Lecture 4 - Random Variables and Discrete Distributions

Sta102/BME102

Colin Rundel

January 27, 2014

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 events 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 events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1

Example - Discrete probability 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 representing your winnings.

Mean and standard deviation of a discrete RVs

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

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

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

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

- Described using Variance and Standard deviation

$$\begin{aligned} Var(X) &= E[(X - E(X))^2] \\ &= \sum_x (x - E(X))^2 \cdot P(X = x) \\ SD(X) &= \sqrt{Var(X)} \end{aligned}$$

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	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$

Properties of Expected Value

- **Constant** - $E(c) = c$ if c is constant
- **Constant Multiplication** - $E(cX) = cE(X)$
- **Constant Addition** - $E(X + c) = E(X) + c$
- **Addition** - $E(X + Y) = E(X) + E(Y)$
- **Subtraction** - $E(X - Y) = E(X) - E(Y)$
- **Multiplication** - $E(XY) = E(X)E(Y)$ if X and Y are independent.

Properties of Variance

- **Constant** - $Var(c) = 0$ if c is constant
- **Constant Multiplication** - $Var(cX) = c^2\ Var(x)$
- **Constant Addition** - $Var(X + c) = Var(X)$
- **Addition** - $Var(X + Y) = Var(X) + Var(Y)$ if X and Y are independent.
- **Subtraction** - $Var(X - Y) = Var(X) + Var(Y)$ if X and Y are independent.

Example - Linear Transformation

The average price of a small cup of coffee to go is \$1.40, with a standard deviation of 30¢. An 8.5% tax is added if you take your coffee to stay. Assume that each time you get a coffee to stay you also tip 50¢. What is the mean, variance, and standard deviation of the amount you spend on coffee when to take it to stay?

Example - Linear Transformation, cont.

We now know that $E(X) = 140$, $SD(X) = 30$, and $Y = 1.085X + 50$,

Example - Adding random variables

The average price of a cup of coffee is \$1.40, with a standard deviation of 30¢. The average price of a muffin is \$2.50, with a standard deviation of 15¢. If you get a cup of coffee and a muffin every day for breakfast, what is the mean, variance, and standard deviation of the amount you spend on breakfast daily? Assume that the price of coffee and muffins are independent.

Simplifying RVs

Random variables do not work like normal algebraic variables:

$$X + X \neq 2X$$

$$\begin{aligned} E(X+X) &= E(X) + E(X) \\ &= 2E(X) \\ E(2X) &= 2E(X) \end{aligned}$$

$$\begin{aligned} Var(X+X) &= Var(X) + Var(X) \\ &= 2\ Var(X) \\ Var(2X) &= 2^2\ Var(X) \\ &= 4\ Var(X) \end{aligned}$$

Clearly $E(X+X) = E(2X)$, but $Var(X+X) \neq Var(2X)$.

Combining RVs

A company has 5 Lincoln Town Cars in its fleet. Historical data show that annual fuel cost for each car is on average \$2,154 with a standard deviation of \$132. What is the mean and the standard deviation of the total annual fuel cost for this fleet?

$$\begin{aligned} E(X_1+X_2+X_3+X_4+X_5) &= E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5) \\ &= 5\times E(X) = 5\times 2,154 = \$10,770 \end{aligned}$$

$$\begin{aligned} Var(X_1+X_2+X_3+X_4+X_5) &= Var(X_1)+Var(X_2)+Var(X_3)+Var(X_4)+Var(X_5) \\ &= 5\times V(X) = 5\times 132^2 = \$^2 87,120 \end{aligned}$$

$$SD(X_1+X_2+X_3+X_4+X_5) = \sqrt{87,120} = \$295.16$$

Continuous Random Variables

Much of what we've discussed so far will not make sense for a continuous random variable (but a lot about how discrete RVs behave is true of continuous RVs as well).

- How do we define the probability of an event when there are (uncountably) infinitely many possible events?
- We won't worry about how to calculate the means and standard deviations of continuous RVs in this course, but we will work with models for continuous RVs where these are specifically defined.
- More on this on Wednesday

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 it needs to be numeric) the random variable takes the value 1 to indicate a success and 0 to indicate a failure.

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

Properties of a Bernoulli Random Variable

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

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

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

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:

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:

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

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

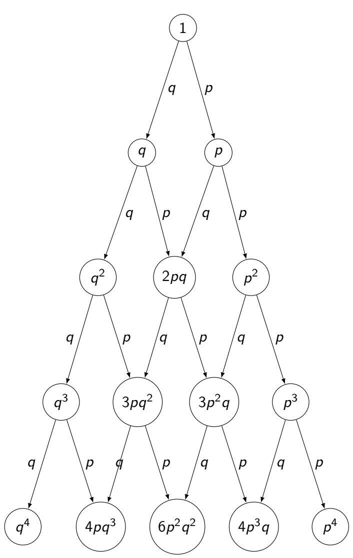
$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

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

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

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

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

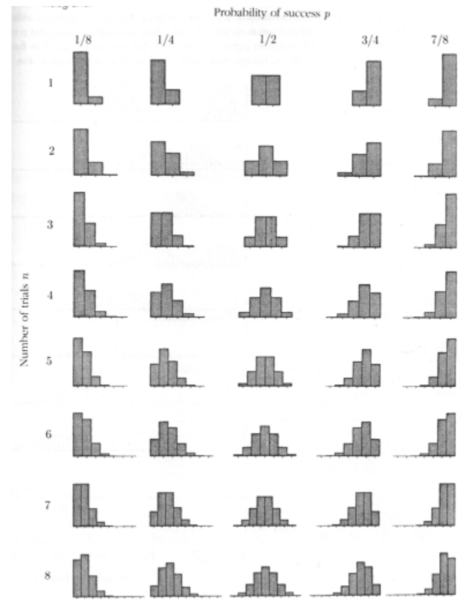
Properties of Binomial RVs

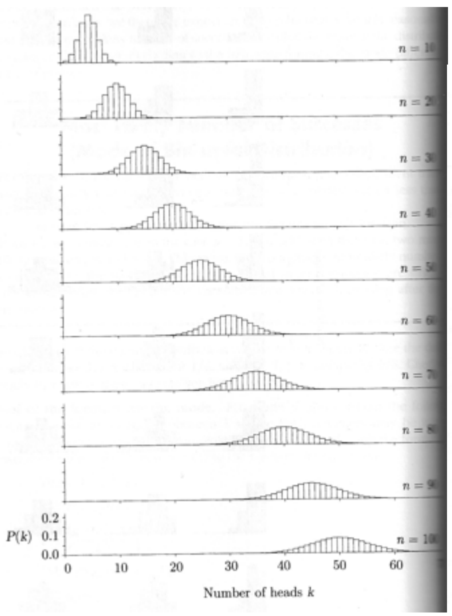
We can actually think of a Binomial random variable as the sum of independent Bernoulli random variables.

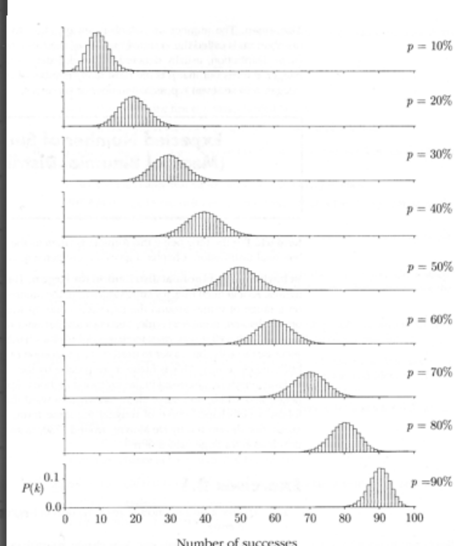
Let $X \sim \text{Binom}(n, p)$ then $X = \sum_{i=1}^{n} Y_i$ where $Y_1, \cdots, Y_n \sim \text{Bern}(p)$.

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i)$$
$$= \sum_{i=1}^{n} p = np$$

$$Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Var(Y_i)$$
$$= \sum_{i=1}^{n} p(1-p) = np(1-p) = npq$$







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.

Clicker Question

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