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 Ω of probability distributions for the observable variable is not compact. The method which is proposed to be used in such situations follows from Theorem 4. The theorem is applied to the following problem. Let $(X_1, X_2, \ldots, X_n)$ be independent normal (θ, σ) where θ and σ 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 $\Sigma(X_i - \bar{X})^2 \geq k$ is fulfilled or not ($\bar{X} = \frac{1}{n} \Sigma X_i$). k is determined such that the level of significance is $\frac{w_1}{w_1 + w_2}$. ($\sigma \leq c$ 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."

2. The General Theory of Statistical Decision Functions.

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

2.1. The General Set-up. Definitions and Notations.

Let $X = \left\{X_1, X_2, \ldots\ldots, \text{ad. inf.}\right\}$ be an infinite sequence of random variables with joint cumulative probability distribution $F(x) = Pr[(X_1 \leq x_1) \cdot (X_2 < x_2) \ldots\ldots]$ where $x = \left\{x_1, x_i, \ldots,\right\}$. It is known that F belongs to a space Ω 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 $d^e = \left\{i_1, i_2, \ldots, i_r\right\}$ where $i_1, \ldots, i_r$ 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 $c(x; d_1^e, \ldots, d_r^e)$ is for each r a real non-negative function of the observations $x = (x_1, x_2, \ldots)$, and of the experimental decisions $d_1^e, \ldots, d_r^e$. It is the cost of observing when the observation is x, the number of stages is r and the stages are $d_1^e, \ldots, d_r^e$. Of course c does only depend on those components of x, the indices of which are contained in $\sum_{j=1}^{r} d_j^e$.

A randomized statistical decision function δ is a statistical procedure for making experimental and terminal decisions.

A Borel-field of measurable sets in D is defined. δ is then defined by means of an infinite sequence of probability measures $\delta(\bar{D}; 0)$, $\delta(\bar{D}; x, d_1^e)$, $\delta(\bar{D}; x, d_1^e, d_2^e)$ ......, where $\bar{D}$ 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 $d_1^e, \ldots, d_r^e$ have been made before, then adapt with probability $\delta(\bar{D}; x, d_1, \ldots, d_r^e)$ a decision belonging to $\bar{D}$. The probability depends of course only on the components of x with indices in $d_1 + \ldots + d_r^e$.

Furthermore $D = \sum_{j=1}^{r} d_j^e$ is measurable and has probability measure 1.

Δ is the space of all decision functions δ 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 <u>space Ω</u>,

$$\lim_{n \to \infty} F_n = F_o$$

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

<u>In the space D^t</u> the distance $\rho(d_1^t, d_2^t)$ is defined by

$$\rho(d_1^t, d_2^t) = \sup_{F \in \Omega} \left| W(F, d_1^t) - W(F, d_2^t) \right| .$$

In the <u>space $D = D^t + D^e$</u>, 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 <u>space Δ</u> of decision functions $\lim_{n\to\infty} \delta_n = \delta_o$ is defined as follows. If X is a discrete random variable, then $\lim \delta_n = \delta_o$ means

$$\lim \delta_n(\bar{D}; x, d_1^e, \ldots, d_r^e) = \delta_o(\bar{D}; x, d_1^e, \ldots, d_r^e)$$

for any x, r, d_1^e and open set $\bar{D}$ 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 ξ is meant a probability measure over the space Ω. $\lim \xi_n = \xi_o$ means that

$$\lim \xi_n(\omega) = \xi_o(\omega)$$

for any open set ω in Ω.

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 δ.

(1) $r(F, \delta) = E(W + c)$

The average risk corresponding to the a priori probability ξ is

(2) $r(\xi, \delta) = \int r(F, \delta)\, d\xi$.

δ_0 is a <u>Bayes solution</u> in the strict sense if corresponding to some ξ

$$r(\xi, \delta_0) = \inf_{\delta \in \Delta} r(\xi, \delta)$$

δ_0 is a <u>Bayes solution in the wide sense</u> if corresponding to some sequence $\xi_1, \xi_2, \ldots,$

(3) $\lim[r(\xi_j, \delta_0) - \inf_{\delta} r(\xi_j, \delta)] = 0$

δ_0 is <u>admissible</u> if there is no δ_1 such that

(4) $r(F, \delta_1) \leq r(F, \delta_0)$ for all F

and

(5) $r(F, \delta_1) < r(F, \delta_0)$ for some F.

δ_0 is a minimax solution if

(6) $\sup_F r(F, \delta_0) = \inf_{\delta} \sup_F r(F, \delta)$

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 δ respectively. The mixed strategy for the first player ("nature") is the a priori probability ξ.

Two decisions δ_1 and δ_2 are said to be equivalent if

$$r(F, \delta_1) = r(F, \delta_2) \text{ for all } F.$$

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

Two theorems are useful in finding admissible minimax solutions.

<u>Theorem 1.</u> If δ_0 is a Bayes solution corresponding to ξ_0 and if

(7) $\Pr[r(F, \delta_0) = \sup_F r(F, \delta_0) \mid \xi_0] = 1$

then δ_0 is a minimax solution.

Theorem 2. If δ_0 is a Bayes solution in the wide sense which is unique relatively to some sequence $(\xi_1, \xi_2, \ldots)$ then δ_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 4 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 δ_0 fulfills (3) but is not admissible. Then there exists a δ_1 fulfilling (4) and (5). From (4) we get

$$\inf_{\delta} r(\xi_i, \delta) \leq r(\xi_i, \delta_1) \leq r(\xi_i, \delta_0)$$

Because of (3) we then have

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

i.e., δ_1 is a Bayes solution. But since δ_0 was unique, δ_1 is equivalent to δ_0 and

$$r(F, \delta_1) = r(F, \delta_0) \text{ for all } F$$

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. Ω 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 one 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 Δ which may be the class of all randomized decisions. If Δ 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 Δ gives essentially a function in Δ, i.e., there is no essential difference between a "pure" and a "mixed" strategy for the statistician. Furthermore δ 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 $\lim_{i\to\infty} \delta_i = \delta_o$, then

$$(8) \qquad \lim_{i=\infty} \inf r(\xi, \delta_i) \geq r(\xi, \delta_o)$$

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 ξ. There exists a minimax solution and any minimax solution is a Bayes solution in the wide sense. Furthermore there exists a minimax solution δ_o which is a limit of Bayes solution δ_j in the strict sense relative to ξ_j and the ξ'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 Ω 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 ξ_o such that the Bayes solution δ_o relative to ξ_o is a minimax solution. Such a ξ_o is called a least favorable a priori probability measure.

3. Some Remarks About Methods of Finding Minimax Solutions.

It was remarked in Section 2 that Theorem 1 is useful in finding minimax solutions. The method used consists in specifying a ξ_0 which is believed to be least favorable, then finding the Bayes solution ξ_0 corresponding to this. If then (7) is fulfilled we have a minimax solution. If in addition δ_0 is a unique Bayes solution then it is admissible.

One may also proceed as follows. Find a δ_0 such that the risk is constant. Then (7) is obviously fulfilled and the only thing we have to do is to find a ξ_0 such that δ_0 is the corresponding Bayes solution. If such a ξ_0 can be found, then δ_0 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 ω_1 against the hypothesis that F belongs to ω_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., $\bar{\omega}_1 \cdot \omega_2 + \omega_1 \cdot \bar{\omega}_2$ 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 case 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 θ in F (which may have any real value) by means of an estimate $d^t = \theta*$ and if $W(F, d^t) = (\theta - \theta*)^2$ 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 ξ_0 is not presumed. We want only to assume existence of a sequence $\xi_1, \xi_2, \ldots$, such that the minimax δ_0 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.

<u>Example 1.</u> Let $X = (X_1, \ldots, X_n)$ have components which are independent normal $(\theta, 1)$. The procedure is non-sequential. We want to test the hypothesis that $\theta = 0$ against $\theta \neq 0$. ω_1 contains the F for which $\theta = 0$, ω_2 consists of all F such that $\theta \neq 0$. The weight function is,

$$
\begin{aligned}
& W(F, \omega_1) = |\theta| && \\
(9) \qquad & W(F, \omega_2) = w && \text{if } \theta = 0 \\
& \qquad\qquad = 0 && \text{if } \theta \neq 0
\end{aligned}
$$

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

$$
(10) \qquad
\begin{aligned}
& \alpha[G(\gamma + \alpha) - G(-\gamma + \alpha)] = 2w[1 - G(\gamma)] \\
& G(\gamma + \alpha) - G(-\gamma + \alpha) = \alpha[g(\gamma + \alpha) - g(-\gamma + \alpha)]
\end{aligned}
$$

and let π be determined by

$$
(11) \qquad \pi w = (1 - \pi) e^{-\frac{1}{2}\alpha^2} \left[e^{\frac{\gamma\alpha}{\sqrt{n}}} + e^{-\frac{\gamma\alpha}{n}} \right]
$$

Then a (admissible) minimax solution is, "accept ω_2 if

$$
(12) \qquad |\bar{X}| \geq \frac{\gamma}{\sqrt{n}}
$$

otherwise ω_1", where $\bar{X}$ is the sample mean. The least favorable a priori distribution ξ_0 (for which δ_0 is a Bayes solution) is such that the probability that $\theta = -\gamma$, $+\gamma$ or 0 is 1. The probability that $\gamma = 0$ is π, the probability that $\theta = -\gamma$ is the same as the probability that $\theta = -\gamma$, namely $\frac{1}{2}(1 - \pi)$. This above result is proved by applying Theorem 1. Furthermore by Theorem 2 the procedure is admissible.

If θ 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 ξ_0 exists.

If θ can take any value between $-\infty$ and $+\infty$ then assumption 3.3 is not fulfilled. Still δ_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 $(\xi_1, \xi_2, \ldots)$, 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. δ_j is a Bayes solution (in the strict sense) relative to ξ_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_ν, $\nu = 1, 2, \ldots$, ad. inf. for which the sequence $r(F_\nu, \delta_j)$ converges for all j it does so uniformly in j.

5.

$$\lim_{j \to \infty} \Pr[r(F, \delta_j) \geq \sup_F r(F, \delta_j) - \alpha_j \mid \xi_j] = 1 \tag{13}$$

6. Assumptions 3.1 - 3.6 of Wald are fulfilled.

Then δ_0 is a minimax solution.

As a special case $\alpha_i = 0$ for all i. However it leaves us with greater freedom in the choice of the sequence $(\xi_1, \xi_2, \ldots)$ fulfilling the assumption in the theorem if we permit α_i to be positive. Note that $\xi_0 = \lim \xi_i$ need not exist (as a probability measure) and even if existing, δ_0 may not be a Bayes solution relative to ξ_o.

Proof of the Theorem. Let

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

and

(15) $$\epsilon_j = 1 - \Pr(F \in \omega_j \mid \xi_j).$$

Then

$$\sup_F r(F, \delta_j) = \int_{\Omega} \sup r(F, \delta_j)\, d\,\xi_j \leq \epsilon_j \sup_F r(F, \delta_j)$$
$$+ \int_{\omega_j} [r(F, \delta_j) + \alpha_j]\, d\,\xi_j \leq \epsilon_j \sup r(F, \delta_j) + r(\xi_j, \delta_j) + \alpha_j$$
$$= \epsilon_j \sup_F r(F, \delta_j) + \inf_\delta r(\xi_j, \delta) + \alpha_j$$

We then have

(16) $$\sup_F r(F, \delta_j) \leq \epsilon_j \sup_F r(F, \delta_j) + \inf_\delta \sup_\xi r(\xi, \delta) + \alpha_j$$

We can always find sequence F_γ, $\gamma = 1, 2, \ldots$, ad. inf. such that

(17) $$\lim r(F_\gamma, \delta_o) = \sup_F r(F, \delta_o)$$

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

$$r(F_\gamma, \delta_j), \quad \gamma = 1, 2, \ldots, \text{ad. inf.}$$

have limits for all j.

We now have, by condition 4 of the theorem,

(18) $$\lim_j \lim_\gamma r(F_\gamma, \delta_j) = \lim_\gamma \lim_j r(F_\gamma, \delta_j)$$

Taking the limit with respect to j on both sides of (16) and combining with (17) and (18) we get

$$(19) \qquad \lim_{\nu \to \infty} \lim_{j \to \infty} r(F_\nu, \delta_j) \leq \inf_{\delta} \sup_{\xi} r(\xi, \delta)$$

By Theorem 3 the quantity on the left hand side of (19) is greater than or equal to

$$\lim_{\nu} r(F_\nu, \delta_o) = \sup_F r(F, \delta_o)$$

and this gives us

$$(20) \qquad \sup_F r(F, \delta_o) \leq \inf_{\delta} \sup_F r(F, \delta)$$

and since the opposite inequality is obviously true, we have proved that δ_o 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 (θ, σ) where θ and σ are unknown

$$(24) \qquad \omega_1 = [\theta, \sigma \mid \sigma \leq c]$$

$$\omega_2 = [\theta, \sigma \mid \sigma > c]$$

$$\Omega = \omega_1 + \omega_2$$

The weight function is

$$(25) \qquad \begin{aligned} W(\theta, \sigma; \omega_1) &= w_1 && \text{if } \sigma > c \\ &= 0 && \text{otherwise} \\ W(\theta, \sigma; \omega_2) &= w_2 && \text{if } \sigma \leq c \\ &= 0 && \text{otherwise} \end{aligned}$$

In order to find a minimax solution we introduce two sequences of numbers $\sigma_{1\nu}$, $\sigma_{2\nu}$, $\nu = 1, 2, \ldots,$ ad. inf. such that

$$(26) \qquad \sigma^2_{1\nu} < c < \sigma^2_{2\nu}$$

and

(27) $\lim_{\nu \to \infty} \sigma_{j\nu} = c$ for $j = 1$ or 2.

Let further π_ν, $\nu = 1, 2, \ldots$, ad. inf. be a sequence such that $0 < \pi_\nu < 1$ and let σ_0 be a number such that $\sigma_0 > \sigma_{2\nu}$ for all ν.

A sequence of a priori probability measures ξ_j for (θ, σ) is defined as follows.

(28)
$$\Pr[\sigma = \sigma_{1j} \mid \xi_j] = \pi_j$$
$$\Pr[\sigma = \sigma_{2j} \mid \xi_j] = 1 - \pi_j$$
$$\Pr[\theta \leq \vartheta \mid \sigma = \sigma_{ij}] = \gamma_{ij}(\vartheta) \quad i = 1, 2;\ j = 1, 2, \ldots, \text{ad. inf.}$$

where

$$\gamma'_{ij} = \text{const.}\ e^{-\frac{n}{2(\sigma_0^2 - \sigma_{ij}^2)}\vartheta^2}$$

The unique Bayes solution δ_j corresponding to ξ_j is then seen to be. "Accept ω_2 with probability 1 if

(29)
$$\Sigma (X_i - \bar{X})^2 \geq k_j$$

otherwise ω_1", where k_j depends in a known fashion on π_j, σ_0, σ_{1j}, σ_{2j}.

The risk function corresponding to δ_j is

(30)
$$r(\theta, \sigma; \delta_j) = w_2[1 - \Gamma(k_j/\sigma^2)] \quad \text{if } \sigma \leq c$$
$$r(\theta, \sigma; \delta_j) = w_1 \Gamma(k_j/\sigma^2) \quad \text{if } \sigma > c$$

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

(31)
$$w_2[1 - \Gamma(k_j^0/\sigma_{1j}^2)] = w_1 \Gamma(k_j^0/\sigma_{2j}^2)$$

and determine π_j such that $k_j = k_j^0$. Let δ_j^0 denote the decision function δ_j with $k_j = k_j^0$.

Let r_j be the largest of the two right hand sides of (30) for $k_j = k_j^0$ and $\sigma = c$. Let α_j, $j = 1, 2, \ldots$, ad. inf. be an arbitrary sequence of positive numbers such that $\lim \alpha_j = 0$. For convenient choice of σ_{1j} and σ_{2j} we have

(32) $$r(\theta, \sigma; \delta_j^0) \geq r_j - \alpha_j$$

for

$$\sigma = \sigma_{1j} \text{ or } \sigma_{2j}.$$ Then

Probability of (32) $\geq$ Probability of σ_{1j} or $\sigma_{2j} = 1$. It is seem from (31) that $\lim k_j^0 = k_o$ where k_o is defined by

(33) $$\Gamma\left(\frac{k^0}{c}\right) = \frac{w_1}{w_1 + w_2}$$

Let now δ_o be the following decision function "accept ω_2 if

(34) $$\sum (X_i - \bar{X})^2 \geq k_o$$

otherwise ω_1". Then, since $r_j = \sup_{\theta, \sigma} r(\theta, \sigma; \delta_j)$, all assumptions of Theorem 4 are fulfilled and δ_o is a minimax solution. Furthermore

(35) $$\inf_{\delta} \sup_{\theta, \sigma} r(\theta, \sigma; \delta) = \frac{w_1 w_2}{w_1 + w_2}$$

It is now easy to see that δ_o is admissible. Suppose that there were a δ' uniformly better than δ_o. Let $\delta_o(x)$ and $\delta'(x)$ be the probability of accepting ω_2, according to δ_o and δ' respectively, if x is the sample point.

Let further

(36) $$P(\theta, \sigma; \delta) = \int \delta(x)\, d\, F(x)$$

Then

(37) $$r(\theta, \sigma; \delta) = w_2 \; P(\theta, \sigma; \delta) \qquad \text{if } \sigma \leq c$$
$$r(\theta, \sigma; \delta) = w_1[1 - P(\theta, \sigma; \delta)] \qquad \text{if } \sigma > c.$$

Let $\bar{\omega} = [F \mid \sigma = c]$. Then δ^0 is similar with respect to $\bar{\omega}$. Suppose that δ' is non-similar. Then there existsa $\theta = \theta_o$ such that

(38) $$P(\theta_o, c; \delta') < \frac{w_2}{w_1 + w_2}$$

Since P for all δ is continuous in θ and σ there exists a $\sigma_o > c$ such that

(39) $\qquad P(\theta_o, \sigma_o; \delta') < \dfrac{w_2}{w_1 + w_2}$

on the other hand we know that

(40) $\qquad P(\theta_o, \sigma_o; \delta_o) > \dfrac{w_2}{w_1 + w_2}$

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

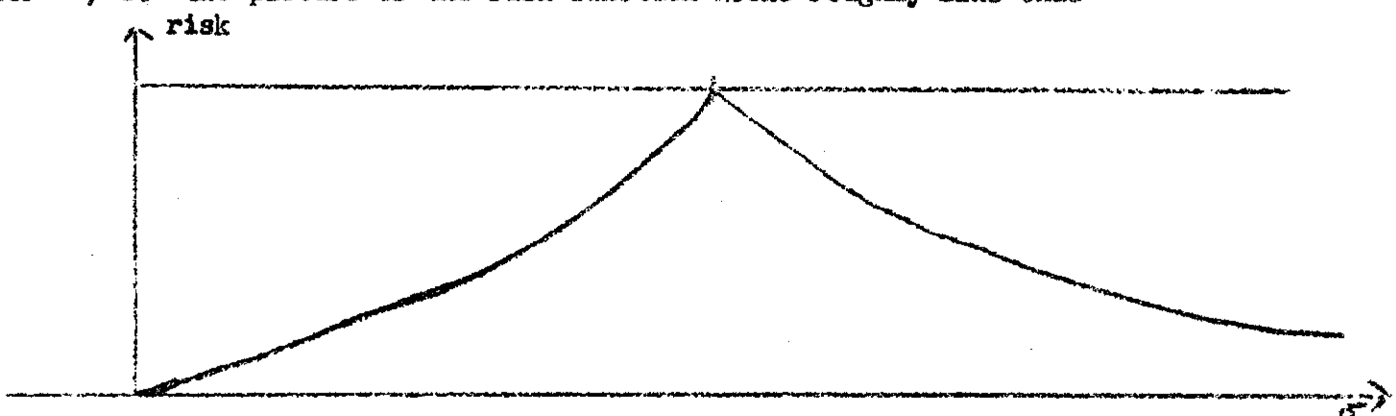
(41) $\qquad r(\theta_o, \sigma_o; \delta') > r(\theta_o, \sigma_o; \delta_o)$

contrary to the assumption that δ' is uniformly better than δ_o. It follows that δ' must be similar. But Neyman and Pearson [6] have proved that $P(\theta, \sigma; \delta_o)$ is uniformly smallest for $\sigma < c$ and uniformly largest for $\sigma > c$ among all $P(\theta, \sigma; \delta)$ which equals $\dfrac{w_2}{w_1 + w_2}$ for $\sigma = c$. Then δ' can't be uniformly better.

It is easy to find a non-admissible minimax solution. Let δ_1 be such that ω_2 is accepted with probability $\dfrac{w_1}{w_1 + w_2}$. Then of course $r(\theta, \sigma; \delta_1) = \dfrac{w_1 w_2}{w_1 + w_2}$ and consequently minimax; but

$$r(\theta, \sigma; \delta_1) > r(\theta, \sigma; \delta_o)$$

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



The horizontal is $r(\theta, \sigma; \delta_1)$ and the peaked curve is $r(\theta, \sigma; \delta_o)$. 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 (θ, σ) components and let F denote the cumulative probability distribution of X. Let σ_{1j} and σ_{2j}, $j = 1, 2, \ldots$, ad. inf. be two sequences of real numbers such that

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

and

$$\sigma_{1j} \leq c \leq c_{2j}$$

for all j. Let ω_i^*, $i = 1, 2$, be the set of all F such that $\sigma = \sigma_{ij}$ for some j, ω_1^{**} the set of all F such that $\sigma \leq c$, and ω_2^{**} the set of all F such that $\sigma \geq c$. Let ω_1 and ω_2 be any sets of F such that

$$\omega_i^* \subset \omega_i \subset \omega_i^{**} \qquad \text{for } i = 1, 2$$

$$\omega_1 \cdot \omega_2 = 0$$

Suppose we want to choose between ω_1 and ω_2 the penalty of wrongly accepting ω_i being w_i, $i = 1, 2$. An admissible minimax statistical procedure is then to accept ω_2 if $\Sigma(X_i - \bar{X})^2 \geq k_o$ otherwise ω_1, where k_o is determined by

$$\Gamma\left(\frac{k^o}{c}\right) = \frac{w_1}{w_1 + w_2}$$

[$\Gamma(z)$ is the cumulative χ^2 distribution with n-1 degrees of freedom, $\bar{X} = \frac{1}{n} \Sigma X_i$].
This procedure is a Bayes solution in the wide sense.

B. Let ω_1 and ω_2 be any two sets of F such that

$$\omega_1 \cdot \omega_2 = 0$$

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

and θ as well and ω_1 as ω_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_o is defined by

$$w_2[1 - \Gamma(k_o/\sigma_1^2)] = w_1\Gamma(k_o/\sigma_2^2)$$

This decision procedure is a Bayes solution in the strict sense. [The least favorable a priori distribution ξ_o is given by (28) when σ_{1j}, σ_{2j} and ξ_j are substituted by σ_1, σ_2 and ξ_o].

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

References

[1]. Abraham Wald: "Statistical Decision Functions", New York, 1950.

[2]. Abraham Wald: "Foundation of a General Theory of Sequential Decision Functions," Econometrica, 1947.

[3]. Abraham Wald: "Statistical Decision Functions," Annals of Mathematical Statistics, 1949

[4]. E. L. Lehmann and C. Stein: "Most Powerful Tests of Composite Hypothesis I. Normal Distributions," The Annals of Mathematical Statistics, 1948.

[5]. Y. L. Hodges, Jr., and E. L. Lehmann: "Some Problems in Minimax Point Estimation," The Annals of Mathematical Statistics, 1950

[6]. J. Neyman and E. S. Pearson: "On the Problems of the Most Efficient Tests of Statistical Hypothesis," Philosophical Transactions of the Royal Society, London, A 231 (1933) p. 289.