Lecture 1

Start by Reviewing smartsite info.

Content of the Course

Essentially a study of random variables (will use R.V. for short)

An important part of this will be to develop an understanding of R.V. dependence.

Dependence btwn R.V.s is really the workhorse of statistics.

e.g. Let X = Apple stock price tomorrow

Y = Apple stock price on Friday.

Both X & Y are R.V.s.

They are dependent.

If X is unusually low then Y will likely be low too.

The goal will be to characterize how X & Y interact so we compute things like

$E(Y|X)$ = "expected value of Y given X"

= "best prediction of Y given you observe X".

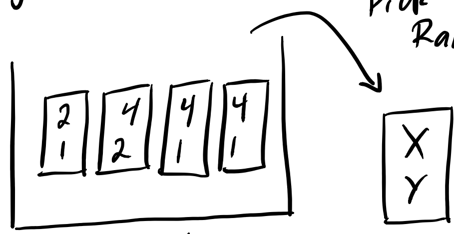
Box Models

I use these a lot to illustrate & understand randomness.

A Box Model is a simple way to represent randomness & R.V.s

e.g.

Pick 1 ticket at Random.



let X be the top number & Y be the bottom number.

4 tickets.
2 numbers on each ticket.

X & Y are R.V.s

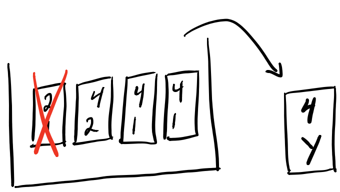
Probability calculations are very easy with Box Models.

$P(X=4)$ = "the probability that X is 4"

$= \dfrac{3}{4}$ ← 3 tickets with $X=4$ / ← 4 tickets total.

$P(X=4 \text{ and } Y=1) = \dfrac{2}{4}$

$P(Y=1 | X=4)$ = "if someone told you that $X=4$ but not what Y is what is the chance $Y=1$".



$= \dfrac{2}{3}$ ← 3 tickets with $X=4$

2 of those 3 tickets have $Y=1$.

Notation

X, Y, Z, etc will generally denote R.V.s

$a, b, c, x, y \ldots$ will denote fixed non-random numbers.

$X_1, X_2, \ldots, X_n$ denote n R.V.s

$x_1, x_2, \ldots, x_n$ denotes n non-random numbers.

$\sum_{i=1}^{n} x_i$ is shorthand notation for $x_1 + x_2 + \cdots + x_n$

e.g. Let $x_1 = 1$
$x_2 = 5$
$x_3 = 7$
$x_4 = 2$

$\therefore \sum_{i=1}^{4} x_i = 1 + 5 + 7 + 2$

Sum the x_i's as i ranges from 1 to 4.

Note: $\sum_{k=1}^{n} x_k = \sum_{i=1}^{n} x_i$

This number (n) tells you when to stop summing

This number ($i=1$) tells you when to start.

$\therefore \sum_{k=1}^{3} x_k = x_1 + x_2 + x_3 = 1 + 5 + 7$

$\sum_{k=2}^{3} x_k = x_2 + x_3 = 5 + 7$

You can also do stuff like this

$$\pi \sum_{i=1}^{3} \sin(x_i) = \pi\left(\sin(x_1) + \sin(x_2) + \sin(x_4)\right)$$
$$= \pi\left(\sin(1) + \sin(5) + \sin(7)\right)$$

The following will be used a lot

$$\frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + \cdots + x_n\right)$$
$$= \bar{x}$$

$\leftarrow$ the regular average of $x_1, x_2, \ldots, x_n$.

Linear properties of Σ

We all know:

$$5(7+9) = 5 \cdot 7 + 5 \cdot 9$$

& $(7+2) + (9+12) = 7 + 2 + 9 + 12$

This is abstractly expressed with Σ as follows

$$a\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} (a x_i)$$

$$\sum_{i=1}^{n} (x_i + y_i) = \left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n} y_i\right)$$

Putting these together

$$\sum_{i=1}^{n} (a x_i + b y_i) = a\left(\sum_{i=1}^{n} x_i\right) + b\left(\sum_{i=1}^{n} y_i\right)$$

Useful when changing units

(5)

e.g. Consider $x_1, x_2, \dots, x_{30}$

Suppose $\bar{x} = -10$ (i.e. the ave of x_i is -10)

For each x_i define y_i as follows

$$y_i = \frac{5}{9}(x_i - 32)$$

Find $\bar{y}$ (i.e. the average of the y_i's)

$$\bar{y} = \frac{1}{30}\left(\sum_{i=1}^{30} y_i\right)$$

$$= \frac{1}{30}\left(\sum_{i=1}^{30} \frac{5}{9}(x_i - 32)\right)$$

$$= \frac{1}{30}\frac{5}{9}\left(\sum_{i=1}^{30}(x_i - 32)\right)$$

$$= \frac{1}{30}\frac{5}{9}\left(\left(\sum_{i=1}^{30} x_i\right) + \underbrace{\sum_{i=1}^{30}(-32)}_{\text{"sum } (-32) \text{ 30 times"}}\right)$$

$$= \frac{1}{30}\frac{5}{9}\left(\left(\sum_{i=1}^{30} x_i\right) - 30 \cdot 32\right)$$

$$= \frac{5}{9}\left(\underbrace{\left(\frac{1}{30}\sum_{i=1}^{30} x_i\right)}_{= \bar{x} = -10} - \frac{1}{\cancel{30}} \cdot \cancel{30} \cdot 32\right)$$

$$= \frac{5}{9}(-10 - 32)$$

(6)

In lab (i.e. discussion.) they will show how to derive:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - (\bar{x})^2.$$