Random Variables and Expectation

Random Variable

A Random Variable (RV) is a variable that probabilistically takes on different values. You can think of an RV as being like a variable in a programming language. They take on values, have types and have domains over which they are applicable. We can define events that occur if the random variable takes one values that satisfy a numerical test (eg does the variable equal 5, is the variable less than 8). We often think of the probabilities of such events.

As an example, let's say we flip three fair coins. We can define a random variable Y to be the total number of "heads" on the three coins. We can ask about the probability of Y taking on different values using the following notation:

- $P(Y = 0) = 1/8$ (T, T, T)

- $P(Y = 1) = 3/8$ (H, T, T), (T, H, T), (T, T, H)

- $P(Y = 2) = 3/8$ (H, H, T), (H, T, H), (T, H, H)

- $P(Y = 3) = 1/8$ (H, H, H)

- $P(Y \geq 4) = 0$

Using random variables is a convenient notation technique that assists in decomposing problems. There are many different types of random variables (indicator, binary, choice, Bernoulli, etc). The two main families of random variable types are discrete and continuous.

Probability Mass Function

For a discrete random variable, the most important thing to know is a mapping between the values that the random variable could take on and the probability of the random variable taking on said value. In mathematics, we call associations functions.
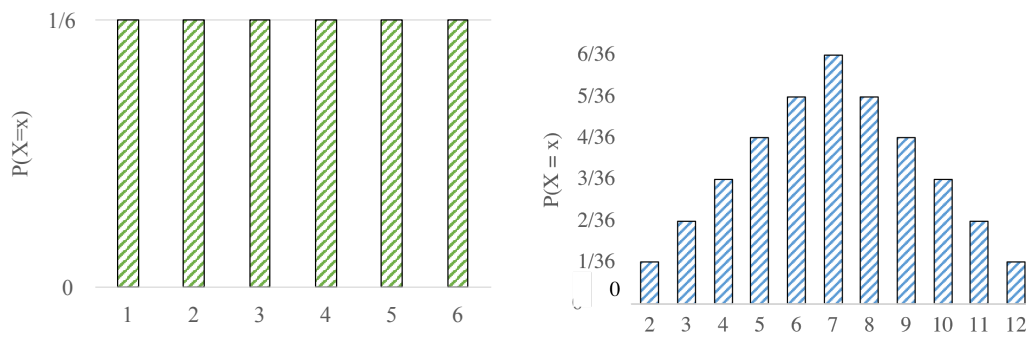

Figure 1: On the left, the PMF of a single 6 sided die roll. On the right, the PMF of the sum of two dice rolls.

The probability mass functions (PMF) maps possible outcomes of a random variable to the corresponding probabilities. Because it is a function, we can plot PMF graphs where the x-axis are the values that the random variable could take on and the y-axis is the probability of the random variable taking on said value:

There are many ways that these Probability Mass Functions can be specified. We could draw a graph. We could have a table (or for you CS folks, a Map) that lists out all the probabilities for all possible events. Or we could write out a mathematical expression.

For example lets consider the random variable X which is the sum of two dice rolls. The probability mass function can be defined by the graph on the right of figure . It could have also been defined using the equation:

$$p_X(x) = \begin{cases} \frac{x-1}{36} & \text{if } x \in \mathbb{Z}, 0 \leq x \leq 6 \\ \frac{13-x}{36} & \text{if } x \in \mathbb{Z}, 7 \leq x \leq 12 \\ 0 & \text{else} \end{cases}$$

The probability mass function, $p_X(x)$, defines the probability of X taking on the value x. The new notation $p_X(x)$ is simply different notation for writing $P(X = x)$. Using this new notation makes it more apparent that we are specifying a function. Try a few values of x, and compare the value of $p_X(x)$ to the graph in figure 1. They should be the same.

Expected Value

A relevant statistic for a random variable is the average value of the random variable over many repetitions of the experiment it represents. This average is called the Expected Value.

The Expected Value for a discrete random variable X is defined as:

$$E[X] = \sum_{x:P(x)>0} xP(x)$$

It goes by many other names: Mean, Expectation, Weighted Average, Center of Mass, 1st Moment.

Example 1

Lets say you roll a 6-Sided Die and that a random variable X represents the outcome of the roll. What is the $E[X]$? This is the same as asking what is the average value.

$$E[X] = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 7/2$$

Example 2

Lets say a school has 3 classes with 5, 10, and 150 students. If we randomly choose a class with equal probability and let X = size of the chosen class:

$$E[X] = 5(1/3) + 10(1/3) + 150(1/3)$$
$$= 165/3 = 55$$

If instead we randomly choose a student with equal probability and let Y = size of the class the student is in

$$E[Y] = 5(5/165) + 10(10/165) + 150(150/165)$$
$$= 22635/165 = 137$$

Example 3

Consider a game played with a fair coin which comes up heads with p = 0.5. Let n = the number of coin flips before the first "tails". In this game you win 2^n. How many dollars do you expect to win? Let X be a

random variable which represents your winnings.

$$E[X] = \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \cdots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i$$

$$= \sum_{i=0}^{\infty} \frac{1}{2} = \infty$$

Properties of Expectation

Expectations preserve linearity which means that

$$E[aX + b] = aE[X] + b$$

It also holds in the case where you are adding random variables. Regardless of the relationship between random variables, the expectation of the sum is equal to the sum of the expectation. For random variables A and B:

$$E[A + B] = E[A] + E[B] \qquad \text{For two random variables}$$

$$E[\sum_i X_i] = \sum_i E[X_i] \qquad \text{For any number of random variables } X_i$$

There is a wonderful law called the Law of the Unconcious Statistician that is used to calculate the expected value of a function $g(X)$ of a random variable X when one knows the probability distribution of X but one does not explicitly know the distribution of $g(X)$.

$$E[g(X)] = \sum_x g(x) \cdot P(X = x)$$

For example, lets apply the law of the unconscious statistician to compute the expectation of the square of a random variable (called the second moment).

$$E[X^2] = E[g(X)] \qquad \text{where } g(X) = X^2$$

$$= \sum_x g(x) \cdot P(X = x) \qquad \text{by the unconscious statistician}$$

$$= \sum_x x^2 \cdot P(X = x) \qquad \text{by the unconscious statistician}$$