

6

SOME LIMIT THEOREMS

12. The Law of Large Numbers

Consider n independent identically distributed random variables $\xi_1, \dots, \xi_n$. In particular, $\xi_1, \dots, \xi_n$ have the same mean $a = \mathbf{E}\xi_k$ and variance $\sigma^2 = \mathbf{D}\xi_k$. If

$$\eta = \frac{1}{n}(\xi_1 + \dots + \xi_n)$$

is the arithmetic mean of the variables $\xi_1, \dots, \xi_n$, then

$$\mathbf{E}\eta = \frac{1}{n} \sum_{k=1}^n \mathbf{E}\xi_k = a,$$

$$\mathbf{D}\eta = \mathbf{E}(\eta - a)^2 = \frac{1}{n^2} \sum_{k=1}^n \mathbf{D}\xi_k = \frac{\sigma^2}{n}.$$

Applying Chebyshev's inequality (see Sec. 9) to the random variable $\eta - a$, we get the inequality

$$\mathbf{P} \{ |\eta - a| > \varepsilon \} \leq \frac{1}{\varepsilon^2} \mathbf{E}(\eta - a)^2 = \frac{\sigma^2}{n\varepsilon^2} \quad (6.1)$$

for arbitrary $\varepsilon > 0$.

THEOREM 6.1 (Weak law of large numbers). *Let $\xi_1, \dots, \xi_n$ be n independent identically distributed random variables with mean a and variance σ^2 . Then, given any $\delta > 0$ and $\varepsilon > 0$, however small, there is an integer*

n such that

$$a - \varepsilon < \frac{1}{n}(\xi_1 + \cdots + \xi_n) < a + \varepsilon$$

with probability greater than $1 - \delta$.

Proof. The theorem is an immediate consequence of (6.1) if we choose $n > \sigma^2/\delta\varepsilon^2$. ■

Remark. Suppose δ and ε are so small that we can practically neglect both the occurrence of events of probability δ and differences between quantities differing by no more than ε . Then Theorem 6.1 asserts that for sufficiently large n , the arithmetic mean

$$\eta = \frac{1}{n}(\xi_1 + \cdots + \xi_n)$$

is an excellent approximation to the mean value $a = \mathbf{E}\xi_k$

Now consider n consecutive Bernoulli trials, in each of which an event A can occur with probability $p = \mathbf{P}(A)$ or fail to occur with probability $q = 1 - p$. Let ξ_k be a random variable equal to 1 if A occurs at the k th trial and 0 if A fails to occur at the k th trial. Then the random variables $\xi_1, \dots, \xi_n$ are independent and identically distributed (by the very meaning of Bernoulli trials). Obviously

$$\mathbf{P}\{\xi_k = 1\} = p, \quad \mathbf{P}\{\xi_k = 0\} = q.$$

Moreover, each random variable ξ_k has mean

$$a = \mathbf{E}\xi_k = p \cdot 1 + q \cdot 0 = p = \mathbf{P}(A).$$

Let $n(A)$ be the number of trials in which A occurs, so that

$$\frac{n(A)}{n}$$

is the relative frequency of the event A . Then clearly

$$n(A) = \xi_1 + \cdots + \xi_n,$$

and hence

$$\frac{n(A)}{n} = \frac{1}{n}(\xi_1 + \cdots + \xi_n).$$

It follows from Theorem 6.1 that $n(A)/n$ virtually coincides with $\mathbf{P}(A)$ for sufficiently large n , more exactly, that given any $\delta > 0$ and $\varepsilon > 0$, however small, there is an integer n such that

$$\mathbf{P}(A) - \varepsilon < \frac{n(A)}{n} < \mathbf{P}(A) + \varepsilon$$

with probability greater than $1 - \delta$. The justification for formula (1.2), p. 3 is now apparent.

Remark. It can be shown¹ that with probability 1 the limit

$$\lim_{n \rightarrow \infty} \frac{n(A)}{A}$$

exists and equals $P(A)$. This result is known as the *strong law of large numbers*.

13. Generating Functions. Weak Convergence of Probability Distributions

Let ξ be a discrete random variable taking the values $0, 1, 2, \dots$ with probabilities

$$P_\xi(k) = P\{\xi = k\}, \quad k = 0, 1, 2, \dots \quad (6.2)$$

Then the function

$$F_\xi(z) = \sum_{k=0}^{\infty} P_\xi(k) z^k, \quad |z| < 1 \quad (6.3)$$

is called the *generating function* of the random variable ξ or of the corresponding probability distributions (6.2). It follows from the convergence of the series (6.3) for $|z| = 1$ and from Weierstrass's theorem on uniformly convergent series of analytic functions² that $F_\xi(z)$ is an analytic function of z in $|z| < 1$, with (6.3) as its power series expansion. Moreover, the probability distribution of the random variable ξ is uniquely determined by its generating function $F_\xi(z)$, and in fact

$$P_\xi(k) = \frac{1}{k!} F_\xi^{(k)}(0), \quad k = 0, 1, 2, \dots,$$

where $F_\xi^{(k)}(z)$ is the k th derivative of $F_\xi(z)$. According to formula (4.10), p. 44, for fixed z the function $F_\xi(z)$ is just the mathematical expectation of the random variable $\varphi(\xi) = z^\xi$, i.e.,

$$F_\xi(z) = E z^\xi, \quad |z| < 1. \quad (6.4)$$

Example 1 (The Poisson distribution). If the random variable ξ has a Poisson distribution with parameter a , so that

$$P_\xi(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, 2, \dots,$$

¹ See e.g., W. Feller, *op. cit.*, p. 203.

² See e.g., R. A. Silverman, *Introductory Complex Analysis*, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1967), p. 191. Also use Weierstrass' *M*-test (*ibid.*, p. 186).

then ξ has the generating function

$$F_{\xi}(z) = \sum_{k=0}^{\infty} \frac{a^k}{k!} e^{-a} z^k = e^{-a} \sum_{k=0}^{\infty} \frac{(az)^k}{k!} = e^{a(z-1)}. \quad (6.5)$$

Suppose the random variable ξ has mean $a = E\xi$ and variance $\sigma^2 = D\xi$. Then, differentiating (6.4) twice with respect to z behind the expectation sign and setting $z = 1$, we get

$$a = E\xi = F'(1), \quad \sigma^2 = E\xi^2 - (E\xi)^2 = F''(1) + F'(1) - [F'(1)]^2. \quad (6.6)$$

The same formulas can easily be deduced from the power series (6.3). In fact, differentiating (6.3) for $|z| < 1$, we get

$$F'(z) = \sum_{k=0}^{\infty} k P_{\xi}(k) z^{k-1},$$

and hence

$$a = \sum_{k=0}^{\infty} k P_{\xi}(k) = \lim_{z \rightarrow 1} F'_{\xi}(z) = F'_{\xi}(1),$$

and similarly for the second of the formulas (6.6).

Next let $\xi_1, \dots, \xi_n$ be n independent random variables taking the values $0, 1, 2, \dots$. Then the random variables $z^{\xi_1}, \dots, z^{\xi_n}$, where z is a fixed number, are also independent. It follows from formula (4.20), p. 47 that

$$Ez^{(\xi_1 + \dots + \xi_n)} = Ez^{\xi_1} \dots z^{\xi_n} = Ez^{\xi_1} \dots Ez^{\xi_n}.$$

Thus we have the formula

$$F_{\xi}(z) = F_{\xi_1}(z) \dots F_{\xi_n}(z), \quad (6.7)$$

expressing the generating function $F_{\xi}(z) = Ez^{\xi}$ of the sum $\xi = \xi_1 + \dots + \xi_n$ of the n random variables $\xi_1, \dots, \xi_n$ in terms of the generating functions $F_{\xi_k}(z) = Ez^{\xi_k}$, $k = 1, \dots, n$ of the separate summands.

Example 2 (The binomial distribution). Suppose the random variable ξ has a binomial distribution with parameters p and n , so that

$$P_{\xi}(k) = C_n^k p^k q^{n-k}, \quad q = 1 - p, \quad k = 0, 1, \dots, n.$$

Then, as already noted in Sec. 10, ξ can be regarded as the sum $\xi = \xi_1 + \dots + \xi_n$ of n independent random variables $\xi_1, \dots, \xi_n$, where

$$\xi_k = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q. \end{cases}$$

The generating function $F_{\xi_k}(z)$ of each summand is clearly $pz + q$, and hence, by (6.7), the generating function of ξ itself is

$$F_{\xi}(z) = (pz + q)^n. \quad (6.8)$$

Now let ξ_n , $n = 1, 2, \dots$ be a sequence of discrete random variables taking the values $0, 1, 2, \dots$, with probability distributions $P_n(k) = P_{\xi_n}(k)$ and generating functions $F_n(z)$, $n = 1, 2, \dots$. Then the sequence of distributions $\{P_n(k)\}$ is said to *converge weakly* to the limiting distribution $P(k)$ if

$$\lim_{n \rightarrow \infty} P_n(k) = P(k) \quad (6.9)$$

for all $k = 0, 1, 2, \dots$

Example 3 (Weak convergence of the binomial distribution to the Poisson distribution). Let $\xi_1, \xi_2, \dots$ be a sequence of random variables such that ξ_n has a binomial distribution $P_n(k)$ with parameters p and n , i.e.,

$$P_n(k) = C_n^k p^k q^{n-k}, \quad q = 1 - p.$$

Suppose p depends on n in such a way that the limit

$$\lim_{n \rightarrow \infty} np = a \quad (6.10)$$

exists. Then, according to formula (5.5), p. 55, the sequence of distributions $\{P_n(k)\}$ converges weakly to the Poisson distribution

$$P(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, 2, \dots$$

with parameter a given by (6.10).

In Example 3, the sequence of generating functions

$$F_n(z) = (pz + q)^n, \quad n = 1, 2, \dots$$

of the random variables $\xi_1, \xi_2, \dots$ converges uniformly to the generating function $F(z) = e^{a(z-1)}$ of the limiting Poisson distribution, i.e.,

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(z) &= \lim_{n \rightarrow \infty} [1 + p(z-1)] = \lim_{n \rightarrow \infty} \left[1 + \frac{np(z-1)}{n} \right]^n \\ &= \lim_{n \rightarrow \infty} \left[1 + \frac{a(z-1)}{n} \right]^n = e^{a(z-1)} \end{aligned}$$

(justify the next-to-the-last step). This is no accident, as shown by

THEOREM 6.2. *The sequence of probability distributions $P_n(k)$, $n = 1, 2, \dots$ with generating functions $F_n(z)$, $n = 1, 2, \dots$ converges weakly to*

the limiting distribution $P(k)$ if and only if

$$\lim_{n \rightarrow \infty} F_n(z) = F(z), \quad (6.11)$$

where

$$F(z) = \sum_{k=0}^{\infty} P(k) z^k$$

is the generating function of $P(k)$ and the convergence is uniform in every disk $|z| \leq r < 1$.

Proof. First suppose (6.9) holds. Clearly,

$$|F_n(z) - F(z)| \leq \sum_{k=0}^K |P_n(k) - P(k)| + \sum_{k=K+1}^{\infty} |z|^k \quad (6.12)$$

for any positive integer K . Given any $\varepsilon > 0$, we first choose K large enough to make

$$\sum_{k=K+1}^{\infty} |z|^k < \frac{r^{K+1}}{1-r} < \frac{\varepsilon}{2},$$

and then find a positive integer N such that

$$|P_n(k) - P(k)| < \frac{\varepsilon}{2(K+1)}$$

holds for $k = 0, \dots, K$ if $n \geq N$. It then follows from (6.12) that

$$|F_n(z) - F(z)| < \varepsilon$$

if $n \geq N$, which immediately proves (6.11).

Conversely, suppose (6.11) holds, where the convergence is uniform in every disk $|z| \leq r < 1$. Then, by Weierstrass' theorem on uniformly convergent sequences of analytic functions,³

$$\lim_{n \rightarrow \infty} F_n^{(k)}(z) = F^{(k)}(z), \quad |z| < 1 \quad (6.13)$$

for all $k = 0, 1, 2, \dots$. But

$$P_n(k) = \frac{1}{k!} F_n^{(k)}(0), \quad P(k) = \frac{1}{k!} F^{(k)}(0),$$

and hence (6.13) implies (6.9) for all $k = 0, 1, 2, \dots$ ■

The following example is typical of the situations where the Poisson distribution is encountered:

Example 4 (Random flow of events). Suppose that events of a given kind occur randomly in the course of time. For example, we can think in terms

³ R. A. Silverman, *op. cit.*, p. 192.

of "service calls" (requests for service) arriving randomly at some "server" (service facility), like inquiries at an information desk, arrival of motorists at a gas station, telephone calls at an exchange, etc. Let $\xi(\Delta)$ be the number of events occurring during the time interval Δ . Then what is the distribution of the random variable $\xi(\Delta)$?

To answer this question, we will assume that our "random flow of events" has the following three properties:

- The events are independent of one another; more exactly, the random variables $\xi(\Delta_1), \xi(\Delta_2), \dots$ are independent if the intervals $\Delta_1, \Delta_2, \dots$ are nonoverlapping.
- The flow of events is "stationary," i.e., the distribution of the random variable $\xi(\Delta)$ depends only on the length of the interval Δ and not on the time of its occurrence (the initial time of Δ , say).
- The probability that at least one event occurs in a small time interval Δt is $\lambda \Delta t + o(\Delta t)$, while the probability that more than one event occurs in Δt is $o(\Delta t)$. Here $o(\Delta t)$ is an infinitesimal of higher order than Δt , i.e.,

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0,$$

and λ is a positive parameter characterizing the "rate of occurrence" or "density" of the events.

Now consider the time interval $\Delta = [0, t]$, and let $\xi(t)$ be the total number of events occurring in $[0, t]$. Dividing $[0, t]$ into n equal parts $\Delta_1, \dots, \Delta_n$, we find that

$$\xi(t) = \sum_{k=1}^n \xi(\Delta_k),$$

where $\xi(\Delta_1), \dots, \xi(\Delta_n)$ are independent random variables and $\xi(\Delta_k)$ is the number of events occurring in the interval Δ_k . Clearly, the generating function of each random variable $\xi(\Delta_k)$ is

$$F_n(z) = \left(1 - \frac{\lambda t}{n}\right) + \frac{\lambda t}{n} z + o\left(\frac{t}{n}\right),$$

where $o(t/n)$ is a term of order higher than t/n . Hence, by (6.7), the generating function of $\xi(t)$ is

$$F(z) = [F_n(z)]^n = \left[1 + \frac{\lambda t(z-1)}{n} + o\left(\frac{t}{n}\right)\right]^n.$$

But $F(z)$ is independent of the subintervals $\Delta_1, \dots, \Delta_n$, and hence we can take the limit as $n \rightarrow \infty$, obtaining

$$F(z) = \lim_{n \rightarrow \infty} \left[1 + \frac{\lambda t(z-1)}{n}\right]^n = e^{\lambda t(z-1)}.$$

Comparing this with (6.5), we find that $F(z)$ is the generating function of a Poisson distribution with parameter $a = \lambda t$, so that

$$\mathbf{P}\{\xi(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \dots$$

Since

$$\lambda t = \mathbf{E}\xi(t),$$

the parameter λ is just the average number of events occurring per unit time.

14. Characteristic Functions. The Central Limit Theorem

Given a real random variable ξ , by the *characteristic function* of ξ is meant the function

$$f_\xi(t) = \mathbf{E}e^{i\xi t}, \quad -\infty < t < \infty. \quad (6.14)$$

Clearly, $f_\xi(t)$ coincides for every fixed t with the mathematical expectation of the complex random variable $\eta = e^{i\xi t}$. For a discrete random variable taking the values $0, 1, 2, \dots$, the characteristic function $f_\xi(t)$ coincides with the values of the generating function $F_\xi(z)$ on the boundary of the unit circle $|z| = 1$, i.e.,

$$f_\xi(t) = F_\xi(e^{it}) = \sum_{k=0}^{\infty} P_\xi(k) e^{ikt}.$$

This formula represents $f_\xi(t)$ as a *Fourier series*, with the probabilities $P_\xi(k) = \mathbf{P}\{\xi = k\}$, $k = 0, 1, 2, \dots$ as its coefficients. Thus these probabilities $P_\xi(k)$ are uniquely determined by the characteristic function $f_\xi(t)$.

If ξ is a continuous random variable with probability density $p_\xi(x)$, then, by formula (4.18), p. 47, the characteristic function is the *Fourier transform* of the density $p_\xi(x)$:

$$f_\xi(t) = \int_{-\infty}^{\infty} e^{ixt} p_\xi(x) dx. \quad (6.15)$$

Inverting (6.15), we find that

$$p_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} f_\xi(t) dt, \quad (6.16)$$

at least at points where $p_\xi(x)$ is suitably well-behaved.⁴ Thus $p_\xi(x)$ is uniquely determined by the characteristic function $f_\xi(t)$.

⁴ If (6.16) fails, another inversion formula can be used, giving the distribution function $\Phi_\xi(x) = \mathbf{P}\{\xi < x\}$ in terms of the characteristic function $f_\xi(t)$ (see e.g., B. V. Gnedenko, *op. cit.*, Sec. 36). We can then deduce $p_\xi(x)$ from $\Phi_\xi(x)$ by differentiation, at least almost everywhere (recall footnote 2, p. 38).

Example 1. Let ξ be a normally distributed random variable, with probability density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (6.17)$$

Then, by (6.15), the characteristic function of ξ is given by

$$\begin{aligned} f_{\xi}(t) &= \int_{-\infty}^{\infty} e^{ixt} p(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt - (x^2/2)} dx \\ &= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx. \end{aligned} \quad (6.18)$$

The function $\varphi(z) = e^{-z^2/2}$ is an analytic function of the complex variable z , and hence, by Cauchy's integral theorem,⁵ the integral of $\varphi(z)$ along the rectangular contour with vertices $(-N, 0)$, $(N, 0)$, $(N, -it)$, $(-N, -it)$ equals zero. Therefore

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx &= \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-N}^N e^{-(x-it)^2/2} dx \\ &= \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-N-it}^{N-it} e^{-z^2/2} dz = \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-N}^N e^{-x^2/2} dx \quad (6.19) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx, \end{aligned}$$

where we use the fact that the integral of $\varphi(z)$ along the vertical sides of the contour vanishes as $N \rightarrow \infty$ (why?). But

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty} p(x) dx = 1,$$

as for any probability density. Hence (6.18) and (6.19) imply

$$f_{\xi}(t) = e^{-t^2/2}. \quad (6.20)$$

Now suppose the random variable ξ is such that $E|\xi|^3$ exists. Then the characteristic function $f_{\xi}(t)$ has the expansion

$$f_{\xi}(t) = 1 + iE\xi \cdot t - \frac{E\xi^2}{2} t^2 + R(t), \quad (6.21)$$

where the remainder $R(t)$ satisfies the estimate

$$|R(t)| \leq CE|\xi|^3 \cdot |t|^3$$

⁵ R. A. Silverman, *op. cit.*, p. 146.

(C denotes a constant). In fact, we need only note that

$$e^{i\xi t} = 1 + i\xi t - \frac{\xi^2}{2} t^2 + \theta, \quad (6.22)$$

by Taylor's formula, where

$$|\theta| \leq C |\xi|^3 t^3.$$

We then get (6.21) by taking the mathematical expectation of both sides of (6.22). In particular, it follows from (6.21) that the mean $a = E\xi$ and variance $\sigma^2 = D\xi$ are given by the formulas

$$a = -if'_\xi(0), \quad \sigma^2 = -f''_\xi(0) + [f'_\xi(0)]^2. \quad (6.23)$$

Example 2. According to (6.23), the normally distributed random variable ξ with probability density (6.17) has mean

$$a = -if'(0) = 0$$

and variance

$$\sigma^2 = -f''(0) = 1.$$

Formula (6.7) has a natural analogue for characteristic functions. In fact, if $\xi_1, \dots, \xi_n$ are independent random variables with sum $\xi = \xi_1 + \dots + \xi_n$, then, by formula (4.20), p. 47, the characteristic function of ξ is

$$f_\xi(t) = f_{\xi_1}(t) \cdots f_{\xi_n}(t). \quad (6.24)$$

Let $\xi_n, n = 1, 2, \dots$ be a sequence of random variables with characteristic functions $f_n(t), n = 1, 2, \dots$. Then the sequence of probability distributions of $\xi_1, \xi_2, \dots$ is said to *converge weakly* to the distribution with density $p(x)$ if

$$\lim_{n \rightarrow \infty} P \{x' \leq \xi_n \leq x''\} = \int_{x'}^{x''} p(x) dx$$

for all x' and x'' ($x' \leq x''$). This should be compared with the definition of weak convergence for discrete random variables taking the values $0, 1, 2, \dots$ given in Sec. 13.

Theorem 6.2 has a natural analogue for characteristic functions, whose proof will not be given here:⁶

THEOREM 6.2.' *The sequence of probability distributions with characteristic functions $f_n(t), n = 1, 2, \dots$ converges weakly to the limiting distribution with density $p(x)$ if and only if*

$$\lim_{n \rightarrow \infty} f_n(t) = f(t), \quad (6.11')$$

⁶ For the proof, see e.g., B. V. Gnedenko, *op. cit.*, Sec. 38.

where

$$f(t) = \int_{-\infty}^{\infty} e^{ixt} p(x) dx$$

is the characteristic function of the limiting distribution and the convergence is uniform in every finite interval $t' \leq t \leq t''$.

We now prove a key proposition of probability theory, called the *central limit theorem*, which has the De Moivre-Laplace theorem (Theorem 5.1, p. 59) as a very special case. Roughly speaking, the central limit theorem asserts that the distribution of the sum of a large number of independent identically distributed random variables is approximately normal.

DEFINITION. Given a sequence of random variables ξ_k , $k = 1, 2, \dots$ with finite means $a_k = E\xi_k$ and variances $\sigma_k^2 = D\xi_k$, consider the "normalized sum"

$$S_n^* = \frac{S_n - ES_n}{\sqrt{DS_n}},$$

where

$$S_n = \sum_{k=1}^n \xi_k.$$

Then the sequence ξ_k , $k = 1, 2, \dots$ is said to satisfy the *central limit theorem* if⁷

$$\lim_{n \rightarrow \infty} P\{x' < S_n^* \leq x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx. \quad (6.25)$$

THEOREM 6.3. Suppose the sequence of independent random variables ξ_k , $k = 1, 2, \dots$ with means a_k and variances σ_k^2 satisfies the *Lyapunov condition*

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^3} \sum_{k=1}^n E|\xi_k - a_k|^3 = 0, \quad (6.26)$$

where

$$B_n^2 = DS_n = \sum_{k=1}^n \sigma_k^2.$$

Then the sequence of random variables satisfies the central limit theorem.

Proof. Equation (6.25) means that the sequence of distributions of the normalized sums S_n^* , $n = 1, 2, \dots$ converges weakly to the normal distribution with probability density (6.16). Hence, according to Theorem 6.2', we need only show that the sequence of characteristic

⁷ Cf. formula (5.8), p. 59. Note that the right-hand side of (6.25) equals $\Phi(x'') - \Phi(x')$, where $\Phi(x)$ is the distribution function of a normal random variable with mean 0 and variance 1.

functions $f_n(t)$, $n = 1, 2, \dots$ of the random variables S_n^* converges uniformly in every finite interval $t' \leq t \leq t''$ to the characteristic function $f(t) = e^{-t^2/2}$ of this normal distribution (recall Example 1). Clearly,

$$S_n^* = \sum_{k=1}^n \frac{\xi_k - a_k}{B_n}, \quad B_n^2 = \text{DS}_n.$$

The random variable $\xi_k - a_k$ has zero mean and variance σ_k^2 , and hence, by (6.21), has characteristic function

$$g_k(t) = 1 - \frac{\sigma_k^2}{2} t^2 + R_k(t),$$

where

$$|R_k(t)| \leq C |t|^3 \mathbf{E} |\xi_k - a_k|^3$$

(C is some constant). Therefore the characteristic function of the random variable $\eta_k = (\xi_k - a_k)/B_n$ is

$$f_{kn}(t) = g_k\left(\frac{t}{B_n}\right) = 1 - \frac{\sigma_k^2}{2B_n^2} t^2 + R_k\left(\frac{t}{B_n}\right),$$

where

$$\left| R_k\left(\frac{t}{B_n}\right) \right| \leq C |t|^3 \frac{\mathbf{E} |\xi_k - a_k|^3}{B_n^3}.$$

It follows from (6.24) that the random variable $S_n^* = \eta_1 + \dots + \eta_n$ has characteristic function

$$f_n(t) = \prod_{k=1}^n f_{kn}(t).$$

Hence

$$\ln f_n(t) = \sum_{k=1}^n \ln f_{kn}(t) \sim \sum_{k=1}^n \left[-\frac{\sigma_k^2}{2B_n^2} t^2 + R_k\left(\frac{t}{B_n}\right) \right],$$

where, because of the hypothesis (6.26),

$$\left| \sum_{k=1}^n R_k\left(\frac{t}{B_n}\right) \right| \leq C |t|^3 \frac{1}{B_n^3} \sum_{k=1}^n \mathbf{E} |\xi_k - a_k|^3 \rightarrow 0$$

as $n \rightarrow \infty$ uniformly in every finite interval $t' \leq t \leq t''$. Therefore

$$\ln f_n(t) \sim -\frac{t^2}{2B_n^2} \sum_{k=1}^n \sigma_k^2 = -\frac{t^2}{2},$$

or equivalently

$$f_n(t) \rightarrow e^{-t^2/2}$$

as $n \rightarrow \infty$. ■

Example 3. The Lyapunov condition is always satisfied if the random variables $\xi_1, \xi_2, \dots$ are identically distributed and if $\alpha = E|\xi_k - a_k|^3$ exists. In fact,

$$B_n^2 = \sum_{k=1}^n D\xi_k = n\sigma^2,$$

where $\sigma^2 = D\xi_k$, and hence

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^3} \sum_{k=1}^n E|\xi_k - a_k|^3 = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \frac{\alpha}{\sigma^3} = 0.$$

PROBLEMS

1. Show that the conclusion of Theorem 6.1 can be written in the form

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \left| \frac{1}{n} \sum_{k=1}^n \xi_k - a \right| < \varepsilon \right\} = 1$$

for arbitrary $\varepsilon > 0$.

2. Let $\xi_1, \dots, \xi_n$ be n independent identically distributed random variables, with common mean $a = E\xi_k$ and variance $\xi^2 = D\xi_k$. Suppose a is known. Can the quantity

$$\frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2$$

be used to estimate σ^2 ?

3. A random variable ξ has probability density*

$$p_\xi(x) = \begin{cases} \frac{x^m}{m!} e^{-x} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where m is a positive integer. Prove that

$$\mathbf{P} \{0 < \xi < 2(m+1)\} > \frac{m}{m+1}.$$

Hint. Use Chebyshev's inequality.

4. The probability of an event A occurring in one trial is $\frac{1}{2}$. Is it true that the probability of A occurring between 400 and 600 times in 1000 independent trials exceeds 0.97?

Ans. Yes.

5. Let ξ be the number of spots obtained in throwing an unbiased die. What is the generating function of ξ ?

* It follows by repeated integration by parts that $\int_0^\infty x^m e^{-x} dx = m!$

6. Use (6.6) and the result of the preceding problem to solve Problem 16, p. 52.
 7. Let ξ be a random variable with the Poisson distribution

$$P_{\xi}(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, 2, \dots \quad (6.27)$$

Use (6.6) to show that $E\xi = D\xi = a$.

8. Find the generating function of the random variable ξ with distribution

$$P\{\xi = k\} = \frac{a^k}{(1+a)^{k+1}} \quad (a > 0).$$

Use (6.6) to find $E\xi$ and $D\xi$.

9. Let η be the sum of two independent random variables ξ_1 and ξ_2 , one with the Poisson distribution (6.27), the other with the Poisson distribution obtained by changing the parameter a to a' in (6.27). Show that η also has a Poisson distribution, with parameter $a + a'$.

10. Let S_n be the number of successes in a series of n independent trials, where the probability of success at the k th trial is p_k . Suppose $p_1, \dots, p_n$ depend on n in such a way that

$$p_1 + \dots + p_n = \lambda,$$

while

$$\max\{p_1, \dots, p_n\} \rightarrow 0$$

as $n \rightarrow \infty$. Prove that S_n has a Poisson distribution with parameter λ in the limit as $n \rightarrow \infty$.

Hint. Use Theorem 6.2.⁹

11. Find the characteristic function $f_{\xi}(t)$ of the random variable with probability density

$$p_{\xi}(x) = \frac{1}{2} e^{-|x|} \quad (-\infty < x < \infty).$$

$$\text{Ans. } f_{\xi}(t) = \frac{1}{1+t^2}.$$

12. Use (6.23) and the result of the preceding problem to solve Problem 13, p. 52.

13. Find the characteristic function of a random variable uniformly distributed in the interval $[a, b]$.

14. A continuous random variable ξ has characteristic function

$$f_{\xi}(t) = e^{-a|t|} \quad (a > 0).$$

Find the probability density of ξ .

⁹ For the details, see W. Feller, *op. cit.*, p. 282.

Ans. $p_{\xi}(x) = \frac{a}{\pi(a^2 + x^2)}.$

15. The derivatives $f'_{\xi}(0)$ and $f''_{\xi}(0)$ do not exist in the preceding problem. Why does this make sense?

Hint. Cf. Problem 24, p. 53.

16. Let ν be the total number of spots which are obtained in 1000 independent throws of an unbiased die. Then $E\nu = 3500$, because of Problem 16, p. 52. Estimate the probability that ν is a number between 3450 and 3550.

17. Let S_n be the same as in Problem 10, and suppose $\sum_{k=1}^{\infty} p_k q_k = \infty$. Prove that

$$\mathbf{P} \left\{ x' < \frac{S_n - \sum_{k=1}^n p_k}{\sqrt{\sum_{k=1}^n p_k q_k}} \leq x'' \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx$$

as $n \rightarrow \infty$.

Hint. Apply Theorem 6.3.