This time: → PDF, CDF, quantiles

next time: → bivariate distributions

read: DS ch. 3

STAT 131 28Apr20 (lecture)

~~office~~ extra 1.5-hour sessions to help with TA 1: Wed 3:30–5pm 29Apr20; Fri; Sat; Sun to be arranged (all Zoom, all recorded)

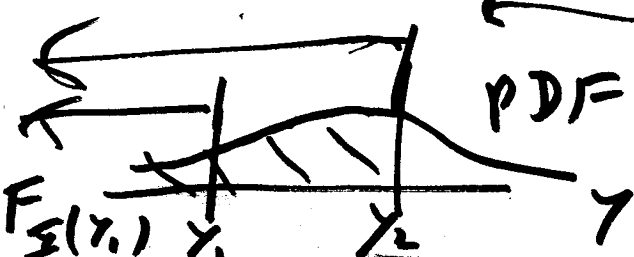
$$F_Y(y) = P(Y \le y) \text{ so}$$

$$P(Y > y) + P(Y \le y) = 1$$

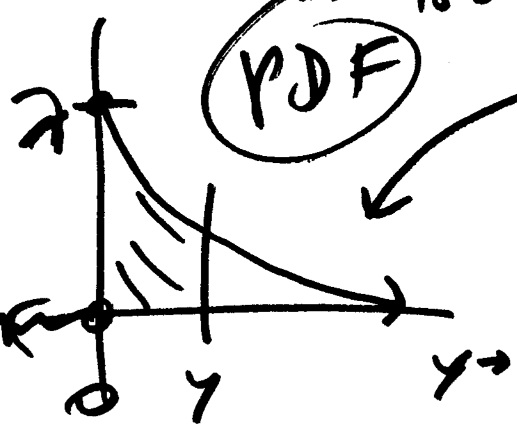
$$P(Y > y) = 1 - P(Y \le y)$$

$$= 1 - F_Y(y)$$

$$P(y_1 < Y \le y_2) = F_Y(y_2) - F_Y(y_1)$$



$f_Y(y)$ decreasing on its support PDF



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

$\mathcal{E}(\lambda)$

Support of rv Y

$(\lambda > 0)$

CDF of $\mathcal{E}(\lambda)$

$$F_Y(y) = \begin{cases} 0 & \text{for } y \le 0 \\ 1-e^{-\lambda y} & y \ge 0 \end{cases}$$

$\boxed{P(Y \le y)}$

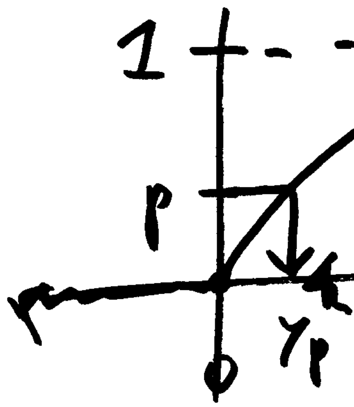
PDF

for $y>0$ $\quad F_Y(y) = \int_0^y \lambda e^{-\lambda t}\, dt$

$$= \left(\cancel{\lambda} \frac{e^{-\lambda t}}{-\cancel{\lambda}} \right)_{t=0}^{t=y} = -\left(e^{-\lambda y} - 1\right)$$

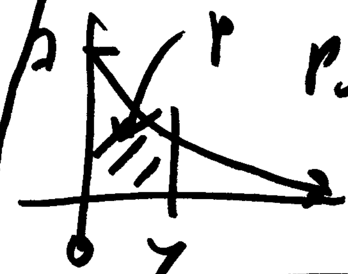
(for $y>0$) $\quad F_Y(y) = 1 - e^{-\lambda y}$

CDF



(increasing on S_Y)

$F_Y(y)$

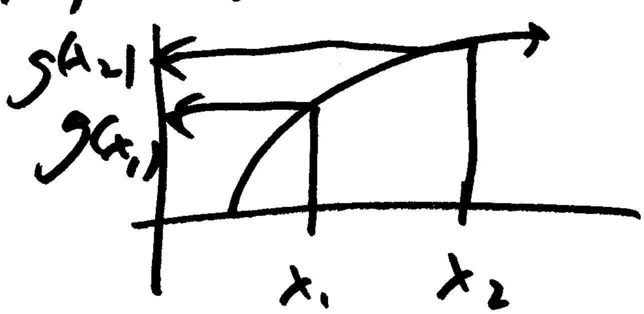


(for $y > 0$)

$$F_Y(y_p) = P(Y \le y_p) = p = 1 - e^{-\lambda y_p}$$

what is the point y_p such that

a function $g(x)$ is (strictly) increasing if



$$x_1 < x_2 \Leftrightarrow g(x_1) < g(x_2)$$

(order-preserving)

and similarly for (strictly) decreasing

$$x_1 < x_2 \rightarrow g(x_1) > g(x_2)$$

to invert $F_Y(y)$ & get $F_Y^{-1}(p)$,

solve

$1 - e^{-\lambda y_p} = p$ for y_p

(ln)

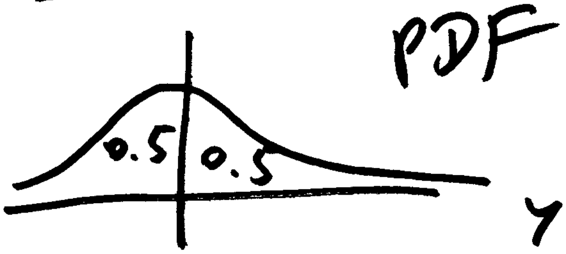
$1 - p = e^{-\lambda y_p} \rightarrow \log(1-p) = -\lambda y_p$

so $y_p = F_Y^{-1}(p) = -\dfrac{\log(1-p)}{\lambda}$

quantile function

for rv Y

PDF of Y (distribution of r.v.)

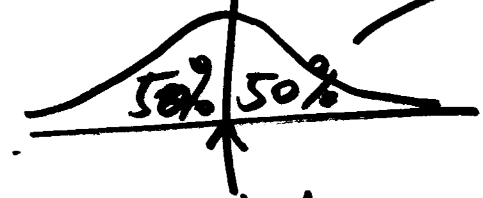


$\tilde{y}$ = median of Y $\quad \tilde{y} = F_Y^{-1}(0.5)$

$P(Y \le \tilde{y}) = F_Y(\tilde{y}) = F_Y\left[F_Y^{-1}(0.5)\right]$
$= 0.5$

shapes of PDFs

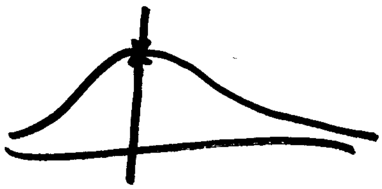
unimodal



this PDF is symmetric

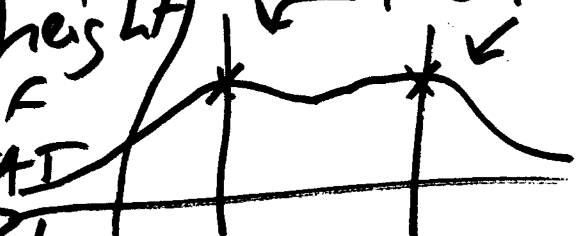
point of symmetry = median = mode (= mean)

(center of gravity) balance point



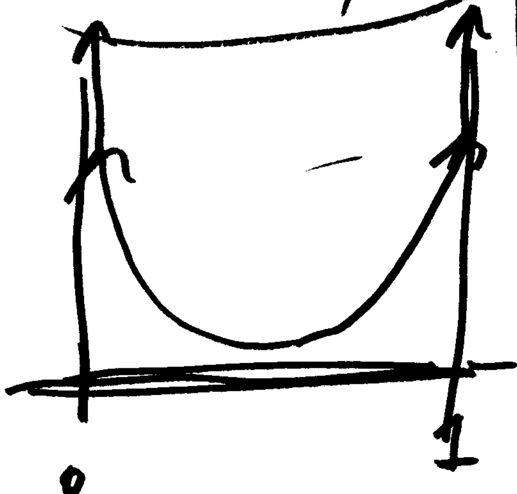
mode = point of highest local density

local modes



height of STAT 131 students

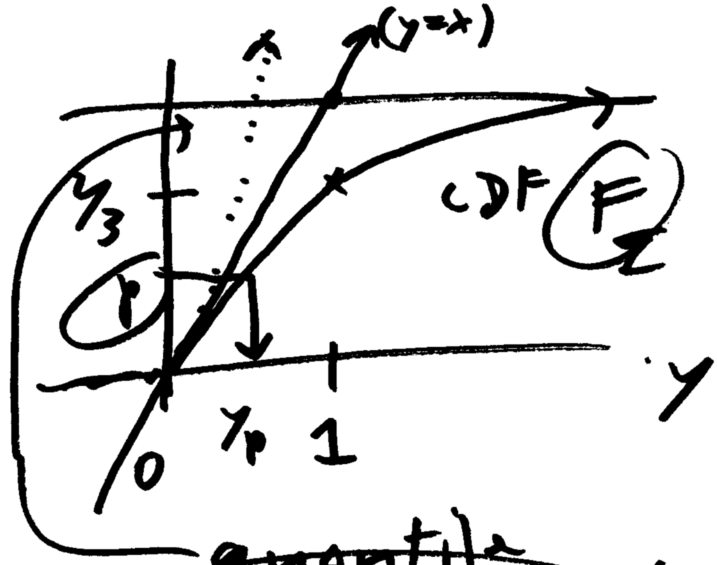
bimodal (multimodal)



quantiles are always in decimal terms, [0, 1]

whereas percentiles are always in percentage terms [0, 100]

median = 0.5 quantile = 50th percentile



(for $y>0$)

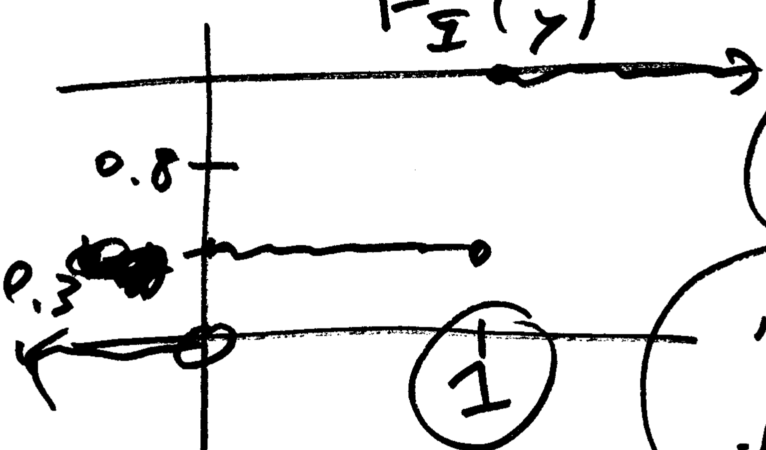
$$F_Z(y) = 1 - e^{-\lambda y}$$

case ($\lambda = 1$)

$$F_Z(y) = 1 - e^{-y}$$

quantile function (F_Z^{-1})

$$F_Z^{-1}(p) = -\frac{\log(1-p)}{\lambda}$$

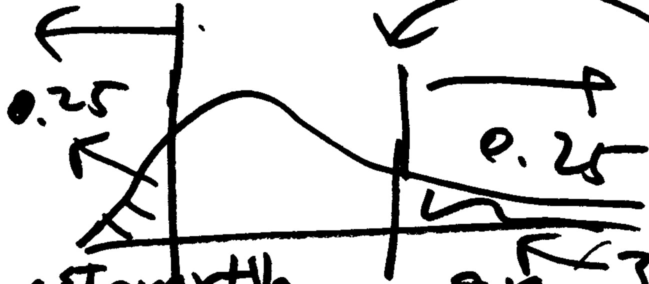


Bernoulli(θ) 0.3

$F_Z^{-1}(p)$ is smallest y value such that $F_Z(y) \geq p$

$$F_Z^{-1}(0.8) = 1$$

$$F_Z^{-1}(0.3) = 0$$



PDF of I

1st quartile 25th percentile = 0.25 quantile

$F_I^{-1}(\underline{\underline{0.25}})$

75th percentile = 0.75 quantile

$F_I^{-1}(0.75)$

3rd quartile

(a) measure of spread: inter-quartile range (IQR).

$$F_I^{-1}(0.75) - F_I^{-1}(0.25)$$

median = 2nd quantile

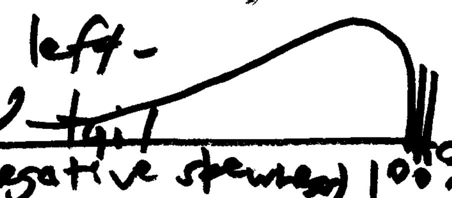
tails

(symmetric)

PDF of I = U.S. annual family income

asymmetric (skewed)

long right-hand tail (positive skewness)

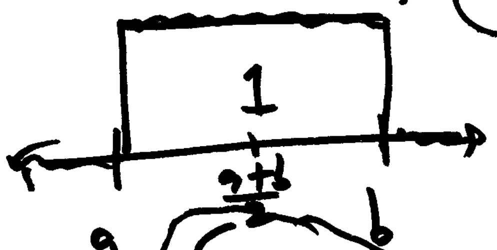


long left-hand tail (negative skewness)

STAT 131 grades

ex. $Y \sim$ Uniform(a, b) $(a<b)$ (8)

(constant density) PDF of Y (continuous)



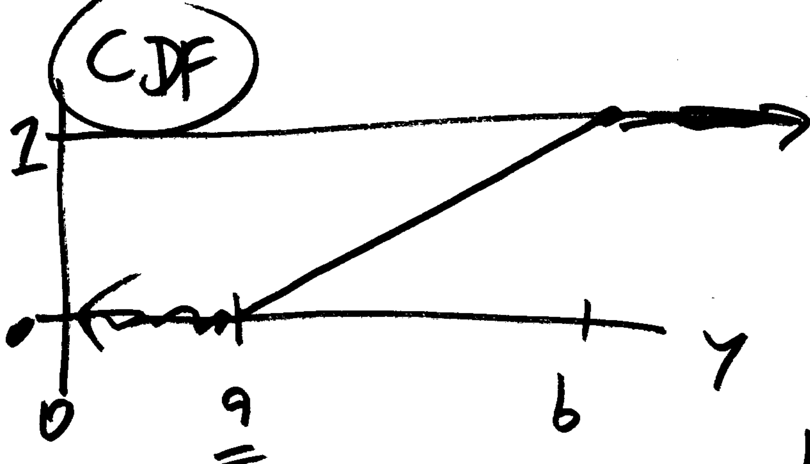
symmetric

$$S' = (a, b)$$

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & \text{for } a<y<b \\ 0 & \text{else} \end{cases}$$

CDF of Y

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0 & y<a \\ \dfrac{y-a}{b-a} & a<y<b \\ 1 & y>b \end{cases}$$



for $a<y<b$

$$F_Y(y) = \int_a^y \frac{1}{b-a}\,dt$$

$$= \left(\frac{t}{b-a}\right)_a^y = \frac{y-a}{b-a}$$

median, of distribution of Y

IQR, need $F_Y^{-1}(p)$

for $a < y < b$ $\quad F_Y(y_p) = \frac{y_p - a}{b - a} = p$

$$y_p = \boxed{a + p(b-a) = F_Y^{-1}(p)}$$

median = 50th percentile

= 0.5 quantile →

$$F_Y^{-1}(0.5) = a + \frac{1}{2}(b-a)$$

$$= \frac{a+b}{2}$$

IQR: $F_X^{-1}(0.75) - F_X^{-1}(0.25)$

$$\left[a + \frac{3}{4}(b-a)\right] - \left[a + \frac{1}{4}(b-a)\right]$$

$$= \frac{b-a}{2}$$