Section 9: Solutions

Review of Main Concepts

- **Realization/Sample**: A realization/sample x of a random variable X is the value that is actually observed.
- **Likelihood**: Let $x_1, \dots x_n$ be iid realizations from probability mass function $p_X(\text{x}\,;\theta)$ (if X discrete) or density $f_X(\text{x}\,;\theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data.

  If X is discrete:
  $$L(x_1, \dots, x_n\,;\,\theta) = \prod_{i=1}^{n} p_X(x_i\,;\,\theta)$$
  If X is continuous:
  $$L(x_1, \dots, x_n\,;\,\theta) = \prod_{i=1}^{n} f_X(x_i\,;\,\theta)$$

- **Maximum Likelihood Estimator (MLE)**: We denote the MLE of θ as $\hat{\theta}_{\text{MLE}}$ or simply $\hat{\theta}$, the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).
  $$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(x_1, \dots, x_n\,;\,\theta) = \arg\max_{\theta} \ln L(x_1, \dots, x_n\,;\,\theta)$$

- **Log-Likelihood**: We define the log-likelihood as the natural logarithm of the likelihood function. Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

  If X is discrete:
  $$\ln L(x_1, \dots, x_n\,;\,\theta) = \sum_{i=1}^{n} \ln p_X(x_i\,;\,\theta)$$
  If X is continuous:
  $$\ln L(x_1, \dots, x_n\,;\,\theta) = \sum_{i=1}^{n} \ln f_X(x_i\,;\,\theta)$$

- **Bias**: The bias of an estimator $\hat{\theta}$ for a true parameter θ is defined as $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$. An estimator $\hat{\theta}$ of θ is unbiased iff $\text{Bias}(\hat{\theta}, \theta) = 0$, or equivalently $\mathbb{E}[\hat{\theta}] = \theta$.
- **Steps to find the maximum likelihood estimator, $\hat{\theta}$:**
  (a) Find the likelihood and log-likelihood of the data.
  (b) Take the derivative of the log-likelihood
  (c) Set it to 0 to find a candidate for the MLE, $\hat{\theta}$. (note: at this step, we change from the θ to the $\hat{\theta}$ because in this step we are solving for the *maximum* likelihood estimator for θ)
  (d) Take the second derivative and show that $\hat{\theta}$ indeed is a maximizer, that $\frac{\partial^2 L}{\partial \theta^2} < 0$ at $\hat{\theta}$. Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

1. Content Review

(a) **True or False**: The Log-Likelihood gives a slightly different estimate, but because it is close enough and easier to compute we use it for our estimate of θ.
Solution:

False: Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

(b) **True or False**: $\hat{\theta}$ is the true parameter and θ is our estimate. **Solution:**

False: It is the other way around. Remember to switch to $\hat{\theta}$ when you set your equation to zero!

(c) **True or False**: An estimator is unbiased if $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta = 0$ or equivalently $\mathbb{E}[\hat{\theta}] = \theta$ **Solution:**

True by definition of

(d) You flip a coin 10 times and observe HHHTHHTHHH (8 heads, 2 tails). What is the MLE of θ, where θ is the true probability of seeing tails?

- ◯ $\hat{\theta} = .2$
- ◯ $\hat{\theta} = .25$
- ◯ $\hat{\theta} = .8$
- ◯ $\hat{\theta} = .3$

Solution:

Option 1: $\hat{\theta} = .2$

0. 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 $\mathbb{E}[X]$?

Solution:

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

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

Hence, by linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{7} (20X_i - 100)\right] = 20 \sum_{i=1}^{7} \mathbb{E}[X_i] - 700 = 140\mu - 700$$

$$= 140 \cdot (n_1 p_1 p_3 + n_2 p_2 (1 - p_3)) - 700.$$

2. Mystery Dish!

A fancy new restaurant has opened up which features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability θ, dish C with probability 2θ, and dish D with probability $0.5 - 3\theta$

Each diner is served a dish independently. Let x_A be the number of people who received dish A, x_B the number of people who received dish B, etc, where $x_A + x_B + x_C + x_D = n$. Find the $\hat{\theta}$, the maximum likelihood estimator for θ.

Solution:

The data tells us, for each diner in the restaurant, what their dish was. We begin by computing the likelihood of seeing the given data given our parameter θ. Because each diner is assigned a dish independently, the likelihood is equal to the product over diners of the chance they got the particular dish they got, which gives us:

$$L(x\,;\,\theta) = 0.5^{x_A}\theta^{x_B}(2\theta)^{x_C}(0.5-3\theta)^{x_D}$$

From there, we just use the MLE process to get the log-likelihood, take the first derivative, set it equal to 0, and solve for $\hat{\theta}$.

$$\ln L(x\,;\,\theta) = x_A \ln(0.5) + x_B \ln(\theta) + x_C \ln(2\theta) + x_D \ln(0.5 - 3\theta)$$

$$\frac{\partial}{\partial\theta} \ln L(x\,;\,\theta) = \frac{x_B}{\theta} + \frac{x_C}{\theta} - \frac{3x_D}{0.5 - 3\theta}$$

$$\frac{x_B}{\hat{\theta}} + \frac{x_C}{\hat{\theta}} - \frac{3x_D}{0.5 - 3\hat{\theta}} = 0$$

Solving yields $\hat{\theta} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$.

3. A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

Solution:

Because each Poisson RV is i.i.d., the likelihood of seeing that data is just the PMF of the Poisson distribution multiplied together for every x_i. From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for $\hat{\theta}$.

$$\begin{aligned}
L(x_1, \ldots, x_n\,;\,\theta) &= \prod_{i=1}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!} \\
\ln L(x_1, \ldots, x_n\,;\,\theta) &= \sum_{i=1}^{n} [-\theta - \ln(x_i!) + x_i \ln(\theta)] \\
\frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_n\,;\,\theta) &= \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\theta}\right] \\
-n + \frac{\Sigma_{i=1}^{n} x_i}{\hat{\theta}} &= 0 \\
\hat{\theta} = \frac{\Sigma_{i=1}^{n} x_i}{n}&.
\end{aligned}$$

Notice that the $-\ln(x_i!)$ term disappears since it is a constant relative to θ, of which we take the derivative.

4. Independent Shreds, You Say?

You are given 100 independent samples $x_1, x_2, \ldots, x_{100}$ from Bernoulli(θ), where θ is unknown. (Each sample is either a 0 or a 1). These 100 samples sum to 30. You would like to estimate the distribution's parameter θ. Give all answers to 3 significant digits.

(a) What is the maximum likelihood estimator $\hat{\theta}$ of θ?

Solution:

Note that $\Sigma_{i\in[n]} x_i = 30$, as given in the problem spec. Therefore, there are 30 **1**s and 70 **0**s. (Note that they come in some specific order.) Therefore, we can setup L as follows, because there is a θ chance of getting a 1, and a $(1-\theta)$ chance of getting a 0 and they are each i.i.d. From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for for $\hat{\theta}$.

$$\begin{aligned}
L(x_1, \ldots, x_n\,;\,\theta) &= (1-\theta)^{70}\theta^{30} \\
\ln L(x_1, \ldots, x_n\,;\,\theta) &= 70\ln(1-\theta) + 30\ln\theta \\
\frac{\partial}{\partial\theta}\ln L(x_1, \ldots, x_n\,;\,\theta) &= -\frac{70}{1-\theta} + \frac{30}{\theta} \\
-\frac{70}{1-\hat{\theta}} + \frac{30}{\hat{\theta}} &= 0 \\
\frac{30}{\hat{\theta}} &= \frac{70}{1-\hat{\theta}} \\
30 - 30\hat{\theta} &= 70\hat{\theta} \\
\hat{\theta} &= \frac{30}{100}
\end{aligned}$$

(b) Is $\hat{\theta}$ an unbiased estimator of θ?

Solution:

An estimator is unbiased if the expectation of the estimator is equal to the original parameter, i.e.: $E[\hat{\theta}] = \theta$. Setting up the expectation of our estimator and plugging it in for the generic case, we get the following, which we can then reduce with linearity of expectation:

$$\begin{aligned}
\mathbb{E}[\hat{\theta}] &= \mathbb{E}\left[\frac{1}{100}\sum_{i=1}^{100} X_i\right] \\
&= \frac{1}{100}\sum_{i=1}^{100}\mathbb{E}[X_i] \\
&= \frac{1}{100}\cdot 100\theta = \theta.
\end{aligned}$$

so it is unbiased.

5. Y Me?

Let $y_1, y_2, \ldots y_n$ be i.i.d. samples of a random variable with density function

$$f_Y(y\,;\,\theta) = \frac{1}{2\theta}\exp\left(-\frac{|y|}{\theta}\right).$$

Find the MLE for θ in terms of $|y_i|$ and n.

Solution:

Since the samples are i.i.d., the likelihood of seeing n samples of them is just their PDFs multiplied together. From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for for $\hat{\theta}$.

$$\begin{aligned}
L(y_1, \ldots, y_n\,;\,\theta) &= \prod_{i=1}^{n} \frac{1}{2\theta} \exp\left(-\frac{|y_i|}{\theta}\right) \\
\ln L\,(y_1, \ldots, y_n \mid \theta) &= \sum_{i=1}^{n} \left[-\ln 2 - \ln\theta - \frac{|y_i|}{\theta}\right] \\
\frac{\partial}{\partial\theta} \ln L(y_1, \ldots, y_n\,;\,\theta) &= \sum_{i=1}^{n} \left[-\frac{1}{\theta} + \frac{|y_i|}{\theta^2}\right] \\
\sum_{i=1}^{n} \left[-\frac{1}{\hat{\theta}} + \frac{|y_i|}{\hat{\theta}^2}\right] &= 0 \\
-\frac{n}{\hat{\theta}} + \frac{\Sigma_{i=1}^{n}|y_i|}{\hat{\theta}^2} &= 0 \\
\hat{\theta} = \frac{\Sigma_{i=1}^{n}|y_i|}{n}
\end{aligned}$$

6. Bird Watching

You are an ornithologist studying a rare species of birds in a nature reserve. Over a period of 50 days, you record the number of sightings of this bird (you see $x_1, x_2, ..., x_{50}$ birds on each day). Your research has shown that the number of sightings on this species depends on the average number of monkeys in the reserve, θ_1, and the average number of eagles in the reserve, θ_2. After years of studying this rare species in other environments, you've found the number of birds observed on a particular day follows the following distribution:

$$p_X(k) = \frac{1}{k!}(\theta_1^k \cdot e^{-\theta_1} \cdot \theta_2^k \cdot e^{-3\theta_2})$$

Find the MLE for θ_1 and θ_2 (i.e., find $\hat{\theta_1}$ and $\hat{\theta_2}$).

(a) What is the likelihood function? **Solution:**

Once again, the likelihood of seeing the above samples is just their PDFs multiplied together.

$$L(x; \theta_1, \theta_2) = \prod_{i=1}^{50} \left(\frac{1}{x_i!}(\theta_1^{x_i} e^{-\theta_1} \cdot \theta_2^{x_i} e^{-3\theta_2})\right)$$

(b) What is the log-likelihood function? **Solution:**

We take the log of the above and simplify:

$$\begin{aligned}\ln(L(x;\theta_1,\theta_2)) &= \ln(\prod_{i=1}^{50}(\frac{1}{x_i!}(\theta_1^{x_i}\cdot e^{-\theta_1}\cdot\theta_2^{x_i}\cdot e^{-3\theta_2}))) &(1)\\ &= \sum_{i=1}^{50}(\ln(\frac{1}{x_i!})+\ln(\theta_1^{x_i})+\ln(e^{-\theta_1})+\ln(\theta_2^{x_i})+\ln(e^{-3\theta_2})) &(2)\\ &= \sum_{i=1}^{50}(\ln(\frac{1}{x_i!})+x_i\ln(\theta_1)-\theta_1+x_i\ln(\theta_2)+-3\theta_2) &(3)\end{aligned}$$

(c) We want to find values of θ_1 and θ_2 that maximize the likelihood function. To do this, we will take the partial derivative with respective to each of these parameters and solve for the values that make them both zero. First, take the partial derivative of the likelihood function with respect to θ_1. **Solution:**

When taking the partial derivative with respect to a certain variable, we take the derivative as usual, but treat other variables as a constant! So here, we treat θ_2 as a constant and derivate with respect to θ_1.

$$\begin{aligned}\frac{\partial}{\partial\theta_1}(ln(L(x;\theta_1,\theta_2)))) &= \sum_{i=1}^{50}(\frac{x_i}{\theta_1}-1) &(4)\\ &= \sum_{i=1}^{50}(\frac{x_i}{\theta_1})-50 &(5)\end{aligned}$$

(d) Now, take the partial derivative with respect to θ_2. **Solution:**

Now, we treat θ_1 as a constant and derivate with respect to θ_2.

$$\begin{aligned}\frac{\partial}{\partial\theta_2}(L(ln(x;\theta_1,\theta_2))) &= \sum_{i=1}^{50}(\frac{x_i}{\theta_2}-3) &(6)\\ &= \sum_{i=1}^{50}(\frac{x_i}{\theta_2})-150 &(7)\end{aligned}$$

(e) Set both these partial derivatives to 0, and solve for $\hat{\theta_1}$ and $\hat{\theta_2}$. **Solution:**

We end up with the equations (notice we added the hats to the thetas at this point - since we set these derivatives to 0, we are now solving for the *maximum* likelihood estimator):

$$\sum_{i=1}^{50}(\frac{x_i}{\hat{\theta_1}})-50=0$$

$$\sum_{i=1}^{50}(\frac{x_i}{\hat{\theta_2}})-150=0$$

We now solve this system of equations for $\hat{\theta_1}$ and $\hat{\theta_2}$. The first equation gives us:

$$\sum_{i=1}^{50}(\frac{x_i}{\hat{\theta_1}}) = 50$$

$$(\frac{\sum_{i=1}^{50} x_i}{\hat{\theta_1}}) = 50$$

$$\hat{\theta_1} = (\frac{\sum_{i=1}^{50} x_i}{50})$$

With similar steps, we get that:

$$\hat{\theta_2} = (\frac{\sum_{i=1}^{50} x_i}{150})$$

7. A biased estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2)\right)$$

where $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

Solution:

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

$$E(\hat{\theta}_2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right)$$

which by linearity of expectation (and distributing the sum) is

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E\left(\frac{2}{n}\overline{X}\sum_{i=1}^{n} X_i\right) + E(\overline{X}^2)$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - 2E(\overline{X}^2) + E(\overline{X}^2)$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2). \qquad (**)$$

We know that for any random variable Y, since $Var(Y) = E(Y^2) - (E(Y))^2$ it holds that

$$E(Y^2) = Var(Y) + (E(Y))^2.$$

Also, we have $E(X_i) = \mu,\ Var(X_i) = \sigma^2\ \forall i$ and $E(\overline{X}) = \mu,\ Var(\overline{X}) = \frac{\sigma^2}{n}$. Combining these facts, we get

$$E(X_i^2) = \sigma^2 + \mu^2\ \ \forall i \quad \text{and} \quad E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2.$$

Substituting these equations into (**) we get

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2)\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)-E(\overline{X}^2)=\sigma^2+\mu^2-\left(\frac{\sigma^2}{n}+\mu^2\right)$$
$$=\left(1-\frac{1}{n}\right)\sigma^2.$$

Thus $\hat{\theta}_2$ is not unbiased.

8. It Means Nothing

Suppose $x_1, x_2, \ldots, x_n$ are samples from a normal distribution whose mean is known to be μ but the variance is unknown. How does the maximum likelihood estimator for the variance differ from the maximum likelihood estimator when both mean and variance are unknown? Which if any is unbiased? **Solution:**

Begin with the same derivation as before, however we now use μ instead of the mean of 0, which gets us:

$$\begin{aligned}
L(x_1,\ldots,x_n\,;\,\sigma^2) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x-\mu)^2}{2\sigma^2}\\
\ln L(x_1,\ldots,x_n\,;\,\sigma^2) &= \sum_{i=1}^{n}-\ln\sqrt{2\pi\sigma^2}-\frac{(x_i-\mu)^2}{2\sigma^2}\\
&= \sum_{i=1}^{n}-\frac{1}{2}\ln 2\pi\sigma^2-\frac{(x_i-\mu)^2}{2\sigma^2}\\
&= \sum_{i=1}^{n}-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^2-\frac{(x_i-\mu)^2}{2\sigma^2}\\
&= -\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\\
\frac{\partial}{\partial\sigma^2}\ln L(x_1,\ldots,x_n\,;\,\sigma^2) &= -\frac{n}{2\sigma^2}+\frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0\\
\frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} &= \frac{n}{2\sigma^2}\\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}$$

Then, we do the same but with two parameters, θ_1, θ_2, the former being the mean, and the latter being the variance. We can take the derivative with respect to θ_2, and do effectively the same as before.

$$\begin{aligned}
L(x_1,\ldots,x_n\,;\,\theta_1,\theta_2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp \frac{-(x_i-\theta_1)^2}{2\theta_2} \\
\ln L(x_1,\ldots,x_n\,;\,\theta_1,\theta_2) &= \sum_{i=1}^{n} -\ln\sqrt{2\pi\theta_2} - \frac{(x_i-\theta_1)^2}{2\theta_2} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\ln 2\pi\theta_2 - \frac{(x_i-\theta_1)^2}{2\theta_2} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\theta_2 - \frac{(x_i-\theta_1)^2}{2\theta_2} \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta_2 - \frac{\Sigma_{i=1}^{n}(x_i-\theta_1)^2}{2\theta_2} \\
\frac{\partial}{\partial\theta_2}\ln L(x_1,\ldots,x_n\,;\,\theta_1,\theta_2) &= -\frac{n}{2\theta_2} + \frac{\Sigma_{i=1}^{n}(x_i-\theta_1)^2}{2\theta_2^2} = 0 \\
\frac{\Sigma_{i=1}^{n}(x_i-\theta_1)^2}{2\theta_2^2} &= \frac{n}{2\theta_2} \\
\hat{\theta_2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\theta}_1)^2
\end{aligned}$$

Now, we just need to find if both estimators are biased or unbiased. We do this by seeing if their expected value is equal to the original parameter or not. Let's start with the former. We move the expectation in with linearity of expectation, and then can identify that the remaining expectation is just the definition of variance (expected deviation from the mean squared) and see that it is unbiased.

$$E[\hat{\sigma^2}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(x_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2$$

We do the same with the other estimator, and find that is biased, since it does not reduce down to the true parameter θ_2:

$$E[\hat{\theta_2}] = E[\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\theta_1})^2] = \frac{1}{n}\sum_{i=1}^{n}E[(x_i-\hat{\theta_1})^2] = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta_2} = \frac{1}{n}n\hat{\theta_2} = \hat{\theta_2}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

vs.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\theta_1})^2$$

(The former turns out to be unbiased, the latter biased.)

9. Covariance Connection

Let X be the network connection status, where $X = 0$ represents a stable connection and $X = 1$ represents an unstable connection. Let Y be the number of successes in data transmission, taking values in the set $\{0,1,2\}$. If $X = 0$, Y follows a Binomial distribution $\text{Bin}(2, 0.8)$, and if $X = 1$, Y follows a Binomial distribution $\text{Bin}(2, 0.3)$. The probabilities for X are given by $P(X = 0) = 0.8$ and $P(X = 1) = 0.2$. Find $\text{Cov}(X, Y)$. (note that we don't know that X and Y are independent here!)

Solution:

To calculate the covariance Cov(X, Y), we need to determine $E[X]$, $E[Y]$, and $E[XY]$. The covariance is then given by the formula:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

First, we calculate $E[X]$: $E[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0 \cdot 0.8 + 1 \cdot 0.2 = 0.2$

Next, we calculate $E[Y]$. First, we calculate $E[Y \mid X = 0]$ and $E[Y \mid X = 1]$. Based on what's given in the problem and using the formula for expectation for a binomial: $E[Y \mid X = 0] = 2 \cdot 0.8 = 1.6$ and $E[Y \mid X = 1] = 2 \cdot 0.3 = 0.6$. Using the law of total expectation:

$$E[Y] = E[Y \mid X = 0]P(X = 0) + E[Y \mid X = 1]P(X = 1) = 1.6 \cdot 0.8 + 0.6 \cdot 0.2 = 1.4$$

To compute $E[XY]$, we first construct the joint PMF for XY and then use the definition of expectation. The possible values for XY are 0, 1, and 2. Let's compute the probabilities for each value:

$$\begin{aligned} P(XY = 0) &= P(X = 0 \cup Y = 0) = P(X = 0) + P(Y = 0) - P(X = 0 \cap Y = 0) \\ &= P(X = 0) + P(Y = 0) - P(X = 0)P(Y = 0 \mid X = 0) = 0.8 + 0.13 - 0.8 \cdot 0.2^2 = 0.898 \end{aligned}$$

$$P(XY = 1) = P(X = 1 \cap Y = 1) = P(X = 1)P(Y = 1 \mid X = 1) = 0.2 \cdot (2 \cdot 0.3 \cdot 0.7) = 0.084$$

$$P(XY = 2) = P(X = 1 \cap Y = 2) = P(X = 1)P(Y = 2 \mid X = 1) = 0.2 \cdot (0.3^2) = 0.018$$

In the above calculations, we use that $P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1) = 0.2^2 \cdot 0.8 + 0.7^2 \cdot 0.2 = 0.13$ Now, using the definition of expectation, we have:

$$E[XY] = 0 \cdot 0.898 + 1 \cdot 0.084 + 2 \cdot 0.018 = 0.12$$

Therefore, $E[XY] = 0.12$. Finally, we calculate the covariance Cov(X, Y):

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.12 - 0.2 \cdot 1.4 = -0.16$$

Therefore, the covariance Cov(X, Y) is -0.16. The negative covariance of -0.16 between the network connection status X and the number of successes in data transmission Y indicates an inverse relationship, suggesting that as the network connection status becomes less stable (i.e., as $X = 1$), the likelihood of successess in data transmission decreases, and vice versa, as expected!

10. 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 $x_1, x_2, x_3 \in \mathbb{R}$. **Solution:**

We use a similar argument as for the binomial PMF. $\binom{n}{x_1,x_2,x_3}$ is the number of ways to select which of the n outcomes result in each of the 3 outcomes. Then, we multiply the probabilities of each trial being the corresponding outcome (e.g., $p_1^{x_1}$ is the probability that all x_1 trials end up being outcome 1). This gives use the following PMF:

11. 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 **Solution:**

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

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

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

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

(b) X_1, X_2, X_3 **Solution:**

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:

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

12. 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 between the first 2 successes. Find the joint pmf of X_1 and X_2. Write an expression for $E[\sqrt{X_1 X_2}]$. You can leave your answer in the form of a sum. **Solution:**

X_1 and X_2 take on two particular values x_1 and x_2, when there are 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

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

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 :

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.

14. 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?

Solution:

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.

15. 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? **Solution:**

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

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