Section 7: Solutions

Review of Main Concepts

- **Normal (Gaussian, "bell curve")**: $X \sim \mathcal{N}(\mu,\ \sigma^2)$ iff X has the following probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \ \ x \in \mathbb{R}$$

  $\mathbb{E}[X] = \mu$ and $Var(X) = \sigma^2$. The "standard normal" random variable is typically denoted Z and has mean 0 and variance 1: if $X \sim \mathcal{N}(\mu,\ \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. The CDF has no closed form, but we denote the CDF of the standard normal as $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leq z)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi(-z) = 1 - \Phi(z)$.

- **Standardizing**: Let X be any random variable (discrete or continuous, not necessarily normal), with $\mathbb{E}[X] = \mu$ and $Var(X) = \sigma^2$. If we let $Y = \frac{X-\mu}{\sigma}$, then $\mathbb{E}[Y] = 0$ and $Var(Y) = 1$.

- **Closure of the Normal Distribution**: Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. That is, linear transformations of normal random variables are still normal.

- **"Reproductive" Property of Normals**: Let $X_1, \ldots, X_n$ be independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$. Let $a_1, \ldots, a_n \in \mathbb{R}$ and $b \in \mathbb{R}$. Then,

$$X = \sum_{i=1}^{n}(a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

  There's nothing special about the parameters – the important result here is that the resulting random variable is still normally distributed.

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

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

- **Central Limit Theorem (CLT)**: Let $X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Let $X = \sum_{i=1}^{n} X_i$, which has $\mathbb{E}[X] = n\mu$ and $Var(X) = n\sigma^2$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which has $\mathbb{E}[\overline{X}] = \mu$ and $Var(\overline{X}) = \frac{\sigma^2}{n}$. $\overline{X}$ is called the *sample mean*. Then, as $n \to \infty$, $\overline{X}$ approaches the normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Standardizing, this is equivalent to $Y = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approaching $\mathcal{N}(0, 1)$. Similarly, as $n \to \infty$, X approaches $\mathcal{N}(n\mu, n\sigma^2)$ and $Y' = \frac{X-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0, 1)$.

  It is no surprise that $\overline{X}$ has mean μ and variance σ^2/n – this can be done with simple calculations. The importance of the CLT is that, for large n, regardless of what distribution X_i comes from, $\overline{X}$ is *approximately normally distributed with mean* μ *and variance* σ^2/n. Don't forget the continuity correction, only when $X_1, \ldots, X_n$ are discrete random variables.

- **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	$F_{X,Y}(x, y) = \sum_{t \leq x, s \leq y} p_{X,Y}(t, s)$	$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(t, s)\, ds dt$

1. Content Review

(a) True or False: For any random variable X, $\mathbb{P}(X = 5) = \mathbb{P}(X - 5 = 0)$. **Solution:**

True. We can think of $X - 5$ as another random variable where we take the output of X and subtract five from it. Then the probability that $X - 5$ is zero is identical to the probability that X is originally five.

(b) True or False: For some continuous random variable X, $\mathbb{P}(X \le 5) \ne \mathbb{P}(X < 5)$. **Solution:**

False. Note that $\mathbb{P}(X \le 5) = \mathbb{P}(X = 5) + \mathbb{P}(X < 5)$. But the first term is zero, so the probabilities are exactly equal. This holds for every continous random variable.

(c) True or False: Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Solution:

True. This follows by the closure of the normal distribution.

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

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

Solution:

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

(e) Select one: Suppose we have n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2. Let $X = \sum_{i=1}^{n} X$. Then as n grows large, the Central Limit Theorem tells us that X behaves similarly to which normal distribution?

- ◯ $X \sim \mathcal{N}(n\mu, n\sigma^2)$
- ◯ $X \sim \mathcal{N}(\mu, n\sigma^2)$
- ◯ $X \sim \mathcal{N}(n\mu, \sigma^2)$
- ◯ $X \sim \mathcal{N}(n\mu, n^2\sigma^2)$

Solution:

The first one. By linearity of expectation, $\mathbb{E}[X] = n\mu$. Now since each of the rvs are independent, we may say that $\text{Var}(() X) = n\sigma^2$. Then as n grows large, X behaves similarly to a normal random variable with the same expectation and variance as itself.

(f) Select one: Given two discrete random variables X and Y, the joint CDF is

- ◯ $F_{X,Y}(x, y) = \sum_{t<x} p_{X,Y}(t, y)$
- ◯ $F_{X,Y}(x, y) = \sum_{s<y} p_{X,Y}(x, s)$

◯ $F_{X,Y}(x,y) = \sum_{t<x}\sum_{s<y} p_{X,Y}(t,s)$

◯ $F_{X,Y}(x,y) = p_{X,Y}(x,y)$

Solution:

The third answer follows directly from the definition of multivariate / joint distributions.

asdfadsfas

2. Will the battery last?

Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with expectation 10,000 miles. If the owner wants to take a 5000 mile road trip, what is the probability that she will be able to complete the trip without replacing the battery, given that the car has already been used for 2000 miles on the same trip?

Solution:

Let N be a r.v. denoting the number of miles until the battery wears out. Then $N \sim \exp(10,000^{-1})$, because N measures the "time" (in this case miles) before an occurrence (the battery wears out) with expectation 10,000. Since this is an exponential distribution, and the expectation of an exponential distribution is $\frac{1}{\lambda}$, $\lambda = \frac{1}{10,000}$. Therefore, via the property of memorylessness of the exponential distribution:

$$\mathbb{P}(N \geq 5000 | N \geq 2000) = \mathbb{P}(N \geq 3000) = 1 - \mathbb{P}(N \leq 3000) = 1 - \left(1 - e^{-\frac{3000}{10000}}\right) \approx 0.741$$

3. Normal questions

(a) Let X be a normal random with parameters $\mu = 10$ and $\sigma^2 = 36$. Compute $\mathbb{P}(4 < X < 16)$. **Solution:**

Let $\frac{X-10}{6} = Z$. By the scale and shift properties of normal random variables $Z \sim \mathcal{N}(0,1)$.

$$\mathbb{P}(4 < X < 16) = \mathbb{P}\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = \mathbb{P}(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.68268$$

(b) Let X be a normal random variable with mean 5. If $\mathbb{P}(X > 9) = 0.2$, approximately what is $Var(X)$? **Solution:**

Let $\sigma^2 = Var(X)$. Then,

$$\mathbb{P}(X > 9) = \mathbb{P}\left(\frac{X-5}{\sigma} > \frac{9-5}{\sigma}\right) = 1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$$

So, $\Phi\left(\frac{4}{\sigma}\right) = 0.8$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{4}{\sigma} = 0.845$. Solving for σ we get $\sigma \approx 4.73$, which means that the variance is about 22.4.

(c) Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that

$$\mathbb{P}(X > c) = 0.10.$$

Solution:

$$\mathbb{P}(X > c) = \mathbb{P}\left(\frac{X-12}{2} > \frac{c-12}{2}\right) = 1 - \Phi\left(\frac{c-12}{2}\right) = 0.1$$

So, $\Phi\left(\frac{c-12}{2}\right) = 0.9$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{c-12}{2} = 1.29$. Solving for c we get $c \approx 14.58$.

Central Limit Theorem Problems
The next few problems are CLT focused problems. Here's a general template for that! Sometimes we'll be trying to solve for the probability of something (e.g., $P(X \leq 10)$, and sometimes, we'll be trying to find a value of some parameter that will allow for the probability to be in a certain range (e.g., $P(X \leq 10) \leq 0.2$). Regardless, we still will want to apply CLT on X, and follow the same process (the only difference is that we may be solving for different things).

(a) Setup the problem - write event you are interested in, in terms of sum of random variables. (what do we want to solve for/what is the probability we want to be true?)
  - Write the random variable we're interested in as a sum of i.i.d., random variables
  - Apply CLT to $X = X_1 + X_2 + ... + X_n$ (we can approximate X as a normal random variable $Y \sim N(\mu, \sigma^2)$)
  - Write the probability we're interested in

(b) *If the RVs are discrete*, apply continuity correction.
(c) Normalize RV to have mean 0 and standard deviation 1: $Z = \frac{Y-\mu}{\sigma}$
(d) Replace RV in probability expression with $Z \sim N(0,1)$
(e) Write in terms of $\Phi(z) = P(Z \leq z)$
(f) Look up in the Phi table (or do a reverse Phi table lookup if we're looking for a value of z that gives us a certain probability)

4. Do it in Reverse

(a) Let X be a normal random variable with parameters $\mu = 8$ and $\sigma^2 = 9$. Find x such that $\mathbb{P}(X \leq x) = 0.6$.

Solution:

Let $\frac{X-8}{3} = Z$. By the scale and shift properties of normal random variables, $Z \sim \mathcal{N}(0,1)$. Thus, we must find z such that $P(Z \leq z) = 0.6$.

$$\Phi(z) = P(Z \leq z) = 0.6$$

$$\Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.6)$$

Thus, $z \approx 0.25$ by looking up the phi values in reverse to undo the Φ function. Then $\frac{x-8}{3} = z \approx 0.25$, so $x \approx 8.75$.

(b) Lots of statistics (like standardized test scores or heights) use *percentiles* to give context to where outcomes fall in a distribution. The nth percentile marks the outcome at which n% of the data points are less than the outcome. Let Y be a normal random variable with parameters $\mu = 15$ and $\sigma^2 = 4$. What value y marks the 85th percentile? What value b marks the 15th percentile?

Solution:

We first find y, which marks the 85th percentile, so $\mathbb{P}(Y \leq y) = 0.85$. Let $\frac{Y-15}{2} = Z$. By the scale and shift properties of normal random variables, $Z \sim \mathcal{N}(0,1)$. Thus, we must find z such that $P(Z \leq z) = 0.85$.

$$\Phi(z) = P(Z \leq z) = 0.85$$

$$\Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.85)$$

Thus, $z \approx 1.04$ by looking up the phi values in reverse to undo the Φ function. Then $\frac{y-15}{2} = z \approx 1.04$, so $y \approx 17.08$.

Recall that normal distributions are symmetric around the mean, where $\mathbb{P}(Y \leq \mu) = 0.5$. Since

$$|\mathbb{P}(Y \leq \mu) - \mathbb{P}(Y \leq y)| = |0.5 - 0.85| = 0.35 = |\mathbb{P}(Y \leq \mu) - \mathbb{P}(Y \leq b)|,$$

$$b = \mu - |b - \mu| = 15 - |17.08 - 15| = 12.92,$$

so $b \approx 12.92$.

5. Round off error

Let X be the sum of 100 real numbers, and let Y be the same sum, but with each number rounded to the nearest integer before summing. If the roundoff errors are independent and uniformly distributed between -0.5 and 0.5, what is the approximate probability that $|X - Y| > 3$? **Solution:**

Let $X = \sum_{i=1}^{100} X_i$, and $Y = \sum_{i=1}^{100} r(X_i)$, where $r(X_i)$ is X_i rounded to the nearest integer. Then, we have

$$X - Y = \sum_{i=1}^{100} X_i - r(X_i)$$

Note that each $X_i - r(X_i)$ is simply the round off error, which is distributed as $Unif(-0.5, 0.5)$. Since $X - Y$ is the sum of 100 i.i.d. random variables with mean $\mu = 0$ and variance $\sigma^2 = \frac{1}{12}$, $X - Y \approx W \sim \mathcal{N}(0, \frac{100}{12})$ by the Central Limit Theorem. For notational convenience let $Z \sim \mathcal{N}(0, 1)$

$$\begin{aligned}
\mathbb{P}(|X - Y| > 3) &\approx \mathbb{P}(|W| > 3) && \text{[CLT]} \\
&= \mathbb{P}(W > 3) + \mathbb{P}(W < -3) && \text{[No overlap between } W > 3 \text{ and } W < -3\text{]} \\
&= 2\,\mathbb{P}(W > 3) && \text{[Symmetry of normal]} \\
&= 2\,\mathbb{P}\left(\frac{W}{\sqrt{100/12}} > \frac{3}{\sqrt{100/12}}\right) \\
&\approx 2\,\mathbb{P}(Z > 1.04) && \text{[Standardize } W\text{]} \\
&= 2\,(1 - \Phi(1.04)) \approx 0.29834
\end{aligned}$$

6. Bad Computer

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

(a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$. **Solution:**

Since X counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that X is binomial with $n = 100$ and $p = 0.9$, or $X \sim \text{Binomial}(100, 0.9)$. Hence, $\mathbb{E}[X] = np = 90$ and $Var(X) = np(1-p) = 9$. Finally,

$$\mathbb{P}(X \geq 87) = \sum_{k=87}^{100} \binom{100}{k} (0.9)^k (1 - 0.9)^{100-k}$$

(b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Use continuity correction.

Important: continuity correction says that if we are using the normal distribution to approximate

$$\mathbb{P}(a \leq \sum_{i=1}^{n} X_i \leq b)$$

where $a \leq b$ are integers and the X_i's are i.i.d. **discrete** random variables, then, as our approximation, we should use

$$\mathbb{P}(a - 0.5 \leq Y \leq b + 0.5)$$

where Y is the appropriate normal distribution that $\sum_{i=1}^{n} X_i$ converges to by the Central Limit Theorem.[1]

For more details see pages 209-210 in the book. **Solution:**

From the previous part, we know that $\mathbb{E}[X] = 90$ and Var $(()\,X) = 9$.

$$\begin{aligned}\mathbb{P}(X \geq 87) &= \mathbb{P}(86.5 < X < 100.5) = \mathbb{P}(\frac{86.5 - 90}{3} < \frac{X - 90}{3} < \frac{100.5 - 90}{3}) \\ &\approx \mathbb{P}(-1.17 < \frac{X - 90}{3} < 3.5) \approx \Phi(3.5) + \Phi(1.17) - 1 \approx 0.9998 + 0.8790 - 1 = 0.8788\end{aligned}$$

Notice that, if you had used $86.5 < X$ in place of $86.5 < X < 100.5$, your answer would have been nearly the same, because $\Phi(3.5)$ is so close to 1.

7. Tweets

A prolific twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!). **Solution:**

Let X be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the ith tweet (since the start of the week). Since X is the sum of 350 i.i.d. rvs with mean $\mu = 75$ and variance $\sigma^2 = 1430$, $X \approx N \sim \mathcal{N}(350 \cdot 75, 350 \cdot 1430)$. Thus,

$$\begin{aligned}\mathbb{P}(26,000 \leq X \leq 27,000) &= \mathbb{P}(25,999.5 \leq X \leq 27,000.5) \\ &\approx \mathbb{P}(25,999.5 \leq N \leq 27,000.5)\end{aligned}$$

[1] The intuition here is that, to avoid a mismatch between discrete distributions (whose range is a set of integers) and continuous distributions, we get a better approximation by imagining that a discrete random variable, say W, is a continuous distribution with density function

$$f_W(x) := p_W(i) \quad \text{when } i - 0.5 \leq x < i + 0.5 \text{ and } i \text{ integer}$$

Standardizing this gives the following formula

$$\begin{aligned}
\mathbb{P}(25,999.5 \le N \le 27,000.5) &= \mathbb{P}\left(\frac{25,999.5 - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \le \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \le \frac{27000.5 - 350 \cdot 75}{\sqrt{350 \cdot 1430}}\right) \\
&\approx \mathbb{P}\left(-0.35 \le \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \le 1.06\right) \\
&\approx \mathbb{P}\left(-0.35 \le Z \le 1.06\right) \\
&= \Phi(1.06) - \Phi(-0.35) \\
&\approx 0.85543 - (1 - 0.63683) \\
&= 0.49226
\end{aligned}$$

So the probability that this user tweets between 26,000 and 27,000 characters in a particular week is approximately 0.4923.

8. Another continuous r.v.

The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & \text{when } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{E}[X] = \frac{3}{5}$, find a and b.

Solution:

To find the value of two variables, we need two equations to solve as a system. We know that $\mathbb{E}[X] = \frac{3}{5}$, so we know, by the definition of expected value, that

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) = \frac{3}{5}$$

Since $f(x)$ is defined to be 0 outside of the given range, we can integrate within only that range, plugging in $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) = \int_{-\infty}^{0} x f(x) + \int_{0}^{1} x f(x) + \int_{1}^{\infty} x f(x) = \int_{0}^{1} x(a + bx^2) = \int_{0}^{1} ax + bx^3 = \frac{ax^2}{2} + \frac{bx^4}{4}\bigg|_0^1 = \frac{a}{2} + \frac{b}{4} = \frac{3}{5}$$

We also know that a valid density function integrates to 1 over all possible values. Thus, we can perform the same process to get a second equation:

$$\int_{-\infty}^{\infty} f(x) = \int_{-\infty}^{0} x f(x) + \int_{0}^{1} x f(x) + \int_{1}^{\infty} x f(x) = \int_{0}^{1} (a + bx^2) = ax + \frac{bx^3}{3}\bigg|_0^1 = a + \frac{b}{3} = 1$$

Solving this system of equations we get that $a = \frac{3}{5}, b = \frac{6}{5}$

9. Point on a line

A point is chosen at random on a line segment of length L. Interpret this statement and find the probability that the ratio of the shorter to the longer segment is less than $\frac{1}{4}$.

Solution:

Define RV X to be the distance of your random point from the leftmost side of the stick. Since we're choosing a point at random, this RV has an equal likelihood of any distance from 0 to L, making it a continuous uniform RV with parameters $a = 0, b = L$. For the ratio to be less than $\frac{1}{4}$, the shorter segment has to be less than $\frac{L}{5}$ in length.

This can happen when $X < \frac{L}{5}$ or $X > \frac{4L}{5}$. Thus, using the CDF of a continuous uniform distribution, the probability that the ratio is less than $\frac{1}{4}$ is

$$\mathbb{P}(X \leq \frac{L}{5}) + \mathbb{P}(X > \frac{4L}{5}) = F_X(\frac{L}{5}) + (1 - F_X(\frac{4L}{5})) = \frac{\frac{L}{5} - 0}{L - 0} + (1 - \frac{\frac{4L}{5} - 0}{L - 0}) = \frac{1}{5} + (1 - \frac{4}{5}) = \frac{2}{5}$$

10. Bitcoin users

There is a population of n people. The number of Bitcoin users among these n people is i with probability p_i, where, of course, $\sum_{0 \leq i \leq n} p_i = 1$. We take a random sample of k people from the population (without replacement). Use Bayes Theorem to derive an expression for the probability that there are i Bitcoin users in the population conditioned on the fact that there are j Bitcoin users in the sample. Let B_i be the event that there are i Bitcoin users in the population and let S_j be the event that there are j Bitcoin users in the sample. Your answer should be written in terms of the p_ℓ's, i, j, n and k.

Solution:

$$\begin{aligned} Pr(B_i|S_j) &= \frac{Pr(S_j|B_i)Pr(B_i)}{Pr(S_j)} && \text{by Bayes Theorem} \\ &= \frac{\frac{\binom{i}{j}\binom{n-i}{k-j}}{\binom{n}{k}} \cdot p_i}{\sum_{\ell=0}^{n} Pr(S_j|B_\ell)Pr(B_\ell)} = \frac{\frac{\binom{i}{j}\binom{n-i}{k-j}}{\binom{n}{k}} \cdot p_i}{\sum_{\ell=0}^{n} \frac{\binom{\ell}{j}\binom{n-\ell}{k-j}}{\binom{n}{k}} \cdot p_\ell} = \frac{\binom{i}{j}\binom{n-i}{k-j} \cdot p_i}{\sum_{\ell=0}^{n} \binom{\ell}{j}\binom{n-\ell}{k-j} \cdot p_\ell}. \end{aligned}$$

Above, we used the fact that $Pr(B_\ell) = p_\ell$ and the fact that $Pr(S_j|B_\ell)$ is the probability of choosing a subset of size k, where j of the selected people are from the subset of ℓ Bitcoin users and $k - j$ are from the remaining $n - \ell$ non-Bitcoin users.

11. Min and max of i.i.d. random variables

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables each with CDF $F_X(x)$ and pdf $f_X(x)$. Let $Y = \min(X_1, \ldots, X_n)$ and let $Z = \max(X_1, \ldots, X_n)$. Show how to write the CDF and pdf of Y and Z in terms of the functions $F_X(\cdot)$ and $f_X(\cdot)$. **Solution:**

First we compute the CDFs of Z and Y as follows:

$$\begin{aligned} F_Z(z) &= P(Z < z) \\ &= P(X_1 < z, ..., X_n < z) && \text{[Definition of max]} \\ &= P(X_1 < z) \cdot ... \cdot P(X_n < z) && \text{[Independence]} \\ &= (F_X(z))^n \end{aligned}$$

$$
\begin{aligned}
F_Y(y) &= P(Y < y) \\
&= 1 - P(Y > y) \\
&= 1 - P(X_1 > y, ..., X_n > y) && \text{[Definition of min]} \\
&= 1 - P(X_1 > y) \cdot ... \cdot P(X_n > y) && \text{[Independence]} \\
&= 1 - (1 - F_X(y))^n
\end{aligned}
$$

Using the fact that $f_X(x) = \frac{d}{dx}F_X(x)$ and the CDFs that we found we can compute the pdfs of Z and Y as follows:

$$
\begin{aligned}
f_Z(z) &= \frac{d}{dz}F_Z(z) \\
&= \frac{d}{dz}(F_X(z))^n \\
&= n \cdot F_X(z)^{n-1} \cdot \left(\frac{d}{dz}F_X(z)\right) \\
&= n \cdot F_X(z)^{n-1} \cdot f_X(z)
\end{aligned}
$$

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \frac{d}{dy}\left(1 - (1 - F_X(y))^n\right) \\
&= -n \cdot (1 - F_X(y))^{n-1} \cdot \frac{d}{dy}(1 - F_X(y)) \\
&= n \cdot (1 - F_X(y))^{n-1} \cdot f_X(y)
\end{aligned}
$$

12. Joint PMF's

Suppose X and Y have the following joint PMF:

X/Y	1	2	3
0	0	0.2	0.1
1	0.3	0	0.4

(a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range ($\Omega_{X,Y}$). **Solution:**

$\Omega_X = \{0, 1\}$, $\Omega_Y = \{1, 2, 3\}$, and $\Omega_{X,Y} = \{(0, 2), (0, 3), (1, 1), (1, 3)\}$

(b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$. **Solution:**

$$p_X(0) = \sum_y p_{X,Y}(0, y) = 0 + 0.2 + 0.1 = 0.3$$

$$p_X(1) = 1 - p_X(0) = 0.7$$

(c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$. **Solution:**

$$p_Y(1) = \sum_x p_{X,Y}(x, 1) = 0 + 0.3 = 0.3$$

$$p_Y(2) = \sum_x p_{X,Y}(x,2) = 0.2 + 0 = 0.2$$

$$p_Y(3) = \sum_x p_{X,Y}(x,3) = 0.1 + 0.4 = 0.5$$

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

No, since a necessary condition is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$.

(e) Find $\mathbb{E}\left[X^3 Y\right]$. **Solution:**

Note that $X^3 = X$ since X takes values in $\{0,1\}$.

$$\mathbb{E}\left[X^3 Y\right] = \mathbb{E}\left[XY\right] = \sum_{(x,y)\in\Omega_{X,Y}} xyp_{X,Y}(x,y) = 1\cdot 1\cdot 0.3 + 1\cdot 3\cdot 0.4 = 1.5$$

13. 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? **Solution:**

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.

14. Continuous Law of Total Probability?

This has not been covered in class yet, but will be soon.

In this exercise, we will extend the law of total probability to the continuous case.

(a) 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 $\mathbb{P}(H)$? **Solution:**

We can use the law of total probability, conditioning on $U = \frac{k}{n}$ for $k = 0, ..., n$.

$$\mathbb{P}(H) = \sum_{k=0}^{n} \mathbb{P}(H|U = \frac{k}{n})\mathbb{P}(U = \frac{k}{n}) = \sum_{k=0}^{n} \frac{k}{n} \cdot \frac{1}{n+1} = \frac{1}{n(n+1)} \sum_{k=0}^{n} k = \frac{1}{n(n+1)} \frac{n(n+1)}{2} = \frac{1}{2}$$

(b) Now suppose $U \sim$ Uniform(0,1) has the *continuous* uniform distribution over the interval $[0, 1]$. Extend the law of total probability to work for this continuous case. (Hint: you may have an integral in your answer instead of a sum). **Solution:**

We can perform basically the same process as above, just using an integral instead of a sum. The values that U can take on are anywhere in the continuous interval $[0, 1]$, so we integrate over that with respect to u. Another change is that we have to use the PDF of U, which in this case is 1 everywhere within our range (since it's uniformly distributed). Plugging that in we can get the same answer of $\frac{1}{2}$ as before.

$$\mathbb{P}(H) = \int_0^1 \mathbb{P}(H|U = u)f_U(u)du = \int_0^1 u \cdot 1du = \frac{1}{2}[u^2]_0^1 = \frac{1}{2}$$

(c) Let's generalize the previous result we just used. Suppose E is an event, and X is a continuous random variable with density function $f_X(x)$. Write an expression for $\mathbb{P}(E)$, conditioning on X. **Solution:**

Set up the same problem as before, only this time we're not actually solving for anything. Note that we have to integrate from negative infinity to infinity. We're technically doing this before as well, however outside of the bounds of $[0, 1]$, the density is equal to 0 so the whole expression is equal to 0. In the general case thought, we don't know the range, so we have to integrate everywhere.

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

15. Transformations

This has not been covered in class yet and probably won't be. But if you're interested, please read Section 4.4.

Suppose $X \sim$ Uniform$(0, 1)$ has the continuous uniform distribution on $(0, 1)$. Let $Y = -\frac{1}{\lambda} \log X$ for some $\lambda > 0$.

(a) What is Ω_Y? **Solution:**

$\Omega_Y = (0, \infty)$ because $\log(x) \in (-\infty, 0)$ for $x \in (0,1)$. Thus, that range times a necessarily negative number $-\frac{1}{\lambda}$, will result in a range from 0 to positive infinity.

(b) First write down $F_X(x)$ for $x \in (0,1)$. Then, find $F_Y(y)$ on Ω_Y. **Solution:**

$F_X(x) = x$ for $x \in (0,1)$ because that is the CDF of the continuous uniform distribution. We find the CDF of Y by plugging in the given definition of Y and getting into a form where we can use the CDF of X. Let $y \in \Omega_Y$.

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(-\frac{1}{\lambda}\log X \le y) = \mathbb{P}(\log X \ge -\lambda y) = \mathbb{P}(X \ge e^{-\lambda y}) = 1 - \mathbb{P}(X < e^{-\lambda y})$$

Then, because $e^{-\lambda y} \in (0,1)$

$$= 1 - F_X(e^{-\lambda y}) = 1 - e^{-\lambda y}$$

(c) Now find $f_Y(y)$ on Ω_Y (by differentiating $F_Y(y)$ with respect to y. What distribution does Y have? **Solution:**

$$f_Y(y) = F_Y'(y) = \lambda e^{-\lambda y}$$

Hence, $Y \sim \mathsf{Exponential}(\lambda)$.

16. Convolutions

This has not been covered in class. We're not yet sure if we will have time for it, but if you're interested, please read Section 5.5.

Suppose $Z = X + Y$, where $X \perp Y$. ($\perp$ is the symbol for independence. In other words, X and Y are independent.) Z is called the convolution of two random variables. If X, Y, Z are discrete,

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x \cap Y = z - x) = \sum_x p_X(x)\, p_Y(z - x)$$

If X, Y, Z are continuous,

$$F_Z(z) = \mathbb{P}(X + Y \le z) = \int_{-\infty}^{\infty} \mathbb{P}(Y \le z - X | X = x) f_X(x) dx = \int_{-\infty}^{\infty} F_Y(z - x) f_X(x) dx$$

Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

(a) Find an expression for $\mathbb{P}(X_1 < 2X_2)$ using a similar idea to convolution, in terms of $F_{X_1}, F_{X_2}, f_{X_1}, f_{X2}$. (Your answer will be in the form of a single integral, and requires no calculations – do not evaluate it). **Solution:**

We use the continuous version of the "Law of Total Probability" to integrate over all possible values of X_2. Take the probability that $X_1 < 2X_2$ given that value of X_2, times the density of X_2 at that value.

$$\mathbb{P}(X_1 < 2X_2) = \int_{-\infty}^{\infty} \mathbb{P}(X_1 < 2X_2 | X_2 = x_2) f_{X_2}(x_2) dx_2 = \int_{-\infty}^{\infty} F_{X_1}(2x_2)\, f_{X_2}(x_2)\, dx_2$$

(b) Find s, where $\Phi(s) = \mathbb{P}(X_1 < 2X_2)$ using the fact that linear combinations of independent normal random variables are still normal. **Solution:**

Let $X_3 = X_1 - 2X_2$, so that $X_3 \sim \mathcal{N}(\mu_1 - 2\mu_2, \sigma_1^2 + 4\sigma_2^2)$ (by the reproductive property of normal distributions)

$$\mathbb{P}(X_1 < 2X_2) = \mathbb{P}(X_1 - 2X_2 < 0) = \mathbb{P}(X_3 < 0) = \mathbb{P}(\frac{X_3 - (\mu_1 - 2\mu_2)}{\sqrt{\sigma_1^2 + 4\sigma_2^2}} < \frac{0 - (\mu_1 - 2\mu_2)}{\sqrt{\sigma_1^2 + 4\sigma_2^2}})$$

$$= \mathbb{P}(Z < \frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}) = \Phi\left(\frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}\right) \to s = \frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}$$