CS 70 Discrete Mathematics and Probability Theory
Fall 2021 HW 15

Due: N/A
Grace period until N/A

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 Final Exam Format

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

2 Condition on an Event

The random variable X has the PDF

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

(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.

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 $a \to 0$?

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 $\mathrm{Var}(D_1)$.

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

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

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

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

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 $\sigma^2 = \mathrm{Var} X = n/4$.

(b) Observe that $X \sim \mathrm{Binomial}(n, 1/2)$ 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.)

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.