08/09/19

AMS 131

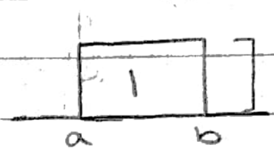
Lecture 6

$Y \sim \text{uniform}(a,b) \longleftrightarrow P\left(\begin{matrix} Y \text{ is in any subinterval} \\ \text{of } (a,b) \end{matrix}\right)$

"is distributed as" iff

= the length of subinterval iff $\longleftrightarrow$

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



indicator function: proposition, true/false, of A

$$I(A) = \begin{cases} 1 & \text{if A true} \\ 0 & \text{if A false} \end{cases} \qquad I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{else} \end{cases} \text{ (sets)}$$

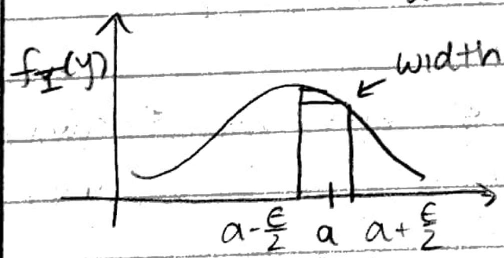
$$Y \sim \text{uniform}(a,b) \longleftrightarrow f_Y(y) = \frac{1}{b-a} I(a \le y \le b) = \frac{I_{[a,b]}(y)}{b-a}$$

$Y \sim \text{uniform}(a,b)$ continuous or $[a,b]$

$\text{uniform}\{a,b\}$ discrete, uniform $\{a, a+1, \dots, b\}$

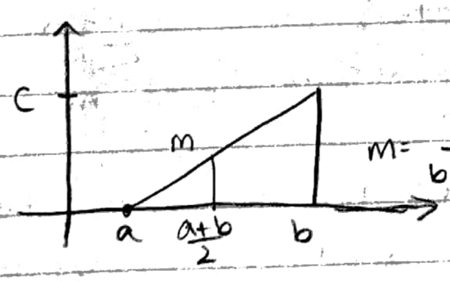
Density values $f_Y(y)$ define probability

$$P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$$



$$P\left(a - \frac{\epsilon}{2} \le Y \le a + \frac{\epsilon}{2}\right) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_Y(y)\,dy - \epsilon f_Y(a)$$

$$f_Y(a) = \frac{1}{\epsilon} P\left(a - \frac{\epsilon}{2} \le Y \le a + \frac{\epsilon}{2}\right) \quad \text{[prob. per unit]}$$



find area = 1

$m = \frac{c}{b-a}$ passes $(a,0)$ $\quad y - y_1 = m(x - x_1)$

$$y = \frac{c}{b-a}(\underset{\substack{\uparrow \\ y}}{x} - a) = f_Y(y)$$

$$\int_a^b \frac{c}{b-a}(x-a)\,dx = 1$$

$\frac{1}{2}bh$

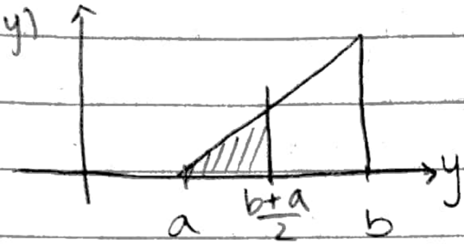
$\frac{1}{2}(b-a)c = 1$

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

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

Lecture 5 (cont.) AMS 131

Triangular distribution $P(a \le Y \le \frac{b+a}{2})$



$$\int_a^{\frac{b+a}{2}} \frac{2(y-a)}{(b-a)^2} = \frac{1}{4}$$

* $x \to y$ sometimes ... notation

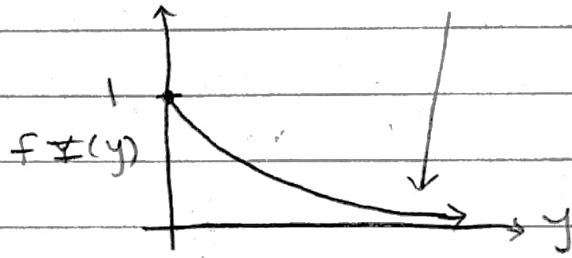
$* = A = \frac{1}{2}(\frac{b+a}{2} - a)h$

Sometimes it's easier to use unbounded continuous R.Vs

Y = voltage in an electrical system

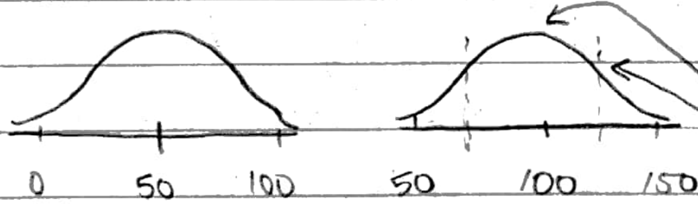
- cannot be infinite, but we want max. practical value
- can have small probability for large values.

$$f_Y(y) = \frac{1}{(1+y)^2} I(y>0)$$



$$\int_0^{\infty} \frac{1}{(1+y)^2} dy = 1$$

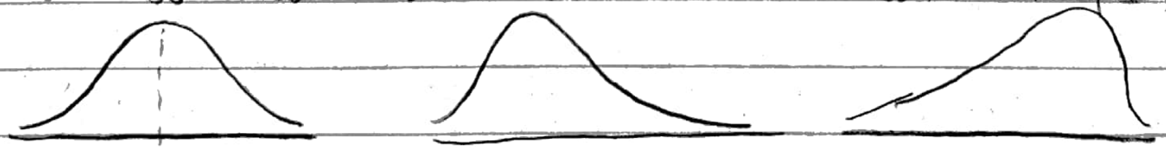
$$\int_{1000}^{\infty} \frac{1}{(1+x)^2} dx = \frac{1}{1001} = 0.001 \leftarrow$$ almost no probability $y = 1000$



PDF sketch

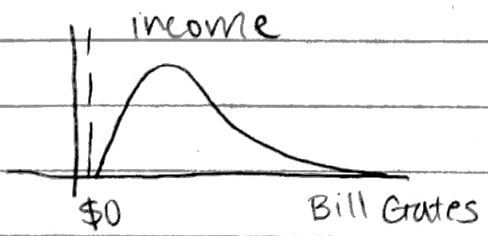
center, tails (left / right)

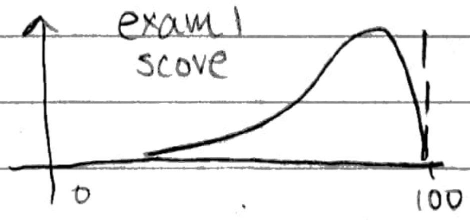
same shape, different centers, same spread



point of symmetry asymmetric = right / left skewed

pos / neg. skew

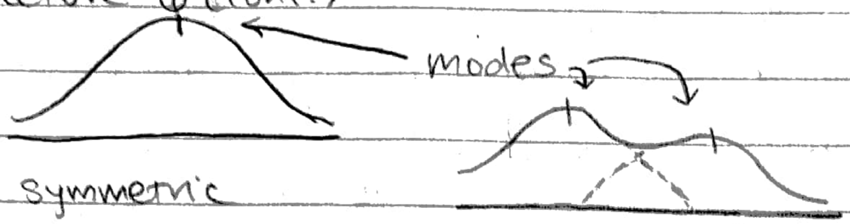




cannot go neg (usually) no extra credit

AMS 131

Lecture 6 (cont.)

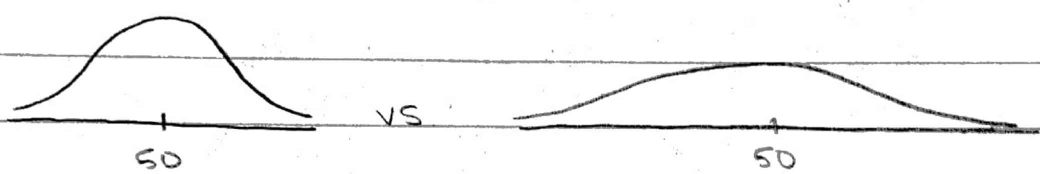


symmetric
unimodal

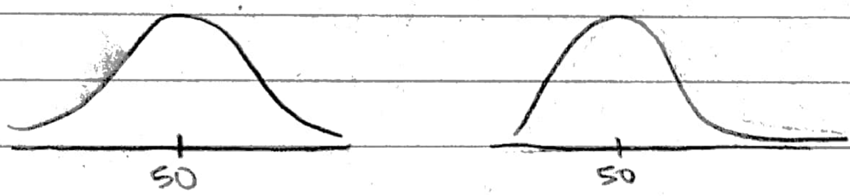
bimodal
multimodal

ex: height of 131 students
mixture distribution

one bell curve for men, one for women.
↳ latent variable : unobserved

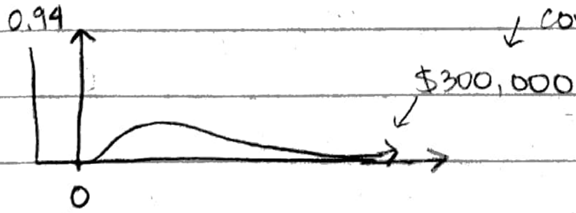


same shape, center   different spread,



same center and spread, different shape

EBAY   Y = total GMB , gross merchandise bought in (say) a 4-week period



↓ construction equipment

0.94 means a lot of people don't buy anything

Mixed distribution: part discrete, part continuous
cannot use PDF or PMF

Scanned by CamScanner

Lecture 6 (cont.) AMS 131

Y_i = survival time from beginning of trial

Some patients could be alive after Tend

continuous part: $0 \le Y \le$ Tend sensoring

discrete part: Y = Tend ← people live longer than Tend, so its attached with prob p.

discrete R.V. continuous R.V. mixed R.V

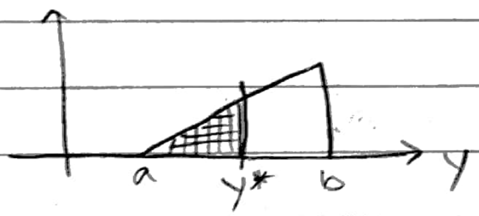
PMF

PDF

cumulative distribution function (cdf)

$F_Y(y)$ CAPITAL F $= P(Y \le y)$ how much

$f_Y(y)$ PDF / PMF probability at or below



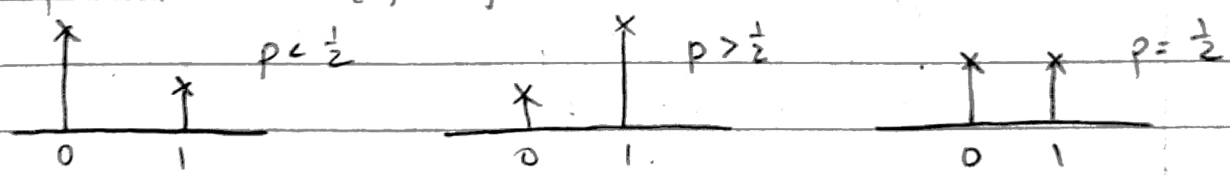
$$F_Y(y) \begin{cases} 0 & \text{for } y \le a \\ \int_a^y f_Y(y)\,dy & a \le y \le b \\ 1 & \text{for } y \ge b \end{cases}$$

$(Y \mid p) \sim$ Bernoulli (p)

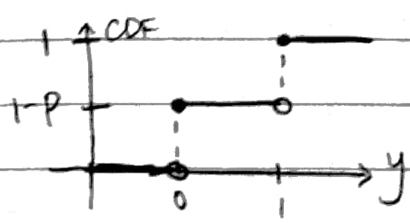
PMF

$$f_Y(y) = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = 0 \\ 0 & \text{else} \end{cases}$$

$= p^y (1-p)^{1-y} \cdot I_{\{0,1\}}(y)$



$$F_Y(y) = P(Y \le y) = \begin{cases} 0 & y < 0 \\ 1-p & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$



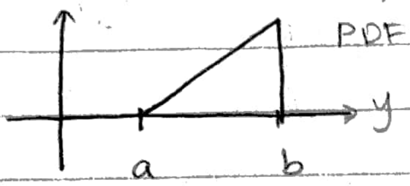
step function - used for all discrete random variables

CDF

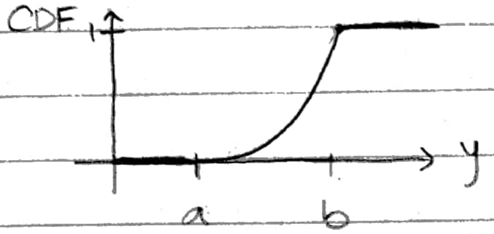
AMS 131

Lecture 6 (cont.)

$(Y \mid a,b) \sim \text{TriangularUp}(a,b)$ ← made up method



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



$$F_Y(y) = \begin{cases} 0 & y < a \\ \square & a \le y < b \\ 1 & y \ge b \end{cases}$$

$$F_Y(y) = P(Y \le y) = \int_a^y \frac{2(t-a)}{(b-a)^2}\,dt = \left(\frac{y-a}{b-a}\right)^2$$

Step function: non decreasing

all F_Y nondecreasing, (or increasing)

all F_Y for continuous R.V.: montonic increasing

$$\lim_{y \to -\infty} F_Y(y) = 0 \qquad \lim_{y \to \infty} F_Y(y) = 1$$

$$F_Y(y^-) \triangleq \lim_{y^* \to y} F_Y(y^*) \triangleq \lim_{y^* \uparrow y} F_Y(y^*)$$

limit from the left side

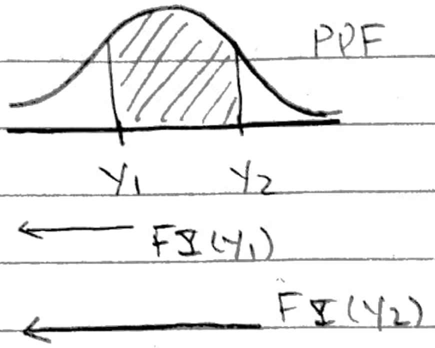
$$F_Y(y^+) \triangleq \lim_{y^* \to y} F_Y(y^*) \triangleq \lim_{y^* \downarrow y} F_Y(y^*)$$

limit from the right side

if $F_Y(y^-) = F_Y(y^+) = F_Y(y) \rightarrow F_Y$ continuous at y

1. $P(Y > y) = 1 - F_Y(y)$
2. For all y_1, y_2 with $y_1 < y_2$
   $P(y_1 < Y \le y_2) = F_Y(y_2) - F_Y(y_1)$
   CDF in between y_1 and y_2

Lecture 6 (cont.)



$$\int_{y_1}^{y_2} f_Y(y)\,dy = F_Y(y_2) - F_Y(y_1)$$

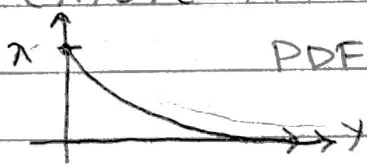
If Y is a continuous R.V. with PDF $f_Y(y)$ and CDF $F_Y(y)$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt \quad \text{and} \quad \frac{d}{dy} F_Y(y) = f_Y(y)$$

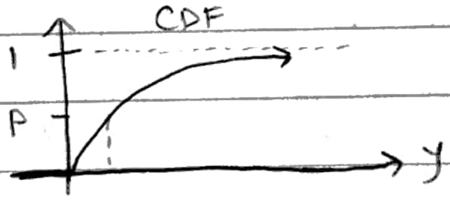
Y follows an exponential $\mathcal{E}(\lambda)$ (rate) distribution with parameter $\lambda > 0$

$(\lambda > 0)\ (Y \mid \lambda) \sim Exp(\lambda)$

$$\underset{PDF}{\longleftrightarrow} \quad f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$



exponential has a fundamental connection to the Poisson distribution



Y exponentially distributed with $\lambda > 0$ $(Y \mid \lambda)$ $\leftrightarrow Y \sim Exp(\lambda)$ $\sim \mathcal{E}(\lambda)$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt = \int_0^y \lambda e^{-\lambda t}\,dt = \lambda \frac{e^{-\lambda t}}{-\lambda}\Big|_0^y = 1 - e^{-\lambda y}$$

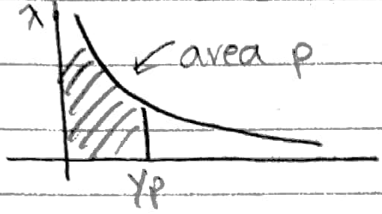
$$= \begin{cases} 0 & \text{for } y \leq 0 \\ 1 - e^{-\lambda y} & y > 0 \end{cases}$$

invertible: find a y value for a given p

Scanned by CamScanner

AMS 131

Lecture 6 (cont.)



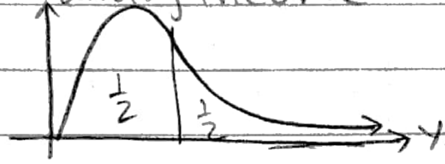
whats the place y_p where $P(0 \le Y \le y_p) = p$

$$P(0 \le Y \le y_p) = F_Y(y_p) = p = 1 - e^{-\lambda y_p} = p$$

$$\downarrow$$

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

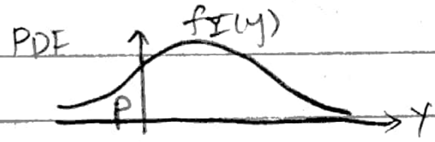
family income



median

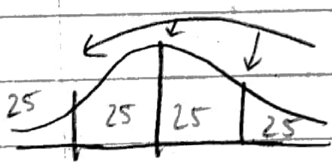
$F_Y(y_p) = \frac{1}{2}$

median $= y_p = F_Y^{-1}(\frac{1}{2})$



$y_p = p^{TH}$ quantile of dist. of Y

$= 100\,p^{TH}$ quantile of dist. of Y



1ST, 2ND, 3RD quartile

25TH 50TH 75TH percentile

Continuous CDF = easy invertible

discrete CDF = harder to define

For all $0 < p < 1$ define

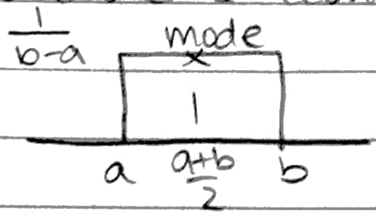
$F_Y^{-1}(p)$ = the smallest y value such that $F_Y(y) \ge p$

F_Y^{-1} is the quantile function

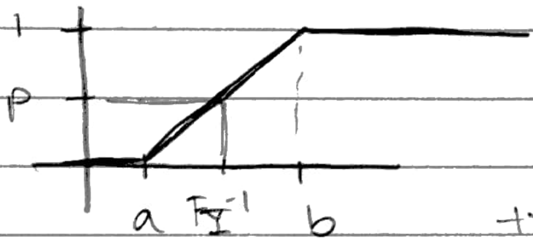
spread: how far the 25TH → 75TH percentile are

interquartile range $= F_Y^{-1}(0.75) - F_Y^{-1}(0.25)$

Lecture 6 (cont.) AMS 131



$$F_Y(y) = \begin{cases} 0 & \text{for } y \le a \\ \frac{y-a}{b-a} & a \le y \le b \\ 1 & y \ge b \end{cases}$$ $Y \sim \text{uniform}(a,b)$



$F_Y^{-1}(p) = (1-p)a + pb$ for $0 < p < 1$

$p = \frac{1}{2}$

the median is $\frac{a+b}{2} = \frac{1}{2}a + \frac{1}{2}b$