Due: Saturday 4/22, 4:00 PM
Grace period until Saturday 4/22, 6:00 PM

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 Just One Tail, Please

Note 17

Let X be some random variable with finite mean and variance which is not necessarily non-negative. The *extended* version of Markov's Inequality states that for a non-negative function $\varphi(x)$ which is monotonically increasing for $x > 0$ and some constant $\alpha > 0$,

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

Suppose $\mathbb{E}[X] = 0$, $\text{Var}(X) = \sigma^2 < \infty$, and $\alpha > 0$.

(a) Use the extended version of Markov's Inequality stated above with $\varphi(x) = (x+c)^2$, where c is some positive constant, to show that:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2 + c^2}{(\alpha + c)^2}$$

(b) Note that the above bound applies for all positive c, so we can choose a value of c to minimize the expression, yielding the best possible bound. Find the value for c which will minimize the RHS expression (you may assume that the expression has a unique minimum).

We can plug in the minimizing value of c you found in part (b) to prove the following bound:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$

This bound is also known as Cantelli's inequality.

(c) Recall that Chebyshev's inequality provides a two-sided bound. That is, it provides a bound on $\mathbb{P}[|X-\mathbb{E}[X]| \geq \alpha] = \mathbb{P}[X \geq \mathbb{E}[X]+\alpha] + \mathbb{P}[X \leq \mathbb{E}[X]-\alpha]$. If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound $\mathbb{P}[X \geq \mathbb{E}[X]+\alpha]$, it is tempting to just divide the bound we get from Chebyshev's by two.

(i) Why is this not always correct in general?

(ii) Provide an example of a random variable X (does not have to be zero-mean) and a constant α such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}[X \geq \mathbb{E}[X]+\alpha] > \frac{\mathrm{Var}(X)}{2\alpha^2}$ or $\mathbb{P}[X \leq \mathbb{E}[X]-\alpha] > \frac{\mathrm{Var}(X)}{2\alpha^2}$.

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov's and Chebyshev's inequality!

(d) Let's try out our new bound on a simple example. Suppose X is a positively-valued random variable with $\mathbb{E}[X] = 3$ and $\mathrm{Var}(X) = 2$.

(i) What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$?

(ii) What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$?

(iii) What bound would Cantelli's Inequality give for $\mathbb{P}[X \geq 5]$? (*Note*: Recall that Cantelli's Inequality only applies for zero-mean random variables.)

2 Coupon Collector Variance

Note 17
Note 19

It's that time of the year again—Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of n different Monopoly Cards with equal probability. You need to collect them all to redeem the grand prize.

Let X be the number of visits you have to make before you can redeem the grand prize. Remember that we've previously shown $\mathrm{Var}(X) = n^2 \left(\sum_{i=1}^{n} i^{-2}\right) - \mathbb{E}[X]$.

The series $\sum_{i=1}^{\infty} i^{-2}$ converges to the constant value $\pi^2/6$. Using this fact and Chebyshev's Inequality, find a lower bound on β for which the probability you need to make more than $\mathbb{E}[X]+\beta n$ visits is less than $1/100$, for large n. *[Hint: Use the approximation $\sum_{i=1}^{n} i^{-1} \approx \ln n$ as n grows large.]*

3 Estimating π

Note 17

In this problem, we discuss some interesting ways that you could probabilistically estimate π, and see how good these techniques are at estimating π.

Technique 1: Buffon's needle is a method that can be used to estimate the value of π. There is a table with infinitely many parallel lines spaced a distance 1 apart, and a needle of length 1. It turns out that if the needle is dropped uniformly at random onto the table, the probability of the needle intersecting a line is $\frac{2}{\pi}$. We have seen a proof of this in the notes.

Technique 2: Consider a square dartboard, and a circular target drawn inscribed in the square dartboard. A dart is thrown uniformly at random in the square. The probability the dart lies in the circle is $\frac{\pi}{4}$.

Technique 3: Pick two integers x and y independently and uniformly at random from 1 to M, inclusive. Let p_M be the probability that x and y are relatively prime. Then

$$\lim_{M\to\infty} p_M = \frac{6}{\pi^2}.$$

Let $p_1 = \frac{2}{\pi}, p_2 = \frac{\pi}{4}$, and $p_3 = \frac{6}{\pi^2}$ be the probabilities of the desired events of **Technique 1**, **Technique 2**, and **Technique 3**, respectively. For each technique, we apply each technique N times, then compute the proportion of the times each technique occurred, getting estimates $\hat{p}_1$, $\hat{p}_2$, and $\hat{p}_3$, respectively.

(a) For each $\hat{p}_i$, compute an expression X_i in terms of $\hat{p}_i$ that would be an estimate of π.

(b) Using Chebyshev's Inequality, compute the minimum value of N such that X_2 is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ.

For X_1 and X_3, computing the minimum value of N will be more tricky, as the expressions for X_1 and X_3 are not as nice as X_2.

(c) For $i = 1$ and 3, compute a constant c_i such that

$$|X_i - \pi| < \varepsilon \iff |\hat{p}_i - p_i| < c_i\varepsilon + o(\varepsilon^2),$$

where the $o(\varepsilon^2)$ represents terms containing powers of ε that are 2 or higher (i.e. $\varepsilon^2, \varepsilon^3$, etc.).

(Hint: You may find the following Taylor series helpful: For x close to 0,

$$\frac{1}{a-x} = \frac{1}{a} + \frac{x}{a^2} + o(x^2)$$
$$\frac{1}{(a-x)^2} = \frac{1}{a^2} + \frac{2x}{a^3} + o(x^2).$$

The $o(x^2)$ represents terms that have x^2 powers or higher.)

In this problem, we assume ε is close enough to 0 such that $o(\varepsilon^2)$ is 0. In other words,

$$\mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon + o(\varepsilon^2)\right] = \mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon\right].$$

Combining with part (c) then gives

$$\mathbb{P}\left[|X_i - \pi| < \varepsilon\right] = \mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon\right].$$

(d) For $i = 1$ and 3, use Chebyshev's Inequality and the above work to compute the minimum value of N such that X_i is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ.

(e) Which technique required the lowest value for N? Which technique required the highest?

4 Short Answer

Note 21

(a) Let X be uniform on the interval $[0,2]$, and define $Y = 4X^2 + 1$. Find the PDF, CDF, expectation, and variance of Y.

(b) Let X and Y have joint distribution

$$f(x,y) = \begin{cases} cxy + \frac{1}{4} & x \in [1,2] \text{ and } y \in [0,2] \\ 0 & \text{otherwise.} \end{cases}$$

Find the constant c. Are X and Y independent?

(c) Let $X \sim \text{Exp}(3)$.

(i) Find probability that $X \in [0, 1]$.

(ii) Let $Y = \lfloor X \rfloor$. For each $k \in \mathbb{N}$, what is the probability that $Y = k$? Write the distribution of Y in terms of one of the famous distributions; provide that distribution's name and parameters.

(d) Let $X_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \ldots, n$ be mutually independent. It is a (very nice) fact that $\min(X_1, \ldots, X_n) \sim \text{Exp}(\mu)$. Find μ.

5 Useful Uniforms

Note 21

Let X be a continuous random variable such that $\mathbb{P}[X \in (a,b)] > 0$ for all $a, b \in \mathbb{R}$ and $a < b$.

(a) Give an example of a distribution that X could have, and one that it could not.

(b) Show that the CDF F of X is strictly increasing. That is, $F(x+\varepsilon) > F(x)$ for any $\varepsilon > 0$. Argue why this implies that $F : \mathbb{R} \to (0,1)$ must be invertible.

(c) Let U be a uniform random variable on $(0,1)$. What is the distribution of $F^{-1}(U)$?

(d) Your work in part (c) shows that in order to sample X, it is enough to be able to sample U. If X was a discrete random variable instead, taking finitely many values, can we still use U to sample X?

6 Waiting For the Bus

Note 21

Edward and Jerry are waiting at the bus stop outside of Soda Hall.

Like many bus systems, buses arrive in periodic intervals. However, the Berkeley bus system is unreliable, so the length of these intervals are random, and follow Exponential distributions.

Edward is waiting for the 51B, which arrives according to an Exponential distribution with parameter λ. That is, if we let the random variable X_i correspond to the difference between the arrival time ith and $(i-1)$st bus (also known as the inter-arrival time) of the 51B, $X_i \sim \text{Exp}(\lambda)$.

Jerry is waiting for the 79, whose inter-arrival times also follows Exponential distributions with parameter μ. That is, if we let Y_i denote the inter-arrival time of the 79, $Y_i \sim \text{Exp}(\mu)$. Assume that all inter-arrival times are independent.

(a) What is the probability that Jerry's bus arrives before Edward's bus?

(b) After 20 minutes, the 79 arrives, and Jerry rides the bus. However, the 51B still hasn't arrived yet. Let D be the additional amount of time Edward needs to wait for the 51B to arrive. What is the distribution of D?

(c) Lavanya isn't picky, so she will wait until either the 51B or the 79 bus arrives. Find the distribution of Z, the amount of time Lavanya will wait before catching her bus.

(d) Khalil doesn't feel like riding the bus with Edward. He decides that he will wait for the second arrival of the 51B to ride the bus. Find the distribution of $T = X_1 + X_2$, the amount of time that Khalil will wait to ride the bus.