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

Name: ____________________

Student ID: ____________________

This exam contains 24 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?

Solution.

(a) There are $\binom{4}{2}$ possible pairs of floors at which the people can get off. For each such pair, e.g. $\{1,2\}$ we need to count the number of possible partitions of $\{P_1, P_2, P_3, P_4, P_5\}$ into 2 different groups. There are 2^5 ways to assign these people into one of the 2 chosen floors, from which we subtract the 2 cases where they all go to one or the other floor. It follows that the number of all possibilities is

$$30 \times \binom{4}{2} = 180$$

The total number of all possibilities is 4^5. Hence, the probability is

$$\frac{180}{1024} = \frac{45}{256} \approx 0.1758.$$

(b) There are 4^5 total combinations of floors at which people can exit, and just 4 possibilities for all to leave at the same floor. Therefore, the probability is

$$\frac{4}{4^5} = \frac{1}{256} \approx 0.0039.$$

(c) The probability that none gets off at the first floor is the same as the probability that they all get off at floors $2, 3$ or 4. Thus, there are 3^5 possibilities and the probability is

$$\frac{3^5}{4^5} = \left(\frac{3}{5}\right)^5 \approx 0.237.$$

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

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

Solution.

(a) Fix $(i, j) \in \{1, ..., n\}^2 : i \neq j$ and consider

$$B_{ij} := \{i \text{ and } j \text{ are born on the same day}\}.$$

It is not difficult to see that counting the sample points in B_{ij} corresponds exactly to counting the number of $(n-1)$-tuplets $(w_1, ..., w_{n-1})$ with each $w_k \in \{1, ..., N\}$. Thus $\mathrm{card}(B_{ij}) = N^{n-1}$ so that

$$P(B_{ij}) = \frac{N^{n-1}}{N^n} = \frac{1}{N}.$$

(b) We count the number of possibilities such that no two people share a birthday. There are N possible birthdays for the first person. Once that is fixed, there are only $N-1$ possible birthdays for the next person, so as not to coincide with the first. Likewise, there are $N-2$ possibilities left for the third person, and so on. We conclude that there are $N(N-1)(N-2)...(N-n+1)$ possible arrangements for the birthdays.

As before, there are N^n possible arrangements in total, and thus

$$
\begin{aligned}
P(A^c) &= \frac{N(N-1)...(N-n+1)}{N^n} \\
&= \frac{N}{N}\frac{N-1}{N}...\frac{N-n+1}{N} \\
&= \left(1-\frac{0}{N}\right)\left(1-\frac{1}{N}\right)...\left(1-\frac{n-1}{N}\right) \\
&= \prod_{j=0}^{n-1}\left(1-\frac{j}{N}\right).
\end{aligned}
$$

(c) We can estimate

$$
\begin{aligned}
\log(P(A_n^c)) &= \sum_{j=0}^{n-1}\log\left(1-\frac{j}{N}\right) \\
&\leq -\frac{1}{N}\sum_{j=0}^{n-1} j \\
&= -\frac{1}{N}\frac{(n-1)n}{2}.
\end{aligned}
$$

This shows the upper bound

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

(d) To show the lower bound, note that

$$
A_n = \bigcup_{1\leq i<j\leq n} B_{ij}.
$$

where B_{ij} is the same event defined before in (a).
Hence,

$$
\begin{aligned}
P(A_n) &= P\left(\bigcup_{1\leq i<j\leq n} B_{ij}\right) \\
&\leq \sum_{1\leq i<j\leq n} P(B_{ij}) \\
&= \frac{1}{N}\left(\sum_{1\leq i<j\leq n} 1\right) \\
&= \binom{n}{2}\times\frac{1}{N} \\
&= \frac{n(n-1)}{2N}.
\end{aligned}
$$

The latter is also equivalent to

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

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?

Solution.

(a)

$$\begin{aligned} P(\text{2 women and 1 man will be hired}) &= \frac{\binom{3}{2} \times \binom{3}{1}}{\binom{6}{3}} \\ &= \frac{3 \times 3}{20} = \frac{9}{20}. \end{aligned}$$

(b)

$$\begin{aligned} P(\text{The 3 hires are men}) &= \frac{\binom{3}{0} \times \binom{3}{3}}{\binom{6}{3}} \\ &= \frac{1}{20}. \end{aligned}$$

(c)

$$\begin{aligned} P(\text{at least a woman is hired}) &= 1 - P(\text{The 3 hires are men}) \\ &= 1 - \frac{1}{20} = \frac{19}{20}. \end{aligned}$$

(d)

$$\begin{aligned} P(\text{all hires were men last year}) &= \frac{\binom{2}{0} \times \binom{2}{2}}{\binom{4}{2}} \\ &\frac{1}{6}. \end{aligned}$$

Since hiring decisions are independent from year to year,

$$P(\text{all hires were men last year and this year}) = \frac{1}{6} \times \frac{1}{20} = \frac{1}{120} < 0.05.$$

Hence, the assumption that the hiring is done at random is dubious.

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 $P(X = 1 \mid S = 3)$.

Solution.

(a) Let X and Y be discrete random variables such that $X \in \{x_1, x_2, ...\}$ and $Y \in \{y_1, y_2, ...\}$. Then, $X \perp\!\!\!\perp Y$ if and only if

$$P(X = x_i, y = y_j) = P(X = x_i)P(Y = y_j), \quad \forall i, j.$$

(b) Take $S = X + Y$. We have $P(S = s) = P(X + Y = s)$. Noting the partition

$$\{X + Y = s\} = \bigcup_{k=0}^{s} \{X = k, Y = s - k\},$$

and using the independence of X and Y, we obtain

$$\begin{aligned}
P(S = s) &= \sum_{k=0}^{s} P(X = k, Y = s - k) \\
&= \sum_{k=0}^{s} P(X = k)P(Y = s - k) \\
&= \sum_{k=0}^{s} \frac{\lambda^k e^{-\lambda}}{k!} \frac{\mu^{s-k} e^{-\mu}}{(s-k)!} \\
&= \frac{e^{-\lambda-\mu}}{s!} \sum_{k=0}^{s} \frac{s!}{k!(s-k)!} \lambda^k \mu^{s-k} \\
&= \frac{e^{-\lambda-\mu}}{s!} (\lambda + \mu)^s,
\end{aligned}$$

using the binomial theorem.
Therefore, we conclude that $S \sim \text{Poi}(\lambda + s)$.

(c)

$$P(X = x \mid S = s) = \begin{cases} \frac{P(X=x, S=s)}{P(S=s)} & \text{if } x \in \{0, ..., s\} \\ 0 & \text{otherwise} \end{cases}$$

with

$$\begin{aligned}
P(X = x, S = s) &= P(X = x, Y = s - x) \\
&= \frac{e^{\lambda} \lambda^x}{x!} \frac{e^{-\mu} \mu^{s-x}}{(s-x)!}.
\end{aligned}$$

Therefore, letting $p = \frac{\lambda}{\lambda+\mu}$ and $q = 1 - p = \frac{\mu}{\lambda+\mu}$, we obtain

$$P(X = x \mid S = s) = \frac{e^{-\lambda-\mu}\lambda^x \mu^{s-x}}{\frac{e^{-\lambda-\mu}(\lambda+\mu)^s}{s!}x!(s-x)!} = \frac{s!}{x!(s-x)!}p^x q^{s-x},$$

and hence $X \mid S = s \sim \mathrm{Bin}(s, p)$ with $p = \frac{\lambda}{\lambda+\mu}$.

(d) Taking $\lambda = \mu = 1$, we have

$$P(X = 0 \mid S = 2) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

and

$$P(X = 1 \mid S = 3) = \binom{3}{1}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}.$$

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?

Solution.

(a)
$$\begin{aligned}
P(X = x) &= p(1-p)^{x-1}, x = 1, 2, ...\\
\Rightarrow P(X > i) &= p \sum_{x=i+1}^{\infty} (1-p)^{x-1}\\
&= p \sum_{x=i}^{\infty} (1-p)^{x}\\
&= p(1-p)^i \frac{1}{1-(1-p)}\\
&= (1-p)^i.
\end{aligned}$$

(b)
$$\begin{aligned}
P(X > j \mid X > i) &= \frac{P(X > j, X > i)}{P(X > i)}\\
&= \frac{P(X > j)}{P(X > i)}\\
&= (1-p)^{j-i}.
\end{aligned}$$

This means that X is "memoryless". In other words, if we interpret X as the time one needs to wait for a certain outcome to happen, and if that outcome hasn't happened yet at times $\{1, ..., i\}$, we have that X just "forgets" how long it has been waiting, and the remaining time is once again exponentially distributed. This also means that the total waiting time, conditional on it being greater than i, is exponentially distributed (up to a shift by i).

(c)

$$\begin{aligned} P(A \text{ and } B \text{ show the same outcome}) &= P(A, B \text{ both get } H) + P(A, B \text{ both get } T) \\ &= p(1-p) + (1-p)p = 2p(1-p). \end{aligned}$$

Hence, $X \sim \text{Geo}(2p(1-p))$.

(d) The probability of winning is

$$P(X = 1) = 2p(1-p).$$

By differentiating, we obtain

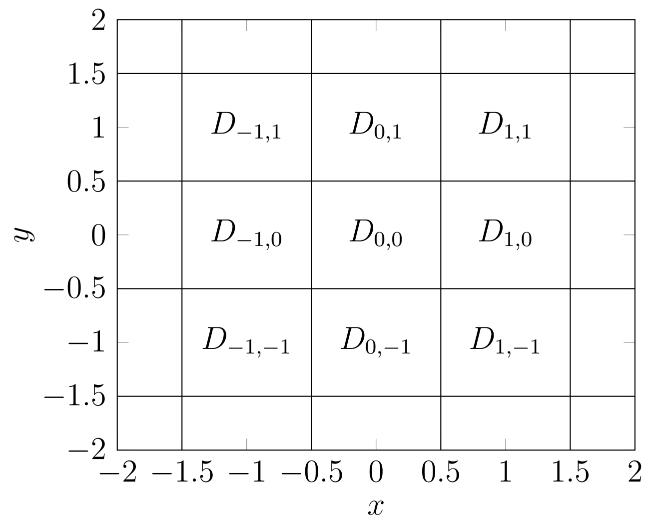
$$2(1-\hat{p}) - 2\hat{p} = 0 \Rightarrow \hat{p} = \frac{1}{2}.$$

By taking a second derivative, we see that $\hat{p} = \frac{1}{2}$ is the maximiser, which leads to a probability of winning equal to $\frac{1}{2}$.

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

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

Consider a random variable $(X,Y)^T$ taking values on $\mathbb{R}^2$. We assume that for a given sequence $(a_k)_{k\ge 0}$, 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 $(a_k)_{k\ge 0}$ 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 $k \ge 0$. 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 $|x| < 1$.

Solution.

(a) Each square $D_{m,n}$ has sides of length 1, therefore, by uniformity, the conditional density of $(X,Y)^T$, given that $(X,Y)^T \in D_{m,n}$, is $\mathbb{1}_{D_{m,n}}$.
By conditioning, we deduce that

$$f_{(X,Y)^T}(x,y) = \sum_{m,n\in\mathbb{Z}} a_{|m|+|n|} \mathbb{1}_{D_{m,n}}(x,y).$$

(b) To check that this is a valid probability distribution, we need to check that the density is non-negative and integrates to 1. Clearly, the density is non-negative if and only if each $a_k \ge 0$. Moreover, we can compute the integral as

$$\begin{aligned}
1 &= \int_{\mathbb{R}^2} f_{(X,Y)^T}(x,y)dxdy \\
&= \int_{\mathbb{R}^2} \sum_{m,n\in\mathbb{Z}} a_{|m|+|n|} \mathbb{1}_{D_{m,n}}(x,y)dxdy \\
&= \sum_{m,n\in\mathbb{Z}} a_{|m|+|n|} \int_{\mathbb{R}^2} \mathbb{1}_{D_{m,n}}(x,y)dxdy \\
&= \sum_{m,n\in\mathbb{Z}} a_{|m|+|n|}
\end{aligned}$$

(note that we used the monotone convergence theorem). To progress further, let $A_k = \{(m,n) \in \mathbb{Z}^2 : |m| + |n| = k\}$. It is clear that

$$1 = \sum_{k=0}^{\infty} |A_k| a_k.$$

We see that the number of such pairs $(m,n) \in A_k$ is given by $4k$, if $k \geq 1$, and 1 if $k = 0$. Therefore we conclude that

$$1 = a_0 + \sum_{k=1}^{\infty} 4k a_k$$

(together with non-negativity) is the condition we need to guarantee that the probability distribution is valid.

(c) We can use the previous equation:

$$\begin{aligned}
1 &= c + \sum_{k=1}^{\infty} 4k \frac{c}{2^k} \\
&= c\left(1 + 4\sum_{k=1}^{\infty} \frac{k}{2^k}\right) \\
&= c\left(1 + 4\frac{\frac{1}{2}}{\left(1 - \frac{1}{2}\right)^2}\right) \\
&= c(1+8) \\
&= 9c,
\end{aligned}$$

so that $c = \frac{1}{9}$.

(d) Let $x \in \mathbb{R}$, and choose the unique $m_0 \in \mathbb{Z}$ such that $m_0 - \frac{1}{2} \leq x < m + \frac{1}{2}$. Therefore, we have that

$$\begin{aligned}
f_X(x) &= \int_{\mathbb{R}} f_{(X,Y)^T}(x,y)dy \\
&= \int_{\mathbb{R}} \sum_{m,n\in\mathbb{Z}} a_{|m|+|n|} \mathbb{1}_{D_{m,n}}(x,y)dy \\
&= \int_{\mathbb{R}} \sum_{n\in\mathbb{Z}} a_{|m_0|+|n|} \mathbb{1}_{D_{m_0,n}}(x,y)dy \\
&= \sum_{n\in\mathbb{Z}} a_{|m_0|+|n|} \int_{\mathbb{R}} \mathbb{1}_{D_{m_0,n}}(x,y)dy \\
&= \sum_{n\in\mathbb{Z}} a_{|m_0|+|n|}
\end{aligned}$$

(once again, using the monotone convergence theorem). We can compute

$$\begin{aligned}
f_X(x) &= \sum_{n\in\mathbb{Z}} a_{|m_0|+|n|} \\
&= a_{|m_0|} + 2a_{|m_0|+1} + 2a_{|m_0|+2} \cdots \\
&= a_{|m_0|} + 2\sum_{j=1}^{\infty} a_{|m_0|+j} \\
&= \frac{1}{9\cdot 2^{|m_0|}} + 2\sum_{j=1}^{\infty} \frac{1}{9\cdot 2^{|m_0|+j}} \\
&= \frac{1}{9\cdot 2^{|m_0|}} + \frac{2}{9\cdot 2^{|m_0|}} \\
&= \frac{1}{3\cdot 2^{|m_0|}}.
\end{aligned}$$

To summarise, the marginal density of X is

$$f_X(x) = \sum_{m\in\mathbb{Z}} \frac{1}{3\cdot 2^{|m|}} \mathbb{1}_{m-\frac{1}{2}\leq x<m+\frac{1}{2}}.$$

We can then easily find the conditional distribution of Y given x: if $m-\frac{1}{2}\leq x<m+\frac{1}{2}$ and $n-\frac{1}{2}\leq y<n+\frac{1}{2}$, we compute

$$f_{Y|X}(y\mid x) = \frac{f_{(X,Y)^T}(x,y)}{f_X(x)} = \frac{\frac{1}{9\cdot 2^{|m|+|n|}}}{\frac{1}{3\cdot 2^{|m|}}} = \frac{1}{3\cdot 2^{|n|}}.$$

Since the conditional density does not depend on x (or m), we conclude that X and Y are independent.

(e) By symmetry, it is enough to calculate $E[X^2] = \frac{1}{2}E[X^2+Y^2]$. Using the marginal density, and applying the monotone convergence theorem, we can compute

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} x^2 \sum_{m\in\mathbb{Z}} \frac{1}{3\cdot 2^{|m|}} \mathbb{1}_{m-\frac{1}{2}\le x<m+\frac{1}{2}} dx \\
&= \sum_{m\in\mathbb{Z}} \frac{1}{3\cdot 2^{|m|}} \int_{m-\frac{1}{2}}^{m+\frac{1}{2}} x^2 dx \\
&= \sum_{m\in\mathbb{Z}} \frac{1}{9\cdot 2^{|m|}} \left(\left(m+\frac{1}{2}\right)^3 - \left(m-\frac{1}{2}\right)^3 \right) \\
&= \sum_{m\in\mathbb{Z}} \frac{1}{9\cdot 2^{|m|}} \left(3m^2+\frac{1}{4}\right) \\
&= \frac{1}{36} + 2\left(\sum_{m\ge 1} \frac{m^2}{3\cdot 2^m} + \frac{1}{36}\sum_{m\ge 1}\frac{1}{2^m} \right) \\
&= \frac{1}{36} + 2\left(\frac{\frac{1}{2}\left(1+\frac{1}{2}\right)}{3\left(1-\frac{1}{2}\right)^3} + \frac{1}{36} \right) \\
&= \frac{1}{36} + 2\left(2+\frac{1}{36}\right) \\
&= 4+\frac{1}{12} \; = \frac{49}{12}.
\end{aligned}
$$

Thus, $E[X^2+Y^2] = \frac{49}{6}$.

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 $N \sim$ Poi(μ) and $X \sim$ Poi(λ). 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 $N_n \sim$ Poi($n\mu$) and $X_n \sim$ 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?

Solution.

(a) For $X \sim$ Poi(λ), we can compute

$$\begin{aligned} G(s) = E[s^X] &= \sum_{k=0}^{\infty} s^k P(X = k) \\ &= \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{\lambda(s-1)} \sum_{k=0}^{\infty} \frac{e^{-s\lambda}(s\lambda)^k}{k!} \\ &= e^{\lambda(s-1)}, \end{aligned}$$

since the sum is equal to 1.

(b) We compute the pgf as follows:

$$
\begin{aligned}
G_Z(s) &= E[s^Z] \\
&= \sum_{n=0}^{\infty} E[s^Z \mid N = n] P(N = n) \\
&= \sum_{n=0}^{\infty} E[s^{\sum_{k=1}^{N} X_k} \mid N = n] P(N = n) \\
&= \sum_{n=0}^{\infty} E[s^{\sum_{k=1}^{n} X_k} \mid N = n] P(N = n) \\
&= \sum_{n=0}^{\infty} E[s^{\sum_{k=1}^{n} X_k}] P(N = n) \\
&= \sum_{n=0}^{\infty} \prod_{k=1}^{n} E[s^{X_k}] P(N = n) \\
&= \sum_{n=0}^{\infty} \prod_{k=1}^{n} G_{X_k}(s) P(N = n) \\
&= \sum_{n=0}^{\infty} G_X(s)^n P(N = n) \\
&= \sum_{n=0}^{\infty} E[G_X(s)^N \mid N = n] P(N = n) \\
&= E[G_X(s)^N] \\
&= G_N(G_X(s)),
\end{aligned}
$$

as required.

(c) Using the formula from the previous part, we obtain

$$
\begin{aligned}
G_Z(s) &= G_N(G_X(s)) \\
&= G_N(e^{\lambda(s-1)}) \\
&= \exp\left(\mu(e^{\lambda(s-1)} - 1)\right).
\end{aligned}
$$

Therefore, Z is not a Poisson random variable: if it were, we would have that $\log(G_Z(s))$ is a linear function of s, which is not the case.

(d) Similarly, we obtain that

$$
G_{Z_n}(s) = \exp\left(n\mu(e^{\frac{\lambda}{n}(s-1)} - 1)\right).
$$

When n is large, we have the Taylor expansion

$$
\begin{aligned}
n\mu(e^{\frac{\lambda}{n}(s-1)} - 1 &= n\mu\left(1 + \frac{\lambda}{n}(s-1) - 1 + O\left(\frac{1}{n^2}\right)\right) \\
&= \mu\left(\lambda(s-1) + O\left(\frac{1}{n}\right)\right) \to \mu\lambda(s-1).
\end{aligned}
$$

Using continuity of the exponential, we obtain that

$$G_{Z_n}(s) = \exp\left(n\mu(e^{\frac{\lambda}{n}(s-1)} - 1)\right) \to \exp(\mu\lambda(s-1)).$$

This is the pgf of a $\mathrm{Poi}(\lambda\mu)$ distribution.

8. (9 points) Let $\mathbf{X} = (X_1, X_2)^T$ be a random variable taking values on the disk $D = \{\mathbf{x} : \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} \leq 1\}$, 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 $E[\|\mathbf{X}\|^2] = E[X_1^2 + X_2^2]$.

Solution.

(a) We use the Jacobian formula. Letting $h(r, \theta) = (r\cos\theta, r\sin\theta)$, we obtain that
$$g_{R,\Theta}(r, \theta) = f_{\mathbf{X}}(r\cos(\theta), r\sin(\theta))|J|,$$
where
$$|J| = \det\begin{pmatrix} \partial_r h_1 & \partial_r h_2 \\ \partial_\theta h_1 & \partial_\theta h_2 \end{pmatrix}$$
is the Jacobian. We can easily calculate
$$|J| = \det\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} = r\cos^2\theta + r\sin^2\theta = r,$$
which gives the formula we wanted:
$$g_{R,\Theta}(r, \theta) = r f_{\mathbf{X}}(r\cos(\theta), r\sin(\theta))|J|.$$

(b) If R, Θ are independent and uniformly distributed, we can write down straight away
$$g_{R,\Theta}(r, \theta) = \frac{1}{2\pi}\mathbb{1}_{r\in(0,1)}\mathbb{1}_{\theta\in[0,2\pi)}.$$
We find
$$E[X_1] = E[R\cos\Theta] = E[R]E[\cos\Theta] = 0$$
by independence of R and Θ, as well as symmetry. Likewise, $E[X_2] = 0$ so $E[\mathbf{X}] = \mathbf{0}$.
Moreover,
$$\begin{aligned} E[\|\mathbf{X}\|^2] &= E[R^2] \\ &= \int_0^1 r^2 dr \\ &= \frac{1}{3}. \end{aligned}$$

(c) If $\mathbf{X}$ is uniformly distributed on D (which has area π), this means that

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\pi}\mathbb{1}_{\mathbf{x}\in D}.$$

Changing to polar coordinates, note that $\mathbf{x} \in D$ if and only if $r \leq 1$ (without restriction on θ). Therefore, part (a) gives that

$$g_{R,\Theta}(r,\theta) = \frac{r}{\pi}\mathbb{1}_{r\in(0,1)}\mathbb{1}_{\theta\in[0,2\pi)}.$$

Since $g_{R,\Theta}$ is factorised into terms that only depend on r or only on θ, this means that R, Θ are independent. By the same argument as in (b) we find that $E[\mathbf{X}] = \mathbf{0}$, and moreover

$$\begin{aligned}
E[\|\mathbf{X}\|^2] &= E[R^2] \\
&= \int_0^1 \int_0^{2\pi} r^2 \frac{r}{\pi} d\theta dr \\
&= \int_0^1 2r^3 dr \\
&= \frac{1}{2}.
\end{aligned}$$

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 $\text{Exp}(\lambda)$-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.

   Solution.

   (a) Since X_1 is exponentially distributed, we have

   $$f_{X_1}(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$$

   and

   $$F_{X_1}(x) = (1 - e^{-\lambda x}) \mathbb{1}_{x \geq 0}.$$

   (b) If lightbulb j takes time $x_j \geq 0$ to break down, the joint density function is

   $$f(x_1, ..., x_k; \lambda) = \prod_{j=1}^{k} \lambda e^{-\lambda x_j} \mathbb{1}_{x_j \geq 0}.$$

   We can write it as

   $$f(x_1, ..., x_k; \lambda) = \lambda^k e^{-\lambda \sum_{j=1}^k x_j} \times \mathbb{1}_{\min_j x_j \geq 0}..$$

   Therefore, we found a factorisation

   $$f(x_1, ..., x_k; \lambda) = g(t, \lambda) h(x_1, ..., x_k),$$

   where $t = \sum_{j=1}^k x_j$, $h(x_1, ..., x_k) = \mathbb{1}_{\min_j x_j \geq 0}$ and

   $$g(t, \lambda) = \lambda^k e^{-\lambda t}.$$

   Therefore, by the factorisation Theorem we conclude that $T = \sum_{j=1}^k X_j$ is a sufficient statistic for λ.

   (c) Similarly to (b), we obtain that the likelihood function is

   $$L(\lambda; x_1, ..., x_k) = \prod_{j=1}^{k} \lambda e^{-\lambda x_j}.$$

   The log-likelihood is

   $$l(\lambda; x_1, ..., x_k) = k \log \lambda - \lambda \sum_{j=1}^{k} x_j,$$

and therefore we can compute the MLE by solving

$$\begin{aligned} 0 &= \frac{\partial l}{\partial \lambda}(\hat{\lambda}) \\ &= \frac{k}{\hat{\lambda}} - \sum_{j=1}^{k} x_j \\ \Leftrightarrow \hat{\lambda} &= \frac{k}{\sum_{j=1}^{k} x_j}. \end{aligned}$$

We can check that this is the MLE, since

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{k}{\lambda^2} < 0,$$

so the log-likelihood is strictly concave.

(d) Assuming that we have sufficient smoothness conditions, the Fisher information is given by

$$I(\lambda) = -E\left[\frac{\partial^2}{\partial \lambda^2} \log f(X \mid \lambda)\right].$$

This can be calculated as

$$\begin{aligned} I(\lambda) &= -E\left[\frac{\partial^2}{\partial \lambda^2}(-\lambda X + \log \lambda)\right] \\ &= E\left[\frac{\partial^2}{\partial \lambda^2}(\lambda X - \log \lambda)\right] \\ &= E\left[\frac{1}{\lambda^2}\right] \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

(e) We have the asymptotic distribution

$$\sqrt{kI(\lambda_0)}(\hat{\lambda} - \lambda_0) \to \mathcal{N}(0, 1),$$

where λ_0 is the true mean. Therefore, we can find an asymptotic confidence interval by noting that

$$P\left(-z_{0.975} \leq \sqrt{kI(\lambda_0)}(\hat{\lambda} - \lambda_0) \leq z_{0.975}\right) \approx 0.95,$$

so that, using $z_{0.975} \approx 1.96$, we have an approximate confidence interval of level 0.975:

$$\left[\hat{\lambda} - \frac{1.96\lambda_0}{\sqrt{k}}, \hat{\lambda} + \frac{1.96\lambda_0}{\sqrt{k}}\right].$$

In order to be able to compute this interval, we approximate the λ_0 in the error term by $\hat{\lambda}$, to obtain

$$\left[\hat{\lambda} - \frac{1.96\hat{\lambda}}{\sqrt{k}}, \hat{\lambda} + \frac{1.96\hat{\lambda}_0}{\sqrt{k}}\right].$$

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. $\sim \mathrm{Exp}(\theta)$, 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).

Solution.

(a) Let $f_0(\mathbf{x}) = \prod_{j=1}^n f_{\theta_0}(x_j)$ be the joint density function under H_0, and let $f_1(\mathbf{x}) = \prod_{j=1}^n f_{\theta_1}(x_j)$ be the joint density function under H_1. Then, in general, the Neyman-Pearson test takes the following form:

$$d_{NP}(\mathbf{x}) = \begin{cases} 1, & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > k_\alpha \\ \gamma_\alpha, & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = k_\alpha \\ 0 & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} < k_\alpha. \end{cases}$$

This can be interpreted as rejecting H_0 if $\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > k_\alpha$, not rejecting if $\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} < k_\alpha$, and randomly rejecting with probability γ_α if $\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = k_\alpha$. Here, $k_\alpha \in \mathbb{R}$ and $\gamma_\alpha \in [0, 1]$ are suitably chosen so that the test has level α.

(b) The Neyman-Pearson test rejects H_0 if

$$\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > k_\alpha.$$

This can be rewritten as

$$\begin{aligned}
\frac{\prod_{j=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_j-1)^2}{2}}}{\prod_{j=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_j^2}{2}}} &> k_\alpha \\
\Leftrightarrow \exp\left(-\sum_{j=1}^n \frac{(x_j-1)^2}{2} + \sum_{j=1}^n \frac{x_j^2}{2}\right) &> k_\alpha \\
\Leftrightarrow -\sum_{j=1}^n \frac{(x_j-1)^2}{2} + \sum_{j=1}^n \frac{x_j^2}{2} &> \log k_\alpha \\
\Leftrightarrow \sum_{j=1}^n x_j - \frac{n}{2} &> \log k_\alpha \\
\Leftrightarrow \sum_{j=1}^n x_j &> \log k_\alpha + \frac{n}{2}.
\end{aligned}$$

Therefore, we can rewrite the test in terms of $\sum_{j=1}^n x_j$. Note that the joint density of $(X_1, ..., X_n)$ is absolutely continuous with respect to Lebesgue measure, and therefore the boundary event has probability 0 and can be ignored, i.e. we can take $\gamma_\alpha = 0$ without loss of generality. The Neyman-Pearson test then takes the new form

$$d_{NP}(\mathbf{x}) = \begin{cases} 1, & \sum_{j=1}^n x_j > t_\alpha \\ 0 & \sum_{j=1}^n x_j \le t_\alpha \end{cases},$$

for a new constant $t_\alpha = \log k_\alpha + \frac{n}{2}$. In order for the test to have level α, we need to check

$$\begin{aligned}
\alpha &= E_{H_0}[d_{NP}(\mathbf{X})] \\
&= P_{H_0}\left(\sum_{j=1}^n X_j > t_\alpha\right).
\end{aligned}$$

Since the X_j are i.i.d. $\sim \mathcal{N}(0,1)$ under H_0, we find that $\sum_{j=1}^n X_j \sim \mathcal{N}(0,n)$ and so

$$P_{H_0}\left(\sum_{j=1}^n X_j > t_\alpha\right) = P_{H_0}\left(\frac{\sum_{j=1}^n X_j}{\sqrt{n}} > \frac{t_\alpha}{\sqrt{n}}\right) = 1 - \Phi\left(\frac{t_\alpha}{\sqrt{n}}\right).$$

We conclude that

$$\begin{aligned}
\alpha &= 1 - \Phi\left(\frac{t_\alpha}{\sqrt{n}}\right) \\
\Leftrightarrow t_\alpha &= \sqrt{n} z_{1-\alpha}.
\end{aligned}$$

(c) To find the power of this test, we need to calculate

$$E_{H_1}[d_{NP}(\mathbf{X})] = P_{H_1}\left(\sum_{j=1}^n X_j > t_\alpha\right).$$

Under H_1, the X_j are i.i.d. $\sim \mathcal{N}(1,1)$, so $\sum_{j=1}^n X_j \sim \mathcal{N}(n,n)$ and we obtain

$$P_{H_1}\left(\sum_{j=1}^n X_j > t_\alpha\right) = P_{H_1}\left(\frac{\sum_{j=1}^n X_j - n}{\sqrt{n}} > \frac{t_\alpha - n}{\sqrt{n}}\right)$$
$$= 1 - \Phi\left(\frac{t_\alpha - n}{\sqrt{n}}\right)$$
$$= 1 - \Phi\left(\frac{\sqrt{n} z_{1-\alpha} - n}{\sqrt{n}}\right)$$
$$= 1 - \Phi\left(z_{1-\alpha} - \sqrt{n}\right).$$

Thus, the power is $\beta = 1 - \Phi\left(z_{1-\alpha} - \sqrt{n}\right) = \Phi\left(\sqrt{n} - z_{1-\alpha}\right)$.

(d) Similarly to part (b), we can rewrite the Neyman-Pearson test:

$$\frac{\prod_{j=1}^n 2e^{-2x_j}}{\prod_{j=1}^n e^{-x_j}} > k_\alpha$$
$$\Leftrightarrow 2^n \exp\left(-2\sum_{j=1}^n x_j + \sum_{j=1}^n x_j\right) > k_\alpha$$
$$\Leftrightarrow n \log 2 - \sum_{j=1}^n x_j > \log k_\alpha$$
$$\Leftrightarrow \sum_{j=1}^n x_j < n \log 2 - \log k_\alpha.$$

Once again, we may ignore the boundary case and so the Neyman-Pearson test has the form

$$d_{NP}(\mathbf{x}) = \begin{cases} 1, & \sum_{j=1}^n x_j < t_\alpha \\ 0 & \sum_{j=1}^n x_j \geq t_\alpha \end{cases},$$

with $t_\alpha = n \log 2 - \log k_\alpha$.

To find the level, we check that

$$\alpha = E_{H_0}[d_{NP}(\mathbf{X})]$$
$$= P_{H_0}\left(\sum_{j=1}^n X_j < t_\alpha\right).$$

Since under H_0, the X_j are i.i.d. $\sim$ Exp(1), we have that $\sum_{j=1}^n X_j \sim G(n,1)$. Therefore,

$$P_{H_0}\left(\sum_{j=1}^n X_j < t_\alpha\right) = F_{G(n,1)}(t_\alpha),$$

where $F_{G(n,1)}$ is the cdf of the $G(n,1)$ distribution. This implicitly gives t_α as the quantile

$$t_\alpha = F_{G(n,1)}^{-1}(\alpha).$$

(e) For the power, we compute

$$E_{H_1}[d_{NP}(\mathbf{X})] = P_{H_1}\left(\sum_{j=1}^{n} X_j < t_\alpha\right).$$

Under H_1, we have that $\sum_{j=1}^{n} X_j \sim G(n, 2)$, and therefore we obtain the power

$$P_{H_1}\left(\sum_{j=1}^{n} X_j < t_\alpha\right) = F_{G(n,2)}(t_\alpha) = F_{G(n,2)}(F^{-1}_{G(n,1)}(\alpha)).$$

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆