FIRST COURSE IN PROBABILITY AND STATISTICS (MS-A0501/2/3/4)
Department of Mathematics and Systems Analysis, Aalto University
Pekka Pere, Natalia Vesselinova, Jukka Kohonen

Exam 16.4.2024

You may use a scientific calculator namely, a calculator that can have operations for trigonometry, power, exponential function, logarithm, and binomial coefficients. However, no other type of calculator is allowed. In particular, you may not use: 1) programmable calculators, which can run program code; 2) symbolic calculators, which can manipulate symbolic expressions; 3) graphical calculators, which are able to plot functions; 4) calculators with Internet connection. You may use your handwritten double-sided A4 note at the exam.

Present the intermediate steps of all your calculations, and justify all your answers in detail. A correct answer alone is worth zero points. All tasks are worth six points. Carefully review all your answers. And once more, the best of luck!

1. Item a) 4 points. Item b) 2 points.

a) Let $\mathsf{P}(A) = 0.5$, $\mathsf{P}(B) = 0.3$, and $\mathsf{P}(A \text{ and } B) = 0.2$. What are the probabilities of the events below?

- Either A or B or both.
- A only (not B).
- At least A.
- At most A (at least not B).

b) Consider the following scenario:[1] John initially took a degree in mathematics, and followed it with a PhD in astrophysics. After that, he worked in the physics department of a university for a while but then found a job in the back room of an algorithmic trading company, developing highly sophisticated models for predicting movements of the financial markets. In his spare time he attends science fiction conventions.

Which of the following do you think has a higher probability?

- John is married with two children.
- John is married with two children, and likes to spend his evenings tackling mathematical puzzles and playing computer games.

[1] The exercise is essentially from D. Hand (2014): *The Improbability Principle.* Transworld. P. 215. The original version is due to Daniel Kahneman. See D. Kahneman (2011): *Thinking, Fast and Slow.* Penguin Books. Chapter 15.

2.

a) The average ecological body mass of a *Tyrannosaurus rex* is 5 200 kg, and the 0.025th and 0.975th quantiles of the distribution of the ecological body mass are 3 700 kg and 6 900 kg, respectively. Can we reason that the distribution of the ecological body mass of *Tyrannosaurus rex* is symmetric?

b) The average ecological body mass above refers to average body mass of *Tyrannosaurus rex*es of different ages. A full-grown *Tyrannosaurus rex* weighs on average 7 100 kg, and the 0.025. and 0.975. quantiles of the distribution of the body mass are 5 100 kg and 9 100 kg. Can we reason that the distribution of the body mass of *Tyrannosaurus rex* is symmetric?

c) Weight of a full-grown *Tyrannosaurus rex* follows the Normal distribution with parameters determined by the figures in item b). What is the standard deviation and variance of weight of a full-grown *Tyrannosaurus rex*?[2] (Hint: The 0.025th and 0.975th quantiles of the Standard Normal distribution are -1.959964 and 1.959964, respectively.)

3. The European Social Survey (ESS) was carried out in 2012. Face-to-face interviews were arranged for 1 700 Finns. 77.11765% or 1 311 of them told the interviewer that they had voted in the parliamentary election 2011. The actual voter turnout at the election was 70.5 %.

From the ESS survey results, calculate a two-sided 95% confidence interval for the voting percentage in the 2011 election. Does it cover the actual voter turnout percentage? (Hint: The 0.95th, 0.975th, 0.99th, and 0.995th quantiles of the Standard Normal distribution are 1.645, 1.960, 2.326, and 2.576, respectively.)

4. The call service requests that arrive at a mobile cellular network are assumed independent of each other and their mean rate (number of calls per time interval) λ is assumed constant. Then, the number of calls can be modelled by a Poisson distribution with parameter $\lambda > 0$. The probability mass function of a Poisson random variable X with a parameter λ, is given by:

$$p_X(x) = \mathsf{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!},$$

where $x = 0, 1, \ldots$ denotes the number of call requests received during an interval and λ takes on positive values.

A mobile operator registers in a certain location of their cellular network $x_1 = 10$, $x_2 = 5$, and $x_4 = 7$ calls during three independent time intervals of the same length. Knowing that the number of call requests per time interval follows a Poisson distribution, the operator wants to estimate their mean rate λ in order to improve the management of the network resources.

a) What is the maximum likelihood estimate (MLE) of the unknown parameter λ? (Hint: multiplicative constants can be ignored.)

b) The mobile operator's beliefs about λ can be described by an exponential distribution of the form

$$e^{-\lambda}.$$

[2] C.R. Marshall *et al.* (2021): Absolute Abundance and Preservation Rate of Tyrannosaurus Rex. *Science*, 372, 284–287. Especially Table 1 and part C of Figure 2.

Using this prior, the observed data and the assumed probabilistic model, formulate the unnormalised posterior (which is proportional to the posterior by a constant) and then find the maximum a posteriori (MAP) estimate of λ.

c) The mobile operator registers new data at the same location: $x_4 = 20$ calls during another time interval of the same fixed length. This additional observation is independent of the earlier observations $(x_1, \ldots, x_4)$ and follows the same Poisson distribution. Update the posterior (again, consider the unnormalised posterior) and use it to calculate the new MAP estimate of λ.