Last two lectures PowerPoint

Stat 202 - 2015S - W8 - Wednesday (Pg 1)

Continuing with Chapter 4

Review of Chapter 4

A random variable is a rule (i.e. function) that assigns a number to each outcome of a random phenomenon

eg

TTT → 0
HTT → 1
THT → 1
HHT → 2

Function X is number of heads in 3 coin tosses

The probability distribution is ~~is~~ described by a table that lists both the possible values of X and their probability

Value of X	x_1	x_2	x_3	...	x_k
Probability	p_1	p_2	p_3	...	p_k

For coin toss example what is this table

What are possible

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

New

The mean ~~x~~ of a random variable

The mean $\bar{x}$ of a set of observations is their ordinary ~~value~~ average

The mean of a random variable is also an average of possible values of X but it takes into account that not all values are equally likely

Values of X	x_1	x_2	...	x_K
Probability	p_1	p_2	...	p_K

Mean of $X = x_1p_1 + x_2p_2 + \cdots + x_Kp_K$
$= \sum x_ip_i$

Find mean of number of heads in 3 tosses of a coin

Values of X	0	1	2	3
Probabilities	1/8	3/8	3/8	1/8

$0 \cdot 1/8 + 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8$

$= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = 1.5$

Law of Large Numbers

If the experiment (random phenonomenon) is repeated many times and each time the value of X (the random variable) is observed then

$$\text{mean of } X \approx \underbrace{\text{average of observations}}$$

* sometimes called estimate of mean of X

The approximation gets better and better with more and more observations

What's the big ~~bo~~ deal?

Suppose your random phenomenon is ~~sample~~ pick a person at random from population and the measure of their height is the random variable

Suppose you want to know

The mean of this random variable: the mean height of all people in population. But suppose there are ~~too~~ many people to measure every one. So you pick n people at random (sample) and measure their heights estimate mean with the average

$$\text{estimate} = \frac{1}{n} \sum X_i$$

X_i ← heights of people in sample

n ← number of people in sample.

The ~~st~~ law of large numbers tells you that as the sample size becomes large, the estimate grows ~~arbitrarily~~ accurate with statistical certainty

Rules for means:

~~A~~ The values of a random variable are numbers

- We can add them
- We can subtract them
- We can transform them

Thus - We can add or subtract two or more random variables.

Example

Consider the following random phenomenon: throw two dices, a red one and a blue one

Let X be number showing on red
Let Y be number showing on blue

Often in many games we are interested in a third random variable

$$Z = X + Y$$

Some notation

The mean of X is written μ_X
" Y " μ_Y
" Z " μ_Z

Formula

$$\mu_Z = \mu_X + \mu_Y$$

Also written as

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Lets do the dice problem

Values of One dice:	1	2	3	4	5	6
Probabilities	1/6	1/6	1/6	1/6	1/6	1/6

$$\mu_X = \mu_Y = \frac{1}{6}\cdot 1 + \tfrac{1}{6}\cdot 2 + \tfrac{1}{6}\cdot 3 + \tfrac{1}{6}\cdot 4 + \tfrac{1}{6}\cdot 5 + \tfrac{1}{6}\cdot 6$$

$$= \frac{1}{6}(1+2+3+4+5+6) = \frac{1}{6}\,7\cdot 3 = \tfrac{7}{2}$$

$$\mu_Z = \mu_{X+Y} = \tfrac{7}{2} + \tfrac{7}{2} = 7$$

Pg 6

Likewise if you do a linear transformation

in book

$$\mu_{a+bX} = a + b\mu_X$$

Difference

these formulas aren't in book but helpfull later

$$\mu_{X-Y} = \mu_X - \mu_Y$$

Linear Combination

$$\mu_{b_1 X_1 + b_2 X_2 + \cdots + b_n X_n}$$

$$= b_1 \mu_{X_1} + b_2 \mu_{X_2} + \cdots + b_n \mu_{X_n}$$

Linear transformation

Random phenomenon: pick a refrigerator at random from a population, measure its average temperature

34°F 37°F 35°F 38°F

average 36°F

Now convert to celcius

linear transformation equation says its the same if you convert obs first then take mean or just convert the mean, Variance of a random variable

Also computed from probability table

Value of X	x_1	x_2	x_3	...	x_K
Probability	p_1	p_2	p_3	...	p_K

$$\sigma_X^2 = (x_1-\mu_X)^2 p_1 + (x_2-\mu_X)^2 p_2 + \dots + (x_K-\mu_X)^2 p_K$$

$$= \sum (x_i-\mu_X)^2 p_i$$

Example

Example toss 3 coins, count # heads

Values of X	0	1	2	3
Probability	1/8	1/8	1/8	1/8

$$\sigma_x^2 = \left(0-\tfrac{3}{2}\right)^2 \cdot \tfrac{1}{8} + \left(1-\tfrac{3}{2}\right)^2 \cdot \tfrac{3}{8}$$

$$+ \left(2-\tfrac{3}{2}\right)^2 \cdot \tfrac{3}{8} + \left(3-\tfrac{3}{2}\right)^2 \tfrac{1}{8}$$

$$= \frac{6}{4}\cdot\frac{1}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{6}{4}\cdot\frac{1}{8}$$

$$= \frac{1}{32}\left(6+3+3+6\right) = \frac{9}{16}$$

The standard deviation is the square root of the variance

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\tfrac{9}{16}} = \frac{3}{4}$$

Rules for variances

Linear transformation

$$\sigma^2_{a+bX} = b^2 \sigma^2_X$$

(adding a constant doesnt change spread

Consequently $\sigma_{a+bx} = b\sigma_x$

Linear ~~tra~~ combination (Independent)

$$\sigma^2_{b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_nX_n}$$
$$= b_1^2\sigma^2_{X_1} + b_2^2\sigma^2_{X_2} + \cdots + b_n^2\sigma^2_{X_n}$$

Sum (Independent)

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

Difference (Independent)

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

not a mistake $b=-1$

$b^2=1$

General addition Rule (not independent)
X and Y have correlation ρ – analogous to correlation in data

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$