Block 3. Probability distribution functions

SOLVED EXAMPLES

Unknown discrete and continuous distributions

Example 3.1. A person goes to the headquarters of a company that has a time-shared computer system and verifies whether users are able or unable to connect to the system. During that day, the person notices that 70% of users are able to connect to the system. Which probability mass function describes this fact?

Solution.

Let X be the random variable

$$X = \begin{cases} 1 & client\ is\ able\ to\ connect \\ 0 & client\ is\ unable\ to\ connect \end{cases}$$

As 70% of users are able to connect during the day, the probability mass function for the random variable X is:

$f(0) = P(X = 0) = 0.3$ (probability that users are unable to connect to the system)

$f(1) = P(X = 1) = 0.7$ (probability that users are able to connect to the system)

$f(x) = P(X = x) = 0, \forall\ x \neq 0\ or\ 1.$

The equivalent description is

$$f(x) = \begin{cases} 0.3 & if\ x = 0 \\ 0.7 & if\ x = 1 \\ 0 & if\ x \neq 0\ or\ 1 \end{cases}$$

Example 3.2. In a communications company, there is a possibility that a bit transmitted via a digital transmission channel is transmitted with errors. Let X be the random variable that is equal to the quantity of bits with errors in the following four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. An expert calculates the probabilities associated with these values, and the results are: $P(X = 0) = 0.6561$, $P(X = 1) = 0.2916$, $P(X = 2) = 0.0486$, $P(X = 3) = 0.0036$, $P(X = 4) = 0.0001$. Assume that you want to know the **probability** that **three** or **fewer** bits are received with errors.

Solution.

You can express this question as: $P(X \leq 3)$. The event $\{X \leq 3\}$ is the union of three events: $\{X = 0\}, \{X = 1\}, \{X = 2\}$ and $\{X = 3\}$. As the events are mutually exclusive, you write:

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.9999.$

It is important to point out that the above expression is cumulative and is represented by the function $F(x) = P(X \leq 3)$. Also, you can use this function to determine the probability at a point, e.g., $P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.0036$ (value given in the statement).

Example 3.3. Let X be a discrete random variable with the corresponding probability distribution function $f(x)$ as indicated in Table 3.1:

Table 3.1.

x	-2	1	2	4
$f(x)$	$1/4$	$1/8$	$1/2$	$1/8$

Find the:

a) probability that the observed value of a random variable X is less than or equal to 1.5.
b) cumulative density function.

Solution.

In this case, the distribution function has been given. If this is not the case, you should determine it.

a) $F(1.5) = P(X \leq 1.5) = f(-2) + f(1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$.
b) To determine the cumulative density function of X, determine:

$$F(-2) = f(-2) = \frac{1}{4} = \frac{2}{8}$$

$$F(1) = f(-2) + f(1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

$$F(2) = f(-2) + f(1) + f(2) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} = \frac{7}{8}$$

$$F(4) = f(-2) + f(1) + f(2) + f(4) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} + \frac{1}{8} = 1$$

Therefore,

$$F(x) = \begin{cases} 0 & x < -2 \\ 2/8 & -2 \le x < 1 \\ 3/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

Figure 3.1 shows the cumulative density function F of X.

Notice that the cumulative density function F is a function

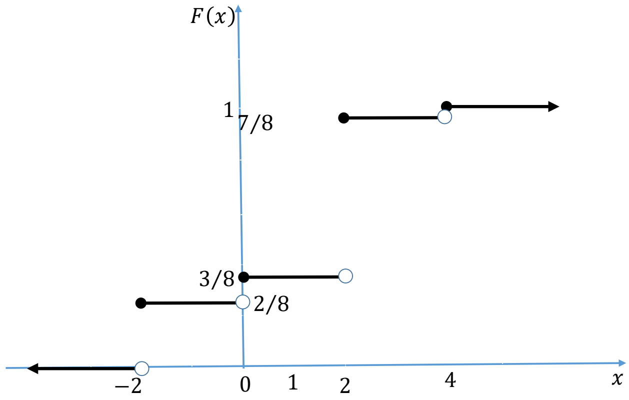

Figure 3.1. Cumulative density function

X has a step in x_i with a height of $f(x_i)$.

Example 3.4. Let $f(x)$ be the function defined in $\mathbb{R}$:

$$f(x) = \begin{cases} \frac{1}{8}x, & 0 \le x \le 4 \\ 0, & \text{in other place} \end{cases}$$

We are interested in calculating the probability, so that $P(X \le x)$, and in graphically representing the cumulative density function.

Solution.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt = \frac{x^2}{16}$$

Therefore,

$$F(x) = \begin{cases} 0, & x \le 0 \\ x^2/16, & 0 \le x \le 4 \\ 1, & x \ge 4 \end{cases}$$

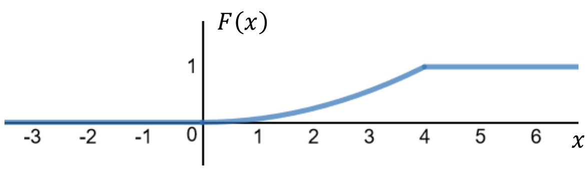

Figure 3.2. Cumulative density function.

This function is also useful for calculating the probability in $0 \le X \le 2$: $P(0 \le X \le 2) = F(2) - F(0) = \frac{1}{4} - 0 = \frac{1}{4}$.

Joint probability distributions

Example 3.1. A person goes to the headquarters of a company that has a time-shared computer system and verifies whether users are able or unable to connect to the system. During that day, the person notices that 70% of users are able to connect to the system. Which probability mass function describes this fact?

Solution.

Let X be the random variable

$$X = \begin{cases} 1 & client\ is\ able\ to\ connect \\ 0 & client\ is\ unable\ to\ connect \end{cases}$$

As 70% of users are able to connect during the day, the probability mass function for the random variable X is:

$f(0) = P(X = 0) = 0.3$ (probability that users are unable to connect to the system)

$f(1) = P(X = 1) = 0.7$ (probability that users are able to connect to the system)

$f(x) = P(X = x) = 0, \forall\ x \neq 0\ or\ 1.$

The equivalent description is

$$f(x) = \begin{cases} 0.3 & if\ x = 0 \\ 0.7 & if\ x = 1 \\ 0 & if\ x \neq 0\ or\ 1 \end{cases}$$

Example 3.2. In a communications company, there is a possibility that a bit transmitted via a digital transmission channel is transmitted with errors. Let X be the random variable that is equal to the quantity of bits with errors in the following four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. An expert calculates the probabilities associated with these values, and the results are: $P(X = 0) = 0.6561$, $P(X = 1) = 0.2916$, $P(X = 2) = 0.0486$, $P(X = 3) = 0.0036$, $P(X = 4) = 0.0001$. Assume that you want to know the **probability** that **three** or **fewer** bits are received with errors.

Solution.

You can express this question as: $P(X \leq 3)$. The event $\{X \leq 3\}$ is the union of three events: $\{X = 0\}, \{X = 1\}, \{X = 2\}$ and $\{X = 3\}$. As the events are mutually exclusive, you write:

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.9999.$

It is important to point out that the above expression is cumulative and is represented by the function $F(x) = P(X \leq 3)$. Also, you can use this function to determine the probability at a point, e.g., $P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.0036$ (value given in the statement).

Example 3.3. Let X be a discrete random variable with the corresponding probability distribution function $f(x)$ as indicated in Table 3.1:

Table 3.1.

x	-2	1	2	4
$f(x)$	$1/4$	$1/8$	$1/2$	$1/8$

Find the:

c) probability that the observed value of a random variable X is less than or equal to 1.5.
d) cumulative density function.

Solution.

In this case, the distribution function has been given. If this is not the case, you should determine it.

c) $F(1.5) = P(X \leq 1.5) = f(-2) + f(1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$.
d) To determine the cumulative density function of X, determine:

$$F(-2) = f(-2) = \frac{1}{4} = \frac{2}{8}$$

$$F(1) = f(-2) + f(1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

$$F(2) = f(-2) + f(1) + f(2) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} = \frac{7}{8}$$

$$F(4) = f(-2) + f(1) + f(2) + f(4) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} + \frac{1}{8} = 1$$

Therefore,

$$F(x) = \begin{cases} 0 & x < -2 \\ 2/8 & -2 \le x < 1 \\ 3/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

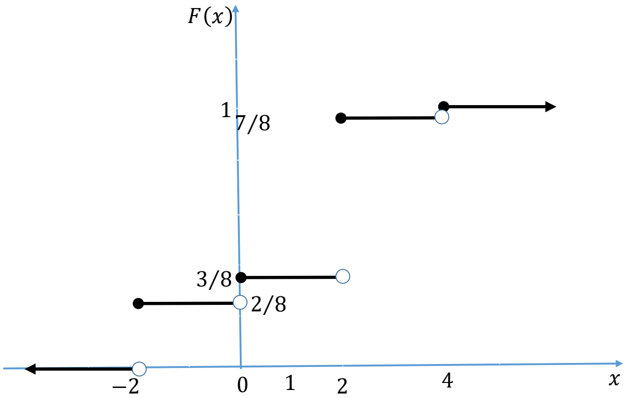

Figure 3.1. Cumulative density function

Figure 3.1 shows the cumulative density function F of X.

Notice that the cumulative density function F is a function X has a step in x_i with a height of $f(x_i)$.

Example 3.4. Let $f(x)$ be the function defined in $\mathbb{R}$:

$$f(x) = \begin{cases} \frac{1}{8}x, & 0 \le x \le 4 \\ 0, & \text{in other place} \end{cases}$$

We are interested in calculating the probability, so that $P(X \le x)$, and in graphically representing the cumulative density function.

Solution.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt = \frac{x^2}{16}$$

Therefore,

$$F(x) = \begin{cases} 0, & x \le 0 \\ x^2/16, & 0 \le x \le 4 \\ 1, & x \ge 4 \end{cases}$$

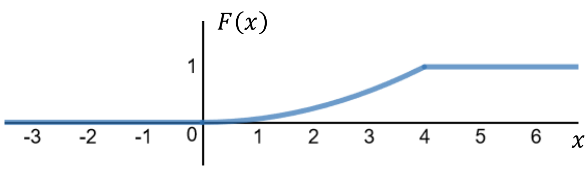

Figure 3.2. Cumulative density function.

This function is also useful for calculating the probability in $0 \le X \le 2$: $P(0 \le X \le 2) = F(2) - F(0) = \frac{1}{4} - 0 = \frac{1}{4}$.

Example 3.8. The shelf life, in years, of a preserved product packed in plastic containers is a random variable whose probability density function is

$$f(x)=\begin{cases} e^{-2\,x}, & x>0 \\ 0, & in\ any\ other\ process \end{cases}.$$

If X_1, X_2 and X_3 are the shelf lives of three of these plastic packages chosen independently: find the value of $P(X_1<1,\ \ 1<X_2<2,\ \ \ X_3>1\ \ )$.

Solution

Because the packages are chosen independently, assume that the random variables X_1, X_2 and X_3 are statistically independent and have a joint probability density

$$f(x_1,x_2,x_3)=f(x_1)\,f(\,x_2)\,f(\,x_3)=\ e^{-2x_1}e^{-2x_2}e^{-2x_3}=e^{-2(x_1+x_2+x_3)},$$

for

$x_1>0,\ \ x_2>0,\ \ x_3>0$, and $f(x_1,x_2,x_3)=0$ in any other case. Therefore,

$$P(X_1<1,\ \ 1<X_2<2,\ \ \ X_3>1\ \ )=\int_1^{\infty}\int_1^2\int_0^1 e^{-2(x_1+x_2+x_3)}\ dx_1\,dx_2\,dx_3=$$

$$=\frac{-1}{2^3}(e^{-2}-1)\,(e^{-4}-e^{-2})\,(-e^{-2})=\frac{1}{2^3}\,(1-e^{-2})^2e^{-4}\cong 1.7\times 10^{-3}$$

Mathematical expectation

Example 3.9. You toss a coin three times. Find the expected value or mathematical hope of getting the maximum number of heads in a row if:

a. the coin is not rigged.
b. the coin is rigged.

Solution

a.

If the coin is not rigged, the space S, which has 8 elements, is equiprobable. The results of the random experiment belong to a discrete space and the random variable is discrete. The values of $f(x_k)$ that correspond to each value of x (Table 3.6):

x	0	1	2	3
$f(x)$	$1/8$	$4/8$	$2/8$	$1/8$

Table 3.6

Example 5.4 shows the distribution of X. With this distribution, you get:

$$E = E(X) = 0 \times \frac{1}{8} + 1 \times \frac{4}{8} + 2 \times \frac{2}{8} + 3 \times \frac{1}{8} = \frac{11}{8}$$

b.

If the weight of the coin is such that: $P(h) = 2/3$ and $P(t) = 1/3.$, then S is not equiprobable.

Therefore, Figure 3.7 shows the distribution f of X:

x	0	1	2	3
$f(x)$	$1/27$	$10/27$	$8/27$	$8/27$

Table 3.7

the weight of the coin is such that:

$$E = E(X) = 0 \times \frac{1}{27} + 1 \times \frac{10}{27} + 2 \times \frac{8}{27} + 3 \times \frac{8}{27} = \frac{50}{27}$$

In both cases, obtaining a greater number of heads in a row is a favourable result, because it is a positive quantity.

Example 3.10. Let X be the random variable denoting the life time of an infrared sensor, in hours. If you write the probability density function as

$$f(x) = \begin{cases} \dfrac{162000}{x^4}, & x > 3 \\ 0, & in\ any\ other\ case \end{cases},$$

What is the expected service life of the sensor?

Solution.

The results of the random experiment belong to a continuous space and the random variable is continuous; therefore, $\mu = E(X) = \int_3^{\infty} x \, \frac{162000}{x^4} dx = 9000.$

You can expect the device to have an average life of 9,000 hours. From a mathematical point of view, this "seems" to be a favourable result, and from a practical point of view it could be favourable or not, as there may be other devices that perform better.

Example 3.11. A shop opens at 9 am, because the owner thinks that they will make a profit on the sales of cartons of milk between the opening time and 10 am. An inspector wants to find out if the shopkeeper does actually make a profit between 9 am and 10 am. To do this, he associates the random variable X with the number of cartons of milk sold in that period, and presents the following probability distribution (Table 3.8):

x	10	11	12	13	14
$P(X = x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

Table 3.8.

If the money (euros) the shopkeeper receives from the sales of cartons of milk can be represented by the random variable $g(X) = X + 2$, do you expect that he will make a profit over that period?

Solution.

The expected value of what the shopkeeper can take is:

$E[g(X)] = E(X+2) = \sum_{x=10}^{14}(x+2)\, f(x)$

$$= 12 \times \frac{1}{8} + 13 \times \frac{1}{4} + 14 \times \frac{1}{8} + 15 \times \frac{1}{4} + 16 \times \frac{1}{4} = €14.25$$

As the value is positive, you can expect the shopkeeper to earn, on average, €14.25.

Example 3.12. For a random experiment, you define a random variable X, whose probability density function is given by

$$f(x) = \begin{cases} \frac{e-1}{e}\, e^{-x}, & 0 < x < 1 \\ 0, & in\ any\ other\ case \end{cases}.$$

Find the expected value of $g(X) = X - 1$.

Solution.

The expected value is:

$E[g(X)] = E(X-1) = \frac{e-1}{e}\int_0^1 (x-1)e^{-x}\, dx = \frac{e-1}{e}\left[\int_0^1 xe^{-x}\, dx - \int_0^1 e^{-x}\, dx\right].$

You determine the first integral by parts and its value is $1 - \frac{2}{e}$. The result of the second integral is $1 - \frac{1}{e}$. If you replace this in the previous expression, you get:

$$E[g(X)] = \frac{e-1}{e}\left(2 - \frac{3}{e}\right) \cong 0.57$$

The expected value is positive. Therefore, according to the nature of the random experiment, you should expect a favourable result. However, the value you obtained is near zero. This expresses the fact that there does not “seem” to be a very strong tendency towards a favourable result.

Example 3.13. In an experiment, you define two random variables, X and Y, with a joint density function:

$$f(x,y) = \begin{cases} 2xye^{-x}, & 0 < x < \infty, \quad 0 < y < 1 \\ 0, & in\ any\ other\ case \end{cases}.$$

Determine $E\left(\frac{Y}{X}\right)$.

Solution.

$$E[g(X,Y)] = E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^\infty \frac{y}{x} 2xye^{-x} dx\, dy = 2\int_0^1 y^2 dy = \frac{2}{3}$$

The mathematical hope is positive and you can expect a favourable result. However, the proximity to zero indicates that there does not "seem" to be a very strong tendency towards a favourable result.

Mean and variances of linear combinations of random variables

Example 3.14. In example 3.9, find the variance and the standard deviation of section a.

Solution.

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \left(0-\tfrac{11}{8}\right)^2 \times \tfrac{1}{8} + \left(1-\tfrac{11}{8}\right)^2 \times \tfrac{4}{8} + \left(2-\tfrac{11}{8}\right)^2 \times \tfrac{2}{8} + \left(3-\tfrac{11}{8}\right)^2 \times \tfrac{1}{8}$$

$$\sigma^2 = \frac{376}{512} \cong 0.7344$$

Also, according to theorem 5.8, you can calculate the variance using the expression: $\sigma^2 = E(X^2) - \mu^2$.

In this case, $\mu = \frac{11}{8}$ and

$$E(X^2) = \sum_x x^2 f(x) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{4}{8} + 2^2 \times \frac{2}{8} + 3^2 \times \frac{1}{8} = \frac{21}{8}$$

$$\sigma^2 = \frac{21}{8} - \left(\frac{11}{8}\right)^2 = \frac{47}{64} \cong 0.7344$$

The standard deviation is $\sigma = \sqrt{\sigma^2} = \sqrt{0.7344} \cong 0.8570$.

Example 3.15. The weekly sales of sand from one company to various construction companies are random. The probability density function of the sales, in tons, is given by:

$$f(x) = \begin{cases} \frac{3}{5}(2-x^2), & 0 \le x \le 1 \\ 0, & in\ other\ cases \end{cases}$$

Find the variability in tons with respect to the mean sales.

Solution.

Notice that the required information is obtained through the meaning of the concept. Moreover, the random variable has not been defined. However, you can assume that it has been, because a probability density function has been assigned. For this reason, we can say: let X be the random variable that represents the weekly sales of sand. As the results of the random experiment belong to a continuous sample space, the random variable is continuous, and according to theorem 5.8: $\sigma^2 = E(X^2) - \mu^2$.

The probability density function takes values other than zero: $\forall\ 0 \leq x \leq 1$. The mean is calculated with:

$$\mu = \int_0^1 x\, f(x)\, dx = \int_0^1 x\, \frac{3}{5}(2 - x^2)\, dx = \frac{3}{5}\left[\int_0^1 2x\, dx - \int_0^1 x^2 dx\right] = \frac{9}{20}$$

$$E(X^2) = \int_0^1 x^2\, \frac{3}{5}(2 - x^2)dx = \frac{3}{5}\left[\int_0^1 2x^2 dx - \int_0^1 x^4\, dx\right] = \frac{7}{25}$$

Therefore, $\sigma^2 = \frac{7}{25} - \left(\frac{9}{20}\right)^2 = \frac{31}{400}$. This value represents the variance, but it is a quadratic variability. To express the variability, in tons, determine the standard deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{31}{400}} \cong 0.278\ tons.$

Example 3.16. Find the variance of the random variable $g(X) = X - 1$ from Example 3.12.

Solution.

$$\sigma_{g(X)}^2 = \int_0^1 [x - 1\ - (0.57)]^2 \frac{e-1}{e}\ e^{-x}\, dx = \frac{e-1}{e} \int_0^1 [x - 0.43]^2\, e^{-x}\, dx =$$

$$\sigma_{g(X)}^2 = \frac{e-1}{e}\left[\int_0^1 x^2\, e^{-x}\, dx - 0.86 \int_0^1 x\, e^{-x}\, dx + 0.43^2 \int_0^1\ e^{-x}\, dx\right]$$

Calculate the first integral by parts. Its value can be expressed as follows:

$$\int_0^1 x^2\, e^{-x}\, dx = -\frac{1}{e} + 2\int_0^1 x\, e^{-x}\, dx$$

When you replace this expression in the variance, you get:

$$\sigma_{g(X)}^2 = \frac{e-1}{e}\left[-\frac{1}{e} + 1.14 \int_0^1 x\, e^{-x}\, dx + 0.43^2 \int_0^1\ e^{-x}\, dx\right]$$

You already calculated the second and third integrals in Example 3.12. Replace their results in the previous expression and $\sigma_{g(X)}^2 \cong 0.032$. Therefore, $\sigma = \sqrt{\sigma_{g(X)}^2} = \sqrt{0.032} \cong 0.18$.

If you observe that the mean of the random experiment is 0.57 and its standard deviation is 0.18; then, the result of Example 3.12 can be considered favourable for the random experiment, because the mean and its variability towards zero are above this value.

Example 3.16.bis Find the mathematical hope for the random variable $g(X) = (X+2)^2$ from Example 3.11.
Solution.

$E[g(X)] = E[(X+2)^2] = E(X^2 + 2X + 4) = E(X^2) + E(2X) + E(4)$, where

$$E(X^2) = \sum_{x=10}^{14} x^2 f(x) = 100 \times \frac{1}{8} + 121 \times \frac{1}{4} + 144 \times \frac{1}{8} + 169 \times \frac{1}{4} + 196 \times \frac{1}{4} = 152$$

$$E(2X) = 2E(X) = 2 \sum_{x=10}^{14} x\, f(x) = 2\left(10 \times \frac{1}{8} + 11 \times \frac{1}{4} + 12 \times \frac{1}{8} + 13 \times \frac{1}{4} + 14 \times \frac{1}{4}\right) = \frac{49}{2}$$

Therefore,

$$E(X^2 + 2X + 4) = 152 + \frac{49}{2} + 4 = 180.5$$

x	10	11	12	13	14
$P(X=x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

Table 3.8.

Known distributions

Example 3.17. You toss a coin 6 times. Find:

a. The probability of getting only two heads.
b. The probability of getting at least four heads.
c. The probability of not getting any heads.
d. The probability of getting one or more heads.
e. The mean, the variance, and the standard deviation from the random variable X.

Solution.

Tossing the coin complies with the properties of Bernoulli's process. Therefore, the random variable X has a corresponding Bernoulli's probability mass function. In this process: the fixed number of tests is $n = 6$, the p probabilities of success and q probabilities of failure are **heads** and **tails**, respectively. Moreover, it is true that: $p = q =$ ½.

a. $P(x = 2\ successes) = \binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4 = \frac{15}{64} \cong 0.23$.

b. $P(at\ least\ 4\ successes) = P(4\ successes) + P(5\ successes) + P(6\ successes) =$

$$= \binom{6}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2 + \binom{6}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^1 + \binom{6}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^0 = \frac{11}{34} \approx 0.32.$$

c. $P(x = 0\ successes) = \binom{6}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^6 = \left(\frac{1}{2}\right)^6 = q^6 = \frac{1}{64}.$

d. $P(x = 1\ or\ more\ successes) = 1 - q^6 = 1 - \frac{1}{64} = \frac{63}{64} \approx 0.98$.

e. As the random variable X has a known binomial distribution, you can calculate the mean and the variance directly using the parameters X:

$\mu = np = 6 \times \frac{1}{2} = 3$ (mean or expected number of successes), $\sigma^2 = npq = 6 \times \frac{1}{2} \times \frac{1}{2} = \frac{3}{2} =$ 1.5 (variance), and $\sigma = \sqrt{npq} = \sqrt{\frac{3}{2}} \approx 1.22$ (standard deviation).

Example 3.18. You roll a die eight times. Find the probability of getting 5 and 6 twice each, and all the other numbers once each.

Solution.

$$p = f\left(2,2,1,1,1,1;\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},8\right) = \frac{8!}{2!\,2!\,1!\,1!\,1!\,1!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{35}{5832} \approx 0.006$$

The probability you get is 0.6%.

Example 3.19. In a quality control exercise, several boxes containing 40 items each are considered to be acceptable if they contain no more than three defective items. In the sampling, you take five items from a box at random, and the box is rejected if you find one defective item.

a. What is the probability that you find **exactly** one defective item in the sample if there are three defective items in the whole box?
b. Find the mean and the variance of the random variable.
c. Use Chebyshev's theorem to interpret the range $\mu \pm 2\sigma$.

Solution.

The conditions of the **hypergeometric experiment** are met: the number of items is finite, when you choose an item it may represent a success or a failure, and there is a determinate number of successes within the items selected. Therefore, you can consider the random variable with a probability mass function as being hypergeometric.

a. The hypergeometric distribution with the parameters $n = 5$, $N = 40$, $M = 3$, and $x = 1$ allows you to find the probability of a defective item:

$$P(X = 1) = h(1;5,3,40) = \frac{\binom{3}{1}\binom{40-3}{5-1}}{\binom{40}{5}} = 0.3011$$

b. From theorem 6.4, you get:

$$\mu = \frac{5\cdot 3}{40} = 0.375 \text{ and } \sigma^2 = \frac{40-5}{39}\cdot 5\cdot\frac{3}{40}\cdot\left(1-\frac{3}{40}\right) = 0.3113.$$

c. $\sigma = 0.558$. Then, $\mu \pm 2\sigma = 0.375 \pm 0.3113$. Therefore, you get the range $[-0.741, 1{,}491]$. The number of defective items has a probability of at least $3/4$ of falling within the indicated range.

Example 3.20. A communications company conducts a study and determines that the probability that a bit transmitted via a digital transmission channel is received with errors is 0.1. What is the probability of transmission until you receive the first bit with errors? What are its mean and variance?

Solution.

Assume that the bit with errors appears in the fifth test. The whole study meets the conditions of a geometric experiment. The random variable X denotes the number of bits transmitted until the first bit with errors appears. Then, $P(X = 5)$ is the probability that the first four bits are correctly transmitted and the last one has errors.

In this situation, you can represent the event as $\{ffffe\}$, where f (bit without errors) and e (bit with errors) are success and failure, respectively. Taking into account the fact that the tests are independent, that $P(X = x) = pq^{x-1}$, and that the probability of correct transmission is 0.9, $P(X = 5) = P(ffffe) = 0.9^4 \cdot 0.1 = 0.066$.

According to theorem 6.5, the mean and the variance are: $\mu = \frac{1}{0.1} = 10$ and $\sigma^2 = \frac{1-0.1}{0.1^2} = 90$, respectively.

Example 3.21. You toss three coins. Find:

a. the probability of getting **only** heads or **only** tails for the second time on the fifth toss.
b. the mean and the variance of the random variable.

Solution.

In this experiment, tossing the three coins meets the characteristics of a negative binomial experiment. Therefore, you can work with the negative binomial probability mass function.

a. The negative binomial distribution with $x = 5$, $r = 2$, and $p = \frac{1}{4}$ lets you find the probability of getting **only** heads or **only** tails for the second time on the fifth toss:

$$nb\left(5; 2, \frac{1}{4}\right) = \binom{5+2-1}{2-1}\left(\frac{1}{4}\right)^2\left(1-\frac{1}{4}\right)^3 = \frac{27}{256}.$$

b. $\mu = \frac{2\left(1-\frac{1}{4}\right)}{\frac{1}{4}} = 6$ and $\sigma^2 = \frac{2\left(1-\frac{1}{4}\right)}{\left(\frac{1}{4}\right)^2} = 24.$

Example 3.22. In a 220-page book, there are 200 misprints, which are randomly distributed throughout the book. Find the probability that any given page has:

a. No misprints.
b. One misprint.
c. Two misprints.
d. 2 or more misprints.

Solution.

Successes: number of misprints in Bernoulli experiments: $n = 220$.

The probability that a given page has a misprint is $p = \frac{1}{200}$.

As p is low, we use the Poisson approximation to the binomial distribution where $\lambda\, t = n\, p = 1.1$:

$$P(X = x) = p(x; \lambda\, t) = \frac{(\lambda\, t)^x e^{-\lambda\, t}}{x!}$$

a. $P(0) = p(0; 1.1) = \frac{(1.1)^0 e^{-1.1}}{0!} = 0.333$
b. $P(1) = p(1; 1.1) = \frac{(1.1)^1 e^{-1.1}}{1!} = 0.366$
c. $P(2) = p(2; 1.1) = \frac{(1.1)^2 e^{-1.1}}{2!} = 0.201$
d. $P(X \geq 2) = 1 - P(0) - P(1) = 1 - 0.333 - 0.366 = 0.301$

Example 3.23. Let X be a random variable with $N\ (70, 4)$. Find:

a. $P(68 \leq X \leq 74)$.

b. $P(72 \leq X \leq 75)$.

c. $P(63 \leq X \leq 68)$

d. $P(X \geq 73)$.

Solution.

The random variable X is not typified because $N(\mu, \sigma^2) = N\,(70, 4)$. Therefore, you have to typify the random variable X before you calculate the probabilities.

a. $P(68 \leq X \leq 74) = P\left(\frac{68-70}{2} \leq \frac{X-\mu}{\sigma} \leq \frac{74-70}{2}\right) =$

$= P(-1 \leq Z \leq 2) = \Phi(2) - \Phi(-1) = 0.9772 - 0.1587 = 0.8185.$

b. $P(72 \leq X \leq 75) = P\left(\frac{72-70}{2} \leq \frac{X-\mu}{\sigma} \leq \frac{75-70}{2}\right) =$

$= P(1 \leq Z \leq 2.5) = \Phi(2.5) - \Phi(1) = 0.9938 - 0.8413 = 0.1525.$

c. $P(63 \leq X \leq 68) = P\left(\frac{63-70}{2} \leq \frac{X-\mu}{\sigma} \leq \frac{68-70}{2}\right) =$

$= P(-3.5 \leq Z \leq 1) = \Phi(1) - \Phi(-3.5) = 0.8413 - 0.0001 = 0.8412.$

d. $P(X \geq 73) = P\left(\frac{X-\mu}{\sigma} \geq \frac{73-70}{2}\right) =$

$= P(Z \geq 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668.$

Example 3.24. In an experiment, you toss a coin 144 times. Find the probability of getting tails:

a. 80 times.
b. Between 58 and 63 times, inclusive.

Less than 65 times.

Solution.

The experiment complies with the characteristics of a Bernoulli experiment. However, the number of tests is high, according to the experiment conducted. Calculating the probabilities with the binomial is tedious, so you try to approximate this using a normal distribution. Therefore, you look for the parameters of the distribution and verify whether you can make the approximation:

$b(x; n, p) = B(x; 144, 0.5)$ and $q = 1 - p = 0.5$. Therefore, $\mu = np = 144 \cdot 0.5 = 72$ and $\sigma^2 = npq = 144 \cdot 0.5 \cdot 0.5 = 36; \sigma = 6.$

$np = 72 > 5$ and $nq = 72 > 5$. Then, to calculate the probabilities, you can make an approximation using the normal distribution, i.e., $b(x;n,p) \approx N(\mu, \sigma^2)$, where $b(80;144{,}0.5) \approx N(72.36)$.

a. $Pb(k) \approx PN(k-0.5 \le X \le k+0.5)$. In this case, $Pb(80) \approx PN(80-0.5 \le X \le 80+0.5) = PN(79.5 \le X \le 80.5)$.

The random variable is not typified and you must typify it: $Pb(80) \approx PN(79.5 \le X \le 80.5) = PN\left(\frac{79.5-72}{6} \le \frac{X-\mu}{\sigma} \le \frac{80.5-72}{6}\right) = PN(1.25 \le Z \le 1.42) = \Phi(1.42) - \Phi(1.25) = 0.9222 - 0.8944 = 0.0278$.

The calculation $b(80;144{,}0.5) = \binom{144}{80}0.5^{80}0.5^{64} = 0.0274$. Notice that the approximation is good.

b. $Pb(58 \le k \le 63) \approx PN(58-0.5 \le X \le 63+0.5) = PN(57.5 \le X \le 63.5) = PN\left(\frac{57.5-72}{6} \le \frac{X-\mu}{\sigma} \le \frac{63.5-72}{6}\right) = PN(-2.42 \le Z \le -1.42) = \Phi(-1.42) - \Phi(-2.42) = 0.7878 - 0.0078 = 0.7800$.

$Pb(k < 65) = Pb(k \le 64) \approx PN(X \le 64+0.5) = PN(X \le 64.5) = PN\left(\frac{X-\mu}{\sigma} \le \frac{64.5-72}{6}\right) = PN(Z \le -1.25) = \Phi(-1.25) = 0.1056$.

Example 3.25. An expert records the reaction times of a randomly chosen businessman when he reacts to variations in the stock market. He notices that the data distribution is skewed and, when he compares this to a graph, he considers that it corresponds to a standard gamma distribution where the parameter $\alpha = 2$. Find the probability that the reaction time of the businessman:

e) Is between 3 and 5 s.
f) Is greater than 4 s.

Solution.

Let X be the random variable describing the reaction times. It is, therefore, a continuous random variable. In this case, the expert observes the distribution of points and, when comparing these to a known graph, he establishes that it is a skewed function of the standard gamma type. So,

a. As X is continuous, $P(a \leq X \leq b) = F(b) - F(a)$. Then, $P(3 \leq X \leq 5) = F(5;2) - F(3;2)$. Using table A-6, find the values of $F(x;\alpha)$. Therefore, $P(3 \leq X \leq 5) = 0.960 - 0.801 = 0.159$.
b. $P(X > 4) = 1 - P(X \leq 4) = 1 - F(4;2) = 1 - 0.908 = 0.092$.

Example 3.26. An electrical company records daily energy use in a city. The collected data are random and, when compared to known function graphs, we can say that they approximate to a gamma distribution with the parameters $\alpha = 3$ and $\beta = 2$. If the company has a daily capacity of 10 kW/h:

a. What is the probability that this supply is not enough on any given day?
b. What is the probability that, with this supply, the energy use is between 4 and 8 million kW/h?
c. Find the mean energy use and the deviation from this.

Solution.

Let X be the random variable representing the daily energy use in the city. To calculate the probabilities, you must take into account the fact that the data distribution corresponds to a gamma function.

a. $P(X > 10) = 1 - P(X \leq 10)$. According to the gamma probability density function, and considering that $P(X \leq x) = \int_0^x \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx$; then, $P(X > 10) = 1 - \frac{1}{\beta^3}\int_0^{10} \frac{x^2 e^{-\frac{x}{\beta}}}{\Gamma(3)} dx$. You can solve the integral with the incomplete gamma function (cumulative density function) by substituting $y = \frac{x}{\beta}$: $P(X > 10) = 1 - \int_0^5 \frac{y^2 e^{-y}}{\Gamma(3)} dy = 1 - F(5;3)$. Use a table to determine the cumulative density function. Therefore, $P(X > 10) = 1 - 0.875 = 0.125$.
b. $P(3 < X < 8) = \frac{1}{\beta^3}\int_4^8 \frac{x^2 e^{-\frac{x}{\beta}}}{\Gamma(3)} dx = \int_2^4 \frac{y^2 e^{-y}}{\Gamma(3)} dy = F(4;3) - F(2;3) = 0.762 - 0.323 = 0.439$.
c. The value of the mean energy use is $E(X) = \alpha\beta = 3 \times 2 = 6$. The variance is $V(X) = \alpha\beta^2 = 3 \times 2^2 = 12$ and its deviation is $\sigma = 3.46$. The units of both are $10^6 kW/h$.

Example 3.27. The fire emergency phone number receives calls according to a Poisson process, with a rate of $\lambda = 0.2$ calls per day. Find the probability that more than three days pass between two calls. What is the expected time between consecutive calls?

Solution.

Due to the relationship that the exponential distribution has with the Poisson distribution, you can consider that the distribution of the time elapsed between two successive events is exponential with the parameter λ (we do not try to prove this here). Therefore, you can consider that the number of days X between successive calls has an exponential distribution where $\lambda = 0.2$.

The probability that more than three days elapse is $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3; 0.2)$. According to the last observation: $P(X > 3) = 1 - F(3; 0.2) = e^{-0.2\times 3} = 0.549$.

The expected time between successive calls is: $\mu = \alpha\beta = \frac{1}{\lambda} = \frac{1}{0.2} = 5\ days$.

Example 3.28. A company performs programmed obsolescence tests and establishes that the failure times in batches of their items are calculated in years. The distribution of the collected data is similar to an exponential function with a mean failure time of $\beta = 4$. After ten years have elapsed, if they sell seven of these items from different batches, what is the probability that at least two of them do not fail?

Solution.

The items may fail or not (dichotomy). Let X be the random variable of the number of items that do not fail after ten years. We associate this random variable with a Bernoulli experiment. Using the binomial distribution, you determine the probability that at least two items do not fail with $P(X \geq 2) = \sum_{x=2}^{7} b(x; 7, p) = 1 - \sum_{x=0}^{1} b(x; 7, p)$, where p is the probability of success (no failure).

To determine p, you consider the random variable T, which is the mean time for failures, $\beta = 4$. If the data are well adjusted to an exponential function, and taking into account the fact that $\lambda = \frac{1}{\beta}$, the probability that an item does not fail after ten years is $P(T > 10) = \frac{1}{4}\int_{10}^{\infty} e^{-\frac{t}{4}}\,dt \approx 0.082 \approx 0.1$.

Now, $p = 0.1$ and the binomials are determined by calculating them through their expression or using a table or a computer program. Then, the probability that at least two items do not fail is

$$P(X \geq 2) = 1 - \sum_{x=0}^{1} b(x; 7, 0.1) = 0.106 \approx 0.1.$$

Example 3.29. A switchboard receives an average of five phone calls per minute. What is the probability that it will take more than a minute for two calls to arrive?

Solution.

The Poisson process is applied to the time that passes until the occurrence of two Poisson events that follow a gamma distribution where $\beta = 1/5$ and $\alpha = 2$. Let X be the random variable representing the time, in minutes, that elapses before two calls arrive. So,

$$P(X \leq x) = \int_0^x \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \, dx$$

$$P(X \leq 1) = 25 \int_0^1 x e^{-5x} \, dx \approx 0.96$$

Whereas the origin of the gamma distribution deals with the time (space) until the occurrence of α Poisson events, there are cases where the distribution fits well, even though there is no clear Poisson-type structure. This is observed in problems of survival time.

Example 3.30. In the study of how the dose of a toxin affects the survival time of rats, it was determined that this survival time, in weeks, has a gamma distribution where $\alpha = 5$ and $\beta = 10$. What is the probability that a rat does not survive longer than 60 weeks?

Solution

Let the random variable X be the survival time. The probability is

$$P(X \leq x) = \int_{0}^{\infty} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}\, dx$$

$$P(X \leq 60) = \frac{1}{\beta^{5}} \int_{0}^{60} \frac{1}{\Gamma(5)} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx$$

With the change $y = \frac{x}{\beta}$, you get the incomplete gamma integral, whose value you find in a statistical table:

$$P(X \leq 60) = \frac{1}{\beta^{5}} \int_{0}^{60} \frac{1}{\Gamma(5)} x^{4} e^{-y}\, dx = F(6; 5) = 0.715.$$