Discussion section, week of 18-22 May 2020

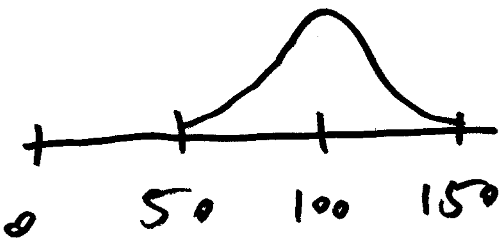
Please read JS ch. 4

STAT 131
18 May 20
(discussion section)

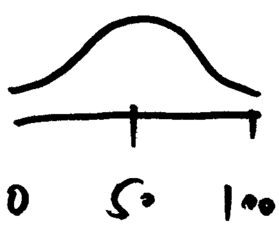
①

similar: same shape, same spread

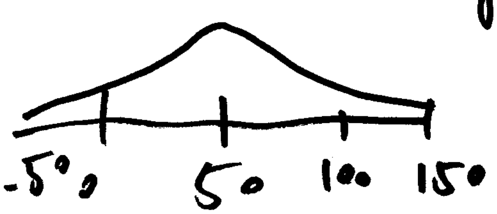
different center



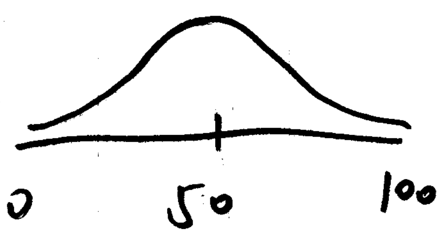
similar: same center, same shape



different spread



same center, same spread, different shape



measures of summary for PMFs & PDFs

- center
- spread
- shape

} graphically, numerically

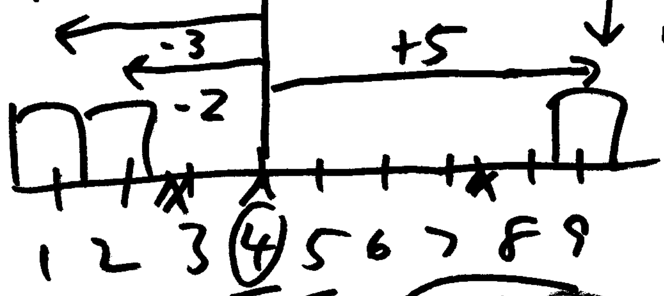
~~pop~~ sample

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ $n = 3$

$y_i \rightarrow$ 2 median

$4 = \bar{y}$

mean



outlier in the right tail

$\bar{y}$ ← center of mass = balance point

$(y_i - \bar{y})$ deviations from mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

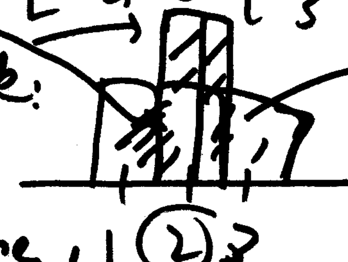
$$= \sum_{i=1}^{n} \left(\frac{1}{n}\right) y_i$$

$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}$

50%

50%

mode: point of highest prob. concentration



mode
median
mean (data)

expectation or expected value of r.V.

$Y \leftarrow \begin{Bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{Bmatrix} \rightarrow$

y	rel freq
1	0.25
2	0.5
3	0.25

PMF

$P(Y = y)$

$$E(Y) = \sum_{\text{all } y \text{ in } S_Y} y \cdot P(Y = y)$$

expected value of random variable Y

Y discrete (PMF) $\xrightarrow{\text{support } S_Y}$ $f_Y(y) = P(Y = y)$

$\rightarrow E(Y) = \sum_{\text{all } y \in S_Y} y \cdot \overbrace{f_Y(y)}^{\text{PMF}}$

$= \sum_{y \in S_Y} y \cdot P(Y = y)$

Y continuous (PDF) $\xrightarrow{\text{support } S_Y}$ $f_Y(y)$

$\rightarrow E(Y) = \int_{y \in S_Y} y \cdot \overbrace{f_Y(y)}^{\text{PDF}}\, dy$

---

ex. T-S case study

Y = # of T-S babies in family of $n = 5$ kids, both parents carriers, so that

$\cancel{\#}$ $p = P(\text{T-S on any single kid}) = \frac{1}{4}$

$$(Y \mid n, p) \sim \text{Binomial}(\overset{5}{n}, p)$$

discrete

$$E(Y) = \sum_{y \in S_Y} y \cdot P(Y = y) = \sum_{y \in S_Y} y \cdot \overset{\text{PMF}}{f_Y(y)}$$

$$S_Y = \{0, 1, \dots, 5\} = \{0, 1, \dots, n\}$$

$$f_Y(y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, \dots, n \\ 0 & \text{else} \end{cases}$$

$$E(Y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

Binomial$(5, \frac{1}{4})$

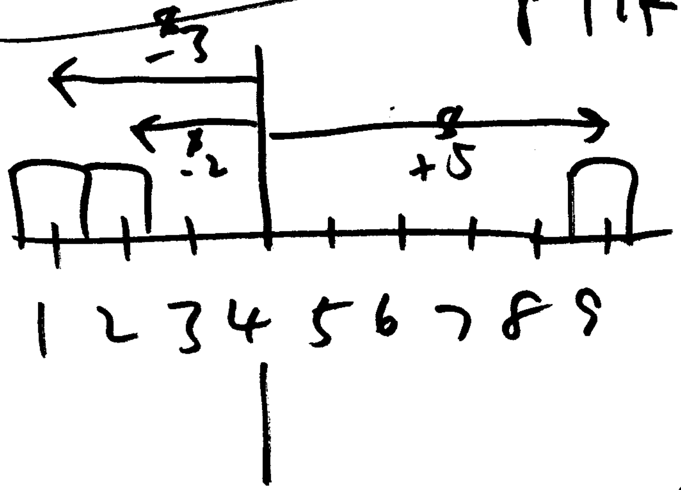
mode 1

~~median 1~~

long right-hand tail

(not symmetric: skewed: positive **skewness**

PMF of Y



$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \to \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}$$

mean 0

$\to \mu = 4$

expected value (mean)

$$Y \to (Y - \mu_Y)$$

$(-2)^2$
$(-3)^2$
$(+5)^2$

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}) = 0 \; (!)$$

~~(deprecated)~~ $\doteq \cancel{8}\, 3.3$ ✓

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}|$$

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \leftarrow s^2 \;✓$$

data | mean | R.V. (PMF)

$$\frac{1}{n}\sum y_i = \bar{y} \qquad E(Y) = \sum_{y \in S_Y} y \cdot f_Y(y)$$

$$= \sum \left(\cancel{\tfrac{1}{n}}\right) y_i \left(\tfrac{1}{n}\right)$$

expected value

(PMF)

$$\frac{1}{n}\sum (y_i - \bar{y})^2 \quad \text{data} \qquad V(Y) = \sum_{y \in S_Y} [y - E(Y)]^2 \cdot f_Y(y)$$

$$= \sum (y_i - \bar{y})^2 \cdot \left(\tfrac{1}{n}\right)$$

variance of Y — μ_Y

Var : DS

if Y is cont: $V(Y) = \int_{y \in S_Y} (y - \mu_Y)^2 \cdot f_Y(y)\,dy$

$= E(Y - \mu_Y)^2$

---

standard deviation of Y =

$$SD(Y) = \sqrt{V(Y)}$$

right data scale — important theory

$(Y \mid n, p) \sim \text{Binomial}(n, p)$

$$\rightarrow E(Y) = np$$

$$V(Y) = -n(p-1)p$$

$$= np(1-p)$$

$$SD(Y) = \sqrt{np(1-p)}$$
