This axiomatic treatment of probability theory is due mainly to the Russian mathematician Kolmogorov.

In statistical theory, the axiomatic definition is of greater help than the classical definition, although the latter is simpler to understand and to deal with.

3.8 Random variable, and its expectation and variance

In most cases, with each elementary event in the sample space we may associate a real number. In tossing a coin, e.g., we may associate the number 1 with the appearance of a head and the number 0 with the appearance of a tail. In throwing a die, we have the numbers 1, 2, ..., 6 corresponding to the six possibilities regarding the face that appears uppermost. We thus define a function on the sample space. A (real-valued) function defined on the sample space is called a *random variable* or a stochastic variable. Obviously, to each value of a random variable x* there corresponds a definite probability. Let $x_1, x_2, \ldots\ldots, x_k$ be the possible values of x, and let $p_1, p_2, \ldots\ldots, p_k$ be the corresponding probabilities. A statement of the possible values, together with the probabilities, gives the *probability distribution* of x. The probability p_i is, of course, to be interpreted as approximately the proportion of cases in which x takes the value x_i in a large series of repetitions of the experiment.

One important characteristic of a random variable is its *expectation*. Thus, let $x(e_\alpha)$ be the value of x corresponding to the elementary event e_α $(\alpha=1, 2, \ldots\ldots, r)$, and let $P(e_\alpha)$ be the probability associated with e_α. Then

$$E(x)=\sum_{\alpha=1}^{r} x(e_\alpha)P(e_\alpha) \qquad \ldots \quad (3.15)$$

is the expectation of x. $E(x)$ is often denoted by μ_x or, simply, μ.

Suppose now that the elementary events are numbered in such a way that $x(e_\alpha)$ for $\alpha=1, 2, \ldots\ldots, r_1$ are all equal to x_1, $x(e_\alpha)$ for $\alpha=r_1+1, r_1+2, \ldots\ldots, r_2$ are all equal to x_2, and so on. Then (3.15) may be written alternatively as

$$E(x)=x_1\sum_{\alpha=1}^{r_1} P(e_\alpha)+x_2\sum_{\alpha=r_1+1}^{r_2} P(e_\alpha)+\ldots\ldots+x_k\sum_{\alpha=r_{k-1}+1}^{r_k} P(e_\alpha).$$

* Actually, here we have a special type of random variable, which is the only appropriate type for a finite sample space. For some other types, see Section 9.2.

But the coefficient of x_i here is nothing but the probability

$$P[x=x_i]=p_i.$$

Hence we have also

$$E(x)=\sum_{i=1}^{k} x_i p_i. \qquad \cdots \quad (3.16)$$

In this form the expectation of x is seen to be the sum of the products of the different possible values of x by their probabilities. Since p_i is the 'long-run relative frequency' with which x assumes the value x_i, $E(x)$ may also be interpreted as the 'long-run average value' of x. (*vide* Section **6.3**).

There is yet a third formula for $E(x)$. Let $A_1, A_2, \ldots\ldots, A_t$ be a set of exhaustive and mutually exclusive events such that x takes the same value x'_j for all elementary events that are favourable to A_j (for each $j=1, 2, \ldots\ldots, t$). Then (3.15) may be expressed in the form

$$E(x)=\sum_{j=1}^{t} x'_j P(A_j). \qquad \cdots \quad (3.17)$$

Indeed, this is the most general formula for $E(x)$. In actual application, we use one of these formulæ—generally the one that best serves our purpose in the given context.

We have the following theorems on expectation :

Theorem 3.5 If $x=a$, a constant, then $E(x)=a$.

Proof : Since $x=a$, we have $x(e_\alpha)=a$ for all α. Hence (3.15) gives

$$E(x)=a\sum_{\alpha=1}^{r} P(e_\alpha)=a,$$

since $\sum_{\alpha=1}^{r} P(e_\alpha)=1$.

Theorem 3.6 If $y=bx$, then $E(y)=bE(x)$.

Proof : Corresponding to the elementary event e_α, $x(e_\alpha)$ is the value of x and $y(e_\alpha)$ is the value of y. Then

$$E(y)=\sum_{\alpha=1}^{r} y(e_\alpha)P(e_\alpha)$$

$$=b\sum_{\alpha=1}^{r} x(e_\alpha)P(e_\alpha)$$

$$=bE(x).$$

Theorem 3.7 If x and y be two random variables and z a third random variable such that $z=x+y$, then $E(z)=E(x)+E(y)$.

Proof: Corresponding to the elementary event e_α, x, y and z have the values $x(e_\alpha)$, $y(e_\alpha)$ and $z(e_\alpha)$, respectively.

Further, $z(e_\alpha)=x(e_\alpha)+y(e_\alpha)$. Hence

$$\begin{aligned} E(z) &= \sum_{\alpha=1}^{r} z(e_\alpha)P(e_\alpha) \\ &= \sum_{\alpha=1}^{r} x(e_\alpha)P(e_q)+\sum_{\alpha=1}^{r} y(e_\alpha)P(e_\alpha) \\ &= E(x)+E(y). \end{aligned}$$

Theorem 3.8 If $y=a+bx$, then $E(y)=a+bE(x)$.

Proof: $E(y)=E(a)+E(bx)$ from *Theorem 3.7*

$=a+bE(x)$ from *Theorems 3.5* and *3.6*.

Another characteristic of a random variable is its *variance*, which serves as a measure of the variation or dispersion of the random variable about its expectation. The variance of a random variable x is defined by

$$\text{var}(x)=E[x-E(x)]^2. \qquad \dots \quad (3.18)$$

Often var(x) is denoted by σ_x^2 or, simply, σ^2. The positive square-root of the variance is called the *standard deviation* (denoted by σ_x or σ).

Since

$$[x-E(x)]^2=x^2-2xE(x)+[E(x)]^2,$$

we have, by virtue of *Theorems 3.5—3.8*,

$$\begin{aligned} \text{var}(x) &= E(x^2)-2E(x)\cdot E(x)+[E(x)]^2 \\ &= E(x^2)-[E(x)]^2. \qquad \dots \quad (3.19) \end{aligned}$$

We have the following theorems on variance :

Theorem 3.9 If $x=a$, a constant, then var$(x)=0$.

Proof: From *Theorem 3.5*, $E(x)=a$. Hence $[x-E(x)]^2=0$ and so, from *Theorem 3.5* again,

$$\begin{aligned} \text{var}(x) &= E[x-E(x)]^2 \\ &= 0. \end{aligned}$$

Theorem 3.10 If $y=bx$, then $\text{var}(y)=b^2\text{var}(x)$.

Proof: From *Theorem 3.6*, $E(y)=bE(x)$.

Hence $$y-E(y)=b[x-E(x)]$$

and $$[y-E(y)]^2=b^2[x-E(x)]^2.$$

On applying *Theorem 3.6* again, we have

$$E[y-E(y)]^2=b^2E[x-E(x)]^2,$$

i.e. $$\text{var}(y)=b^2\text{var}(x).$$

Theorem 3.11 If $y=a+bx$, then $\text{var}(y)=b^2\text{var}(x)$.

Proof: *Theorem 3.8* gives $E(y)=a+bE(x)$. Hence $y-E(y)=b[x-E(x)]$. Next, proceeding as in the proof of *Theorem 3.10*, we have the stated result.

Ex. 3·12 Suppose two players, A and B, agree to play a game under the condition that A will get from B a rupees if he wins and will pay to B b rupees if he loses. Let the probability of A's winning the game be p and that of B's winning the game be $q=1-p$.

A's gain is then a random variable x, assuming two values, a (with probability p) and $-b$ (with probability q). Hence

$$E(x)=ap-bq$$

and
$$\begin{aligned}\text{var}(x)&=[a-E(x)]^2p+[-b-E(x)]^2q\\&=(a+b)^2q^2p+(a+b)^2p^2q\\&=(a+b)^2pq.\end{aligned}$$

Ex. 3·13 Let a die be thrown repeatedly till the first six appears. The number of throws needed to get the first six is then a random variable x, taking the values 1, 2, 3, ... *ad inf.* Further, if the die is perfect, the probability of getting a six in a throw is 1/6 and that of getting one of the other values (viz. 1, 2, ..., 5) is 5/6. Hence

$$P[x=k]=\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6},$$

for the throws are made independently of each other and $x=k$ if, and only if, a six is obtained in the kth throw but in none of the earlier throws.

The mathematical expectation of x is, therefore,

$$E(x)=\sum_{k=1}^{\infty} k\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}$$

$$=\frac{1}{6}\left[1+2\left(\frac{5}{6}\right)+3\left(\frac{5}{6}\right)^2+\ldots\ldots\right]$$

$$=\frac{1}{6}\left(1-\frac{5}{6}\right)^{-2}=6.$$

Thus 'on the average' six throws will be needed to get the first six.

Again,

$$x^2=x(x-1)+x,$$

so that

$$E(x^2)=E[x(x-1)]+E(x)$$

$$=\sum_{k=1}^{\infty} k(k-1)\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}+6$$

$$=\frac{1}{6}\left[2\times1\left(\frac{5}{6}\right)+3\times2\left(\frac{5}{6}\right)^2+4\times3\left(\frac{5}{6}\right)^3+\ldots\ldots\right]$$

$$+6$$

$$=2\times\frac{1}{6}\times\frac{5}{6}\left[1+3\left(\frac{5}{6}\right)+\frac{3\times4}{2!}\left(\frac{5}{6}\right)^2+\frac{3\times4\times5}{3!}\left(\frac{5}{6}\right)^3+\ldots\ldots\right]$$

$$+6$$

$$=2\times\frac{1}{6}\times\frac{5}{6}\left(1-\frac{5}{6}\right)^{-3}+6=60+6=66.$$

Hence

$$\text{var}(x)=E(x^2)-[E(x)]^2$$

$$=66-36=30.$$

3.9 Joint distribution of two random variables