FROM PROBABILITY BY SEYMOUR LIPSCHUTZ

Chapter 5

Random Variables

INTRODUCTION

We recall the concept of a function. Let S and T be arbitrary sets. Suppose to each $s \in S$ there is assigned a unique element of T; the collection f of such assignments is called a *function* (or: *mapping* or *map*) from S into T, and is written $f: S \to T$. We write $f(s)$ for the element of T that f assigns to $s \in S$, and call it the *image* of s under f or the *value* of f at s. The *image* $f(A)$ of any subset A of S, and the *preimage* $f^{-1}(B)$ of any subset B of T are defined by

$$f(A) = \{f(s) : s \in A\} \quad \text{and} \quad f^{-1}(B) = \{s : f(s) \in B\}$$

In words, $f(A)$ consists of the images of points of A and $f^{-1}(B)$ consists of those points whose images belong to B. In particular, the set $f(S)$ of all the image points is called the *image set* (or: *image* or *range*) of f.

Now suppose S is the sample space of some experiment. As noted previously, the outcomes of the experiment, i.e. the sample points of S, need not be numbers. However, we frequently wish to assign a specific number to each outcome, e.g. the sum of the points on a pair of dice, the number of aces in a bridge hand, or the time (in hours) it takes for a lightbulb to burn out. Such an assignment is called a random variable; more precisely,

Definition: A *random variable* X on a sample space S is a function from S into the set $\mathbf{R}$ of real numbers such that the preimage of every interval of $\mathbf{R}$ is an event of S.

We emphasize that if S is a discrete space in which every subset is an event, then every real-valued function on S is a random variable. On the other hand, it can be shown that if S is uncountable then certain real-valued functions on S are not random variables.

If X and Y are random variables on the same sample space S, then $X+Y$, $X+k$, kX and XY (where k is a real number) are the functions on S defined by

$$(X+Y)(s) = X(s) + Y(s) \qquad (kX)(s) = kX(s)$$
$$(X+k)(s) = X(s) + k \qquad (XY)(s) = X(s)\,Y(s)$$

for every $s \in S$. It can be shown that these are also random variables. (This is trivial in the case that every subset of S is an event.)

We use the short notation $P(X = a)$ and $P(a \le X \le b)$ for the probability of the events "X maps into a" and "X maps into the interval $[a, b]$." That is,

$$P(X = a) = P(\{s \in S : X(s) = a\})$$

and

$$P(a \le X \le b) = P(\{s \in S : a \le X(s) \le b\})$$

Analogous meanings are given to $P(X \le a)$, $P(X = a,\ Y = b)$, $P(a \le X \le b,\ c \le Y \le d)$, etc.

DISTRIBUTION AND EXPECTATION OF A FINITE RANDOM VARIABLE

Let X be a random variable on a sample space S with a finite image set; say, $X(S) = \{x_1, x_2, \ldots, x_n\}$. We make $X(S)$ into a probability space by defining the probability of x_i to be $P(X = x_i)$ which we write $f(x_i)$. This function f on $X(S)$, i.e. defined by $f(x_i) = P(X = x_i)$, is called the *distribution* or *probability function* of X and is usually given in the form of a table:

x_1	x_2	$\cdots$	x_n
$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$

The distribution f satisfies the conditions

$$\text{(i)}\ f(x_i) \geqq 0 \quad \text{and} \quad \text{(ii)}\ \sum_{i=1}^{n} f(x_i) = 1$$

Now if X is a random variable with the above distribution, then the *mean* or *expectation* (or: *expected value*) of X, denoted by $E(X)$ or μ_X, or simply E or μ, is defined by

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

That is, $E(X)$ is the *weighted average* of the possible values of X, each value weighted by its probability.

Example 5.1: A pair of fair dice is tossed. We obtain the finite equiprobable space S consisting of the 36 ordered pairs of numbers between 1 and 6:

$$S = \{(1,1), (1,2), \ldots, (6,6)\}$$

Let X assign to each point (a, b) in S the maximum of its numbers, i.e. $X(a, b) = \max(a, b)$. Then X is a random variable with image set

$$X(S) = \{1, 2, 3, 4, 5, 6\}$$

We compute the distribution f of X:

$$f(1) = P(X=1) = P(\{(1,1)\}) = \tfrac{1}{36}$$

$$f(2) = P(X=2) = P(\{(2,1), (2,2), (1,2)\}) = \tfrac{3}{36}$$

$$f(3) = P(X=3) = P(\{(3,1), (3,2), (3,3), (2,3), (1,3)\}) = \tfrac{5}{36}$$

$$f(4) = P(X=4) = P(\{(4,1), (4,2), (4,3), (4,4), (3,4), (2,4), (1,4)\}) = \tfrac{7}{36}$$

Similarly,

$$f(5) = P(X=5) = \tfrac{9}{36} \quad \text{and} \quad f(6) = P(X=6) = \tfrac{11}{36}$$

This information is put in the form of a table as follows:

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

We next compute the mean of X:

$$E(X) = \sum x_i f(x_i) = 1 \cdot \tfrac{1}{36} + 2 \cdot \tfrac{3}{36} + 3 \cdot \tfrac{5}{36} + 4 \cdot \tfrac{7}{36} + 5 \cdot \tfrac{9}{36} + 6 \cdot \tfrac{11}{36}$$

$$= \tfrac{161}{36} = 4.47$$

Now let Y assign to each point (a, b) in S the sum of its numbers, i.e. $Y(a, b) = a + b$. Then Y is also a random variable on S with image set

$$Y(S) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

The distribution g of Y follows:

y_i	2	3	4	5	6	7	8	9	10	11	12
$g(y_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

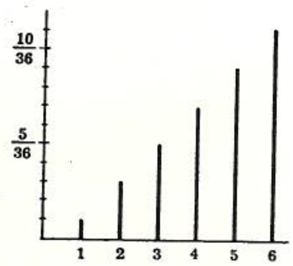
We obtain, for example, $g(4) = \frac{3}{36}$ from the fact that $(1,3)$, $(2,2)$, and $(3,1)$ are those points of S for which the sum of the components is 4; hence

$$g(4) = P(Y=4) = P(\{(1,3),(2,2),(3,1)\}) = \tfrac{3}{36}$$

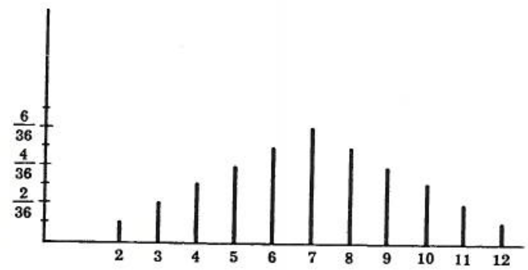
The mean of Y is computed as follows:

$$E(Y) = \sum y_i\, g(y_i) = 2\cdot\tfrac{1}{36} + 3\cdot\tfrac{2}{36} + \cdots + 12\cdot\tfrac{1}{36} = 7$$

The charts which follow graphically describe the above distributions:



Distribution of X



Distribution of Y

Observe that the vertical lines drawn above the numbers on the horizontal axis are proportional to their probabilities.

Example 5.2: A coin weighted so that $P(\text{H}) = \frac{2}{3}$ and $P(\text{T}) = \frac{1}{3}$ is tossed three times. The probabilities of the points in the sample space $S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$ are as follows:

$$P(\text{HHH}) = \tfrac{2}{3}\cdot\tfrac{2}{3}\cdot\tfrac{2}{3} = \tfrac{8}{27} \qquad P(\text{THH}) = \tfrac{1}{3}\cdot\tfrac{2}{3}\cdot\tfrac{2}{3} = \tfrac{4}{27}$$
$$P(\text{HHT}) = \tfrac{2}{3}\cdot\tfrac{2}{3}\cdot\tfrac{1}{3} = \tfrac{4}{27} \qquad P(\text{THT}) = \tfrac{1}{3}\cdot\tfrac{2}{3}\cdot\tfrac{1}{3} = \tfrac{2}{27}$$
$$P(\text{HTH}) = \tfrac{2}{3}\cdot\tfrac{1}{3}\cdot\tfrac{2}{3} = \tfrac{4}{27} \qquad P(\text{TTH}) = \tfrac{1}{3}\cdot\tfrac{1}{3}\cdot\tfrac{2}{3} = \tfrac{2}{27}$$
$$P(\text{HTT}) = \tfrac{2}{3}\cdot\tfrac{1}{3}\cdot\tfrac{1}{3} = \tfrac{2}{27} \qquad P(\text{TTT}) = \tfrac{1}{3}\cdot\tfrac{1}{3}\cdot\tfrac{1}{3} = \tfrac{1}{27}$$

Let X be the random variable which assigns to each point in S the largest number of successive heads which occurs. Thus,

$$X(\text{TTT}) = 0$$
$$X(\text{HTH}) = 1, \quad X(\text{HTT}) = 1, \quad X(\text{THT}) = 1, \quad X(\text{TTH}) = 1$$
$$X(\text{HHT}) = 2, \quad X(\text{THH}) = 2$$
$$X(\text{HHH}) = 3$$

The image set of X is $X(S) = \{0,1,2,3\}$. We compute the distribution f of X:

$$f(0) = P(\text{TTT}) = \tfrac{1}{27}$$
$$f(1) = P(\{\text{HTH, HTT, THT, TTH}\}) = \tfrac{4}{27}+\tfrac{2}{27}+\tfrac{2}{27}+\tfrac{2}{27} = \tfrac{10}{27}$$
$$f(2) = P(\{\text{HHT, THH}\}) = \tfrac{4}{27}+\tfrac{4}{27} = \tfrac{8}{27}$$
$$f(3) = P(\text{HHH}) = \tfrac{8}{27}$$

This information is put in the form of a table as follows:

x_i	0	1	2	3
$f(x_i)$	$\frac{1}{27}$	$\frac{10}{27}$	$\frac{8}{27}$	$\frac{8}{27}$

The mean of X is computed as follows:

$$E(X) = \sum x_i f(x_i) = 0\cdot\tfrac{1}{27} + 1\cdot\tfrac{10}{27} + 2\cdot\tfrac{8}{27} + 3\cdot\tfrac{8}{27} = \tfrac{50}{27} = 1.85$$

Example 5.3: A sample of 3 items is selected at random from a box containing 12 items of which 3 are defective. Find the expected number E of defective items.

The sample space S consists of the $\binom{12}{3} = 220$ distinct equally likely samples of size 3. We note that there are:

$$\binom{9}{3} = 84 \text{ samples with no defective items;}$$

$$3\cdot\binom{9}{2} = 108 \text{ samples with 1 defective item;}$$

$$\binom{3}{2}\cdot 9 = 27 \text{ samples with 2 defective items;}$$

$$\binom{3}{3} = 1 \text{ sample with 3 defective items.}$$

Thus the probability of getting 0, 1, 2 and 3 defective items is respectively 84/220, 108/220, 27/220 and 1/220. Thus the expected number E of defective items is

$$E = 0\cdot\tfrac{84}{220} + 1\cdot\tfrac{108}{220} + 2\cdot\tfrac{27}{220} + 3\cdot\tfrac{1}{220} = \tfrac{165}{220} = .75$$

Remark: Implicitly we have obtained the expectation of the random variable X which assigns to each sample the number of defective items in the sample.

In a gambling game, the expected value E of the game is considered to be the value of the game to the player. The game is said to be *favorable* to the player if E is positive, and *unfavorable* if E is negative. If $E = 0$, the game is *fair*.

Example 5.4: A player tosses a fair die. If a prime number occurs he wins that number of dollars, but if a non-prime number occurs he loses that number of dollars. The possible outcomes x_i of the game with their respective probabilities $f(x_i)$ are as follows:

x_i	2	3	5	−1	−4	−6
$f(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The negative numbers −1, −4 and −6 correspond to the fact that the player loses if a non-prime number occurs. The expected value of the game is

$$E = 2\cdot\tfrac{1}{6} + 3\cdot\tfrac{1}{6} + 5\cdot\tfrac{1}{6} - 1\cdot\tfrac{1}{6} - 4\cdot\tfrac{1}{6} - 6\cdot\tfrac{1}{6} = -\tfrac{1}{6}$$

Thus the game is unfavorable to the player since the expected value is negative.

Our first theorems relate the notion of expectation to operations on random variables.

Theorem 5.1: Let X be a random variable and k a real number. Then (i) $E(kX) = kE(X)$ and (ii) $E(X+k) = E(X) + k$.

Theorem 5.2: Let X and Y be random variables on the same sample space S. Then $E(X+Y) = E(X) + E(Y)$.

A simple induction argument yields

Corollary 5.3: Let $X_1, X_2, \ldots, X_n$ be random variables on S. Then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$$

VARIANCE AND STANDARD DEVIATION

The mean of a random variable X measures, in a certain sense, the "average" value of X. The next concept, that of the variance of X, measures the "spread" or "dispersion" of X.

Let X be a random variable with the following distribution:

x_1	x_2	$\cdots$	x_n
$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$

Then the *variance* of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = E((X-\mu)^2)$$

where μ is the mean of X. The *standard deviation* of X, denoted by σ_X, is the (nonnegative) square root of $\mathrm{Var}(X)$:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

The next theorem gives us an alternate and sometimes more useful formula for calculating the variance of the random variable X.

Theorem 5.4: $\mathrm{Var}(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = E(X^2) - \mu^2$.

Proof. Using $\sum x_i f(x_i) = \mu$ and $\sum f(x_i) = 1$, we have

$$\begin{aligned}\sum (x_i - \mu)^2 f(x_i) &= \sum (x_i^2 - 2\mu x_i + \mu^2) f(x_i) \\ &= \sum x_i^2 f(x_i) - 2\mu \sum x_i f(x_i) + \mu^2 \sum f(x_i) \\ &= \sum x_i^2 f(x_i) - 2\mu^2 + \mu^2 = \sum x_i^2 f(x_i) - \mu^2\end{aligned}$$

which proves the theorem.

Example 5.5: Consider the random variable X of Example 5.1 (which assigns the maximum of the numbers showing on a pair of dice). The distribution of X is

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

and its mean is $\mu_X = 4.47$. We compute the variance and standard deviation of X. First we compute $E(X^2)$:

$$\begin{aligned}E(X^2) = \sum x_i^2 f(x_i) &= 1^2\cdot\tfrac{1}{36} + 2^2\cdot\tfrac{3}{36} + 3^2\cdot\tfrac{5}{36} + 4^2\cdot\tfrac{7}{36} + 5^2\cdot\tfrac{9}{36} + 6^2\cdot\tfrac{11}{36} \\ &= \tfrac{791}{36} = 21.97\end{aligned}$$

Hence

$$\mathrm{Var}(X) = E(X^2) - \mu_X^2 = 21.97 - 19.98 = 1.99 \quad \text{and} \quad \sigma_X = \sqrt{1.99} = 1.4$$

Now consider the random variable Y of Example 5.1 (which assigns the sum of the numbers showing on a pair of dice). The distribution of Y is

y_i	2	3	4	5	6	7	8	9	10	11	12
$g(y_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

and its mean is $\mu_Y = 7$. We compute the variance and standard deviation of Y. First we compute $E(Y^2)$:

$$E(Y^2) = \sum y_i^2\, g(y_i) = 2^2 \cdot \tfrac{1}{36} + 3^2 \cdot \tfrac{2}{36} + \cdots + 12^2 \cdot \tfrac{1}{36} = \tfrac{1974}{36} = 54.8$$

Hence

$$\mathrm{Var}(Y) = E(Y^2) - \mu_Y^2 = 54.8 - 49 = 5.8 \quad \text{and} \quad \sigma_Y = \sqrt{5.8} = 2.4$$

We establish some properties of the variance in

Theorem 5.5: Let X be a random variable and k a real number. Then (i) $\mathrm{Var}(X+k) = \mathrm{Var}(X)$ and (ii) $\mathrm{Var}(kX) = k^2\mathrm{Var}(X)$. Hence $\sigma_{X+k} = \sigma_X$ and $\sigma_{kX} = |k|\sigma_X$.

Remark 1. There is a physical interpretation of mean and variance. Suppose at each point x_i on the x axis there is placed a unit with mass $f(x_i)$. Then the mean is the center of gravity of the system, and the variance is the moment of inertia of the system.

Remark 2. Many random variables give rise to the same distribution; hence we frequently speak of the mean, variance and standard deviation of a distribution instead of the underlying random variable.

Remark 3. Let X be a random variable with mean μ and standard deviation $\sigma > 0$. The *standardized random variable* X^* corresponding to X is defined by

$$X^* = \frac{X - \mu}{\sigma}$$

We show (Problem 5.23) that $E(X^*) = 0$ and $\mathrm{Var}(X^*) = 1$.

JOINT DISTRIBUTION

Let X and Y be random variables on a sample space S with respective image sets

$$X(S) = \{x_1, x_2, \ldots, x_n\} \quad \text{and} \quad Y(S) = \{y_1, y_2, \ldots, y_m\}$$

We make the product set

$$X(S) \times Y(S) = \{(x_1, y_1), (x_1, y_2), \ldots, (x_n, y_m)\}$$

into a probability space by defining the *probability* of the ordered pair (x_i, y_j) to be $P(X = x_i, Y = y_j)$ which we write $h(x_i, y_j)$. This function h on $X(S) \times Y(S)$, i.e. defined by $h(x_i, y_j) = P(X = x_i, Y = y_j)$, is called the *joint distribution* or *joint probability function* of X and Y and is usually given in the form of a table:

X \ Y	y_1	y_2	$\cdots$	y_m	Sum
x_1	$h(x_1, y_1)$	$h(x_1, y_2)$	$\cdots$	$h(x_1, y_m)$	$f(x_1)$
x_2	$h(x_2, y_1)$	$h(x_2, y_2)$	$\cdots$	$h(x_2, y_m)$	$f(x_2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$h(x_n, y_1)$	$h(x_n, y_2)$	$\cdots$	$h(x_n, y_m)$	$f(x_n)$
Sum	$g(y_1)$	$g(y_2)$	$\cdots$	$g(y_m)$	

The above functions f and g are defined by

$$f(x_i) = \sum_{j=1}^{m} h(x_i, y_j) \quad \text{and} \quad g(y_j) = \sum_{i=1}^{n} h(x_i, y_j)$$

i.e. $f(x_i)$ is the sum of the entries in the ith row and $g(y_j)$ is the sum of the entries in the jth column; they are called the *marginal distributions* and are, in fact, the (individual) distributions of X and Y respectively (Problem 5.12). The joint distribution h satisfies the conditions

$$\text{(i)}\ h(x_i, y_j) \geq 0 \quad \text{and} \quad \text{(ii)}\ \sum_{i=1}^{n} \sum_{j=1}^{m} h(x_i, y_j) = 1$$

Now if X and Y are random variables with the above joint distribution (and respective means μ_X and μ_Y), then the *covariance of X and Y*, denoted by $\mathrm{Cov}(X, Y)$, is defined by

$$\boxed{\mathrm{Cov}(X, Y)} = \sum_{i,j} (x_i - \mu_X)(y_j - \mu_Y)\, h(x_i, y_j) = E[(X - \mu_X)(Y - \mu_Y)]$$

or equivalently (see Problem 5.18) by

$$\mathrm{Cov}(X, Y) = \sum_{i,j} x_i y_j\, h(x_i, y_j) - \mu_X \mu_Y = E(XY) - \mu_X \mu_Y$$

The *correlation* of X and Y, denoted by $\rho(X, Y)$, is defined by

$$\boxed{\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}}$$

The correlation ρ is dimensionless and has the following properties:

(i) $\rho(X, Y) = \rho(Y, X)$ (iii) $\rho(X, X) = 1,\ \rho(X, -X) = -1$

(ii) $-1 \leq \rho \leq 1$ (iv) $\rho(aX + b, cY + d) = \rho(X, Y)$, if $a, c \neq 0$

We show below (Example 5.7) that pairs of random variables with identical (individual) distributions can have distinct covariances and correlations. Thus $\mathrm{Cov}(X, Y)$ and $\rho(X, Y)$ are measurements of the way that X and Y are interrelated.

Example 5.6: A pair of fair dice is tossed. We obtain the finite equiprobable space S consisting of the 36 ordered pairs of numbers between 1 and 6:

$$S = \{(1, 1), (1, 2), \ldots, (6, 6)\}$$

Let X and Y be the random variables on S in Example 5.1, i.e. X assigns the maximum of the numbers and Y the sum of the numbers to each point of S. The joint distribution of X and Y follows:

X \ Y	2	3	4	5	6	7	8	9	10	11	12	Sum
1	$\frac{1}{36}$	0	0	0	0	0	0	0	0	0	0	$\frac{1}{36}$
2	0	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0	0	0	$\frac{3}{36}$
3	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0	$\frac{5}{36}$
4	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	$\frac{7}{36}$
5	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	$\frac{9}{36}$
6	0	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{11}{36}$
Sum	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	

The above entry $h(3,5) = \frac{2}{36}$ comes from the fact that $(3,2)$ and $(2,3)$ are the only points in S whose maximum number is 3 and whose sum is 5; hence

$$h(3,5) = P(X=3, Y=5) = P(\{(3,2), (2,3)\}) = \tfrac{2}{36}$$

The other entries are obtained in a similar manner.

We compute the covariance and correlation of X and Y. First we compute $E(XY)$:

$$\begin{aligned} E(XY) &= \sum x_i y_j \, h(x_i, y_j) \\ &= 1\cdot 2\cdot \tfrac{1}{36} + 2\cdot 3\cdot \tfrac{2}{36} + 2\cdot 4\cdot \tfrac{1}{36} + \cdots + 6\cdot 12\cdot \tfrac{1}{36} \\ &= \tfrac{1232}{36} = 34.2 \end{aligned}$$

By Example 5.1, $\mu_X = 4.47$ and $\mu_Y = 7$, and by Example 5.5, $\sigma_X = 1.4$ and $\sigma_Y = 2.4$; hence

$$\operatorname{Cov}(X,Y) = E(XY) - \mu_X\mu_Y = 34.2 - (4.47)(7) = 2.9$$

and
$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{2.9}{(1.4)(2.4)} = .86$$

Example 5.7: Let X and Y, and X' and Y' be random variables with the following joint distributions:

X \ Y	4	10	Sum
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
Sum	$\frac{1}{2}$	$\frac{1}{2}$	

X' \ Y'	4	10	Sum
1	0	$\frac{1}{2}$	$\frac{1}{2}$
3	$\frac{1}{2}$	0	$\frac{1}{2}$
Sum	$\frac{1}{2}$	$\frac{1}{2}$	

Observe that X and X', and Y and Y' have identical distributions:

x_i	1	3
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

Distribution of X and X'

y_i	4	10
$g(y_i)$	$\frac{1}{2}$	$\frac{1}{2}$

Distribution of Y and Y'

We show that $\operatorname{Cov}(X,Y) \neq \operatorname{Cov}(X',Y')$ and hence $\rho(X,Y) \neq \rho(X',Y')$. We first compute $E(XY)$ and $E(X'Y')$:

$$\begin{aligned} E(XY) &= 1\cdot 4\cdot \tfrac{1}{4} + 1\cdot 10\cdot \tfrac{1}{4} + 3\cdot 4\cdot \tfrac{1}{4} + 3\cdot 10\cdot \tfrac{1}{4} = 14 \\ E(X'Y') &= 1\cdot 4\cdot 0 + 1\cdot 10\cdot \tfrac{1}{2} + 3\cdot 4\cdot \tfrac{1}{2} + 3\cdot 10\cdot 0 = 11 \end{aligned}$$

Since $\mu_X = \mu_{X'} = 2$ and $\mu_Y = \mu_{Y'} = 7$,

$$\operatorname{Cov}(X,Y) = E(XY) - \mu_X\mu_Y = 0 \quad \text{and} \quad \operatorname{Cov}(X',Y') = E(X'Y') - \mu_{X'}\mu_{Y'} = -3$$

Remark: The notion of a joint distribution h is extended to any finite number of random variables $X, Y, \ldots, Z$ in the obvious way; that is, h is a function on the product set $X(S) \times Y(S) \times \cdots \times Z(S)$ defined by

$$h(x_i, y_j, \ldots, z_k) = P(X = x_i,\ Y = y_j,\ \ldots,\ Z = z_k)$$

INDEPENDENT RANDOM VARIABLES

A finite number of random variables $X, Y, \ldots, Z$ on a sample space S are said to be *independent* if

$$P(X = x_i,\ Y = y_j,\ \ldots,\ Z = z_k) = P(X = x_i)\,P(Y = y_j)\cdots P(Z = z_k)$$

for any values $x_i, y_j, \ldots, z_k$. In particular, X and Y are independent if

$$P(X = x_i,\ Y = y_j) \;=\; P(X = x_i)\,P(Y = y_j)$$

Now if X and Y have respective distributions f and g, and joint distribution h, then the above equation can be written as

$$h(x_i, y_j) \;=\; f(x_i)\,g(y_j)$$

In other words, X and Y are independent if each entry $h(x_i, y_j)$ is the product of its marginal entries.

Example 5.8: Let X and Y be random variables with the following joint distribution:

X \ Y	2	3	4	Sum
1	.06	.15	.09	.30
2	.14	.35	.21	.70
Sum	.20	.50	.30	

Thus the distributions of X and Y are as follows:

x	1	2
$f(x)$	.30	.70

Distribution of X

y	2	3	4
$g(y)$	.20	.50	.30

Distribution of Y

X and Y are independent random variables since each entry of the joint distribution can be obtained by multiplying its marginal entries; that is,

$$P(X = x_i,\ Y = y_j) \;=\; P(X = x_i)\,P(Y = y_j)$$

for each i and each j.

We establish some important properties of independent random variables which do not hold in general; namely,

Theorem 5.6: Let X and Y be independent random variables. Then:

(i) $E(XY) = E(X)E(Y)$,

(ii) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$,

(iii) $\text{Cov}(X, Y) = 0$.

Part (ii) in the above theorem generalizes to the very important

Theorem 5.7: Let $X_1, X_2, \ldots, X_n$ be independent random variables. Then

$$\text{Var}(X_1 + \cdots + X_n) \;=\; \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

FUNCTIONS OF A RANDOM VARIABLE

Let X and Y be random variables on the same sample space S. Then Y is said to be a *function* of X if Y can be represented $Y = \Phi(X)$ for some real-valued function Φ of a real variable; that is, if $Y(s) = \Phi[X(s)]$ for every $s \in S$. For example, kX, X^2, $X + k$ and $(X + k)^2$ are all functions of X with $\Phi(x) = kx$, x^2, $x + k$ and $(x + k)^2$ respectively. We have the fundamental

Theorem 5.8: Let X and Y be random variables on the same sample space S with $Y = \Phi(X)$. Then

$$E(Y) = \sum_{i=1}^{n} \Phi(x_i)\, f(x_i)$$

where f is the distribution function of X.

Similarly, a random variable Z is said to be a function of X and Y if Z can be represented $Z = \Phi(X, Y)$ where Φ is a real-valued function of two real variables; that is, if

$$Z(s) = \Phi[X(s), Y(s)]$$

for every $s \in S$. Corresponding to the above theorem, we have

Theorem 5.9: Let X, Y and Z be random variables on the same sample space S with $Z = \Phi(X, Y)$. Then

$$E(Z) = \sum_{i,j} \Phi(x_i, y_j)\, h(x_i, y_j)$$

where h is the joint distribution of X and Y.

We remark that the above two theorems have been used implicitly in the preceding discussion and theorems. We also remark that the proof of Theorem 5.9 is given as a supplementary problem, and that the theorem generalizes to a function of n random variables in the obvious way.

DISCRETE RANDOM VARIABLES IN GENERAL

Now suppose X is a random variable on S with a countably infinite image set; say $X(S) = \{x_1, x_2, \ldots\}$. Such random variables together with those with finite image sets (considered above) are called *discrete* random variables. As in the finite case, we make $X(S)$ into a probability space by defining the *probability* of x_i to be $f(x_i) = P(X = x_i)$ and call f the *distribution* of X:

x_1	x_2	x_3	$\cdots$
$f(x_1)$	$f(x_2)$	$f(x_3)$	$\cdots$

The *expectation* $E(X)$ and *variance* Var (X) are defined by

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots = \sum_{i=1}^{\infty} x_i f(x_i)$$

$$\mathrm{Var}(X) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \cdots = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i)$$

when the relevant series converge absolutely. It can be shown that Var (X) exists if and only if $\mu = E(X)$ and $E(X^2)$ both exist and that in this case the formula

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$

is valid just as in the finite case. When Var (X) exists, the *standard deviation* σ_X is defined as in the finite case by

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

The notions of joint distribution, independent random variables and functions of random variables carry over directly to the general case. It can be shown that if X and Y are defined on the same sample space S and if Var (X) and Var (Y) both exist, then the series

$$\operatorname{Cov}(X, Y) = \sum_{i,j} (x_i - \mu_X)(y_j - \mu_Y)\, h(x_i, y_j)$$

converges absolutely and the relation

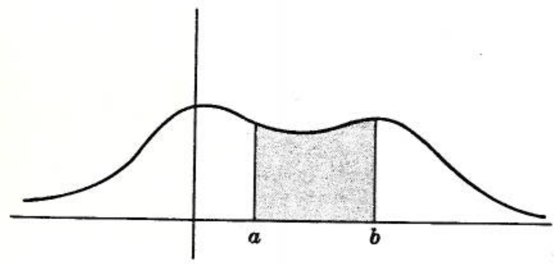
$$\operatorname{Cov}(X, Y) = \sum_{i,j} x_i y_j\, h(x_i, y_j) - \mu_X \mu_Y = E(XY) - \mu_X \mu_Y$$

holds just as in the finite case.

Remark: To avoid technicalities we will establish many theorems in this chapter only for finite random variables.

CONTINUOUS RANDOM VARIABLES

Suppose that X is a random variable whose image set $X(S)$ is a continuum of numbers such as an interval. Recall from the definition of random variables that the set $\{a \leq X \leq b\}$ is an event in S and therefore the probability $P(a \leq X \leq b)$ is well defined. We assume that there is a piecewise continuous function $f: \mathbf{R} \to \mathbf{R}$ such that $P(a \leq X \leq b)$ is equal to the area under the graph of f between $x = a$ and $x = b$ (as shown on the right). In the language of calculus,



$P(a \leq X \leq b)$ = area of shaded region

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

In this case X is said to be a *continuous random variable*. The function f is called the *distribution* or the *continuous probability function* (or: *density function*) of X; it satisfies the conditions

$$\text{(i)}\ f(x) \geq 0 \quad \text{and} \quad \text{(ii)}\ \int_{\mathbf{R}} f(x)\, dx = 1$$

That is, f is nonnegative and the total area under its graph is 1.

The *expectation* $E(X)$ is defined by

$$E(X) = \int_{\mathbf{R}} x\, f(x)\, dx$$

when it exists. Functions of random variables are defined just as in the discrete case; and it can be shown that if $Y = \Phi(X)$, then

$$E(Y) = \int_{\mathbf{R}} \Phi(x)\, f(x)\, dx$$

when the right side exists. The *variance* $\operatorname{Var}(X)$ is defined by

$$\operatorname{Var}(X) = E((X-\mu)^2) = \int_{\mathbf{R}} (x-\mu)^2 f(x)\, dx$$

when it exists. Just as in the discrete case, it can be shown that $\operatorname{Var}(X)$ exists if and only if $\mu = E(X)$ and $E(X^2)$ both exist and then

$$\operatorname{Var}(X) = E(X^2) - \mu^2 = \int_{\mathbf{R}} x^2 f(x)\, dx - \mu^2$$

The *standard deviation* σ_X is defined by $\sigma_X = \sqrt{\mathrm{Var}(X)}$ when Var (X) exists.

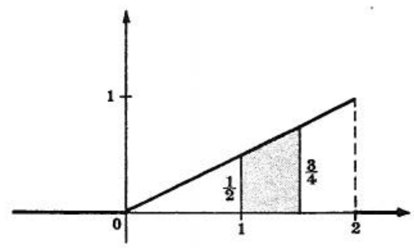
We have already remarked that we will establish many results for finite random variables and take them for granted in the general discrete case and in the continuous case.

Example 5.9: Let X be a continuous random variable with the following distribution:

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$P(1 \leq X \leq 1.5) = \text{area of shaded region in diagram}$$
$$= \tfrac{1}{2} \cdot \tfrac{1}{2}(\tfrac{1}{2} + \tfrac{3}{4}) = \tfrac{5}{16}$$



Graph of f

We next compute the expectation, variance and standard deviation of X:

$$E(X) = \int_{\mathbf{R}} x\, f(x)\, dx = \int_0^2 \tfrac{1}{2}x^2\, dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3}$$

$$E(X^2) = \int_{\mathbf{R}} x^2 f(x)\, dx = \int_0^2 \tfrac{1}{2}x^3\, dx = \left[\frac{x^4}{8}\right]_0^2 = 2$$

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = 2 - \frac{16}{9} = \frac{2}{9} \quad \text{and} \quad \sigma_X = \sqrt{\frac{2}{9}} = \frac{1}{3}\sqrt{2}$$

A finite number of continuous random variables, say $X, Y, \ldots, Z$, are said to be *independent* if for any intervals $[a, a'], [b, b'], \ldots, [c, c']$,

$$P(a \leq X \leq a', b \leq Y \leq b', \ldots, c \leq Z \leq c') = P(a \leq X \leq a')P(b \leq Y \leq b') \cdots P(c \leq Z \leq c')$$

Observe that intervals play the same role in the continuous case as points did in the discrete case.

CUMULATIVE DISTRIBUTION FUNCTION

Let X be a random variable (discrete or continuous). The *cumulative distribution function* F of X is the function $F: \mathbf{R} \to \mathbf{R}$ defined by

$$F(a) = P(X \leq a)$$

If X is a discrete random variable with distribution f, then F is the "step function" defined by

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

On the other hand, if X is a continuous random variable with distribution f, then

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

In either case, F is monotonic increasing, i.e.

$$F(a) \leq F(b) \quad \text{whenever} \quad a \leq b$$

and the limit of F to the left is 0 and to the right is 1:

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1$$