Lecture 5 - Discrete Distributions

Sta102 / BME102

February 1, 2016

Colin Rundel

Random Variables

Random variables

A *random variable* is a numeric quantity whose value depends on the outcome of a random event

- We use a capital letter, X, to denote a random variables
- The values of a random variable will be denoted with a lower case letter, in this case 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 RVs

Discrete Probability distributions

A discrete *probability distribution* is a list or formula that enumerates all possible *outcomes* and their probabilities.

Discrete Probability distributions

A discrete *probability distribution* is a list or formula that enumerates all possible *outcomes* and their probabilities.

- The probability distribution for flipping a coin:

Outcome	H	Y
Probability	0.5	0.5

Discrete Probability distributions

A discrete *probability distribution* is a list or formula that enumerates all possible *outcomes* and their probabilities.

- The probability distribution for flipping a coin:

Outcome	H	Y
Probability	0.5	0.5

- Rules for probability distributions:
  1. All outcomes must be disjoint
  2. Probability of each outcome must be between 0 and 1
  3. The sum of probabilities of all outcome must add up to 1

Discrete Probability distributions

A discrete *probability distribution* is a list or formula that enumerates all possible *outcomes* and their probabilities.

- The probability distribution for flipping a coin:

Outcome	H	Y
Probability	0.5	0.5

- Rules for probability distributions:
    1. All outcomes must be disjoint
    2. Probability of each outcome must be between 0 and 1
    3. The sum of probabilities of all outcome must add up to 1

- The probability distribution for flipping two coins:

Outcome	HH	TT	TH	HT
Probability	0.25	0.25	0.25	0.25

Discrete RVs Examples

We can reformulate the outcome spaces from the previous slide as random variables by quantifying the outcomes in some way.

For example, let X be the number of heads in one coin flip and Z be the number of heads in two coin flips:

Example - Discrete RV model

In a game of cards you win \$1 if you draw a heart (except the ace), \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and \$0 for any other card.

Let W be the amount you win playing this game, what the probability distribution of W?

Expected Value

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

- We call this the expected value, it is the (probability) weighted average of the possible outcomes

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

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

Variance

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 - E(W) and Var(W)

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

Example - E(W) and Var(W)

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

X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
0	$\frac{35}{52}$			
1	$\frac{12}{52}$			
5	$\frac{4}{52}$			
10	$\frac{1}{52}$			

Example - E(W) and Var(W)

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

X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$		
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$		
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$		
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$		

Example - E(W) and Var(W)

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

X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$		
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$		
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$		
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$		
		$E(X) = \frac{42}{52} = 0.81$		

Example - E(W) and Var(W)

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

X	$P(X)$	$X \cdot P(X)$	$(X - E(X))^2$	$P(X) \cdot (X - E(X))^2$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	
		$E(X) = \frac{42}{52} = 0.81$		

Example - E(W) and Var(W)

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

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	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.4561 = 1.6242$
		$E(X) = \frac{42}{52} = 0.81$		

Example - E(W) and Var(W)

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

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	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.4561 = 1.6242$
		$E(X) = \frac{42}{52} = 0.81$		$Var(X) = 3.4246$

Example - E(W) and Var(W)

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

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	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.4561 = 1.6242$
		$E(X) = \frac{42}{52} = 0.81$		$Var(X) = 3.4246$ $SD(X) = \sqrt{3.4246} = 1.85$

Sampling and RVs

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

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 $(x_1, x_2, \ldots, x_n)$.

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 $(x_1, x_2, \ldots, x_n)$.

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,

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

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

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?

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?

Bernoulli RVs

Bernoulli Random Variable

A random variable for modeling binary events,

- Two possible outcomes:
  - Success - value 1
  - Failure - value 0
- Single parameter p, probability of a success

Bernoulli Random Variable

A random variable for modeling binary events,

- Two possible outcomes:
  - Success - value 1
  - Failure - value 0
- Single parameter p, probability of a success

X	P(X=x—p)
0	1-p
1	p

Bernoulli Random Variable

A random variable for modeling binary events,

- Two possible outcomes:
  - Success - value 1
  - Failure - value 0
- Single parameter p, probability of a success

X	P(X=x—p)
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

Geometric RVs

Geometric Random Variables

A random variable that models the number of (identical) Bernoulli trials needed to obtain the first success.

- Infinite outcomes - $\{1, 2, 3, \ldots, \infty\}$
- Single parameter p, probability of a success for each trial

Geometric Random Variables

A random variable that models the number of (identical) Bernoulli trials needed to obtain the first success.

- Infinite outcomes - $\{1, 2, 3, \ldots, \infty\}$
- Single parameter p, probability of a success for each trial

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

Geometric Random Variables

A random variable that models the number of (identical) Bernoulli trials needed to obtain the first success.

- Infinite outcomes - $\{1, 2, 3, \ldots, \infty\}$
- Single parameter p, probability of a success for each trial

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

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

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

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

Geometric RV - Example 1

What is the probability of flipping a coin and getting the first heads on the third flip?

Geometric RV - Example 2

What is the probability of flipping a coin more than 4 times before getting a heads?

Geometric RV - Example 3

The expected number of rolls it will take to get a 7 when rolling two dice?

Binomial RVs

Permutations

A common counting problem in probability asks - if we have n items and want to select k of them how many possible unique draws are there?

Permutations

A common counting problem in probability asks - if we have n items and want to select k of them how many possible unique draws are there?

For example, how many permutations of *2* numbers between *1* and *6* are there:

Permutations

A common counting problem in probability asks - if we have n items and want to select k of them how many possible unique draws are there?

For example, how many permutations of *2* numbers between *1* and *6* are there:

$$\begin{array}{ccccc} \{1,2\} & \{1,3\} & \{1,4\} & \{1,5\} & \{1,6\} \\ \{2,1\} & \{2,3\} & \{2,4\} & \{2,5\} & \{2,6\} \\ \{3,1\} & \{3,2\} & \{3,4\} & \{3,5\} & \{3,6\} \\ \{4,1\} & \{4,2\} & \{4,3\} & \{4,5\} & \{4,6\} \\ \{5,1\} & \{5,2\} & \{5,3\} & \{5,4\} & \{5,6\} \\ \{6,1\} & \{6,2\} & \{6,3\} & \{6,4\} & \{6,5\} \end{array}$$

Permutations

A common counting problem in probability asks - if we have n items and want to select k of them how many possible unique draws are there?

For example, how many permutations of *2* numbers between *1* and *6* are there:

$$\begin{array}{ccccc} \{1,2\} & \{1,3\} & \{1,4\} & \{1,5\} & \{1,6\} \\ \{2,1\} & \{2,3\} & \{2,4\} & \{2,5\} & \{2,6\} \\ \{3,1\} & \{3,2\} & \{3,4\} & \{3,5\} & \{3,6\} \\ \{4,1\} & \{4,2\} & \{4,3\} & \{4,5\} & \{4,6\} \\ \{5,1\} & \{5,2\} & \{5,3\} & \{5,4\} & \{5,6\} \\ \{6,1\} & \{6,2\} & \{6,3\} & \{6,4\} & \{6,5\} \end{array}$$

Generically, this is given by the permutation formula:

$$P(n,k) = {}_nP_k = \frac{n!}{(n-k)!}$$

Combinations

Another option for those n items is if we select k of them and want to know how many possible groupings there are (here we ignore draw order).

Combinations

Another option for those n items is if we select k of them and want to know how many possible groupings there are (here we ignore draw order).

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

Combinations

Another option for those n items is if we select k of them and want to know how many possible groupings there are (here we ignore draw order).

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

$$\{1,2\}\ \{1,3\}\ \{1,4\}\ \{1,5\}\ \{1,6\}$$
$$\{2,3\}\ \{2,4\}\ \{2,5\}\ \{2,6\}$$
$$\{3,4\}\ \{3,5\}\ \{3,6\}$$
$$\{4,5\}\ \{4,6\}$$
$$\{5,6\}$$

Combinations

Another option for those n items is if we select k of them and want to know how many possible groupings there are (here we ignore draw order).

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

$$\{1,2\}\ \{1,3\}\ \{1,4\}\ \{1,5\}\ \{1,6\}$$
$$\{2,3\}\ \{2,4\}\ \{2,5\}\ \{2,6\}$$
$$\{3,4\}\ \{3,5\}\ \{3,6\}$$
$$\{4,5\}\ \{4,6\}$$
$$\{5,6\}$$

Generically, this is given by the *combination* formula (binomial coefficient):

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

Pascal's Triangle

It is interesting to note that there is a connection between the binomial coefficient and Pascal's triangle:

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 5 \quad 15 \quad 20 \quad 15 \quad 5 \quad 1 \\
1 \quad 6 \quad 20 \quad 35 \quad 35 \quad 20 \quad 6 \quad 1 \\
1 \quad 7 \quad 26 \quad 55 \quad 70 \quad 55 \quad 26 \quad 7 \quad 1 \\
\vdots
\end{array}
$$

Pascal's Triangle

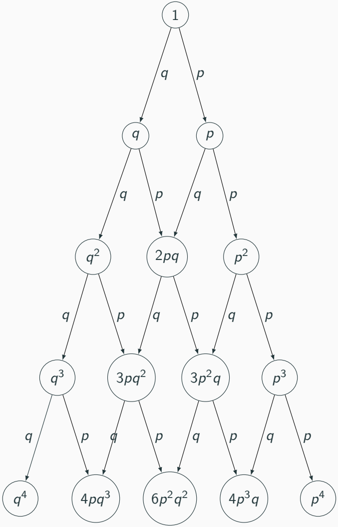
It is interesting to note that there is a connection between the binomial coefficient and Pascal's triangle:

$$
\begin{array}{ccccccccc}
 & & & & \binom{0}{0} & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & & & \\
 & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
 & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
\binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \\
 & & & & \vdots & & & &
\end{array}
$$

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

A random variable that models the *number of successes* in a *fixed number* of *independent Bernoulli trials*.

- Discrete outcomes - $\{0, 1, 2, 3, \ldots, n\}$
- Two parameters:
  - p - probability of a success for each trial
  - n - number of trials

X	$P(X = x)$
0	$\binom{n}{0} p^0 (1-p)^n$
1	$\binom{n}{1} p^1 (1-p)^{n-1}$
2	$p(1-p)$
$\vdots$	$\vdots$
$n-1$	$\binom{n}{n-1} p^{n-1} (1-p)^1$
n	$\binom{n}{n} p^n (1-p)^0$

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

$$E(X) = np$$
$$Var(X) = np(1-p)$$

Binomial - Example 1

Imagine you roll four 6-sided dice, find the probability of

- getting 4 dice showing a 5.

Binomial - Example 2

Imagine you roll four 6-sided dice, find the probability of

- Getting 2 dice showing a 5 or 6.

Binomial - Example 3

Imagine you roll four 6-sided dice, find the probability of

- Getting more than one die showing a 5 or 6.

Binomial - Example 4

Imagine you roll four 6-sided dice, find the probability of

- Getting five or less dice showing an even number.