A UNIFIED THEORY OF ESTIMATION, I[1]

By Allan Birnbaum

Institute of Mathematical Sciences, New York University

0. Introduction and summary. This paper extends and unifies some previous formulations and theories of estimation for one-parameter problems. The basic criterion used is admissibility of a point estimator, defined with reference to its full distribution rather than special loss functions such as squared error. Theoretical methods of characterizing admissible estimators are given, and practical computational methods for their use are illustrated.

Point, confidence limit, and confidence interval estimation are included in a single theoretical formulation, and incorporated into estimators of an "omnibus" form called "confidence curves." The usefulness of the latter for some applications as well as theoretical purposes is illustrated.

Fisher's maximum likelihood principle of estimation is generalized, given exact (non-asymptotic) justification, and unified with the theory of tests and confidence regions of Neyman and Pearson. Relations between exact and asymptotic results are discussed.

Further developments, including multiparameter and nuisance parameter problems, problems of choice among admissible estimators, formal and informal criteria for optimality, and related problems in the foundations of statistical inference, will be presented subsequently.

1. A broad formulation of the problem of point estimation. We consider problems of estimation with reference to a specified experiment E, leaving aside here questions of experimental design including those of choice of a sample size or a sequential sampling rule; some definite sampling rule, possibly sequential, is assumed specified as part of E. Let $S = \{x\}$ denote the sample space of possible outcomes x of the experiment. Let $f(x, \theta)$ denote one of the elementary probability functions on S which are specified as possibly true. Let $\Omega = \{\theta\}$ denote the specified parameter space. For each θ in Ω and for each subset of A of S, the probability that E yields an outcome x in A is given by

$$\text{Prob}\,\{X \,\varepsilon\, A \mid \theta\} = \int_A f(x, \theta)\,d\mu(x),$$

where μ is a specified σ-finite measure on S. (We assume tacitly here and below that consideration is appropriately restricted to measurable sets and functions only.)

If $\gamma = \gamma(\theta)$ is any function defined on Ω(e.g., $\gamma(\theta) \equiv \theta$ or $\gamma(\theta) \equiv \theta^2$), with range Γ, a point estimator of γ is any measurable function $g = g(x)$ taking values in Γ (or in $\bar{\Gamma}$, its closure, if, for example, Γ is an open interval). The problem of

Received May 29, 1959; revised October 10, 1960.

[1] Research supported in part by the Office of Naval Research, Contracts No. Nonr-285 (38) and Nonr-266 (33).

choosing a good estimator, that is an estimator which tends to take values close to the true unknown value of γ, has been formulated mathematically in various ways. Most formulations achieve mathematical definiteness by introducing criteria of closeness which appear somewhat arbitrary from some standpoints of application and undesirably schematic as expressions of the intuitive notion of closeness.

If Ω is given no specific (parametric) structure, then the latter features can be fully avoided only by a very broad formulation which specifies only that if γ is true, then an exactly correct estimate ($g = \gamma$) is closer than any incorrect estimate ($g \neq \gamma$). If Ω is finite, $\Omega = \{\theta_1, \cdots \theta_k\}$, and $\gamma(\theta) = \theta$, this leads to the formulation of Lindley [1] in which estimators are compared only on the basis of their error probabilities

$$p_{ij} = \text{Prob}\,\{\theta^*(X) = \theta_i \mid \theta_j\}, \quad i, j, = 1, \cdots k, i \neq j,$$

where $\theta^*(x)$ is any estimator of θ. This formulation has no very useful extension to typical estimation problems in which, for example, Ω is an interval, and in which the event $\theta^*(X) = \theta$ exactly has typically negligible probability and little interest.

The case in which Ω is any set of real numbers, for example an interval, and $\gamma(\theta) \equiv \theta$, may be termed the central problem of theory of point-estimation, although very important generalizations of this problem have been treated extensively. For this problem, closeness of θ^* to θ has been specified by the introduction of specific loss functions: The absolute error criterion, $|\theta^* - \theta|$, was introduced by Laplace. Gauss replaced this by the squared error criterion $(\theta^* - \theta)^2$, which proved mathematically much more tractable and provided a definite formulation of the problem that seemed equally reasonable.

Each such definite specification of closeness can be criticized as somewhat arbitrary, except in a context where one postulates the reality of the indicated costs of errors of each possible kind. To avoid such features it is evidently necessary and sufficient to adopt the following weak specification of closeness: If $\theta_1^* < \theta_2^* \leqq \theta$ or if $\theta \leqq \theta_2^* < \theta_1^*$, the estimate θ_2^* is called closer than θ_1^* to θ; if $\theta_1^* < \theta < \theta_2^*$, no comparison as to closeness is to be made. (The latter point was put forth by Galileo in an exchange which retains interest in connection with questions of formulation of estimation problems, particularly distinctions between errors of inference and economic valuations, and the historical origins of unbiasedness criteria. Cf. [2].)

This specification of closeness leads to comparisons between estimators on the basis of all of their probabilities of errors of over-estimation and under-estimation by various amounts $d = |\theta^* - \theta|$:

$$(1.1) \quad a(u, \theta, \theta^*) = \begin{cases} F(u, \theta, \theta^*) \equiv \text{Prob}\ \{\theta^*(X) \leqq u \mid \theta\} \text{ for } u < \theta, \\ 1 - F(u - 0, \theta, \theta^*) \equiv \text{Prob}\ \{\theta^*(X) \geqq u \mid \theta\} \text{ for } u > \theta. \end{cases}$$

That is, estimators are compared only on the basis of their complete cumulative distribution functions (c.d.f's) $F(u, \theta, \theta^*)$ for each $\theta \varepsilon \Omega$, rather than on the basis of certain "summaries" (functionals) of these c.d.f's such as mean squared error. The function $a(u, \theta, \theta^*)$, defined for any estimator $\theta^*(x)$ at each $\theta \varepsilon \Omega$ and each $u \neq \theta$, will be called the *risk curve* of θ^* at θ (or, more precisely, of $\theta^*(\cdot)$ at θ).

The family of distributions under consideration may be viewed as having a parametric structure only in the sense that it is ordered by the labeling of each function $f(x, \theta)$ of x by a different real number θ. From this standpoint, the problem of estimating θ is equivalent to that of estimating $\gamma = \gamma(\theta)$ if the latter is any specified strictly monotone function. The formulation adopted above is clearly unaffected by (invariant under) such transformations of the parameter space ($\Omega \rightarrow \gamma(\Omega) \equiv \Gamma$), as contrasted with most other formulations referred to above.

A theory of point estimation based on this broad formulation seems appropriate for typical problems of inference occurring in empirical research, since various kinds of errors of inference and their probabilities admit simple direct interpretations, whereas other formulations introduce specifications akin to costs of various errors which seem somewhat hypothetical or arbitrary in such situations. The present theory also has theoretical and technical relevance for estimation theories based on more restrictive formulations, since it includes such theories in a formal sense that will be elaborated in a following section.

2. Admissible point estimators. An estimator $\theta^*(x)$ of θ is naturally considered a good one if its error-probabilities are suitably small, i.e., if (the ordinates of) its risk curves $a(u, \theta, \theta^*)$, for each $\theta \varepsilon \Omega$ and each $u \neq \theta$, are suitably small. This leads to a natural partial ordering of estimators, under which some but not all pairs of estimators can be compared. As a basis for systematic evaluations and comparisons of estimators we require the following

DEFINITIONS: For a given estimation problem, an estimator θ^* is called *at least as good as* an estimator θ^{**} if $a(u, \theta, \theta^*) \leqq a(u, \theta, \theta^{**})$ for all $\theta \varepsilon \Omega$ and all $u \neq \theta$. If θ^* and θ^{**} are each at least as good as the other, then $a(u, \theta, \theta^*) \equiv a(u, \theta, \theta^{**})$, and the estimators are called *equivalent*. If neither of θ^*, θ^{**} is at least as good as the other, the two estimators are called *not comparable*. If θ^* is at least as good as θ^{**} and if $a(u, \theta, \theta^*) < a(u, \theta, \theta^{**})$ for some $\theta \varepsilon \Omega$ and some $u \neq \theta$, θ^* is called *better than* θ^{**}. As estimator θ^* is called *admissible* if no other estimator is better than θ^*. The class of admissible estimators is called *the admissible class*. A class of estimators is called *complete* if, for each estimator outside the class, there is a better one in the class. The *minimal* (smallest) *complete class*, if one exists, coincides with the admissible class. A class of estimators is called *essentially complete* if, for each estimator not in the class, there is one at least as good in the class. A *minimal essentially complete class*, if one exists, is a subclass of the admissible class.

The above definition of admissibility was included in a list of criteria for point estimators by Savage [3] (pp. 224–225), but it has not previously been used systematically.

The criterion of closeness of estimators introduced by Pitman [4] is based not on the full c.d.f's of individual estimators, but on the joint distribution of absolute errors for each pair of estimators; this criterion does not give a partial ordering of estimators, and does not lend itself to our present purposes.

For the probabilities of under-estimation and over-estimation, we define also

$$\begin{aligned} a(\theta-, \theta, \theta^*) &= \text{Prob}\,\{\theta^*(X) < \theta \mid \theta\} = \lim_{\epsilon \downarrow 0} a(\theta - \epsilon; \theta, \theta^*), \\ a(\theta+, \theta, \theta^*) &= \text{Prob}\,\{\theta^*(X) > \theta \mid \theta\} = \lim_{\epsilon \downarrow 0} a(\theta + \epsilon; \theta, \theta^*). \end{aligned} \tag{2.1}$$

For formal convenience, we also define $a(\theta, \theta, \theta^*) \equiv 0$. When reference to a given estimator θ^* is understood, we may write simply $a(u, \theta)$, $a(\theta-, \theta)$, or $a(\theta+, \theta)$. The functions $a(\theta-, \theta)$ and $a(\theta+, \theta)$ of θ play a useful technical role, and will be called respectively the *lower* and *upper location functions* of θ^*.

In many problems, estimators for which Prob $\{\theta^*(X) = \theta \mid \theta\} > 0$ for some θ are found not useful. The remaining estimators have continuous c.d.f's, and have $a(\theta-, \theta) \equiv 1 - a(\theta+, \theta)$. No two such estimators having different location functions can be comparable; for $a(\theta-, \theta, \theta^*) < a(\theta-, \theta, \theta^{**})$ is equivalent to $a(\theta+, \theta, \theta^*) > a(\theta+, \theta, \theta^*)$; this shows that neither estimator is at least as good as the other.

The broad and "weak" definition of admissibility adopted here leads to very large admissible classes in typical problems. However it does not seem unreasonable to conceive of the problem of point estimation as one in which the investigator chooses an estimator on the basis of consideration of the risk curves of all estimators in some essentially complete class. In principle this consideration should be complete, but of course the practical counterpart of this can be at most a more or less extensive familiarity with an essentially complete class, developed by study of the risk-curves of a variety of specific estimators, possibly strengthened by some general theoretical considerations (including envelope risk-curves, discussed below), and perhaps also by reference to one or several loss functions and criteria of optimality which may seem more or less appropriate in specific applications. Such an approach is not so difficult to carry out as might be anticipated, as will be illustrated. Of course difficulties of computation or complexity may sometimes dictate that an inadmissable estimator must be adopted; even in such cases, the most general basis on which any particular estimator might be justified, as not too inefficient, is evidently the comparison of its risk-curves with those of other estimators, especially admissible ones.

Knowledge of the admissible class or of an essentially complete class of estimators in the present broad sense can be useful in applying other formulations of the estimation problem. For example, every estimator which is admissible with respect to a squared error loss function must clearly be admissible in the present sense; hence the search for estimators good in the former sense can be restricted without loss to any class known to be essentially complete in the broader sense. In this way, a hierarchy of definitions of admissibility leads to a corresponding nested hierarchy of admissible or essentially complete classes of estimators. (The latter concepts, and that of vector-valued risk functions, were introduced in other contexts by L. Weiss [5].)

3. Admissible confidence limits. If $\theta'' = \theta''(x)$ is a point estimator of θ in a specified problem, with $a(\theta-, \theta, \theta'')$ relatively small (typically, appreciably less than $\frac{1}{2}$) for all θ, then θ'' may be used as an *upper estimator* of θ; if $a(\theta-, \theta, \theta'') \equiv \alpha < \frac{1}{2}$ for all θ, then θ'' is an *upper* $(1 - \alpha)$-level *confidence limit* estimator. Lower estimators are defined similarly.

The merits of any upper estimator depend upon the following considerations, in suitable combination:

(a) Since $a(\theta-, \theta, \theta'')$ is the probability of an error in inferences of the form: "θ is not greater than the observed value $\theta''(x)$," the values $a(\theta-, \theta, \theta'')$ should be suitably small for all θ.

(b) For each θ and each $u > \theta$, $a(u, \theta, \theta'')$ is the probability that θ'' will be larger than necessary to provide a valid upper limit for θ; hence such values $a(u, \theta, \theta'')$ should be suitably small. Such properties in general have been termed *shortness* properties by Neyman [6], and, for confidence limits, *accuracy* properties by Lehmann [7].

(c) For each θ and each $u < \theta$, $a(u, \theta, \theta'')$ is the probability that θ'' will be in error, as an upper limit for θ, by $(\theta - u)$ or more; such values $a(u, \theta, \theta'')$ should be suitably small, since, at least when other things are equal, θ'' should be misleading by as little as possible.

These considerations lead to definitions of admissibility and of complete classes of upper estimators (and, similarly, lower estimators) which coincide formally with the definitions found above for point estimators. Hence there is no necessary formal distinction between the formulations, theories, and techniques of point estimation on the one hand and confidence limit estimation on the other; a single formal theory of point estimation suffices, and the distinctions required are only those of qualitative emphasis and quantitative degree which reflect the variety of possible purposes for which a point or confidence limit estimator may be chosen from, say, an essentially complete class.

4. Admissible interval estimators. If $J = J(x) = (\theta', \theta'') = (\theta'(x), \theta''(x))$ is a pair of point estimators such that $\theta'(x) \leqq \theta''(x)$ for each x in S, then J is an *interval estimator* of θ. In particular, if Prob $\{\theta'(X) \leqq \theta \leqq \theta''(X) \mid \theta\} = 1 - \alpha$ for each θ, then J is a *confidence interval* with confidence coefficient $1 - \alpha$, or a $(1 - \alpha)$ *confidence interval.* (Typically a value $(1 - \alpha) \gg .5$ is chosen.) If θ' and θ'' are respectively lower and upper $[(1 - \alpha)/2]$ confidence limit estimators, then it is natural to call J a *median-unbiased* $(1 - \alpha)$ confidence interval.

The merits of any interval estimator J depend upon the following considerations, in suitable combination:

(a) For each θ, the probabilities $a(\theta-, \theta, \theta'')$ and $a(\theta+, \theta, \theta')$, of underestimation and overestimation of θ by J, should be suitably small. (As with point estimators, it seems desirable to avoid a formulation implying comparability of these two kinds of errors.)

(b) For each set of values $u' < \theta < u''$, the values $a(u', \theta, \theta')$ and $a(u'', \theta, \theta'')$ should be suitably small, representing shortness properties of J corresponding to shortness properties of the lower and upper estimators θ' and θ'' respectively.

(c) For each θ and each $u > \theta$, $a(u, \theta, \theta')$ should be suitably small; and for each θ and each $u < \theta$, $a(u, \theta, \theta'')$ should be suitably small; since, at least when other things are equal, J should be misleading by as little as possible.

To represent all of these properties, we define the *risk* curves of an interval estimator $J = (\theta', \theta'')$, at each θ, as the *pair* of functions $[a(u', \theta, \theta'), a(u'', \theta, \theta'')]$ of u', u''; that is, the risk curves of θ' and θ''. Thus the risk curves of J at θ are a representation of the bivariate cumulative distribution of $\theta'(X)$ and $\theta''(X)$ when θ is true.

These considerations lead us to formulate the following basic *definitions*: An interval estimator $J = (\theta', \theta'')$ will be called at least as good as another $J^* = (\theta^*, \theta^{**})$ if θ' is at least as good as θ^* and θ'' is at least as good as θ^{**} in the sense defined for point estimators in Section 2 above. Similarly, J will be called better than J^* if it is at least as good as J^* and also θ' is better than θ^* and/or θ'' is better than θ^{**}. J will be called admissible if no other interval estimator is better. Complete classes are defined in the usual way. It is convenient to refer to the pair of functions $\alpha(\theta-, \theta, \theta'')$, $\alpha(\theta+, \theta, \theta')$ of θ as the location functions of $J = (\theta', \theta'')$.

If two interval estimators have different location functions, they are not comparable (neither is at least as good as the other); this follows immediately from the corresponding property for point estimators. *A simple sufficient condition for admissibility* of $J = (\theta', \theta'')$ is that θ' and θ'' be admissible point estimators.

5. Confidence curve estimators. The selection of an estimator of one of the above kinds for purposes of informative inference, including typical applications in scientific research, is generally admitted to involve elements of choice which are in some degree arbitrary. Such elements include the choice of a particular confidence level for an interval estimator, and the choice of location functions for an interval estimator with given confidence coefficient. In addition, a point estimate is sometimes desired along with an interval. Such considerations and related ones have led to proposals for use simultaneously of a point estimator and a set of confidence limit or interval estimators having various confidence coefficients. Such estimators may be regarded as a modern formulation of a long-standing practice of reporting estimates in the form $\theta^* \pm k\sigma_{\theta^*}$, where k is some constant and $\sigma^2_{\theta^*} = Var\ (\theta^*(X))$. The latter form may be interpreted as an ordered set of three point estimators. For example, if $\theta^*(X)$ has a normal distribution with a known constant variance, and $k = 1$, then the "estimator" $\theta^*(x) \pm k\sigma_{\theta^*}$ may be written as the ordered set of estimators

$$[\theta^*(x) - \sigma_{\theta^*}, \theta^*(x), \theta^*(x) + \sigma_{\theta^*}] \equiv [\theta(x, .84), \theta(x, .5), \theta(x, .16)].$$

Estimates of this "omnibus" kind can be interpreted flexibly but validly, in any context of application for informative inferences, in the ways customary for (a) point estimates such as $\theta(x, .5)$, (b) confidence limits such as $\theta(x, .84)$ and $\theta(x, .16)$, and (c) confidence intervals such as $[\theta(x, .84), \theta(x, .16)]$.

Tukey [8] proposed that for typical general purposes it would be advantageous to use a set of five point estimators at standard levels: $\theta(x, \alpha)$, with $\alpha = 2\frac{1}{2}\%$, $16\frac{2}{3}\%$, 50%, $83\frac{1}{3}\%$, and $97\frac{1}{2}\%$. Cox [9] proposed use of the full continuous family of confidence limits $\theta(x, \alpha)$, $0 \leqq \alpha \leqq 1$. Such an omnibus estimator includes formally, as elements, not only confidence limits at all levels and a median-unbiased point estimator, but also median-unbiased confidence intervals at all levels. Whether such estimators should be used in practice, rather than more standard methods, is a matter of judgment and taste which can perhaps be decided best in specific contexts of application. It is often convenient, as will be illustrated below, to discuss estimation theory and techniques for estimators of this omnibus form, since such discussion includes conveniently and compactly a treatment of estimators of the various kinds mentioned.

Any such estimator, consisting of a specified set of confidence limit estimators $\theta(x, \alpha)$, α in some specified subset of the closed unit interval (possibly the whole interval), ordered in the sense that $\alpha < \alpha'$ implies $\theta(x, \alpha) \geqq \theta(x, \alpha')$ for each x in S, will be called a *confidence curve estimator*. We shall usually consider the inclusive case, $0 \leqq \alpha \leqq 1$, so as to include formally all other cases. In many problems it is convenient to give such estimators a form which can be reported graphically: if for each $x \varepsilon S$, $\theta(x, \alpha)$ increases continuously from $\underline{\theta}$ to $\bar{\theta}$ as α decreases from 1 to 0, then we define the confidence curve estimator $c(\theta, x)$, for each $x \varepsilon S$, as the continuous curve (function of $\theta \varepsilon \bar{\Omega}$)

$$c(\vartheta, x) = \min [\alpha, 1 - \alpha \mid \theta(x, \alpha) = \theta]. \tag{5.1}$$

For example, if X is normally distributed with unit variance and mean θ, then the confidence curve estimator of θ is

$$c(\theta, x) = \begin{cases} \Phi(\theta - x), -\infty \leqq \theta \leqq x, \\ 1 - \Phi(\theta - x), x \leqq \theta \leqq \infty; \end{cases} \tag{5.2}$$

for any observed value x, the estimate $c(\theta, x)$ can be described by a more or less complete sketch of its graph when convenient.

The definitions of admissibility and of complete classes for confidence curve estimators parallel those above for confidence interval estimators. A simple sufficient (but not, in general, necessary) condition that a confidence curve estimator be admissible is that for each α, its element $\theta^*(x, \alpha)$ be an admissible point estimator. In problems for which there exists a uniformly best confidence limit estimator for each confidence coefficient, this condition is necessary as well as sufficient, and there is a unique (a.e.) admissible confidence curve estimator which consists simply of the family of these best confidence limit estimators.

6. Elementary theory of admissible point estimators. An important part of the general theory of admissible point estimators, and of corresponding practical techniques of estimation, can be developed conveniently by an essentially elementary use of the theory of tests of one-sided hypotheses as originated by Neyman and Pearson and as extended (by simple use of their Fundamental Lemma) to generate a variety of admissible tests of such hypotheses. In problems for which uniformly best one-sided tests exist, the complete theory of admissible estimators is obtained in this way; for other problems, the development of the remaining parts of the theory requires more general methods introduced in Section 10 below.

For each θ_o in Ω, we consider two one-sided testing problems: (a) the problem of testing the hypothesis $H(\theta_o):\theta \leqq \theta_o$ (against the general alternative $H'(\theta_o):\theta > \theta_o$); and (b) the problem of testing $H(\theta_o -):\theta < \theta_o$ (against the general alternative $H'(\theta_o -):\theta \geqq \theta_o$). In case θ_o is a minimum value in Ω, consideration of $H(\theta_o -)$ is to be omitted; if θ_o is a maximum in Ω, $H(\theta_o)$ is omitted.

Any given point estimator $\theta^* = \theta^*(x)$ of θ can be used in the following way to define a test of each of the hypotheses mentioned: Accept the hypothesis if and only if the observed value $\theta^*(x)$ is consistent with the hypothesis. Such a test of the hypothesis $H(\theta_o)$ has the acceptance region $A(\theta_o) = \{x \mid \theta^*(x) \leqq \theta_o\}$; such a test of $H(\theta_o-)$ has acceptance region $A(\theta_o-) = \{x \mid \theta^*(x) < \theta_o\}$. If $\theta_1 < \theta$, then $A(\theta_1-) \subset A(\theta_1) \subset A(\theta -) \subset A(\theta)$; for brevity, we shall say that such a sequence of sets $A(\theta)$ is nondecreasing in θ, with the understanding the argument θ may take a value $(\theta-)$ which is considered smaller than θ and larger than $\theta - \epsilon$ for each positive ϵ.

Such a test of $H(\theta_o-)$ has probabilities of errors of Type I given by

$$1 - \text{Prob } (A(\theta_o-)\mid \theta) = a(\theta_o , \theta, \theta^*) \quad \text{for each} \quad \theta < \theta_o ,$$

and of Type II given by

$$\text{Prob } (A(\theta_o-)\mid \theta) = a(\theta_o-, \theta, \theta^*) \quad \text{for each} \quad \theta \geqq \theta_o .$$

Such a test of $H(\theta_o)$ has probabilities of errors of Type I given by

$$1 - \text{Prob } (A(\theta_o)\mid\theta) = a(\theta_o+, \theta, \theta^*) \quad \text{for each} \quad \theta \leqq \theta_o ,$$

and of Type II given by

$$\text{Prob } (A(\theta_o)\mid \theta) = a(\theta_o , \theta, \theta^*) \quad \text{for each} \quad \theta > \theta_o .$$

Thus each of the error-probabilities $a(u, \theta, \theta^*)$, upon which depend the admissibility of any given point estimator θ^*, appears as an error-probability of a test of a one-sided hypothesis based upon use of θ^*. These relationships provide the following simple sufficient condition for admissibility of a point estimator.

LEMMA 1. *For any specified family of probability density functions $f(x, \theta)$ (with respect to an underlying σ-finite measure $\mu(x)$ defined on the sample space $S = \{x\}$), $\theta \varepsilon \Omega$ (a subset of the real line), a given estimator $\theta^* = \theta^*(x)$ (any measurable function taking values in the closure $\bar{\Omega}$ of Ω) is admissible if each of the acceptance* *regions* $A(\theta_o)$, $A(\theta_o-)$, *based on* θ^* *as defined above, gives an admissible test of the corresponding one-sided hypotheses* $H(\theta_o)$, $H(\theta_o-)$ *defined above.*

Proof. (A test is called admissible if no other test has all error-probabilities at least as small, with at least one strictly smaller.) If θ^* satisfies the assumptions of the Lemma but is inadmissible, let θ^{**} be an estimator better than θ^*. Then $a(\theta_o, \theta, \theta^{**}) \leqq a(\theta_o, \theta, \theta^*)$ for each $\theta \varepsilon \Omega$ and each $\theta_o \neq \theta$, and the inequality is strict for some $\theta = \theta' \varepsilon \Omega$ and some $\theta_o = \theta_o' \varepsilon \bar{\Omega}$, $\theta_o' \neq \theta'$. Assume for definiteness that $\theta_o' > \theta'$ (the other case can be discussed in the same way). Then the acceptance region $\{x \mid \theta^{**}(x) < \theta_o'\}$ gives a better test of the hypothesis $H(\theta_o'-)$ than does $\{x \mid \theta^*(x) < \theta_o'\}$. This contradicts the assumed admissibility of the test based on the latter region, completing the proof.

Many estimators of interest can be conveniently investigated theoretically and constructed practically by the device of using as indicated below a function $v(x, \theta)$, defined for each sample point x and each $\theta \varepsilon \Omega$. If, for each fixed θ, $v(x, \theta)$ is a measurable function of x, it is a *statistic*; and as θ varies, $v(x, \theta)$ represents a family of statistics. We term such a function v a *quasistatistic.*

Corollary 1. *A sufficient condition for admissibility of an estimator* $\theta^*(x)$ *is that it be defined, for each* x, *as the solution* θ *of the equation* $v(x, \theta) = 0$, *where* v *is a quasistatistic such that*:

(a) *For each* x *in* S, $v(x, \theta) = 0$ *holds for a unique* θ *in* $\bar{\Omega}$.

(b) *If* $\theta_1 < \theta_2$ *and* θ_1, θ_2 *are in* $\bar{\Omega}$, *then* $\{x \mid v(x, \theta_1) \leqq 0\} \subset \{x \mid v(x, \theta_2) < 0\}$. (*A simple sufficient condition for* (b) *is that for each* x, $v(x, \theta)$ *be decreasing in* θ. *If* (a) *holds, it suffices that* $v(x, \theta)$ *be nonincreasing in* θ, *for each* x.)

(c) *For each* θ_o *in* Ω, *the acceptance regions* $\{x \mid v(x, \theta_o) \leqq 0\}$ *and* $\{x \mid v(x, \theta_o) < 0\}$ *are admissible respectively for testing the one-sided hypotheses* $H(\theta_o)$ *and* $H(\theta_o-)$.

Proof. If $v(x, \theta)$ satisfies the stated conditions, the conclusion follows immediately from Lemma 1 upon observing that

$$\{x \mid v(x, \theta_o) \leqq 0\} = \{x \mid \theta^*(x) \leqq \theta_o\} \quad \text{and} \quad \{x \mid v(x, \theta_o) < 0\} = \{x \mid \theta^*(x) < \theta_o\}$$

When an estimator θ^* is defined implicitly, by use of a quasistatistic $v(x, \theta)$, as the solution θ of the equation $v(x, \theta) = 0$, in applications it is not necessary to have an explicit formula for $\theta^*(x)$ since for any observed sample point x it suffices merely to determine the corresponding root θ of the defining equation; and in the cases of many such estimators of practical and theoretical interest, no explicit formula for $\theta^*(x)$ is available. The preceding lemma shows that basic qualitative properties of efficiency can be established for such estimators without use of any explicit formula for $\theta^*(x)$. Their quantitive properties can also be determined without such explicit formulas: Since $v(x, u) < 0$ is equivalent to $\theta^*(x) < u$, and $v(x, u) = 0$ is equivalent to $\theta^*(x) = u$, we have

$$(6.1) \quad a(u, \theta, \theta^*) = \begin{cases} \text{Prob}\,[\theta^*(X) \leqq u \mid \theta] = \text{Prob}\,[v(X, u) \leqq 0 \mid \theta] \text{ for } u < \theta \\ \text{Prob}\,[\theta^*(X) \geqq u \mid \theta] = \text{Prob}\,[v(X, u) \geqq 0 \mid \theta] \text{ for } u > \theta. \end{cases}$$

Thus all quantitative properties of such estimators θ^* can be determined, when convenient, by determining

$$\text{Prob } [v(X, u) \leqq 0 \mid \theta] \quad \text{and} \quad \text{Prob } [v(X, u) = 0 \mid \theta] \quad \text{for each} \quad u \neq \theta.$$

Some theoretical properties of such estimators are also conveniently treated in terms of the c.d.f's. of v. For example, if for each $n = 1, 2, \cdots, \theta_n^*$ is an estimator determined by a quasistatistic $v_n = v_n(x_n, \theta)$, then the condition that the sequence of estimators θ_n^* be *consistent* (that is, that $\lim_n a(u, \theta, \theta_n^*) = 0$, for each $\theta \varepsilon \Omega$ and each $u \neq \theta$), can be stated, and in many cases conveniently proved, in the form: $\lim_n$ Prob $[v_n(X_n, u) \leqq 0 \mid \theta] = 0$ or 1, according as $u < \theta$ or $u > \theta$, for each $\theta \varepsilon \Omega$.

7. Uniformly best estimators. Any estimator $\theta^*(x)$ of θ will be called a *uniformly best estimator* if each of the tests of one-sided hypotheses based on θ^*, in the manner of the preceding section, is a uniformly best test (uniformly most powerful on H' and uniformly least powerful on H). Since each such test is admissible, each such estimator is admissible.

It is well known that for a one-sided testing problem there exist uniformly best tests of all sizes, if there exists a sufficient statistic $t(x)$ with the monotone likelihood ratio property (m.l.r.) ([7], Sect. 3.2).

LEMMA 2. *If the family of density functions $f(x, \theta)$, $\theta \varepsilon \Omega$, admits a sufficient statistic $t = t(x)$ having the monotone likelihood ratio property, then an essentially complete class of admissible estimators is constituted by estimators of the form $\theta^* = \theta^*(t, y)$, any nondecreasing function of t and of y, where y is an observed value of an auxiliary randomization variable Y having under each θ the same uniform distribution on the unit interval $0 \leqq y < 1$, and such that $t' < t''$ implies $\theta^*(t', y') \leqq \theta^*(t'', y'')$ for all y', y''. If $t(x)$ has a continuous* c.d.f., *for each θ, then estimators of this form but not depending upon y constitute an essentially complete class of estimators.*

PROOF. Let $\theta^*(x, y)$ be any estimator (possibly depending on an auxiliary randomization variable Y), let $G(\theta) =$ Prob $\{\theta^* \leqq \theta \mid \theta\}$, let $G(\theta-) =$ Prob $\{\theta^*(X) < \theta \mid \theta\}$, let $F(t, \theta) =$ Prob $\{t(X) \leqq t \mid \theta\}$, where $t(x)$ is a sufficient statistic with the m.l.r. property, and let $z(t(x), y, \theta) = yF(t(x), \theta) + (1 - y)F(t(x)-, \theta)$. Consider the quasistatistic

$$v = v(x, y, \theta) = z(t(x), y, \theta) - G(\theta).$$

For each θ_o, $A(\theta_o) = \{(x, y) \mid v(x, y, \theta_o) < 0\}$ is clearly a uniformly best acceptance region for testing $H(\theta_o)$ at level $1 - G(\theta_o) = a(\theta_o+, \theta_o, \theta^*)$. Consider the quasistatistic $v' = v'(x, y, \theta) = z(t(x), y, \theta) - G(\theta-) \leqq v + [G(\theta) - G(\theta-)]$. For each θ_o, $A(\theta_o-) = \{(x, y) \mid v'(x, y, \theta_o) < 0\}$ is clearly a uniformly best acceptance region for testing $H(\theta_o-)$; at $\theta = \theta_o$ it has Type II error-probability $G(\theta_o-) = a(\theta_o-, \theta_o, \theta^*)$.

To verify that these acceptance regions constitute a sequence of sets which is nondecreasing in θ in the sense defined in Section 6, we note that obviously

$A(\theta_o-) \subset A(\theta_o)$, and we proceed to prove that $\theta_1 < \theta_2$ implies $A(\theta_1) \subset A(\theta_2-)$: Assume that $(x', y') \varepsilon A(\theta_1)$; but $(x', y') \not\varepsilon A(\theta_2-)$; then

$$z' \equiv z(t(x'), y', \theta_1) < G(\theta_1)$$

and

$$z'' \equiv z(t(x'), y', \theta_2) \geqq G(\theta_2-).$$

A best test of $H(\theta_1)$ of size $(1 - z')$ (the test which rejects when $z(t(x), y, \theta_1) \geqq z'$) has maximum power at $\theta = \theta_2$, namely $1 - z''$; the test with acceptance region $\{x \mid \theta^*(x) \leqq \theta_1\}$ has size $1 - G(\theta_1) < (1 - z')$ and hence has power $\text{Prob}\ \{\theta^*(X) > \theta_1 \mid \theta_2\} < 1 - z''$. Hence $z'' < \text{Prob}\ \{\theta^*(X) \leqq \theta_1 \mid \theta_2\} \leqq \text{Prob}\ \{\theta^*(X) < \theta_2\} = G(\theta_2-)$, a contradiction which proves that $A(\theta_1) \subset A(\theta_2-)$.

For each (x, y), let $\theta^{**} = \theta^{**}(x, y)$ be defined by

$$\theta^{**}(x, y) = \inf \{\theta \mid \theta \varepsilon \bar{\Omega}, (x, y) \varepsilon A(\theta)\}.$$

Then θ^{**} is a nondecreasing function of $t(x)$ and of y, and is a uniformly best estimator having the same location functions as the arbitrarily given θ^*. Since θ^{**} is admissible, it is strictly better than θ^* or else is equivalent to θ^*, completing the proof.

If for each θ, $F(t, \theta)$ is continuous and increasing in t, and if for each t, $F(t, \theta)$ is continuous and decreasing in θ, then we have the admissible confidence curve estimator

$$c(\theta, t) = \min [F(t, \theta), 1 - F(t, \theta)], \tag{7.1}$$

where $t = t(x)$ is an observed value.

8. Score quasistatistics and generalized maximum likelihood estimators. For a given family $f(x, \theta)$, $\theta \varepsilon \Omega$, let $\theta_1(\theta)$, $\theta_2(\theta)$ be two functions defined on Ω, taking values in $\bar{\Omega}$, and satisfying $\theta_1(\theta) < \theta_2(\theta)$ and $\theta_1(\theta) \leqq \theta \leqq \theta_2(\theta)$ for $\theta \varepsilon \Omega$. Then for each $\theta' \varepsilon \Omega$, a best test at level $\alpha(\theta')$ of $H_1: \theta = \theta_1(\theta')$ against $H_2: \theta = \theta_2(\theta')$ is one which accepts H_1 when the quasistatistic

$$S(x, \theta_1(\theta), \theta_2(\theta)) \equiv [\log f(x, \theta_2(\theta)) - \log f(x, \theta_1(\theta))]/[\theta_2(\theta) - \theta_1(\theta)] \tag{8.1}$$

satisfies $S(x, \theta_1(\theta'), \theta_2(\theta')) \leqq G(\theta', \alpha(\theta'))$, where $G(\theta', \alpha(\theta'))$ is a constant such that $\alpha(\theta')$ is the probability, when θ' is true, that this inequality will be satisfied. For many problems the functions $\theta_1(\theta)$, $\theta_2(\theta)$, and $\alpha(\theta)$ can be chosen so that the *generalized score quasistatistic* $v(x, \theta) = S(x, \theta_1(\theta), \theta_2(\theta)) - G(\theta, \alpha(\theta))$, $\theta \varepsilon \Omega$, satisfies the conditions of Corollary 1 and hence defines an admissible estimator $\theta^*(x)$ as the solution θ of the equation $v(x, \theta) = 0$. If, for example, $\text{Prob}\ \{v(X, \theta) = 0 \mid \theta\} \equiv 0$ for $\theta \varepsilon \Omega$, and the set $\{x \mid f(x, \theta) > 0\}$ is independent of $\theta \varepsilon \Omega$, then each acceptance region $\{x \mid v(x, \theta) \leqq 0\}$ gives a best test which is essentially unique (a.e. P_θ, $\theta \varepsilon \Omega$), and hence admissible for testing $H(\theta)$ and $H(\theta-)$.

Again, as

$$\theta_2(\theta) - \theta_1(\theta) \to 0, S(x, \theta_1(\theta), \theta_2(\theta)) \to S(x, \theta) = \frac{\partial}{\partial\theta} \log f(x, \theta), \tag{8.2}$$

if the derivative exists at each x, for each $\theta \varepsilon \Omega$; consider as above the (*locally-best*) *score quasistatistic* $v(x, \theta) = S(x, \theta) - G(\theta, \alpha(\theta))$. If this $v(x, \theta)$ satisfies the conditions of Corollary 1, then an admissible estimator $\theta^*(x)$ is defined as the solution θ of the equation $v(x, \theta) = 0$. It is well known [7] that if for every set A we have

$$\frac{\partial}{\partial\theta} \int_A f(x, \theta)\, d\mu = \int_A \frac{\partial}{\partial\theta} f(x, \theta)\, d\mu,$$

then an acceptance region $\{x \mid v(x, \theta) \leqq 0\}$ gives a locally-best test of $H(\theta)$ and of $H(\theta)$; under additional mild restrictions, such as those mentioned above, these tests are also admissible. Such estimators will be called *locally-best estimators*. Estimators of this form were proposed on different theoretical grounds by Tukey [8], in connection with the methods discussed in Section 5 above, and by Wald [10], who showed that under broad regularity conditions they are asymptotically efficient. The case $G(\theta, \alpha(\theta)) \equiv 0$ determines (through the equation $S(x, \theta) = 0$) the maximum likelihood estimator $\hat{\theta}(x)$, which is thus shown to be admissible and locally-best under the conditions mentioned.

To illustrate the meaning of the locally-best property in terms of the risk-curves of an estimator, consider a median-unbiased locally-best estimator, for which $a(\theta-, \theta) \equiv a(\theta+, \theta) \equiv \frac{1}{2}$; for convenience here we define $a(\theta, \theta) \equiv \frac{1}{2}$. The locally-best property has been defined in terms of the operating characteristics of tests, represented by $a(u, \theta)$ as a function of θ, for each fixed u; and by a maximum condition on the (absolute values of the right and left) derivatives of $a(u, \theta)$ with respect to θ, at $\theta = u$. This condition, when realized, clearly implies a similar maximum condition on the derivatives of $a(u, \theta)$ with respect to u, for each fixed θ, at $u = \theta$, when continuous partial derivatives of $a(u, \theta)$ exist. And the latter maximum condition directly represents concentration of the distribution of the estimator around θ.

Estimators defined by use of the various score quasistatistics mentioned may be called *generalized maximum likelihood estimators*. (If score statistics have discontinuous distributions, their use can be supplemented if desired by use of randomization variables; we omit discussion of this complication.)

If Prob $\{v(X, \theta) = 0 \mid \theta\} = 0$ for each $\theta \varepsilon \Omega$, then each such estimator has the location functions $a(\theta-, \theta) \equiv 1 - a(\theta+, \theta) \equiv \alpha(\theta)$. If $\alpha(\theta) \equiv \alpha$, a constant, such an estimator is a confidence limit; if $\alpha(\theta) \equiv \frac{1}{2}$, such an estimator is a median-unbiased point estimator. In the important case that $X = (Y_1, \cdots, Y_n)$, a sample of independent observations Y_i, we have $S(X, \theta) = \sum_{i=1}^n S(Y_i, \theta)$; the normal approximation (based on the Central Limit Theorem)

$$a(\theta-, \theta, \hat{\theta}) = \text{Prob}\,\{S(X, \theta) < 0 \mid \theta\} \doteq \Phi(0) = \tfrac{1}{2}$$

(using that $E(S(X, \theta) | \theta) = 0$) is often close; hence in such cases the maximum likelihood estimator $\hat{\theta}(x)$ is approximately median-unbiased. If $S(X, \theta)$ has a symmetrical distribution under θ, then clearly $\hat{\theta}$ is exactly median-unbiased.

In some cases, as illustrated below, a family of score quasistatistics, e.g.

$$v(x, \theta, \alpha) = S(x, \theta) - G(\theta, \alpha), \qquad 0 \leqq \alpha \leqq 1, \tag{8.3}$$

or

$$v(x, \theta, \alpha) = S(x, \theta_1(\theta), \theta_2(\theta)) - G(\theta, \alpha), \qquad 0 \leqq \alpha \leqq 1, \tag{8.4}$$

can be used to determine admissible confidence curve estimators $\theta(x, \alpha)$, $0 \leqq \alpha \leqq 1$, as solutions of equations $v(x, \theta, \alpha) = 0$.

8.1 *Large-sample approximations.* If $x = (y_1, \cdots, y_n)$ is a sample of n independent identically distributed observations (non-identical distributions can be discussed similarly),

$$S(x, \theta_1(\theta), \theta_2(\theta)) = \sum_{i=1}^{n} S(y_i, \theta_1(\theta), \theta_2(\theta)).$$

Let

$$\mu(u, \theta) = E[S(Y_1, \theta_1(u), \theta_2(u)) | \theta]$$

and

$$\sigma^2(u, \theta) = \text{Var}\,[S(Y_1, \theta_1(u), \theta_2(u)) | \theta]$$

exist for each θ, $u \in \bar{\Omega}$. We allow $\theta_1(\theta) = \theta_2(\theta) = \theta$ here, taking $S(X, \theta, \theta) \equiv S(X, \theta)$ in this case, and assume that $\theta_1(\theta)$, $\theta_2(\theta)$ are fixed, while n may vary, in the present discussion.

In the special case $v_n(x, \theta) = \sum_{i=1}^{n} S(y_i, \theta)$, if $v_n(x, \theta)$ satisfies the conditions of Cor. 1, then the maximum likelihood estimator $\hat{\theta}_n(x)$ is the solution θ of $v_n(x, \theta) = 0$. We have by Khintchine's Theorem (even if $\sigma^2(u, \theta)$'s do not exist) that $n^{-1}v_n(X, u)$ converges in probability to $\mu(u, \theta)$ when θ is true. If $u' < \theta < u''$ implies $\mu(u', \theta) < \mu(\theta, \theta) \equiv 0 < \mu(u'', \theta)$, then $\lim_n a(u, \theta, \hat{\theta}_n) = 0$ for $u \neq \theta$; that is, $\hat{\theta}_n$ is consistent.

Returning to the general case, for large n the Central Limit Theorem gives the normal approximation to the distributions of

$$v_n(X, u, \alpha) = \sum_{i=1}^{n} S(Y_i, \theta_1(u), \theta_2(u)) - G_n(u, \alpha):$$

$$\text{Prob}\,\{v_n(X, u, \alpha) \leqq 0 | \theta\} \doteq \Phi\left(\frac{G_n(u, \alpha) - n\mu(u, \theta)}{n^{\frac{1}{2}}\sigma(u, \theta)}\right); \tag{8.1.1}$$

and for $u = \theta$, the approximate determination of $G_n(\theta, \alpha)$:

$$\alpha \doteq \Phi\left(\frac{G_n(\theta, \alpha)}{n^{\frac{1}{2}}\sigma(\theta, \theta)}\right), \text{ or } G_n(\theta, \alpha) \doteq n^{\frac{1}{2}}\sigma(\theta, \theta)\Phi^{-1}(\alpha), \tag{8.1.2}$$

which in the preceding formula gives

$$(8.1.3)\quad \text{Prob}\,\{v_n(X, u) \leqq 0 \mid \theta\} \doteq \Phi\left(-n^{\frac{1}{2}}\frac{\mu(u, \theta)}{\sigma(u, \theta)} + \frac{\sigma(\theta, \theta)}{\sigma(u, \theta)}\Phi^{-1}(\alpha)\right).$$

For the maximum likelihood estimator, $G_n \equiv 0$, corresponding to $\alpha = \frac{1}{2}$ in these formulae. Thus the risk curves of the confidence limit estimator $\theta^* = \theta_n(x, \alpha)$ determined by $v_n(x, \theta, \alpha) = 0$ are approximately

$$(8.1.4)\qquad a(u, \theta, \theta_n(\cdot, \alpha)) = \begin{cases}\Phi(h(u, \theta, \alpha, n)), & u < \theta,\\ 1 - \Phi(h(u, \theta, \alpha, n)), & u > \theta, 0 < \alpha < 1,\end{cases}$$

where

$$h(u, \theta, \alpha, n) = -n^{\frac{1}{2}}\frac{\mu(u, \theta)}{\sigma(u, \theta)} + \frac{\sigma(\theta, \theta)}{\sigma(u, \theta)}\Phi^{-1}(\alpha).$$

Here the sufficient (and necessary) condition for consistency of $\theta_n(x, \alpha)$, for a fixed α, $0 < \alpha < 1$, is again that $u' < \theta < u''$ imply $\mu(u', \theta) < 0 < \mu(u'', \theta)$.

These approximations are of some theoretical and practical use in connection with the sometimes-difficult problem of verification of the conditions of Corollary 1, as illustrated in the discussion of Example 1 in Section 9 below.

8.2 *Local approximations for locally best estimators*. In cases where there exist precise estimators, that is estimators whose risk curves are small except for u very near θ, it is natural to center attention on small neighborhoods of the possible true values θ, and to consider estimators whose risk curves are *relatively* small in such neighborhoods, such as those based on score quasistatistics with $\theta_2(\theta) - \theta_1(\theta)$ small or zero for all θ. If $\mu'(u, \theta) \equiv [\partial/(\partial u)]\mu(u, \theta)$ and $\sigma'(u, \theta) = [\partial/(\partial u)]\sigma(u, \theta)$ exist, then $h'(u, \theta, \alpha, n) = [\partial/(\partial u)]h(u, \theta, \alpha, n)$ gives the Taylor series approximation

$$(8.2.1)\qquad h(u, \theta, \alpha, n) \doteq h(\theta, \theta, \alpha, n) + h'(\theta, \theta, \alpha, n)(u - \theta)$$

and a corresponding alternative form of the above approximation to $a(u, \theta, \theta_n(\cdot, \alpha))$. In the special case of locally-best score quasistatistics, since $\mu(\theta, \theta) \equiv 0$ and $\mu'(\theta, \theta) = \sigma^2(\theta, \theta)$, we find

$$(8.2.2)\quad h(u, \theta, \alpha, n) \doteq n^{\frac{1}{2}}\sigma(\theta, \theta)(\theta - u) + \Phi^{-1}(\alpha)\left[1 + \frac{\sigma'(\theta, \theta)}{\sigma(\theta, \theta)}(\theta - u)\right].$$

In the first term, the coefficient $n^{\frac{1}{2}}\sigma(\theta, \theta)$ of the error $(\theta - u)$ is $(I(\theta))^{\frac{1}{2}}$, where $I(\theta)$ is Fisher's "Information in X at θ." The second term is zero for $\alpha = \frac{1}{2}$ and for the maximum likelihood estimator; for other estimators, the first term dominates the second as n increases. The indicated approximations to risk curves are

$$(8.2.3)\qquad a(u, \theta, \hat{\theta}_n) \doteq a(u, \theta, \theta_n(\cdot, .5)) \doteq \Phi(-n^{\frac{1}{2}}\sigma(\theta, \theta)|u - \theta|),$$

and for $\alpha \neq \frac{1}{2}$

$$a(u, \theta, \theta_n(\cdot, \alpha))$$

$$(8.2.4) \quad \doteq \begin{cases} \Phi\left\{ - n^{\frac{1}{2}}\sigma(\theta,\theta)(\theta - u) + \Phi^{-1}(\alpha)\left[\dfrac{\sigma'(\theta,\theta)}{\sigma(\theta,\theta)}(\theta - u) + 1\right]\right\}, & u < \theta \\ 1 - \Phi\{ \cdots \text{same argument} \cdots \}, & u > \theta, \end{cases}$$

$$\doteq \text{(more roughly) } \Phi\{- n^{\frac{1}{2}}\sigma(\theta, \theta)\ |u - \theta|\}.$$

These approximations exhibit the approximate normality of distribution of these estimators for large n. While locally best estimators are in general not comparable with other estimators (e.g., those above with $\theta_1(\theta) < \theta_2(\theta)$ for all θ) having similar location functions except in problems of a simple structure, the designation "Information" for $I(\theta)$ is clearly appropriate and useful for cases in which so much precision is attainable that interest is practically restricted to very small $|u - \theta|$, in which case an appropriate choice of an estimator will usually be one which is locally best or perhaps one defined as above with $\theta_2(\theta) - \theta_1(\theta)$ small for all θ.

It should be noted that the preceding approximations which utilize a Taylor series approximation are not accompanied by bounds on errors of approximations. Even in cases where such approximations are very close, under a severely nonlinear transformation of the parameter space ($\theta \to \eta = \eta(\theta)$ with $\eta(\theta)$ differentiable and increasing) such approximations can become very inaccurate. Hence the principal concrete value of such approximation formulae seems to be that they provide convenient quantitative conjectures which are more or less plausible but which require independent confirmation or disconfirmation for specific problems and sample sizes. Similar remarks apply to the preceding approximation formulae based on the Central Limit Theorem only, with the qualification that such approximations could be termed "less asymptotic" than those which also use the Taylor series approximation, in the sense that the former approximations are unaffected by monotone transformations of the parameter space, and their use can in principle be accompanied by use of the known bounds on errors in the Central Limit Theorem approximation.

8.3 *Remarks on asymptotic efficiency of estimators*. The theory of the asymptotic efficiency of maximum likelihood estimators (cf., for example, Cramér [11], pp. 489–490, 500–504) utilizes a criterion of asymptotic efficiency which is restrictive in that it applies only to estimators having asymptotically normal distributions with means equal to the parameter estimated; such estimators are clearly asymptotically median-unbiased (probability of underestimation approaches $\frac{1}{2}$ as n increases). It is advantageous to use a less restrictive criterion of asymptotic efficiency, one which applies to all (sequences of) estimators which are asymptotically median-unbiased. In order to embrace confidence limit estimation as well as point estimation, it is advantageous to define a criterion of asymptotic efficiency which can be applied to any sequence of estimators whose probabilities of underestimation (at each θ) converge with increasing n to a fixed constant α, $0 < \alpha < 1$; any such sequence may be termed an *asymptotically valid* sequence of confidence limit estimators (of specified coefficient α).

Under broad conditions (some simple ones were given above) consistent estimators exist; it is then natural to define asymptotic efficiency of estimators in terms of the properties of risk curves of estimators in the neighborhood of the true value of θ: an asymptotically efficient sequence of confidence limit estimators may be defined informally as one which is asymptotically valid and asymptotically locally best. The estimators defined above and illustrated in the following section based upon quasistatistics of the form $v_n(x_n, \theta, \alpha) = S(x_n, \theta) - G_n(\theta, \alpha)$ provide examples of such estimators, and have the further properties of being exactly (non-asymptotically) valid and locally-best (and typically admissible). Additional examples are based on quasistatistics of the form $v_n(x_n, \theta, \alpha) = S(x_n, \theta_{1,n}(\theta), \theta_{2,n}(\theta)) - G_n(\theta, \alpha)$ where as n increases $\theta_{2,n}(\theta) - \theta_{1,n}(\theta)$ decreases to zero rapidly enough to give the asymptotically locally-best property; such estimators have the further properties of exact validity and admissibility, and the functions $\theta_{i,n}(\theta)$ can be chosen so that for any finite sample size a suitable emphasis is given to avoiding errors exceeding specified positive magnitudes; for practical applications, such estimators seem preferable in principle to (exactly) locally-best estimators.

The usual asymptotic theory (Cramér, l.c.) is free of the important assumption (b) of Corollary 1 above. From the present standpoint it may be observed that the principal role of the regularity assumptions of the usual theory is to guarantee that with increasing n, for each θ, the probability of the set of points x_n on which $S(x_n, u)$ is decreasing in u (at least for u near θ) approaches unity (that is, our condition (b) "tends to hold" as n increases).

The remarks of Lehmann [12] on the limited value of any exclusively-asymptotic theory of optimum tests apply with equal force to estimation theory. Asymptotically efficient estimators may approach efficiency at arbitrarily slow rates as n increases. Only on the basis of an auxiliary non-asymptotic investigation of the quantitative and /or qualitative (optimality) properties of an asymptotically efficient estimator can it be recommended in an application with a specified (finite) sample size.

9. Examples.

EXAMPLE 1. *Normal mean.* Let $x = (y_1, \cdots, y_n)$ be a sample of n independent observations from a normal distribution with known variance, say $\sigma^2 = 1$, and unknown mean θ, $-\infty < \theta < \infty$. Then

$$f(x, \theta) = (2\pi)^{-n/2} \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{n} (y_i - \theta)^2 \right\}.$$

Since this example, with the statistic $t(x) = \bar{y} = \sum_{i=1}^{n} y_i/n$, satisfies the conditions of Lemma 2, estimators of the form $\theta^*(\bar{y})$ which are nondecreasing functions of $\bar{y}$ constitute an essentially complete class of admissible estimators. In the general case where σ has any known positive value, letting $\Phi(u)$ denote the standard normal c.d.f, we have uniformly best confidence limit estimators at each level α or $1 - \alpha$:

$$\theta(x, \alpha) = \bar{y} - \Phi^{-1}(\alpha)\sigma n^{-\frac{1}{2}}. \tag{9.1}$$

When $\alpha = \frac{1}{2}$, we obtain the classical estimator $\bar{y}$, which is seen to be a uniformly best median-unbiased estimator. Since $\bar{y}$ is independent of σ, it can be used even when σ is unknown, in which case it remains median-unbiased and is uniformly best over all values of θ and σ. (The latter property can be represented formally in term of risk curves $a(u, \theta, \sigma, \theta^*)$, representing the distributions of any estimator $\theta^*(x)$ as they depend upon θ and σ. This illustrates a general method of extending the treatment of the present paper to problems involving nuisance parameters.) The same property clearly holds for each of the classical least squares estimators of an estimable function in linear regression theory under normality assumptions, and for the classical estimator of each component of the mean of a multivariate normal distribution. (In a different formulation of the estimation problem, Stein [13] has shown that in general the latter classical estimators are inadmissible; this result is based upon a decision-theoretic formulation in which the particular form adopted for the loss function plays a crucial role.)

EXAMPLE 2. *Logistic mean.* Let $x = (y_1, \cdots, y_n)$ be a sample of n independent observations from a logistic distribution with unknown mean θ:

$$\text{Prob}\,(Y \leqq y \mid \theta) = \Psi(y - \theta) = (1 + \exp\{-(y - \theta)\})^{-1},$$
$$-\infty < y < \infty, \qquad -\infty < \theta < \infty;$$

Y has the density function

$$\psi(y - \theta) = \exp\{-(y - \theta)\}/(1 + \exp\{-(y - \theta)\})^2, \quad -\infty < y < \infty. \tag{9.2.1}$$

For any fixed $\Delta > 0$, taking $\theta_1(\theta) = \theta - \Delta$, $\theta_2(\theta) = \theta + \Delta$, determines a score quasistatistic

$$\begin{aligned} S(x, \theta - \Delta, \theta + \Delta) &= \frac{1}{2\Delta} \\ &\cdot \left[\sum_{i=1}^{n} (\log \psi(y_i - \theta - \Delta) - \log \psi(y_i - \theta + \Delta)\right]. \end{aligned} \tag{9.2.2}$$

For any fixed α, $0 \leqq \alpha \leqq 1$, taking $\alpha(\theta) \equiv \alpha$ determines a score quasistatistic

$$v(x, \theta, \alpha) = S(x, \theta - \Delta, \theta + \Delta) - G(\theta, \alpha) \tag{9.2.3}$$

which satisfies the conditions of Corollary 1 of Section 6 above, and hence determines an admissible confidence limit estimator $\theta^* = \theta(x, \alpha)$ as the solution θ of the equation $v(x, \theta, \alpha) = 0$. Since θ is a translation parameter, $G(\theta, \alpha)$ is independent of θ, and may be written $G(\alpha)$. By symmetry, $G(.5) = 0$. $G(\alpha)$ can be determined approximately, except for α very near 0 or 1 and for very small n, by use of the Central Limit Theorem. A locally best confidence limit estimator $\theta^* = \theta(x, \alpha)$ is determined as the solution θ of the equation

$$v(x, \theta, \alpha) \equiv S(X, \theta) - G(\alpha) = 0. \tag{9.2.4}$$

Here $S(y, \theta) = [\partial/(\partial\theta)] \log \psi(y - \theta) = 2\Psi(y - \theta) - 1$; $\Psi(Y - \theta)$ has, when θ is true, a uniform distribution on the unit interval; hence when θ is true the c.d.f. of $\sum_{i=1}^{n} \Psi(Y_i - \theta)$ (and hence that of $S(X, \theta)$) can be calculated as in Cramér [11], pp. 244–246. The normal approximation gives (since

$$\sigma^2(0) = \text{Var}\,[S(Y, \theta) | \theta] = \tfrac{1}{3}, \qquad \text{Var}\,[S(X, \theta) | \theta] = n/3),$$

$$G(\alpha) \doteq \Phi^{-1}(\alpha)(n/3)^{-\frac{1}{2}};$$

$\alpha = \frac{1}{2}$ gives exactly $G(.5) = 0$ and determines the maximum likelihood estimator $\hat{\theta} = \theta(x, .5)$. In general, a locally best confidence limit estimator $\theta(x, \alpha)$ is determined (approximately, except fos $\alpha = .5$) as the root θ of the equation $S(x, \theta) = \Phi^{-1}(\alpha)(n/3)^{\frac{1}{2}}$, or

$$\sum_{i=1}^{n} \Psi(y_i - \theta) \doteq (n/2) + \Phi^{-1}(\alpha)(n/3)^{\frac{1}{2}}/2. \tag{9.2.5}$$

Such an equation is easily solved numerically by use of Berkson's tables of $\Psi(u)$ ([14]).

The present example serves also to illustrate the determination of an admissible confidence curve estimator by use of a family of quasistatistics as described at the end of Section 6 above. Each of the families of quasistatistics $v(x, \theta, \alpha)$, $0 \leqq \alpha \leqq 1$ considered here (each based upon a fixed $\Delta \geqq 0$) has the property that $\theta(x, \alpha)$ is, for each fixed x, decreasing in α; in fact, for each x, $\theta(x, \alpha)$ decreases continuously from ∞ to $-\infty$ as α increases from 0 to 1. Thus for each observed x, each θ $(-\infty \leqq \theta \leqq \infty)$ will be a confidence limit $\theta(x, \alpha)$ for some α; we can conveniently determine the required solutions $\theta(x, \alpha)$ of

TABLE I

i	θ_i	S_i	approx. α_i	exact α_i
1	2.0	−0.559	.288	
2	1.44	−0.256	.399	
3	1.18	−0.758	.470	
4	1.12	−0.031	.488	
5	1.08	−0.0005	.4998	.4998
6	3.08	−0.927	.177	
7	4.0	−1.166	.122	
8	5.0	−1.511	.065	
9	6.0	−2.0	.023	
10	7.0	−2.462	.007	
11	−1.0	1.924	.973	
12	−2.0	2.523	.994	.998
13	0.0	1.0	.841	.833

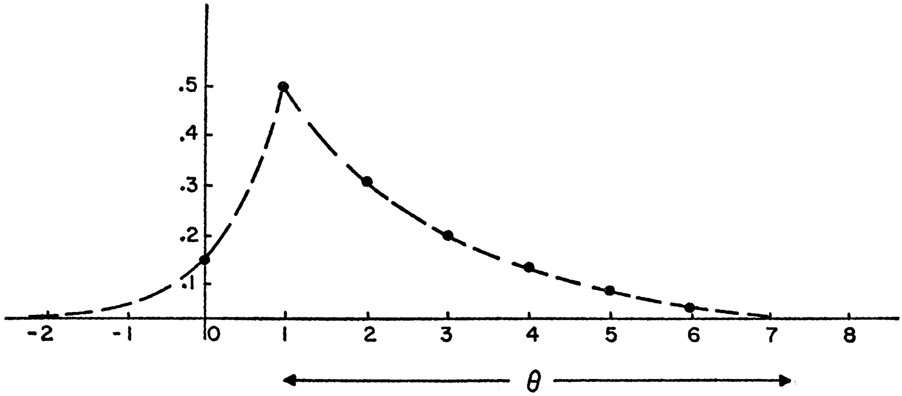


FIG. 1

$v(x, \theta, \alpha) = 0$ in the form

$$(9.2.6) \qquad \alpha(x, \theta) = \text{Prob}\,\{S(X, \theta) \leqq S(x, \theta) \mid \theta\} \doteq \Phi(S(x, \theta)(3/n)^{\frac{1}{2}})$$

for as many values of θ as desired.

NUMERICAL EXAMPLE. Let $x = (y_1, y_2, y_3) = (0, 0, 6)$. Letting θ_i denote a trial value of θ, $S_i = S(x, \theta_i)$, and $\alpha_i = \alpha(x, \theta_i) = \text{Prob}\,\{S(X, \theta_i) \leqq S(x, \theta_i) \mid \theta_i\}$, $i = 1, 2, \cdots$, and taking $\theta_1 = \bar{y} = 2$ as a trial value plausibly near $\theta(x, .5) = \hat{\theta}$, we obtain

$$S_1 = 2\sum_{i=1}^{3} \Psi(y_i - 2) - 3 = -.559, \qquad \alpha_1 \doteq \Phi(-0.559) = .288.$$

Further similar computations are summarized in Table I and in Fig. 1, a sketch of the confidence curve $c(\theta, x) = \min\,[\alpha(x, \theta), 1 - \alpha(x, \theta)]$.

The closeness of the normal approximations can be checked in the present case by use of the exact formula (based on Cramér, l.c.)

$$(9.2.7) \qquad \alpha(x, \theta) = \begin{cases} z^3/6, & 0 \leqq z \leqq 1, \\ z^3/6 - (z - 1)^3/2, & 1 \leqq z \leqq 2, \\ 1 - (3 - z)^3/6, & 2 \leqq z \leqq 3, \end{cases}$$

where $z = z(x, \theta) = (S(x, \theta) + 3)/2$. The approximation is seen to be quite adequate here. In other examples, if exact values of $\alpha(x, \theta)$ cannot be obtained by use of standard tables or tractable integrals, one may consider checking approximate values of $\alpha(x, \theta)$, for a few values of θ of particular interest, by use of (a) the error-bound on the normal approximation, (b) numerical integration, (c) empirical sampling (Monte Carlo), or possibly (d) an asymptotic expansion. For (a) and (d), see Wallace [15].

The values θ_i above, for $i = 2, \cdots, 5$, were determined by $\theta_{i+1} = \theta_i + S_i$, based on Fisher's formula $\theta_{i+1} = \theta_i + S(x, \theta_i)/\text{Var}\,[S(X, \theta_i) \mid \theta_i]$ for iterative calculation of maximum likelihood estimates.

The values θ_6 and θ_{11} above were chosen as trial approximations to the confidence limits $\theta(x, .025)$, $\theta(x, .975)$ respectively, by use of the asymptotic formula for such confidence limits:

$$\hat{\theta} \pm \Phi^{-1}(.975)/(\text{Var}\,[S(X, \theta) \mid \hat{\theta}])^{\frac{1}{2}} \doteq \hat{\theta} \pm 2.$$

The poor approximations obtained provide a limited illustration of the fact that such approximations are "more asymptotic," i.e., may be expected to be often less close, than the normal approximations to distributions of score statistics.

EXAMPLE 3. *Rectangular mean.* Let $x = (y_1, \cdots, y_n)$ be a sample of n independent observations on a random variable Y with density

$$h(y, \theta) = \begin{cases} 1 \text{ if } \theta - \frac{1}{2} \leqq y \leqq \theta + \frac{1}{2} \\ 0 \textit{ otherwise.} \end{cases} \tag{9.3.1}$$

with $\theta = E(Y)$ unknown. Let r and s denote respectively the smallest and the largest of the observed values y_i. Let $\theta^* = \theta^*(r, s)$ be any function, defined for all r, s such that $r \leqq s \leqq r + 1$, which satisfies $s - \frac{1}{2} \leqq \theta^*(r, s) \leqq r + \frac{1}{2}$ and which is nondecreasing in r and in s. Then $\theta^*(r, s)$ satisfies the conditions of Lemma 1 since, for each θ_o, $\{x \mid \theta^* \leqq \theta_o\}$ and $\{x \mid \theta^* < \theta_o\}$ satisfy the (necessary and) sufficient condition given by Pratt [16] for admissibility of one-sided tests on θ. (It can be shown that such estimators constitute an essentially complete class.)

For samples of size $n = 2$, each of the following estimators is admissible and median-unbiased:

$$\theta^*(x) = (r + s)/2, \text{ the usual mean-unbiased estimator.}$$

$$\theta'(x) = \begin{cases} s - \frac{1}{2}, & \text{if } s \geqq r + 2^{-\frac{1}{2}}, \\ r + (2^{\frac{1}{2}} - 1)/2, & \text{if } s \leqq r + 2^{-\frac{1}{2}}, \end{cases}$$

$$\theta_2''(x) = \begin{cases} r + \frac{1}{2}, & \text{if } r \leqq s - 2^{-\frac{1}{2}}, \\ s - (2^{\frac{1}{2}} - 1)/2, & \text{if } r \geqq s - 2^{-\frac{1}{2}}. \end{cases}$$

Among median-unbiased admissible estimators, θ' is uniformly best with respect to errors of under-estimation, and θ'' is uniformly best with respect to errors of over-estimation. Analogous confidence curve estimators are easily constructed.

For any fixed k, $0 \leqq k \leqq .5$, for testing hypotheses of the form $H(\theta_o): \theta \leqq \theta_o$ or $H(\theta_o-): \theta < \theta_o$, there is an admissible acceptance region

$$A(\theta_o) = \{x \mid (r + s)/2 \leqq \theta_o + k, \; s \leqq \theta_o + .5\}$$

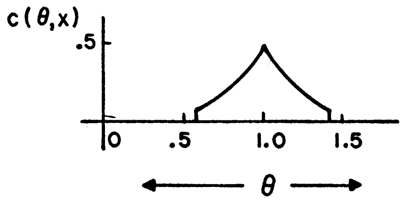
and another admissible acceptance region

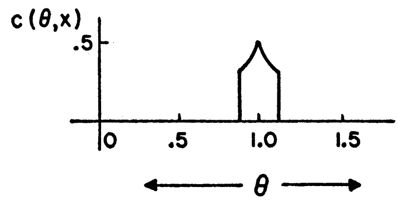
$$A'(\theta_o) = \{x \mid (r + s)/2 \leqq \theta_o - k, \quad \text{or} \quad r \leqq \theta_o - .5\}.$$

From such tests we obtain admissible confidence limit estimators at each level, and the corresponding admissible confidence curve estimator:

$$c(\theta, x) = \begin{cases} 0, \text{ if } \theta \geqq r + .5 \text{ or } \theta \leqq s - .5, \\ 2[.5 - |\theta - (r + s)/2]^2, \text{ otherwise.} \end{cases} \tag{9.3.2}$$

If $x = (0.9, 1.1) = (r, s)$, or alternatively if $x = (0.6, 1.4) = (r, s)$, we obtain respective confidence curve estimates which reflect that the "amount of information in a sample" increases with $(s - r)$:





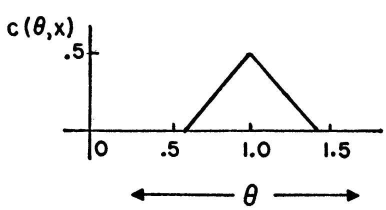
Alternatively, for any fixed k, $-.5 \leqq k \leqq .5$, there is for each $H(\theta_o)$ and $H(\theta_o-)$ an admissible acceptance region

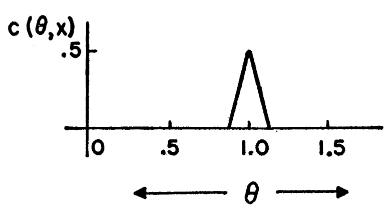
$$A(\theta_o) = \{x \mid (.5 - k)r + (.5 + k)s \leqq \theta_o + k\}.$$

From such tests we obtain admissible confidence limit estimators at each confidence level, and the corresponding admissible confidence curve estimator:

$$c(\theta, x) = \begin{cases} 0, \text{ if } \theta \geqq r + .5 \text{ or } \theta \leqq s - .5, \\ [1 - |r + s - 2\theta|/(1 - s + r)]/2, \text{otherwise.} \end{cases} \tag{9.3.3}$$

For the two samples considered above, we obtain the respective confidence curve estimates:





Since the last curve lies under that given by the first estimator for the same sample, it provides stronger inferences about θ. This is not inconsistent with the admissibility of the first estimator, which provides (at most confidence levels) stronger inferences (shorter confidence intervals) from relatively uninformative samples like the first sample.

EXAMPLE 4. *Cauchy median.* Let Y have the Cauchy density function $h(y, \theta) =$

$1/\pi(1 + (y - \theta)^2)$, $-\infty < y < \infty$, $-\infty < \theta < \infty$. Then $S(y, \theta) = 2(y - \theta)/[1 + (y - \theta)^2]$. Taking $v(x, \theta) = S(y, \theta)$, the conditions of Corollary 1 are satisfied, and $v(x, \theta) = 0$ defines the median-unbiased locally-best estimator $\theta^*(y) = y$. However for $\alpha \neq .5$, $0 < \alpha < 1$, the conditions of Corollary 1 are not satisfied by $v(x, \theta) = S(y, \theta) - G(\alpha)$. For $x = (y_1, y_2)$, even for $\alpha = .5$, $v(x, \theta) = S(x, \theta) = \sum_{i=1}^{2} S(y_i, \theta)$ fails to satisfy the conditions of Corollary 1. (For $|y_2 - y_1|$ large, $S(x, \theta) = 0$ has three roots θ.) Thus *in general there do not exist confidence limit estimators (nor median-unbiased estimators) which are locally-best uniformly in θ.*

Detailed treatment of other examples will be reported elsewhere.

10. Introduction to general theory of admissible estimators. To illustrate the general theory of admissible estimators, and the place of the methods introduced above within the general theory, we consider the case in which Ω is finite: $\Omega = \{\theta \mid \theta = 1, 2, \cdots k\}$. The principal features of the general case (in which Ω is any subset of the real line) can be illustrated conveniently in this case, for which the complete theory can be developed by relatively elementary methods. For any such estimation problem, we have a specified family of density functions $f(x, \theta)$, $\theta = 1, \cdots, k$. For each estimator $\theta^*(x)$, let

$$b(u, \theta, \theta^*) = \begin{cases} \text{Prob } [\theta^*(X) = u \mid \theta], & \text{if } u \neq \theta, \\ 0, & \text{if } u = \theta. \end{cases}$$

We may interpret such an estimation problem in relation to a different *multi-decision problem*, that of choosing, on the basis of an observed value x, one of k specified simple hypotheses. Any measurable function $\theta^*(x)$ taking only the values $1, \cdots k$, represents both a possible solution to the multidecision problem and an estimator.

For the multidecision problem, the merits of each decision function $\theta^*(x)$ are represented completely by its error-probabilities $b(j, \theta, \theta^*)$. A decision function θ^* is called admissible if there is no other for which all corresponding error-probabilities are at least as small, with at least one strictly smaller. Complete classes, minimal essentially complete classes, etc., are defined correspondingly (cf., Lindley [1] and Wolfowitz [17]). It is readily seen that a necessary condition that $\theta^*(x)$ be admissible for the estimation problem is that it be admissible for the multidecision problem.

The relations between the estimation and multidecision problems can be illustrated further in terms of techniques, related to Bayes' formula, which play basic roles in the theory of each problem: For any estimation problem specified as above, let $q = q(u, \theta)$ be an arbitrary real-valued function such that $q(u, \theta) \geqq 0$ for $u, \theta = 1, \cdots k$; any such function will be called a *weight function* (for the estimation problem). For any such q and any estimator θ^*, we define the *Bayes risk*:

$$R(q, \theta^*) = \sum_{\theta=1}^{k} \sum_{u=1}^{k} q(u, \theta) a(u, \theta, \theta^*). \tag{10.1}$$

For any multidecision problem specified as above let $Q = Q(u, \theta) \geqq 0$ be an arbitrary weight-function; then for any multidecision function θ^* the corresponding Bayes risk is:

$$R'(Q, \theta^*) = \sum_{\theta=1}^{k} \sum_{u=1}^{k} Q(u, \theta) b(u, \theta, \theta^*). \tag{10.2}$$

For any given θ^* and $q(u, \theta)$, it is easily verified that

$$R(q, \theta^*) = \sum_{\theta} \sum_{j} Q(j, \theta), b(j, \theta, \theta^*), \tag{10.3}$$

where

$$Q(j, \theta) = \begin{cases} \sum_{j \geqq u > \theta} q(u, \theta), & \text{for } j > \theta, \\ 0, & \text{for } j = \theta, \\ \sum_{j \leqq u < \theta} q(u, \theta), & \text{for } j < \theta. \end{cases}$$

For each θ, $Q(j, \theta)$ is nondecreasing in j for $j \geqq \theta$, and nonincreasing in j for $j \leqq \theta$. Thus each weight-function $q(u, \theta)$ for the estimation problem determines uniquely a weight-function $Q(j, \theta)$ for the multidecision problem which has, for each θ, a single relative minimum; and conversely each such Q determines a unique q. Thus the Bayes solutions θ^* for the estimation problem (i.e., the functions θ^* which, for some given q, minimize $R(q, \theta^*)$) are a subclass of the Bayes solutions for the multidecision problems, characterized by the preceding restriction on the possible forms of the weight function $Q(u, \theta)$ for the latter problem.

For any given weight-function q, the determination of Bayes estimators is conveniently carried out as follows: Let Q be determined by q as above. Then $R(q, \theta^*) = R'(Q, \theta^*)$ is minimized if, for each x, $\theta^*(x)$ takes the (a) value u which minimizes $\sum_{\theta=1}^{k} Q(u, \theta) f(x, \theta)$. Various simple conditions for admissibility of such Bayes multidecision functions, when applicable, immediately imply admissibility of the same functions as estimators.

Various specific formulations of the estimation problem can be exhibited as special cases of the present formulation, corresponding to various choices of the weight-function q. This applies in particular to mean-squared error and other loss function formulations; choice of suitable simple loss functions, taking at most two positive values for each θ, leads to estimators defined by score quasi-statistics.

REFERENCES

[1] D. V. Lindley, "Statistical inference," *J. Roy. Stat. Soc.*, Ser. B, Vol. 15 (1953), pp. 30–76.

[2] Correa Moylan Walsh, *The Problem of Estimation. A Seventeenth-Century Controversy and Its Bearing on Modern Statistical Questions, Especially Index-numbers*, London, P. S. King and Son, 1921.

[3] Leonard J. Savage, *The Foundations of Statistics*, New York, John Wiley & Sons, 1954.

[4] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any form," *Biometrika*, Vol. 30 (1939), pp. 391–421.

[5] Lionel Weiss, "A higher order complete class theorem," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 677–680.

[6] Jerzy Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. of London*, Ser. A, Vol. 236 (1937), pp. 333–380.

[7] E. L. Lehmann, *Testing Statistical Hypotheses*, New York, J. Wiley & Sons, 1959.

[8] John W. Tukey, "Standard confidence points," Memorandum Report 26, Statistical Research Group, Princeton University, July 26, 1949.

[9] D. R. Cox, "Some problems connected with statistical inference." *Ann. Math. Stat.*, Vol. 29 (1958), pp. 357–372.

[10] A. Wald, "Asymptotically shortest confidence intervals," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 127–137.

[11] Harald Cramér, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[12] E. L. Lehmann, "Some comments on large sample tests," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1949, pp. 451–458.

[13] Charles Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1956, pp. 197–206.

[14] Joseph Berkson, "Tables for the maximum likelihood estimate of the logistic function," *Biometrics*, Vol. 13 (1957), pp. 28–34.

[15] David L. Wallace, "Asymptotic approximations to distributions," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 635–654.

[16] J. W. Pratt, "Admissible one-sided tests for the mean of a rectangular distribution," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1268–1271.

[17] J. Wolfowitz, Review of [1] by D. V. Lindley, *Math. Reviews*, Vol. 15 (1954), p. 242.