

STATISTICAL ESTIMATION*

BY

J. L. DOOB

Let $\{x(p)\}$ be a family of chance variables depending on the parameter p . Suppose that it is known that a certain chance variable x belongs to the family $\{x(p)\}$, and suppose that a sample of n values of x : x_1, \dots, x_n , has been obtained. The problem of estimation is to find a function $p(x_1, \dots, x_n)$ which in some sense is a suitable estimate of the value of p for which $x(p) \equiv x$. Any function $p(x_1, \dots, x_n)$ will be called an estimate. If $\{p_n(x_1, \dots, x_n)\}$ are estimates of p for samples of 1, 2, \dots , the sequence of estimates will be called a statistic.[†] A statistic $\{p_n(x_1, \dots, x_n)\}$ is called consistent if the probability (based on the value p_0 of p) that $|p_n(x_1, \dots, x_n) - p_0| < \epsilon$ approaches 1 as n becomes infinite for every positive ϵ and for every value p_0 of the parameter in the range considered. It will frequently be true that $\lim_{n \rightarrow \infty} p_n = p_0$ with probability 1 for every p_0 in the range.

It is evident that if there is one consistent statistic there are infinitely many, since if N is any positive integer, the consistency of a statistic is independent of its first N estimates. This shows a fact, sometimes not made clear, that the problem of estimation from a sample of n for n fixed throughout the discussion has nothing whatsoever to do with the idea of consistency. In the problem of estimation from a sample of n , what is desired is an estimate $p_n(x_1, \dots, x_n)$ which is close to the value p_0 of the parameter which determines the probabilities, and this closeness should depend as little as possible on p_0 . In this paper ideas connected with the estimates given by the method of maximum likelihood, developed by R. A. Fisher, will be examined in detail.

If $f(x_1, \dots, x_n)$ is non-negative and integrable over r -dimensional space,

$$\int \cdots \int_E f(x_1, \dots, x_r) dx_1 \cdots dx_r$$

can be considered as the probability that a sample (x_1, \dots, x_r) of a chance variable (whose range is the set of points of r -dimensional space) is in the set E , and $f(x_1, \dots, x_r)$ will be called a probability density in r dimensions. The value of the above integral will be called the probability of E . We can

* Presented to the Society, December 30, 1934; received by the editors April 4, 1935.

Research under a grant-in-aid from the Carnegie Corporation.

[†] This differentiation between the terms estimate and statistic is not customary, but allows the usual language to be preserved while making the discussion more precise.

suppose that f is defined at every point of r -dimensional space, defining f as 0 where it was undefined originally.* Let $f(x_1, \dots, x_r; p)$ be such a probability density in r dimensions, depending on the parameter p . The set of these densities for all values of p in the range considered will be called a family of probability densities. If $f(x_1, \dots, x_r; p)$ has the property that there is a domain D in the r -dimensional space of (x_1, \dots, x_r) with the property that, for each value of p , $f > 0$ on D except possibly for a set of points of Lebesgue measure 0 and $f = 0$ on the complement of D except possibly for a set of points of Lebesgue measure 0, the family will be said to have the property D . Let $r = 1$. If $p_n(x_1, \dots, x_n)$ is defined as a value of p which maximizes $\prod_{j=1}^n f(x_j; p)$ if such a value exists, p_n is called the n th maximum likelihood estimate of p . The statistic $\{p_n\}$ is called the maximum likelihood statistic. If $r > 1$, the definition is the same except that x_j becomes a complex of r numbers.

It has been shown that maximum likelihood statistics $\{p_n\}$ are consistent in a very wide class of cases and that in a somewhat less wide class the distribution of $n^{1/2}(p_n - p)$ as n becomes infinite approaches the normal distribution with mean 0 and variance σ^2 , where

$$\begin{aligned} \frac{1}{\sigma^2} &= - \int \dots \int f \frac{\partial^2 \log f}{\partial p^2} dx_1 \dots dx_r \\ (1) \quad &= \int \dots \int \left[\frac{\partial \log f}{\partial p} \right]^2 f dx_1 \dots dx_r. \dagger \end{aligned}$$

In the case of a family of discrete-valued chance variables: when there is an integer N such that $x(p)$ takes on the values $1, \dots, N$ with probabilities $f(1; p), \dots, f(N; p)$, and takes on no other values, the maximum likelihood estimates are obtained, as above, by maximizing $\prod_{j=1}^n f(x_j; p)$, where the x 's are now positive integers less than $N + 1$. The distribution of $n^{1/2}(p_n - p)$ as n becomes infinite approaches the normal distribution with mean 0 and variance σ^2 where $1/\sigma^2$ is the negative of the expectation of $\partial^2 \log f / \partial p^2$ or the expectation of $[\partial \log f / \partial p]^2$.

* Quotients with f in the denominator will frequently appear in the formulas. These quotients are defined as 0 when $f = 0$. Integration will always be over all space when the limits of integration are not stated explicitly.

† J. L. Doob, these Transactions, vol. 36 (1934), pp. 766-775. A proof given by H. Hotelling, *ibid.*, vol. 32 (1930), pp. 847-859, holds for discrete-valued chance variables. The study of maximum likelihood statistics was started by R. A. Fisher in the Philosophical Transactions of the Royal Society of London, (A), vol. 222 (1921), pp. 309-368, and continued in the Proceedings of the Cambridge Philosophical Society, vol. 22 (1925), pp. 700-725, and in the Proceedings of the Royal Society, (A), vol. 144 (1934), pp. 285-307. The work of Fisher's second paper (which will be referred to as Fisher II) is dealt with systematically in this paper, and most of the results in this paper were stated in a general way by Fisher.

These and other facts suggested to R. A. Fisher that in each case the quantity $1/\sigma^2$ be called the amount of information in the original family of chance variables, or the amount of information in a sample of 1, pertinent to estimating p . Since in the continuous case the information is obtained from values of x_1, \dots, x_r and in the discrete case from values of x (which can only take on one of N values), the amount of information $1/\sigma^2$ in the two cases will be denoted by $I(x_1, \dots, x_r)$, $I(x)$, respectively. R. A. Fisher showed* using the first integral in (1) that $I(x_1, \dots, x_r; y_1, \dots, y_r)$: the amount of information in a sample of 2, i.e., the amount of information in a sample of 1 from the distribution with density

$$f(x_1, \dots, x_r; p) \cdot f(x_{r+1}, \dots, x_{2r}; p)$$

in $2r$ dimensions, is $2I(x_1, \dots, x_r)$, and more generally that the amount of information in a sample of s is $sI(x_1, \dots, x_r)$. The corresponding theorem holds for the discrete case. The quantity $I(x_1, \dots, x_r)$ is a function of p , but this function can be made a constant by a suitable transformation to a new parameter.

THEOREM 1. *Let $\{f(x_1, \dots, x_r; p)\}$ be a family of probability densities with the property D. We suppose that $f_p = \partial f(x_1, \dots, x_r; p)/\partial p$ exists, that*

$$(2) \quad I(x_1, \dots, x_r) = \int \dots \int \frac{f_p^2}{f} dx_1 \dots dx_r$$

is finite, and that, if E is any Lebesgue measurable point set in the space of (x_1, \dots, x_r) ,

$$(3) \quad \frac{d}{dp} \int_E \dots \int f dx_1 \dots dx_r = \int_E \dots \int f_p dx_1 \dots dx_r.$$

Let $\{\alpha_j(x_1, \dots, x_r)\}$, $j=1, \dots, s$, be Lebesgue measurable functions, and suppose that the s -variate distribution of $\alpha_1, \dots, \alpha_s$ is determined by a probability density $\phi(\alpha_1, \dots, \alpha_s; p)$. Then the family $\{\phi\}$ has the property D. Suppose that $\phi(\alpha_1, \dots, \alpha_s; p)$ satisfies the same regularity conditions in $(\alpha_1, \dots, \alpha_s; p)$ that $f(x_1, \dots, x_r; p)$ does in $(x_1, \dots, x_r; p)$. Then†

$$(4) \quad \begin{aligned} I(\alpha_1, \dots, \alpha_s) &= \int \dots \int \frac{\phi_p^2}{\phi} d\alpha_1 \dots d\alpha_s \\ &= \text{L.U.B.} \sum_{i=1}^N \frac{\left[\int_E \dots \int f_p dx_1 \dots dx_r \right]^2}{\int_E \dots \int f dx_1 \dots dx_r}, \end{aligned}$$

* Fisher II, pp. 709-710.

† The notation L.U.B. will be used for "least upper bound."

where E_1, \dots, E_N are any Lebesgue measurable point sets which are mutually exclusive, such that the denominators in the sum are positive, and which have the property that there are Borel measurable sets E'_1, \dots, E'_N in the s -dimensional space of $\alpha_1, \dots, \alpha_s$ such that E_j is the set of points (x_1, \dots, x_r) corresponding to the set E'_j .

To simplify the notation we prove the theorem for $r=s=1$. The proof that the family $\{\phi\}$ has the property D is simple and will be omitted. Divide the α -axis into mutually exclusive Borel-measurable sets E'_1, \dots, E'_N having positive probability, which cover the α -axis, except possibly for a set of probability 0,* and which have the property that either $\phi_p \geq 0$, or $\phi_p \leq 0$, on each set. Since the family $\{\phi\}$ has the property D , if $\phi(\alpha; p) = 0$ on a set \mathcal{E}' of values of α of positive Lebesgue measure, $\phi_p(\alpha; p) = 0$ almost everywhere on \mathcal{E}' . Then we can apply the inequality of Schwarz to obtain

$$(5) \quad \sum_{j=1}^N \frac{[\int_{E'_j} \phi_p(\alpha; p) d\alpha]^2}{\int_{E'_j} \phi(\alpha; p) d\alpha} \leq \sum_{j=1}^N \int_{E'_j} \frac{\phi_p^2}{\phi} d\alpha = I(\alpha).$$

The first sum can be considered as $I(z)$, where z is the chance variable taking on the value j on E'_j , i.e., z takes on the value j with probability $\int_{E'_j} \phi(\alpha; p) d\alpha$, $j=1, \dots, N$. If m_j is the greatest lower bound of $|\phi_p|/\phi$ on E'_j ,

$$(6) \quad \sum_{j=1}^N m_j^2 \int_{E'_j} \phi(\alpha; p) d\alpha \leq I(z) \leq I(\alpha).$$

The least upper bound of the sum on the left for all possible subdivisions of the α -axis is precisely $I(\alpha)$,† and is the same as the least upper bound of the sum leaving out the restriction that ϕ_p has only one sign on each set E'_j . The least upper bound of $I(z)$, the first sum in (5), must also be $I(\alpha)$. Let E_j be the set of points x corresponding to the α -set E'_j . Since

$$(7) \quad \int_{E'_j} \phi(\alpha; p) d\alpha = \int_{E_j} f(x; p) dx, \quad \int_{E'_j} \phi_p(\alpha; p) d\alpha = \int_{E_j} f_p(x; p) dx,$$

the first sum in (5) becomes the sum in (4).

This theorem suggests a more generally applicable definition of amount of information. From now on, if the family $f(x_1, \dots, x_r; p)$ satisfies the conditions of Theorem 1, the amount of information in a set of measurable func-

* The parameter p is considered fixed throughout this discussion.

† If $\phi(\alpha; p) = 1$ for $0 < \alpha < 1$ and 0 otherwise, this can be inferred from the definition of the Lebesgue integral, and is in fact a definition of integration suggested by Young; cf. E. W. Hobson, *The Theory of Functions of a Real Variable*, 2d edition, vol. 2, pp. 369-373. The general case can be treated in a similar way or reduced to this one by a transformation of the α -axis.

tions $\alpha_1(x_1, \dots, x_r), \dots, \alpha_s(x_1, \dots, x_r)$ will be defined by the last expression in (4). More general definitions can be made, but this generality will suffice for the purposes of this paper.

The following lemma will be needed below.

LEMMA 1. *Let $\{x(p)\}$, $\{x_1(p)\}$, $\{x_2(p)\}$, \dots be families of chance variables whose ranges are the points of the x -axis. Let $P(p, E)$, $P_1(p, E)$, $P_2(p, E)$, \dots be the probabilities that these variables have their values in the set E . We suppose that $I_p, I_p^{(1)}, \dots$, the amounts of information in the families, defined in accordance with the definition just given, are finite-valued, and that there is a value p_0 of p such that*

$$(8) \quad \lim_{n \rightarrow \infty} P_n(p_0, E) = P(p_0, E),$$

$$(9) \quad \lim_{n \rightarrow \infty} \frac{d}{dp} P_n(p_0, E) = \frac{d}{dp} P(p_0, E),$$

for all Borel measurable sets E . Then

$$(10) \quad \liminf_{n \rightarrow \infty} I_{p_0}^{(n)} \geq I_{p_0}.$$

If E_1, \dots, E_N is any division of the x -axis into Borel measurable sets such that $P(p, E_j) > 0, j = 1, \dots, N$,

$$(11) \quad I_{p_0}^{(n)} \geq \sum_{j=1}^N \frac{\left\{ \frac{d}{dp} P_n(p_0, E_j) \right\}^2}{P_n(p_0, E_j)}.$$

This implies that

$$(12) \quad \liminf_{n \rightarrow \infty} I_{p_0}^{(n)} \geq \sum_{j=1}^N \frac{\left\{ \frac{d}{dp} P(p_0, E_j) \right\}^2}{P(p_0, E_j)}.$$

But the sum can be made arbitrarily close to I_{p_0} by choosing E_1, \dots, E_N properly, thus proving (10).

In the proof of this lemma, (8) and (9) are only needed for sets E of the type needed in (12) to have the sum near I_{p_0} . Therefore if the distribution of $x(p)$ has a continuous density function, $f(x; p)$, such that $f_p(x; p)$ is continuous and that (2) holds (where $r = 1$), the sum in (12) has least upper bound

$$\int \frac{f_p^2}{f} dx,$$

and it is sufficient to know that (8) and (9) are true for intervals.

THEOREM 2. Let $\{f(x_1, \dots, x_r; p)\}$ be a family of probability densities satisfying the conditions of Theorem 1, and let $\alpha_1(x_1, \dots, x_r), \dots, \alpha_s(x_1, \dots, x_r)$ be Lebesgue measurable functions. Then

$$(13) \quad I(\alpha_1, \dots, \alpha_s) \leq I(x_1, \dots, x_r)$$

and there is equality for a value p_0 of p if and only if a function $\phi(\alpha_1, \dots, \alpha_s)$ exists such that, with probability 1,*

$$(14) \quad \frac{f_p(x_1, \dots, x_r; p_0)}{f(x_1, \dots, x_r; p_0)} = \phi(\alpha_1, \dots, \alpha_s).$$

There are infinitely many functions $\alpha(x_1, \dots, x_r)$ such that $I(\alpha) = I(x_1, \dots, x_r)$ for all values of p .

The inequality (13) can be obtained by applying Schwarz's inequality to (4). In order to discuss the case of equality, a new expression will be given for $I(\alpha_1, \dots, \alpha_s)$. We assume, to simplify the notation, that $r=s=1$. Let E' be a Borel measurable set of values of a new variable α , and let E be the set of values of x for which $\alpha(x)$ is in E' . Then, a fixed value of p being assumed throughout the discussion,

$$(15) \quad Q(E') = \int_E f_p(x; p) dx$$

defines a set function $Q(E')$ on the α -axis. This set function is completely additive. Let the measure of E' be defined as the probability that $\alpha(x)$ is in E' , i.e., the probability of E . Then if the measure of E' vanishes, E has probability 0. The probability of E is

$$\int_E f(x; p) dx$$

so that $f(x; p)$ must vanish almost everywhere on E . Since this is true for a single value of p it must be true for all values of p in the range considered, because the family $f(x; p)$ has the property D .† Then it is readily seen that $f_p=0$ almost everywhere on E . Thus $Q(E')=0$. Since the vanishing of the measure of E' implies the vanishing of $Q(E')$, $Q(E')$ has a density of distribution $\phi(\alpha)$ on the α -axis.‡ Writing the integrals in terms of x ,

$$(16) \quad Q(E') = \int_E f_p(x; p) dx = \int_E \phi[\alpha(x)] f(x; p) dx.$$

* As elsewhere in this paper, the expression "with probability 1" means, in terms of the measure (probability) defined on the space in question, "on a set of measure 1."

† The exceptional set of measure 0 may vary with p .

‡ O. Nikodym, *Fundamenta Mathematicae*, vol. 15 (1930), p. 179.

The expression for $I(\alpha)$ can be rewritten:

$$(17) \quad I(\alpha) = \text{L. U. B.} \sum_{i=1}^N \frac{\{\int_{E_i} \phi[\alpha(x)] f(x; p) dx\}^2}{\int_{E_i} f(x; p) dx},$$

and therefore by an argument used above,

$$(18) \quad I(\alpha) = \int \phi[\alpha(x)]^2 f(x; p) dx.$$

If the set E_i is chosen so that $\phi[\alpha(x)]$ does not change sign on it, and if m_i is the greatest lower bound of $|\phi[\alpha(x)]|$ on E_i ,

$$(19) \quad \begin{aligned} I(\alpha) &\geq \sum_{i=1}^N \int_{E_i} \phi[\alpha(x)]^2 f(x; p) dx \geq \sum_{i=1}^N m_i \int_{E_i} |\phi[\alpha(x)]| f(x; p) dx \\ &= \sum_{i=1}^N m_i \left| \int_{E_i} f_p(x; p) dx \right|. \end{aligned}$$

Using the fact that

$$(20) \quad \begin{aligned} I(\alpha) &= \text{L. U. B.} \sum_{i=1}^N m_i \int_{E_i} |\phi[\alpha(x)]| f(x; p) dx \\ &= \text{L. U. B.} \sum_{i=1}^N m_i \left| \int_{E_i} f_p(x; p) dx \right|, \end{aligned}$$

we shall show that

$$(21) \quad I(\alpha) = \int \phi[\alpha(x)] f_p(x; p) dx.$$

We can suppose the subscripts so chosen that $\phi[\alpha(x)] \geq 0$ on E_1, \dots, E_r , $\phi \leq 0$ on E_{r+1}, \dots, E_N . Then (20) becomes

$$(22) \quad I(\alpha) = \text{L. U. B.} \left\{ \sum_{i=1}^r m_i \int_{E_i} f_p(x; p) dx - \sum_{i=r+1}^N m_i \int_{E_i} f_p(x; p) dx \right\},$$

and we know that to find the least upper bound of the sum we must make a finer and finer subdivision of the α -axis. By reasoning used before, (22) becomes ϕf_p integrated over the set where $\phi \geq 0$ less $-\phi f_p$ integrated over the set where $\phi \leq 0$, which proves (21). By Schwarz's inequality,

$$(23) \quad I(\alpha)^2 \leq \left\{ \int \phi^2 f dx \right\} \cdot \left\{ \int (f_p^2 / f) dx \right\} = I(\alpha) \cdot I(x).$$

This inequality shows that $I(\alpha) \leq I(x)$ and that there can be equality only

when $\phi f^{1/2}$, $f_p/f^{1/2}$ are linearly dependent, i.e., when (14) is true, the ratio of the constants of dependence being determined by (16).

Conversely if there is a function $\phi_1(\alpha)$ such that, for some value of p , $f_p/f = \phi_1(\alpha)$ with probability 1, (16) becomes

$$(16') \quad \int_E \phi_1[\alpha(x)]f(x; p)dx = \int_E \phi[\alpha(x)]f(x; p)dx,$$

and since this equation determines ϕ , ϕ and ϕ_1 are identical except possibly for a set of values of α of probability 0. Equation (18) now shows that $I(\alpha) = I(x)$.

The condition (14) is satisfied if the correspondence between the x and α spaces is one-to-one, excluding sets of 0 probability if necessary. To prove the last part of the theorem it is thus sufficient to show that there are measurable functions $\alpha(x_1, \dots, x_r)$ which determine a one-to-one correspondence between the values of α and the points (x_1, \dots, x_r) , excluding a set of values of α of probability 0 and a set of points (x_1, \dots, x_r) of probability 0. If these exceptional point sets are of probability 0 for one value of p they will be for all values of p , since the family $f(x_1, \dots, x_r; p)$ has the property D . Such functions have been discussed in detail.*

If equation (14) can be integrated, it becomes an equation of the form

$$(24) \quad f(x_1, \dots, x_r; p) = u(\alpha_1, \dots, \alpha_s; p) \cdot v(x_1, \dots, x_r).$$

An estimate $\alpha(x_1, \dots, x_r)$ which satisfies an equation of this type (for $s=1$) is called a sufficient estimate.† It has been shown by B. O. Koopman that if $f(x_1, \dots, x_r)$ is of the form $\prod_{j=1}^r g(x_j; p)$, and if there is a sufficient estimate satisfying certain regularity conditions, $g(x; p)$ must be of the form

$$(25) \quad e^{A(x)B(p)+C(x)+D(p)}. \ddagger$$

THEOREM 3. Let $\{f(x; p)\}$ be a family of probability densities, and suppose that for each value of the integer r , $\prod_{j=1}^r f(x_j; p) = f(x_1, \dots, x_r; p)$ satisfies the regularity conditions of Theorem 1. Let $\{p_n\}$ be the maximum likelihood statistic, and suppose that $f(x; p)$ satisfies regularity conditions insuring that

$$(26) \quad (nI(x))^{1/2}(p_n - p) = \frac{1}{(nI(x))^{1/2}} \sum_{i=1}^n \frac{f_p(x_i; p)}{f(x_i; p)} + R_n,$$

where for each value of p , $\lim_{n \rightarrow \infty} R_n = 0$ with probability 1. § Then

* Cf. for instance B. Jessen, Acta Mathematica, vol. 63 (1934), pp. 260-263.

† R. A. Fisher II, pp. 712-714.

‡ These Transactions, vol. 39 (1936), pp. 399-409. Koopman treats the general case where there is more than one parameter.

§ Such conditions were determined by J. L. Doob, these Transactions, vol. 36 (1934), pp. 770-771; cf. equation (47) on p. 773.

$$(27) \quad I(p_n) \leq nI(x), \quad \lim_{n \rightarrow \infty} \frac{I(p_n)}{n} = I(x).$$

The amount of information from a sample of n is $nI(x)$, so that, using Theorem 2, $I(p_n) \leq nI(x)$. Lemma 1 will be used to prove the second part of the theorem. Let $q_n = (nI(x))^{1/2} p_n$, where p will be supposed fixed throughout this discussion. Equation (26) is used* to show that the distribution of $q_n - p(nI(x))^{1/2}$ for large n is nearly normal, with mean 0 and variance 1: i.e., that if $P_n(q, E)$ is the probability that q_n is in the interval E , $\mu_1 < q_n < \mu_2$, where $q = p(nI(x))^{1/2}$,

$$(28) \quad \lim_{n \rightarrow \infty} P_n(q, E) = \frac{1}{(2\pi)^{1/2}} \int_{\mu_1}^{\mu_2} e^{-(x-q)^2/2} dx.$$

This corresponds to equation (8). In order to obtain the equation corresponding to (9), we evaluate $P_n(q, E)$. Let E_n be the set of points $\vec{x}_n(x_1, \dots, x_n)$ at which $\mu_1 < q_n < \mu_2$. Then

$$(29) \quad P_n(q, E) = \int \cdots \int_{E_n} \prod_{i=1}^n f(x_i; p) dx_i,$$

$$(30) \quad \frac{\partial}{\partial q} P_n(q, E) = \int \cdots \int_{E_n} A_n \prod_{i=1}^n f(x_i; p) dx_i,$$

where

$$(31) \quad A_n = \frac{1}{(nI(x))^{1/2}} \sum_{i=1}^n \frac{f_p(x_i; p)}{f(x_i; p)}.$$

The set E_n is determined by $\mu_1 < q + A_n + R_n < \mu_2$ and the difference between the integral (30) over E_n and over E'_n : the set determined by

$$(32) \quad \mu_1 < q + A_n < \mu_2$$

is the integral (30) over $E_n - E_n \cdot E'_n$ less the integral over $E'_n - E_n \cdot E'_n$. Now by definition of $I(x_1, \dots, x_n)$, and by Fisher's result stated on page 412,

$$I(x_1, \dots, x_n) = nI(x) = \int \cdots \int nI(x) A_n^2 \prod_{i=1}^n f(x_i; p) dx_i.$$

Therefore

$$\left\{ \int \cdots \int A_n \prod_{i=1}^n f(x_i; p) dx_i \right\}^2 \leq \left\{ \int \cdots \int A_n^2 \prod_{i=1}^n f(x_i; p) dx_i \right\}$$

* Ibid., pp. 773-774.

$$(33) \quad \cdot \left\{ \int \cdots \int \prod_{i=1}^n f(x_i; p) dx_i \right\} \\ \leq \int \cdots \int \prod_{i=1}^n f(x_i; p) dx_i.$$

Then if the last integral on the right, the probability of e , is small, the integral on the left is also. To show that as n becomes infinite, E_n' can be substituted for E_n in (30), it is therefore only necessary to show that the probabilities of $E_n - E_n \cdot E_n'$, $E_n' - E_n \cdot E_n'$ approach 0 with $1/n$. The set $E_n - E_n \cdot E_n'$ is the set of points satisfying

$$(34a) \quad q + A_n \leq \mu_1 < q + A_n + R_n < \mu_2$$

or

$$(34b) \quad \mu_1 < q + A_n + R_n < \mu_2 \leq q + A_n,$$

and $E_n' - E_n \cdot E_n'$ is the sum of two sets defined analogously. The proofs that the probabilities of each of these four sets approach 0 with $1/n$ are similar, and only that for the set determined by (34a) will be given. This set is included in the set e_n determined by

$$(35) \quad q + A_n \leq \mu_1 < q + A_n + R_n.$$

These inequalities can be rewritten

$$(36) \quad 0 \leq \mu_1 - q - A_n < R_n.$$

Let ϵ be a positive number. The part of e_n where $\mu_1 - q - A_n > \epsilon$ has probability approaching 0 since $\lim_{n \rightarrow \infty} R_n = 0$ with probability 1. The part of e_n where $\mu_1 - q - A_n \leq \epsilon$ is included in the set where

$$(37) \quad 0 \leq \mu_1 - q - A_n \leq \epsilon.$$

Since $\{f_p(x; p)/f(x; p)\}^2$ is integrable, the probability of the set determined by (37) approaches

$$\frac{1}{(2\pi)^{1/2}} \int_{\mu_1 - q - \epsilon}^{\mu_1 - q} e^{-x^2/2} dx \leq \frac{\epsilon}{(2\pi)^{1/2}}$$

as n becomes infinite, by the Laplace-Liapounoff theorem.* We have thus shown, since ϵ is arbitrary, that the probability of e_n approaches 0 with $1/n$, and thus that as n becomes infinite, E_n' can be substituted for E_n in (30).

* A proof of this theorem, with the required amount of generality, is given by A. Khintchine, *Ergebnisse der Mathematik*, vol. 2, No. 4: *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, pp. 1-8.

Let $\lambda_1, \dots, \lambda_N$ be a set of real numbers, where

$$\mu_1 - q = \lambda_1 < \dots < \lambda_N = \mu_2 - q,$$

and let $e_n^{(i)}$ be the set of points where

$$\lambda_j \leq A_n < \lambda_{j+1} \quad (j = 1, \dots, N-2),$$

defining $e_n^{(N-1)}$ similarly except that equality is allowed on the right. Then if $p(e_n^{(i)})$ is the probability of $e_n^{(i)}$,

$$(38) \quad \int \dots \int_{E_n'} A_n \prod_{j=1}^n f(x_j; p) dx_j \geq \sum_{j=1}^{N-1} \lambda_j p(e_n^{(j)}).$$

But, using the Laplace-Liapounoff theorem again,

$$(39) \quad \lim_{n \rightarrow \infty} p(e_n^{(i)}) = \frac{1}{(2\pi)^{1/2}} \int_{\lambda_j}^{\lambda_{j+1}} e^{-x^2/2} dx.$$

Then

$$(40) \quad \liminf_{n \rightarrow \infty} \frac{d}{dq} P_n(q, E) \geq \sum_{j=1}^{N-1} \lambda_j \frac{1}{(2\pi)^{1/2}} \int_{\lambda_j}^{\lambda_{j+1}} e^{-x^2/2} dx.$$

Since the division of the axis determined by $\lambda_1, \dots, \lambda_N$ can be made arbitrarily fine,

$$(41) \quad \liminf_{n \rightarrow \infty} \frac{d}{dq} P_n(q, E) \geq \frac{1}{(2\pi)^{1/2}} \int_{\mu_1 - q}^{\mu_2 - q} x e^{-x^2/2} dx.$$

A slight modification of the discussion shows that

$$(42) \quad \limsup_{n \rightarrow \infty} \frac{d}{dq} P_n(q, E) \leq \frac{1}{(2\pi)^{1/2}} \int_{\mu_1 - q}^{\mu_2 - q} x e^{-x^2/2} dx.$$

Hence

$$(43) \quad \lim_{n \rightarrow \infty} \frac{d}{dq} P_n(q, E) = \frac{1}{(2\pi)^{1/2}} \int_{\mu_1 - q}^{\mu_2 - q} x e^{-x^2/2} dx = \frac{d}{dq} \frac{1}{(2\pi)^{1/2}} \int_{\mu_1}^{\mu_2} e^{(x-q)^2/2} dx.$$

Lemma 1 can now be applied. If a family of densities is determined by

$$g(x, q) = \frac{1}{(2\pi)^{1/2}} e^{-(x-q)^2/2},$$

the amount of information obtainable from a sample of 1 is readily seen to be 1. Therefore by Lemma 1,

$$(44) \quad \liminf_{n \rightarrow \infty} I(q_n) \geq 1.$$

The quantity $I(q_n)$ refers to the distribution of $q_n = (nI(x))^{1/2} p_n$ when $p(nI(x))^{1/2}$ is the parameter. Therefore $I(q_n) = I(p_n)/(nI(x))$, so that (44) becomes $\liminf_{n \rightarrow \infty} I(p_n)/n \geq I(x)$. This, taken in conjunction with the fact that $I(p_n) \leq nI(x)$, shows the truth of (27).

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.