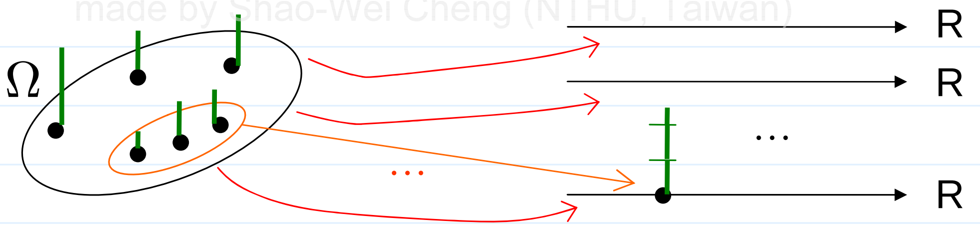
Random Variables

- A Motivating Example
  - ➢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} = $\{\{i_1, \ldots, i_k\}: 1 \le i_1 < \cdots < i_k \le n\}$
  - ➢A probability measure P can be defined on Ω, e.g, when there is an equally likely chance of being chosen for each students,

$$P(\{i_1, \ldots, i_k\}) = 1/\binom{n}{k}.$$

  - ➢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

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



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

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

  - ➢The P defined on Ω would be transformed into a *new* probability measure defined on 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*}

$$\begin{array}{lcccccccc}
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
X_1: & 3, & 2, & 2, & 2, & 1, & 1, & 1, & 0. \\
X_2: & 1, & 1, & 1, & 0, & 1, & 0, & 0, & 0. \\
X_3: & 3, & 1, & 1, & 1, & -1, & -1, & -1, & -3.
\end{array}$$

  - ➢ **Q**: Why particularly interested in functions that map to "R"?

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

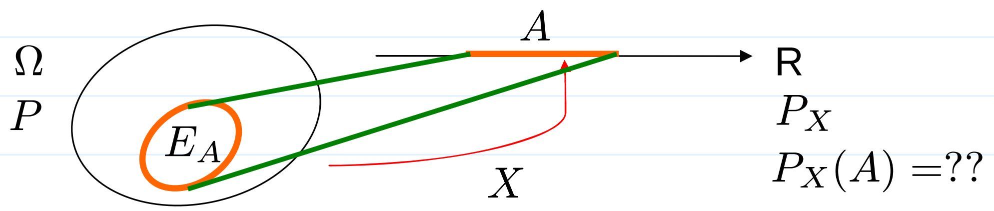
Ans: For a set (an event) $A \subset \mathbb{R}$,

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

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

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

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



Discrete Random Variables

- 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.,

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

➢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)

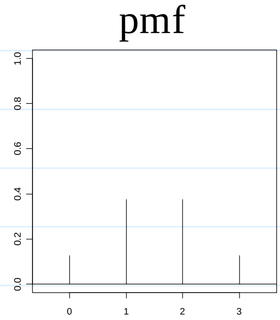
➢Definition: If X is a discrete r.v., then the *probability mass function* of X is defined by

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

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

- Example. For the X_1 in the Coin Tossing example,
  - $\mathcal{X} = \{0, 1, 2, 3\}$
  - $f_{X_1}(0) = 1/8, \quad f_{X_1}(1) = 3/8,$
    $f_{X_1}(2) = 3/8, \quad f_{X_1}(3) = 1/8.$

    and $f_{X_1}(x) = 0, \quad \text{for } x \notin \mathcal{X}.$
  - Graphical display



- Example (Committees). A committee of size n=4 is selected from 5 men and 5 women. Then,
  - Ω={combination of 4}, $\#\Omega = \binom{10}{4} = 210$, $P(A)$=$\#A/\#\Omega$

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

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

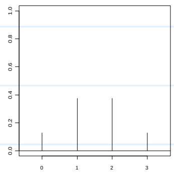
Theorem. If f_X is the pmf of r.v. X with range $\mathcal{X}$, then

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

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

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

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



- 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.

- 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).
- **Q**: Suppose that X and Y are two r.v.'s with same pmf. Is it always true that $X(\omega) = Y(\omega)$ for $\omega\in\Omega$?

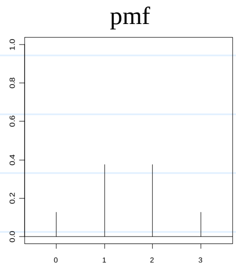
➢ Definition: A function F_X is called the *cumulative distribution function* of a random variable X if

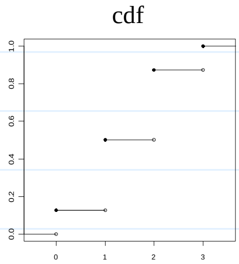
$$F_X(x) = P_X(X \leq x),\ x \in \mathbb{R}.$$

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

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

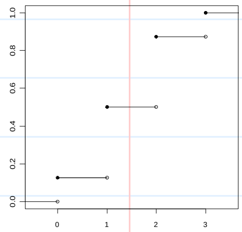
$$F_{X_1}(x) = \begin{cases} 0, & x < 0, \\ 1/8, & 0 \leq x < 1, \\ 4/8, & 1 \leq x < 2, \\ 7/8, & 2 \leq x < 3, \\ 1, & 3 \leq x. \end{cases}$$

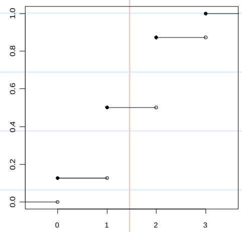


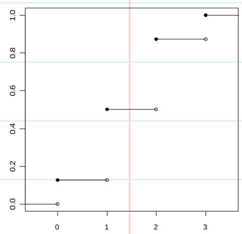


- **Q**: What should a cdf look like?

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







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

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

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

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

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

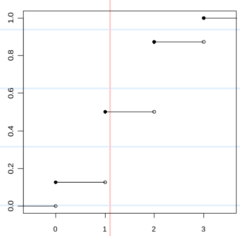
(3) For any $x \in \mathbb{R}$, $F_X(x)$ is continuous from the right, i.e.,

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

proof. Let x_n be a sequence s.t. $x_n \downarrow x$.

Let $E_n = (-\infty, x_n]$, then $E_n \downarrow (-\infty, x]$.

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



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

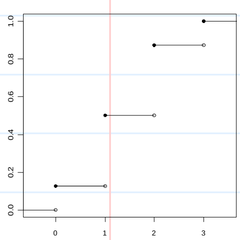
proof. Let $x_n \downarrow -\infty$, then $E_n \equiv (-\infty, x_n] \downarrow \emptyset$.

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

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

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

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



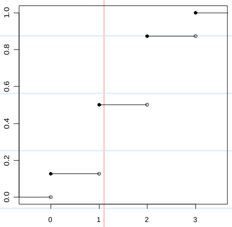
proof.
$$\begin{aligned} P_X(X > x) &= 1 - P_X\left(\{X > x\}^c\right) \\ &= 1 - P_X(X \leq x) = 1 - F_X(x). \end{aligned}$$

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

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

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

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



proof. $F_X(x) = P_X(X \in (-\infty, x]) = \displaystyle\sum_{x_i \in (-\infty, x] \cap \mathcal{X}} f_X(x_i).$

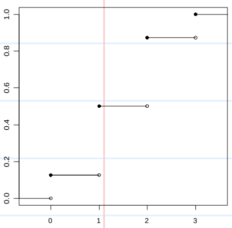
For $x_n \uparrow x$, $(-\infty, x_n] \uparrow (-\infty, x)$, and

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

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

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

(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 $T_x = (F_X(x-), F_X(x))$.

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

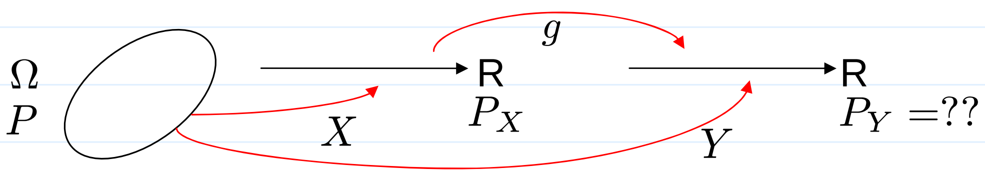
$\exists$ a rational number, denoted by r_x, in T_x.

Because the set of rational numbers is a countable set,

$\mathbb{D}$ is either finite or countably infinite.

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

- Transformation



➢ 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

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

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

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

proof. Since $\{\omega \in \Omega : Y(\omega) = y\} = \displaystyle\bigcup_{\substack{x \in \mathcal{X} \\ g(x) = y}} \{\omega \in \Omega : X(\omega) = x\},$

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

- 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

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

provided that the sum converges absolutely.

  - Example. If all value in $\mathcal{X}$ are equally likely, then $E(X)$ is simply the average of the possible values of X.
  - Example (Committees). In the committees example,

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

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

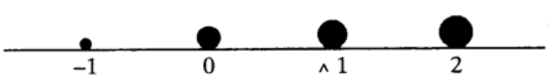
$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

    - Its range is {0, 1} and its pmf is

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

for a p.m. P defined on Ω.

    - So, $E(\mathbf{1}_A) = 0 \cdot [1 - P(A)] + 1 \cdot P(A) = P(A)$.
  - 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.
    - It is helpful to think of the expectation as the "center" of mass of the pmf
      - 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.



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

$\wedge$ = center of gravity = .9

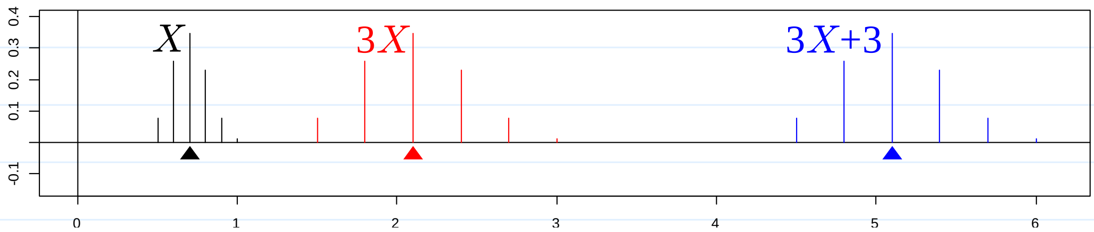
    - Expectation can be interpreted as a long-run average (Chapter 8)

- 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.

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

  - Example. $E(X^2) = \sum_{x\in\mathcal{X}} x^2 f_X(x)$.
  - ➢Theorem. For $a, b \in \mathrm{R}$, $E(aX+b) = a\cdot E(X)+b$.

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



- Mean and Variance.
  - ➢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
  $$Var(X) \equiv E[(X-\mu_X)^2] = \sum_{x\in\mathcal{X}} (x-\mu_X)^2 f_X(x).$$
  provided that the sum converges.
    - Example (Committees)

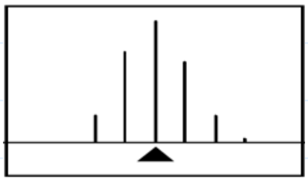
x	$f(x)$	$xf(x)$	$(x-\mu)^2 f(x)$	$x^2 f(x)$
0	5/210	0/210	20/210	0/210
1	50/210	50/210	50/210	50/210
2	100/210	200/210	0/210	400/210
3	50/210	150/210	50/210	450/210
4	5/210	20/210	20/210	80/210
Totals	1	2	2/3	14/3

So, $\mu = 2$ and $\sigma^2 = 2/3$

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

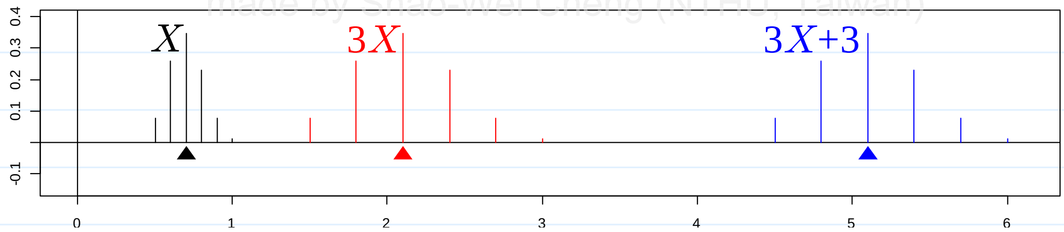
➢ Some properties of variance.

- 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.

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

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

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



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

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

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

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

- Corollary. $E[(X-c)^2]$ is minimized by letting $c=\mu_X$; and the minimum value is σ_X^2.
- Corollary. $\sigma_X^2 = E(X^2) - (E(X))^2$.
  (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

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

Some Common Discrete Distributions

- Bernoulli and Binomial Distributions

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

- Example. (a) Sampling with replacement (b) Coin Tossing (c) Roulette
- The sample space for the n trials is

$$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^n$$

- **Assume** that events depending on different trials are independent
- **Q**: Given an event $A_0 \subset \Omega_0$, what is the probability that A_0 occurs k times in the n trials?
- Problem Formulation: Let $A_i \subset \Omega$ be

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

$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n},$$

Q: What is $P(X=k)$?

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

- Example (Roulette, n=4, k=2, LNp.2-3).
  - Let $W_i = \{$Win on i^{th} Game$\}$

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

    Then, $P(W_i)=9/19 \equiv p$ and $P(L_i)=10/19=1-p \equiv q$
  - Let $X = \mathbf{1}_{W_1} + \mathbf{1}_{W_2} + \mathbf{1}_{W_3} + \mathbf{1}_{W_4}$, then

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

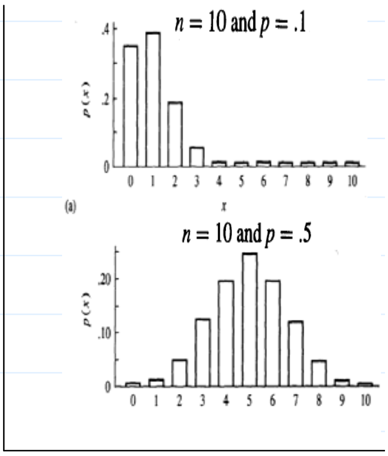
  - So,

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

➢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 $k=0, 1, \ldots, n$,

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



<u>proof</u>. **We may choose k trials in $\binom{n}{k}$ ways.**

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

□ □ ••• □

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

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

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

- 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.

- Notice that a binomial r.v. can be regarded as the sum of n independent Bernoulli r.v.'s.
- The binomial distribution is called after the Binomial Theorem: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

- Example (Bridge). **Q**: What is the probability that South gets no Aces on at least k=5 of n=9 hands?
  - Let A_i={no Aces on the i^{th} hand}, i=1, 2, …, 9, and

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

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

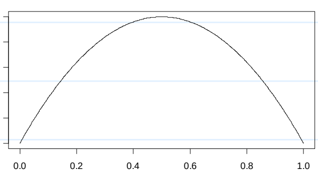
$$P(X=k) = \binom{9}{k} p^k (1-p)^{n-k}.$$

  - 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

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



proof.

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

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

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

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

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

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

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

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

$$= n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

➢ Summary for $X \sim$ 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: $Var(X)=np(1-p)$

• Geometric and Negative Binomial Distributions

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

- The sample space is
$$\Omega = \Omega_0 \times \Omega_0 \times \Omega_0 \times \cdots$$
- **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 $\binom{9}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^7 \times \left(\frac{1}{3}\right) = \binom{9}{2}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^7$.

- Problem Formulation:
  - Let $A_1, A_2, \ldots \subset \Omega$ be

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

$$X_n = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}, \text{ for } n = 1, 2, 3, \ldots.$$

  - Let Y_1 = smallest n with $X_n \geq 1$,
    Y_2 = smallest n with $X_n \geq 2$,
    $\ldots$,
    Y_r = smallest n with $X_n \geq r$,
  - **Q**: What is $P(Y_r=k)$?

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 $r=1$, $P(Y_1 = k) = P(\{X_{k-1} = 0\} \cap A_k)$
$$= P(\{X_{k-1}=0\}) \cdot P(A_k) = (1-p)^{k-1}p$$

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

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

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

- 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.

- A negative binomial r.v. can be regarded as the sum of r independent geometric r.v.'s.

□ □ ••• □ □ | □ □ ••• □ □ | ••• | □ □ ••• □ □ |

- The negative binomial distribution is called after the Negative Binomial Theorem:

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

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

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

proof.

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

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

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

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

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

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

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

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

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

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

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

➢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

➢ 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 $0 < \lambda < \infty$
- Then, $p_n \approx \lambda/n$ when n is large enough.
- And, $\binom{n}{k} p_n^k (1-p_n)^{n-k}$

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

- Here, for each fixed k,

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

- 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 n large, $n \gg k$, and $p_n \approx 0$,

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \approx \frac{1}{k!}\lambda^k e^{-\lambda}.$$

□□□□□□□□□□□□ • • • □□□

➢ 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.
- **Q**: P(5 or more errors in n=2500 letters)=??
- Ans.
  - Let X be the number of errors, then X~binomial(2500, 0.001) and

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

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

- 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?)

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

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

➢Probability Mass Function

- Theorem. Let

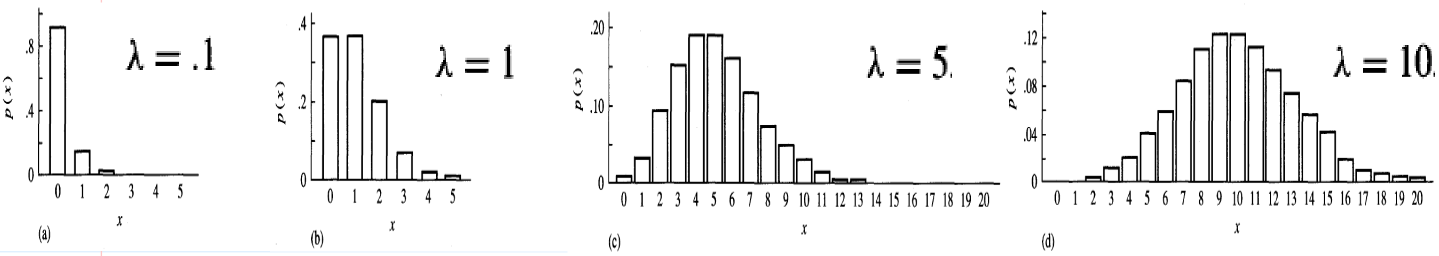
$$f(k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!}, & k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

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*.



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

$$\mu = \lambda \quad \text{and} \quad \sigma^2 = \lambda.$$

proof.

$$E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \cdot \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \cdot \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda$$

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

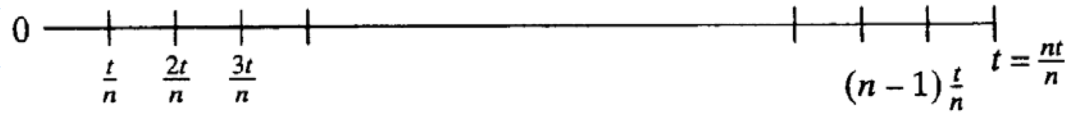
$$= \sum_{x=0}^{\infty} x(x-1) \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2 \cdot \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} = \lambda^2 \cdot \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda^2$$

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

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

➢Poisson Process

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



To model them, we can

  - Divide the time period, say [0, t], into n small intervals
  - 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

  - 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

  - The distribution for the number of events occurring in [0, t] can be approximated by Poisson($n \cdot p_n$)

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

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

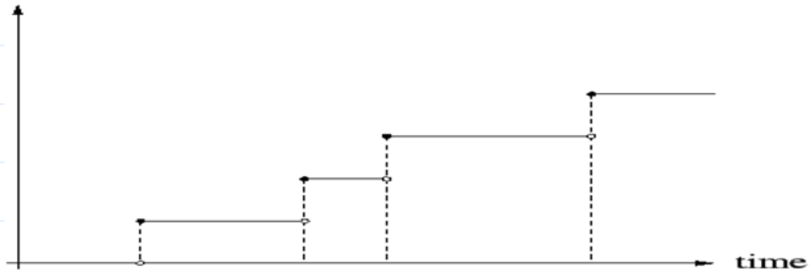
for $0 \le s < t < \infty$, and

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

are independent whenever

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

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



- 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
  - **Q**: What is the probability of 3 or more accidents occur in a 2 month periods?
  - Here, λt = 5.5×2 = 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)
  - Range: $\mathcal{X} = \{0, 1, 2, \ldots\}$
  - Pmf: $f_X(x) = \lambda^x e^{-\lambda}/x!$, for $x \in \mathcal{X}$
  - Parameter: $0<\lambda<\infty$
  - Mean: $E(X)=\lambda$
  - Variance: $Var(X)=\lambda$

- Hypergeometric Distribution
  - Experiment: Draw a sample of n ($\leq N$) balls *without replacement* from a box containing R red balls and $N-R$ white balls
    - Let X be the number of red balls in the sample
    - **Q**: What is $P(X=k)$?
    - Example. The Committee Example.
    - (c.f.) If drawn *with replacement*, what is the distribution of X?
  - Probability Mass Function
    - Theorem. For k = 0, 1, 2, …, 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 $0 \leq k \leq R$ and $0 \leq n-k \leq N-R$,

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.

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

- 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: $\binom{a+b}{r} = \sum_{k=0}^{r} \binom{a}{k}\binom{b}{r-k}$.

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

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

proof.

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

$$= \frac{nR}{N} \sum_{x=1}^{n} \frac{\binom{R-1}{x-1}\binom{(N-1)-(R-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}} = \frac{nR}{N} \sum_{y=0}^{n-1} \frac{\binom{R-1}{y}\binom{(N-1)-(R-1)}{(n-1)-y}}{\binom{N-1}{n-1}} = \frac{nR}{N}$$

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

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

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

$$= \frac{n(n-1)R(R-1)}{N(N-1)} \sum_{y=0}^{n-2} \frac{\binom{R-2}{y}\binom{(N-2)-(R-2)}{(n-2)-y}}{\binom{N-2}{n-2}} = \frac{n(n-1)R(R-1)}{N(N-1)}$$

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

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

➢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

$$\frac{\binom{R_i}{k}\binom{N_i-R_i}{n-k}}{\binom{N_i}{n}} \to \binom{n}{k} p^k (1-p)^{n-k}.$$

proof.

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

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

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

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

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

➢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: $Var(X) = nR(N-R)(N-n)/(N^2(N-1))$

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