2A Continuous random variables

Class problems

2A1 (A model for wage distribution.) We decide to model the monthly wage (in euros) earned by a randomly chosen employee with a continuous random variable X that has density

$$f(x) = \begin{cases} \alpha\, c^{\alpha}\, x^{-\alpha-1}, & \text{when } x > c, \\ 0, & \text{otherwise,} \end{cases}$$

with $\alpha = 1.6$ and $c = 1500$.

(a) What is the cumulative distribution function of X? Draw a graph of it.

(b) What are the possible values of X? In particular, how small and how large can wages be?

(c) Calculate the probability that a randomly chosen employee earns more than 15 000 eur.

(d) Find a real number x such that 90% of employees earn at most x eur.

Solution.

(a) The CDF at point x (if $x > c$) is obtained from the integral

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{c}^{x} f(t)\,dt = \alpha c^{\alpha} \int_{c}^{x} t^{-\alpha-1}\,dt = \alpha c^{\alpha} \Big[-\alpha^{-1} t^{-\alpha}\Big]_{t=c}^{x} = 1 - c^{\alpha} x^{-\alpha}.$$

(For points $x \le c$ we have $F(x) = 0$). Thus

$$F(x) = \begin{cases} 1 - c^{\alpha} x^{-\alpha}, & x > c, \\ 0, & \text{otherwise.} \end{cases}$$

(Graph not shown here.)

(b) Because $f(x) > 0$ for all $x > c$, and because f is continuous on the interval (c, ∞), X can take any values in $(1500, \infty)$; wages are above 1500, but can be arbitrarily large.

(c) By the complement rule, the probability of $\{X > x\}$ is $1 - P(X \le x) = 1 - F(x) = c^{\alpha} x^{-\alpha}$. Plugging in $x = 15000$ and the parameters c, α, we have $P(X > 15000) = 1 - F(15000) = 1500^{1.6} \times 15000^{-1.6} \approx 2.51\%$.

(d) Let $u = 0.90$ and solve x from the equation $F(x) = u$, that is $1 - c^{\alpha} x^{-\alpha} = u$, so we get

$$x = \frac{c}{(1-u)^{1/\alpha}} = \frac{1500}{0.10^{1/1.6}} \approx 6325.45.$$

Thus according to our model, 90% of employees earn at most 6325.45 eur (and 10% earn more than that).

The point $x = 6325.45$ is called the 0.9-quantile of X. This distribution is called the Pareto distribution, with shape parameter $\alpha = 1.6$ and scale parameter $c = 1500$.

2A2 (Both late.) In this problem we study a *continuous joint distribution* of two variables in a simple case. From multivariate calculus, we only need the following fact: the integral of a *constant* function over a region equals that constant times the area of the region.

Ursula and Vera have agreed to meet for lunch exactly at noon (12:00). However, Ursula arrives U minutes late, and Vera arrives V minutes late. We assume their arrival times are independent, and both are uniformly distributed over the interval $[0, 60]$.

(a) Write down the density functions f_U and f_V, and the joint density function $f_{U,V}$.
Start from the individual densities and then deduce the joint density. Read lecture slides 1B and/or Ross's section 4.3.1. Be careful to express where your functional expression is valid. Where is the joint density zero and where is it nonzero?

(b) Calculate the probability that Ursula arrives before 12:20.
Use the distribution of U only.

(c) Find geometrically the probability that Ursula arrives before 12:15 and Vera arrives between 12:30 and 12:45.
Color the relevant rectangle in the square $[0, 60] \times [0, 60]$. Calculate the area of this rectangle. Recall the joint density and apply the basic fact from multivariate calculus.

(d) Calculate the probability in (c) again, without resorting to geometry, from the probabilities of the individual events $\{U < 15\}$ and $\{30 < V < 45\}$. Exploit the independence of U and V.

(e) Find geometrically the probability that Ursula arrives at least 30 minutes after Vera.
You need a bit of reasoning to find which region contains the points corresponding to the event. Try fixing a value of V, and think what must U then be for the event to occur. After you have found and drawn the region, calculate its area. Recall the area of a triangle.

Solution.

(a) Both U and V have the uniform distribution over $[0, 60]$, which has density

$$f_U(x) = f_V(x) = \begin{cases} \frac{1}{60}, & 0 < x < 60, \\ 0, & \text{otherwise.} \end{cases}$$

Because U and V are independent, the joint density function is the product $f_{U,V}(x, y) = f_U(x) f_V(y)$, thus

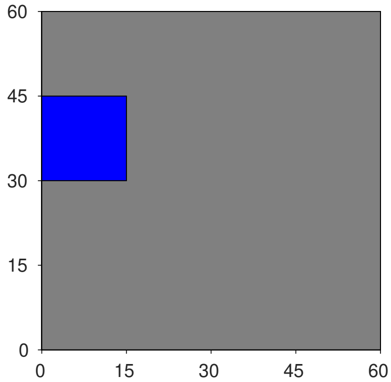
$$f_{U,V}(x, y) = \begin{cases} \frac{1}{3600}, & 0 < x < 60,\ 0 < y < 60, \\ 0, & \text{otherwise.} \end{cases}$$

This joint distribution is the uniform distribution over the square $[0, 60] \times [0, 60]$. Observe that the joint density is constant (uniform) everywhere in the square, but zero outside the square.

(b) The probability is

$$\int_0^{20} f_U(x)\,dx \;=\; 20 \times \frac{1}{60} \;=\; \frac{1}{3}.$$

(c) The event corresponds to the square $A = [0, 15] \times [30, 45]$ colored blue in this figure. If (U, V) is anywhere in the blue square, then the event has occurred.



Now the area of A is $15 \times 15 = 225$, and the joint density function has constant value $1/3600$ everywhere in A, so the integral of the joint density over A is

$$225 \times \frac{1}{3600} = \frac{1}{16} = 6.25\%.$$

In fact, the probability we found is the *ratio* between two areas; that of the small square, and that of the big square. Think: why is that true?

(d) Because U and V are independent,

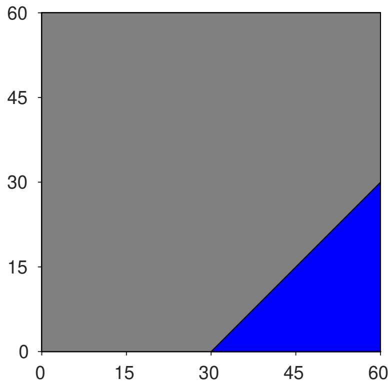
$$\mathrm{P}(U < 15 \,\textbf{and}\; 30 < V < 45) \;=\; \mathrm{P}(U < 15) \times \mathrm{P}(30 < V < 45) \;=\; \frac{1}{4} \times \frac{1}{4} \;=\; \frac{1}{16}.$$

The result is the same as in (c), as it should be.

(e) The event can be written as $\{(U, V) \in A\}$, where

$$A = \{(x, y) : x \geq y + 30\}.$$

If the region A is not immediately clear, we can start reasoning as follows. At height $y = 0$, the area contains the points where $x \geq 30$. At height $y = 10$ it contains the points where $x \geq 40$, and so on. In fact, the area A appears to the right from the straight line $x = y + 30$ (equivalently, $y = x - 30$).



From the figure we see that A is a triangle with base 30 and height 30. Its area is then $\frac{1}{2} \times 30^2 = 450$. Because the joint density is constant over A, the probability is

$$P(U \geq V + 30) \;=\; 450 \times \frac{1}{3600} = 1/8 = 12.5\%.$$

Home problems

2A3 (Device lifetime) A satellite orbiting the Earth contains a device whose lifetime X (in years) has the *exponential distribution* with parameter $\lambda = 0.5$. Find its density function from lecture slides (or look up Ross's section 5.6).

(a) Show how the cumulative distribution function $F(t)$ is obtained by integrating the density.

(b) Directly from the CDF, calculate the probabilities of the events $A = \{X \leq 1\}$, $B = \{X > 5\}$, and $C = \{5 < X \leq 6\}$. Give the results with at least 6 decimals. Explain in words what this events are.

(c) From the numerical results in (b), and using the definition of conditional probability, calculate $P(C|B)$. Compare to $P(A)$ and explain.

(d) Consider a very short interval of $h = 0.01$ years. If the device has lasted to a certain point in time, what is the probability that it breaks during the *next* 0.01 years? Compare your numerical result to the value of λh, and explain why λ is called the *rate parameter*.

Grading. 0.5 points per item. Total points rounded up to an integer.

Solution. The density function is $\lambda e^{-\lambda x}$ when $x > 0$, and zero when $x \leq 0$.

(a) At points $t > 0$, by integrating the density $f(t)$ we get the CDF at t as

$$F(t) \;=\; \int_{-\infty}^{t} f(s)\, ds \;=\; \int_{0}^{t} \lambda e^{-\lambda s}\, ds \;=\; 1 - e^{-\lambda t}.$$

At points $t \leq 0$ the CDF is $F(t) = 0$. Altogether

$$F(t) \;=\; \begin{cases} 1 - e^{-\lambda t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

(b)

$$\begin{aligned} P(A) &= P(X \leq 1) = F(1) = 1 - e^{-0.5 \cdot 1} \;\approx\; 0.393470 \\ P(B) &= P(X > 5) = 1 - F(5) = e^{-0.5 \cdot 5} \;\approx\; 0.082085 \\ P(C) &= P(5 < X \leq 6) = F(6) - F(5) = (1 - e^{-0.5 \cdot 6}) - (1 - e^{-0.5 \cdot 5}) \;\approx\; 0.032298 \end{aligned}$$

Explanation:

- A = the device breaks during its first year
- B = the device lasts over 5 years
- C = the lifetime is between 5 and 6 years; or equivalently: the device lasts over 5 years **and** breaks during the next year.

(c) Observe that $C \subseteq B$, so $\mathrm{P}(C \cap B) = \mathrm{P}(C)$.

$$\mathrm{P}(C|B) = \frac{\mathrm{P}(C \cap B)}{\mathrm{P}(B)} \approx \frac{0.032298}{0.082085} \approx 0.393470.$$

Your last decimals may vary depending on how you rounded the intermediate values. The value 0.393470 was computed from intermediate values in full machine precision.

We observe that $\mathrm{P}(C|B) = \mathrm{P}(A)$. That is, the device had this probability of breaking during the *first year*; but if it lasts five years, then it has the *same* probability of breaking during the *next year*. For this reason, the exponential distribution is called *memoryless*. This property is very special to this distribution; you can easily imagine other situations and distributions (perhaps of a lifetime of a device) that are not memoryless, for example, if old parts are worn out and more likely to break.

(d) Either applying the conditional probabilities, or by the memorylessness property, we observe it does not matter how long the device has lasted: the probability of breaking during the next $h = 0.01$ years is the same as during the first $h = 0.01$ years, namely

$$\mathrm{P}(X \leq 0.01) = 1 - e^{0.5 \cdot 0.01} \approx 0.0049875.$$

This is very close to $\lambda h = 0.5 \cdot 0.01 = 0.005$. The parameter $\lambda = 0.5$ characterizes the *rate* of failures per time interval; if the device is still functioning, then during a short interval h it fails with probability λh.

2A4 (Fuzzy logic.) In fuzzy logic, a proposition (a claim that something is true) has, instead of a binary truth value (0 or 1), a real-valued truth value in the interval $[0, 1]$. Researchers are modelling a particular proposition's truth value with the random variable X that has density

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

(a) Determine the constant c. The density must integrate to 1.

(b) Determine the cumulative distribution function of X, and draw it.

(c) Calculate the probability that the truth value is at least 0.75.

(d) Find the *mode* of the distribution, that is, the point x where the density $f(x)$ attains its maximum.

Grading. 0.5 points per item. If constant c was not found in (a), items (b)–(d) still receive points if solved (up to the unknown constant). In (b), points are given even if the answer does not explicitly tell where the CDF has values zero and one. Total points rounded up to integer.

Solution.

(a) The integral of the density (over the full range) must equal 1, so we require

$$1 = \int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 cx(1-x)\, dx.$$

Because

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

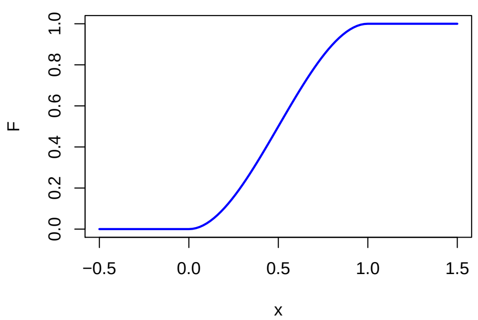
we thus require $c = 6$.

(b) Because X never takes values outside $[0,1]$, we have $F(x) = 0$, when $x < 0$, and $F(x) = 1$, when $x > 1$. In points $0 \leq x \leq 1$ the CDF is obtained by integrating density

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_0^x f(t)\, dt = 6 \int_0^x t(1-t)\, dt = 6 \cdot \left[\frac{1}{2}t^2 - \frac{1}{3}t^3\right]_0^x = x^2(3-2x).$$

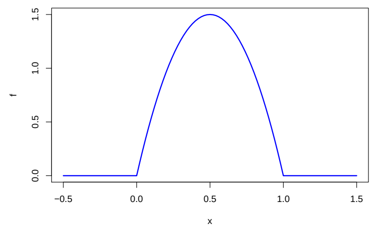
Thus the cumulative distribution function is

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2(3-2x), & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$$



(c) By the complement rule, this probability is $P(X \geq 0.75) = 1 - P(X < 0.75) = 1 - F(0.75) = 1 - (0.75)^2 \cdot (3 - 2 \cdot 0.75) \approx 15.6\%$.

(d) Outside the interval $[0,1]$ the density is zero, so the maximum must be found within the interval. On this interval, the density is a downward-opening parabola $f(x) = -6x^2 + 6x$, that goes to zero at the endpoints. The maximum is found at the point where the derivative f' is zero.



On the interval $[0, 1]$ we have

$$f'(x) = -12x + 6,$$

which is zero when $x = 1/2$, so the maximum is here. Thus the *mode* of this distribution is $1/2$.