Recall: prob. space (LN p.2-17) $(\Omega, \mathfrak{F}, P)$

Random Variables — 隨机變數.

- A Motivating Example — a special case of survey sampling.
  - ➢ Experiment: Sample k students without replacement from the population of all n students (labeled as 1, 2, …, n, respectively) in our class.
  - ➢ Ω = {all combinations} = $\{\{i_1, \ldots, i_k\}: 1 \le i_1 < \cdots < i_k \le n\}$
    $\mathfrak{F} = 2^{\Omega}$
  - ➢ A probability measure P can be defined on Ω, e.g, when there is an equally likely chance of being chosen for each students,

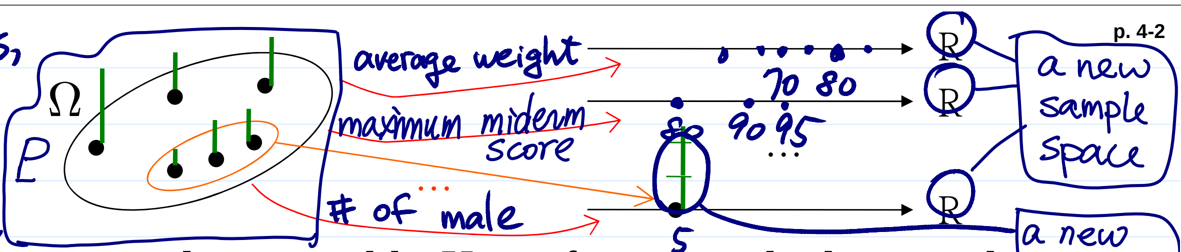
$$P(\{i_1, \ldots, i_k\}) = 1/\underbrace{\binom{n}{k}}_{\#\Omega}$$

  - ➢ For an outcome $\omega \in \Omega$, the experimenter may be more interested in some quantitative attributes of ω, rather than the ω itself, e.g.,
    - The average weight of the k sampled students
    - The maximum of their midterm scores
    - The number of male students in the sample

    ∵ the outcome ω is random

  - **Q**: What mathematical structure would be useful to characterize the *random* quantitative attributes of ω's?

In some cases, might be difficult to observe



- Definition: A *random variable* X is a function which maps the sample space Ω to the real numbers $\mathbb{R}$, i.e. measurable.

  a new prob. measure

$$X: \Omega \to \mathbb{R}.$$

$X(\omega)$ random.

  - ➢ The P defined on Ω would be transformed into a *new* probability measure defined on $\mathbb{R}$ through the mapping X
    $\Rightarrow$ the outcome of X is random, but the map X is deterministic
  - ➢ Example (Coin Tossing): Toss a fair coin 3 times, and let
    - X_1 = the total number of heads
      X_2 = the number of heads on the first toss
      X_3 = the number of heads minus the number of tails
    - Ω = {*hhh, hht, hth, thh, htt, tht, tth, ttt*}

can do numerical calculation, "+" "−" "×" "÷" exp, ln, … & >, < …

	↓1/8	↓1/8	↓1/8	↓1/8	↓1/8	↓1/8	↓1/8	↓1/8 $P(\omega)$
X_1 :	3,	2,	2,	2,	1,	1,	1,	0.
X_2 :	1,	1,	1,	0,	1,	0,	0,	0.
X_3 :	3,	1,	1,	1,	−1,	−1,	−1,	−3.

$P_{X_1}(0) = 1/8$
$\quad\;\; (1) = 3/8$
$\quad\;\; (2) = 3/8$
$\quad\;\; (3) = 1/8$

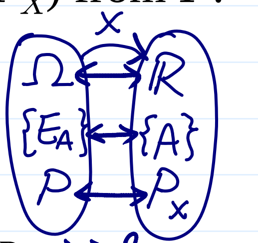
  - ➢ **Q**: Why particularly interested in functions that map to "$\mathbb{R}$"?

➢ **Q**: How to define the probability measure of X (P_X) from P?

Ans: For a set (an event) $A \subset \mathbb{R}$,

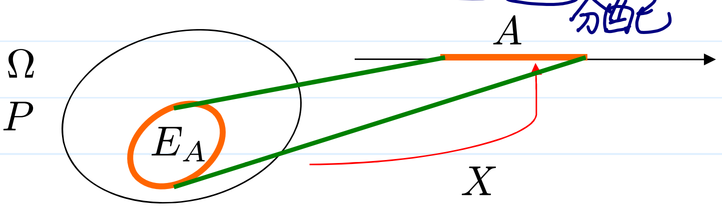
$$P_X(X \in A) \equiv P(\underbrace{\{\omega : X(\omega) \in A\}}_{E_A}).$$

The P_X is often called the *distribution* of X. 分配



$\boxed{A \text{ occurs} \Leftrightarrow E_A \text{ occurs}}$

$\boxed{P_X(A) = P(E_A)}$



original prob. space   new prob. space

P_X

$P_X(A) = ??$

離散 ← Discrete Random Variables ←c.f.→ discrete sample space

• Definition: For a random variable (r.v.) X, let

$$\mathcal{X} = \{X(\omega) : \omega \in \Omega\},$$

be the range of X. Then, X is called *discrete* if $\mathcal{X}$ is a finite or countably infinite set, i.e.,

$$\mathcal{X} = \{x_1, \ldots, x_n\} \text{ or } \mathcal{X} = \{x_1, x_2, \ldots\}.$$

$X_1: \mathcal{X} = \{0,1,2,3\} \subset \mathbb{R}$

$X_2: \mathcal{X} = \{0,1\}$

$X_3: \mathcal{X} = \{-3,-1,1,3\}$

➢Example. The X_1, X_2, X_3 in the Coin Tossing example.

• **Q**: The sample space of a r.v. is the **real line** $\mathbb{R}$. Are there some particular ways to depict a probability measure (p.m.) on $\mathbb{R}$? [c.f., for general sample space Ω, a p.m. is defined on (all) subsets of Ω]

Ans: 3 commonly used tools to depict the p.m. of discrete r.v.'s:

1. Probability mass function (pmf)
2. Cumulative distribution function (cdf)
3. Moment generating function (mgf, Chapter 7)

→ When any of them is known, the rest 2 can be obtained

➢Definition: If X is a discrete r.v., then the *probability mass function* of X is defined by

$$f_X(x) \equiv P_X(\{X = x\}) = P(\{\omega \in \Omega : X(\omega) = x\})$$

for $x \in \mathbb{R}$. (c.f., the p: Ω→R in LNp.2-5)

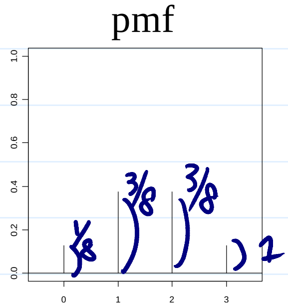
← discrete sample space.

▪ Example. For the X_1 in the Coin Tossing example,

- $\mathcal{X} = \{0, 1, 2, 3\}$ ← ≠ R
- $f_{X_1}(0) = 1/8$, $f_{X_1}(1) = 3/8$,
  $f_{X_1}(2) = 3/8$, $f_{X_1}(3) = 1/8$,
  and $f_{X_1}(x) = 0$, for $x \notin \mathcal{X}$.

LNp.2-5: $p: \Omega \to \mathbb{R}$

- Graphical display



➢Example (Committees). A committee of size n=4 is selected from 5 men and 5 women. Then,

#Ω is finite
each ω is equally likely

- Ω={combination of 4}, $\#\Omega = \binom{10}{4} = 210$, $P(A) = \#A/\#\Omega$

$m_1 \vdots m_5$ $w_1 \vdots w_5$ $4-x$ x (男, 女)

- Let X be the number of women on the committee, then
  - $f_X(x) = P_X(X=x) = \binom{5}{x}\binom{5}{4-x}/\binom{10}{4}$ ← $P(A)$ defined on Ω
  - $f_X(0) = f_X(4) = \frac{5}{210}$, $f_X(1) = f_X(3) = \frac{50}{210}$, $f_X(2) = \frac{100}{210}$.

  $\mathcal{X} = \{0,1,2,3,4\}$ $f_X(x) = 0$, if $x \notin \mathcal{X}$

- **Q**: What should a pmf look like?

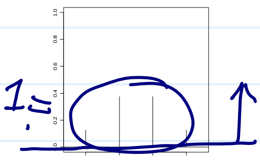
  Theorem. If f_X is the pmf of r.v. X with range $\mathcal{X}$, then

  (i) $f_X(x) \geq 0$, for all $x \in \mathbb{R}$,

  only finite or countably infinite $x \in \mathbb{R}$ s.t. $f(x) > 0$.

  (ii) $f_X(x) = 0$, for $x \notin \mathcal{X}$,

  (iii) $\sum_{x \in \mathcal{X}} f_X(x) = 1$.



$1 =$

To define P_X, it's enough to define f_X

  (iv) moreover, for $A \subset \mathbb{R}$, $P_X(X \in A) = \sum_{x \in A \cap \mathcal{X}} f_X(x)$.

proof. (i)(ii) by pmf definition ($\because f_X(x) = P(\{\omega \in \Omega : X(\omega) = x\})$)

(iv) For $A \subset \mathbb{R}$, let $A \cap \mathcal{X} = \{x_1', x_2', \cdots\}$.

$P_X(A) = P_X(X \in A \cap \mathcal{X}) + P_X(X \in A \cap \mathcal{X}^c)$

$= \sum_k P_X(X = x_k') = \sum_k f_X(x_k') = \sum_{x \in A \cap \mathcal{X}} f_X(x)$

(iii) follow (iv) by letting A in (iv) be $\mathcal{X}$

$\sum_{x \in \mathcal{X}} f_X(x) = P_X(X \in \mathcal{X}) = P(\Omega) = 1$.

- Theorem. Any function f that satisfies (i), (ii), and (iii) for some finite or countably infinite set $\mathcal{X}$ is the pmf of some random variable X.

proof. For given $\mathcal{X}$ & f, let $\Omega = \mathcal{X}$. For $\forall A \subset \mathcal{X}$, $P(A) = \sum_{x \in A} f(x)$.
then, $(\Omega, \{A\}, P)$ is prob. space. Let $X: \Omega \to \mathbb{R}$, $X(\omega) = \omega$.
Then, X is a r.v. (discrete) & for any $x \in \mathbb{R}$, $f_X(x) = P_X(X=x) = f(x)$.

sort of ignore the prob. space $(\Omega, \mathcal{F}, P)$

- Henceforth, we can define "pmf" as any function that satisfies (i), (ii), and (iii).
- We can specify a distribution by giving $\mathcal{X}$ and f, subject to the three conditions (i), (ii), (iii).

  defined on Ω

  - **Q**: Suppose that X and Y are two r.v.'s with same pmf. Is it always true that $X(\omega) = Y(\omega)$ for $\omega \in \Omega$? Ans. No.

$1/2$ $1/2$ $\mathbb{R}$ 0 1 $P(A) = 1/2$ A A^c Ω $1/2$ $1/2$ $\mathbb{R}$ 0 1

➢ Definition: A function F_X is called the *cumulative distribution function* of a random variable X if

$\mathbb{R}, P_X$ x

$$F_X(x) = P_X(X \leq x),\ x \in \mathbb{R}.$$

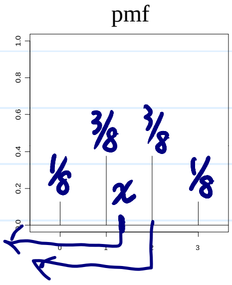
not restricted to discrete r.v.

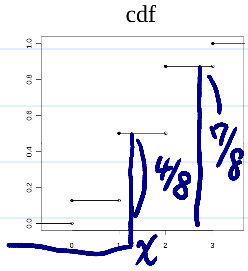
c.f. pmf defined only for discrete r.v.

(**Note**. The definition of cdf can be applied to arbitrary r.v.'s)

- Example. For the X_1 in the Coin Tossing example,

$$F_{X_1}(x) = \begin{cases} 0, & x < 0, \\ 1/8, & 0 \leq x < 1, \\ 4/8, & 1 \leq x < 2, \\ 7/8, & 2 \leq x < 3, \\ 1, & 3 \leq x. \end{cases}$$

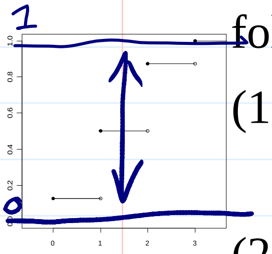




- **Q**: What should a cdf look like?

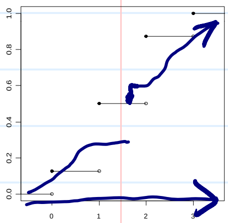
Theorem. If F_X is the cdf of a r.v. X, then it must satisfy the following properties:

If not specified, prove for any r.v. (discrete, continuous, mixed)



(1) $0 \le F_X(x) \le 1$.

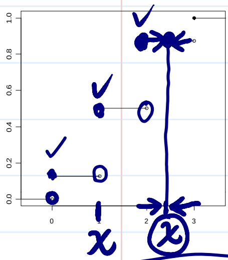
proof. $0 \le F_X(x) = P(\{\omega \in \Omega : X(\omega) \in (-\infty, x]\}) \le 1$. ($= P_X((-\infty,x])$)



(2) $F_X(x)$ is nondecreasing, i.e., $F_X(a) \le F_X(b)$ for $a<b$.

proof. For $a<b$, $(-\infty, a] \subset (-\infty, b]$,

$$F_X(a) = P_X((-\infty, a]) \le P_X((-\infty, b]) = F_X(b).$$



(3) For any $x \in \mathbb{R}$, $F_X(x)$ is continuous from the right, i.e.,

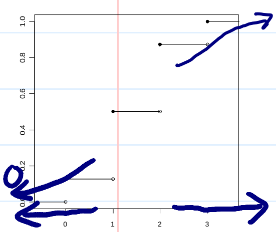
$$F_X(x) = F_X(x+) \equiv \lim_{t \downarrow x} F_X(t),$$

proof. Let x_n be a sequence s.t. $x_n \downarrow x$. (i.e., $x_1 \ge x_2 \ge x_3 \ge \cdots$, $\lim x_n = x$)

Let $E_n = (-\infty, x_n]$, then $E_n \downarrow (-\infty, x]$ $\cap E_n$ (∵ E_n decreasing)

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) = P_X\left(\lim_{n\to\infty} E_n\right) \\ &= \lim_{n\to\infty} P_X(E_n) = \lim_{n\to\infty} P_X((-\infty, x_n]) \\ &= \lim_{n\to\infty} F_X(x_n) \end{aligned}$$

Recall: $E_n \uparrow E$, $E_n \downarrow E$
$P(\lim E_n) = \lim P(E_n)$
LNp.2-14



(4) $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to -\infty} F_X(x) = 0$,

proof. Let $x_n \downarrow -\infty$, then $E_n \equiv (-\infty, x_n] \downarrow \emptyset$. ($\leftarrow \cap E_n$.)

$$\begin{aligned} \lim_{n\to\infty} F_X(x_n) &= \lim_{n\to\infty} P_X((-\infty, x_n]) \\ &= P_X\left(\lim_{n\to\infty} E_n\right) = P_X(\emptyset) = 0. \end{aligned}$$

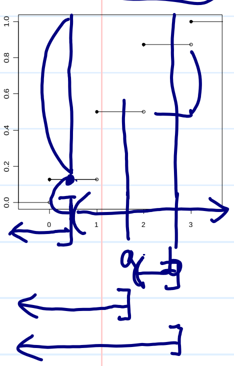
Similarly, if $x_n \uparrow \infty$, then $E_n \equiv (-\infty, x_n] \uparrow \mathbb{R}$, and ($\leftarrow \cup E_n$.)

$$\begin{aligned} \lim_{n\to\infty} F_X(x_n) &= \lim_{n\to\infty} P_X((-\infty, x_n]) \\ &= P_X\left(\lim_{n\to\infty} E_n\right) = P_X(\mathbb{R}) = 1. \end{aligned}$$

Recall
$A \subset B$
$P(B \setminus A)$
$= P(B) - P(A)$
LNp.2-8

(5) $P_X(X>x) = 1 - F_X(x)$ and $P_X(a<X\le b) = F_X(b) - F_X(a)$.

$= P_X((a,b])$
$= P_X(X \in (a,b])$



proof.
$$\begin{aligned} P_X(X > x) &= 1 - P_X(\{X > x\}^c) \\ &= 1 - P_X(X \le x) = 1 - F_X(x). \end{aligned}$$

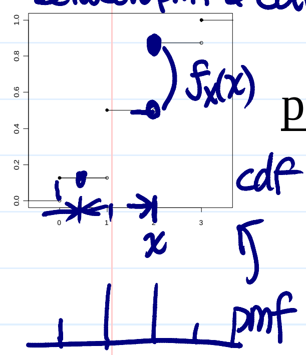
For $a < b$, $(-\infty, a] \subset (-\infty, b]$, and

$$\begin{aligned} P_X(a < X \le b) &= P_X((-\infty, b] \setminus (-\infty, a]) \\ &= P_X((-\infty, b]) - P_X((-\infty, a]) = F_X(b) - F_X(a). \end{aligned}$$

transformation between pmf & cdf.

(6) Moreover, if X is discrete with pmf f_X, then for $x \in \mathbb{R}$,

$$F_X(x) = \sum_{x_i \in \mathcal{X}, x_i \le x} f_X(x_i), \text{ and } f_X(x) = F_X(x) - F_X(x-).$$



proof. $F_X(x) = P_X(X \in (-\infty, x]) \underset{\text{by (iv) in LNp.4-5}}{=} \sum_{x_i \in (-\infty, x] \cap \mathcal{X}} f_X(x_i)$. Prob $= F_X(x-)$

For $x_n \uparrow x$, $\underbrace{(-\infty, x_n]}_{E_n} \uparrow \underbrace{(-\infty, x)}_{\cup E_n}$ and

$$F_X(x-) = \lim_{n\to\infty} F_X(x_n) = P_X((-\infty, x)).$$

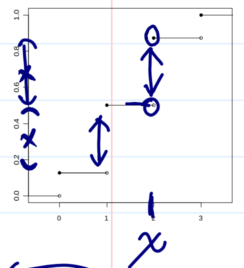
x prob $= F_X(x)$

So, $f_X(x) = P_X(\{x\}) = P_X((-\infty, x] \setminus (-\infty, x))$

$$= P_X((-\infty, x]) - P_X((-\infty, x)) = F_X(x) - F_X(x-)$$

Q: at most how many jumps?

(7) F_X has at most countably many discontinuity points.



proof. Let $\mathbb{D}$ be the collection of discontinuity points.

For $x \in \mathbb{D}$, let $T_x = (F_X(x-), F_X(x))$. $x \in \mathbb{D} \leftarrow$ one-to-one $r_x \in \mathbb{Q}$

Because $F_X(x-) \ne F_X(x)$, $\Rightarrow F_X(x-) < F_X(x)$

$\exists$ a rational number, denoted by r_x, in T_x. if $r_x \ne r_{x'}$, $x \ne x'$.

Because the set of rational numbers is a countable set,

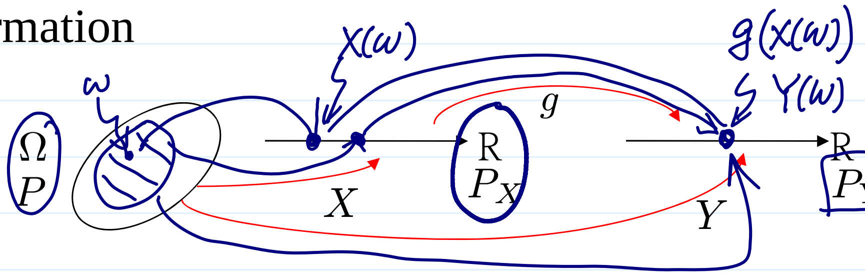
$\mathbb{D}$ is either finite or countably infinite. LNp. 4-7~8.

Last Thm in LNp4-5 cf.

▪ Theorem. If a function F satisfies (2), (3), and (4), then F is a cumulative distribution function of some random variable.

proof: skip, out of the scope of the course.

- Transformation

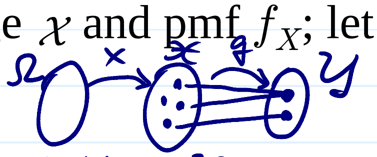


$Y: \Omega \to \mathbb{R}$

$Y: \mathcal{X} \to \mathbb{R}$

➢Theorem. Let X be a discrete r.v. with range $\mathcal{X}$ and pmf f_X; let

$$Y = g(X)$$



then, the range of Y is $= \{Y(\omega) : \omega \in \Omega\}$

countable? $\mathcal{Y} = \{g(x) : x \in \mathcal{X}\}$,

For $\forall y \in \mathcal{Y}$ $\exists x_y \in \mathcal{X}$ s.t. $g(x) = y$ one to one

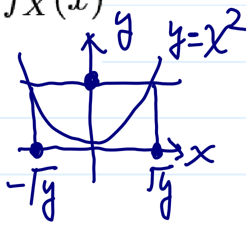
i.e., Y is a discrete r.v., and the pmf of Y is

$$f_Y(y) = \underbrace{\sum_{x \in \mathcal{X}, g(x) = y}}_{\text{countable}} f_X(x).$$

proof. Since $\underbrace{\{\omega \in \Omega : Y(\omega) = y\}}_{P(\cdot)} = \underbrace{\bigcup_{x \in \mathcal{X}, g(x)=y}}_{\text{countable}} \underbrace{\{\omega \in \Omega : X(\omega) = x\}}_{\text{mutually exclusive}}$,

$$f_Y(y) = \sum_{x \in \mathcal{X}, g(x)=y} P(\{\omega \in \Omega : X(\omega) = x\}) = \sum_{x \in \mathcal{X}, g(x)=y} f_X(x)$$

▪ Example. If $Y = X^2$, then $f_Y(y) = f_X(\sqrt{y}) + f_X(-\sqrt{y})$.



❖ **Reading**: textbook, Sec 4.1, 4.2, 4.10

Expectation (Mean) and Variance

- **Q**: We often characterize a person by his/her height, weight, hair color, …. How can we "roughly" characterize a distribution?
- Definition: If X is a discrete r.v. with pmf f_X and range $\mathcal{X}$, then the *expectation* (or called *expected value*) of X is

期望值

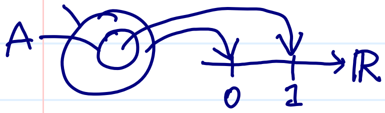
$$E(X) = \sum_{x\in\mathcal{X}} x f_X(x),$$

provided that the sum converges absolutely. → i.e. $\sum_{x\in\mathcal{X}} |x| f_X(x) < \infty$.

  - ➢Example. If all value in $\mathcal{X}$ are equally likely, then $E(X)$ is simply the average of the possible values of X. $\mathcal{X}=\{x_1,\cdots,x_n\}$, $\frac{1}{n},\cdots,\frac{1}{n}$, $E(X)=\sum_{i=1}^{n} x_i\cdot\frac{1}{n} = \frac{x_1+\cdots+x_n}{n}$
  - ➢Example (Committees). In the committees example, → LNp.4-5

$$E(X) = 0\cdot\frac{5}{210} + 1\cdot\frac{50}{210} + 2\cdot\frac{100}{210} + 3\cdot\frac{50}{210} + 4\cdot\frac{5}{210} = 2.$$

  - ➢Example (Indicator Function).
    - Ω ▪ For an event $A\subset\Omega$, the indicator function of A is the r.v.



$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega\in A, \\ 0, & \text{if } \omega\notin A. \end{cases}$$

$\{\mathbb{1}_A=1\}=\{A \text{ occurs}\}$
$\{\mathbb{1}_A=0\}=\{A \text{ not occurs}\}$

- ▪ Its range is $\{0, 1\}$ and its pmf is → $f(x)=0$, if $x\notin\mathcal{X}=\{0,1\}$.

$$f(0)=P(A^c)=1-P(A) \quad\text{and}\quad f(1)=P(A),$$

for a p.m. P defined on Ω.

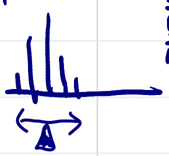
- ▪ So, $E(\mathbf{1}_A) = 0\cdot[1-P(A)] + 1\cdot P(A) = P(A)$

Note: Expectation may not be a value that the r.v. can occur.

➢Intuitive Interpretation of Expectation

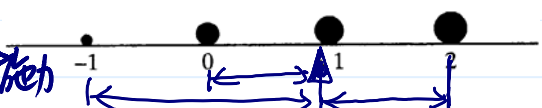
- ▪ Expectation of a r.v. parallels the notion of a weighted average, where more likely values are weighted higher than less likely values. 加權平均 — $\sum_{x\in\mathcal{X}} f(x) = 1$
- ▪ It is helpful to think of the expectation as the "center" of mass of the pmf $0=\sum x f_X(x) - E(X)\cdot\sum f_X(x) = \sum (x-E(X))\cdot f_X(x)$

pmf



- 重心 ▫ center of gravity: If we have a rod with weights f_X at each possible point x_i then the point at which the rod is balanced is called the center of gravity.

槓桿原理
力矩=距離×施力



$p(-1) = .10,\quad p(0) = .25,\quad p(1) = .30,\quad p(2) = .35$

$-1\times 0.1+0\times 0.25+1\times 0.3+2\times 0.35 =$

$\wedge$ = center of gravity = .9

- ▪ Expectation can be interpreted as a long-run average (Chapter 8)

e.g. repeat 10000 times
≈ 1000 times −1
2500 0
3000 1
3500 2
random
average ≈ 0.9.
deterministic

- Expectation of Transformation

➢ Theorem. If X is a discrete r.v. with range $\mathcal{X}$ and pmf f_X; let

$$Y = g(X),$$

and $\mathcal{Y}$ be the range of Y, f_Y be the pmf of Y, then

$$E(Y) \equiv \sum_{y\in\mathcal{Y}} y f_Y(y) = \sum_{x\in\mathcal{X}} g(x) f_X(x),$$

provided that the sum converges absolutely.

To calculate $E(Y)$, There is no need to calculate $f_Y(y)$

proof.
$$\sum_{x\in\mathcal{X}} g(x)f_X(x) = \sum_{y\in\mathcal{Y}} \sum_{\substack{x\in\mathcal{X} \\ g(x)=y}} g(x)f_X(x) = \sum_{y\in\mathcal{Y}} y \sum_{\substack{x\in\mathcal{X} \\ g(x)=y}} f_X(x) = \sum_{y\in\mathcal{Y}} y f_Y(y)$$

by Thm (LNp.4-10)

$\mathcal{X}$ $\xrightarrow{g}$ $\mathbb{R}$ (Y)

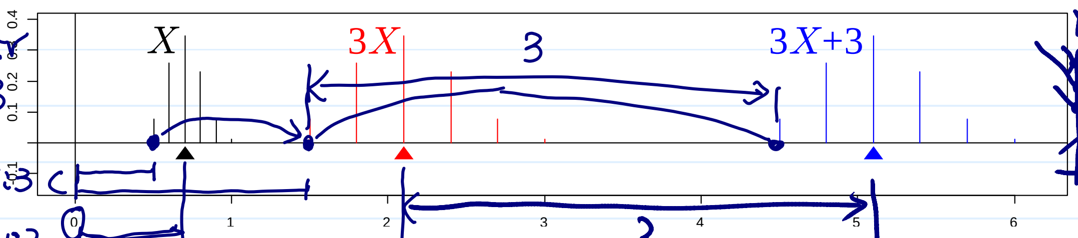
- Example. $E(X^2) = \sum_{x\in\mathcal{X}} x^2 f_X(x)$.

➢ Theorem. For $a, b \in \mathbb{R}$, $E(aX+b) = a\cdot E(X)+b$.

fixed.

proof. $E(aX+b) = \sum_{x\in\mathcal{X}} (ax+b)f_X(x) = a[\sum_{x\in\mathcal{X}} x f_X(x)] + b[\sum_{x\in\mathcal{X}} f_X(x)]$

ratio of their length = 1:3



3

$\mathbb{R}(Y)$

ax+b

$\mathbb{R}$

X

1:3

1:3

3

- Mean and Variance. 變異數

平均值

➢ Definition. The expectation of X is also called the *mean* of X and/or f_X. The variance of X (and/or f_X) is defined by

another name.

$$Var(X) \equiv E[(X-\mu_X)^2] = \sum_{x\in\mathcal{X}} (x-\mu_X)^2 f_X(x).$$

provided that the sum converges.

- Example (Committees)

LNp.4-5

x	$x-\mu$	$f(x)$	$xf(x)$	$(x-\mu)^2 f(x)$	$x^2 f(x)$
0	-2	5/210	0/210	20/210	0/210
1	-1	50/210	50/210	50/210	50/210
2	0	100/210	200/210	0/210	400/210
3	1	50/210	150/210	50/210	450/210
4	2	5/210	20/210	20/210	80/210
Totals		1	2	2/3	14/3

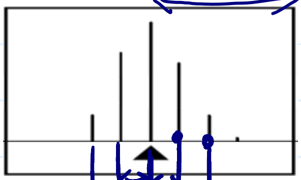
So, $\mu_X = 2$ and $\sigma^2 = 2/3$, $\sigma = \sqrt{2/3}$

$E(X^2)$

- The $E(X)$ is often denoted by μ_X and $Var(X)$ by σ_X^2. Also, $\sigma_X = \sqrt{\sigma_X^2}$ is called the *standard deviation* of X.

- Note.
  - μ_X and σ_X^2 only depends on f_X. They are fixed constants, not random.
  - If X has units, then μ_X and σ_X have the same unit as X and variance has unit squared.

➢ Intuitive Interpretation of Variance



- Variance is the weighted average value of the squared deviation of X from μ_X.
- Variance is related to how the pmf is spread out

$\sigma^2 > 0$

➢ Some properties of variance. $Var(X) \geq 0 \Leftarrow$ why?

- The variance of a r.v. is always non-negative
- The only r.v. with variance equal to zero is a r.v. which can only take on a single value.

For discrete r.v.'s | pf $\sum_{x\in\mathcal{X}}(x-\mu_X)^2 f_X(x)=0$ $\Leftrightarrow (x-\mu_X)^2 f_X(x)=0, \forall x\in\mathcal{X}$

$\Leftrightarrow (x-\mu_X)^2=0$ if $f_X(x)>0$

$\Leftrightarrow P(X=\mu_X)=1$. pmf μ_X 1

➢ Theorem. For $a, b \in \mathbb{R}$, $Var(aX+b) = a^2\, Var(X)$

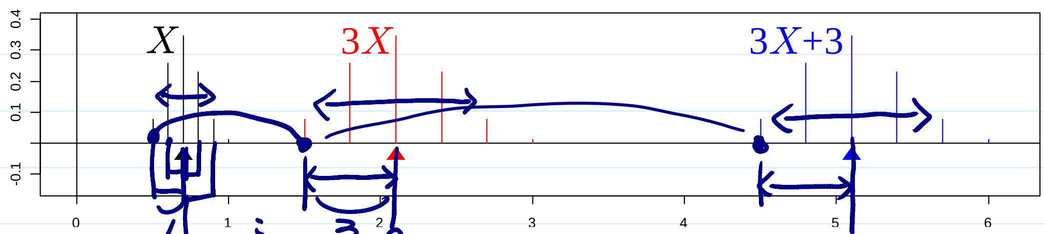
proof. Let $Y = aX + b$, then $E(Y) = a\cdot\mu_X + b \equiv \mu_Y$.

$$\begin{aligned} Var(Y) &= E(Y-\mu_Y)^2 = E[(aX+\cancel{b})-(a\mu_X+\cancel{b})]^2 \\ &= E[a^2(X-\mu_X)^2] = a^2E(X-\mu_X)^2 = a^2Var(X) \end{aligned}$$

$a>1$

$0<a<1$

$a<0$



fixed

➢ Theorem. If X is a discrete r.v. with mean μ_X, then for any $c\in\mathbb{R}$,

Mean Square Error = Var + bias^2

$$(*)\quad E[(X-c)^2] = \sigma_X^2 + (c-\mu_X)^2.$$

proof. $E[(X-c)^2] = E[(X-\mu_X+\mu_X-c)^2] = \sum_{x\in\mathcal{X}}[(x-\mu_X+\mu_X-c)^2]f_X(x)$

$$= \sum_{x\in\mathcal{X}}[(x-\mu_X)^2 + 2(x-\mu_X)(\mu_X-c) + (\mu_X-c)^2]f_X(x)$$

$$= \underbrace{\sum_{x\in\mathcal{X}}(x-\mu_X)^2 f_X(x)}_{\sigma_X^2} + 2(\mu_X-c)\underbrace{\sum_{x\in\mathcal{X}}(x-\mu_X)f_X(x)}_{0} + (\mu_X-c)^2\sum_{x\in\mathcal{X}}f_X(x)$$

- Corollary. $E[(X-c)^2]$ is minimized by letting $c=\mu_X$; and the minimum value is σ_X^2. proof: clear from (*)
- Corollary. $\sigma_X^2 = E(X^2) - (E(X))^2$. proof: Let the c in (*) be zero.

  useful for calculating σ_X^2

  (Recall: $E(X^2) = \sum_{x\in\mathcal{X}} x^2 f_X(x)$)
  - Example (Committees). $Var(X)=14/3-2^2=2/3$.

➢ $E(X^n)$ is often called the n^{th} *moment* of X Var = (2nd moment) − (1st moment)^2

❖ **Reading**: textbook, Sec 4.3, 4.4, 4.5

Some Common Discrete Distributions

伯努利

- Bernoulli and Binomial Distributions 二項
  - ➢Experiment: A basic experiment with sample space Ω_0 is repeated n times. P_0?
    - ▪ Example. (a) Sampling with replacement (b) Coin Tossing (c) Roulette

      $\Omega=\{\omega=(\omega_1,\omega_2,\cdots,\omega_n):\ \omega_i\in\Omega_0\}$

    - ▪ The sample space for the n trials is

      $$\Omega=\Omega_0\times\cdots\times\Omega_0=\Omega_0^{\,n} \leftarrow \text{List}$$

      e.g. different $k_1,k_2,\cdots,k_r\le n$

      A: k_1,k_3,k_{10} trials

      B: k_2,k_5,k_8 trials

    - ▪ **Assume** that events depending on different trials are independent — ith

      e.g. ith trial, $A\subset\Omega_0\Rightarrow\Omega_0\times\cdots\times\Omega_0\times A\times\Omega_0\times\cdots\times\Omega_0\equiv A'$

      jth trial, $B\subset\Omega_0\Rightarrow\Omega_0\times\cdots\times\Omega_0\times B\times\Omega_0\times\cdots\times\Omega_0\equiv B'$, $j\neq i$

    - ▪ **Q**: Given an event $A_0\subset\Omega_0$, what is the probability that A_0 occurs k times in the n trials?

      $\Omega_0=\mathbb{R}$

    - ▪ Problem Formulation: Let $A_i\subset\Omega$ be

      $A_1: A_0\times\Omega_0\times\cdots\times\Omega_0$

      $A_2: \Omega_0\times A_0\times\cdots\times\Omega_0$

      $A_n: \Omega_0\times\cdots\times\Omega_0\times A_0$

      A', B' indep

      $A_i=\{A_0$ occurs on the i^{th} trial$\}$, and

      $\Omega\to\mathbb{R}$, r.v.

      $$X=\underset{0/1}{\mathbf{1}_{A_1}}+\cdots+\underset{0/1}{\mathbf{1}_{A_n}}\Rightarrow \#\text{ that } A_0 \text{ occurs in the } n \text{ trials.}$$

      Q: What is $P(X=k)$? LNp.4-11

      (Note. $A_1,\ldots,A_n$ are assumed to be independent events.)

- ▪ Example (Roulette, n=4, k=2, LNp.2-3). $\Omega_0\to\Omega_0^4=\Omega$, A_0

  mutually indep: W_1,W_2,W_3,W_4

  W_1, W_1^c $\quad W_2, W_2^c$

  $B_1\ B_2\ B_3\ B_4$ mutually indep.

  - ▫ Let $W_i=\{\text{Win on } i^{\text{th}} \text{ Game}\}$

    $L_i=W_i^c=\{\text{Lose on } i^{\text{th}} \text{ Game}\}$.

    Then, $P(W_i)$=9/19 ≡ p and $P(L_i)$=10/19=1−p ≡ q

    Let $X=\mathbf{1}_{W_1}+\mathbf{1}_{W_2}+\mathbf{1}_{W_3}+\mathbf{1}_{W_4}$, then

    $$\begin{aligned}\{X=2\}=&(W_1\cap W_2\cap L_3\cap L_4)\cup(W_1\cap L_2\cap W_3\cap L_4)\\ &\cup(W_1\cap L_2\cap L_3\cap W_4)\cup(L_1\cap W_2\cap W_3\cap L_4)\\ &\cup(L_1\cap W_2\cap L_3\cap W_4)\cup(L_1\cap L_2\cap W_3\cap W_4)\end{aligned}$$

  - ▫ So, ⇒ mutually exclusive

    $$\begin{aligned}P(\{X=2\}) &= P(W_1\cap W_2\cap L_3\cap L_4)+\cdots\\ &\quad +P(L_1\cap L_2\cap W_3\cap W_4)\\ &\overset{\downarrow}{=} P(W_1)P(W_2)P(L_3)P(L_4)+\cdots\\ &\quad +P(L_1)P(L_2)P(W_3)P(W_4)\\ &= ppqq+pqpq+pqqp\\ &\quad +qppq+qpqp+qqpp\\ &= 6p^2q^2.\end{aligned}$$

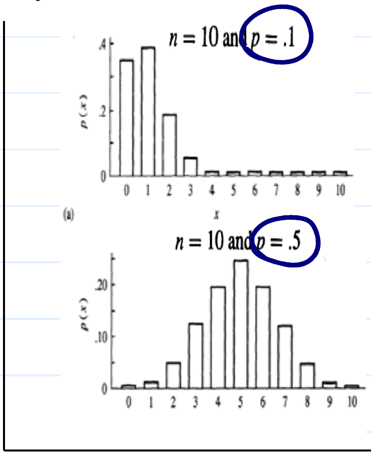
    ∵ mutually indep

    symmetry

    $\binom{4}{2}$

➢ Probability Mass Function

- Let $A_1, \ldots, A_n$ be independent events and $P(A_i)=p$, $i=1, \ldots, n$.
- Let $X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}$.
- Then, for $k=0, 1, \ldots, n$,



$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

proof. We may choose k trials in $\binom{n}{k}$ ways.

Say, $\{1, 2, 3, \ldots, k\}$ is chosen.

$0+1+1 \cdots +0 = X$

$\boxed{1}\ \boxed{2} \bullet\bullet\bullet \boxed{n}$

$1/0 \quad 1/0 \quad 1/0$

$$\begin{aligned} & P(A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c) \\ = \; & P(A_1) \times \cdots \times P(A_k) \times P(A_{k+1}^c) \times \cdots \times P(A_n^c) \\ = \; & p^k (1-p)^{n-k} \end{aligned}$$

- (exercise) Show that the following function is a pmf.

check (i)(ii)(iii) in LN p.4-5
For (iii), use △

$$f(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

- The distribution of the r.v. X is called the *binomial* distribution with parameters n and p. In particular, when $n=1$, it is called the *Bernoulli* distribution with parameter p.

each 1_{A_i} follow Bernoulli dist.

- Notice that a binomial r.v. can be regarded as the sum of n independent Bernoulli r.v.'s.

  For definition of indep. r.v.'s, see chapter 6 (future lecture)

- The binomial distribution is called after the Binomial Theorem: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$. — △

  a: p, b: $1-p$

- Example (Bridge). **Q**: What is the probability that South gets $\boxed{\text{no Aces}}$ on at least $k=5$ of $n=9$ hands? $\Omega_0 \to \Omega_0^9 = \Omega$

  - Let A_i={no Aces on the i^{th} hand}, $i=1, 2, \ldots, 9$, and

    no Aces's

$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_9}.$$

Bernoulli(p)

$A_1, \cdots, A_9$ mutually indep & have identical prob.

  - Then, $P(A_i) = \binom{48}{13} \Big/ \binom{52}{13} \approx 0.3038 \equiv p$.
  - So, for $k = 0, 1, \ldots, 9$,

$$P(X=k) = \binom{9}{k} p^k (1-p)^{n-k}.$$

  - And,

$$P(X \geq 5) = \sum_{k=5}^{9} \binom{9}{k} p^k (1-p)^{n-k} \approx 0.1035.$$

➢ Theorem. The mean and variance of the Binomial(n, p) distribution are

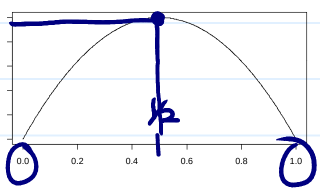
$$\mu = np \quad \text{and} \quad \sigma^2 = np(1-p).$$

intuitive interpretation

intuition

Why?

n/4



proof.

Sum-To-One (STO) method

$$E(X)=\sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=1}^{n} \cancel{x}\cdot\frac{n!}{\cancel{x}!(n-x)!}p^x(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n\cdot(n-1)!}{(x-1)!(n-x)!}p^x(1-p)^{n-x}$$

let $y=x-1$ $= np\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}$ — pmf of binomial$(n-1,p)$

$$= np\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{(n-1)-(x-1)} = np$$

Note: $E[g_1(X)+g_2(X)] = \sum(g_1(x)+g_2(x))f_X(x) = \sum g_1(x)f_X(x) + \sum g_2(x)f_X(x) = E[g_1(X)] + E[g_2(X)]$

$$E[X(X-1)] = E(X^2-X) = E(X^2)-E(X)$$

$$= \sum_{x=0}^{n} x(x-1)\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=2}^{n}\cancel{x}\cancel{(x-1)}\cdot\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$= \sum_{x=2}^{n}\frac{n(n-1)\cdot(n-2)!}{(x-2)!(n-x)!}p^x(1-p)^{n-x}$$

let $y=x-2$ $= n(n-1)p^2\sum_{y=0}^{n-2}\binom{n-2}{y}p^y(1-p)^{(n-2)-y}$ — pmf of binomial$(n-2,p)$

$$= n(n-1)p^2\sum_{x=2}^{n}\binom{n-2}{x-2}p^{x-2}(1-p)^{(n-2)-(x-2)} = n(n-1)p^2$$

$$Var(X) = E(X^2)-[E(X)]^2 = [E(X^2)-E(X)]+E(X)-[E(X)]^2$$
$$= n(n-1)p^2+np-n^2p^2 = np(1-p)$$

➢Summary for X ~ Binomial(n, p)

- Range: $\mathcal{X}=\{0,1,2,\ldots,n\}$
- Pmf: $f_X(x)=\binom{n}{x}p^x(1-p)^{n-x}$, for $x\in\mathcal{X}$
- Parameters: $n\in\{1, 2, 3, \ldots\}$ and $0\le p\le 1$
- Mean: $E(X)=np$
- Variance: $Var(X)=np(1-p)$

regarded as fixed constant in the distribution

c.f.

• Geometric and Negative Binomial Distributions 幾何 負二項

➢Experiment: A basic experiment with sample space Ω_0 is repeated *infinite* times.

- The sample space is $\Omega=\{\omega=(\omega_1,\omega_2,\cdots):\omega_i\in\Omega_0\}$

$$\Omega=\Omega_0\times\Omega_0\times\Omega_0\times\cdots$$

- **Assume** that events depending on different trials are independent
- For a given event $A_0\subset\Omega_0$, we continue performing the trials until A_0 occurs exactly r times
- **Q**: What is the probability that we need to perform k trials?

- Example.
  - A company must hire 3 engineers.
  - Each interview results in a hire with probability 1/3
  - **Q**: What is the probability that 10 interviews are required?
  - We need: (i) 2 hires on the first 9 interview (ii) Success on the 10th interview
  - So, the probability is $\underbrace{\binom{9}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^7}_{\text{binomial pmf.}} \times \left(\frac{1}{3}\right) = \binom{9}{2}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^7.$

two "1" & seven "0"

- Problem Formulation:
  - Let $A_1, A_2, \ldots \subset \Omega$ be $A_i = \Omega_0 \times \cdots \times \Omega_0 \times \underset{i\text{th}}{A_0} \times \Omega_0 \times \cdots$

    $A_i = \{A_0$ occurs on the i^{th} trial$\}$, and $\{X_n \geq r\} = \{Y_r \leq n\}$

    Binomial(n,p) $\sim$ $X_n = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}$, for $n = 1, 2, 3, \ldots$.
  - Let Y_1 = smallest n with $X_n \geq 1$,

    Y_2 = smallest n with $X_n \geq 2$,

    …,

    # of trials to produce r {A_0 occurs} = Y_r = smallest n with $X_n \geq r$,
  - **Q**: What is $P(Y_r = k)$?

$X_1 = 0,\ X_2 = 1,\ X_3 = 2,\ X_4 = 2,\ X_5 = 2,\ X_6 = 3,\ X_7 = 3$

0 1 1 0 0 1 0

Y_1 Y_2 Y_3

➢ Probability Mass Function

- Let $A_1, A_2, \ldots$ be independent and $P(A_i)=p$, $i=1, 2, 3, \ldots$.
- Then, for $k=r, r+1, r+2, \ldots$,

$$P(Y_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

proof. If $r = 1$, $P(Y_1 = k) = P(\{X_{k-1} = 0\} \cap A_k)$

$$= P(\{X_{k-1} = 0\}) \cdot P(A_k) = \underbrace{(1-p)^{k-1} p}_{\text{pmf of geometric}}$$

binomial(k-1, p)

In general, $P(Y_r = k) = P(\{X_{k-1} = r-1\} \cap A_k)$

$$= P(\{X_{k-1} = r-1\}) \cdot P(A_k)$$

$$= \underbrace{\binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} p}_{\text{pmf of negative binomial.}}$$

- (exercise) Show that the following function is a pmf.

For (iii) in LNp4-5, apply △

$$f(k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & k = r, r+1, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

- The distribution of the r.v. Y_r is called the *negative binomial* distribution with parameters r and p. In particular, when $r=1$, it is called the *geometric* distribution with parameter p.

Bernoulli Binomial c.f.

- A negative binomial r.v. can be regarded as the sum of r independent geometric r.v.'s. $R_1+R_2+\cdots+R_r=R$

$$\underbrace{\overset{0}{\boxed{1}}\ \overset{0}{\boxed{2}}\ \cdots\ \overset{0}{\boxed{\ }}\ \overset{1\ (1\text{st})}{\boxed{R_1}}}_{\text{geometric}}\ \Big|\ \underbrace{\overset{0}{\boxed{1}}\ \overset{0}{\boxed{2}}\ \cdots\ \overset{0}{\boxed{\ }}\ \overset{1\ (2\text{nd})}{\boxed{R_2}}}_{\text{geometric}}\ \Big|\ \cdots\ \Big|\ \underbrace{\overset{0}{\boxed{1}}\ \overset{0}{\boxed{2}}\ \cdots\ \overset{0}{\boxed{\ }}\ \overset{r\text{th}}{\boxed{R_r}}}_{\text{geometric}}\ \Big|$$

- The negative binomial distribution is called after the Negative Binomial Theorem: $(x+a)^{-r}=\sum_{k=0}^{\infty}\binom{r+k-1}{k}(-x)^k a^{-r-k}$ ($-x \to -t$, $a \to 1$)

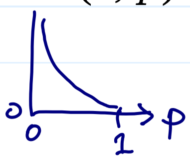
geometric distribution ↑ pmf is a geometric sequence (幾何數列, 等比數列)

$$2.\quad \frac{1}{(1-t)^r}=\sum_{k=0}^{\infty}\binom{r+k-1}{k}t^k,\quad \text{for } |t|<1.$$

➢ Theorem. The mean and variance of negative binomial(r, p) is

$$\mu = r/p \quad \text{and} \quad \sigma^2 = r(1-p)/p^2.$$

intuitive interpretation



proof.

$$E(X)=\sum_{x=r}^{\infty}x\binom{x-1}{r-1}p^r(1-p)^{x-r}=\frac{r}{p}\sum_{x=r}^{\infty}\frac{x\cdot(x-1)!}{r\cdot(r-1)!(x-r)!}p^{r+1}(1-p)^{x-r}$$

$x\cdot(x-1)! = ((x+1)-1)!$, $r\cdot(r-1)! = ((r+1)-1)!$, $x = (x+1)-1$, $r = (r+1)-1$

$$=\frac{r}{p}\sum_{x=r}^{\infty}\binom{(x+1)-1}{(r+1)-1}p^{r+1}(1-p)^{(x+1)-(r+1)}$$

let $y=x+1$

$$=\frac{r}{p}\sum_{y=r+1}^{\infty}\underbrace{\binom{y-1}{(r+1)-1}p^{r+1}(1-p)^{y-(r+1)}}_{\text{pmf of negative binomial}(r+1,\,p)}=r/p$$

$$E[X(X+1)]=E(X^2+X)=E(X^2)+E(X)$$

$$=\sum_{x=r}^{\infty}x(x+1)\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

$$=\frac{r(r+1)}{p^2}\sum_{x=r}^{\infty}\frac{(x+1)x\cdot(x-1)!}{(r+1)r\cdot(r-1)!(x-r)!}p^{r+2}(1-p)^{(x+2)-(r+2)}$$

$(x+1)x\cdot(x-1)! = ((x+2)-1)!$, $(r+1)r\cdot(r-1)! = ((r+2)-1)!$, $x = (x+2)-1$, $r = (r+2)-1$

$$=\frac{r(r+1)}{p^2}\sum_{x=r}^{\infty}\binom{(x+2)-1}{(r+2)-1}p^{r+2}(1-p)^{(x+2)-(r+2)}$$

$y=x+2$

$$=\frac{r(r+1)}{p^2}\sum_{y=r+2}^{\infty}\underbrace{\binom{y-1}{(r+2)-1}p^{r+2}(1-p)^{y-(r+2)}}_{\text{pmf of negative binomial}(r+2,\,p)}$$

$$=r(r+1)/p^2$$

$$Var(X)=E(X^2)-[E(X)]^2=[E(X^2)+E(X)]-E(X)-[E(X)]^2$$

$$=\frac{r(r+1)}{p^2}-\frac{r}{p}-\frac{r^2}{p^2}=\frac{r(1-p)}{p^2}$$

➢ Summary for $X \sim$ Negative Binomial(r, p)

- Range: $\mathcal{X}=\{r, r+1, r+2, \ldots\}$
- Pmf: $f_X(x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$, for $x\in\mathcal{X}$
- Parameters: $r\in\{1, 2, 3, \ldots\}$ and $0\le p\le 1$
- Mean: $E(X)=r/p$
- Variance: $Var(X)=r(1-p)/p^2$

- Poisson Distribution

卜卜松 卜瓦松

$X_1 \sim$ binomial$(1, p_1)$ ... EX: p_1
$X_m \sim$ binomial(m, p_m) mp_m
$X_{m+1} \sim$ $(m+1, p_{m+1})$ $(m+1)p_{m+1}$
X_n : (n, p_n) np_n
$\downarrow \infty$ $\downarrow 0$ $\downarrow \lambda$

➢ Recall: Expression for e^x, e=2.7183⋯
  - First Expression: $e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$
  - Second Expression: $e^x = \sum_{k=0}^{\infty}\frac{1}{k!}x^k$.

➢ The Derivation
  - Consider a sequence of binomial(n, p_n) distributions satisfying
    (a) $p_n \to 0$ when $n \to \infty$
    (b) $n\cdot p_n \to \lambda$ when $n \to \infty$, where $0 < \lambda < \infty$

    $\lim_{n\to\infty} P(X_n = k) = ??$
  - Then, $p_n \approx \lambda/n$ when n is large enough.
  - And, $\binom{n}{k} p_n^k (1-p_n)^{n-k}$

    difficult to calculate when n is large & $k \ll n$.

    $\frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$

    $$\approx \frac{1}{k!}(n)_k\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{1}{k!}\lambda^k\frac{(n)_k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

    $\left(1+\frac{-\lambda}{n}\right)^n \cdot \left(1+\frac{-\lambda}{n}\right)^{-k} \to 1$
  - Here, for each fixed k,

    $\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n} =$, each factor $\to 1$

    $$\lim_{n\to\infty}\frac{(n)_k}{n^k} = 1 \quad \text{and} \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}.$$
  - So, when n large and $n \gg k$, $\frac{(n)_k}{n^k} \approx 1$ and $\left(1-\frac{\lambda}{n}\right)^{n-k} \approx e^{-\lambda}$.

  - In other words, when $\boxed{n \text{ large}}$, $\boxed{n \gg k}$, and $\boxed{p_n \approx 0}$,

Poisson($\lambda = np_n$)
ss pmf similar for $k \ll n$.

$$\binom{n}{k}p_n^k(1-p_n)^{n-k} \approx \boxed{\frac{1}{k!}\lambda^k e^{-\lambda}}.$$ ← pmf of Poisson.

binomial(n, p_n) a lot of "0", relatively few "1" $p_n \approx 0$

$X = k = 0+0+0+0+0+0+0+1+0 \ 0 \ 0 \ 0 \ 0+1+0+0 \ldots 0+0+0$

1 2 3 4 5 6 7 8 9 10 11 12 • • • n-2 n-1 n ← n large

➢ Example. - - -
  - A professor hits the wrong key with probability p=0.001 each time he types a letter. Assume independence for the occurrence of errors between different letter typings.

    $p \approx 0$

    typing: indep. Bernoulli(p)
  - **Q**: P(5 or more errors in n=2500 letters)=??

    n large.
  - Ans.
    - Let X be the number of errors, then X~binomial(2500, 0.001) and $P(X \geq 5)$

      $$P(5 \text{ or more errors}) = 1 - P(X \leq 4)$$

      $k \ll n$ (*)

      $$= 1 - \sum_{k=0}^{4}\binom{2500}{k}(0.001)^k(0.999)^{2500-k}.$$

      may be difficult to calculate.

- The probability can be approximated by $\lambda^k e^{-\lambda}/k!$ with

$$\lambda = 2500 \times 0.001 = 2.5 \text{ times of errors,}$$

where 2.5 is the *expected number* of the errors that would occur in the 2500 typings. (**Q**: What should the λ's be for 5000 typings, 7500 typings, and 10000 typings?)

$\lambda=5$ $\lambda=7.5$ $\lambda=10.$

- So, $P(X = k) \approx (2.5)^k e^{-2.5}/k!$, for k=0,1,2,3,4, and

$$1 - P(X \le 4) \approx 1 - \sum_{k=0}^{4} \underbrace{\frac{(2.5)^k e^{-2.5}}{k!}} = 0.1088.$$

cf. (*) in LNp.4-28 which is easier to obtain or calculate?

Probability Mass Function

- Theorem. Let

$$f(k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!}, & k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

then, $f(k)$ is a pmf.

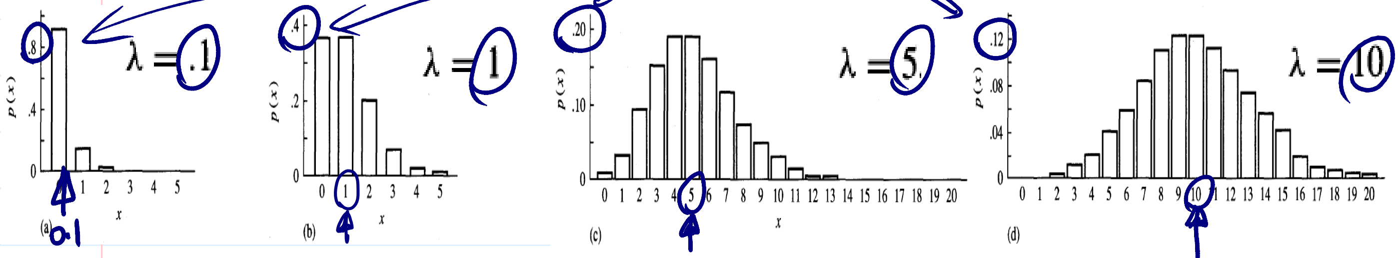
proof. LNp.4-5, (i) & (ii) are straightforward. For (iii),

$$\sum_{k=0}^{\infty} f(k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = e^{-\lambda} \cdot e^{\lambda} = 1.$$

- The pmf is called the *Poisson* pmf with parameter λ. The distribution is named after Simeon Poisson, who derived the approximation of Poisson pmf to binomial pmf.

- The λ can be interpreted as the *average occurrence frequency*.

different



Theorem. The mean and variance of Poisson(λ) is

$$\boxed{\text{intuitive interpretation}} \quad \mu \stackrel{\downarrow}{=} \lambda \quad \text{and} \quad \sigma^2 = \lambda.$$

proof.

let $y=x-1$

$$E(X) = \sum_{x=0}^{\infty} \cancel{x} \cdot \frac{e^{-\lambda}\lambda^x}{\underset{x-1}{\cancel{x}}!} = \lambda \cdot \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \stackrel{\downarrow}{=} \lambda \cdot \sum_{y=0}^{\infty} \underbrace{\frac{e^{-\lambda}\lambda^y}{y!}}_{\text{pmf of Poisson}(\lambda)} = \lambda$$

$$E[X(X-1)] = E(X^2 - X) = E(X^2) - E(X) \quad y=x-2$$

$$= \sum_{x=0}^{\infty} \cancel{x(x-1)} \cdot \frac{e^{-\lambda}\lambda^{\overset{x-2}{x}}}{\underset{x-2}{x}!} = \lambda^2 \cdot \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} \stackrel{\downarrow}{=} \lambda^2 \cdot \sum_{y=0}^{\infty} \frac{\overline{e^{-\lambda}\lambda^y}}{y!} = \lambda^2$$

# of times some event occurring during a period of time

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= \quad [E(X^2) - E(X)] + E(X) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

- Note: For X~binomial(n, p), where (i) n large; (ii) p small,
  - distribution of $X \approx$ Poisson(λ=np)
  - $E(X) = np$ = mean of the Poisson = λ
  - $Var(X)$= $\underbrace{np}_{\approx\lambda}\underbrace{(1-p)}_{\approx 1}$ ≈ variance of the Poisson = λ

➢ Poisson Process

an event

a period of time

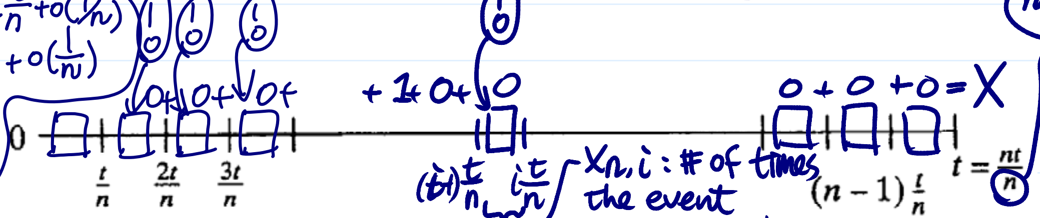
- Example:
  - (1) # of earthquakes occurring during some fixed time span
  - (2) # of people entering a bank during a time period

$P(X_{n,i}=0) = 1 - \lambda \frac{t}{n} + o(\frac{1}{n})$

$P(X_{n,i}=1) = \lambda \cdot \frac{t}{n} + o(\frac{1}{n})$

$P(X_{n,i}\geq 2) = o(\frac{1}{n})$

n large



$X_{n,i}$: # of times the event occurs in the ith small interval

To model them, we can

- Divide the time period, say [0, t], into n small intervals of equal length
- Make the intervals so small (i.e., n is large) that at most one event can occur in each interval

$\Rightarrow$ Then, we can treat the number of events in a single interval as a Bernoulli r.v. with a small p_n $= P(X_{n,i}=1) \approx \lambda \cdot \frac{t}{n}$

- Assume that the number of events to occur in non-overlapping intervals are independent

$\Rightarrow$ Now, the number of events in the whole period of time [0, t] is binomial(n, p_n), where n is a quite large number and p_n is a small probability & $n \cdot p_n \approx n \cdot \lambda \frac{t}{n} = \lambda t$

- The distribution for the number of events occurring in [0, t] can be approximated by Poisson($n \cdot p_n$) $\approx \lambda t \Rightarrow \lambda = \frac{np_n}{t}$

- Definition. A *Poisson process* with *rate* λ is a family of r.v.'s N_t, $0 \leq t < \infty$, for which

=次/單位時間

$$N_0 = 0 \quad \text{and} \quad N_t - N_s \sim \text{Poisson}(\lambda \cdot (t-s)),$$

↕ c.f. previous λ.

for $0 \leq s < t < \infty$, and

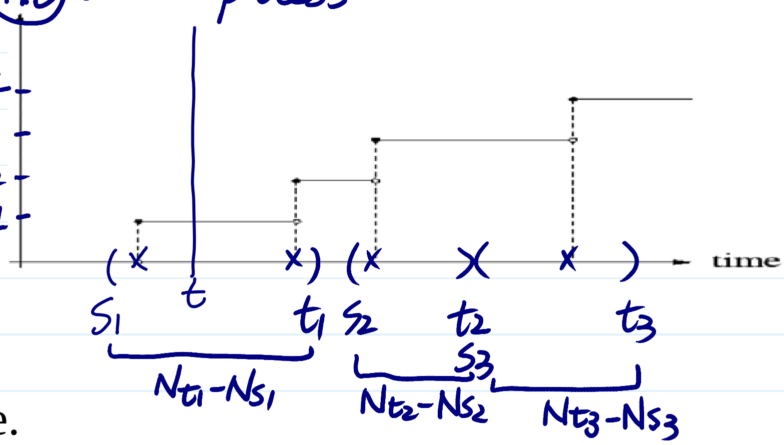
$$N_{t_i} - N_{s_i},\ i = 1, 2, \ldots, m$$

are independent whenever

$$0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_m < t_m.$$

- Here, N_t denotes the # of events that occurs by time t
- λ is the average # of events occurring per unit time

the X_n in LN p.4-23 c.f. N_t ← count process



- Example.
  - Traffic accident occurs (at 光復路&建功路口, e.g.) according to a Poisson process at a rate of λ=5.5 per month
    an event
  - **Q**: What is the probability of 3 or more accidents occur in a 2 month periods?
    a period of time
  - Here, $\lambda t = 5.5\times 2 = 11$. (**Q**: What should λt be for one and half months? for a year?)
  - So, $P(N_2 = k) = 11^k \cdot e^{-11}/k!$ and

$$P(N_2 \geq 3) = 1 - P(N_2 \leq 2) = 1 - \sum_{k=0}^{2} \frac{e^{-11}\cdot 11^k}{k!}$$

- Summary for $X\sim$ Poisson(λ) ← c.f. → Q: What is the sample space behind the X? ↔ binomial Ω, negative binomial Ω
  - Range: $\mathcal{X} = \{0, 1, 2, \ldots\}$
  - Pmf: $f_X(x) = \lambda^x e^{-\lambda}/x!$, for $x\in\mathcal{X}$
  - Parameter: 0<λ< ∞
  - Mean: $E(X)$=λ
  - Variance: $Var(X)$=λ

期中考考到此

- Hypergeometric Distribution 超幾何
  - Experiment: Draw a sample of n ($\leq N$) balls *without replacement* from a box containing R red balls and $N-R$ white balls (the balls are equally likely to be drawn)
    - Let X be the number of red balls in the sample
    - **Q**: What is $P(X=k)$?
    - Example. The Committee Example ← LN p.4-4 & 5
    - (c.f.) If drawn *with replacement*, what is the distribution of X?

$\underbrace{\underset{\boxed{1}}{0} + \underset{\boxed{2}}{1} + 1 + \cdots + \underset{\boxed{n-1}}{0} + \underset{\boxed{n}}{1}}$ $= X$

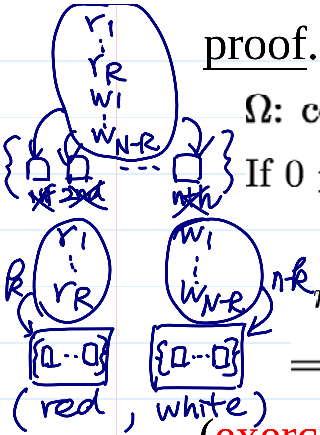
with replacement ⇒ indep. case → Binomial

without replacement ⇒ dependent case → hypergeometric

  - Probability Mass Function
    - Theorem. For $k = 0, 1, 2, \ldots, n$,

$$P(X=k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

(Notice that $\binom{r}{t}\equiv 0$ when either t<0 or r<t.)



proof. Label the N balls as $r_1, \ldots, r_R, w_1, \ldots, w_{N-R}$.

Ω: combinations of size n from N different balls. $\Rightarrow \#\Omega = \binom{N}{n}$

If $0 \le k \le R$ and $0 \le n-k \le N-R$,

k red balls may be chosen in $\binom{R}{k}$ ways.

$n-k$ white balls may be chosen in $\binom{N-R}{n-k}$ ways.

$$\Longrightarrow \#\{X=k\} = \binom{R}{k}\binom{N-R}{n-k}$$

- (exercise) Show that the following function is a pmf.

LN p.4-5 (i)(ii) straightforward

(iii) $\sum f(k)=1$ by $\triangle 3$

$$f(k) = \begin{cases} \binom{R}{k}\binom{N-R}{n-k}\Big/\binom{N}{n}, & k = 0, 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

- The distribution of the r.v. X is called the *hypergeometric* distribution with parameters n, N, and R.
  - The hypergeometric distribution is called after the hypergeometric identity: $\binom{a+b}{r} = \sum_{k=0}^{r} \binom{a}{k}\binom{b}{r-k}$. — $\triangle 3$

$\left(\sum_{i=0}^{a}\binom{a}{i}x^i\right)\left(\sum_{j=0}^{b}\binom{b}{j}x^j\right) = (1+x)^a(1+x)^b = (1+x)^{a+b} = \sum_{r=0}^{a+b}\binom{a+b}{r}x^r$

➢ Theorem. The mean and variance of hypergeometric(n, N, R) are

intuitive interpretation

$$\mu = \frac{nR}{N} \quad \text{and} \quad \sigma^2 = \frac{nR(N-R)(N-n)}{N^2(N-1)}.$$

proof.

$$E(X) = \sum_{x=0}^{n} \cancel{x} \cdot \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} \cancel{x} \cdot \frac{\frac{R!}{x!(R-x)!} \cdot \binom{N-R}{n-x}}{\binom{N}{n}}$$

$$= \frac{nR}{N} \sum_{x=1}^{n} \frac{\binom{R-1}{x-1}\binom{(N-1)-(R-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}} \overset{y=x-1}{=} \frac{nR}{N} \sum_{y=0}^{n-1} \underbrace{\frac{\binom{R-1}{y}\binom{(N-1)-(R-1)}{(n-1)-y}}{\binom{N-1}{n-1}}}_{\text{pmf of hypergeometric }(n-1,\,N-1,\,R-1)} = \frac{nR}{N}$$

$$E[X(X-1)] = E(X^2 - X) = E(X^2) - E(X)$$

$$= \sum_{x=0}^{n} x(x-1) \cdot \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} = \sum_{x=2}^{n} \cancel{x(x-1)} \cdot \frac{\frac{R!}{x!(R-x)!} \cdot \binom{N-R}{n-x}}{\binom{N}{n}}$$

$$= \frac{n(n-1)R(R-1)}{N(N-1)} \sum_{x=2}^{n} \frac{\binom{R-2}{x-2}\binom{(N-2)-(R-2)}{(n-2)-(x-2)}}{\binom{N-2}{n-2}}$$

$$\overset{y=x-2}{=} \frac{n(n-1)R(R-1)}{N(N-1)} \sum_{y=0}^{n-2} \underbrace{\frac{\binom{R-2}{y}\binom{(N-2)-(R-2)}{(n-2)-y}}{\binom{N-2}{n-2}}}_{\text{pmf of hypergeometric }(n-2,\,N-2,\,R-2)} = \frac{n(n-1)R(R-1)}{N(N-1)}$$

$$Var(X) = E(X^2) - [E(X)]^2 = [E(X^2) - E(X)] + E(X) - [E(X)]^2$$

$$= \frac{n(n-1)R(R-1)}{N(N-1)} + \frac{nR}{N} - \left(\frac{nR}{N}\right)^2 = \frac{nR(N-R)(N-n)}{N^2(N-1)}$$

➢Theorem. Let $N_i \to \infty$ and $R_i \to \infty$ in such a way that

$$p_i \equiv R_i/N_i \to p,$$

where $0 < p < 1$, then

$$\underset{\text{pmf of hypergeometric}}{\frac{\binom{R_i}{k}\binom{N_i-R_i}{n-k}}{\binom{N_i}{n}}} \to \underbrace{\binom{n}{k}p^k(1-p)^{n-k}}_{\text{pmf of binomial}}.$$

intuition: when # of red & white balls are large, without replacement ≈ with replacement.

proof.

$$\frac{\binom{R_i}{k}\binom{N_i-R_i}{n-k}}{\binom{N_i}{n}} = \frac{R_i!}{k!(R_i-k)!}\cdot\frac{(N_i-R_i)!}{(n-k)![(N_i-R_i)-(n-k)]!}\cdot\frac{n!(N_i-n)!}{N_i!}$$

$$= \frac{n!}{k!(n-k)!}\cdot\left[\underbrace{\frac{R_i}{N_i}}_{\to p}\times\underbrace{\frac{R_i-1}{N_i}}_{\to p}\times\cdots\times\underbrace{\frac{R_i-k+1}{N_i}}_{\to p}\right]\cdot$$

$$\left[\underbrace{\frac{N_i-R_i}{N_i}}_{\to 1-p}\times\underbrace{\frac{(N_i-R_i)-1}{N_i}}_{\to 1-p}\times\cdots\times\underbrace{\frac{(N_i-R_i)-(n-k)+1}{N_i}}_{\to 1-p}\right]\cdot$$

$$\left[\underbrace{\frac{N_i}{N_i}}_{\to 1}\times\underbrace{\frac{N_i}{N_i-1}}_{\to 1}\times\cdots\times\underbrace{\frac{N_i}{N_i-n+1}}_{\to 1}\right]$$

$$\to \binom{n}{k}p^k(1-p)^{n-k}$$

➢Summary for $X \sim$ Hypergeometric(n, N, R)

- Range: $\mathcal{X} = \{0, 1, 2, \ldots, n\}$
- Pmf: $f_X(x) = \binom{R}{x}\binom{N-R}{n-x}/\binom{N}{n}$, for $x \in \mathcal{X}$
- Parameters: $n, N, R \in \{1, 2, 3, \ldots\}$ and $n \le N$, $R \le N$
- Mean: $E(X)=nR/N$
- Variance: $Var(X)=nR(N-R)(N-n)/(N^2(N-1))$

Summary

Q: how to characterize the distribution of a (discrete) r.v.?

random variable

$\Omega \xrightarrow{X} \mathbb{R}$

$P \leftrightarrow P_X$

discrete r.v.
continuous r.v.
mixed r.v.

LNp.49 (6)

1. pmf ⟷ LNp.4-5 (i)(ii)(iii)
2. cdf ⟷ LNp.4-7&8 (2)(3)(4)
3. mgf

Expectation
- mean — weighted average — $E(aX+b)=aE(X)+b$
- variance — how the distribution spread out — $Var(aX+b)=a^2Var(X)$, $Var(X)=E(X^2)-(EX)^2$

Transformation

Some common discrete r.v.
- Bernoulli, Binomial
- Geometric, negative binomial
- Poisson
- hypergeometric

❖ **Reading**: textbook, Sec 4.6, 4.7, 4.8.1~4.8.3