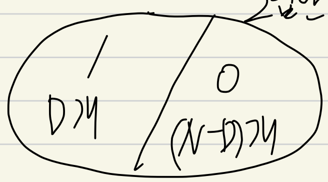
Chapter 3

① 초기하분포 (hypergeometric distribution)

1) bernoulli trial .. 순차적으로 n번 시행

2) without replacement

모집단 (population) N개



$$p := \frac{D}{N}$$

$$f(x) = P(X=x) = \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}, \quad \begin{pmatrix} 0 \le x \le D \\ 0 \le n-x \le N-D \end{pmatrix}$$

the number of 1 ⇒ discrete!

→ $X \sim H(n; N, D)$: hypergeometric distribution.

$$E[X] = \sum_{x=0}^{D} x \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}$$

$$= \sum_{x=1}^{D} D \binom{D-1}{x-1}\binom{N-1-(D-1)}{n-1-(x-1)} \Big/ \binom{N}{n} \quad (*)$$

$$= D\binom{N-1}{n-1} \Big/ \binom{N}{n} = D \frac{(N-1)!}{(n-1)!(N-n)!} \times \frac{n!(N-n)!}{N!} = n\frac{D}{N}$$

$$x\binom{D}{x} = x\frac{D!}{x!(D-x)!} = \frac{D(D-1)!}{(x-1)!(D-1-(x-1))!} = D\binom{D-1}{x-1}$$

$(*)$ $(1+t)^{N-1} = (1+t)^{D-1}(1+t)^{N-1-(D-1)}$ ← 첫번째 항에서 (t^{x-1}의 계수)과 둘째 항에서 ($t^{n-1-(x-1)}$)의 계수를 곱하고 모든 가능한 $x=1, \dots, D$에 대해 더한다.

(t^{n-1}의 계수): $\binom{N-1}{n-1} = \sum_{x=1}^{D}\binom{D-1}{x-1}\binom{N-1-(D-1)}{n-1-(x-1)}$

e.g. $(1+t)^3(1+t)^2$에서 t^3의 계수

$$\binom{3}{1}t\binom{2}{2}t^2$$

$$+\binom{3}{2}t^2\binom{2}{1}t$$

$$+\binom{3}{3}t^3\binom{2}{0}t^0$$

$$x(x-1)\binom{D}{x}$$

$$= x(x-1)\frac{D!}{x!\,(D-x)!}$$

$$= \frac{D(D-1)(D-2)!}{(x-2)!\,(D-2-(x-2))!}$$

$$= D(D-1)\binom{D-2}{x-2}$$

$$E[X(X-1)] = \sum_{x=2}^{D} x(x-1)\binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}$$

$$= \sum_{x=2}^{D} D(D-1)\binom{D-2}{x-2}\binom{N-2-(D-2)}{n-2-(x-2)} \Big/ \binom{N}{n}$$

$$= D(D-1)\binom{N-2}{n-2} \Big/ \binom{N}{n} = D(D-1)\frac{(N-2)!}{(N-n)!\,(n-2)!}\quad\frac{n!\,(N-n)!}{N!}$$

$$= D(D-1)\frac{n(n-1)}{N(N-1)}$$

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

$$= D(D-1)\frac{n(n-1)}{N(N-1)} + E[X](1-E[X])$$

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

$N >> n$

$$\binom{D}{x}\binom{N-D}{n-x}\Big/\binom{N}{n} = \frac{D!}{x!\,(D-x)!}\frac{(N-D)!}{(n-x)!\,(N-D-n+x)!} \times \frac{n!\,(N-n)!}{N!}$$

$$= \binom{n}{x}\frac{D(D-1)\cdots(D-x+1)}{N(N-1)\cdots(N-n+1)} \times (N-D)\cdots(N-D-n+x+1)$$

$$= \binom{n}{x}\frac{D(D-1)\cdots(D-x+1)}{N(N-1)\cdots(N-x+1)} \times \frac{(N-D)\cdots(N-D-n+x+1)}{(N-x)\cdots(N-n+1)}$$

$$\simeq \binom{n}{x}\left(\frac{D}{N}\right)^{x}\left(\frac{1-D}{N}\right)^{n-x}$$ … "binomial distribution"

$N>>1$ 이면 hypergeometric distribution ≃ binomial dist.

$$\frac{D(D-1)\cdots(D-x+1)}{N(N-1)\cdots(N-x+1)} = \frac{\frac{D}{N}\left(\frac{D}{N}-\frac{1}{N}\right)\cdots\left(\frac{D}{N}-\frac{x+1}{N}\right)}{1\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{x+1}{N}\right)}$$

$$\to \left(\frac{D}{N}\right)^{x} \text{ as } N>>1 \text{ and } D>>1.$$

with replacement.

② Binomial & Multinomial distribution.

$$f(x) = P(X=x) = \binom{n}{x} p^{x}(1-p)^{x} \quad (x=0,\dots,n)$$

$\rightarrow X \sim B(n, p)$

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

$*$ binomial theorem.

$$= np\,(p+1-p)^{n-1} = np$$

$$E[X(X-1)] = \sum_{x=2}^{n} x(x-1)\binom{n}{x} p^{x}(1-p)^{n-x}$$
$$= \sum_{x=2}^{n} n(n-1)\binom{n-2}{x-2} p^{x}(1-p)^{n-x}$$
$$= n(n-1)p^{2} \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2}(1-p)^{n-2-(x-2)}$$
$$= n(n-1)p^{2}\,(p+(1-p))^{n-2}$$
$$= n(n-1)p^{2}$$

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

Prop. $X \sim B(n,p)$

$\Rightarrow E[X] = np \quad Var(X) = np(1-p).$

representational definition.

$P(Z_i = 1) = p \quad P(Z_i = 0) = 1-p \quad Z_i \sim \text{Bernoulli}(p)$

$\therefore X \sim B(n,p) \Longleftrightarrow X \overset{d}{=} Z_1 + \cdots + Z_n \quad Z_i \overset{iid}{\sim} \text{Bernoulli}(p).$

distribution 으로서 같다.

independent and identically distributed.

Thm. 2.3.1

(a) $X \sim B(n,p)$

$\Rightarrow mgf_X(t) = (pe^t + q)^n, \quad t \in \mathbb{R}, \; q = 1-p.$

(b) $X_1 \sim B(n_1, p) \quad X_2 \sim B(n_2, p)$ independent.

$\Rightarrow X_1 + X_2 \sim B(n_1+n_2, p)$

$Z_i \overset{iid}{\sim} Bernoulli(p)$

proof)

$\because$ independent

(a) $mgf_X(t) = mgf_{Z_1}(t) \times \cdots \times mgf_{Z_n}(t) = (mgf_{Z_1}(t))^n$

$\because$ identically distributed.

$$mgf_{Z_1}(t) = E[e^{tZ_1}]$$
$$= e^{t\cdot 0} q + e^{t\cdot 1} p$$
$$= pe^t + q.$$

$$\therefore mgf_X(t) = (pe^t + q)^n$$

(b) $mgf_{X_i}(t) = (pe^t + q)^{n_i} \quad (i=1,2)$

$$mgf_{X_1+X_2}(t) = mgf_{X_1}(t)\, mgf_{X_2}(t)$$
$$= (pe^t + q)^{n_1} (pe^t + q)^{n_2}$$

적률생성함수의 분포 결정성.

$$\therefore X_1 + X_2 \sim B(n_1+n_2, p)$$

$\square$

Note $E[e^{tX}] = E[e^{t(Z_1 + \cdots + Z_n)}]$

$= E[e^{tZ_1} \cdots e^{tZ_n}]$

$= E[e^{tZ_1}] \cdots E[e^{tZ_n}]$

$= mgf_{Z_1}(t) \cdots mgf_{Z_n}(t)$

X, Y: indep.

$\because E[XY] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_{1,2}(x,y)\,dy\,dx$

$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_1(x)\, y f_2(y)\,dy\,dx$

$= \int_{-\infty}^{+\infty} x f_1(x)\,dx \int_{-\infty}^{+\infty} y f_2(y)\,dy$

$= E[X]\,E[Y].$

Multinomial Distribution.

$n = 5$

$N = 300$

100	150	50
바나나	딸기	배
x_1	x_2	x_3

$p_1 = \frac{100}{300}$, $p_2 = \frac{150}{300}$, $p_3 = \frac{50}{300}$

$x_1 + x_2 + x_3 = 5$

$$\frac{5!}{x_1!\,x_2!\,x_3!} =: \binom{5}{x_1, x_2, x_3}$$

	1 바나나	2 딸기	3 배
Z_1	Z_{11} 1	Z_{12} 0	Z_{13} 0
Z_2	Z_{21} 0	Z_{22} 1	Z_{23} 0
Z_3	Z_{31} 0	Z_{32} 0	Z_{33} 1
Z_4	Z_{41} 1	Z_{42} 0	Z_{43} 0
Z_5	Z_{51} 1	Z_{52} 0	Z_{53} 0
X	x_1 3	x_2 1	x_3 1

(# of features) $= k$.

$X = (X_1, \dots, X_k)^t$

$$f(x_1, \dots, x_k) = \binom{n}{x_1, \dots, x_k} p_1^{x_1} \cdots p_k^{x_k} \qquad x_i = 0, \dots, n \ (i = 1, \dots, k), \quad x_1 + \cdots + x_k = n$$

$\rightarrow X \sim \text{Multi}(n, (p_1, \dots, p_k)^t)$

⟨Representational Definition⟩

$Z_i = (Z_{i1}, \dots, Z_{ik})^t \quad (i = 1, \dots, n)$

$P(Z_i = e_j) = p_j \quad (j = 1, \dots, k)$

$\Leftrightarrow P(Z_{i1} = z_1, \dots, Z_{ik} = z_k) = p_1^{z_1} \cdots p_k^{z_k}$

$z_i = 0$ or 1 $\quad z_1 + \cdots + z_k = 1$

$\therefore Z_i \sim \text{Multi}(1, (p_1, \dots, p_k)^t)$

$\therefore X \sim \text{Multi}(n, (p_1, \dots, p_k)^t)$

$\Leftrightarrow X \overset{d}{\equiv} Z_1 + \cdots + Z_n, \quad Z_i \overset{iid}{\sim} \text{Multi}(1, (p_1, \dots, p_k)^t)$

Prop.

$X = (X_1, \dots, X_k)^t \sim \mathrm{Multi}(n, (p_1, \dots, p_k)^t)$

(a) $E[X_\ell] = n p_\ell \quad (\ell = 1, \dots, k)$

$\mathrm{Var}(X_\ell) = n p_\ell (1 - p_\ell)$

$\mathrm{Cov}(X_\ell, X_m) = -n p_\ell p_m \quad (\ell \neq m)$

(b) $\mathrm{mgf}_X(t) = (p_1 e^{t_1} + \cdots + p_k e^{t_k})^n \quad t = (t_1, \dots, t_k)^t$

proof)

$X \overset{d}{=} Z_1 + \cdots + Z_n \quad Z_i \overset{iid}{\sim} \mathrm{Multi}(1, (p_1, \dots, p_k)^t)$

$(E[X_1], \dots, E[X_k]) = E[X] = E[Z_1 + \cdots + Z_n] \overset{\because \text{ iid}}{=} n E[Z_1]$

$\mathrm{Var}(X) = \mathrm{Var}(Z_1 + \cdots + Z_n) = n \mathrm{Var}(Z_1)$

$Z_\ell \perp Z_m \Rightarrow \mathrm{Cov}(Z_\ell, Z_m) = 0 \quad \ell \neq m$

$Z_1 = (Z_{11}, \dots, Z_{1k})^t \sim \mathrm{Multi}(1, (p_1, \dots, p_k)^t)$

$\Rightarrow P(Z_{1\ell} = 1) = p_\ell \quad P(Z_{1\ell} = 0) = 1 - p_\ell$

$Z_{1m} Z_{1\ell} = 0 \quad (\ell \neq m)$

$\therefore E[Z_{1\ell}] = 1 \cdot p_\ell + 0 \cdot (1 - p_\ell) = p_\ell.$

$\mathrm{Cov}(Z_{1\ell}, Z_{1m}) = E[Z_{1\ell} Z_{1m}] - E[Z_{1m}] E[Z_{1\ell}]$

$= -p_\ell \cdot p_m$

$\therefore E[X] = n E[Z_\ell] = n (p_1, \dots, p_k)^t$

$\Rightarrow E[X_\ell] = n p_\ell.$

$$\mathrm{Var}(X) = \mathrm{Var}(Z_1 + \cdots + Z_n) = n\,\mathrm{Var}(Z_1) \in M_{k\times k}(\mathbb{R})$$

$$\mathrm{Var}(Z_{1i}) = E[Z_{1i}^2] - E[Z_{1i}]^2$$

$$= 1^2 \cdot p_i + 0\cdot(1-p_i) - (p_i + 0\cdot(1-p_i))^2$$

$$= p_i(1-p_i)$$

$$\Rightarrow (\mathrm{Var}(Z_1))_{ij} = \begin{cases} p_i(1-p_i) & \text{if } i=j \\ -p_i p_j & \text{if } i\neq j \end{cases}$$

$$\mathrm{Var}(X) = n\,\mathrm{Var}(Z_1)$$

$$\mathrm{Var}(X_e) = n\,\mathrm{Var}(Z_{1e}) = n p_e(1-p_e)$$

<u>Note</u> $p = (p_1, \dots, p_k)^t$

$$E[X] = np \quad (\because E[X_e] = np_e)$$

$$\mathrm{Var}(X) = \mathrm{diag}(p_1, \dots, p_k) - pp^t \qquad \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} (p_1 \cdots p_k)$$

i) $i=j$

$$(\mathrm{Var}(X))_{ii} = p_i(1-p_i) = p_i - p_i^2$$

$$= [\mathrm{diag}(p_1, \dots, p_k)]_i - [pp^t]_{ii}$$

ii) $i \neq j$ $\qquad$ i th component.

$$(\mathrm{Var}(X))_{ij} = -p_i p_j = -[pp^t]_{ij}$$

$$\therefore \mathrm{Var}(X) = \mathrm{diag}(p_1, \dots, p_k) - pp^t$$

$X \overset{d}{=} Z_1 + \cdots + Z_n \quad Z_i \overset{iid}{\sim} \mathrm{Multi}(1, (p_1, \ldots, p_k)^t)$

$X = (X_1, \ldots, X_k)^t \quad X \sim \mathrm{Multi}(n, (p_1, \ldots, p_k)^t)$

(b) $mgf_X(t) = E[e^{t_1 X_1 + \cdots + t_k X_k}]$

$$= E[e^{t_1(Z_{11} + \cdots + Z_{n1})} \cdots e^{t_k(Z_{1k} + \cdots + Z_{nk})}]$$

$$= E[e^{t_1 Z_{11} + \cdots + t_k Z_{1k}} \cdots e^{t_1 Z_{n1} + \cdots + t_k Z_{nk}}]$$

$\because$ independent

$$= E[e^{t_1 Z_{11} + \cdots + t_k Z_{1k}}] \cdots E[e^{t_1 Z_{n1} + \cdots + t_k Z_{nk}}]$$

$$= mgf_{Z_1}(t) \cdots mgf_{Z_n}(t)$$

$\because$ identically distributed.

$$= (mgf_{Z_1}(t))^n$$

$mgf_{Z_1}(t) = E[e^{t_1 Z_{11} + \cdots + t_k Z_{1k}}]$

$$= e^{t_1} p_1 + \cdots + e^{t_k} p_k$$

$\therefore mgf_X(t) = (e^{t_1} p_1 + \cdots + e^{t_k} p_k)^n$

$\square$

4/4.

기하 분포 (geometric distribution)

서로 독립, bernoulli trial w/ prob. p.

$X_1, X_2, \ldots$ 성공할 때까지의 횟수를 확률변수.

첫번째 성공이 관찰될 때까지의 시행횟수 W_1

e.g - p.38

$$P(W_1 = x) = (1-p)^{x-1} p \quad (x=1, 2, \ldots)$$

Write $W_1 \sim Geo(p)$

Thm 3.3.1.

$W_1 \sim Geo(p)$

(a) $mgf_{W_1}(t) = (1-qe^t)^{-1} e^t p \quad t < -\log q$

(b) $E[W_1] = \frac{1}{p} \quad Var(W_1) = \frac{q}{p^2}$

proof)

(a) $mgf_{W_1}(t) = E[e^{tW_1}]$

$$= \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p \quad q := 1-p$$

$$= \sum_{x=1}^{\infty} (qe^t)^{x-1} e^t p$$

$$= e^t p (1-qe^t)^{-1} \quad -1 < qe^t < 1$$

$qe^t < 1$

$t < -\log q.$

(b) $cgf_{W_1}(t) = t + \log p - \log(1-qe^t)$

$$c_1(t) = 1 + \frac{+qe^t}{1-qe^t}$$

$$c_1(0) = 1 + \frac{+q}{1-q} = 1 + \frac{+q}{p} = \frac{p+q}{p} = \frac{1}{p}$$

$E[W_1] = \frac{1}{p}$

$$C_2(t) = \frac{d}{dt}\left(1 + \frac{qe^t}{1-qe^t}\right)$$

$$= \frac{qe^t(1-qe^t) - qe^t(-qe^t)}{(1-qe^t)^2} = \frac{qe^t}{(1-qe^t)^2}$$

$$C_2(0) = \frac{q}{(1-q)^2} = \frac{q}{p^2} \quad \therefore \ Var(W_1) = \frac{q}{p^2}$$

□

음이항분포 (Negative Binomial distribution)

서로 독립인 성공률이 p인 베르누이 시행

$X_1, X_2, \ldots$

r번째 성공이 관찰 될 때까지의 시행횟수 W_r

$(1+t)^n$ - 음수일수도 있다. (유리수도 가능)

$$P(W_r = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad (x = r, r+1, \ldots)$$

p.120.

Write $W_r \sim Negbin(r, p)$.

Representational definition.

$W_1, \ W_2 - W_1, \ W_3 - W_2, \ \ldots, \ W_r - W_{r-1}$

$P(W_1 = x_1, W_2 - W_1 = x_2, \ldots, W_r - W_{r-1} = x_r)$

$=$ P(연속된 $(x_i - 1)$회 실패 후 성공 $i = 1, \ldots, r$)

$= (1-p)^{x_1-1} p \, (1-p)^{x_2-1} p \cdots (1-p)^{x_r-1} p$

$\therefore$ $W_1, W_2 - W_1, \ldots, W_r - W_{r-1}$은 서로 독립이고

동일한 기하분포를 따름. $\therefore X \sim Negbin(r, p) \iff X \overset{d}{=} Z_1 + \cdots + Z_r$

$Z_i \overset{iid}{\sim} Geo(p)$

Thm 3.3.2

(a) $X \sim \text{Negbin}(r, p) \Rightarrow E[X] = \frac{r}{p}$ $\quad Var(X) = \frac{rq}{p^2}$

(b) $X \sim \text{Negbin}(r, p) \Rightarrow mgf_X(t) = \left(pe^t(1-qe^t)^{-1}\right)^r$

(c) $X_1 \sim \text{Negbin}(r_1, p)$, $X_2 \sim \text{Negbin}(r_2, p)$ indep.

$\Rightarrow X_1 + X_2 \sim \text{Negbin}(r_1 + r_2, p)$

proof) $X \overset{d}{=} Z_1 + \cdots + Z_r \quad Z_i \overset{iid}{\sim} Geo(p)$

(a) $E[X] = E[Z_1 + \cdots + Z_r] = r E[Z_1] = \frac{r}{p} \quad (\because E[Z_1] = \frac{1}{p})$

(b) $mgf_X(t) = E[e^{tX}] = E[e^{t(Z_1 + \cdots + Z_r)}]$

$= E[e^{tZ_1}] \cdots E[e^{tZ_r}]$

$= (mgf_{Z_1}(t))^r$

$= \left(pe^t(1-qe^t)^{-1}\right)^r$

(c) $mgf_{X_1+X_2}(t) = E[e^{tX_1 + tX_2}]$ $\quad \because$ independant

$= E[e^{tX_1}]\, E[e^{tX_2}]$

$= mgf_{X_1}(t) \times mgf_{X_2}(t)$

$= \left(pe^t(1-qe^t)^{-1}\right)^{r_1+r_2}$

적률생성함수의 분포결정성에 의해

$X_1 + X_2 \sim \text{Negbin}(r_1 + r_2, p)$. $\quad \square$

⟨이항분포의 포아송 근사⟩

Poisson Distribution.

$B(n,p)$ $\quad n>>1 \quad p<<1 \rightarrow np$가 어떠한 양수로 수렴함을 전제

$\binom{n}{x}(1-p)^{n-x}p^{x} = n(n-1)\cdots(n-x+1)(1-p)^{n-x}p^{x}/x!$

$= \underbrace{\frac{n}{n}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}_{\simeq 1\ (\because n>>1)}(np)^{x}\left(1-\frac{np}{n}\right)^{n-x}/x!$

$\simeq (np)^{x}e^{-np}/x!$

$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} = e$

$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^{n} = \lim_{n\to\infty}\left(1+\frac{a}{n}\right)^{\frac{n}{a}\cdot a}$

$= e^{a}$

$\left(1-\frac{np}{n}\right)^{n-x} \simeq \left(1-\frac{np}{n}\right)^{n}$

$\lim_{\substack{n\to\infty \\ p<<1}}\left(1-\frac{np}{n}\right)^{-\frac{n}{np}\cdot(-np)} = e^{-np}$

$f(x) := \frac{e^{-\lambda}\lambda^{x}}{x!} \quad (x=0,1,2,\ldots),\ \lambda>0$

$\boxed{\text{We write } X \sim \text{Poisson}(\lambda)}$ $\quad \lambda$: 발생율.

Thm 3.4.1

(a) $X\sim \text{Poisson}(\lambda) \Rightarrow mgf_X(t) = e^{-\lambda+\lambda e^{t}}$

(b) $X\sim \text{Poisson}(\lambda) \Rightarrow E[X]=\lambda \quad Var(X)=\lambda$

(c) $X_1\sim \text{Poisson}(\lambda_1),\ X_2\sim \text{Poisson}(\lambda_2)$ indep.

$\Rightarrow X_1+X_2 \sim \text{Poisson}(\lambda_1+\lambda_2)$

$X \sim \text{Poisson}(\lambda) \quad f(x) := \dfrac{e^{-\lambda}\lambda^x}{x!} \quad (x = 0, 1, 2, \dots)$

(a) $mgf_X(t) = E[e^{tX}] = \displaystyle\sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!}$

$\because \exp(x) = \displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$

$$= \sum_{x=0}^{\infty} (\lambda e^t)^x / x! \; e^{-\lambda}$$

$$= \exp(-\lambda + \lambda e^t)$$

(b) $cgf_X(t) = -\lambda + \lambda e^t$

$C_1(t) = \lambda e^t \quad C_2(t) = \lambda e^t$

$\therefore E[X] = \lambda \quad Var(X) = \lambda$

(c) $mgf_{X_1+X_2}(t) = E[e^{tX_1 + tX_2}]$

$\because$ independent

$$= E[e^{tX_1}]\,E[e^{tX_2}]$$

$$= mgf_{X_1}(t)\, mgf_{X_2}(t)$$

$$= \exp(-\lambda_1 + \lambda_1 e^t)\,\exp(-\lambda_2 + \lambda_2 e^t)$$

$$= \exp(-(\lambda_1 + \lambda_2) + (\lambda_1 + \lambda_2)e^t)$$

$\therefore$ 적률생성함수의 분포결정성에 의해 $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$ $\square$

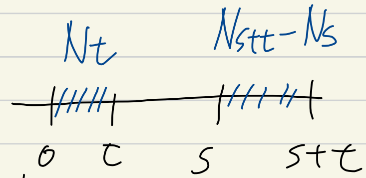
Poisson Process.

시각 t까지 특정 현상이 발생하는 횟수 N_t $(t\geq 0)$

$\{N_t : t\geq 0\}$: "발생률이 λ인 Poisson process."

if (a) 정상성(stationarity)

특정 현상이 발생한 횟수(N_t)의 분포는 시작하는 시각과 관계가 없다.

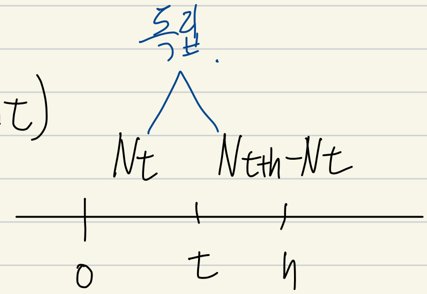


i.e. N_t의 분포와 $N_{t+s}-N_s$의 분포는 같다.

& $N_0=0$

(b) 독립증분성 (independent increment)

N_t와 $N_{t+h}-N_t$는 서로 독립이다.



(c) 비례성(proportionality)

$$P(N_h=1)=\lambda h+o(h) \quad h\to 0$$

(d) 희귀성(rareness) 짧은 시간동안에 현상이 두 번 이상

$$P(N_h=2)=o(h) \quad h\to 0.$$

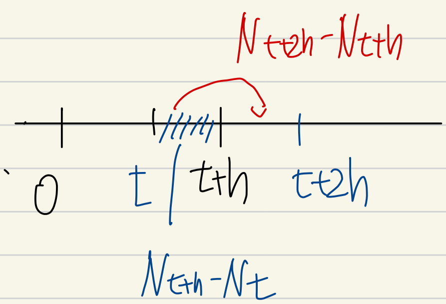
Thm 3.4.2. 발생률이 λ인 poisson process $\{N_t : t \geq 0\} \Rightarrow N_t \sim \text{Poisson}(\lambda t)$

proof) $P(N_{t+h}=x) = P(N_{t+h}=x, N_t=x) + P(N_{t+h}=x-1, N_t=x-1)$
$+P(N_{t+h}=x, N_t \leq x-2)$

$= P(N_t=x, N_{t+h}-N_t=0) + P(N_t=x-1, N_{t+h}-N_t=1)$
$+P(N_{t+h}=x, N_{t+h}-N_t \geq 2)$

$\because$ (a), (b)

$= P(N_t=x)P(N_h=0) + P(N_t=x-1)P(N_h=1)$
$+o(h)$



$P(N_{t+h}=x, N_{t+h}-N_t \geq 2) = P(N_{t+h}=x, N_{t+2h}-N_{t+h} \geq 2)$ ($\because$ (a))

$= P(N_{t+h}=x)P(N_{t+2h}-N_{t+h} \geq 2)$ $\because$ (b)

$= P(N_{t+h}=x)P(N_h \geq 2)$ $\because$ (a)

$= P(N_{t+h}=x)\,o(h)$ $\because$ (d)

$= o(h)$

$g(x,t) := P(N_t = x)$ $\quad P(N_h=0) = \text{Tot} - (N_h = 1) - (N_h \geq 2)$

$$\Rightarrow g(x, t+h) = g(x,t)\left(\left(1-(\lambda h + o(h)\right) - o(h)\right)$$
$$+ g(x-1,t)(\lambda h + o(h)) + o(h)$$

$$\frac{d}{dt}g(x,t) = \lim_{h\to 0}\frac{g(x,t+h)-g(t)}{h}$$

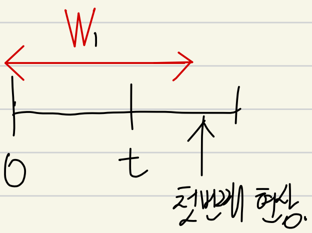
$$\frac{d}{dt}g(x,t) = -\lambda g(x,t) + \lambda g(x-1,t)$$

$$\frac{d}{dt}\left(\frac{e^{\lambda t}g(x,t)}{\lambda^x}\right) = \frac{\lambda e^{\lambda t}g(x,t) + e^{\lambda t}\left(-\lambda g(x,t) + \lambda g(x-1,t)\right)}{\lambda^x}$$
$$= \frac{e^{\lambda t}g(x-1,t)}{\lambda^{x-1}} \quad \cdots\cdots \qquad \square$$

O.D.E를 공부하라...!

지수분포

Exponential dist



발생률이 λ인 Poisson Process $\{N_t : t \geq 0\}$

첫번째 현상이 발생할 때까지의 걸린 시간 $W_1 = \min\{t : N_t \geq 1\}$

$$(W_1 > t) = (N_t = 0)$$

$N_t \sim \text{Poisson}(\lambda t)$

$$P(N_t = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

$$P(W_1 > t) = P(N_t = 0)$$

$$= e^{-\lambda t} \quad t \geq 0.$$

$$P(W_1 \leq t) = \begin{cases} 1 - e^{-\lambda t} & (t \geq 0) \\ 0 & (t < 0) \end{cases}$$

$$= \int_{-\infty}^{t} \lambda e^{-\lambda x} I(x \geq 0)\, dx$$

$$f(x) := \lambda e^{-\lambda x} I(x \geq 0)$$

$$W_1 \sim \text{Exp}(1/\lambda) \quad (\lambda > 0)$$

e.g. $W_1 \sim \text{Exp}(1)$ $\quad \text{pdf}_{W_1}(x) = e^{-x} I(x \geq 0)$

정리 3.5.1

각자!

Thm 3.5.1

$W_1 \sim Exp(1/\lambda)$ $(\lambda > 0)$

(a) $mgf_{W_1}(t) = (1 - t/\lambda)^{-1}$ $\quad (t < \lambda)$

(b) $E[W_1] = 1/\lambda$, $Var(W_1) = 1/\lambda^2$

proof)

(a) $mgf_{W_1}(t) = E[e^{tW_1}] = \int_{-\infty}^{+\infty} e^{tx} \lambda e^{-\lambda x} I_{(x \geq 0)} dx$

$$= \lambda \int_0^\infty e^{-(\lambda - t)x} dx$$

$$= \lambda \left[\frac{-1}{\lambda - t} e^{-(\lambda - t)x} \right]_0^\infty \quad t < \lambda$$

$$= \frac{\lambda}{\lambda - t} = (1 - t/\lambda)^{-1}$$

(b) $cgf_{W_1}(t) = \log(1 - t/\lambda)^{-1}$

$$= -\log(1 - t/\lambda)$$

$C_1(t) = + \frac{+1/\lambda}{1 - t/\lambda}$, $C_1(0) = \frac{1}{\lambda}$

$C_2(t) = \frac{-1/\lambda(-1/\lambda)}{(1 - t/\lambda)^2}$, $C_2(0) = \frac{1}{\lambda^2}$

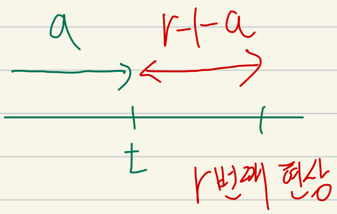
$\therefore E[W_1] = \frac{1}{\lambda}$ $\quad Var(W_1) = \frac{1}{\lambda^2}$ $\quad \square$

Gamma Distribution

발생율 λ인 poisson process $\{N_t : t \ge 0\}$

r번째 현상이 발생할 때까지의 걸린 시간 $W_r = \min\{t : N_t \ge r\}$

$$(W_r > t) = (N_t \le r-1)$$



$r-1-a \ge 0$

$a \le r-1.$

$$P(W_r \le t) = 1 - P(W_r > t)$$
$$= 1 - P(N_t \le r-1)$$
$$= 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$\times(\lambda t)^k$

$$\frac{d}{dt}\left(1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}\right) = -\sum_{k=0}^{r-1}\left(\frac{-\lambda e^{-\lambda t}}{k!} + \frac{\lambda k \lambda^{k-1} t^{k-1} e^{-\lambda t}}{k!}\right)$$
$$= \lambda e^{-\lambda t}\left(\sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} - \frac{k(\lambda t)^{k-1}}{k!}\right)$$
$$= \lambda e^{-\lambda t}\left(\sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} - \sum_{k=1}^{r-1} \frac{(\lambda t)^{k-1}}{(k-1)!}\right)$$

$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$

$\Gamma(n) = (n-1)!, \; n \in \mathbb{N}$

$$= \lambda e^{-\lambda t}\left(\lambda^{r-1} t^{r-1} / (r-1)!\right)$$
$$= \lambda^r t^{r-1} e^{-\lambda t} / (r-1)! \qquad t > 0$$

$\Gamma(r)$

$\therefore W_r \sim Gamma(r, \frac{1}{\lambda})$ 형상모수(shape param.) 척도모수(scale param.)

In general, $X \sim Gamma(\alpha, \beta)$ $(\alpha > 0, \beta > 0)$

$$\Leftrightarrow pdf_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} I(x > 0)$$

$X \sim \text{Gamma}(\alpha, \beta) \quad f(x) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} I(x>0)$

Thm 3.5.2.

(a) $E[X] = \alpha\beta \quad Var(X) = \alpha\beta^2$

(b) $mgf_X(t) = (1-\beta t)^{-\alpha} \quad t < 1/\beta$

(c) $X_1 \sim \text{Gamma}(\alpha_1, \beta) \quad X_2 \sim \text{Gamma}(\alpha_2, \beta)$
indep.
$\Rightarrow X_1 + X_2 \sim \text{Gamma}(\alpha_1+\alpha_2, \beta)$

$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$

$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

proof)

(a) $E[X] = \int_0^{+\infty} x \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} dx$

$y = \dfrac{x}{\beta} \quad = \dfrac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^\alpha e^{-y} \beta dy$

$= \dfrac{\beta}{\Gamma(\alpha)} \Gamma(\alpha+1) = \alpha\beta$

(b) $E[X^2] = \int_0^{+\infty} x^2 \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$

$= \dfrac{\beta}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha+1} e^{-y} \beta dy$

$= \dfrac{\beta^2}{\Gamma(\alpha)} \Gamma(\alpha+2) = \alpha(\alpha+1)\beta^2$

$$\therefore Var(X) = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2$$
$$= \alpha\beta^2(\alpha+1-\alpha)$$
$$= \alpha\beta^2$$

(b) $mgf_X(t) = \int_0^{+\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx$

$y = (1/\beta - t)x \quad = \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{1}{\beta^\alpha}(1/\beta - t)^{-\alpha+1} y^{\alpha-1} e^{-y} dx$

$dy = (1/\beta - t)dx$

$$= \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{1}{\beta^\alpha}(1/\beta - t)^{-\alpha+1} e^{-y} y^{\alpha-1} dy$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{1}{\beta^\alpha}(1/\beta - t)^{-\alpha} e^{-y} y^{\alpha-1} dy$$

$$= \frac{(1/\beta - t)^{-\alpha}}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty y^{\alpha-1} e^{-y} dy$$

$$= \frac{(1/\beta - t)^{-\alpha}}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)$$

$$= (\beta(1/\beta - t))^{-\alpha}$$

$$= (1 - \beta t)^{-\alpha}$$

(c) $mgf_{X_1+X_2}(t) = mgf_{X_1}(t)\, mgf_{X_2}(t)$

$$= (1-\beta t)^{-\alpha_1}(1-\beta t)^{-\alpha_2}$$

$$= (1-\beta t)^{-(\alpha_1+\alpha_2)}$$

$\therefore\ X_1 + X_2 \sim Gamma(\alpha_1+\alpha_2, \beta)$ $\square$

Representational definition

$$X \sim Gamma(\alpha, \beta) \iff X \overset{d}{=} Z_1 + \cdots + Z_r$$

$$Z_i \overset{iid}{\sim} Exp(\beta)$$

Normal Distribution (정규분포)

$$\phi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty < z < +\infty$$

$$\int_{-\infty}^{+\infty} \phi(z)\,dz = 1$$

$$f(x) = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

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

Thm 3.6.1

(a) $E[X] = \mu \qquad Var(X) = \sigma^2$

(b) $mgf_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \qquad -\infty < t < \infty$

(c) $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ indep.

$\Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

proof) (a)

$$E[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx$$

$z := (x-\mu)/\sigma$

$$= \int_{-\infty}^{+\infty} (\sigma z + \mu) \frac{1}{\cancel{\sigma}} \phi(z) \cancel{\sigma} dz$$

$$= \sigma \int_{-\infty}^{+\infty} z \phi(z) dz + \mu$$

$y := \frac{z^2}{2}$

$$= \sigma\left(\int_{0}^{+\infty} z \phi(z) dz + \int_{-\infty}^{0} z \phi(z) dz\right) + \mu$$

$dy = z dz$

$$= \sigma\left(\int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y} dy + \int_{+\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-y} dy\right) + \mu$$

$$= \sigma\left(\int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y} dy - \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y} dy\right) + \mu$$

$$= \mu.$$

$$Var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx$$

$$z := (x-\mu)/\sigma$$

$$= \int_{-\infty}^{+\infty} \sigma^2 z^2 \frac{1}{\cancel{\sigma}}\phi(z)\cancel{\sigma}\,dz$$

$$= \sigma^2 \int_{-\infty}^{+\infty} z^2 \phi(z)\,dz$$

$\because$ even function

$$= 2\sigma^2 \int_{0}^{+\infty} z^2 \phi(z)\,dz$$

$$y := \frac{z^2}{2}$$

$$dy = z\,dz$$

$$= 2\sigma^2 \int_0^{\infty} \sqrt{2y}\,\frac{1}{\sqrt{2\pi}} e^{-y}\,dy$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} \sqrt{y}\, e^{-y}\,dy$$

$$\Gamma(\alpha) := \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \sigma^2.$$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$$\Gamma\left(1+\frac{1}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

$$\therefore Var(X) = \sigma^2$$

(b)

$$mgf_X(t) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx$$

$z = \frac{x-\mu}{\sigma}$

$$= \int_{-\infty}^{+\infty} e^{t(\sigma z+\mu)} \frac{1}{\cancel{\sigma}} \phi\left(\frac{x-\mu}{\sigma}\right) \cancel{\sigma}\, dz$$

$$= \int_{-\infty}^{+\infty} e^{t(\sigma z+\mu)} \phi(z) dz$$

$$\int_{-\infty}^{+\infty} e^{t\sigma z} \phi(z) dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + t\sigma z} dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2t\sigma z + t^2\sigma^2) + \frac{1}{2}t^2\sigma^2} dz$$

$y := z - \sigma t$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t\sigma)^2} dz \; e^{\frac{t^2\sigma^2}{2}}$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \; e^{\frac{t^2\sigma^2}{2}}$$

$$= e^{\frac{t^2\sigma^2}{2}}$$

$$\therefore \ mgf_X(t) = \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)} \phi(z)\, dz$$

$$= e^{\mu t} \int_{-\infty}^{+\infty} e^{t\sigma z} \phi(z)\, dz$$

$$= e^{\mu t} e^{t^2\sigma^2/2}$$

$$= \exp(\mu t + t^2\sigma^2/2)$$

(c) $X_1 \sim (\mu_1, \sigma_1^2)$, $X_2 \sim (\mu_2, \sigma_2^2)$ indep.

$$mgf_{X_1+X_2}(t) = E[e^{tX_1+tX_2}]$$

$$= mgf_{X_1}(t)\, mgf_{X_2}(t)$$

$$= \exp(-\mu_1 t + t^2\sigma_1^2/2 - \mu_2 t + t^2\sigma_2^2/2)$$

$$= \exp(-(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2)$$

$$\therefore X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$\square$

$$mgf_X(t) = \exp(\mu t + t^2\sigma^2/2)$$

Thm 3.6.2.	(a) $X \sim N(\mu, \sigma^2)$ $\Rightarrow aX+b \sim N(a\mu+b, a^2\sigma^2)$
Representational Definition	(b) $X \sim (\mu, \sigma^2)$ $\quad X \overset{d}{=} \sigma Z + \mu \quad Z \sim N(0,1)$
proof)	$E[e^{t(aX+b)}] = e^{tb}[e^{taX}] = mgf_X(at)\, e^{bt}$ $= \exp(\mu(at) + (a^2t^2)\sigma^2/2 + bt)$ $= \exp((a\mu+b)t + t^2(a\sigma)^2/2)$ $Y \sim N(a\mu+b, a^2\sigma^2)$ $mgf_Y(t) = \exp((a\mu+b)t + t^2(a\sigma)^2/2)$ $\therefore aX+b \sim N(a\mu+b, a^2\sigma^2)$ (b) $Z \sim N(0,1)$ $aZ+b \sim N(a\cdot 0+b, a^2 1^2) =$ $a := \sigma \quad b := \mu$ $X := \sigma Z + b \sim N(\mu, \sigma^2)$ $\square$