

Homework 6

Many of these questions are taken from Grinstead and Snell or from Feller; some have been lightly edited.

1. A certain lottery has the following rules: you buy a ticket, choose 3 different numbers from 1 to 100, and write them on the ticket. The lottery has a box with 100 balls numbered 1 to 100. Three (different) balls are chosen. If any of the balls has one of the numbers you have chosen, you win. What is the probability of winning?
2. You are dealt five cards from a standard deck. What is the probability that the first four are aces and the fifth is a king?
3. A lady wishes to color her fingernails on one hand using at most two of the colors red, yellow, and blue. How many ways can she do this?
4. Each of the 50 states has two senators, so there are 100 senators in all. Suppose a committee of 50 senators is chosen at random from this group of 100.
 - (a) What is the probability that California is represented?
 - (b) What is the probability that every state is represented?
5. Let X_1, X_2, \dots, X_{100} be the outcomes of 100 independent rolls of a fair die.
 - (a) What are $\mathbb{E}(X_1)$ and $\text{var}(X_1)$?
 - (b) Define the random variable X to be $X_1 - X_2$. What are $\mathbb{E}(X)$ and $\text{var}(X)$?
 - (c) Define the random variable Y to be $X_1 - 2X_2 + X_3$. What is $\mathbb{E}(Y)$ and $\text{var}(Y)$?
 - (d) Define $Z = X_1 - X_2 + X_3 - X_4 + \dots + X_{99} - X_{100}$. What are $\mathbb{E}(Z)$ and $\text{var}(Z)$?
6. In class we have focused mostly on the case of *discrete* random variables, which have a finite number of possible values. For instance, the roll of a die has six possible values, and a coin toss has two possible values. Often, however, it is useful to consider *continuous* random variables, which can take on *infinitely* many values.

For example, suppose random variable X is distributed uniformly in $[0, 1]$. That is, X can take on any value in the range $[0, 1]$, and all values are equally likely. Some extra care is needed in specifying the distribution of X , because the probability of any specific value is zero; for instance, $\Pr(X = 0.3) = 0$.

A common way to describe the distribution of a continuous random variable is by a *cumulative distribution function*, which specifies $\Pr(X \leq a)$ for all possible values a . For instance, $\Pr(X \leq 0.3) = 0.3$. Here's the full cumulative distribution function of X : for any $0 \leq a \leq 1$, $\Pr(X \leq a) = a$.

Now, suppose you throw a dart at a dartboard of radius 1, and it lands at a random location on the board, all locations being equally likely. Let Y denote the distance of the dart from the center of the board. What is the cumulative distribution function of Y ?

7. An enormous array $S[1 \dots n]$ contains mostly gibberish words (ie. words that aren't in the dictionary). But it also contains some real words w_1, \dots, w_k ; in fact, each of these words appears m times in the array. These words together reveal an important secret, so you want to find them. But you don't want to have to scan through the entire array, which would take $O(n)$ time. So instead you repeat the following procedure until you find *all* of the k words:
- Pick a position $1 \leq i \leq n$ at random.
 - Check whether the word $S[i]$ is in the dictionary. If so, you have found one of the secret words.
- (a) What is the expected number of tries before you find your first real word?
- (b) What is, roughly, the expected number of tries before you find all k words?
8. The most popular method for testing whether a number is prime is a Monte Carlo randomized algorithm \mathcal{A} with the following behavior:
- If x is prime, $\mathcal{A}(x)$ always says "prime"
 - If x is not prime, $\mathcal{A}(x)$ says "not prime" with probability $1/2$; otherwise it says "prime"

You want to find out whether a certain x is prime or not. However, you want the probability of getting the wrong answer to be at most 1 in a billion (10^9). How would you do this?

9. In general, sorting n elements takes $O(n \log n)$ time, but we saw in class that in the special case where the distribution of elements is known to be uniform, it is possible to sort in *expected* $O(n)$ time.

What if the distribution of elements follows some distribution function F that we know? That is, the n elements to be sorted—call them X_1, \dots, X_n —are chosen independently, and

$$\Pr(X_i \leq a) = F(a) \quad \text{for any } a.$$

Show that it is still possible to sort in expected linear time. *Hint:* Once again, use n buckets, but instead of making them equal width, allow them to have different widths and equalize some other criterion...

10. In class, we talked about a randomized algorithm for finding the minimum cut of a graph $G = (V, E)$. Sometimes we are interested in finding the *maximum cut*, that is, a partition of V into two groups V_1 and V_2 such that the number of edges between V_1 and V_2 is maximized. Here's a possible algorithm for maximum cut:
- Select V_1 as follows: for each $u \in V$, include u in V_1 with probability $1/2$.
 - Let $V_2 = V - V_1$.
- (a) What is the expected size of the resulting cut, as a function of $|E|$?
- (b) This algorithm might not return the maximum cut. However, can you show that it is, in expectation, at least half as big as the maximum cut?

Answer key 1. 0.088. 2. $1/3248700$. 3. 96. 4 (a) $149/198$ (b) $2^{50}/\binom{50}{100}$. 5 (a) $7/2$, $35/12$ (b) 0, $35/6$ (c) 0, $35/2$ (d) 0, $875/3$. 6. $\Pr(X \leq a) = a^2$ for any $0 \leq a \leq 1$. 7 (a) n/mk (b) $(nk \log k)/m$. 8. Run \mathcal{A} 30 times. If it ever says "not prime", output "not prime"; else output "prime". 10 (a) $|E|/2$ (b) Yes, because the min cut has size at most $|E|$.