CS 70   Discrete Mathematics and Probability Theory
Summer 2018   Sinho Chewi and Vrettos Moulos   Homework 7

Due: Friday, August 3, 2018 at 10 PM

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.

1   Darts with Friends

Michelle and Alex are playing darts. Being the better player, Michelle's aim follows a uniform distribution over a circle of radius r around the center. Alex's aim follows a uniform distribution over a circle of radius $2r$ around the center.

(a) Let the distance of Michelle's throw be denoted by the random variable X and let the distance of Alex's throw be denoted by the random variable Y.

  - What's the cumulative distribution function of X?
  - What's the cumulative distribution function of Y?
  - What's the probability density function of X?
  - What's the probability density function of Y?

(b) What's the probability that Michelle's throw is closer to the center than Alex's throw? What's the probability that Alex's throw is closer to the center?

(c) What's the cumulative distribution function of $U = \min\{X, Y\}$?

(d) What's the cumulative distribution function of $V = \max\{X, Y\}$?

(e) What is the expectation of the absolute difference between Michelle's and Alex's distances from the center, that is, what is $\mathbb{E}[|X - Y|]$? [*Hint*: Use parts (c) and (d), together with the continuous version of the tail sum formula.]

2 Lunch Meeting

Alice and Bob agree to try to meet for lunch between 12 PM and 1 PM at their favorite sushi restaurant. Being extremely busy, they are unable to specify their arrival times exactly, and can say only that each of them will arrive (independently) at a time that is uniformly distributed within the hour. In order to avoid wasting precious time, if the other person is not there when they arrive they agree to wait exactly fifteen minutes before leaving. What is the probability that they will actually meet for lunch?

3 Darts (Again!)

Alvin is playing darts. His aim follows an exponential distribution; that is, the probability density that the dart is x distance from the center is $f_X(x) = \exp(-x)$. The board's radius is 4 units.

(a) What is the probability the dart will stay within the board?

(b) Say you know Alvin made it on the board. What is the probability he is within 1 unit from the center?

(c) If Alvin is within 1 unit from the center, he scores 4 points, if he is within 2 units, he scores 3, etc. In other words, Alvin scores $\lfloor 5 - x \rfloor$, where x is the distance from the center. What is Alvin's expected score after one throw?

4 Noisy Love

Due to the Central Limit Theorem, the Gaussian distribution is often used as a model for noise. In this problem, we will see how to perform calculations with Gaussian noise models.

Suppose you have confessed to your love interest on Valentine's Day and you are waiting to hear back. Your love interest is trying to send you a binary message: "0" means that your love interest is not interested in you, while "1" means that your love interest reciprocates your feelings. Let X be your love interest's message for you. Your current best guess of X has $\mathbb{P}(X = 0) = 0.7$ and $\mathbb{P}(X = 1) = 0.3$. Unfortunately, your love interest sends you the message through a noisy channel, and instead of receiving the message X, you receive the message $Y = X + \varepsilon$, where ε is independent Gaussian noise with mean 0 and variance 0.49.

(a) First, you decide upon the following rule: if you observe $Y > 0.5$, then you will assume that your love interest loves you back, whereas if you observe $Y \leq 0.5$, then you will assume that

your love interest is not interested in you. What is the probability that you are correct using this rule? (Express your answer in terms of the CDF of the standard Gaussian distribution $\Phi(z) = \mathbb{P}(\mathcal{N}(0,1) \le z)$, and then evaluate your answer numerically.)

(b) Suppose you observe $Y = 0.6$. What is the probability that your love interest loves you back? [*Hint*: This problem requires conditioning on an event of probability 0, namely, the event $\{Y = 0.6\}$. To tackle this problem, think about conditioning on the event $\{Y \in [0.6, 0.6 + \delta]\}$, where $\delta > 0$ is small, so that $f_Y(0.6) \cdot \delta \approx \mathbb{P}(Y \in [0.6, 0.6 + \delta])$, and then apply Bayes Rule.]

(c) Suppose you observe $Y = y$. For what values is it more likely than not that your love interest loves you back? [*Hint*: As before, instead of considering $\{Y = y\}$, you can consider the event $\{Y \in [y, y + \delta]\}$ for small $\delta > 0$. So, when is $\mathbb{P}(X = 1 \mid Y \in [y, y + \delta]) \ge \mathbb{P}(X = 0 \mid Y \in [y, y + \delta])$?]

(d) Your new rule is to assume that your love interest loves you back if (based on the value of Y that you observe) it is more likely than not that your love interest loves you back. Under this new rule, what is the probability that you are correct?

5   Confidence Interval Introduction

We observe a random variable X which has mean μ and standard deviation $\sigma \in (0, \infty)$. Assume that the mean μ is unknown, but σ is known.

We would like to give a 95% confidence interval for the unknown mean μ. In other words, we want to give a random interval (a, b) (it is random because it depends on the random observation X) such that the probability that μ lies in (a, b) is at least 95%.

We will use a confidence interval of the form $(X - \varepsilon, X + \varepsilon)$, where $\varepsilon > 0$ is the width of the confidence interval. When ε is smaller, it means that the confidence interval is narrower, i.e., we are giving a more *precise* estimate of μ.

(a) Using Chebyshev's Inequality, calculate $\mathbb{P}\{|X - \mu| \ge \varepsilon\}$.

(b) Explain why $\mathbb{P}\{|X - \mu| < \varepsilon\}$ is the same as $\mathbb{P}\{\mu \in (X - \varepsilon, X + \varepsilon)\}$.

(c) Using the previous two parts, choose the width of the confidence interval ε to be large enough so that $\mathbb{P}\{\mu \in (X - \varepsilon, X + \varepsilon)\}$ is guaranteed to exceed 95%.

[Note: Your confidence interval is allowed to depend on X, which is observed, and σ, which is known. Your confidence interval is not allowed to depend on μ, which is unknown.]

(d) The previous three parts dealt with the case when you observe one sample X. Now, let n be a positive integer and let $X_1, \ldots, X_n$ be i.i.d. samples, each with mean μ and standard deviation $\sigma \in (0, \infty)$. As before, assume that μ is unknown but σ is known.

Here, a good estimator for μ is the *sample mean* $\bar{X} := n^{-1} \sum_{i=1}^{n} X_i$. Calculate the mean and variance of $\bar{X}$.

(e) We will now use a confidence interval of the form $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$ where $\varepsilon > 0$ again represents the width of the confidence interval. Imitate the steps of (a) through (c) to choose the width ε to be large enough so that $\mathbb{P}\{\mu \in (\bar{X} - \varepsilon, \bar{X} + \varepsilon)\}$ is guaranteed to exceed 95%.

To check your answer, your confidence interval should be *smaller* when n is larger. Intuitively, if you collect more samples, then you should be able to give a more *precise* estimate of μ.

(f) It's New Year's Eve, and you are re-evaluating your finances for the next year. You know that each month's expenditure is independently and identically distributed with a standard deviation of $500. Over the past twelve months, you spent an average of $1500 per month. How much money must your job make per month so that your income exceeds your mean monthly expenditures with probability at least 95%?

6 Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \frac{1}{n} S_n - \mathbb{E}\left[\frac{1}{n} S_n \right] \right| > \varepsilon \right) = 0.$$

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of p and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not, and give a brief justification of your answer. (Whenever convenient, you can assume that n is even.)

(a) **Yes** or **No**: Each packet is sent on a completely different route.

(b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) **Yes** or **No**: All the packets are sent on one route.

7 [Optional] 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 $\sigma^2 = \text{var} X = n/4$.

(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 $\sqrt{\dfrac{n}{k(n-k)}}$ to be close to $\sqrt{\dfrac{n}{(n/2) \times (n/2)}} = \dfrac{2}{\sqrt{n}}$. 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 c to find an approximation for $\mathbb{P}[Y = t]/d$. Show that the end result is equivalent to:

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

(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.)