MATH 4710 **FINAL (SOLUTIONS)** **FALL 2020**

Instructions

(1) There are 5 independent problems. The point total is 30. Start early so that you have time to come back to those questions that require you to work harder and read some documents. Take care to solve correctly those questions that are easiest for you. Even when it is not required, ask yourself: what probability space am I using? what are the elementary outcomes used to describe the problem?

(2) You can write your answers on any reasonable media that is convenient to you as long as you can produce a clean pdf file to upload on gradescope. Write your name and Cornell NetID on the top of the first page before you begin.

(3) Make sure you clearly indicate the questions you are addressing and separate them neatly from each other. Write clearly using a black or blue pen or pencil if you write on paper. Use extra pages when needed.
When you upload your exam on gradescope, please assign problems to your pages.

(4) Provide reasons for your answers and explain your computations. For numerical answers, give either a simplified fraction or a decimal answer, which ever comes more easily. You can use a basic calculator (e.g., Desmos) if needed.

(5) You can use your notes, our canvas website including all documents provided there and the book. Do not use other websites or the internet (except for Desmos or a simple electronic calculator). Do not discuss prelim problems with other students. Do not discuss prelim problems with anyone except Pr. Saloff-Coste (ask Professor Saloff-Coste privately on Piazza, by email, or in office hours if you have questions. It is Ok to do so).

(6) Academic integrity is expected of all Cornell University students at all times, whether in the presence or absence of members of the faculty.

Problem 1: (6 pts) Two long racks are holding balls. Rack 1 holds red balls and rack 2 holds white balls. Along both racks, each holding hole is either empty, with probability $1-p$, or holds one ball, with probability $p \in (0,1)$, independently. The integers m and n are fixed and are smaller than the number of places on each rack. Now, the balls contained in the first m holding holes from rack 1 and the first n holding holes from rack 2 are put in an urn. Let X be the total number of red balls in the urn, Y be the number of white balls in the urn, and Z the total number of balls in the urn.

(a 1pt) What is the expected number of balls in the urn? What is the variance?

The total number of balls in the urn is a binomial with parameter $n+m$ and p. So the expected number of balls in the urn is $(n+m)p$ and the variance is $(n+m)p(1-p)$.

(b 2pts) Fix $\ell \in \{0, \dots, n+m\}$. Given that the urn contains a total of ℓ balls, what is the probability that it contains k red balls?

We want to compute

$$\begin{aligned} P(X=k|Z=\ell) &= \frac{P(X=k \text{ and } Z=\ell)}{P(Z=\ell)} \\ &= \frac{P(X=k \text{ and } Y=\ell-k)}{P(Z=\ell)} \\ &= \frac{p^k(1-p)^{m-k}\binom{m}{k} \times p^{\ell-k}(1-p)^{n-(\ell-k)}\binom{n}{\ell-k}}{p^\ell(1-p)^{n+m-\ell}\binom{n+m}{\ell}} \\ &= \frac{\binom{m}{k}\binom{n}{\ell-k}}{\binom{n+m}{\ell}} \end{aligned}$$

(c 1pt) Which of the pairs (X,Y) and (X,Z) are independent and which are not?

The random variables X and Y are independent by construction. The random variables X and Z are not. For instance, $P(X=01, Z=0)=0$ but $P(X=1)>0$ and $P(Z=0)>0$.

(d 2pts) Compute $E(XZ)$.

Write $Z=X+Y$ and $XZ=X^2+XY$. It follows that $E(XZ)=E(X^2)+E(X)E(Y)$ (because X and Y are independent). Now $E(X^2)=\mathrm{Var}(X)+E(X)^2=mp(1-p)+m^2p^2$ and $E(X)=mp$, $E(Y)=np$. So

$$E(XZ)=mp(1-p)+m^2p^2+nmp^2.$$

Problem 2: (6 pts) The random variables X and Y are independent exponential random variables with parameter λ.

(a 3pts) Find the joint density function g of of the pair (U, V) where $U = X + Y$ and $V = X/Y$.

The pair (X, Y) had density $f(x, y) = \lambda^2 e^{-\lambda(x+y)} \mathbf{1}_{(0,+\infty)^2}(x, y)$. The transformation $(x, y) \mapsto (u, v)$: $\begin{cases} u = x + y \\ v = x/y \end{cases}$ sends $(0, \infty)^2$ to $(0, \infty)^2$ with inverse transformation $(u, v) \mapsto (x, y)$: $\begin{cases} x = uv/(1 + v) \\ y = u/(1 + v) \end{cases}$
The pair (U, V) has density

$$g(u, v) = f(uv/(1 + v), u/(1 + v))|J(u, v)|\mathbf{1}_{(0,\infty)}(u)\mathbf{1}_{(0,+\infty)}(v)$$

where $J(u, v) = \det \begin{bmatrix} \frac{v}{1+v} & \frac{u}{(1+v)^2} \\ \frac{1}{1+v} & \frac{-u}{(1+v)^2} \end{bmatrix} = -\frac{uv+u}{(1+v)^3} = -\frac{u}{(1+v)^2}$. This gives

$$g(u, v) = \lambda^2 e^{-\lambda u} \frac{u}{(1 + v)^2} \mathbf{1}_{(0,\infty)}(u)\mathbf{1}_{(0,+\infty)}(v) = \lambda^2 u e^{-\lambda u} \frac{1}{(1 + v)^2} \mathbf{1}_{(0,\infty)}(u)\mathbf{1}_{(0,+\infty)}(v).$$

(b 3pts) Find the density function f_U of the marginal U and the density function f_V of the marginal V. Are U and V independent?

The density of U is $f_U(u) = \lambda^2 u e^{-\lambda u} \mathbf{1}_{(0,+\infty)}(u)$, this is a gamma density with parameter $\alpha = 2$ and λ. We have used $\int_0^\infty \frac{dv}{(1+v)^2} = 1$.
The density of V is $f_V(v) = \frac{1}{(1+v)^2} \mathbf{1}_{(1,+\infty)}(v)$. The random variables U and V are independent.

Problem 3: (6 pts) The random variables X, Y have joint density function

$$f(x,y) = \begin{cases} \frac{1}{4}(y-x)e^{-y} & \text{if } -y < x < y \text{ and } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that that for any integer k, $\int_0^\infty y^k e^{-y} dy = \Gamma(k+1) = k!$.

(a 2pts) Compute the covariance of X and Y.

We compute $E(X)$, $E(Y)$. and $E(XY)$ using the joint density of X and Y and the density of Y (because it is a known distribution).

$$E(X) = \frac{1}{4}\int_0^\infty \int_{-y}^{y} x(y-x)e^{-y} dx dy = -\frac{1}{4}\int_0^\infty \frac{2y^3}{3} e^{-y} dy = -\frac{3! \times 2}{12} = -1.$$

$$E(Y) = \frac{1}{4}\int_0^\infty \int_{-y}^{y} y(y-x)e^{-y} dx dy = \frac{1}{2}\int_0^\infty y^3 e^{-y} dy = \frac{3!}{2} = 3.$$

$$E(XY) = \frac{1}{4}\int_0^\infty \int_{-y}^{y} xy(y-x)e^{-y} dx dy = -\frac{1}{4}\int_0^\infty \frac{2y^4}{3} e^{-y} dy = -\frac{2 \times 4!}{12} = -4.$$

It follows that $\mathrm{Cov}(X,Y) = -4 + 3 = -1$.

(b 2pts) Find the density function of the second marginal Y.

For $y > 0$, (the density is 0 otherwise),

$$f_Y(y) = \int_{-y}^{y} \frac{1}{4}(y-x)e^{-y} dx = \frac{1}{2} y^2 e^{-y}.$$

This is a Gamma distribution with $\lambda = 1$ and $\alpha = 3$ ($\Gamma(3) = (3-1)! = 2$).

(c 2pts) Find the density function of the first marginal X. Are X and Y independent?

$$\begin{aligned} f_X(x) &= \int_{|x|}^\infty \frac{1}{4}(y-x)e^{-y} dy = \frac{1}{4}\left(\int_{|x|}^\infty y e^{-y} dy - x\int_{|x|}^\infty e^{-y} dy\right) \\ &= \frac{1}{4}\left(e^{-|x|} + (|x| - x)e^{-|x|}\right) = \frac{1}{4}\left(1 + |x| - x\right) e^{-|x|}. \end{aligned}$$

Let's compute $f(1,2) = \frac{1}{4}e^{-2}$ and $f_X(1)f_Y(2) = \frac{1}{4}e^{-1}\frac{1}{2}4e^{-2} = \frac{1}{2}e^{-3}$. This shows that X and Y are not independent. We can also use the covariance for the same conclusion.

Problem 4: (6 pts)

Let X and Y be independent exponential random variables with parameters $\lambda > 0$ and $\mu > 0$, respectively.

(a 2pts) Find the probability $P(X < Y)$.

$$\begin{aligned} P(X < Y) &= \int_0^\infty \left(\int_x^\infty \lambda\mu e^{-\lambda x} e^{-\mu y} dy \right) dx \\ &= \int_0^\infty \lambda e^{-(\lambda+\mu)x} dx \\ &= \frac{\lambda}{\lambda+\mu} \end{aligned}$$

(b 2pts) Find the density functions of $U = \max\{X, Y\}$ and $V = \min\{X, Y\}$.

We first compute, for $u > 0$,

$$\begin{aligned} P(U \le u) &= P(X \le u \text{ and } V \le u) = P(X \le u)P(Y \le u) = \int_0^u \lambda e^{-\lambda x} dx \int_0^u \mu e^{-\mu y} dy \\ &= (1 - e^{-\lambda u})(1 - e^{-\mu u}). \end{aligned}$$

Taking derivatives, it follows that. $f_U(u) = 0$ if $u \le 0$ and, for $u > 0$,

$$f_U(u) = \lambda e^{-\lambda u}(1 - e^{-\mu u}) + \mu e^{-\mu u}(1 - e^{-\lambda u}) = \lambda e^{-\lambda u} + \mu e^{-\mu u} - (\lambda+\mu)e^{-(\lambda+\mu)u}.$$

For V, let $v > 0$ and compute

$$\begin{aligned} P(V \ge v) &= P(X \ge v \text{ and } Y \ge v) = \int_v^\infty \lambda e^{-\lambda x} dx \int_v^\infty \mu e^{-\mu y} dy \\ &= e^{-\lambda v} e^{-\mu v} = e^{-(\lambda+\mu)v}. \end{aligned}$$

It follows that V is exponential with parameter $\lambda + \mu$, i.e., $f_V(v) = (\lambda+\mu)e^{-(\lambda+\mu)v}\mathbf{1}_{(0,+\infty)}(v)$.

(c 2pts) Find the density function $f_{U,V}$ of the pair (U, V).

Solution 1: Fix $0 < v < u < \infty$ and $0 < \epsilon < (u - v)$ and compute $P(U \in (u - \epsilon/2, u + \epsilon/2), V \in (v - \epsilon/2, v + \epsilon/2))$ which we know is approximately equal to $\epsilon^2 f_{U,V}(u, v)$.

We have

$$\begin{aligned} P(U \in (u - \epsilon/2, u + \epsilon/2), V \in (v - \epsilon/2, v + \epsilon/2)) &= P(X \in (u - \epsilon/2, u + \epsilon/2), Y \in (v - \epsilon/2, v + \epsilon/2)) \\ &\quad +P(Y \in (u - \epsilon/2, u + \epsilon/2), X \in (v - \epsilon/2, v + \epsilon/2)) \\ &\approx \epsilon^2 (f_X(u)f_Y(v) + f_X(v)f_Y(v)) \\ &= \epsilon^2 \lambda\mu(e^{-\lambda u - \mu v} + e^{-\lambda v - \mu u}) \end{aligned}$$

So, for $0 < v < u < +\infty$, $f_{U,V}(u, v) = \lambda\mu(e^{-\lambda u - \mu v} + e^{-\lambda v - \mu u})$. Note that $f_{U,V}(u, v) = 0$ otherwise.

Solution 2: For $0 < v < u < +\infty$, compute

$P(U \le u, V \le v) = P(U \le u) - P(U \le u \text{ and } V \ge v) = P(X \le u \text{ and } V \le v) - P(v \le X \le u \text{ and } v \le Y \le u)$ and compute the mixed second partial derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial v} P(U \le u, V \le V) = f_{U,V}(u, v)$.

Problem 5: (6 pts). Let $X_1, \dots, X_n$ be n independent geometric random variables, all with the same parameter $p \in (0, 1]$. Say a pair $\{i, j\}$, $1 \le i < j \le n$, is a match if $X_i = X_j$. Let N denotes the total number of matches.

(a 2pts) For $1 \le i < j \le n$, let $E_{i,j}$ be the event that $\{i, j\}$ is a match. What is $P(E_{i,j})$ and which of the pairs of events E_{i_1,j_1}, E_{i_2,j_2} are independent?

For (i, j), $i < j$ to be a match, X_i can be any positive integer k and then X_j must also be equal to k. This gives

$$P(E_{i,j}) = \sum_{k=1}^{\infty} p^2(1-p)^{2(k-1)} = p^2 \sum_{0}^{\infty} (1-p)^{2k} = \frac{p^2}{(1-(1-p)^2)} = \frac{p}{2-p}.$$

If distinct pairs (i_1, j_1) and (i_2, j_2) have no integers in common, E_{i_1,j_1} and E_{i_2,j_2} are independent. If they have exactly one common integer, then $E_{i_1,j_1} \cap E_{i_2,j_2}$ corresponds to the equality of three of the X_is and

$$P(E_{i_1,j_1} \cap E_{i_2,j_2}) = \sum_{k=1}^{\infty} p^3(1-p)^{3(k-1)} = \frac{p^3}{(1-(1-p)^3)}.$$

This is not equal to $P(E_{i_1,j_1})P(E_{i_2,j_2})$ and these events are not independents in this case.

(b 2pts) Compute the expectation of N, $E(N)$.

For $1 \le i < j \le n$, let $Y_{i,j}$ be equal to 1 if $E_{i,j}$ occurs and 0 otherwise. The random variable N is then the sum $N = \sum_{1 \le i < j \le n} Y_{i,j}$ and $E(N) = E(\sum_{1 \le i < j \le n} Y_{i,j}) = \sum_{1 \le i < j \le n} E(Y_{i,j})$. This gives

$$E(N) = \binom{n}{2} \frac{p^2}{(1-(1-p)^2)} = \binom{n}{2} \frac{p}{2-p}.$$

(c 2pts) For $p = 1/2$, compute the variance of N.

To compute the variance in general, we write

$$\mathrm{Var}(N) = \mathrm{Var}\left(\sum_{1 \le i < j \le n} Y_{i,j}\right) = \sum_{1 \le i < j \le n} \mathrm{Var}\ (Y_{i,j}) + \sum_{\substack{(i_1,j_1) \neq (i_2,j_2) \\ 1 \le i_1 < j_1 \le n, 1 \le i_2 < j_2 \le n}} \mathrm{Cov}(Y_{i_1,j_1}, Y_{i_2,j_2}).$$

The sum of variance gives $\binom{n}{2} \frac{p^2}{(1-(1-p)^2)} \left(1 - \frac{p^2}{(1-(1-p)^2)}\right)$.

In the sum of covariances, only those pairs of pairs with one integer in common contribute. There are exactly $\binom{n}{3}$ order triplets $1 \le k < \ell < m \le n$ and each of then contribute 3 pairs of pairs (k, ℓ) and (k, m), (k, ℓ) and (ℓ, m), (k, m) and $\ell, m)$. Each of these pairs appears twice in the sum. So the sum of covariances equal

$$6\binom{n}{3} \mathrm{Cov}(Y_{1,2}, Y_{2,3}) = 6\binom{n}{3} \left(\frac{p^3}{(1-(1-p)^3)} - \left(\frac{p^2}{(1-(1-p)^2)}\right)^2\right).$$

Putting things together,

$$\begin{aligned} \mathrm{Var}(N) \quad &= \quad \binom{n}{2} \frac{p^2}{(1-(1-p)^2)} \left(1 - \frac{p^2}{(1-(1-p)^2)}\right) \\ &\quad + 6\binom{n}{3} \left(\frac{p^3}{(1-(1-p)^3)} - \left(\frac{p^2}{(1-(1-p)^2)}\right)^2\right). \end{aligned}$$

This computation simplifies if $p = 1/2$ but the structure stay the same. When $p = 1/2$, $P(E_{i,j}) = 1/3$. If the pairs $(i_1, j_1), (i_2, j_2)$ have exactly one common integer, then $\mathrm{Cov}(Y_{1,2}, Y_{2,3}) = 2/63$. It follows that $\mathrm{Var}(N) = \frac{2}{9}\binom{n}{2} + \frac{4}{21}\binom{n}{3}$.