
\]

and $p_1 = \frac{d_{18}}{d_{13} + d_{18}} (1 - 2p_3)$.

Thus:

\[
\begin{align*}
& (a) \quad u_3 = u_5 \iff \\
& \quad \begin{cases} 
  p_1 = \frac{d_{18}}{d_{13} + d_{18}} (1 - 2p_3) \\
  p_2 = \frac{d_{13}}{d_{13} + d_{18}} (1 - 2p_3)
\end{cases}
\end{align*}
\]

\[
\begin{align*}
u_1 = u_3 \iff (t_8 - t_7)p_2 + t_1 p_1 = (2t_4 - t_6)p_3
\end{align*}
\]

(3) (a) \& $u_3 = u_5 \iff \\
\quad \begin{cases} 
  p_1 = \frac{d_{18}}{d_{13} + d_{18}} (1 - 2p_3) \\
  p_2 = \frac{d_{13}}{d_{13} + d_{18}} (1 - 2p_3)
\end{cases}
\]

\[
\begin{align*}
& t_1 d_{18} + d_{78} d_{13} (1 - 2p_3) = (2t_4 - t_6)(1 - 2p_3)
\end{align*}
\]

Let $c = t_1 d_{18} + d_{78} d_{13}$ and $d = (d_{13} + d_{18})(2t_4 - t_6)$

Since $t_8 \geq t_7$, $c > 0$.

If $2t_4 < t_6$ then (4) is impossible since $p_3 \in [0, \frac{1}{2}]$.

If $2t_4 = t_6$ then (4) is possible if and only if $p_3 = \frac{1}{2}$, e.g. $p_1 = p_2 = 0$; whence $u_1 = t_4$ and $u_2 = t_5$.

(i) now follows from corollary 3.2.

If $2t_4 = 2t_5 = t_6$, it follows in the same way as before (proposition 2.4) that: $\delta = \delta(\varphi_1, \ldots, \varphi_4)$ unbiased $\Rightarrow \varphi_1(x) = \varphi_2(x) = 0$ a.e. ($\mu$), which proves (ii).
Suppose that $2t_4 > t_6$:

(4) $\iff c(1-2p_3) = dp_3 \iff$

(5) $p_3 = \frac{c}{d+2c}$ (whence $1-2p_3 = \frac{d}{d+2c}$).

Thus:

(6) (a) $u_1 = u_3 = u_5 \iff (3) \& (5)$.

We see that since $c > 0$, $p_1, \ldots, p_4$ are uniquely determined by (a) and $u_1 = u_3 = u_5$. It also follows that $S(W_3)$ contains a point with equal coordinates if and only if $u_2 = u_3$ where the determined values are substituted for $p_1, \ldots, p_4$:

\[
\begin{align*}
(7) \quad \left\{ \begin{array}{c}
t_2 \frac{d}{d+2c} \frac{d}{d+2c} + t_5 \frac{2c}{d+2c} = \\
t_1 \frac{d}{d+2c} \frac{d}{d+2c} + t_8 \frac{d}{d+2c} \frac{d}{d+2c} + t_6 \frac{c}{d+2c} \\
t_2 \frac{d}{d+2c} (2t_4 - t_6) + 2t_5 \frac{d}{d+2c} \frac{d}{d+2c} + t_6 \frac{d}{d+2c} \frac{d}{d+2c} + t_8 \frac{d}{d+2c} \\
\end{array} \right. \\
\end{align*}
\]

\[
\begin{align*}
(8) \quad \left\{ \begin{array}{c}
t_1 \frac{d}{d+2c} + t_8 \frac{d}{d+2c} - t_2 \frac{d}{d+2c} (2t_4 - t_6) = \\
= (t_1 \frac{d}{d+2c} + t_8 \frac{d}{d+2c} - t_7 \frac{d}{d+2c} (2t_5 - t_6) \\
\end{array} \right. \\
\end{align*}
\]

Substituting $v = \frac{t_1 \frac{d}{d+2c} + t_8 \frac{d}{d+2c}}{d+2c} = \frac{t_3 t_8 - t_1^2}{d+2c}$ (lemma 2.3),

we see that $(8) \iff (3.1)$.

Q.E.D.

Corollary 3.4

(i) If $W_3$ satisfy the conditions in proposition 3.3 (iii) and $t_3 > t_1$ then every unbiased decision rule is tolerant.

(ii) W.r.t. the modified matrix $W_2$, no decision rule is unbiased.

(iii) Lehmann's loss matrix $W_3$ satisfy the conditions in proposition 3.3 (iii).

For this matrix

$$p^* = \left( \frac{a_2}{a_2+b_2}(1-\frac{2b_1}{a_1+b_1}), \frac{b_2}{a_2+b_2}(1-\frac{2b_1}{a_1+b_1}), \frac{b_1}{a_1+b_1}, \frac{b_1}{a_1+b_1} \right),$$
(i) follows immediately from corollary 3.2 and the proof above. (ii) and (iii) follow by substitution.

Thus, for $W_2$, neither the minimax principle nor the unbiasedness principle can be applied. Therefore if we want to use $W_2$, it is necessary to find other and less compelling requirements on a decision rule. (It is necessary to have some requirements since the trivial procedure which with probability 1 concludes $\theta \in \hat{H}_i$, obviously minimizes the risque uniformly for $\theta \in \hat{H}_i$.)

From proposition 3.3 (iii) we see that the unbiasedness principle may be used for a somewhat (but not essentially) more general loss function than the type indicated in (2.10) (Lehmann's loss functions).

This "sensitivity" of the unbiasedness principle may first of all be due to the special structure of a m.d.p of type I.

In the following sections we shall investigate in more detail the consequences of applying the unbiasedness principle in connection with $W_1$.

4. Unbiasedness for $W_1$

Let the situation be as described in section 1 with loss function $W_1(\theta, d)$.

$\hat{H} = \{(\alpha, \gamma, \beta, \frac{1}{\beta}, \sigma); (\alpha, \gamma, \beta, \frac{1}{\beta}) \in \mathbb{R}^4, \sigma > 0\}$

$D = \{\hat{H}^", "\alpha > 0", "\gamma < 0", "\gamma > 0\}$

For $D(x) = \delta(q_1(x), \ldots, q_4(x))$, let $p_j(\theta) = E_\theta q_j(x)$; $j = 1, \ldots, 4$.

$P = \{p_\theta; \theta \in \hat{H}\}$ consists of normal distributions an constitutes a regular exponential family. By using $\sum_{j=1}^{4} p_j(\theta) = 1$, we obtain:
As in section 2 we let $W_\gamma$ satisfy:

\[(4.2) \quad a_1 \geq b_1 + 2b_2 \quad \text{and} \quad a_2 \geq b_1 + b_2\]

Proposition 4.1 below, motivates to a certain degree the course chosen by Lehmann [3] and Sverdrup [1] when decision procedures are constructed for various two-step decision problems.

We write $\varepsilon_1 = \frac{b_1}{a_1 + b_1}$, $\varepsilon_2 = \frac{b_2}{a_2 + b_2}$.

Let $\mathcal{R}$ denote the set of unbiased decision rules and $\mathcal{M}$, as before, the MR-rules.

In section 6 we will make use of the fact that proposition 4.1 does not depend on condition (4.2).

**Proposition 4.1**  A decision rule $\delta = \delta(p_1, \ldots, p_4)$ is unbiased w.r.t. $W_\gamma$ if and only if for all $\theta \in \mathbb{H}$:

\[
\begin{align*}
(i) \quad & \frac{p_3(\theta)}{p_1(\theta) + p_2(\theta)} \begin{cases} 
\leq \varepsilon_2 & \text{for } \alpha < 0 \\
= \varepsilon_2 & \text{for } \alpha = 0 \\
\geq \varepsilon_2 & \text{for } \alpha > 0
\end{cases} \\
(ii) \quad & p_3(\theta) \geq \max\left\{ \frac{\varepsilon_1}{2}, p_4(\theta) \right\} \quad \text{for } \gamma < 0 \\
& p_4(\theta) \geq \max\left\{ \frac{\varepsilon_1}{2}, p_3(\theta) \right\} \quad \text{for } \gamma > 0 \\
& p_3'(\theta) = p_4'(\theta) = \frac{\varepsilon_1}{2} \quad \text{for } \gamma = 0
\end{align*}
\]

(Thus, by dividing the decision problem into two steps, unbiasedness implies similarity for the component tests.)
Proof: Let \( r(\theta, \delta) = h_i(\theta) \) for \( \theta \in \mathbb{H}_i \); \( i = 1, \ldots, 6 \).

Proposition 3.1 implies:

\[
\delta \in \mathcal{R} \iff h_1(\theta) = \min_{1 \leq j \leq 6} h_j(\theta) = h(\theta) \forall \theta \in \mathbb{H}_1; \ i=1,\ldots,6
\]

Let \((i, j)\) denote the statement that \( h_i(\theta) \leq h_j(\theta) \).

By (4.1) we obtain:

\[
(1, 2) \iff a_2 p_2(\theta) \leq b_2 p_1(\theta)
\]

\[
(1, 3) \iff p_3(\theta) \leq \frac{\xi_1}{2}
\]

\[
(1, 4) \iff p_4(\theta) \leq \frac{\xi_1}{2}
\]

\[
(1, 5) \iff a_2 p_2(\theta) + (a_1 + b_1) p_4(\theta) \leq b_2 p_1(\theta) + b_1
\]

\[
(1, 6) \iff a_2 p_2(\theta) + (a_1 + b_1) p_3(\theta) \leq b_2 p_1(\theta) + b_1
\]

We see that \((1, 2), (1, 3), (1, 4) \Rightarrow (1, 5), (1, 6)\)

Thus:

\[
h_1(\theta) = h(\theta) \iff a_2 p_2(\theta) \leq b_2 p_1(\theta); \ p_3(\theta) \leq \frac{\xi_1}{2}; \ p_4(\theta) \leq \frac{\xi_1}{2}
\]

Similarly we obtain:

\[
h_2(\theta) = h(\theta) \iff b_2 p_1(\theta) \leq a_2 p_2(\theta); \ p_3(\theta) \leq \frac{\xi_1}{2}; \ p_4(\theta) \leq \frac{\xi_1}{2}
\]

\[
h_3(\theta) = h(\theta) \iff a_2 p_2(\theta) \leq b_2 p_1(\theta); \ p_3(\theta) \geq \max \left\{ \frac{\xi_1}{2}, \ p_4(\theta) \right\}
\]

\[
h_4(\theta) = h(\theta) \iff a_2 p_2(\theta) \leq b_2 p_1(\theta); \ p_4(\theta) \geq \max \left\{ \frac{\xi_1}{2}, \ p_3(\theta) \right\}
\]

\[
h_5(\theta) = h(\theta) \iff b_2 p_1(\theta) \leq a_2 p_2(\theta); \ p_4(\theta) \geq \max \left\{ \frac{\xi_1}{2}, \ p_3(\theta) \right\}
\]

\[
h_6(\theta) = h(\theta) \iff b_2 p_1(\theta) \leq a_2 p_2(\theta); \ p_3(\theta) \geq \max \left\{ \frac{\xi_1}{2}, \ p_4(\theta) \right\}
\]

Also:

\[
a_2 p_2(\theta) \geq b_2 p_1(\theta) \iff \frac{p_2(\theta)}{p_1(\theta)+p_2(\theta)} \geq \frac{b_2}{a_2+b_2} = \xi_2
\]

(We cannot have that \( p_1(\theta) + p_2(\theta) = 0 \) here because:

\[
\exists \theta_0; \ p_1(\theta_0) = p_2(\theta_0) = 0 \implies p_1(\theta) = p_2(\theta) = 0 \forall \theta \in \mathbb{H} \implies \delta \notin \mathcal{R}
\]

(corollary 3.4))

Combining, we now have:

\[
\delta \in \mathcal{R} \iff
\]
\[
\frac{p_2(\theta)}{p_1(\theta)+p_2(\theta)} \begin{cases} 
\leq \varepsilon_2 & \text{for } \alpha \leq 0 \\
\geq \varepsilon_2 & \text{for } \alpha > 0
\end{cases}
\]

\[p_3(\theta), p_4(\theta) \leq \frac{\varepsilon_1}{2} \text{ for } \gamma = 0\]

\[p_3(\theta) \geq \max \left\{ \frac{\varepsilon_1}{2}, p_4(\theta) \right\} \text{ for } \gamma < 0\]

\[p_4(\theta) \geq \max \left\{ \frac{\varepsilon_1}{2}, p_3(\theta) \right\} \text{ for } \gamma > 0\]

By the continuity of \(p_j(\theta); j = 1, \ldots, 4\), the rest follows immediately. \(\text{Q.E.D.}\)

By proposition 4.1 and corollary 3.2 now follows:

**Corollary 4.2** If \(a_1 \geq b_1 + 2b_2\) and \(a_2 \geq b_1 + b_2\) then no unbiased decision rule minimizes the maximal risque.

As will be clarified below, this shows that it is not always reasonable to apply the unbiasedness principle in a situation like this. ("reasonable" must be understood in relation to the purpose of constructing procedures that in some sense minimize the risque.)

This will be most clear if the model contains assumptions of the type \(|\frac{\alpha}{\sigma}| \leq k_1, |\frac{\lambda}{\sigma}| \leq k_2\), where \(k_1\) and \(k_2\) are not too great. The argument consists of comparing unbiasedness with the minimax risque property:

Following proposition 2.4 we have:

\[\delta \in \mathcal{M} \iff p_1(\theta)+p_2(\theta) = 1 \quad \& \quad (4.3)\]

\[p_2(\theta) \begin{cases} 
\geq \varepsilon_2 & \text{for } \alpha > 0 \\
\leq \varepsilon_2 & \text{for } \alpha \leq 0
\end{cases}\]

For \(\delta \in \mathcal{M}\) the risque becomes:
The uniformly worst MR-rule is thus given by $\delta^* = \delta(\varphi_1^*, \ldots, \varphi_4^*)$ where $\forall x; \varphi_1^*(x) + \varphi_2^*(x) = 1$ & $\varphi_2^*(x) = \varepsilon_2$.

Let $\delta \in \mathcal{F}$ be arbitrary and assume first that $\gamma = 0$. From proposition 4.1 we have $(p_j(\epsilon)$ referring to $\delta)$:

$$r(\theta, \delta) = \begin{cases}
  a_2p_2(\theta) + b_2p_1(\theta) & \alpha \leq 0, \gamma = 0 \\
  b_1 + a_2p_2(\theta) & \alpha \leq 0, \gamma \neq 0 \\
  b_1 + b_2p_1(\theta) & \alpha > 0, \gamma \neq 0
\end{cases}$$

and $\sup_{\theta \in \mathcal{F}} r(\theta, \delta) = v = b_1 + a_2\varepsilon_2$

Hence, by the continuity of $r(\theta, \delta) - r(\theta, \delta^*)$ for $\gamma = 0$, $\delta^*$ strictly dominates $\delta$ for all $\alpha$ when $\gamma = 0$.

If $\varepsilon_1 > \frac{a_2}{a_1} \varepsilon_2$, we find for $\alpha = \gamma = 0$:

$$r(\theta, \delta) - r(\theta, \delta^*) = \frac{2b_1(a_2(a_1-b_2)+a_1b_2)}{(a_1+b_1)(a_2+b_2)} > 0$$

Hence, by the continuity of $r(\theta, \delta) - r(\theta, \delta^*)$ for $\gamma = 0$, $\delta^*$ dominates $\delta$ strictly in a region surrounding $\alpha = 0$ when $\gamma = 0$.

Now, assume $\gamma \neq 0$. To facilitate the argument (now and later) we state the following trivial lemma:

**Lemma 4.3** Let $\delta \in \mathcal{F}$ be arbitrary and $\delta^*$ as above. If $\gamma \neq 0$ then

$$r(\theta, \delta^*) - r(\theta, \delta) = r_i(\theta) \quad \text{for} \quad \theta \in \mathcal{H}_i; \quad i = 3, \ldots, 6,$$

where

$$r_3(\theta) = a_2(1-p_3(\theta) - p_4(\theta))(\varepsilon_2 - \frac{p_2(\theta)}{p_1(\theta) + p_2(\theta)})(a_1 + b_1)(a_1p_3(\theta) - a_2p_4(\theta))$$

($\gamma < 0, \alpha \leq 0$)
\[ r_4(\theta) = a_2 (1-p_3(\theta) - p_4(\theta))(\xi_2 - \frac{p_2(\theta)}{p_1(\theta) + p_2(\theta)}) + (a_1 + b_1)(c_1 p_4(\theta) - c_2 p_3(\theta)) \]

\[ r_5(\theta) = b_2 (1-p_3(\theta) - p_4(\theta))(\frac{p_2(\theta)}{p_1(\theta) + p_2(\theta)} - \xi_2) + (a_1 + b_1)(c_1 p_4(\theta) - c_2 p_3(\theta)) \]

\[ r_6(\theta) = b_2 (1-p_3(\theta) - p_4(\theta))(\frac{p_2(\theta)}{p_1(\theta) + p_2(\theta)} - \xi_2) + (a_1 + b_1)(c_1 p_3(\theta) - c_2 p_4(\theta)) \]

Here \( c_1 = \frac{b_1}{a_1 + b_1} + \frac{a_2 b_2}{(a_1 + b_1)(a_2 + b_2)} \), \( c_2 = \frac{a_1}{a_1 + b_1} - \frac{a_2 b_2}{(a_1 + b_1)(a_2 + b_2)} \)

with \( c_1 + c_2 = 1, c_1 < \frac{1}{2}, c_2 > \frac{1}{2} \) (for \( a_1 > b_1 + 2b_2 \))

**Proof:** For \( \gamma \neq 0 \), \( r(\theta, \delta^*) = b_1 + \frac{a_2 b_2}{a_2 + b_2} = v. \)

(4.4) We find: \( \frac{1}{2} - c_1 = \frac{a_2(a_1 - b_1 - 2b_2) + b_2(a_1 - b_2)}{2(a_1 + b_1)(a_2 + b_2)} > 0 \)

Therefore \( c_1 < \frac{1}{2} \) and \( c_2 > \frac{1}{2} \).

From (4.1) we have \( (p_j = p_j(\theta)) \):

\[ \theta \in \mathbb{H}_3 \Rightarrow v-r(\theta, \delta) = \frac{a_2 b_2}{a_2 + b_2} + b_1 p_3 - a_2 p_2 - a_1 p_4 = \]

\[ = \frac{a_2 b_2}{a_2 + b_2} + (b_1 + \frac{a_2 b_2}{a_2 + b_2})p_3 - a_2 p_2 - (a_1 - \frac{a_2 b_2}{a_2 + b_2})p_4 - \frac{a_2 b_2}{a_2 + b_2}(p_3 + p_4) = \]

\[ = a_2 \xi_2 (p_1 + p_2) - a_2 p_2 + (a_1 + b_1)(c_1 p_3 - c_2 p_4) = r_3(\theta). \]

The other cases are treated similarly. \( \text{Q.E.D.} \)

It now follows from proposition 4.1 that if \( \delta \in \mathbb{R} \) then the first terms of \( r_j(\theta) \) \((j=3, \ldots, 6)\) equal zero for all \( \theta \) such that \( \alpha = 0 \), and greater or equal zero otherwise. The second terms of the \( r_j(\theta) \) 's will be less than zero for all \( \theta \) such that \( \gamma \) belongs to a region around \( \gamma = 0 \).

Hence we see that \( \delta^* \) (which is the worst MR-rule) strictly dominates an arbitrary \( \delta \in \mathbb{R} \) in a region around \( \alpha = \gamma = 0 \) and at best in a region of the type indicated on the figure below.
(Note that unbiasedness implies similarity for $p_3(\theta)$ and $p_4(\theta)$.)

By lemma 4.3 it easily follows (see section 5) that there exists a $\delta_0 \in \mathcal{R}$ which among all $\delta \in \mathcal{R}$ uniformly minimizes the risque for $\alpha = 0$. The length of $AB$ must therefore attain a minimum value which is greater than zero.

Because of the argument given in section 2, it may seem uninteresting to compare a $\delta \in \mathcal{R}$ with $\delta^*$ since all MR-rules are intolerant. It is, however, easy to construct a sequence of tolerant decision rules, $\langle \delta_n \rangle_{n=1}^{\infty}$, whose performance functions converge (w.r.t. $\theta$) to the performance function of a MR-rule ($\delta'$).

As a corollary of this it follows moreover that if $\mathcal{T}$ denotes the class of tolerant decision rules then there is no $\delta \in \mathcal{T}$ which uniformly minimizes the risque among the tolerant rules because:

Suppose $\delta \in \mathcal{T}$ uniformly minimizes the risque in $\mathcal{T}$, Let $\delta_n$ and $\delta'$ be as described above. Then:

$$r(\theta, \delta) \leq r(\theta, \delta_n) \quad \forall \theta \in \mathcal{H}, \quad n=1,2,...$$

But since $r(\theta, \delta_n) \rightarrow r(\theta, \delta') \quad \forall \theta \in \mathcal{H}$, then

$$r(\theta, \delta) \leq r(\theta, \delta') \quad \forall \theta \in \mathcal{H},$$

whence

$$r(\theta, \delta) \leq \sup_{\theta \in \mathcal{H}} r(\theta, \delta') \quad \forall \theta \in \mathcal{H}.$$
Hence $\bar{g} \in M$ and $\bar{g}$ is intolerant.

This result, obviously is also true for the loss matrix $W_3$, if $W_3$ satisfies (1.5) and $t_3 > t_1$. It thus supports the search for stronger conditions on a decision rule than just tolerance.

With assumptions of the type $|\frac{\alpha}{\sigma}| \leq k_1, |\frac{\gamma}{\sigma}| \leq k_2$, it is thus doubtful whether the use of the unbiasedness principle is reasonable. It seems more natural to choose a decision rule of which the properties in some way constitute a compromise between the minimax property and unbiasedness. From what is said up to now, it is reasonable to think that such a choice must depend on to what extent there is an interest in the parameters for their own sake, e.g. on the contextual meaning of the parameters.

On the other hand, if there is little knowledge a priori (e.g. $\alpha, \gamma$ can take any value) then the requirement that a decision rule should be "impartial" (unbiased), seems more important. Thus the dominance described above, will more have the character of being a price that has to be paid in order to meet such a requirement.

In addition we have had an illustration of how the nature of a priori information can influence the choice of requirements for a decision rule.

5. Some Specific Decision Rules Considered in the Light of $W_1$

Until now we have ignored an important, practical advantage of the unbiasedness principle in this situation: Because of the theory of regular exponential families and proposition 4.1, the principle leads to a possible solution of the problem of nuisance parameters and to a natural way of constructing unbiased procedures.
The following consideration thus leads to an unbiased decision rule:

Corresponding to Lehmann's method (see [21, [3]), the problem described in (1.1) decomposes into 3 hypotheses: \( \omega_1 = (\gamma \leq 0) \), \( \omega_2 = (\gamma > 0) \) and \( \omega_3 = (\alpha \leq 0) \). \( \omega_1 \) and \( \omega_2 \) lead to two disjoint critical regions \( R^+ \) and \( R^- \) respectively with indicator functions

\[
\varphi_4 = \chi_{R^+} \quad \text{and} \quad \varphi_5 = \chi_{R^-} \quad (\chi_A(y) = 1 \text{ if } y \in A \text{ and } 0 \text{ otherwise})
\]

Since \( \overline{P} \) is regular exponential, each of the tests \( \varphi_3 \) and \( \varphi_4 \) can be made uniformly most powerful among all similar tests (the level being \( \frac{\xi_1}{2} = \frac{b_1}{a_1+b_1} < \frac{1}{2} \)).

The conditional class of distributions, given \( X \in A = (R^- \cup R^+)^c \), is still regular exponential. In the usual way we can then construct an optimal unbiased (relatively to the conditional class) test for \( \omega_3 \). This is carried out in Sverdrup [1]. We call the resulting decision rule \( \delta_0 = \delta(\varphi_1^0, \ldots, \varphi_4^0) \). By proposition 4.1 it follows that \( \delta_0 \in \mathcal{R} \). (This also follows directly from the construction of \( W_1 \) (see Lehmann [3], p 559).)

We shall describe \( \delta_0 \) a little closer (see Sverdrup [1]):

Let the situation be as in (1.1). The least square estimators then become:

\[
\hat{\xi} = \bar{X} \quad \hat{\alpha}_i = X_i - \bar{X} \quad \hat{\beta}_j = X_j - \bar{X} \quad \hat{\gamma}_{ij} = X_{ij} - X_i - X_j + \bar{X} \quad \text{and}
\hat{\delta} = \hat{\alpha}_1 - \hat{\alpha}_2 \quad \hat{\beta} = \hat{\beta}_1 - \hat{\beta}_2 \quad \hat{\gamma} = \hat{\gamma}_{11} - \hat{\gamma}_{12} - \hat{\gamma}_{21} + \hat{\gamma}_{22}
\]

The usual estimator for \( \sigma^2 \) is:

\[
S^2 = \frac{1}{n-4} \sum_{ijv} (X_{ijv} - \bar{X}_{ij})^2 \quad \text{Let} \quad Z = \sum_{ijv} X_{ijv}^2
\]

By letting \( T_\gamma = \frac{\hat{\gamma} \sqrt{n}}{S} \), \( T_\alpha = \frac{\hat{\alpha} \sqrt{n}}{S} \) then \( T_\alpha, T_\gamma \) are distributed according to the t-distribution with \( n-4 \) degrees of freedom when \( \alpha = 0 \) and \( \gamma = 0 \) respectively. Let \( t_\mu, \nu \) denote the \( \mu \)-fractile in the t-distribution (\( \nu \)).
Let $t_1 = t_{1-\frac{1}{2}, n-4}$, where $\frac{1}{2} = \frac{b_1}{a_1 + b_1} < \frac{1}{2}$.

At step one we obtain:

$\varphi^0_3(x) = \mathcal{N}(t_1, x < t_1)(x)$ and
$\varphi^0_4(x) = \mathcal{N}(t_1, x > t_1)(x)$ whence $A = (T_2^2 \leq t_1^2)$.

$\varphi^0_3$ and $\varphi^0_4$ satisfy:

For all tests $\varphi$ such that $E_{\theta} \varphi(X) = \frac{\xi_1}{2}$, $\forall \theta; \gamma = 0$, then
(5.1) $E_{\theta} \varphi^0_3(X) \geq E_{\theta} \varphi(X)$ and $E_{\theta} \varphi^0_4(X) \leq E_{\theta} \varphi(X)$, $\forall \theta; \gamma < 0$, and
(5.2) $E_{\theta} \varphi^0_3(X) \leq E_{\theta} \varphi(X)$ and $E_{\theta} \varphi^0_4(X) \geq E_{\theta} \varphi(X)$, $\forall \theta; \gamma > 0$.

Step two:

Let $W = \frac{(n-4)S^2 + \bar{x}^2}{(n-4)n\bar{x}^2} = \frac{(n-4) + T_2^2}{(n-4)T_2^2}$

and $r(W) = \begin{cases} 
\sqrt{(n-4)(t_1^2W - 1)} & \text{for } W \geq \frac{1}{t_1^2} \\
0 & \text{for } W < \frac{1}{t_1^2} 
\end{cases}$

Let $f(W)$ be the solution of the equation

$\mathcal{N}_{n-4}(f(W)) = \xi_2 + (1-2\xi_2) \mathcal{N}_{n-4}(r(W))$

(5.3) We see that $0 \leq f(W) \leq t_1 - \xi_2, n-4$.

The critical region for $\omega_0$, given that $T_2^2 \leq t_1^2$, then becomes
($T_2^2 > f(W)$). In the conditional distribution class, this is the uniformly most powerful unbiased test.

The performance function for $\delta_0$ is now given by

$(P^0_1(\theta) = E_{\theta} \varphi^0_1(X)):

P^0_1(\theta) = P_{\theta}((T_2^2 \leq f(W)) \cap (T_{\gamma}^2 \leq t_1^2))$

$P^0_2(\theta) = P_{\theta}((T_2^2 > f(W)) \cap (T_{\gamma}^2 \leq t_1^2))$

$P^0_3(\theta) = P_{\theta}(T_{\gamma} < -t_1), P^0_4(\theta) = P_{\theta}(T_{\gamma} > t_1)$
According to lemma 4.3 the minimization of the risque for $\Theta \in \mathcal{H}_4$ is equivalent to the maximization of $r_i(\Theta)$ for $i = 3, \ldots, 6$. Since the first terms in $r_i(\Theta)$; $i = 3, 4$ vanish for unbiased decision rules when $\alpha = 0$, it follows immediately from proposition 4.1, lemma 4.3 and (5.1):

**Proposition 5.1** $\delta_0 \in \mathcal{R}$ and among all unbiased procedures $\delta_0$ minimizes the risque uniformly for all $\Theta$ such that $\alpha = 0$.

We also have:

**Proposition 5.2** $\delta_0$ is admissible among the unbiased procedures.

**Proof:** Suppose that $\delta'$ dominates $\delta_0$, e.g. $\forall \Theta; \ r(\Theta, \delta') \leq r(\Theta, \delta_0)$ & $\exists \Theta; \ r(\Theta, \delta') < r(\Theta, \delta_0)$.

We must have $r(\Theta, \delta') = r(\Theta, \delta_0)$ for $\alpha = 0$, whence ((5.1) and lemma 4.3):

$p_3'(\Theta) = p_3^0(\Theta)$ and $p_4'(\Theta) = p_4^0(\Theta)$ for all $\Theta$ such that $\alpha = 0$. Since to any decision rule there corresponds a decision rule based on a sufficient statistic with the same risque function at least where the sample space is Euclidian (see Ferguson [5], p 120), we can restrict the attention to the tests based on the sufficient statistic $(\hat{\alpha}, \hat{\beta}, \hat{\xi}, Z)$ where $Z = \sum_{i,j,v} x_{ijv}^2$.

Relatively to this class the following holds:

Among the conditional tests, the optimal $\varphi_3^0$ and $\varphi_4^0$ are uniquely determined (a.e. $(\mu)$). On the other hand, $\delta' \in \mathcal{R}$ implies that the tests at step one are similar and therefore conditional since the exponential family is regular.

Hence: $\varphi_j'(x) = \varphi_j^0(x)$ a.e. $(\mu)$; $j = 3, 4$.

By proposition 4.1 and lemma 4.3 it follows that among all $\delta(\varphi_1, \ldots, \varphi_4) \in \mathcal{R}$ such that $\varphi_3 = \varphi_3^0$, $\varphi_4 = \varphi_4^0$, $\delta_0$ uniformly...
minimizes the risque. Thus \( r(\theta, \delta_0) = r(\theta, \delta') \) for all \( \theta \in \Theta \), which completes the proof.

Thus if there exists an unbiased procedure uniformly minimizing the risque, it must be \( \delta_0 \). On the other hand, in view of lemma 4.3 there is no reason for believing in the existence of an optimal unbiased procedure since the choice of test method at step one must influence the choice of test at step two (see also Lehmann [3], p 559-560). Because of proposition 5.1 and 5.2 a possible proof of the nonexistence of an optimal unbiased procedure might consist in constructing a \( \delta \in \mathcal{R} \) such that \( r(\theta, \delta) < r(\theta, \delta_0) \) for at least one \( \theta \in \Theta \). For example define \( \delta \) as follows:

\[
\varphi_3(x) = \varphi_4(x) = \frac{\varepsilon_1}{2} \quad \forall x \in \mathcal{X}
\]

\[
\varphi_2(x) = (1-\varepsilon_1)\chi_{(T_\alpha > t_1-\varepsilon_2, n-4)}(x)
\]

\[
\varphi_1(x) = (1-\varepsilon_1)\chi_{(T_\alpha \leq t_1-\varepsilon_2, n-4)}(x)
\]

Proposition 4.1 \( \Rightarrow \delta \in \mathcal{R} \). It would be natural to compare \( \delta \) and \( \delta_0 \) for \( \gamma = 0 \) which is the same as to compare

\[
p_2(\theta) = (1-\varepsilon_1)p_\theta(T_\alpha > t_1-\varepsilon_2, n-4)
\]

with

\[
p_2^o(\theta) = p_\theta((T_\alpha > f(\mathcal{W})) \cap (T_\gamma \leq t^2_1)) \quad \text{for } \gamma = 0.
\]

This ought to be possible at least by numerical tools.

It follows from section 4 that particularly when there is little a priori knowledge about \( \alpha \) and \( \gamma \) that it is reasonable to use \( \delta_0 \). If it is known a priori that \( |\frac{\alpha}{\gamma}| < k_1 \), \( |\frac{\alpha}{\gamma}| < k_2 \) where \( k_1, k_2 \) are not too great then the application of the unbiasedness principle is more doubtful. In a situation like this it may be natural to apply Anderson's method (described in Sverdrup [1]) which by [1] (p 38), leads to \( \delta_1 = \delta(\varphi_1, \ldots, \varphi_4) \) where \( \varphi_j = \varphi_j^o \)
for \( j = 3,4 \), but where step two is based on \( T_{\alpha} = \frac{\sqrt{n}}{S} \). Here \( S^2 = (n-4)S_n^2 + n \bar{Y}^2 \) so that \( T_{\alpha} \) is t-distributed with \( n-3 \) degrees of freedom for \( \alpha = 0 \).

Then \( \varphi_2'(x) = \chi(T_{\alpha} > t_{1-\frac{1}{2}}, n-3) \wedge (T_{-\alpha} \leq t_{1-\frac{1}{2}})(x) \) and \( \varphi_1'(x) \) is analogous. Since \( E \varphi_2'(X) = E \varphi_1'(X) = \frac{\varepsilon_1}{2} \) for \( \gamma = 0 \), we have in view of the optimum property shown in [1], p 34-35, that \( r(\theta, \delta_1) < r(\theta, \delta_0) \) when \( \gamma = 0, \alpha \neq 0 \).

Moreover \( r(\theta, \delta_1) = r(\theta, \delta_0) \) for all \( \gamma \) when \( \alpha = 0 \).

(Hence \( \delta_1 \notin \mathcal{R} \).)

Anderson's method is motivated by a situation in which it does not make any difference whether one comes to wrong conclusions or not at step two when there already was a mistake at step one; e.g. we only know that the procedure is good for \( \gamma = 0 \). Because of continuity this must be true in a region about the \( \alpha \)-axis, but in practice numerical methods would have to be used to compare \( \delta_0 \) and \( \delta_1 \) in the region given by \( |\frac{\gamma}{\sigma}| < k_1, |\frac{\alpha}{\sigma}| < k_2 \), in order to decide which one is more appropriate. See also the end of section 4.

Another possibility is the following:

Lehmann ([2], [3]) has developed a theory that shows how it is possible to construct uniformly least risky unbiased procedures for certain decision problems by decomposing the problem into test situations for which there exist uniformly most powerful similar tests. This theory is not valid for our example, but a natural idea would be to try a procedure taken from the decision problem of this kind which is most similar to our problem. Thus, consider the hypotheses \( \omega_1 = (\gamma \leq 0), \omega_2 = (\gamma > 0) \) and \( \omega_3 = (\alpha \leq 0) \) with loss functions given in (1.2) and (1.3). This leads to the following free decision space:
\( D^1 = \{ d_1', \ldots, d_6' \} = \{ "(\bigcap \), "\leq 0", "\geq 0", "\geq 0, \alpha > 0", "\leq 0, \alpha > 0" \} . \)

The decision rule which consists in stating that \( \gamma < 0, \gamma > 0, \alpha > 0 \) according to as \( T_\gamma < \frac{t_{1-\xi_3}}{2}, n_4 \), \( T_\gamma > \frac{t_{1-\xi_1}}{2}, n_4 \), \( T_\alpha > t_{1-\xi_2}, n_4 \) respectively, is uniformly least risky among the unbiased decision rules when the total losses are obtained by adding the component ones. From this we obtain in a natural way the procedure
\[
\delta_2 = \delta(\phi_1^{(2)}, \ldots, \phi_4^{(2)}) \quad \text{where} \quad \phi_j^{(2)} = \phi_j^0, \quad j = 3, 4 \quad \text{and} \quad \phi_2^{(2)}(x) = \chi(T_\alpha > t_{1-\xi_2}, n_4) \land (T_\gamma \leq t_{1-\xi_1})(x) \]
\[
\phi_1^{(2)}(x) = \chi(T_\alpha \leq t_{1-\xi_2}, n_4) \land (T_\gamma \leq t_{1-\xi_1})(x). \]

By (4.1) we then have:
\[
r(\theta, \delta_2) - r(\theta, \delta_0) = \begin{cases} a_2(p_2^{(2)}(\theta) - p_2^0(\theta)) & \text{for } \alpha \leq 0 \\ b_2(p_1^{(2)}(\theta) - p_1^0(\theta)) & \text{for } \alpha > 0 \end{cases} \]

Thus, by (5.3) and \( f(W) = t_{1-\xi_2}, n_4 \Leftrightarrow T_\gamma = 0 \), we obtain
\[
r(\theta, \delta_2) - r(\theta, \delta_0) = \begin{cases} -a_2 p_\theta((f(W) < T_\alpha \leq t_{1-\xi_2}, n_4) \land (T_\gamma \leq t_{1-\xi_1})) & \text{for } \alpha \leq 0 \\ b_2 p_\theta((f(W) < T_\alpha \leq t_{1-\xi_2}, n_4) \land (T_\gamma \leq t_{1-\xi_1})) & \text{for } \alpha > 0 \end{cases} \]

Hence if the model and the context for the investigation are such that the unbiasedness property will be the most important one, then \( \delta_0 \) should be preferred. Also from the point of view of the performance function, one would probably prefer \( \delta_0 \) since this is more sensitive to alternatives of the type \( \alpha > 0 \) than \( \delta_2 (p_2^{(o)}(\theta) > p_2^{(2)}(\theta), \forall \alpha) \).

With Sheffe's multiple comparison method we find the situation very similar to the one above, which was to be expected since this
method takes into consideration many possible statements about \( \alpha, \gamma \) which we are not interested in.

The sections 2-5 seem to reveal a problem which is particularly inclined to arise in decision problems where several different parameters are of interest (contrary to for example the 3-decision problem about \( \gamma \) at step one). This we may perhaps call "the problem of optimal disposition of the amount of information in the observations with respect to the various different parameters", where "optimal" must be understood in relation to the nature and strength of interest in each of the parameters (except the nuisance parameters) which mainly depends on the outer circumstances of the investigation. In many multiple situations this optimum requirement seems to be reasonably well met with by the unbiasedness property which in a certain sense expresses the wish not to favour any single parameter at the expense of all the others. However, in the light of our example, situations can be imagined where it would be reasonable to relax this wish a little. The following argument may support this:

Proposition 4.1 shows that unbiasedness recommends a division into two steps where we on step one only consider \( \gamma \) and on step two (if it arises) only \( \alpha \). It is possible that we could achieve a more efficient "disposition of the amount of information" by dividing the problem into three steps where we in addition to the two already mentioned introduce an introductory step at which we so to speak take a general view of the situation before we eventually say something substantial at the second and third step. This idea is suggested by Sverdrup. To be more specific: At the first step a look is taken at \( \hat{\alpha} \) in order to obtain an idea whether it is "important" or not that step three arises, so that there are
two alternative methods at step two to judge \( \gamma \) depending on the result at the first step, and where step three arises or not depending on the result at step two. Thus, one may possibly be able to secure a better basis for the last step than by dividing the problem into only two steps (see Sverdrup [1], section 1.4.A).

Now, this objection against the unbiasedness principle is probably first of all relevant in a situation where the loss function \( (W_1) \) in a natural way arises out of the outer circumstances of the problem, a situation which perhaps is somewhat unusual. In a situation where the loss function primarily arises in order to assist the statistician in clarifying the problem, the objection could just as well aim at the loss function as well as against the unbiasedness property. When we constructed the loss functions \( W_1, W_2 \) and \( W_3 \), an important part of the argument was e.g. the division of the problem beforehand into two steps so that unbiasedness rather becomes a confirmation of this through proposition 4.1.

6. A Sufficient Condition that \( \mathcal{R} \in \mathcal{M} \) w.r.t. \( W_1 \)

Until now an essential assumption for \( W_1 \) is that \( a_1 \geq b_1 + 2b_2 \). If this is not satisfied then we shall see below that the meaning and consequences of the concepts of unbiasedness and minimax risque may change considerably. This section will also illustrate another technique given by Schaafsma [4]:

Let \( W \) and \( S(W) \) be as in (2.3). Then ([4], theorem 4.1):

**Theorem 6.1** For any m.d.p of type I with lossfunction \( W \), the following holds: If there exists a point in \( S(W) \) for which all the coordinates are equal, and which is both a minimax point and a maximin point (2.4-5), then \( \mathcal{R} \in \mathcal{M} \) (e.g. any unbiased procedure minimizes the maximal risque).
In checking whether the assumptions in theorem 6.1 are fulfilled, the following lemma ([4], lemma 4.2) may be useful:

**Lemma 6.2** Let \( w^* = \left( \begin{array}{c} w^* \\ \vdots \\ w^* \end{array} \right) \in S(W) \) be such that there exists a strategy \( g' = (g_1, \ldots, g_m) \) for player I with \( g_i > 0 \) for \( i = 1, \ldots, m \) and \( \sum g_i = 1 \) such that \( g'_w = w^* \) for \( j = 1, \ldots, n \).

\( (W = (w_1, \ldots, w_n)) \). Then \( w^* \) is both the only minimax point and the only maximin point in \( S(W) \).

We now return to our situation:

\[
W_1 = \begin{pmatrix}
0 & a_2 & a_1 & a_1 \\
b_2 & 0 & a_1 & a_1 \\
b_1 & a_2 + b_1 & 0 & a_1 + b_1 \\
b_1 & a_2 + b_1 & a_1 + b_1 & 0 \\
b_1 + b_2 & b_1 & a_1 + b_1 & 0 \\
b_1 + b_2 & b_1 & 0 & a_1 + b_1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
0 & t_7 & t_4 & t_4 \\
t_2 & 0 & t_5 & t_5 \\
t_1 & t_8 & 0 & t_6 \\
t_1 & t_8 & t_6 & 0 \\
t_3 & t_1 & t_6 & 0 \\
t_3 & t_1 & 0 & t_6
\end{pmatrix}
\]

For the moment we assume no restriction on \( W_1 \) (except \( a_1 > 0 \), \( b_i > 0 \)): Because of the structure of \( W_1 \), however, we have:

\[
(6.1) \begin{cases}
t_3 > t_1 \text{ and } t_2 \\
t_8 > t_1 \text{ and } t_7 \\
t_6 > t_4, t_5 \text{ and } t_1 \\
t_4 = t_5
\end{cases}
\]

We then have:

**Proposition 6.3** (i) If \( a_1 < b_1 \), no procedure is unbiased.

(ii) If \( a_1 = b_1 \) then every unbiased procedure is intolerant w.r.t \( \alpha \).
(iii) If \( a_1 > b_1 \) then \( \delta = \delta(\varphi_1, \ldots, \varphi_4) \in \mathcal{K} \) if and only if \( \delta \) satisfies the requirements in proposition 4.1.

(iv) If \( a_1 > b_1 \), \( a_2 \geq b_2 \) and
\[
b_2 \leq a_1 < b_2 + \frac{a_2}{b_2 + a_2} (b_1 + b_2)
\]
then \( \mathcal{R} \subseteq \mathcal{M} \).

(v) If \( \frac{\varepsilon_1}{2} = \frac{b_1}{a_1 + b_1} \), \( \varepsilon_2 = \frac{b_2}{a_2 + b_2} \) then
\[
a_1 < b_2 + \frac{a_2}{b_2 + a_2} (b_1 + b_2) \iff \frac{\varepsilon_1}{2} > \frac{b_1}{b_1 + b_2} \frac{1}{2 - \varepsilon_2}.
\]

Proof: We use (6.1) and earlier proofs. Consider proposition 3.3 we have \( 2t_4 \leq t_6 \iff a_1 \leq b_1 \). Then by (6.1) and the first part of the proof of proposition 3.3, (i) and (ii) follow. The proof of proposition 4.1 depends only on the assumption \( a_1 > b_1 \) (since \( a_1 \leq b_1 \Rightarrow p_3(\theta) + p_4(\theta) \geq 1 \)).
Hence (iii).

(iv): According to lemma 6.2 we will try to construct a strategy \( g' = (g_1, \ldots, g_6) \) for I with \( g_i > 0; i = 1, \ldots, 6 \) such that \( g^i w_j = w^* \), \( j = 1, \ldots, 4 \), where \( W_1 = (w_1, \ldots, w_4) \),
\[
w^* = b_1 + \frac{a_1 - b_1}{a_1 + b_1} v, \quad v = b_1 + \frac{a_2 b_2}{a_2 + b_2},
\]
(see corollary 3.4).

Let \( u_j = g^i w_j; \ j = 1, \ldots, 4 \).

Thus:
\[
\begin{align*}
\sum_{i=1}^{6} & g_i = 1 \\
u_1 &= b_2 (g_2 + g_5 + g_6) + b_1 (g_3 + g_4 + g_5 + g_6) = w^* \\
u_2 &= a_2 (g_1 + g_3 + g_4) + b_1 (g_3 + g_4 + g_5 + g_6) = w^* \\
u_3 &= a_1 (g_1 + g_2 + g_4 + g_5) + b_1 (g_4 + g_5) = w^* \\
u_4 &= a_1 (g_1 + g_2 + g_6) + b_1 (g_3 + g_6) = w^*
\end{align*}
\]
(6.2)
We have:

(1) \( u_3 = u_4 \iff \varepsilon_3 + \varepsilon_6 = \varepsilon_4 + \varepsilon_5 \)

Let \( q_2 = \varepsilon_3 + \varepsilon_6 \). We find:

\( u_1 = u_2 \iff (b_2 + a_2)g_2 + (b_2 + a_2)(g_5 + g_6) = a_2 \)

Thus:

(2) \( u_1 = u_2 \iff \varepsilon_2 + \varepsilon_5 + \varepsilon_6 = \frac{a_2}{b_2 + a_2} \iff \varepsilon_1 + \varepsilon_3 + \varepsilon_4 = \frac{b_2}{a_2 + b_2} \)

As in lemma 4.3 we introduce:

\[
\begin{align*}
 c_1 &= \frac{b_1}{a_1 + b_1} + \frac{a_2 b_2}{(a_1 + b_1)(a_2 + b_2)}, \\
 c_2 &= \frac{a_1}{a_1 + b_1} - \frac{a_2 b_2}{(a_1 + b_1)(a_2 + b_2)}
\end{align*}
\]

(\( c_1 c_2 = 1 \)). We have:

\[
a_2(\varepsilon_1 + \varepsilon_3 + \varepsilon_4) + b_1(\varepsilon_3 + \varepsilon_6) = a_1(\varepsilon_1 + \varepsilon_2 + \varepsilon_4 + \varepsilon_5)
\]

Hence by (2):

\[
\frac{a_2 b_2}{a_2 + b_2} + b_1(\varepsilon_3 + \varepsilon_6) = a_1(\varepsilon_1 + \varepsilon_2) + a_1(\varepsilon_3 + \varepsilon_6)
\]

Put \( q_1 = \varepsilon_1 + \varepsilon_2 \) (\( \iff q_1 + 2q_2 = 1 \)) whence:

\[
\frac{a_2 b_2}{a_2 + b_2} = a_1 q_1 + a_1 q_2 = a_1(1 - 2q_2) + (a_1 - b_1) q_2 = a_1 q_2(a_1 + b_1)
\]

which gives:

\[ q_2 = \frac{a_1}{a_1 + b_1} - \frac{a_2 b_2}{(a_1 + b_2)(a_1 + b_1)} = c_2 \]

Hence \( q_1 = 1 - 2q_2 = c_1 + c_2 - 2c_2 = c_1 - c_2 \)

Thus, by \( \sum \varepsilon_i = 1 \):

\[
\begin{align*}
 u_1 &= u_2 = u_3 = u_4 \iff \varepsilon_3 + \varepsilon_6 = \varepsilon_4 + \varepsilon_5 = c_2 \\
 \varepsilon_1 + \varepsilon_2 &= c_1 - c_2 \quad \& \quad \varepsilon_2 + \varepsilon_5 + \varepsilon_6 = \frac{a_2}{a_2 + b_2}
\end{align*}
\]

We have: \( w* = b_1 + (a_1 - b_1)c_1 = a_1 c_1 + (1 - c_1) = a_1 c_1 + b_1 c_2 \)

so that: \( u_1 = w* \iff \)

\[
\frac{b_2(\varepsilon_2 + \varepsilon_5 + \varepsilon_6) + 2b_1 c_2}{a_2 + b_2} = a_1 c_1 + b_1 c_2 \iff \frac{b_2 a_2}{a_2 + b_2} = a_1 c_1 - b_1 c_2
\]
But \( a_1c_1 - b_1c_2 = \frac{a_1b_1}{a_1+b_1} + a_1\frac{a_2b_2}{a_1(a_1+b_1)(a_2+b_2)} - \frac{b_1a_1}{a_1+b_1} + b_1\frac{a_2b_2}{a_1(a_1+b_1)(a_2+b_2)} \)

\[= \frac{a_2b_2}{a_2+b_2} \quad \text{which agrees with (3).} \]

Thus:

\[
\begin{align*}
\begin{cases}
g_1 + g_2 = c_1 - c_2 \\ g_3 + g_6 = c_2 \\ g_4 + g_5 = c_2 \\ g_1 + g_3 + g_4 = \varepsilon_2
\end{cases}
\end{align*}
\]

\[(6.2) \Leftrightarrow \begin{cases}
g_2 = c_1 - c_2 - g_1 \\ g_4 = \varepsilon_2 - g_1 - g_3 \\ g_5 = c_2 - \varepsilon_2 + g_1 + g_3 \\ g_6 = c_2 - g_3
\end{cases}
\]

\[(6.3) \]

In order to be able to choose \( g' \) as in lemma 6.2, we must have

\[c_2 > 0 \quad (\Leftrightarrow a_1 > \frac{a_2}{a_2+b_2} b_2) \quad \text{and} \quad c_1 > c_2 \quad (\Leftrightarrow c_1 > \frac{1}{2}). \]

By (4.4):

\[
\frac{1}{2} - c_1 < 0 \quad \Leftrightarrow \quad a_2(a_1-b_1-2b_2) + b_2(a_1-b_2) < 0
\]

\[\Leftrightarrow \quad a_1 - b_2 < \frac{a_2}{a_2+b_2}(b_1+b_2) \]

Hence: \( c_2 > 0 \) \& \( c_1 > c_2 \)

\[(4) \quad \frac{a_2}{b_2+a_2} b_2 < a_1 < b_2 + \frac{a_2}{a_2+b_2} (b_1+b_2) \]

which is possible and implied by the assumption in (iv).

Now assume that

\[(5) \quad b_2 \leq a_1 < b_2 + \frac{a_2}{a_2+b_2} (b_1+b_2) \quad \text{and} \quad a_2 \geq b_2 \]
First we show that (5) \[\Rightarrow\]
(6) \(\varepsilon_2 - c_2 < \min\{2c_2, 2(c_1 - c_2), \varepsilon_2\}\):

Since \(c_2 > 0\), \(\varepsilon_2 - c_2 < \varepsilon_2\).

We have: \(\varepsilon_2 - c_2 < 2c_2 \iff\)

\[\frac{b_2}{a_2 + b_2} < \frac{3a_1}{a_1 + b_1} - \frac{3a_2 b_2}{(a_1 + b_1)(a_2 + b_2)} \iff\]

\[3a_2(a_1 - b_2) + b_2(2a_1 - b_1) > 0\] which is implied by (5).

We have: \(\varepsilon_2 - c_2 < 2c_1 - 2c_2 \iff \varepsilon_2 < 3c_1 - 1\)

but \(c_1 > c_2 \Rightarrow c_1 > \frac{1}{2} \Rightarrow 3c_1 - 1 > \frac{3}{2} \geq \varepsilon_2 (a_2 \geq b_2)\)

and (6) follows.

Choose \(\eta > 0\) such that

\(\varepsilon_2 - c_2 < \eta < \min\{2c_2, 2(c_1 - c_2), \varepsilon_2\}\)

Let \(\delta_2 = \varepsilon_2 - \frac{\eta}{2}, \quad \delta_1 = c_1 - c_2 - \frac{\eta}{2}\)

\(\Rightarrow 0 < \delta_2 < c_2\) and \(0 < \delta_1 < c_1 - c_2\)

Let \(g_1 = g_3 = \frac{\eta}{2}\) whence \(g_1 = c_1 - c_2 - \delta_1, \quad g_3 = c_2 - \delta_2\)

Thus:

\(g_6 = c_2 - g_3 = \delta_2 > 0\)

\(g_5 = c_2 - \varepsilon_2 + g_1 + g_3 = \eta - (\varepsilon_2 - c_2) > 0\)

\(g_4 = \varepsilon_2 - \eta > 0\)

\(g_2 = c_1 - c_2 - g_1 = c_1 - c_2 - \frac{\eta}{2} > 0\)

which completes the proof of (iv).

(v): \(a_1 < b_2(1 - \varepsilon_2)(b_1 + b_2) = b_2(2 - \varepsilon_2) + b_1(1 - \varepsilon_2)\)

\(\iff\) \(a_1 < b_1\left(\frac{b_2}{b_1}(2 - \varepsilon_2) + 1 - \varepsilon_2\right) = b_1k\)

\(\iff\) \(a_1 + b_1 < b_1(k + 1)\) \(\iff\) \(\frac{b_1}{a_1 + b_1} > \frac{1}{1 + k}\)

\(\iff\) \(\frac{\varepsilon_1}{2} > \frac{1}{b_2(2 - \varepsilon_2) + 2 - \varepsilon_2} = \frac{b_1}{b_1 + b_2} \frac{1}{2 - \varepsilon_2}\)

Q.E.D.
(i) and (ii) of proposition 6.3 seem to be relevant primarily in a situation where the decision problem at step one is forced, e.g. where one is obliged a priori to conclude that \( \gamma = 0 \) when there is no reason to state that \( \gamma < 0 \) or \( \gamma > 0 \).

A connection between the result in proposition 6.3 and the performance function may be established in the following ways:

\[ \frac{\mu_1}{2}, \mu_2 \] correspond to the levels at step one and two respectively when we use the unbiasedness principle. On the one hand, if the losses due to errors of the second kind are of the same magnitude for both steps, we can achieve \( \mathcal{R} \leq \mathcal{M} \) if the level at step one is sufficiently increased. If for example \( b_1 = b_2 \), we must have

\[ \frac{\mu_1}{2} > \frac{1}{4 - \frac{1}{2}} > \frac{1}{4} = 0.25. \]

(For \( \mu_2 = 0.05 \) this is achieved by letting \( \mu_1 = 0.26 \) whereby the conditions in proposition 6.3 (iv) are satisfied.)

On the other hand, if there is a wish to control \( \frac{\mu_1}{2}, \mu_2 \) and also that \( \mathcal{R} \leq \mathcal{M} \) then this can be achieved by using a special loss function of the type \( \mathcal{W}_1 \): If for example \( \frac{\mu_1}{2} = \mu_2 = 0.05 \), we must have \( a_1 = 1.9b_1 \), \( a_2 = 19b_2 \). According to proposition 6.3 (v) it is necessary that \( \frac{b_1}{b_1 + b_2} < \frac{\mu_1}{2}(2 - \mu_2) = \frac{39}{400} \). Then

\[
\frac{b_1}{b_1 + b_2} = \frac{39}{429} \approx \frac{1}{11}, \text{ say; e.g. } b_2 = 10b_1. \text{ Hence } a_1 > b_2.
\]

Thus, if e.g. \( \mathcal{W}_1 \) is of the type \( b_1 = b, b_2 = 10b, a_1 = 19b, a_2 = 190b \ (b > 0) \), then \( \mathcal{R} \leq \mathcal{M} \).
7. References

[1] E. Sverdrup, "Multiple Decision Theory"
Lecture Notes, Series no. 15, Aarhus univ. 1969.

AMS Vol 28, 1957 (p 1).

AMS Vol. 28, 1957 (p 547-572)

Decision Problems of type I"

Approach"

[6] S. Karlin, "Mathematical Methods and Theory in Games,
Programming and Economics"