5/15/19

RG Section 7

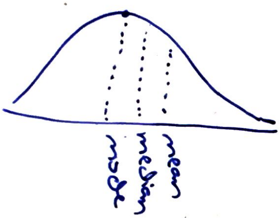
Characteristics of a distribution

- Center
- Spread
- Shape

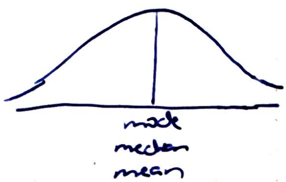
in order of how much is used

Measures of Center

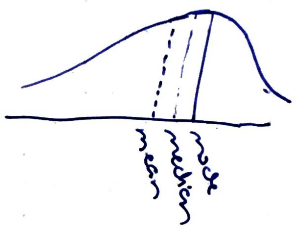
Positive skewness

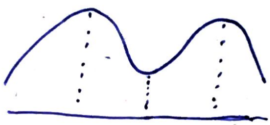


Symmetric



Negative skewness





Examples in real life

- Mode: Talking about your sales with your boss
- Median: Discussing Gross Domestic Product of a country
- Mean: used most often — so often it has its own symbol E

Ex: $E[x]$ Easiest: Bernoulli (p)

PMF: $f_X(x) = p^x(1-p)^{1-x}$

$S_X = \{0, 1\}$

$$E[X] = \sum_{x \in S_X} x f_X(x)$$ The weight is the probabilities

$$= (0)(1-p) + (1)(p) = p \text{ : mean}$$

Poisson(λ)

PMF: $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\quad S_X = \{0, 1, \ldots\}$

$$E[X] = \sum_{x \in S_X} x f_X(x)$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

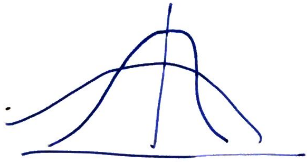
$\sum_{x \in S_X} f_X(x) = 1$ Adding all probabilities with PMF

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1$$

$$\Rightarrow \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$$

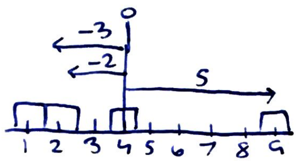
$$E[X] = \sum_{x=0}^{\infty} \frac{x e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{x e^{-\lambda}\lambda^x}{x!} = \sum_{x \geq 1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!}$$

$$= e^{-\lambda}\lambda \sum_{x=1}^{\infty} \frac{\lambda^{(x-1)}}{(x-1)!} = e^{-\lambda}\lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \cancel{e^{-\lambda}} \lambda \cancel{e^{\lambda}} = \lambda \text{ : mean}$$



2 normal distributions with the same mean

Measures of Spread



MAD
Mean Absolute Difference

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \xrightarrow{\text{difference to the mean}} \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \rightarrow \begin{bmatrix} |y_1 - \bar{y}| \\ |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix} = \sum_{i=1}^{n} |y_i - \bar{y}| / n$$

$$\downarrow$$

$$\begin{bmatrix} (y_1 - \bar{y})^2 \\ (y_2 - \bar{y})^2 \\ \vdots \\ (y_n - \bar{y})^2 \end{bmatrix} = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n$$

variance

This one is more likely to be ~~#s~~ far from the mean

$$\downarrow$$

$$\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n}$$

standard deviation ← most common because of Central Limit Theorem

R code!

dpois(n, lambda) : PMF