MAD 2018-19, Assignment 1

Jonas Peters, Fabian Gieseke, Aasa Feragen

hand in until: 04.12.2018 at 23:59

General comments: The assignments in MAD must be completed and written individually. You are allowed (and encouraged) to discuss the exercises in small groups. If you do so, you are required to list your group partners in the submission. The report must be written completely by yourself. In order to pass the assignment, you will need to get at least 40% of the available points. The data needed for the assignment can be found in the assignment folder that you download from Absalon.

Submission instructions: Submit your report as a PDF, not zipped up with the rest. Please add your source code to the submission, both as executable files and as part of your report. To include it in your report, you can use the `lstlisting` environment in LaTeX, or you can include a "print to pdf" output in your pdf report.

Exercise 1 (4 points. Pdf and cdf). We model the life span of a chip (in years) with a distribution that has the following cumulative distribution function (cdf).

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - \exp(-(\beta x)^\alpha), & x > 0 \end{cases}$$

with $\alpha > 0$ and $\beta > 0$ being parameters.

a) Determine the probability density function (pdf) of the distribution.
b) Suppose $\alpha = 2$ and $\beta = \frac{1}{4}$. What is the prob. that the chip works longer than four years? What is the prob. that the chip stops working in the time interval $[5, 10]$ years?
c) How large is the median of a life span (for general α, β)?
d) Let X and Y be independent random variables, where X has cdf F with $\alpha = 2$ and $\beta = \frac{1}{4}$ and Y has cdf F with $\alpha = 3$ and $\beta = \frac{1}{6}$. Compute $\mathbb{P}(\min(X, Y) \leq 4)$.

Exercise 2 (4 points. Joint pdf and sampling). Consider a random vector with pdf $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ with

$$f_{X,Y}(x, y) = \begin{cases} 1/4 & \text{if } -1 \leq x, y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

a) Draw the graph of this density (by hand is fine).
b) Compute $\mathbb{P}(X^2 + Y^2 \leq 1)$. (Hint: recall that the two-dimensional integral of a function equals the volume under its graph.)
c) First, generate a sample of n iid realizations (X_1, Y_1), ..., (X_n, Y_n) of the random vector (X, Y) using `numpy.random.uniform`, for example. (You may use that X and Y are independent.) Then estimate $\mathbb{P}(X^2 + Y^2 \leq 1)$ by

$$\mathbb{P}(X^2 + Y^2 \leq 1) \approx \hat{\theta}_n := \frac{\#\{i \,|\, X_i^2 + Y_i^2 \leq 1\}}{n}.$$

Use $\hat{\theta}_n$ to estimate π and report the values for $n = 10000$, $n = 100000$ and $n = 1000000$. Show the code.

Exercise 3 (4 points. (from 22.11.) Estimating mean and variance). a) Let $X_1, \ldots, X_n$, $n > 3$, be iid random variables. Prove that the estimator $\hat{\theta}_{\text{stupid}} := \frac{1}{3} \sum_{i=1}^{3} X_i$ is an unbiased estimate of the true mean $\mathbb{E}X_1$.

b) Repeat the following two steps 1000 times: (1) generate iid realizations $X_1, \ldots, X_5$ from a Gaussian distribution $\mathcal{N}(0, 1)$ using `np.random.normal`. (2) Compute $\hat{\theta}_{\text{stupid}}$, the sample mean $\bar{X} = \sum_{i=1}^{5} X_i$, and the following estimator of the variance[1] $\frac{1}{5} \sum_{i=1}^{5} (X_i - \bar{X})^2$. Show your code.
c) Plot the 1000 values of $\hat{\theta}_{\text{stupid}}$ in black, $\bar{X}$ in red, and include a horizontal line for the true mean $0 = \mathbb{E}X_1$. Which one has smaller variance (according to the plot)?
d) Plot the 1000 values of the estimated variance and include a horizontal line for the true variance $1 = \text{var}(X_1)$. Can you spot a BIAS?

[1] In fact, this estimator is the maximum likelihood estimator of the variance.

e) In the years 2001–2018, Jeff Bezos estimated net worth (in March, in Billion Dollars) has increased by

$$\begin{array}{lllllll} X_1 = -4.1, & X_2 = -0.5, & X_3 = +1.0, & X_4 = +2.6, & X_5 = -1.0, & X_6 = +0.2, & X_7 = +4.4, \\ X_8 = -0.5, & X_9 = -1.4, & X_{10} = +5.8, & X_{11} = +5.5, & X_{12} = +5.1, & X_{13} = +5.7, & X_{14} = +1.6, \\ X_{15} = +19.8, & X_{16} = -5.1, & X_{17} = +27.6, & X_{18} = +39.2. & & & \end{array}$$

We now want to estimate how much his net worth will increase in 2019. Despite all the deliberations above, using $X_{18} = +39.2$ seems a better estimate than taking an average over the last 18 years. Why?

Exercise 4 (4 points. (from 27.11.) Maximum Likelihood). a) Let $X_1, \dots, X_n$ be i.i.d. with geometric distribution $\mathcal{G}eo(\theta)$. Compute the maximum likelihood estimator (MLE) for θ. (Hint: sometimes it is easier to maximize the log of a function than a function itself.)

b) Let $X_1, \dots, X_n$ be i.i.d. with uniform distribution $\mathcal{U}[1, b]$. Compute the MLE for b.