Section 6 – Solutions

Review

- **Discrete to Continuous:**

	Discrete	Continuous
PMF/PDF	$p_X(x) = \mathbb{P}(X = x)$	$f_X(x) \neq \mathbb{P}(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leqslant x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
Expectation	$\mathbb{E}[X] = \sum_x x p_X(x)$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

- **Uniform:** $X \sim$ Uniform(a, b) iff X has the following probability density function:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

  $\mathbb{E}[X] = \frac{a+b}{2}$ and $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

- **Exponential:** $X \sim$ Exponential(λ) iff X has the following probability density function:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

  $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$. The exponential random variable is the continuous analog of the geometric random variable: it represents the waiting time to the next event, where $\lambda > 0$ is the average number of events per unit time. Note that the exponential measures how much time passes until the next event (any real number, continuous), whereas the Poisson measures how many events occur in a unit of time (nonnegative integer, discrete). The exponential random variable X is memoryless:

$$\text{for any } s, t \geqslant 0,\ \mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

  The geometric random variable also has this property.

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

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

  $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. 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 $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leqslant z)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi(-z) = 1 - \Phi(z)$.

  Here is the **Standard normal table (link found on the course website).**

- **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 $X_1, \dots, X_n$ be independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$. Let $a_1, \dots, a_n \in \mathbb{R}$ 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.

- **Central Limit Theorem (CLT)**: Let $X_1, \dots, X_n$ 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.

- **Continuity Correction**: This is a technique for getting a better estimate when applying CLT to the sum $X = \sum_{i=1}^{n} X_i$ or the average of a set of random variables $X_1, \dots, X_n$ that are discrete. Specifically, if asked to compute $\mathbb{P}(a \leqslant X \leqslant b)$ where $a \leqslant b$ are integers, you should compute $\mathbb{P}(a - 0.5 \leqslant X \leqslant b + 0.5)$ so that the width of the interval being integrated is the same as the number of terms you are summing over $(b - a + 1)$. Note that if you applying the CLT to sums/averages of continuous RVs instead, you should not apply the continuity correction.

- **Continuous Law of Total Probability:**

  Suppose that E is an event, and X is a continuous random variable with density function $f_X(x)$. Then

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

Task 1 – Content Review

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

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, $\mathbb{P}(X \leqslant 5) \neq \mathbb{P}(X < 5)$.

False. Note that $\mathbb{P}(X \leqslant 5) = \mathbb{P}(X = 5) + \mathbb{P}(X < 5)$. 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 $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

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 $f_X(x)$,

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

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

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?

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

The first one. By linearity of expectation, $\mathbb{E}X = n\mu$. Now since each of the rvs are independent, we may say that $\text{Var}\left(()\, X\right) = n\sigma^2$. 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

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

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

Task 2 – The exponential distribution is memoryless (problem from lecture)

Show that the exponential distribution is memoryless. Specifically, suppose that X is exponential with parameter λ. Show that $\mathbb{P}\left(X > t + s \mid X > s\right) = \mathbb{P}\left(X > t\right)$.

We have

$$\begin{aligned}
\mathbb{P}\left(X > t + s | X > s\right) &= \frac{\mathbb{P}\left(X > t + s \cap X > s\right)}{\mathbb{P}\left(X > s\right)} = \frac{\mathbb{P}\left(X > t + s\right)}{\mathbb{P}\left(X > s\right)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F_X(t) \\
&= \mathbb{P}(X > t)\ .
\end{aligned}$$

Task 3 – More practice with exponentials (problem from lecture)

The time it takes to check someone out at a grocery store is exponential with an expected value of 10 minutes. Suppose that when you arrive at a grocery store, there is one person in the middle of being served. What is the probability that you will have to wait between 10 and 20 minutes before that person is done being served?

Since the expected value of an exponential random variable is $1/\lambda$, we have $1/\lambda = 10$ minutes, so $\lambda = 1/10$. In addition, since the exponential distribution is memoryless (that is, it doens't matter how long the person being served has already been there), the time that you will have to wait is exponential with parameter $1/10$. Thus

$$\mathbb{P}(10 \leqslant T \leqslant 20) = \int_{10}^{20} \frac{1}{10} e^{-x/10}\, dx = e^{-1} - e^{-2} .$$

Task 4 – Will the battery last?

The owner of a car starts on a 5000 mile road trip. Suppose that the number of miles that the car will run before its battery wears out is exponentially distributed with expectation 10,000 miles. After successfully driving for 2000 miles on the trip without the battery wearing out, what is the probability that she will be able to complete the trip without replacing the battery?

Let N be a r.v. denoting the number of miles until the battery wears out. Then $N \sim \text{exp}(10,000^{-1})$, 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:

$$\mathbb{P}(N \geqslant 5000 \mid N \geqslant 2000) = \mathbb{P}(N \geqslant 3000) = 1 - \mathbb{P}(N \leqslant 3000) = 1 - \left(1 - e^{-\frac{3000}{10000}}\right) \approx 0.741$$

Task 5 – Batteries and exponential distributions (from Section 6)

Let X_1, X_2 be independent exponential random variables, where X_i has parameter λ_i, for $1 \leqslant i \leqslant 2$. Let $Y = \min(X_1, X_2)$.

a) Show that Y is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Hint: Start by computing $\mathbb{P}(Y > y)$. Two random variables with the same CDF have the same pdf. Why?

We start with computing $\mathbb{P}(Y > y)$ by substituting in the definition of Y:

$$\mathbb{P}(Y > y) = \mathbb{P}(\min\{X_1, X_2\} > y) .$$

The probability that the minimum of two values is above a value is the chance that both of them are above that value. From there, we can separate them further because X_1 and X_2 are independent.

$$\mathbb{P}(X_1 > y \cap X_2 > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-\lambda_1 y} e^{-\lambda_2 y}$$
$$= e^{-(\lambda_1 + \lambda_2) y} = e^{-\lambda y} .$$

So $F_Y(y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-\lambda y}$ and $f_Y(y) = \lambda e^{-\lambda y}$ so $Y \sim \text{Exp}(\lambda)$, since this is the same CDF and PDF as an exponential distribution with parameter $\lambda = \lambda_1 + \lambda_2$.

b) What is $\mathbb{P}(X_1 < X_2)$? (Use the continuous version of the law of total probability, conditioning on the probability that $X_1 = x$.)

By the law of total probability,

$$\mathbb{P}(X_1 < X_2) = \int_0^\infty \mathbb{P}(X_1 < X_2 \mid X_1 = x) f_{X_1}(x)\, \mathrm{d}x = \int_0^\infty \mathbb{P}(X_2 > x) \lambda_1 e^{-\lambda_1 x}\, \mathrm{d}x =$$

$$\int_0^\infty e^{-\lambda_2 x} \lambda_1 e^{-\lambda_1 x}\, \mathrm{d}x = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labelled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ, independently of all other batteries. Initially, you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process, you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

Let T be the time until the end of the process. We are trying to find $\mathbb{E}[T]$. $T = Y_1 + \ldots + Y_{n-1}$ where Y_i is the time until we have to replace a battery from the ith pair. The reason it there are only $n-1$ RVs in the sum is because there are $n-1$ times where we have two batteries and wait for one to fail. By part (a), the time for one to fail is the min of exponentials, so $Y_i \sim$ Exponential(2λ). Hence the expected time for the first battery to fail is $\frac{1}{2\lambda}$. By linearity and memorylessness, $\mathbb{E}[T] = \sum_{i=1}^{n-1} \mathbb{E}[Y_1] = \frac{n-1}{2\lambda}$.

d) In the scenario of the previous part, what is the probability that battery i is the last remaining battery as a function of i? (You might want to use the memoryless property of the exponential distribution that has been discussed.)

If there are two batteries i, j in the flashlight, by part (b), the probability each outlasts each other is $1/2$. Hence, the last battery n has probability $1/2$ of being the last one remaining. The second to last battery $n-1$ has to beat out the previous battery and the n^{th}, so the probability it lasts the longest is $(1/2)^2 = 1/4$. Work down inductively to get that the probability the i^{th} is the last remaining is $(1/2)^{n-i+1}$ for $i \geqslant 3$. Finally the first two batteries share the remaining probability as they start at the same time, with probability $(1/2)^{n-1}$ each.

Task 6 – Grading on a curve

In some classes (not CSE classes) an examination is regarded as being good (in the sense of determining a valid spread for those taking it) if the test scores of those taking it are well approximated by a normal density function. The instructor often uses the test scores to estimate the normal parameters μ and σ^2 and then assigns a letter grade of A to those whose test score is greater than $\mu + \sigma$, B to those whose score is between μ and $\mu + \sigma$, C to those whose score is between $\mu - \sigma$ and μ, D to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$ and F to those getting a score below $\mu - 2\sigma$. If the instructor does this and a student's grade on the test really is normally distributed with mean μ and variance σ^2, what is the probability that student will get each of the possible grades A,B,C,D and F? (Use a table for anything you can't calculate.)

We can solve for each of these probabilities by standardizing the normal curve and then looking up each bound in the Z-table. Let X be the students score on the test. Then we have

$$\mathbb{P}(A) = \mathbb{P}(X \geqslant \mu + \sigma) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \geqslant 1\right) = 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} < 1\right)$$

By the closure properties of the normal random variable, $\frac{X-\mu}{\sigma}$ is distributed as a normal random variable with mean 0 and variance 1. Since this is the standard normal, we can plug it into our Φ-table to get the following:

$$\mathbb{P}(A) = 1 - \Phi(1) = 1 - 0.84134 = 0.15866$$

The other probabilities can be found using a similar approach:

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(\mu < X < \mu + \sigma) = \Phi(1) - \Phi(0) = 0.34134 \\
\mathbb{P}(C) &= \mathbb{P}(\mu - \sigma < X < \mu) = \Phi(0) - \Phi(-1) = 0.34134 \\
\mathbb{P}(D) &= \mathbb{P}(\mu - 2\sigma < X < \mu - \sigma) = \Phi(-1) - \Phi(-2) = 0.13591 \\
\mathbb{P}(F) &= \mathbb{P}(X < \mu - 2\sigma) = \Phi(-2) = 0.02275
\end{aligned}$$

Task 7 – Normal questions at the table (from Section 6)

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

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

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

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.68268$$

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

Let $\sigma^2 = \text{Var}(X)$. 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 $\mathbb{P}(X > c) = 0.10$.

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

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., $P(X \leqslant 10)$), 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., $P(X \leqslant 10) \leqslant 0.2$). Regardless, we still will want to apply CLT on X (the only difference is that we may be solving for different things).

1. 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 $X = X_1 + X_2 + \ldots + X_n$ (we can approximate X as a normal random variable $Y \sim N(\mu, \sigma^2)$)
   - Write the probability we're interested in
2. *If the RVs are discrete*, apply continuity correction.
3. Normalize RV to have mean 0 and standard deviation 1: $Z = \frac{Y-\mu}{\sigma}$
4. Replace RV in probability expression with $Z \sim N(0,1)$
5. Write in terms of $\Phi(z) = P(Z \leqslant z)$
6. Look up in the Phi table (or a reverse Phi table lookup if we're for a value of z that gives a certain probability)

Task 8 – 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 $|X - Y| > 3$?

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 $\text{Unif}(-0.5, 0.5)$. 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 $Z \sim \mathcal{N}(0, 1)$

$$\begin{aligned}
\mathbb{P}(|X - Y| > 3) &\approx \mathbb{P}(|W| > 3) && \text{[CLT]} \\
&= \mathbb{P}(W > 3) + \mathbb{P}(W < -3) && \text{[No overlap between } W > 3 \text{ and } W < -3] \\
&= 2\,\mathbb{P}(W > 3) && \text{[Symmetry of normal]} \\
&= 2\,\mathbb{P}\left(\frac{W}{\sqrt{100/12}} > \frac{3}{\sqrt{100/12}}\right) \\
&\approx 2\,\mathbb{P}(Z > 1.039) && \text{[Standardize } W] \\
&= 2\,(1 - \Phi(1.039)) \approx 0.29834
\end{aligned}$$

Task 9 – 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!).

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,

$$\mathbb{P}(26,000 \leqslant X \leqslant 27,000) \approx \mathbb{P}(26,000 \leqslant N \leqslant 27,000)$$

Now, we apply continuity correction:

$$\mathbb{P}(26,000 \leqslant N \leqslant 27,000) \approx \mathbb{P}(25,999.5 \leqslant N \leqslant 27,000.5)$$

Standardizing this gives the following formula

$$\begin{aligned}
\mathbb{P}(25,999.5 \leqslant N \leqslant 27,000.5) &\approx \mathbb{P}\left(-0.3541 \leqslant \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \leqslant 1.0608\right) \\
&= \mathbb{P}\left(-0.3541 \leqslant Z \leqslant 1.0608\right) \\
&= \mathbb{P}(Z \leqslant 1.0608) - \mathbb{P}(Z \leqslant -0.3541) \\
&= \Phi(1.0608) - \Phi(-0.3541) \\
&\approx 0.4923
\end{aligned}$$

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

Task 10 – Confidence interval

Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from a normal distribution with unknown mean μ and variance 36. How big does n need to be so that μ is in

$$[\overline{X} - 0.11, \overline{X} + 0.11]$$

with probability at least 0.97?

Recall that

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

You may use the fact that $\Phi^{-1}(0.985) = 2.17$.

Our goal is to find n such that μ lies within 0.11 of $\bar{X}$ 97% of the time. This is equivalent to finding n such that the probability that μ lies outside the range is less than 3%.

$$\mathbb{P}(|\bar{X} - \mu| > 0.11) \leqslant 0.03$$

Let us define $Z = \frac{\bar{X}-\mu}{\sigma}$. We can solve for σ by using the Properties of Variance. Since

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

we can say that

$$\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n} \sum_{i=1}^{n} X_i)$$

Using the Properties of Variance and the fact that X_i's are i.i.d., $\text{Var}(\bar{X}) = \frac{1}{n^2} \cdot n \cdot 36 = \frac{36}{n}$, so $\sigma = \frac{6}{\sqrt{n}}$.

$$\begin{aligned}
\mathbb{P}(|\bar{X} - \mu| > 0.11) &\leqslant 0.03 \\
\mathbb{P}(|Z| \cdot \sigma > 0.11) &\leqslant 0.03 && \text{[Definition of } Z\text{]} \\
\mathbb{P}\left(|Z| > \frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.03 \\
\mathbb{P}\left(Z < -\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[Symmetry of Normal Dist.]} \\
\Phi\left(-\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[CDF of Standard Norm.]} \\
-\frac{0.11}{6}\sqrt{n} &\leqslant -\Phi^{-1}(0.985) \\
\sqrt{n} &\geqslant \frac{6 \cdot \Phi^{-1}(0.985)}{0.11} \\
n &\geqslant \left(\frac{6 \cdot \Phi^{-1}(0.985)}{0.11}\right)^2 \\
&\approx 14009.95
\end{aligned}$$

Then n must be at least 14010.

Task 11 – Normal Approximation of a Sum

Imagine that we are trying to transmit a signal. During the transmission, there are 100 sources independently making low noise. Each source produces an amount of noise that is uniformly distributed between $a = -1$ and $b = 1$. If the total amount of noise is greater than 10 or less than -10, then it corrupts the signal. However, if the absolute value of the total amount of noise is under 10, then it is not a problem. What is the approximate probability that the absolute value of the total amount of noise from the 100 signals is less than 10?

Let S be the total amount of noise. We want to find $\mathbb{P}(|S| < 10) = \mathbb{P}(-10 < S < 10)$. Let X_i be the noise from source i. Then, we have

$$S = \sum_{i=1}^{100} X_i.$$

Since the X_i are uniformly distributed, we have that $\mathbb{E}[X_i] = \frac{a+b}{2} = 0$ and $\text{Var}(X_i) = \frac{(b-a)^2}{12} = \frac{1}{3}$. Since the X_i are i.i.d, by the Central Limit Theorem, we find that S is approximately distributed according to $N\left(0, 100 \cdot \frac{1}{3}\right)$. Now, we standardize to get

$$\begin{aligned}\mathbb{P}(-10 < S < 10) &= \mathbb{P}\left(\frac{-10-0}{\sqrt{100/3}} < \frac{S-0}{\sqrt{100/3}} < \frac{10-0}{\sqrt{100/3}}\right) \\ &= 2\Phi(\sqrt{3}) - 1 \approx 0.91\end{aligned}$$

Task 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 ($\Omega_{X,Y}$).

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

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

Note that $\Omega_X = \{0, 1\}$.

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

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

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

Note that $\Omega_Y = \{1, 2, 3\}$.

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

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

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

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

X and Y are not independent. Recall that a *necessary* condition for X and Y to be independent is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$. The joint range $\Omega_{X,Y}$ does not satisfy this criteria, so it cannot be independent.

e) Find $\mathbb{E}[X^3Y]$.

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

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

Task 13 – 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 ($\Omega_{X,Y}$).

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

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

Note that $\Omega_X = \{0, 1\}$.

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

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

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

Note that $\Omega_Y = \{1, 2, 3\}$.

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

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

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

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

X and Y are not independent. Recall that a *necessary* condition for X and Y to be independent is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$. The joint range $\Omega_{X,Y}$ does not satisfy this criteria, so it cannot be independent.

e) Find $\mathbb{E}[X^3Y]$.

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

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

Task 14 – Do You "Urn" to Learn More About Probability?

Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let $X_i = 1$ if the i-th ball selected is white and let it be equal to 0 otherwise. Give the joint probability mass function of

a) X_1, X_2

Here is one way of defining the joint pmf of X_1, X_2

$$p_{X_1,X_2}(1,1) = \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 1 \mid X_1 = 1) = \frac{5}{13} \cdot \frac{4}{12} = \frac{20}{156}$$

$$p_{X_1,X_2}(1,0) = \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 0 \mid X_1 = 1) = \frac{5}{13} \cdot \frac{8}{12} = \frac{40}{156}$$

$$p_{X_1,X_2}(0,1) = \mathbb{P}(X_1 = 0)\,\mathbb{P}(X_2 = 1 \mid X_1 = 0) = \frac{8}{13} \cdot \frac{5}{12} = \frac{40}{156}$$

$$p_{X_1,X_2}(0,0) = \mathbb{P}(X_1 = 0)\,\mathbb{P}(X_2 = 0 \mid X_1 = 0) = \frac{8}{13} \cdot \frac{7}{12} = \frac{56}{156}$$

b) X_1, X_2, X_3

Instead of listing out all the individual probabilities, we could write a more compact formula for the pmf. In this problem, the denominator is always $P(13, k)$, where k is the number of random variables in the joint pmf. And the numerator is $P(5, i)$ times $P(8, j)$ where i and j are the number of 1s and 0s, respectively.

If we wish to compute $p_{X_1,X_2,X_3}(x_1, x_2, x_3)$, then the number of 1s (i.e., white balls) is $x_1 + x_2 + x_3$, and the number of 0s (i.e., red balls) is $(1 - x_1) + (1 - x_2) + (1 - x_3)$. Then, we can write the pmf as follows:

$$p_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{10!}{13!} \cdot \frac{5!}{(5 - x_1 - x_2 - x_3)!} \cdot \frac{8!}{(5 + x_1 + x_2 + x_3)!}$$

Task 15 – Trinomial Distribution

A generalization of the Binomial model is when there is a sequence of n independent trials, but with three outcomes, where $\mathbb{P}(\text{outcome } i) = p_i$ for $i = 1, 2, 3$ and of course $p_1 + p_2 + p_3 = 1$. Let X_i be the number of times outcome i occurred for $i = 1, 2, 3$, where $X_1 + X_2 + X_3 = n$. Find the joint PMF $p_{X_1,X_2,X_3}(x_1, x_2, x_3)$ and specify its value for all $x_1, x_2, x_3 \in \mathbb{R}$.

Are X_1 and X_2 independent?

In a similar argument with the binomial PMF, we have

$$p_{X_1,X_2,X_3}(x_1, x_2, x_3) = \binom{n}{x_1}\binom{n - x_1}{x_2}\binom{n - x_1 - x_2}{x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}\ .$$

This may also be interpreted as multinomial coefficients (reference), and so we may rewrite as

$$p_{X_1,X_2,X_3}(x_1,x_2,x_3) = \binom{n}{x_1,x_2,x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3} = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \ ,$$

where $x_1 + x_2 + x_3 = n$ and are nonnegative integers.

X_1 and X_2 are not independent. For example $\mathbb{P}(X_1 = n) > 0$ and $\mathbb{P}(X_2 = n) > 0$, but $\mathbb{P}(X_1 = n, X_2 = n) = 0$. In other words, $\Omega_{X_1,X_2,X_3} \neq \Omega_{X_1} \times \Omega_{X_2} \times \Omega_{X_3}$, which is a necessary condition for independence.

Task 16 – Successes

Consider a sequence of independent Bernoulli trials, each of which is a success with probability p. Let X_1 be the number of failures preceding the first success, and let X_2 be the number of failures after the first success but preceding the second success. Find the joint pmf of X_1 and X_2. Write an expression for $\mathbb{E}[\sqrt{X_1X_2}]$. You can leave your answer in the form of a sum.

In order for X_1 to take on a particular value, say x_1, it must have x_1 failures until the first success, i.e., the next trial is a success. To that end, for X_1 and X_2 to take on two particular values x_1 and x_2, there must be x_1 failures followed by one success, and then x_2 failures followed by one success. Since the Bernoulli trials are independent, the joint pmf is

$$p_{X_1,X_2}(x_1,x_2) = (1-p)^{x_1} p \cdot (1-p)^{x_2} p = (1-p)^{x_1+x_2} p^2$$

for $(x_1,x_2) \in \Omega_{X_1,X_2} = \{0,1,2,\ldots\} \times \{0,1,2,\ldots\}$. By the definition of expectation and LOTUS,

$$\mathbb{E}[\sqrt{X_1X_2}] = \sum_{(x_1,x_2)\in\Omega_{X_1,X_2}} \sqrt{x_1x_2} \cdot p_{X_1,X_2}(x_1,x_2) = \sum_{(x_1,x_2)\in\Omega_{X_1,X_2}} \sqrt{x_1x_2} \cdot (1-p)^{x_1+x_2} p^2 \ .$$

Task 17 – 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 $X_1 \sim \text{Geometric}(p_1)$. Let $X_2 \sim \text{Geometric}(p_2)$. 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:

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

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 18 – 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 :

section8˙plot.png

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.

Task 19 – Grades and homework turn-in time

Suppose we're currently trying to find a relationship between the time a student turns in their homework and the grade that they receive on the respective homework. Let T denote the amount of time *prior* to the deadline that the homework is submitted. We have observed that no student submits the homework more than 2 days earlier than the deadline, and also no student submits their assignment late, so $0 \leqslant T \leqslant 2$. Now let G be a random variable, indicating the percentage that the student receives on the homework assignment, that is, $0 \leqslant G \leqslant 1$. Suppose G and T are continuous random variables, and their joint pdf is given by

$$f_{G,T}(g,t) = \begin{cases} \frac{9}{10}g^2t + \frac{1}{5} & \text{when } 0 \leqslant g \leqslant 1 \text{ and } 0 \leqslant t \leqslant 2 \\ 0 & \text{otherwise .} \end{cases}$$

For both parts, round your solution to three decimal places.

a) What is the probability that a randomly selected student gets a grade above 50% on the homework?

We are looking for $\mathbb{P}(G > 0.5)$. To do this, we must first compute the marginal density function $f_G(g)$. Applying by definition,

$$\begin{aligned} f_G(g) &= \int_{-\infty}^{\infty} f_{G,T}(g,t)\,dt \\ &= \int_0^2 f_{G,T}(g,t)\,dt = \int_0^2 \frac{9}{10}g^2t + \frac{1}{5}\,dt = \left(\frac{9}{10}\frac{1}{2}t^2g^2 + \frac{1}{5}t\right)\Bigg|_0^2 = \frac{9}{5}g^2 + \frac{2}{5}\,. \end{aligned}$$

Then

$$\mathbb{P}(G > 0.5) = \int_{0.5}^{\infty} f_G(g)\,dg = \int_{0.5}^{1} \frac{9}{5}g^2 + \frac{2}{5}\,dg = \frac{29}{40} = 0.725\,.$$

b) What is the probability that a student gets a grade above 50%, given that the student submitted less than a day before the deadline?

We are looking for

$$\mathbb{P}(G > 0.5 \mid T < 1) = \frac{\mathbb{P}(G > 0.5 \cap T < 1)}{\mathbb{P}(T < 1)}\,,$$

which follows by the definition of conditional probability. The numerator can be computed using the joint pdf. However, the denominator needs us to calculate the marginal pdf. We can follow a similar approach to the previous part and get

$$f_T(t) = \int_0^1 f_{G,T}(g,t)\,dg = \int_0^1 \frac{9}{10}g^2t + \frac{1}{5}\,dg = \frac{3}{10}t + \frac{1}{5}\,.$$

Thus,

$$\mathbb{P}(G > 0.5 \mid T < 1) = \frac{\int_{0.5}^1 \int_0^1 f_{G,T}(g,t)\,dt\,dg}{\int_0^1 f_T(t)\,dt} = \frac{\int_{0.5}^1 \int_0^1 \frac{9}{10}g^2t + \frac{1}{5}\,dt\,dg}{\int_0^1 \frac{3}{10}t + \frac{1}{5}\,dt} \approx 0.661\,.$$