Lecture 4

- Cont. pmf, cdf
- Introduce continuous r.v. and pdfs
- Distributions of functions of r.v.s
- If time univariate $\rightarrow$ bivariate

Recap

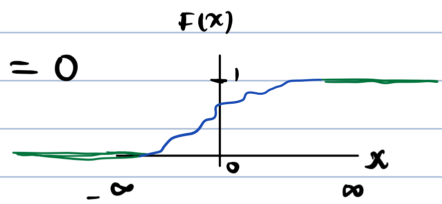
- Random variables $X : S \rightarrow \mathbb{R}$
- Distribution collection of values of $\mathbb{P}(X \in C)$ for all "valid" $C \subseteq S$
- Discrete random variables + pmf
  $p(x) = \mathbb{P}(X = x)$
- Cdf : $F(x) = \mathbb{P}(X \leq x)$

Cdf (cont.)

Properties of a cdf :

(i) $F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \mathbb{P}(X \leq -x) = 0$

$F(\infty) = \lim_{x \to \infty} F(x) = 1$



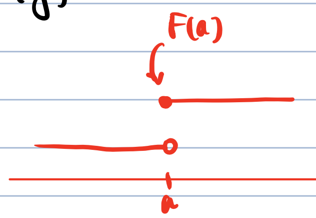
(ii) $F(x)$ is nondecreasing

e.g., X is discrete

Then $F(x) = \sum_{y \leq x} \underbrace{p(y)}_{\geq 0}$

(iii) $F(x)$ is right continuous

$\lim_{x \to a^+} F(x) = F(a)$



e.g., We repeatedly toss a coin, which has prob. p of being H, until a H appears

X = # of times the coin is flipped

What is the cdf of X? (Geometric distr.)

① Specify the possible values of X : $\{1, 2, 3, \ldots\}$

② Work out $\mathbb{P}(X \leq b)$ for $b = 1, 2, 3, \ldots$ Goal: general expression

$\mathbb{P}(X \leq b) = \sum_{i=1}^{b} p(i) = \sum_{i=1}^{b} (1-p)^{i-1} p \rightarrow p \underbrace{\sum_{j=0}^{b-1} (1-p)^j}_{\text{sum of geom. series}} = p \cdot$

$\underbrace{\ldots}_{\text{Exercise}} \; 1 - (1-p)^b$

③ $F(b) = \begin{cases} 1 - (1-p)^b & \text{for } b \leq y < b+1 \\ 0 & b < 1 \end{cases}$

How to answer questions about probs using cdf, pmf, ...

e.g., say we are asked what is $P(a < X \le b)$ $a < b$ $P(X<2)$? $P(a \le X \le b)$?

1. $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$

(A3)

2. $\mathbb{P}(X \le b) = \mathbb{P}(\{X \le a\} \cup \{a < X \le b\}) = \mathbb{P}(X \le a) + \mathbb{P}(a < X \le b)$
3. $2 \Rightarrow \mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F(b) - F(a)$

e.g., $\mathbb{P}(X < b) = \mathbb{P}\left(\bigcup_{i=1}^{\infty}\{X \le b - \frac{1}{i}\}\right) = \lim_{n\to\infty} \mathbb{P}(X \le b - \frac{1}{n}) = \lim_{n\to\infty} F(b-\frac{1}{n}) \overset{\text{Def}}{=} F(b-)$

$b - \frac{1}{n} \to b$ $F(b) = \mathbb{P}(X \le b)$

$F(b-) = \mathbb{P}(X < b)$

usually what you do in practice if $p(x)$ is known $\Big\{ = \begin{array}{l} F(b) - p(b) \\ \mathbb{P}(X \le b) - \mathbb{P}(X = b) \end{array}$

Summary: $\mathbb{P}(a < X \le b) = F(b) - F(a)$

$\mathbb{P}(X < b) = F(b) - p(b)$ (if $p(b)$ is known)

e.g., X is an r.v.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & 0 \le x < 1 \\ 2/3 & 1 \le x < 2 \\ 11/12 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

$\mathbb{P}(X < 3) = 11/12$ ✓
$= F(3-)$

$\mathbb{P}(X = 1) = \cancel{2/3}$
$= \mathbb{P}(X \le 1) - \mathbb{P}(X < 1) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$ ✓

$\mathbb{P}(X > \frac{1}{2}) = 1 - \mathbb{P}(X \le \frac{1}{2}) = 1 - F(\frac{1}{2}) = 1 - \frac{1}{2} = \frac{1}{2}$ ✓

$\mathbb{P}(2 < X \le 4) = 1/12$
$= F(4) - F(2) = 1 - \frac{11}{12} = \frac{1}{12}$ ✓

Continuous random variables

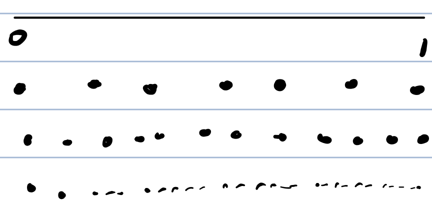
* The possible values of cont. r.v.s is not countable
e.g., Height $[0, \infty)$

Discrete $\to$ continuous   Example: Uniform distribution

$P(X = x) \to$ ✗

$P(X=x)$

Discrete r.v.s $\left\{ \begin{array}{l} X_1 \\ X_2 \\ X_3 \\ \vdots \end{array} \right.$



1/7
1/14
1/28
⋮

Continuous r.v. $\rightarrow$ Unif(0,1)   0

Takeaway : $\mathbb{P}(X = x) = 0$ for X continuous

$\mathbb{P}(X \geq 5)$ or $\mathbb{P}(X \in [a,b])$ ... $\leftarrow$ How do we do this for continuous r.v.s?
Answer : pdf

<u>DEF</u> : A r.v. X is continuous if $\exists$ non-negative function $f : \mathbb{R} \rightarrow \mathbb{R}$ s.t. for any $B \subset \mathbb{R}$, $\mathbb{P}(X \in B) = \int_B f(x)\,dx$

f is called the probability density function (pdf of X)

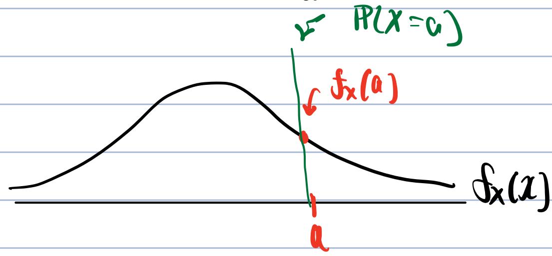
Notation : $f_X(x)$

r.v.   where we eval. the pdf

e.g., $B = [a,b]$ then $\mathbb{P}(X \in B) = \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

e.g., $B = \{a\}$ $\mathbb{P}(X \in B) = \mathbb{P}(X = a) = \int_a^a f_X(x)\,dx = 0$

Note : $\mathbb{P}(X = a) \neq f_X(a)$



e.g., $B = \{x \in \mathbb{R} : x \leq a\}$ then $\mathbb{P}(X \leq a) = \int_{-\infty}^{a} f_X(x)\,dx = F(a)$

<u>Properties</u>

(1) $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$ $f_X(x)$ need not be ≤ 1 whereas $p(x) \leq 1$ (pmf)

(2) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

e.g., Suppose X is a continuous r.v. with p.d.f.

$$f_X(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is C? 3/8   We know $\int_{-\infty}^{\infty} C(4x - 2x^2)\,dx = \int_0^2 C(4x-2x^2)\,dx = 1$

$$C = \frac{1}{\int_0^2 (4x - 2x^2)\,dx} \;\ldots\; 3/8$$

(b) What is $\mathbb{P}(X > 1)$? $f_X(x) = \frac{3}{8}(4x - 2x^2)$ for $0 < x < 2$ (0 o/w)

$\int_1^{\infty} f_X(x)\,dx = \int_1^2 \frac{3}{8}(4x - 2x^2)\,dx \;\ldots\; 1/2$

How to answer questions about probs using cdf, ~~pmf~~, ...

(i) $F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(y)\,dy = \mathbb{P}(X < x)$

$$\left.\begin{aligned} (*) &= \mathbb{P}(a \le X \le b) \\ &= \mathbb{P}(a < X \le b) \\ &= \mathbb{P}(a \le X < b) \\ &= \mathbb{P}(a < X < b) \end{aligned}\right\} = F(b) - F(a)$$

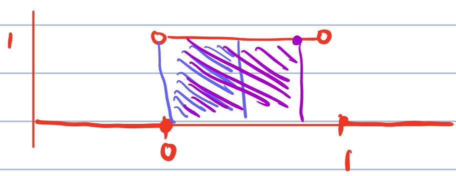
Reason: e.g., $\mathbb{P}(X < x) = \mathbb{P}(X \le x) - \cancelto{0}{\mathbb{P}(X = x)}$

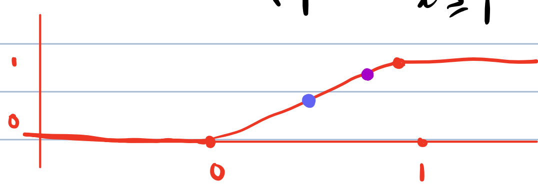
(ii) Relationship b/w cdf $\to$ pdf $\quad \frac{d}{dx} F(x) = f_X(x)$

Examples of continuous distributions

(i) Uniform distribution $\quad \text{Unif}(0,1)$

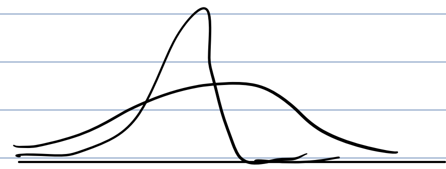
$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{o/w} \end{cases} \qquad F(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$





The above generalizes to $\text{Unif}(a,b)$

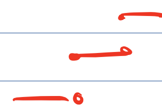
(ii) Normal (Gaussian) distribution $\quad N(\mu, \sigma^2)$



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \qquad x \in \mathbb{R}$$

"spread" "location"

Quantile function ("inverse of the cdf")

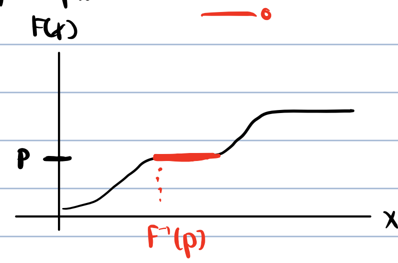


DEF: X is a r.v. with cdf F. For each $p \in (0,1)$, define

$$\underbrace{F^{-1}(p)} = \inf\{x : F(x) \ge p\}$$

$F^{-1}(p)$ is called the p-quantile of X

$F^{-1}(\cdot)$ is called the quantile function



e.g., Median is the 50% quantile

e.g., Test scores are sometimes reported in quantiles (e.g., 80% quantile = you did better than 80% of test takers)

e.g., Hypothesis testing

Distributions of functions of random variables

Goal: Suppose X is a r.v. and we know its pdf or cdf or pmf...

What is the distribution (pdf or cdf or pmf) of $g(X)$ for some function g?

<u>Discrete random variable</u> : Suppose X is discrete with pmf $p(x)$ and cdf $F(x)$
Suppose g is a one-to-one function and non decreasing

PMF

(i) $\mathbb{P}(g(X) = y) = \mathbb{P}(X = g^{-1}(y))$

$$p(x) = \mathbb{P}(X = x) = \mathbb{P}(g(X) = \underbrace{g(x)}_{y})$$

CDF

(ii) $F_{g(x)}(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F(g^{-1}(y))$

<u>In general</u> (g may not be one-to-one), $P_{g(x)}(y) = \sum_{x \in g^{-1}(y)} p_X(x)$ ⟵ Pmf of $g(x)$

where $g^{-1}(A) = \{x : g(x) \in A\}$
(inverse image)

If g is one-to-one $= p_X(g^{-1}(y))$

$F_{g(x)}(y) = \mathbb{P}(g(X) \leq y) = \sum_{g(x) \leq y} P_{g(x)}(y)$

<u>Continuous random variable</u> : Suppose X is continuous with pdf f and cdf F
Suppose g is one-to-one and non decreasing

(i) $F_{g(x)}(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$ ⟵ Same as discrete case

(ii) $f_{g(x)}(y) = \frac{d}{dy} F_{g(x)}(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \times \frac{d}{dy} g^{-1}(y)$

e.g., $g(X) = cX$ for some $c > 0$

$F_{g(x)}(y) = \mathbb{P}(cX \leq y) = \mathbb{P}(X \leq y/c) = F_X(y/c)$

$f_{g(x)}(y) = \frac{d}{dy} F_X(\frac{y}{c}) = f_X(\frac{y}{c}) \cdot \frac{1}{c}$

In general, we have the following result:

<u>THM</u> : Suppose X is continuous and $g(X)$ is $\overbrace{\text{strictly monotone}}^{\Rightarrow g \text{ is one-to-one}}$. Then $Y = g(X)$ has pdf given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

$g^{-1}(y)$ is the value of x s.t. $g(x) = y$