Foundations of data science

AMIT RANA, BASTIAN SCHNEIDER, GERD MUND, PENELOPE MÜCK,
JONATHAN LENNARTZ, ANNIKA TARNOWSKI & MICHAEL NÜSKEN

1. Exercise sheet
Handin as announced on eCampus

fits to 01

Exercise 1.1 (Computer scientist's random variables). ($\boxed{8}$ points)

Consider the following algorithm:

1. Throw a coin $C \xleftarrow{⚄} \{0, 1\}$.
2. Choose $X \xleftarrow{⚄} [0, 1]$ uniformly.
3. Roll a die $D \xleftarrow{⚄} \{1, 2, 3, 4, 5, 6\}$.
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 $\mathrm{var}(Z)$. $\boxed{1}$

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

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

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

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

This translates to expected values:

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

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

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

and so

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

(ii) Similarly we get

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

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

fits to 02

Exercise 1.2 (Apply Markov's and Chebyshev's inequalities). (12 points)

In this exercise you shall test the quality of Markov's and Chebyshev's inequalities.

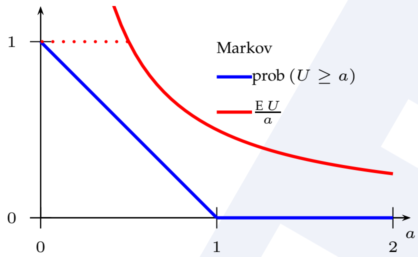
(i) Consider a uniform real number from the unit interval, ie. $U \xleftarrow{\text{🎲}} [0, 1]$ uniformly chosen.

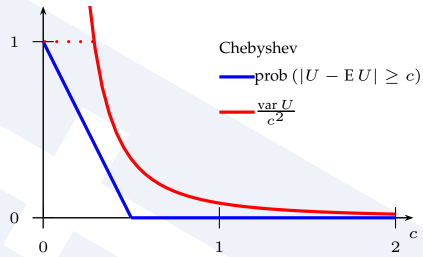
1 (a) Plot prob $(U \geq a)$ and $\frac{\mathrm{E}\,U}{a}$ as functions of a.

1 (b) Plot prob $(|U - \mathrm{E}\,U| \geq c)$ and $\frac{\operatorname{var} U}{c^2}$ as functions of c.

2 (c) Compare and interpret.

Solution.





We can see that the Markov bound for the variable X is very loose and does not contain much information. For example, in the interval $[0, \frac{1}{2}]$ the values of $\mathrm{E}(U)$ are larger than 1, even if we know that all probabilities are naturally bounded by 1. The tail bound is also extremely loose, as even if $X > 1$ is impossible, we still see a significant value of the Markov bound there.

The Chebychev inequality is also quite meaningless for small values of c. Even for c close to 0.5, which is the end of the range that c can take, we still have a lot of difference between the bound and the actual function. The tail bounds here are much better than in the Markov case and tend to 0 very quickly. ○

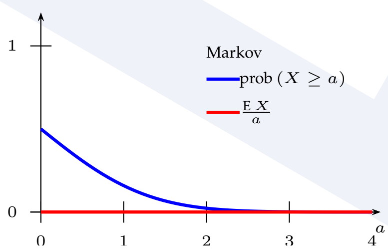
(ii) Consider a normally distributed real number with mean 0 and variance 1, ie. $X \longleftarrow \mathcal{N}(0, 1)$.

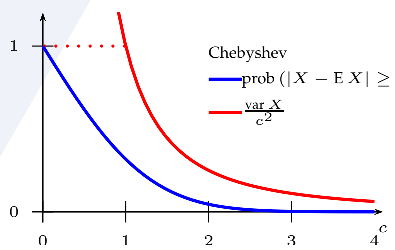
(a) Plot prob $(X \geq a)$ and $\frac{\mathrm{E}\,X}{a}$ as functions of a. 1

(b) Plot prob $(|X - \mathrm{E}\,X| \geq c)$ and $\frac{\operatorname{var} X}{c^2}$ as functions of c. 1

(c) Compare and interpret. 2

Solution.





The most important thing to note is that Markov's inequality does not apply here. This is due to the fact, that X is a random variable that can take negative values. Markov is only applicable for non-negative random variables.

Chebychev's inequality yields a decent bound, except for very small c.

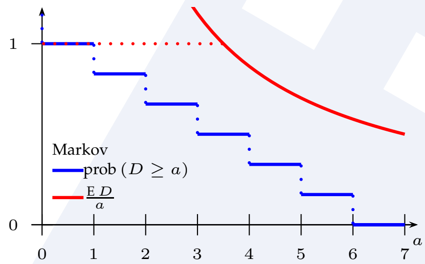
(iii) Consider a fair die, ie. $D \xleftarrow{\text{die}} \{1, 2, 3, 4, 5, 6\}$ with uniform distribution.

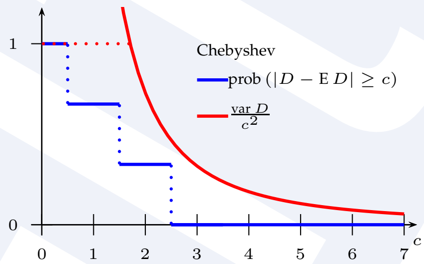
1 (a) Plot $\operatorname{prob}(D \geq a)$ and $\frac{\operatorname{E} D}{a}$ as functions of a.

1 (b) Plot $\operatorname{prob}(|D - \operatorname{E} D| \geq c)$ and $\frac{\operatorname{var} D}{c^2}$ as functions of c.

2 (c) Compare and interpret.

Solution.





We see again that the Markov inequality is a quite rough bound. The tail bound of Chebychev is better, but it's still not very good.

fits to 02

Exercise 1.3 (When is Markov sharp?). (0+4 points)

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

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

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

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

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

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