Foundations of data science

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

2. Exercise sheet
Handin as announced on eCampus

fits to 03

Exercise 2.1 (Prove a Chernoff bound). (0+10 points)

In several steps you shall prove the

Theorem (Chernoff, positive part). *Let $X = \sum_{i<n} X_i$ with independent, p-biased Bernoulli variables $X_i \xleftarrow{\text{iid}} \{0,1\}$ with* prob $(X_i = 1) = p$.

Let $\varepsilon \in \left]0, \frac{1}{2}\right[$. Then for some constant $c > 0$ we have

$$\text{prob}\left(X - \mathrm{E}\, X \geq \varepsilon\, \mathrm{E}\, X\right) \leq e^{-c\varepsilon^2\, \mathrm{E}\, X}.$$

(i) Compute $\mathrm{E}\, X$. +1

Solution. By the linearity of the expected value, we have that

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i).$$

A simple computation shows that $\mathrm{E}(X_i) = p$, such that

$$\mathrm{E}(X) = \sum_{i=1}^{n} p = np.$$ ○

(ii) For $\alpha > 0$ compute $\mathrm{E}\, e^{\alpha X}$. +3
[Do not forget to reason!]

Solution. At first, one can compute

$$e^{\alpha X} = e^{\alpha \sum_{i=1}^{n} X_i} = \prod_{i=1}^{n} e^{\alpha X_i}.$$

Now as the X_i are independent, so are the variables $e^{\alpha X_i}$. Hence

$$\mathrm{E}(e^{\alpha X}) = \mathrm{E}\left(\prod_{i=1}^{n} e^{\alpha X_i}\right) = \prod_{i=1}^{n} \mathrm{E}\left(e^{\alpha X_i}\right).$$

It remains to compute

$$\begin{aligned}\mathrm{E}\left(e^{\alpha X_i}\right) &= \mathrm{prob}\left(X_i = 1\right) \cdot e^{\alpha \cdot 1} + \mathrm{prob}\left(X_i = 0\right) \cdot e^{\alpha \cdot 0} \\ &= p \cdot e^{\alpha} + (1-p) \cdot 1 = 1 + p(e^{\alpha} - 1),\end{aligned}$$

so that overall we obtain

$$\mathrm{E}\left(e^{\alpha X}\right) = \prod_{i=1}^{n} (1 + p(e^{\alpha} - 1)) = (1 + p(e^{\alpha} - 1))^n$$

○

+1 (iii) Use $1 + x \leq e^x$ to estimate $\mathrm{E}\, e^{\alpha X}$ by $e^{\beta np}$ for some β.

Solution. If we use $x = p(e^{\alpha} - 1)$ and apply the inequality, we obtain

$$1 + p(e^{\alpha} - 1) \leq e^{p(e^{\alpha} - 1)}.$$

Hence

$$\mathrm{E}(e^{\alpha X}) = (1 + p(e^{\alpha} - 1))^n \leq (e^{p(e^{\alpha} - 1)})^n = e^{np(e^{\alpha} - 1)},$$

so for $\beta = (e^{\alpha} - 1)$ we obtain the desired result. ○

+3 (iv) Prove that $e^{\alpha} - 1 - \alpha(1 + \varepsilon) \leq -\frac{1}{3}\varepsilon^2$ for a suitable α and for each $\varepsilon \in \left]0, \frac{1}{2}\right[$.

Solution. We start by noting, that $e^{\alpha} - 1 - \alpha(1 + \varepsilon) \leq -\frac{1}{3}\varepsilon^2$ is equivalent to

$$e^{\alpha} - 1 - \alpha(1 + \varepsilon) + \frac{1}{3}\varepsilon^2 \leq 0.$$

If we consider the left hand side as a function of α, we can compute that the minimum is obtained at $\alpha = \ln(1 + \varepsilon)$. Because of this we will consider $\alpha = \ln(1 + \varepsilon)$ and it remains to show that for any ε in question

$$f(\varepsilon) = e^{\ln(1+\varepsilon)} - 1 - \ln(1 + \varepsilon)(1 + \varepsilon) + \frac{1}{3}\varepsilon^2 = \varepsilon - \ln(1 + \varepsilon)(1 + \varepsilon) + \frac{1}{3}\varepsilon^2 \leq 0.$$

We can compute that $f(0) = 0$, so our aim will be to show, that $f'(\varepsilon) \leq 0$ for $\varepsilon \in \left]0, \frac{1}{2}\right[$ such that the function $f(\varepsilon)$ will decrease (or stay constant) and hence be 0 or negative in the given interval.

Computing f' yields

$$f'(\varepsilon) = 1 - \frac{1 + \varepsilon}{1 + \varepsilon} - \ln(1 + \varepsilon) + \frac{2}{3}\varepsilon = \frac{2}{3}\varepsilon - \ln(1 + \varepsilon).$$

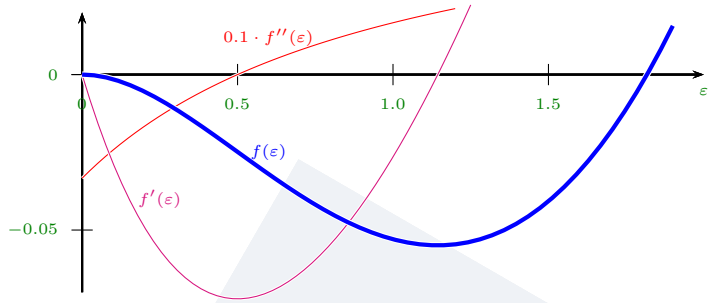
Again we have $f'(0) = 0$ and it's not immediate to see that $f'(\varepsilon) \leq 0$. But if we compute the second derivative, we have

$$f''(\varepsilon) = \frac{2}{3} - \frac{1}{1 + \varepsilon},$$

which is negative for $\varepsilon \in \left]0, \frac{1}{2}\right[$. So in this interval, $f'(\varepsilon)$ will be a decreasing function, and as it starts at 0, it will be negative. So also f will be a decreasing function and as it starts at 0 as well, it will be negative throughout $\varepsilon \in \left]0, \frac{1}{2}\right[$.

So indeed, the given inequality holds for $\alpha = \ln(1 + \varepsilon)$ and $\varepsilon \in \left]0, \frac{1}{2}\right[$. ○

For confirmation:



(v) Apply Markov's inequality for $Z = e^{\alpha X}$ to prove the theorem. +2

Solution. What we want to prove is that

$$\text{prob}\left(X - \mathrm{E}\,X \geq \varepsilon\,\mathrm{E}\,X\right) \leq e^{-c\varepsilon^2\,\mathrm{E}\,X}.$$

To transform the probability into a form such that we can apply Markov, we calculate

$$\text{prob}\left(X - \mathrm{E}\,X \geq \varepsilon\,\mathrm{E}\,X\right) = \text{prob}\left(X \geq (1+\varepsilon)\,\mathrm{E}\,X\right) = \text{prob}\left(e^{\alpha X} \geq e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}\right),$$

where the last equality holds as $\alpha > 0$, such that $e^{\alpha x}$ is a monotone increasing function in x.

But $Z = e^{\alpha X}$ is now a positively distributed random variable, such that we can apply Markov's inequality:

$$\text{prob}\left(e^{\alpha X} \geq e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}\right) = \text{prob}\left(Z \geq e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}\right) \leq \frac{\mathrm{E}(Z)}{e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}}.$$

But we have already estimated $\mathrm{E}(Z)$ and inserting this leads to

$$\frac{\mathrm{E}(Z)}{e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}} \leq \frac{e^{(e^\alpha - 1)\,\mathrm{E}\,X}}{e^{\alpha(1+\varepsilon)\,\mathrm{E}\,X}} = e^{((e^\alpha - 1) - (\alpha(1+\varepsilon)))\,\mathrm{E}\,X}.$$

Now we can use our inequality from the previous part and the fact that $e^{t\,\mathrm{E}(X)}$ is a positive monotone increasing function in t, to conclude that

$$e^{((e^\alpha - 1) - (\alpha(1+\varepsilon)))\,\mathrm{E}\,X} \leq e^{-\frac{1}{3}\varepsilon^2\,\mathrm{E}\,X} = e^{-c\varepsilon^2\,\mathrm{E}\,X},$$

for $c = \frac{1}{3}$. ○

You may want to consider the negative part, namely that also +0

$$\text{prob}\left(X - \mathrm{E}\,X \leq -\varepsilon\,\mathrm{E}\,X\right) \leq e^{-c\varepsilon^2\,\mathrm{E}\,X}.$$

Hint: $\alpha < 0$ does help.

fits to 03

Exercise 2.2 (Annuli). (7 points)

(i) Compute and estimate the volume of the $\frac{1}{100}$-annulus compared to the volume of the d-dimensional ball B^d. 1

Solution. By the formula from the lecture we know that the volume of the annulus of width ε of the d-dimensional ball, here denoted A^d_ε, is

$$\operatorname{vol}\left(A^d_\varepsilon\right) = \operatorname{vol}(B^d) - (1-\varepsilon)^d \operatorname{vol}(B^d).$$

So for $\varepsilon = \frac{1}{100}$ we obtain

$$\operatorname{vol}\left(A^d_{\frac{1}{100}}\right) = \operatorname{vol}(B^d) - \left(1 - \frac{1}{100}\right)^d \operatorname{vol}(B^d),$$

so compared to the d-dimensional ball we have

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{100}}\right)}{\operatorname{vol}(B^d)} = 1 - \left(1 - \frac{1}{100}\right)^d \geq 1 - e^{-\frac{1}{100}d}.$$

So especially we have

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{100}}\right)}{\operatorname{vol}(B^d)} \xrightarrow{d\to\infty} 1.$$

○

1 (ii) Compute and estimate the volume of the $\frac{1}{\sqrt{d}}$-annulus compared to the volume of the d-dimensional ball B^d.

Solution. With the same formula as above, for $\varepsilon = \frac{1}{\sqrt{d}}$ we obtain

$$\operatorname{vol}\left(A^d_{\frac{1}{\sqrt{d}}}\right) = \operatorname{vol}(B^d) - \left(1 - \frac{1}{\sqrt{d}}\right)^d \operatorname{vol}(B^d),$$

so compared to the d-dimensional ball we have

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{\sqrt{d}}}\right)}{\operatorname{vol}(B^d)} = 1 - \left(1 - \frac{1}{\sqrt{d}}\right)^d \geq 1 - e^{-\frac{1}{\sqrt{d}}d} = 1 - e^{-\sqrt{d}}.$$

So especially, we have

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{\sqrt{d}}}\right)}{\operatorname{vol}(B^d)} \xrightarrow{d\to\infty} 1.$$

○

1 (iii) Compute and estimate the volume of the $\frac{1}{d^2}$-annulus compared to the volume of the d-dimensional ball B^d.

Solution. With the same formula as above, for $\varepsilon = \frac{1}{d^2}$ we obtain

$$\operatorname{vol}\left(A^d_{\frac{1}{d^2}}\right) = \operatorname{vol}(B^d) - \left(1 - \frac{1}{d^2}\right)^d \operatorname{vol}(B^d),$$

so compared to the d-dimensional ball we have

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{d^2}}\right)}{\operatorname{vol}(B^d)} = 1 - \left(1 - \frac{1}{d^2}\right)^d \geq 1 - e^{-\frac{1}{d^2}d} = 1 - e^{-\frac{1}{d}}.$$

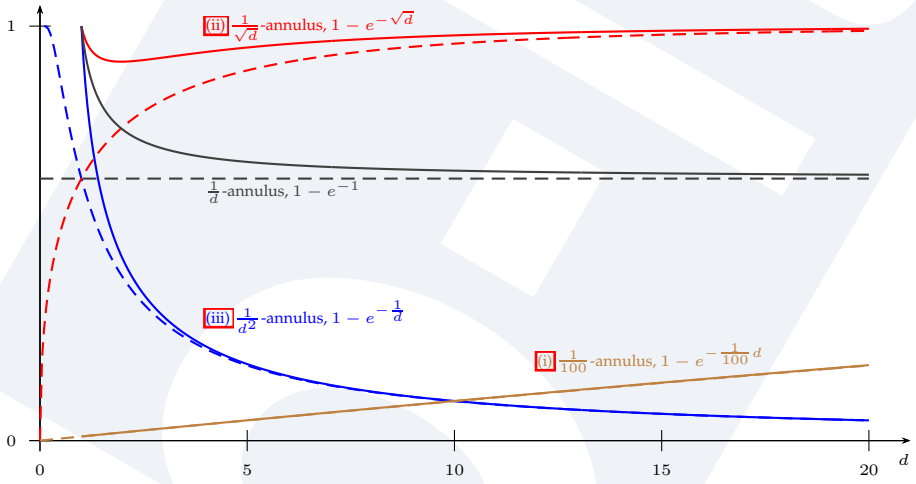
Actually applying an appropriate upper bound we find

$$\frac{\operatorname{vol}\left(A^d_{\frac{1}{d^2}}\right)}{\operatorname{vol}(B^d)} \xrightarrow{d\to\infty} 0.$$

○

(iv) Plot the three functions and the relative volume of the $\frac{1}{d}$-annulus for $d = 1..20$. [2]

Solution.



Here, the true functions are drawn as solid lines, the estimates as dashed lines. ○

(v) For which ε does the ε-annulus have at least 99% of the ball volume? [2]

Solution. If we consider the previous picture, the graph in which we have control of both lower and higher dimensions, is the graph of $\varepsilon = \frac{1}{d}$. The idea now is to consider $\varepsilon = \frac{c}{d}$ to get a different constant. We compute that the volume of the annulus compared to the ball itself is

$$(1 - \left(1 - \frac{c}{d}\right)^d) \geq 1 - e^{-c}.$$

To obtain $1 - e^{-c} \geq 0.99$, we need to solve $e^{-c} \leq 0.01$. As the exponential function is monotone increasing, this is equivalent to

$$-c \leq \ln(0.01) \Leftrightarrow c \geq \ln(100).$$

So especially for $c = \ln(100) = 4.6051$ we have that an $\varepsilon = \frac{c}{d}$-annulus contains more than 99% of the volume of any d-dimensional ball. ○

fits to 04

Exercise 2.3 (Gamma). (8+4 points)

4 (i) Prove that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Hint: Change the variable and use $\int_{-\infty}^{\infty} e^{-x^2}\ \mathrm{d}x = \sqrt{\pi}$.

Solution. We are combining a few elementary transformations on integrals to obtain that. It needs one 'standard' substitution and one trick.

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right) &= \int_0^{\infty} x^{\frac{1}{2}-1} e^{-x}\ \mathrm{d}x \\ &\overset{\substack{x=y^2,\\ \mathrm{d}x=2y\ \mathrm{d}y}}{=\!=\!=\!=\!=\!=\!=\!=} \int_0^{\infty} y^{-1} e^{-y^2} 2y\ \mathrm{d}y \\ &= \int_{-\infty}^{\infty} e^{-y^2}\ \mathrm{d}y = \sqrt{\pi}.\end{aligned}$$

○

+4 (ii) Prove that $\int_{-\infty}^{\infty} e^{-x^2}\ \mathrm{d}x = \sqrt{\pi}$.

Hint: To that end consider $\int_{\mathbb{R}^2} e^{-(y^2+z^2)}\ \mathrm{d}y\ \mathrm{d}z$. Use the substitution $y = r\cos(\varphi)$, $z = r\sin(\varphi)$ to express this with polar coordinates.

Solution. For computing this integral the simplest way I know is to square it:

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-y^2}\ \mathrm{d}y \cdot \int_{-\infty}^{\infty} e^{-z^2}\ \mathrm{d}z &= \int_{\mathbb{R}^2} e^{-(y^2+z^2)}\ \mathrm{d}y\ \mathrm{d}z \\ &\overset{\substack{(y,z)=(r\cos\varphi, r\sin\varphi)\\ \mathrm{d}y\ \mathrm{d}z = r\ \mathrm{d}r\ \mathrm{d}\varphi}}{=\!=\!=\!=\!=\!=\!=\!=\!=\!=} \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\ \mathrm{d}r\ \mathrm{d}\varphi \\ &= \int_0^{\infty} 2re^{-r^2}\ \mathrm{d}r \cdot \frac{1}{2}\int_0^{2\pi} \mathrm{d}\varphi \\ &= \underbrace{\left[-e^{-r^2}\right]_0^{\infty}}_{=1} \cdot \pi \\ &= \pi\end{aligned}$$

and thus $\int_{-\infty}^{\infty} e^{-y^2}\ \mathrm{d}y = \sqrt{\pi}$. ○

4

(iii) Recall the formula for integration by parts (look it up if need be…), $\int_a^b f(x)g'(x)\,\mathrm{d}x = \ldots$, where f and g are any suitably nice functions. Use this formula to show that indeed $\Gamma(z+1) = z\Gamma(z)$.

Solution. We split the integrand into the parts x^z with derivate zx^{z-1} and e^{-x} with anti-derivative $-e^{-x}$:

$$\begin{aligned}\Gamma(z+1) &= \int_0^\infty \underbrace{x^{(z+1)-1}}_{=:f\downarrow}\underbrace{e^{-x}}_{=:g\uparrow}\,\mathrm{d}x \\ &= \underbrace{\left[-x^z e^{-x}\right]_0^\infty}_{=0} + z\int_0^\infty x^{z-1}e^{-x}\,\mathrm{d}x = z\Gamma(z).\end{aligned}$$

○

fits to 04

Exercise 2.4 (Simplices). (4 points)

4

Consider an n-simplex

$$\Delta_n = \left\{ x \in \mathbb{R}^n \,\middle|\, \exists x_i \in \mathbb{R}_{\geq 0} \colon x = \sum_{0\leq i<n} x_i e_i,\ \sum_{0\leq i<n} x_i \leq 1 \right\}.$$

Estimate its near surface volume $\Delta_n \setminus (c + (1-\varepsilon)(\Delta_n - c))$ with $c = \text{centroid}(e_0, \ldots, e_{n-1}, 0)$ relative to its volume.

Solution. From the lecture we know that for $A \subset \mathbb{R}^d, \varepsilon \in [0,1]$ with $(1-\varepsilon)A \subset A$

$$\begin{aligned}\mathrm{vol}((1-\varepsilon)A) &= (1-\varepsilon)^d \,\mathrm{vol}(A) \text{ and} \\ \mathrm{vol}(A \setminus (1-\varepsilon)A) &= \mathrm{vol}(A) - \mathrm{vol}((1-\varepsilon)A).\end{aligned}$$

To proof the claim shift the coordinate system such that the centroid c of the simplex is the

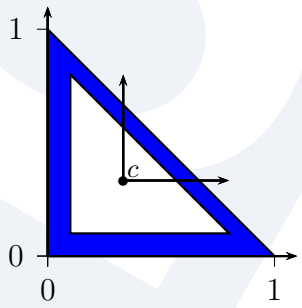


Figure 2.1: 2-simplex with centroid c

origin. In this shifted coordinate system $c = 0$ and we have $(c+(1-\varepsilon)(\Delta_n - c)) = (1-\varepsilon)\Delta_n$

and $(1-\varepsilon)\Delta_n \subset \Delta_n$. If we consider the relative volume and use the above equations we get

$$\begin{aligned}\frac{\mathrm{vol}(\Delta_n \setminus (c+(1-\varepsilon)(\Delta_n - c))}{\mathrm{vol}(\Delta_n)} &= 1 - \frac{\mathrm{vol}((1-\varepsilon)\Delta_n)}{\mathrm{vol}(\Delta_n)} \\ &= 1-(1-\varepsilon)^d \\ &\geq 1 - e^{-\varepsilon d}.\end{aligned}$$

○