Applying Equation (14.39) we see that a joint density g of (U, V) is given by the formula

$$g(u, v) = f\left(\frac{u+v}{2}, \frac{u-v}{2}\right)\left|\frac{\partial(Q, R)}{\partial(u, v)}\right| = \frac{1}{2} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

To obtain a density $f_U = f_{X+Y}$ we integrate with respect to v and find

$$f_{X+Y}(u) = \frac{1}{2}\int_{-\infty}^{\infty} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) dv .$$

The change of variable $x = \frac{1}{2}(u + v)$, $dx = \frac{1}{2}\, du$, transforms this to

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f(x, u - x)\, dx .$$

Similarly, we find

$$f_{X-Y}(v) = \frac{1}{2}\int_{-\infty}^{\infty} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) du = \int_{-\infty}^{\infty} f(x, x - v)\, dx .$$

An important special case occurs when X and Y are independent. In this case the joint probability density factors into a product,

$$f(x, y) = f_X(x) f_Y(y),$$

and the integrals for f_{X+Y} and f_{X-Y} become

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(x) f_Y(u - x)\, dx, \qquad f_{X-Y}(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(x - v)\, dx .$$

EXAMPLE ***2. The sum of two exponential distributions.*** Suppose now that each of X and Y has an exponential distribution, say $f_X(t) = f_Y(t) = 0$ for $t < 0$, and

$$f_X(t) = \lambda e^{-\lambda t}, \qquad f_Y(t) = \mu e^{-\mu t} \qquad \text{for } t \geq 0 .$$

Determine the density of X + Y when X and Y are independent.

Solution. If $u < 0$ the integral for $f_{X+Y}(u)$ is 0 since the factor $f_X(x) = 0$ for x < 0, and the factor $f_Y(u - x) = 0$ for x $\geq$ 0. If $u \geq 0$ the integral for $f_{X+Y}(u)$ becomes

$$f_{X+Y}(u) = \int_0^u \lambda e^{-\lambda x} \mu e^{-\mu(u-x)}\, dx = \lambda\mu e^{-\mu u}\int_0^u e^{(\mu-\lambda)x}\, dx .$$

To evaluate the last integral we consider two cases, $\mu = \lambda$ and $\mu \neq 1$.

If $\mu = \lambda$ the integral has the value u and we obtain

$$f_{X+Y}(u) = \lambda^2 u e^{-\lambda u} \qquad \text{for } u \geq 0 .$$

If $\mu \neq \lambda$ we obtain

$$f_{X+Y}(u) = \lambda\mu e^{-\mu u}\,\frac{e^{(\mu-\lambda)u} - 1}{\mu - \lambda} = \lambda\mu\,\frac{e^{-\lambda u} - e^{-\mu u}}{\mu - \lambda} \qquad \text{for } u \geq 0 .$$

EXAMPLE 3. *The maximum and minimum of two independent random variables.* Let X and Y be two independent one-dimensional random variables with densities f_X and f_Y and corresponding distribution functions F_X and F_Y. Let U and V be the random variables

$$U = \max\{X, Y\}, \qquad V = \min\{X, Y\}.$$

That is, for each ω in the sample space, $U(\omega)$ is the maximum and $V(\omega)$ is the minimum of the two numbers $X(\omega)$, $Y(\omega)$. The mapping $u = \max\{x, y\}$, $v = \min\{x, y\}$ is not one-to-one, so the procedure used to deduce Equation (14.39) is not applicable. However, in this case we can obtain the distribution functions of U and V directly from first principles.

First we note that $U \le t$ if, and only if, $X \le t$ and $Y \le t$. Therefore $P(U \le t) = P(X \le t, Y \le t)$. By independence this is equal to $P(X \le t)P(Y \le t) = F_X(t)F_Y(t)$. Thus, we have

$$F_U(t) = F_X(t)F_Y(t).$$

At each point of continuity of f_X and f_Y we can differentiate this relation to obtain

$$f_U(t) = f_X(t)F_Y(t) + F_X(t)f_Y(t).$$

Similarly, we have $V > t$ if and only if $X > t$ and $Y > t$. Therefore

$$F_V(t) = P(V \le t) = 1 - P(V > t) = 1 - P(X > t, Y > t) = 1 - P(X > t)P(Y > t)$$
$$= 1 - (1 - F_X(t))(1 - F_Y(t)) = F_X(t) + F_Y(t) - F_X(t)F_Y(t).$$

At points of continuity of f_X and f_Y we differentiate this relation to obtain

$$f_V(t) = f_X(t) + f_Y(t) - f_X(t)F_Y(t) - F_X(t)f_Y(t).$$

14.24 Exercises

1. Let X and Y be two independent one-dimensional random variables, each with a uniform distribution over the interval [0, 1]. Let $U = X + Y$ and let $V = X - Y$.
   (a) Prove that U has a continuous density f_U given by

$$f_U(u) = \begin{cases} u & \text{if } 0 < u \le 1, \\ 2 - u & \text{if } 1 < u < 2, \\ 0 & \text{otherwise.} \end{cases}$$

   (b) Describe, in a similar way, a continuous density f_V for V.
   (c) Determine whether or not U and V are independent.
2. Let X and Y be as in Exercise 1, and let $U = \max\{X, Y\}$, $V = \min\{X, Y\}$.
   (a) Prove that U has a density function such that $f_U(t) = 2t$ for $0 \le t < 1$, and $f_U(t) = 0$ otherwise.
   (b) Describe a density function f_V for V.
   (c) Determine whether or not U and V are independent.

3. Let X and Y be two independent one-dimensional random variables, each having an exponential distribution with parameter $\lambda = 1$, and let $f(x, y) = f_X(x) f_Y(y)$, the product of the densities of X and Y.
(a) Let A denote the set of points in the xy-plane at which $f(x, y) > 0$. Make a sketch of A and of its image A' under the mapping defined by $u = x + y$, $v = x/(x + y)$.
(b) Let $U = X + Y$ and $V = X/(X + Y)$ be two new random variables, and compute a probability density g of (U, V).
(c) Compute a probability density&.
(d) Compute a probability density f_V.
4. Let X and Y be two independent random variables, each with a standard normal distribution (mean = 0, variance = 1). Introduce new random variables U and V by the equations $U = X/Y$, $V = Y$.
(a) Show that a probability density function of (U, V) is given by the formula

$$g(u, v) = -\frac{v}{2\pi} e^{-(1+u^2)v^2/2} \qquad \text{if} \quad v < 0.$$

(b) Find a similar formula for computing $g(u, v)$ when $v \geq 0$.
(c) Determine a probability density function of U.
5. Assume X has the density function given by

$$f_X(x) = \begin{cases} \dfrac{1}{\pi\sqrt{1 - x^2}} & \text{if} \quad -1 < x < 1, \\ 0 & \text{if } |x| \geq 1. \end{cases}$$

If an independent random variable Y has density

$$f_Y(y) = \begin{cases} y e^{-y^2/2} & \text{if} \quad y \geq 0, \\ 0 & \text{if} \quad y < 0, \end{cases}$$

find a density function of Z = X Y.
6. Given two independent one-dimensional random variables X and Y with continuous densities f_X and f_Y. Let U and V be two random variables such that $X = U \cos V$, $Y = U \sin V$, with $U > 0$ and $-\pi < V \leq \pi$.
(a) Prove that U has a density such that $f_U(u) = 0$ for $u < 0$, and

$$f_U(u) = u \int_{-\pi}^{\pi} f_X(u \cos v) f_Y(u \sin v)\, dv \qquad \text{for } u \geq 0.$$

(b) Determine f_U and the corresponding distribution F_U explicitly when each of X and Y has a normal distribution with mean $m = 0$ and variance σ^2.
7. (a) Assume $\sigma_1 > 0$ and $\sigma_2 > 0$. Verify the algebraic identity

$$\left(\frac{x - m_1}{\sigma_1}\right)^2 + \left(\frac{t - x - m_2}{\sigma_2}\right)^2 = \left(\frac{x - m_0}{\sigma_0}\right)^2 + \left(\frac{t - (m_1 + m_2)}{\sigma}\right)^2,$$

where

$$\sigma^2 = \sigma_1^2 + \sigma_2^2, \qquad \sigma_0^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma^2}, \qquad \text{and} \qquad m_0 = \frac{m_1\sigma_2^2 + (t - m_2)\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

(b) Given two independent one-dimensional random variables X and Y. Assume X has a normal distribution with mean m_1 and variance σ_1^2, and that Y has a normal distribution with mean m_2 and variance σ_2^2. Prove that X + Y has a normal distribution with mean $m = m_1 + m_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

8. Given two one-dimensional random variables X and Y with densities f_X and f_Y and joint density f. For each fixed y, define

$$f_X(x \mid Y = y) = \frac{f(x, y)}{f_Y(y)} \qquad \text{whenever } f_Y(y) > 0.$$

This is called the conditional probability density of X, given that Y = y. Similarly, we define the conditional probability density of Y, given that X = x, by the equation

$$f_Y(y \mid X = x) = \frac{f(x,y)}{f_X(x)} \qquad \text{whenever } f_X(x) > 0.$$

(a) If f_Y and f_X are positive, prove that $\int_{-\infty}^{\infty} f_X(x \mid Y = y)\, dx = \int_{-\infty}^{\infty} f_Y(y \mid X = x)\, dy = 1$.
(b) If f_Y and f_X are positive, prove that

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_X(x \mid Y = y)\, dy \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_X(x) f_Y(y \mid X = x)\, dx.$$

9. A random variable (X, Y) is said to have a ***normal bivariate distribution*** if its density function f is given by the formula

$$f(x, y) = \frac{\sqrt{D}}{2\pi} e^{-Q(x,y)/2},$$

where $Q(x, y)$ is the quadratic form

$$Q(x, y) = A_{11}(x - x_0)^2 + 2A_{12}(x - x_0)(y - y_0) + A_{22}(y - y_0)^2.$$

The numbers A_{11}, A_{12}, A_{22} are constants with $A_{11} > 0$. The number $D = A_{11}A_{22} - A_{12}^2$ is called the ***discriminant*** of Q and is assumed to be positive. The numbers x_0 and y_0 are arbitrary.
(a) Show that $Q(x, y)$ can be expressed as a sum of squares as follows:

$$Q(x, y) = A_{11}\left(u + \frac{A_{12}}{A_{11}} v\right)^2 + \frac{D}{A_{11}} v^2, \qquad \text{where } u = x - x_0,\ v = y - y_0.$$

(b) Define the "improper" double integral $\iint_{-\infty}^{+\infty} f(x, y)\, dx\, dy$ to be the limit

$$\iint_{-\infty}^{+\infty} f(x, y)\, dx\, dy = \lim_{t \to +\infty} \iint_{R(t)} f(x, y)\, dx\, dy,$$

where $R(t)$ is the square $[-t, t] \times [-t, t]$. Show that

$$\iint_{-\infty}^{+\infty} f(x,y)\, dx\, dy = 1.$$

[*Hint:* Use part (a) to transform the double integral over R(t) into a double integral in the uu-plane. Then perform a linear change of variables to simplify the integral and, finally, let $t \to +\infty$.]

10. If a two-dimensional random variable (X, Y) has a normal bivariate distribution as described in Exercise 9, show that X and Y themselves are normal one-dimensional random variables with means x_0 and y_0, respectively, and with variances $\sigma^2(X) = A_{22}/D$, $\sigma^2(Y) = A_{11}/D$.
11. If (X, Y) has a normal bivariate distribution as described in Exercise 9, show that the random variable Z = X + Y has a one-dimensional normal distribution with mean $x_0 + y_0$ and variance $(A_{11} - 2A_{12} + A_{22})/D$.

14.25 Expectation and variance

The mass interpretation of probability distributions may be carried a step further by introducing the concepts of ***expectation*** and ***variance.*** These play the same role in probability theory that "center of mass" and "moment of inertia" play in mechanics. Without the Stieltjes integral we must give separate definitions for the discrete and continuous cases.

DEFINITIONS OF EXPECTATION AND VARIANCE. ***Let X be a one-dimensional random variable. The expectation of X and the variance of X are real numbers denoted by E(X) and* Var (X) *respectively, and are defined as follows:***

(a) ***For a continuous random variable with density function f_X,***

$$E(X) = \int_{-\infty}^{+\infty} t\, f_X(t)\, dt\,,$$

$$\operatorname{Var}(X) = \int_{-\infty}^{+\infty} [t - E(X)]^2 f_X(t)\, dt\,.$$

(b) ***For a discrete random variable with mass points $x_1, x_2, \ldots$ having probabilities $p_k = P(X = x_k)$, we define***

$$E(X) = \sum_{k=1}^{\infty} x_k p_k\,,$$

$$\operatorname{Var}(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k\,.$$

Note: We say that *E(X)* and Var (X) *exist* only when the integral or series in question is ***absolutely convergent.*** It is understood that the series is a finite sum when the sample space is finite; in this case *E(X)* and Var (X) always exist. They also exist when f_X is 0 outside some finite interval.

The mathematical expectation E(X) is a theoretically computed value associated with the random variable X. In some respects, the distribution acts as though its entire mass were concentrated at a single point, ***E(X).*** The true significance of mathematical expectation in probability theory will be discussed in Section 14.29 in connection with the so-called "laws of large numbers."

In mechanics, a knowledge of the center of mass alone gives no indication of how the mass is spread or dispersed about its center. A measure of this dispersion is provided by the "second moment" or "moment of inertia." In probability theory, this second moment is the variance. It measures the tendency of a distribution to spread out from its expected value. In Section 14.28 we shall find that a small variance indicates that large deviations from the expected value are unlikely.

Although the expectation $E(X)$ may be positive or negative, the variance Var (X) is always nonnegative. The symbol σ^2 is also used to denote the variance. Its positive square root is called the standard deviation and is denoted by σ. The standard deviation is a weighted average; in fact, σ is a weighted root mean square of the distance of each value of X from the expected value $E(X)$. The analogous concept in mechanics is the "radius of gyration."

EXAMPLE 1. ***Uniform distribution.*** Let X have a uniform distribution over an interval $[a, b]$. Then $f(t) = 1/(b - a)$ if $a < t < b$, and $f(t) = 0$ otherwise. Therefore the expectation of X is given by

$$E(X) = \int_{-\infty}^{+\infty} tf(t)\,dt = \frac{1}{b-a}\int_a^b t\,dt = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Thus the mean is the mid-point of the interval. If we write m for $(a + b)/2$ and note that $m - a = b - m = (b - a)/2$ we find

$$\operatorname{Var}(X) = \frac{1}{b-a}\int_a^b (t-m)^2\,dt = \frac{1}{b-a}\int_{a-m}^{b-m} u^2\,du = \frac{(b-a)^2}{12}.$$

Note that the variance depends only on the *length* of the interval.

EXAMPLE 2. ***Binomial distribution.*** If X has a binomial distribution with parameters n and p we have

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k q^{n-k},$$

where $q = 1 - p$. To evaluate this sum, let

$$f(x, y) = (x + y)^n = \sum_{k=0}^{n}\binom{n}{k} x^k y^{n-k}$$

and note that

$$\sum_{k=0}^{n} k\binom{n}{k} x^{k-1} y^{n-k} = \frac{\partial f(x, y)}{\partial x} = n(x + y)^{n-1}.$$

If we multiply both sides of this last equation by x and put $x = p$ and $y = q$, we obtain $E(X) = np$.

By a similar argument we may deduce the formula

$$\operatorname{Var}(X) = \sum_{k=0}^{n}(k - np)^2\binom{n}{k} p^k q^{n-k} = npq.$$

Two proofs of this formula are suggested in Exercise 6 of Section 14.27.

EXAMPLE 3. *Normal distribution.* The terms "mean" and "variance" have already been introduced in connection with our description of the normal distribution in Section 14.14. These terms are justified by the formulas

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-[(t-m)/\sigma]^2/2}\, dt = m$$

and

$$\operatorname{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (t - m)^2 e^{-[(t-m)/\sigma]^2/2}\, dt = \sigma^2$$

Proofs of these formulas are requested in Exercise 7 of Section 14.27

Gamblers often use the concept of expectation to decide whether a given game of chance is favorable or unfavorable. As an illustration we shall consider the game of betting on "red" or "black" in roulette.

EXAMPLE 4. *Roulette.* A roulette wheel carries the numbers from 0 to 36. The number 0 appears on a gray background, half of the remaining 36 numbers on a red background, and the other half on a black background. The usual methods of betting are:

(1) Bet \$1 on a color (red or black). Possible return: \$2.
(2) Bet \$1 on a single number (0 excepted). Possible return: \$36.
(3) Bet \$1 on any dozen numbers (0 excepted). Possible return: \$3.

If 0 is the winning number the house wins and all other players lose.

Let X be the random variable which measures the financial outcome of betting by method (1). The possible values of X are $x_1 = -1$ and $x_2 = +1$. The point probabilities are $P(X = x_1) = \frac{19}{37}$, $P(X = x_2) = \frac{18}{37}$. Therefore the expectation is

$$E(X) = (-1)\tfrac{19}{37} + (+1)\tfrac{18}{37} = -\tfrac{1}{37};$$

this is usually interpreted to mean that the game is unfavorable to those who play it. The mathematical justification for this interpretation is provided by one of the *laws of large numbers,* to be discussed in Section 14.29 The reader may verify that the expectation has the same value for methods (2) and (3) as well.

EXAMPLE 5. *A coin-tossing game.* In a coin-tossing game there is a probability p that heads (H) will come up and a probability q that tails (T) will come up, where $0 \leq p \leq 1$ and $q = 1 - p$. The coin is tossed repeatedly until the first outcome occurs a second time; at this point the game ends. If the first outcome is H we are paid \$1 for each T that comes up until we get the next H. For example, $HTTTH$ pays \$3, but HH pays \$0. If the first outcome is T the same rules apply with H and T interchanged. The problem is to determine how much we should pay to play this game. For this purpose we shall consider the random variable which counts the number of dollars won and compute its expected value.

For the sample space we take the collection of all possible games that can be played in this manner. This set can be expressed as the union of two sets A and B, where

$$A = \{TT, THT, THHT, THHHT, \ldots\} \quad \text{and} \quad B = \{HH, HTH, HTTH, HTTTH, \ldots\}.$$

We denote the elements of set A (in the order listed) as $a_0, a_1, a_2, a_3, \ldots$ and those of set B as $b_0, b_1, b_2, b_3, \ldots$. Next, we assign the point probabilities as follows:

$$P(a_n) = p^n q^2 \qquad \text{and} \qquad P(b_n) = q^n p^2 .$$

(When $p = 0$, we put $P(a_0) = 1$ and let $P(x) = 0$ for all other x in $A \cup B$. When $q = 0$ we put $P(b_0) = 1$ and let $P(x) = 0$ for all other x.) In Section 13.21 it was shown that this is an acceptable assignment of probabilities.

The random variable X in which we are interested is defined on the sample space $A \cup B$ as follows :

$$X(a_n) = X(b_n) = n \qquad \text{for } n = 0, 1, 2, \ldots .$$

The event "$X = n$" consists of the two games a_n and b_n, so we have

$$P(X = n) = p^n q^2 + q^n p^2 ,$$

where p^0 and q^0 are to be interpreted as 1 when $p = 0$ or $q = 0$. The expectation of X is given by the sum

$$E(X) = \sum_{n=0}^{\infty} nP(X = n) = q^2 \sum_{n=0}^{\infty} np^n + p^2 \sum_{n=0}^{\infty} nq^n . \tag{14.40}$$

If either $p = 0$ or $q = 0$, we obtain $E(X) = 0$. Otherwise we may compute the sums of the series in (14.40) by noting that for $0 < x < 1$ we have

$$\sum_{n=0}^{\infty} nx^n = x \frac{d}{dx} \sum_{n=0}^{\infty} x^n = x \frac{d}{dx}\left(\frac{1}{1 - x}\right) = \frac{x}{(1 - x)^2} .$$

Using this in (14.40) with $x = p$ and $x = q$ we obtain, for $0 < p < 1$,

$$E(X) = \frac{q^2 p}{(1 - p)^2} + \frac{p^2 q}{(1 - q)^2} = p + q = 1 .$$

We interpret this result by saying that the game is unfavorable to those who pay more than \$1 to play it.

This particular example is of special interest because the expectation $E(X)$ is independent of p when $0 < p < 1$. In other words, loading the coin in favor of heads or tails does not affect the expected value except in the extreme cases in which it is so loaded that it always falls heads or always falls tails. Note that, as a function of p, the expectation $E(X)$ is discontinuous at the points $p = 0$ and $p = 1$. Otherwise it has the constant value 1. This interesting example was suggested to the author by H. S. Zuckerman.

14.26 Expectation of a function of a random variable

If a new random variable Y is related to a given one X by an equation of the form $Y = \varphi(X)$, its expectation is given (in the continuous case) by the equation

$$E(Y) = \int_{-\infty}^{+\infty} t f_Y(t)\, dt. \tag{14.41}$$

The expectation E(Y) can be computed directly in terms of the density f_X without determining the density of Y. In fact, the following formula is equivalent to (14.41):

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(t) f_X(t)\, dt\,. \tag{14.42}$$

A proof of (14.42) in the most general case is difficult and will not be attempted here. However, for many special cases of importance the proof is simple. In one such case, φ is differentiable and strictly increasing on the whole real axis, and takes on every real value. For a continuously distributed random variable X with density f_X we have the following formula for the density function f_Y (derived in Section 14.17):

$$f_Y(t) = f_X[\psi(t)] \cdot \psi'(t),$$

where ψ is the inverse of φ. If we use this in (14.41) and make the change of variable $u = \psi(t)$ [so that $t = \varphi(u)$], we obtain

$$E(Y) = \int_{-\infty}^{+\infty} t f_Y(t)\, dt = \int_{-\infty}^{+\infty} t f_X[\psi(t)] \cdot \psi'(t)\, dt = \int_{-\infty}^{+\infty} \varphi(u) f_X(u)\, du\,,$$

which is the same as (14.42).

When Equation (14.42) is applied to $Y = (X - m)^2$, where $m = E(X)$, we obtain

$$E(Y) = \int_{-\infty}^{+\infty} (t - m)^2 f_X(t)\, dt = \mathrm{Var}\,(X).$$

This shows that variance is itself an expectation. A formula analogous to (14.42) also holds, of course, in the discrete case. More generally, it can be shown that

$$E[\varphi(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f(x, y)\, dx\, dy$$

if (X, Y) is a continuous random variable with joint density f.

Note: For two-dimensional random variables, expectation and variance may be defined in a manner similar to that used for the one-dimensional case, except that double integrals and double sums are employed. We shall not discuss this extension here.

14.27 Exercises

1. A die is rolled. Let X denote the number of points on the upturned face. Compute E(X) and Var (X).
2. Assume that X is a continuous random variable with a probability density function. Let $Y = (X - m)/\sigma$, where $m = E(X)$ and $\sigma = \sqrt{\mathrm{Var}\,(X)}$. Show that $E(Y) = 0$ and $E(Y^2) = 1$.
3. Derive the following general properties of expectation and variance for either the discrete or the continuous case.
   (a) $E(cX) = cE(X)$, where c is a constant.
   (b) $\mathrm{Var}\,(cX) = c^2\, \mathrm{Var}\,(X)$, where c is a constant.
   (c) $E(X + Y) = E(X) + E(Y)$.
   (d) $\mathrm{Var}\,(X) = E(X^2) - [E(X)]^2$.
   (e) $\mathrm{Var}\,(X + Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) + 2E[(X - E(X))(Y - E(Y))]$.
   (f) $E[\varphi_1(X) + \varphi_2(Y)] = E[\varphi_1(X)] + E[\varphi_2(Y)]$. [Part (c) is a special case.]

4. If X and Y are independent random variables, show that
   (a) Var (X + Y) = Var (X) + Var (Y).
   (b) $E[\varphi(X)\cdot\psi(Y)] = E[\varphi(X)]\cdot E[\psi(Y)]$.
   (c) If $X_1, X_2, \ldots, X_n$ are independent random variables with $E(X_k) = m_k$, show that

$$\operatorname{Var}\left[\sum_{k=1}^{n}(X_k - m_k)\right] = \sum_{k=1}^{n}\operatorname{Var}(X_k - m_k) = \sum_{k=1}^{n}\operatorname{Var}(X_k).$$

5. Let $X_1, X_2, X_3, \ldots, X_n$ be n independent random variables, each having the same expectation, $E(X_k) = m$, and the same variance, Var $(X_k) = \sigma^2$. Let $\bar{X}$ denote the arithmetic mean, $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. Use Exercises 3 and 4 to prove that $E(X) = m$ and Var $(\bar{X}) = \sigma^2/n$.
6. (a) If $q = 1 - p$, prove the formula

$$\sum_{k=0}^{n}(k - np)^2\binom{n}{k}p^k q^{n-k} = npq,$$

   thereby showing that Var (X) = npq for a random variable X having a binomial distribution with parameters n andp. [Hint: $k^2 = k(k - 1) + k$.]
   (b) If X has a binomial distribution with parameters n and p, show that X can be expressed as a sum of n independent random variables $X_1, X_2, \ldots, X_n$, each assuming the possible values 0 and 1 with probabilities p and y, respectively, and each having a binomial distribution. Use this result and Exercise 5 to show that $E(X) = np$ and Var (X) = npq.
7. Determine the expectation and variance (whenever they exist) for a random variable X having
   (a) a Poisson distribution with parameter λ.
   (b) a Cauchy distribution.
   (c) an exponential distribution with parameter λ.
   (d) a normal distribution.
8. A random variable X has a probability density function given by

$$f(t) = \frac{C(r)}{|t|^r} \quad \text{if } |t| > 1, \qquad f(t) = 0 \quad \text{if } |t| \le 1,$$

   where $r > 1$ and $C(r)$ is independent of t.
   (a) Express $C(r)$ in terms of r and make a sketch to indicate the nature of the graph off.
   (b) Determine the corresponding distribution function F_X and make a sketch to indicate the nature of its graph.
   (c) Compute $P(X < 5)$ and $P(5 < X < 10)$ in terms of r.
   (d) For what values of r does Xhave a finite expectation? Compute $E(X)$ in terms of r when the expectation is finite.
   (e) For what values of r does X have a finite variance? Compute Var (X) in terms of r when the variance is finite.
9. A gambler plays roulette according to the following "system." He plays in sets of three games. In the first and second games he always bets \$1 on red. For the third game he proceeds as follows :
   (a) If he wins in the first and second games, he doesn't bet.
   (b) If he wins in one of the first or second and loses in the other, he bets \$1 on the color opposite to the outcome of the second game.
   (c) If he loses in both the first and second, he bets \$3 on red.
   Let X, Y, and Z denote, respectively, the financial outcomes of the first, second, and third games. Compute $E(X)$, $E(Y)$, $E(Z)$, and $E(X + Y + Z)$.

10. ***(Petersburg Problem).*** A player tosses a coin and wins \$1 if his first toss is heads. If he tosses heads again he wins another dollar. If he succeeds in tossing heads a third time he gets another \$2 (for a total of \$4). As long as he tosses heads in succession n times, his accumulated winnings are 2^{n-1} dollars. The game terminates when he tosses tails. Let X denote the number of dollars won in any particular game. Compute $E(X)$. In view of your result, how much would you be willing to pay Harold's Club in Reno for the privilege of playing this game?
11. (a) Assume X is a continuous random variable with probability density f_X. Let $Y = (X - m)/\sigma$, where $m = E(X)$ and $\sigma = \sqrt{\mathrm{Var}\,(X)}$. Prove that

$$E(e^Y) = e^{-m/\sigma}\int_{-\infty}^{\infty} e^{t/\sigma} f_X(t)\, dt .$$

(b) Let X be a discrete random variable having a Poisson distribution with parameter λ. Define Y as in part (a) and prove that

$$E(e^Y) = e^{-\lambda G(\lambda)}, \qquad \text{where } G(\lambda) = 1 + \frac{1}{\sqrt{\lambda}} - e^{1/\sqrt{\lambda}} .$$

12. A random variable X has a standard normal distribution. Compute: (a) $E(|X|)$, (b) $E(e^X)$, (c) Var (e^X), (d) $E(\sqrt{X^2 + Y^2})$. In part (d), Y also has a standard normal distribution but is independent of X.

14.28 Chebyshev's inequality

As mentioned earlier, a small value for the variance means that it is unlikely that a random variable X will deviate much from its expected value. To make this statement more precise we introduce the absolute value $|X - E(X)|$ which measures the actual distance between X and $E(X)$. How likely is it that this distance is more than a given amount? To answer this question we must determine the probability

$$P[|X - E(X)| > c],$$

where c is a given positive number. In the continuous case we have

$$P[|X - E(X)| > c] = 1 - P[|X - E(X)| \le c] = 1 - P[E(X) - c \le X \le E(X) + c]$$

$$= \int_{-\infty}^{+\infty} f_X(t)\, dt - \int_{E(X)-c}^{E(X)+c} f_X(t)\, dt$$

$$= \int_{-\infty}^{E(X)-c} f_X(t)\, dt + \int_{E(X)+c}^{+\infty} f_X(t)\, dt\,; \tag{14.43}$$

therefore, the calculation of this probability can be accomplished once the density function f_X is known, Of course, if f_X is unknown this method gives no information. However, if the ***variance*** is known, we can obtain an upper bound for this probability. This upper bound is provided by the following theorem of P. L. Chebyshev (1821–1894), a famous Russian mathematician who made many important contributions to probability theory and other branches of mathematics, especially the theory of numbers.

THEOREM 14.11. **CHEBYSHEV'S INEQUALITY.** *Let X be a one-dimensional random variable withjnite expectation E(X) and variance* Var (X). *Then for every positive number c we have*

$$P[|X - E(X)| > c] \leq \frac{\mathrm{Var}\,(X)}{c^2}. \tag{14.44}$$

Proof. In the continuous case we have

$$\begin{aligned}\mathrm{Var}\,(X) &= \int_{-\infty}^{+\infty} [t - E(X)]^2 f_X(t)\, dt \\ &\geq \int_{-\infty}^{E(X)-c} [t - E(X)]^2 f_X(t)\, dt + \int_{E(X)+c}^{+\infty} [t - E(X)]^2 f_X(t)\, dt \\ &\geq c^2\left(\int_{-\infty}^{E(X)-c} f_X(t)\, dt + \int_{E(X)+c}^{+\infty} f_X(t)\, dt\right).\end{aligned}$$

Because of (14.43), the coefficient of c^2 on the right is $P[|X - E(X)| > c]$. Therefore, when we divide by c^2 we obtain (14.44). This completes the proof for the continuous case; the discrete case may be similarly treated.

Chebyshev's inequality tells us that the larger we make c the smaller the probability is that $|X - E(X)| > c$. In other words, it is unlikely that X will be very far from E(X); it is even more unlikely if the variance Var (X) is small.

If we replace c by $k\sigma$, where $k > 0$ and σ denotes the standard deviation, $\sigma = \sqrt{\mathrm{Var}\,(X)}$, Chebyshev's inequality becomes

$$P[|X - E(X)| > k\sigma] \leq \frac{1}{k^2}.$$

That is, the probability that X will differ from its expected value by more than k standard deviations does not exceed $1/k^2$. For example, when $k = 10$ this inequality tells us that the probability $P[|X - E(X)| > 10\sigma]$ does not exceed 0.010. In other words, the probability is no more than 0.010 that an observed value of X will differ from the expected value by more than ten standard deviations. Similarly, when $k = 3$ we find that the probability does not exceed $\frac{1}{9} = 0.111\ldots$ that an observed value will differ from the mean by more than three standard deviations.

Chebyshev's inequality is a general theorem that applies to all distributions. In many applications the inequality can be strengthened when more information is known about the particular distribution. For example, if X has a binomial distribution with parameters n andp it can be shown (by use of the normal approximation to the binomial distribution) that for large n the probability is about 0.003 that an observed value will differ from the mean by more than three standard deviations. (For this result, $n \geq 12$ suffices.) This is much smaller than the probability 0.11 I provided by Chebyshev's inequality.

EXAMPLE. *Testing a coin for fairness.* We want to decide whether or not a particular coin is fair by tossing it 10,000 times and recording the number of heads. For a fair coin the random variable X which counts the number of heads has a binomial distribution with parameters $n = 10{,}000$ and $p = \frac{1}{2}$. The mean of X is $np = 5{,}000$ and the standard deviation is $\sigma = \sqrt{npq} = 50$. (See Example 2 in Section 14.25.) As mentioned above, the probability for a binomially distributed random variable to differ from its expected value by more than 3σ is about 0.003. Therefore, let us agree to say that a coin is ***not fair*** if the number of heads in 10,000 tosses differs from the mean by more than 30. Since $E(X) = 5{,}000$ and $3\sigma = 150$, we would say the coin is unfair if the number of heads in 10,000 tosses is less than 4,850 or more than 5,150.

14.29 Laws of large numbers

In connection with coin-tossing problems, it is often said that the probability of tossing heads with a perfectly balanced coin is $\frac{1}{2}$. This does not mean that if a coin is tossed twice it will necessarily come up heads exactly once. Nor does it mean that in 1000 tosses heads will appear exactly 500 times. Let us denote by $h(n)$ the number of heads that occur in n tosses. Experience shows that even for very large n, the ratio $h(n)/n$ is not necessarily $\frac{1}{2}$. However, experience also shows that this ratio does seem to ***approach*** $\frac{1}{2}$ as n increases, although it may oscillate considerably above and below $\frac{1}{2}$ in the process. This suggests that it might be possible to prove that

$$\lim_{n\to\infty} \frac{h(n)}{n} = \frac{1}{2}, \tag{14.45}$$

Unfortunately, this cannot be done. One difficulty is that the number $h(n)$ depends not only on n but also on the particular experiment being performed. We have no way of knowing in advance how $h(n)$ will vary from one experiment to another. But the real trouble is that it ***is*** possible (although not very likely) that in some particular experiment the ratio $h(n)/n$ may ***not*** tend to $\frac{1}{2}$ at all. For example, there is no reason to exclude the possibility of getting heads on ***every*** toss of the coin, in which case $h(n) = n$ and $h(n)/n \to 1$. Therefore, instead of trying to prove the formula in (14.45), we shall find it more reasonable (and more profitable) to ask how likely it is that $h(n)/n$ will differ from $\frac{1}{2}$ by a certain amount. In other words, given some positive number c, we seek the probability

$$P\left(\left|\frac{h(n)}{n} - \frac{1}{2}\right| > c\right).$$

By introducing a suitable random variable and using Chebyshev's inequality we can get a useful ***upper bound*** to this probability, a bound which does not require an explicit knowledge of $h(n)$. This leads to a new limit relation that serves as an appropriate substitute for (14.45).

No extra effort is required to treat the more general case of a Bernoullian sequence of trials, in which the probability of "success" is p and the probability of "failure" is q. (In coin tossing, "success" can mean "heads" and for p we may take $\frac{1}{2}$.) Let X denote the random variable which counts the number of successes in n independent trials. Then X has a binomial distribution with expectation $E(X) = np$ and variance $\operatorname{Var}(X) = npq$. Hence Chebyshev's inequality is applicable; it states that

$$P(|X - np| > c) \le \frac{npq}{c^2}. \tag{14.46}$$

Since we are interested in the ratio X/n, which ***we*** may call the ***relative frequency*** of success, we divide the inequality $|X - np| > c$ by n and rewrite (14.46) as

$$P\left(\left|\frac{X}{n} - p\right| > \frac{c}{n}\right) \leq \frac{npq}{c^2}. \tag{14.47}$$

Since this is valid for every $c > 0$, we may let c depend on n and write $c = \epsilon n$, where ϵ is a fixed positive number. Then (14.47) becomes

$$P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) \leq \frac{pq}{n\epsilon^2}.$$

The appearance of n in the denominator on the right suggests that we let $n \to \infty$. This leads to the limit formula

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) = 0 \quad \text{for every fixed } \epsilon > 0, \tag{14.48}$$

called the ***law of large numbers for the Bernoulli distribution.*** It tells us that, given any $\epsilon > 0$ (no matter how small), the probability that the relative frequency of success differs from p by more than ϵ is a function of n which tends to 0 as $n \to \infty$. This limit relation gives a mathematical justification to the assignment of the probability $\frac{1}{2}$ for tossing heads with a perfectly balanced coin.

The limit relation in (14.48) is a special case of a more general result in which the "relative frequency" X/n is replaced by the arithmetic mean of n independent random variables having the same expectation and variance. This more general theorem is usually referred to as the ***weak law of large numbers;*** it may be stated as follows:

THEOREM 14.12. WEAK LAW OF LARGE NUMBERS. *Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each having the same expectation and the same variance, say*

$$E(X_k) = m \quad \text{and} \quad \operatorname{Var}(X_k) = \sigma^2 \quad \text{for } k = 1, 2, \ldots, n.$$

Define a new random variable $\bar{X}$ (called the arithmetic mean of $X_1, X_2, \ldots, X_n$) by the equation

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Then, for every fixed $\epsilon > 0$, we have

$$\lim_{n\to\infty} P(|\bar{X} - m| > \epsilon) = 0. \tag{14.49}$$

An equivalent statement is

$$\lim_{n\to\infty} P(|\bar{X} - m| \leq \epsilon) = 1. \tag{14.50}$$