ON THE PROBABILITY OF LARGE DEVIATIONS OF RANDOM VARIABLES*

by

I. N. Sanov (Moscow)

(Translation by Dana E. A. Quade, University of North Carolina)

This translation 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)-458. Reproduction in whole or in part is permitted for any purpose of the United States Government.

* This article appeared in Matematicheskii Sbornik, 42 (84):1, pp. 11-44.

Institute of Statistics
Mimeograph Series No. 192
March, 1958

ON THE PROBABILITY OF LARGE DEVIATIONS OF RANDOM VARIABLES[1]

I. N. Sanov (Moscow)
(Translation by Dana E. A. Quade)

The problem of determining the probability of large deviations of random variables attracts continual attention. There is a whole series of important investigations concerned with this question. Among these works one should first of all name those of N. V. Smirnov [1], H. Cramer [4], V. V. Petrov [5], and a whole series of others. The problem of estimating the probability of large deviations of the empirical distribution function of a random variable ξ from its theoretical distribution function after N independent observations has, as it turns out, a close connection with this problem; and also that of estimating the probability of large deviations of the order statistics from their means in a random sample from a population. The distributions of what might be called "normal" deviations in the preceding two cases have been obtained in papers by A. N. Kolmogorov [7] and N. V. Smirnov [2].

It will become clear that in problems about large deviations an important role is played by the concept of the integral $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ which is presented in this paper. This integral is defined for any two distribution functions $F(x)$ and $\Phi(x)$.

In this paper we present an approach to the problem for deviations which are sufficiently large. Thus, for example, with respect to the deviation of the normalized number m of successes in N independent trials in a Bernouilli scheme with p $(0 < p < 1)$ as the probability of success in one

1 This translation 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)-83. Reproduction in whole or in part is permitted for any purpose of the United States Government.

trial asymptotic expressions have been obtained for the following probabilities:

$$P\left(\frac{m - Np}{\sqrt{Npq}} \geq A\right), \quad A \text{ a constant, } q = 1-p$$

(integral theorem of Laplace).

Asymptotic formulas for the probabilities

$$P\left(\frac{m - Np}{\sqrt{Npq}} \geq AN^{\alpha}\right), \quad 0 < A, \quad 0 < \alpha < \frac{1}{2}$$

are derived in papers by N. V. Smirnov [1], H. Cramér [4], and V. V. Petrov [5]. In this paper a method is developed which enables one to obtain, for example, in the Bernouilli scheme, asymptotic formulas for the probabilities

$$P\left(\frac{m - Np}{\sqrt{Npq}} \geq \gamma(N)\right) \text{ where } \gamma(N) = O(N^{1/2})^{*}.$$

Hence the methods developed in this paper should be looked on, not as replacements for the methods of the papers cited above, but as a certain extension of their results; although, perhaps, it would be of some interest to develop these methods to such a degree that they give all the basic results of the aforementioned investigations. There is a real possibility of this. The value of such a method will consist in a unique approach to a whole circle of different questions.

---

* The case where $\gamma(N) = o(N^{1/2})$ is considered in the paper by V. V. Petrov [5].

I. DISCRETE CASE

We consider a series of $N > 1$ independent trials according to the simple scheme with n outcomes $a_1, a_2, \ldots, a_n$ and corresponding probabilities for the appearance of these outcomes

$$(1) \qquad \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \qquad \sum_{i=1}^{n} p_i = 1.$$

Let $p_i > 0$ for $i = 1,2,\ldots,n$ and let the result of these N trials be that the corresponding outcomes occur $m_1, m_2, \ldots, m_n$ times where $m_1 + m_2 + \ldots + m_n = N$ and $m_i \geq 0$ for $i = 1,2,\ldots,n$. Denote the relative frequency of the outcome a_i by ν_i:

$$\nu_i = \frac{m_i}{N} \quad \text{where } \nu_1 + \nu_2 + \ldots + \nu_n = 1.$$

It is well-known that the probability $P(m_1, m_2, \ldots, m_n)$ of occurrence of the set of numbers of outcomes $m_1, m_2, \ldots, m_n$ is given by the formula

$$(2) \qquad P(m_1, m_2, \ldots, m_n) = \frac{N!}{m_1! m_2! \ldots m_n!} p_1^{m_1} p_2^{m_2} \ldots p_n^{m_n}$$

where we take $0! = 1$.

In Stirling's formula

$$m! = \sqrt{2\pi m}\; e^{\frac{\Theta_m}{12m}} m^m e^{-m}, \quad |\Theta_m| < 1,$$

we denote the first two factors by ρ_m:

$$\rho_m = \sqrt{2\pi m}\; e^{\frac{\Theta_m}{12m}}.$$

Furthermore, we shall put $\rho_0 = 1$, $0^0 = 1$, $0! = \rho_0 0^0 e^{-0} = 1$.

Putting $N\nu_i$ in place of m_i we obtain:

$$(3) \qquad P(m_1,m_2,\dots,m_n) = \frac{\rho_N}{\rho_{m_1}\,\rho_{m_2}\,\cdots\,\rho_{m_n}}\, e^{N\sum_{i=1}^{n}\nu_i \ln\frac{p_i}{\nu_i}} = e^{N\sum_{i=1}^{n}\nu_i \ln\frac{p_i}{\nu_i} + R},$$

where $R = \ln \rho_N - \sum_{i=1}^{n} \ln \rho_{m_i}$ and we take that $0 \cdot \ln \frac{p_i}{0} = 0$. Evidently $1 \leq \rho_{m_i} < \sqrt{2\pi N}\, e^{\frac{1}{12N}}$ and $1 < \rho_N < \sqrt{2\pi N}\, e^{\frac{1}{12N}}$; and hence, because of the monotonicity of the function $\ln x$, we have the estimate

$$(4) \qquad |R| < (n+1)\left(\frac{1}{2}\ln N + \frac{1}{2}\ln 2\pi + \frac{1}{12N}\right).$$

The number T of all possible sets $(m_1,m_2,\dots,m_n)$ is equal to the number of different non-negative integral solutions of the indefinite equation $x_1+x_2+\dots+x_n = N$. From this it follows that

$$(5) \qquad T = \frac{N^{n-1}}{(n-1)!} + O(N^{n-2}).$$

We denote by Ω a region of points in the $(n-1)$-dimensional simplex $\Sigma = \{x_1 + x_2 + \dots + x_n = 1,\ x_i \geq 0 \text{ for } i = 1,2,\dots,n\}$ such that it includes at least one interior point of this simplex together with its ϵ-neighborhood. Let T_Ω be the number of sets $(\nu_1,\nu_2,\dots,\nu_n) \in \Omega$ for a fixed N. Then from (5) we obtain:

$$(6) \qquad T_\Omega = O(N^{n-1}) = O(e^{\min N})$$

and for N sufficiently large

$$T_\Omega > 1.$$

We take a set of real numbers $q_1, q_2, \dots, q_n$ which satisfy the conditions $q_1 + q_2 + \dots + q_n = 1$ and $q_i \geq 0$ for $i = 1,2,\dots,n$. And we take an arbitrary $\epsilon > 0$. Then if we take as the region Ω the region where $|x_i - q_i| \leq \epsilon$ for $i = 1,2,\dots,n$ and use formulas (3) and (6) and approximation (4) it is not difficult to prove

Theorem 1. After N independent trials according to the scheme (1), for any $\epsilon > 0$

$$(7) \qquad P(|\nu_1 - q_1| \leq \epsilon,\ |\nu_2 - q_2| \leq \epsilon, \dots, |\nu_n - q_n| \leq \epsilon) = e^{N\left[\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} + O(\epsilon \ln \frac{1}{\epsilon}) + O(\frac{\ln N}{N})\right]}.$$

If $q_i > 0$ for $i = 1,2,\dots,n$ then $O(\epsilon \ln \frac{1}{\epsilon})$ may be replaced by $O(\epsilon)$.

It is possible to introduce the following interpretation for the region Ω. We shall say that the region Ω is distinguishable from the scheme (1) if for every $\eta > 0$ there exist two open sets Ω_1 and Ω_2 of points of the simplex Σ such that $\Omega_1 \subset \Omega \subset \Omega_2$ and

$$\sup_{x \in \Omega_2} \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} - \eta < \sup_{x \in \Omega} \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} < \sup_{x \in \Omega_1} \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} + \eta,$$

where $x = (x_1, x_2, \dots, x_n)$ is a point of the simplex Σ.

Theorem 2. If the region Ω is closed and distinguishable from the scheme (1) then after N independent trials according to scheme (1) the asymptotic formula

$$(8) \qquad P[(\nu_1, \nu_2, \dots, \nu_n) \in \Omega] = e^{N\left[\sum_{i=1}^{n} \lambda_1 \ln \frac{p_i}{\lambda_i} + o(1)\right]}$$

holds, where $(\lambda_1, \lambda_2, \dots, \lambda_n)$ is that point of the region Ω at which the function $\sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i}$ attains its maximum for all $(x_1, x_2, \dots, x_n) \in \Omega$.

It is not difficult to verify that $\max \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} = 0$ on the simplex Σ, this maximum being attained when $x_i = p_i$ for $i = 1,2,\ldots,n$. The remaining points of the simplex Σ are not extremal. Hence it follows that if $(p_1,p_2,\ldots,p_n) \notin \Omega$ then $\max_{x\in\Omega} \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} = c$, where $c < 0$, and the point $(\lambda_1,\lambda_2,\ldots,\lambda_n)$ where the maximum is attained is not an interior point of Ω; i.e., it is a point of the boundary.

We note a particular case of a region Ω distinguishable from scheme (1) which allows us to obtain more precise results. Let the region Ω include at least one interior point of the simplex Σ, but not include the point $(p_1,p_2,\ldots,p_n)$, and be defined by the inequality $F(x_1,x_2,\ldots,x_n) \geq 0$ where $F(x_1,x_2,\ldots,x_n)$ has all its derivatives of second order continuous in the simplex Σ. Clearly, $F(p_1,p_2,\ldots,p_n) < 0$.

Thus, for example, given $n = 2$ and the linear function $F(x_1,x_2,\ldots,x_n) = a_1x_1 + a_2x_2 - b$, the condition which defines the region Ω becomes the following:

$$F(x_1,x_2) = a_1x_1 + a_2(1-x_1) - b \geq 0$$

or, if $a_2 < a_1$,

$$x_1 \geq \frac{b - a_2}{a_1-a_2} = p_1' \quad \text{or} \quad \nu_1 \geq p_1' \, .$$

Here the point (p_1,p_2) lies outside this region, i.e. $p_1 < p_1'$. In this case

$$P[(\nu_1,\nu_2)\in\Omega] = \sum_{\frac{m_1}{N} \geq p_1'} P(m_1,m_2)$$

and this probability is different from zero only if $p_1' \leq 1$. In the following we assume that $p_1' < 1$ and

$$P(m_1,m_2) = \frac{1}{\sqrt{2\pi N\nu_1\nu_2}}\, e^{N(\nu_1 \ln \frac{p_1}{\nu_1} + \nu_2 \ln \frac{p_2}{\nu_2}) + \frac{\gamma}{12N}}$$

where $\gamma = \Theta_N - \frac{\Theta_{m_1}}{\nu_1} - \frac{\Theta_{m_2}}{\nu_2}$.

The function $\nu_1 \ln \frac{p_1}{\nu_1} + (1-\nu_1) \ln \frac{p_2}{\nu_2}$ decreases monotonically for $\nu_1 > p_1$. Hence for every $\epsilon > 0$ $(\epsilon \geq \frac{1}{N})$

$$P[(\nu_1,\nu_2)\epsilon\Omega] = \sum_{p_1'+\epsilon > \frac{m_1}{N} \geq p_1'} \frac{1}{\sqrt{2\pi N p_1' p_2'}}\, e^{N(\nu_1 \ln \frac{p_1}{\nu_1} + \nu_2 \ln \frac{p_2}{\nu_2})}$$

$$\cdot [1 + O(\epsilon) + O(\tfrac{1}{N}) + O(e^{-N\epsilon c})],$$

where $c > 0$, $p_2' = 1 - p_1'$. In the interval $[p_1', p_1'+\epsilon)$ where ν_1 varies (we are assuming that $0 < p_1' < p_1' + \epsilon < 1$)

$$\nu_1 \ln \frac{p_1}{\nu_1} + (1-\nu_1) \ln \frac{p_2}{1-\nu_1} = p_1' \ln \frac{p_1}{p_1'} + (1-p_1') \ln \frac{p_2}{1-p_1'}$$

$$+ (\nu_1 - p_1') \ln \left(\frac{p_1 p_2'}{p_1' p_2}\right) + O(\epsilon^2).$$

Hence we obtain

$$P[(\nu_1,\nu_2)\epsilon\Omega] = \frac{1}{\sqrt{2\pi p_1' p_2'}}\, e^{N(p_1' \ln \frac{p_1}{p_1'} + p_2' \ln \frac{p_2}{p_2'})}$$

$$\cdot \sum_{Np_1'+N\epsilon > m_1 \geq Np_1'} \left(\frac{p_1 p_2'}{p_2 p_1'}\right)^{m_1 - Np_1'}$$

$$\cdot [1 + O(\epsilon) + O(N\epsilon^2) + O(\tfrac{1}{N}) + O(e^{-N\epsilon c})].$$

If we put $\epsilon = N^{-1+\eta}$, where $\eta > 0$, and note that $0 < \frac{p_1 p_2'}{p_2 p_1'} < 1$, we obtain the following theorem:

Theorem 3. In the case of N independent Bernouilli trials with probability of success p_1 $(0 < p_1 < 1)$ we have the following asymptotic formula for m, the number of successes, when $p < p_1' < 1$:

$$(9) \quad P\left(\frac{m}{N} \geq p_1'\right) = \frac{1}{\sqrt{2\pi N p_1' p_2'}} \, e^{N\left(p_1' \ln \frac{p_1}{p_1'} + p_2' \ln \frac{p_2}{p_2'}\right)} \frac{\left(\frac{p_1 p_2'}{p_2 p_1'}\right)^{(-Np_1')}}{1 - \frac{p_1 p_2'}{p_2 p_1'}} \cdot \left[1 + O(N^{-1+2\eta})\right].$$

Here $(-Np_1')$ is the fractional part of the number $-Np_1'$; $0 \leq (-Np_1') < 1$, $p_2 = 1 - p_1$, $p_2' = 1 - p_1'$.

In the general case of a region Ω defined by the inequality $F(\nu_1, \nu_2, \ldots, \nu_n) \geq 0$, where $F(p_1, p_2, \ldots, p_n) < 0$, one can without any difficulty prove

Theorem 4. If the function $F(x_1, x_2, \ldots, x_n)$ has continuous derivatives up to and including the second order in the simplex Σ, then after N independent trials according to scheme (1)

$$(10) \quad P\left[F(\nu_1, \nu_2, \ldots, \nu_n) \geq 0\right] = \frac{a(N)}{\sqrt{N}} \, e^{N \sum_{i=1}^{n} \lambda_i \ln \frac{p_i}{\lambda_i}} \left[1 + o(1)\right],$$

where a(N) is a certain function of N bounded above and below by positive constants; $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the point at which max $\sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i}$ is attained subject to the constraints $\sum_{i=1}^{n} x_i = 1$, $F(x_1, x_2, \ldots, x_n) = 0$, and $x_i \geq 0$ for $i = 1, 2, \ldots, n$.

In the preceding case the role of the function a(N) was played by

$$a(N) = \frac{1}{\sqrt{2\pi p_1' p_2'}} \cdot \frac{\left(\frac{p_1 p_2'}{p_2 p_1'}\right)^{(-Np_1')}}{1 - \left(\frac{p_1 p_2'}{p_2 p_1'}\right)} .$$

We shall show how to find $\max \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i}$ subject to the constraints $F(x_1,x_2,\ldots,x_n) = 0$ and $\sum_{i=1}^{n} x_i = 1$.

According to the rule for finding a constrained extremum we write the system of equations:

$$(11) \quad \frac{\partial}{\partial x_k} \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} - \lambda \frac{\partial}{\partial x_k} F(x_1,x_2,\ldots,x_n) - \mu \frac{\partial}{\partial x_k} \left(\sum_{i=1}^{n} x_i - 1\right) = 0$$

for $k = 1,2,\ldots,n$; or

$$\ln p_k - \ln x_k - 1 - \lambda \frac{\partial F}{\partial x_k} - \mu = 0 \quad \text{for} \quad k = 1,2,\ldots,n.$$

Expressing x_k in terms of λ and μ from this system, we shall determine λ and μ from the equations of constraint

$$F(x_1,x_2,\ldots,x_n) = 0 \quad \text{and} \quad \sum_{i=1}^{n} x_i - 1 = 0.$$

For example, if $F(x_1,x_2,\ldots,x_n)$ is the linear form $F(x_1,x_2,\ldots,x_n) = a_1x_1 + a_2x_2 + \ldots + a_nx_n - L$, the indicated system becomes the following:

$$\ln p_k - \ln x_k - 1 - \lambda a_k - \mu = 0 \quad \text{for } k = 1,2,\ldots,n.$$

Hence $x_k = p_k\, e^{-1-\mu-\lambda a_k}$. We have the equalities

$$e^{-1-\mu} \sum_{k=1}^{n} p_k\, e^{-\lambda a_k} = 1 \quad \text{and} \quad e^{-1-\mu} \sum_{k=1}^{n} a_k\, p_k\, e^{-\lambda a_k} = L$$

for determining λ and μ. Dividing the second equality by the first we obtain for λ the equation

$$(12) \quad \sum_{k=1}^{n} (a_k - L) p_k\, e^{-\lambda a_k} = 0.$$

This equation has a real solution only if

$$\min_k a_k \leq L \leq \max_k a_k .$$

Under these conditions the real solution of equation (12) is unique. After finding λ_0, the root of equation (12), we obtain:

$$e^{-1-\mu} = \frac{1}{\sum_{k=1}^{n} p_k e^{-\lambda_0 a_k}} \qquad \text{and}$$

(13)

$$x_s = \frac{p_s e^{-\lambda_0 a_s}}{\sum_{k=1}^{n} p_k e^{-\lambda_0 a_k}} .$$

We note here that if we substitute the solution x_s into the expression $\sum_{s=1}^{n} x_s \ln \frac{p_s}{x_s}$ we shall have

$$\sum_{s=1}^{n} x_s \ln \frac{p_s}{x_s} = \sum_{s=1}^{n} p_s e^{-1-\mu-\lambda_0 a_s} \ln (e^{1+\mu+\lambda_0 a_s})$$

$$= e^{-1-\mu}(1+\mu) \sum_{s=1}^{n} p_s e^{-\lambda_0 a_s} + e^{-1-\mu} \lambda_0 \sum_{s=1}^{n} a_s p_s e^{-\lambda_0 a_s}$$

$$= e^{-1-\mu}(1+\mu)e^{1+\mu} + \lambda_0 L$$

$$= 1 + \mu + \lambda_0 L .$$

Thus, subject to the constraints $\sum_{i=1}^{n} x_i = 1$ and $\sum_{i=1}^{n} a_i x_i = L$,

(14) $$\max \sum_{i=1}^{n} x_i \ln \frac{p_i}{x_i} = \lambda_0 L + \ln \left(\sum_{i=1}^{n} p_i e^{-\lambda_0 a_i} \right),$$

where λ_0 is the real solution of the following equation with respect to λ:

$$\sum_{k=1}^{n} (a_k - L) p_k e^{-\lambda a_k} = 0.$$

On applying the approximation just investigated to the case of the most powerful test between two simple hypotheses we obtain the theorem:

Theorem 4'. Let Hypothesis 1 be that the N independent trials with numbers $m_1, m_2, \ldots, m_n$ of outcomes $a_1, a_2, \ldots, a_n$ came from the simple scheme

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0.$$

And let Hypothesis 2 be that the same trials came from the scheme

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n \\ q_1 & q_2 & \dots & q_n \end{pmatrix}, \quad \sum_{i=1}^{n} q_i = 1, \quad q_i \geq 0.$$

Then the probability β of the inequality

$$\sum_{i=1}^{n} m_i \ln \frac{p_i}{q_i} \geq NL$$

when Hypothesis 2 is true (the probability of an error of the second kind in the most powerful test, the critical value being NL: see [8], p. 312) is given by the asymptotic formula

$$\beta = \frac{a(N)}{\sqrt{N}} e^{cN} [1 + o(1)], \tag{15}$$

where the constant $c \leq 0$, and $0 < c_1 \leq a(N) \leq c_2$, c_1 and c_2 also being constants. Here $\min_i \ln \frac{p_i}{q_i} \leq L \leq \max_i \ln \frac{p_i}{q_i}$ and c is given by the formula

$$c = -t_0 L + \ln \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}\right)^{t_0} \tag{16}$$

where t_0 is the root of the equation

$$\sum_{i=1}^{n} \left(\ln \frac{p_i}{q_i} - L\right) q_i \left(\frac{p_i}{q_i}\right)^{t} = 0. \tag{17}$$

The proof of this theorem is based on Theorem 4 and the preceding discussion, if we put $a_k = \ln \frac{p_k}{q_k}$. The only difference is that we have taken $\lambda = -t$.

It is not difficult to see that if

$$\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} \leq L \leq \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

then the solution to equation (17) will lie in the interval $[0, 1]$. In fact, putting 0 for t in the left hand side of equation (17) gives $\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} - L \leq 0$, and putting 1 for t gives $\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} - L \geq 0$; hence the root t_0 must lie between the indicated limits.

II. GENERAL CASE

1. Properties of the Integral $\int \ln \frac{dF}{d\Phi}\, d\Phi$.

In what follows let $F(x)$ be the arbitrary but fixed distribution function of a random variable ξ. Let $\Phi(x)$ be another distribution function. We shall require that the following condition be satisfied:

(G). In every interval in which $\Delta F = 0$, $\Delta\Phi = 0$ also.

In such cases we put $\Delta\Phi \ln \frac{\Delta F}{\Delta\Phi} = 0$.

We shall take an arbitrary subdivision $(x)_n$ of the interval $(-\infty, +\infty)$:

$$-\infty = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = +\infty \tag{18}$$

and another subdivision $(y)_r$:

$$-\infty = y_0 < y_1 < y_2 < \dots < y_{r-1} < y_r = +\infty .$$

For the first subdivision we form the sum $S_{(x)} = \sum_{i=1}^{n} \Delta_i\Phi \ln \frac{\Delta_i F}{\Delta_i\Phi}$ and for the second the sum $S_{(y)} = \sum_{i=1}^{r} \Delta_i'\Phi \ln \frac{\Delta_i' F}{\Delta_i'\Phi}$, where

$$\Delta_i\Phi = \Phi(x_i) - \Phi(x_{i-1}), \quad \Delta_i F = F(x_i) - F(x_{i-1}) \quad \text{for } i = 1,2,\dots,n,$$

$$\Delta_i'\Phi = \Phi(y_i) - \Phi(y_{i-1}), \quad \Delta_i' F = F(y_i) - F(y_{i-1}) \quad \text{for } i = 1,2,\dots,r.$$

Then we have the theorem:

<u>Theorem 5.</u> If the subdivision $(y)_r$ is a refinement of the subdivision $(x)_n$ then $0 \geq S_{(x)} \geq S_{(y)}$.

This theorem follows from the following inequality: if $x,y,u,v \geq 0$,

$$(x+y) \ln \frac{u+v}{x+y} \geq x \ln \frac{u}{x} + y \ln \frac{v}{y} \tag{19}$$

(see [9], p. 97, Inequality no. 117).

Two cases are possible.

1) For all subdivisions $(x)_n$ like (18) the sum $S_{(x)}$ is bounded below. Then there exists a greatest lower bound A on the quantity $S_{(x)}$ for all subdivisions $(x)_n$:

(20) $$A = \inf_{(x)_n} S_{(x)_n} .$$

In this case, for every $\epsilon > 0$ there exists a subdivision $(x)_n$ such that for this subdivision

(21) $$A + \epsilon > S_{(x)} \geq A.$$

It is evident that if $(y)_r$ is any refinement of the subdivision $(x)_n$ then the following inequality also holds:

$$A + \epsilon > S_{(y)} \geq A.$$

In this case we say that by definition there exists the integral

(22) $$\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi = \inf_{(x)_n} S_{(x)_n} = A.$$

2) The sums $S_{(x)}$ are not bounded below, or condition (G) is not satisfied. Then we say that the integral $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ does not exist, but sometimes we shall write that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} = -\infty$.

We shall note several properties of $\int \ln \frac{dF}{d\Phi}\, d\Phi$.

This integral is an additive function of the interval of integration. From the definition of the integral it is clear that if it exists over the whole interval $(-\infty, +\infty)$ it exists also over every subinterval.

It is not difficult to show that, for all distribution functions $F(x)$ and $\Phi(x)$, $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi \leq 0$, and equality is attained only if $F(x) \equiv \Phi(x)$. This inequality follows from the corresponding inequality for finite sums:

$$\sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi \leq 0$$

and from the fact that

$$\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi = \inf_{(x)_n} \sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi .$$

From the last equality we obtain that if $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi = 0$ then

$$0 \leq \sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi \leq 0$$

and therefore

$$\sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi = 0$$

for every subdivision $(x)_n$, so that $\Delta_i F = \Delta_i \Phi$. (This is true, as was shown above, for all corresponding finite sums.) Hence without difficulty we conclude that $F(x) \equiv \Phi(x)$.

We shall show how to solve several problems concerning the determination of $\sup_{\Phi} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi$ in those cases where the function Φ runs through a class of functions Ω .

First of all we take for the class Ω the set of all distribution functions which pass through a certain finite number of fixed points in the (x,y) plane. Since the functional $\int \ln \frac{dF}{d\Phi} d\Phi$ is additive with respect to the interval of integration it is sufficient to solve the problem by parts in each interval where x varies between two neighboring fixed points.

After having added the points $(-\infty, 0)$ and $(+\infty, 1)$ to the number of fixed points, we shall find the curve which maximizes $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi$ among all functions Φ in the class Ω . We shall call this curve <u>piecewise-extremal</u>

for $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$, since the segment of it between two neighboring fixed points (a,c) and (b,d) maximizes $\int_{a}^{b} \ln \frac{dF}{d\Phi}\, d\Phi$.

So we seek a non-decreasing function $\Phi(x)$ which passes through the two points (a,c) and (b,d), $a < b$, $c \leq d$. This function satisfies the condition $\int_{a}^{b} d\Phi(x) = d-c$.

We shall look among all such functions for that one which maximizes $\int_{a}^{b} \ln \frac{dF}{d\Phi}\, d\Phi$.

For this we shall take any subdivision $(x)_n$ of the interval (a,b) into n parts and with respect to that subdivision we shall seek

$\max \sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi$ among all choices of $\Delta_i \Phi \geq 0$ such that $\sum_{i=1}^{n} \Delta_i \Phi = d-c$.

The problem reduces to finding the maximum of the function $f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \ln \frac{\Delta_i F}{x_i} x_i$ subject to the conditions $x_i \geq 0$ and $f_1(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i - d + c = 0$ for fixed $\Delta_i F$. According to the rule for finding a maximum subject to constraints we write the system of equations in x_i

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial f_1}{\partial x_i} = 0 \quad \text{for } i = 1, 2, \ldots, n,$$

or

$$\ln \Delta_i F - \ln x_i - 1 - \lambda = 0$$

or

$$x_i = e^{-1-\lambda} \Delta_i F.$$

We define λ by the condition that

$$\sum_{i=1}^{n} x_i = d-c \, ;$$

or

$$e^{-1-\lambda} \sum_{i=1}^{n} \Delta_i F = d-c \, ;$$

but $$\sum_{i=1}^{n} \Delta_i F = F(b) - F(a)$$

whence $$e^{-1-\lambda} = \frac{d-c}{F(b)-F(a)}$$

and $$x_i = \frac{d-c}{F(b)-F(a)} \Delta_i F.$$

$$\Phi(a + \sum_{i=1}^{k} \Delta_i x) = \Phi(a) + \sum_{i=1}^{k} \Delta_i \Phi$$

$$= \Phi(a) + \sum_{i=1}^{k} x_i$$

$$= c + \frac{d-c}{F(b)-F(a)} \sum_{i=1}^{k} \Delta_i F.$$

Because of the continuity of $\Phi(x)$ and $F(x)$ on one side we obtain, upon infinitely refining the subdivision,

$$\Phi(x) = c + \frac{d-c}{F(b)-F(a)} [F(x)-F(a)].$$

Without difficulty we conclude that the function so found is the solution to the stated problem.

We shall find the actual value of $\max_{\Phi} \int_a^b \ln \frac{dF}{d\Phi} d\Phi$ under the condition that $\int_a^b d\Phi = d-c$. To do this, we substitute into the integral the extremal function already found, $\Phi(x) = c + \frac{d-c}{F(b)-F(a)} [F(x)-F(a)]$, or $d\Phi(x) = \frac{d-c}{F(b)-F(a)} dF(x)$, and the integral becomes the expression

$$\int_a^b (\ln \frac{F(b)-F(a)}{d-c}) \frac{d-c}{F(b)-F(a)} dF(x) = (d-c) \ln \frac{F(b)-F(a)}{d-c} .$$

We consider the class $\Omega(a_0,a_1,a_2,\ldots,a_n;\ b_0,b_1,b_2,\ldots,b_n)$ of all distribution functions which pass through the points with coordinates

$$(a_0,b_0),\ (a_1,b_1),\ \ldots,\ (a_n,b_n),$$

where $\quad -\infty = a_0 < a_1 < a_2 < \ldots < a_n = +\infty$

and $\quad 0 = b_0 \leq b_1 \leq b_2 \leq \ldots \leq b_{n-1} \leq b_n = 1.$

This is a so-called "cylinder". If $b_i = \frac{s_i}{N}$ for $i = 0,1,\ldots,n$, where s_i and N are integers, we then call this cylinder an N-cylinder which in what follows we shall denote by Δ_N.

The set of points $a_0 < a_1 < \ldots < a_n$ we shall call the subdivision $(x)_n$ on which the cylinder is constructed. We shall call the cylinder F-admissible if, for each interval (a_i, a_{i+1}) where $F(a_{i+1}) - F(a_i) = 0$, $b_{i+1} - b_i = 0$ also.

The results we have obtained may be formulated as the following theorem:

Theorem 6. Among all the distribution functions Φ of the F-admissible cylinder $\Omega(-\infty, a_1,a_2,\ldots,a_{n-1}, +\infty\,;\ 0, b_1,b_2,\ldots,b_{n-1}, 1)$ there exists one function $\Phi_0(x)$ which maximizes $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$. This function is defined in each interval (a_{i-1}, a_i) in the following manner:

$$(23) \qquad \Phi_0(x) = b_{i-1} + \frac{b_i - b_{i-1}}{F(a_i)-F(a_{i-1})}\left[F(x)-F(a_{i-1})\right], \quad i = 1,2,\ldots,n.$$

Also
$$\max_{\Phi\in\Omega} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi = \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi_0}\, d\Phi_0$$

$$(24) \qquad = \sum_{i=1}^{n} (b_i - b_{i-1}) \ln \frac{F(a_i)-F(a_{i-1})}{b_i - b_{i-1}}$$

$$= \sum_{i=1}^{n} \Delta_i b \ln \frac{\Delta_i F}{\Delta_i b}$$

where $\Delta_i b = b_i - b_{i-1}$ and $\Delta_i F = F(a_i) - F(a_{i-1})$.

We shall call $\Phi_0(x)$ the <u>piecewise-extremal function corresponding to the cylinder Ω</u>.

Another sort of class Ω of distribution functions may be given in the following manner. We take a function $f(x)$ on the interval $(-\infty, +\infty)$ and consider the set Ω_α of all those distribution functions $\Phi(x)$ for which $\int_{-\infty}^{\infty} f(x)d\Phi(x) \geq \alpha$. If $F(x) \notin \Omega_\alpha$, i.e. $\int_{-\infty}^{\infty} f(x)dF(x) < \alpha$, then one is easily convinced that $\max_{\Phi\in\Omega_\alpha} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ can be attained only by a function Φ for which $\int_{-\infty}^{\infty} f(x)d\Phi(x) = \alpha$.

Hence we see that the problem becomes that of finding $\max_{\Phi} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ subject to two constraints:

$$(25) \qquad \int_{-\infty}^{\infty} f(x)d\Phi(x) = \alpha \quad \text{and} \quad \int_{-\infty}^{\infty} d\Phi(x) = 1.$$

The solution to this problem can be outlined in the following manner. We write the equation

$$\frac{\partial}{\partial(d\Phi)} \left[\ln \frac{dF}{d\Phi}\, d\Phi\right] - \lambda \frac{\partial}{\partial(d\Phi)} \left[f(x)\, d\Phi\right] - \mu \frac{\partial}{\partial(d\Phi)}\, d\Phi = 0$$

or

$$\ln dF - \ln d\Phi - 1 - \lambda f(x) - \mu = 0$$

whence

$$d\Phi = e^{-1-\mu-\lambda f(x)} dF$$

or

$$\Phi(x) = e^{-1-\mu} \int_{-\infty}^{x} e^{-\lambda f(u)}\, dF(u),$$

the constants λ and μ being found from the two conditions (25), which give

$$e^{-1-\mu} \int_{-\infty}^{\infty} f(x)\, e^{-\lambda f(x)}\, dF(x) = \alpha$$

and $$e^{-1-\mu} \int_{-\infty}^{\infty} e^{-\lambda f(x)} \, dF(x) = 1.$$

Dividing the second equality by the first, we eliminate μ; and for the determination of λ we obtain the equation

$$\int_{-\infty}^{\infty} [f(x) - \alpha] \, e^{-\lambda f(x)} \, dF(x) = 0. \tag{26}$$

Taking any root λ_0 of this equation, we find for $e^{-1-\mu}$ the expression $\left(\int_{-\infty}^{\infty} e^{-\lambda_0 f(x)} \, dF(x) \right)^{-1}$, and thence we have for $\Phi(x)$ the following expression:

$$\Phi(x) = \frac{\int_{-\infty}^{x} e^{-\lambda_0 f(u)} \, dF(u)}{\int_{-\infty}^{\infty} e^{-\lambda_0 f(u)} \, dF(u)} \,. \tag{27}$$

We shall not treat here such questions as whether these integrals converge, or whether there is a unique solution to equation (26) which determines λ_0, for different functions $f(x)$ and $F(x)$. Each case must be studied separately.

As examples we present a few results obtained by this method for different classes Ω_α and functions $F(x)$. (We omit the intermediate calculations.)

1. Let $$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} \, du,$$

and let Ω_α be defined by the inequality

$$\int_{-\infty}^{\infty} x \, d\,\Phi(x) \geq \alpha > 0;$$

then $\max\limits_{\Phi\in\Omega_\alpha} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ is attained for

$$\Phi(x) = F(x-a),$$

and

$$\int_{-\infty}^{\infty} \ln \frac{dF(x)}{dF(x-a)}\, dF(x-a) = -\frac{\alpha^2}{2}\ .$$

2. Let
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du,$$

and let Ω_α be defined by the inequality

$$\int_{-\infty}^{\infty} x^2\, d\Phi(x) \geq \alpha > 1\ ;$$

then $\max\limits_{\Phi\in\Omega_\alpha} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ is attained for

$$\Phi(x) = F\left(\frac{x}{\sqrt{\alpha}}\right).$$

3. Let $F(x)$ be the discrete Poisson distribution with mean $Ex = a$, i.e. the step function with jumps of height $p_i = \frac{e^{-a}a^i}{i!}$ at the points $i = 0,1,2\ldots$, and let Ω_b be the class of all step functions $\Phi(x)$ with jumps at the points $i = 0,1,2,\ldots$ such that $\int_{-\infty}^{\infty} x\, d\Phi(x) \geq b > a$; then $\max\limits_{\Phi\in\Omega_b} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ is attained for the function $\Phi_0(x)$ with jumps of height $q_i = \frac{e^{-b}b^i}{i!}$ at the points $i = 0,1,2,\ldots$, i.e. the Poisson distribution with mean b.

In conclusion we shall say a few words about finding $\max\limits_{\Phi\in\Omega} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ for other classes Ω, in particular when Ω is a geometrical region (i.e. a region of functions bounded above and below by certain curves.)

The function Φ which maximizes the stated functional must always either coincide with F or belong to the "border" of the region Ω. To determine the points where the extremal curve $\Phi(x)$ leaves the geometrical boundary it may be of great advantage to consider two vector fields which give directions at each point of the strip (x,y), $-\infty < x < +\infty$, $0 \leq y \leq 1$. One field consists of the directions at the points (x,y) of the non-decreasing curves $\Phi_1(x)$ joining the points (x,y) to $(+\infty, 1)$ which maximize $\int_x^{\infty} \ln \frac{dF}{d\Phi} \, d\Phi$. The other field consists of the directions at the points (x,y) of the curves $\Phi_2(x)$ joining the points (x,y) to $(-\infty, 0)$ which maximize $\int_{-\infty}^{x} \ln \frac{dF}{d\Phi} \, d\Phi$. If the function $F(x)$ has a continuous derivative $F'(x)$ then the angular coefficients of the vectors of the first field for the points (x,y) have the form:

$$\alpha^{(1)}_{(x,y)} = \frac{1-y}{1-F(x)} F'(x).$$

The corresponding angular coefficients of the vectors of the second field for the points (x,y) have the form:

$$\alpha^{(2)}_{(x,y)} = \frac{y}{F(x)} F'(x).$$

2. Definition of Probabilities in the General Case.

Suppose we have N successive independent observations on a random variable ξ with distribution function $F(x)$ and the empirical distribution function of these observations is $F_N(x)$. Let Ω be a certain class of distribution functions. We shall define $P(F_N \in \Omega)$. For this purpose we shall use the method set forth in the book by A. N. Kolmogorov [6].

We shall introduce certain classes of distribution functions which in what follows we shall call <u>neighborhoods of type ϵ</u>. To determine them we take any subdivision $(x)_n$ of the interval $(-\infty, +\infty)$ into n parts:

(28) $$-\infty = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = +\infty .$$

On this subdivision we construct two step functions $U_1(x)$ and $U_2(x)$ of the following form:

(29) $$\begin{cases} U_1(x) = a_1 \text{ in the interval } (x_0, x_1), \\ \qquad = a_2 \text{ in } [x_1, x_2), \\ \qquad \dots \\ \qquad = a_n \text{ in } [x_{n-1}, x_n), \end{cases}$$

where $$0 = a_1 \leq a_2 \leq \dots \leq a_{n-1} \leq a_n \leq 1,$$

and

(30) $$\begin{cases} U_2(x) = b_1 \text{ in the interval } (x_0, x_1), \\ \qquad = b_2 \text{ in } [x_1, x_2), \\ \qquad \dots \\ \qquad = b_n \text{ in } [x_{n-1}, x_n), \end{cases}$$

where $$0 \leq b_1 \leq b_2 \leq \dots \leq b_{n-1} \leq b_n = 1.$$

Let the following conditions also be satisfied:

$$a_i \leq b_i \quad \text{for} \quad i = 1, 2, \dots, n$$

and

(31) $$\begin{cases} \omega_1 = b_1 - a_2 \geq 0 \\ \omega_2 = b_2 - a_3 > 0 \\ \omega_3 = b_3 - a_4 > 0 \\ \dots \\ \omega_{n-2} = b_{n-2} - a_{n-1} > 0 \\ \omega_{n-1} = b_{n-1} - a_n \geq 0 . \end{cases}$$

Then we say that the set of all distribution functions $\Phi(x)$ which satisfy the condition

$$(32) \qquad U_1(x) \leq \Phi(x) \leq U_2(x),$$

with the exception of the function $U_1(x)$, forms a class of distribution functions which we shall call a neighborhood of type ϵ or simply an ϵ-neighborhood.

For a given subdivision $(x)_n$ of the interval $(-\infty, +\infty)$ and a fixed N we have already considered another sort of class Δ_N of distribution functions which we called an N-cylinder. This class is defined analogously to the one of type ϵ by the choice of a certain subdivision $(x)_n$ for which it is supposed that every

$$(33) \qquad \omega_i = 0 \quad \text{for} \quad i = 1,2,\ldots,n-1$$

and, furthermore, every a_i and b_j is of the form $\frac{s}{N}$ where s is a non-negative integer, i.e., a_i and b_j are multiples of $\frac{1}{N}$.

We notice that within each fixed ϵ-neighborhood there will fall, for sufficiently large N, a certain finite number of N-cylinders. Every empirical distribution function belongs to some N-cylinder.

We shall call a class Ω of distribution functions a Borel class or a B-class if it is the sum of a countable number of ϵ-neighborhoods.

We are now in a position to define the probability $P(F_N \epsilon \Delta_N)$ when Δ_N is an N-cylinder.

To this end, if we define Δ_N by putting

$$a_1 = \frac{0}{N} = \frac{s_1}{N},\ a_2 = \frac{s_2}{N},\ \ldots,\ a_{n-1} = \frac{s_{n-1}}{N},\ a_n = \frac{s_n}{N},\ b_n = \frac{s_{n+1}}{N} = 1,$$

where s_i is an integer for $i = 1,2,\ldots,n+1$,

and introduce the new notations:

$$
\begin{aligned}
s_2 - s_1 &= m_1 & \qquad F(x_1) - F(x_0) &= p_1\\
s_3 - s_2 &= m_2 & F(x_2) - F(x_1) &= p_2\\
&\dots & &\dots\\
s_n - s_{n-1} &= m_{n-1} & F(x_{n-1}) - F(x_{n-2}) &= p_{n-1}\\
s_{n+1} - s_n &= m_n & F(x_n) - F(x_{n-1}) &= p_n\\
\sum_{i=1}^{n} m_i &= N & \sum_{i=1}^{n} p_i &= 1,
\end{aligned}
$$

then we can give

Definition 1.* We define that

$$P(F_N \in \Delta_N) = \frac{N!}{m_1!m_2!\dots m_n!}\, p_1^{m_1}\, p_2^{m_2} \dots p_n^{m_n} \,. \tag{34}$$

If we have a finite sum $\sum_k \Delta_N^{(k)}$, where the classes $\Delta_N^{(k)}$ include no functions in common, but are built on a common subdivision $(x)_n$, then we suppose that

$$P(F_N \in \sum_k \Delta_N^{(k)}) = \sum_k P(F_N \in \Delta_N^{(k)}). \tag{35}$$

Now we are in a position to define the probability $P(F_N \in \Omega)$ for an arbitrary class of distribution functions Ω . For this we take any subdivision $(x)_n$ of the interval $(-\infty, +\infty)$ and construct all possible

---

* Denote by $F_N^{(x_1,x_2,\dots,x_n)}(x)$ the empirical distribution function of the observations $x_1,x_2,\dots,x_N$; and by Ω_N the set of all points in N-dimensional real space such that $F_N^{(x_1,x_2,\dots,x_N)} \in \Omega$. It is not difficult to verify that the definition just given coincides with the definition of the N-dimensional distribution $P[(x_1,x_2,\dots,x_N) \in \Omega_N]$ where the random variables x_i are independent and have the distribution function $F(x)$. The latter definition is found in the book by A. N. Kolmogorov [6]. However, for our purposes the definition we have presented is more convenient.

N-cylinders on this subdivision. The number of different cylinders for this subdivision is finite and does not exceed $N^{n-1} = O(e^{n \ln N})$. We take the set-theoretic sum $\sum_k \Delta_N^{(k)}$ of all the N-cylinders so constructed which lie entirely within Ω. We introduce the notation:

$$S_x = P(F_N \in \sum_k \Delta_N^{(k)}) = \sum_k P(F_N \in \Delta_N^{(k)}).$$

It is not difficult to prove the following theorem:

<u>Theorem 7</u>. If the subdivision $(y)_m$ is a refinement of the subdivision $(x)_n$, then $S_y \geq S_x$. And, since $S_x \leq 1$, the $\sup\limits_{(x)_n} S_x = A \leq 1$ exists and has this property: for every $\epsilon > 0$ there exists a subdivision $(x)_n$ such that for it and all its refinements the following inequality holds:

$$A - \epsilon \leq S_x \leq A.$$

Proof. As may be seen easily, the stated property follows from the fact that in refining the subdivision $(x)_n$ into the subdivision $(y)_m$ every cylinder which belongs to Ω is divided into a finite sum of disjoint "refined" (i.e. belonging to the refined subdivision) cylinders and, in addition, there may appear in Ω a number of new "refined" cylinders. That the probability of F_N falling into a cylinder of the subdivision $(x)_n$ is equal to the sum of the probabilities of F_N falling into the "refined" cylinders into which this first cylinder is divided can be shown by induction on the number of points added by refinement. When one point is added to the subdivision the equality in the probabilities stated above follows from the identity: if $p_i = p_i' + p_i''$ then

$$\sum_{\substack{m_i'+m_i''=m_i \\ 0\leq m_i',\,0\leq m_i''}} \frac{N!}{m_1!\ldots m_{i-1}!m_i'!m_i''!m_{i+1}!\ldots m_n!} p_1^{m_1}\ldots p_{i-1}^{m_{i-1}} p_i'^{\,m_i'} p_i''^{\,m_i''} p_{i+1}^{m_{i+1}}\ldots p_n^{m_n}$$

$$= \frac{N!}{m_1!\ldots m_{i-1}!m_i!m_{i+1}!\ldots m_n!} p_1^{m_1}\ldots p_{i-1}^{m_{i-1}} p_{i+1}^{m_{i+1}}\ldots p_n^{m_n} \sum_{m_i'=0}^{m_i} C_{m_i}^{m_i'} p_i'^{\,m_i'} p_i''^{\,m_i-m_i'}$$

$$= \frac{N!}{m_1!m_2!\ldots m_i!\ldots m_n!} p_1^{m_1}\ldots p_{i-1}^{m_{i-1}}(p_i' + p_i'')^{m_i}\cdot p_{i+1}^{m_{i+1}}\ldots p_n^{m_n}.$$

The theorem just proven allows us to give

Definition 2. For an arbitrary class of functions Ω we define

$$(36) \qquad P(F_N \in \Omega) = \sup_{(x)_n} S_x = \sup_{(x)_n} \sum_{\Delta_N \in \Omega} P(F_N \in \Delta_N),$$

where Δ_N runs through the set of all disjoint cylinders constructed on the subdivision $(x)_n$ which lie entirely within Ω. In case Ω does not contain even one cylinder Δ_N for any subdivision $(x)_n$ we say that $P(F_N \in \Omega) = 0$.

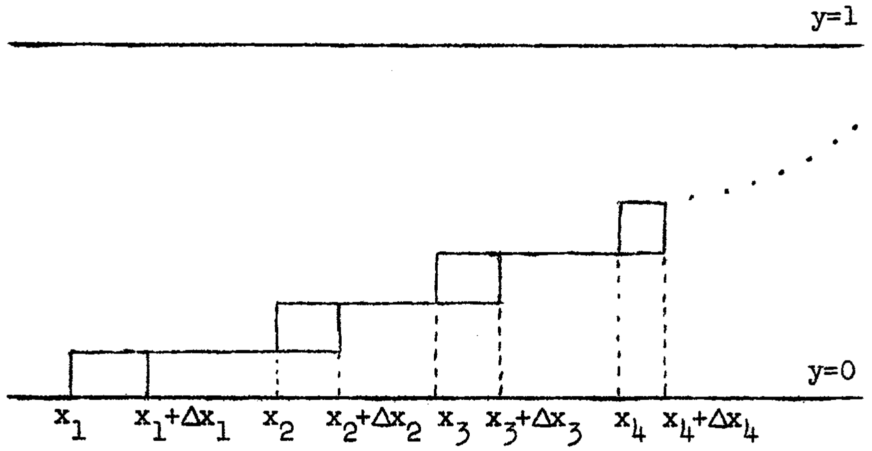
It is not difficult to put this expression for the probability $P(F_N \in \Omega)$ into the form of an integral. As the intervals (except those at the ends) of the subdivision $(x)_n$ are made infinitesimally small, S_x becomes for fixed N an integral. In order to write down this integral, we shall suppose that the intervals have been made so small that F_N does not have more than one jump in any one of them, but in all there are N jumps of height $\frac{1}{N}$ each. Let $F(x)$ have a continuous bounded derivative $p(x) = F'(x)$, where $p(x) \longrightarrow 0$ as $x \longrightarrow +\infty$ and as $x \longrightarrow -\infty$. When the intervals are so small we have, for the cylinder Δ_N with the subdivision $(x)_n$:

$$P(F_N \in \Delta_N) = N!\, \Delta_1 F(x_1) \Delta_2 F(x_2) \ \dots\ \Delta_N F(x_N)$$

$$= N!\, p(\xi_1) p(\xi_2) \ \dots\ p(\xi_N)\, \Delta_1 x_1\, \Delta_2 x_2 \ \dots\ \Delta_N x_N$$

where $\qquad \xi_i \in [x_i,\ x_i + \Delta_i x_i]$ and $x_1 < x_2 < \dots < x_N$.

In this case the cylinder Δ_N looks as follows :



Hence, for example, for a Borel class of functions, we obtain

$$P(F_N \in \Omega) = N! \overbrace{\int\int \quad \dots \int}^{N}_{\substack{F_N^{(x_1, x_2, \dots, x_N)} \in \Omega \\ x_1 < x_2 < \dots < x_N}} p(x_1)p(x_2)\dots p(x_N)dx_1 dx_2 \dots dx_N$$

(37) $$= \overbrace{\int\int \dots \int}^{N}_{F_N^{(x_1, x_2, \dots, x_N)} \in \Omega} p(x_1)p(x_2)\dots p(x_N)dx_1 dx_2 \dots dx_N$$

By $F_N^{(x_1,x_2,\dots,x_N)}$ we mean the empirical distribution function of the observations $x_1, x_2, \dots, x_N$.

It will be convenient to denote by Ω_N the N-dimensional domain of variation of the variables $x_1, x_2, \dots, x_N$ defined by the inclusion $F_N^{(x_1,x_2,\dots,x_N)} \in \Omega$. It is evident that the region Ω_N is invariant with respect to any permutation of the variables $x_1, x_2, \ldots, x_N$. In these terms

$$(38) \qquad P(F_N \in \Omega) = \overbrace{\iint\cdots\int}^{N}_{(x_1, x_2, \ldots, x_N) \in \Omega_N} p(x_1)p(x_2)\ldots p(x_N)dx_1 dx_2 \ldots dx_N \ .$$

The methods by which the classes of functions Ω may be defined are quite varied. We shall note several of these possibilities.

1. Let there be given any distribution function $\Phi_0(x)$ such that $\Phi_0(x) = 1$ for $x \geq \alpha$.

Ω is the class of all distribution functions $\Phi(x)$ such that $\Phi(x) \leq \Phi_0(x)$ for all x.

2. Let there be given any non-negative function f(x) which is continuous in the interval $(-\infty, +\infty)$.

Ω is the class of all distribution functions $\Phi(x)$ such that $\int_{-\infty}^{\infty} f(x)d\Phi(x) \geq \alpha$ or this integral diverges.

3. Let there be given several pairs of numbers

$$\alpha_1 < \beta_1, \ \alpha_2 < \beta_2, \ldots, \alpha_k < \beta_k; \ \beta_1 < \beta_2 < \ldots < \beta_k \leq 1; \ \alpha_i \geq 0$$

and let there be chosen k values x, $x_1 < x_2 < \ldots < x_k$.

Ω is the class of all distribution functions $\Phi(x)$ which satisfy the condition

$$\alpha_i \leq \Phi(x_i) \leq \beta_i \quad \text{for} \quad i = 1,2,\ldots,k.$$

4. Consider a class of distribution functions Ω_α of the following sort: it consists of all distribution functions $\Phi(x)$ such that $\Phi(x) = 0$ for $x \leq \lambda$ and $\Phi(x) = 1$ for $x \geq \lambda + \alpha$, where α is a constant but λ depends on $\Phi(x)$.

5. Let Ω_α be the class of all distribution functions $\Phi(x)$ such that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi < -\alpha$, where $\alpha > 0$ is a fixed number and $F(x)$ is a given distribution function.

One is easily convinced that in all these examples the class of functions Ω is a Borel class, and if there exists a continuous bounded derivative $F'(x) = p(x)$ then $P(F_N \in \Omega)$ can be written in the form of an N-dimensional integral.

We shall calculate, for example, the probability $P(F_N \in \Omega_\alpha)$, where $\alpha < 1$, for the class Ω_α of Example 4 in the case where

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases} .$$

In order to solve this problem we shall proceed in the following manner. We divide the interval $[0, 1]$ into n equal parts (n an integer). We consider a series of N-cylinders. The cylinder A_s is defined by two vertices $(\frac{s-1}{n}, 0)$ and $(\frac{s-1-[-\alpha n]}{n}, 1)$. The cylinder B_s is defined by two vertices $(\frac{s-1}{n}, 0)$ and $(\frac{s-[-\alpha n]}{n}, 1)$. Evidently

$$\bigcup_{s=1}^{n+2+[-\alpha n]} A_s \subset \Omega_\alpha \subset \bigcup_{s=1}^{n+[-\alpha n]} B_s .$$

Hence

$$\text{(H)} \qquad P(F_N \in \bigcup_{s=1}^{n+2+[-\alpha n]} A_s) \leq P(F_N \in \Omega_\alpha) \leq P(F_N \in \bigcup_{s=1}^{n+[-\alpha n]} B_s),$$

where $\bigcup_{s=1}^{n+2+[-\alpha n]} A_s = A_1 + (A_2 - A_1) + (A_3 - A_2) + \dots + (A_{n+2+[-\alpha n]} - A_{n+1+[-\alpha n]})$.

The last sum consists of disjoint sets. Thus

$$P(F_N \in \bigcup_s A_s) = P(F_N \in A_1) + \sum_{s=2}^{n+2+[-\alpha n]} P[F_N \in (A_s - A_{s-1})].$$

On the basis of the definition of the probability $P(F_N \in \Delta_N)$, where Δ_N is an N-cylinder,

$$P(F_N \in A_1) = \left(\frac{-[-\alpha n]-1}{n}\right)^N.$$

The probability $P[F_N \in (A_s - A_{s-1})]$ is the same for all s, so we calculate this expression for $s = 2$.

$A_2 - A_1$ is divisible into the sum of N disjoint N-cylinders K_t. K_t has the vertices

$$\left(\frac{1}{n}, 0\right), \left(\frac{-[-\alpha n]-1}{n}, \frac{t}{N}\right) \text{ and } \left(\frac{-[-\alpha n]}{n}, 1\right), \text{ where } t = 0,1,\ldots,N-1,$$

and so

$$P[F_N \in (A_2 - A_1)] = \sum_{t=0}^{N-1} P(F_N \in K_t)$$

$$= \sum_{t=0}^{N-1} \frac{N!}{t!N-t!} \left(\frac{-[-\alpha n]-2}{n}\right)^t \left(\frac{1}{n}\right)^{N-t}.$$

Hence

$$P(F_N \in \bigcup_{s=1}^{n+2+[-\alpha n]} A_s) = \left(\frac{-[-\alpha n]-1}{n}\right)^N + \frac{n+1+[-\alpha n]}{n}$$

$$\cdot N\left(\frac{-[-\alpha n]-2}{n}\right)^{N-1} + O\left(\frac{N!}{n}\right).$$

This last expression, as $n \longrightarrow \infty$, tends to

$$\alpha^N + N\alpha^{N-1}(1-\alpha).$$

Similarly we obtain

$$P(F_N \in \bigcup_{s=1}^{n+[-\alpha n]} B_s) = \left(\frac{-[-\alpha n]+1}{n}\right)^N + \frac{n-1+[-\alpha n]}{n}$$

$$\cdot N\left(\frac{-[-\alpha n]}{n}\right)^{N-1} + O\left(\frac{N!}{n}\right).$$

This expression also tends to $\alpha^N + N\alpha^{N-1}(1-\alpha)$ as $n \to \infty$. It then follows from inequality (H) that

$$P(F_N \in \Omega_\alpha) = \alpha^N + N\alpha^{N-1}(1-\alpha). \tag{39}$$

3. F-Properties of the Class of Functions Ω.

We consider the class of functions Ω. Let $F(x)$ be the distribution function of a random variable ξ which does not take on only a finite set of values. Let $F_N(x)$ be the empirical distribution function of the results of N independent observations on the random variable ξ.

We introduce the notation:

$$\omega_F(\Omega) = \sup_{\Phi \in \Omega} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi.$$

Let $\Omega^{(N)}$ be the class of all possible empirical distribution functions belonging to the class Ω which could result from N independent observations on the random variable ξ under consideration.

In order to consider the covering of the class Ω by ϵ-neighborhoods it is natural to introduce the following:

Definition 3. We say that the class of functions Ω is F-distinguishable if the following conditions are satisfied:

1) $\omega_F(\Omega) > -\infty$.

2) For every $\eta > 0$ there is an ϵ-neighborhood V such that $V \subset \Omega$ and $\omega_F(V) > \omega_F(\Omega) - \eta$.

3) For every $\eta > 0$ and every N there is a finite number $K = K(\eta, n)$ of ϵ-neighborhoods $V_1, V_2, \ldots, V_K$ such that $\Omega^{(N)} \subset \bar{\Omega} = \bigcup_{i=1}^{K} V_i$ and $\omega_F(V_i) < \omega_F(\Omega) + \eta$ for $i = 1,2,\ldots,K$. Moreover, if we denote by n_i the

number of points in the subdivision on which the ϵ-neighborhood V_i is constructed, and put $\bar{n} = \max_{1 \leq i \leq K} n_i$, $\bar{n} = \bar{n}(\eta, n)$, then for every fixed η the following conditions are satisfied:

$$\ln K = o(N) \quad \text{and} \quad \bar{n} = o\left(\frac{N}{\ln N}\right).$$

When the classes Ω of Examples 1-5 are studied, it is not difficult to see that those of Examples 1-4 are F-distinguishable, but the class Ω_α, where $\alpha > 0$, of Example 5 is not F-distinguishable. This last we shall show later.

Furthermore, let $\Phi(x)$ be any distribution function such that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ exists.

<u>Definition 4</u>. We shall say that we have <u>a sequence of ϵ-neighborhoods V_n which include $\Phi(x)$ and F-converge to $\Phi(x)$</u> if the following conditions are satisfied:

1) There exists a sequence of subdivisions $(x)_n$ of type (28) of the interval $(-\infty, +\infty)$ such that for this sequence of subdivisions

$$\lim_{n \to \infty} \sum_{i=1}^{n} \ln \frac{\Delta_i F}{\Delta_i \Phi} \Delta_i \Phi = \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi, \tag{40}$$

where $\Delta_i \Phi > 0$ and $\Delta_i F > 0$ for $i = 1,2,\ldots,n$, and the points x_i of the subdivision $(x)_n$ are points of continuity of $\Phi(x)$ and $F(x)$.

2) There exist two sequences of positive numbers η_n and τ_n such that

$$\lim_{n \to \infty} \eta_n = \lim_{n \to \infty} \tau_n = 0.$$

3) The ϵ-neighborhood V_n is constructed on the subdivision $(x)_n$, and the two step-functions of (29) and (30) are so constructed that the inequalities

$$a_{i+1} \leq \Phi(x_i) \leq b_i, \quad i = 1,2,\ldots,n-1,$$

are satisfied, where unless i=1 or i=n-1 the sign of equality is excluded.

4) For any distribution function $\Phi_1(x) \epsilon V_n$ and for the subdivision $(x)_n$ we have the inequalities

$$(41) \qquad \left| \sum_{i=1}^{n} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1} \right| \leq \eta_n \quad \text{and} \quad \left| \sum_{i=1}^{n} (\Delta_i\Phi_1 - \Delta_i\Phi) \ln \frac{\Delta_i F}{\Delta_i\Phi} \right| \leq \tau_n .$$

The last condition puts bounds only on the values of $\Phi_1(x)$ at the points x_i for $i = 1,2,\ldots,n$; hence it can always be satisfied by choosing sufficiently small values for the differences $\omega_i = b_i - a_{i+1}$, $i = 1,2,\ldots,n-1$. We shall also put $\omega_0 = 0$ and $\omega_n = 0$.

It is possible to prove the following theorem:

Theorem 8. On any sequence of subdivisions $(x)_n$ of the interval $(-\infty, +\infty)$ which satisfies condition 1) of Definition 4 it is possible to construct a sequence of ϵ-neighborhoods V_n which include $\Phi(x)$ and F-converge to $\Phi(x)$ by taking for each n two sequences of numbers a_i and b_i which satisfy condition 3) of Definition 4 and inequality (31) and are such that, furthermore, the sequence ω_i satisfies the following conditions as n increases:

$$(42) \qquad \lim_{n\to\infty} \sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} = 0 \text{ and } \lim_{n\to\infty} \sum_{i=1}^{n} (\omega_{i-1}+\omega_i) \left| \ln \frac{\Delta_i F}{\Delta_i\Phi} \right| = 0.$$

Proof. We introduce the notations:

$$\sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} = \gamma_n, \quad \text{and} \quad \sum_{i=1}^{n} (\omega_{i-1}+\omega_i) \left| \ln \frac{\Delta_i F}{\Delta_i\Phi} \right| = \rho_n,$$

where $\gamma_n > 0$, $\rho_n > 0$, and γ_n and $\rho_n \to 0$ as $n \to \infty$.

By proper choice of the ϵ-neighborhood V_n at the points x_i we have, for any function $\Phi_1(x) \epsilon V_n$,

$$|\Phi_1(x_i) - \Phi(x_i)| \leq \omega_i \quad \text{for } i = 1,2,\ldots,n-1$$

and
$$\Phi_1(x_0) = \Phi(x_0),\ \Phi_1(x_n) = \Phi(x_n).$$

Hence

$$|\Delta_i\Phi_1 - \Delta_i\Phi| = |\Phi_1(x_i) - \Phi_1(x_{i-1}) - \Phi(x_i) + \Phi(x_{i-1})| \leq \omega_{i-1} + \omega_i .$$

For any function $\Phi_1(x) \in V_n$ we have:

$$\sum_{i=1}^{n} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1} = \sum_{i\in S_1} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1} + \sum_{i\in S_2} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1} ,$$

where S_1 is the set of all i for which $\Delta_i\Phi_1 \geq \Delta_i\Phi$ and S_2 is the set of all i for which $\Delta_i\Phi_1 < \Delta_i\Phi$.

We use the inequality

(43) $$0 \leq \ln (1+x) \leq x \quad \text{for} \quad x \geq 0.$$

For $i \in S_1$,

$$\left|\Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| = \left|\Delta_i\Phi_1 \ln \frac{\Delta_i\Phi_1}{\Delta_i\Phi}\right|$$

$$= \Delta_i\Phi_1 \ln\left(1 + \frac{\Delta_i\Phi_1 - \Delta_i\Phi}{\Delta_i\Phi}\right)$$

$$\leq \Delta_i\Phi_1 \frac{\Delta_i\Phi_1 - \Delta_i\Phi}{\Delta_i\Phi}$$

$$= \frac{(\Delta_i\Phi_1 - \Delta_i\Phi)^2}{\Delta_i\Phi} + \Delta_i\Phi_1 - \Delta_i\Phi.$$

Hence we obtain

$$\left|\sum_{i\in S_1} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| \leq \sum_{i\in S_1} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} + \sum_{i\in S_1} (\omega_{i-1}+\omega_i).$$

Analogously, for $i\in S_2$,

$$\left|\Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| \leq \Delta_i\Phi_1 \ln\left(1 + \frac{\Delta_i\Phi - \Delta_i\Phi_1}{\Delta_i\Phi_1}\right)$$

$$\leq \Delta_i\Phi_1 \frac{\Delta_i\Phi - \Delta_i\Phi_1}{\Delta_i\Phi_1}$$

$$= |\Delta_i\Phi_1 - \Delta_i\Phi|.$$

This yields:

$$\left|\sum_{i\in S_2} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| \leq \sum_{i\in S_2} (\omega_{i-1}+\omega_i).$$

Combining these inequalities, we obtain:

$$\left|\sum_{i=1}^{n} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| \leq \sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} + \sum_{i=1}^{n} (\omega_{i-1}+\omega_i) .$$

Here we notice that on applying the Schwarz-Bunyakovsky inequality we have:

$$\sum_{i=1}^{n} (\omega_{i-1}+\omega_i) = \sum_{i=1}^{n} \frac{\omega_{i-1}+\omega_i}{(\Delta_i\Phi)^{1/2}} (\Delta_i\Phi)^{1/2} \leq \left(\sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} \sum_{i=1}^{n} \Delta_i\Phi\right)^{1/2} .$$

And since $\sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} = \gamma_n$ and $\sum_{i=1}^{n} \Delta_i\Phi = 1$ we obtain finally:

$$\left|\sum_{i=1}^{n} \Delta_i\Phi_1 \ln \frac{\Delta_i\Phi}{\Delta_i\Phi_1}\right| \leq \gamma_n + (\gamma_n)^{1/2} \longrightarrow 0 \text{ as } n \longrightarrow \infty,$$

and

$$\left|\sum_{i=1}^{n} (\Delta_i\Phi_1-\Delta_i\Phi) \ln \frac{\Delta_i F}{\Delta_i\Phi}\right| \leq \sum_{i=1}^{n} (\omega_{i-1}+\omega_i)\left|\ln \frac{\Delta_i F}{\Delta_i\Phi}\right| = \rho_n \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Thus conditions (2) and (4) of Definition 4 are satisfied if we take $\eta_n = \gamma_n + (\gamma_n)^{1/2}$ and $\tau_n = \rho_n$.

<u>Remark</u>. Under certain quite reasonable conditions the second inequality of (42) is superfluous. In fact, if for all subdivisions $(x)_n$ the sums $\sum_{i=1}^{n} (\ln \frac{\Delta_i F}{\Delta_i\Phi})^2 \Delta_i \Phi$ remain bounded above (which will, for example, always be the case if the integral $\int_{-\infty}^{\infty} (\ln \frac{dF}{d\Phi})^2 d\Phi$ exists), then the second inequality follows from the first. In fact,

$$\sum_{i=1}^{n} (\omega_{i-1}+\omega_i)\left|\ln \frac{\Delta_i F}{\Delta_i\Phi}\right| = \sum_{i=1}^{n} \frac{\omega_{i-1}+\omega_i}{(\Delta_i\Phi)^{1/2}} \left|\ln \frac{\Delta_i F}{\Delta_i\Phi}\right| (\Delta_i\Phi)^{1/2}$$

$$\leq \left\{\sum_{i=1}^{n} \frac{(\omega_{i-1}+\omega_i)^2}{\Delta_i\Phi} \sum_{i=1}^{n} (\ln \frac{\Delta_i F}{\Delta_i\Phi})^2 \Delta_i\Phi\right\}^{1/2} ,$$

and as $\gamma_n \longrightarrow 0$ then, under the condition stated above, this last expression tends to zero.

The concept of F-convergence of a sequence of ϵ-neighborhoods V_n to a curve $\Phi(x)$ will be clarified by the following theorem.

Theorem 9. Let $F(x)$ and $\Phi(x)$ be two distribution functions, where $F(x)$ is the distribution function of a random variable ξ which can take on infinitely many values. Assume that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi$ exists. Then, for every sequence of ϵ-neighborhoods V_n which include $\Phi(x)$ and F-converge to it, if the distribution function $\Phi^{(n)} \epsilon V_n$, and $(x)_n$ is the subdivision of condition 1) of Definition 4 on which the ϵ-neighborhood V_n is constructed, for this subdivision

$$(44) \qquad \sum_{i=1}^{n} \Delta_i\Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i\Phi^{(n)}} = \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi + \mu_n,$$

where $|\mu_n| < \lambda_n$ and the sequence of numbers $\lambda_n \longrightarrow 0$ and does not depend on the choice of the function $\Phi^{(n)}$ in the neighborhood V_n.

Proof. We have

$$\sum_{i=1}^{n} \Delta_i\Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i\Phi^{(n)}} = \sum_{i=1}^{n} \Delta_i\Phi \ln \frac{\Delta_i F}{\Delta_i \Phi} + \sum_{i=1}^{n} (\Delta_i\Phi^{(n)} - \Delta_i\Phi) \ln \frac{\Delta_i F}{\Delta_i\Phi}$$
$$+ \sum_{i=1}^{n} \Delta_i\Phi^{(n)} \ln \frac{\Delta_i\Phi}{\Delta_i\Phi^{(n)}} .$$

By inequality (41), the second term is in absolute value less than τ_n and the third is in absolute value less than η_n. Hence

$$\left|\sum_{i=1}^{n} \Delta_i\Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i\Phi^{(n)}} - \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi\right| \leq \left|\sum_{i=1}^{n} \Delta_i\Phi \ln \frac{\Delta_i F}{\Delta_i\Phi} - \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi\right| + \tau_n + \eta_n$$
$$= \lambda_n \longrightarrow 0$$

Q.E.D.

4. Basic Theorems for the General Case.

Having performed the necessary preliminary investigations sections 1-3 we are in a position to formulate basic theorems for the general case.

Theorem 10. Let $F(x)$ be the distribution function of a random variable ξ, and $F_N(x)$ the empirical distribution function after N independent observations. Let $\Phi(x)$ be another distribution function such that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ exists. Let V_n be a sequence of ϵ-neighborhoods which include $\Phi(x)$ and F-converge to it. Then

$$(45) \qquad P(F_N \in V_n) = e^{N\left[\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi + \delta_n + O\left(\frac{n \ln N}{N}\right)\right]},$$

where there exists a sequence of positive numbers λ_n such that $\lim_{n \to \infty} \lambda_n = 0$ and $|\delta_n| < \lambda_n$.

Proof. For every sufficiently large fixed N V_n contains at least one and no more than $N^{n-1} = e^{O(n \ln N)}$ different N-cylinders Δ_N which are constructed on the same subdivision $(x)_n$ as is V_n. Among these cylinders there exists a cylinder Δ_N' such that for it $P(F_N \in \Delta_N)$ is maximized. Then

$$(45') \qquad P(F_N \in \Delta_N') \leq P(F_N \in V_n) = \sum_{\Delta_N \in V_n} P(F_N \in \Delta_N) \leq N^{n-1} P(F_N \in \Delta_N').$$

The piecewise-extremal curve $\Phi^{(n)}$ passes through the vertices of the N-cylinder Δ_N'. Evidently $\Phi^{(n)} \in V_n$.

If in formula (34) of Definition 1 we put $p_i = \Delta_i F$ and $m_i = N \Delta_i \Phi^{(n)}$, where the difference Δ_i is based on the subdivision $(x)_n$, and use formula (3) with estimate (4) we obtain:

$$(46) \qquad P(F_N \in \Delta_N') = e^{N \sum_{i=1}^{n} \Delta_i \Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i \Phi^{(n)}} + R},$$

where

$$|R| < (n+1)\left(\frac{1}{2}\ln N + \frac{1}{2}\ln 2\pi + \frac{1}{12N}\right).$$

Since $\Phi^{(n)} \in V_n$ and the sum $\sum_{i=1}^{n} \Delta_i \Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i \Phi^{(n)}}$ is based on the subdivision $(x)_n$, we may apply Theorem 9, which shows that there exists a sequence of numbers $\lambda_n > 0$ such that $\lim_{n \to \infty} \lambda_n = 0$ and the difference

$$(47) \qquad \delta_n = \sum_{i=1}^{n} \Delta_i \Phi^{(n)} \ln \frac{\Delta_i F}{\Delta_i \Phi^{(n)}} - \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$$

satisfies the condition $|\delta_n| < \lambda_n$ for all distribution functions $\Phi^{(n)} \in V_n$. Using inequality (45) and equalities (46) and (47) we can prove Theorem 10 without difficulty.

A second important theorem of the general case is:

Theorem 11. Let $F(x)$ be the distribution function of a random variable ξ, and let $F_N(x)$ be the empirical distribution function of the results of N independent observations on the random variable ξ. Let Ω be an F-distinguishable class of distribution functions. Then

$$(48) \qquad P(F_N \in \Omega) = e^{N[\omega_F(\Omega) + o(1)]}$$

where

$$\omega_F(\Omega) = \sup_{\Phi \in \Omega} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi.$$

Proof. We shall first prove this theorem for the case where Ω coincides with an ϵ-neighborhood ($\Omega = V$). For N sufficiently large and fixed there enter into the ϵ-neighborhood V each time the largest possible number, say s, of disjoint cylinders $\Delta_{N,k}$ with the same subdivision $(x)_n$ as V, so that

$$P(F_N \in V) = \sum_{k=1}^{s} P(F_N \in \Delta_{N,k}).$$

By definition (34) and inequality (4)

$$P(F_N \in \Delta_{N,k}) = e^{N\left[\sum_{i=1}^{n} \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i) + O\left(\frac{n \ln N}{N}\right)\right]}.$$

Here we have used the notation adopted for the definition of the cylinder ($a_{n+1} = b_n$).

We notice that $\sum_{i=1}^{n} \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i)$ is identical with $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi$ for the case where Φ is the piecewise-extremal function of this integral which passes through the vertices of the cylinder $\Delta_{N,k}$.

Since $\Phi \in \Delta_{N,k} \subset V = \Omega$,

$$\sum_{i=1}^{n} \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i) \leq \sup_{\Phi \in V} \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi = \omega_F(V).$$

Hence $$P(F_N \in \Delta_{N,k}) \leq e^{N\left[\omega_F(V) + O\left(\frac{n \ln N}{N}\right)\right]}.$$

Since $s < N^{n-1}$ we obtain:

$$P(F_N \in V) \leq e^{N\left[\omega_F(V) + O\left(\frac{n \ln N}{N}\right)\right]}.$$

The function which maximizes $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi} d\Phi$ for $\Phi \in V$ is the piecewise-extremal one passing through n points. At all the points of the subdivision $(x)_n$, for N sufficiently large, it differs from the vertices of some cylinder Δ'_N of the subdivision by not more than $1/N$. We consider those intervals $[x_n, x_{k+1}]$ of our subdivision where $\Delta_k F \neq 0$. (Here we require that in those intervals where $\Delta_k F = 0$ all the corresponding terms

$\sum_{i=1}^{n} \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i)$ must be simply equal to zero for the cylinder Δ_N', the choice of such a cylinder Δ_N' being possible for N sufficiently large.)

For the cylinder Δ_N'

$$\sum_{i=1}^{n} \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i) - \omega_F(V) = 0(\frac{1}{N}) \ ,$$

$$P(F_N \in V) \geq P(F_N \in \Delta_N') = e^{N[\Sigma \ln \frac{\Delta_i F}{a_{i+1}-a_i} (a_{i+1}-a_i) + 0(\frac{n \ln N}{N})]}$$

$$= e^{N[\omega_F(V) + 0(\frac{1}{N}) + 0(\frac{n \ln N}{N})]} \ .$$

Combining this with the preceding, we obtain

$$P(F_N \in V) = e^{N[\omega_F(V) + 0(\frac{n \ln N}{N})]} \ .$$

We now pass on to the general case. We shall write:

$$P(F_N \in \Omega) = e^{N[\omega_F(\Omega) + \gamma_N]} \ ,$$

$$\gamma_N = \frac{\ln P(F_N \in \Omega)}{N} - \omega_F(\Omega).$$

We shall show that $\gamma_N \longrightarrow 0$. For this it is necessary to show that for every $\eta > 0$ there is an N_0 such that if $N > N_0$ then $|\gamma_N| < \eta$. For every $\eta > 0$, according to Definition 3, there is a neighborhood $\mathcal{E} \subset \Omega$ corresponding to a subdivision of the interval $(-\infty, +\infty)$ into n parts such that $\omega_F(\mathcal{E}) > \omega_F(\Omega) - \frac{\eta}{2}$, and also a set-theoretic sum $\overline{\Omega}_\eta$ of $k = k_{\frac{\eta}{2}}$ ϵ-neighborhoods $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ such that $\Omega \subset \overline{\Omega}_\eta$ and $\omega_F(\mathcal{E}_i) < \omega_F(\Omega) + \frac{\eta}{2}$ for $i = 1,2,\ldots,k$. Now because of the inclusions $\mathcal{E} \subset \Omega \subset \overline{\Omega}_\eta$ we have

$$P(F_N \in \mathcal{E}) \leq P(F_N \in \Omega) \leq P(F_N \in \overline{\Omega}_\eta),$$

and thus

$$\frac{\ln P(F_N \in \mathcal{E})}{N} - \omega_F(\Omega) \leq \frac{\ln P(F_N \in \Omega)}{N} - \omega_F(\Omega)$$

$$\leq \frac{\ln P(F_N \in \overline{\Omega}_\eta)}{N} - \omega_F(\Omega).$$

We denote by n_1 the maximum number of sub-intervals in any of the subdivisions $(x)_n$ which underlie the ϵ-neighborhoods $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$;

$$P(F_N \in \mathcal{E}) = e^{N[\omega_F(\mathcal{E}) + O(\frac{n \ln N}{N})]},$$

$$P(F_N \in \overline{\Omega}_\eta) \leq \sum_{i=1}^{k} P(F_N \in \mathcal{E}_i)$$

$$\leq k \max_i P(F_N \in \mathcal{E}_i)$$

$$= k\, e^{N[\max_i \omega_F(\mathcal{E}_i) + O(\frac{n_1 \ln N}{N})]}.$$

Hence

$$\omega_F(\mathcal{E}) - \omega_F(\Omega) + O(\frac{n \ln N}{N}) \leq \frac{\ln P(F_N \in \Omega)}{N} - \omega_F(\Omega)$$

$$\leq \max_i \omega_F(\mathcal{E}_i) - \omega_F(\Omega) + O(\frac{n_1 \ln N}{N}) + \frac{\ln k}{N},$$

and furthermore

$$-\frac{\eta}{2} + O(\frac{n \ln N}{N}) \leq \frac{\ln P(F_N \in \Omega)}{N} - \omega_F(\Omega) \leq \frac{\eta}{2} + O(\frac{n_1 \ln N}{N}) + \frac{\ln k}{N}.$$

For sufficiently large $N (N > N_0)$ it is evident that the inequality $-\eta < \gamma_N < \eta$ will be satisfied, Q.E.D.

In order to determine the order of convergence of the quantities γ_N to zero a greater concretization of the character of the covering of the class Ω by neighborhoods of type ϵ seems to be necessary.

In the case where the function $F(x)$ has a continuous bounded derivative $p(x) = F'(x)$ it is possible to present this theorem in a different form. We denote by $F_N^{(x_1,x_2,\ldots,x_N)}(x)$ the empirical distribution function corresponding to the set of observed values $x_1,x_2,\ldots,x_N$. We denote by Ω_N the set of all those points in N-dimensional space such that $F_N^{(x_1,x_2,\ldots,x_N)} \epsilon \Omega$, where Ω is some class of functions. Then we have

Theorem 12. Let $F(x)$ be a continuous distribution function with a continuous bounded derivative $p(x)$, and let Ω be an F-distinguishable class of distribution functions. Then

$$\lim_{N\to\infty}\left(\overbrace{\iint\cdots\int}^{N}_{F_N^{(x_1,x_2,\ldots,x_N)}\epsilon\,\Omega} p(x_1)p(x_2)\ldots p(x_N)dx_1\, dx_2\, \ldots\, dx_N\right)^{\frac{1}{N}}$$

$$(49) \qquad = \lim_{N\to\infty}\left(\overbrace{\iint\cdots\int}^{N}_{\Omega_N} p(x_1)p(x_2)\ldots p(x_N)dx_1\, dx_2\, \ldots\, dx_N\right)^{\frac{1}{N}}$$

$$= e^{\sup\limits_{\Phi\epsilon\Omega}\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi} .$$

This theorem is a direct consequence of the preceding theorem.

From Theorem 12 it follows that the class of functions Ω_α of Example 5 of Section 2, for $\alpha > 0$, is not F-distinguishable. Indeed, for this class $\omega_F(\Omega) = -\alpha$. However, $P(F_N \epsilon \Omega_\alpha) = 1$ for every α and every N. This can be seen from the fact that every function F_N can be included in a cylinder Δ_N which lies entirely within Ω_α and therefore every empirical distribution function F_N belongs to Ω_α for every $\alpha \geq 0$.

Because of this circumstance Theorem 12 is not true for the class Ω_α, and therefore the class Ω_α is not F-distinguishable.

5. Continuous Case.

We mention several peculiarities of the special case where the functions $F(x)$ and $\Phi(x)$ have continuous or piecewise-continuous derivatives $F'(x)$ and $\Phi'(x)$. We shall suppose that in every interval $[\alpha, \beta]$ where $F'(x) = 0$ it will also occur that $\Phi'(x) = 0$. In these intervals we shall put $\ln \frac{F'(x)}{\Phi'(x)} \Phi'(x) = 0$. Then the singular points of the function $\ln \frac{F'(x)}{\Phi'(x)} \Phi'(x)$ can only be zeros of $F'(x)$.

For the cases important in practice it is usually sufficient to consider the case where $F'(x)$ has only a finite number of zeros in its interval of variation $[a, b]$. (We assume that $F'(x) = 0$ outside the interval $[a, b]$.) We shall call a zero of $F'(x)$ or an endpoint of the interval of variation $[a, b]$ an <u>admissible singular point</u> of $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ if for every interval $[x', x'']$, $x'' - x' = \delta > 0$, which contains this singular point,

$$\lim_{\delta \to 0} \ln \frac{F(x'')-F(x')}{\Phi(x'')-\Phi(x')} \left[\Phi(x'')-\Phi(x')\right] = 0.$$

Then it is not difficult to show that if the function $\ln \frac{F'(x)}{\Phi'(x)} \Phi'(x)$ has a finite number of admissible singular points and if the improper integral (improper both as to range and to singular points) $\int_{-\infty}^{\infty} \ln \frac{F'(x)}{\Phi'(x)} \Phi'(x)dx$ converges, then $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ converges and equals

$$\int_{-\infty}^{\infty} \ln \frac{F'(x)}{\Phi'(x)} \Phi'(x)dx .$$

For the case of such a pair of functions $F'(x)$ and $\Phi'(x)$ it is not difficult to formulate all the basic results just as in the general case, replacing $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ by $\int_{-\infty}^{\infty} \ln \frac{F'(x)}{\Phi'(x)}\, \Phi'(x)dx$. All this can be done simply enough.

III. PROBLEMS OF LARGE DEVIATIONS

The results obtained in Theorems 10 and 11 have a direct application to the problem of the definition of the probability of large deviations. They permit the evaluation of the character of the tendency to zero of this probability. For the case of independent trials in the simple scheme with n outcomes, Theorems 2 and 4 permit the estimation of $P[F(\nu_1,\nu_2,\ldots,\nu_n) \geq \alpha]$; here $\nu_1,\nu_2,\ldots,\nu_n$ are the observed proportions of each outcome after N independent trials and $F(x_1,x_2,\ldots,x_n)$ is a function which has continuous third derivatives with respect to $x_1,x_2,\ldots,x_n$ in the simplex where $x_1+x_2+\ldots+x_n \leq 1$, $x_i \geq 0$. In this way one may obtain asymptotic formulas for this probability.

For the continuous case Theorems 10 and 11 open up the possibility of studying the probability of large deviations of the empirical distribution function $F_N(x)$ of the random variable ξ with continuous distribution function $F(x)$ from its theoretical distribution function.

In particular, in this way it is easy to obtain the following facts:

For every $\epsilon > 0$

$$\lim_{N\to\infty} P(\sup_x |F_N(x) - F(x)| > \epsilon) = 0, \tag{50}$$

and for every $A > 0$ and $\eta > 0$,

$$\lim_{N\to\infty} P(\sup_x |F_N(x) - F(x)| > AN^{-\frac{1}{2}+\eta}) = 0. \tag{51}$$

An important consequence of the basic theorems is

Theorem 13. Let $F_N(x)$ be the empirical distribution function of the random variable ξ after N independent trials; let $F(x)$ be the distribution function of the random variable ξ; and let $\Phi(x)$ be another distribution function such that $\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi$ exists. Then, for every $\epsilon > 0$,

$$(52) \qquad P(\sup_x |F_N(x)-\Phi(x)| < \epsilon) = e^{N\left[\int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi + \gamma_\epsilon + \mu_{N,\epsilon}\right]}$$

where

$$\lim_{\epsilon \to 0} \gamma_\epsilon = 0, \quad \lim_{\epsilon \to 0} \lim_{N \to \infty} \mu_{N,\epsilon} = 0;$$

and hence

$$(53) \qquad \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \lim_{N \to \infty} \frac{\ln P(\sup_x |F_N(x)-\Phi(x)| < \epsilon)}{N} = \int_{-\infty}^{\infty} \ln \frac{dF}{d\Phi}\, d\Phi.$$

We notice that certain problems about the distributions of various random variables can be considered as problems of types which have been treated in the present paper.

For example, the distribution of the random variable

$$\eta = x_1 + x_2 + \dots + x_N,$$

where the x_i's are independent and identically distributed random variables with the distribution function $F(x)$, can be considered in the following manner:

$$\frac{\eta}{N} = \sum_{i=1}^{N} \frac{x_i}{N} = \int_{-\infty}^{\infty} x \, d\, F_N(x),$$

where $F_N(x)$ is the empirical distribution function of the random variable ξ with distribution function $F(x)$ after N independent observations. Hence

$$(54) \qquad P(\eta \geq \alpha N) = P(\frac{\eta}{N} \geq \alpha) = P(\int_{-\infty}^{\infty} x dF_N(x) \geq \alpha) = P(F_N \in \Omega_\alpha),$$

where Ω_α is the class of all distribution functions $\Phi(x)$ for which $\int_{-\infty}^{\infty} x \, d\,\Phi(x) \geq \alpha$. The asymptotic probability of large deviations of the random variable η from its mean has been studied by H. Cramér [4] and V. V. Petrov [5].

Another type of problem amenable to the methods of this paper is the estimation of the probability of large deviations of the order statistics. Let $x_1 \leq x_2 \leq \cdots \leq x_N$ be N ordered independent observations on the random variable ξ with distribution function $F(x) = x$, $0 \leq x \leq 1$. Then

$$P(x_k < \alpha) = P(F_N \in \Omega_{N,k,\alpha}),$$

where $\Omega_{N,k,\alpha}$ is the class of all distribution functions which "touch" the rectangle $0 \leq x \leq \alpha$, $\frac{k-1}{N} \leq y \leq \frac{k}{N}$. The distribution of such order statistics can be found in the paper by N. V. Smirnov [3].

It is necessary to remark that in this paper we have mainly considered the case of classes Ω which do not depend on N, the number of independent trials. Hence the results obtained have been formulated under conditions which give asymptotic expressions not for the probability P of large deviations itself but only for the logarithm of that probability. In order to obtain asymptotic expressions for the probability itself it apparently will be necessary to narrow the set of classes Ω being considered and apply the more exact methods for obtaining asymptotic expressions which are worked out in A. N. Kolmogorov [7], H. Cramér [4], and N. V. Smirnov [2], [3].

Furthermore, a problem which at present is of great interest is that of obtaining results of the type here considered for the case of a class Ω which depends on N.

IV. DEFINITION OF ENTROPY IN THE CONTINUOUS CASE

We notice that the results of this paper permit one to give a probabilistic definition of entropy in the continuous or mixed case.

In fact, if $\Phi(x)$ is the distribution function of a random variable on the interval $[0,1]$ we may look upon the entropy of the function $\Phi(x)$ as a certain limit. Namely, let $F_N(x)$ be the empirical distribution function of the random variable ξ with uniform distribution $F(x)$, $0 \leq x \leq 1$. Then we can give

Definition 5. The entropy of the distribution function $\Phi(x)$ on the interval $[0,1]$ is the double limit

$$E(\Phi) = -\lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \lim_{N \to \infty} \frac{\ln P(\sup_x |F_N - \Phi| \leq \epsilon)}{N} . \tag{55}$$

From Theorem 13 it follows that

$$E(\Phi) = \int_0^1 \ln \frac{d\Phi}{dx} \, d\Phi .$$

In the case where $\Phi(x)$ has a continuous derivative $\Phi'(x) = p(x)$,

$$E(\Phi) = \int_0^1 p(x) \ln p(x) dx . \tag{56}$$

It is not difficult to verify that the entropy, if it exists, is non-negative. $E(\Phi) = 0$ if and only if $\Phi(x) = x$, $p(x) = 1$.

The definition of entropy usually given for other intervals apparently does not have such a simple probabilistic interpretation.

(Received by the editor February 24, 1956)

References

1. N. V. Smirnov, "On the Probabilities of Large Deviations," Mat. Sbornik, 40 (1933), 443-54.

2. N. V. Smirnov, "On the Deviation of the Empirical Distribution Curve," Mat. Sbornik, 6 (48) (1939), 3-26.

3. N. V. Smirnov, "Limiting Distributions for the Members of an Ordered Series," Trudy Matem. Inst. V. A. Steklov, XXVI (1949).

4. H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," Act. Scient. et Industr., Paris, 1938.

5. V. V. Petrov, "A Generalization of a Limit Theorem of Cramér," Uspekhi Matem. Nauk, IX, 4 (62) (1954), 195-202.

6. A. N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933.

7. A. N. Kolmogorov, "Sulla determinazione empirica di una leggi di distribuzione," Giorn. Ist. Ital. Attuari, IV (1933).

8. B. V. Gnedenko, A Course in the Theory of Probability, M.-L., Gostekhizdat, 1950.

9. G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge, 1934.