MAT 105 - Summary of Lecture 13
Tuesday, February 23, 2016

Chance, too, which seems to rush along with slack reins, is bridled and governed by law. (Boethius)

Discrete random variables

Consider the following example, using a two-step experiment approach to conditional probability. I roll a 3-sided die. Let X be the value that appears on the die, with $X \in \{1, 2, 3\}$. Now we toss X coins and let Y count the number of heads. Of course, Y depends on X, but its values could be $Y \in \{0, 1, 2, 3\}$. We find the distribution of Y by looking at the following tree:

$$\begin{cases} X = 1, \quad \text{prob. } 1/3 \begin{cases} Y = 0, & \text{prob. } 1/2 \quad (T) \\ Y = 1, & \text{prob. } 1/2 \quad (H) \end{cases} \\ X = 2, \quad \text{prob. } 1/3 \begin{cases} Y = 0, & \text{prob. } 1/4 \quad (TT) \\ Y = 1, & \text{prob. } 2/4 \quad (HT, TH) \\ Y = 2, & \text{prob. } 1/4 \quad (HH) \end{cases} \\ X = 3, \quad \text{prob. } 1/3 \begin{cases} Y = 0, & \text{prob. } 1/8 \quad (TTT) \\ Y = 1, & \text{prob. } 3/8 \quad (HTT, THT, TTH) \\ Y = 2, & \text{prob. } 3/8 \quad (THH, HTH, HHT) \\ Y = 3, & \text{prob. } 1/8 \quad (HHH) \end{cases} \end{cases}$$

Thus we have

$$\begin{aligned} P(Y = 3) &= P(X = 3, Y = 3) = \frac{1}{3} \cdot \frac{1}{8} = \frac{1}{24} \\ P(Y = 2) &= P(X = 3, Y = 2) + P(X = 2, Y = 2) = \frac{1}{3} \cdot \frac{3}{8} + \frac{1}{3} \cdot \frac{1}{4} = \frac{5}{24} \\ P(Y = 1) &= P(X = 3, Y = 1) + P(X = 2, Y = 1) + P(X = 1, Y = 1) = \frac{1}{3} \cdot \frac{3}{8} + \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{3} \cdot \frac{1}{2} = \frac{11}{24} \\ P(Y = 0) &= P(X = 3, Y = 0) + P(X = 2, Y = 0) + P(X = 1, Y = 0) = \frac{1}{3} \cdot \frac{1}{8} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{2} = \frac{7}{24} \end{aligned}$$

What if we know that $Y = 1$? What is the probability that we rolled a 2? That is, find

$$P(X = 2 \mid Y = 1) = \frac{P(X = 2, Y = 1)}{P(Y = 1)} = \frac{\frac{1}{3} \cdot \frac{2}{4}}{\frac{11}{24}} = \frac{1}{3} \cdot \frac{2}{4} \cdot \frac{24}{11} = \frac{4}{11}.$$

Note that once we knew the value of X, we had to compute the probabilities of Y by listing all outcomes in the event of tossing X coins. What if we first roll a 100 sided die and then toss up to 100 coins? We surely do not want to list 2^{100} outcomes for this experiment. There must be an easier way! We will recognize Y as a common random variable and we will be able to describe its distribution without having to list all possible outcomes.

We will focus mostly on discrete random variables. What do we mean by a **discrete** random variable? We mean that the random variable takes on "countably many" values. That is, either a finite number like $\{1, 2, 3, 4, 5, 6\}$ or an infinite number of values that we can order, such as $\{0, 1, 2, 3, 4, \dots\}$. It never takes on all real values in an interval! That would make the random variable **continuous**. While continuous random variables are important, and we will study at least one of them this semester, they are hard to work with without knowing some calculus. So we will focus our attention on the discrete random variables that are most often used in games.

1. **Bernoulli Random Variables**

   Suppose we run an experiment **ONCE** and it can result in **success or failure**. Suppose the probability of success is p for some constant p satisfying $0 \le p \le 1$. Let

$$\begin{array}{lll} X = 1 & \text{if experiment results in success} & \Rightarrow P(X=1) = p, \\ X = 0 & \text{if experiment results in failure} & \Rightarrow P(X=0) = 1-p. \end{array}$$

   We call X a Bernoulli random variable with parameter p (which measures success), and denote it by $X = Bernoulli(p)$. We compute the expectation and variance as follows:

$$E[X] = 0 \cdot P(X=0) + 1 \cdot P(X=1) = p$$
$$Var(X) = E[X^2] - E[X]^2 = 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) - E[X]^2 = p - p^2 = p(1-p).$$

   Examples:

   (a) Flip one coin and wish for a Heads. Then getting H is considered a success, so we let $X = 1$ when that occurs. Thus, $P(X = 1) = P(\text{Heads}) = 1/2$.

   (b) Take MAT105 and wish to pass. Suppose 95% of students pass the class, so if we let $X = 1$ if you pass the class, then $P(X = 1) = P(\text{pass}) = .95$.

   (c) Run code and wish to not have any errors. About 70% of the time your code runs without errors, so if we let $X = 1$ for an error-free test, then $P(X = 1) = P(\text{no errors}) = .7$.

2. **Binomial Random Variables:**

   Consider the example: roll four 6-sided dice and let X = number of 6's. Then $X \in \{0, 1, 2, 3, 4\}$, so it is discrete since it takes on 5 values. Find the distribution of X.

$$\begin{aligned} P(X=0) &= P(\text{no 6's}) = \frac{5^4}{6^4} \\ P(X=1) &= P(\text{one 6}) = \frac{\binom{4}{1} \times 1 \times 5^3}{6^4} \\ P(X=2) &= P(\text{two 6's}) = \frac{\binom{4}{2} \times 1^2 \times 5^2}{6^4} \\ P(X=3) &= P(\text{three 6's}) = \frac{\binom{4}{3} \times 1^3 \times 5^1}{6^4} \\ P(X=4) &= P(\text{four 6's}) = \frac{\binom{4}{4} \times 1^4}{6^4} \end{aligned}$$

   We reasoned as follows (look at $P(X = 2)$, all other work the same):

   - there are 4 places to fill in, 6 choices for each place, so a total of 6^4 outcomes. This is the denominator in each probability above.
   - out of 4 places we choose 2 where to place 6's. This is done in $\binom{4}{2}$ ways.
   - there is 1 way to choose a 6 for each of the two 6's. So 1^2 ways to choose 6's.
   - there are 5 ways to choose a value different than 6 to place in the remaining two places. This is done in 5^2 ways.
   - the resulting probability is then $P(X = 2) = \dfrac{\binom{4}{2} \times 1^2 \times 5^2}{6^4} = \binom{4}{2} \left(\dfrac{1}{6}\right)^2 \left(\dfrac{5}{6}\right)^2$.

But one can think of this in another way:

$$P(X=2) = (\text{\# of ways to place two 6's in four spots})(\text{prob. of 6})^2(\text{prob. of non-6})^2$$

This agrees with the general formula we will use for such a random variable that counts the number of successes in a given number of trials.

Definition: The **binomial random variable** with parameters n and p denoted by $Bin(n,p)$ counts the number of successes in n independent trials, with probability of each success being p. Its distribution is given for $k \in \{0,1,2,\dots,n\}$ by the formula

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Remark: the reason why we call it a binomial random variable is because its distribution has the combinations $\binom{n}{k}$, which are also called binomials, due to their use in the Binomial Theorem. The Binomial Theorem states that the coefficients of $(x+y)^n$ are given by $\binom{n}{k}$. More precisely,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

You may have encountered the Binomial Theorem as an application of Pascal's triangle (or vice versa):

$$\begin{array}{ccccccccccc}
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

Examples of binomial random variables:

(a) Count the number of heads in 10 coin flips. Then if X is the counter, $X = Bin(10, 0.5)$.

(b) A certain device is functioning with probability 0.9. We test 50 devices and let X be the number that are functioning. Then $X = Bin(50, 0.9)$.

(c) A medical screening test is positive with probability 0.01. If we let X be the number of positive cases out of 1000 people screened for the disease, $X = Bin(1000, 0.01)$.

The expectation and variance of $X = Bin(n,p)$ are $E[X] = np$ and $Var(X) = np(1-p)$.

Remark: Note that this agrees with our result for Bernoulli for two reasons. First, $Bernoulli(p) = Bin(1,p)$. Second, if one averages μ in one trial, then they should average $n\mu$ in n trials, since $Bin(n,p)$ is just a sequence of n $Bernoulli(p)$ trials.