ON THE ESTIMATION OF PARAMETERS RESTRICTED BY INEQUALITIES[1]

By H. D. Brunk

University of Missouri

1. Summary. There are collected in this paper several observations and results more or less loosely related by their connections with the subject mentioned in the title. The discussion moves from the general to the specific, beginning with some remarks on minimization of convex functions subject to side conditions, and ending with a discussion of uniform consistency of estimators of linearly ordered parameters.

Section 2 deals with one aspect of the problem of minimizing a function of several variables, subject to side conditions which specify that the variables must satisfy certain inequalities. It is frequently true in such problems that information as to which of the restricting sets contain the minimizing point on their boundaries is of great assistance in finding this point. Theorem 2.1 provides the basis for a stepwise procedure leading to this information when both the function to be minimized and the restricting sets are convex. It makes no contribution, however, to the problem of finding the minimizing point on a given boundary or intersection of boundaries.

Brief mention is made in Section 3 of some examples of estimation problems for which the remark to which Section 2 is devoted is appropriate.

Section 4 is concerned with a situation in which samples are taken from k populations, each known to belong to a given one-parameter "exponential family". The problem is the maximum likelihood estimation of the k parameters determining the populations, subject to certain restrictions. Methods are discussed of finding the minimizing point on a given intersection of boundaries of restricting sets. In the particular case when all populations belong to the same exponential family and when the restrictions on the parameters are order restrictions, it is observed that the maximum likelihood estimators (MLE's) of the means are independent of the particular exponential family.

In Section 5 is discussed a property, related to sufficiency, of the MLE's discussed in Section 4. Let y denote a vector representing a set of possible values of the MLE's, E a Borel subset of the sample space, τ a parameter point, S_0 the intersection of the restricting sets. If S_0 is bounded by hyperplanes, there is a determination of the conditional probability $p_T(E \mid y)$ which is independent of τ when y is interior to S_0, and, when y lies on a face, edge, or vertex of S_0, is independent of τ on the closure of that face, edge, or vertex. This result may

Received September 10, 1956; revised December 16, 1957.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18(600)-1108.

be regarded as a generalization of a remark ([16], p. 77) to the effect that if X and Y are normally distributed random variables with unit standard deviation and means ξ and η respectively, and if ξ and η are known to satisfy a linear equation, then the foot of the perpendicular from the observation point (x, y) to the line is a sufficient estimator.

Section 6 is devoted to the same problem as are Sections 4 and 5, except that the parameters are linearly ordered, and that the populations need not belong to exponential families. Conditions are obtained for the strong uniform consistency of an estimator which is the MLE when the populations do belong to the same exponential family. An asymptotic lower bound is given for the probability of achieving a given precision uniformly.

2. Minimizing a convex function on the intersection of closed convex sets. (The author's thanks are due the referee, whose suggestions have materially improved the exposition in this section.) Let $y = (y_1, y_2, \cdots, y_k)$ denote the generic point of R_k, Euclidean space of k dimensions, and let $G(y)$ be a lower semi-continuous function such that $\{y : G(y) \leqq a\}$ is bounded for each a, satisfying

$$G[\lambda y' + (1 - \lambda)y''] \leqq \max [G(y'), G(y'')] \tag{2.1}$$

for $0 \leqq \lambda \leqq 1$, and for all y', y'' in its (convex) domain of definition. (This form of condition (2.1) is due to the referee.) In particular, G satisfies (2.1) if G is convex.

For an arbitrary set $A \subset R_k$, let $\mathcal{B}(A)$ denote its boundary. We write $A \subset B$ if A is properly contained in B or if $A = B$. Let ϕ denote the empty set. Let there be given a finite number of intersecting closed convex sets A_i $(i = 1, 2, \cdots, N)$. We assume G defined on a convex set containing $\bigcup_{i=1}^{N} A_i$. We define Q_i to be the set on which $G(y)$ achieves its minimum value for $y \in A_i$, $i = 1, 2, \cdots, N$. For a set $i_1, i_2, \cdots, i_n$ of distinct positive integers not greater than N we define $Q_{i_1, i_2, \cdots, i_n}$ to be the set on which $G(y)$ achieves its minimum value for $y \in A_{i_1} A_{i_2} \cdots A_{i_n}$.

THEOREM 2.1. *Let A_1, A_2 be intersecting closed sets, A_1 convex. Then either $Q_{12} \subset Q_1$ or $Q_{12}\mathcal{B}(A_2) \neq \phi$.*

PROOF. If $Q_1 A_2 \neq \phi$ then obviously $Q_1 \supset Q_{12}$. It remains to consider the situation in which $Q_1 A_2 = \phi$. Let $p \in Q_1$, $q \in Q_{12}$. Since A_1 is convex, the segment pq lies in A_1. Since $p \notin A_2$, $q \in A_2$, there is a point r on pq such that $r \in A_1\mathcal{B}(A_2)$. By property (2.1), $G(r) \leqq G(q)$, hence $r \in Q_{12}$. This completes the proof of Theorem 2.1.

COROLLARY 2.1. *If $G(y)$ is lower semi-continuous, if $\{y : G(y) \leqq a\}$ is bounded for each a and if G satisfies*

$$G[\lambda y' + (1 - \lambda)y''] < \max [G(y'), G(y'')] \tag{2.2}$$

for $0 < \lambda < 1$, and for all y', y'' in its (convex) domain of definition, and if A_1, A_2 are intersecting closed convex sets in its domain of definition, then Q_1 and Q_{12} consist of single points, q_1 and q_{12}; either $q_1 = q_{12}$ or $q_{12} \in \mathcal{B}(A_2)$. We note that a strictly convex function G satisfies (2.2).

Corollary 2.1 justifies the procedure outlined in the following paragraph for minimizing G subject to the condition $y \varepsilon A_1A_2, \cdots, A_N$, where $A_1, A_2, \cdots, A_N$ are given intersecting closed convex sets. In many particular instances of this problem, one of the chief difficulties is that of determining which of the sets A_i contain the solution (a point minimizing G) on their boundaries, when the point at which G attains its unrestricted minimum is not in $A_1A_2, \cdots, A_N$. The procedure described below can be used to determine those sets among $A_1, A_2, \cdots, A_N$ on whose boundaries the solution lies. We remark that G need not be convex in order for the method to apply, provided it is lower semi-continuous and satisfies (2.2).

The first step is to determine the point at which G assumes its unrestricted minimum. If this point lies in $A_1A_2, \cdots, A_N$, it is the solution. If not, one of the sets is selected in which it does not lie, and designated as A_1 (relabelling, if necessary). Now consider the problem of minimizing G subject to $y \varepsilon A_1$. Applying Corollary 2.1, with A_1 there replaced by the whole space in this application, and A_2 there by A_1 in this application, we find that the solution, q_1, lies on $\mathcal{B}(A_1)$. It may be that q_1 lies in $A_1A_2, \cdots, A_N$, in which case it is the solution. If not, we designate as A_2 (relabelling, if necessary) one of the sets which does not contain q_1. We now consider the problem of minimizing G subject to $y \varepsilon A_1A_2$. By Corollary 2.1, the solution q_{12} lies on $\mathcal{B}(A_2)$. We find first the point q_2 where G is minimized subject to $y \varepsilon A_2$. If $q_2 \varepsilon A_1A_2$, then $q_2 = q_{12}$ is the solution of the present limited problem. Otherwise, by another application of Corollary 2.1, $q_{12} \varepsilon \mathcal{B}(A_1)\mathcal{B}(A_2)$, etc.

This stepwise procedure was introduced in situations involving certain functions G and convex sets A_i described by inequalities of the form $y_j \leqq y_k$ by van Eeden ([13], Theorem I, p. 445; [14], Theorem II, p. 134). The stepwise procedure outlined above makes no contribution to the problem of finding the point where G is minimized on a given "extended hyperface" $\mathcal{B}(A_{i_1}) \cdots \mathcal{B}(A_{i_n})$. Further, in special cases it may even occur that one will determine the minimizing point on each of the 2^N-1 "extended hyperfaces" before finding a minimizing point in $A_1A_2, \cdots, A_N$. Usually, however, one will expect the procedure to terminate with the solution long before all "extended hyperfaces" have been examined.

Non-linear programming methods have been developed for solving certain problems of this class (see, for example, [3]). Problems arising from some of the applications discussed below are such that it is relatively easy to find the minimizing point on a given "extended hyperface", and some trial calculations with such problems using the above stepwise procedure resulted in far less lengthy calculations than did those using general nonlinear programming methods.

3. Examples.

(i) In the bioassay type of problem, one is required to minimize a convex function of the form

$$-\sum_{i=1}^{N} [a_i \log y_i + b_i \log (1 - y_i)], \tag{3.1}$$

where the a_i and b_i are given numbers, and the y_i are subject to the restriction

$$0 \leqq y_1 \leqq y_2 \leqq \cdots \leqq y_N \leqq 1.$$

Even if one is not willing to assume a particular form for the distribution function and is thus led to this nonparametric formulation, he may feel that, for example, the distribution function should not rise too rapidly, and be led to impose further conditions of the form

$$y_{i+1} - y_i \leqq c_i \text{ or } y_{i+2} - 2y_{i+1} + y_i \leqq d_i \tag{3.2}$$

where the c_i and d_i are prescribed numbers. The problem remains in the class discussed in Section 2; however, the minimizing point on the boundary of a set described by inequalities of the form (3.2) is not in general so easily found as is that on a boundary $y_i = y_{i+1}$. The fact that the partial derivatives of the function (3.1) are so readily determined suggests that the method of Lagrange's multipliers, together with Newton's (multivariate) method for solution of simultaneous equations may prove appropriate.

A similar but simpler problem might conceivably arise in connection with ordinary random sampling. Let $x_1, \cdots, x_n$ be sample values of a sample of size n from a population with unknown distribution function F, and let $p_1, p_2, \cdots, p_n$ be the salti or jumps of F at the sample values. The MLE's of $p_1, p_2, \cdots, p_n$ maximize $\prod_{i=1}^n p_i$ or minimize $-\sum_{i=1}^n \log p_i$ subject to the restriction $\sum_{i=1}^n p_i = 1$, and are given by $p_i = 1/n$, $i = 1, 2, \cdots, n$, furnishing the empiric distribution function. But now if we suppose further conditions put on F, perhaps of the form $F(x_{i+1}) - F(x_i) \leqq c(x_{i+1} - x_i)$ or $p_i \leqq c(x_{i+1} - x_i)$, $i = 1, 2, \cdots, n - 1$, the remark of Section 2 may prove useful.

(ii) In the example on page 833 in [6], one is given $\{\alpha_{ij}\}$, $\{n_{ij}\}$, and required to choose $\{p_{ij}\}$ so as to minimize

$$-\sum_{i=1}^{n} \sum_{j=1}^{k} [\alpha_{ij} \log p_{ij} + (n_{ij} - \alpha_{ij}) \log (1 - p_{ij})].$$

Here $p_{ij} = 1 - F(x_i, y_j)$, where x_i, $i = 1, 2, \cdots, n$, and y_j, $j = 1, 2, \cdots, k$ are given, and where $F(x, y)$ is an unknown bivariate distribution function, so that not only is it required to be monotone in the two variables separately, but also second differences are to be positive.

(iii) Let a person chosen at random from a group have a probability U of contracting a certain disease in unit time; U is to be considered a random variable, with distribution function F. If a particular person has probability u_0, then the probability that he will be infected for the first time during a second unit of time is $(1 - u_0)u_0$, infected for the first time during a third is $(1 - u_0)^2 u_0$, etc. Thus the probability that a person chosen at random will become infected during the first unit of time is

$$p_1 = \int_0^1 u \, dF(u) = \int_0^\infty (1 - e^{-t}) \, dG(t),$$

where $G(t) = F(1 - e^{-t})$; the probability that he will first become infected during the jth unit of time is $p_j = \int_0^1 (1-u)^{j-1} u\, dF(u) = \int_0^\infty e^{-(j-1)t} (1 - e^{-t})\, dG(t)$, $j = 1, 2, \cdots$. If we set $q_j = \int_0^\infty e^{-jt}\, dG(t)$, $j = 0, 1, 2, \cdots$, then $p_j = q_{j-1} - q_j$, $j = 1, 2, \cdots$, and $q_j = 1 - \sum_{i=1}^{j} p_i$, $j = 1, 2, \cdots$. Since G is a distribution function, we have

$$\Delta_j q = q_{j+1} - q_j = -p_{j+1} \leqq 0,$$
$$\Delta_j^2 q = q_{j+2} - 2q_{j+1} + q_j = -(p_{j+2} - p_{j+1}) \geqq 0, \text{ etc.}$$

Suppose that of n persons initially chosen at random, x_j first become infected during the jth unit of time, $j = 1, 2, \cdots, k$, and that $x_{k+1} = n - \sum_{j=1}^{k} x_j$ fail to become infected during the first k units of time. The MLE's of the probabilities p_j ($j = 1, 2, \cdots, k$) and $1 - \sum_{j=1}^{k} p_j$ are the solutions $y_1, y_2, \cdots, y_k, y_{k+1}$ of the following problem: to minimize

$$-\sum_{j=1}^{k+1} x_j \log y_j,$$

subject to

$$\sum_{j=1}^{k+1} y_j = 1, \tag{3.3}$$

and

$$\begin{cases} 0 \leqq y_j \leqq 1, j = 1, 2, \cdots, k+1, \\ y_{j+1} - y_j \leqq 0, j = 1, 2, \cdots, k-1, \\ y_{j+2} - 2y_{j+1} + y_j \geqq 0, j = 1, 2, \cdots, k-2, \text{ etc.} \end{cases} \tag{3.4}$$

The problem may be made to fit precisely the pattern of Section 2 if we replace (3.3) by

$$\sum_{j=1}^{k+1} y_j \leqq 1;$$

the altered problem clearly has the same solution.

4. Exponential families. The remark to which Section 2 is devoted is especially appropriate for the problem of estimating parameters using samples from populations belonging to exponential families (cf. [2]; [4]; [17], pp. 64, 68; [24]); more particularly, when the restrictions on the parameter point are expressed by inequalities which are linear in its coordinates.

Let $F(x)$ be a distribution function. The integral

$$\phi(\tau) = \int_{-\infty}^{\infty} e^{x\tau}\, dF(x),$$

giving its moment-generating function, converges to 1 for $\tau = 0$; we shall suppose its interval of convergence contains the origin as an interior point. It then converges in a vertical strip of the complex τ plane containing the origin to an analytic function which is positive on the real axis. We set

$$\Theta(\tau) = \log \phi(\tau),$$

using the principal value of the logarithm (which is real when $\phi(\tau) > 0$, hence when τ is real); $\Theta(\tau)$ is analytic for real τ in the interval of convergence.

DEFINITION. The distribution functions $F(x; \tau)$ form an *exponential family*, the *family of exponential type determined by* $F(x)$ *or by* $\Theta(\tau)$, if, for τ in the interval of convergence,

$$F(x; \tau) = \int_{(-\infty,x)} \exp [u\tau - \Theta(\tau)] \, dF(u). \tag{4.1}$$

It was shown by Koopman [20] and by Pitman [23] that, except for change in variable or change in parameter, a (sufficiently regular) one-parameter family of distributions over a common, fixed (possibly infinite) interval admits a sufficient statistic *only* if the parameter enters as does the parameter τ in (4.1). Further, it is clear from the derivation in [11] of the Cramer-Rao inequality that in the above statement the term "sufficient" may be replaced by "efficient" (as defined in [11]).

If X_τ is a random variable whose distribution function $F(x; \tau)$ is given by (4.1), then its expectation and variance are given by

$$E(X_\tau) = \theta(\tau), \; V(X_\tau) = \theta'(\tau), \tag{4.2}$$

where

$$\theta(\tau) = \Theta'(\tau). \tag{4.3}$$

Since $V(X_\tau) \geqq 0$ it follows that $\theta(\tau)$ is increasing and $\Theta(\tau)$ convex; indeed, $\theta(\tau)$ is *strictly* increasing, and $\Theta(\tau)$ *strictly* convex unless $F(x)$ is degenerate, a possibility we shall rule out from further consideration.

We define $\tau(\theta)$ as the inverse function of $\theta(\tau)$, and $T(\theta)$ by

$$T(\theta) = \int_{\theta_0}^{\theta} \tau(v) \, dv,$$

where $\theta_0 = \theta(0)$. Evidently $T(\theta)$ is convex, and assumes its minimum value, 0, at θ_0. According to an inequality of W. H. Young ([18], p. 111), we have

$$T(x) + \Theta(y) - xy \geqq 0, \tag{4.4}$$

with equality holding if and only if $y = \tau(x)$ $(x = \theta(y))$. This becomes geometrically obvious on interpreting T and Θ relative to the graph of $y = \tau(x)$ or $x = \theta(y)$ in the xy plane.

We note that (i) a normal distribution with variable mean and fixed standard deviation, (ii) a Poisson distribution with variable mean, (iii) the distribution of the square of a normally distributed random variable having zero mean and variable variance, (iv) a binomial distribution with variable mean, and (v) a negative binomial distribution with variable parameter p, are examples of exponential families. If X_0 is any random variable whose moment generating function exists on an open interval containing the origin, there is an exponential family of distributions admitting the distribution of X_0 as a member for one parameter value. In random sampling from a population of this family, the sample mean is the MLE of $E(X_\tau)$ (cf. discussion of (4.6) below); it is also the least squares estimator; it is unbiased, consistent, sufficient, and efficient.

Let us now consider an estimation problem. Let k be a positive integer. For $i = 1, 2, \cdots, k$, consider a population whose distribution belongs to the exponential family determined by a given distribution function $F_i(x)$ for a particular parameter value τ_i, regarded as unknown. Let $z_i = (x_{1,i}, x_{2,i}, \cdots, x_{n_i,i})$ denote the set of sample values of a sample of size n_i from the ith population, and set $z = (z_1, \cdots, z_k)$. Let $\bar{x}_i$ denote the sample mean ($i = 1, 2, \cdots, k$), and let $\bar{x}$ denote the point $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ in the Euclidean space R_k of k dimensions. If E is an event in the sample space, its probability is given by

$$P_\tau(E) = \int_E \exp\left\{\sum_{i=1}^{k} n_i[\bar{x}_i \tau_i - \Theta_i(\tau_i)]\right\} dP_0(z), \tag{4.5}$$

where

$$P_0(E) = \int_E \prod_{i=1}^{k} \prod_{j=1}^{n_i} dF_i(x_{j,i}).$$

Set $\tau = (\tau_1, \cdots, \tau_k)$, $y = (y_1, \cdots, y_k)$. The MLE of τ is that point $y = \tau^*$ which maximizes $\sum_{i=1}^k n_i[\bar{x}_i y_i - \Theta_i(y_i)]$; or equivalently, which minimizes

$$G(y) \equiv \sum_{i=1}^{k} n_i[T_i(\bar{x}_i) + \Theta_i(y_i) - \bar{x}_i y_i]. \tag{4.6}$$

This function is convex in y. It is clear from inequality (4.4) and the remark following it that the unrestricted minimum is afforded by $\tau^* = (\tau_1^*, \tau_2^*, \cdots, \tau_k^*)$, where $\tau_i^* = \tau_i(\bar{x}_i)$ (the special case $k = 1$ was mentioned above). Suppose that restrictions on τ may be expressed by $\tau \in A_1 A_2 \cdots A_N$, where A_i is the closure of an open convex subset of $R_k (i = 1, 2, \cdots, N)$. We consider now the subproblem of minimizing G on a given intersection of boundaries of some of the sets A_i. Assuming the boundaries of the sets A_i sufficiently regular, if the unrestricted minimum of G is attained outside A_j, then the point $\tau^* = (\tau_1^*, \cdots, \tau_k^*)$ at which G assumes its minimum on $\mathcal{B}(A_j)$ satisfies

$$\sum_{\nu=1}^{k} \alpha_\nu^r(\tau^*) n_\nu[\theta_\nu^* - \bar{x}_\nu] = 0, r = 1, 2, \cdots, k - 1, \tag{4.7}$$

where, for $r = 1, 2, \cdots, k - 1$, $\alpha^r(\tau^*) = [\alpha_1^r(\tau^*), \cdots, \alpha_k^r(\tau^*)]$ is one of $k - 1$ independent vectors tangent at τ^* to $\mathcal{B}(A_j)$, and where $\theta_i^* = \theta_i(\tau_i^*)$. Similarly, the condition that G assume its minimum on an "edge" $\mathcal{B}(A_{i_1})\mathcal{B}(A_{i_2}) \cdots \mathcal{B}(A_{i_n})$ is (4.7) for $r = 1, 2, \cdots, k - n$, where $\alpha^r(\tau^*)$ is one of $k - n$ independent vectors tangent at τ^* to $\mathcal{B}(A_{i_1})\mathcal{B}(A_{i_2}) \cdots \mathcal{B}(A_{i_n})$. Thus the point minimizing G

on a given boundary or intersection of boundaries is a solution of equations of form (4.7). If, in particular, the boundaries $\mathcal{B}(A_i)$ are all hyperplanes, then the α_i^r are constant on a given intersection of boundaries, and values θ_i^* of the means corresponding to the coordinates τ_i^* of the minimizing point are solutions of linear equations of the form

$$\sum_{i=1}^{k} \alpha_i^r n_i[\theta_i^* - \bar{x}_i] = 0.$$

If the restricting conditions require that $\theta = (\theta_1, \cdots, \theta_k)$, rather than τ, belong to the intersection of closed convex sets, their maps in the τ-space need not in general be convex, and the above discussion need not apply. There are a number of situations of interest, however, in which the above technique for finding the minimizing point on a given intersection of boundaries will still be applicable.

(i) The function of θ, $G[\tau(\theta)]$, obtained by replacing y in (4.6) by $\tau(\theta)$, may be convex in θ. For example, this will be the case if each population is normal with known variance, or binomial, or Poisson. Since the transformation from θ-space to τ-space is 1–1 and analytic, the above discussion for finding the minimizing point in τ-space will apply even though the restricting sets in τ-space may not be convex.

(ii) All populations belong to the same exponential family, and only order restrictions are made on the parameters; that is, the regions A_i are defined by inequalities of the form $\theta_r \leqq \theta_s$. In this case $\Theta(\tau) \equiv \Theta_i(\tau)$ is independent of i, and $\tau_r \leqq \tau_s$ if and only if $\theta_r = \theta(\tau_r) \leqq \theta(\tau_s) = \theta_s$; since $\theta(\tau)$ and $\tau(\theta)$ are strictly increasing. The independent vectors α^r for a given "edge" in this case are determined by the indices i of the boundaries intersecting in the edge, independently of the particular function. *The* MLE (cf. Section 6 for a specific description in a special case) *of θ is therefore independent of the particular exponential family to which the populations belong, provided they all belong to the same exponential family, and provided only order restrictions are made on the parameters* θ_i, $i = 1, 2, \cdots, k$. In particular, for the purpose of determining the MLE's of the means, one could in such a situation assume without loss of generality that the populations are all normal with standard deviation 1, but with possibly different means, satisfying the specified order restrictions. (In the special case where the order restrictions specify a simple ordering of the means, the failure of the MLE's to depend on the particular exponential family was noted in [6] and in [7]). Thus in this situation the problem of finding the MLE reduces to that of minimizing the function

$$\sum_{i=1}^{k} n_i(\bar{x}_i - \theta_i)^2$$

subject to specified restrictions of the form $\theta_r \leqq \theta_s$. With an obvious linear change of variable, it can be expressed as the problem of finding the foot of the segment of smallest length from a given point onto a set bounded by hyperplanes passing through the origin.

5. A sufficiency property. Let us consider for a moment the simplest case of estimating a restricted parameter. We sample from a single population, belonging to an exponential family. The parameter θ is known to lie in a proper subinterval of its natural range. The MLE, $\bar{x}$, of the unrestricted parameter is known to be consistent, efficient, sufficient, and unbiased. It seems to the author that a "reasonable" estimator of the restricted parameter is $\bar{x}$, appropriately truncated, which is also the MLE. This estimator is not sufficient (nor unbiased). Likewise, in the more general situation discussed in Section 4, the MLE is not sufficient. However, it does possess a certain "sufficiency-like" property, expressed in Theorem 5.1. Referring to the general problem formulated in Section 4, we suppose that the parameter point $\tau = (\tau_1, \cdots, \tau_k)$ is subject to the restriction $\tau \varepsilon S_0 = A_1A_2, \cdots, A_N$, where now each A_j is a closed set bounded by a hyperplane. (In the event that all populations are normal with the same standard deviation, or that all populations belong to the same exponential family and the equation of the boundary of each A_j is of the form $\tau_r \leqq \tau_s$; the corresponding sets in θ-space will also be bounded by hyperplanes.) Let z denote a point of the sample space, and let $Y(z) = [Y_1(z), Y_2(z), \cdots, Y_k(z)]$ denote the corresponding MLE of τ, subject to $\tau \varepsilon S_0$. For a Borel set E in the sample space, let $p_\tau(E \mid y)$ denote the conditional probability of E for a given value y (in S_0) of $Y(z)$. That is, $p_\tau(E \mid y)$ is to be defined so that for each Borel set $B \subset S_0$ we have

$$P_\tau(E \cap Y^{-1}(B)) = \int_B p_\tau(E \mid y)\, dP_\tau Y^{-1}(y), \tag{5.1}$$

where $P_\tau(E)$ is given by (4.5) for each event E in the sample space, where $Y^{-1}(B)$ denotes the inverse image of B under the map Y from the sample space into S_0, and where $P_\tau Y^{-1}(B) = P_\tau[Y^{-1}(B)]$.

THEOREM 5.1. *Let S_0 be bounded by hyperplanes. There is a determination of $p_\tau(E \mid y)$ which is independent of τ when y is interior to S_0, and, when y lies interior to a $(k - 1)$-dimensional face or $(k - j)$-dimensional $(j = 2, 3, \cdots, k)$ edge or vertex of S_0, is independent of τ on the closure of that face, edge, or vertex.*

PROOF. For x in θ-space, define $y(x)$ by $y(x) = (\tau_1(x_1), \tau_2(x_2), \cdots, \tau_k(x_k))$. For z in the sample space, define $V(z) = y(\bar{x})$. We have $Y(z) = V(z)$ if $y(\bar{x}) \varepsilon S_0$. Define $q(E \mid y)$ to be the conditional probability of E given a value y of $V(z)$; this conditional probability may be taken to be independent of τ, since $V(z)$ is a sufficient estimator of τ. For y interior to S_0, we define

$$p_\tau(E \mid y) = q(E \mid y), \qquad \text{for all } \tau \varepsilon S_0.$$

Then if B is interior to S_0, and if $\tau \varepsilon S_0$, we have

$$P_\tau(E \cap Y^{-1}(B)) = P_\tau(E \cap V^{-1}(B)) = \int_B q(E \mid y)\, dP_\tau V^{-1}(y)$$

$$= \int_B p_\tau(E \mid y)\, dP_\tau Y^{-1}(y).$$

Now suppose y is on a $(k-1)$-dimensional face or $(k-j)$-dimensional ($j = 2, 3, \cdots, k$) edge, W, of S_0, which is open in its relative topology. For τ not on the closure, W^{cl}, of W, let $p_\tau(E \mid y)$ denote any determination of the conditional probability satisfying (5.1). Choose a fixed $\beta \in W$, and let $p_\beta(E \mid y)$ denote any determination of the conditional probability satisfying (5.1). For τ on W^{cl}, define $p_\tau(E \mid y)$ to be equal to $p_\beta(E \mid y)$. We now wish to verify that p so defined satisfies (5.1) when B is a Borel subset of W. For such B we have, by definition,

$$P_\beta[E \cap Y^{-1}(B)] = \int_{E \cap Y^{-1}(B)} \exp\left\{\sum_{i=1}^{k} n_i[\bar{x}_i \beta_i - \Theta_i(\beta_i)]\right\} dP_0(z)$$

$$= \int_B p_\beta(E \mid y)\, dP_\beta Y^{-1}(y).$$

Also

$$P_\tau[E \cap Y^{-1}(B)] = \int_{E \cap Y^{-1}(B)} \exp\left\{\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) - n_i[\Theta_i(\tau_i) - \Theta_i(\beta_i)]\right\}\cdot$$

$$\exp\left\{\sum_{i=1}^{k} n_i[\bar{x}_i \beta_i - \Theta_i(\beta_i)]\right\} dP_0(z).$$

If the MLE, Y, of τ is in W, then, by (4.7),

$$\sum_{i=1}^{k} \alpha_i^r n_i \bar{x}_i = \sum_{i=1}^{k} \alpha_i^r n_i \theta_i(Y_i),$$

where the α^r are independent vectors spanning W. If $\tau \in W^{cl}$, then $\tau - \beta$ is a linear combination of the α^r; hence

$$\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) = \sum_{i=1}^{k} n_i \theta_i(Y_i)(\tau_i - \beta_i),$$

a function of Y for fixed τ, β. So also, then, is

$$\exp\left\{\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) - [\Theta_i(\tau_i) - \Theta_i(\beta_i)]\right\}$$

a function, $\psi[Y(z)]$, of $Y(z)$. We have then

$$P_\tau[E \cap Y^{-1}(B)] = \int_{E \cap Y^{-1}(B)} \psi[Y(z)]\, dP_\beta(z)$$

$$= \int_B \psi(y) p_\beta(E \mid y)\, dP_\beta Y^{-1}(y)$$

$$= \int_B p_\beta(E \mid y)\, dP_\tau Y^{-1}(y),$$

since

$$dP_\tau Y^{-1}(y) = \psi(y)\, dP_\beta Y^{-1}(y).$$

One now verifies from the appropriate definitions above that (5.1) holds for arbitrary $\tau \varepsilon S_0$, and Borel set $B \subset S_0$. This completes the proof of Theorem 5.1.

Theorem 5.1 may be regarded as a generalization of a remark ([16], p. 77) to the effect that if X and Y are normally distributed random variables with unit standard deviation and means ξ and η respectively, and if ξ and η are known to satisfy a linear equation, then the foot of the perpendicular from the observation point (x, y) is a sufficient estimator.

Theorem 5.1 may be interpreted somewhat as follows. Given the value of $Y(z)$, the exact knowledge of the observed sample point would imply no additional information as to how to select τ on the face (or edge) on which $Y(z)$ lies, since the conditional distribution, given $Y(z)$, is independent of τ on this face.

6. Uniform consistency of a class of estimators. In Section 4, an estimation problem of the following kind was considered. Let k be a positive integer. To each positive integer $i \leqq k$ corresponds a population whose distribution is known except for the unknown value of its mean, θ_i. The means θ_i are known to satisfy certain inequalities. The problem of estimating a distribution function from all-or-none data (bioassay) is of this kind, in which the populations are binomial and the inequalities are of the form $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k$ (cf. Section 3; also [1], [12]). Even if the populations are not binomial, but all belong to a common exponential family, the MLE's subject to $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k$ are very easily determined, as follows (cf [1], [7]). Let $\bar{x}_i$ denote the sample mean of a sample of size n_i from the i-th population, whose mean is θ_i, $i = 1, 2, \cdots, k$. If $\bar{x}_1 \leqq \bar{x}_2 \leqq \cdots \leqq \bar{x}_k$, these are the MLE's of the parameters θ_i, $i = 1, 2, \cdots, k$. If for some i we have $\bar{x}_i > \bar{x}_{i+1}$, these two means are replaced by the single ratio $(n_i\bar{x}_i + n_{i+1}\bar{x}_{i+1}) / (n_i + n_{i+1})$, obtaining an ordered set of only $k. - 1$ ratios $(k - 2$ of which are sample means). This procedure is repeated until an ordered set of ratios is obtained which are monotone non-decreasing. Then for each i, the MLE, $\hat{\theta}_i$, of θ_i is equal to that one of the final set of ratios to which the original ratio $\bar{x}_i$ contributed.

If the number, k, of observation points is held fixed, while the number of observations at each point increases indefinitely, classical theory assures the strong consistency of the $\hat{\theta}$, and yields their asymptotic distribution; the $\hat{\theta}$ will asymptotically coincide with the sample means. We shall be interested here chiefly in situations in which there are a large number of observation points, but only a few observations, perhaps only one, at each. In [1] and in [7] the *local consistency* of the MLE's is proved. It is assumed that there is an unknown function $\theta(t)$ (as in bioassay, for example), known to be non-decreasing and continuous, such that $\theta_i = \theta(t_i)$, $i = 1, 2, \cdots, k$. Then if t is held fixed, one can achieve an arbitrarily high probability of an arbitrarily great precision at t by selecting enough observation points in the neighborhood of t, even if only one observation is made at each. In [1] and [7] it was assumed that the populations all belonged to the same exponential family; but it is clear that the estimators $\hat{\theta}$ can be formed without regard to the distributions of the k populations; they are determined by the sample means alone (of course, they will not in general be MLE's). Indeed, the proof of the local consistency of the estimators $\hat{\theta}$ does not require an assumption that the populations belong to an exponential family.

Theorem 6.2 below gives conditions sufficient for the *strong uniform consistency* of the estimators $\hat{\theta}$, *without assuming the populations belong to an exponential family.* The proof requires a somewhat strengthened form of the strong law of large numbers, which is presented in Theorem 6.1.

THEOREM 6.1. *Let r be a fixed positive number. Let $Y_1, Y_2, \cdots$, be independent random variables with $E(Y_i) = 0$, $E(|Y_i|^{2r}) < \infty$, and*

$$\sum_i E(|Y_i|^{2r})/i^{r+1} < \infty. \tag{6.1}$$

Corresponding to each positive integer $n \geqq 2$, let $i_{1,n}, i_{2,n}, \cdots, i_{n,n}$ be a permutation of the positive integers $1, 2, \cdots, n$, obtained by assigning a place to the integer n between some two successive integers, or at the beginning, or at the end, of the permutation corresponding to the integer $n - 1$. Define $S_{j,n} = \sum_{\nu=1}^{j} Y_{i_{\nu,n}}$, $j = 1, 2, \cdots, n$. Then

$$\Pr\left\{\lim_{n\to\infty} \max_{j=1,2,\cdots,n} \frac{1}{n} | S_{j,n} | = 0\right\} = 1.$$

INDICATION OF PROOF OF THEOREM 6.1. The situation is more complicated than that of the classical strong law, but familiar arguments suffice. For $\nu = 0$, $1, \cdots$, arrange the terms Y_i having indices i such that $2^{\nu-1} < i < 2^{\nu}$ in the order given by the permutation for 2^{ν}, and let $\mathfrak{Z}(\nu)$ denote the family of partial sums containing the first of these terms, the sum of the first two, the sum of the first three, etc. Now consider partial sums $S_{j,n}$, $j \leqq n$. For each n, choose $k = k(n)$ so that $2^{k-1} < n \leqq 2^k$. To avoid complicated subscripts, let $p = p(n) = 2^{k-1}$. Let $Z_1, Z_2, \cdots, Z_{2p-n}$ denote the random variables $Y_{n+1}, Y_{n+2}, \cdots, Y_{2p}$ written in the order given by the permutation for $2p = 2^k$. Let $\mathfrak{U}(n)$ denote the family of partial sums: $\{Z_1, Z_1 + Z_2, \cdots, Z_1 + Z_2 + \cdots + Z_{2p-n}\}$. For fixed j, n, and for $\nu = 0, 1, 2, \cdots, k - 1$, let $T_\nu = T_\nu(j, n)$ denote the sum of terms Y_i which appear in the sum $S_{j,n}$ and which have indices i such that $2^{\nu-1} < i \leqq 2^{\nu}$. Then $T_\nu \varepsilon \mathfrak{Z}(\nu)$ for $\nu = 0, 1, 2, \cdots, k - 1$. Let $T_k = T_k(j, n)$ denote the minimal member of $\mathfrak{Z}(k)$ containing all terms appearing in $S_{j,n}$ whose indices i satisfy $2^{k-1} < i \leqq 2^k$ (minimal in the sense of containing the fewest possible terms). Let $U = U(j, n)$ be the sum of terms appearing in T_k of index greater than n; then $U \varepsilon \mathfrak{U}(n)$, and $S_{j,n} = \sum_{\nu=0}^{k} T_\nu - U$. Let $\mathfrak{V}(k)$ denote the family of all sums of the form $\sum_{\nu=0}^{k} W_\nu$, where $W_\nu \varepsilon \mathfrak{Z}(\nu)$, $\nu = 0, 1, 2, \cdots, k$. Let $V = V(j, n) = \sum_{\nu=0}^{k} T_\nu$. Then $V \varepsilon \mathfrak{V}(k)$ and

$$S_{j,n} = V - U.$$

Let ϵ be positive. Let A_n denote the event: $\{\max_{0\leqq j\leqq n}|S_{j,n}| > 2^{k+1}\epsilon\}$, B_n the event: $\{\max_{0\leqq j\leqq n} | U | \leqq 2^k\epsilon\}$, and C_k the event: $\{\max_{V\varepsilon\mathfrak{V}(k)} | V | > 2^k\epsilon\}$. Then

$$A_nB_n \subset C_k \qquad (k = k(n)). \tag{6.2}$$

It follows from Chung's inequality ([10], p. 348) and the generalized Kolmogorov inequality ([21], p. 265), that

$$P(B_n) > 1 - A\left[\sum_{i=p+1}^{2p} E(|Y_i|^{2r})\right] \Big/ (2^k\epsilon)^{r+1},$$

where A is a constant depending only on r. From hypothesis (6.1), using Kronecker's lemma, we conclude that

$$\lim_{k\to\infty}\left[\sum_{i=p+1}^{\infty} E(|Y_i|^{2r})\right] \Big/ (2^k\epsilon)^{r+1} = 0,$$

hence there is a positive integer k_0 such that $P(B_n) > \frac{1}{2}$ for $n > p_0 = 2^{k_0-1}$ Further, A_{p_0+1} and B_{p_0+1} are independent, and if A^c denotes the complement of A, then for $n = p_0 + 2, p_0 + 3, \cdots$, we have that $A^c_{p_0+1} \cap A^c_{p_0+2} \cap \cdots A^c_{n-1} \cap A_n$ and B_n are independent. It follows from the "Lemma for Events", [21], p. 246, that

$$P(\cup_{n=p_0+1}^{\infty} A_n B_n) \geqq \tfrac{1}{2} P(\cup_{n=p_0+1}^{\infty} A_n),$$

so that from (6.2) we have

$$P(\cup_{n=p_0+1}^{\infty} A_n) \leqq 2P(\cup_{\nu=k_0}^{n} C_\nu),$$

hence

$$P(\limsup_{n\to\infty} A_n) \leqq 2\, P(\limsup_{\nu\to\infty} C_\nu),$$

or

$$\Pr\{\max_{0\leqq j\leqq n} |S_{j,n}| > 2^{k+1}\epsilon \quad \text{for infinitely many } n\}$$

$$\leqq 2 \Pr\{\max_{V\epsilon\mathfrak{V}(\nu)} |V| > 2^\nu\epsilon \quad \text{for infinitely many } \nu\}.$$

Kolmogorov's method ([19], cf. also [25], p. 202), with Chung's inequality and the generalized Kolmogorov inequality can be used to show that the right hand member is 0. Since $2^{k+1} < 4n(k = k(n), 2^{k-1} < n \leqq 2^k)$, we have

$$\Pr\left\{\max_{0\leqq j\leqq n} \frac{1}{n}\left|S_{j,n}\right| > 4\epsilon \quad \text{for infinitely many } n\right\} = 0.$$

A standard argument completes the proof.

We return now to the estimation problem. For $i = 1, 2, \cdots, k$, $\bar{x}_i$ is the sample mean of a sample of size n_i from a population whose mean is $\theta(t_i)$. It is known that $\theta(t)$ is non-decreasing. We are concerned with the estimator $\hat{\theta}(t)$ obtained as described above. It is given ([1], [7]) by

$$(6.3)\quad \begin{cases} \hat{\theta}(t) = \max\limits_{t_r\leqq t}\min\limits_{t_s\geqq t}\left(\sum\limits_{\nu=r}^{s} n_\nu\bar{x}_\nu\right)\Big/\left(\sum\limits_{\nu=r}^{s} n_\nu\right), \\ \qquad = \min\limits_{t_s\geqq t}\max\limits_{t_r\leqq t}\left(\sum\limits_{\nu=r}^{s} n_\nu\bar{x}_\nu\right)\Big/\left(\sum\limits_{\nu=r}^{s} n_\nu\right). \end{cases}$$

THEOREM 6.2. *Let $\theta(t)$ be continuous and non-decreasing on (a, b). Let $\{s_n\}$ be a sequence of observation points dense in (a, b). Let one observation be made at each point (the observation points need not be distinct). Let the variances of the observed random variables be bounded. Let $\hat{\theta}_n(t)$ denote the estimate of $\theta(t)$ based on observations made at the first n observation points, defined to be constant between observation points, and continuous from the left. If $c > a$, $d < b$, then*

$$\Pr\{\lim_{n\to\infty} \max_{c \leqq t \leqq d} | \hat{\theta}_n(t) - \theta(t) | = 0\} = 1.$$

PROOF. The original proof used Theorem 6.1 and a geometrical interpretation of $\hat{\theta}$ due to W. T. Reid [5] which is also used in the proof of Theorem 6.3. It required as additional hypothesis that the norm (maximum distance between adjacent points of subdivision) of the subdivision of (a, b) formed by the first n observation points be $O(1/n)$, and required the less restrictive hypothesis (6.1) on the variances of the observable random variables. The present proof uses an approach suggested by the referee. This proof also could be modified to use the hypothesis (6.1) on the variances instead of boundedness, together with a uniformity condition on the distribution of the observation points, but the above formulation appears more natural and useful.

We observe first that if $\hat{\theta}_n(u_i) - \theta(u_i) \to 0$ for each u_i of a sequence $\{u_i\}$ dense in (a, b), then it follows from the monotonicity of $\hat{\theta}_n$ and the continuity of θ that $\max_{c \leqq t \leqq d} |\hat{\theta}_n(t) - \theta(t)| \to 0$. Consequently it suffices to show that, for each individual $t \in (a, b)$, $\Pr\{\hat{\theta}_n(t) - \theta(t) \to 0\} = 1$, since it then follows that $\Pr\{\hat{\theta}_n(u_i) - \theta(u_i) \to 0 \text{ for all } u_i\} = 1$, if $\{u_i\}$ is any countable sequence of points in (a, b).

We now prove that for fixed $t \in (a, b)$, $\Pr\{\hat{\theta}_n(t) - \theta(t) \to 0\} = 1$. It suffices to prove that for every $\epsilon > 0$ we have

$$\Pr\{\liminf_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \geqq -\epsilon\} = 1 \tag{6.4}$$

and

$$\Pr\{\limsup_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \leqq \epsilon\} = 1. \tag{6.5}$$

We prove the first; the proof of the second is similar.

We suppose the sequence $\{s_i\}$ of observation points chosen, not necessarily distinct nor ordered according to increasing index, and an observation Z_i made at each, so that $E(Z_i) = \theta(s_i)$. Let $\sigma_i^2 = V(Z_i)$, the variance of the random variable Z_i observable at s_i. For fixed n, let $t_1, t_2, \cdots, t_k$ denote the $k = k(n)$ distinct observation points among $s_1, s_2, \cdots, s_n$, arranged in increasing order, and let n_i denote the number of observations made at t_i, so that $\sum_{i=1}^k n_i = n$.

Let $t \in (a, b)$. Given $\epsilon > 0$, choose n sufficiently large that there is a $t_r < t$ such that $|\theta(t_r) - \theta(t)| < \epsilon$. By (6.3),

$$\hat{\theta}_n(t) - \theta(t) \geqq \min_{t_s \geqq t} \frac{\sum_{\nu=r}^s n_\nu[\bar{x}_\nu - \theta(t_r)]}{\sum_{\nu=r}^s n_\nu} - [\theta(t) - \theta(t_r)].$$

Since θ is non-decreasing, we have $\theta(t_r) \leqq \theta(t_\nu)$ for $\nu \geqq r$, hence

$$\hat{\theta}_n(t) - \theta(t) > \min_{t_s \geqq t} \frac{\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)]}{\sum_{\nu=r}^{s} n_\nu} - \epsilon. \tag{6.6}$$

For $\rho = 1, 2, \cdots$, let s_{i_ρ} denote the ρth of the members of the sequence $\{s_i\}$ which lie at, or to the right of, t_r. Consider the sequence of observable random variables, centered at means, $Z_{i_\rho} - \theta(s_{i_\rho})$. The sums $\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)]$ are *not* successive partial sums of this sequence or of any sequence, since as ρ increases new observation points are interspersed among the old. However, in applying Theorem 6.1 with $Y_\rho = Z_{i_\rho} - \theta(s_{i_\rho})$, we find that the ratios $\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)] / \sum_{\nu=r}^{s} n_\nu$ are just such ratios $S_{j,n} / n$ as are considered there. We conclude from (6.6) that $\Pr\{\liminf_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \geqq -\epsilon\} = 1$. A similar argument shows that for $\epsilon > 0$,

$$\Pr\{\limsup_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \leqq \epsilon\} = 1,$$

whence

$$\Pr\{\lim_{n\to\infty} \hat{\theta}_n(t) - \theta(t) = 0\} = 1.$$

Together with the earlier remarks, this completes the proof of Theorem 6.2.

Theorem 6.3, below, gives an asymptotic lower bound for the probability of achieving a given uniform precision on a closed subinterval of (a, b).

THEOREM 6.3. *For a fixed positive integer n, let n observations be made at observation points $t_1 \leqq t_2 \leqq \cdots \leqq t_k$ in (a, b), n_i observations being made at t_i, $i = 1, 2, \cdots, k$, so that $n = \sum_{i=1}^{k} n_i$. Let $\Delta = \max_{i=0,1,2,\cdots,k}(t_{i+1} - t_i)$; $t_0 = a$, $t_{k+1} = b$. Let the populations be such as to permit the application of the Central Limit Theorem as required in* [15] (cf. *also* [9]; *for an appropriate Lindeberg condition, see* [22], p. 127). *Let $\theta(t)$ have a bounded derivative, $|\theta'(t)| \leqq K$, $K > 0$, for $t \in (a, b)$, and let $\sigma^2 = \sum_{i=1}^{k} n_i\sigma_i^2$, where σ_i^2 is the variance of an observation made at t_i, $i = 1, 2, \cdots, k$. For $z > 0$, let $h = [2z\sigma\Delta / K]^{\frac{1}{2}}$, $c = a + h$, $d = b - h$. Then*

$$\Pr\{\max_{c \leqq t \leqq d} |\hat{\theta}(t) - \theta(t)| < 2\sqrt{2Kz\sigma\Delta}\} \ \dot{\geqq}\ \frac{4}{\pi}\sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{2\nu + 1} \exp[-(2\nu + 1)^2\pi^2/8z^2].$$

The symbol "$\dot{\geqq}$" is to be interpreted as "asymptotically (as $n \to \infty$) at least as large as". The estimate is most nearly accurate if only one observation is made at each point, if the observation points are distributed uniformly over (a, b), and if $\theta'(t)$ is constant.

PROOF. Order the observations according to increasing t, ordering in an arbitrary way those occurring at the same observation point. Let $Z_{\nu,n}$ denote the νth observation, $\nu = 1, 2, \cdots, n$; its mean is $\theta(t_j)$ and its variance σ_j^2 if it is made at the observation point t_j. For positive integers $j \leqq k$, define $N_j = \sum_{t_\nu \leqq t_j} n_\nu$, $s(N_j) = \sum_{t_\nu \leqq t_j} n_\nu\theta(t_\nu)$, and $s^*(N_j) = \sum_{t_\nu \leqq t_j} n_\nu\bar{x}_\nu$. We have $s^*(N_j)$ as one of the partial sums of the sequence $Z_{\nu,n}$, and $s(N_j)$ as its expectation. If $S_{\nu,n}$ denotes the νth partial sum, and s_ν^2 its variance, then it is known that

$$\lim_{n\to\infty} \Pr\{\max_{\nu\leqq n} |S_{\nu,n} - E(S_{\nu,n})| < zs_n\} = \frac{4}{\pi}\sum_{i=0}^{\infty}\frac{(-1)^i}{2i+1}\exp[-(2i+1)^2\pi^2/8z^2]$$

(strictly speaking, we require that the theorem as developed in [15] and [9] be generalized so as to apply to sums of the form $\sum_{\nu=1}^{n} X_{\nu,n}$, where the $X_{\nu,n}$ are independent for distinct ν, rather than to sums of the form $\sum_{\nu=1}^{n} X_\nu$; but only trivial modifications are required in the proofs). Define $s(u)$ and $s^*(u)$ to be linear between successive integers; then

$$(6.7)\quad \Pr\{\max_{0\leqq u\leqq n} |s^*(u) - s(u)| \leqq z\sigma\} \dot{\geqq} \frac{4}{\pi}\sum_{i=0}^{\infty}\frac{(-1)^i}{2i+1}\exp[-(2i+1)^2\pi^2/8z^2].$$

We observe that $s(u)$ is a convex function whose graph consists of line segments: for $N_{i-1} < u < N_i$ we have $s'(u) = \theta(t_i)$, $i = 1, 2, \cdots, k$. The graph of $s^*(u)$ also consists of line segments, but it need not be convex, since $\bar{x}_i$ need not increase with i.

Let $g(u)$ denote the greatest convex function not greater than $s^*(u)$; the graph of this function consists of line segments. We denote by $g'(u)$ ($s'(u)$) the derivative of $g(u)$ ($s(u)$) where it is defined, and the left-hand limit of the derivative at a corner. One verifies from formulas (6.3) that $\hat{\theta}(t_i) = g'(N_i)$, $i = 1, 2, \cdots, k$. Now let u' be fixed, so that $u' \leqq \sum_{t_\nu\leqq d} n_\nu$ Suppose $\max_{0\leqq u\leqq n} |s^*(u) - s(u)| < z\sigma$. Then for $u \geqq u'$ we have

$$g(u) \leqq s^*(u) < s(u) + z\sigma.$$

Since the point $(u', g(u'))$ is on a line segment whose endpoints are at vertical distance less than $z\sigma$ from the graph of s (or else it is itself such an endpoint), and since s is convex, we have also

$$g(u') > s(u') - z\sigma.$$

Hence $g(u) - g(u') < s(u) - s(u') + 2z\sigma$. Therefore

$$g'(u') \leqq \frac{g(u) - g(u')}{u - u'} < \frac{s(u) - s(u')}{u - u'} + \frac{2z\sigma}{u - u'}.$$

Choose i, j so that $N_{i-1} < u' \leqq N_i$, $N_{j-1} < u' \leqq N_j$. We have $[s(u) - s(u')] / (u - u') \leqq s'(u) = \theta(t_j) \leqq \theta(t_i) + K(t_j - t_i) = s'(u') + K(t_j - t_i)$. But $t_j - t_i \leqq (N_{j-1} - N_i)\,\Delta + \Delta \leqq (u - u' + 1)\Delta$, so that

$$g'(u') < s'(u') + K(u - u' + 1)\Delta + 2z\sigma / (u - u')$$

for $u' < u \leqq n$. We choose $u = u' + [2z\sigma / K\Delta]^{\frac{1}{2}}$, and find that $g'(u') - s'(u') < 2[2Kz\sigma\Delta]^{\frac{1}{2}} + K\Delta \doteq 2[2Kz\sigma\Delta]^{\frac{1}{2}}$. Similarly, $g'(u') - s'(u') \dot{>} -2[2Kz\sigma\Delta]^{\frac{1}{2}}$, if $u' \geqq \sum_{t_\nu\leqq c} n_\nu$. Since $\hat{\theta}(t_i) = g'(N_i)$ and $\theta(t_i) = s'(N_i)$, $i = 1, 2, \cdots, k$, we have

$$\max_{c\leqq t\leqq d} |\hat{\theta}(t) - \theta(t)| \dot{<} 2[2Kz\sigma\Delta]^{\frac{1}{2}}$$

if

$$\max_{0 \leqq u \leqq n} |s^*(u) - s(u)| < z\sigma.$$

The conclusion of the theorem follows from (6.7).

If $K = 0$, or if we wish a lower bound on the probability for uniform precision over a larger subinterval $[c, d]$, we must simply take u in the above discussion equal to $u' + h / \Delta$, where $h = \max [b - d, c - a]$, obtaining

$$|g'(u') - s'(u')| \leqq K(h + \Delta) + 2z\sigma\Delta / h,$$

or

$$\max_{c \leqq t \leqq d} | \hat{\theta}(t) - \theta(t) | \leqq K(h + \Delta) + 2z\sigma\Delta/h.$$

To get an idea of the rate of convergence guaranteed with at least a certain probability, suppose $\theta(t) \equiv t$ on $(0, 1)$, and that $\Delta = 1 / (n + 1)$, one binomial observation being made at each observation point $i / (n + 1)$, $i = 1, 2, \cdots, n$. We find $\sigma^2 \doteq n / 6$, $K = 1$, $h = (2z^2 / 3n)^{1/4}$, and

$$\Pr\{\max_{c \leqq t \leqq d} | \hat{\theta}(t) - \theta(t) | < 2(2z^2/3n)^{1/4}\} \gtrapprox \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{2\nu + 1} \exp\left[-\frac{(2\nu + 1)^2\pi^2}{8z^2} \right],$$

which suggests that the minimum precision (reciprocal of error) assured with a given probability increases like $n^{1/4}$. On the other hand, if the observations are concentrated near a given point, Theorem 3.1 of [1] suggests that the precision at that point increases like $n^{1/3}$.

REFERENCES

[1] MIRIAM AYER, H. D. BRUNK, G. M. EWING, W. T. REID, AND EDWARD SILVERMAN, "An empirical distribution function for sampling with incomplete information", *Ann. Math. Stat.*, Vol. 26 (1955), pp. 641–647.

[2] D. E. BARTON, "A class of distributions for which the maximum likelihood estimate is unbiased and of minimum variance for all sample sizes", *Biometrika*, Vol. 43(1956), pp. 200–202.

[3] E. M. L. BEALE, "On minimizing a convex function subject to linear inequalities", *J. Roy. Stat. Soc.*, Ser. B, Vol. 17(1955), pp. 173–184.

[4] DAVID BLACKWELL AND M. A. GIRSHICK. "Theory of games and statistical decisions," *John Wiley and Sons, Inc.*, 1954.

[5] H. D. BRUNK, G. M. EWING, AND W. T. REID, "The minimum of a certain definite integral suggested by the maximum likelihood estimate of a distribution function", (Abstract) *Bull. Amer. Math. Soc.*, Vol. 60 (1954), p. 535.

[6] H. D. BRUNK, G. M. EWING, AND W. R. UTZ, "Minimizing integrals in certain classes of monotone functions", *Pacific J. Math.*, Vol. 7(1957), pp. 833–847.

[7] H. D. BRUNK, "Maximum likelihood estimates of monotone parameters", *Ann. Math. Stat.*, Vol. 26(1955), pp. 607–616.

[8] H. D. BRUNK, "The strong law of large numbers", *Duke Math. J.*, Vol. 15(1948), pp. 181–195.

[9] K. L. CHUNG, "On the maximum partial sums of sequences of independent random variables", *Trans. Amer. Math. Soc.*, Vol. 64 (1948), pp. 205–233.

[10] K. L. Chung, "The strong law of large numbers", *Proc. Second Berkeley Symposium on Math. Stat. and Prob.*, Univ. of California Press, 1951, pp. 341–352.

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

[12] W. F. Dixon and F. J. Massey, "Introduction to Statistical Analysis", *McGraw-Hill*, 1951.

[13] Constance van Eeden, "Maximum likelihood estimation of ordered probabilities", *Proc. Koninkl. Nederl. Akad. van Wetenschappen, Amsterdam*, Ser. A, Vol. 59, and *Indag. Math.*, Vol. 18 (1956), pp. 444–455.

[14] Constance van Eeden, "Maximum likelihood estimation of partially or completely ordered parameters", *Proc. Koninkl. Nederl. Akad. van Wetenschappen, Amsterdam*, Ser. A, Vol. 60 and *Indag. Math.*, Vol. 19 (1957), pp. 128–211.

[15] P. Erdös and M. Kac, "On certain limit theorems of the theory of probability", *Bull. Amer. Math. Soc.*, Vol. 52 (1946), pp. 292–302.

[16] R. A. Fisher, "Statistical methods and scientific induction", *J. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955), pp. 69–78.

[17] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions", *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.

[18] G. H. Hardy, J. E. Littlewood, and G. Pólya. "Inequalities," *Cambridge University Press*, Second edition, 1952.

[19] A. Kolmogorov, "Sur la loi forte des grands nombres", *Comptes Rendus Paris*, Vol. 191 (1930), pp. 910–911.

[20] B. O. Koopman, "On distributions admitting a sufficient statistic", *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.

[21] M. Loève, "Probability Theory", *D. van Nostrand Co.*, 1955.

[22] W. G. Madow, "Limiting distributions of quadratic and bilinear forms", *Ann. Math. Stat.*, Vol. 11 (1940), pp. 125–146.

[23] E. J. G. Pitman, "Sufficient statistics and intrinsic accuracy", *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 567–579.

[24] M. C. K. Tweedie, "Functions of a statistical variate with given means with special reference to Laplacian distributions", *Proc. Camb. Phil. Soc.*, Vol. 43 (1947), pp. 41–49.

[25] J. V. Uspensky, "Introduction to Mathematical Probability", *McGraw-Hill*, New York, 1937.