Instructor: **Paolo Zacchia / Yaroslav Korobka**[1] Updated: October 27, 2025
Course: **Statistics**

1 Basic probability theory

Exercise 1.1. For each of the following experiments, describe the sample space.

(a) Toss a coin four times.

(b) Count the number of insect-damaged leaves on a plant.

(c) Measure the lifetime (in hours) of a particular brand of light bulb.

(d) Record the weights of 10-day-old rats.

(e) Observe the proportion of defectives in a shipment of electronic components.

Solution. The sample space S is the set of all possible outcomes.

(a) For each coin toss in the series we receive either Head (H) or Tail (T). The example of the realization of the series of four tosses: THHH, meaning 1st toss gives T and the remaining tosses give H. The sample space S is the set of all such combinations. One may want to calculate the cardinality of this set. It is $2^4 = 16$: on each position (toss) we have either T or H (2 options) and we observe four positions.

(b) The number of leaves should be a nonnegative integer. Thus, $S = \{0, 1, 2, 3, \ldots\}$.

(c) The lifetime may be less than an hour or some integer number of hours plus the remaining fraction of an hour. Moreover, the lifetime cannot be negative. Thus, $S = \{0, 1, 2, 3, \ldots\}$ in case the lifetime is measured in hours, and $S = \{t : t \geq 0\}$ in case the lifetime is measured in infinitesimal units of time.

(d) We need to choose measurement units. Consider grams, then we may say $S = (0, +\infty)$. We may also define some upper bound, for example 1 kilogram. Then $S = (0, 1000]$.

(e) One can define n to be the number of items in the shipment. Since we need a proportion, the sample space is $S = \{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n}{n}\}$.

Exercise 1.2. Verify the following identities:

(a) $A \setminus B = A \setminus (A \cap B) = A \cap B^c$,

(b) $B = (B \cap A) \cup (B \cap A^c)$,

(c) $B \setminus A = B \cap A^c$,

(d) $A \cup B = A \cup (B \cap A^c)$.

[1]The author thanks Arsenii Scherbov for providing materials for the first half of the course.

Solution. When working with set expressions one may use Venn diagrams to understand how to proceed. However, the illustration does not constitute a formal proof and thus is not sufficient as an answer. To proof that the expression is true one must explicitly show both directions: if $x \in$ LHS (left-hand side) then $x \in$ RHS (right-hand side) and if $x \in$ RHS then $x \in$ LHS.

(a) $A \setminus B \Longleftrightarrow x \in A$ and $x \notin B \Longleftrightarrow x \in A$ and $x \notin A \cap B \Longleftrightarrow x \in A \setminus (A \cap B)$. At the same time, $x \in A$ and $x \notin B \Longleftrightarrow x \in A$ and $x \in B^c \Longleftrightarrow x \in A \cap B^c$.

(b) By definition $A \cup A^c = S$. Then using the Distributive Law $(B \cap A) \cup (B \cap A^c) = B \cap (A \cup A^c) = B \cap S = B$. One may also show the equality by the technique used in point (a).

(c) $B \setminus A \Longleftrightarrow x \in B$ and $x \notin A \Longleftrightarrow x \in B$ and $x \in A^c \Longleftrightarrow x \in B \cap A^c$.

(d) $A \cup B = A \cup [(B \cap A) \cup (B \cap A^c)] = A \cup (B \cap A) \cup A \cup (B \cap A^c) = A \cup [A \cup (B \cap A^c)] = A \cup (B \cap A^c)$.

Exercise 1.3. For events A and B, find formulas for the probabilities of the following events in terms of the quantities $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(A \cap B)$:

(a) either A or B or both,

(b) either A or B but not both,

(c) at least one of A or B,

(d) at most one of A or B.

Solution.

(a) "either A or B or both" means $A \cup B$, so we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

This is because $A \cup B = A \cup (B \cup A^c)$ (Exercise 1.2. point (d)), from it $\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \cup A^c)) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

In the last expression, the second equality uses the fact that A and $B \cap A^c$ are disjoint; the last equality comes from rearranging the following $\mathbb{P}(B) = \mathbb{P}((B \cap A) \cup (B \cap A^c)) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$ (combines Exercise 1.2. point (b) and the fact that two sets are disjoint).

It is easier to draw the picture and note the fact that we count the area of the intersection two times if we do not subtract $\mathbb{P}(A \cap B)$.

(b) "either A or B but not both" is $(A \cap B^c) \cup (B \cap A^c)$, so we have

$$\begin{aligned}\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(B \cap A^c) \\ &= [\mathbb{P}(A) - \mathbb{P}(A \cap B)] + [\mathbb{P}(B) - \mathbb{P}(B \cap A)] \\ &= \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).\end{aligned}$$

The easy way is to draw a picture and note that this probability is equal to $\mathbb{P}(A \cup B) - \mathbb{P}(A \cap B)$.

(c) "at least one of A or B" is $A \cup B$, so we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(d) "at most one of A or B" is $(A \cap B)^c$, so we have

$$\mathbb{P}((A \cap B)^c) = 1 - \mathbb{P}(A \cap B).$$

Note that this event also includes the possibility of not A and not B.

Exercise 1.4. Consider two different setups:

(a) A fair dice is cast until a 6 appears. What is the probability that it must be cast more than five times?

(b) Prove that if $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then:

- if A and B are mutually exclusive, they cannot be independent,
- if A and B are independent, they cannot be mutually exclusive.

Solution.

(a) It must be cast more than five times if we did not observe the appearance of 6 on first five casts. We do not observe 6 on each cast with probability $5/6$. Since casts of a die are independent (adequate additional interpretation of the setup which you should explicitly mention in your solution) the probability to not receive 6 in first five casts is $(5/6)^5 \approx 0.4$.

(b)
- Let A and B be mutually exclusive. Suppose, for the sake of contradiction, that A and B are independent, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. However, from the initial property $A \cap B = \emptyset$ and $\mathbb{P}(A \cap B) = 0$ (mutually exclusive). Since we are given events A and B with positive probabilities, we come to a contradiction $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = 0$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Thus, mutually exclusive events cannot be independent.
- Independence of A and B together with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$ gives $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) > 0$. It means that $A \cap B \neq \emptyset$. Therefore, A and B are not mutually exclusive by definition. Thus, independent events cannot be mutually exclusive.

  One may also proceed by contradiction: let A and B be independent. Suppose, for the sake of contradiction, that A and B are mutually exclusive, then $\mathbb{P}(A \cap B) = 0$. However, from independence and $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$ we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) > 0$. This is a desired contradiction.

Exercise 1.5. Two coins, one with $\mathbb{P}(\text{head}) = u$ and one with $\mathbb{P}(\text{head}) = w$, are to be tossed together independently. Let

$$p_0 = \mathbb{P}(0 \text{ heads occur}), \quad p_1 = \mathbb{P}(1 \text{ heads occur}), \quad p_2 = \mathbb{P}(2 \text{ heads occur}).$$

Can u and w be chosen such that $p_0 = p_1 = p_2$? Prove your answer.

Solution. Start with probabilities: $p_0 = (1-u)(1-w)$, that is, first is T and second is T; $p_1 = (1-u)w + u(1-w)$, that is, first is T and second is H or first is H and second is T; and $p_2 = uw$, that is, both are H. Equating these probabilities, we receive

$$p_0 = p_2 \Rightarrow u + w = 1, \quad p_1 = p_2 \Rightarrow uw = \frac{w+u}{3}.$$

These together imply

$$u(1-u) = \frac{1}{3}.$$

This equation has no solution in the real numbers (more specifically, the right-hand side must be $\leq \frac{1}{4}$ for the solution in real numbers to exist). Thus, we cannot choose legitimate u and w to satisfy the conditions.

Exercise 1.6. Consider telegraph signals "dot" and "dash" sent in the proportion 3:4, where erratic transmissions cause a dot to become a dash with probability $1/4$ and a dash to become a dot with probability $1/3$.

(a) If a dash is received, what is the probability that a dash has been sent? If a dot is received, what is the probability that a dot has been sent?

(b) Assuming independence between signals, if the message dot-dot was received, what is the probability distribution of the four possible messages that could have been sent?

Solution. Let "DA" mean "dash", "DO" mean "dot", "R" mean "received", "S" mean "sent". From dot/dash 3:4 proportion we can calculate unconditional probabilities to observe dots and dashes:

$$\mathbb{P}(\text{dot sent}) := \mathbb{P}(\text{DOS}) = \frac{3}{3+4} = \frac{3}{7}, \quad \mathbb{P}(\text{dash sent}) := \mathbb{P}(\text{DAS}) = \frac{4}{3+4} = \frac{4}{7}.$$

From probabilities of mistakes we deduce other important objects, such as

$$\mathbb{P}(\text{DOR}|\text{DOS}) = 1 - \frac{1}{4} = \frac{3}{4}, \quad \mathbb{P}(\text{DAR}|\text{DAS}) = 1 - \frac{1}{3} = \frac{2}{3},$$

and

$$\mathbb{P}(\text{DOR}|\text{DAS}) = \frac{1}{3}, \quad \mathbb{P}(\text{DAR}|\text{DOS}) = \frac{1}{4}.$$

(a) Using the Bayes' rule, we calculate the following:

$$\begin{aligned}
\mathbb{P}(\text{DAS}|\text{DAR}) &= \frac{\mathbb{P}(\text{DAR}|\text{DAS}) \cdot \mathbb{P}(\text{DAS})}{\mathbb{P}(\text{DAR}|\text{DAS}) \cdot \mathbb{P}(\text{DAS}) + \mathbb{P}(\text{DAR}|\text{DOS}) \cdot \mathbb{P}(\text{DOS})} \\
&= \frac{(\frac{2}{3})(\frac{4}{7})}{(\frac{2}{3})(\frac{4}{7}) + (\frac{1}{4})(\frac{3}{7})} = \frac{32}{41},
\end{aligned}$$

which is precisely the probability that a dash has been sent if a dash is received. Next,

$$\begin{aligned}
\mathbb{P}(\text{DOS}|\text{DOR}) &= \frac{\mathbb{P}(\text{DOR}|\text{DOS}) \cdot \mathbb{P}(\text{DOS})}{\mathbb{P}(\text{DOR}|\text{DOS}) \cdot \mathbb{P}(\text{DOS}) + \mathbb{P}(\text{DOR}|\text{DAS}) \cdot \mathbb{P}(\text{DAS})} \\
&= \frac{(\frac{3}{4})(\frac{3}{7})}{(\frac{3}{4})(\frac{3}{7}) + (\frac{1}{3})(\frac{4}{7})} = \frac{27}{43},
\end{aligned}$$

which is the probability that a a dot has been sent if a dot is received.

(b) We need yet another probability with DOR condition,

$$\begin{aligned}
\mathbb{P}(\text{DAS}|\text{DOR}) &= \frac{\mathbb{P}(\text{DOR}|\text{DAS}) \cdot \mathbb{P}(\text{DAS})}{\mathbb{P}(\text{DOR}|\text{DAS}) \cdot \mathbb{P}(\text{DAS}) + \mathbb{P}(\text{DOR}|\text{DOS}) \cdot \mathbb{P}(\text{DOS})} \\
&= \frac{(\frac{1}{3})(\frac{4}{7})}{(\frac{1}{3})(\frac{4}{7}) + (\frac{3}{4})(\frac{3}{7})} = \frac{16}{43}.
\end{aligned}$$

Independence greatly simplifies calculations. We have DOR two times, so the possible signals sent with corresponding probabilities are,

$$\begin{aligned}
\mathbb{P}(\text{DOS}|\text{DOR}) \cdot \mathbb{P}(\text{DOS}|\text{DOR}) &= \frac{27}{43} \cdot \frac{27}{43}, \\
\mathbb{P}(\text{DOS}|\text{DOR}) \cdot \mathbb{P}(\text{DAS}|\text{DOR}) &= \frac{27}{43} \cdot \frac{16}{43}, \\
\mathbb{P}(\text{DAS}|\text{DOR}) \cdot \mathbb{P}(\text{DOS}|\text{DOR}) &= \frac{16}{43} \cdot \frac{27}{43}, \\
\mathbb{P}(\text{DAS}|\text{DOR}) \cdot \mathbb{P}(\text{DAS}|\text{DOR}) &= \frac{16}{43} \cdot \frac{16}{43},
\end{aligned}$$

which is the final distribution we are looking for.

Exercise 1.7. A students takes a test that consists of 20 multiple-choice questions, each with 4 possible answers. It is necessary to answer 10 questions correctly to pass the test. Find the probability that the student passes the test, given that she is guessing.

Solution. "At least 10 questions correct" means that we need to account for the situations when there are 10 correct guesses or up to all 20 guesses are correct. Thus, we will have a sum of probabilities.

What is the probability to correctly guess k times? For each question the probability of a correct answer is $\frac{1}{4}$, there are 20 questions and the order of correct answers does not matter (all questions are awarded equally). Each test results with k correct guesses has a probability of

$$\left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k}.$$

There are $\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$ ways to choose the order of correct answers (here $n = 20$). Thus, the probability to correctly guess exactly k answers is

$$\binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k}.$$

To account for different values of $k = 10, \ldots, 20$, we add up these probabilities and receive the final answer

$$\sum_{k=10}^{n=20} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} = 0.014.$$

2 Random variables

Exercise 2.1. Consider a random variable X with the following probability density function and $c < \infty$ being some constant:

$$f_X(x) = \begin{cases} cx(1-x), & \text{if } x \in (0,1) \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine the value of the parameter c that makes $f_X(x)$ a pdf. Derive the cumulative density function (cdf) of X.

(b) Derive mean, variance, skewness, and kurtosis of X.

(c) Consider another random variable $Y = X^3$, and derive its pdf.

(d) Derive mean, variance, skewness, and kurtosis of Y.

(e) For $g(x) = x^3$ compare $\mathbb{E}[g(X)]$ and $g(\mathbb{E}[X])$. What about $z(x) = x^{\frac{1}{3}}$?

(f) Determine the value of c that makes $f(x) = c \cdot e^{-|x|}$, $-\infty \leq x \leq \infty$ a pdf.

Solution.

(a) By the definition of the pdf,

$$\int_0^1 x(1-x)dx = \int_0^1 (x-x^2)dx = \int_0^1 xdx - \int_0^1 x^2 dx$$
$$= \frac{x^2}{2}\Big|_0^1 - \frac{x^3}{3}\Big|_0^1 = \frac{1}{2} - 0 - \frac{1}{3} + 0 = \frac{1}{6}.$$

Thus, we need $c = 6$ for the pdf to integrate to 1 over the entire support.

By the definition of the cdf,

$$F_X(x) = \int_0^x 6x(1-x)dx$$
$$= 6\left(\frac{x^2}{2}\Big|_0^x - \frac{x^3}{3}\Big|_0^x\right) = 6\left(\frac{3x^2 - 2x^3}{6}\right) = 3x^2 - 2x^3,$$

so that

$$F_X(x) = \begin{cases} 1, & \text{if } x \geq 1, \\ 3x^2 - 2x^3, & \text{if } x \in (0,1), \\ 0, & \text{if } x \leq 0. \end{cases}$$

(b) Again by the definition, we derive

- mean: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 6x^2(1-x)dx = 6\left(\frac{x^3}{3}\Big|_0^1 - \frac{x^4}{4}\Big|_0^1\right) = 6\left(\frac{1}{3} - \frac{1}{4}\right) = 0.5,$
- variance[2]: $\text{var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 0.3 - (0.5)^2 = 0.05$, because the second moment is
$$\mathbb{E}[X^2] = \int_0^1 6x^3(1-x)dx = 6\left(\frac{x^4}{4}\Big|_0^1 - \frac{x^5}{5}\Big|_0^1\right) = 6\left(\frac{1}{4} - \frac{1}{5}\right) = 0.3,$$
- skewness, κ_3, which is a rescaled version of the third *centered* moment:

$$\mathbb{E}[(X - \mathbb{E}[X])^3] = \int_0^1 (x - \mathbb{E}[x])^3 f_X(x)dx = \int_0^1 (x-0.5)^3 6x(1-x)dx$$
$$= 6\int_0^1 (x^3 - 1.5x^2 + 0.75x - 0.125)x(1-x)dx$$
$$= 6\left(-\frac{x^6}{6} + \frac{x^5}{2} - \frac{9x^4}{16} + \frac{7x^3}{24} - \frac{x^2}{16}\right)\Bigg|_0^1 = 0,$$

[2]Note that you can also proceed by the definition $\text{var}[X] = \int_0^1 (x - \mathbb{E}[x])^2 f_X(x)dx$.

so that the skewness is

$$\kappa_3 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^{\frac{3}{2}}} = 0.$$

- kurtosis, κ_4, which is a rescaled version of the *forth* moment,

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^4] &= \int_0^1 (x - \mathbb{E}[x])^4 f_X(x)dx = \int_0^1 (x - 0.5)^4 6x(1-x)dx \\
&= 6\left(-\frac{x^7}{7} + \frac{x^6}{2} - \frac{7x^5}{10} + \frac{x^4}{2} - \frac{3x^3}{16} + \frac{x^2}{32}\right)\Bigg|_0^1 \approx 0.005,
\end{aligned}$$

so that the kurtosis is

$$\kappa_4 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^2} \approx 2.14.$$

Note that the skewness is equal to zero, the fact that follows from the symmetry of the distribution. If the skewness is positive, the distribution is positively skewed or skewed right, meaning that the right tail of the distribution is longer than the left. If the skewness is negative, the distribution is negatively skewed or skewed left, meaning that the left tail is longer.

The kurtosis is a measure of the combined sizes of the two tails. If the kurtosis is greater than 3, then the distribution has heavier tails than a normal distribution. If the kurtosis is less than 3, then the distribution has lighter tails than a normal distribution. Note that kurtosis is always greater or equal to 0.

(c) Let us proceed by the definition of a cdf,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(X^3 \le y) = \mathbb{P}(X \le y^{1/3}) \\
&= \int_0^{y^{1/3}} 6x(1-x)dx = (3x^2 - 2x^3)\Big|_0^{y^{1/3}} = 3y^{2/3} - 2y.
\end{aligned}$$

Note that the support does not change, since $g(0) = 0$ and $g(1) = 1$. The full answer would be:

$$F_Y(y) = \begin{cases} 1, & \text{if } y \ge 1, \\ 3y^{2/3} - 2y, & \text{if } y \in (0,1), \\ 0, & \text{if } y \le 0. \end{cases}$$

The first equality is the definition; the second equality uses our transformation $g(x) = x^3$; the third inequality applies the inverse $g^{-1}(x) = x^{1/3}$ which is increasing on the support of X $(0,1)$ (and thus, preserves the order); the fourth inequality is again a definition.

Using the definition, we can also calculate the pdf,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = 2y^{-1/3} - 2,$$

for $y \in (0,1)$ and 0 otherwise.

(d) Again by the definition,

- mean: $\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)dy = 0.2,$
- variance: $\text{var}\,[Y] = \int_{-\infty}^{\infty} (y - \mathbb{E}[y])^2 f_Y(y)dy = \dfrac{13}{300} \approx 0.04,$
- skewness: $\kappa_3[Y] = \dfrac{\mathrm{E}[(X - \mathrm{E}[X])^3]}{\mathrm{E}[(X - \mathrm{E}[X])^2]^{\frac{3}{2}}} = \dfrac{\frac{63}{5500}}{(\frac{13}{300})^{\frac{3}{2}}} \approx 1.27,$
- kurtosis: $\kappa_4[Y] = \dfrac{\mathrm{E}[(X - \mathrm{E}[X])^4]}{\mathrm{E}[(X - \mathrm{E}[X])^2]^2} = \dfrac{\frac{713}{96250}}{(\frac{13}{300})^2} \approx 3.95.$

(e) Here, the following result is useful:

Theorem 2.2 (Jensen's inequality)**.** *For any random variable X, if $g(x)$ is a convex function, then*

$$g(\mathbb{E}[X]) \le \mathbb{E}[g(X)].$$

If $g(x)$ is a concave function, the inequality is reversed.

Proof. We focus on the convex case. Let $a + bx$ be the tangent line to $g(x)$ at $x = \mathbb{E}[X]$. Since $g(x)$ is convex, $g(x) \ge a + bx$. Evaluating at $x = X$ and taking expectations, we find

$$\mathbb{E}[g(X)] \ge a + b\mathbb{E}[X] = g(\mathbb{E}[X])$$

as claimed. □

Now, we have

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_0^1 x^3 6x(1-x) dx = 0.2,$$
$$g(\mathbb{E}[X]) = g(0.5) = 0.5^3 = 0.125.$$

so $\mathbb{E}[g(X)] > g(\mathbb{E}[X])$ in line with Theorem 2.2.
For $z(x) = x^{1/3}$ we have

$$\mathbb{E}[z(X)] = \int_{-\infty}^{\infty} z(x) f_X(x) dx = \int_0^1 x^{1/3} 6x(1-x) dx \approx 0.77,$$
$$z(\mathbb{E}[X]) = z(0.5) = 0.5^{\frac{1}{3}} \approx 0.79,$$

so $\mathbb{E}[z(X)] < z(\mathbb{E}[X])$. It is, again, in line with Theorem 2.2 because $z(x)$ is concave.

(f) $\int_{-\infty}^{\infty} e^{-|x|} dx = \int_{-\infty}^{0} e^x dx + \int_0^{\infty} e^{-x} dx = 1 + 1 = 2$, so we need $c = \dfrac{1}{2}$ for pdf to integrate to 1 over the entire support. Non-negativity is obvious.

Exercise 2.3. In each of the following find the pdf of Y, and show that the pdf integrates to 1.

(a) $Y = X^3$ and $f_X(x) = 42x^5(1-x)$, $0 < x < 1$,

(b) $Y = 4X + 3$ and $f_X(x) = 7e^{-7x}$, $0 < x < \infty$,

(c) $Y = X^2$ and $f_X(x) = 30x^2(1-x)^2$, $0 < x < 1$.

Solution. Here, two following results are useful (we state them separately even though Theorem 2.5 is used in the proof of Theorem 2.4):

Theorem 2.4 (Monotonic transformation)**.** *If X has a pdf $f_X(x)$, $f_X(x)$ is continuous on its support $\mathcal{X}$, $g(x)$ is strictly monotone, and $g^{-1}(y)$ is continuously differentiable on $\mathcal{Y}$, then for $y \in \mathcal{Y}$*

$$f_Y(y) = f_X\left(g^{-1}(y)\right) J(y), \quad J(y) := \left|\frac{d}{dy} g^{-1}(y)\right|.$$

Theorem 2.5. *Let X and $Y = g(X)$ be two random variables, $\mathcal{X}$ and $\mathcal{Y}$ be their supports, and let $F_X(x)$ be the cdf of X.*

(a) If g is increasing in $\mathcal{X}$, we have $F_Y(y) = F_X(g^{-1}(y))$ for all $y \in \mathcal{Y}$.

(b) If g is decreasing in $\mathcal{X}$ and X is a continuous random variable, we have $F_Y(y) = 1 - F_X(g^{-1}(y))$ for all $y \in \mathcal{Y}$.

(a) We have that $y = g(x) = x^3$, $g'(x) = 3x^2$ so $g(x)$ is monotone and increasing. If $0 < x < 1$ then $[g(0) = 0] < [g(x) = y] < [g(1) = 1]$, thus $0 < y < 1$ is our new support. $g^{-1}(y) = y^{\frac{1}{3}}$, $\dfrac{d}{dy} g^{-1}(y) = \dfrac{1}{3} y^{-\frac{2}{3}}$.

Applying Theorem 2.4, we have

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy} g^{-1}(y)\right| = 42 y^{\frac{5}{3}}\left(1 - y^{\frac{1}{3}}\right)\left(\frac{1}{3} y^{-\frac{2}{3}}\right) = 14y - 14y^{\frac{4}{3}}$$

for $0 < y < 1$ and 0 otherwise. It is a valid pdf, since

$$\int_0^1 14y - 14y^{\frac{4}{3}} dy = 7y^2 - 6y^{\frac{7}{3}}|_0^1 = 1 - 0 = 1.$$

(b) An alternative way how to compute transformations is by using the cdf of the random variable. First, we derive the cdf as

$$F_X(x) = \int_0^x 7e^{-7x} dx = -e^{-7x}\Big|_0^x = -e^{-7x} + e^0 = 1 - e^{-7x},$$

and then $y = g(x) = 4x + 3$, $g'(x) = 4 > 0$ so $g(x)$ is monotone and increasing, and $g^{-1}(y) = \dfrac{y-3}{4}$. If $0 < x < \infty$ then $[g(0) = 3] < [g(x) = y] < [g(\infty) = \infty]$, thus $3 < y < \infty$ is our new support. Now apply Theorem 2.5 to get

$$F_Y(y) = F_X(g^{-1}(y)) = 1 - e^{-\frac{7}{4}(y-3)}$$

for $3 < y < \infty$ and 0 otherwise. From here, we can differentiate to come to the pdf,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{7}{4}e^{-\frac{7}{4}(y-3)}$$

for $3 < y < \infty$ and 0 otherwise. It is a valid pdf, since

$$\int_3^\infty \frac{7}{4}e^{-\frac{7}{4}(y-3)} dy = -e^{-\frac{7}{4}(y-3)}|_3^\infty = 0 - (-1) = 1.$$

(c) Yet another alternative, is to proceed "by definition": note that with $g(x) = x^2$ and $0 < x < 1$ we have $[g(0) = 0] < [g(x) = y] < [g(1) = 1]$, thus $0 < y < 1$ is our new support. By definition, $F_Y(y) = \mathbb{P}(Y < y) = \mathbb{P}(X < \sqrt{y}) = F_X(\sqrt{y})$. In this step we find a region in X such that $Y < y$ holds. Differentiate both sides with respect to y (come to pdfs) to receive

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{2\sqrt{y}} 30(\sqrt{y})^2(1-\sqrt{y})^2 = 15y^{\frac{1}{2}}(1-y^{\frac{1}{2}})^2$$

for $0 < y < 1$ and 0 otherwise. It is a valid pdf, since

$$\int_0^1 15y^{\frac{1}{2}}(1-y^{\frac{1}{2}})^2 dy = \int_0^1 15y^{\frac{1}{2}} - 30y + 15y^{\frac{3}{2}} dy = 15\frac{2}{3} - 30\frac{1}{2} + 15\frac{2}{5} = 1.$$

3 Parametric distributions

Exercise 3.1. Consider the following:

(a) Derive the moment-generating function (mgf) of $\mathcal{B}(n, p)$ distribution.

(b) It has been determined that 5% of drivers checked at a road stop show traces of alcohol and 10% of drivers checked do not wear seat belts. Assume that the two infractions are independent from one another. If an officer stops five drivers at random: *(i)* calculate the probability that exactly three of the drivers have committed any one of the two offenses, *(ii)* calculate the probability that at least one of the drivers checked has committed at least one of the two offenses.

(c) Derive the mgf of the Poisson(λ) distribution with parameter λ.

(d) Find mean and variance of the Poisson distribution.

(e) Discuss the connection between Poisson(λ) and $\mathcal{B}(n, p)$.

(f) In the manufacture of glassware, bubbles can occur in the glass which reduces the status of the glassware to that of a low quality. If, on average, one in every 1000 items produced has a bubble, calculate the probability that exactly six items in a batch of three thousand are of a low quality.

Solution.

(a) First, note that the pmf of $\mathcal{B}(n, p)$ is $f(x) = C_n^x p^x(1-p)^{n-x}$, where $x \geq 0$ is an integer, and p is a probability of a success. The mgf is then

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{x=0}^n e^{tx} C_n^x p^x (1-p)^{n-x} = \sum_{x=0}^n C_n^x (pe^t)^x (1-p)^{n-x} = [pe^t + (1-p)]^n,$$

where the last equality uses the binomial formula $\sum_{x=0}^n C_n^x u^x v^{n-x} = (u+v)^n$ with $u = pe^t$ and $v = 1 - p$. You can check that $\mathbb{E}[X] = np$ and var $[X] = np(1-p)$ by using this mgf.

(b) Denote X the number of drivers who have committed any one of the two offenses; A meaning a driver shows traces of alcohol, and B meaning a driver does not wear a seat belt. We have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.05 + 0.1 - 0.05 \cdot 0.1 = 0.145 := p.$$

n is 5 so we are having $\mathcal{B}(n = 5, p = 0.145)$. Then for *(i)*, we have $\mathbb{P}(X = 3) = C_5^3 (0.145)^3 (1 - 0.145)^{5-3} = 0.0223$.

For *(ii)*, use the rule that $\mathbb{P}(\text{at least one}) = 1 - \mathbb{P}(\text{no one})$. The resulting probability is $1 - C_5^0 (1 - 0.145)^5 = 0.543$.

(c) First, note that the pmf is $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, where $x \geq 0$ is an integer and λ is a positive constant. We have

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} e^{tx} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)},$$

where the fourth equality uses power series expansion of exponential function $e^x = \sum_{n=0}^{\infty} \dfrac{x^n}{n!} = 1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \ldots$, which is valid $\forall x \in \mathbb{R}$.

(d) To compute any nth moment, differentiate the mgf w.r.t. t n times and evaluate the derivative at $t = 0$,

$$\mathbb{E}[X^n] = \frac{d^n M_X(t)}{dt^n}\Big|_{t=0}.$$

For our needs, we require the first and the second moment. We have

$$\begin{aligned}
\mathbb{E}[X] &= \frac{de^{\lambda(e^t-1)}}{dt}\Big|_{t=0} = e^{\lambda(e^t-1)} \lambda e^t |_{t=0} = e^{\lambda(1-1)} \lambda e^0 = \lambda, \\
\mathbb{E}[X^2] &= \frac{d^2 e^{\lambda(e^t-1)}}{dt^2}\Big|_{t=0} = e^{\lambda(e^t-1)} \lambda e^t \lambda e^t + e^{\lambda(e^t-1)} \lambda e^t |_{t=0} = \lambda^2 + \lambda, \\
\operatorname{var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}$$

(e) Binomial probabilities can be approximated by Poisson probabilities (the latter are easier to calculate). Approximation is valid if n is large and np is small. To get the result, consider $X \sim \mathcal{B}(n, p)$ and $Y \sim \text{Poisson}(\lambda)$ with $\lambda = np$. The approximation is then defined as having $P(X = x) \approx P(Y = x)$ for large n and small np. Formally, we show the result if we show that their mgfs converge, thus having $F_X(u) = F_Y(u)$ for all u (see Theorem 2.3.11 in *Casella & Berger*).

Define $p = \dfrac{\lambda}{n}$, let $n \to \infty$, and look at the limit of $M_X(t)$,

$$M_X(t) = [pe^t + (1-p)]^n = [1 + p(e^t - 1)]^n = \left[1 + \frac{\lambda(e^t - 1)}{n}\right]^n \to e^{\lambda(e^t-1)} = M_Y(t),$$

where the limit uses the fact that $\lim_{n\to\infty} \left(1 + \dfrac{a_n}{n}\right)^n = e^a$ with $a = \lim_{n\to\infty} a_n$.

(f) Denote X the number of items with bubbles (low quality). Clearly, $X \sim \mathcal{B}(n = 3000, p = 0.001)$, so we are in the context where n is large and np is small. Thus, the Poisson approximation should be accurate. Then $\lambda = np = 3000 \cdot 0.001 = 3$ and $\mathbb{P}(X = 6) \approx \dfrac{\lambda^x e^{-\lambda}}{x!} = \dfrac{3^6 e^{-3}}{6!} \approx 0.0498 \cdot 1.0125 \approx 0.05041$. Using any statistical programming language, we can compute the exact probability, which is 0.050384.

Note: remember the usual interpretation of Poisson as the number of occurrences in a given interval for which the average rate of occurrences is λ (think of an example with a student waiting for a bus at a bus stop).

Exercise 3.2. A random point (X, Y) is distributed uniformly on the square with vertices $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$. That is, the joint pdf is $f(x, y) = \frac{1}{4}$ on the square.

(a) Determine the probabilities of the following events: *(i)* $X^2 + Y^2 \leq 1$, *(ii)* $2X - Y \geq 0$, *(iii)* $|X + Y| \leq 2$.

(b) Derive the marginal and the conditional densities of X. What would you conclude regarding the same densities of Y?

(c) A and B agree to meet at a certain place between 1 PM and 2 PM. Suppose they arrive at the meeting place independently and randomly during the hour. Find the distribution of the length of time A waits for B. If B arrives before A, define A's waiting time as 0.