CS 70 Discrete Mathematics and Probability Theory
Fall 2021 HW 15

1 Final Exam Format

Please fill out this form to choose how you will take the final exam.

Solution: Please fill out the form here.

2 Condition on an Event

The random variable X has the PDF

$$f_X(x) = \begin{cases} cx^{-2}, & \text{if } 1 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine the value of c.

(b) Let A be the event $\{X > 1.5\}$. Calculate $\mathbb{P}(A)$ and the conditional PDF of X given that A has occurred.

Solution:

(a) Integrate:

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = c\int_1^2 x^{-2}\,\mathrm{d}x = -cx^{-1}\Big|_{x=1}^2 = -c\left(\frac{1}{2} - 1\right) = \frac{c}{2} = 1$$

so $c = 2$.

(b) To find $\mathbb{P}(A)$,

$$\mathbb{P}(A) = \int_{1.5}^2 f_X(x)\,\mathrm{d}x = 2\int_{1.5}^2 x^{-2}\,\mathrm{d}x = -2x^{-1}\Big|_{x=1.5}^2 = -2\left(\frac{1}{2} - \frac{2}{3}\right) = \frac{1}{3}.$$

The conditional PDF is thus

$$f_{X|A}(x) = \frac{f_X(x)}{\mathbb{P}(A)} = 6x^{-2}, \qquad x \in [1.5, 2].$$

3 Exponential LLSE

Let $X \sim U[0,a]$ and let $Y = e^X$. Compute $L[Y \mid X]$. What does $L[Y \mid X]$ approach as $a \to 0$?

Solution:

Compute all of the necessary terms.

$$\begin{aligned}
\mathbb{E}[X] &= \frac{a}{2} \\
\text{var}(X) &= \frac{a^2}{12} \\
\mathbb{E}[Y] &= \int_0^a e^x \cdot \frac{1}{a}\,\mathrm{d}x = \frac{1}{a}(e^a - 1) \\
\mathbb{E}[XY] &= \int_0^a xe^x \cdot \frac{1}{a}\,\mathrm{d}x = e^a - \frac{1}{a}(e^a - 1) \\
\text{cov}(X,Y) &= e^a - (e^a - 1)\left(\frac{1}{2} + \frac{1}{a}\right) = \frac{2ae^a - (a+2)(e^a - 1)}{2a}
\end{aligned}$$

Therefore, one has

$$L[Y \mid X] = \frac{1}{a}(e^a - 1) + 6\frac{2ae^a - (a+2)(e^a - 1)}{a^3}\left(X - \frac{a}{2}\right).$$

As $a \to 0$, the LLSE should approach the tangent line to the curve e^x at the point $x = 0$. Hence, as $a \to 0$, $L[Y \mid X] \to 1 + X$. (This can be verified by explicitly computing the limits, but this is tedious.

4 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,

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

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

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

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.

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

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.

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

(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

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

5 Coins of LLSE

There are 3 coins in a bag, with biases $1/3$, $1/2$, $2/3$ (bias means the chance the coin will be heads). After picking a coin, you flip the coin 4 times. Let X_i be the indicator variable that the ith flip is heads. Let $X = \sum_{1\le i\le 2} X_i$ and $Y = \sum_{3\le i\le 4} X_i$. Find $L(Y \mid X)$. Recall that

$$L(Y \mid X) = \mathbb{E}[Y] + \frac{\operatorname{cov}(Y,X)}{\operatorname{Var}(X)}(X - \mathbb{E}[X]).$$

Solution:

$$\begin{aligned}
\mathbb{E}[X] &= 2\cdot\mathbb{E}[X_i] \\
&= 2\cdot\mathbb{P}[X_i = 1] \\
&= 2\cdot\left(\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{2}{3}\right) \\
&= 2\cdot\left(\frac{2+3+4}{18}\right) = 1
\end{aligned}$$

$$\mathbb{E}[Y] = \mathbb{E}[X] = 1$$

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \operatorname{cov}\left(\sum_{1\le i\le 2} X_i, \sum_{3\le i\le 4} X_i\right) \\
&= 4\cdot\operatorname{cov}(X_1,X_3) \\
&= 4\cdot(\mathbb{E}[X_1X_3] - \mathbb{E}[X_1]\,\mathbb{E}[X_3])
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X_1X_3] - \mathbb{E}[X_1]\,\mathbb{E}[X_3] &= \mathbb{P}(X_1 = 1, X_3 = 1) - \mathbb{P}(X_1 = 1)^2 \\
&= \left[\frac{1}{3}\cdot\left(\frac{1}{3}\right)^2+\frac{1}{3}\cdot\left(\frac{1}{2}\right)^2+\frac{1}{3}\cdot\left(\frac{2}{3}\right)^2\right] - \left[\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{2}{3}\right]^2 \\
&= \frac{1}{54}
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}(X) &= \operatorname{cov}\left(\sum_{1\le i\le 2} X_i, \sum_{1\le i\le 2} X_i\right) \\
&= 2\cdot\operatorname{Var}(X_1) + 2\cdot\operatorname{cov}(X_1,X_2) \\
&= 2\cdot(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) + 2\cdot\frac{1}{54} \\
&= 2\cdot\left(\frac{1}{2}-\frac{1}{4}\right) + 2\cdot\frac{1}{54} = \frac{29}{54}
\end{aligned}$$

$$
\begin{aligned}
L(Y \mid X) &= 1 + \frac{4/54}{29/54} \cdot (X-1) \\
&= 1 + \frac{4}{29} \cdot (X-1) \\
&= \frac{4}{29} X + \frac{25}{29}
\end{aligned}
$$

6 Bernoulli CLT

In this question we will explicitly see why the central limit theorem holds for the Bernoulli distribution as we add up more and more coin tosses.

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) Observe that $X \sim \text{Binomial}(n, 1/2)$ and $\mathbb{P}[X = k] = \binom{n}{k}/2^n$. 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 $2/\sqrt{n}$.

(c) 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?

(d) 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 b to find an approximation for $\mathbb{P}[Y = t]/d$. Show that the end result is equivalent to:

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

(e) 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.

7 That's an Odd Program

(a) Is there a program `Even1` which takes as input a program `P` and returns whether or not `P` contains an even number of lines of code? Either explain how the program would work or show that the problem is uncomputable.

(b) Is there a program `Even2` which takes as inputs a program `P` and an input `x`, and returns whether or not `P(x)` runs an even number of distinct lines of code? Either explain how the program would work or show that the problem is uncomputable.

Solution:

(a) Yes. We can run through the source code of `P` and count the number of lines, and then output whether or not that count is even.

(b) No. The idea here is to exploit the fact that self-reference can induce contradictions. In particular, we proceed in a manner similar to the proof that `TestHalt` is uncomputable - we assume that `Even2` is computable, and build a program using it that is designed to break under self-reference. Consider the following program:

```
define TuringE(P):
    if Even2(P, P):
        print("Filler");
    return;
```

Note that we choose not to consider the define statement as a line of code, but the logic is the same even if we did. The critical part of this program is the following: if `P(P)` runs an even number of lines of code, then `TuringE(P)` runs an odd number of lines of code, and if `P(P)` runs an odd number of lines of code, then `TuringE(P)` runs an even number of lines of code. In other words, `TuringE(P)` behaves in the exact opposite way that `P(P)` behaves. With this observation, we can ask what the behavior of `TuringE(TuringE)` is. By definition, if `TuringE(TuringE)` runs an even number of lines of code, then `TuringE(TuringE)` runs an odd number of lines of code. If `TuringE(TuringE)` runs an odd number of lines of code, then `TuringE(TuringE)` runs an even number of lines of code. In either case, this is a contradiction, as there is no number that is both even and odd. Thus, we can conclude that no implementation for `Even2` can exist.