Ch 5. Special Discrete Distributions

便箋標題 2009/10/16

5.1 Bernoulli and binomial random variables

- The sample space of a Bernoulli trial contains two points, s and f

The r.v. defined by $X(s) = 1$ and $X(f) = 0$

$\Downarrow$

Bernoulli r.v.

p : the prob. of a success

$1-p$ (q) : the prob. of a failure

The prob. mass function of X is

$$p(x) = \begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = 0 \cdot p(X=0) + 1 \cdot p(X=1) = p$$

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

$$E[X^2] = 0^2 \cdot p(X=0) + 1^2 \cdot p(X=1) = p$$

$$\therefore Var[X] = p - p^2 = p(1-p) \qquad \sigma_X = \sqrt{Var[X]} = \sqrt{p(1-p)}$$

Ex 5-1

a throw of a fair dice

the event of obtaining 4 or 6 $\Rightarrow$ success

event of 1, 2, 3, 5 $\Rightarrow$ failure

$\Rightarrow$ Bernoulli trial with $p = \frac{2}{6} = \frac{1}{3}$

$$p(x) = \begin{cases} \frac{2}{3} & x = 0 \\ \frac{1}{3} & x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

$E[x] = p = \frac{1}{3}$ $\quad Var[x] = p(1-p) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$

$X_1, X_2, \ldots$ a sequence of Bernoulli r.v.

If $\{X_1 = j_1\}$ $\{X_2 = j_2\}$, $\{X_3 = j_3\}$ ---

events are independent

$\Rightarrow$ we say that $\{X_1, X_2, \ldots\}$ and the corresponding

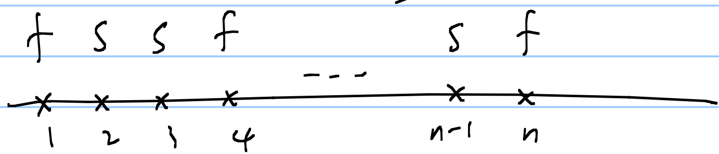
Bernoulli trials are independent

- Binomial r.v.

If n Bernoulli trials all with prob. of success p

are performed independently,

then X (the number of successes) is called "a binomial" with parameters n and p



The set of possible values of X is $\{0, 1, 2, \dots, n\}$

$X \sim B(n, p)$

n: # of Bernoulli trials

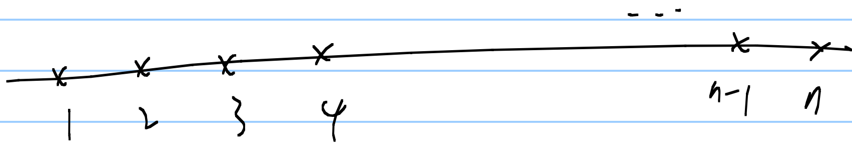
Bernoulli , Binomial

Thm 5.1

$$p(x) = P(X = x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$ ← binomial prob. mass function

prob. mass function

of a Binomial r.v. X with n and p.



Ex 5.3

hospital 10 babies, of whom six were boys,

prob. that the first six births were all boys ?

A: the event that the first six births were all boys, and the last four all girls.

X: the number of boys within the 10 babies

$$P(A \mid X=6) = \frac{P(A, X=6)}{P(X=6)} = \frac{P(A)}{P(X=6)} = \frac{\left(\frac{1}{2}\right)^{10}}{\binom{10}{6}\left(\frac{1}{2}\right)^{10}} = \frac{1}{\binom{10}{6}}$$

$$\simeq 0.0048$$

$p(X=6) = ? \quad \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = \binom{10}{6}\cdot\left(\frac{1}{2}\right)^{10}$

$p(A) = \left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4$

$= \left(\frac{1}{2}\right)^{10}$

Ex 5.4

In a small town, out of 12 accidents that occurred in June 1986, four happened on Friday the 13th.

30天 → each accident occurs on Friday the 13th $\frac{1}{30}$

the prob. of at least four accidents on Friday the 13th is

$$1 - \sum_{i=0}^{3} \binom{12}{i} \left(\frac{1}{30}\right)^{i} \left(\frac{29}{30}\right)^{12-i} \cong 0.000493$$

Gambler's Ruin

$$\lfloor t \rfloor$$

$$\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \cdots \left(\frac{1}{2}\right)^{\lfloor t \rfloor}$$

- Find the value of x at which $p(x)$ is maximum

$\lfloor \; \rfloor$

$x = \lfloor (n+1)p \rfloor$ (floor function)

$$\frac{p(x)}{p(x-1)} > 1 \ ?$$

$$= \frac{\frac{n!}{(n-x)!\,x!} p^{x} (1-p)^{n-x}}{\frac{n!}{(n-x+1)!\,(x-1)!} p^{x-1} (1-p)^{n-x+1}} = \frac{(n-x+1)p}{x(1-p)}$$

$$= \frac{(n+1)p - xp + x - x}{x(1-p)}$$

$$= \frac{[(n+1)p - x] + x(1-p)}{x(1-p)}$$

$$\frac{p(x)}{p(x-1)} = \frac{(n+1)p - x}{x(1-p)} + 1 > 1 \Rightarrow (n+1)p - x > 0$$

$\therefore x < (n+1)p$ 保证 $p(x) > p(x-1)$

$p(x)$ increases : x changes from 0 to $\lfloor (n+1)p \rfloor$

$p(x)$ decreases : x changes from $\lfloor (n+1)p \rfloor$ to n

- Expectation and Variance of a Binomial r.v.

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

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

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

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

$$= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^{i} (1-p)^{n-1-i}$$

$$= np \left[p + (1-p) \right]^{n-1} = np$$

$E[X] = np$

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

$\sigma_X = \sqrt{np(1-p)}$

5.2 Poisson r.v.

$n \to \infty$ (Bernoulli trials)

$p \to 0$

$\underline{np = \lambda}$ of moderate value

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

binomial prob. mass function

$$\Downarrow$$

$$p(X=i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}$$

$n \to \infty$

$$= \left[\frac{n \cdot (n-1) \cdots (n-i+1)}{n^i} \right] \frac{\lambda^i}{i!} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

$n(n-1)\cdots(n-i+1) \to n^i$

$$\left[\frac{n \cdot (n-1) \cdots (n-i+1)}{n^i} \right] \to 1$$

$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$

$\left(1-\frac{\lambda}{n}\right)^i \to 1$

$$= e^{-\lambda}\frac{\lambda^i}{i!}$$ Poisson approximation to Binomial

L.V. Bortkiewicz

$$\sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}\cdot e^{\lambda} = 1$$

$\therefore$ $e^{-\lambda}\frac{\lambda^i}{i!}$ is a prob. mass function

$$P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!}$$ $X \sim$ Poisson r.v.

$E[X] = \lambda \qquad Var[X]$

$$E[X] = \sum_{i=0}^{\infty} i\,p(X=i) = \sum_{i=1}^{\infty} i \frac{e^{-\lambda}\lambda^{i}}{i!} = \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!}}_{=e^{\lambda}}$$

$\lambda e^{-\lambda}$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$Var[X] = \underline{E[X^2]} - (E[X])^2$$

$$E[X^2] = \underbrace{E[X(X-1)]} + E[X]$$

$$\sum_{i=2}^{\infty} i(i-1)\underline{p(X=i)} = \lambda^2 e^{-\lambda} \frac{\lambda^{i-2}}{(i-2)!} = \lambda^2$$

$$= \frac{e^{-\lambda}\lambda^{i}}{i!}$$

$$= \lambda^2 + \lambda$$

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

For a Poisson r.v. X, $E[X] = Var[X] = \underline{\underline{\lambda}}$

$$\sigma_X = \sqrt{\lambda}$$

- Poisson approximation to Binomial

$n \to \infty$ $p \to 0$ $np = \lambda$ moderate

$p < 0.1$ $\underset{\text{mean}}{np} \le 10$

$np > 10 \Rightarrow$ normal approximation

Poisson Example

- Let X be the number of misprints on a document page typed by a secretary.

$\Rightarrow$ X is binomial r.v. if a word is called a success, provided that this is misprinted.

n : the number of words on the document page $\rightarrow$ large

misprints are rare events

np : the average number of misprints

$\Rightarrow$ X is approximately a Poisson r.v.

Ex 5.11

on average, every three pages of a book there is one typographical error.

If the number of typo- errors on a single pages is a Poisson r.v., the prob. of at least one error on a specific page ?

$$p(X=n) = \frac{e^{-\lambda}(\lambda)^n}{n!} \quad \text{where } \lambda = ? \ \tfrac{1}{3}$$

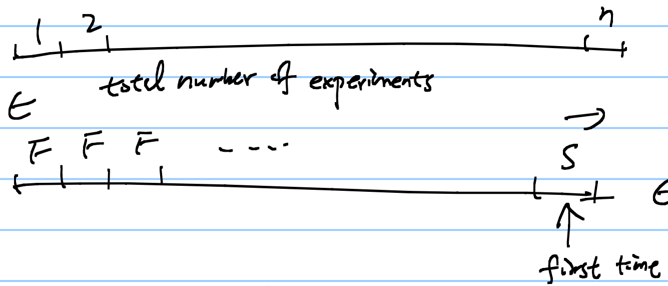
$$p(X=n) = \frac{e^{-(\frac{1}{3})}(\frac{1}{3})^n}{n!}$$

$$p(X \geq 1) = 1 - p(X=0) = 1 - e^{-\frac{1}{3}} \approx 0.28$$

5.3 Geometric R.V.s

• a sequence of independent Bernoulli trials, each with prob of success p, are performed

X: the number of experiments (Bernoulli trials) until the first success occurs. ⇒ Geometric



Binomial: number of successes among the n experiments

Geometric

$$p(X=n) = (1-p)^{n-1} \cdot p \qquad n = 1, 2, 3, \cdots$$

mass function

$$p(x) = \begin{cases} (1-p)^{x-1} p & 0 < p < 1 \quad x = 1, 2, 3 \ldots \\ 0 & \text{elsewhere} \end{cases}$$

geometric mass function

$E[x] = ?$ $\quad Var[x] = ?$ $\quad \sigma_x = ?$

$$E[x] = \sum_{n \geq 1}^{\infty} n q^{n-1} p = p \underbrace{\sum_{n \geq 1}^{\infty} n q^{n-1}}$$

$(q = 1-p)$

$$= p \frac{d}{dq} \overbrace{\sum_{n \geq 0}^{\infty}} q^n = p \frac{d}{dq} \underbrace{\sum_{n \geq 0}^{\infty} q^n}$$

$$= p \frac{d}{dq} \left(\frac{1}{1-q}\right) = p \frac{1}{(1-q)^2} = p \frac{1}{p^2}$$

$$= \frac{1}{p}$$

$$E[x^2] = \sum_{n \geq 1}^{\infty} n^2 q^{n-1} p = p \sum_{n \geq 1}^{\infty} \frac{d}{dq}(n q^n) = p \frac{d}{dq} \sum_{n = 1}^{\infty} n q^n = p \frac{d}{dq} \left[E[x] \frac{q}{1-q} \right]$$

$$= p\frac{d}{dq}\left[\frac{1}{1-q}\cdot\frac{q}{1-q}\right] = p\frac{d}{dq}\left[\frac{q}{(1-q)^2}\right]$$

$$= p\left[\frac{1}{p^2}+\frac{2(1-p)}{p^3}\right] = \frac{2}{p^2}-\frac{1}{p}$$

$$Var[X] = E[X^2]-\left(E[X]\right)^2$$

$$= \frac{1}{p^2}-\frac{1}{p} = \frac{1-p}{p^2}$$

Ex 5.18

draw cards at random, with replacement, until an ace is drawn

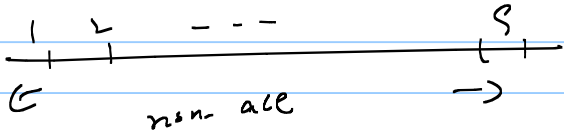
the prob. that at least 10 draws are needed ?

Ans : X is geometric with $p = \frac{1}{13}$

$$P(X \geq 10) = \sum_{n=10}^{\infty}\left(\frac{12}{13}\right)^{n-1}\left(\frac{1}{13}\right) = \frac{1}{13}\sum_{n=10}^{\infty}\left(\frac{12}{13}\right)^{n-1}$$

$$= \frac{1}{13} \frac{\left(\frac{12}{13}\right)^9}{1-\frac{12}{13}} = \left(\frac{12}{13}\right)^9 \approx 0.49$$

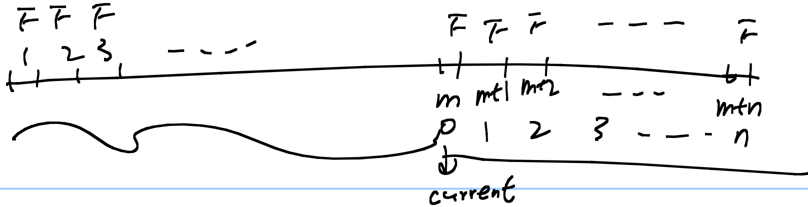
shortcut



$$\Rightarrow \left(\frac{12}{13}\right)^9 \approx 0.49$$

Memoryless property

$$P(X > n+m \mid X > m) = P(X > n)$$

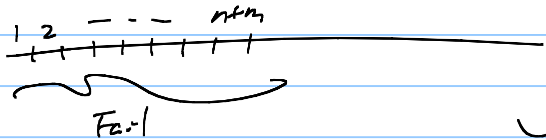


Geometric mass function 성질

$$P(X>n+m \mid X>m) = \frac{P(X>n+m,\ X>m)}{P(X>m)} = \frac{P(X>n+m)}{P(X>m)}$$

For a geometric X

$$P(X>n+m) = (1-p)^{n+m}$$



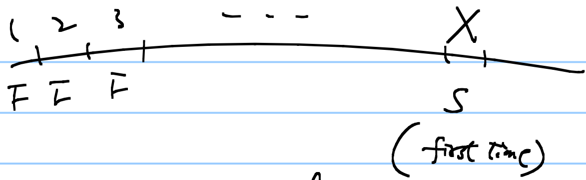
$$P(X>m) = (1-p)^{m}$$

→ 성질 ②

$$P(X>n+m \mid X>m) = \frac{(1-p)^{n+m}}{(1-p)^{m}}$$

$$= (1-p)^{n} = P(X>n)$$

Why Geometric is memoryless ?



independent Bernoulli trials

- Geometric is the only memoryless discrete r.v.

$$X : \quad P(X>n+m \mid X>m) = P(X>n)$$

$$\Rightarrow \quad P(X=n) = p(1-p)^{n-1}$$

Ex 5.19

A father asks his 3 sons to cut their backyard lawn.

each boy tosses a coin to determine the odd person

In the case that all three get heads or tails $\Rightarrow$ continue tossing

p : heads

$q = 1-p$ tails

(a) prob. that they reach a decision in less than n tosses ?

2 heads 1 tail

1 head 2 tails

$$\binom{3}{2} p^2 q + \binom{3}{2} p q^2 = 3pq\underbrace{(p+q)}_{=1} = 3pq$$

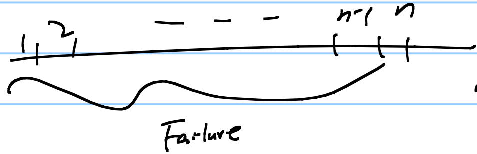
$$p(X<n) = 1 - p(X \geq n)$$
$$= 1 - [1-3pq]^{n-1}$$

[$p(X \geq n)$ ⇒ n-1 rounds ⇒ failures]

(b) If $p = 1/2$, what is the minimum number of tosses required to reach a decision with prob. 0.95?

$$p(X \leq n) \geq 0.95 \Rightarrow p(X > n) \leq 0.05$$

$$(1-3pq)^n \leq 0.05$$



$(1-3pq)$

$$\because 1-3pq = 1 - 3 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\therefore \left(\frac{1}{4}\right)^n \leq 0.05$$

$$n \geq 2.16$$

$\therefore$ n is 3 ※

- Negative Binomial r.v.s

$\Updownarrow$

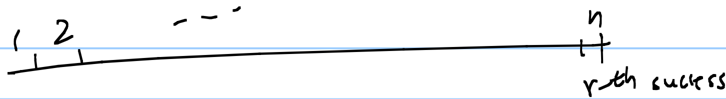
Geometric

a sequence of independent Bernoulli trials is performed.

X: the number of Bernoulli trials until the r-th success occurs.

$$P(X=n) = \binom{n-1}{r-1} p^{r} (1-p)^{n-r} \quad \text{where } n = r, r+1, r+2, \dots$$

negative binomial

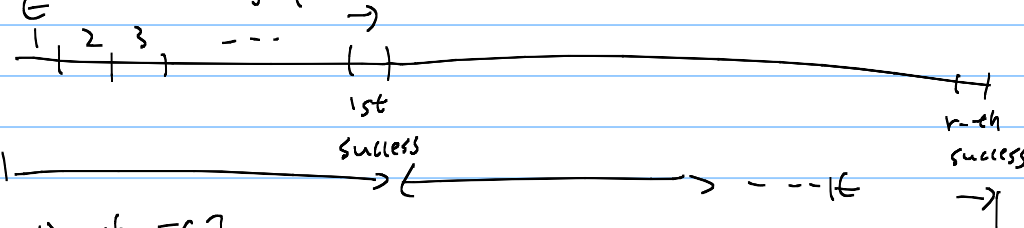


$E[X] = r\frac{1}{p}$ Var[X]

Geometry

$E[X] = \frac{1}{p}$

$E[X] = \frac{1}{p}$



Derivation of $E[X]$

$$E[X^k] = \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

$$\because\ n\binom{n-1}{r-1} = r\binom{n}{r} \Rightarrow \ = \frac{r}{p}\sum_{n=r}^{\infty} n^{k-1}\binom{n}{r} p^{r+1} (1-p)^{n-r}$$

Setting $m = n+1$

$$\Rightarrow E[X^k] = \frac{r}{p}\sum_{m=r+1}^{\infty} (m-1)^{k-1} \underbrace{\binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)}}_{\text{negative binomial mass function} \Rightarrow \text{for } Y \text{ parameters } (r+1,\, p)}$$

negative binomial (r, p)

$$= \frac{r}{p} E\left[(Y-1)^{k-1}\right]$$

negative binomial $(r+1, p)$

$$E[X] = \frac{r}{p}$$

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

$$E[X^2] = \frac{r}{p} E\left[(Y-1)^1\right]$$

$$= \frac{r}{p} E[Y-1]$$

$$= \frac{r}{p}\left\{\underbrace{E[Y]}_{\frac{r+1}{p}} - 1\right\}$$

$$\therefore E[X^2] = \frac{r}{p}\left[\frac{r+1}{p} - 1\right]$$

$$Var[X] = \frac{r}{p}\left[\frac{r+1}{p} - 1\right] - \left(\frac{r}{p}\right)^2$$

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

Ex 5.20

Sharon and Ann play a series of backgammon games until one of them wins five games

the prob. Sharon wins a game is $\underline{\underline{0.58}}$

(c) the prob. that the series ends in seven games

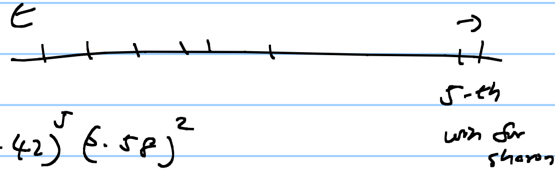
Ans: Let X be the number of games until Sharon wins five games

Y . . . . Ann wins five games

$X \sim$ Negative binomial $(5, 0.58)$ for Sharon

$Y \sim$ Negative binomial $(5, 0.42)$ for Ann

$$p(X=7) + p(Y=7) \cong 0.24$$



mass function

$$\binom{7-1}{5-1}(0.58)^5(0.42)^2 \qquad \binom{7-1}{5-1}(0.42)^5(0.58)^2$$

(b) If the series ends in seven games,

the prob. that Sharon wins ?

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(X=7)}{P(X=7)+P(Y=7)} \approx 0.71$$

B: event that the series ends in seven games

A: the event that Sharon wins

$P(AB) = P(X=7)$

$P(B) = P(X=7) + P(Y=7)$

Negative binomial

Ex 7.21 (Attrition Ruin Problem)

Two gamblers

A beats B with prob. p

B beats A $1-p$

each play results in a forfeiture of \$1 for the loser
in no change for the winner

A initially has a dollars

B b dollars

$\Rightarrow$ B ruined ?

prob.

Ans: when B is ruined ?

b plays $\rightarrow$ B 全輸

$b+1$ plays $\rightarrow$ B 輸了 b 次 and b-th loss on $(b+1)$-st play

⋮

$b+a-1$ plays $\rightarrow$ ... b 次 b-th loss on $(b+a-1)$-st play

E_i : the event that, in the first $b+i$ plays,

B loses b times and the b-th loss occurs

in the $(b+i)$-th play

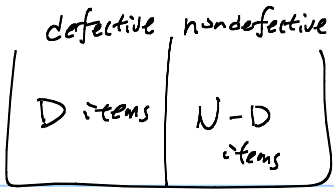
A^* : the event that A wins

$$p(A^*) = \sum_{i=0}^{a-1} p(E_i)$$

$$p(E_i) = \binom{i+b-1}{b-1} p^b q^i$$

$$p(A^*) = \sum_{i=0}^{a-1} \binom{i+b-1}{b-1} p^b q^i$$ ← negative binomial

- Hypergeometric Random Variables



⇒ n items are drawn at random and without replacement

$n \le \min(D, N-D)$

X: the # of defective items drawn

⇒ hypergeometric r.v.s

prob. mass function of X $\quad p(x) = p(X=x) = \dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$

⇑ hypergeometric prob. mass function

$E[X] = ?$ $\qquad Var[X] = ?$

$$E[X^k] = \sum_{i=0}^{n} i^k p(X=i) = \sum_{i=1}^{n} i^k \binom{D}{i}\binom{N-D}{n-i} \Big/ \binom{N}{n}$$

$$i\binom{D}{i} = D\binom{D-1}{i-1}$$

$$n\binom{N}{n} = N\binom{N-1}{n-1}$$

$$E[X^k] = \frac{nD}{N}\sum_{i=1}^{n} i^{k-1}\binom{D-1}{i-1}\binom{N-D}{n-i} \Big/ \binom{N-1}{n-1}$$

Let $j = i-1$

$$E[X^k] = \frac{nD}{N}\sum_{j=0}^{n-1} \left[(j+1)^{k-1}\binom{D-1}{j}\binom{N-D}{n-1-j} \Big/ \binom{N-1}{n-1}\right]$$

$p(Y=j)$

Y: hypergeometric r.v. with parameters $n-1,\ N-1,\ D-1$

$$E[X^k] = \frac{nD}{N} E\left[(Y+1)^{k-1}\right]$$

$$E[X] = \frac{nD}{N}$$ #

$E[X^2]$ (for $Var(X)$)

$$E[X^2] = \frac{nD}{N} \underline{E[Y+1]}$$

$$\left[(n-1)(D-1)/(N-1)\right]$$

$$E[X^2] = \frac{nD}{N}\left[\frac{(n-1)(D-1)}{N-1} + 1\right]$$

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

$$= \frac{nD}{N}\left[\frac{(n-1)(D-1)}{N-1} + 1 - \frac{nD}{N}\right]$$

$$p = \frac{D}{N} \Rightarrow Var[X] = \frac{N-n}{N-1} np(1-p)$$

Ex 5.23

500 independent calculations, a scientist has made 25 errors

a second scientist checks 7 of these calculations randomly.

$\Rightarrow$ prob. that he detects 2 errors ?

25	475
errors	no errors

$\rightarrow$ 7 are selected

X : # of errors found by the second scientist

$\Rightarrow$ X hypergeometric $N = 500$, $D = 25$

$n = 7$

$$p(X=2) = \frac{\binom{25}{2}\binom{475}{5}}{\binom{500}{7}} \doteq 0.04$$

Ex 5.24

abortion

community of a+b potential voters

① a for abortion, b are against it (b<a)

② a vote : determine the will of the majority (legalizing abortion

③ n random persons of these a+b voters do not vote

⇒ what is the prob. that those potential against abortion will win ?

Ans : $a-X < b-(n-X)$

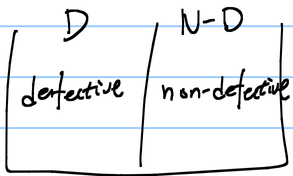
X : the number of those who do not vote and for abortion

$$a-X < b-n+X \Rightarrow 2X > a-b+n \Rightarrow X > \frac{a-b+n}{2}$$

X: hypergeometric

$$P\left(X > \frac{a-b+n}{2}\right) = \sum_{i=\lfloor \frac{a-b+n}{2}\rfloor+1}^{n} P(X=i) = \sum_{i=\lfloor \frac{a-b+n}{2}\rfloor+1}^{n} \frac{\binom{a}{i}\binom{b}{n-i}}{\binom{a+b}{n}}$$

Remark 5-2



n items are chosen without replacement

$\Rightarrow$ hypergeometric

# of defective items

$P(X=x)$

with replacement

$\Rightarrow$? binomial $\sim (n, \frac{D}{N})$

$$p(X=x) = \binom{n}{x}\left(\frac{D}{N}\right)^{x}\left(1-\frac{D}{N}\right)^{n-x} \qquad x=0,\cdots n$$

N is very large , D is very large

hypergeometric ~ binomial

$$\lim_{\substack{N\to\infty \\ D\to\infty \\ \frac{D}{N}\to p}} \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^{x} (1-p)^{n-x}$$

hypergeometric ,

binomial ,

poisson