CHAPTER 1

Relevant Elements of Probability Theory

1.0 Introduction

The aim of the first two sections of this chapter is to provide a survey of the basic notions and results of probability theory which can be found in many textbooks. The concepts and theorems mentioned in Sections 1 and 2 are of an auxiliary nature and are included more for reference than for primary study. For this reason the majority of statements are given without proofs, the single exception being the theorem by Curtiss [21], which will be used frequently throughout the book.

Section 3 deals with typical examples of applications of various limit theorems to the analysis of asymptotic distributions in combinatorial problems. In terms of the properties of double generating functions we formulate rather general conditions providing asymptotic normality of certain classes of probability distributions which are met in combinatorial probability. Section 4 contains a comprehensive description of limiting distributions of random variables specified by double generating functions of the form $\exp\{x\,g(t)\}$, where $g(t)$ is a polynomial. In subsequent chapters, the method used to obtain the description will be extended to double generating functions of the form $\exp\{g(x,t)\}$, where the function $g(x,t)$ is not necessarily a polynomial in t.

1.1 Probability distributions and random variables

1.1.1 Probability space

A *space* of elementary events is an arbitrary nonempty set Ω representing all possible outcomes of an experiment which may be repeated, at least in principle, infinitely many times. Elements of Ω are called *elementary events*.

An *event* is any subset $A \subseteq \Omega$. An event A is said to be a particular case of an event B if the inclusion $A \subseteq B$ occurs. The *sum* $A \cup B$ of two events A and B is the event consisting of all elementary events belonging to at least one of the events A or B. The *product* $A \cap B$ is defined as the event consisting of all elementary events belonging to both A and B. The *difference* $A \setminus B$ is the event consisting of the elements of the set A not belonging to B. The event $\bar{A} = \Omega \setminus A$ is called the *complement* of A (that is, the *opposite* of the event A). We call the event Ω *certain*; the complement to Ω is the empty set $\varnothing = \overline{\Omega}$ and is called the *impossible* event. Events A and B are *mutually exclusive* or *disjoint* if $A \cap B = \varnothing$. Events $B_1, B_2, \ldots, B_n, \ldots$ constitute a *complete set of events* if

$$B_1 \cup B_2 \cup \ldots \cup B_n \cup \ldots = \Omega, \qquad B_i \cap B_j = \varnothing, \quad i \neq j.$$

A space of elementary events is called *discrete* if it is finite or countably infinite: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ or $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$. On a discrete space Ω, let a nonnegative function $\mathbf{P}(\omega)$, $\omega \in \Omega$, be given such that $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$. The *probability* $\mathbf{P}(A)$ of an event $A \subseteq \Omega$ is determined by the formula

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

It follows directly from the definition of $\mathbf{P}(A)$ that

$$\mathbf{P}(\Omega) = 1, \qquad \mathbf{P}(\varnothing) = 0,$$

and, for any event A,

$$0 \leq \mathbf{P}(A) \leq 1, \qquad \mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A).$$

Throughout, we make use of the following notations:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n, \qquad \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

For any finite number of events $A_1, A_2, \ldots, A_n$, *Boole's inequality*

$$\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mathbf{P}(A_i)$$

is valid. If $A_1, A_2, \ldots, A_n$ are pairwise disjoint events, we have

$$\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mathbf{P}(A_i).$$

For arbitrary events $A_1, A_2, \dots, A_n$ the formula

$$\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k-1} S_k \tag{1.1}$$

is valid, where

$$S_k := \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbf{P}\big(A_{i_1} A_{i_2} \dots A_{i_k}\big),$$

and summation is accomplished over all possible combinations of k numbers taken from $1, 2, \dots, n$ without repetition, and

$$A_{i_1} A_{i_2} \dots A_{i_k} = \bigcap_{j=1}^{k} A_{i_j}.$$

Let Ω be a discrete space of elementary events. Any numerical function $\xi := \xi(\omega)$ of $\omega \in \Omega$ is called a *random variable*. A random variable is said to be *discrete* if the set of its values is finite or countably infinite.

The *distribution law* $\mathbf{P}_\xi(B)$ of the discrete random variable ξ is the probability

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\omega : \xi(\omega) \in B\},$$

defined for each number set B. In particular, for $x \in (-\infty, \infty)$ we write

$$\mathbf{P}_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) = x\} = \mathbf{P}\{\xi = x\},$$

where the notation $\{\xi = x\}$ stands for the event $\{\omega \colon \xi(\omega) = x\}$.

In what follows we will use mainly discrete spaces of elementary events and discrete random variables. However, in order to study limiting probability laws, we must consider arbitrary spaces of elementary events.

A system of subsets $\mathfrak{F}$ of an arbitrary space of elementary events Ω is called a *σ-algebra* if the following conditions are satisfied:

(1) $\Omega \in \mathfrak{F}$;

(2) If $A \in \mathfrak{F}$ then $\bar{A} \in \mathfrak{F}$;

(3) If $\{A_n\}$ is a sequence of sets from $\mathfrak{F}$ then

$$\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}, \qquad \bigcap_{i=1}^{\infty} A_i \in \mathfrak{F}$$

(obviously, the fulfilment of only one of the conditions is sufficient since the others will follow as a corollary).

If $\mathfrak{F}$ is a σ-algebra of Ω then the pair $(\Omega, \mathfrak{F})$ is said to be a *measurable space*. Elements of the σ-algebra $\mathfrak{F}$ are called the events corresponding to the measurable space $(\Omega, \mathfrak{F})$.

A numerical function $\mathbf{P}$ defined on the σ-algebra $\mathfrak{F}$ of a measurable space $(\Omega, \mathfrak{F})$ is called a *probability* if it satisfies the following axioms:

(1) $\mathbf{P}(A) \geq 0$ for any $A \in \mathfrak{F}$.

(2) For any sequence $\{A_n\}$ of pairwise disjoint events the following equality is valid:

$$\mathbf{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

(3) $\mathbf{P}(\Omega) = 1$.

The triple $(\Omega, \mathfrak{F}, \mathbf{P})$ is called a *probability space.* Any probability space satisfying the axioms mentioned possesses all the properties listed for discrete spaces of elementary events. Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be a probability space and let $A, B \in \mathfrak{F}$. If $\mathbf{P}(B) > 0$ then the number $\mathbf{P}(A \mid B)$, defined by the formula

$$\mathbf{P}(A \mid B) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)},$$

is called the *conditional* probability of the event A under the condition B.

If events $B_1, B_2, \ldots, B_n, \ldots$ are such that $B_i \cap B_j = \varnothing$, $\mathbf{P}(B_i) > 0$ and $A \subseteq \cup B_i$, then the following formula is valid (the *total probability* formula):

$$\mathbf{P}(A) = \sum_{i=1}^{\infty} \mathbf{P}(B_i)\, \mathbf{P}(A \mid B_i).$$

Events $A_1, A_2, \ldots, A_n$ are said to be *mutually independent* if, for all combinations of the subscripts $1 \leq i_1 < \cdots < i_k \leq n$, $k = 2, 3, \ldots, n$, we have

$$\mathbf{P}\Big(\bigcap_{s=1}^{k} A_{i_s}\Big) = \prod_{s=1}^{k} \mathbf{P}(A_{i_s}).$$

We now give the definition of a random variable for an arbitrary probability space. A *random variable* ξ on a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ is a numerical function $\xi := \xi(\omega)$, $\omega \in \Omega$, mapping Ω onto the set of real numbers $\mathbf{R}$ such that for any $x \in (-\infty, \infty)$ the set $\{\omega\colon \xi(\omega) < x\}$ belongs to the σ-algebra $\mathfrak{F}$. The function $F_\xi(x) := \mathbf{P}\{\xi < x\}$ is called the *distribution function* or, simply, the *distribution* of the random variable ξ. Any distribution function $F_\xi(x)$ is defined for all $x \in (-\infty, \infty)$; it is nondecreasing and left-continuous. Also,

$$\lim_{x\to-\infty} F_\xi(x) = 0, \qquad \lim_{x\to\infty} F_\xi(x) = 1.$$

If ξ is a discrete random variable taking values $x_1 < x_2 < \cdots$ with positive probabilities, then

$$\mathbf{P}\{\xi = x_k\} = F_\xi(x_{k+1}) - F_\xi(x_k), \qquad k = 1, 2, \ldots.$$

A random variable ξ is said to be *continuous* if there exists a function $p_\xi(x) \geq 0$ such that, for all $x \in (-\infty, \infty)$,

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(y)\, dy\,, \qquad \int_{-\infty}^{\infty} p_\xi(y)\, dy = 1\,.$$

The function $p_\xi(x)$ is called the *density* of the distribution of the random variable ξ.

A random variable ξ is said to be normally distributed with parameters (m, σ) if its distribution function $\Phi_{m,\sigma}(x)$ has the form

$$\Phi_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} \exp\left\{ -\frac{(u-m)^2}{2\,\sigma^2} \right\} du \tag{1.2}$$

with $\sigma > 0$. If $m = 0$ and $\sigma = 1$ we say that ξ has *standard normal distribution*.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be random variables defined on a common probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. The vector $(\xi_1, \xi_2, \ldots, \xi_n)$ is called an *n-dimensional random variable* or a *random vector*. The function

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = \mathbf{P}\{\xi_1 < x_1, \ldots, \xi_n < x_n\}\,,$$
$$x_i \in (-\infty, \infty)\,, \qquad i = 1, 2, \ldots, n\,,$$

is the distribution function of the n-dimensional random variable $(\xi_1, \xi_2, \ldots, \xi_n)$. A random variable (vector) $(\xi_1, \xi_2, \ldots, \xi_n)$ is said to be *continuous* if there exists a function $p_{\xi_1 \ldots \xi_n}(y_1, \ldots, y_n) \geq 0$, $y_i \in (-\infty, \infty)$, $i = 1, 2, \ldots, n$, such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{\xi_1 \ldots \xi_n}(y_1, \ldots, y_n)\, dy_1 \ldots dy_n = 1$$

and

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1 \ldots \xi_n}(y_1, \ldots, y_n)\, dy_1 \ldots dy_n\,.$$

The function $p_{\xi_1 \ldots \xi_n}(y_1, \ldots, y_n)$ is called the *density* of the distribution of the n-dimensional random variable $(\xi_1, \xi_2, \ldots, \xi_n)$ or, simply, the density of $(\xi_1, \xi_2, \ldots, \xi_n)$.

The distribution of a continuous random variable $(\xi_1, \xi_2, \ldots, \xi_n)$ is said to be a *nondegenerate normal distribution* if its density has the form

$$p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = (\det \mathrm{A})^{1/2} \, (2\pi)^{-n/2} \times \exp\left\{-\tfrac{1}{2} Q(x_1, \ldots, x_n)\right\}, \tag{1.3}$$

where

$$Q(x_1, \ldots, x_n) := \sum_{i,j=1}^{n} a_{ij} \, x_i \, x_j$$

is a positive definite quadratic form and $\det \mathrm{A}$ is the determinant of the matrix $\mathrm{A} = \| a_{ij} \|$.

The random variables $\xi_1, \xi_2, \ldots, \xi_n$ are said to be *independent* if

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$$

for all tuples $(x_1, \ldots, x_n)$, $x_i \in (-\infty, \infty)$, $i = 1, \ldots, n$.

The random variables of an infinite sequence $\{\xi_k\}$ are said to be independent if the preceding equality holds for any n. If an n-dimensional random vector $(\xi_1, \xi_2, \ldots, \xi_n)$ is continuous and the random variables $\xi_1, \xi_2, \ldots, \xi_n$ are independent then the densities of the variables are related by the formula

$$p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \cdots p_{\xi_n}(x_n), \qquad x_i \in (-\infty, \infty), \quad i = 1, 2, \ldots, n.$$

For discrete independent random variables $\xi_1, \xi_2, \ldots, \xi_n$ the equality

$$\mathbf{P}\{\xi_1 = x_1, \ldots, \xi_n = x_n\} = \mathbf{P}\{\xi_1 = x_1\} \cdots \mathbf{P}\{\xi_n = x_n\}$$

is valid for all $x_i \in (-\infty, \infty)$, $i = 1, \ldots, n$.

1.1.2 Moments of random variables

First we recall the definition of the Stieltjes integral for a distribution function $F(x)$ and a function $f(x)$ continuous on an interval $[a, b]$ of the real line. We partition the interval $[a, b]$ into n subintervals $[x_i, x_{i+1}]$ such that $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ and calculate a sum

$$S := \sum_{i=1}^{n} f(\tilde{x}_i) \left[F(x_i) - F(x_{i-1}) \right],$$

where $\tilde{x}_i$ is any number from the interval $[x_i, x_{i+1}]$. If the sum S tends to a finite limit as

$$\max_{1\le i\le n} |x_{i+1} - x_i| \longrightarrow 0, \qquad n \to \infty,$$

and this limit does not depend on the particular sequence of partitions and on the choice of the points $\tilde{x}_i$, then it is called the Stieltjes integral of $f(x)$ with respect to the distribution function $F(x)$ and is denoted by

$$\int_a^b f(x)\, dF(x).$$

In what follows we assume that the Stieltjes integral of a function $f(x)$ with respect to a function $F(x)$ exists if and only if the corresponding integral of the function $|f(x)|$ exists. By definition,

$$\int_{-\infty}^{\infty} f(x)\, dF(x) = \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b f(x)\, dF(x).$$

Let ξ be a random variable defined on a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. The *mathematical expectation* (or mean value or, simply, mean) of the random variable ξ is the number

$$\mathbf{E}\,\xi := \int_{-\infty}^{\infty} x\, dF(x),$$

where $F(x)$ is the distribution function of ξ. The mathematical expectation of a discrete random variable ξ taking values $\cdots < x_{-1} < x_0 < x_1 < \cdots < x_n < \cdots$ is calculated by the formula

$$\mathbf{E}\,\xi := \sum_{k=-\infty}^{\infty} x_k\, \mathbf{P}\{\xi = x_k\};$$

if ξ is a continuous random variable with density $p_\xi(x)$, the integral in the definition of expectation reduces to the usual Riemann integral, namely,

$$\mathbf{E}\,\xi := \int_{-\infty}^{\infty} x\, p_\xi(x)\, dx.$$

We list the basic properties of expectation:

(1) $\mathbf{E}\,(C\,\xi) = C\,\mathbf{E}\,\xi$ for any constant C;

(2) $\mathbf{E}\,(\xi_1 + \xi_2) = \mathbf{E}\,\xi_1 + \mathbf{E}\,\xi_2$ if the mathematical expectations $\mathbf{E}\,\xi_1$ and $\mathbf{E}\,\xi_2$ exist;

(3) If ξ_1 and ξ_2 are independent random variables then

$$\mathbf{E}\,\xi_1\xi_2 = \mathbf{E}\,\xi_1 \cdot \mathbf{E}\,\xi_2 .$$

The *variance* of a random variable ξ is defined by the formula

$$\operatorname{Var}\xi := \mathbf{E}\left(\xi - \mathbf{E}\,\xi\right)^2.$$

Variance has the following basic properties:
(1) $\operatorname{Var}(C\,\xi) = C^2 \operatorname{Var}\xi$ for any constant C;
(2) If ξ_1 and ξ_2 are independent random variables then

$$\operatorname{Var}(\xi_1 + \xi_2) = \operatorname{Var}\xi_1 + \operatorname{Var}\xi_2 ;$$

(3) Let ξ be a nonnegative random variable: $\xi \geq 0$, and let ε be an arbitrary positive number. The following inequality is valid:

$$\mathbf{P}\{\xi \geq \varepsilon\} \leq \frac{\mathbf{E}\,\xi}{\varepsilon} . \tag{1.4}$$

This inequality implies, for any random variable ξ, the *Chebyshev inequality*:

$$\mathbf{P}\{|\,\xi - \mathbf{E}\,\xi\,| \geq \varepsilon\} \leq \frac{\operatorname{Var}\xi}{\varepsilon^2} . \tag{1.5}$$

The *moment of the kth order* of a random variable ξ is defined to be the quantity $M_k := \mathbf{E}\,\xi^k$ (if the mathematical expectation exists). The numbers $\mathbf{E}\,|\xi|^k$ and $\mu_k := \mathbf{E}(\xi - \mathbf{E}\,\xi)^k$ are called, respectively, the *absolute* and *central moments* of order k. The following relations are valid:

$$\mu_k = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} M_j\, M_1^{k-j} , \qquad k = 0, 1, \dots ,$$

$$M_k = \sum_{j=0}^{k} \binom{k}{j} \mu_j\, M_1^{k-j} , \qquad k = 0, 1, \dots ,$$

where $\mu_0 = M_0 = 1$. The *factorial* and *binomial moments* of order k are defined, respectively, by the equalities

$$[M]_k := \mathbf{E}\,(\xi)_k , \qquad k = 0, 1, \dots ,$$

$$B_k := \mathbf{E}\binom{\xi}{k} , \qquad k = 0, 1, \dots ,$$

where

$$(\xi)_0 := 1\,, \quad (\xi)_k := \xi\,(\xi-1)\cdots(\xi-k+1)\,, \quad k>0\,, \quad \binom{\xi}{k} := \frac{(\xi)_k}{k\,!}\,.$$

Obviously,

$$[M]_k = k\,!\;B_k\,, \qquad k=0,\,1,\,\ldots\,.$$

The following relations are valid:

$$[M]_k = \sum_{j=0}^{k} s(k,j)\;M_j\,, \qquad k=0,\,1,\,\ldots\,,$$

$$M_k = \sum_{j=0}^{k} \sigma(k,j)\;[M]_j\,, \qquad k=0,\,1,\,\ldots\,,$$

where $s(k,j)$ and $\sigma(k,j)$ are the *Stirling numbers* of the *first* and *second kind* respectively. These numbers are defined by the equalities

$$(x)_k := \sum_{j=0}^{k} s(k,j)\;x^j, \qquad k=0,\,1,\,\ldots\,,$$

$$x^k := \sum_{j=0}^{k} \sigma(k,j)\;(x)_j\,, \qquad k=0,\,1,\,\ldots\,.$$

1.1.3 Integer-valued random variables

Discrete random variables taking only integer values will be of special importance in this book. Such random variables are called *integer-valued.* Below we consider the case of nonnegative integer-valued random variables.

Let ξ be a nonnegative integer-valued random variable with

$$q_k := \sum_{j=k+1}^{\infty} \mathbf{P}\{\,\xi=j\}\,, \qquad k=0,\,1,\ldots\,.$$

Then we have

$$\mathbf{E}\,\xi = \sum_{k=0}^{\infty} q_k\,. \tag{1.6}$$

The *generating function* of an integer-valued random variable ξ is defined by the equality

$$P(x) := \sum_{k=0}^{\infty} P_k \, x^k, \qquad P_k := \mathbf{P}\{\xi = k\}.$$

It is clear that $P(x)$ is an analytic function within the circle $|x| \leq 1$ and, in view of the equality

$$P_k = \frac{1}{k!} P^{(k)}(0), \qquad k = 0, 1, \dots,$$

where $P^{(k)}(0)$ is the value of the kth derivative of $P(x)$ at the point $x = 0$, it determines the distribution of ξ uniquely. One can use, as an inversion formula, *Cauchy's integral formula*:

$$P_k = \frac{1}{2\pi i} \oint_C P(z) \frac{dz}{z^{n+1}},$$

where C is a contour in the complex plane enclosing the origin and lying inside the circle where $P(x)$ is analytic.

Let $\xi_1, \xi_2, \dots, \xi_n$ be independent random variables and let $P_1(x)$, $P_2(x), \dots, P_n(x)$ be the corresponding generating functions. The generating function $P(x)$ of the random variable $\zeta = \xi_1 + \xi_2 + \cdots + \xi_n$ is given by the formula

$$P(x) = P_1(x)\, P_2(x) \cdots P_n(x).$$

For a random variable ξ the functions

$$M(x) := \sum_{k=0}^{\infty} M_k \frac{x^k}{k!},$$

$$\bar{M}(x) := \sum_{k=0}^{\infty} [M]_k \frac{x^k}{k!},$$

$$B(x) := \sum_{k=0}^{\infty} B_k \, x^k$$

are called, respectively, the *moment generating function*, the *factorial moment generating function* and the *binomial moment generating function* of ξ. These generating functions are expressed by the generating function of the random variable ξ as follows:

$$M(x) = P(e^x), \tag{1.7}$$

$$\bar{M}(x) = B(x) = P(x+1). \tag{1.8}$$