HW 13

Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (*signature here*)

1 Normal Distribution

Recall the following facts about the normal distribution: if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the random variable $Z = (X-\mu)/\sigma$ is standard normal, i.e. $Z \sim \mathcal{N}(0,1)$. There is no closed-form expression for the CDF of the standard normal distribution, so we define $\Phi(z) = \mathbb{P}[Z \leq z]$. You may express your answers in terms of $\Phi(z)$.

The average jump of a certain frog is 3 inches. However, because of the wind, the frog does not always go exactly 3 inches. A zoologist tells you that the distance the frog travels is normally distributed with mean 3 and variance $1/4$.

(a) What is the probability that the frog jumps more than 4 inches?

(b) What is the probability that the distance the frog jumps is between 2 and 4 inches?

Solution:

(a) First, we write down the probability we want to find, then transform the probability in order to work with the standard normal.

$$\mathbb{P}[X > 4] = \mathbb{P}[X-3 > 1] = \mathbb{P}\left[\frac{X-3}{1/2} > 2\right] = \mathbb{P}[Z > 2] = 1 - \Phi(2) \approx 0.0228$$

(b) Since the mean of the jump is 3, and the normal distribution is symmetric, we can rewrite the desired probability as

$$\mathbb{P}[2 < X < 4] = 1 - (\mathbb{P}[X > 4] + \mathbb{P}[X < 2]) = 1 - 2 \cdot \mathbb{P}[X > 4].$$

We have computed $\mathbb{P}[X > 4] = 0.0228$ in Part (a), so we can plug this in to get 0.9544.

2 Binomial CLT

In this question we will explicitly see why the central limit theorem holds for the binomial distribution as the number of coin tosses grows.

Let X be the random variable showing the total number of heads in n independent coin tosses.

(a) Compute the mean and variance of X. Show that $\mu = \mathbb{E}[X] = n/2$ and $\sigma^2 = \operatorname{var} X = n/4$.

(b) Prove that $\mathbb{P}[X = k] = \binom{n}{k}/2^n$.

(c) Show by using Stirling's formula that

$$\mathbb{P}[X = k] \simeq \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2k}\right)^k \left(\frac{n}{2(n-k)}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}}.$$

In general we expect $2k$ and $2(n-k)$ to be close to n for the probability to be non-negligible. When this happens we expect $\sqrt{\frac{n}{k(n-k)}}$ to be close to $\sqrt{\frac{n}{(n/2) \times (n/2)}} = \frac{2}{\sqrt{n}}$. So replace that part of the formula by $2/\sqrt{n}$.

(d) In order to normalize X, we need to subtract the mean, and divide by the standard deviation. Let $Y = (X - \mu)/\sigma$ be the normalized version of X. Note that Y is a discrete random variable. Determine the set of values that Y can take. What is the distance d between two consecutive values?

(e) Let $X = k$ correspond to the event $Y = t$. Then $X \in [k - 0.5, k + 0.5]$ corresponds to $Y \in [t - d/2, t + d/2]$. For conceptual simplicity, it is reasonable to assume that the mass at point t is distributed uniformly on the interval $[t - d/2, t + d/2]$. We can capture this with the idea of a "probability density" and say that the probability density on this interval is just $\mathbb{P}[Y = t]/d = \mathbb{P}[X = k]/d$.

Compute k as a function of t. Then substitute that for k in the approximation you have from part c to find an approximation for $\mathbb{P}[Y = t]/d$. Show that the end result is equivalent to:

$$\frac{1}{\sqrt{2\pi}} \left[\left(1 + \frac{t}{\sqrt{n}}\right)^{1 + t/\sqrt{n}} \left(1 - \frac{t}{\sqrt{n}}\right)^{1 - t/\sqrt{n}} \right]^{-n/2}$$

(f) As you can see, we have expressions of the form $(1+x)^{1+x}$ in our approximation. To simplify them, write $(1+x)^{1+x}$ as $\exp((1+x)\ln(1+x))$ and then replace $(1+x)\ln(1+x)$ by its Taylor series.

The Taylor series up to the x^2 term is $(1+x)\ln(1+x) \simeq x + x^2/2 + \cdots$ (feel free to verify this by hand). Use this to simplify the approximation from the last part. In the end you should get the familiar formula that appears inside the CLT:

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right).$$

(The CLT is essentially taking a sum with lots of tiny slices and approximating it by an integral of this function. Because the slices are tiny, dropping all the higher-order terms in the Taylor expansion is justified.)

Solution:

(a) We can write X as a sum: $X = Y_1 + \cdots + Y_n$ where each Y_i is a Bernoulli random variable; i.e. $Y_i = 1$ when the i-th coin toss is heads and 0 if it is tails. Then from linearity of expectation we have

$$\mathbb{E}[X] = \mathbb{E}[Y_1] + \cdots + \mathbb{E}[Y_n] = n\,\mathbb{E}[Y_1] = \frac{n}{2}$$

where we used the fact that all Y_i have the same expectation which is $1/2$.

To compute the variance, note that because $Y_1, \ldots, Y_n$ are independent, we can decompose the variance into a sum of variances. Therefore we have the following:

$$\operatorname{var} X = \operatorname{var} Y_1 + \cdots + \operatorname{var} Y_n$$

Now in order to compute $\operatorname{var} Y_i$, note that by definition $\operatorname{var} Y_i = \mathbb{E}[(Y_i - \mathbb{E}[Y_i])^2]$. We know that $\mathbb{E}[Y_i] = 1/2$, so $\operatorname{var} Y_i = \mathbb{E}[(Y_i - 1/2)^2]$. But note that Y_i takes the values 0 and 1, therefore $(Y_i - 1/2)^2$ always takes the value $1/4$. So its expectation is also $1/4$. This means that

$$\operatorname{var} X = \frac{1}{4} + \cdots + \frac{1}{4} = \frac{n}{4}.$$

(b) The number of configurations of heads/tails for the coins that result in k coins being heads is $\binom{n}{k}$, since there are this many ways to pick the positions of the heads. Each configuration of heads/tails is equally likely and they each have probability $1/2^n$, because the coins are independent and the probability of each coin being in a specific state is $1/2$. So the total probability for the event $X = k$ is $\binom{n}{k}/2^n$.

(c) We need Stirling's formula to approximate the $\binom{n}{k}$ part. Remember that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

and Stirling's approximation says that $m! \simeq \sqrt{2\pi m}(m/\mathrm{e})^m$. Therefore we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \simeq \frac{\sqrt{2\pi n}(n/\mathrm{e})^n}{\sqrt{2\pi k}(k/\mathrm{e})^k\sqrt{2\pi(n-k)}((n-k)/\mathrm{e})^{n-k}}.$$

We can break the $(n/\mathrm{e})^n$ part into $(n/\mathrm{e})^k(n/\mathrm{e})^{n-k}$ and then combine these with the denominator. By doing this the part $(n/\mathrm{e})^k/(k/\mathrm{e})^k$ becomes $(n/k)^k$ and the part $(n/\mathrm{e})^{n-k}/((n-k)/\mathrm{e})^{n-k}$ becomes $(n/(n-k))^{n-k}$.

As for the parts under the square root, one of the $\sqrt{2\pi}$'s in the denominator cancels the one in the numerator and therefore only one remains in the denominator. We also get

$$\frac{\sqrt{n}}{\sqrt{k}\sqrt{n-k}} = \sqrt{\frac{n}{k(n-k)}}.$$

Therefore we have

$$\binom{n}{k} \simeq \frac{1}{\sqrt{2\pi}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}\sqrt{\frac{n}{k(n-k)}}.$$

Now we need to divide both sides by 2^n to get to $\mathbb{P}[X=k]$. We can write $2^n = 2^k 2^{n-k}$ and merge each term into the corresponding power. We get

$$\binom{n}{k}2^{-n} \simeq \frac{1}{\sqrt{2\pi}}\left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k}\sqrt{\frac{n}{k(n-k)}}$$

which is what we wanted to prove.

If we replace the $\sqrt{\frac{n}{k(n-k)}}$ part with $2/\sqrt{n}$ we get

$$\binom{n}{k}2^{-n} \simeq \frac{1}{\sqrt{2\pi}}\left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k}\frac{2}{\sqrt{n}}.$$

(d) The set of values that X can take is $\{0,\ldots,n\}$. Therefore the set of values that Y can take is $(i-n/2)/(\sqrt{n}/2) = (2i-n)/\sqrt{n}$ for $i = 0,\ldots,n$. Originally (for X) the distance between consecutive values is 1, but since we are dividing by $\sigma = \sqrt{n}/2$, this distance becomes $1/\sigma = 2/\sqrt{n}$. Note that subtracting the mean has no effect on the distance between consecutive points.

(e) We know how to compute t as a function of k. We simply do what we do to X to get to Y, i.e. subtract the mean of X and divide by its standard deviation. Therefore

$$t = \frac{k-n/2}{\sqrt{n}/2} = \frac{2k-n}{\sqrt{n}}.$$

Now to reverse this process and go from t to k we need to do the reverse, i.e. first multiply by σ and then add the mean of X. Therefore $k = \sqrt{n}t/2 + n/2 = (\sqrt{n}t+n)/2$.

Now note that $n/(2k) = n/(\sqrt{n}t+n) = ((\sqrt{n}t+n)/n)^{-1} = (1+t/\sqrt{n})^{-1}$. Similarly we have $n/(2(n-k)) = n/(2n-n-\sqrt{n}t) = ((n-\sqrt{n}t)/n)^{-1} = (1-t/\sqrt{n})^{-1}$.

Now we can write $(n/(2k))^k$ as $(1+t/\sqrt{n})^{-k}$ and $(n/(2(n-k)))^{n-k}$ as $(1-t/\sqrt{n})^{-(n-k)}$. To get rid of k even in the exponent we need to write it in terms of t. We have $-k = -(\sqrt{n}t + n)/2 = -(n/2)(1+t/\sqrt{n})$. Similarly we have $-(n-k) = -(n-n/2-\sqrt{n}t/2) = -(n/2)(1-t/\sqrt{n})$.

Now it's time to assemble the pieces. Remember that we had

$$\mathbb{P}[X=k] = \mathbb{P}[Y=t] \simeq \frac{1}{\sqrt{2\pi}}\left(\frac{n}{2k}\right)^k\left(\frac{n}{2(n-k)}\right)^{n-k}\frac{2}{\sqrt{n}}.$$

Replacing the parts $(n/(2k))^k$ and $(n/(2(n-k)))^{n-k}$ the way we described gives us

$$\mathbb{P}[Y=t] \simeq \frac{1}{\sqrt{2\pi}}\left(1+\frac{t}{\sqrt{n}}\right)^{-(n/2)(1+t/\sqrt{n})}\left(1-\frac{t}{\sqrt{n}}\right)^{-(n/2)(1-t/\sqrt{n})}\frac{2}{\sqrt{n}}.$$

We need to approximate $\mathbb{P}[Y=t]/d$, and note that $d = 2/\sqrt{n}$ which is exactly the last term appearing in our approximation of $\mathbb{P}[Y=t]$. So by dividing by d, that term simply cancels out and we get

$$\frac{\mathbb{P}[Y=t]}{d} \simeq \frac{1}{\sqrt{2\pi}}\left[\left(1+\frac{t}{\sqrt{n}}\right)^{1+t/\sqrt{n}}\left(1-\frac{t}{\sqrt{n}}\right)^{1-t/\sqrt{n}}\right]^{-n/2}.$$

(f) The term $(1+x)^{1+x}$ as suggested can be written as $\exp((1+x)\ln(1+x))$ and then $(1+x)\ln(1+x)$ can be replaced by its Taylor series up to the first few terms, i.e. by $x+x^2/2$. Now if we also do this for $-x$, we get $(1-x)^{1-x} = \exp((1-x)\ln(1-x)) \simeq \exp(-x+x^2/2)$. By multiplying our approximation for x and $-x$ we get

$$(1+x)^{1+x}(1-x)^{1-x} \simeq \exp\left(x+\frac{x^2}{2}\right)\exp\left(-x+\frac{x^2}{2}\right) = \exp(x^2).$$

Now if we let $x = t/\sqrt{n}$ we get an approximation for the term inside parenthesis from last part. We get

$$\left(1+\frac{t}{\sqrt{n}}\right)^{1+t/\sqrt{n}}\left(1-\frac{t}{\sqrt{n}}\right)^{1-t/\sqrt{n}} \simeq \exp\left\{\left(\frac{t}{\sqrt{n}}\right)^2\right\} = \exp\frac{t^2}{n}.$$

Therefore we have

$$\frac{\mathbb{P}[Y=t]}{d} \simeq \frac{1}{\sqrt{2\pi}}\left(\exp\frac{t^2}{n}\right)^{-n/2} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)$$

which is the formula for the probability density function of the standard normal random variable.

3 Why Is It Gaussian?

Let X be a normally distributed random variable with mean μ and variance σ^2. Let $Y = aX+b$, where a and b are non-zero real numbers. Show explicitly that Y is normally distributed with mean $a\mu+b$ and variance $a^2\sigma^2$. (Your proof should be more explicit than what's in the class notes.

One approach is to start with the cumulative distribution function of Y and use it to derive the probability density function of Y.)

Solution:

The reference for this solution is *A First Course in Probability* by Sheldon Ross, 8th edition.

Let $a > 0$.

We start with the cumulative distribution function (CDF) of Y, F_Y.

$$\begin{aligned}
F_Y(x) &= \mathbb{P}[Y \leq x] && \text{By definition of CDF} \\
&= \mathbb{P}[aX + b \leq x] && \text{Plug in } Y = aX + b \\
&= \mathbb{P}\left[X \leq \frac{x-b}{a}\right] && \text{Because } a > 0 \\
&= F_X\left(\frac{x-b}{a}\right) && \text{By definition of CDF. } F_X \text{ denotes the CDF of } X.
\end{aligned} \tag{1}$$

Let f_Y denote the probability density function (PDF) of Y.

$$\begin{aligned}
f_Y(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_Y(x) && \text{The PDF is the derivative of the CDF.} \\
&= \frac{\mathrm{d}}{\mathrm{d}x} F_X\left(\frac{x-b}{a}\right) && \text{Plug in the result from (1)} \\
&= \frac{1}{a} \cdot f_X\left(\frac{x-b}{a}\right) && \begin{array}{l}\text{PDF is the derivative of CDF.} \\ \text{Apply chain rule, } \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x-b}{a}\right) = \frac{1}{a}.\end{array} \\
&= \frac{1}{a} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-((x-b)/a-\mu)^2/(2\sigma^2)} && X \sim \mathcal{N}(\mu, \sigma^2). \\
&= \frac{1}{a\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-(x-b-a\mu)^2/(2\sigma^2 a^2)} && \frac{x-b}{a} - \mu = \frac{1}{a}(x - b - a\mu)
\end{aligned} \tag{2}$$

We have shown that f_Y equals the probability density function of a normal random variable with mean $b + a\mu$ and variance $\sigma^2 a^2$. So, Y is normally distributed with mean $b + a\mu$ and variance $\sigma^2 a^2$. The proof is done for $a > 0$. The proof for $a < 0$ is similar.

4 Deriving Chebyshev's Inequality

Recall Markov's Inequality, which applies for non-negative X and $\alpha > 0$:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$$

Use an appropriate substitution for X and α to derive Chebyshev's Inequality, where μ denotes the expected value of Y.

$$\mathbb{P}[|Y - \mu| \geq k] \leq \frac{\operatorname{var}(Y)}{k^2}$$

Solution:

Let $X = (Y-\mu)^2$. Note that this satisfies the criterion that X is non-negative. Let $\alpha = k^2$ for $k > 0$. Again, this satisfies the criterion that $\alpha > 0$. Note also that the event $|Y-\mu| \geq k$ is equivalent to the event $(Y-\mu)^2 \geq k^2$. Then

$$\mathbb{P}(|Y-\mu| \geq k) = \mathbb{P}\big((Y-\mu)^2 \geq k^2\big) = \mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha} = \frac{\mathbb{E}((Y-\mu)^2)}{k^2} = \frac{\mathrm{var}(Y)}{k^2}.$$

This is equivalent to Chebyshev's Inequality.

5 Markov's Inequality and Chebyshev's Inequality

A random variable X has variance $\mathrm{var}(X) = 9$ and expectation $\mathbb{E}[X] = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

(a) $\mathbb{E}\left[X^2\right] = 13$.

(b) $\mathbb{P}[X = 2] > 0$.

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

(d) $\mathbb{P}[X \leq 1] \leq 8/9$.

(e) $\mathbb{P}[X \geq 6] \leq 9/16$.

(f) $\mathbb{P}[X \geq 6] \leq 9/32$.

Solution:

(a) TRUE. Since $9 = \mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2^2$, we have $\mathbb{E}[X^2] = 9 + 4 = 13$.

(b) FALSE. Construct a random variable X that satisfies the conditions in the question but does not take on the value 2. A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 1/2$, and $a \neq b$. The expectation must be 2, so we have $a/2 + b/2 = 2$. The variance is 9, so $\mathbb{E}[X^2] = 13$ (from Part (a)) and $a^2/2 + b^2/2 = 13$. Solving for a and b, we get $\mathbb{P}[X = -1] = \mathbb{P}[X = 5] = 1/2$ as a counterexample.

(c) FALSE. Construct a random variable X that satisfies the conditions in the question but does not have an equal chance of being less than or greater than 2. A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = p, \mathbb{P}[X = b] = 1 - p$. Here, we use the same approach as part (b) except with a generic p, since we want $p \neq 1/2$. The expectation must be 2, so we have $pa + (1-p)b = 2$. The variance is 9, so $\mathbb{E}\left[X^2\right] = 13$ and $pa^2 + (1-p)b^2 = 13$. Solving for a and b, we find the relation $b = 2 \pm 3/\sqrt{x}$, where $x = (1-p)/p$. Then, we can find an example by plugging in values for x so that $a, b \leq 10$ and $p \neq 1/2$. One such counterexample is $\mathbb{P}[X = -7] = 1/10, \mathbb{P}[X = 3] = 9/10$.

(d) TRUE. Let $Y = 10 - X$. Since X is never exceeds 10, Y is a non-negative random variable. By Markov's inequality,

$$\mathbb{P}[10 - X \geq a] = \mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a} = \frac{\mathbb{E}[10 - X]}{a} = \frac{8}{a}.$$

Setting $a = 9$, we get $\mathbb{P}[X \leq 1] = \mathbb{P}[10 - X \geq 9] \leq 8/9$.

(e) TRUE. Chebyshev's inequality says $\mathbb{P}[|X - \mathbb{E}[X]| \geq a] \leq \text{var}(X)/a^2$. If we set $a = 4$, we have

$$\mathbb{P}[|X - 2| \geq 4] \leq \frac{9}{16}.$$

Now we observe that $\mathbb{P}[X \geq 6] \leq \mathbb{P}[|X - 2| \geq 4]$, because the event $X \geq 6$ is a subset of the event $|X - 2| \geq 4$.

(f) FALSE. We use the same approach as in Part (c) and find a counterexample. One example is $\mathbb{P}[X = 0] = 9/13, \mathbb{P}[X = 13/2] = 4/13$.

6 Practical Confidence Intervals

(a) It's New Year's Eve, and you're re-evaluating your finances for the next year. Based on previous spending patterns, you know that you spend \$1500 per month on average, with a standard deviation of \$500, and each month's expenditure is independently and identically distributed. As a poor college student, you also don't have any income. How much should you have in your bank account if you don't want to go broke this year, with probability at least 95%?

(b) As a UC Berkeley CS student, you're always thinking about ways to become the next billionaire in Silicon Valley. After hours of brainstorming, you've finally cut your list of ideas down to 10, all of which you want to implement at the same time. A venture capitalist has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability p, generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability $1 - p$ (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on p that you need to satisfy to secure the venture capitalist's funding?

(c) One of your start-ups uses error-correcting codes, which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99%?

Solution:

(a) Let T be the random variable representing the amount of money we spend in the year.

We have $T = \sum_{i=1}^{12} X_i$, where X_i represents the spending in the i-th month. So, $\mathbb{E}[T] = 12 \cdot \mathbb{E}[E_1] = 18000$.

And, since the X_is are independent, $\text{var}(T) = 12 \cdot \text{var}(X_1) = 12 \cdot 500^2 = 3,000,000$.

We want to have enough money in our bank account so that we don't finish the year in debt with 95% confidence. So, we want to keep some money ε more than the mean expenditure such that the probability of deviating above the mean by more than ε is less than 0.05.

Let's use Chebyshev's inequality here to express this.

$$\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon) \leq \frac{\text{var}(T)}{\varepsilon^2} \leq 0.05$$

This gives us $\varepsilon^2 \geq \dfrac{3,000,000}{0.05}$. So, $\varepsilon \geq 7746$. This means that we want to have a balance of $\geq \mathbb{E}[T] + \varepsilon = 25746$.

Observe that here, while we wanted to estimate $\mathbb{P}(T - \mathbb{E}[T] \geq \varepsilon)$, Chebyshev's inequality only gives us information about $\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon)$. But since

$$\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon) \geq \mathbb{P}(T - \mathbb{E}[T] \geq \varepsilon),$$

this is fine. We just get a more conservative estimate.

(b) For this question, to keep the numbers from exploding, let's work in thousands of dollars. Let X_i be the profit made from idea i, and T be the total profit made. We have $T = \sum_{i=1}^{10} X_i$.

Here, $\mathbb{E}[X_1] = 100p - 50(1-p) = 150p - 50$.

And $\text{var}(X_1) = 150^2 p(1-p)$ as the distribution of X_1 is a shifted and scaled Bernoulli distribution. Using $\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$ yields the same answer.

We have, $\mathbb{E}[T] = 10 \cdot \mathbb{E}[X_1]$. Similarly, $\text{var}(T) = 10 \cdot \text{var}(X_1)$.

Now, we want to bound the probability of T going below 0 by 0.05. In other words, we want $\mathbb{P}(T < 0) \leq 0.05$.

But, in order to apply Chebyshev's inequality, we need to look at deviation from the mean. We use the assumption that to get our funding we obviously need $\mathbb{E}[T] > 0$. Then:

$$\mathbb{P}(T < 0) \leq \mathbb{P}(T \leq 0 \cup T \geq 2\,\mathbb{E}[T]) = \mathbb{P}(|T - \mathbb{E}[T]| \geq \mathbb{E}[T]) \leq \frac{\text{var}(T)}{\mathbb{E}[T]^2} \leq 0.05$$

Looking at just the last inequality, we have:

$$\frac{\text{var}(T)}{\mathbb{E}[T]^2} = \frac{10 \cdot \text{var}(X_1)}{100 \cdot \mathbb{E}[X_1]^2} = \frac{\text{var}(X_1)}{10 \cdot \mathbb{E}[X_1]^2} \leq 0.05$$

$$\therefore \frac{\text{var}(X_1)}{\mathbb{E}[X_1]^2} \leq 0.5$$

Now, substituting what we have for variance and expectation, we get the following:

$$-22500p^2 + 22500p \leq 0.5(150p - 50)^2$$

which gives us the quadratic:

$$33750p^2 - 30000p + 1250 \geq 0$$

The solutions for p are $p \geq \frac{1}{9}(4+\sqrt{13})$ and $p \leq \frac{1}{9}(4-\sqrt{13})$. So $p \geq 0.845$ or ≤ 0.0438.

The relevant solution here is to pick $p \geq 0.845$, since the other solution yields negative expectation (contradicting the earlier assumption of positive expectation).

(c) We want $k = 1000$ packets to get across without being erased. Say we send n packets. Let X_i be the indicator random variable representing whether the ith packet got across or not.

Let the total number of unerased packets sent across be T. We have $T = \sum_{i=1}^{n} X_i$ and we want $T \geq 1000$.

We want $\mathbb{P}(T < 1000) \leq 0.01$. Now, let's try to get this in a form so that we can use Chebyshev's inequality. We know that $\mathbb{E}[T] > 1000$, so we can say that

$$\begin{aligned}\mathbb{P}(T < 1000) &\leq \mathbb{P}\big(T \leq 1000 \cup T \geq \mathbb{E}[T] + (\mathbb{E}[T] - 1000)\big) \\ &= \mathbb{P}\big(|T - \mathbb{E}[T]| \geq (\mathbb{E}[T] - 1000)\big) \leq \frac{\text{var}(T)}{(\mathbb{E}[T] - 1000)^2} \leq 0.01.\end{aligned}$$

What is $\mathbb{E}[T]$? $\mathbb{E}[T] = n\mathbb{E}[X_1] = n(1-p) = 0.2n$.

Next, what is $\text{var}(T)$? $\text{var}(T) = n\,\text{var}(X_1) = np(1-p) = 0.16n$.

Now, $\frac{\text{var}(T)}{(\mathbb{E}[T]-k)^2} \leq 0.01 \implies 16n \leq (0.2n - 1000)^2$. This gives us the quadratic:

$$0.04n^2 - 416n + 1000000 \geq 0$$

Solving the last quadratic, we get $n \geq 6629$ or $n \leq 3774$. Since the second inequality doesn't make sense for our situation, our answer is $n \geq 6629$.

7 Quadratic Regression

In this question, we will find the best quadratic estimator of Y given X. First, some notation: let μ_i be the ith moment of X, i.e. $\mu_i = \mathbb{E}[X^i]$. Also, define $\beta_1 = \mathbb{E}[XY]$ and $\beta_2 = \mathbb{E}[X^2Y]$. For simplicity, we will assume that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$. (Note that this poses no loss of generality, because we can always transform the random variables by subtracting their means and dividing by their standard deviations.) We claim that the best quadratic estimator of Y given X is

$$\hat{Y} = \frac{1}{\mu_3^2 - \mu_4 + 1}(aX^2 + bX + c)$$

where

$$\begin{aligned} a &= \mu_3\beta_1 - \beta_2, \\ b &= (1-\mu_4)\beta_1 + \mu_3\beta_2, \\ c &= -\mu_3\beta_1 + \beta_2. \end{aligned}$$

Your task is to prove the Projection Property for $\hat{Y}$.

(a) Prove that $\mathbb{E}[Y - \hat{Y}] = 0$.

(b) Prove that $\mathbb{E}[(Y - \hat{Y})X] = 0$.

(c) Prove that $\mathbb{E}[(Y - \hat{Y})X^2] = 0$.

Any quadratic function of X is a linear combination of 1, X, and X^2. Hence, these equations together imply that $Y - \hat{Y}$ is orthogonal to any quadratic function of X, and so $\hat{Y}$ is the best quadratic estimator of Y.

Solution:

(a) By linearity of expectation:

$$\mathbb{E}[Y - \hat{Y}] = \mathbb{E}[Y] - \frac{a\,\mathbb{E}[X^2] + b\,\mathbb{E}[X] + c}{\mu_3^2 - \mu_4 + 1} = \frac{-a-c}{\mu_3^2 - \mu_4 + 1} = 0$$

since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\mathbb{E}[X^2] = 1$.

(b)

$$\begin{aligned} \mathbb{E}[(Y - \hat{Y})X] &= \mathbb{E}[XY] - \frac{a\,\mathbb{E}[X^3] + b\,\mathbb{E}[X^2] + c\,\mathbb{E}[X]}{\mu_3^2 - \mu_4 + 1} \\ &= \beta_1 - \frac{(\mu_3\beta_1 - \beta_2)\mu_3 + ((1-\mu_4)\beta_1 + \mu_3\beta_2)}{\mu_3^2 - \mu_4 + 1} \end{aligned}$$

which, after a little algebra, gives 0.

(c)

$$\begin{aligned} \mathbb{E}[(Y - \hat{Y})X^2] &= \mathbb{E}[X^2Y] - \frac{a\,\mathbb{E}[X^4] + b\,\mathbb{E}[X^3] + c\,\mathbb{E}[X^2]}{\mu_3^2 - \mu_4 + 1} \\ &= \beta_2 - \frac{\mu_4(\mu_3\beta_1 - \beta_2) + \mu_3((1-\mu_4)\beta_1 + \mu_3\beta_2) - \mu_3\beta_1 + \beta_2}{\mu_3^2 - \mu_4 + 1} \end{aligned}$$

which, after a little algebra, gives 0.

8 LLSE and Graphs

Consider a graph with n vertices numbered 1 through n, where n is a positive integer ≥ 2. For each pair of distinct vertices, we add an undirected edge between them independently with probability p. Let D_1 be the random variable representing the degree of vertex 1, and let D_2 be the random variable representing the degree of vertex 2.

(a) Compute $\mathbb{E}[D_1]$ and $\mathbb{E}[D_2]$.

(b) Compute $\text{var}(D_1)$.

(c) Compute $\text{cov}(D_1, D_2)$.

(d) Using the information from the first three parts, what is $L(D_2 \mid D_1)$?

Solution:

Throughout this problem, let $X_{i,j}$ be an indicator random variable for whether the edge between vertex i and vertex j exists, for $i, j = 1, \ldots, n$. Note that $X_{i,j} = X_{j,i}$.

(a) Observing that $D_1, D_2 \sim \text{Binomial}(n-1, p)$, we obtain $\mathbb{E}[D_1] = \mathbb{E}[D_2] = (n-1)p$.

Anyway, it is good to review how we derived the expectation of the binomial distribution in the first place. By linearity of expectation,

$$\mathbb{E}[D_1] = \mathbb{E}\left[\sum_{i=2}^{n} X_{1,j}\right] = \sum_{i=2}^{n} \mathbb{E}[X_{1,j}] = (n-1)\,\mathbb{E}[X_{i,j}] = (n-1)p.$$

By symmetry, $\mathbb{E}[D_2] = (n-1)p$ also.

(b) Since $D_1, D_2 \sim \text{Binomial}(n-1, p)$, then $\text{var}\, D_1 = \text{var}\, D_2 = (n-1)p(1-p)$.

Again, it is good to review how we calculated the variance of the binomial distribution.

Solution 1: Write the variance of D_1 as a sum of covariances.

$$\begin{aligned}\text{var}(D_1) &= \text{cov}\left(\sum_{i=2}^{n} X_{1,i}, \sum_{i=2}^{n} X_{1,i}\right) = (n-1)\,\text{var}(X_{1,2}) + \big((n-1)^2 - (n-1)\big)\,\text{cov}(X_{1,2}, X_{1,3}) \\ &= (n-1)p(1-p) + 0 = (n-1)p(1-p).\end{aligned}$$

We used the fact that $X_{1,i}$ and $X_{1,j}$ are independent if $i \neq j$, so their covariance is zero.

Solution 2: Compute the variance directly.

$$\begin{aligned}\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \mathbb{E}\left[\left(\sum_{i=2}^{n} X_{1,i}\right)^2\right] - (n-1)^2 p^2 \\ &= (n-1)\,\mathbb{E}[X_{1,2}^2] + \big((n-1)^2 - (n-1)\big)\,\mathbb{E}[X_{1,2}X_{1,3}] - (n-1)^2 p^2 \\ &= (n-1)p + (n^2 - 3n + 2)p^2 - (n-1)^2 p^2 \\ &= (n-1)p + (n-1)(n-2)p^2 - (n-1)^2 p^2 = (n-1)p\big(1 + (n-2)p - (n-1)p\big) \\ &= (n-1)p(1-p)\end{aligned}$$

(c) We can write

$$\operatorname{cov}(D_1, D_2) = \operatorname{cov}\Big(\sum_{i=2}^{n} X_{1,i}, \sum_{i=1, i\neq 2}^{n} X_{2,i}\Big) = \sum_{i=2}^{n} \sum_{j=1, j\neq 2}^{n} \operatorname{cov}(X_{1,i}, X_{2,j}).$$

Note that all pairs of $X_{1,i}, X_{2,j}$ are independent except for when $i = 2$ and $j = 1$, so all terms in the sum are zero except for $\operatorname{cov}(X_{1,2}, X_{2,1})$, and our covariance is just equal to $\operatorname{cov}(X_{1,2}, X_{2,1}) = \operatorname{var}(X_{1,2}) = p(1-p)$.

(d) Since

$$L(D_2 \mid D_1) = \mathbb{E}[D_2] + \frac{\operatorname{cov}(D_1, D_2)}{\operatorname{var}(D_1)}(D_1 - \mathbb{E}[D_1]),$$

we plug in our values from the first three parts to get that

$$\begin{aligned} L(D_2 \mid D_1) &= (n-1)p + \frac{p(1-p)}{(n-1)p(1-p)}\big(D_1 - (n-1)p\big) \\ &= (n-1)p + \frac{1}{n-1}\big(D_1 - (n-1)p\big) = \frac{1}{n-1}D_1 + (n-2)p. \end{aligned}$$