EPFL
Probability and Statistics for IC
2022, Spring semester

Do not turn this page until instructed

Probability and Statistics: Exam

June 20, 2022

Duration: The exam will start at 15:15 and end at 18:15 (unless special arrangements).

Family name:

First name:

SCIPER number:

Exercise	Points	Indicative marks
1		/4 points
2		/2 points
3		/3 points
4		/3 points
5		/3 points
6		/7 points
7		/3 points
Total:		/25 points

PROTOCOL:

- If caught cheating, you will get a 0 and we will report to the section.
- No personal documents, cheat sheets or calculators are allowed during the exam.
- Justify all your answers! Unjustified answers will not get full score even if correct. However, partial reasoning might get partial points.
- Try to simplify numerical expressions but no need to give exact decimal expressions (e.g., you can leave factorial expressions).
- Ask a question only if it is crucial. There should be no assistance needed with the exam questions, so if your question is unnecessary, we will not reply to it.
- There are 7 problems. After each problem, there is some blank space for the solution. If you need to add extra pages, please write your name on all sheets you add and make sure they are stapled together with the rest of the exam when collected.

Solutions

Exercise 1. We consider 5 urns: 3 of type A and 2 of type B. Each urn of type A contains 1 red ball and 3 white balls; each urn of type B contains 2 red balls and 2 white balls. We sample an urn uniformly at random; then, in the chosen urn, we sample two balls uniformly at random, *with replacement* (i.e., we replace the first sampled ball back in the urn before sampling the second ball).
Let U_A (resp. U_B) denote the event "the chosen urn is of type A (resp. of type B)". Let X denote number of red balls among the sampled balls.

1. Compute $\mathbb{P}(U_A)$, $\mathbb{P}(U_B)$ and $\mathbb{P}(X = k|U_B)$ for $k = 0, 1, 2$.

2. What is the probability mass function of X?

3. Compute $\mathbb{E}[X]$.

4. Knowing that we have sampled a red ball and a white ball, what is the probability that we have sampled an urn of type A?

5. Are the events $\{X = 1\}$ and U_A independent? As always, justify your answer.

Exercise 2. Let $(X, Y) \in \mathbb{R}^2$ be a random point, sampled uniformly in the unit disk. Said differently, (X, Y) has density

$$f_{(X,Y)}(x, y) = \frac{1}{\pi} I(x^2 + y^2 \leq 1) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1\,, \\ 0 & \text{if } x^2 + y^2 > 1\,. \end{cases}$$

1. What is the density of X?

2. Are X and Y independent? (As always, justify your answer.)

Exercise 3. Let U_1, U_2, U_3 be independent identically distributed (i.i.d.) $\mathcal{N}(0, 1)$ random variables.

1. What is the distribution of $X = \begin{pmatrix} U_1 - 2U_2 \\ U_1 + U_2 + U_3 \\ U_2 - U_3 \end{pmatrix}$?

2. Are the random variables $Y = U_1 + U_2 + U_3$ and $Z = U_2 - U_3$ independent?

3. Show that the random variable $T = \frac{Y^2}{3} + \frac{Z^2}{2}$ has a chi-square distribution χ_2^2 with 2 degrees of freedom.

For the following exercise, we recall this proposition:

Proposition 1. *Let X be a random variable with density f_X and $g : \mathbb{R} \to \mathbb{R}$ be a monotone increasing (or decreasing) map, with differentiable inverse g^{-1}. Then $Y = g(X)$ has density*

$$f_Y(y) = \left| \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} \right| f_X(g^{-1}(y)) \, .$$

Exercise 4. Let X be a random variable with density

$$f(x;\alpha) = \frac{2}{\alpha} x \exp\left(-\frac{x^2}{\alpha}\right) I(x > 0) \, ,$$

where $I(x > 0)$ denotes the indicator of $x > 0$, and $\alpha > 0$ is a parameter of the distribution.

1. What is the distribution of X^2?

2. We observe an independent identically distributed (i.i.d.) sample $(x_1, \ldots, x_n)$ from $f(\,.\,;\alpha)$, but α is unknown.

   (a) Give the maximum likelihood estimate $\hat{\alpha}$ for α.

   (b) Is $\hat{\alpha}$ unbiased in estimating α?

Exercise 5. We perform ten blood tests on a single individual. For each one of them, we obtain the following cholesterol levels (in grams):

$$245, 248, 247, 247, 249, 247, 247, 246, 246, 248.$$

We model this sequence of measurements as independent realizations of a random variable X "cholesterol level" whose distribution is assumed to be normal with mean μ and variance σ^2.
In this exercise, we seek to build confidence intervals using *exact* (not asymptotic) pivots for the parameters μ and σ^2. Try to give confidence intervals in terms of numerical expressions, but no need to simplify those (e.g. you can leave fractions, square roots, etc..). Please, write the full derivation of the confidence intervals asked; formulas with no derivation might not get full score, even if correct.

1. Assume first that σ^2 is known, and equal to 1.5. Build an equi-tailed confidence interval for μ, with confidence level 95%.

2. Assume now that σ^2 is unknown.

   (a) Build an equi-tailed confidence interval for μ, with confidence level 95%.

   (b) Build an equi-tailed confidence interval for σ^2, with confidence level 95%.

For the following exercise, you can use this fact:

Fact 1. *Let* $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ *be two sequences of random variables and let* x, y *be two real numbers. If* $X_n \xrightarrow[n\to\infty]{} x$ *almost surely and* $Y_n \xrightarrow[n\to\infty]{} y$ *almost surely, then* $X_n + Y_n \xrightarrow[n\to\infty]{} x + y$ *almost surely.*

Exercise 6. Let $X_1, X_2, X_3, \dots$ be an infinite sequence of independent identically distributed (i.i.d.) Ber$(1/2)$ random variables.

1. Define $I_{1,n}$ the number of ones in the sequence $X_1, \dots, X_n$, i.e.,

$$I_{1,n} = \#\{1 \le i \le n \,|\, X_i = 1\}\,.$$

(a) Give a number b_1 such that $\frac{1}{n} I_{1,n} \xrightarrow[n\to\infty]{} b_1$ almost surely. Justify your answer.

(b) Give a sequence $(c_n)_{n\ge 1}$ of real numbers, and a real number d, such that

$$c_n \left(\frac{1}{n} I_{1,n} - d \right) \xrightarrow[n\to\infty]{} Z \quad \text{in distribution,}$$

where Z is a standard Gaussian random variable. Justify your answer.

2. Define $I_{2,n}$ the number of pairs of consecutive ones in the sequence $X_1, \dots, X_n$, i.e.,

$$I_{2,n} = \#\{1 \le i \le n-1 \,|\, X_i = X_{i+1} = 1\}\,.$$

Give a number b_2 such that $\frac{1}{n} I_{2,n} \xrightarrow[n\to\infty]{} b_2$ almost surely. Justify your answer.

3. Define $I_{k,n}$ the number of sub-sequences of k consecutive ones in the sequence $X_1, \dots, X_n$, i.e.,

$$I_{k,n} = \#\{1 \le i \le n-k+1 \,|\, X_i = X_{i+1} = \dots = X_{i+k-1} = 1\}\,.$$

Give a number b_k such that $\frac{1}{n} I_{k,n} \xrightarrow[n\to\infty]{} b_k$ almost surely. Justify your answer.

4. Let K_n be the length of the largest sub-sequence of ones in the sequence $X_1, \dots, X_n$, i.e.,

$$K_n = \max\{k \in \mathbb{N} \,|\, I_{k,n} > 0\}\,.$$

(By convention, we assume $I_{0,n} = n$ so that K_n is well defined.)

(a) Let k, n be two positive integers such that $k \le n$. Show that $\mathbb{P}(K_n \ge k) \le \frac{n}{2^k}$.

(b) Let k, n be two positive integers such that $k \le n$. Show that $\mathbb{P}(K_n < k) \le \left(1 - \frac{1}{2^k}\right)^{\lfloor n/k \rfloor}$, where $\lfloor n/k \rfloor$ denotes the integer part of n/k.

(c) Using the two questions above, show that $\frac{\log 2}{\log n} K_n \xrightarrow[n\to\infty]{} 1$ in probability.

Exercise 7. Alice has 10 CHF and Bob has 5 CHF. They play a game. They repeatedly throw a fair coin (i.e. a coin such that $\mathbb{P}(\{\text{head}\}) = \mathbb{P}(\{\text{tail}\}) = 1/2$). Each throw is independent from the others. When a tail comes up, Alice gives 1 CHF to Bob; and conversely when a head comes up, Bob gives 1 CHF to Alice. The first player that reaches 0 CHF loses, and the other player wins.

1. What is the probability that Alice wins?

2. What is the expected number of coin throws in a game?