Randomness

Outline

- Introduction to probability
- Conditional probability
- Random variables
- Some useful probability distributions

Random variables

Random variables

- A random variable is a numeric quantity whose value depends on the outcome of a random event.

  $X : \underbrace{\Omega}_{\text{sample space}} \to \mathbb{R}$ (mapping)

  - We use a capital letter, like X, to denote a random variable
  - The values of a random variable are denoted with a lowercase letter, in this case x
  - For example, $P(X = x)$

Random variables

- Discrete random variables often take only integer values 當 random trial 只有有限多種可能結果時
    - Example: number of students present, gender of an unborn baby $X = \begin{cases} 1, & \text{male} \\ 2, & \text{female} \end{cases}$
- Continuous random variables take real (decimal) values
    - Example: tomorrow's PM 2.5 level, your final grade
      $\in (0, \infty)$ $[0, 100]$

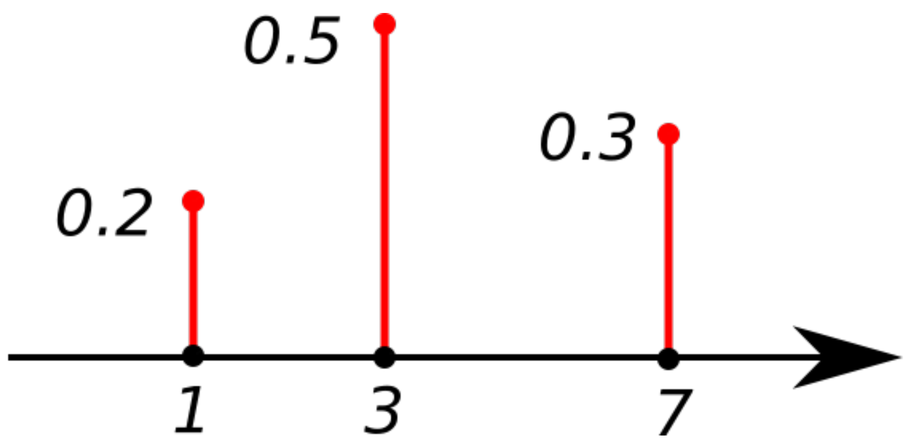
Probability Mass Functions

A probability mass function (pmf), $P(X = x) = f(x)$, of a discrete random variable X, is a function that gives the probability that a discrete random variable is exactly equal to some value. It should satisfy the following properties:

1. $P(X = x) = f(x) \geq 0$
2. $\sum f(x) = 1$

Probability Mass Functions

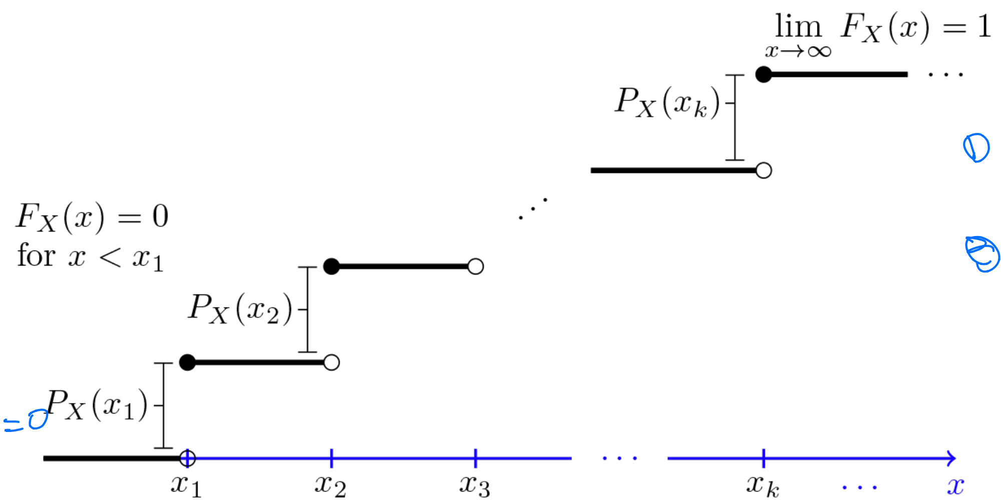
- Since the image of X is countable, the probability mass function f(x) is zero for all but a countable number of values of x.



Cumulative Distribution Function

- The cumulative distribution function (c.d.f.) for a discrete random variable is

$$F(x) = Prob(X \leq x) = \sum_{t \leq x} f(t)$$



① increasing

② $F(x) \in [0, 1]$

$\lim_{x\to-\infty} F(x) = 0$

Probability density function

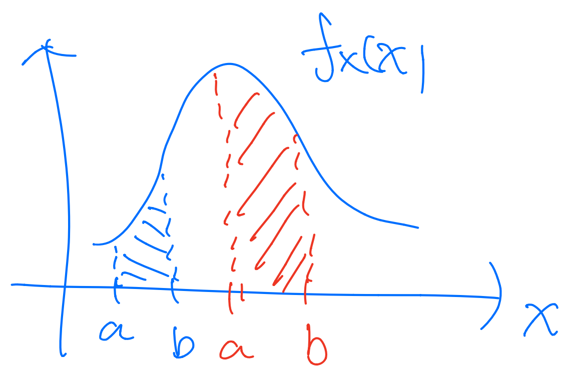
A probability density function (pdf) of a continuous random variable is a function satisfying the following properties:

Note that $f_X(x) \neq P(X=x)$

(likelihood)

1. $f_X(x) \geq 0, -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $Prob(a < X \leq b) = \int_a^b f_X(x)dx$

$P(X \in (a,b])$



Probabilities from continuous distributions

- For a continuous random variable X, the probability $P(X = x)$ is defined as 0. otherwise $\sum_x P(X=x) = \infty$

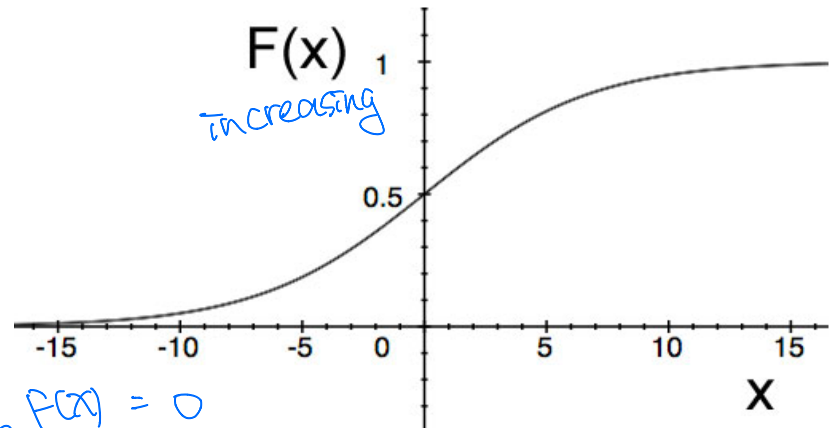
- $P(a < X \leq b) = \cancel{F_X(a) - F_X(b)} = \int_a^b f(x)dx$

$F_X(b) - F_X(a)$

C.d.f. for a continuous random variable

- The c.d.f. for a continuous random variable is

$$\frac{d}{dx}F_X(x) = f_X(x) \qquad F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad F_X(x) = P(X \le x)$$



increasing

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

$\lim_{x \to -\infty} F(x) = 0$

Useful probability distributions

Bernoulli distribution

來自只有 2 種可能結果的 random trial

(Bernoulli trial)

$$X = \begin{cases} +1, & \text{with prob. } p \\ -1, & \text{"} \quad 1-p \end{cases}$$

$$X = \begin{cases} 1, & \text{with prob. } p \\ 0, & \text{"} \quad 1-p \end{cases}$$

Bernoulli random variable

Binomial distribution

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, $X_i = \begin{cases} 1, & \text{with prob. } p \\ 0, & \text{" } \quad 1-p \end{cases}$

$\Rightarrow Y = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$

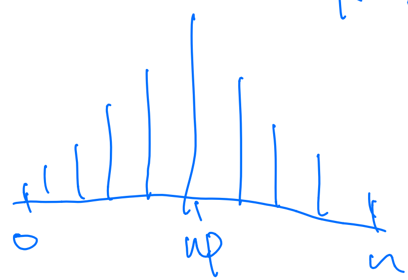
$Y \in \{0, 1, 2, \cdots, n\}$

$P(Y=0) = (1-p)^n$

$P(Y=n) = p^n$

$P(Y=10) = \binom{n}{10} p^{10} (1-p)^{n-10}$

p.m.f.

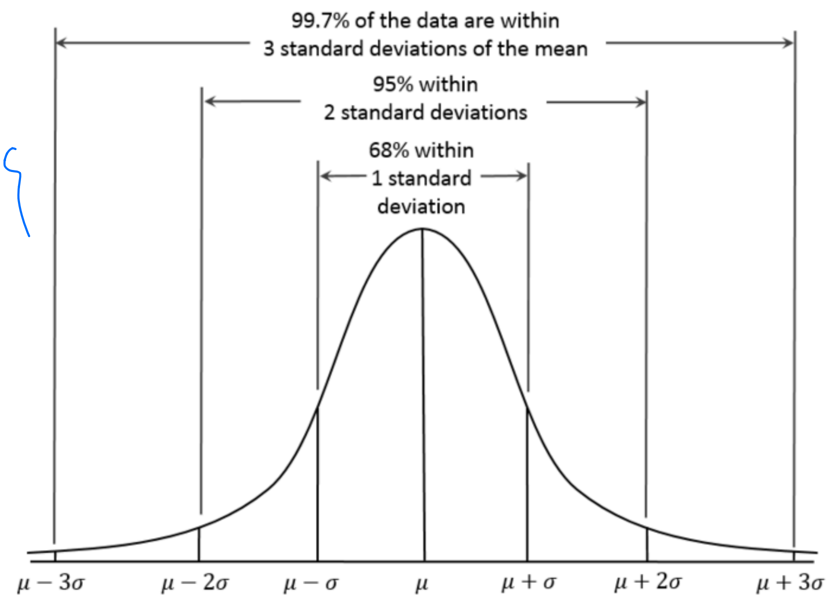


Normal distribution

X : continuous random variable

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$X \sim N(\mu, \sigma^2)$$



Central limit theorem

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} f_X(x)$ 任意 p.d.f.

If $Y = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow Y \sim N(\mu, \sigma^2)$ if $n \to \infty$

or $Y \xrightarrow{d} N(\mu, \sigma^2)$

分配收斂至 (converge in distribution)

Distributions derived from normal random variables

① $Z_1, Z_2, \cdots, Z_n \overset{iid}{\sim} N(0,1) \Rightarrow X = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$

(卡方分配)

② If $Z \sim N(0,1)$ and $V \sim \chi^2_\nu$, $Z \perp V$

$\Rightarrow T = \dfrac{Z}{\sqrt{V/\nu}} \sim t_\nu$ (student's t distribution)

③ If $U \sim \chi^2_\mu$ and $V \sim \chi^2_\nu$, $U \perp V$

$\Rightarrow F = \dfrac{U/\mu}{V/\nu} \sim F_{\mu,\nu}$ (F-distribution)

Generating pseudo random numbers in Python

- random table
- numpy.random
- seed

Readings

- Appendix A and Chapter 2.5–2.7 of Introductory Statistics with Randomization and Simulation
- Chapter 9 of our first reference