CS 70   Discrete Mathematics and Probability Theory
Spring 2017   Rao
HW 10

1 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.

2 How Many Kings?

Suppose that you draw 3 cards from a standard deck without replacement. Let X denote the number of kings you draw.

(a) What is $\Pr(X = 0)$?

(b) What is $\Pr(X = 1)$?

(c) What is $\Pr(X = 2)$?

(d) What is $\Pr(X = 3)$?

(e) Do the answers you computed in parts (a) through (d) add up to 1, as expected?

(f) Compute $\mathbf{E}(X)$ from the definition of expectation.

(g) Suppose we define indicators X_i, $1 \le i \le 3$, where X_i is the indicator variable that equals 1 if the ith card is a king and 0 otherwise. Compute $\mathbf{E}(X)$.

(h) Are the X_i indicators independent? How does this affect your answer to part (g)?

3 Sisters

Consider a family with n children, each with a 50% chance of being male or female. Let X be the total number of sisters that the male children have, and let Y be the total number of sisters that the female children have (for example, if $n = 3$ and there are two boys and one girl, then $X = 2$ and $Y = 0$). Find expressions for $\mathbf{E}(X)$ and $\mathbf{E}(Y)$ in terms of n. Do we expect that boys have more sisters or girls have more sisters?

[Hint: Define a random variable B to denote the number of boys, find an expression for X as a function of B, and apply linearity of expectation. Use a similar approach for girls.]

4 Unbiased Variance Estimation

We have a random variable X and want to estimate its variance, σ^2 and mean, μ, by sampling from it. In this problem, we will derive an "unbiased estimator" for the variance.

(a) We define a random variable Y that corresponds to drawing n values from the distribution for X and averaging, or $Y = (X_1 + \ldots + X_n)/n$. What is $\mathbf{E}(Y)$? Note that if $\mathbf{E}(Y) = \mathbf{E}(X)$ then Y is an unbiased estimator of $\mu = \mathbf{E}(X)$. *Hint: This should not be difficult.*

(b) Now let's assume the actual mean is 0 as variance doesn't change when one shifts the mean.

   Before attempting to define an estimator for variance, show that $\mathbf{E}(Y^2) = \sigma^2/n$.

(c) In practice, we don't know the mean of X so following part (a), we estimate it as Y. With this in mind, we consider the random variable $Z = \sum_{i=1}^{n}(X_i - Y)^2$. What is $\mathbf{E}(Z)$?

(d) What is a good unbiased estimator for the $\text{Var}(X)$?

(e) How does this differ from what you might expect? Why? (Just tell us your intuition here, it is all good!)

5 Markov Bound for Coupon Collectors

Suppose you are trying to collect a set of n different baseball cards. You get the cards by buying boxes of cereal: each box contains exactly one card, and it is equally likely to be any of the n cards. You are interested in finding m, a lower bound on the number of boxes you should buy to ensure that the probability of you collecting all n cards is at least $\frac{1}{2}$. In class, we used the Union Bound to show that it suffices to have $m \geq n \ln(2n)$.

Use Markov's Inequality to find a different (weaker) lower bound on m.

6 Safeway Monopoly Cards

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.

(a) Let X be the number of visits you have to make before you can redeem the grand prize. Show that $\text{Var}(X) = n^2 \left(\sum_{i=1}^{n} i^{-2} \right) - \mathbf{E}(X)$. *[Hint: Does this remind you of a particular problem? What is the expectation for this problem?]*

(b) 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 $\mathbf{E}(X) + \beta n$ 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.]*

7 Dice

In this problem, let $X_1, X_2, \ldots X_n$ each denote the outcomes of standard six-sided dice rolls. Let A denote the average of the outcomes $\left(\sum_{i=1}^{n} X_i \right)/n$.

(a) For $n = 100$, find some a and b such that A is in the interval $[a, b]$ with probability at least 90% (Don't use trivial intervals like $[1, 6]$).

(b) For $n = 30$, find a lower bound on $\Pr[3 \le A \le 4]$.

(c) Find the minimum n for which you can guarantee that A is within the range $[3, 4]$ with at least 99% probability.

8 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 1000 packets are 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%?

9 Entropy

Let X_i, $1 \le i \le n$, be a sequence of i.i.d. Bernoulli random variables with parameter p, i.e. $\Pr(X_i = 1) = p$ and $\Pr(X_i = 0) = 1 - p$.

(a) Express $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ in terms of p and n_0, where $x_i \in \{0, 1\}$ for all $1 \le i \le n$ and n_0 depends on $(x_1 \ldots x_n)$ and represents the number of 1's in $(x_1 \ldots x_n)$. Call this result $f(x_1, x_2, \ldots, x_n)$.

(b) Define the random variable Y_n as $Y_n = -n^{-1} \log f(X_1, X_2, \ldots, X_n)$. What does Y_n tend to as n grows large?