CSE 312: Foundations of Computing II

Section 4: Random Variables, Linearity of Expectation

0. Balls in Bins

Let X be the number of bins that remain empty when m balls are distributed into n bins randomly and independently. For each ball, each bin has an equal probability of being chosen. (Notice that two bins being empty are not independent events: if one bin is empty, that decreases the probability that the second bin will also be empty. This is particularly obvious when $n = 2$ and $m > 0$.) Find $\mathbb{E}[X]$.

1. Fair Game?

You flip a fair coin independently and count the number of flips until the first tail, including that tail flip in the count. If the count is n, you receive 2^n dollars. What is the expected amount you will receive? How much would you be willing to pay at the start to play this game?

2. Symmetric Difference

Suppose A and B are random, independent (possibly empty) subsets of $\{1, 2, \ldots, n\}$, where each subset is equally likely to be chosen as A or B. Consider $A \Delta B = (A \cap B^C) \cup (B \cap A^C) = (A \cup B) \cap (A^C \cup B^C)$, i.e., the set containing elements that are in exactly one of A and B. Let X be the random variable that is the size of $A \Delta B$. What is $\mathbb{E}[X]$?

3. Negative Binomial Random Variable

Recall that $W \sim Geo(p)$ (W has a geometric distribution with success parameter p) if it is the number of independent coin flips up to and including the first head, where Pr(HEAD) $= p$. The probability mass function is $p_W(k) = (1-p)^{k-1}p$ and $\mathbb{E}[W] = \frac{1}{p}$. What if we wanted to flip until the r^{th} head, and not just the first? We say X is a **negative binomial** random variable with parameters r a positive integer and $p =$ Pr(HEAD) (written $X \sim NegBin(r, p)$) if X is the number of independent coin flips up to and including the r^{th} head.

(a) What is the codomain Ω_X, and the probability mass function $p_X(k)$, if $X \sim NegBin(r, p)$?

(b) Find $\mathbb{E}[X]$ (hint: use linearity of expectation with r appropriate random variables, which are not necessarily indicator variables).

4. Hypergeometric Random Variable

Recall the trick or treating scenario: Suppose on Halloween, someone is too lazy to keep answering the door, and leaves a jar of exactly N total candies. You count that there are exactly K of them which are kit kats (and the rest are not). The sign says to please take exactly n candies. Each item is equally likely to be drawn. Let X be the number of kit kats we draw (out of n). We say X is a **hypergeometric** random variable, and write $X \sim HypGeo(N, K, n)$.

(a) Find $p_X(k) =$ Pr$(X = k)$.

(b) Compute $\mathbb{E}[X]$ (hint: define appropriate indicator variables and use linearity of expectation).

(c) Suppose we have the same setup: N candies total, K of which are kit kats, and we plan to draw n of them. This time, however, we just want to sniff the candies. We will draw a candy, sniff the candy, **put it back**, and draw another,.... We do this n times total. Let Y be the number of kit kats sniffed. What distribution does Y have, and what is $\mathbb{E}[Y]$? Compare it to the expectation from the previous part from when we didn't return the candies.