Foundations of data science, summer 2020

JONATHAN LENNARTZ, MICHAEL NÜSKEN, ANNIKA TARNOWSKI

2. Exercise sheet
Hand in solutions until Thursday, 30 April 2020, 12:00

Exercise 2.1 (When is Markov sharp?). (4 points)

Show that for any $a > 0$ there exists a probability distribution such that the Markov inequality is sharp, ie. $\text{prob}\,(X \geq a) = \frac{\mathrm{E}(X)}{a}$. Use the $\text{prob}\,(\ldots)$-notation to write down the distribution explicitly. $\boxed{4}$

Hint: Recall the proof from the lecture. If it helps you, restrict to the discrete setting.

Solution. We use that $\text{prob}\,(X \in I) = E(\mathbb{1}_{\{X\in I\}})$ since

$$\begin{aligned}\mathrm{E}(\mathbb{1}_{\{X\in I\}}) &= \sum_{i=0}^{1} i \cdot \text{prob}\left(\mathbb{1}_{\{X\in I\}} = i\right)\\ &= 1 \cdot \text{prob}\left(\mathbb{1}_{\{X\in I\}} = 1\right)\\ &= \text{prob}\,(X \in I)\,.\end{aligned}$$

Applied to the equation, we find that the following are equivalent:

$$\begin{gathered}\text{prob}\,(X \geq a) = \frac{\mathrm{E}(X)}{a}.\\ a \cdot \text{prob}\,(X \geq a) = \mathrm{E}(X).\\ \mathrm{E}(a \cdot \mathbb{1}_{\{X\geq a\}}) = \mathrm{E}(X \cdot \mathbb{1}_{\{X\geq 0\}}).\\ \mathrm{E}(X \cdot \mathbb{1}_{\{X\geq 0\}} - a \cdot \mathbb{1}_{\{X\geq a\}}) = 0.\\ \mathrm{E}(X \cdot \mathbb{1}_{\{0\leq X<a\}} + (X-a) \cdot \mathbb{1}_{\{X\geq a\}}) = 0.\\ \mathrm{E}(X \cdot \mathbb{1}_{\{0\leq X<a\}}) = 0 \quad \wedge \quad \mathrm{E}((X-a) \cdot \mathbb{1}_{\{X\geq a\}}) = 0.\\ \text{prob}\,(0 < X < a) = 0 \quad \wedge \quad \text{prob}\,(0 < X - a) = 0.\end{gathered}$$

For the last step we use: for any set M and any random variable Z with $Z \geq 0$ on M we have that $\mathrm{E}(Z \cdot \mathbb{1}_M) = 0$ implies $\text{prob}\,(Z \in M \setminus \{0\}) = 0$. As a consequence Markov is sharp iff $\text{prob}\,(X \in \{0, a\}) = 1$.

In other words, each distribution that satisfies above requirement is given by some $p \in [0, 1]$ and

$$\text{prob}\,(X = a) = p, \quad \text{prob}\,(X = 0) = 1 - p.$$

○

Exercise 2.2 (Maximum of two dies). (8 points)

Take two independent fair dies $D_i \xleftarrow{⚄⚄} \mathcal{D} := \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$ and consider the larger outcome

$$M := \max(D_0, D_1).$$

$\boxed{2}$ (i) Compute $\operatorname{prob}(M \leq a)$ for $a \in \mathbb{N}_{\leq 6}$.

Solution. We have by definition

$$\operatorname{prob}(M \leq a) = \operatorname{prob}(D_0 \leq a \wedge D_1 \leq a),$$

and as the dice are independent

$$\operatorname{prob}(D_0 \leq a \wedge D_1 \leq a) = \operatorname{prob}(D_0 \leq a) \cdot \operatorname{prob}(D_1 \leq a) = \frac{a}{6} \cdot \frac{a}{6} = \frac{a^2}{36}.$$

$\boxed{2}$ (ii) Compute $\operatorname{prob}(M = a)$ for $a \in \mathbb{N}_{\leq 6}$.

Solution. As the distribution is discrete (meaning: there are just finitely many outputs for M) we have

$$\operatorname{prob}(M = a) = \operatorname{prob}(M \leq a) - \operatorname{prob}(M \leq (a-1)),$$

such that we can insert from (i), that

$$\operatorname{prob}(M = a) = \frac{a^2}{36} - \frac{(a-1)^2}{36} = \frac{2a-1}{36}.$$

$\boxed{2}$ (iii) Compute $\mathrm{E}(M)$.

Solution. We can now sum up

$$\begin{aligned}
\mathrm{E}(X) &= 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} \\
&= \frac{1 + 6 + 15 + 28 + 45 + 66}{36} \\
&= \frac{161}{36} = 4.47\!\perp
\end{aligned}$$

$\boxed{2}$ (iv) Compute $\mathrm{E}(M^2)$ and derive $\mathrm{var}(M)$.[1]

Solution. We can similarly sum up

$$\begin{aligned}\mathrm{E}(X^2) &= 1 \cdot \frac{1}{36} + 4 \cdot \frac{3}{36} + 9 \cdot \frac{5}{36} + 16 \cdot \frac{7}{36} + 25 \cdot \frac{9}{36} + 36 \cdot \frac{11}{36} \\ &= \frac{1 + 12 + 45 + 112 + 225 + 396}{36} \\ &= \frac{791}{36} = 21.97\bot\end{aligned}$$

and derive the variance

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{764 \cdot 36 - (161)^2}{(36)^2} = \frac{2555}{1296} = 1.97\bot$$

This leads to the standard deviation $1.40\bot$. ○

Exercise 2.3 (Random variables). (0+8 points)

Consider the following experiment:

1. Throw a coin $C \xleftarrow{\text{⚄}} \{0, 1\}$.
2. Choose $X \xleftarrow{\text{⚄}} [0, 1]$ uniformly.
3. Roll a die $D \xleftarrow{\text{⚄}} \{1, 2, 3, 4, 5, 6\}$.
4. If $C = 0$ then
5. Let $Z \leftarrow D + X$.
6. Else
7. Let $Z \leftarrow D - X$.
8. Return Z

The output Z of this algorithm is a random variable.

(i) Compute its expection $\mathrm{E}(Z)$. $\boxed{+3}$

(ii) Compute its second moment $\mathrm{E}(Z^2)$. $\boxed{+4}$

(iii) Compute its variance $\mathrm{var}(Z)$. $\boxed{+1}$

[1] Side remark: to indicate how a real number was rounded we append a special symbol. Examples: $\pi = 3.14\bot = 3.142\bot = 3.1416\top = 3.14159\bot$. The height of the platform shows the size of the left-out part and the direction of the antenna indicates whether actual value is larger or smaller than displayed. We write, say, $e = 2.72\top = 2.71\sqcap$ as if the shorthand were exact.

Solution. Notice

- $\mathrm{E}(X) = \int_0^1 x \,\mathrm{d}x = \frac{1}{2}$,
- $\mathrm{E}(X^2) = \int_0^1 x^2 \,\mathrm{d}x = \frac{1}{3}$ and
- $\mathrm{E}(D) = \sum_{d\in\{1,2,3,4,5,6\}} d \operatorname{prob}(D = d) = \frac{21}{6} = \frac{7}{2}$,
- $\mathrm{E}(D^2) = \sum_{d\in\{1,2,3,4,5,6\}} x^2 \operatorname{prob}(D = d) = \frac{1+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$.

Thus conditioned on $C = 0$ we find

$$\mathrm{E}(Z \mid C = 0) = \mathrm{E}(D + X) = \mathrm{E}(D) + \mathrm{E}(X) = 4$$

and conditioned on $C = 1$ we find

$$\mathrm{E}(Z \mid C = 0) = \mathrm{E}(D - X) = \mathrm{E}(D) - \mathrm{E}(X) = 3.$$

Interludium *total probability*: We can always split probalities into pieces via conditional probabilities on the pieces of a partition of the universe: If $\Omega = \biguplus_i \Omega_i$ then

$$\operatorname{prob}(V \in \mathcal{A}) = \sum_i \operatorname{prob}(V \in \mathcal{A} \mid \Omega_i) \operatorname{prob}(\Omega_i)$$

This translates to expected values:

$$\mathrm{E}(V) = \sum_i \mathrm{E}(V \mid \Omega_i) \operatorname{prob}(\Omega_i).$$

You can easily check this by using the definitions. This is *the* tool to deal with case distinctions in programs in general.

(i) In the case at hand the interludium means that we simply compute

$$\mathrm{E}(Z) = \frac{1}{2}\mathrm{E}(Z \mid C = 0) + \frac{1}{2}\mathrm{E}(Z \mid C = 1).$$

and so

$$\mathrm{E}(Z) = \mathrm{E}(D) = \frac{7}{2}.$$

(ii) Similarly we get

$$\begin{aligned}
\mathrm{E}(Z^2) &= \mathrm{E}((D+X)^2)\,\text{prob}\,(C=0) + \mathrm{E}((D-X)^2)\,\text{prob}\,(C=1) \\
&= \frac{1}{2}\left(\mathrm{E}(D^2) + 2\,\mathrm{E}(D)\,\mathrm{E}(X) + \mathrm{E}(X^2)\right) \\
&\quad + \frac{1}{2}\left(\mathrm{E}(D^2) - 2\,\mathrm{E}(D)\,\mathrm{E}(X) + \mathrm{E}(X^2)\right) \\
&= \mathrm{E}(D^2) + \mathrm{E}(X^2) = \frac{91}{6} + \frac{1}{3} = \frac{31}{2}.
\end{aligned}$$

(iii) Consequently, $\mathrm{var}(Z) = \mathrm{E}(Z^2) - \mathrm{E}(Z)^2 = \frac{31}{2} - \left(\frac{7}{2}\right)^2 = \frac{13}{4}$. ○