CSE 312: Foundations of Computing II

Section 4: Random Variables

1. Review of Main Concepts

(a) **Random Variable (rv)**: A numeric function $X : \Omega \to \mathbb{R}$ of the outcome.

(b) **Range/Support:** The support/range of a random variable X, denoted Ω_X, is the set of all possible values that X can take on.

(c) **Discrete Random Variable (drv)**: A random variable taking on a countable (either finite or countably infinite) number of possible values.

(d) **Probability Mass Function (pmf) for a discrete random variable $\mathbf{X}$**: a function $p_X : \Omega_X \to [0, 1]$ with $p_X(x) = \mathbb{P}(X = x)$ that maps possible values of a discrete random variable to the probability of that value happening, such that $\sum_x p_X(x) = 1$.

(e) **Cumulative Distribution Function (CDF) for a random variable $\mathbf{X}$**: a function $F_X : R \to R$ with $F_X(x) = \mathbb{P}(X \le x)$

(f) **Expectation (expected value, mean, or average)**: The expectation of a discrete random variable is defined to be $\mathbb{E}[X] = \sum_x x p_X(x) = \sum_x x\mathbb{P}(X = x)$. The expectation of a function of a discrete random variable $g(X)$ is $\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$.

(g) **Linearity of Expectation**: Let X and Y be random variables, and $a, b, c \in \mathbb{R}$. Then, $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$. Also, for any random variables $X_1, \dots, X_n$,

$$\mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + + \ldots + \mathbb{E}[X_n] .$$

(h) **Variance**: Let X be a random variable and $\mu = \mathbb{E}[X]$. The variance of X is defined to be $Var(X) = \mathbb{E}\left[(X - \mu)^2\right]$. Notice that since this is an expectation of a non-negative random variable ($(X - \mu)^2$), variance is always non-negative. With some algebra, we can simplify this to $Var(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.

(i) **Standard Deviation**: Let X be a random variable. We define the standard deviation of X to be the square root of the variance, and denote it $\sigma = \sqrt{Var(X)}$.

(j) **Property of Variance**: Let $a, b \in \mathbb{R}$ and let X be a random variable. Then, $Var(aX + b) = a^2 Var(X)$.

(k) **Independence**: Random variables X and Y are independent iff

$$\forall x \forall y, \quad \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

In this case, we have $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ (the converse is not necessarily true).

(l) **i.i.d. (independent and identically distributed)**: Random variables $X_1, \dots, X_n$ are i.i.d. (or iid) iff they are independent and have the same probability mass function.

(m) **Variance of Independent Variables**: If X is independent of Y, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. This depends on independence, whereas linearity of expectation always holds. Note that this combined with the above shows that $\forall a, b, c \in \mathbb{R}$ and if X is independent of Y, $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$.

2. 3-sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the probability mass function of X?

(b) Find $\mathbb{E}[X]$ directly from the definition of expectation.

(c) Find $\mathbb{E}[X]$ again, but this time using linearity of expectation.

(d) What is $Var(X)$? (Use LOTUS to compute $\mathbb{E}\left[X^2\right]$.)

3. Kit Kats

Suppose we have N candies in a jar, K of which are kit kats. Suppose we draw (without replacement) until we have (exactly) k kit kats, $k \leq K \leq N$. Let X be the number of draws until the k^{th} kit kat. What is Ω_X, the range of X? What is $p_X(n) = \mathbb{P}(X = n)$?

4. Hungry Washing Machine

You have 10 pairs of socks (so 20 socks in total), with each pair being a different color. You put them in the washing machine, but the washing machine eats 4 of the socks chosen at random. Every subset of 4 socks is equally probable to be the subset that gets eaten. Let X be the number of complete pairs of socks that you have left.

(a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

(b) Find $\mathbb{E}[X]$ from the definition of expectation.

(c) Find $\mathbb{E}[X]$ using linearity of expectation.

(d) Which way was easier? Doing both (a) and (b), or just (c)?

5. Practice

(a) Let X be a random variable with $p_X(k) = ck$ for $k \in \{1, \ldots, 5\} = \Omega_X$, and 0 otherwise. Find the value of c that makes X follow a valid probability distribution and compute its mean and variance ($E[X]$ and $Var(X)$).

(b) Let X be *any* random variable with mean $E[X] = \mu$ and variance $Var(X) = \sigma^2$. Find the mean and variance of $Z = \dfrac{X - \mu}{\sigma}$. (When you're done, you'll see why we call this a "standardized" version of X!)

(c) Let X, Y be independent random variables. Find the mean and variance of $X - 3Y - 5$ in terms of $E[X], E[Y], Var(X)$, and $Var(Y)$.

(d) Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables each with mean μ and variance σ^2. The sample mean is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find the mean and variance of $\bar{X}$. If you use the independence assumption anywhere, **explicitly label** at which step(s) it is necessary for your equalities to be true.

6. Hat Check

At a reception, n people give their hats to a hat-check person. When they leave, the hat-check person gives each of them a hat chosen at random from the hats that remain. What is the expected number of people who get their own hats back? (Notice that the hats returned to two people are not independent events: if a certain hat is returned to one person, it cannot also be returned to the other person.)

7. 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 $\mathbb{E}[X]$.

For this section, we expect to end here (or before!). The rest of these problems can be done at home for extra practice, or if you finish 1-5 early. Solutions will be posted.

8. 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?

9. 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 $\mathbb{E}[X]$?

10. Identify that Range!

Identify the support/range Ω_X of the random variable X, if X is...

(a) The sum of two rolls of a six-sided die.

(b) The number of lottery tickets I buy until I win it.

(c) The number of heads in n flips of a coin with $0 < \mathbb{P}(\text{head}) < 1$.

(d) The number of heads in n flips of a coin with $\mathbb{P}(\text{head}) = 1$.

(e) The time I wait at the bus stop for the next bus.