CONTRIBUTIONS TO THE THEORY OF STATISTICAL ESTIMATION AND TESTING HYPOTHESES[1]

By Abraham Wald

1. **Introduction.** Let us consider a family of systems of n variates $X_1(\theta^{(1)}, \cdots, \theta^{(k)}), \cdots, X_n(\theta^{(1)}, \cdots, \theta^{(k)})$ depending on k parameters $\theta^{(1)}, \cdots, \theta^{(k)}$. A system of k values $\theta^{(1)}, \cdots, \theta^{(k)}$ can be represented in the k-dimensional parameter space by the point θ with the co-ordinates $\theta^{(1)}, \cdots, \theta^{(k)}$. Denote by Ω the set of all possible points θ. For any point θ of Ω we shall denote by $P(E \epsilon w|\theta)$ the probability that the sample point $E = (x_1, \cdots, x_n)$ falls into the region w of the n-dimensional sample space, where x_j denotes the observed value of the variate $X_j(\theta)(j = 1, \cdots, n)$. The distribution $P(E \epsilon w|\theta)$ is supposed to be known for any point θ of Ω. In the theory of testing hypotheses and of statistical estimation we have to deal with problems of the following type: A sample point $E = (x_1, \cdots, x_n)$ of the n-dimensional sample space is given. We know that x_j is the observed value of $X_j(\theta)$ but we do not know the parameter point θ, and we have to draw inferences about θ by means of the sample point observed. The assumption that θ belongs to a certain subset ω of Ω is called a hypothesis. We shall deal in this paper with the following general problem: Let us consider a system S of subsets of Ω. Denote by H_ω the hypothesis corresponding to the element ω of S, and by H_S the system of all hypotheses corresponding to all elements of S. We have to decide by means of the observed sample point E which hypothesis of the system H_S should be accepted. That is to say for each H_ω we have to determine a region of acceptance M_ω in the n-dimensional sample space. The hypothesis H_ω will be accepted if and only if the sample point E falls in the region M_ω. M_ω and $M_{\omega'}$ are disjoint if $\omega \neq \omega'$. The statistical problem is the question as to how the system M_S of all regions M_ω should be chosen.

The problem in this formulation is very general. It contains the problems of testing hypotheses and of statistical estimation treated in the literature.[2] For instance if we want to test the hypothesis H_ω corresponding to a certain subset ω of Ω, the system of hypotheses H_S consists only of the two hypotheses H_ω and $H_{\bar{\omega}}$ where $\bar{\omega}$ denotes the subset of Ω complementary to ω. If we want to estimate θ by a unique point, then S is the system of all points of Ω. In the theory of confidence intervals we estimate one of the parameter co-ordinates $\theta^{(1)}, \cdots, \theta^{(k)}$,

[1] Research under a grant-in-aid from the Carnegie Corporation of New York.

[2] See, for instance, J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Transactions of the Royal Society*, London, Vol. 231 (1937), pp. 333–380.

say $\theta^{(1)}$, by an interval. In this case S is a certain system of subsets ω of the following type: ω is the set of all points $\theta = (\theta^{(1)}, \cdots, \theta^{(k)})$ for which $\theta^{(1)}$ lies in a certain interval $[a, b]$. The problem in our formulation covers also cases which, as far as I know, have not yet been treated. Consider for instance 3 subsets ω_1, ω_2 and ω_3 of Ω such that the sum of them is equal to Ω. It may be that we are interested only to know in which of the subsets ω_1, ω_2, ω_3 the unknown parameter point lies. In this case the system of hypotheses H_S consists only of the 3 hypotheses H_{ω_1}, H_{ω_2} and H_{ω_3}. Cases like this might be of practical interest.

For the determination of the "best" system (in a certain sense) of regions of acceptance we shall use methods and principles which are closely related to those of the Neyman-Pearson theory of testing hypotheses. In the Neyman-Pearson theory two types of error are considered. Let $\theta = \theta_1$ be the hypothesis to be tested, where θ_1 denotes a certain point of the parameter space. Denote this hypothesis by H_1 and the hypothesis $\theta \neq \theta_1$ by $\bar{H}$. The type I error is that which is made by rejecting H_1 when it is true. The type II error is made by accepting H_1 when it is false. The fundamental principle in the Neyman-Pearson theory can be formulated as follows: among all critical regions (regions of rejection of H_1, i.e. regions of acceptance of $\bar{H}$) for which the probability of type I error is equal to a given constant α, we have to choose that region for which the probability of type II error is a minimum. The difficulty which arises here lies in the circumstance that the probability of type II error depends on the true parameter point θ. That is to say, if the critical region is given the probability of type II error will be a function of the true parameter point θ. Since we do not know the true parameter point θ, we want to have a critical region which minimizes the probability of type II error with respect to any possible alternative hypothesis $\theta = \theta_2 \neq \theta_1$. If such a common best critical region exists, then the problem is solved. But such cases are rather exceptional. If a common best critical region does not exist, Neyman and Pearson consider unbiased critical regions of different types,[3] which minimize the type II error locally, that is to say with respect to alternative hypotheses in the neighborhood of the hypothesis considered. In this paper we develop methods for the determination of a system of regions of acceptance taking in account type II errors also relative to alternative hypotheses not lying in the neighborhood of the hypothesis to be tested.

2. **Some Definitions.** Let us denote by Ω the set of all possible parameter points θ and by S a system of subsets of Ω. If ρ denotes the sum of the elements of a subset σ of S, then we shall denote ΣM_ω by M_ρ, where M_ω denotes the

[3] J. Neyman and E. S. Pearson: *Statistical Research Memoirs*, Volumes I and II. The authors consider also unbiased regions of type A_1 for which the probability of type II error with respect to every alternative hypothesis is not greater than for any other unbiased region of the same size. However regions of type A_1 do not always exist (the existence of such regions has been proved for a special but important class of cases).

region of acceptance of H_ω and the summation is to be taken over all elements ω of σ.

Definition 1. Denote by M_S and M_S' two different systems of regions of acceptance corresponding to the same system H_S of hypotheses. The systems M_S and M_S' are said to be equivalent if for each point θ of Ω and for every ρ which is a sum of elements of S which does not contain θ, the equation

$$P(E \in M_\rho' \mid \theta) = P(E \in M_\rho \mid \theta)$$

holds, where M_ρ' denotes the region according to the system M_S' and M_ρ denotes the region according to the system M_S.

Definition 2. Denote by M_S and M_S' two different systems of regions of acceptance corresponding to the same system of hypotheses. The system M_S' is said to be absolutely better than the system M_S if they are not equivalent and if for each θ and for every ρ which is a sum of elements of S which does not contain θ the inequality

$$P(E \in M_\rho' \mid \theta) \leq P(E \in M_\rho \mid \theta)$$

holds.

Definition 3. A system M_S of regions of acceptance is said to be admissible if no absolutely better system of regions exists.

3. **The problem of the choice of M_S.** The choice of M_S will in general be affected by the following two circumstances:

(1) We do not attribute the same importance to each error. For instance the acceptance of the hypothesis that θ lies in a certain interval I has in general more serious consequences if θ is far from I than if θ is near to I. The choice of M_S will in general depend on the relative importance of the different possible errors.

(2) In some cases we have a priori more confidence that the true parameter point lies in a certain interval I than in some other cases. The choice of M_S will in general be affected also by this fact. Let us illustrate this by an example. We have two coins, a new and an old one and we want to test for both coins whether the probability p of tossing head is equal to $\frac{1}{2}$. Let us assume that we make 100 tosses with each of the coins and we get head 40 times in each case. Since we have a priori no very great confidence that the old coin is unbiased, the fact that head occured only 40 times will suffice to reject the hypothesis that for the old coin $p = \frac{1}{2}$. But in the case of the new coin, having much greater a priori confidence that it is unbiased, we shall perhaps not reject the hypothesis $p = \frac{1}{2}$ and we shall rather assume that a somewhat improbable event occurred. That is to say, we do not choose the same critical region in both cases due to the fact that our a priori confidence for $p = \frac{1}{2}$ is in the case of the new coin greater than in the case of the old one.

In order to study the dependence of the choice of M_S on the two circumstances

mentioned, let us introduce a weight function for the possible errors and an a priori probability distribution for the unknown parameter θ. The weight function $W(\theta, \omega)$ is a real valued non-negative function defined for all points θ of Ω and for all elements ω of S, which expresses the relative importance of the error committed by accepting H_ω when θ is true. If θ is contained in ω, $W(\theta, \omega)$ is, of course, equal to zero. The question as to how the form of the weight function $W(\theta, \omega)$ should be determined, is not a mathematical or statistical one. The statistician who wants to test certain hypotheses must first determine the relative importance of all possible errors which will entirely depend on the special purposes of his investigation. If that is done, we shall in general be able to give a more satisfactory answer to the question as to how the system of regions of acceptance should be chosen. In many cases, especially in statistical questions concerning industrial production, we are able to express the importance of an error in terms of money, that is to say, we can express the loss caused by the error considered in terms of money. We shall also say that $W(\theta, \omega)$ is the loss caused by accepting H_ω when θ is true.

The situation regarding the introduction of an a priori probability distribution of θ is entirely different. First, the objection can be made against it, as Neyman has pointed out, that θ is merely an unknown constant and not a variate, hence it makes no sense to speak of the probability distribution of θ. Second, even if we may assume that θ is a variate, we have in general no possibility of determining the distribution of θ and any assumptions regarding this distribution are of hypothetical character. On account of these facts the determination of the system of regions of acceptance should be independent of any a priori probability considerations. The "best" system of regions of acceptance, which we shall define later, will depend only on the weight function of the errors. The reason why we introduce here a hypothetical probability distribution of θ is simply that it proves to be useful in deducing certain theorems and in the calculation of the best system of regions of acceptance.

Let us denote by $f(\theta)$ a distribution function of θ. For the sake of simplicity let us assume that the probability density of the distribution $P(E \in w \mid \theta)$ exists in any point E of the sample space for any θ and denote it by $p(E \mid \theta)$. The expected value of the loss is given by

$$I = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE \tag{1}$$

where ω_E denotes the element of S corresponding to E (that is to say, ω_E is that element of S for which E is a point of the region of acceptance M_{ω_E}), and the integral is to be taken over the product of the sample space M with the parameter space Ω. The expected value I of the loss depends on the system M_S of regions of acceptance. The system M_S for which I becomes a minimum, can be regarded as the best system of regions relative to the given weight function and to the given a priori distribution of θ.

One can easily show the following: If M_S' is an absolutely better system of regions (in sense of the definition 2) than the system M_S, then for any weight

function $w(\theta, \omega)$ and for any a priori distribution $f(\theta)$ the expected value I' of the loss corresponding to M_S' is less than the expected value I of the loss corresponding to M_S. (For some exceptional weight and a priori distribution functions I' may be equal to I.)

Hence we can give the following rule: *We have to choose an admissible system of regions of acceptance.*

Now let us consider the question whether besides admissibility further restrictions upon the choice of M_S can be made. In order to see this, let us consider two admissible systems of regions M_S and M_S' which are not equivalent. One can easily show that there exist two weight functions $W_1(\theta, \omega)$, $W_2(\theta, \omega)$ and two a priori distributions $f_1(\theta)$ and $f_2(\theta)$ such that for $W_1(\theta, \omega)$ and $f_1(\theta)$ the expected value of the loss corresponding to M_S is less than that corresponding to M_S', and for $W_2(\theta, \omega)$ and $f_2(\theta)$ the expected value of the loss corresponding to M_S is greater than that corresponding to M_S'. Hence no absolute criteria can be given as to which of the systems M_S and M_S' should be chosen. In order to be able to make further restrictions upon the choice of M_S, we have to make assumptions regarding the form of the weight function. We shall deal with this question in section 6.

4. **Calculation of admissible systems of regions.** As we have seen, we have to choose an admissible system of regions. The question arises as to how we can find admissible systems of regions.

Provided that $p(E \mid \theta)$ is continuous in E and θ jointly, one can easily show that M_S' is an admissible system of regions if there exists a bounded, uniformly continuous and everywhere positive (except if θ is contained in ω) weight function $W(\theta, \omega)$ and an a priori distribution $f(\theta)$ such that every open subset of Ω has a positive probability and the expected value of the loss

$$I(M_S) = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE. \tag{2}$$

becomes a minimum for $M_S = M_S'$. (ω_E denotes that element of S for which M_{ω_E} contains E). In fact if there existed an absolutely better system M_S'' of regions, then $I(M_S'')$ would be less than $I(M_S')$ in contradiction to our assumption that $I(M_S)$ becomes a minimum for $M_S = M_S'$.

In order to obtain an admissible system M_S we may choose any bounded, uniformly continuous and everywhere positive (except if θ is contained in ω) weight function $W(\theta, \omega)$ and any arbitrary a priori distribution $f(\theta)$ (subject only to the condition that every open subset of Ω should have a positive probability) and then the system M_S which makes

$$I(M_S) = \int_M \int_\Omega W(\theta, \omega_E) p(E \mid \theta)\, df(\theta)\, dE$$

a minimum is an admissible one. In order to determine M_S we have only to determine for each E the corresponding element ω_E of S. Let us consider the integral

$$I_E = \int_\Omega W(\theta, \omega) p(E \mid \theta)\, df(\theta).$$

The integral I_E is for a fixed E only a function of ω. It is obvious that ω_E must be that element of S for which I_E becomes a minimum.

5. **Admissible systems M_S and the Neyman-Pearson best critical regions.** Let us consider the case that the system H_S of hypotheses consists only of the following two hypotheses: 1) $\theta = \theta_0$ where θ_0 is a certain point of Ω. 2) θ belongs to the set complementary to θ_0. Let us denote by ω_1 the set consisting only of the point θ_0, and by ω_2 the set complementary to ω_1. S consists in this case only of two elements ω_1 and ω_2. The system M_S of regions consists of two regions of acceptance M_{ω_1} and M_{ω_2} corresponding to the hypotheses H_{ω_1} and H_{ω_2}. If a common best critical region in the sense of Neyman-Pearson exists and if M_S is admissible, then M_{ω_2} is obviously a common best critical region. This leads to the following remarkable conclusion: If a common best critical region exists and if the system M_S of regions consisting of the two regions M_{ω_1} and M_{ω_2} minimizes the expectation of the loss (formula 2) for a weight function and for an a priori distribution subject to some weak conditions mentioned in paragraph 4, then M_{ω_2} is a common best critical region. That is to say, the form of the weight function and of the a priori distribution affects only the size of the region M_{ω_2} but it will always be a common best critical region.

6. **The choice of M_S if a weight function is given.** We shall now consider the case in which a weight function $W(\theta, \omega)$ is given and we shall deal with the question as to how M_S in this case is to be chosen.

If the parameter point is an unknown constant and if θ denotes the true parameter point, then the expected value of the loss is given by

$$r(\theta) = \int_M W(\theta, \omega_E) p(E \mid \theta)\, dE \tag{3}$$

where the integration is to be taken over the whole sample space M and H_{ω_E} denotes the hypothesis accepted if E is the observed sample point. That is to say ω_E is that element of S for which E is contained in the region of acceptance M_{ω_E}. We shall call the expression (3) the risk of accepting a false hypothesis if θ is the true parameter point. Since we do not know the true parameter point θ, we shall have to study the risk $r(\theta)$ as a function of θ. We shall call this function the risk function. The form of the risk function depends on the system M_S of regions and on the form of the weight function. In order to express this fact, we shall denote the risk function corresponding to the system M_S and to the weight function $W(\theta, \omega)$ also by

$$r[\theta \mid M_S, W(\theta, \omega)].$$

Definition 4. Denote by M_S and M'_S two systems of regions of acceptance corresponding to the same system H_S of hypotheses. We shall say that M_S and M'_S are equivalent relative to the weight function $W(\theta, \omega)$ if the risk function

$r[\theta \mid M_S, W(\theta, \omega)]$ is identically equal to the risk function $r[\theta \mid M_S', W(\theta, \omega)]$, that is to say if for each point θ,

$$r[\theta \mid M_S', W(\theta, \omega)] = r[\theta \mid M_S, W(\theta, \omega)].$$

Definition 5. Denote by M_S and M_S' two systems of regions corresponding to the same system H_S of hypotheses. We shall say that M_S is uniformly better than M_S' relative to the weight function $W(\theta, \omega)$ if M_S and M_S' are not equivalent and for each θ

$$r(\theta \mid M_S, W(\theta, \omega)] \leq r[\theta \mid M_S', W(\theta, \omega)].$$

Definition 6. A system M_S of regions of acceptance is said to be admissible relative to the weight function $W(\theta, \omega)$ if no uniformly better system of regions exists relative to the weight function considered.

It is obvious that *we have to choose a system M_S of regions which is admissible relative to the weight function considered.*

There exist in general many systems M_S which are admissible relative to the weight function given. The question arises as to how can we distinguish among them. Denote by r_{M_S} the maximum of the risk function corresponding to the system M_S of regions and to the given weight function. If we do not take into consideration a priori probabilities of θ, then it seems reasonable to choose that system M_S for which r_{M_S} becomes a minimum. We shall see in section 8 that the system M_S for which r_{M_S} becomes a minimum has some important properties which justify the distinction of this particular system of regions among all admissible systems.

Definition 7. We shall call an admissible system M_S' of regions for which r_{M_S} becomes a minimum a best system of regions of acceptance relative to the weight function given.[4]

Now we shall have to deal with the question of determining a best system M_S of regions and what special properties this system M_S has.

7. Reduction of the problem to the case when the system H_S of hypotheses is the system of all simple hypotheses. A hypothesis H_ω is said to be a simple hypothesis if ω contains exactly one point of the parameter space Ω. We assume that each element ω of S is a closed subset of Ω. Hence the power of S is not greater than the power of the continuum and therefore we can always set up a correspondence between the elements ω of S and the points θ of Ω such that to each point θ corresponds a certain element ω_θ of S and to each element ω of S at least one point θ exists for which $\omega_\theta = \omega$. For instance if S consists of the two elements ω_1 and ω_2 then we can set up a correspondence as follows: the element ω_θ of S corresponding to θ is ω_1 if θ is contained in ω_1 and ω_2 otherwise.

[4] As we shall see later (Theorem 3), the best system of regions is uniquely determined if some regularity conditions are fulfilled.

If Ω is one dimensional and S is the system of all intervals of a certain length ϵ then we can define the interval ω_θ corresponding to θ as the interval of which the initial point is θ and the terminal point $\theta + \epsilon$.

Let us denote the weight function by $W(\theta, \omega)$ defined for all values of θ and for all elements ω of S. Consider the system $H_{\bar{S}}$ of all simple hypotheses and the following weight function

$$W(\theta, \bar{\theta}) = W(\theta, \omega_{\bar{\theta}}) \tag{4}$$

where θ denotes the true parameter point and $\bar{\theta}$ denotes the estimated point. A system $M_{\bar{S}}$ of regions of acceptance for $H_{\bar{S}}$ is given by a vector function $\bar{\theta}(E)$ of the observations such that to each point $E = (x_1, \cdots, x_n)$ of the sample space M corresponds a certain point $\bar{\theta}(E)$ of the parameter space. For each point θ_0 the region M_{θ_0} of the acceptance of the hypothesis $\theta = \theta_0$ is given by the equation $\bar{\theta}(E) = \theta_0$. We shall call the function $\bar{\theta}(E)$ an estimate of θ, the system of regions $M_{\bar{S}}$ is uniquely determined by the estimate. We shall call $\bar{\theta}(E)$ a best estimate relative to a given weight function if the system of regions determined by $\bar{\theta}(E)$ is a best system of regions relative to the weight function considered.

Let us denote by $\bar{\theta}(E)$ a best estimate of θ relative to the weight function $W(\theta, \bar{\theta})$ defined in (4). A best system M_S of regions of acceptance in the original problem can obviously be obtained in the following way: Denote by ω an element of S. The region M_ω of acceptance of the hypothesis H_ω consists of the points E for which

$$\omega_{\bar{\theta}(E)} = \omega.$$

Hence we can restrict our considerations to the case when the system of hypotheses is the system of all simple hypotheses. We shall deal with the problem of how a best estimate of θ can be found and what properties this estimate has.

8. **Some theorems concerning the best estimate.** In order to study the properties of a best estimate $\bar{\theta}(E)$ it is useful to consider hypothetical a priori distributions of θ. We shall especially consider point distributions of θ, that is to say, distributions where a finite number of points $\theta_1, \cdots, \theta_s$ of the parameter space Ω exist such that the probability of any subset of Ω not containing any of the points $\theta_1, \cdots, \theta_s$ is zero. If $\theta_1, \cdots, \theta_s$ are given, a point distribution is characterized by a vector $\rho = (\rho_1, \cdots, \rho_s)$ where ρ_i denotes the probability of θ_i and $\Sigma\rho_i = 1$.

If $\theta(E)$ denotes an estimate of θ and if $f(\theta)$ denotes a distribution function of θ then the expected value of the loss, that is to say the expected value of the weight function $W[\theta, \theta(E)]$ is obviously given by

$$\int_M \int_\Omega W[\theta, \theta(E)]p(E \mid \theta)\, df(\theta)\, dE \tag{5}$$

where $p(E \mid \theta)$ denotes the probability density in E if θ is the true parameter point and the integration is to be taken over the product of the sample space M and parameter space Ω.

Let us assume that for every sample point E there exists a parameter point $\theta_f(E)$ such that the expression

$$\int_\Omega W(\theta, \bar{\theta}) p(E \mid \theta)\, df(\theta) \tag{6}$$

becomes a minimum with respect to $\bar{\theta}$ for $\bar{\theta} = \theta_f(E)$. We shall call the estimate $\theta_f(E)$ a minimum risk estimate with respect to the distribution $f(\theta)$, since also the expression (5) becomes a minimum for the estimate $\theta_f(E)$.

We shall make the following assumptions:

Assumption 1. The parameter space is a bounded and closed subset of the k-dimensional Euclidean space.

Assumption 2. The weight function $W(\theta, \bar{\theta})$ is continuous in θ and $\bar{\theta}$ jointly.

Assumption 3. The probability density $p(E \mid \theta)$ is continuous in E and θ jointly. That is to say if $\lim E_i = E$ and $\lim \theta_i = \theta$ then $\lim p(E_i \mid \theta_i) = p(E \mid \theta)$.

Assumption 4. For any distribution $f(\theta)$ of θ there exists at most one minimum risk estimate $\theta_f(E)$.[5]

Assumption 5. If $f(\theta)$ and $f'(\theta)$ denote two different point distributions of θ and if $\theta_f(E)$ and $\theta_{f'}(E)$ are minimum risk estimates corresponding to $f(\theta)$ and $f'(\theta)$ respectively, then $\theta_f(E)$ is not identically equal to $\theta_{f'}(E)$.

The assumptions 1–5, with addition of an assumption 6 which we shall formulate later, enables us to deduce important properties of the best estimate $\bar{\theta}(E)$. First we shall prove some Lemmas by means of the assumptions 1–5.

LEMMA 1. *For any a priori distribution $f(\theta)$ of θ there exists exactly one minimum risk estimate $\theta_f(E)$.*

According to Assumption 2 $W(\theta, \bar{\theta})$ is continuous. Since the parameter space Ω is compact on account of Assumption 1, $W(\theta, \bar{\theta})$ is uniformly continuous. According to Assumption 3 $p(E \mid \theta)$ is continuous; hence for any fixed sample point E, $p(E \mid \theta)$ is bounded. From these facts it follows easily that the expression (6) is a continuous function of $\bar{\theta}$ for any fixed sample point E. Hence there exists at least one parameter point $\theta_f(E)$ such that (6) becomes a minimum for $\bar{\theta} = \theta_f(E)$. Since, according to Assumption 4, at most one parameter point exists for which (6) becomes a minimum, Lemma 1 is proved.

If a distribution $f(\theta)$ of θ is given then the distribution of each of the components $\theta^{(1)}, \cdots, \theta^{(k)}$ of θ can be found. Denote by Q_j the set of real numbers which are discontinuities of the distribution of the component $\theta^{(j)} (j = 1, \cdots, k)$ and form the set $Q = Q_1 + \cdots + Q_k$. As is well known, Q is at most denumerable. A k-dimensional interval J of the parameter space given by

$$a_j \leq \theta^{(j)} \leq b_j \qquad (j = 1, \cdots, k)$$

is called a continuity interval of the distribution $f(\theta)$ if no a_j and no b_j belongs to Q. A sequence $\{f_n(\theta)\}$ of distributions is said to be convergent towards the

[5] As will be shown in Section 10, Assumption 4 is not as restrictive as it would appear. It will be satisfied in the great majority of practical cases.

distribution $f(\theta)$, i.e. in symbols $\lim f_n(\theta) = f(\theta)$, if for any continuity interval J of $f(\theta)$ the probability of J corresponding to the distribution $f_n(\theta)$ converges with increasing n towards the probability of J corresponding to the distribution $f(\theta)$.

LEMMA 2. *If* $\{f_n(\theta)\}$ $(n = 1, \cdots, \text{ad inf.})$ *denotes a sequence of distributions, then there exists a subsequence* $\{f_{n_m}(\theta)\}$ $(m = 1, \cdots, \text{ad inf.})$ *which converges towards a distribution.*

As is well known, there exists a completely additive set function $P(\omega)$ defined for all Borel measurable subsets ω of Ω and a subsequence $\{n_m\}$ of $\{n\}$, such that for any continuity interval J of $P(\omega)$ the probability of J corresponding to the distribution $f_{n_m}(\theta)$ converges with increasing m towards $P(J)$. Since Ω is bounded, there exists a continuity interval J such that for all n the probability of J according to $f_n(\theta)$ is equal to 1. Hence $P(\Omega) = 1$, that is to say, $P(\omega)$ is a probability set function which proves Lemma 2.

LEMMA 3. *If* $\{f_n(\theta)\}$ $(n = 1, \cdots, \text{ad inf.})$ *denotes a sequence of distributions which converges towards the distribution* $f(\theta)$ *and if* $\lim E_n = E$ *then*

$$\lim_{n=\infty} \theta_{f_n}(E_n) = \theta_f(E),$$

where $\theta_{f_n}(E)$ *denotes the minimum risk estimate corresponding to* $f_n(\theta)$ *and* $\theta_f(E)$ *denotes the minimum risk estimate corresponding to* $f(\theta)$.

If $\{\varphi_n(\theta)\}$ denotes a sequence of real valued functions which converges uniformly towards a continuous function $\varphi(\theta)$ then

$$\lim \int_\Omega \varphi_n(\theta)\, df_n(\theta) = \int_\Omega \varphi(\theta)\, df(\theta). \tag{7}$$

Since $\{\varphi_n(\theta)\}$ converges uniformly towards $\varphi(\theta)$, (7) is obviously true if

$$\lim \int_\Omega \varphi(\theta)\, df_n(\theta) = \int_\Omega \varphi(\theta)\, df(\theta)$$

holds. The latter equality follows easily from the fact that Ω is compact.

Consider a subsequence $\{n_m\}$ of $\{n\}$ such that $\lim_{m=\infty} \theta_{f_{n(m)}}(E_{n_m})$ exists. Denote this limit by θ^*. In order to prove Lemma 3, we have only to show that $\theta^* = \theta_f(E)$. If $\theta_f(E) \neq \theta^*$ then on account of Assumption 4

$$\int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df(\theta) < \int_\Omega W(\theta, \theta^*)p(E \mid \theta)\, df. \tag{8}$$

$W(\theta, \bar{\theta})$ is uniformly continuous since Ω is compact. On account of Assumption 3 also $p(E \mid \theta)$ is uniformly continuous in the product of Ω with a bounded subset of the sample space. Hence

$$W[\theta, \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)$$

converges uniformly in θ towards

$$W(\theta, \theta^*)p(E \mid \theta)$$

and we have on account of (7) and (8)

$$\lim_{m=\infty} \int_\Omega W[\theta, \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} = \int_\Omega W(\theta, \theta^*)p(E \mid \theta)\, df > \int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df, \tag{9}$$

and

$$\lim_{m=\infty} \int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df_{n_m} = \int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df. \tag{10}$$

From (9) and (10) it follows that there exists a positive δ such that for sufficiently large m

$$\int_\Omega W[\theta, \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} > \int_\Omega W[\theta, \theta_f(E)]p(E \mid \theta)\, df_{n_m} + \delta.$$

Since the sequence of functions $\{p(E_n \mid \theta)\}$ converges uniformly in θ towards $p(E \mid \theta)$, we have for sufficiently large m

$$\int_\Omega W[\theta, \theta_{f_{n(m)}}(E_{n_m})]p(E_{n_m} \mid \theta)\, df_{n_m} > \int_\Omega W[\theta, \theta_f(E)]p(E_{n_m} \mid \theta)\, df_{n_m}.$$

But this is a contradiction, since $\theta_{f_n}(E)$ is a minimum risk estimate. Hence the assumption $\theta^* \neq \theta_f(E)$ is proved to be an absurdity. This proves Lemma 3.

LEMMA 4. *To each positive ϵ a bounded and closed subset M_ϵ of the sample space M can be given such that*

$$\int_{M_\epsilon} p(E \mid \theta)\, dE \geq 1 - \epsilon$$

for every point θ of the parameter space Ω.

Let us assume that Lemma 4 is not true and we shall deduce a contradiction. Denote by $M_\nu (\nu = 1, 2, \cdots, \text{ad inf.})$ the sphere in the sample space M whose center is the origin and whose radius is equal to ν. Since Lemma 4 is supposed to be not true, to each ν there exists a parameter point θ_ν such that

$$\int_{M_\nu} p(E \mid \theta_\nu)\, dE < 1 - \epsilon \qquad (\nu = 1, \cdots, \text{ad inf.}). \tag{11}$$

Since Ω is compact, there exists a subsequence $\{\theta_{\nu_\mu}\}$ of the sequence $\{\theta_\nu\}$ such that $\lim_{\mu=\infty} \theta_{\nu_\mu}$ exists. Denote $\lim \theta_{\nu_\mu}$ by θ. Since

$$\int_M p(E \mid \theta)\, dE = 1$$

there exists a positive integer ν' such that

$$\int_{M_{\nu'}} p(E \mid \theta)\, dE > 1 - \frac{\epsilon}{2}.$$

On account of Assumption 3 we get easily

$$\lim_{\mu=\infty} \int_{M_{\nu'}} p(E \mid \theta_{\nu_\mu})\, dE = \int_{M_{\nu'}} p(E \mid \theta)\, dE.$$

Hence for sufficiently large μ we get

$$\int_{M_{\nu_\mu}} p(E \mid \theta_{\nu_\mu})\, dE \geq \int_{M_{\nu'}} p(E \mid \theta_{\nu_\mu})\, dE > 1 - \epsilon,$$

in contradiction to (11). This proves Lemma 4.

For any estimate $\theta(E)$ we shall call the integral

$$r(\theta) = \int_M W[\theta, \theta(E)]p(E \mid \theta)\, dE$$

the risk function of the estimate $\theta(E)$. The value of the risk function $r(\theta)$ is for any θ equal to the expected value of the loss (of the weight function) if θ is the true parameter point.

LEMMA 5. *To any positive η a positive δ can be given such that for any estimate $\theta(E)$ and for any pair θ, θ' of parameter points whose Euclidean distance is less than δ the inequality*

$$|r(\theta) - r(\theta')| = \left| \int_M W[\theta, \theta(E)]p(E \mid \theta)\, dE - \int_M W[\theta', \theta(E)]p(E \mid \theta')\, dE \right| < \eta$$

holds.

Since $W(\theta, \bar{\theta})$ is uniformly continuous, to any $\epsilon > 0$ a positive δ can be given such that for any pair of points θ, θ' whose Euclidean distance is less than δ the relation

$$|W(\theta, \bar{\theta}) - W(\theta', \bar{\theta})| < \epsilon \tag{12}$$

holds for every $\bar{\theta}$. On account of Assumption 3 δ can be chosen in such a way that also the inequality

$$|p(E \mid \theta) - p(E \mid \theta')| < \epsilon \tag{13}$$

is satisfied for any sample point E of a bounded subset M' of M and for any pair θ, θ' whose Euclidean distance is less than δ.

Since $W(\theta, \bar{\theta})$ is continuous and Ω is compact, $W(\theta, \bar{\theta})$ must be bounded. Denote by A an upper bound of $W(\theta, \bar{\theta})$. According to Lemma 4 there exists a bounded and closed subset M' of the sample space M such that

$$\int_{M'} p(E \mid \theta)\, dE \geq 1 - \frac{\eta}{2A} \text{ for any } \theta.$$

It is obvious that

$$\left| \int_{M-M'} W[\theta, \theta(E)]p(E \mid \theta)\, dE - \int_{M-M'} W[\theta', \theta(E)]p(E \mid \theta')\, dE \right| \leq \frac{\eta}{2}.$$

In order to prove Lemma 5 we have only to show that

$$\left| \int_{M'} W[\theta, \theta(E)]p(E \mid \theta)\, dE - \int_{M'} W[\theta', \theta(E)]p(E \mid \theta')\, dE \right| < \frac{\eta}{2}. \tag{14}$$

On account of (12) and (13), (14) is certainly true for sufficiently small ϵ. Hence Lemma 5 is proved.

LEMMA 6. *If the sequence $\{f_n(\theta)\}$ of distributions converges towards the distribution $f(\theta)$ and if $r_{f_n}(\theta)$ denotes the risk function of the minimum risk estimate $\theta_{f_n}(E)$ then $\{r_{f_n}(\theta)\}$ converges uniformly towards the risk function $r_f(\theta)$ of the minimum risk estimate $\theta_f(E)$.*

According to Lemma 4 to any positive ϵ a bounded and closed subset M_ϵ of M can be given such that

$$\int_{M_\epsilon} p(E \mid \theta)\, dE \geq 1 - \epsilon \tag{15}$$

for every θ. From Lemma 3 it follows easily that $\{\theta_{f_n}(E)\}$ converges uniformly towards $\theta_f(E)$ in M_ϵ. Hence

$$\lim_{n=\infty} \int_{M_\epsilon} W[\theta, \theta_{f_n}(E)]p(E \mid \theta)\, dE = \int_{M_\epsilon} W[\theta, \theta_f(E)]p(E \mid \theta)\, dE$$

holds for every θ and for every positive ϵ. Since $W(\theta, \bar{\theta})$ is bounded and ϵ can be chosen arbitrarily small, we get on account of (15) that

$$\lim_{n=\infty} \int_{M} W[\theta, \theta_{f_n}(E)]p(E \mid \theta)\, dE = \int_{M} W[\theta, \theta_f(E)]p(E \mid \theta)\, dE,$$

that is to say

$$\lim r_{f_n}(\theta) = r_f(\theta).$$

The uniformity of the convergence follows easily from Lemma 5.

In the following argument we shall consider an arbitrary but fixed system of s parameter points $\theta_1, \cdots, \theta_s$, and point distributions such that no point $\theta \neq \theta_1, \cdots, \theta_s$ has positive probability. Such a point distribution is characterized by a vector $\rho = (\rho_1, \cdots, \rho_s)$ where ρ_i denotes the probability of θ_i $(i = 1, \cdots, s)$ and $\Sigma\rho_i = 1$. The points $\theta_1, \cdots, \theta_s$ are kept constant and only ρ will vary. Hence if we speak about different distributions $\rho = (\rho_1, \cdots, \rho_s)$, $\rho' = (\rho_1', \cdots, \rho_s')$ they are always related to the same points $\theta_1, \cdots, \theta_s$ unless we state explicitly the contrary.

LEMMA 7. *If $\rho = (\rho_1, \cdots, \rho_s)$ and $\rho' = (\rho_1 + \Delta\rho_1, \cdots, \rho_s + \Delta\rho_s)$ denote two different distributions then*

$$\sum_{i=1}^{s} [(\lambda - 1)\rho_i + \lambda\Delta\rho_i][r_i(\rho') - r_i(\rho)] < 0$$

holds for any positive λ, *where*

$$r_i(\rho) = \int_M W[\theta_i, \theta_\rho(E)]p(E \mid \theta_i)\, dE \qquad (i = 1, \cdots, s),$$

$$r_i(\rho') = \int_M W(\theta_i, \theta_{\rho'}(E)]p(E \mid \theta_i)\, dE,$$

and $\theta_\rho(E)$ *and* $\theta_{\rho'}(E)$ *denote the minimum risk estimates corresponding to* ρ *and* ρ' *respectively.*

We have

$$\sum_i (\rho_i + \Delta\rho_i)r_i(\rho) = \int_M \sum_i W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i)\, dE = I_1$$

and

$$\sum_i (\rho_i + \Delta\rho_i)r_i(\rho') = \int_M \sum W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i)\, dE = I_2.$$

Since $\theta_{\rho'}(E)$ is the minimum risk estimate corresponding to ρ', we have $I_1 \geq I_2$. We shall show that $I_1 > I_2$. According to Assumption 5 $\theta_\rho(E)$ is not identically equal to $\theta_{\rho'}(E)$. Hence there exists a point E' such that $\theta_\rho(E') \neq \theta_{\rho'}(E')$. On account of Assumption 4

$$\Sigma W[\theta_i, \theta_\rho(E')]\rho_i' p(E' \mid \theta_i) > \Sigma W[\theta_i, \theta_{\rho'}(E')]\rho_i' p(E' \mid \theta_i).$$

From Lemma 3 it follows that $\theta_\rho(E)$ and $\theta_{\rho'}(E)$ are continuous functions of E. Hence there exists a positive δ and a sphere s with center in E' such that

$$\Sigma W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i) > \Sigma W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i) + \delta$$

for every point E of S. Since $\theta_{\rho'}(E)$ is the minimum risk estimate corresponding to ρ' we have

$$\Sigma W[\theta_i, \theta_\rho(E)]\rho_i' p(E \mid \theta_i) \geq \Sigma W[\theta_i, \theta_{\rho'}(E)]\rho_i' p(E \mid \theta_i)$$

for every point E outside S. Hence $I_1 > I_2$ that is to say

$$\Sigma(\rho_i + \Delta\rho_i)r_i(\rho) > \Sigma(\rho_i + \Delta\rho_i)r_i(\rho'). \tag{16}$$

Analogously we get

$$\Sigma\rho_i r_i(\rho) < \Sigma\rho_i r_i(\rho'). \tag{17}$$

Multiplying (16) by an arbitrary positive value λ and subtracting (17) we get

$$\Sigma[\lambda(\rho_i + \Delta\rho_i) - \rho_i]r_i(\rho) > \Sigma[\lambda(\rho_i + \Delta\rho_i) - \rho_i]r_i(\rho').$$

Hence

$$\Sigma[(\lambda - 1)\rho_i + \lambda\Delta\rho_i][r_i(\rho') - r_i(\rho)] < 0.$$

Let us denote for any ρ the maximum of the numbers

$$r_1(\rho), \cdots, r_s(\rho)$$

by $r(\rho)$. We shall call a distribution ρ for which $r(\rho)$ becomes a minimum, a risk-minimizing distribution. We shall say that the risk-minimizing distribution $\rho = (\rho_1, \cdots, \rho_s)$ is not degenerate if $\rho_1 > 0, \cdots, \rho_s > 0$. Otherwise we shall say that ρ is degenerate.

LEMMA 8. *There exists at least one risk-minimizing distribution* ρ.

From Lemma 6 it follows that $r_1(\rho), \cdots, r_s(\rho)$ are continuous functions of ρ. Hence also $r(\rho)$ is continuous. Since the set of all possible distributions ρ is bounded and closed, there must be at least one distribution ρ for which $r(\rho)$ becomes a minimum.

LEMMA 9. *If* $\rho = (\rho_1, \cdots, \rho_s)$ *denotes a risk-minimizing distribution which is not degenerate then*

$$r_1(\rho) = r_2(\rho) = \cdots = r_s(\rho).$$

Let us assume that there are two integers i and j, for instance 1 and 2, such that $r_1(\rho) < r_2(\rho)$. We shall deduce a contradiction from this assumption. Let us consider two different distributions $\rho' = (\rho_1', \cdots, \rho_s')$ and $\rho'' = (\rho_1'', \cdots, \rho_s'')$ where $\rho_1'' > 0$. Hence at least one of the quantities

$$(\rho_1' - \rho_1''), \cdots, (\rho_s' - \rho_s'')$$

is unequal to zero. Since $\Sigma\rho_i' = \Sigma\rho_i'' = 1$, also at least one of the quantities

$$(\rho_2' - \rho_2''), \cdots, (\rho_s' - \rho_s'')$$

must be unequal to zero. On account of Lemma 7 we have

$$\sum_{i=1}^{s} [(\lambda - 1)\rho_i' + \lambda(\rho_i'' - \rho_i')][r_i(\rho'') - r_i(\rho')] < 0.$$

If we put $\lambda = \frac{\rho_1'}{\rho_1''}$ we get

$$\sum_{i=2}^{s} \left[\left(\frac{\rho_1'}{\rho_1''} - 1\right)\rho_i' + \frac{\rho_1'}{\rho_1''}(\rho_i'' - \rho_i')\right][r_i(\rho'') - r_i(\rho')] < 0.$$

Hence at least one of the quantities

$$r_2(\rho'') - r_2(\rho'), \cdots, r_s(\rho'') - r_s(\rho')$$

must be unequal to zero.

Since $\rho_1 > 0$, there exists a closed sphere S_ρ with center at ρ such that for any point ρ' of S_ρ $\rho_1' > 0$. Hence for any two different points ρ' and ρ'' of S_ρ at least one of the quantities

$$r_2(\rho'') - r_2(\rho'), \cdots, r_s(\rho'') - r_s(\rho')$$

is unequal to zero. Denote by $\bar{S}_\rho$ the projection of S_ρ on the $s - 1$ dimensional space given by $\rho_1 = 0$. Consider the transformation according to which the image of the point $\bar{\rho}' = (\rho_2', \cdots, \rho_s')$ of $\bar{S}_\rho$ is the point $\bar{q}(\bar{\rho}') = [r_2(\rho'), \cdots, r_s(\rho')]$. It is obvious that the images of two different points of $\bar{S}_\rho$ are different.

Since $r_i(\rho)$ $(i = 1, \cdots, s)$ is continuous, the transformation is continuous and therefore topological. Denote the image of $\bar{S}_\rho$ by $\bar{R}_\rho$. Since $\bar{\rho} = (\rho_2, \cdots, \rho_s)$ is an interior point of $\bar{S}_\rho$, according to the Brouwer-Jordan theorem[6] on domain invariance the image $\bar{q}(\bar{\rho}) = [r_2(\rho), \cdots, r_s(\rho)]$ of $\bar{\rho}$ must also be an interior point of $\bar{R}_\rho$. Hence for sufficiently small $\epsilon > 0$ the point

$$t(\epsilon) = [r_2(\rho) - \epsilon, \cdots, r_s(\rho) - \epsilon]$$

is contained in $\bar{R}_\rho$. Denote by $\bar{\rho}(\epsilon) = [\rho_2(\epsilon), \cdots, \rho_s(\epsilon)]$ the point of $\bar{S}_\rho$ whose image is $t(\epsilon)$. It is obvious that

$$\lim_{\epsilon=0} \bar{\rho}(\epsilon) = \bar{\rho} = (\rho_2, \cdots, \rho_s). \tag{18}$$

Consider the point $\rho(\epsilon)$ of S_ρ whose projection is $\bar{\rho}(\epsilon)$ that is to say $\rho(\epsilon)$ has the co-ordinates $1 - \Sigma\bar{\rho}_i(\epsilon), \bar{\rho}_2(\epsilon), \cdots, \bar{\rho}_s(\epsilon)$. From (18) it follows that also

$$\lim_{\epsilon=0} \rho(\epsilon) = \rho = (\rho_1, \rho_2, \cdots, \rho_s). \tag{19}$$

Since $r_1[\rho(\epsilon)], \cdots, r_s[\rho(\epsilon)]$ are continuous functions of ϵ and since $r_1(\rho) < r_2(\rho)$, for sufficiently small ϵ the maximum of the numbers

$$r_1[\rho(\epsilon)], r_2[\rho(\epsilon)] = r_2(\rho) - \epsilon, \cdots, r_s[\rho(\epsilon)] = r_s(\rho) - \epsilon$$

is certainly smaller than the maximum $r(\rho)$ of the numbers

$$r_1(\rho), \cdots, r_s(\rho),$$

in contradiction to our assumption that ρ is a risk minimizing distribution. Hence the assumption $r_1(\rho) < r_2(\rho)$ is proved to be an absurdity and Lemma 9 is proved.

In the previous arguments we have considered an arbitrary but fixed system of s parameter points $\theta_1, \cdots, \theta_s$ and all distributions ρ were related to these points. In the following arguments we shall vary the points $\theta_1, \cdots, \theta_s$ and therefore we shall have to state the parameter points to which the distribution ρ is related.

Let us consider a sequence $\{\theta_\nu\}$ $(\nu = 1, \cdots, \text{ad inf.})$ of parameter points which is dense in Ω. We say that a subset ω of Ω is dense in Ω if for each point θ of Ω any arbitrarily small open neighborhood of θ contains at least one point of ω. Since Ω is compact, a sequence $\{\theta_\nu\}$ which is dense in Ω certainly exists. Let us consider the first s points $\theta_1, \cdots, \theta_s$ of the sequence $\{\theta_\nu\}$. According to Lemma 8 there exists for any s a risk-minimizing distribution $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ related to $\theta_1, \cdots, \theta_s$.

Assumption 6. There exists a sequence $\{\theta_s\}$ $(s = 1, \cdots, \text{ad inf.})$ of parameter points which is dense in Ω and such that for almost any s[7] the risk-minimizing

[6] See for instance Alexandroff and Hopf, *Topologie*, Berlin 1935, p. 396.

[7] By "almost any s" we understand "for all s greater than a sufficiently large integer."

distribution $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ related to the first s points $\theta_1, \cdots, \theta_s$, is not degenerate.

LEMMA 10. *Denote by* $\{\theta_s\}$ $(s = 1, 2, \cdots, \text{ad inf.})$ *a sequence of parameter points for which the conditions of Assumption* 6 *are fulfilled. Denote by* $\rho(s) = [\rho_1(s), \cdots, \rho_s(s)]$ *the risk-minimizing distribution related to the first* s *points* $\theta_1, \cdots, \theta_s$. *Then there exists a non-negative constant* c *such that for any arbitrarily small positive* ϵ *the inequality*

$$c - \epsilon \leq \int_M W[\theta, \theta_{\rho(s)}(E)]p(E \mid \theta)\, dE \leq c + \epsilon$$

holds identically in θ *for almost every* s. *That is to say the risk function of the minimum risk estimate* $\theta_{\rho(s)}(E)$ *lies entirely between* $c - \epsilon$ *and* $c + \epsilon$ *for almost every* s.

Denote the risk function

$$\int_M W[\theta, \theta_{\rho(s)}(E)]p(E \mid \theta)\, dE$$

of the estimate $\theta_{\rho(s)}(E)$ by $r(\theta, s)$. First we shall prove that there exists a sequence $\{c_s\}$ $(s = 1, \cdots, \text{ad inf.})$ of non-negative numbers such that for every $\epsilon > 0$ the inequality

$$c_s - \epsilon \leq r(\theta, s) \leq c_s + \epsilon \tag{20}$$

holds for almost every s. In fact to any positive η a positive integer s_η can be given such that for any $s > s_\eta$ the points $\theta_1, \cdots, \theta_s$ are η-dense in Ω. That is to say every point θ of Ω lies in a sphere with radius η and center in one of the points $\theta_1, \cdots, \theta_s$. Since for sufficiently large s $\rho(s)$ is not degenerate, we have on account of Lemma 9 for sufficiently large s

$$r(\theta_1, s) = \cdots = r(\theta_s, s) = c_s. \tag{21}$$

Since for sufficiently large s $\theta_1, \cdots, \theta_s$ is η-dense in Ω, we get easily from Lemma 5 that (20) holds for any positive ϵ for almost every s.

In order to prove Lemma 10 we have only to show that $\lim_{s=\infty} c_s$ exists and is finite. First we see that for no estimate $\theta(E)$ can the corresponding risk function

$$r(\theta) = \int_M W[\theta, \theta(E)]p(E \mid \theta)\, dE$$

lie entirely below $r(\theta, s)$ that is to say

$$r(\theta) < r(\theta, s) \tag{22}$$

cannot hold for any θ. In fact if (22) were true for a certain estimate $\theta(E)$ then

$$\Sigma\rho_i(s)r(\theta_i) = \int_M \Sigma W[\theta_i, \theta(E)]\rho_i(s)p(E \mid \theta_i)\, dE < \Sigma\rho_i(s)r(\theta_i, s)$$

$$= \int_M \Sigma W[\theta_i, \theta_{\rho(s)}(E)]\rho_i(s)p(E \mid \theta_i)\, dE,$$

which is not possible since $\theta_{\rho(s)}(E)$ is a minimum risk estimate. Hence (22) cannot hold for any θ. From this fact follows easily that $\lim c_s$ exists and is finite. This proves Lemma 10.

LEMMA 11. *Denote $f(\theta)$ a distribution of θ and let $\theta_f(E)$ be the corresponding minimum risk estimate. If $\theta(E)$ denotes an arbitrary estimate then*

$$r(\theta) \equiv r_f(\theta)$$

if $\theta_f(E) \neq \theta(E)$ only in a set of measure 0, and

$$\int_\Omega r(\theta)\, df(\theta) > \int_\Omega r_f(\theta)\, df(\theta)$$

if $\theta_f(E) \neq \theta(E)$ in a set of positive measure. $r(\theta)$ denotes the risk function of $\theta(E)$ and $r_f(\theta)$ denotes the risk function of $\theta_f(E)$.

If $\theta_f(E) \neq \theta(E)$ only in a set of measure zero, then we have obviously $r(\theta) \equiv r_f(\theta)$. Consider the case that $\theta_f(E) \neq \theta(E)$ in a set M' of positive measure. According to Assumption 4 we have

$$\int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df(\theta) > \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df(\theta)$$

for any point E of M'. Since

$$\int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df(\theta) = \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df(\theta)$$

for any other point E of the sample space M, we get

$$\int_\Omega r(\theta)\, df = \int_M \int_\Omega W[\theta, \theta(E)]\, p(E \mid \theta)\, df\, dE$$

$$> \int_M \int_\Omega W[\theta, \theta_f(E)]\, p(E \mid \theta)\, df\, dE = \int_\Omega r_f(\theta)\, df.$$

Hence Lemma 11 is proved.

We are now able to prove some theorems about the best estimate $\bar{\theta}(E)$ relative to a given weight function. An estimate $\bar{\theta}(E)$ is a best estimate according to our definition 7, if the maximum of the risk function of $\bar{\theta}(E)$ is less than or equal to the maximum of the risk function of any other estimate $\theta(E)$ and if $\bar{\theta}(E)$ is an admissible estimate (that is to say there exists no estimate $\theta(E)$ such that the risk function $r(\theta)$ of $\theta(E)$ is not identical to the risk function $\bar{r}(\theta)$ of $\bar{\theta}(E)$ and in every point θ $\bar{r}(\theta) \geq r(\theta)$.

THEOREM 1. *If $\bar{\theta}(E)$ is a best estimate and if the Assumptions 1–6 are fulfilled then the risk function $\bar{r}(\theta)$ of $\bar{\theta}(E)$ is constant, that is to say*

$$\bar{r}(\theta) \equiv c.$$

According to Assumption 6 there exists a sequence $\{\theta_s\}$ ($s = 1, \cdots$, ad inf.) of parameter points such that $\{\theta_s\}$ is dense in Ω and for almost every s the risk-

minimizing distribution $\rho(s)$ related to $\theta_1, \cdots, \theta_s$ is not degenerate. On account of Lemma 10 there exists a non-negative constant c such that for any $\epsilon > 0$ the inequality

$$c - \epsilon \leq r(\theta, s) \leq c + \epsilon \tag{23}$$

holds for almost every s. $r(\theta, s)$ denotes the risk function of the estimate $\theta_{\rho(s)}(E)$. According to Lemma 2 there exists a subsequence $\{s_n\}$ $(n = 1, \cdots,$ ad inf.) of integers such that the sequence $\{\rho(s_n)\}$ of distributions converges towards a distribution $f(\theta)$. From Lemma 6 it follows that

$$\lim_{n=\infty} r(\theta, s_n) = r_f(\theta)$$

where $r_f(\theta)$ denotes the risk function of the minimum risk estimate $\theta_f(E)$. On account of (23) we have

$$r_f(\theta) \equiv c.$$

From Lemma 11 it follows that for any other estimate $\theta(E)$ either

$$r(\theta) \equiv r_f(\theta) \equiv c$$

or

$$\int_\Omega r(\theta)\, df > \int_\Omega r_f(\theta)\, df,$$

where $r(\theta)$ denotes the risk function of $\theta(E)$. In the latter case there exists at least one point θ for which $r(\theta) > r_f(\theta)$. Hence $\theta_f(E)$ is a best estimate. If $\bar{\theta}(E)$ is also a best estimate, we get on account of Lemma 11 that $\bar{\theta}(E)$ can differ from $\theta_f(E)$ only in a set of measure 0 and the risk function of $\bar{\theta}(E)$ is identically equal to c. Hence we have proved Theorem 1 and also the following Theorems 2–3:

THEOREM 2. *If the Assumptions* 1–6 *are fulfilled there exists a distribution* $f(\theta)$ *of* θ *such that the corresponding minimum risk estimate* $\theta_f(E)$ *is a best estimate.*

THEOREM 3. *If Assumptions* 1–6 *are fulfilled and* $\bar{\theta}(E)$, $\theta^*(E)$ *are best estimates, then* $\bar{\theta}(E) = \theta^*(E)$ *almost everywhere and the corresponding risk functions are identically equal.*

Now we shall prove (without making the Assumptions 1–6)

THEOREM 4. *If* $W(\theta, \bar{\theta})$ *and* $p(E \mid \theta)$ *are continuous and* Ω *is compact, and if* $f(\theta)$ *denotes a distribution of* θ *such that any open set has a positive probability, then the minimum risk estimate* $\theta_f(E)$ *is a best estimate if its risk function* $r_f(\theta)$ *is identically equal to a constant.*

Let $r_f(\theta)$ be identically equal to c and consider an arbitrary estimate $\theta(E)$. Since $W(\theta, \bar{\theta})$ and $p(E \mid \theta)$ are continuous and Ω is compact, the risk function $r(\theta)$ of $\theta(E)$ is a continuous function of θ. Since $\theta_f(E)$ is a minimum risk estimate we have

$$\int_\Omega r(\theta)\, df \geq \int_\Omega r_f(\theta)\, df = c. \tag{24}$$

In order to prove Theorem 4, we have to show that either

$$r(\theta) \equiv c \tag{25}$$

or there exists a point θ' such that

$$r(\theta') > c. \tag{26}$$

If (25) does not hold there exists a point θ^* such that $r(\theta^*) \neq c$. If $r(\theta^*) > c$ our statement is proved. Consider the case $r(\theta^*) < c$. On account of the continuity of $r(\theta)$ there exists a positive δ and an open neighborhood U of θ^* such that

$$r(\theta) < c - \delta$$

for every θ in U. Since $\int_U df$ is assumed to be positive, the inequality (24) can hold only if there exists at least a point θ' for which $r(\theta') > c$. This proves Theorem 4.

9. **Determination of the best estimate $\bar{\theta}(E)$ for a certain class of distributions $p(E \mid \theta)$.** In this paragraph we shall prove two theorems which enable us to calculate very easily the best estimate $\bar{\theta}(E)$ for a certain special but important class of distributions.

The risk function of an estimate $\bar{\theta}(E)$ is given by

$$r(\theta) = \int_M W[\theta, \bar{\theta}(E)]\, p(E \mid \theta)\, dE,$$

where the integration is to be taken over the whole sample space M. We consider the integral equation

$$\int_M W[\theta, \bar{\theta}(E)]\, p(E \mid \theta)\, dE \equiv c, \tag{27}$$

where c denotes an arbitrary constant. If we can find an estimate $\bar{\theta}(E)$ which satisfies (27) for a certain c and which is an admissible estimate relative to the weight function considered, then $\bar{\theta}(E)$ is certainly a best estimate. If Assumptions 1–6 are fulfilled, an admissible estimate satisfying (27) certainly exists. As we shall see, a best estimate can very easily be determined by the above procedure if the conditions in the following theorem 5 are fulfilled.

THEOREM 5. *Let us assume that the following conditions are fulfilled:*

I. *The parameter space Ω is one dimensional and θ can take any real value from $-\infty$ to $+\infty$.*

II. *The probability density $p(E \mid \theta)$ depends only on the differences $x_1 - \theta, \cdots, x_n - \theta$, that is to say $p(E \mid \theta) = p(x_1 - \theta, \cdots, x_n - \theta)$, where $x_1, \cdots, x_n$ denote the co-ordinates of E.*

III. *The value of the weight function depends only on the difference $u = \theta - \bar{\theta}$ and is uniformly continuous in u.*

IV. *For any value $\bar{\theta}$ and for any sample point E the integral*

$$\psi(\bar{\theta}, E) = \int_{-\infty}^{+\infty} W(\theta - \bar{\theta}) p(E \mid \theta)\, d\theta \tag{28}$$

has a finite value.

V. *For every E there exists a finite value $\theta'(E)$ such that $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \theta'(E)$.*

Then there exists an estimate $\bar{\theta}(E)$ such that for any E, $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \bar{\theta}(E)$ and $\bar{\theta}(E'') - \bar{\theta}(E') = \lambda$ for any $E' = (x_1', \cdots, x_n')$ and $E'' = (x_1'', \cdots, x_n'')$ for which $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. An estimate with these properties is a best estimate.

Let us consider two sample points $E' = (x_1', \cdots, x_n')$ and $E'' = (x_1'', \cdots, x_n'')$ such that $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. From the conditions II and III follows that if $\psi(\bar{\theta}, E')$ becomes a minimum for $\bar{\theta} = \theta_1$, then $\psi(\bar{\theta}, E'')$ becomes a minimum for $\bar{\theta} = \theta_2 = \theta_1 + \lambda$. Hence there exists an estimate $\bar{\theta}(E) = \bar{\theta}(x_1, \cdots, x_n)$ such that for any E, $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \bar{\theta}(E)$ and $\bar{\theta}(E'') - \bar{\theta}(E') = \lambda$ if $x_1'' - x_1' = \cdots = x_n'' - x_n' = \lambda$. We shall show that such an estimate $\bar{\theta}(E)$ is a best estimate. First we shall show that the risk function

$$r(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta - \bar{\theta}(E)]\, p(x_1 - \theta, \cdots, x_n - \theta')\, dx_1 \cdots dx_n$$

is constant. Let us consider two arbitrary parameter values θ' and θ''. Then we have

$$r(\theta') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta' - \bar{\theta}(E)]\, p(x_1 - \theta', \cdots, x_n - \theta')\, dx_1 \cdots dx_n,$$

$$r(\theta'') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta'' - \bar{\theta}(E)]\, p(x_1 - \theta'', \cdots, x_n - \theta'')\, dx_1 \cdots dx_n.$$

Making in the second integral the transformation

$$y_1 = x_1 - (\theta'' - \theta'), \cdots, y_n = x_n - (\theta'' - \theta'),$$

we get

$$\begin{aligned} r(\theta'') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} & W\{\theta'' - \bar{\theta}[y_1 + (\theta'' - \theta'), \cdots, y_n \\ & + (\theta'' - \theta')]\}\, p(y_1 - \theta', \cdots, y_n - \theta')\, dy_1 \cdots dy_n \\ = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} & W[\theta' - \bar{\theta}(y_1, \cdots, y_n)]\, p(y_1 - \theta', \cdots, y_n - \theta')\, dy_1 \cdots dy_n. \end{aligned}$$

Hence $r(\theta') = r(\theta'')$ and our statement that $r(\theta)$ is constant is proved. In order to prove Theorem 5, we have only to show that $\bar{\theta}(E)$ is an admissible estimate. For this purpose let us consider an arbitrary estimate $\theta^*(E)$ and

denote the corresponding risk function by $r^*(\theta)$. Since $\bar{\theta}(E)$ minimizes the integral (28), we have

$$\psi[\theta^*(E), E] \geq \psi[\bar{\theta}(E), E] \tag{29}$$

for all sample points E. Let us consider the integral

$$I = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \{W[\theta - \bar{\theta}(E)] - W[\theta - \theta^*(E)]\, p(E \mid \theta)\, d\theta\, dx_1 \cdots dx_n. \tag{30}$$

Integrating (30) with respect to θ we get

$$I = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \{\psi[\bar{\theta}(E), E] - \psi[\theta^*(E) - E]\}\, dx_1 \cdots dx_n. \tag{31}$$

Integrating (30) with respect to E, we get

$$I = \int_{-\infty}^{+\infty} [r(\theta) - r^*(\theta)]\, d\theta. \tag{32}$$

On account of (29) and (31) we have $I \leq 0$, hence

$$\int_{-\infty}^{+\infty} [r(\theta) - r^*(\theta)]\, d\theta \leq 0. \tag{33}$$

From (33) it follows that if $r^*(\theta) \leq r(\theta)$ for every θ then $r^*(\theta) < r(\theta)$ can hold only for the points of a set of measure zero. In case $r^*(\theta)$ is continuous, this means that $r^*(\theta) \equiv r(\theta)$. Hence if $r^*(\theta)$ is continuous, then either $r^*(\theta) \equiv r(\theta)$ or there exists at least one point θ' such that $r^*(\theta') > r(\theta)$. The risk function $r^*(\theta)$ is continuous if the estimate $\theta^*(E)$ is uniformly continuous in the whole sample space. In fact, we have

$$r^*(\theta + t) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta + t - \theta^*(E)]\, p(x_1 - \theta - t, \cdots, x_n - \theta - t)\, dx_1 \cdots dx_n.$$

Making the transformation

$$y_i = x_i - t \qquad (i = 1, \cdots, n)$$

we get

$$r^*(\theta + t) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} W[\theta + t - \theta^*(y_1 + t, \cdots, y_n + t)\, p(y_1 - \theta, \cdots, y_n - \theta)\, dy_1 \cdots dy_n.$$

Since $W(u)$ and $\theta^*(E)$ are uniformly continuous, from the latter equation we get easily

$$\lim_{t=0} r^*(\theta + t) = r^*(\theta)$$

that is to say $r^*(\theta)$ is continuous. Considering only continuous estimates the admissibility of $\bar{\theta}(E)$, and therefore also Theorem 5, is proved. If $\theta^*(E)$ is not

uniformly continuous we have only proved that if $r^*(\theta) \leq r(\theta)$ for every θ, then $r^*(\theta) < r(\theta)$ can hold only in a set of measure zero. I should like to mention without proof that even if $\theta^*(E)$ is not continuous, $r^*(\theta) \leq r(\theta)$ implies $r^*(\theta) \equiv r(\theta)$.

An estimate $\hat{\theta}(E)$ is called a maximum likelihood estimate if for any fixed E $p(E \mid \theta)$ becomes a maximum with respect to θ for $\theta = \hat{\theta}(E)$.

THEOREM 6. *Consider the following conditions*:

VI. *There exists exactly one maximum likelihood estimate* $\hat{\theta}(E)$ *with the following properties*:

a) *For any* E $p(E \mid \theta)$ *is non-decreasing with increasing* θ *for* $\theta < \hat{\theta}(E)$ *and non-increasing with increasing* θ *for* $\theta > \hat{\theta}(E)$.

b) *For any* E $p(E \mid \theta)$ *is a symmetric function of* θ *about* $\hat{\theta}(E)$ *that is to say, for for any real value* λ $p[E \mid \hat{\theta}(E) - \lambda] = p[E \mid \hat{\theta}(E) + \lambda]$.

VII. *The value of the weight function depends only on the absolute value of the difference* $u = \theta - \bar{\theta}$ *and* $\frac{dw(u)}{du}$ *exists, is uniformly continuous and* > 0 *for* $u > 0$.

If the conditions I–V of Theorem 5 *and the above condition VII are fulfilled, and if* $\hat{\theta}(E)$ *is a maximum likelihood estimate satisfying VI, then* $\hat{\theta}(E)$ *is a best estimate.*

Assume that the conditions I–V and VII are satisfied and that $\hat{\theta}(E)$ is a maximum likelihood estimate satisfying VI. It is obvious that $\hat{\theta}(E'') - \hat{\theta}(E') = \lambda$ for $E' = (x_1, \cdots, x_n)$ and $E'' = (x_1 + \lambda, \cdots, x_n + \lambda)$. In order to prove Theorem 6, we have, according to Theorem 5, only to show that the integral in (28)

$$\psi(\bar{\theta}, E) = \int_{-\infty}^{+\infty} W(\theta - \bar{\theta}) p(E \mid \theta)\, d\theta$$

becomes a minimum for $\bar{\theta} = \hat{\theta}(E)$. Denote $\theta - \bar{\theta}$ by u. Since $\frac{dW(u)}{du}$ is uniformly continuous, we have

$$\frac{\partial \psi(\bar{\theta}, E)}{\partial \bar{\theta}} = \int_{-\infty}^{+\infty} -\left[\frac{dW(u)}{du}\right] p(E \mid \theta)\, d\theta.$$

Since $\frac{dW(u)}{du} = -\frac{dW(-u)}{du}$ we have

$$\frac{\partial \psi(\bar{\theta}, E)}{\partial \bar{\theta}} = \int_0^{\infty} \left[\frac{dW(u)}{du}\right] [p(E \mid \bar{\theta} - u) - p(E \mid \bar{\theta} + u)]\, du. \tag{34}$$

From condition VI it follows easily that for any fixed E and $\bar{\theta}$ the function of u $(0 \leq u \leq \infty)$

$$p(E \mid \bar{\theta} - u) - p(E \mid \bar{\theta} + u)$$

does not change its sign and if $\bar{\theta} \neq \hat{\theta}(E)$ there exists an interval J such that the above expression is unequal to zero for every point u of J. Hence on account of $\frac{dW(u)}{du} > 0$ for $u > 0$, the integral in (34) vanishes only for $\bar{\theta} = \hat{\theta}(E)$. Since according to the condition V there exists a finite value θ' such that $\psi(\bar{\theta}, E)$ becomes a minimum for $\bar{\theta} = \theta'$, θ' must be equal to $\hat{\theta}(E)$. This proves Theorem 6.

The condition VI is seldom exactly fulfilled. But for large n, in the great majority of practical cases, VI will be fulfilled with good approximation and the best estimate approaches the maximum likelihood estimate with increasing n.

10. **Two examples.** As a first example we consider a normal distribution with the variance 1. The mean value θ is unknown and we have to estimate it by means of a sample $E = (x_1, \cdots, x_n)$. In this case

$$p(E \mid \theta) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\Sigma(x_i-\theta)^2}.$$

It is obvious that for a very broad class of weight functions the conditions I–V of Theorem 5 are fulfilled. The maximum likelihood estimate $\hat{\theta}(x_1, \cdots, x_n) = \frac{x_1 + \cdots + x_n}{n}$ satisfies the condition VI of Theorem 6. Hence if the weight function satisfies also the condition VII, then the best estimate of θ is the maximum likelihood estimate $\hat{\theta}(x_1, \cdots, x_n) = \frac{x_1 + \cdots + x_n}{n}$.

Let us now consider a weight function defined as follows:

$$W(\theta, \bar{\theta}) = 2(\bar{\theta} - \theta) \quad \text{if} \quad \bar{\theta} \geq \theta$$

and

$$W(\theta, \bar{\theta}) = \theta - \bar{\theta} \qquad \text{if} \quad \bar{\theta} < \theta.$$

Since for this weight function, the conditions I–V satisfied, according to Theorem 5 the best estimate of θ is the value $\bar{\theta}$ for which the integral

$$\int_{-\infty}^{+\infty} W(\theta, \bar{\theta}) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta = \int_{-\infty}^{\bar{\theta}} 2(\bar{\theta} - \theta) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta + \int_{\bar{\theta}}^{\infty} (\theta - \bar{\theta}) e^{-\frac{1}{2}\Sigma(x_i-\theta)^2} d\theta$$

becomes a minimum. As an easy calculation shows, the estimate obtained in this way is not the arithmetic mean.

As a second example we consider the family of variates $X(\theta)$ with the probability density $f(x, \theta)$ defined as follows:

$$f(x, \theta) = 1 \quad \text{if} \quad \theta - \tfrac{1}{2} \leq x \leq \theta + \tfrac{1}{2}$$

and

$$f(x, \theta) = 0 \text{ for all other values of } x.$$

If $E = (x_1, \cdots, x_n)$ denotes a sample point where x_1 denotes the smallest and x_n denotes the greatest value in the sample, then

$$p(E \mid \theta) = \prod_{i=1}^{n} f(x_i, \theta) = 1 \quad \text{if} \quad x_n - \tfrac{1}{2} \leq \theta \leq x_1 + \tfrac{1}{2}$$

and

$$p(E \mid \theta) = 0 \quad \text{for all other values of } \theta.$$

The classical method of maximum likelihood cannot be applied here, since $p(E \mid \theta)$ is maximum for every value θ for which $x_n - \frac{1}{2} \leq \theta \leq x_1 + \frac{1}{2}$. It is obvious that for a broad class of weight functions the conditions I–V are satisfied. The estimate $\bar{\theta}(E) = \dfrac{x_1 + x_n}{2}$, where x_1 denotes the smallest and x_n the greatest value in the sample, satisfies the condition VI. Hence if the weight function satisfies also the condition VII, the best estimate of θ is given by $\bar{\theta}(E) = \dfrac{x_1 + x_n}{2}$.

Let us now calculate the best estimate of θ if the weight function is given as follows:

$$W(\theta, \bar{\theta}) = \theta - \bar{\theta} \qquad \text{if} \quad \bar{\theta} \leq \theta$$

and

$$W(\theta, \bar{\theta}) = 2(\bar{\theta} - \theta) \quad \text{if} \quad \bar{\theta} > \theta.$$

In this case the conditions I–V are satisfied but not the condition VII. We have to calculate the integral $\psi(\bar{\theta}, E)$ given in (28), which reduces in this case to

$$\psi(\bar{\theta}, E) = \int_{x_n - \frac{1}{2}}^{x_1 + \frac{1}{2}} W(\theta, \bar{\theta})\, d\theta = \int_{x_n - \frac{1}{2}}^{\bar{\theta}} 2(\bar{\theta} - \theta)\, d\theta + \int_{\bar{\theta}}^{x_1 + \frac{1}{2}} (\theta - \bar{\theta})\, d\theta$$

$$= 1.5\bar{\theta}^2 - [(x_1 + \tfrac{1}{2}) + 2(x_n - \tfrac{1}{2})]\bar{\theta} + \tfrac{1}{2}(x_1 + \tfrac{1}{2})^2 + (x_n - \tfrac{1}{2})^2.$$

This expression becomes a minimum for

$$\bar{\theta} = \frac{x_1 + 2x_n - \frac{1}{2}}{3}.$$

Hence the best estimate of θ is given by this expression.

11. **Miscellaneous remarks.** Assumptions 1–6 of paragraph 8 are sufficient but not necessary for the proof of the Theorems 1–3 (Theorems 4–6 have been deduced without Assumptions 1–6). They can be weakened in many respects. The assumption that the parameter space is bounded can be dropped if we impose certain conditions on the weight function $W(\theta, \bar{\theta})$ and the probability density $p(E \mid \theta)$. It is certainly not necessary to assume that $W(\theta, \bar{\theta})$ and $p(E \mid \theta)$ are everywhere continuous. It is however doubtful whether Theorems 1–3 remain valid in the form in which they are stated, if we admit discon-

tinuities in a set of measure zero without imposing any other restrictions. Also Assumptions 4–6 can in all probability be essentially weakened.

I should like to mention that Assumption 4 is not as restrictive as it would appear. Let us make this clear in the case that the parameter space is a one-dimensional interval $[a, b]$. If we assume that $W(\theta, \bar{\theta})$ is a polynomial of the second degree in $\bar{\theta}$ and the coefficient of $\bar{\theta}^2$ is positive for every θ, and if $p(E \mid \theta) > 0$ for every E and θ, the Assumption 4 can easily be proved. In fact,

$$\psi(\bar{\theta}, E) = \int_a^b W(\theta, \bar{\theta}) p(E \mid \theta)\, df(\theta) = A(E) + B(E)\bar{\theta} + C(E)\bar{\theta}^2.$$

Since the coefficient of $\bar{\theta}^2$ in $W(\theta, \bar{\theta})$ is positive and since $p(E \mid \theta) > 0$ for every E and θ, $C(E) > 0$ for every E and for any arbitrary distribution $f(\theta)$. From this fact follows easily that for every E there exists a value $\bar{\theta}(E)$ in the interval $[a, b]$ such that

$$\psi[\bar{\theta}(E), E] < \psi(\bar{\theta}, E)$$

for every $\bar{\theta}$ contained in $[a, b]$ and unequal to $\bar{\theta}(E)$. Hence Assumption 4 is proved.

Let us consider a system S of subsets of the parameter space Ω and the corresponding system H_S of hypotheses. The weight function $W(\theta, \omega)$ is defined for all points θ of Ω and for all elements ω of S and expresses the weight of the error committed by accepting H_ω when θ is true. If θ is an element of ω then $W(\theta, \omega)$ is of course equal to zero. Let us assume that $W(\theta, \omega)$ has the special form: $W(\theta, \omega) = 1$ if θ is not contained in ω, and $W(\theta, \omega) = 0$ if θ is an element of ω. It is obvious that in this case for any θ the value of the risk function $r(\theta)$ is equal to the probability of accepting a false hypothesis if θ is the true parameter point. Because of this fact the theory developed here has close relation to the theory of confidence intervals. Let us first make this clear for the case when the parameter space is one dimensional, that is to say θ is a real number.

In the theory of confidence intervals we estimate the unknown parameter θ by an interval $I(E)$ extending from $\theta_1(E)$ to $\theta_2(E)$ where $\theta_1(E)$ and $\theta_2(E)$ are certain functions of the sample point E. The interval $I(E)$ is defined in such a way that the following probability statement holds: If we perform an experiment, the probability that we shall obtain a sample point E such that $I(E)$ will cover the true parameter point θ, is equal to a given constant α (called confidence coefficient) and is independent of the value of θ. Let us consider a certain example of such an inference with the confidence coefficient α and denote by $I(E)$ the interval corresponding to E. We define a system S of intervals as follows: An interval I is an element of S if and only if there exists a sample point E for which $I(E) = I$. Consider the corresponding system H_S

of hypotheses and the weight function $W(\theta, I)$ defined for all values θ and all elements I of S as follows:

$$W(\theta, I) = 0 \quad \text{if} \quad \theta \text{ is a point of } I$$

$$W(\theta, I) = 1 \quad \text{if} \quad \theta \text{ is not contained in } I.$$

Denote by M_S a best system of regions of acceptance relative to the weight function defined above. Denote by $I'(E)$ the element of S which we accept according to M_S if E is the sample point. On account of the special form of the weight function, the risk is obviously equal to the probability of accepting a false interval. From the definition of the best system of regions it follows that for any θ the probability that $I'(E)$ will cover θ is greater than or equal to α. If the risk function is constant, that is to say, if the probability that $I'(E)$ will cover the true parameter point θ is independent of the value of θ, then the intervals $I'(E)$ are confidence intervals corresponding to a confidence coefficient $\alpha' \geq \alpha$.

Similar observations can be made if the parameter space is k-dimensional ($k > 1$) that is to say, θ is a system of k numbers $\theta^{(1)}, \cdots, \theta^{(k)}$. An important case is that when we have to estimate only one of the components, say $\theta^{(1)}$, by an interval. As the investigations of W. Feller[8] have shown, confidence intervals in such cases do not exist always. That is to say, it is not always possible to determine $I(E)$ such that the probability that $I(E)$ will cover $\theta^{(1)}$ is equal to a given constant α independently of the values of $\theta^{(1)}, \cdots, \theta^{(k)}$. It is of great interest to know under what conditions confidence intervals exist. I should like to mention that a further development of the theory given in paragraph 8 may contribute much to the solution of this problem. In order to make this clear, let us consider a system S_1 of one-dimensional intervals. To each element I of S_1 let there correspond the subset ω of the k-dimensional parameter space Ω consisting of all points $\theta = (\theta^{(1)}, \cdots, \theta^{(k)})$ for which $\theta^{(1)}$ lies in I. Consider the system S of subsets ω of Ω corresponding to all elements of S_1 and the system H_S of hypotheses corresponding to S. The weight function is to be chosen as follows: $W(\theta, \omega) = 1$ if θ is not an element of ω and $W(\theta, \omega) = 0$ if θ is an element of ω. Consider a best system M_S of regions of acceptance and the corresponding risk function $r(\theta)$. On account of the special definition of $W(\theta, \omega)$, $r(\theta)$ is equal to the probability of accepting a false hypothesis if θ is the true parameter point. If the risk function $r(\theta)$ is identically equal to a constant α, we have confidence intervals corresponding to the confidence coefficient α. In order to see under what conditions the risk function is constant, we have to consider an equivalent problem (see paragraph 7) where the system of hypotheses is the system of all simple hypotheses and the weight function $W(\theta, \bar{\theta})$

[8] W. Feller, "Note on Regions Similar to the Sample Space," *Statistical Research Memoirs*, Vol. II, 1938.

is given according to formula (4). If $W(\theta, \bar{\theta})$ satisfies Assumptions 1–6, the risk function of the best estimate is constant. As we have mentioned, Assumptions 1–6 can be weakened. In order to get valuable results concerning the problem of the existence of confidence intervals, we have to weaken especially Assumption 2. In fact $W(\theta, \omega)$ takes only the values 1 and 0 and therefore $W(\theta, \bar{\theta})$ cannot be continuous.

Finally I should like to mention that the most stringent test as defined by Robert W. B. Jackson[9] is contained as special case in our general definition of the best system of regions of acceptance. Jackson considers a discontinuous parameter space Ω. Consider the problem of testing the hypothesis $\theta = \theta_0$ where θ_0 denotes a point of Ω. According to Jackson's definition we have the most stringent test if the critical region w_0 satisfies the condition: the maximum of the numbers A and B

$$A = P(E \in w \mid \theta_0), \quad B = \text{least upper bound of } P(E \in \bar{w} \mid \theta) \text{ formed for all } \theta \neq \theta_0,$$

becomes a minimum for $w = w_0$. $\bar{w}$ denotes the region complementary to w. It is easy to see that Jackson's definition of the most stringent test coincides with our definition of the best system of regions of acceptance in the following special case:

1) Ω is discontinuous 2) S consists only of two elements.

3) The weight function $W(\theta, \omega)$ is equal to 1 if θ is not contained in ω.

COLUMBIA UNIVERSITY.

[9] Robert W. Jackson, "Testing Statistical Hypotheses," *Statistical Research Memoirs*, Vol. I, 1936.