Some Remarks on Admissible Minimax Solutions of Statistical Decision Problems

by Erling Sverdrup

November 6, 1950

1. Introduction.

This paper deals with methods of finding minimax solutions of statistical decision problems. The method mostly used in statistical literature depends on the existence of a least favorable a priori distribution. In many situations of practical importance we don't know if such a distribution exists, notably in the case where the weight function is discontinuous or and the space $\mathcal{F}$ of probability distributions for the observable variable is not compact. The method which is proposed to be used in such situations follows from Theorem 1. The theorem is applied to the following problem. Let $(X_1, X_2, \ldots, X_n)$ be independent normal $(\theta, \sigma^2)$ where $\theta$ and $\sigma$ are unknown. It is to be decided whether $\sigma \leq c$ or $\sigma \geq c$ is true (c known). The penalty for the two kinds of error are $w_2$ and $w_1$ respectively. An admissible minimax solution is to let the decision depend on whether $\sum (X_i - \bar{X})^2 \geq k$ is fulfilled or not ($\bar{X} = \frac{1}{n} \sum X_i$). $k$ is determined such that the level of significance is $\frac{w_1}{w_1 + w_2}$ (where $\sigma \leq c$ being the "0-hypothesis"). The example also demonstrates of how little value it may be only to know that a statistical procedure is minimax unless this property is coupled with the property "admissibility."


For the convenience of the reader I shall in this section very briefly review A. Wald's General Theory [1].

Let \( X = \{X_1, X_2, \ldots, \text{ad. inf.}\} \) be an infinite sequence of random variables with joint cumulative probability distribution \( F(x) = \Pr[(X_1 \leq x_1) \wedge (X_2 < x_2) \ldots] \) where \( x = (x_1, x_2, \ldots) \). It is known that \( F \) belongs to a space \( \mathcal{F} \) of probability distributions.

\( D^t \) is the so-called terminal decision space consisting of elements \( d^t \).

\( D^e \) is the so-called experimental decision space consisting of elements \( d^e \) each of which has the form \( d^e = (i_1, i_2, \ldots, i_r) \) where \( i_1, i_2, \ldots, i_r \) is a finite sequence of integers. \( d^e \) will also be interpreted as the decision to observe \( X_{i_1}, \ldots, X_{i_r} \).

\( D = D^e + D^t \) is the decision space.

After the statistician has observed (in one or several stages) some components of \( X \) he has to make a choice between the different \( d^t \) in \( D^t \).

The weight function is \( W(F, d^t) \). It is the "penalty" for making a wrong decision. It is a real non-negative function of \( F \) and \( d^t \). The cost function \( c(x; d^e_1, \ldots, d^e_r) \) is for each \( r \) a real non-negative function of the observations \( x = (x_1, x_2, \ldots) \), and of the experimental decisions \( d^e_1, \ldots, d^e_r \). It is the cost of observing when the observation is \( x \), the number of stages is \( r \) and the stages are \( d^e_1, \ldots, d^e_r \). Of course \( c \) does only depend on those components of \( x \), the indices of which are contained in \( \sum_{j=1}^r d^e_j \).

A randomized statistical decision function \( \delta \) is a statistical procedure for making experimental and terminal decisions.

A Borel-field of measurable sets in \( D \) is defined. \( \delta \) is then defined by means of an infinite sequence of probability measures \( \delta(\overline{D}; 0), \delta(\overline{D}; x, d^e_1), \delta(\overline{D}; x, d^e_1, d^e_2) \ldots, \) where \( \overline{D} \) is an arbitrary measurable set in \( D \). This probability measure depends on the observation \( x \) and the experimental decisions previously taken.

If \( x \) has been observed and experimental decisions \( d^e_1, \ldots, d^e_r \) have been made before, then adapt with probability \( \delta(\overline{D}; x, d^e_1, \ldots, d^e_r) \) a decision belonging to \( \overline{D} \). The probability depends of course only on the components of \( x \) with indices in \( d^e_1 + \ldots + d^e_r \).
Furthermore \( D = \sum_{j=1}^{r} d_j^e \) is measurable and has probability measure \( 1 \).

\( \Delta \) is the space of all decision functions \( \delta \) which we wish to consider.

In the spaces defined above certain topologies are defined, either by means of a limit definition or a distance definition.

In the space \( \mathcal{F} \),
\[
\lim_{n \to \infty} F_n = F_0
\]
means that for any \( k \) and for any Borel-set \( S_k \) in the \( k \)-dim. Euclidian space \( \lim_{n \to \infty} \Pr[(X_1, \ldots, X_k) \in S_k | F_n] = \Pr[(X_1, \ldots, X_k) \in S_k | F_0] \), uniformly in \( S_k \).

In the space \( D^t \) the distance \( \rho(d_1^t, d_2^t) \) is defined by
\[
\rho(d_1^t, d_2^t) = \sup_{F \in \mathcal{F}} |W(F, d_1^t) - W(F, d_2^t)|
\]
In the space \( D = D^t + D^e \), open sets mean open set in \( D^t \) or sets in \( D^e \) consisting of a single element or any union of such sets in \( D^t \) and/or \( D^e \).

In the space \( \Delta \) of decision functions \( \lim_{n \to \infty} \delta_n = \delta_0 \) is defined as follows.

If \( X \) is a discrete random variable, then \( \lim \delta_n = \delta_0 \) means
\[
\lim \delta_n(\bar{B}; x, d_1^e, \ldots, d_r^e) = \delta(\bar{B}; x, d_1^e, \ldots, d_r^e)
\]
for any \( x, r, d_1^e \) and open set \( \bar{B} \) whose boundary has probability measure \( 0 \). If \( X \) admits a probability density the definition is more complicated and will not be given here. (Wald [1] page 65-66)

In any space in which a class of measurable sets is needed, it is the smallest Borel-field containing all open sets.

By an a priori probability measure \( \xi \) is meant a probability measure over the space \( \mathcal{F} \). \( \lim \xi_n = \xi_0 \) means that
\[
\lim \xi_n(\omega) = \xi_0(\omega)
\]
for any open set \( \omega \) in \( \mathcal{F} \).

Besides the "regular" topologies defined above, which are used in formulating the definitions, Wald introduces some auxiliary topologies, which however are only
needed for convenience in the course of some of the proofs.

The risk function \( r \) is the expected value of \( W + c \) corresponding to a specification of \( F \) and \( \xi \).

\[
(1) \quad r(F, \xi) = E(W + c)
\]

(2) \( \xi \) is the a priori probability

\[
\delta_0 \quad \text{is a Bayes solution in the strict sense if corresponding to some} \xi
\]

\[
r(\xi, \delta) = \inf_{\delta \in \Delta} r(\xi, \delta)
\]

\( \delta_0 \) is a Bayes solution in the wide sense if corresponding to some sequence \( \xi_1, \xi_2, \ldots \),

\[
(3) \quad \lim_{J} (r(\xi_j, \delta) - \inf_{\delta} r(\xi_j, \delta)) = 0
\]

\( \delta_0 \) is admissible if there is no \( \delta_1 \) such that

\[
(4) \quad r(F, \delta_1) \leq r(F, \delta_0) \quad \text{for all} \ F
\]

and

\[
(5) \quad r(F, \delta_1) < r(F, \delta_0) \quad \text{for some} \ F.
\]

(6) \( \delta_0 \) is a minimax solution if

\[
\sup_{F} r(F, \delta_0) = \inf_{\delta} \sup_{F} r(F, \delta)
\]

A statistical decision problem may be considered as a game in von Neumann's sense where the pure strategies for the two players are \( F \) and \( \delta \) respectively. The mixed strategy for the first player ("nature") is the a priori probability \( \xi \).

Two decisions \( \delta_1 \) and \( \delta_2 \) are said to be equivalent if

\[
r(F, \delta_1) = r(F, \delta_2) \quad \text{for all} \ F.
\]

The use of the term "unique solution" is relative to this definition of equivalence.

Two theorems are useful in finding admissible minimax solutions.

Theorem 1. If \( \delta_0 \) is a Bayes solution corresponding to \( \xi \) and if

\[
(7) \quad \Pr[r(F, \delta_0) = \sup_{F} r(F, \delta_0) | \xi] = 1
\]
then \( \delta_0 \) is a minimax solution.

**Theorem 2.** If \( \delta_0 \) is a Bayes solution in the wide sense which is unique relatively to some sequence \((\delta_1, \delta_2, \ldots)\) then \( \delta_0 \) is admissible.

These theorems are contained indirectly in several remarks made by Wald in [1], [2] and [3]. Special cases of Theorem 1 have been applied by Lehmann and Stein [4], Lehmann and Hodge [5] and others in finding minimax solutions. The proof of the theorem is almost the same as of Theorem 1 below except that condition 6 and the last part of the proof is not needed. I submit a proof of Theorem 2 (although some mathematician will undoubtedly call it "trivial")

**Proof of Theorem 2.** Suppose that \( \delta_0 \) fulfills (3) but is not admissible. Then there exists a \( \delta_1 \) fulfilling (4) and (5). From (4) we get

\[
\inf_{\delta} r(\delta, \delta) \leq r(\delta, \delta_1) \leq r(\delta, \delta_0)
\]

Because of (3) we then have

\[
\lim[r(\delta, \delta_1) - \inf_{\delta} r(\delta, \delta)] = 0
\]

i.e., \( \delta_1 \) is a Bayes solution. But since \( \delta_0 \) was unique, \( \delta_1 \) is equivalent to \( \delta_0 \) and

\[
r(F, \delta_1) = r(F, \delta_0)
\]

contrary to (5).

**Corollary of Theorem 2.** Any unique Bayes solution in the strict sense is admissible.

---

2.2. **Fundamental Assumptions Made by A. Wald.**

Wald's assumptions are (with his numbering),

**Assumption 3.1.** The stochastic process \( X = (X_1, X_2, \ldots) \) is either discrete or absolutely continuous.

**Assumption 3.2.** \( \Omega \) is separable. (This is a consequence of assumption 3.1).

**Assumption 3.3.** The weight function \( W(F, d^t) \) is a bounded function of \( F \) and \( d^t \).

**Assumption 3.4.** The space \( D^t \) is compact.
Assumption 3.5 is a restriction on the cost function. The restriction assures that the probability is that a terminal decision will be taken after a finite number of components of $X$ has been observed.

Assumption 3.6 restricts the form of $\Delta$ which may be the class of all randomized decisions. If $\Delta$ is a subset of this class then it must be closed in the topological sense. Furthermore the assumptions assures that superposing of a randomization on the randomized decision in $\Delta$ gives essentially a function in $\Delta$, i.e., there is no essential difference between a "pure" and a "mixed" strategy for the statistician. Furthermore $\delta$ is closed under truncation of the process.

The precise formulation of assumption 3.5-3.6 is given in Wald [1] page 63 and 68. Besides these assumptions some assumptions concerning measurability of the functions involved, are made.

From these assumptions Wald infers

Theorem 3 (Wald's Theorem 3.2). If assumptions 3.1 - 3.6 hold and if $\lim_{i \to \infty} \delta_i = \delta_0$, then

$$\lim_{i \to \infty} \inf \{ r(\xi, \delta_i) : \xi \in \Sigma \} \geq r(\xi, \delta_0)$$

For proof see Wald [1] page 77.

Under the assumptions 3.1 - 3.6 Wald [1] proves that there exists a Bayes solution relative to any a priori probability measure $\xi$. There exists a minimax solution and any minimax solution is a Bayes solution in the wide sense. Furthermore there exists a minimax solution $\delta_0$ which is a limit of Bayes solution $\delta_j$ in the strict sense relative to $\xi_j$ and the $\xi_j$'s may be chosen discrete and such that the probability mass may be concentrated in a finite number of points. (Wald's Theorem 3.12). In order to reach some further results Wald makes the

Assumption 3.7. The space $\Lambda$ is compact and $W(F, d^t)$ is a continuous function of $F$ uniformly in $d^t$.

If assumptions 3.1 - 3.7 are fulfilled Wald proves that there exists an a priori probability measure $\xi_0$ such that the Bayes solution $\delta_0$ relative to $\xi_0$ is a minimax solution. Such a $\xi_0$ is called a least favorable a priori probability measure.

It was remarked in Section 2 that Theorem 1 is useful in finding minimax solutions. The method used consists in specifying a $\xi_o$ which is believed to be least favorable, then finding the Bayes solution $\xi_o$ corresponding to this. If then (7) is fulfilled we have a minimax solution. If in addition $\xi_o$ is a unique Bayes solution then it is admissible.

One may also proceed as follows. Find a $\xi_o$ such that the risk is constant. Then (7) is obviously fulfilled and the only thing we have to do is to find a $\xi_o$ such that $\xi_o$ is the corresponding Bayes solution. If such a $\xi_o$ can be found, then $\xi_o$ is minimax.

It is obvious that the above procedure can only be applied in very special cases. If assumption 3.7 is not fulfilled then you have no guarantee that there exists a least favorable a priori distribution. There are many important situations in which 3.7 is not fulfilled. If for instance you want to test that $F$ belongs to a set $\omega_1$ against the hypothesis that $F$ belongs to $\omega_2$ and the penalty for the two kinds of errors are $w_1$ and $w_2$ respectively and if further the intersection of one of the sets with the closure of the other is non-empty, i.e., $\omega_1 \cap \omega_2$ is non-empty, then the weight function is obviously discontinuous and 3.7 is not fulfilled.

Below we shall be concerned with non-sequential testing of hypothesis in which cases assumptions 3.1 - 3.6 are almost always fulfilled. (In the case of point estimation however, there are important situations where 3.3 is not fulfilled. If we want to estimate a scalar parameter $\Theta$ in $F$ (which may have any real value) by means of an estimate $d^* - \Theta$ and if $W(F, d^*) = (\Theta - \Theta)^2$ then 3.2 is not fulfilled).

If only assumptions 3.1 - 3.6 and not assumption 3.7 is fulfilled then we would wish to generalize Theorem 1 such that the existence of a $\xi_o$ is not presumed. We want only to assume existence of a sequence $\xi_1, \xi_2, \ldots$, such that the minimax $\xi_o$ is a limit of its corresponding Bayes solutions. Such a sequence is known to exist (by Wald [1] Theorem 3.12). Theorem 4 below is a useful theorem.
which only assumes existence of such a sequence.

It should be pointed out that Assumption 3.7 is by no means a necessary condition for the existence of a least favorable a priori distribution, nor are assumptions 3.1 - 3.6 necessary for the existence of a sequence of a priori distributions which is "asymptotically least favorable." This is seen from the following example.

**Example 1.** Let \( X = (X_1, \ldots, X_n) \) have components which are independent normal \((0, 1)\). The procedure is non-sequential. We want to test the hypothesis that \( \theta = 0 \) against \( \theta \neq 0 \). \( \omega_1 \) contains the \( F \) for which \( \theta = 0 \), \( \omega_2 \) consists of all \( F \) such that \( \theta \neq 0 \). The weight function is,

\[
W(F, \omega_1) = |\theta| \\
W(F, \omega_2) = \begin{cases} 
\nu & \text{if } \theta = 0 \\
0 & \text{if } \theta \neq 0
\end{cases}
\]

Let \( \sigma(v) = \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt \), \( g(v) = \sigma'(v) \). Let \( \alpha \) and \( \gamma \) be the solution of the following equations.

\[
\alpha[g(\gamma + \alpha) - g(-\gamma + \alpha)] = 2\nu[1 - g(\gamma)] \\
g(\gamma + \alpha) - g(-\gamma + \alpha) = \alpha[g(\delta + \alpha) - g(-\delta + \alpha)]
\]

and let \( \eta \) be determined by

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

Then a (admissible) minimax solution is, "accept \( \omega_2 \) if

\[
\left| \frac{\bar{x}}{\sqrt{n}} \right| > \frac{\gamma}{\sqrt{n}}
\]

otherwise \( \omega_1 \), where \( \bar{x} \) is the sample mean. The least favorable a priori distribution \( \delta_o \) (for which \( \delta_o \) is a Bayes solution) is such that the probability that \( \theta = -\gamma \), \( +\gamma \) or 0 is 1. The probability that \( \theta = 0 \) is \( \eta \), the probability that \( \theta = -\gamma \) is the same as the probability that \( \theta = +\gamma \), namely \( \frac{1}{2}(1 - \eta) \). This above result is proved by applying Theorem 1. Furthermore by Theorem 2 the procedure is admissible.

If \( \theta \) were limited to an interval \((-A, A)\) then Assumptions 3.1 - 3.6 would be
fulfilled, but 3.7 is not fulfilled. The weight function is continuous, still a \( \delta_0 \) exists.

If \( G \) can take any value between \(-\infty\) and \(+\infty\) then assumption 3.3 is not fulfilled. Still \( \delta_0 \) is a Bayes solution in the wide sense (since it is a Bayes solution in the strict sense).

4. A General Theorem. We shall prove the following theorem.

Theorem 4. Suppose that there exists a sequence of a priori distribution \((\tilde{\delta}_1, \tilde{\delta}_2, \ldots)\), a sequence of statistical decision functions \((\delta_0, \delta_1, \delta_2, \ldots)\) and a sequence of real non-negative numbers \((\alpha_1, \alpha_2, \ldots)\) such that

1. \( \delta_j \) is a Bayes solution (in the strict sense) relative to \( \tilde{\xi}_j \)
   for \( j = 1, 2, \ldots \).

2. \( \delta_0 = \lim_{j \to \infty} \delta_j \) in the regular sense.

3. \( \lim_{j \to \infty} \alpha_j = 0 \).

4. \( r(F, \delta_j) \) is a bounded function of \( F \) and \( j \). For any sequence \( F_j \), \( \forall j = 1, 2, \ldots, \) ad. inf. for which the sequence \( r(F_j, \delta_j) \) converges for all \( j \) it does so uniformly in \( j \).

5. \[
\lim_{j \to \infty} \Pr[r(F, \delta_j) \geq \sup_{F} r(F, \delta_j) - \alpha_j \mid \tilde{\xi}_j] = 1
\]

6. Assumptions 3.1 - 3.6 of Wald are fulfilled.

Then \( \delta_0 \) is a minimax solution.

As a special case \( \alpha_1 = 0 \) for all \( i \). However it leaves us with greater freedom in the choice of the sequence \((\tilde{\xi}_1, \tilde{\xi}_2, \ldots)\) fulfilling the assumption in the theorem if we permit \( \alpha_1 \) to be positive. Note that \( \tilde{\xi}_0 = \lim \tilde{\xi}_1 \) need not exist (as a probability measure) and even if existing, \( \delta_0 \) may not be a Bayes.
Proof of the Theorem. Let

\[(14) \quad \omega_j = \sup_F r(F, \delta_j) \geq \sup_F r(F, \delta_j) - \alpha_j \]

and

\[(15) \quad \varepsilon_j = 1 - \Pr(F \notin \omega_j | \hat{\delta}_j). \]

Then

\[
\sup_F r(F, \delta_j) = \sup_F r(F, \delta_j) d \times \varepsilon_j \leq \varepsilon_j \sup_F r(F, \delta_j) \\
+ \int [r(F, \delta_j) + \alpha_j] d \times \varepsilon_j \leq \varepsilon_j \sup_F r(F, \delta_j) + r(\hat{\delta}_j, \delta_j) + \alpha_j \\
- \varepsilon_j \sup_F r(F, \delta_j) + \inf_{\delta_j} r(\hat{\delta}_j, \delta_j) + \alpha_j
\]

We then have

\[(16) \quad \sup_F r(F, \delta_j) \leq \varepsilon_j \sup_F r(F, \delta_j) + \inf_{\delta_j} r(\hat{\delta}_j, \delta_j) + \alpha_j \]

We can always find sequence \( F_y \) \( y = 1, 2, \ldots \), ad. inf. such that

\[(17) \quad \lim r(F_y, \delta_0) = \sup_F r(F, \delta_0) \]

By, if necessary, taking a subsequence of \( F_y \), \( y = 1, 2, \ldots \), (diagonal procedure) we can always secure that the sequences

\[r(F_y, \delta_j), \quad y = 1, 2, \ldots, \text{ad. inf.}\]

have limits for all \( j \).

We now have, by condition 4 of the theorem,

\[(18) \quad \lim_j \lim_y r(F_y, \delta_j) = \lim_y \lim_j r(F_y, \delta_j) \]

Taking the limit with respect to \( j \) on both sides of (16) and combining with (17) and (18) we get
\[(19) \quad \lim_{\gamma \to \infty} \lim_{j \to \infty} r(F_{\gamma}, \delta_j) \leq \inf_{\delta} \sup_{\xi} r(F, \delta, \xi) \]

By Theorem 3 the quantity on the left hand side of (19) is greater than or equal to
\[\lim_{\gamma} r(F_{\gamma}, \delta_\gamma) = \sup_{F} r(F, \delta_\gamma)\]
and this gives us
\[(20) \quad \sup_{F} r(F, \delta_\gamma) \leq \inf_{F} r(F, \delta)\]
and since the opposite inequality is obviously true, we have proved that \(\delta_\gamma\) is a minimax solution.

[Equation numbers (21) - (23) not used]

5. Examples.

Example 2. Let \((X_1, X_2, \ldots, X_n)\) be independent normal \((n, \sigma)\) where \(n\) and \(\sigma\) are unknown.

\[(24) \quad \omega_1 = \{n, \sigma | \sigma \leq c\} \]
\[(25) \quad \omega_2 = \{n, \sigma | \sigma > c\} \]
\[\Omega = \omega_1 + \omega_2 \]

The weight function is
\[(25) \quad W(n, \sigma; \omega_1) = w_1 \quad \text{if} \quad \sigma > c \]
\[= 0 \quad \text{otherwise} \]
\[W(n, \sigma; \omega_2) = w_2 \quad \text{if} \quad \sigma \leq c \]
\[= 0 \quad \text{otherwise} \]

In order to find a minimax solution we introduce two sequences of numbers \(\sigma_{1\gamma}, \sigma_{2\gamma}\), \(\gamma = 1, 2, \ldots, \) ad. inf. such that
\[(26) \quad \sigma_{1\gamma}^2 < c < \sigma_{2\gamma}^2\]
and

\[ \lim_{\nu \to \infty} \sigma_{j\nu} = c \quad \text{for } j = 1 \text{ or } 2. \]

Let further \( \nu_j, j = 1, 2, \ldots, \) be a sequence such that \( 0 < \nu_j < 1 \) and let \( \sigma_0 \)
be a number such that \( \sigma_0 > \sigma_{2j} \) for all \( j \).

A sequence of a priori probability measures \( j_j \) for \( (\theta, \sigma) \) is defined as follows.

\[ \Pr[ \sigma = \sigma_{1j} | j_j ] = \frac{1}{\nu_j} \]
\[ \Pr[ \sigma = \sigma_{2j} | j_j ] = 1 - \frac{1}{\nu_j} \]
\[ \Pr[ \sigma \leq \sigma_j | \sigma = \sigma_{1j} ] = \gamma_{1j}(\sigma_j) \quad 1 = 1, 2; j = 1, 2, \ldots, \text{ad. inf.} \]

where

\[ \gamma_j = \frac{n}{2} \frac{\sigma^2}{2(\sigma_j - \sigma_{ij})} \]

The unique Bayes solution \( j_j \) corresponding to \( j_j \) is then seen to be: "Accept \( \omega_j \)
with probability 1 if

\[ \sum (x_i - \bar{x})^2 \geq k_j \]

otherwise \( \omega_j \), where \( k_j \) depends in a known fashion on \( \nu_j, \sigma_0, \sigma_{1j}, \sigma_{2j} \).

The risk function corresponding to \( j_j \) is

\[ r(\theta, \sigma; j_j) = \left[ 1 - \gamma_j(k_j/\sigma^2) \right] \quad \text{if } \sigma \leq c \]
\[ r(\theta, \sigma; j_j) = \gamma_j(k_j/\sigma^2) \quad \text{if } \sigma > c \]

where \( \gamma_j(z) \) is the cumulative \( \chi^2 \) distribution with \( (n-1) \) degrees of freedom. Let \( k_j^0 \) be such that

\[ \left[ 1 - \gamma_j(k_j^0/\sigma_{1j}^2) \right] = \gamma_j(k_j^0/\sigma_{2j}^2) \]

and determine \( \nu_j \) such that \( k_j = k_j^0 \). Let \( j_j^0 \) denote the decision function \( j_j \)
with \( k_j = k_j^0 \).
Let \( r_j \) be the largest of the two right hand sides of (30) for \( k_j = k_j^o \) and \( \delta = c \). Let \( \delta_j, j = 1, 2, \ldots, \) be an arbitrary sequence of positive numbers such that \( \lim \delta_j = 0 \). For convenient choice of \( \delta_{1j} \) and \( \delta_{2j} \) we have

\[
(32) \quad r(\Theta, \sigma; \delta_j^o) \geq r_j = \delta_j
\]

for

\[ \delta = \delta_{1j} \text{ or } \delta_{2j} \].

Then

\[
\text{Probability of (32) } \geq \text{Probability of } \delta_{1j} \text{ or } \delta_{2j} = 1.
\]

It is seen from (31) that

\[
\lim k_j^o = k_0 \text{ where } k_0 \text{ is defined by}
\]

\[
(33) \quad \int (k_j^o)^c \frac{w_1}{w_1 + w_2}
\]

Let now \( \delta_0 \) be the following decision function "accept \( \omega_2 \) if

\[
(34) \quad \Sigma(x_1 - \bar{x})^2 \geq k_0
\]

otherwise \( \omega_1 \). Then, since \( r_j = \sup_{\Theta, \sigma} \sup \delta_j \) for all assumptions of Theorem 4 are fulfilled and \( \delta_0 \) is a minimax solution. Furthermore

\[
(35) \quad \inf \sup_{\delta} \sup_{\Theta, \sigma} r(\Theta, \sigma; \delta_j^o) = \frac{w_1 w_2}{w_1 + w_2}
\]

It is now easy to see that \( \delta_0 \) is admissible. Suppose that there were a \( \delta^\prime \) uniformly better than \( \delta_0 \). Let \( \delta^\prime_0(x) \) and \( \delta^\prime_0(x) \) be the probability of accepting \( \omega_2 \), according to \( \delta_0 \) and \( \delta^\prime \) respectively, if \( x \) is the sample point.

Let further

\[
(36) \quad F(\Theta, \sigma; \delta) = \int \delta(x) \ d F(x)
\]

Then

\[
(37) \quad r(\Theta, \sigma; \delta) = \begin{cases} w_2 \ P(\Theta, \sigma; \delta) & \text{if } \sigma \leq c \\ w_1 [1 - P(\Theta, \sigma; \delta)] & \text{if } \sigma > c. \end{cases}
\]

Let \( \tilde{\omega} = \{ F \mid \sigma = c \} \). Then \( \delta^o \) is similar with respect to \( \tilde{\omega} \). Suppose that \( \delta^\prime \) is non-similar. Then there exists a \( \Theta_0 = \Theta_0 \) such that

\[
(38) \quad P(\Theta_0, c; \delta^\prime) < \frac{w_2}{w_1 + w_2}
\]

Since \( P \) for all \( \delta \) is continuous in \( \Theta \) and \( \sigma \) there exists a \( \sigma_0 > c \) such that
(39) \[ P(\theta_0, \sigma_0; \delta') < \frac{w_2}{w_1 + w_2} \]

on the other hand we know that

(40) \[ P(\theta_0, \sigma_0; \delta_0) > \frac{w_2}{w_1 + w_2} \]

Since \( r(\theta, \sigma; \delta_0) \) obtains \( i \) to sup only for \( \sigma = c \). By (39), (40) and the second equation (37) we then obtain

(41) \[ r(\theta_0, \sigma_0; \delta') > r(\theta_0, \sigma_0; \delta_0) \]

contrary to the assumption that \( \delta' \) is uniformly better than \( \delta_0 \). It follows that \( \delta' \) must be similar. But Neyman and Pearson [6] have proved that \( P(\theta, \sigma; \delta_0) \) is uniformly smallest for \( \sigma < c \) and uniformly largest for \( \sigma > c \) among all \( P(\theta, \sigma; \delta) \) which equals

\[ \frac{w_2}{w_1 + w_2} \]

for \( \sigma = c \). Then \( \delta' \) can't be uniformly better.

It is easy to find a non-admissible minimax solution. Let \( \delta_1 \) be such that \( \Omega_2 \) is accepted with probability \( \frac{w_1}{w_1 + w_2} \). Then of course \( r(\theta, \sigma; \delta_1) = \frac{w_1 w_2}{w_1 + w_2} \)
and consequently minimax; but

\[ r(\theta, \sigma; \delta_1) > r(\theta, \sigma; \delta_0) \]

for \( \sigma \neq c \). The picture of the risk function looks roughly like this

![Risk Function Diagram]

The horizontal is \( r(\theta, \sigma; \delta_1) \) and the peaked curve is \( r(\theta, \sigma; \delta_0) \). Obviously most statisticians would in this case dislike a statistical procedure which disregards the statistical material and only takes into account some "lottery" number which is quite irrelevant to the problem.
By considering the arguments advanced above, it is easily seen that we can make the following general statement.

**Theorem 5.** A. Let $x = (x_1, \ldots, x_n)$ have independent normal $(\theta, \sigma)$ components and let $F$ denote the cumulative probability distribution of $x$. Let $\sigma_{1j}$ and $\sigma_{2j}$, $j = 1, 2, \ldots$, ad. inf. be two sequences of real numbers such that

$$\sigma_{1j} < \sigma_{2j}$$

and

$$\sigma_{1j} \leq \sigma \leq \sigma_{2j}$$

for all $j$. Let $\omega_1^*$, $i = 1, 2$, be the set of all $F$ such that $\sigma = \sigma_{1j}$ for some $j$, $\omega_1^{**}$ the set of all $F$ such that $\sigma \leq \sigma_1$, and $\omega_2^{**}$ the set of all $F$ such that $\sigma \geq \sigma_2$.

Let $\omega_1$ and $\omega_2$ be any sets of $F$ such that

$$\omega_1^* \subset \omega_1 \subset \omega_1^{**}$$

for $i = 1, 2$

$$\omega_1 \cap \omega_2 = \emptyset$$

Suppose we want to choose between $\omega_1$ and $\omega_2$ the penalty of wrongly accepting $\omega_1$ being $\omega_1^*$, $i = 1, 2$. An admissible minimax statistical procedure is then to accept $\omega_2$ if $\sum (x_i - \bar{x})^2 \geq k_0$ otherwise $\omega_1$, where $k_0$ is determined by

$$\int \frac{k_0^{\frac{1}{2n}}}{\sqrt{w_1/w_2}} = \frac{w_1}{w_1^{\frac{1}{2}} - w_2}$$

($\chi^2$ is the cumulative $\chi^2$ distribution with $n-1$ degrees of freedom, $\bar{x} = \frac{1}{n} \sum x_i$).

This procedure is a Bayes solution in the wide sense.

B. Let $\omega_1$ and $\omega_2$ be any two sets of $F$ such that

$$\omega_1 \cap \omega_2 = \emptyset$$

$$\sigma_1 = \sup_{F \in \omega_1} \sigma < \sigma_2 = \inf_{F \in \omega_2} \sigma$$

and 0 as well and $\omega_1$ as $\omega_2$ runs through all real numbers. The weight function is as under A. An admissible statistical decision is then as under A, except that $k_0$ is defined by
\[ w_2[1 - \hat{f}(k_0 / \sigma_1^2)] = w_1 \hat{f}(k_0 / \sigma_2^2) \]

This decision procedure is a Bayes solution in the strict sense. [The least favorable a priori distribution \( \xi_0 \) is given by (28) when \( \sigma_{1j}^j, \sigma_{2j}^j \) and \( \hat{\xi}_j \) are substituted by \( \sigma_1, \sigma_2 \) and \( \xi_0 \)].

(The statement that \( \delta_0 \) under A is a Bayes solution in the wide sense requires proof; but will not be given here).
References


