Section 7 – Solutions

Review

- **Multivariate: Discrete to Continuous:**

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leqslant x, s \leqslant y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(t,s)\,ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
Expectation	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y)p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$
Independence must have	$\forall x, y, p_{X,Y}(x,y) = p_X(x)p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x, y, f_{X,Y}(x,y) = f_X(x)f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$
Conditional PMF/PDF	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$\mathbb{E}[X\|Y = y] = \sum_x x \cdot p_{X\|Y}(x\|y)$	$\mathbb{E}[X\|Y = y] = \int_{-\infty}^{\infty} x f_{X\|Y}(x\|y)dx$

- **Law of Total Probability (Continuous):** A is an event, and X is a continuous random variable with density function $f_X(x)$.

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) dx$$

- **Conditional Expectation**: Let X and Y be random variables. Then, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y) \qquad X \text{ discrete}$$

and for any event A,

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|A) \qquad X \text{ discrete}$$

Note that linearity of expectation still applies to conditional expectation: $\mathbb{E}[X + Y|A] = \mathbb{E}[X|A] + \mathbb{E}[Y|A]$

- **Law of Total Expectation (Event Version)**: Let X be a random variable, and let events $A_1, ..., A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

- **Law of Total Expectation (RV Version)**: Suppose X and Y are random variables. Then,

$$\mathbb{E}[X] = \sum_y \mathbb{E}[X|Y = y] p_Y(y) \qquad Y \text{ discrete r.v..}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy \qquad Y \text{ continuous r.v.}$$

- **Markov's Inequality**: Let X be a non-negative random variable, and $\alpha > 0$. Then,

$$\mathbb{P}(X \geqslant \alpha) \leqslant \frac{\mathbb{E}[X]}{\alpha}$$

- **Chebyshev's Inequality**: Suppose Y is a random variable with $\mathbb{E}Y = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for any $\alpha > 0$,

$$\mathbb{P}(|Y - \mu| \geqslant \alpha) \leqslant \frac{\sigma^2}{\alpha^2}$$

- **(Multiplicative) Chernoff Bound**: Let $X_1, X_2, ..., X_n$ be *independent* Bernoulli random variables.

  Let $X = \Sigma_{i=1}^{n} X_i$, and $\mu = \mathbb{E}X$. Then, for any $0 \leqslant \delta \leqslant 1$,

  - $$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant (1+\delta)\mu\right) \leqslant \exp\left(-\frac{\delta^2 \mu}{3}\right)$$

  - $$\mathbb{P}\left(\sum_{i=1}^{n} X_i \leqslant (1-\delta)\mu\right) \leqslant \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

Task 1 – Content Review

a) Select one: For an event A and a continuous random variable X with density $f_X(x)$,

- ◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)\mathbb{P}(X = x)dx$
- ◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)f_X(x)dx$
- ◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} xf_X(x)dx$
- ◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)dx$

The second choice follows directly by definition of continuous law of total probability.

b) True or false: the Union Bound always gives a result in $[0, 1]$.

False. Consider X and Y, which are independent indicator random variables.
Suppose $p_X(x) = \begin{cases} 0.75 & x = 0 \\ 0.25 & x = 1 \end{cases}$ and $p_Y(y) = \begin{cases} 0.75 & y = 0 \\ 0.25 & y = 1 \end{cases}$.
Then we may apply the Union Bound to place a bound on $P(X = 0 \cup Y = 0)$:

$$P(X = 0 \cup Y = 0) \leqslant P(X = 0) + P(Y = 0) = 0.75 + 0.75 = 1.5.$$

In these cases, the Union Bound tells us very little, since the probability of any event occurring is at most 1.

c) True or false: Markov's Inequality always gives a non-negative result.

True. Markov's Inequality is

$$\mathbb{P}\left(() X \geqslant \alpha\right) \leqslant \frac{\mathbb{E}X}{\alpha}$$

as long as X is a non-negative random variable and $\alpha > 0$. Since X is a non-negative random variable, $\mathbb{E}X \geqslant 0$, so $\frac{\mathbb{E}X}{\alpha} \geqslant 0$.

d) Suppose C and D are discrete random variables. Then $\mathbb{E}[C|D = d] =$

- ◯ $\sum_d dp_{D|C}(d|c)$
- ◯ $\sum_c cp_{C|D}(c|d)$
- ◯ $\int_{-\infty}^{\infty} cf_{c|d}dx$
- ◯ $\frac{\mathbb{E}[C]}{\mathbb{E}[D]}$

Choice b is the correct answer from the definition of conditional expectation for discrete random variables.

e) Suppose X and Y are random variables and A is an event. Given that $\mathbb{E}[X|A] = 4$ and $\mathbb{E}[Y|A] = 10$, what is $\mathbb{E}[2X + Y/2|A]$?

- ◯ 14
- ◯ 18
- ◯ 9
- ◯ 13

Choice d is the correct answer since linearity of expectation still applies to conditional expectation:

$$\mathbb{E}[2X + Y/2|A] = \mathbb{E}[2X|A] + \mathbb{E}[Y/2|A] = 2\mathbb{E}[X|A] + \mathbb{E}[Y|A]/2 = 2 \cdot 4 + 10/2 = 8 + 5 = 13.$$

f) True or false: Chebyshev's Inequality can best be described as giving an upper bound on the distribution's right tail.

False. Chebyshev's Inequality gives an upper bound on the sum of the probabilties of the left and right tails of the distribution.

Joint Distributions

Task 2 – Who fails first?

Here's a question that commonly comes up in industry, but isn't immediately obvious. You have a disk with probability p_1 of failing each day. You have a CPU which independently has probability p_2 of failing each day. What is the probability that your disk fails *before* your CPU?

a) Compute the probability by summing over the relevant part of the probability space.

We model the problem by considering two Geometric random variables and deriving the probability that one is smaller than the other. Let $X_1 \sim$ Geometric(p_1). Let $X_2 \sim$ Geometric(p_2). Assume X_1 and X_2 are independent. We want $\mathbb{P}(X_1 < X_2)$.

$$\begin{aligned}
\mathbb{P}(X_1 < X_2) &= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1,X_2}(k, k_2) \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1}(k) \cdot p_{X_2}(k_2) && \text{(by independence)} \\
&= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} (1-p_1)^{k-1} p_1 \cdot (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 \sum_{k_2=k+1}^{\infty} (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 (1-p_2)^k \sum_{k_2=1}^{\infty} (1-p_2)^{k_2-1} p_2 \\
&= \sum_{k=1}^{\infty} (1-p_1)^{k-1} p_1 (1-p_2)^k \cdot 1 \\
&= p_1(1-p_2) \sum_{k=1}^{\infty} [(1-p_2)(1-p_1)]^{k-1} \\
&= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.
\end{aligned}$$

b) Try to provide an intuitive reason for the answer.

Think about X_1 and X_2 in terms of coin flips. Notice that all the flips are irrelevant until the final flip, since before the final flip, both the X_1 coin and the X_2 coin only yield tails. $\mathbb{P}(X_1 < X_2)$ is the probability that on the final flip, where by definition at least one coin comes up heads, it is the case that the X_1 coin is heads and the X_2 coin is tails. So we're looking for the probability that the X_1 coin produces a heads and the X_2 coin produces a tails, conditioned on the fact that they're not both tails, which is derived as:

$$\begin{aligned}\mathbb{P}(\text{Coin 1 } = H\& \text{ Coin 2 } = T \mid \text{not both } T) &= \frac{\mathbb{P}(\text{Coin 1 } = H\& \text{ Coin 2 } = T)}{\mathbb{P}(\text{not both } T)} \\ &= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.\end{aligned}$$

Another way to approach this problem is to use conditioning. Recall that in computing the probability of an event, we saw in Chapter 2 that it is often useful to condition on other events. We can use this same idea in computing probabilities involving random variables, because $X = k$ and $Y = y$ are just events.

c) Recompute the probability using the law of total probability, conditioning on the value of X_1.

Again, let $X_1 \sim \text{Geometric}(p_1)$ and $X_2 \sim \text{Geometric}(p_2)$, where X_1 and X_2 are independent. Then

$$\begin{aligned}\mathbb{P}(X_1 < X_2) &= \sum_{k=1}^{\infty} \mathbb{P}(X_1 < X_2 \mid X_1 = k) \cdot \mathbb{P}(X_1 = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(k < X_2 \mid X_1 = k) \cdot \mathbb{P}(X_1 = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(X_2 > k) \cdot \mathbb{P}(X_1 = k) && \text{(by independence)} \\ &= \sum_{k=1}^{\infty} (1-p_2)^k \cdot (1-p_1)^{k-1} \cdot p_1 \\ &= p_1(1-p_2) \sum_{k=1}^{\infty} [(1-p_2)(1-p_1)]^{k-1} \\ &= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}.\end{aligned}$$

Task 3 – Continuous joint density

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w,v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

For two random variables X, Y to be independent, we must have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x \in \Omega_X,\ y \in \Omega_Y$. Let's start with X and Y by finding their marginal PDFs. By definition, and using the fact that the joint PDF is 0 outside of $y > 0$, we get:

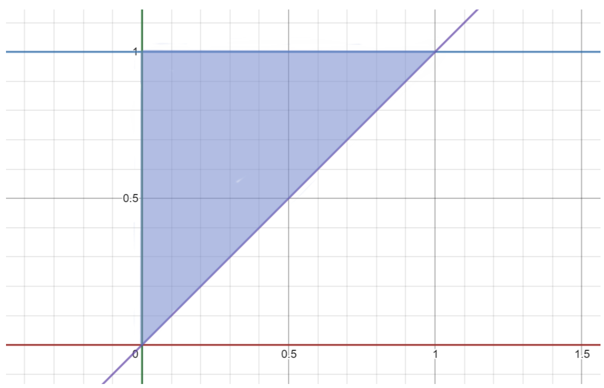
$$f_X(x) = \int_0^\infty xe^{-(x+y)}dy = e^{-x}x$$

We do the same to get the PDF of Y, again over the range $x > 0$:

$$f_Y(y) = \int_0^\infty xe^{-(x+y)}dx = e^{-y}$$

Since $e^{-x}x \cdot e^{-y} = xe^{-x-y} = xe^{-(x+y)}$ for all $x, y > 0$, X and Y are independent.

We can see that W and V are not independent simply by observing that $\Omega_W = (0,1)$ and $\Omega_V = (0,1)$, but $\Omega_{W,V}$ is not equal to their Cartesian product. Specifically, looking at their range of $f_{W,V}(w,v)$. Graphing it with w as the "x-axis" and v as the "y-axis", we see that :



The shaded area is where the joint pdf is strictly positive. Looking at it, we can see that it is not rectangular, and therefore it is not the case that $\Omega_{W,V} = \Omega_W \times \Omega_V$. Remember, the joint range being the Cartesian product of the marginal ranges is not sufficient for independence, but it is *necessary*. Therefore, this is enough to show that they are not independent.

Conditional Distributions, Law of total expectation, Continuous LoTP

Task 4 – A Dysfunctional Family

Rick and his grandson Morty are set to meet at a certain time. Since their relationship is a little strained, neither of them wants to be there on time. Let $X \sim Unif(0,10)$ be the amount of minutes Morty is going to be late. Rick has cameras around the meeting spot and will observe Morty's arrival time $X = x$. Then, he will arrive at the meeting spot $Unif(x, 5x)$ minutes late. Let Y be the random variable indicating how late Rick will be.

a) Using the above definitions determine f_X, $f_{Y|X}$, and f_{XY}. (You will want to determine f_{YX} and use it to determine f_{XY}.).

Since X is a uniform RV on $(0,10)$, we have

$$f_X(x) = \begin{cases} \frac{1}{10} & x \in (0,10) \\ 0 & \text{otherwise} \end{cases}$$

Also given that $X = x$, Y is also uniform on $(x, 5x)$ so

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{4x} & y \in (x, 5x) \\ 0 & \text{otherwise} \end{cases}$$

Since $f_{Y|X}(y|x) = \frac{f_{YX}(y,x)}{f_X(x)}$, we have

$$f_{YX}(y, x) = f_{Y|X}(y|x) \cdot f_X(x) = \begin{cases} \frac{1}{40x} & x \in (0, 10) \text{ and } y \in (x, 5x) \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f_{XY}(x, y) = f_{YX}(y, x) = \begin{cases} \frac{1}{40x} & x \in (0, 10) \text{ and } y \in (x, 5x) \\ 0 & \text{otherwise} \end{cases}$$

b) Compute $\mathbb{E}[Y]$.

By definition since Y conditioned on $X = x$ is a uniform RV on $(x, 5x)$, we have

$$\mathbb{E}[Y \mid X = x] = \frac{x + 5x}{2} = 3x.$$

(Alternatively, we could compute it from first principles using $f_{Y|X}$ and the definition

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)dy = \int_{x}^{5x} \frac{y}{4x}dy = \frac{y^2}{8x}\Big|_{y=x}^{y=5x} = \frac{25x^2}{8x} - \frac{x^2}{8x} = \frac{24x}{8} = 3x.)$$

Then, using the Law of Total Expectation we get:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] \cdot f_X(x)dx = \int_0^{10} \frac{3x}{10}dx = \frac{3}{10} \cdot \frac{x^2}{2}\Big|_0^{10} = \frac{300}{20} - 0 = 15$$

Note: This is a place where the Law of Total Expectation makes things **much** easier than figuring out the PDF of Y and doing direct calculation of the expectation:
Just to see how bad it would get... by definition we have:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx = \int_0^{10} f_{XY}(x, y)dx$$

since we know that $f_{XY}(x, y)$ is 0 when $x \leqslant 0$ or $x \geqslant 10$. However, we can't just plug in $\frac{1}{40x}$ for $f_{XY}(x, y)$ because we also need to satisfy that $x \leqslant y \leqslant 5x$ for that value to be correct. For a fixed value of y, which values of x could work? We need to have $x \leqslant y$, but we also need to have $y \leqslant 5x$ or in other worse $x \geqslant y/5$. In particular, this is equivalent to $y/5 \leqslant x \leqslant y$. Therefore we need $\max(0, y/5) \leqslant x \leqslant \min(10, y)$. Therefore

$$f_Y(y) = \int_0^{10} f_{XY}(x, y)dx = \int_{\max(0,y/5)}^{\min(10,y)} \frac{1}{40x}dx.$$

This would have non-zero contributions for all y with $0 \leqslant y \leqslant 50$ and would be a big mess to calculate since the integral involves $1/x$ which would have logarithms in it...

Task 5 – Law of Total Probability Review

a) (Discrete version) Suppose we flip a coin with probability U of heads, where U is equally likely to be one of $\Omega_U = \{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$ (notice this set has size $n+1$). Let H be the event that the coin comes up heads. What is $\mathbb{P}(H)$?

We can use the law of total probability, conditioning on $U = \frac{k}{n}$ for $k = 0, ..., n$. Note that the probability of getting heads conditioning on a fixed U value is U, and that the probability of U taking on any value in its range is $\frac{1}{n+1}$ since it is discretely uniform.

$$\begin{aligned}\mathbb{P}(H) &= \sum_{k=0}^{n} \mathbb{P}\left(H \mid U = \frac{k}{n}\right)\mathbb{P}\left(U = \frac{k}{n}\right) \\ &= \sum_{k=0}^{n} \frac{k}{n} \cdot \frac{1}{n+1} \\ &= \frac{1}{n(n+1)} \sum_{k=0}^{n} k \\ &= \frac{1}{n(n+1)} \frac{n(n+1)}{2} = \frac{1}{2}\end{aligned}$$

b) Now suppose $U \sim$ Uniform(0,1) has the *continuous* uniform distribution over the interval $[0, 1]$. What is $\mathbb{P}(H)$?

Use the continuous version of the law of total probability: suppose E is an event, and X is a continuous random variable with density function $f_X(x)$. Then

$$\mathbb{P}(E) = \int_{-\infty}^{\infty} \mathbb{P}(E \mid X = x) f_X(x) dx$$

We do the same thing, this time using the continuous law of total probability. Note, this time, that we're conditioning on $U = u$ and taking the integral with respect to u, and that the density of U for any value in its range is 1 because it is uniformly random.

$$\mathbb{P}(H) = \int_{-\infty}^{\infty} \mathbb{P}(H \mid U = u) f_U(u) du$$

We can take the integral from 0 to 1 instead because outside of that range the density of U is 0.

$$= \int_{0}^{1} \mathbb{P}(H \mid U = u) f_U(u) du = \int_{0}^{1} u \cdot 1 \, du = \frac{1}{2}[u^2]_0^1 = \frac{1}{2}$$

Task 6 – 3 points on a line

Three points X_1, X_2, X_3 are selected at random on a line L (continuous independent uniform distributions). What is the probability that X_2 lies between X_1 and X_3?

Let $X_1, X_2, X_3 \sim Unif(0,1)$.

$$\begin{aligned}
\mathbb{P}\left(()\, X_1 < X_2 < X_3\right) &= \int_{-\infty}^{\infty} \mathbb{P}\left(()\, X_1 < X_2 < X_3 \mid X_2 = x\right) f_{X_2}(x)\, dx && \text{Continuous LoTP} \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(()\, X_1 < x, X_3 > x\right) f_{X_2}(x)\, dx && \text{Independence of } X_1, X_2, X_3 \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(()\, X_1 < x\right) \mathbb{P}\left(()\, x < X_3\right) f_{X_2}(x)\, dx && \text{Independence of } X_1, X_3 \\
&= \int_{-\infty}^{\infty} F_{X_1}(x)\, (1 - F_{X_3}(x))\, f_{X_2}(x)\, dx \\
&= \int_0^1 x\, (1-x)\, 1\, dx \\
&= \frac{x^2}{2} - \frac{x^3}{3}\Big|_0^1 = \frac{1}{6}
\end{aligned}$$

Task 7 – Lemonade Stand

Suppose I run a lemonade stand, which costs me \$100 a day to operate. I sell a drink of lemonade for \$20. Every person who walks by my stand either buys a drink or doesn't (no one buys more than one). If it is raining, n_1 people walk by my stand, and each buys a drink independently with probability p_1. If it isn't raining, n_2 people walk by my stand, and each buys a drink independently with probability p_2. It rains each day with probability p_3, independently of every other day. Let X be my profit over the next week. In terms of n_1, n_2, p_1, p_2 and p_3, what is $\mathbb{E}[X]$?

Let R be the event it rains. Let X_i be how many drinks I sell on day i for $i = 1, \ldots, 7$. We are interested in $X = \sum_{i=1}^{7} (20X_i - 100)$. We have $X_i|R \sim$ Binomial(n_1, p_1), so $\mathbb{E}[X_i|R] = n_1 p_1$. Similarly, $X_i|R^C \sim$ Binomial(n_2, p_2), so $\mathbb{E}[X_i|R^C] = n_2 p_2$. By the law of total expectation,

$$\mu = \mathbb{E}[X_i] = \mathbb{E}[X_i|R]\mathbb{P}\left(R\right) + \mathbb{E}[X_i|R^C]\mathbb{P}\left(R^C\right) = n_1 p_1 p_3 + n_2 p_2 (1 - p_3)$$

Hence, by linearity of expectation,

$$\begin{aligned}
\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{7} (20X_i - 100)] = 20 \sum_{i=1}^{7} \mathbb{E}[X_i] - 700 = 140\mu - 700 \\
= 140 \cdot (n_1 p_1 p_3 + n_2 p_2 (1 - p_3)) - 700.
\end{aligned}$$

Task 8 – Trapped Miner

A miner is trapped in a mine containing 3 doors.

- D_1: The 1^{st} door leads to a tunnel that will take him to safety after 3 hours.
- D_2: The 2^{nd} door leads to a tunnel that returns him to the mine after 5 hours.
- D_3: The 3^{rd} door leads to a tunnel that returns him to the mine after a number of hours that is Binomial with parameters $(12, \frac{1}{3})$.

At all times, he is equally likely to choose any one of the doors. What is the expected number of hours for this miner to reach safety?

Let T = number of hours for the miner to reach safety. (T is a random variable)
Let D_i be the event the i^{th} door is chosen. $i \in \{1, 2, 3\}$. Finally, let T_3 be the time it takes to return to the mine in the third case only (a random variable). Note that the expectation of T_3 is $12 * \frac{1}{3}$ because it is binomially distributed with parameters $n = 12, p = \frac{1}{3}$. By Law of Total Expectation, linearity of expectation, and by applying the conditional expectations given by the problem statement:

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[T|D_1]\mathbb{P}(D)_1 + \mathbb{E}[T|D_2]\mathbb{P}(D)_2 + \mathbb{E}[T|D_3]\mathbb{P}(D)_3 \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (\mathbb{E}[T_3 + T]) \cdot \frac{1}{3} \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (\mathbb{E}[T_3] + \mathbb{E}[T]) \cdot \frac{1}{3} \\
&= 3 \cdot \frac{1}{3} + (5 + \mathbb{E}[T]) \cdot \frac{1}{3} + (4 + \mathbb{E}[T]) \cdot \frac{1}{3}
\end{aligned}$$

Solving this equation for $\mathbb{E}[T]$, we get

$$\mathbb{E}[T] = 12$$

Therefore, the expected number of hours for this miner to reach safety is 12.

Task 9 – Elevator rides

[This is the problem we did in class.] The number X of people who enter an elevator on the ground floor is a Poisson random variable with mean 10. If there are N floors above the ground floor, and if each person is equally likely to get off at any one of the N floors, independently of where others get off, compute the expected number of stops the elevator will make before discharging all the passengers.

Let S be the number of stops the elevator makes, and $X \sim Poi(10)$. We shall calculate $E[S]$.
By the law of total expectation, partitioning on the value of X, we have

$$E[S] = \sum_{i=0}^{\infty} E[S|X = i]P(X = i)$$

By the definition of Poison distribution, we know

$$P(X = i) = e^{-10}\frac{10^i}{i!}$$

To calculate $E[S|X = i]$, let $S = Y_1 + Y_2 + ... + Y_N$, where

$$Y_j = \begin{cases} 1 & \text{if someone gets off at the } j^{th} \text{ floor} \\ 0 & \text{otherwise} \end{cases}$$

Then, by the linearity of conditional expectation, we have

$$E[S|X = i] = E[Y_1 + Y_2 + ... + Y_N|X = i] = \sum_{j=1}^{N} E[Y_j|X = i] = \sum_{j=1}^{N} \mathbb{P}(Y_j = 1|X = i)$$

To figure out $\mathbb{P}(Y_j = 1|X = i)$, it would be more convenient to find its complement, $\mathbb{P}(Y_j = 0|X = i)$, which represents the probability that nobody gets off at j^{th} floor. Since each person is equally likely to get off at any one of N floor, we know $\mathbb{P}(Y_j = 0|X = i) = (\frac{N-1}{N})^i$. Thus, we have

$$E[S|X = i] = \sum_{j=1}^{N} \mathbb{P}(Y_j = 1|X = i) = \sum_{j=1}^{N} 1 - \mathbb{P}(Y_j = 0|X = i) = \sum_{j=1}^{N} 1 - (\frac{N-1}{N})^i$$

Finally, we find

$$E[S] = \sum_{i=0}^{\infty} E[S|X=i]P(X=i) = \sum_{i=0}^{\infty}\left(\sum_{j=1}^{N} 1-(\frac{N-1}{N})^i\right)e^{-10}\frac{10^i}{i!}$$

Tail Bounds

Task 10 – Tail bounds

Suppose $X \sim$ Binomial$(6, 0.4)$. We will bound $\mathbb{P}(X \geqslant 4)$ using the tail bounds we've learned, and compare this to the true result.

a) Give an upper bound for this probability using Markov's inequality. Why can we use Markov's inequality?

> We know that the expected value of a binomial distribution is np, so: $\mathbb{P}(X \geqslant 4) \leqslant \frac{\mathbb{E}[X]}{4} = \frac{2.4}{4} = 0.6$. We can use it since X is nonnegative.

b) Give a lower bound for $P(X < 4)$ using Markov's inequality.

> Taking the complement of $P(X < 4)$, we get $1 - P(X \geqslant 4)$. Since we found the upper bound for $P(X \geqslant 4)$, the lower bound for $P(X < 4) \geqslant 1 - 0.6 = 0.4$.

c) Give an upper bound for this probability using Chebyshev's inequality. You may have to rearrange algebraically and it may result in a weaker bound.

> $\mathbb{P}(X \geqslant 4) = \mathbb{P}(X - 2.4 \geqslant 1.6) \leqslant \mathbb{P}(|X - 2.4| \geqslant 1.6)$ we can add those absolute value signs because that only adds more possible values, so it is an upper bound on the probability of $X - 2.4 \geqslant 1.6$. Then, using Chebyshev's inequality we get:
> $\mathbb{P}(|X - 2.4| \geqslant 1.6) \leqslant \frac{Var(X)}{1.6^2} = \frac{1.44}{1.6^2} = 0.5625$

d) Give an upper bound for this probability using the Chernoff bound.

> Since $E[X] = 6 \cdot 0.4 = 2.4$,
> $\mathbb{P}(X \geqslant 4) = \mathbb{P}(X - 2.4 \geqslant 4 - 2.4) = (X - E[X] \geqslant \frac{2}{3}E[X]) \leqslant e^{-(\frac{2}{3})^2\mathbb{E}[X]/4} = e^{-4\times 2.4/36} \approx 0.77$

e) Give an upper bound for $P(X \leqslant 2)$ using the Chernoff bound.

> $\mathbb{P}(X \leqslant 2) = \mathbb{P}(X - 2.4 \leqslant -0.4) = \mathbb{P}(-(X - 2.4) \geqslant 0.4) \leqslant \mathbb{P}(|X - 2.4| \geqslant 0.4)$
> $= \mathbb{P}(|X - E[X]| \geqslant \frac{1}{6}E[X]) \leqslant e^{-(\frac{1}{6})^2\mathbb{E}[X]/4} = e^{-2.4/144} \approx 0.98$

f) Give the exact probability of $P(X \geqslant 4)$.

> Since X is a binomial, we know it has a range from 0 to n (or in this case 0 to 6). Thus, the possible values to satisfy $X \geqslant 4$ are 4, 5, or 6. We plug in the PMF for each to get: $\mathbb{P}(X \geqslant 4) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) = \binom{6}{4}(0.4)^4(0.6)^2 + \binom{6}{5}(0.4)^5(0.6) + \binom{6}{6}0.4^6 \approx 0.1792$

Task 11 – How many samples?

Let $X = X_1 + \ldots X_n$ be the sum of n independent $Poisson(\lambda)$ random variables. Recall that the Poisson distribution has expectation and variance both equal to λ and has the summation property that X is a $Poisson(n\lambda)$ random variable.

a) How large a value of n would Chebyshev's inequality need to guarantee that $\mathbb{P}(X \leqslant \mathbb{E}[X]/2) \leqslant 0.01$?

We have

$$\mathbb{P}(X \leqslant \mathbb{E}[X]/2) = \mathbb{P}(X - \mathbb{E}[X] \leqslant -\mathbb{E}[X]/2) \leqslant \mathbb{P}(|X - \mathbb{E}[X]| \geqslant \mathbb{E}[X]/2).$$

Applying Chebyshev's inequality we have

$$\mathbb{P}(X \leqslant \mathbb{E}[X]/2) \leqslant \mathbb{P}(|X - \mathbb{E}[X]| \geqslant \mathbb{E}[X]/2) \leqslant \frac{4\text{Var}(X)}{\mathbb{E}[X]^2} = \frac{4n\lambda}{n^2\lambda^2} = \frac{4}{n\lambda}.$$

In order for this to be at most 0.01, we require $\boxed{n \geqslant 400/\lambda}$.

b) How large a value of n would Markov's inequality need to guarantee that $\mathbb{P}(X \geqslant \mathbb{E}[X]/2) \leqslant 0.01$?

X is non-negative so Markov's inequality applies to X, but no value of n will guarantee any probability less than 1.

Task 12 – Claris's Late!

Suppose the probability Claris is late to teaching lecture on a given day is at most 0.01. Do not make any independence assumptions.

a) Use a Union Bound to bound the probability that Claris is late at least once over a 30-lecture quarter.

Let R_i be the event Claris is late to lecture on day i for $i = 1, ..., 30$. Then, by the union bound,

$$\begin{aligned}
\mathbb{P}\left(\text{late at least once}\right) &= \mathbb{P}\left(\bigcup_{i=1}^{30} R_i\right) \\
&\leqslant \sum_{i=1}^{30} \mathbb{P}\left(() \, R_i\right) && \text{[union bound]} \\
&\leqslant \sum_{i=1}^{30} 0.01 && [\mathbb{P}\left(R_i\right) \leqslant 0.01] \\
&= 0.30
\end{aligned}$$

b) Use a Union Bound to bound the probability that Claris is **never** late over a 30-lecture quarter.

As in the previous part, let R_i be the event Claris is late to lecture on day i for $i = 1, ..., 30$. Then, by the union bound, we found that

$$\mathbb{P}\left(\text{late at least once}\right) \leqslant 0.30$$

The probability Claris is never late is the complement of the probability she is late at least once over the 30 lectures. Taking the complement and doing algebra:

$$\begin{aligned}
\mathbb{P}\left(\text{late at least once}\right) &\leqslant 0.30 \\
-\mathbb{P}\left(\text{late at least once}\right) &\geqslant -0.30 && \text{[multiplying by negative flips the inequality]} \\
1-\mathbb{P}\left(\text{late at least once}\right) &\geqslant 1-0.30 \\
\mathbb{P}\left(\text{never late}\right) &\geqslant 0.70
\end{aligned}$$

Note that we have now found a *lower* bound for this probability using the union bound because of taking the complement.

c) Use a Union Bound to bound the probability that Claris is late at least once over a 120-lecture quarter.

Let R_i be the event Claris is late to lecture on day i for $i = 1, ..., 120$. Then, by the union bound,

$$\begin{aligned}
\mathbb{P}\left(\text{late at least once}\right) &= \mathbb{P}\left(\bigcup_{i=1}^{120} R_i\right) \\
&\leqslant \sum_{i=1}^{120} \mathbb{P}\left(R_i\right) && \text{[union bound]} \\
&\leqslant \sum_{i=1}^{120} 0.01 && [\mathbb{P}\left(R_i\right) \leqslant 0.01] \\
&= 1.20
\end{aligned}$$

Notice that $\mathbb{P}\left(\text{late at least once}\right) \leqslant 1.20$ is not a very helpful bound since probabilities have to be at most 1 already.

Task 13 – Exponential Tail Bounds

Let $X \sim \text{Exp}(\lambda)$ and $k > 1/\lambda$. Recall that $\mathbb{E}\left[X\right] = \frac{1}{\lambda}$ and $\text{Var}\left(X\right) = \frac{1}{\lambda^2}$.

a) Use Markov's inequality to bound $\text{P}(X \geqslant k)$.

$$\mathbb{P}(X \geqslant k) \leqslant \frac{1}{\lambda k}$$

b) Use Markov's inequality to bound $\text{P}(X < k)$.

From Markov's inequality (and our answer in (a)), we know that $P(X \geqslant k) \leqslant \frac{1}{\lambda k}$. Then,

$$\begin{aligned}
P(X \geqslant k) &\leqslant \frac{1}{\lambda k} \\
-P(X \geqslant k) &\geqslant -\frac{1}{\lambda k} && \text{multiplying be a negative flips the inequality} \\
1 - P(X \geqslant k) &\geqslant 1 - \frac{1}{\lambda k} \\
P(X < k) &\geqslant 1 - \frac{1}{\lambda k} && \text{by definition of complement}
\end{aligned}$$

Note that because we took the complement and the sign flipped, we have now found a *lower* bound for $P(X < k)$.

c) Use Chebyshev's inequality to bound $\mathsf{P}(X \geqslant k)$.

$$\mathbb{P}(X \geqslant k) = \mathbb{P}\left(X - \frac{1}{\lambda} \geqslant k - \frac{1}{\lambda}\right) \leqslant \mathbb{P}\left(\left|X - \frac{1}{\lambda}\right| \geqslant k - \frac{1}{\lambda}\right) \leqslant \frac{1}{\lambda^2(k - 1/\lambda)^2} = \frac{1}{(\lambda k - 1)^2}$$

d) What is the exact formula for $\mathsf{P}(X \geqslant k)$?

$$\mathbb{P}(X \geqslant k) = e^{-\lambda k}$$

e) For $\lambda k \geqslant 3$, how do the bounds given in parts (a), (b), and (c) compare?

$$e^{-\lambda k} < \frac{1}{(\lambda k - 1)^2} < \frac{1}{\lambda k}$$

so Markov's inequality gives the worst bound.