Foundations of data science, summer 2020

JONATHAN LENNARTZ, MICHAEL NÜSKEN, ANNIKA TARNOWSKI

1. Exercise sheet
Hand in solutions until Thursday, 23 April 2020, 12:00

A word on the exercises. They are important. Of course, you know that. You need 50% of the credits for being well-prepared for the final exam.

The following exercises on this sheet are not graded. You may achieve credits by handing in any solution. We will discuss solutions in the tutorial.

Exercise 1.1 (Random variables). (8 points)

(i) Consider the course' example: A player has a fair die D resulting in one of $\mathcal{D} :=$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅} with probability $\frac{1}{6}$ each, ie. $D \xleftarrow{\text{⚄}} \mathcal{D}$ uniformly, and a fair coin $C \xleftarrow{\text{⚄}} \{0, 1\}$ uniformly.

Actually, our player is a cheater and the coin merely tells whether he decides to cheat or not. In case the coin comes up heads, say that's encoded 1, he changes the die's outcome to ⚅. Denote the variable describing the faked die by F.

(a) Describe F as a function of C and D. $\boxed{2}$

Solution. We have (if we identify the die's results with the numbers 1 to 6)

$$F = \begin{cases} D & \text{if } C = 0 \\ 6 & \text{else.} \end{cases}$$

(b) Note that the pair (C, F) is again a random variable, namely with outputs in $\{0, 1\} \times \mathcal{D}$. Write down a 2×6-table with its distribution. $\boxed{2}$

Solution. We obtain for all combinations of $C \in \{0, 1\}$ and $F \in \{1, 2, 3, 4, 5, 6\}$ that the probability of $\operatorname{prob}((C, F) = (c, f))$ takes the values of:

(c, f)	$f = 1$	2	3	4	5	6
$c = 0$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
1	0	0	0	0	0	$\frac{1}{2}$

$\boxed{2}$ (c) Prove that (C, F) is not independent.

Solution. If the variables were independent, it should hold that for all $c \in \{0, 1\}$, $f \in \{1, 2, 3, 4, 5, 6\}$ we would have

$$\text{prob}\,(C = c \wedge F = f) = \text{prob}\,(C = c) \cdot \text{prob}\,(F = f)\,.$$

Now we can choose one of the many counterexamples:

$$\text{prob}\,(C = 1 \wedge F = 1) = 0 \neq \frac{1}{24} = \text{prob}\,(C = 1) \cdot \text{prob}\,(F = 1)\,,$$

or just as well

$$\text{prob}\,(C = 1 \wedge F = 6) = \frac{1}{2} \neq \frac{7}{24} = \text{prob}\,(C = 1) \cdot \text{prob}\,(F = 6)\,.$$

(ii) Consider a continuous random variable U with outcomes in $[0, 1] \subset \mathbb{R}$ with uniform density, ie. $U \xleftarrow{} [0, 1]$ with density $p(x) = 1$ for $x \in [0, 1]$ and $p(x) = 0$ otherwise.

$\boxed{2}$ Determine the density of U^2, ie. a function q such that

$$\text{prob}\,\left(a < U^2 < b\right) = \int_a^b q(x)\ \mathrm{d}x\,.$$

Solution. A nice way to see this is to consider the function $Q(x) = \text{prob}\,(U^2 < x)$. As it holds that

$$\int_a^b q(x)\ \mathrm{d}x = Q(b) - Q(a)$$

we know that Q must be the anti derivative of q. We compute

$$Q(x) = \text{prob}\,\left(U^2 < x\right) = \begin{cases} 0 & x \leq 0 \\ \text{prob}\,(U < \sqrt{x}) = \sqrt{x} & 0 < x \leq 1 \\ \text{prob}\,(U < \sqrt{x}) = 1 & \text{else.} \end{cases}$$

Now we can compute

$$q(x) = \text{prob}\left(U^2 < x\right) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2}x^{-\frac{1}{2}} & 0 < x \leq 1 \\ 0 & \text{else.} \end{cases}$$

○