Probability and Statistics
FS 2019
Session Exam
23.01.2020
Time Limit: 180 Minutes

Name: ____________________

Student ID: ____________________

This exam contains 11 pages (including this cover page) and 10 problems. A formulae sheet is provided with the exam.

Grade Table (for grading use only, please leave empty)

Question	Points	Score
1	8	
2	10	
3	9	
4	10	
5	8	
6	12	
7	10	
8	9	
9	12	
10	12	
Total:	100	

1. (8 points) In a building with 4 floors (plus the ground floor), an elevator starts with 5 people at the ground floor.

   (a) (3 points) What is the probability that these people get off at exactly 2 different floors?

   (b) (3 points) What is the probability that all these people get off at the same floor?

   (c) (2 points) What is the probability that none of these people gets off at the first floor?

2. (10 points) Consider $n \geq 2$ people, assumed to have the same probability of being born on a given day of the calendar year. For simplicity, we assume that the calendar has $N = 365$ days.

   Consider the event

   $$A_n = \{\text{at least 2 people are born the same day}\}.$$

   (a) (2 points) Fix two people $1 \leq i \neq j \leq n$ from the group. Show that

   $$P(i \text{ and } j \text{ have the same birthday}) = \frac{1}{N}.$$

   (b) (3 points) Show that

   $$P(A_n^c) = \prod_{j=0}^{n-1} \left(1 - \frac{j}{N}\right).$$

   (c) (3 points) Using (b), and the inequality $\log(1+x) \leq x$ for $x \in (-1, +\infty)$, show that

   $$P(A_n^c) \leq \exp\left(-\frac{n(n-1)}{2N}\right).$$

   (d) (2 points) Using (a), show that

   $$P(A_n^c) \geq 1 - \frac{n(n-1)}{2N}.$$

3. (9 points) A firm wants to make 3 new hires. 6 applicants are interviewed, among which there are 3 women and 3 men.

   We assume that the applicants are equally qualified and that the hiring will be done at random.

   (a) (3 points) What is the probability that 2 women and 1 man will be hired?

   (b) (2 points) What is the probability that all 3 hires are men?

   (c) (1 point) What is the probability that at least 1 woman is hired?

   (d) (3 points) Assume that no woman was hired. Last year, the same firm also searched for 2 new hires from 4 applicants, among which 2 were women and 2 were men. The firm also hired all men last year.

   Assuming that the hires are independent from year to year, do you think that the hiring process is really done at random?

4. (10 points) Consider two random variables X and Y, such that X and Y are independent and $X \sim \text{Poi}(\lambda)$, $Y \sim \text{Poi}(\mu)$ for some $\lambda > 0, \mu > 0$.

   (a) (2 points) State the definition of independence of X and Y, when both X and Y are discrete random variables.

   (b) (3 points) Put $S = X + Y$. For $s \in \mathbb{N}_0$, compute $P(S = s)$. What is the distribution of S?

   (c) (3 points) For $s \in \mathbb{N}_0$ and $x \in \mathbb{N}_0$, compute $P(X = x \mid S = s)$. Deduce the distribution of X conditionally on $X + Y = s$.

   (d) (2 points) Take $\lambda = \mu = 1$. Compute $P(X = 0 \mid S = 2)$ and $P(X = 1 \mid S = 3)$.

5. (8 points) Consider a random variable X such that $X \sim \text{Geo}(p)$, for a given $p \in (0,1)$.

   (a) (2 points) Compute $P(X > i)$, for $i \in \mathbb{N}_{>0}$.

   (b) (2 points) Show that for any $j \geq i$,

   $$P(X > j \mid X > i) = P(X > j - i).$$

   How do you interpret this result?

   Consider the following experiment. You have 2 coins, A and B, such that

   $$P(A \text{ shows } H) = p$$

   and

   $$P(B \text{ shows } H) = 1 - p,$$

   for some $p \in (0,1)$. You toss A and B until you obtain the same outcome on both coins. When this happens, you stop. Let X denote the number of tosses required to stop the experiment.

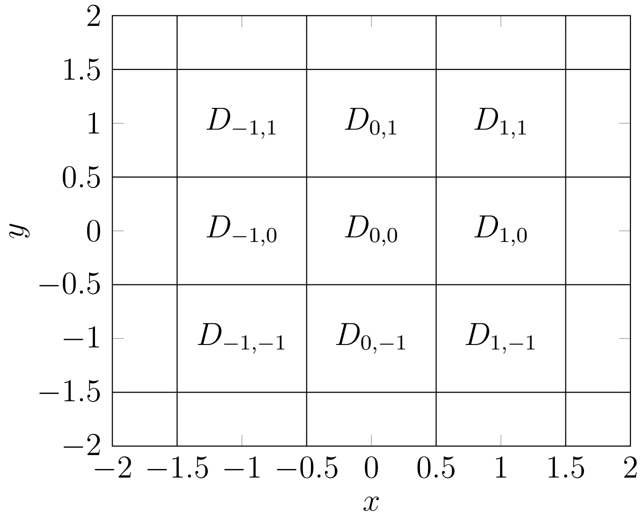
   (c) (2 points) What is the distribution of X?

   (d) (2 points) You win a prize if $X = 1$. What is the value of p which maximises your chance of winning? What is that chance?

6. (12 points) For each pair of integers $m, n \in \mathbb{Z}$, consider the square

$$D_{m,n} = \left\{ (x, y)^T \in \mathbb{R}^2 : |x - m| \leq \frac{1}{2}, |y - n| \leq \frac{1}{2} \right\}.$$

Consider a random variable $(X, Y)^T$ taking values on $\mathbb{R}^2$. We assume that for a given sequence $(a_k)_{k \geq 0}$, we have $P((X, Y)^T \in D_{m,n}) = a_{|m|+|n|}$. Conditionally on the event $\{(X, Y)^T \in D_{m,n}\}$, we assume that $(X, Y)^T$ is uniformly distributed on $D_{m,n}$.



(a) (2 points) Write down the joint density of the random vector $(X, Y)^T$ with respect to Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$.

(b) (3 points) Write down the conditions on $(a_k)_{k \geq 0}$ such that this is a valid probability distribution.

(c) (2 points) Suppose from now on that $a_k = \frac{c}{2^k}$ for some $c > 0$ and all $k \geq 0$. Calculate c.
Hint. You may use the identity $\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}$, for $|x| < 1$.

(d) (3 points) Find the marginal distribution of X and the conditional distribution of Y given X. Are X, Y independent?

(e) (2 points) Compute $E[X^2 + Y^2]$.
Hint. You may use the identity $\sum_{k=1}^{\infty} k^2 x^k = \frac{x(1+x)}{(1-x)^3}$, for $|x| < 1$.

7. (10 points) Recall the definition of the probability generating function (pgf): if X is a random variable taking values in $\{0, 1, ...\}$, then $G_X(s) = E[s^X]$, for any $s \in \mathbb{R}$ such that this is well-defined (in particular for $s \in [0, 1]$).

   (a) (2 points) By computing it explicitly, show that the pgf of the Poi(λ) distribution is
   $$G(s) = \exp(\lambda(s - 1)).$$

   (b) (3 points) Let $N, X_1, X_2, ...$ be independent random variables taking values on $\{0, 1, ...\}$, and such that the X_k are i.i.d.. Consider
   $$Z = \sum_{k=1}^{N} X_k.$$
   Show that, assuming both sides are well defined,
   $$G_Z(s) = G_N(G_X(s)),$$
   where X has the same distribution as X_1.
   Hint. Try taking a conditional expectation with respect to N first.

   (c) (2 points) Consider now $N \sim \text{Poi}(\mu)$ and $X \sim \text{Poi}(\lambda)$. Compute the pgf of
   $$Z = \sum_{k=1}^{N} X_k.$$
   Is Z a Poisson random variable?

   (d) (3 points) Let
   $$Z_n = \sum_{k=1}^{N_n} X_{k,n},$$
   where for each n, we have $N_n \sim \text{Poi}(n\mu)$ and $X_n \sim \text{Poi}(\frac{\lambda}{n})$ are i.i.d. (and independent from N_n). Compute the limit of G_{Z_n} as $n \to \infty$. Can you identify this limit as the pgf of a random variable?

8. (9 points) Let $\mathbf{X} = (X_1, X_2)^T$ be a random variable taking values on the disk $D = \{\mathbf{x} : \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} \le 1\}$, with density $f_{\mathbf{X}}(x_1, x_2)$ with respect to the Lebesgue measure on $\mathbb{R}^2$.

   Let $(R, \Theta)^T$ be the polar coordinates of $\mathbf{X}$, i.e. $R \in [0, 1]$ and $\Theta \in [0, 2\pi)$ such that $\mathbf{X} = (R\cos(\Theta), R\sin(\Theta))$.

   Let $g_{R,\Theta}$ be the joint density of $(R, \Theta)^T$ with respect to the Lebesgue measure.

   (a) (3 points) Using the Jacobian formula, show that

   $$g_{R,\Theta}(r, \theta) = r f_{\mathbf{X}}(r\cos(\theta), r\sin(\theta)).$$

   Now, we will consider two different ways of generating a random point $\mathbf{X}$ in the disk D. The notation $\mathbf{X}, R, \Theta$ refers to the previously defined random variables.

   (b) (3 points) Suppose that R, Θ are independent, with $\Theta \sim \mathrm{U}([0, 2\pi))$ and $R \sim \mathrm{U}([0, 1])$. Write down $g_{R,\Theta}$, and find $E[\mathbf{X}] = E[(X_1, X_2)^T]$ as well as $E[\|\mathbf{X}\|^2] = E[X_1^2 + X_2^2]$.

   (c) (3 points) Suppose that $\mathbf{X}$ is uniformly distributed on D. Compute $g_{R,\Theta}$, and find $E[\mathbf{X}] = E[(X_1, X_2)^T]$ as well as $E[\|\mathbf{X}\|^2] = E[X_1^2 + X_2^2]$.

9. (12 points) You buy and sell lightbulbs. The lightbulbs you order from the factory break down after a length of time (in months) that is Exp(λ)-distributed, for some unknown parameter λ.

   (a) (1 point) If X_1 is the length of time for which a lightbulb works until it breaks, write down the density and cdf of X_1.

   (b) (3 points) In order to estimate λ, you tested k lightbulbs, which broke after times $X_1, ..., X_k$. Justifying your answer, give a sufficient statistic for λ based on $X_1, ..., X_k$.

   (c) (3 points) Compute the MLE for λ.

   (d) (2 points) Compute the Fisher information for λ.

   (e) (3 points) Find an asymptotic confidence interval of level 0.95 for λ.

10. (12 points) Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with density $x \mapsto f_\theta(x), x \in \mathbb{R}$ with respect to Lebesgue measure. Here, θ is some unknown parameter in $\mathbb{R}$.

We want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ for some $\theta_0 \neq \theta_1$. We fix $\alpha \in (0, 1)$.

(a) (2 points) Recall the Neyman-Pearson test for this testing problem, at level α.

Remark. In the following parts, you can give your answer in terms of the cdf (and inverse cdf) of any standard distribution covered in the lectures. In the case of normal distributions, you must use the cdf Φ of the standard normal distribution.

(b) (3 points) Suppose $X_1, ..., X_n$ are i.i.d. $\sim \mathcal{N}(\theta, 1)$. Give the exact form of the Neyman-Pearson test for $(\theta_0, \theta_1) = (0, 1)$, specifying clearly the rejection region.

(c) (2 points) Find an expression for the power of the test constructed in (b).

(d) (3 points) We suppose now that $X_1, ..., X_n$ are i.i.d. $\sim \text{Exp}(\theta)$, with $\theta > 0$.

Give the exact form of the Neyman-Pearson test for $(\theta_0, \theta_1) = (1, 2)$ at level α, specifying clearly the rejection region.

(e) (2 points) Find an expression for the power of the test constructed in (d).

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆