Final exam probability theory (WBMA046-05)

16 June 2021, 8.30 – 12.00

- ▷ Before the start of the exam, everybody taking part in the exam must sign the student declaration in the exam environment.
- ▷ To check for possible fraud, an unannounced sample of students will be contacted soon after the exam.
- ▷ The answers need to be written by hand, scanned and submitted within the time limit. You must upload your exam in a single pdf file.
- ▷ Every exercise needs to be handed in on a separate sheet.
- ▷ Write your name and student number **on every sheet**.
- ▷ It is forbidden to communicate with other persons during the exam, except with the course instructor.
- ▷ The only tools and aids that you are allowed to use are a non-programmable calculator (not a phone!), and the following material from the nestor course environment:
  a) The pdf file of the lecture notes (not videos, not scribbles).
  b) The pdf files of the tutorial problems.
  c) The pdf files of the homework problems.
  d) The pdf files of the solutions to the homework problems.
- ▷ Always give a short proof of your answer or a calculation to justify it, or clearly state the facts from the lecture notes or homework you are using.
- ▷ Phones are only allowed for the purpose of scanning the handwritten solutions.
- ▷ Simplify your final answers as much as possible.
- ▷ **NOTA BENE.** Using separate sheets for the different exercises, solving the exam corresponding to your student number, writing your name and student number on all sheets, and submitting all sheets in a single pdf is worth 10 out of the 100 points.

Problem 1 (a:6, b:6, c:4, d:4 pts).
You have a coin that shows tails with probability $p \in [0,1]$. For problems a) and b) you flip the coin until you have obtained either 2 heads or 2 tails (not necessarily in a row). Let Y denote the total number of flips.

a) Determine the pmf of Y and compute $\mathbb{E}[Y]$.

b) Compute $\mathsf{Var}(Y)$, and determine the value of $p \in [0,1]$ that maximizes $\mathsf{Var}(Y)$.

Now, you take the same coin but flip it precisely 20 times. Say that an *alternation* occurs at position $i \leqslant 19$ if the result of the ith flip is different from the result of the $(i+1)$th flip.

c) Compute the expected number of alternations during the 20 flips as a function of $p \in [0,1]$.

d) Now assume that $p = 1/2$. Compute the probability that there are exactly 13 alternations during the 20 flips.

Problem 2 (a:3, b:6, c:5, d:6 pts).

a) You are given two 6-sided dice. The first is a standard die labeled with the integers from 1 to 6, the second is a special die which has the integers 1 to 4, and then the 6 twice. You select one of the two dice at random.You roll the die three times and see ‘6’ in the first roll, ‘6’ in the second roll, and ‘1’ in the third roll. What is the probability that you picked the standard die?

b) Albert, Beatrix, Claudia and Daniel play a card game. A deck consists of 52 cards containing precisely 4 Queens. The cards are shuffled at random and each player receives 13 cards. You know that Beatrix has the Queen of Spades. What is the probability that each player receives precisely one Queen?

You have a box containing 10 pieces of dark chocolate and 80 pieces of milk chocolate. You empty the box in 45 turns where in each turn you remove at random two of the remaining pieces of chocolate. Let Y be the number of turns where you pick two pieces of dark chocolate.

c) Compute $\mathbb{P}(Y=0)$ and $\mathbb{P}(Y=1)$.

d) Compute $\mathbb{E}[Y]$ and $\mathsf{Var}(Y)$.

Problem 3 (a:5, b:6, c:6, d:6, e:7 pts)
Let X be a positive continuous random variable and Y be a discrete random variable with values in $\{0,1,\dots\}$. The joint pdf of a random vector (X,Y) equals

$$f_{X,Y}(x,k) := 9e^{-cx}\frac{x^{k+1}}{k!}.$$

You may use without proof that $\int_0^\infty x^{\alpha-1}e^{-\beta x}\mathrm{d}x = \Gamma(\alpha)/\beta^\alpha$ for $\alpha,\beta>0$.

a) Determine c.

b) Determine the marginal pdf's $f_X(x)$ and $f_Y(k)$.

c) Determine $\mathsf{Var}(X|Y=2)$ and $\mathsf{Var}(Y|X=3)$.

d) Let $(B_i)_{i\geqslant 1}$ be iid Bernoulli random variables with parameter $p\in[0,1]$ that are independent of Y. Compute the mgf of B_i and the mgf of $Y' := \#\{i\leqslant Y\colon B_i=1\}$.

e) Compute the pdf of $X+\min\{Y,1\}$.

Problem 4 (a:6, b:6, c:3, d:5 pts).
Let $t>0$ and $\{F_m\}_{m\geqslant 1}$ be independent $\mathcal{N}(0,t)$-distributed random variables. Say that the position $n\geqslant 2$ is a *peak* if $F_n > \max\{F_1,\dots,F_{n-1}\}$.

a) Show that $\mathbb{P}(n \text{ is a peak}) = 1/n$ and that $\mathbb{E}[\min\{n\geqslant 2\colon n \text{ is a peak}\}] = \infty$.

b) Let $M_n := \#\{2\leqslant j\leqslant n\colon j \text{ is a peak}\}$ be the number of peaks at positions at most n. Show that $\mathbb{E}[M_n] = \sum_{2\leqslant i\leqslant n} 1/i$ and $\mathsf{Var}(M_n) = \sum_{2\leqslant i\leqslant n}(i-1)/i^2$.

c) Show that $F_1^2+F_2^2$ has a pdf given by $f(x) = (2t)^{-1}\exp(-x/(2t))$.

d) Show that $\mathbb{P}\big(\lim_{n\to\infty}\sup_{m\geqslant n} F_m/\sqrt{\log m}\leqslant\sqrt{2t}\big)=1$. You may use without proof that if $Z\sim\mathcal{N}(0,1)$, then for all $z>1$, $\sqrt{2\pi}\mathbb{P}(Z>z)\leqslant e^{-z^2/2}$.

Solutions

Problem 1.a

- ▷ $\mathbb{P}(Y=2) = p^2 + (1-p)^2$, $\mathbb{P}(Y=3) = 2p(1-p)$.
- ▷ $\mathbb{E}[Y] = 2\mathbb{P}(Y=2) + 3\mathbb{P}(Y=3) = 2 + 2p(1-p)$.

Problem 1.b

- ▷ $\mathbb{E}[Y^2] = 4\mathbb{P}(Y=2) + 9\mathbb{P}(Y=3) = 4 + 10p(1-p)$.
- ▷ $\mathsf{Var}(Y) = 4 + 10p(1-p) - \big(4 + 8p(1-p) + 4p^2(1-p)^2\big) = 2p(1-p) - 4p^2(1-p)^2 = 2p(1-p)(1-2p(1-p))$.
- ▷ The latter expression is maximal if $2p(1-p) = 1/2$, i.e., if $p = 1/2$.

Problem 1.c

- ▷ By linearity of expectation, the expected number of alternations is given by

$$\mathbb{E}[\sum_{i\leqslant 19} \mathbb{1}\{i \text{ is alternation}\}] = \sum_{i\leqslant 19} \mathbb{P}(i \text{ is alternation}) = 38p(1-p).$$

Problem 1.d

- ▷ For $p = 1/2$, the alternations are binomially distributed with parameter 19 and 1/2.
- ▷ Hence, the desired probability is given by $\binom{19}{13}2^{-19}$.

Problem 2.a

- ▷ Apply Bayes' formula: $\mathbb{P}(\text{standard die} \,|\, \text{‘661‘}) = \frac{\mathbb{P}(\text{standard die}, \text{‘661‘})}{\mathbb{P}(\text{standard die}, \text{‘661‘}) + \mathbb{P}(\text{special die}, \text{‘661‘})} = \frac{0.5 \cdot 6^{-3}}{0.5\cdot 6^{-3} + 0.5\cdot 4\cdot 6^{-3}} = 0.2$

Problem 2.b

- ▷ Let $A := \{$ Beatrix has queen of spades$\}$; $B := \{$ each player receives precisely one queen$\}$.
- ▷ The queen of spades can be any one of Beatrix 13 cards, thus $\#A = 13\cdot 51!$.
- ▷ There are 3! ways to distribute the remaining queens; the queen can be any one of each player's 13 cards; thus $\#(A\cap B) = 3!13^4 48!$.
- ▷ Then, $\mathbb{P}(B\,|\,A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \frac{3!13^4 48!}{13\cdot 51!} = \frac{3!13^3}{49\cdot 50\cdot 51}$.

Problem 2.c

- ▷ In total, 10 dark chocolates can appear in $\binom{90}{10}$ positions.
- ▷ $\boldsymbol{Y=0}$. We need to choose 10 of the 45 turns, where the dark chocolates appear. In each turn there are 2 options to choose the position of the dark chocolate. Hence, $\mathbb{P}(Y=0) = \binom{45}{10}2^{10}/\binom{90}{10}$
- ▷ $\boldsymbol{Y=1}$. We need to choose 8 of the 45 turns, where a single dark chocolates appear and we need to choose 1 turn where two dark chocolates appear. In each turn there are 2 options to choose the position of the dark chocolate. Hence, $\mathbb{P}(Y=1) = \frac{45!}{8!1!36!}2^8/\binom{90}{10}$.

Problem 2.d

- ▷ Let A_i be the event that two pieces of dark chocolate are selected in round i so that $Y = \sum_i \mathbb{1}\{A_i\}$.
- ▷ Then, $\mathbb{P}(A_1) = \frac{10\cdot 9}{90\cdot 89}$ so that $\mathbb{E}[Y] = 45\mathbb{P}(A_1) = 45\frac{10\cdot 9}{90\cdot 89} = \frac{45}{89}$.
- ▷ The multilinearity of covariances and the identical distributions give that

$$\mathsf{Cov}(Y,Y) = \sum_{i\leqslant 45}\mathsf{Cov}(A_i,A_i) + \sum_{i\neq j\leqslant 45}\mathsf{Cov}(A_i,A_j) = 45(\mathbb{P}(A_1) - \mathbb{P}(A_1)^2) + 45\cdot 44\mathsf{Cov}(A_1,A_2).$$

▷ Thus, $\mathsf{Var}(Y) = 45(\mathbb{P}(A_1) - \mathbb{P}(A_1)^2) + 45 \cdot 44(\mathbb{P}(A_1)\frac{8\cdot 7}{88\cdot 87} - \mathbb{P}(A_1)^2)$.

Problem 3.a

▷ From $\int_0^\infty x^{\alpha-1}e^{-\beta x}\mathrm{d}x = \Gamma(\alpha)/\beta^\alpha$ we get that $1 = \int_0^\infty \sum_{k\geqslant 0} 9e^{-cx}\frac{x^{k+1}}{k!}\mathrm{d}x = \int_0^\infty 9e^{-cx+x}x\mathrm{d}x = \frac{9}{(c-1)^2} \Rightarrow c = 4$.

▷ Note that the value $c = -2$ is not possible as the integral is divergent.

Problem 3.b

▷ $f_X(x) = \sum_{k\geqslant 0} f_{X,Y}(x,k) = \sum_{k\geqslant 0} 9e^{-4x}\frac{x^{k+1}}{k!} = 9xe^{-3x}$ and $f_Y(k) = \int_0^\infty 9e^{-4x}\frac{x^{k+1}}{k!}\mathrm{d}x = \frac{9\Gamma(k+2)}{k!4^{k+2}} = \frac{9(k+1)}{4^{k+2}}$.

Problem 3.c

▷ We obtain a $\Gamma(k+2,4)$-pdf, $f_{X|Y}(x|k) = \frac{f_{X,Y}(x,k)}{f_Y(k)} = \frac{9e^{-4x}\frac{x^{k+1}}{k!}}{\frac{9(k+1)}{4^{-k-2}}} = \frac{4^{k+2}}{(k+1)!}x^{k+1}e^{-4x} \Rightarrow \mathsf{Var}(X|k) = (k+2)/4^2$.

▷ We obtain a $\mathsf{Poi}(k)$-pmf, $f_{Y|X}(k|y) = f_{X,Y}(x,k)/f_X(x) = \frac{9e^{-4x}\frac{x^{k+1}}{k!}}{9xe^{-3x}} = e^{-x}\frac{x^k}{k!} \Rightarrow \mathsf{Var}(Y|x) = x$.

Problem 3.d

▷ Set $g(z) := (pz + (1-p))/4$

▷ Since the $(B_i)_{i\geqslant 1}$ are independent and identically distributed, we get as in Homework 3.3.1 that

$$\begin{aligned}\mathbb{E}[z^{Y'}] &= \mathbb{E}\Big[\sum_{k\geqslant 0}\mathbb{P}(Y=k)z^{\sum_{i\leqslant k}B_i}\Big] = \sum_{k\geqslant 0}\mathbb{P}(Y=k)\mathbb{E}[z^{B_1}]^k = \sum_{k\geqslant 0}\frac{9(k+1)}{4^{k+2}}(pz+(1-p))^k \\ &= \frac{9}{4^2}\frac{1}{(1-g(z))^2}.\end{aligned}$$

Problem 3.e

▷ We first compute the cdf

$$\begin{aligned}\mathbb{P}(X + (Y\wedge 1) \leqslant z) &= \mathbb{P}(X \leqslant z, Y = 0) + \sum_{k\geqslant 1}\mathbb{P}(X+1\leqslant z, Y=k) \\ &= \int_0^\infty 9xe^{-4x}\mathbb{1}\{x\leqslant z\}\mathrm{d}x + \sum_{k\geqslant 1}\int_0^\infty 9e^{-4x}\frac{x^{k+1}}{k!}\mathbb{1}\{x+1\leqslant z\}\mathrm{d}x \\ &= \int_0^z 9xe^{-4x}\mathrm{d}x + \sum_{k\geqslant 1}\mathbb{1}\{z\geqslant 1\}\int_0^{z-1} 9e^{-4x}\frac{x^{k+1}}{k!}\mathrm{d}x \\ &= \int_0^z 9xe^{-4x}\mathrm{d}x + \mathbb{1}\{z\geqslant 1\}\sum_{k\geqslant 1}\int_1^z 9e^{-4(x-1)}\frac{(x-1)^{k+1}}{k!}\mathrm{d}x \\ &= \int_0^z 9xe^{-4x}\mathrm{d}x + \mathbb{1}\{z\geqslant 1\}\int_1^z 9e^{-4(x-1)}(x-1)(e^{x-1}-1)\mathrm{d}x\end{aligned}$$

▷ Thus, the pdf is given by $f(x) = 9\big(xe^{-4x} + \mathbb{1}\{x\geqslant 1\}e^{-4(x-1)}(x-1)(e^{x-1}-1)\big)$.

Problem 4.a

▷ Write $A_n := \{n \text{ is a peak}\}$.

▷ By symmetry, each of $F_1,\dots,F_n$ has the same probability to be the largest in $\{F_1,\dots,F_n\} \Rightarrow \mathbb{P}(A_n) = 1/n$.

▷ Set $N := \min\{n\geqslant 2\colon n \text{ is a peak}\}$.

▷ Then, $\mathbb{E}[N] = \sum_{n\geqslant 1}\mathbb{P}(N\geqslant n) = \sum_{n\geqslant 1}\mathbb{P}(X_1 > \max\{X_2,\dots,X_{n-1}\}) = \sum_{n\geqslant 1} n^{-1} = \infty$.

Problem 4.b

▷ $\mathbb{E}[M_n] = \sum_{i\leqslant n}\mathbb{P}(A_i) = \sum_{i\leqslant n} 1/i$.

- ▷ For $i < j$ we have $\mathbb{P}(A_i \,|\, A_j) = \mathbb{P}(A_i)$ so that $\mathsf{Cov}(A_i, A_j) = 0$. Hence,

$$\mathsf{Var}(M_n) = \sum_{i \leqslant n} \mathsf{Cov}(A_i, A_i) + 2 \sum_{i < j \leqslant n} \mathsf{Cov}(A_i, A_j) = \sum_{i \leqslant n} (1/i - 1/i^2).$$

Problem 4.c
We check equality of the mgf's. By independence, we compute the mgf

$$M_{F_1^2 + F_2^2}(s) = M_{F_1^2}(s) M_{F_2^2}(s) = \Big((1 - 2st)^{-1/2}\Big)^2 = (1 - 2st)^{-1},$$

which is the mgf of an exponential random variable with mean $2t$, as asserted.
Problem 4.d

- ▷ $F'_m = F_m / \sqrt{t}$ is a standard normal random variable.
- ▷ Let $\varepsilon > 0$ be arbitrary.
- ▷ Then, by the hint, and since the infinite sum converges,

$$\mathbb{P}\big(\lim_{n \to \infty} \sup_{m \geqslant n} F'_m / \sqrt{\log m} > \sqrt{2}(1 + \varepsilon)\big) \leqslant \lim_{n \to \infty} \sum_{m \geqslant n} \mathbb{P}(F'_m > \sqrt{2 \log m}(1 + \varepsilon)) \leqslant \lim_{n \to \infty} \sum_{m \geqslant n} m^{-(1+\varepsilon)^2} = 0.$$