Practice Final

Formula Sheet

$$Var(X) = E(X^2) - (E(X))^2, Cov(X,Y) = E[XY] - E[X]E[Y]$$

$$Cov\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j Cov(X_i, Y_j)$$

$$E\{E[X|Y]\} = E[X]$$

$$Var(X) = E\{Var(X|Y)\} + Var(E[X|Y])$$

The formula sheet will also include pdfs and pmfs for densities, so these don't have to be memorized.

1. (a) Suppose I roll a 6-sided die. If I get a six, then I roll again and keep rolling until I get something other than a six. If get something other than a six, I stop. Let X be the result. Show that $P(X = i) = 1/5$ for $i \in \{1, 2, 3, 4, 5\}$ and $P(X = i) = 0$ otherwise.

Let A be the event that you roll a 6 on the first roll. Then for $i \in \{1, 2, 3, 4, 5\}$,

$$\begin{aligned}
P(X = i) &= P(X = i|A)P(A) + P(X = i|A^c)P(A^c) \\
&= P(X = i)(1/6) + P(X = i, A^c) \\
&= P(X = i)(1/6) + P(X = i) \\
&= P(X = i)(1/6) + 1/6 \\
\Rightarrow P(X = i)(1 - 1/6) &= 1/6 \\
\Rightarrow P(X = i) &= \frac{1/6}{5/6} \\
&= 1/5
\end{aligned}$$

You can also do an infinite series approach.

(b) Suppose I roll a four sided die and a six sided die and that both dice are fair and independent. If the two rolls are identical, then I roll both dice again until I get distinct values. Let X be the value of the four sided die and Y the value of the six sided die. Find the marginal distributions for X and Y. Also find $P(X < Y)$.

First let's write the sample space, and note that all outcomes are equally likely

$$
\begin{aligned}
S = \{&(1,2),(1,3),(1,4),(1,5),(1,6)\\
&(2,1),(2,3),(2,4),(2,5),(2,6)\\
&(3,1),(3,2),(3,4),(3,5),(3,6)\\
&(4,1),(4,2),(4,3),(4,5),(4,6)\}
\end{aligned}
$$

The marginal distribution for X is

$$P(X = i) = 1/4I(i \in \{1,2,3,4\})$$

The marginal distribution for Y is

$$P(Y = i) = \begin{cases} 3/20 & i \in \{1,2,3,4\} \\ 4/20 & i \in \{5,6\} \\ 0 & \text{otherwise} \end{cases}$$

Thus, X has a discrete uniform distribution, but Y does not.

For $P(X < Y)$, we count how many cases in the Sample space satisfy the condition and get $\frac{14}{20} = \frac{7}{10}$.

(c) Under the conditions of (b), what is $P(X = 1|Y = 6)$?

$$P(X = 1|Y = 6) = P(X = 1, Y = 6)/P(Y = 6) = (1/20)/(4/20) = 1/4 = P(X = 1)$$

(d) Are X and Y independent? Why or why not?

Solution. No, even though $P(X = 1|Y = 6) = P(X = 1)$ other examples fail tests for independence, such as $P(X = 1|Y = 2) = 1/3 \neq P(X = 1)$ and $P(X = 1|Y = 1) = 0$.

2. Let $A = \{1,2,3,\ldots,n\}$. How many subsets of A are there? Solution: 2^n.

3. Suppose a class has 40 students, of whom 13 are undergraduate math majors, 12 are undergraduate statistics majors, and 15 of whom are graduate students. Suppose a subset of 5 students is chosen at random. What is the probability the subset has at least one of each type of student (undergrad math, undergrad stat, and graduate student)? You can give an expression in terms of binomial coefficients rather than a specific number at the end.

Solution. Let M be the number of math students, S the number of stat students, and G the number of graduate students. Then we could have (M, S, G) have the following numbers

$$(1,1,3),(1,2,2),(1,3,1),(2,1,2),(2,2,1),(3,1,1)$$

and these are the only possibilities. The probability of the first case is

$$\frac{\binom{13}{1}\binom{12}{1}\binom{15}{3}}{\binom{40}{5}}$$

and we get similar expressions for the other cases. Thus the total probability is

$$\frac{\binom{13}{1}\binom{12}{1}\binom{15}{3}+\binom{13}{1}\binom{12}{2}\binom{15}{2}+\binom{13}{1}\binom{12}{3}\binom{15}{1}+\binom{13}{2}\binom{12}{1}\binom{15}{2}+\binom{13}{2}\binom{12}{2}\binom{15}{1}+\binom{13}{3}\binom{12}{1}\binom{15}{1}}{\binom{40}{5}} \approx 0.65$$

4. Suppose $Y|X \sim \text{Pois}(X)$ and, $X \sim \exp(\lambda)$ (i.e., X has rate λ). Find

(a) $E[Y]$

Solution. $E[Y] = E[E[Y|X]] = E[X] = 1/\lambda$

(b) $Var(Y)$

Here $Var(Y|X) = X$ (because $Y|X$ is Poisson(X)), and $E[Y|X] = X$

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X]) = E[X] + Var(X) = \frac{1}{\lambda} + \frac{1}{\lambda^2} = \frac{\lambda+1}{\lambda^2}$$

5. Suppose $X \sim \text{gamma}(\alpha_1, 1)$ and $Y \sim \text{gamma}(\alpha_2, 1)$ are independent gamma random variables. Find the density for $Z = \frac{X}{X+Y}$.

Solution. It helps to know in advance that the answer is a beta distribution. This makes sense because $X + Y > X$ and therefore $0 < \frac{X}{X+Y} < 1$. We'll try a bivariate transformation with $U = \frac{X}{X+Y}$ and $V = X$. We have $x = v$ and $u = v/(v + y) \Rightarrow= uv + uy = v \Rightarrow y = v/u - v$. Note that

$$\frac{d}{dx}x(x+y)^{-1} = (x+y)^{-1} - x(x+y)^{-2} = \frac{x+y-x}{(x+y)^2} = \frac{y}{x+y} = 1 - \frac{x}{x+y} = 1-u$$

and

$$\frac{d}{dy}x(x+y)^{-1} = -\frac{x}{x+y}^{-2} = -u^2/v$$

The Jacobian gives us

$$\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{y}{x+y} & -\frac{x}{(x+y)^2} \\ 1 & 0 \end{vmatrix} = \frac{x}{(x+y)^2} = u^2/v$$

To save space, I'll leave the indicator notation until the end. The joint density is (I'll just write the indicator function at the last step):

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(v, v/u - v)|J|^{-1} \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 - 1} e^{-v} (v/u - v)^{\alpha_2 - 1} e^{-v/u + v} v/u^2 \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 - 1} e^{-v} v^{\alpha_2 - 1} (1/u - 1)^{\alpha_2 - 1} e^{-v/u + v} v/u^2 \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 + \alpha_2 - 1} e^{-v} [(1-u)/u]^{\alpha_2 - 1} e^{-v/u + v} \cdot 1/u^2 \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{-\alpha_2 + 1 - 2} (1-u)^{\alpha_2 - 1} v^{\alpha_1 + \alpha_2 - 1} e^{-v/u} I(0 < u < 1, v > 0) \end{aligned}$$

Integrating with respect to v, we get

$$f_U(u) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{-\alpha_2 + 1 - 2} (1-u)^{\alpha_2 - 1} \int_0^\infty v^{\alpha_1 + \alpha_2 - 1} e^{-v/u} \, dv$$

Let's just do the integral recognizing that it corresponds to integrating a gamma density:

$$\begin{aligned} &\int_0^\infty v^{\alpha_1 + \alpha_2 - 1} e^{-v/u} \, dv \\ &= \Gamma(\alpha_1 + \alpha_2) u^{\alpha_1 + \alpha_2} \int_0^\infty \frac{1}{\Gamma(\alpha_1 + \alpha_2) u^{\alpha_1 + \alpha_2}} v^{\alpha_1 + \alpha_2 - 1} e^{-v/u} \, dv \\ &= \Gamma(\alpha_1 + \alpha_2) u^{\alpha_1 + \alpha_2} \cdot 1 \end{aligned}$$

Plugging this result into the previous expression for $f_U(u)$, we get for $0 < u < 1$,

$$\begin{aligned} f_U(u) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{-\alpha_2 + 1 - 2} (1-u)^{\alpha_2 - 1} \int_0^\infty v^{\alpha_1 + \alpha_2 - 1} e^{-v/u} \, dv \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{-\alpha_2 + 1 - 2} (1-u)^{\alpha_2 - 1} \cdot \Gamma(\alpha_1 + \alpha_2) u^{\alpha_1 + \alpha_2} \\ &= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 + \alpha_2} u^{-\alpha_2 + 1 - 2} (1-u)^{\alpha_2 - 1} \\ &= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1} (1-u)^{\alpha_2 - 1} \end{aligned}$$

and we recognize this as the density for a beta(α_1,α_2) random variable. I notice that the original question used Z instead of U, so I'll write

$$f_Z(z) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1-1}(1-z)^{\alpha_2-1} I(0 < z < 1)$$

If you try the cdf method for this problem, you end up trying to integrate a gamma density but over an interval that is not 0 to infinity, which a mess, so the bivariate approach seems easier for this problem than the cdf method even when you are only interested in one of the marginal distributions.

6. Suppose the time until my car breaks down is exponential with mean 10 days (i.e., $\lambda = 1/10$) and the time until my wife's car breaks down is exponential with mean 15 days. What is the probability that her car breaks down first?

Solution. For competing exponentials with rates λ_i and λ_j, the probability that $X_i < X_j$ is $\frac{\lambda_i}{\lambda_i+\lambda_j}$. Therefore, for this problem, we have $\lambda_1 = 1/10$ and $\lambda_2 = 1/15$. Thus, $P(X_2 < X_1) = \frac{1/15}{1/10+1/15} = \frac{2}{3+2} = \frac{2}{5}$. Thus, there is a 40% chance that her car will break down first.

Another approach to this problem is to integrate the joint density

$$P(X_2 < X_1) = \int_0^\infty \int_0^{x_1} (1/10)(1/15)e^{-x_1/10}e^{-x_2/15}\, dx_2 dx_1$$

7. (a) X be exponential with rate λ and let Y be uniform(0,1) with X and Y being independent. Let $U = X + Y$ and $V = X$. Find the joint distribution of U and V and the marginal distribution of U.

Solution. We have $x = v$ and $y = u - v$. The Jacobian is

$$\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

Thus,

$$f_{U,V}(u,v) = f_{X,Y}(v, u-v) = f_X(v)f_Y(u-v) = \lambda e^{-\lambda v} I(0 < u-v < 1, v > 0)$$

The limits here gaurantee that the argument of $f_Y(u-v)$ is between 0 and 1 and that v is positive.

In particular, we need $v < u$, which is implied by $0 < u - v$, and which also makes sense because U is obtained by taxing V and adding something positive. We also have $u - v < 1 \Rightarrow v > u - 1$. Thus, we need $\max\{u-1, 0\} < v < u$. For the joint density, we could have used this expression in the indicator instead of $I(0 < u - v < 1, v > 0)$.

To find the marginal for U, we need to worry about whether $u < 1$ or not. If $u < 1$, then $u - 1 < 0$, so we use $v > 0$. If $u > 1$, then $u - 1 > 0$ and we use $v > u - 1$.

Thus, for $u < 1$,

$$f_U(u) = \int_0^u \lambda e^{-\lambda v} dv = 1 - e^{-\lambda u}$$

For $u > 1$, we have

$$f_U(u) = \int_{u-1}^u \lambda e^{-\lambda v} dv = e^{-\lambda v} \Big|_{u-1}^{u} = -e^{-\lambda u} + e^{-\lambda(u-1)} = e^{-\lambda u}(e - 1)$$

Generally,

$$f_U(u) = \begin{cases} 1 - e^{-\lambda u} & 0 < u < 1 \\ e^{-\lambda u}(e-1) & u \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) Find $Cov(X, Z)$

Solution. I think I meant $Cov(X, U)$ because usually I use $Z = X + Y$.

$$Cov(X, U) = Cov(X, X + Y) = Cov(X, X) + Cov(X, Y) = Var(X) + 0 = \frac{1}{\lambda^2}$$

8. If there is no snow, the rate of accidents at an intersection is Poisson with rate $\lambda = 0.1$ for an entire day. If there is snow, there rate is $\lambda = 0.2$. If the chance of snow tomorrow is 30%, what is the probability of at least one accident? For this intersection, if there is at least one accident on a given day, what is the probability that it was snowing that day?

We use the fact that $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda}$ for a Poisson random variable.

Let S be the event that there is snow. Then

$$P(X \geq 1) = P(X > 1|S)P(S) + P(X > 1|S^c)P(S^c) = (1 - e^{-.2})(0.3) + (1 - e^{-.1})(0.7) \approx 0.12$$

$$P(S|X \geq 1) = \frac{P(X > 1|S)P(S)}{P(X \geq 1)} = \frac{(1 - e^{-.2})(0.3)}{(1 - e^{-.2})(0.3) + (1 - e^{-.1})(0.7)} \approx 0.45.$$