CSE 312: Foundations of Computing II

Section 4: Random Variables, Linearity of Expectation Solutions

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 $\mathbb{E}[X]$.

Solution:

For $i \in [n]$, let X_i be 1 if bin i is empty, and 0 otherwise. Then, $X = \sum_{i=1}^{n} X_i$. We first compute $\mathbb{E}[X_i] = 1 \cdot \Pr(X_i = 1) + 0 \cdot \Pr(X_i = 0) = \Pr(X_i = 1) = (\frac{n-1}{n})^m$. Hence,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot (\frac{n-1}{n})^m$$

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?

Solution:

The expected amount is ∞. Let N be the number of flips until the first tail, so $p_N(n) = \frac{1}{2^n}$ for $n \in \mathbb{N}$. Hence $\mathbb{E}\left[2^N\right] = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty$. In theory, you should be willing to pay any finite amount of money to play this game, but I admit I would be nervous to pay a lot. For instance, if you pay \$1000, you will lose money unless the first 9 flips are all heads. With high probability you will lose money, and with low probability you will win a lot of money.

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 $\mathbb{E}[X]$?

Solution:

For $i = 1, 2, \ldots, n$, let X_i be the indicator of whether $i \in A \Delta B$. Then $\mathbb{E}[X_i] = \Pr(X_i = 1) = \frac{1}{2}$, and $X = \sum_{i=1}^{n} X_i$, so

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{n}{2}$$

.

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 $\Pr(\text{HEAD}) = p$. 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 $p = \Pr(\text{HEAD})$ (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 $X \sim NegBin(r, p)$?

Solution:

We must flip at least r times, and can flip any number of times, so $\Omega_X = \{r, r+1, ...\}$. To get the r^{th} head on the k^{th} flip, the first $k-1$ must have exactly $r-1$ heads and $k-r$ tails, followed by a head; there are $\binom{k-1}{r-1}$ ways to choose positions of these heads and tails. There are r heads and $k-r$ tails total, with probability $p^r(1-p)^{k-r}$ for any particular sequence. Hence,

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

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

Solution:

Let $X_1, ..., X_r$ be independent $Geo(p)$ random variables. Then, $X = \sum_{i=1}^{r} X_i$. Hence, by linearity of expectation,

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

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 $p_X(k) = \Pr(X = k)$.

Solution:

$$\boxed{p_X(k) = \Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}}$$

We choose k out of the K kit kats, and $n-k$ out of the $N-K$ other candies. The denominator is the total number of ways to choose n candies out of N total.

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

Solution:

For $i = 1, ..., n$, let X_i be 1 if candy i was a kit kat, and 0 otherwise. Then, $\mathbb{E}[X_i] = 1 \cdot \Pr(X_i = 1) + 0 \cdot \Pr(X_i = 0) = \Pr(X_i = 1) = \frac{K}{N}$. So,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{K}{N} = n\frac{K}{N}$$

(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.

Solution:

$Y \sim Bin(n, \frac{K}{N})$, and we know $\mathbb{E}[Y] = n\frac{K}{N}$. It is the same!