Lecture 5 - Discrete Distributions

Sta102 / BME102

Colin Rundel

January 23, 2015

Random variables

- A *random variable* is a numeric quantity whose value depends on the outcome of a random event
  - We use a capital letter, like X, to denote a random variables
  - The values of a random variable will be denoted with a lower case letter, in this case x
  - For example, $P(X = x)$
- There are two types of random variables:
  - *Discrete random variables* take on only integer values
    - Example: Number of credit hours, Difference in number of credit hours this term vs last
  - *Continuous random variables* take on real (decimal) values
    - Example: Cost of books this term, Difference in cost of books this term vs last

Discrete Probability distributions

A *discrete probability distribution* lists all possible *outcomes* and the probabilities with which they occur.

- The probability distribution for the gender of one child:

Event	B	G
Probability	0.5	0.5

- Rules for probability distributions:
  1. The outcomes listed must be disjoint
  2. Each outcome probability must be between 0 and 1
  3. The sum of the outcome probabilities must add up to 1

Example - Discrete RV model

In a game of cards you win $1 if you draw a heart, $5 if you draw an ace (including the ace of hearts), $10 if you draw the king of spades and nothing for any other card you draw.

Write the probability distribution for the random variable of your winnings.

Mean and standard deviation of a discrete RVs

We are often interested in the value we expect to arise from a RV.

- We call this the expected value, it is a weighted average of the possible outcomes

$$E(X) = \sum_x x \cdot P(X = x)$$

$$E(f(X)) = \sum_x f(x) \cdot P(X = x)$$

We are also often interested in the variability in the values of a RV.

- Described using Variance and Standard deviation

$$Var(X) = E[(X - E(X))^2] = \sum_x (x - E(X))^2 \cdot P(X = x)$$

$$SD(X) = \sqrt{Var(X)}$$

Example - Discrete RV - Mean and SD

For the previous example what is the expected value and the standard deviation of your winnings?

X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$

Sampling and RVs

Imagine that you don't just play the game once you play it repeatedly, for a total of n times.

We can think of each play as being a *sample* from the winnings distribution, giving us n samples $(x_1, x_2, \ldots, x_3)$.

We can then calculate *summary statistics* for this sample:

$$\bar{x}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

$$s_n^2 = \frac{1}{n-1}\left[(x_1 - \bar{x}_n)^2 + (x_2 - \bar{x}_n)^2 + \cdots + (x_n - \bar{x}_n)^2\right]$$

Sampling and RV (Cont.)

We care about the expected value and variance of a RV's distribution are important because,

$$\lim_{n \to \infty} \bar{x}_n = E(X)$$

$$\lim_{n \to \infty} s_n^2 = Var(X).$$

As such, the expected value gives the long run winnings (or losses) per game and variance give the uncertainty in each observation.

Expected value and Roulette

What is the expected value of betting on $1 on black in roulette? The variance?

Bernoulli Random Variable

A Bernoulli random variable describes a trial with only two possible outcomes, one of which we will label a success and the other a failure and where the probability of a success is given by the parameter p. Since RVs must be numeric, the random variable is defined to be 1 for a success and 0 for a failure.

X	P(X=x)
0	1-p
1	p

$$P(X = x|p) = \begin{cases} p & \text{if x=1} \\ 1-p & \text{if x=0} \end{cases}$$

Properties of a Bernoulli Random Variable

Let $X \sim \text{Bern}(p)$ then

$$\begin{aligned} E(X) &= \sum_x x\ P(X=x) \\ &= 0 \cdot P(X=0) + 1 \cdot P(X=1) \\ &= P(X=1) \\ &= p \end{aligned}$$

$$\begin{aligned} Var(X) &= E(X-p)^2 \\ &= E(X^2 - 2Xp + p^2) \\ &= E(X^2) - 2p\ E(X) + p^2 \\ &= (0^2 \cdot P(X=0) + 1^2 \cdot P(X=1)) - p^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

Geometric Random Variable

A Geometric random variable describes the number of (identical) Bernoulli trials that occur before the first success is observed. The distribution has a single parameter, the probability of a success p. There is another slightly different characterization that counts the number of failures before the first success. We will focus on the former for now.

X	$P(X=x)$
1	p
2	$p(1-p)$
3	$p(1-p)^2$
4	$p(1-p)^3$
$\vdots$	$\vdots$
∞	$p(1-p)^\infty = 0$

$$P(X = x|p) = p(1-p)^{x-1}$$

$$E(X) = 1/p$$

$$Var(X) = \frac{1-p}{p^2}$$

Validity of a Geometric RV

If $|r| < 1$ then,

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

We can use the first result to show that X has a valid probability distribution,

$$\begin{aligned}
\sum_{x=1}^{\infty} P(X=x) &= \sum_{x=1}^{\infty} p(1-p)^{x-1} = p \sum_{x=1} (1-p)^{x-1} \\
&= \frac{p}{(1-p)} \sum_{x=1} (1-p)^x = \frac{p}{(1-p)} \left(\frac{1}{1-(1-p)} - 1 \right) \\
&= \left(\frac{1}{1-p} - \frac{p}{(1-p)} \right) \\
&= \frac{1-p}{1-p} = 1
\end{aligned}$$

Properties of a Geometric Random Variable

Similarly, if $|r| < 1$ then,

$$\sum_{k=1}^{\infty} k r^k = \frac{r}{(1-r)^2}.$$

If we define $X \sim \text{Geo}(p)$ then

$$\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x \; P(X=x) = \sum_{x=1}^{\infty} x \; p(1-p)^{x-1} \\
&= \frac{p}{(1-p)} \sum_{x=1}^{\infty} x \; (1-p)^x = \frac{p}{(1-p)} \frac{(1-p)}{(1-(1-p))^2} \\
&= 1/p
\end{aligned}$$

$$Var(X) = E((X-E(X))^2) = \sum_{x=1}^{\infty} (x-1/p)^2 \; P(X=x) = \frac{1-p}{p^2}$$

Permutations

Another option for those n items is if we select k of them and want to know how many possible unique orderings there are.

Given by

$$\frac{n!}{(n-k)!}$$

How many permutations of two numbers between 1 and 6 are there:

Combinations

A common problem in probability asks - if we have n items and want to select k of them how many possible groupings (order does not matter) are there?

Given by the binomial coefficient

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

How many combinations of two numbers between 1 and 6 are there:

Derivation

Pascal's Triangle

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 5 15 20 15 5 1
1 6 20 35 35 20 6 1
1 7 26 55 70 55 26 7 1
⋮

$$\binom{0}{0}$$
$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$
$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$

Some properties of the Combinations / Binomial coefficient

Pascal's Triangle is symmetric:

$$\binom{n}{k} = \binom{n}{n-k}$$

Each element is the sum of the two above it:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \text{ for } 0 < k < n$$

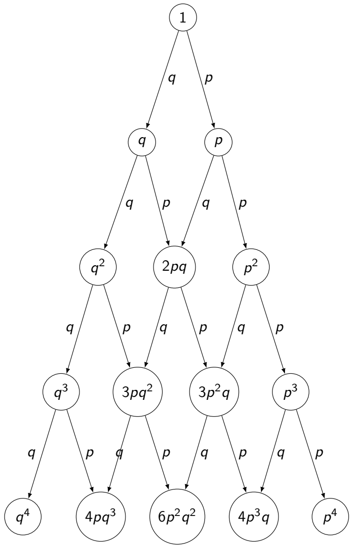
The sum of row n is 2^n if we start counting rows at 0:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Example - Cell Culture

A researcher is working with a new cell line, if there is a 10% chance of a single culture becoming contaminated during the week what is the probability that if the researcher has four cultures that only one of them will be contaminated at the end of the week? What about the probability k cultures lasting the week?

Binomial Distribution



Binomial Distribution

We define a random variable X that reflects the *number of successes* in a *fixed number* of *independent trials* with the *same probability of success* as having a binomial distribution.

If there are n trials then

$$X \sim \text{Binom}(n, p)$$

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

Properties of Binomial RVs

Let $X \sim \text{Binom}(n, p)$ then, $\left[\text{Hint: } (a+b)^m = \sum_{k=0}^{m} \binom{n}{m} x^m\, y^{n-m}\right]$

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} x\ P(X=x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} = \sum_{x=1}^{n} \frac{n!}{(n-x)!(x-1)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x'=0}^{n-1} \frac{(n-1)!}{(n-(x'+1))!(x')!} p^{x'} (1-p)^{n-(x'+1)} \\
&= np \sum_{x'=0}^{n-1} \frac{(n-1)!}{(n-1-x')!(x')!} p^{x'} (1-p)^{n-1-x'} \\
&= np(p + (1-p))^n = np
\end{aligned}$$

Properties of Binomial RVs

Let $X \sim \text{Binom}(n, p)$ then,

$$\begin{aligned}
Var(X) &= E\left[(X - E(X))^2\right] \\
&= \sum_{x=0}^{n} (x - np)^2\ P(X = x) \\
&= \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&\vdots \qquad\qquad \text{(lots of awfulness)} \\
&= np(1-p)
\end{aligned}$$

We'll see an simple and elegant way of solving this on Wednesday.

Binomial Examples

Imagine you roll four 6-sided dice, find the following probabilities:

- Getting 4 dice showing a 5.
- Getting 2 dice showing a 5 or 6.
- Getting >1 dice showing a 5 or 6.
- Getting 5 or less dice showing an even number.

St. Petersburg Lottery

We start with $1 on the table and a coin.

At each step: Toss the coin; if it shows Heads, take the money. If it shows Tails, I double the money on the table.

How much would you pay me to play this game? i.e. what is the expected value?