

Problem Set 11

Due: December 4, 7pm

Reading Assignment: Chapters 18, 19

Problem 1. [15 points]

In this problem, we will (hopefully) be making tons of money! Use your knowledge of probability and statistics to keep from going broke!

Suppose the stock market contains N types of stocks, which can be modelled by independent random variables. Suppose furthermore that the behavior of these stocks is modelled by a double-or-nothing coin flip. That is, stock S_i has half probability of doubling its value and half probability of going to 0. The stocks all cost a dollar, and you have N dollars. Say you only keep these stocks for one time-step (that is, at the end of this timestep, all stocks would have doubled in value or gone to 0).

(a) [3pts] What is your expected amount of money if you spend all your money on one stock? Your variance?

(b) [3pts] Suppose instead you diversified your purchases and bought N shares of all different stocks. What is your expected amount of money then? Your variance?

(c) [3pts] The money that you have invested came from your financially conservative mother. As a result, your goals are much aligned with hers. Given this, which investment strategy should you take?

(d) [3pts] Now instead say that you make money on rolls of dice. Specifically, you play a game where you roll a standard six-sided dice, and get paid an amount (in dollars) equal to the number that comes up. What is your expected payoff? What is the variance?

(e) [3pts] We change the rules of the game so that your payoff is the cube of the number that comes up. In that case, what is your expected payoff? What is its variance?

Problem 2. [8 points] Here are seven propositions:

$$\begin{array}{llll} x_1 & \vee & x_3 & \vee & \neg x_7 \\ \neg x_5 & \vee & x_6 & \vee & x_7 \\ x_2 & \vee & \neg x_4 & \vee & x_6 \\ \neg x_4 & \vee & x_5 & \vee & \neg x_7 \\ x_3 & \vee & \neg x_5 & \vee & \neg x_8 \\ x_9 & \vee & \neg x_8 & \vee & x_2 \\ \neg x_3 & \vee & x_9 & \vee & x_4 \end{array}$$

Note that:

1. Each proposition is the OR of three terms of the form x_i or the form $\neg x_i$.
2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables x_1, \dots, x_9 independently and with equal probability.

- (a) [4 pts] What is the expected number of true propositions?
- (b) [4 pts] Use your answer to prove that there exists an assignment to the variables that makes *all* of the propositions true.

Problem 3. [10 points] We have two coins: one is a fair coin and the other is a coin that produces heads with probability $3/4$. One of the two coins is picked, and this coin is tossed n times. Explain how to calculate the number of tosses to make us 95% confident which coin was chosen. You do not have to calculate the minimum value of n , though we'd be pleased if you did.

Problem 4. [22 points]

Suppose n balls are thrown randomly into n boxes, so each ball lands in each box with uniform probability. Also, suppose the outcome of each throw is independent of all the other throws.

- (a) [5 pts] Let X_i be an indicator random variable whose value is 1 if box i is empty and 0 otherwise. Write a simple closed form expression for the probability distribution of X_i . Are X_1, X_2, \dots, X_n independent random variables?
- (b) [2 pts] Find a constant, c , such that the expected number of empty boxes is asymptotically equal (\sim) to cn .
- (c) [5 pts] Show that

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \leq \binom{n}{k} \left(\frac{1}{n}\right)^k.$$

- (d) [5 pts] Let R be the maximum of the numbers of balls that land in each of the boxes. Conclude from the previous parts that

$$\Pr\{R \geq k\} \leq \frac{n}{k!}.$$

(e) [5 pts] Conclude that

$$\lim_{n \rightarrow \infty} \Pr \{R \geq n^\epsilon\} = 0$$

for all $\epsilon > 0$.

Problem 5. [13 points] The goal of this problem will be to study the expectations of quotients of positive random variables.

Throughout, you may assume the following inequality: for all $\lambda \in (0, 1)$, and all positive reals x and y ,

$$\frac{1}{\lambda x + (1 - \lambda)y} \leq \frac{\lambda}{x} + \frac{1 - \lambda}{y}.$$

(This property has a name: the function $f(x) = \frac{1}{x}$ is *convex*.)

(a) [10 pts] Let X be a positive random variable with finitely many outcomes. Prove that

$$E \left[\frac{1}{X} \right] \geq \frac{1}{E[X]}.$$

(Hint: Try induction on the number of outcomes of X .)

(b) [3 pts] Let R, T be positive independent random variables with finitely many outcomes each. Prove that $E[\frac{R}{T}] \geq \frac{E[R]}{E[T]}$.

Problem 6. [8 points] We roll a fair die until we have rolled all 6 numbers. The rolls are independent. What is the expected number of rolls until this happens?

Problem 7. [10 points] We are given a random vector of n distinct numbers in random order. We then determine the maximum of these numbers using the following procedure:

Pick the first number. Call it the *current maximum*. Go through the rest of the vector (in order) and each time we come across a number (call it x) that exceeds our current maximum, we update the current maximum with x .

What is the expected number of times we update the current maximum?

(Hint: Let X_i be the indicator variable for the event that the i th element in the vector is larger than all the previous elements.)

Problem 8. [14 points] Suppose we are trying to estimate some physical parameter p . When we run our experiments and process the results, we obtain an estimator of p , call it p_e . But if our experiments are probabilistic, then p_e itself is a random variable. We call the random variable p_e an *unbiased* estimator if $E[p_e] = p$.

For example, say we are trying to estimate the height, h , of Green Hall. However, each of our measurements has some noise that is, say, Gaussian with zero mean. So each measurement can be viewed as a sample from a random variable X . The expected value of each

measurement is thus $E[X] = h$, since the probabilistic noise has zero mean. Then, given n independent trials, x_1, \dots, x_n , an unbiased estimator for the height of Green Hall would be

$$h_e = \frac{x_1 + \dots + x_n}{n},$$

since

$$E[h_e] = E\left[\frac{x_1 + \dots + x_n}{n}\right] = \frac{E[x_1] + \dots + E[x_n]}{n} = E[x_1] = h.$$

Now say we take n independent observations of a random variable Y . Let the true (but unknown) variance of Y be $\text{Var}[Y] = \sigma^2$. Then we can define the following estimator σ_e^2 for $\text{Var}[Y]$ using the data from our observations:

$$\sigma_e^2 = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \left(\frac{y_1 + y_2 + \dots + y_n}{n}\right)^2.$$

Is this an unbiased estimator of the variance? In other words, is $E[\sigma_e^2] = \sigma^2$? If not, can you suggest how to modify this estimator to make it unbiased?