MATH 505a Spring 2022 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

(a)

Let X_1, X_2, X_3 be independent exponential random variables with parameter $\lambda = 1$. So $\mathbb{P}(X_i > x) = e^{-x}, x > 0$. Find

$$\mathbb{E}\left(\frac{X_1}{X_1 + X_2 + X_3}\right).$$

Solution. Let $Y = X_1 + X_2$, then compute the pdf of Y:

$$\begin{aligned}
\mathbb{P}(X_2 + X_3 \leq y) &= \int_0^y \mathbb{P}(X_2 \leq y - x) f_{X_3}(x) dx \\
&= \int_0^y (1 - e^{-(y-x)}) e^{-x} dx \\
&= \left[-e^{-x} - xe^{-x}\right]_0^y \\
&= 1 - e^{-y}(y+1),\ \ y \geq 0
\end{aligned}$$

So by differentiating the cdf, we have

$$f_Y(y) = ye^{-y},\ \ y \geq 0$$

Now consider the probability:

$$\begin{aligned}
\mathbb{P}\left(\frac{X_1}{X_1 + Y} \leq z\right) &= \mathbb{P}\left(X_1 \leq \frac{zY}{1-z}\right) \\
&= \int_0^\infty \mathbb{P}\left(X_1 \leq \frac{zy}{1-z}\right) f_Y(y) dy \\
&= \int_0^\infty \left[1 - exp\left(-\frac{zy}{1-z}\right)\right] \cdot y \cdot exp(-y) dy \\
&= 1 - (1-z)^2,\ \ 0 \leq z \leq 1
\end{aligned}$$

And from the fact that $\frac{X_1}{X_1+Y} \geq 0$ (only takes non-negative value), we can compute the expectation by the complementary cdf:

$$\mathbb{E}\left(\frac{X_1}{X_1 + X_2 + X_3}\right) = \int_0^1 (1-z)^2 dz = \frac{1}{3}$$

(b)

Let (X, Y) be independent uniforms on $[0, 1]$. Find the joint density function of X and $V = X + Y$. Find $f(x|v)$, the density function of X conditional on $V = v$. Also, find $\mathbb{E}(X|V)$.

Proof. First we should compute the pdf of V. When $0 < v < 1$:

$$\mathbb{P}(X + Y \leq v) = \int_0^v (v - y)dy = \frac{1}{2}v^2$$

When $1 \leq v < 2$:

$$\mathbb{P}(X + Y \leq v) = \int_{v-1}^1 (v - y)dy + (v - 1) = -\frac{1}{2}v^2 + 2v - 1$$

So we have the pdf:

$$f_V(v) = \begin{cases} v & 0 < v < 1 \\ 2 - v & 1 \leq v < 2 \end{cases}$$

Note that the conditional pdf is

$$f_{X|V}(x|v) = \frac{f_{X,V}(x, v)}{f_V(v)} = \frac{f_X(x)f_Y(y)}{f_V(v)},\ y = v - x$$

plug in the previous results,

$$f_{X|V}(x|v) = \begin{cases} \frac{1}{v} & 0 < x < v < 1 \\ \frac{1}{2-v} & 1 \leq v < 2,\ v - 1 < x < 1 \end{cases}$$

So the expectation follows:

$$\mathbb{E}(X|V) = \frac{1}{2}V \cdot \mathbb{1}_{(0,1)}(V) + \frac{2V - V^2}{4 - 2V} \cdot \mathbb{1}_{[1,2)}(V)$$

□

Problem 2

In an election, candidates A receives n votes, and candidate B receives m votes, where $n > m$. Assuming that all $\binom{n+m}{m}$ orderings are equally likely, show that the probability that A is always ahead in the count of votes is $(n - m)/(n + m)$.

Proof. Denote $P_{i,j} = \mathbb{P}(\text{A is always ahead} \mid \text{A received } i \text{ votes, B received } j \text{ votes})$. By conditioning on the last vote, we have

$$\begin{aligned} P_{n,m} &= \mathbb{P}(\text{A received the last vote})P_{n-1,m} + \mathbb{P}(\text{B received the last vote})P_{n,m-1} \\ &= \frac{n}{n+m}P_{n-1,m} + \frac{m}{n+m}P_{n,m-1} \end{aligned}$$

Then we construct an induction: $\forall k \in \mathbb{N}$, if $n + m = k$ and $n > m$, then $P_{n,m} = \frac{n-m}{n+m}$.

1. $k = 1,\ n > m \implies n = 1,\ m = 0$ and $P_{1,0} = 1$

2. Given that the statement is true for $n + m = k$, let $n + m = k + 1,\ n > m$. Then we have

$$P_{n,m} = \frac{n}{n+m} P_{n-1,m} + \frac{m}{n+m} P_{n,m-1}$$

Note that $n + m - 1 = k \implies P_{n,m-1} = \frac{n-m+1}{n+m-1}$ and $P_{n-1,m} = \frac{n-m-1}{n+m-1}$. So

$$\begin{aligned} P_{n,m} &= \frac{n}{n+m}\frac{n-1-m}{n-1+m} + \frac{m}{n+m}\frac{n-m+1}{n+m-1} \\ &= \frac{(n-1+m)(n-m)}{(n-1+m)(n+m)} \\ &= \frac{n-m}{n+m} \end{aligned}$$

In the case that $n = m+1$, $P_{n-1,m} = 0$ since B eventually will catch up A. Therefore, although it's not in our assumption, the equation $P_{n-1,m} = 0 = (n-1-m)/(n+m-1)$ is still true.

□

Problem 3

Let n be a positive integer with prime factorization $n = p_1^{m_1} \cdots p_k^{m_k}$ for distinct primes $p_1, \cdots, p_k$ with $m_1, \cdots, m_k > 0$. Choose an integer N uniformly at random from the set $\{1, 2, \cdots, n\}$.. Show that the probability that N shares no common prime factor with n is equal to

$$\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_k}\right).$$

(Hint: use inclusion-exclusion)

Proof. Let $A_i = \{N = ap_i, a \in \mathbb{N}\}$ and for $I \subset \{1, 2, \cdots, n\}$

$$\begin{aligned} \mathbb{P}(\cap_{i\in I} A_i) &= \mathbb{P}(N = a\prod_{i\in I} p_i,\ a \in \mathbb{N}) \\ &= \frac{\frac{n}{\prod_{i\in I} p_i}}{n} \\ &= \prod_{i\in I} \frac{1}{p_i} \end{aligned}$$

Now by inclusion-exclusion theorem,

$$\begin{aligned}
\mathbb{P}(\text{N is co-prime to n}) &= 1 - \mathbb{P}(\cup_{i=1}^{k} A_i) \\
&= 1 - \left(\sum_{i=1}^{k} \frac{1}{p_i} - \sum_{i,j=1}^{k} \frac{1}{p_i p_j} + \cdots + (-1)^{k+1} \frac{1}{\prod_{i=1}^{k} p_i} \right) \\
&= 1 + \sum_{i=1}^{k} \left(-\frac{1}{p_i} \right) + \sum_{i,j=1}^{k} \left(-\frac{1}{p_i} \right) \left(-\frac{1}{p_j} \right) + \cdots + \prod_{i=1}^{k} \left(-\frac{1}{p_i} \right) \\
&= \left(1 - \frac{1}{p_1} \right) \left(1 - \frac{1}{p_2} \right) \cdots \left(1 - \frac{1}{p_k} \right)
\end{aligned}$$

$\square$

MATH 505a Fall 2021 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you find errata.

Problem 1

(a)

Let X be a non-negative random variable with finite expectation. Show that

$$\sum_{i=1}^{\infty} \mathbb{P}(X \geq i) \leq E[X] < 1 + \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof. Since X is non-negative,

$$\begin{aligned}\mathbb{E}[X] &= \int_{[0,\infty)} xf(x)dx \\ &= \sum_{i=0}^{\infty} \int_{[i,i+1)} xf(x)dx\end{aligned}$$

Then notice that,

$$i \int_{[i,i+1)} f(x)dx \leq \int_{[i,i+1)} xf(x)dx \leq (i+1) \int_{[i,i+1)} f(x)dx$$

That is,

$$i\mathbb{P}(i \leq X \leq i+1) \leq \int_{[i,i+1)} xf(x)dx \leq (i+1)\mathbb{P}(i \leq X \leq i+1)$$

Plugging into the sum, the lower bound becomes:

$$\begin{aligned}&\sum_{i=0}^{\infty} i\mathbb{P}(i \leq X \leq i+1) \\ =&\sum_{i=1}^{\infty} \mathbb{P}(X \geq i)\end{aligned}$$

Similarly, the upper bound:

$$\begin{aligned}
&\sum_{i=0}^{\infty}(i+1)\mathbb{P}(i \leq X \leq i+1) \\
&= \sum_{i=0}^{\infty} i\mathbb{P}(i \leq X \leq i+1) + \sum_{i=0}^{\infty}\mathbb{P}(i \leq X \leq i+1) \\
&= \sum_{i=1}^{\infty}\mathbb{P}(X \geq i) + 1
\end{aligned}$$

□

(b)

Show that if X takes values only in $\{0, 1, \cdots, n\}$ for some n, then the first inequality in (a) is an equality:

$$\sum_{i=1}^{\infty}\mathbb{P}(X \geq i) = \mathbb{E}[X].$$

Proof. Note that if X only take natural number values, we have $\mathbb{P}(X \geq i) = \sum_{k=i}^{\infty}\mathbb{P}(X = k)$. Plug this into the left hand side:

$$\begin{aligned}
\sum_{i=1}^{\infty}\mathbb{P}(X \geq i) &= \sum_{i=1}^{\infty}\sum_{k=i}^{\infty}\mathbb{P}(X = k) \\
&= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \cdots \\
&\qquad + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \cdots \\
&\qquad\qquad + \mathbb{P}(X = 3) + \cdots \\
&\qquad\qquad + \cdots \\
&= \sum_{i=1}^{\infty} i\mathbb{P}(X = i) \\
&= \mathbb{E}[X]
\end{aligned}$$

□

(c)

Let M be the minimum value seen in 4 die rolls. Find $\mathbb{E}[M]$. You don't need to simplify to one number, just get an expression in terms of numbers only.

Proof. Note that M only takes values in $\{1, 2, 3, 4, 5, 6\}$, let X_i denotes the value of i-th dice roll we can use the conclusion from (b) that:

$$
\begin{aligned}
\mathbb{E}[M] &= \sum_{i=1}^{6} \mathbb{P}(M \geq i) \\
&= \sum_{i=1}^{6} \prod_{j=1}^{4} \mathbb{P}(X_j \geq i) \\
&= \sum_{i=1}^{6} \left(1 - \frac{i-1}{6}\right)^4
\end{aligned}
$$

□

Problem 2

Suppose X and Y are independent continuous random variables with uniform distribution on $[0, 1]$.

(a)

Find the density function of $X + 2Y$.

Solution. By conditioning on X, we have:
Case 1. $z \in [0, 1)$,

$$
\begin{aligned}
\mathbb{P}(X + 2Y \leq z) &= \int_0^z \frac{1}{2}(z - x)\, dx \\
&= \frac{z^2}{4}
\end{aligned}
$$

Case 2. $z \in [1, 2)$,

$$
\begin{aligned}
\mathbb{P}(X + 2Y \leq z) &= \int_0^1 \frac{1}{2}(z - x)\, dx \\
&= \frac{2z - 1}{4}
\end{aligned}
$$

Case 3. $z \in [2, 3]$,

$$
\begin{aligned}
\mathbb{P}(X + 2Y \leq z) &= (z - 2) + \int_{z-2}^1 \frac{1}{2}(z - x)\, dx \\
&= z - 2 - \frac{z^2 - 2z - 3}{4}
\end{aligned}
$$

So compute the pdf by differentiating:

$$
f_{X+Y}(z) = \begin{cases} \frac{z}{2} & 0 \leq z < 1 \\ \frac{1}{2} & 1 \leq z < 2 \\ -\frac{z}{2} + \frac{3}{2} & 2 \leq z \leq 3 \end{cases}
$$

(b)

Find the joint density function for $X - Y$, $X + Y$.

Solution. Let $U = X + Y$, $V = X - Y$. Then $X = \frac{U+V}{2}$, $Y = \frac{U-V}{2}$. We can compute the absolute value of Jacobian of the map $(u, v) \mapsto (x, y)$:

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = \frac{1}{2}$$

then We have the following:

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(\frac{u+v}{2}, \frac{u-v}{2}) \cdot |J| \\ &= \frac{1}{2} \mathbb{1}_{0 \le \frac{u+v}{2} \le 1, 0 \le \frac{u-v}{2} \le 1}(u, v) \\ &= \frac{1}{2} \mathbb{1}_{0 \le u+v \le 2, 0 \le u-v \le 2}(u, v) \end{aligned}$$

Problem 3

Consider Bernoulli trials with success probability $p \in (0, 1)$. Let p_n be the probability of an odd number of successes in n trials.

(a)

Express p_n in terms of p_{n-1}.

$$p_n = p \cdot (1 - p_{n-1}) + (1 - p) \cdot p_{n-1}$$

(b)

Based on (a), for what value λ does $p_{n-1} = \lambda$ imply $p_n = \lambda$?

$$\begin{aligned} \lambda &= p \cdot (1 - \lambda) + (1 - p) \cdot \lambda \\ \implies \lambda &= \frac{1}{2} \end{aligned}$$

(c)

Show that $lim_n p_n = \lambda$, the value you found in (b). HINT: Write p_n as $\lambda + \epsilon_n$, for the λ you found in (b).

$$\begin{aligned} \lambda + \epsilon_n &= p \cdot (1 - (\lambda + \epsilon_{n-1})) + (1 - p) \cdot (\lambda + \epsilon_{n-1}) \\ \epsilon_n &= (1 - 2p)\epsilon_{n-1} \\ \overset{(*)}{\implies} \lim_{n \to \infty} \epsilon_n &= \lim_{n \to \infty} (1 - 2p)^{n-1} \epsilon_1 = 0 \end{aligned}$$

$(*) : |1 - 2p| < 1$

MATH 505a Spring 2021 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

A permutation π on n symbols is said to have i as a fixed point if $\pi(i) = i$.

(a)

Find the probability p_n that a random permutation of n symbols has no fixed points. HINT: Principle of inclusion and exclusion. (Your answer may involve a finite sum, which you don't need to simplify.)

Solution. Use inclusion-exclusion: let A_i be the set of permutations that $\pi(i) = i$.

$$\begin{aligned}
p_n &= 1 - \mathbb{P}(\cup_{i=1}^n A_i) \\
&= 1 - (\sum_i \mathbb{P}(A_i) - \sum_{i,j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n+1}\mathbb{P}(\cap_{i=1}^n A_i) \\
&= 1 - n \cdot \frac{(n-1)!}{n!} + \binom{n}{2} \cdot \frac{(n-2)!}{n!} + \cdots + (-1)^n \binom{n}{n} \cdot \frac{1}{n!} \\
&= \sum_{p=0}^n (-1)^p \frac{1}{p!}
\end{aligned}$$

(b)

Let S be a subset of $\{1, 2, \cdots, n\}$ of size k. Find the probability that the set of fixed points of a random permutation on n symbols is equal to S, and find the probability that a permutation has exactly k fixed points. HINT: If you didn't find the values p_j in part(a), you can still give answers for (b) expressed in terms of one or more p_j's.

Solution.

$$\begin{aligned}\mathbb{P}(\{\text{fixed points}\} = S) &= \mathbb{P}(\pi(i) = i,\ \forall i \in S) \cdot \mathbb{P}(\pi(j) \neq j,\ \forall j \in S^c | \pi(i) = i,\ \forall i \in S) \\ &= \frac{(n-k)!}{n!} p_{n-k} \\ \mathbb{P}(\text{k fixed points}) &= \binom{n}{k} \cdot \mathbb{P}(\{\text{fixed points}\} = S) \\ &= \frac{p_{n-k}}{k!}\end{aligned}$$

We get the second probability knowing that there are $\binom{n}{k}$ many sets with k fixed points.

(c)

Show that as n tends to infinity, the distribution of the number of fixed points converges to a Poisson(1) distribution.

Proof.

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(\text{k fixed points}) &= \frac{1}{k!} \lim_{n\to\infty} \sum_{p=0}^{n-k} \frac{(-1)^p}{p!} \\ &= \frac{1}{k!} \sum_{p=0}^{\infty} \frac{(-1)^p}{p!} \\ &= \frac{e^{-1}}{k!} \\ &\sim Poisson(1)\end{aligned}$$

□

Problem 2

Let $\{S_n,\ n \geq 0\}$ be symmetric simple random walk, that is, $S_n = \sum_{i=1}^{n} \xi_i$ with $\xi_1, \xi_2, \cdots$ i.i.d. satisfying $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = \frac{1}{2}$. Let $T = min\{n : S_n = 0\}$, and write $\mathbb{P}_a$ for probabilities when the walk starts at $S_0 = a$. By *basic probabilities* for $\{S_n\}$ we mean probabilities of the form $\mathbb{P}_0(S_n = k)$, $\mathbb{P}_0(S_n \geq k)$, or $\mathbb{P}_0(S_n \leq k)$, all of which corresponding to starting at $S_0 = 0$.

(a)

For $a \geq 1$, $i \geq 1$, $n \geq 1$, express $\mathbb{P}_a(S_n = i, T \leq n)$ and $\mathbb{P}_a(S_n = i, T > n)$ in terms of finitely many basic probabilities. HINT: Reflection principle.

Solution. Use reflection(reflect the part of path after the first approach at 0, with respect to 0), we have:

$$\mathbb{P}_a(S_n = i,\ T \leq n) = \mathbb{P}_a(S_n = -i) = \mathbb{P}_0(S_n = i + a)$$

Use conditional probability, we have:

$$\begin{aligned}
\mathbb{P}_a(S_n = i, T > n) &= \mathbb{P}_a(S_n = i) \cdot \mathbb{P}_a(T > n | S_n = i) \\
&= \mathbb{P}_a(S_n = i) \cdot (1 - \mathbb{P}_a(T \le n | S_n = i)) \\
&= \mathbb{P}_a(S_n = i) \cdot (1 - \frac{\mathbb{P}_a(T \le n, S_n = i)}{\mathbb{P}_a(S_n = i)}) \\
&= \mathbb{P}_a(S_n = i) - \mathbb{P}_a(S_n = i, T \le n) \\
&= \mathbb{P}_0(S_n = i - a) - \mathbb{P}_0(S_n = i + a)
\end{aligned}$$

(b)

For $a \ge 1$, $i \ge 1$, $n \ge 1$, show that

$$\mathbb{P}_a(T > n) = \sum_{j=1-a}^{a} \mathbb{P}_0(S_n = j).$$

HINT: use (a) and look for cancellation

Proof.

$$\begin{aligned}
\mathbb{P}_a(T > n) &= \sum_{i=1}^{a+n} \mathbb{P}_a(S_n = i,\ T > n) \\
&= \sum_{i=1}^{a+n} \mathbb{P}_0(S_n = i - a) - \mathbb{P}_0(S_n = i + a) \\
&= \sum_{i=1-a}^{n} \mathbb{P}_0(S_n = i) - \sum_{j=1+a}^{n} \mathbb{P}_0(S_n = j) \\
&= \sum_{j=1-a}^{a} \mathbb{P}_0(S_n = j)
\end{aligned}$$

□

(c)

You may take as given that $\mathbb{P}_0(S_{2m} = 2j) \sim 1/\sqrt{\pi m}$ as $m \to \infty$ for each fixed $j \in \mathbb{Z}$; here $\sim$ means that ratio converges to 1. Use this to find c, α such that $\mathbb{P}_a(T > n) \sim c/n^\alpha$ as $n \to \infty$, where $a > 0$. Does c or α depend on a? HINT: It's enough to consider even n - why?

Proof. Assume n is even where $n = 2m$. For very large n, we have:

$$\begin{aligned}
\mathbb{P}_a(T > 2m) &= \sum_{j=1-a}^{a} \mathbb{P}_0(S_{2m} = j) \\
&= \sum_{j \in A} \mathbb{P}_0(S_{2m} = j), \ A = \{\text{even numbers in } \{1-a, 2-a, \cdots, a\}\} \\
&\sim a \cdot \frac{1}{\sqrt{\pi m}} \\
&= \frac{a\sqrt{\frac{2}{\pi}}}{n^{1/2}}
\end{aligned}$$

So we get $c = a\sqrt{\frac{2}{\pi}}$ and $\alpha = \frac{1}{2}$, where c depends on a, α does not.

Now we assume n is odd, and we will prove the convergence by squeezing. First by inclusion, we have the inequality:

$$\mathbb{P}_a(T > n-1) \geq \mathbb{P}_a(T > n) \geq \mathbb{P}_a(T > n+1)$$

divide the expected limit:

$$\frac{\mathbb{P}_a(T > n-1)}{c/n^{\alpha}} \geq \frac{\mathbb{P}_a(T > n)}{c/n^{\alpha}} \geq \frac{\mathbb{P}_a(T > n+1)}{c/n^{\alpha}}$$

normalize both sides:

$$\frac{\mathbb{P}_a(T > n-1)}{c/(n-1)^{\alpha}} \cdot \left(\frac{n}{n-1}\right)^{\alpha} \geq \frac{\mathbb{P}_a(T > n)}{c/n^{\alpha}} \geq \frac{\mathbb{P}_a(T > n+1)}{c/(n+1)^{\alpha}} \cdot \left(\frac{n}{n+1}\right)^{\alpha}$$

Now, notice $n-1$ and $n+1$ are even, so if we let n go to infinity, both upper and lower bound above will converge to 1. □

Problem 3

Let X, Y be independent standard normal $(0, 1)$ random variables.

(a)

Find a for which $U = X + 2Y$, $V = aX + Y$ are independent.

Solution. Note that $U = (1, 2) \cdot (X, Y)^T$, $V = (a, 1) \cdot (X, Y)^T$, and $(X, Y)^T \sim \mathcal{N}(0, I)$. (U, V) are normal vector, so U, V are independent if and only if $Cov(U, V) = 0$.

$$\begin{aligned}
Cov(U, V) &= (1, 2) \cdot I \cdot (a, 1)^T \\
&= a + 2 \\
a &= -2
\end{aligned}$$

(b)

Find $\mathbb{E}(XY|X+2Y=a)$ for all $a \in \mathbb{R}$. HINT: Use(a).

Solution. Note that $X = \frac{U-2V}{5}$ and $Y = \frac{2U+V}{5}$. So the expectation turns into:

$$\frac{1}{25}\mathbb{E}(2U^2 - 3UV - 2V^2|U=a) = \frac{1}{25}(2a^2) - 3a \cdot \mathbb{E}(V) - 2 \cdot \mathbb{E}(V^2)) = \frac{2a^2 - 10}{25}$$

MATH 505a Fall 2020 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Let X_n have binomial $B(n,p)$ distribution.

(a)

Find $\mathbb{E}(\frac{1}{X_n+1})$. Simplify your answer so it does not involve a sum to n, $n+1$, etc.

Solution.

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{X_n+1}\right) &= \sum_{k=0}^{n} \frac{1}{k+1}\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \frac{n!}{(n-k)!(k+1)!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \frac{(n+1)!}{(n-k)!(k+1)!} \frac{1}{n+1} p^{k+1} (1-p)^{n-k} \frac{1}{p} \\
&= \frac{1}{(n+1)p} \sum_{k=1}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \\
&= \frac{1}{(n+1)p} \left(1-(1-p)^{n+1}\right)
\end{aligned}$$

(b)

Suppose $p = p_n$ and $np_n \to \lambda$ as $n \to \infty$, with $\lambda \in (0,\infty)$. Find $\lim_n \mathbb{E}(\frac{1}{X_n+1})$. Is it the same as $\lim_n \frac{1}{\mathbb{E}(X_n+1)}$?

Solution.

$$\begin{aligned}\lim_n \mathbb{E}\left(\frac{1}{X_n+1}\right) &= \lim_n \frac{1-(1-p)^{n+1}}{(n+1)p} \\ &= \frac{1}{\lambda}\left(1-\lim_n\left(1-\frac{np}{n}\right)^{n+1}\right) \\ &= \frac{1-e^{-\lambda}}{\lambda}\end{aligned}$$

It is not same as $\lim_n \frac{1}{\mathbb{E}(X_n+1)} = \frac{1}{\lambda+1}$.

Problem 2

Let X, Y be independent with $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$ distribution.

(a)

Find $\mathbb{P}(X = k|X+Y = n)$ for $0 \leq k \leq n$. Simplify your answer so it does not involve a sum. Do the actual calculation, don't just cite a theorem.

Solution.

$$\begin{aligned}\mathbb{P}(X=k|X+Y=n) &= \frac{\mathbb{P}(X=k, Y=n-k)}{\sum_{l=0}^n \mathbb{P}(X=l, Y=n-l)} \\ &= \frac{e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}}{\sum_{l=0}^n e^{-\lambda}\frac{\lambda^l}{l!}e^{-\mu}\frac{\mu^{n-l}}{(n-l)!}} \\ &= \frac{\binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}}{\sum_{l=0}^n \binom{n}{l}\left(\frac{\lambda}{\lambda+\mu}\right)^l\left(\frac{\mu}{\lambda+\mu}\right)^{n-l}} \\ &= \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}\end{aligned}$$

(b)

Find $\mathbb{E}(X^2+Y^2|X+Y=n)$.

Solution.

$$\begin{aligned}\mathbb{E}(X^2+Y^2|X+Y=n) &= \sum_{k=0}^n (k^2+(n-k)^2)\binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k} \\ &= \mathbb{E}(M^2)+\mathbb{E}((n-M)^2)\end{aligned}$$

where $M \sim Binomial(n, \frac{\lambda}{\lambda+\mu})$. Given that $\mathbb{E}(M) = \frac{n\lambda}{\lambda+\mu}$ and $Var(M) = \frac{n\lambda\mu}{(\lambda+\mu)^2}$, we have:

$$\mathbb{E}(X^2+Y^2|X+Y=n) = 2\left(\frac{n\lambda\mu}{(\lambda+\mu)^2}+\left(\frac{n\lambda}{\lambda+\mu}\right)^2\right)+n^2-\frac{2\lambda n^2}{\lambda+\mu}$$

Problem 3

The county hospital is located at the center of a square whose sides are 2 miles wide. If an accident occurs within this square, then the hospital sends out an ambulance. The road network is rectangular, so the travel distance from the hospital, at $(0,0)$, to the point (x, y) is $|x| + |y|$. If an accident occurs at a point that is uniformly distributed in the square, find the mean and variance of the travel distance of the ambulance.

Solution.

$$\begin{aligned}\mathbb{E}(|X| + |Y|) &= \int_{-1}^{1}\int_{-1}^{1}(|x| + |y|)\cdot\frac{1}{4}dxdy \\ &\overset{(*)}{=} \int_{0}^{1}\int_{0}^{1}(x+y)dxdy \\ &= 1\end{aligned}$$

$(*)$ by symmetry.

$$\begin{aligned}Var(|X| + |Y|) &= \mathbb{E}((|X| + |Y|)^2) - (\mathbb{E}(|X| + |Y|))^2 \\ &= \int_{0}^{1}\int_{0}^{1}(x+y)^2 dxdy - 1 \\ &= \frac{7}{6} - 1 \\ &= \frac{1}{6}\end{aligned}$$

Problem 4

Let X be a finite set X, and let P and Q be probabilities on X. Define the total variation distance between P and Q by

$$||P - Q||_{TV} = \frac{1}{2}\sum_{x\in X}|P(x) - Q(x)|.$$

Prove that

$$||P - Q||_{TV} = \max_{A\subset X}|P(A) - Q(A)|,$$

where the maximum is over subsets A of X.

Proof. Let $S = \{x \in X : P(X) \geq Q(x)\}$, then,

$$\begin{aligned}||P - Q||_{TV} &= \frac{1}{2}\left(\sum_{x\in S}(P(x) - Q(x)) + \sum_{x\in S^c}(Q(x) - P(x))\right) \\ &= \frac{1}{2}\left(P(S) - Q(S) + Q(S^c) - P(S^c)\right) \\ &\overset{(*)}{=} P(S) - Q(S)\end{aligned}$$

$(*)$ for any $A \subset X$,

$$P(A) + P(A^c) = Q(A) + Q(A^c) = 1 \implies P(A) - Q(A) = Q(A^c) - P(A^c)$$

Now it suffices to show that $\max_{A \subset X} |P(A) - Q(A)| = P(S) - Q(S)$. Given $A \subset X$,

$$\begin{aligned}
|P(A) - Q(A)| &= |(P(A \cap S) + P(A \cap S^c)) - (Q(A \cap S) + Q(A \cap S^c))| \\
&= |(P(A \cap S) - Q(A \cap S)) - (Q(A \cap S^c) - P(A \cap S^c))| \\
&\overset{(**)}{\leq} \max\{P(A \cap S) - Q(A \cap S), Q(A \cap S^c) - P(A \cap S^c)\} \\
&\overset{(***)}{\leq} \max\{P(S) - Q(S), Q(S^c) - P(S^c)\} \\
&\overset{(*)}{=} P(S) - Q(S)
\end{aligned}$$

$(**)$ $P(A \cap S) - Q(A \cap S) \geq 0$, $Q(A \cap S^c) - P(A \cap S^c) \geq 0$ by the definition of S.
$(* * *)$ Any subset $B \subset S$, $0 \leq P(B) - Q(B) \leq P(S) - Q(S)$, by the definition of S. Similarly, any $C \subset S^c$, $0 \leq Q(C) - P(C) \leq Q(S^c) - P(S^c)$. $\square$

MATH 505a Spring 2020 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Each pack of bubble gum contains one of n types of coupon, equally likely to be each of the types, independently from one pack to another. Let T_j be number of packs you must buy to obtain coupons of j different types. Note that $T_1 = 1$ always.

(a)

Find the distribution and expected value of $T_2 - T_1$ and of $T_3 - T_2$.

Solution. Notice that these differences are geometric distributions:

$$\mathbb{P}(T_2 - T_1 = t) = \left(\frac{n-1}{n}\right)\left(\frac{1}{n}\right)^{t-1},\ t \geq 1.$$

$$\mathbb{P}(T_3 - T_2 = t) = \left(\frac{n-2}{n}\right)\left(\frac{2}{n}\right)^{t-1},\ t \geq 1.$$

$$\mathbb{E}(T_2 - T_1) = \frac{n}{n-1}$$

$$\mathbb{E}(T_3 - T_2) = \frac{n}{n-2}$$

(b)

Compute $\mathbb{E}T_n$.

Solution.

$$\begin{aligned}\mathbb{E}(T_n) &= \mathbb{E}[(T_n - T_{n-1}) + (T_{n-1} - T_{n-2}) + \cdots + (T_2 - T_1) + T_1] \\ &= n + \frac{n}{2} + \cdots + \frac{n}{n-1} + 1 \\ &= \sum_{k=1}^{n} \frac{n}{k}\end{aligned}$$

(c)

Fix k and let A_i be the event that none of the first k packs you buy contain coupon i. Find $\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4)$. Then fix $\alpha > 0$, take $k = \lfloor \alpha n \rfloor$ and find the limit of this probability as $n \to \infty$. Here $\lfloor x \rfloor$ denotes the largest integer $\leq x$. HINT: Consider probabilities $\mathbb{P}(A_i), \mathbb{P}(A_i \cap A_j)$, etc.

Solution. By inclusion-exclusion theorem,

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4) &= \sum_{i=1}^{4} \mathbb{P}(A_i) - \sum_{i,j=4}^{4} \mathbb{P}(A_i \cap A_j) + \sum_{i,j,k=1}^{4} \mathbb{P}(A_i \cap A_j \cap A_k) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) \\
&= \binom{4}{1}\left(\frac{n-1}{n}\right)^k - \binom{4}{2}\left(\frac{n-2}{n}\right)^k + \binom{4}{3}\left(\frac{n-3}{n}\right)^k - \binom{4}{4}\left(\frac{n-4}{n}\right)^k \\
&= 4\left(1-\frac{1}{n}\right)^k - 6\left(1-\frac{2}{n}\right)^k + 4\left(1-\frac{3}{n}\right)^k - \left(1-\frac{4}{n}\right)^k
\end{aligned}$$

Now, take $k = \lfloor \alpha n \rfloor$ and $n \to \infty$, we have:

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4) &= \lim_{n\to\infty} \left(4\left(1-\frac{1}{n}\right)^{\lfloor \alpha n \rfloor} - 6\left(1-\frac{2}{n}\right)^{\lfloor \alpha n \rfloor} + 4\left(1-\frac{3}{n}\right)^{\lfloor \alpha n \rfloor} - \left(1-\frac{4}{n}\right)^{\lfloor \alpha n \rfloor}\right) \\
&= \lim_{n\to\infty} \left(4\left(1-\frac{1}{n}\right)^{\alpha n} - 6\left(1-\frac{2}{n}\right)^{\alpha n} + 4\left(1-\frac{3}{n}\right)^{\alpha n} - \left(1-\frac{4}{n}\right)^{\alpha n}\right) \\
&= 4e^{-\alpha} - 6e^{-2\alpha} + 4e^{-3\alpha} - e^{-4\alpha}
\end{aligned}$$

(d)

Assume there are $n = 4$ coupon types; find $\mathbb{P}(T_4 > k)$ for all $k \geq 4$. HINT: This is short if you use what you've already done.

Solution. From (c), we have:

$$\mathbb{P}(T_4 > k) = 4\left(\frac{3}{4}\right)^k - 6\left(\frac{1}{2}\right)^k + 4\left(\frac{1}{4}\right)^k, \ k \geq 4$$

Problem 2

Let X be exponential(λ) (that is, density $f(x) = \lambda e^{-\lambda x}$). The *integer part* of X is $\lfloor X \rfloor = \max\{k \in \mathbb{N} : k \leq X\}$. The *fractional part* of X is $X - \lfloor X \rfloor$. Show that $\lfloor X \rfloor$ and $X - \lfloor X \rfloor$ are independent. HINT: In general, two random variables U, V are independent if the distribution of V conditioned on $U = u$ doesn't depend on u.

Proof. It suffices to show the conditional probability is same as the unconditioned one:

$$\begin{aligned}
\mathbb{P}(\lfloor X \rfloor = n | X - \lfloor X \rfloor = \alpha) &= \frac{f_X(n+\alpha)}{\sum_{k=0}^{\infty} f_X(k+\alpha)} \\
&= \frac{\lambda e^{-\lambda(n+\alpha)}}{\sum_{k=0}^{\infty} \lambda e^{-\lambda(k+\alpha)}} \\
&= \frac{e^{-\lambda n}}{\sum_{k=0}^{\infty} e^{-\lambda k}} \\
&= \frac{e^{-\lambda n}}{\frac{1}{1-e^{-\lambda}}} \\
&= e^{-\lambda n} - e^{-\lambda(n+1)} \\
&\overset{(*)}{=} \mathbb{P}(n \leq X < n+1) \\
&= \mathbb{P}(\lfloor X \rfloor = n)
\end{aligned}$$

$(*)$ X is continuous. □

Problem 3

Let X_1, X_2, X_3 be i.i.d. uniform in [0,1]. Let $X_{(1)}$ be the smallest of the 3 values, $X_{(2)}$ the second smallest, and $X_{(3)}$ the largest.

(a)

Find the distribution function and expected value for $X_{(1)}$.

Solution. Let $F_{(i)}$ denotes the cdf of $X_{(i)}$.

$$\begin{aligned}
F_{(1)}(x) &= \mathbb{P}(X_{(1)} \leq x) \\
&= 1 - \mathbb{P}(X_{(1)} > x) \\
&= 1 - \prod_{i=1}^{3} \mathbb{P}(X_i > x) \\
&= 1 - (1-x)^3, \ 0 < x < 1
\end{aligned}$$

Since $X_{(1)} \geq 0$, we can compute expectation using complementary cdf:

$$\begin{aligned}
\mathbb{E}(X_{(1)}) &= \int_0^1 (1-x)^3 dx \\
&= \frac{1}{4}
\end{aligned}$$

(b)

Find the distribution function and the density of $X_{(2)}$.

Solution.

$$\begin{aligned}
\mathbb{P}(X_{(2)} \leq x) &\overset{(*)}{=} (\mathbb{P}(X_1 \leq x))^3 + \binom{3}{2}(\mathbb{P}(X_1 \leq x))^2(\mathbb{P}(X_1 > x)) \\
&= x^3 + 3x^2(1-x),\ 0 < x < 1 \\
f_{(2)}(x) &= 6x - 6x^2,\ 0 < x < 1
\end{aligned}$$

MATH 505a Fall 2019 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Suppose A, B, C are pairwise independent, $A \cap B \cap C = \emptyset$, and $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = p$.

(a)

What is the largest possible value of p?

Solution. First, notice that $A \cap B \cap C = \emptyset \implies (A \cap B) \cap (B \cap C) = \emptyset$. Then

$$p = \mathbb{P}(B) \geq \mathbb{P}(A \cap B) + \mathbb{P}(B \cap C) = \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(B)\mathbb{P}(C) = 2p^2$$

We have

$$p \leq \frac{1}{2}$$

Then, we let $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$. ($\mathbb{P}(A \cap B) = \frac{1}{4}$ and $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{3}{4}$). And let $C = (A \cup B) \backslash (A \cap B)$. Now we want to show that such A, B, C satisfies the assumptions.

$$\mathbb{P}(C) = \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B) = \frac{1}{2}$$

$$\mathbb{P}(C \cap A) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \frac{1}{4}$$

$$\mathbb{P}(C \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{1}{4}$$

$$A \cap B \cap C = (A \cap B) \cap ((A \cup B) \backslash (A \cap B)) = \emptyset$$

Therefore $\frac{1}{2}$ is the largest possible value for p.

(b)

Is it possible that $\mathbb{P}(A \cup B \cup C) = 1$? Prove or disprove.

Solution. Impossible.

Proof. By inclusion-exclusion theorem,

$$\begin{aligned}\mathbb{P}(A\cup B\cup C) &= \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-(\mathbb{P}(A\cap B)+\mathbb{P}(A\cap C)+\mathbb{P}(B\cap C))+\mathbb{P}(A\cap B\cap C)\\ &= 3p-3p^2\end{aligned}$$

which reaches maximum at $p = 0.5$ and $\mathbb{P}(A\cup B\cup C) = 0.75$. □

Problem 2

Consider two coins: coin 1 shows heads with probability p_1 and coin 2 shows heads with probability p_2. Each coin is tossed repeatedly. Let T_i be the time of first heads for coin i, and define the event $A = \{T_1 < T_2\}$.

(a)

Find $\mathbb{P}(A)$. HINT: One possible method is to condition on one of the variables.

Solution. Notice that $T_i \sim \text{Geometric}(p_i)$.

$$\begin{aligned}\mathbb{P}(A) &= \sum_{t=2}^{\infty}\mathbb{P}(T_1\le t-1)\mathbb{P}(T_2=t)\\ &= \sum_{t=2}^{\infty}(1-(1-p_1)^{t-1})(p_2(1-p_2)^{t-1})\\ &= p_2\left(\sum_{t=1}^{\infty}(1-p_2)^t-\sum_{t=1}^{\infty}[(1-p_2)(1-p_1)]^t\right)\\ &= \frac{p_1(1-p_2)}{1-(1-p_1)(1-p_2)}\end{aligned}$$

(b)

Find $\mathbb{P}(T_1 = k|A)$ for all $k \ge 1$.

Solution.

$$\begin{aligned}\mathbb{P}(T_1=k|A) &= \frac{\mathbb{P}(T_1=k)\mathbb{P}(T_2>k)}{\mathbb{P}(A)}\\ &= \frac{[1-(1-p_1)(1-p_2)](1-p_1)^{k-1}p_1(1-p_2)^k}{p_1(1-p_2)}\\ &= [1-(1-p_1)(1-p_2)](1-p_1)^{k-1}(1-p_2)^{k-1}\end{aligned}$$

Problem 3

Player A and B are having a table tennis match; the first player to win 3 games wins the match. One of the players is better than the other; this better player wins each game with probability 0.7. Carl comes to watch the match. He does not know who is the better player so (based on Carl's information) A, B each initially have probability 0.5 to be the better player. Then Carl sees A win 2 of the first 3 games.

(a)

What is now the probability (after the 3 games, based on Carl's information) that A is the better player? Simplify your answer to a single fraction or decimal.

Solution. Let $C = \{\text{A won 2 of the first 3 games}\}$, $A = \{\text{A is better}\}$, and $B = \{\text{B is better}\}$.

$$\begin{aligned}
\mathbb{P}(A|C) &= \mathbb{P}(C|A)\frac{\mathbb{P}(A)}{\mathbb{P}(C)} \\
&= \binom{3}{2}(0.7)^2(0.3)\frac{0.5}{\mathbb{P}(C \cap A) + \mathbb{P}(C \cap B)} \\
&= \binom{3}{2}(0.7)^2(0.3)\frac{0.5}{\binom{3}{2}(0.7)^2(0.3)0.5 + \binom{3}{2}(0.3)^2(0.7)0.5} \\
&= \frac{7}{10}
\end{aligned}$$

(b)

What is now the probability (after 3 games, based on Carl's information) that A will go on to win the match? NOTE: Express your answer for (b) in terms of numbers; you do not need to simplify to a single number. Ananswer in a form like $\frac{5}{4} + 7(2 - \frac{9}{5})$ is OK.

Solution. Let $W_A = \mathbb{P}(\text{A win a game}|C) = 0.7 \cdot 0.7 + 0.3 \cdot 0.3 = 0.58$, so $W_B = \mathbb{P}(\text{B win a game}|C) = 0.42$. So,

$$\begin{aligned}
\mathbb{P}(\text{A win the match}|C) &= W_A + W_B W_A \\
&= 0.58 + 0.42 \cdot 0.58 \\
&= 0.8236
\end{aligned}$$

Problem 4

Suppose X_n is binomial with parameters (n, p) with $0 \leq p \leq 1$, and X is Poisson(λ).

(a)

Find the moment generating function of X_n.

Solution.

$$\begin{aligned}
\mathbb{E}(e^{X_n t}) &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}p(pe^t)^k(1-p)^{n-k} \\
&\overset{(*)}{=} (1-p+pe^t)^n
\end{aligned}$$

$(*)$ Binomial formula.

(b)

Suppose $n \to \infty$ and $p = p_n \to 0$ with $np \to \lambda \in (0, \infty)$. Show that $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$ as $n \to \infty$, for all $k \geq 0$. HINT:$(1 - \frac{c_n}{n})^n \to e^{-c}$ if $c_n \to c$.

Proof.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_n = k) &= \lim_{n\to\infty} \binom{n}{k}p^k(1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{1}{k!}\frac{n!}{(n-k)!}p^k(1-\frac{np}{n})^n(1-\frac{np}{n})^{-k} \\
&= \frac{1}{k!}\lim_{n\to\infty}\frac{n-k+1}{n}\frac{n-k+2}{n}\cdots\frac{n-1}{n}\frac{n}{n}(np)^k(1-\frac{np}{n})^n(1-\frac{np}{n})^{-k} \\
&= \frac{1}{k!}\cdot 1\cdot 1\cdots 1\cdot 1\cdot \lambda^k\cdot e^{-\lambda}\cdot 1 \\
&= e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}$$

□

(c)

For n, p as in part(b), show that $\mathbb{P}(X_n > k) \to \mathbb{P}(X > k)$ as $n \to \infty$, for all $k \geq 0$.

Proof.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_n > k) &= \lim_{n\to\infty}(1 - \mathbb{P}(X_n \leq k) \\
&= 1 - \lim_{n\to\infty}\sum_{x=0}^{k}\mathbb{P}(X_n = x) \\
&= 1 - \sum_{x=0}^{k}\lim_{n\to\infty}\mathbb{P}(X_n = x) \\
&\overset{(**)}{=} 1 - \sum_{x=0}^{k}\mathbb{P}(X = x) \\
&= \mathbb{P}(X > k)
\end{aligned}$$

$(**)$ Conclusion of (b). □

MATH 505a Spring 2019 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Suppose that each of 5 jobs is assigned at random to one of three servers A,B and C. [For example, one possible outcome would be that job 1 goes to server B, job 2 goes to server C, job 3 goes to server C, job 4 goes to server B and job 5 goes to server A. "At random" here means that there are 3^5 equally likely outcomes]

(a)

Find the probability that server C gets all 5 jobs.

$$\mathbb{P}(\text{C gets 5 jobs}) = \left(\frac{1}{3}\right)^5$$

(b)

Let S be the number of servers that get exactly one job. Find $\mathbb{E}S$.

Solution. Let I_A, I_B, I_C denote the indicator functions of A, B, C get exactly one job, respectively. Then,

$$\begin{aligned}
\mathbb{E}(S) &= \mathbb{E}(I_A + I_B + I_C) \\
&= \mathbb{P}(I_A = 1) + \mathbb{P}(I_B = 1) + \mathbb{P}(I_C = 1) \\
&= 3\binom{5}{1}\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^4 \\
&= \frac{80}{81}
\end{aligned}$$

(c)

Find the probability that no server gets more than 2 jobs.

$$\begin{aligned}\mathbb{P}(\text{no server gets more than 2 jobs}) &= \mathbb{P}(\text{2 servers get 2 jobs each, 1 server get 1 job})\\ &= \binom{3}{1}\left(\frac{1}{3}\right)^2\left(\frac{1}{3}\right)^2\left(\frac{1}{3}\right)\\ &= 3\left(\frac{1}{3}\right)^5\\ &= \frac{1}{81}\end{aligned}$$

(d)

Take the same story, but with m in pace of 5 for the number of jobs, and n in place of 3 for the number of servers. Find the variance of S, in terms of m and n.

Solution. Let I_i denotes the indicator function that the i-th server gets exactly 1 job.

$$\begin{aligned}Var(S) &= \mathbb{E}(S^2) - (\mathbb{E}S)^2\\ &= \mathbb{E}(\sum_{i=1}^{n} I_i^2 + 2\sum_{i\neq j}^{n} I_i \cdot I_j) - \left(\sum_{i=1}^{n}\mathbb{E}(I_i)\right)^2\\ &= \sum_{i=1}^{n}\mathbb{P}(I_i = 1) + 2\sum_{i\neq j}^{n}\mathbb{P}(I_i = 1, I_j = 1) - \left(\sum_{i=1}^{n}\mathbb{P}(I_i = 1)\right)^2\\ &= m\left(\frac{n-1}{n}\right)^{m-1} - \left(m\left(\frac{n-1}{n}\right)^{m-1}\right)^2 + 2\cdot n(n-1)\cdot m(m-1)\cdot\left(\frac{1}{n}\right)^2\left(\frac{n-2}{n}\right)^{m-2}\end{aligned}$$

Problem 2

(a)

Suppose that X is Poisson with parameter λ. Find the characteristic function of X.

$$\begin{aligned}\mathbb{E}(e^{itX}) &= \sum_{k=0}^{\infty} e^{itk}e^{-\lambda}\frac{\lambda^k}{k!}\\ &= e^{-\lambda(1-e^{it})}\sum_{k=0}^{\infty} e^{-e^{it}\lambda}\frac{(e^{it}\lambda)^k}{k!}\\ &= e^{-\lambda(1-e^{it})}\end{aligned}$$

(b)

Suppose that X_n is Poisson with parameter λ_n and that $\lambda_n \to \infty$. Show using characteristic functions that $(X_n - \lambda_n)/\sqrt{\lambda}$ converges in distribution, and describe the limiting distribution.

Proof.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}(e^{it(X_n-\lambda_n)/\sqrt{\lambda}}) &= \lim_{n\to\infty} \exp\left(-it\sqrt{\lambda_n} - \lambda_n + \lambda_n e^{i\frac{t}{\sqrt{\lambda_n}}}\right) \\
&= \lim_{n\to\infty} \exp\left(-it\sqrt{\lambda_n} - \lambda_n + \lambda_n \sum_{k=0}^{\infty} \frac{(it/\sqrt{\lambda_n})^k}{k!}\right) \\
&= \lim_{n\to\infty} \exp\left(-it\sqrt{\lambda_n} + \lambda_n(it/\sqrt{\lambda_n}) + \lambda_n \frac{(it/\sqrt{\lambda_n})^2}{2} + \lambda_n \frac{(it/\sqrt{\lambda_n})^3}{6} + \cdots\right) \\
&= \lim_{n\to\infty} \exp\left(-\frac{t^2}{2} - \frac{it^3}{6\lambda^{1/2}} + \cdots\right) \\
&= \exp\left(-\frac{t^2}{2}\right)
\end{aligned}$$

It converges to the standard normal distribution. □

Problem 3

A stick of length 1 is broken into two pieces at a uniformly distributed random point.

(a)

Find the expected length of the smaller piece.

Solution. Let X_1, X_2, U_1 denote the length of the smaller stick, the length of the larger stick, and the location of the first break point, respectively.

$$\begin{aligned}
F_{X_1}(x) &= \mathbb{P}(X_1 \le x, U_1 \le 1/2) + \mathbb{P}(X_1 \le x, U_1 > 1/2) \\
&= x + (1 - (1-x)) \\
&= 2x
\end{aligned}$$

$$f_{X_1}(x) = 2,\ 0 < x < 1/2$$

Similarly, we have

$$F_{X_2}(x) = 1 - \mathbb{P}(X_1 < 1 - x) = 2x - 1$$

$$f_{X_2}(x) = 2,\ 1/2 < x < 1$$

From the pdf of X_1, we have,

$$\mathbb{E}(X_1) = \frac{1}{4}$$

(b)

Find the expected value of the ratio of the smaller length over the larger.

Solution.

$$\begin{aligned}
\mathbb{P}(\frac{X_1}{X_2} \leq t) &= \mathbb{P}(\frac{X_1}{1-X_1} \leq t) \\
&= \mathbb{P}(X_1 \leq \frac{t}{t+1}) \\
&= \frac{2t}{t+1}, \ \ 0 < t < 1
\end{aligned}$$

Since the ratio only takes non-negative value, we can use complementary cdf to compute expectation:

$$\begin{aligned}
\mathbb{E}\left(\frac{X_1}{X_2}\right) &= \int_0^1 1 - \frac{2t}{1+t} dt \\
&= 1 - 2 + 2\int_0^1 \frac{1}{1+t} dt \\
&= -1 + 2\ln(2)
\end{aligned}$$

(c)

Suppose the larger piece is then broken at a random point, uniformly distributed over its length, independent of the first break point. There are then three pieces. Find the probability the longest of the three has length more than 1/2.

Solution. Let X_3, U_2 be the length of the larger piece of the previous larger piece and the second break point location (start from the left end of the X_2). First notice that the pdf of U_2:

$$F_{U_2|X_2=x}(t) = \frac{t}{x}$$

Since $\mathbb{P}(X_1 > 1/2) = 0$, so the desired probability is

$$\begin{aligned}
\mathbb{P}(X_3 > 1/2) &= \int_{1/2}^1 \mathbb{P}(X_3 > 1/2, U_2 > x/2) + \mathbb{P}(X_3 > 1/2, U_2 \leq x/2) \cdot f_{X_2}(x) dx \\
&= \int_{1/2}^1 [\mathbb{P}(U_2 > 1/2|X_2 = x) + \mathbb{P}(U_2 < X_2 - 1/2|X_2 = x) \cdot 2dx \\
&= 2\int_{1/2}^1 1 - \frac{1}{2x} + \frac{1}{x}(x - \frac{1}{2}) dx \\
&= 2 + 2\ln(1/2)
\end{aligned}$$

MATH 505a Fall 2018 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Let X be exponentially distributed random variable with $\mathbb{P}(X > t) = e^{-rt}$ for $t > 0$. Write X as the sum of its integer and fractional parts: $X = Y + Z$ with $Y = \lfloor X \rfloor \in \mathbb{Z}$ and $Z \in [0, 1)$.

(a)

Find $\mathbb{E}(X)$

Solution. Since X only takes non-negative value,

$$\mathbb{E}(X) = \int_0^\infty e^{rt} dt = \frac{1}{r}$$

(b)

Find $\mathbb{P}(Y = n),\ n = 0, 1, 2, ...$

Solution.

$$\mathbb{P}(Y = n) = \mathbb{P}(n \leq X < n+1) = e^{-rn} - e^{-r(n+1)}$$

(c)

Find $\mathbb{E}(Y)$ and $\mathbb{E}(Z)$.

Solution.

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{n=1}^{\infty} \mathbb{P}(Y \geq n) \\
&= \sum_{n=1}^{\infty} e^{rn} \\
&= \frac{e^{-r}}{1 - e^{-r}}
\end{aligned}$$

$$\mathbb{E}(Z) = \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = \frac{1}{r} - \frac{e^{-r}}{1 - e^{-r}}$$

(d)

Show that Y and Z are independent.

Proof. It suffices to show that $\mathbb{P}(Y = n|Z = a) = \mathbb{P}(Y = n),\ \forall n.$

$$\begin{aligned}
\mathbb{P}(Y = n|Z = a) &= \frac{\mathbb{P}(X = n + a)}{\sum_{i=0}^{\infty} \mathbb{P}(X = i + a)} \\
&= \frac{re^{-r(n+a)}}{\sum_{i=0}^{\infty} re^{-r(n+i)}} \\
&= e^{-rn} \cdot (1 - e^{-r}) \\
&= e^{-rn} - e^{-r(n+1)} \\
&= \mathbb{P}(Y = n)
\end{aligned}$$

□

Problem 2

Let f and g be bounded nondecreasing functions on $\mathbb{R}$, and let X, Y be independent and identically distributed random variables.

(a)

Show that

$$\mathbb{E}[(f(X) - f(Y))\,(g(X) - g(Y))] \geq 0$$

Proof. By the nondecreasing monotonicity,

$$\mathbb{P}(f(X) - f(Y) \geq 0|X > Y) = \mathbb{P}(g(X) - g(Y) \geq 0|X > Y) = 1$$

$$\mathbb{P}(f(X) - f(Y) \leq 0|X \leq Y) = \mathbb{P}(g(X) - g(Y) \leq 0|X \leq Y) = 1$$

So we can argue that,

$$\begin{aligned}
\mathbb{P}((f(X) - f(Y))\,(g(X) - g(Y)) \geq 0) &= \mathbb{P}((f(X) - f(Y))\,(g(X) - g(Y)) \geq 0|X > Y)\mathbb{P}(X > Y) \\
&\quad + \mathbb{P}((f(X) - f(Y))\,(g(X) - g(Y)) \geq 0|X \leq Y)\mathbb{P}(X \leq Y) \\
&= 1
\end{aligned}$$

Therefore, it follows that

$$\mathbb{E}[(f(X) - f(Y))\,(g(X) - g(Y))] \geq 0$$

□

(b)

Show that $f(X)$ and $g(X)$ are positively correlated, that is,

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)] \cdot \mathbb{E}[g(X)].$$

Proof.

$$\begin{aligned}
\mathbb{E}[(f(X)-f(Y))(g(X)-g(Y))] &= \mathbb{E}(f(X)g(X) - f(X)g(Y) - f(Y)g(X) + g(Y)f(Y)) \\
&\overset{(*)}{=} \mathbb{E}(f(X)g(X)) - \mathbb{E}(f(X))\mathbb{E}(g(Y)) - \mathbb{E}(f(Y))\mathbb{E}(g(X)) + \mathbb{E}(g(Y)f(Y)) \\
&\overset{(**)}{=} 2\mathbb{E}(f(X)g(X)) - 2\mathbb{E}(f(X))\mathbb{E}(g(X)) \\
&= 2\mathrm{Cov}(f(X), g(X)) \\
&\overset{(***)}{\geq} 0
\end{aligned}$$

$(*)$ X, Y independent.
$(**)$ X, Y identically distributed.
$(***)$ by the result from (a) □

Problem 3

Suppose that X and Y have joint density $f(x,y)$ given by $f(x,y) = ce^{-x}$ for $x > 0$ and $-x < y < x$ and $f(x,y) = 0$ otherwise.

(a)

Show that $c = 1/2$.

Solution.

$$\begin{aligned}
\int_0^\infty \int_{-x}^{x} f(x,y)dydx &= 1 \\
\int_0^\infty \int_{-x}^{x} ce^{-x}dydx &= 1 \\
2c\int_0^\infty xe^{-x}dx &= 1 \\
2c &= 1 \\
c &= \frac{1}{2}
\end{aligned}$$

(b)

Find the marginal densities of X and Y, and the conditional density of Y given X.

Solution.

$$\begin{aligned}
f_X(x) &= \int_{-x}^{x} \frac{1}{2}e^{-x}dy \\
&= xe^{-x},\ x > 0 \\
f_Y(y) &= \int \frac{1}{2}e^{-x}\mathbf{1}_{(-x,x)}(y)dx \\
&= \int_{|y|}^{\infty} \frac{1}{2}e^{-x}dx \\
&= \frac{1}{2}e^{-|y|} \\
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
&= \frac{1}{2x},\ x > 0,\ -x < y < x.
\end{aligned}$$

(c)

Find $\mathbb{P}(X > 2Y)$

Solution.

$$\begin{aligned}
\mathbb{P}(X \geq 2Y) &= \int_0^{\infty} \mathbb{P}\left(Y \leq \frac{X}{2}|X = x\right) f_X(x)dx \\
&= \int_0^{\infty} \int_{-\infty}^{x/2} \frac{1}{2x}\mathbf{1}_{(-x,x)}(y) \cdot xe^{-x}dydx \\
&= \frac{3}{4}\int_0^{\infty} xe^{-x}dx \\
&= \frac{3}{4}
\end{aligned}$$

MATH 505a Spring 2018 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Let X and Y be independent standard normal random variables and define $V = \min(X, Y)$. Compute the probability density function of V^2. The final answer should be an elementary function.

Solution. Let ϕ denote the cdf of standard normal, and by the symmetry of the standard normal distribution:

$$\begin{aligned}
\mathbb{P}(V \leq t) &= 1 - \mathbb{P}(V > t) \\
&= 1 - \mathbb{P}(X > t)\mathbb{P}(Y > t) \\
&\overset{(*)}{=} 1 - \phi(-t)^2
\end{aligned}$$

For $V^2,\ t > 0$:

$$\begin{aligned}
\mathbb{P}(V^2 \leq t^2) &= \mathbb{P}(-t \leq V \leq t) \\
&= \mathbb{P}(V \leq t) - \mathbb{P}(V \leq -t) \\
&= \phi(t)^2 - \phi(-t)^2
\end{aligned}$$

By differentiate, we have:

$$\begin{aligned}
f_{V^2}(x) &= \frac{d}{dx}(\phi(\sqrt{x})^2 - \phi(-\sqrt{x})^2) \\
&= 2\phi(\sqrt{x})f_X(\sqrt{x})\frac{1}{2\sqrt{x}} - 2\phi(-\sqrt{x})f_X(-\sqrt{x})\frac{-1}{2\sqrt{x}} \\
&= \frac{1}{\sqrt{2\pi x}}e^{-x/2}(\phi(\sqrt{x}) + \phi(-\sqrt{x})) \\
&\overset{(*)}{=} \frac{1}{\sqrt{2\pi x}}e^{-x/2}
\end{aligned}$$

$(*)$ Symmetry of the standard normal.

Problem 2

Consider positions 1 to n arranged in a circle, so that 2 comes after 1, 3 comes after 2, ..., n comes after $n-1$, and 1 comes after n. Similarly, take 1 to n as values, with cyclic order, and consider all $n!$ ways to assign values to positions, bijectively, with all $n!$ possibilities equally likely. For $i = 1$ to n, let X_i be the indicator that position i and the one following are filled in with two consecutive values in increasing order, and define

$$S_n = \sum_{i=1}^{n} X_i, \quad T_n = \sum_{i=1}^{n} iX_i$$

For example, with $n = 6$ and the circular arrangement 314562, we get $X - 3 = 1$ since 45 are consecutive in increasing order, and similarly $X_4 = X_6 = 1$, so that $S_6 = 3,\ T_6 = 13$.

(a)

Compute the mean and the variance of S_n.

Solution.

$$\begin{aligned}
\mathbb{E}(S_n) &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} \mathbb{P}(X_i = 1) \\
&= n \cdot \frac{n-1}{n(n-1)} \\
&= 1 \\
\mathbb{E}(X_i^2) &= \mathbb{E}(X_i) = 1 \\
\mathbb{E}(X_i X_j) &= \begin{cases} \frac{n-2}{n(n-1)(n-2)}, & |i-j| = 1 \\ \frac{n-2}{n(n-1)(n-2)}, & |i-j| > 1 \end{cases}
\end{aligned}$$

$$\begin{aligned}
Var(S_n) &= \mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j}^{n} \mathbb{E}(X_i X_j) - \mathbb{E}(S_n)^2 \\
&= 1 + n(n-1)\frac{n-2}{n(n-1)(n-2)} - 1 \\
&= 1
\end{aligned}$$

(b)

Compute the mean and variance of T_n.

Solution.

$$
\begin{aligned}
\mathbb{E}(T_n) &= \sum_{i=1}^{n} i\mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} i \cdot \frac{1}{n} \\
&= \frac{1}{n}\frac{n(n+1)}{2} \\
&= \frac{1+n}{2} \\
\mathbb{E}(T_n^2) &= \mathbb{E}(\sum_{i=1}^{n} X_i)^2 \\
&= \mathbb{E}(\sum_{i,j}^{n} ijX_iX_j) \\
&= \sum_{i,j}^{n} ij\mathbb{E}(X_iX_j) \\
&= \frac{1}{n(n-1)}\sum_{i,j}^{n} ij \\
&= \frac{1}{n(n-1)}(\sum_{i=1}^{n} i)^2 \\
&= \frac{1}{n(n-1)}(\frac{(n+1)n}{2})^2 \\
&= \frac{n(n+1)^2}{4(n-1)} \\
Var(T_n) &= \mathbb{E}(T_n^2) - \mathbb{E}(T_n)^2 \\
&= \frac{n(n+1)^2}{4(n-1)} - \frac{(1+n)^2}{4}
\end{aligned}
$$

Problem 3

A box is filled with coins, each giving heads with some probability p. The value of p varies from one coin to another, and it is uniform in [0,1]. A coin is selected at random; that one coin is tossed multiple times. HINT: $\int_0^1 x^m(1-x)^l dx = \frac{m!l!}{(m+l+1)!}$ for nonnegative integers $m,\ l$.

(a)

Compute the probability that the first two tosses are both heads.

Solution.

$$\begin{aligned}\mathbb{P}(\text{head twice}) &= \int_0^1 \mathbb{P}(\text{head twice}|p=t)f_p(t)dt \\ &= \int_0^1 t^2 dt \\ &= \frac{1}{3}\end{aligned}$$

(b)

Let X_n be the number of heads in the first n tosses. Compute $\mathbb{P}(X_n = k)$ for all $0 \leq k \leq n$.

Solution. By the hint,

$$\begin{aligned}\mathbb{P}(X_n) &= \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} dp \\ &= \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1}\end{aligned}$$

(c)

Let N be the number of tosses needed to get heads for the first time. Compute $\mathbb{P}(N = n)$ for all $n \leq 1$.

Solution.

$$\begin{aligned}\mathbb{P}(N=n) &= \int_0^1 (1-p)^{n-1} p \; dp \\ &= \frac{(n-1)!}{(n+1)!} \\ &= \frac{1}{n(n+1)}\end{aligned}$$

(d)

Compute the expected value of N.

Solution.

$$\begin{aligned}\mathbb{E}(N) &= \sum_{n=1}^{\infty} \frac{1}{n+1} \\ &= \infty\end{aligned}$$

MATH 505a Fall 2017 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Let X be uniform on [1,5], let Y be uniform on [0,1], and assume that X and Y are independent.

(a)

Compute the probability density function of the product XY.

Solution.

$$\begin{aligned}\mathbb{P}(XY \le t) &= \mathbb{P}(Y \le \frac{t}{X}) \\ &= \int \mathbb{P}(Y \le \frac{t}{x}) f_X(x)dx\end{aligned}$$

Case 1. $0 < t \le 1$,

$$\begin{aligned}\mathbb{P}(Y \le \frac{t}{X}) &= \int_1^5 \frac{t}{x}\frac{1}{4}dx \\ &= \frac{t}{4}\ln(5)\end{aligned}$$

Case 2. $1 < t \le 5$,

$$\begin{aligned}\mathbb{P}(Y \le \frac{t}{X}) &= \int_1^t 1 \cdot \frac{1}{4}dx + \int_t^5 \frac{t}{x} \cdot \frac{1}{4}dx \\ &= \frac{1}{4}(t-1) + \frac{t}{4}\ln(\frac{5}{t})\end{aligned}$$

Therefore, by differentiating:

$$f_{XY}(t) = \begin{cases} \frac{\ln(5)}{4}, & t \in (0,1], \\ \frac{1}{4}\ln\left(\frac{5}{t}\right), & t \in (1,5], \\ 0, & \text{otherwise.} \end{cases}$$

(b)

Compute the cumulative distribution function of the ratio X/Y.

Solution.

$$\begin{aligned}\mathbb{P}(X/Y \le t) &= \mathbb{P}(Y \ge X/t) \\ &= \int \mathbb{P}(Y \ge x/t) f_X(x) dx\end{aligned}$$

Case 1. $1 < t \le 5$,

$$\begin{aligned}\mathbb{P}(Y \ge X/t) &= \int_1^t (1 - \frac{x}{t}) \cdot \frac{1}{4} dx \\ &= \frac{t}{8} + \frac{1}{8t} - \frac{1}{4}\end{aligned}$$

Case 2. $t \ge 5$,

$$\begin{aligned}\mathbb{P}(Y \ge X/t) &= \int_1^5 (1 - \frac{x}{t}) \cdot \frac{1}{4} dx \\ &= 1 - \frac{3}{t}\end{aligned}$$

(c)

Compute the characteristic function of the sum $X + Y$.

Solution.

$$\begin{aligned}\mathbb{E}(e^{it(X+Y)}) &= \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) \\ &= \int_1^5 e^{itx} \frac{1}{4} dx \int_0^1 e^{ity} dy \\ &= -\frac{1}{4t^2}(e^{5it} - e^{it})(e^{it} - 1)\end{aligned}$$

(d)

Compute the moment generating function of the random variable $X - \ln(Y)$.

Solution.

$$\begin{aligned}\mathbb{E}(e^{t(X - \ln Y)}) &= \mathbb{E}(e^{tX} \cdot e^{-t \ln Y}) \\ &= \mathbb{E}(e^{tX}) \cdot \mathbb{E}(Y^{-t}) \\ &= \int_1^5 e^{tx} \frac{1}{4} dx \int_0^1 y^{-t} dy\end{aligned}$$

Notice that the right multiplicand's integrability depends on t, so

$$\mathbb{E}(e^{t(X-\ln Y)}) = \begin{cases} \frac{e^{5t}-e^t}{4t(1-t)}, & t < 1, \\ \infty, & t \geq 1 \end{cases}$$

Problem 2

An urn contains $2n$ balls, coming in pairs: two balls are labeled "1", two balls are labeled "2",..., two balls are labeled "n". A sample of size n is taken without replacement. Denote by N the number of pairs in the sample. Compute the expected value and the variance of N. **You do not need to simplify the expression for the variance.**

Solution. Let X_i be the indicator function of the pair of balls labeled "i" are selected. And the probability of any pair being selected is the ratio of the number of combinations to select $n-2$ balls from the rest of $2n-2$ balls and the total number of combinations to select n balls from $2n$ balls.

$$\begin{aligned} \mathbb{E}(N) &= \mathbb{E}(\sum_{i=1}^{n} X_i) \\ &= \sum_{i=1}^{n} \mathbb{P}(X_i = 1) \\ &= n \cdot \frac{\binom{2n-2}{n-2}}{\binom{2n}{n}} \\ &= \frac{n(n-1)}{2(2n-1)} \end{aligned}$$

$$\begin{aligned} \mathbb{E}(X_i^2) &= \mathbb{E}(X_i) \\ &= \frac{n-1}{2(2n-1)} \end{aligned}$$

Notice that the probability of two pairs being selected is the ratio of the number of combinations to select $n-4$ balls from the rest of $2n-4$ balls and the total number of combinations to select n balls from $2n$ balls. so for $i \neq j$,

$$\begin{aligned} \mathbb{E}(X_i X_j) &= \mathbb{P}(X_i = 1,\ X_j = 1) \\ &= \frac{\binom{2n-4}{n-4}}{\binom{2n}{n}} \\ &= \frac{n(n-1)(n-2)(n-3)}{2n(2n-1)(2n-2)(2n-3)} \end{aligned}$$

$$\begin{aligned} Var(N) &= \mathbb{E}(N^2) - \mathbb{E}(N)^2 \\ &= \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j}^{n} \mathbb{E}(X_i X_j) - \mathbb{E}(N)^2 \\ &= n \cdot \frac{(n-1)}{2(2n-1)} + n(n-1) \cdot \frac{n(n-1)(n-2)(n-3)}{2n(2n-1)(2n-2)(2n-3)} - \left(\frac{n(n-1)}{2(2n-1)}\right)^2 \end{aligned}$$

Problem 3

Let $U_1, U_2, \dots$ be iid random variables, uniformly distributed on [0,1], and let N be a Poisson random variable with mean value equal to one. Assume that N is independent of $U_1, U_2, \dots$ and define

$$Y = \begin{cases} 0, & \text{if } N = 0, \\ \max_{1 \le i \le N} U_i, & \text{if } N > 0. \end{cases}$$

Compute the expected value of Y.

Solution. First we compute the expectation of $max_{1 \le i \le k} U_i$ for some $k \ge 1$. For $0 < t < 1$,

$$\begin{aligned} \mathbb{P}(\max_{1 \le i \le k} U_i \le t) &= \prod_{i=1}^{k} \mathbb{P}(U_i \le t) \\ &= t^k \end{aligned}$$

Since U_i only takes nonnegative value,

$$\begin{aligned} \mathbb{E}(\max_{1 \le i \le k} U_i) &= \int_0^1 (1 - t^k) dt \\ &= \frac{k}{k+1} \end{aligned}$$

We compute the expectation of Y by conditioning on N,

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{k=0}^{\infty} \mathbb{E}(Y|N = k)\mathbb{P}(N = k) \\ &= 0 + \sum_{k=1}^{\infty} \frac{k}{k+1} e^{-1} \frac{1}{k!} \\ &= e^{-1} \sum_{k=1}^{\infty} \left(\frac{1}{k!} - \frac{1}{(k+1)!} \right) \\ &= e^{-1} \end{aligned}$$

MATH 505a Spring 2017 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

Three points are chosen independently and uniformly inside the unit square in the plane. Find the expected area of the smallest closed rectangle that has sides parallel to the coordinate axes and that contains the three points. HINT: Consider what happens with just one coordinate.

Solution. Let $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ be the coordinate of the three points, A be the area of the rectangle. Also let $X_{(i)}$ be the ith smallest among X_1, X_2, X_3. Since X_i's and Y_i's are iid,

$$\begin{aligned}\mathbb{E}(A) &= \mathbb{E}((X_{(3)} - X_{(1)})(Y_{(3)} - Y_{(1)})) \\ &= (\mathbb{E}(X_{(3)} - X_{(1)}))^2\end{aligned}$$

$$\mathbb{P}(X_{(3)} \le x) = \prod_{i=1}^{3} \mathbb{P}(X_i \le x) = x^3,\ f_{X_{(3)}}(x) = 3x^2,\ 0 < x < 1$$

$$\mathbb{P}(X_{(1)} \le x) = 1 - \prod_{i=1}^{3} \mathbb{P}(X_i > x) = 1 - (1-x)^3,\ f_{X_{(1)}}(x) = 3(1-x)^2,\ 0 < x < 1$$

$$\begin{aligned}\mathbb{E}(A) &= \left(\int_0^1 3x^3 dx - \int_0^1 3(1-x)^2 \cdot x dx\right)^2 \\ &= \left(\frac{1}{2}\right)^2 \\ &= \frac{1}{4}\end{aligned}$$

Problem 2

Suppose (X, Y) has joint density of the form $f(x, y) = g(\sqrt{x^2 + y^2})$ for $(x, y) \in \mathbb{R}^2$, for some function g. Show that $Z = Y/X$ has the Cauchy density $h(t) = 1/(\pi(1 + t^2)),\ t \in \mathbb{R}$. HINT: Polar coordinates.

Proof. Use polar coordinate (draw the graph to help visualizing), denoting $\theta = arctan(t)$,

$$\begin{aligned}
\mathbb{P}\left(\frac{Y}{X} \le t\right) &= \mathbb{P}(X > 0, Y \le tX) + \mathbb{P}(X < 0, Y \ge tX) \\
&= \int_{(-\pi/2,\theta)\cup(\pi/2,\theta+\pi)} \int_0^\infty g(r) r dr d\theta \\
&= \left(\int_0^\infty g(r) r dr\right)\left(\int_{(-\pi/2,\theta)\cup(\pi/2,\theta+\pi)} d\theta\right) \\
&= \left(\int_0^\infty g(r) r dr\right)\cdot 2\left(\theta - \frac{\pi}{2}\right) \\
&\overset{(*)}{=} \frac{2\theta - \pi}{2\pi}
\end{aligned}$$

$(*)$ Notice that $\mathbb{P}(Y/X \le \infty) = 1 = \left(\int_0^\infty g(r) r dr\right)\cdot 2\pi$

By differentiating,

$$f_{Y/X}(t) = \frac{d}{dt}\frac{2\cdot arctan(t) - \pi}{2\pi} = \frac{1}{\pi}\frac{1}{t^2+1}$$

□

Problem 3

Assume $\sqrt{3} < C < 2$. Consider a sequence $X_1, X_2, X_3, \ldots$ of random variables where X_1 is uniform on [0,1], and where the conditional distribution of X_{m+1} given X_n is uniform on $[0, CX_n]$.

(a)

Find the conditional expectation of $(X_{n+1})^r$ given X_n, for $r \ge 1$.

Solution. Given X_n,

$$f_{X_{n+1}|X_n}(x) = \frac{1}{CX_n},\ 0 < x < CX_n$$

$$\begin{aligned}
\mathbb{E}(X_{n+1}^r | X_n) &= \int_0^{CX_n} \frac{x^r}{CX_n} dx \\
&= \frac{(CX_n)^r}{r+1}
\end{aligned}$$

(b)

Show that X_n converges to 0 in mean but not in mean square.

Proof. By the result from part(a),

$$\begin{aligned}
\mathbb{E}(X_n) &= \mathbb{E}(\mathbb{E}(X_n|X_{n-1})) \\
&= \frac{C}{2}\mathbb{E}(X_{n-1}) \\
&= \left(\frac{C}{2}\right)^2 \mathbb{E}(X_{n-2}) \\
&\dots \\
&= \left(\frac{C}{2}\right)^{n-1} \mathbb{E}(X_1) \\
&= \left(\frac{C}{2}\right)^{n-1} \cdot \frac{1}{2}
\end{aligned}$$

$$\sqrt{3} < C < 2 \implies \frac{C}{2} < 1 \implies \mathbb{E}(X_n) \to 0$$

$$\begin{aligned}
\mathbb{E}(X_n^2) &= \frac{C^2}{3}\mathbb{E}(X_{n-1}^2) \\
&= \left(\frac{C^2}{3}\right)^{n-1} \cdot \frac{1}{3}
\end{aligned}$$

$$\frac{C^2}{3} > 1 \implies \mathbb{E}(X_n^2) \to \infty$$

□

(c)

Show that X_n converges to 0 almost surely.

Proof. Given $\epsilon > 0$,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(X_n > \epsilon) &\le \sum_{n=1}^{\infty} \frac{\mathbb{E}(X_n)}{\epsilon} \\
&= \frac{1}{\epsilon}\sum_{n=1}^{\infty} \frac{1}{2}\left(\frac{C}{2}\right)^{n-1} \\
&< \infty
\end{aligned}$$

since $\frac{C}{2} < 1$. Then by Borel-Cantelli lemma,

$$\mathbb{P}(\text{limsup}_n\{X_n > \epsilon\}) = 0$$

Note that for any m, $\mathbb{P}(\cup_{n\ge m}\{X_m > \epsilon\}) \ge \mathbb{P}(\lim_n X_n > \epsilon)$. Therefore,

$$\begin{aligned}
\mathbb{P}(\lim_n X_n > \epsilon) &\le \lim_m \mathbb{P}(\cup_{n\ge m}\{X_m > \epsilon\}) \\
&= \mathbb{P}(\text{limsup}_n\{X_n > \epsilon\}) \\
&= 0
\end{aligned}$$

□

Problem 4

Suppose that n boys and m girls are arranged in a row, and assume that all possible orderings of the $n+m$ children are equally likely.

(a)

Find the probability that all n boys appear in a single block.

Solution. The total number of combinations is $\binom{n+m}{n}$, since we can index the positions from 1 to $n+m$, and for each combination we assign different choice of positions to boys/girls. So when all the boys are in a single block, we only need to choose different positions for the left most boy from 1 to $m+1$.

$$\mathbb{P}(\text{boys in a single block}) = \frac{m+1}{\binom{m+n}{n}} = \frac{n!(m+1)!}{(n+m)!}$$

(b)

Find the probability that no two boys are next to each other.

Solution. We are essentially assigning boys to the $m+1$ "gaps" between girls including the left and right ends. Therefore, we are choosing n positions from m+1 positions.

$$\begin{aligned}\mathbb{P}(\text{no two boys are next to each other}) &= \frac{\binom{m+1}{n}}{\binom{n+m}{n}} \\ &= \frac{(m+1)!m!}{(m-n+1)!(n+m)!}\end{aligned}$$

And obviously the probability is 0 when $n > m+1$.

(c)

Find the expected number of boys who have a girl next to them on both sides.

Solution. Let X_i be the indicator function of ith boy having two girls next to him on both sides.

$$\begin{aligned}\mathbb{E}(N) &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n} \mathbb{P}(X_i = 1) \\ &= n \cdot \mathbb{P}(\text{not at position 1 or position n+m}) \cdot \mathbb{P}(\text{left is a girl}) \cdot \mathbb{P}(\text{right is a girl}|\text{left is a girl}) \\ &= n \cdot \frac{m+n-2}{m+n} \cdot \frac{m}{n+m-1} \cdot \frac{m-1}{n+m-2} \\ &= \frac{nm(m-1)}{(n+m-1)(n+m)}\end{aligned}$$

MATH 505a Fall 2016 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

You have a choice to roll a fair die either 100 times. For each of the following outcomes, state whether it is more likely with 100 rolls, or with 1000 rolls. Justify your answer, but you do not need to give a full formal proof.

(a)

The number 1 shows on the die between 15% and 20% of the time.

Solution. $n = 1000$. (not a proof) Let X_i be the indicator function for the ith roll being 1. By the Weak Law of Large Numbers (WLLN),

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_i) = \frac{1}{6} \text{ in probability.}$$

That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\frac{1}{6} - \epsilon < \frac{1}{n}\sum_{i=1}^{n} X_i < \frac{1}{6} + \epsilon) = 1$$

Notice that $15\% < \frac{1}{6} < 20\%$. Therefore,

$$\lim_{n\to\infty} \mathbb{P}(15\% < \frac{1}{n}\sum_{i=1}^{n} X_i < 20\%) = 1$$

In another word, we should expect that the above probability is larger (closer to 1) when n larger.

(b)

The number showing is at most 3, at least half the time.

Solution. $n = 100$.

Proof. Let X_i be the indicator function for the ith roll at most 3. Compute the probability for $2n$ rolls:

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{2n}\sum_{i=1}^{2n} X_i \geq \frac{1}{2}\right) &= \sum_{i=n}^{2n} \binom{2n}{i} \left(\frac{1}{2}\right)^{2n} \\
&= \frac{1}{2} \cdot 2 \sum_{i=n}^{2n} \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \\
&\overset{(*)}{=} \frac{1}{2}\left(\binom{2n}{n}\left(\frac{1}{2}\right)^{2n} + \sum_{i=0}^{2n}\binom{2n}{i}\left(\frac{1}{2}\right)^{2n}\right) \\
&= \frac{1}{2} + \left(\frac{1}{2}\right)^{2n+1}\binom{2n}{n}
\end{aligned}$$

$(*)$ By the symmetry that $\binom{n}{m} = \binom{n}{n-m}$.

Compare the above probability between $2n$ and $2n+2$ notice that for the right term,

$$\begin{aligned}
\frac{\binom{2n}{n}(\frac{1}{2})^{2n+1}}{\binom{2n+2}{n+1}(\frac{1}{2})^{2n+3}} &= \frac{4 \cdot (n+1)^2}{(2n+1)(2n+2)} \\
&= \frac{4n^2+8n+4}{4n^2+6n+2} \\
&> 1, \text{ for } n > 0
\end{aligned}$$

Therefore the probability is decreasing with n. □

(c)

The number showing is 2 or 5, at least half the time.

Solution. $n = 100$. (not a proof) Let X_i be the indicator function for ith roll being 2 or 5. By the WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_i) = \frac{1}{3}, \quad \text{in probability.}$$

Since $\frac{1}{2} > \frac{1}{3}$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{1}{2}\Big) = 0$$

Therefore, we should expect that this probability is smaller when n is larger.

Problem 2

let X and Y be independent exponential random variables with parameters λ and μ (that is, $\mathbb{E}(X) = 1/\lambda$ and $\mathbb{E}(Y) = 1/\mu$), and let $Z = \min(X, Y)$.

(a)

Show that Z is independent of the event $X < Y$. In other words, show the event $Z \le t$ is independent of $X < Y$ for all t.

Proof.

$$\begin{aligned}\mathbb{P}(Z \le t | X < Y) &= \mathbb{P}(X \le t | X < Y) \\ &= \frac{\mathbb{P}(X \le t, X < Y)}{\mathbb{P}(X < Y)}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X \le t, X < Y) &= \int \mathbb{P}(x \le t, x < Y | X = x) f_X(x) dx \\ &\overset{(*)}{=} \int \mathbf{1}_{x \le t} \cdot \mathbb{P}(Y > x) f_X(x) dx \\ &= \int_0^t e^{-\mu x} \cdot \lambda e^{-\lambda x} dx \\ &= \frac{\lambda}{\lambda + \mu}(1 - e^{-(\lambda+\mu)t})\end{aligned}$$

$(*)$ By the independence of X and Y.

$$\begin{aligned}\mathbb{P}(X < Y) &= \int \mathbb{P}(X < y) f_Y(y) dy \\ &= \int_0^\infty (1 - e^{-\lambda y}) \cdot \mu e^{-\mu y} dy \\ &= \frac{\lambda}{\lambda + \mu}\end{aligned}$$

According to what we computed above,

$$\mathbb{P}(Z \le t | X < Y) = 1 - e^{-(\lambda+\mu)t}$$

$$\begin{aligned}\mathbb{P}(Z \le t) &= 1 - \mathbb{P}(Z > t) \\ &= 1 - \mathbb{P}(Y > t)\mathbb{P}(X > t) \\ &= 1 - e^{-(\lambda+\mu)t}\end{aligned}$$

□

(b)

Find the distribution of $\max(X - Y, 0)$.

Solution. For $t > 0$

$$\begin{aligned}\mathbb{P}(\max(X - Y, 0) \le t) &= \mathbb{P}(X - Y \le t, X \ge Y) + \mathbb{P}(0 \le t, X < Y) \\ &= \int_0^\infty \mathbb{P}(y \le X \le y + t) \cdot \mu e^{-\mu y} dy + \mathbb{P}(X < Y) \\ &= \int_0^\infty (e^{-\lambda y} - e^{-\lambda(y+t)}) \mu e^{-\mu y} dy \\ &= \frac{\mu}{\mu + \lambda}(1 - e^{-\lambda t})\end{aligned}$$

Problem 3

Let $X_1, X_2, \ldots$ be iid with characteristic function ψ. Let N be independent of the X_i's with $\mathbb{P}(N = n) = 2^{-n}$ for all $n \geq 1$. Let $Y = \sum_{i=1}^{n} X_i$. Find the characteristic function of Y.

Solution.

$$\begin{aligned}
\mathbb{E}(e^{itY}) &= \mathbb{E}(e^{it\sum_j=1^N X_j}) \\
&= \mathbb{E}(\prod_{j=1}^{N} \mathbb{E}(e^{itX_j})) \\
&= \mathbb{E}(\psi(t)^N) \\
&= \sum_{n=1}^{\infty} \psi(t)^n 2^{-n} \\
&= \begin{cases} \frac{\psi(t)}{2-\psi(t)}, & |\psi(t)/2| < 1 \\ \infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

Problem 4

$n \geq 4$ men, among whom are Alfred, Bill, Charles and David, stand in a row. Assume that all possible orderings of the men are equally likely.

(a)

Find the probability that Charles stands somewhere between Alfred and Bill. (Note this does not mean they are necessarily adjacent—there might be other people between Alfred and Bill.)

Solution. Let A, B, C, D denote Alfred, Bill, Charles, and David, respectively. And let $C_{A,B}$ denotes the event that Charles is between A and B. Now, since all the orderings are equally likely, to find the desired probability, we only need to consider the ordering for A, B, and C. They are: ABC, ACB, BAC, BCA, CAB, CBA. So

$$\mathbb{P}(C_{A,B}) = \frac{1}{3}$$

(b)

Find the probability that David stands somewhere between Alfred and Bill given that Charles stands somewhere between Alfred and Bill.

Solution. Again, we only need to consider the relative ordering of A, B, C, D. For the cases that C, D are both between A, B: $ACDB, ADCB, BCDA, BDCA$

$$\begin{aligned}\mathbb{P}(D_{A,B}|C_{A,B}) &= \frac{\mathbb{P}(D_{A,B} \cap C_{A,B})}{\mathbb{P}(C_{A,B})} \\ &= \frac{\frac{4}{4!}}{\frac{1}{3}} \\ &= \frac{1}{2}\end{aligned}$$

(c)

Find the expected value and variance of the number of men out of n who stand between Alfred and Bill. (Note Alfred and Bill themselves are not counted in this number.)

Solution. Let X_i be the indicator function of ith person(not including A and B) is between A and B. Use the results from (a) and (b):

$$\begin{aligned}\mathbb{E}(\sum_{i=1}^{n-2} X_i) &= \sum_{i=1}^{n-2} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n-2} \mathbb{P}(X_i = 1) \\ &= \frac{n-2}{3}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\sum_{i=1}^{n-2} X_i) &= \mathbb{E}(\sum_{i=1}^{n-2} X_i)^2 - (\mathbb{E}(\sum_{i=1}^{n-2} X_i))^2 \\ &= \sum_{i=1}^{n-2} \mathbb{E}(X_i^2) + \sum i \neq j^{n-2} \mathbb{E}(X_i X_j) - \left(\frac{n-2}{3}\right)^2 \\ &= \left(\frac{n-2}{3}\right) + \frac{1}{6} \cdot (n-2)(n-3) - \left(\frac{n-2}{3}\right)^2\end{aligned}$$

MATH 505a Spring 2016 Qual Solution Attempts

Troy Tao

August 4, 2022

Contact yntao@usc.edu if you think this document needs revision.

Problem 1

A stick of length 1 is broken at a point uniformly distributed over its length.

(a)

Find the mean and variance of the sum S of the squares of the lengths of the two pieces.

Solution. Let U be the location of the breaking point.

$$\begin{aligned}
\mathbb{E}(S) &= \mathbb{E}(U^2 + (1-U)^2) \\
&= \mathbb{E}(2U^2 - 2U + 1) \\
&= 2 \cdot \frac{1}{3} - 2 \cdot \frac{1}{2} + 1 \\
&= \frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(S) &= \mathbb{E}(S^2) - (\mathbb{E}(S))^2 \\
&= \mathbb{E}((2U^2 - 2U + 1)^2) - \left(\frac{2}{3}\right)^2 \\
&= \mathbb{E}(4U^4 - 8U^3 + 8U^2 - 4U + 1) - \frac{4}{9} \\
&= 4 \cdot \frac{1}{5} - 8 \cdot \frac{1}{4} + 8 \cdot \frac{1}{3} - 4 \cdot \frac{1}{2} + 1 - \frac{4}{9} \\
&= \frac{1}{45}
\end{aligned}$$

(b)

Find the density function of the product M of the lengths of the two pieces. Note that $M \in [0, \frac{1}{4}]$.

Solution.

$$\begin{aligned}
\mathbb{P}(M \le t) &= \mathbb{P}(U(1-U) \le t) \\
&= \mathbb{P}(U \le \frac{1}{2}(1-\sqrt{1-4t})) + \mathbb{P}(U \ge \frac{1}{2}(1+\sqrt{1-4t})) \\
&= 1-\sqrt{1-4t},\ 0 \le t \le \frac{1}{4} \\
f_M(t) &= \frac{d}{dt}(1-\sqrt{1-4t}) \\
&= \frac{2}{\sqrt{1-4t}},\ 0 \le t \le \frac{1}{4}
\end{aligned}$$

Problem 2

There are two types of batteries in a bin. The life span of type i is an exponential random variable with mean $\mu_i, i = 1, 2$. The probability of type i battery to be chosen is p_i, with $p_1 + p_2 = 1$. Suppose a randomly chosen battery is still operating after t hours. What is the probability that it will still be operating after an additional s hours?

Solution. Denote the life span for type 1 and 2 battery as B_1, B_2. Let B be the life span of the chosen battery.

$$\begin{aligned}
\mathbb{P}(B > s+t | B > t) &= \mathbb{P}(B_1 > s+t | B_1 > t)p_1 + \mathbb{P}(B_2 > s+t | B_2 > t)p_2 \\
&\overset{(*)}{=} \mathbb{P}(B_1 > s)p_1 + \mathbb{P}(B_2 > s)p_2 \\
&= p_1 e^{-\mu_1 s} + p_2 e^{-\mu_2 s}
\end{aligned}$$

$(*)$ Exponential random variables are memory-less. Proof is omitted.

Problem 3

Fix positive integers $m \le n$ with $n > 4$. Suppose m people sit at a circular table with n seats, with all $\binom{n}{m}$ seating equally likely. A seat is called *isolated* if it is occupied and both adjacent seats are vacant. Find the mean and variance of the number of isolated seats.

Solution. Let X_i be the indicator function of ith seat being isolated. Let $N = \sum_{i=1}^{n}$ be the total number of isolated seats. Clearly when $n < m+2$, $N = 0$, $\mathbb{E}(N) = 0$, $\mathrm{Var}(N) = 0$. Assume $n \ge m+2$,

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} \binom{n-3}{m-1} \Big/ \binom{n}{m} \\
&= n \cdot \binom{n-3}{m-1} \Big/ \binom{n}{m}
\end{aligned}$$

$$\mathbb{E}(X_i^2) = \mathbb{E}(X_i) = \binom{n-3}{m-1} \Big/ \binom{n}{m}$$

For $i \neq j$, we want to compute the probability when both ith and jth seats are isolated.

Case 1. When $|i-j| = 1$ or $|i-j| = n-1$(since the end is connect to the start), it's impossible since they are next to each other and both being occupied, so

$$\mathbb{E}(X_i X_j) = 0$$

Case 2. When $|i-j| = 2$ or $|i-j| = n-2$, we need at least 3 vacant seats otherwise it's impossible, so

$$\mathbb{E}(X_i X_j) = \begin{cases} \binom{n-5}{m-2} \Big/ \binom{n}{m} & n \geq m+3 \\ 0 & o.w. \end{cases}$$

Case 3. When $2 < |i-j| < n-2$, we need at least 4 vacant seats otherwise it's impossible, so

$$\mathbb{E}(X_i X_j) = \begin{cases} \binom{n-6}{m-2} \Big/ \binom{n}{m} & n \geq m+4 \\ 0 & o.w. \end{cases}$$

Now, we compute the variance for different range of $n-m$:

When $n - m \geq 4$,

$$\begin{aligned}
\mathrm{Var}(N) &= \mathbb{E}(N^2) - (\mathbb{E}(N))^2 \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j}^{n} \mathbb{E}(X_i X_j) - (\mathbb{E}(N))^2 \\
&= \mathbb{E}(N) - (\mathbb{E}(N))^2 + \sum_{\substack{|i-j|=2 \\ \text{or } n-2}} \binom{n-5}{m-2} \Big/ \binom{n}{m} + \sum_{\substack{2<|i-j| \\ |i-j|<n-2}} \binom{n-6}{m-2} \Big/ \binom{n}{m} \\
&= \mathbb{E}(N) - (\mathbb{E}(N))^2 + 2n \cdot \binom{n-5}{m-2} \Big/ \binom{n}{m} + (n(n-1) - 4n) \cdot \binom{n-6}{m-2} \Big/ \binom{n}{m} \\
&= \binom{n-3}{m-1} \Big/ \binom{n}{m} - \left[\binom{n-3}{m-1} \Big/ \binom{n}{m} \right]^2 + 2n \cdot \binom{n-5}{m-2} \Big/ \binom{n}{m} + (n(n-1) - 4n) \cdot \binom{n-6}{m-2} \Big/ \binom{n}{m}
\end{aligned}$$

When $n - m = 3$, *Case 3* is impossible, so

$$\begin{aligned}
\mathrm{Var}(N) &= \mathbb{E}(N) - (\mathbb{E}(N))^2 + \sum_{i \neq j}^{n} \mathbb{E}(X_i X_j) \\
&= \binom{n-3}{m-1} \Big/ \binom{n}{m} - \left[\binom{n-3}{m-1} \Big/ \binom{n}{m} \right]^2 + \sum_{\substack{|i-j|=2 \\ \text{or } n-2}} \binom{n-5}{m-2} \Big/ \binom{n}{m}
\end{aligned}$$

When $n - m = 2$, it is impossible to have more than one isolated seats, which means we won't have any nonzero $\mathbb{E}(X_i X_j),\ i \neq j$

$$\begin{aligned}
\mathrm{Var}(N) &= \mathbb{E}(N) - (\mathbb{E}(N))^2 \\
&= \binom{n-3}{m-1} \Big/ \binom{n}{m} - \left[\binom{n-3}{m-1} \Big/ \binom{n}{m} \right]^2
\end{aligned}$$