AMS 131 30 Apr 19 ①

this time: PDFs & CDFs &
next time: inverse CDFs

(see syllabus for reading)

distributional shapes

ex. Binomial(5, $\frac{1}{4}$) distribution (Tay-Sachs)
$n \rightarrow$ $\leftarrow p$ Y:

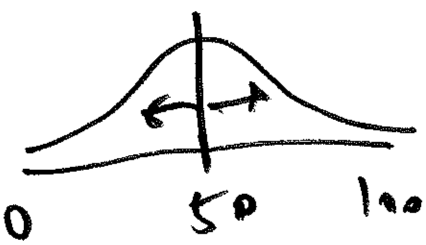
$\uparrow p(Y = y)$

this is a PMF

$f_Y(y)$ $\quad y \rightarrow$ (discrete)

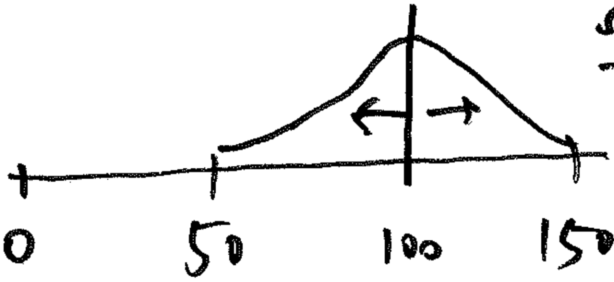
How characterize PMFs & PDFs?

tails

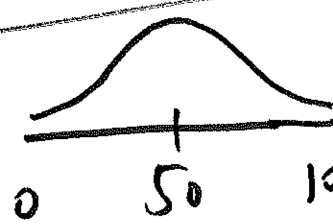


same spread

same shape

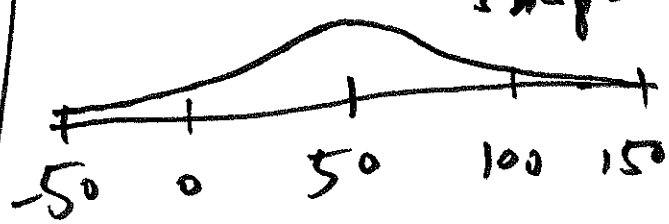


different center

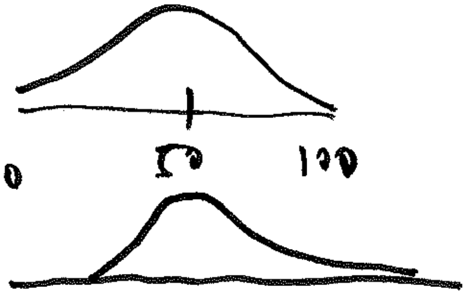


same center

same shape



different spread



same center

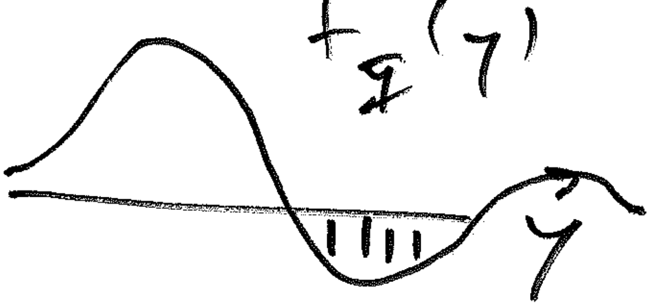
same spread

different shape

$$P\left(a-\tfrac{\epsilon}{2} \le X \le a+\tfrac{\epsilon}{2}\right) \underset{\text{prob.}}{\doteq} \cancel{\epsilon}\, \boxed{f_X(a)}$$

ϵ per units along $\mathbb{R}$

density = concentration of probability



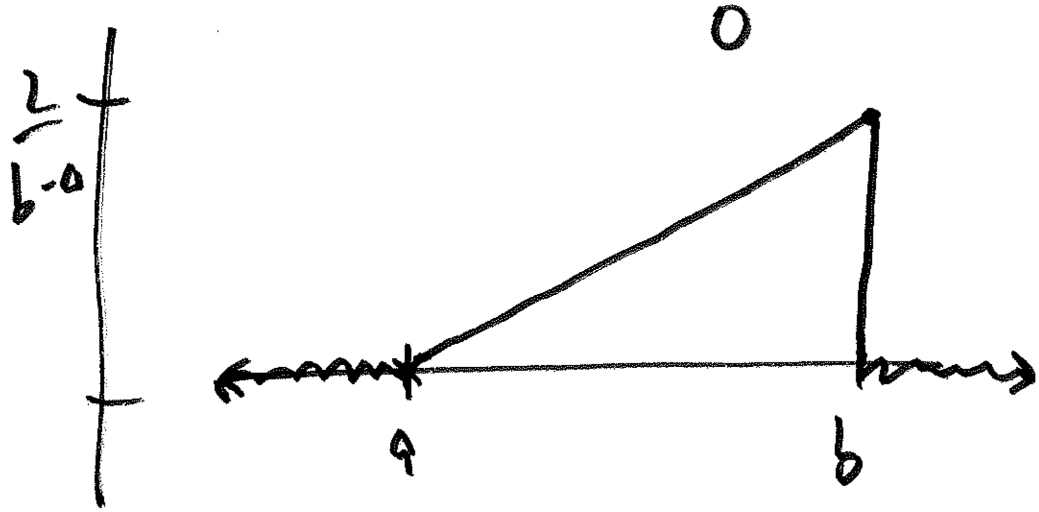
① $f_X(y) \ge 0$

② $\int_{-\infty}^{\infty} f_X(y) = 1$

$$\int_a^b c\,\frac{x-a}{b-a} = \cancel{[illegible]}$$

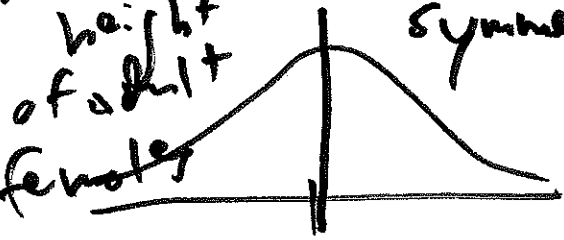
$$\frac{x}{b-a} = \frac{c\,(\cancel{b-a})}{\cancel{x}}$$

$$f_Y(y) = \begin{cases} \dfrac{2(y-a)}{(b-a)^2} & \text{for } a \le y \le b \\ 0 & \text{else} \end{cases}$$



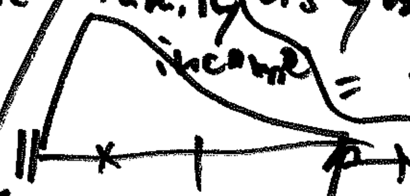
symmetric | family | asymmetric = skewed

height of adult females



point of symmetric

income



barrier

Bill Gates

positive skewness

long right-hand tail

scores on THT 1 (%)

negative skew

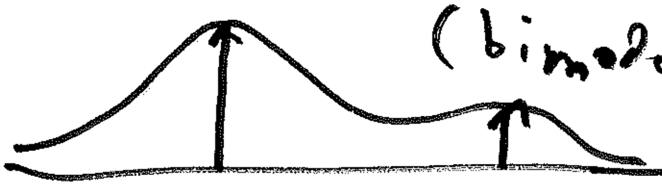
long left-hand tail

---

unimodal

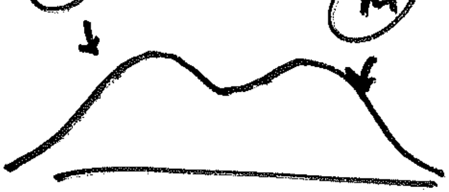
the mode

multimodal (bimodal)



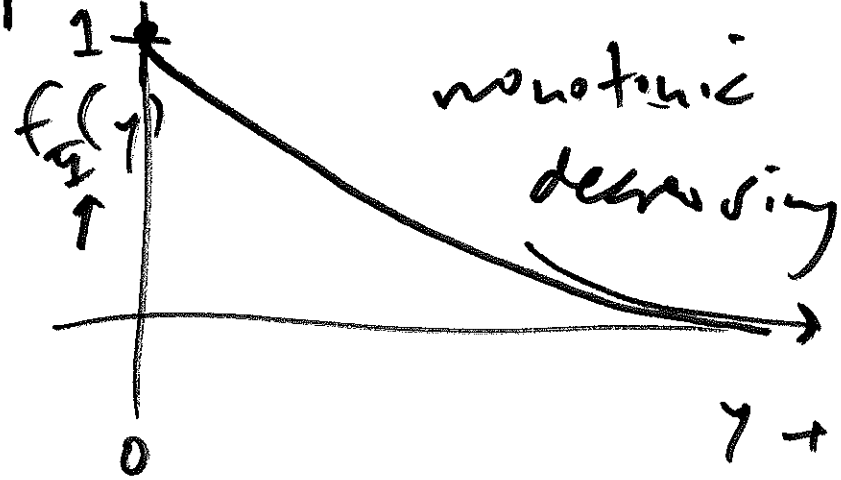
a mode

another mode

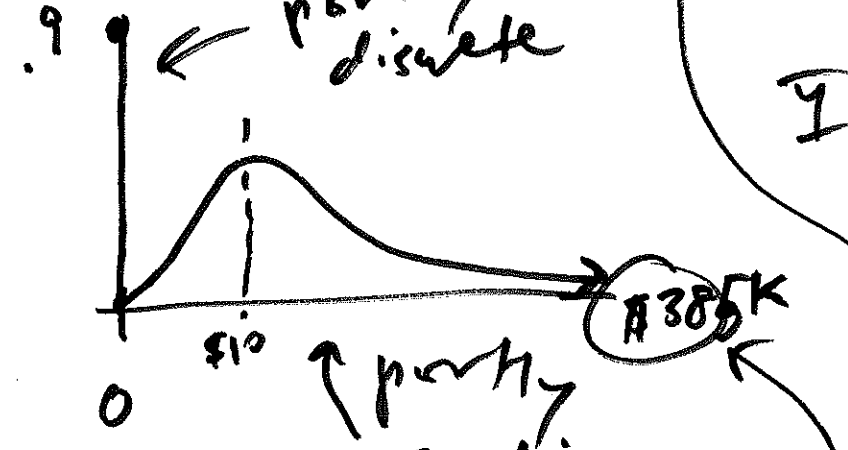


ht. of all ANS131 students

(mixture distribution)



$$f_Y(y) = \begin{cases} \dfrac{1}{(1+y)^2} & \text{for } y > 0 \\ 0 & \text{else} \end{cases}$$

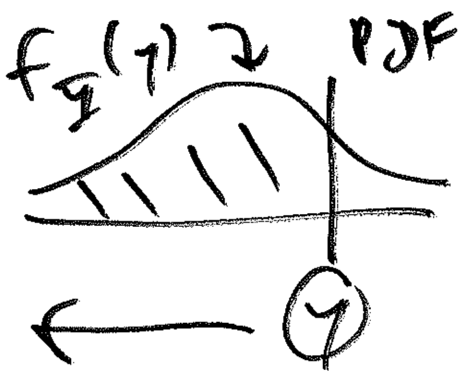


name? "mixed" dist.? ✓

eBay

Y = GMB (gross merchandise bought) in $(0, T)$

e.g. 2 weeks



CDF

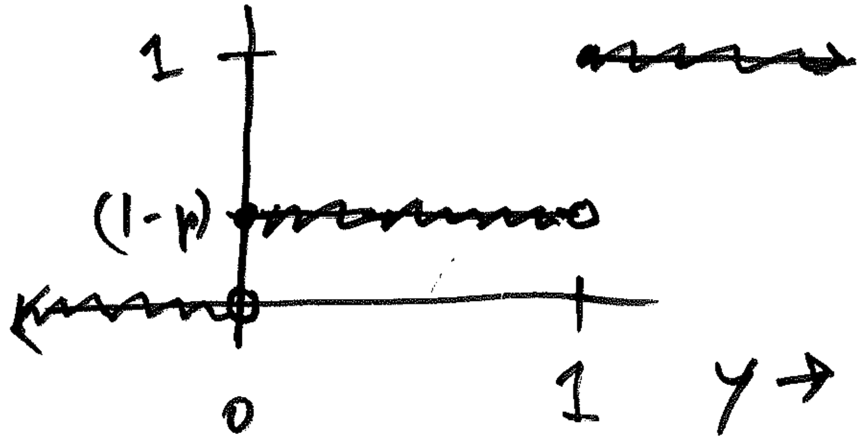
cont.

$$F_Y(y) = P(Y \le y)$$

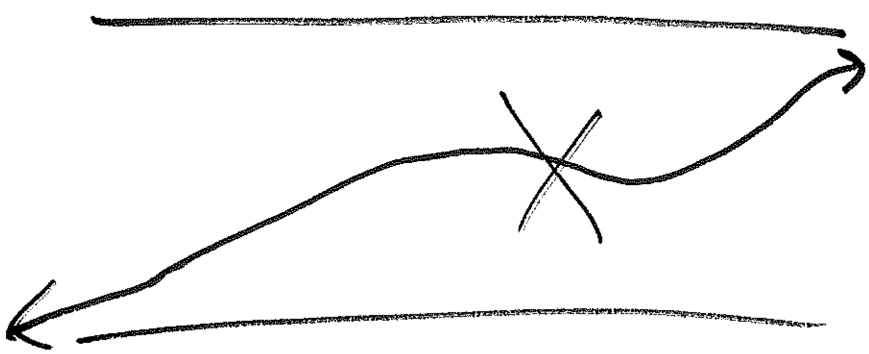
$$= \int_{-\infty}^{y} f_Y(t)\, dt$$

dummy variable of integration

$0 \le F_Y(y) \le 1$



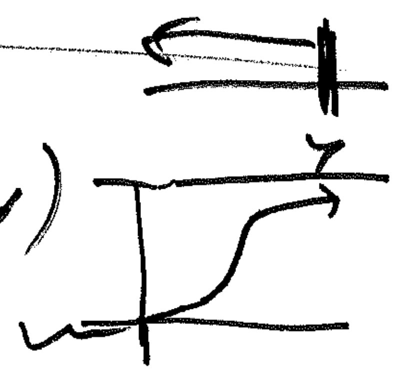
$$F_{\underline{x}}(y) = P(\underline{x} \le y)$$



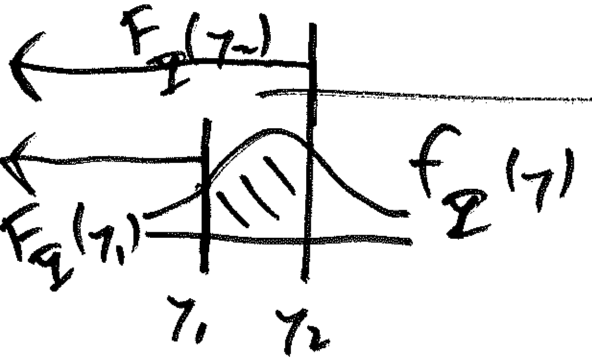
$y \rightarrow$

$F_{\underline{x}}(y)$ has to be non-decreasing

$$F_{\underline{x}}(y) = P(\underline{x} \le y)$$



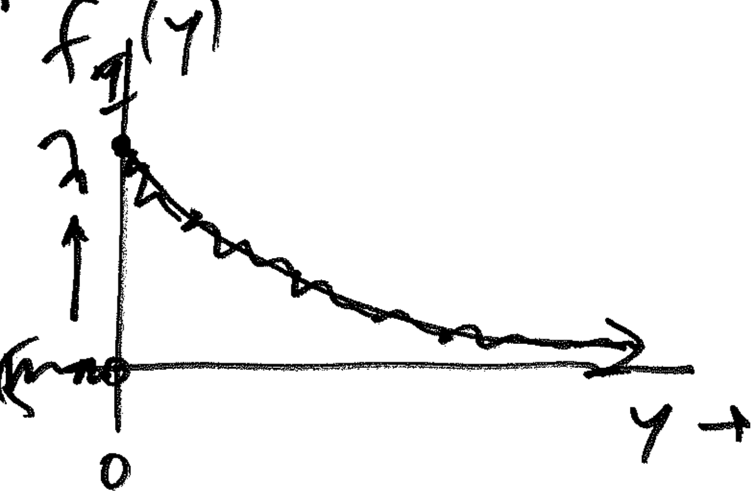
$$\rightarrow \quad P(\underline{x} > y) = 1 - F_{\underline{x}}(y)$$



cont.

$$P(y_1 < \underline{x} \le y_2) = \int_{y_1}^{y_2} f_{\underline{x}}(y)\,dy$$

$$F_{\underline{x}}(y_2) - F_{\underline{x}}(y_1)$$

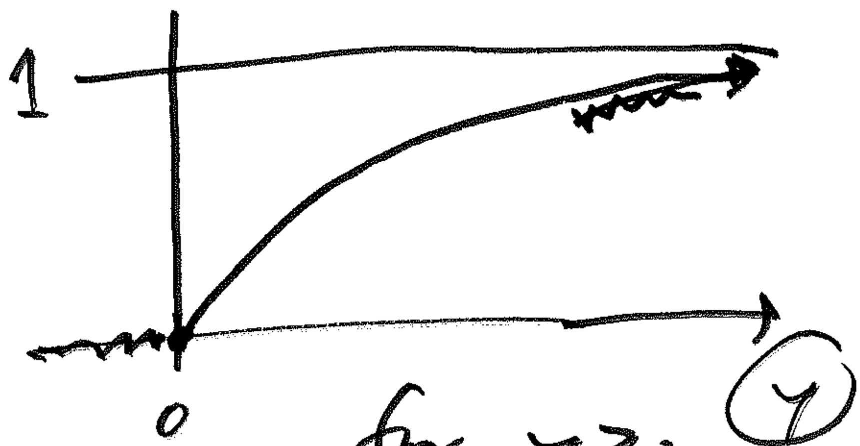


$(\lambda > 0)$

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{for } y \geq 0 \\ 0 & \text{else} \end{cases}$$

CDF of Y:

$$F_Y(y) = \begin{cases} 0 & \text{for } y < 0 \\ 1 - e^{-\lambda y} & y \geq 0 \\ (1 & \text{as } y \uparrow \infty) \end{cases}$$

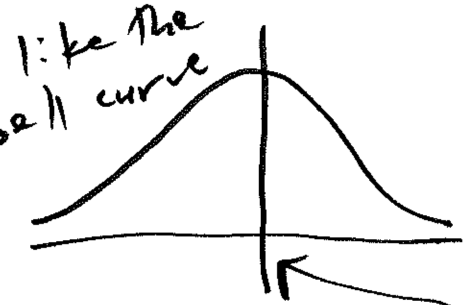


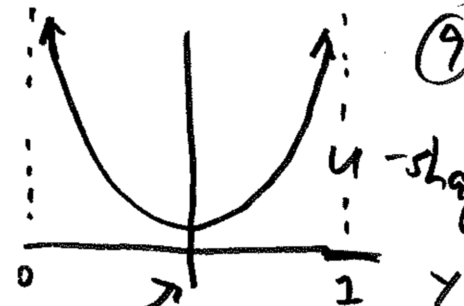
for $y \geq 0$

$$F_Y(y) = \int_0^y \lambda e^{-\lambda t}\, dt$$

$$= 1 - e^{-\lambda y}$$

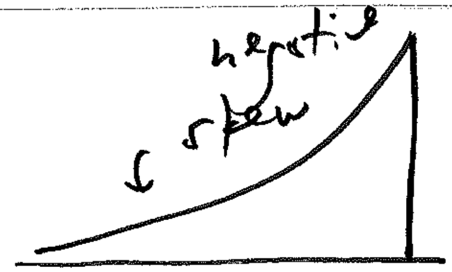
Some common distributional shapes





both of these are symmetric about a point of symmetry

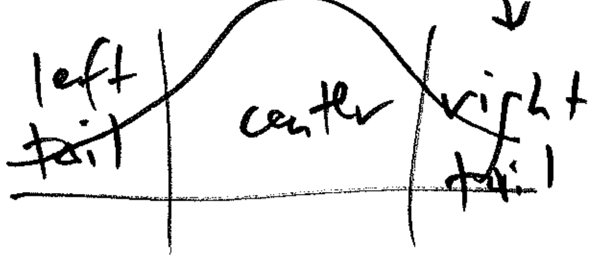
reverse-J shaped



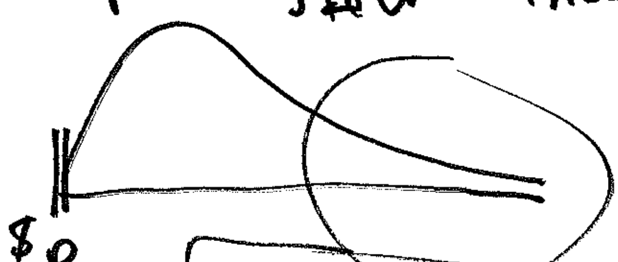
J-shaped

these are asymmetric or skewed

tails



positive skew



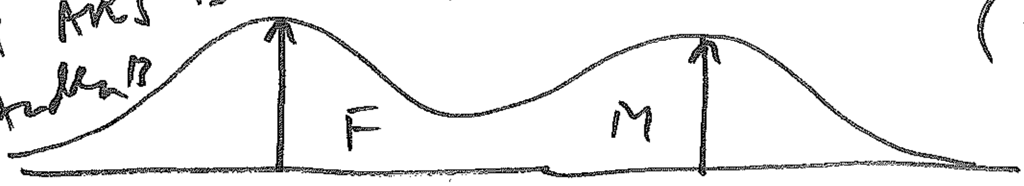
family income in U.S. in 2015

long right-hand tail

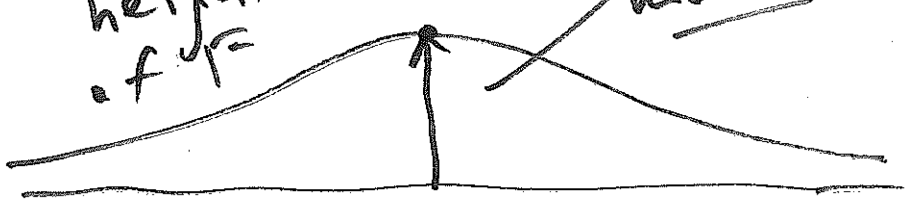
% correct on take home final

long left-hand tail

height of all AMS 131 students — symmetric bimodal (multimodes)



height of ppl



mode = highest point on dist.

symmetric unimodal

---

eBay) Randomized controlled trial (RCT),

T = { slightly larger pictures }

C = { standard size }

Y_i in T = (total amount in \$ of stuff bought in 4 weeks)

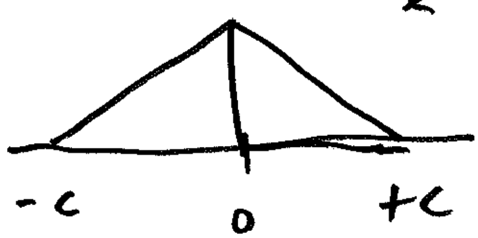
90% $Y_i = \$0$

10% $\$0 \le Y_i \le \$160{,}000$

$$P\left(a - \frac{\epsilon}{2} \le \underline{Y} \le a + \frac{\epsilon}{2}\right) \doteq \epsilon \cdot f_{\underline{Y}}(a)$$

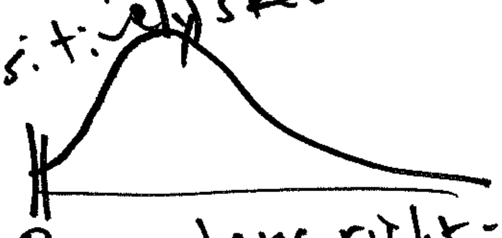
$$f_{\underline{Y}}(a) \doteq \boxed{\frac{P\left(a - \frac{\epsilon}{2} \le \underline{Y} \le a + \frac{\epsilon}{2}\right)}{\epsilon}}$$

$$= \lim_{\epsilon \to 0}$$



pdf for astronomy (telescope) measurement errors

(asymmetric)
(positively) skewed

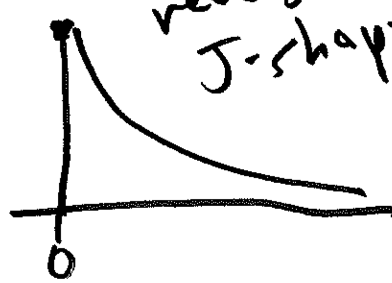


barrier in left tail

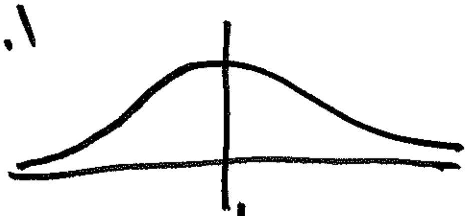
long right-hand tail

Poisson ($\lambda = 3$)

reverse J-shape



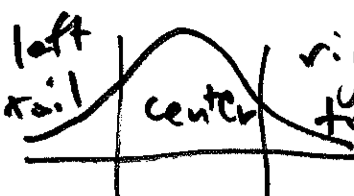
Poisson ($\lambda = 0.5$)



point of symmetry

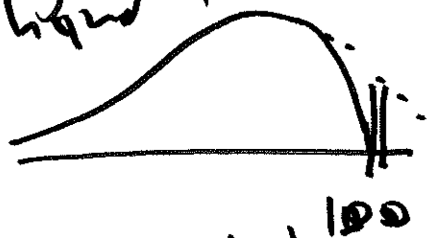
symmetric

Poisson ($\lambda = 50$)



tails

long left-hand tail

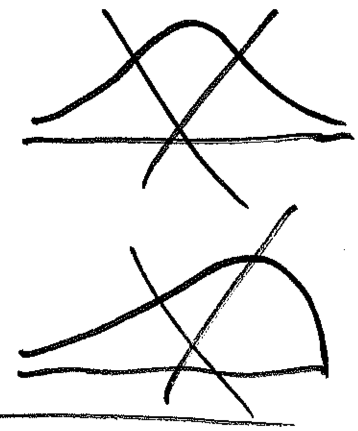


% correct on T-H test 1

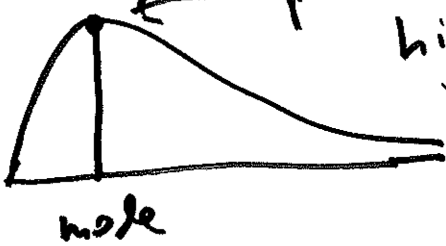
(negatively) skewed

barrier in right tail

u.s. family income in 2016



unimodal



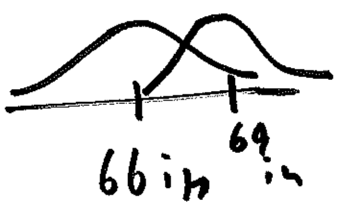
point of highest density

bimodal

(n = 132) height of AMS 131 students

(multimodal)

mixture of F & M



randomized controlled trials in eCommerce

= A/B test

T treatment (bigger picture) $\quad$ C control

n_T = # people in T

n_C = ______ C

$n_T \doteq$ 12 million

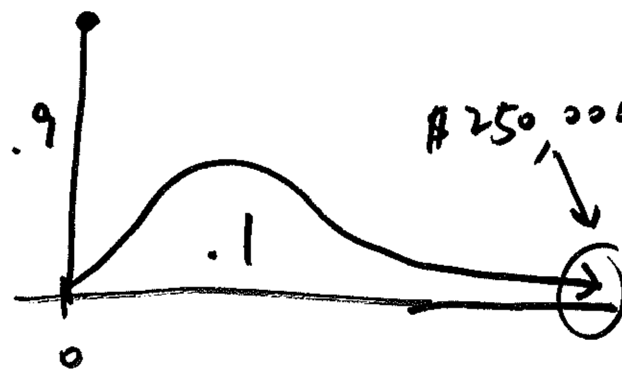
$n_C \doteq$ 12 million

$k =$ (4 weeks)

Y_i = (total bought by person i during $k = 4$ weeks)

with prob. p, $(Y_i = 0)$ $\quad$ $\boxed{0.9}$

with prob. $(1 - p)$



Y_i has a mixed discrete-continuous dist.

pmf? no

pdf? no

eBuy