CONTINUOUS RANDOM VARIABLES

Introduction

Continuous random variables

- A continuous random variable is a random variable that takes on an uncountably infinite number of possible values.
- Instead of speaking of computing the probability of *specific values* (as we do with discrete random variables) we speak of computing the probability of *intervals*.
- Examples:
  - Weight
  - Blood pressure
  - Viral load

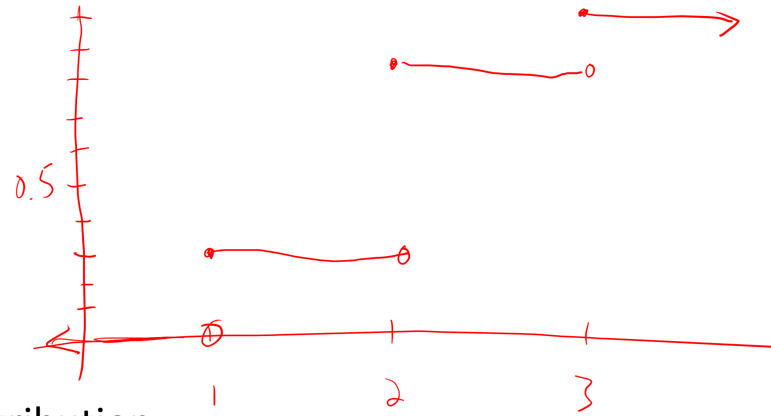
$P(50 < Y < 60)$

Cumulative distribution function

- Definition 1: The CUMULATIVE DISTRIBUTION FUNCTION (cdf) of a random variable Y, denoted $F(y)$ is

$$F(y) = P(Y \leq y), -\infty < y < \infty$$

- Both discrete and continuous random variables have cdfs.



Cumulative distribution function

- Example 1: A (discrete) random variable Y has distribution

y	1	2	3
$p(y)$	0.3	0.5	0.2

Find the cdf of Y and sketch its graph.

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

$P(Y \le y)$ if $y < 1$ is 0

$$F(y) = \begin{cases} 0 & y < 1 \\ 0.3 & 1 \le y < 2 \\ 0.8 & 2 \le y < 3 \\ 1 & y \ge 3 \end{cases}$$

$P(Y \le 1) = P(Y = 1) = 0.3$

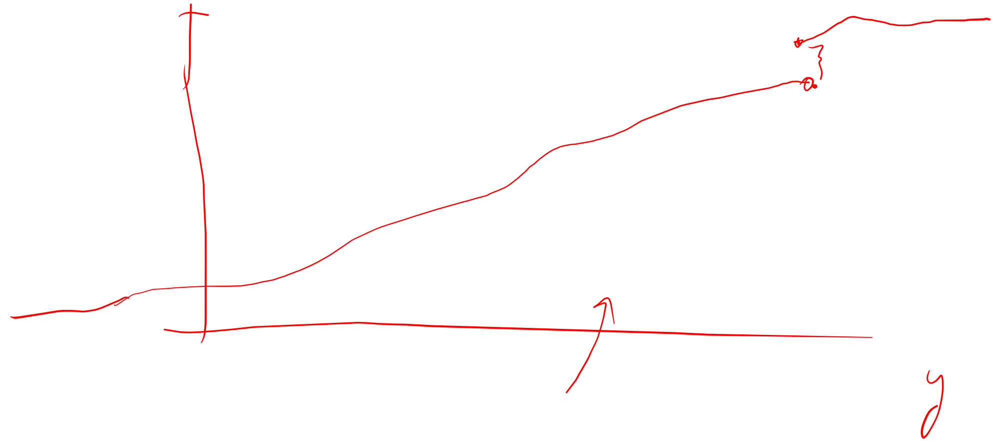
$P(Y \le y)$, $1 < y < 2$ $= 0.3$

$P(Y \le 2) = P(Y=1) + P(Y=2) = 0.3 + 0.5 = 0.8$

$P(Y \le y)$, $2 < y < 3$ $= 0.8$

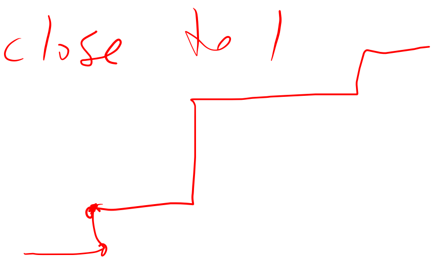
$P(Y \le 3) = 1$

$P(Y \le y)$ $y > 3$ is 1

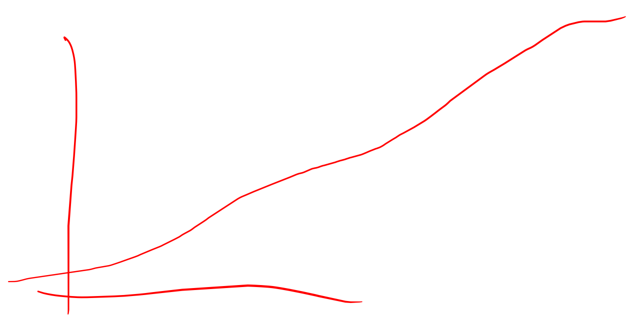


Distribution functions

- Properties of (cumulative) distribution functions:
  - $F(-\infty) = \lim_{y \to -\infty} F(y) = 0$ P(Y ≤ -1,000,000) small
  - $F(\infty) = \lim_{y \to \infty} F(y) = 1$ P(Y ≤ 1000 000) close to 1
  - $F(y)$ is a nondecreasing function of y.
- The distribution function of a discrete random variable is a step function.



- Definition 2: A random variable is a CONTINUOUS RANDOM VARIABLE if its cdf is continuous (no steps!).



Probability mass functions, probability density functions

- A discrete random variable has a *probability mass function* – it puts mass on a countable number of values.

$\sum p(y_i) = 1$

- A continuous random variable doesn't put mass on any value – instead, it has a *probability density function*.
- Definition 3: The PROBABILITY DENSITY FUNCTION of a random variable Y with distribution function $F(y)$ is

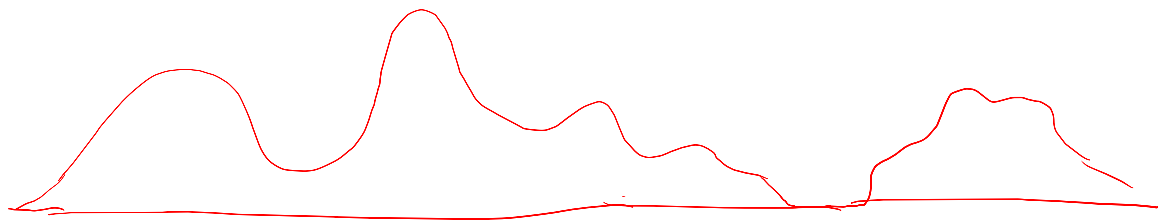
$$f(y) = \frac{d}{dy}F(y)$$

- Therefore, $P(Y \leq y) = \int_{-\infty}^{y} f(t)dt$.

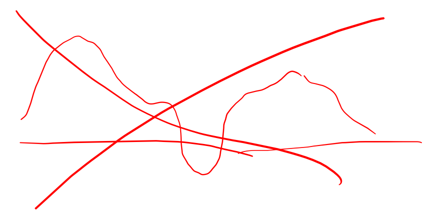
FTOC

$F(y) =$

Density functions

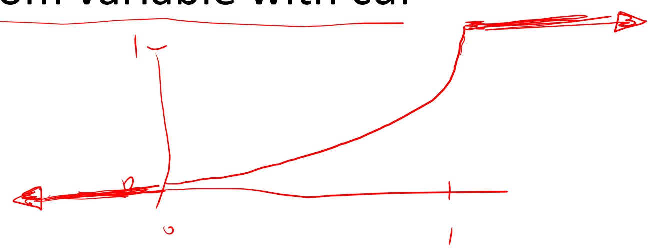


- Properties of a density function:
  - $f(y) \geq 0$ for all y, $-\infty < y < \infty$ ✓ F nondecreasing
  - $\int_{-\infty}^{\infty} f(y)dy = 1$ total area
- Example 2: Suppose Y is a random variable with cdf



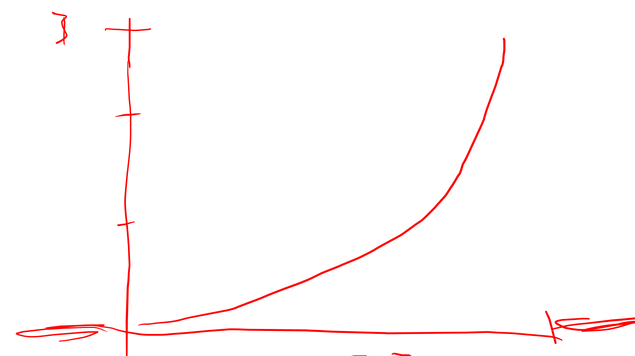
$(-\infty, \infty)$

$$F(y) = \begin{cases} 0, & y < 0 \\ y^3, & 0 \leq y < 1 \\ 1, & y > 1 \end{cases}$$



Find the pdf of Y.

$$f(y) = \frac{d}{dy}F(y) = \begin{cases} 0 & y < 0 \\ 3y^2 & 0 \leq y \leq 1 \\ 0 & y > 1 \end{cases}$$



Area under curve $= 1$?? check

Calculating probabilities of intervals

- With a continuous random variable Y, we can compute the probability that Y falls into a particular interval $[a, b]$, i.e., $P(a \leq Y \leq b)$.
- Theorem 1: If Y is a continuous random variable with pdf f, then for $a < b$,

$$P(a \leq Y \leq b) = \boxed{\int_a^b f(y)dy}$$

Proof

$$P(a \leq Y \leq b) = P(Y \leq b) - P(Y \leq a)$$

$$= \boxed{F(b) - F(a)}$$

$$= \int_{-\infty}^{b} f(y)\,dy - \int_{-\infty}^{a} f(y)\,dy$$

$$= \cancel{\int_{-\infty}^{a} f(y)\,dy} + \int_{a}^{b} f(y)\,dy - \cancel{\int_{-\infty}^{a} f(y)\,dy} \quad \square$$

Calculating probabilities of intervals

$f(y) = 3y^2, \quad 0 < y < 1$

- Example 2 (continued): for the previous example, calculate. $P(0.4 \leq Y \leq 0.8)$

$$\int_{0.4}^{0.8} 3y^2\,dy = y^3\Big|_{0.4}^{0.8} = \underbrace{(0.8)^3}_{F(0.8)} - \underbrace{(0.4)^3}_{F(0.4)} = \ldots = 0.448$$

$F(y) = y^3 \quad 0 < y < 1$

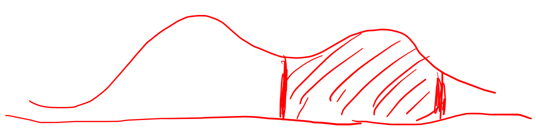
Important for discrete RVs

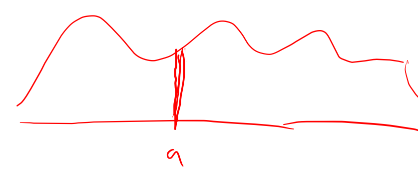
Some notes about continuous random variables

(a, b) [a, b) (a, b] [a, b]

- The same way, we can compute probabilities of "one-sided" intervals, e.g., $[a, \infty)$, $(-\infty, b]$.

- With continuous random variables, we don't have to worry about "open" or "closed" intervals, i.e.,
  $P(a \leq Y \leq b) = P(a \leq Y < b) = P(a < Y \leq b) = P(a < Y < b)$

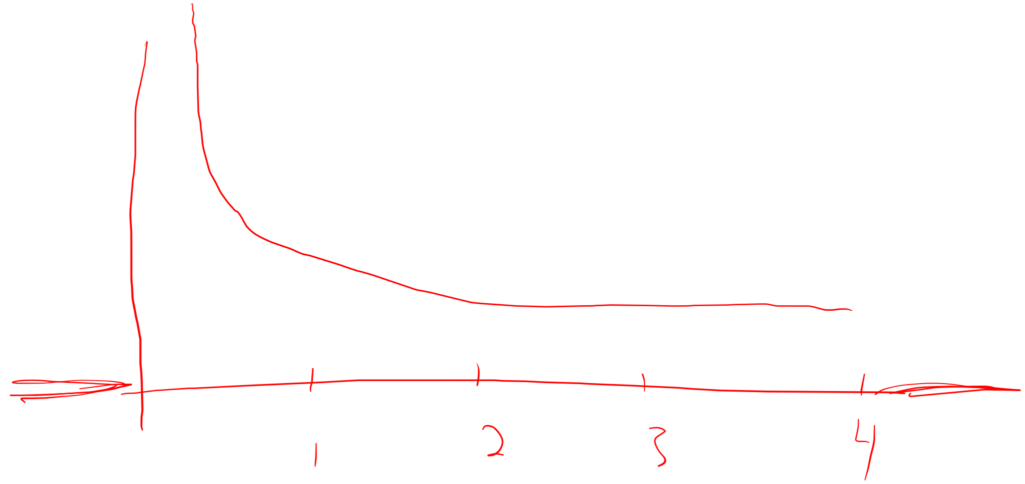




- So for a continuous random variable, $P(Y = a) =$????

$$P(Y = a) = P(a \leq Y \leq a) = \int_a^a f(y)\,dy = 0$$

Another example



- Example 3: Suppose Y is a (continuous) random variable with pdf

$$f(y) = \begin{cases} cy^{-1/2}, & 0 < y \le 4 \\ 0, & \text{otherwise} \end{cases}$$

Find c, find the distribution function F, and calculate $P(Y < 1)$.

Total prob must be 1

$$\int_0^4 f(y)\,dy = 1$$

$$\int_0^4 c\,y^{-1/2} = c\left(2y^{1/2}\Big|_0^4\right) \overset{\text{has to}}{=} 1$$

$$c\left(2\sqrt{4} - 2\cdot 0\right) = 1 \qquad c\,4 = 1 \qquad \boxed{c = \frac{1}{4}}$$

$$F(y) = \int_{-\infty}^{y} \frac{1}{4} t^{-1/2}\,dt = \frac{1}{4}\int_0^y t^{-1/2}\,dt = \frac{1}{4}\left(2t^{1/2}\Big|_0^y\right) = \frac{1}{4}\left(2\sqrt{y} - 0\right) = \frac{1}{2}\sqrt{y}$$

$$\boxed{F(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2}\sqrt{y} & 0 \le y \le 4 \\ 1 & y > 4 \end{cases}}$$

$$P(Y < 1) = F(1) = \frac{1}{2}\sqrt{1} = \boxed{\frac{1}{2}}$$

$$0 < y < 4$$