1 Random variables and probability distributions

The goal of this exercise is to refresh the basic concepts of probability which are necessary for the treatment of stochastic processes. You are expected to be familiar with these topics from previous studies such as `MS-A05XX First course in probability and statistics`. Suggested background reading for refreshing the basics of probability is available for example in the freely downloadable book [GS97] or the lecture notes [Les18].

Classroom exercises

1.1 *Six javelin throws.* In an athletics event Tero Pitkämäki throws a javelin six times. Suppose that the lengths of the throws (in meters) are independent random numbers $Z_1, \ldots, Z_6$ following the uniform distribution on the continuous interval $(80, 92)$.

(a) Calculate the expected value and variance of Tero's first throw.

Solution. Denote $a = 80$, $b = 92$, and $h = b - a = 12$. The length of the first throw can be written as $Z_1 = a + hU_1$, where U_1 follows the continuous uniform distribution on the interval $(0, 1)$. The probability density function (pdf) of U_1 is $1_{(0,1)}(u)$, so the expected value is

$$\mathbb{E}U_1 \;=\; \int_0^1 u\,du \;=\; \frac{1}{2},$$

the second moment is

$$\mathbb{E}U_1^2 \;=\; \int_0^1 u^2\,du \;=\; \frac{1}{3},$$

and the variance is

$$\operatorname{Var}(U_1) \;=\; \mathbb{E}U_1^2 - (\mathbb{E}U_1)^2 \;=\; \frac{1}{3} - \left(\frac{1}{2}\right)^2 \;=\; \frac{1}{12}.$$

Thus,

$$\mathbb{E}Z_1 = a + h\mathbb{E}U_1 = 86,$$

and

$$\operatorname{Var}(Z_1) = h^2 \operatorname{Var}(U_1) = 12.$$

Additional information. Since the length of the throw is Z_1 meters, the variance is 12 *square meters.* The uncertainty related to Tero's throws might be better described by the *standard deviation* $\sqrt{\operatorname{Var}(Z_1)} \approx 3.46m$.

(b) Determine the cumulative distribution function F_Y and the density function f_Y of the random number $Y = \max(Z_1, \ldots, Z_6)$, and calculate the probability that Tero's longest throw is at least 91 meters.

Solution. The cumulative distribution function (cdf) F_Y of a random number Y is defined as

$$F_Y(y) = \mathbb{P}(Y \le y),$$

so the cdf of $Y = \max(Z_1, \dots, Z_6)$ is

$$\begin{aligned} F_Y(y) &= \mathbb{P}(\max(Z_1, \dots, Z_6) \le y) \\ &= \mathbb{P}(Z_1 \le y, \dots, Z_6 \le y) \\ &= \mathbb{P}(Z_1 \le y)^6 \\ &= ((y-a)/h)^6 \\ &= h^{-6}(y-a)^6, \end{aligned}$$

for $y \in [a, b]$. Moreover, $F_Y(y) = 0$ for $y < a$, and $F_Y(y) = 1$ for $y > b$, so that

$$F_Y(y) = \begin{cases} 0, & y \le a \\ h^{-6}(y-a)^6, & a \le y \le b \\ 1, & y \ge b. \end{cases}$$

The pdf is obtained by differentiating:

$$f_Y(y) = \begin{cases} 0, & y \le a \\ 6h^{-6}(y-a)^5, & a < y < b \\ 0, & y \ge b. \end{cases} \tag{1}$$

We see that $F_Y(y) = \int_{-\infty}^{y} f_Y(x)dx$, i.e., f_Y indeed is the pdf corresponding to F_Y. Tero's longest throw exceeds 91 meters with probability

$$\mathbb{P}(Y > 91) = 1 - \mathbb{P}(Y \le 91) = 1 - F_Y(91) = 0.407.$$

Additional information. In general, $f_Y : \mathbb{R} \to \mathbb{R}_{\ge 0}$ is the pdf of a random number of Y, if for all $y \in \mathbb{R}$ it holds that

$$F_Y(y) = \int_{-\infty}^{y} f_Y(x)dx. \tag{2}$$

Not all random numbers have pdfs. In particular, by the Fundamental Theorem of Calculus it follows from (2) that

(a) the pdf can exist only if F_Y is continuous
(b) if F_Y is continuous and the derivative F_Y' exists and is continuous (at all but finitely many points), then Y has the pdf $f_Y = F_Y'$ (for all but finitely many points).

In practice, one may guess that $f_Y = F_Y'$ and show that it satisfies (2). Differentiating the cdf of $Y = \max(Z_1, \dots, Z_6)$ gives

$$F_Y'(y) = \begin{cases} 0, & y < a \\ 6h^{-6}(y-a)^5, & a < y < b \\ 0, & y > b. \end{cases}$$

The derivative exists and is continuous, except possibly at a and b. By choosing $f_Y(a) = f_Y(b) = 0$ we get the solution (1). Other choices could be made, since the values at the two points do not affect the integral in (2).

(c) Find out the cumulative distribution function F_X of the random number $X = \min(Z_1, \ldots, Z_6)$, and calculate the probability that at least one of Tero's throws is shorter than 85 meters.

Solution. $F_X(x) = 0$ when $x < 80$, and $F_X(x) = 1$ when $x > 92$. Otherwise,

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \le x) \\ &= 1 - \mathbb{P}(Z_1 > x \ \& \ \ldots \ \& \ Z_6 > x) \\ &= 1 - [1 - \mathbb{P}(Z_1 \le x)]^6 \\ &= 1 - [1 - (x-a)/h]^6. \end{aligned}$$

This cdf is continuous, so that

$$\mathbb{P}(X < 85) = \mathbb{P}(X \le 85) = F_X(85) = 1 - [1 - (85-80)/12]^6 \approx 0.96.$$

(d) Does the joint distribution of X and Y have a density function $f_{X,Y}$? Are the random numbers X and Y independent? Justify your answer.

Solution. The distribution of X and Y does have a (joint) pdf. We show this by finding it explicitly. First we study the joint cdf: for $x \ge y$ we have

$$F_{(X,Y)}(x,y) = \mathbb{P}(X \le x, Y \le y) = \mathbb{P}(Y \le y) = F_Y(y),$$

because $X = \min(Z_1, \ldots, Z_6) \le \max(Z_1, \ldots, Z_6)) = Y$. For $x < y$ we obtain

$$\begin{aligned} F_{(X,Y)}(x,y) &= \mathbb{P}(X \le x, Y \le y) \\ &= \mathbb{P}(Y \le y) - \mathbb{P}(X > x, Y \le y) \\ &= F_Y(y) - \mathbb{P}(x < Z_1 \le y, \ldots, x < Z_6 \le y) \\ &= F_Y(y) - \mathbb{P}(x < Z_1 \le y)^6 \\ &= F_Y(y) - [F_{Z_1}(y) - F_{Z_1}(x)]^6. \end{aligned}$$

By differentiating we find that in the set $A := \{(x,y) \in \mathbb{R}^2 : a < x < y < b\}$,

$$\partial_x \partial_y F_{(X,Y)}(x,y) = 30h^{-6}(y-x)^4.$$

One may verify that integrating the this function over A gives 1. Especially, we may choose

$$f_{(X,Y)}(x,y) = \begin{cases} 30h^{-6}(y-x)^4, & a < x < y < b, \\ 0, & \text{otherwise.} \end{cases}$$

Additional information. The joint cdf of a random vector $(X_1, \ldots, X_n)$ is defined as

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n).$$

The random vector $(X_1, \ldots, X_n)$ has a joint pdf $f_{(X_1,\ldots,X_n)} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, if for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$ it holds

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{(X_1,\ldots,X_n)}(a_1, \ldots, a_n) da_n \ldots da_1.$$

Not all random vectors have joint pdfs. In particular,

(a) the joint cdf can only exist if $F_{(X_1,\ldots,X_n)}$ is continuous

(b) where $F_{(X_1,\ldots,X_n)}$ is continuous and the derivatives up to nth order exist and are continuous, the joint pdf $f_{(X_1,\ldots,X_n)} = \partial_1 \cdots \partial_n F_{(X_1,\ldots,X_n)}$ describes the distribution of $(X_1, \ldots, X_n)$ *locally.*

As (b) above suggests, in practice we may differentiate the joint cdf with respect to all variables and verify that the resulting function integrates to 1.

X and Y are not independent: since the longest throw cannot be shorter than the shortest throw,

$$\mathbb{P}(X \geq 90, Y \leq 89) = 0 \neq \mathbb{P}(X \geq 90)\mathbb{P}(Y \leq 89) > 0.$$

Additional information. By definition, random variables X and Y are independent if for all sets $B_X, B_Y \subset \mathbb{R}$ it holds that

$$\mathbb{P}(X \in B_X, Y \in B_Y) = \mathbb{P}(X \in B_X)\mathbb{P}(Y \in B_Y).$$

(e) Does the joint distribution of Z_1 and Y have a density function $f_{Z_1,Y}$?

Solution. The joint pdf does not exist. Every throw has the same probability of being the shortest, i.e., $\mathbb{P}(Z_1 = Y) = 1/6$. If there was a joint pdf $f_{Z_1,Y}$, its integral over the line $Z_1 = Y$ would be $1/6$. This is not possible, because the area of the line is zero.

Homework problems

1.2 *Memoryless distributions.* Let X be a random number following the geometric distribution on the set $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ with a success probability $p \in (0, 1)$, so that X has the probability mass function

$$\pi_X(k) = (1-p)^k p, \qquad \text{for } k = 0, 1, 2, \dots$$

(a) Find the conditional probability $\mathbb{P}\big[X \geq t + h \,\big|\, X \geq t\big]$ for integers $t, h \geq 0$.

Solution. By the definition of conditional probability, $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$. We denote $q = 1 - p$, so that

$$\begin{aligned}
\mathbb{P}(X \geq t+h | X \geq t) &= \frac{\mathbb{P}(X \geq t+h)}{\mathbb{P}(X \geq t)} \\
&= \frac{p\sum_{k=t+h}^{\infty}(1-p)^k}{p\sum_{k=t}^{\infty}(1-p)^k} \\
&= \frac{pq^{t+h}\frac{1}{1-q}}{pq^t\frac{1}{1-q}} \\
&= q^h = (1-p)^h.
\end{aligned}$$

Additional information. The result above does not depend on t. This is the memorylessness property of the geometric distribution: waiting for a time t does not change the distribution of the remaining waiting time. This memorylessness property is also sufficient to characterize the geometric distribution. Assume that a $\mathbb{Z}_+$-valued random variable X has a memoryless distribution. Then for all t

$$\begin{aligned}
\mathbb{P}(X = 1+t \mid X \geq 1) &= \mathbb{P}(X = t) \\
\Rightarrow \quad \mathbb{P}(X = 1+t)\frac{1}{\mathbb{P}(X \geq 1)} &= \mathbb{P}(X = t),
\end{aligned}$$

so that the ratio between successive point masses is the constant $\frac{1}{\mathbb{P}(X\geq 1)}$, i.e., X has a geometric distribution.

(b) Calculate the expected value and variance of X.

Solution. We calculate the expected value and variance directly from the definition. An alternative way would be to use the moment generating function.

(i)

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{\infty} k\mathbb{P}(X = k) \\
&= \sum_{k=0}^{\infty} kp(1-p)^k.
\end{aligned}$$

The sum can be evaluated using the derivative. Denote $q = 1 - p$, so that

$$\begin{aligned} \sum_{k=0}^{\infty} kq^k &= q\frac{\partial}{\partial q}\sum_{k=0}^{\infty} q^k \\ &= q\frac{\partial}{\partial q}\frac{1}{1-q} \\ &= q\frac{1}{(1-q)^2} = \frac{1-p}{p^2}, \end{aligned} \tag{3}$$

and we obtain

$$\begin{aligned} \mathbb{E}X &= p\sum_{k=0}^{\infty} k(1-p)^k \\ &= p\frac{1-p}{p^2} = \frac{1-p}{p}. \end{aligned}$$

(ii) We find the variance using the formula $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$, so what's left is to find the second moment:

$$\mathbb{E}X^2 = p\sum_{k=0}^{\infty} k^2(1-p)^k.$$

The sum is again evaluated by differentiating:

$$\begin{aligned} \sum_{k=0}^{\infty} k^2q^k &= q\frac{\partial}{\partial q}\left(\sum_{k=0}^{\infty} kq^k\right) \\ &= q\frac{\partial}{\partial q}\left(\frac{q}{(1-q)^2}\right) \\ &= q\frac{q+1}{(1-q)^3}, \end{aligned}$$

and so

$$\mathbb{E}X^2 = p(1-p)\frac{2-p}{p^3} = \frac{(1-p)(2-p)}{p^2}.$$

Finally

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ &= \frac{(1-p)(2-p)}{p^2} - \frac{(1-p)^2}{p^2} \\ &= \frac{1-p}{p^2}. \end{aligned}$$

Additional information. The sums (3) ja (4) can also be evaluated by only using the formula for the sum of the geometric series. Since all the summands are nonnegative, we may change the order of summation freely:

$$\begin{aligned}\sum_{k=0}^{\infty} kq^k &= \sum_{k=1}^{\infty} q^k + \sum_{k=2}^{\infty} q^k + \sum_{k=3}^{\infty} q^k + \dots \\ &= \frac{q}{1-q} + \frac{q^2}{1-q} + \frac{q^3}{1-q} + \dots \\ &= \frac{1}{1-q}\sum_{k=1}^{\infty} q^k \\ &= \frac{q}{(1-q)^2}.\end{aligned}$$

Similarly for (4):

$$\begin{aligned}\sum_{k=0}^{\infty} k^2 q^k &= \sum_{k=1}^{\infty} kq^k + \sum_{k=2}^{\infty} kq^k + \sum_{k=3}^{\infty} kq^k + \dots \\ &= \sum_{\ell=0}^{\infty}\sum_{k=1}^{\infty} (\ell+k) q^{\ell+k} \\ &= \sum_{\ell=0}^{\infty}\sum_{k=1}^{\infty} \ell q^{\ell+k} + \sum_{\ell=0}^{\infty}\sum_{k=1}^{\infty} k q^{\ell+k} \\ &= \sum_{\ell=0}^{\infty} \ell q^{\ell} \sum_{k=1}^{\infty} q^k + \sum_{\ell=0}^{\infty} q^{\ell} \sum_{k=1}^{\infty} kq^k \\ &= \frac{q}{(1-q)^2}\frac{q}{1-q} + \frac{1}{1-q}\frac{q}{(1-q)^2} \\ &= q\frac{q+1}{(1-q)^3}.\end{aligned}$$

These are seen to be identical to the expressions we got by differentiation.

Let Y be a random number following the exponential distribution with a rate parameter $\lambda > 0$, so that Y has the density function

$$f_Y(x) \;=\; \lambda e^{-\lambda x} 1_{(0,\infty)}(x) \;=\; \begin{cases} \lambda e^{-\lambda x}, & \text{for } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(c) Find the conditional probability $\mathbb{P}\big[Y > t+h \,\big|\, Y > t\big]$ for real numbers $t, h > 0$.

Solution. It holds that

$$\mathbb{P}\big[Y > t+h \,\big|\, Y > t\big] \;=\; \mathbb{P}\big[Y > t+h\big] / \mathbb{P}\big[Y > t\big].$$

To evaluate this we need the integral of the exponential function:

$$\int_a^\infty \lambda e^{-\lambda x} dx = e^{-a\lambda}.$$

It follows that

$$\mathbb{P}\big[Y > t+h \,\big|\, Y > t\big] \;=\; e^{-(t+h)\lambda}/e^{-t\lambda} = e^{-h\lambda}.$$

Additional information. Since the result does not depend on t, the exponential distribution is memoryless like the geometric distribution, but in *continuous time*. This property is sufficient to characterize the exponential distribution, i.e., there are no other memoryless and continuous distributions on $\mathbb{R}_+$.

(d) Calculate the expected value and variance of Y.

Solution.

(i) Integration by parts gives

$$\begin{aligned}
\mathbb{E}Y &= \int_{\mathbb{R}} y f_Y(y) dy \\
&= \lambda \int_{\mathbb{R}_+} y e^{-\lambda y} dy \\
&= \lambda \left[-\int_{\mathbb{R}_+} -\frac{1}{\lambda} e^{-\lambda y} dy + \Big/_{y=0}^{\infty} \left(-\frac{1}{\lambda} e^{-\lambda y} \times y \right) \right] \\
&= \frac{1}{\lambda} \int_{\mathbb{R}_+} \lambda e^{-\lambda y} dy = 1/\lambda.
\end{aligned}$$

(ii) We modify the above calculation to yield a recursive formula for the moments of the exponential distribution: $\mathbb{E}Y^0 = 1$ and

$$\begin{aligned}
\mathbb{E}[Y^k] &= \int_{\mathbb{R}} y^k f_Y(y) dy \\
&= \lambda \int_{\mathbb{R}_+} y^k e^{-\lambda y} dy \\
&= \lambda \left[-\int_{\mathbb{R}_+} -\frac{1}{\lambda} e^{-\lambda y} \times k y^{k-1} dy + \Big/_{y=0}^{\infty} \left(-\frac{1}{\lambda} e^{-\lambda y} \times y^k \right) \right] \\
&= \frac{k}{\lambda} \int_{\mathbb{R}_+} \lambda e^{-\lambda y} y^{k-1} dy \\
&= \frac{k}{\lambda} \mathbb{E}[Y^{k-1}].
\end{aligned}$$

Thus, $\mathbb{E}[Y^k] = k!/\lambda^k$ and in particular

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 1/\lambda^2.$$

1.3 *Robot football finals.* Otaniemi Eulers and the Leppävaara Algebra are fighting for the championship of the robot football league by playing a series of matches in a best-of-three system, where the winner of the whole league is the team that gets two wins. Suppose that the results of the matches are independent and that Eulers win each match with probability $p = 0.55$.

(a) What is the probability that Eulers take the championship?
Solution.
Let X_k be the number of matches won by Eulers among the k first matches. Denote

$$\theta_j = \begin{cases} 1, & \text{if Eulers win match } j, \\ 0, & \text{otherwise.} \end{cases}$$

For example, $X_3 = \theta_1 + \theta_2 + \theta_3$ is a sum of three independent Ber(p)-distributed random variables with $p = 0.55$. We recall that such a random variable has distribution Bin$(3, 0.55)$, and so

$$\mathbb{P}(X_3 \geq 2) = \binom{3}{2} p^2(1-p)^1 + \binom{3}{3} p^3(1-p)^0 = 3p^2(1-p) + p^3 = 0.575.$$

(b) What is the probability that it takes three matches to determine the champion?
Solution. The championship is determined by the third match if $(\theta_1, \theta_2) = (1, 0)$ or $(\theta_1, \theta_2) = (0, 1)$, so the probability is given by

$$2(1-p)p = 0.495.$$

(c) What is the expected number of matches required to determine the champion?
Solution. The championship is determined by the second or the third match. By the previous part, the latter probability is $p_3 = 2(1-p)p$. The complement probability is then $p_2 = 1 - p_3$, and so

$$2p_2 + 3p_3 = 2(1-p_3) + 3p_3 = 2 + p_3 = 2.495.$$

Now consider a change in the rules where a best-of-seven system is used, so that the champion is the team that gets four wins.

(d) What is the probability that Eulers take the championship?
Solution. Eulers take the championship exactly when $X_7 \geq 4$:

$$\mathbb{P}(X_7 \geq 4) = \sum_{j=4}^{7} \binom{7}{j} p^j (1-p)^{7-j} = 0.608.$$

(e) What is the probability that it takes seven matches to determine the champion?
Solution. A seventh game is needed exactly when the wins after six games are 3–3. Again from the Binomial distribution we get

$$\mathbb{P}(X_6 = 3) = \binom{6}{3} p^3(1-p)^3 \approx 0.303.$$

References

[GS97] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability.* American Mathematical Society, http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html, 1997.

[Les18] Lasse Leskelä. Stokastiikka ja tilastollinen ajattelu. https://math.aalto.fi/~lleskela/LectureNotes003.html, 2018.