Chapter 3 : Probability Distributions

3.1. HYPERGEOMETRIC DISTRIBUTIONS

Definition: population, population distribution, Sampling without replacement, Simple random sampling, random sample, Sample

Let N : Size of population
D : Size of "True" | 1 in population
n : sample Size

$$f(x) = \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n} \qquad 0 \le x \le D,\ 0 \le n-x \le N-D$$

is called the hypergeometric distribution and is denoted as:

$$X \sim H(n; N, D)$$

$$E[X] = \sum_x x \cdot \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}$$

$$= \sum_{x=1}^{D} x \cdot \frac{D!}{x!(D-x)!} \cdot \frac{(N-D)!}{(n-x)!(N-D-(n-x))!} \Big/ \frac{N!}{n!(N-n)!}$$

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

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

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

$$= n \cdot \frac{D}{N}$$

* $x=1$ 에서 시작하는 이유는 $E[X] = \sum x \cdot f(x)$ 이기 때문에 $x=0$ 인 때 $0 \cdot f(x=0) = 0$. 즉 식이 같음.

* [Textbook]

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

$$= \sum_{y=0}^{D-1} \binom{D-1}{y}\binom{(N-1)-(D-1)}{(n-1)-y} \Big/ \binom{N}{n} \qquad * \sum_{k=0}^{l} \binom{x}{k}\binom{y}{l-k} = \binom{x+y}{l}$$

$$= \binom{N-1}{n-1} \Big/ \binom{N}{n} = 1 \Big/ \frac{N}{n} = \frac{n}{N}$$

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

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

$$= \sum_{y=0}^{D-2} D(D-1) \binom{D-2}{y}\binom{(N-2)-(D-2)}{(n-2)-(x-2)} \Big/ \binom{N}{n}$$

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

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

$$\therefore E[X^2] - E[X] = D(D-1)\frac{n(n-1)}{N(N-1)}, \quad E[X^2] = E[X(X-1)] + E[X] = D(D-1)\frac{n(n-1)}{N(N-1)} + n\cdot\frac{D}{N}$$

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

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

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

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

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

$$= n\cdot\frac{D}{N}\left\{\frac{nDN - nN - DN + N + N^2 - N - nDN + nD}{(N-1)N}\right\}$$

$$= n\cdot\frac{D}{N}\left\{\frac{N^2 - nN + nD - DN}{(N-1)N}\right\} = n\cdot\frac{D}{N}\left\{\frac{(N-n)N - (N-n)D}{(N-1)N}\right\}$$

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

Let $D/N = p$. $\mathrm{Var}(X) = \frac{N-n}{N-1}\cdot n\cdot p(1-p)$ Q.E.D.

Expectation and variance of hypergeometric distribution.

$X \sim H(n; N, D)$, let $p = D/N$ be the population proportion.

$$E[X] = np, \quad \mathrm{Var}(X) = \frac{N-n}{N-1}\,np(1-p)$$

Ex. 3.1.1

For quality control we randomly sample $n=5$ from a population $N=60$. If the number of faulty products exceed the threshold, we disqualify the batch(N). If the batch has 6 faulty products, meaning $D=6$, what should the threshold be for the probability of disqualification to be lesser than or equal to 1%?

$X \sim H(n=5; N=60, D=6)$, $f_X(x) = \binom{D}{x}\binom{N-D}{n-x} \Big/ \binom{N}{n}$

$$P(X \ge K) = 1 - P(X < K) = 1 - P(X \le K-1) = 1 - \sum_{x=0}^{K-1} \binom{6}{x}\binom{54}{5-x} \Big/ \binom{60}{5}$$

	K=0	K=1	K=2	K=3	K=4	K=5
P(X=K)	0.579053	0.347432	0.068123	0.005240	0.000148	0.000001
P(X≤K)	0.579053	0.926485	0.994608	0.999848	0.999999	1.0

$\therefore$ Threshold: $P(X \ge 3) = 1 - P(X \le 2) = 0.005402 < 0.01$

$\therefore$ Disqualify batch if 3 or more faulty products are found.

In a hypergeometric distribution, if the population size N is significantly greater than the sample size n, we can approximate the probability:

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

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

$$\fallingdotseq \binom{n}{x}\left(\frac{D}{N}\right)^x\left(1-\frac{D}{N}\right)^{n-x}$$

The probability on the RHS is the probability of sampling with replacement

In other words, if the population size is large enough,

Sampling without replacements approximates to sampling with replacement

3.2. BINOMIAL AND MULTINOMIAL DISTRIBUTION.

Suppose that the components of a population is either '0' or '1', and the proportion of '1' in the population is p.
If X denotes the number of '1's in a sample of size n that was sampled with replacement, the pdf of X is:

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x},\quad x=0,1,2,\cdots,n$$

We call this distribution the binomial distribution and denote it as $X\sim B(n,p)$

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

$$E[X(X-1)]=\sum_{x=0}^{n}x(x-1)\cdot\binom{n}{x}p^x(1-p)^{n-x}=n(n-1)p^2\sum_{x=2}^{n}\binom{n-2}{x-2}p^{x-2}(1-p)^{n-x}=n(n-1)p^2$$

$$\begin{aligned}Var(X)=E[X^2]-E[X]^2&=E[X(X-1)]+E[X]-E[X]^2=n(n-1)p^2+np-n^2p^2\\&=\cancel{n^2p^2}-np^2+np-\cancel{n^2p^2}\\&=np(1-p)\end{aligned}$$

[Textbook proof]

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

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

Definition: Mean and Variance of a Binomial Distribution

If $X\sim B(n,p)$, $\quad E[X]=np$, $\quad Var(X)=np(1-p)$

(설명) 이항분포의 유도과정에서 복원추출의 역할에 대하여 생각해보자. 각 개체가 '0' 또는 '1'의 두 가지로 분류되고 '1'의 비율이 p인 모집단에서 한 개씩 복원추출 하는 경우에는 각각의 추출결과가 서로 독립이고 같은 분포를 따르는 확률변수로 나타내어진다. 즉 n개씩의 추출결과를 $Z_1, Z_2, \cdots, Z_n, \cdots$ 이라고 하면 이들은 서로 독립이고 각각의 분포도

$$P(Z_i=1)=p,\ P(Z_i=0)=1-p\quad (i=1,2,\cdots,n)$$

로서 동일하다. 여기에서 각 Z_i의 분포는 베르누이(Bernoulli) 분포라고 하며 모집단의 분포를 뜻하고, 기호로는 $Z_i\sim Bernoulli(p)$ 와 같이 나타낸다.

이와 같이 두 가지로 분류된 모집단을 관측하는 것을 베르누이 시행 (Bernoulli trial) 이라고 하며, 서로 독립인 베르누이 시행이란 서로 독립이고 베르누이분포를 따르는 확률변수들을 관측하는 것을 뜻한다. 또한 '1'과 '0'을 각각 '성공'과 '실패'로 부르고, 이항분포란 서로 독립인 베르누이 시행에서 나오는 '성공' 횟수의 분포라고도 한다. 즉 서로 독립인 베르누이 시행을 이용하여 이항분포를 다음과 같이 정의할 수 있고, 이를 이용하여 이항분포의 평균이나 분산의 공식을 쉽게 유도할 수 있을 뿐만 아니라 이항분포의 여러가지 성질을 쉽게 밝힐 수 있다.

Definition: Representational Definition (대의적 정의) of the Binomial Distribution

$$B(n,p)\overset{d}{=}Bernoulli_1(p)\oplus\cdots\oplus Bernoulli_n(p)$$

i.e. $X\sim B(n,p)\Leftrightarrow X\overset{d}{=}Z_1+\cdots+Z_n,\ Z_i\overset{i.i.d}{\sim}Bernoulli(p)\ (i=1,\cdots,n)$

⟨Theorem 3.2.1. Properties of the Binomial Distribution⟩

(a) If $X\sim B(n,p)$, $mgf_X(t)=(pe^t+q)^n$, $-\infty<t<\infty$ $(q=1-p)$

(b) If $X_1\sim B(n_1,p)$, $X_2\sim B(n_2,p)$ and X_1 and X_2 are independent then,

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

(proof) $Z_i\sim Bernoulli(p)\quad (i=1,\cdots,n)$

(a) pdf of Z_i : $f_{Z_i}(z)=p^z(1-p)^{1-z}\quad (z=0,1)$

$$M_Z(t)=E[e^{tz}]=\sum_{z=0}^{1}e^{tz}p^z(1-p)^{1-z}$$

$$=(1-p)+e^tp=pe^t+q,\quad q=1-p$$

$$M_X(t)=E[e^{tX}]=E[e^{t(z_1+\cdots+z_n)}]=E[e^{tz_1}]\cdots E[e^{tz_n}]$$

$$=\underbrace{(pe^t+q)\cdots(pe^t+q)}_{n\text{개}}=(pe^t+q)^n$$

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

$$=(pe^t+q)^{n_1}(pe^t+q)^{n_2}$$

$$=(pe^t+q)^{n_1+n_2}:\ mgf\ of\ X\sim B(n_1+n_2,p)$$

* distribution-determining property of mgf.

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

Ex. 3.2.1. For i.i.d r.v.s $X_1,\cdots,X_K$ that follow the Bernoulli(p) $(0<p<1)$ distribution, show that

$$(X_1+\cdots+X_D\mid X_1+\cdots+X_N=n)\sim H(n;N,D)\quad (D\le N)$$

i.e. given that $X_1+\cdots+X_N=n$, the conditional distribution of $X_1+\cdots+X_D$ is the hypergeometric distribution $H(n;N,D)$

$X_1+\cdots+X_N\sim Bin(N,p)$, $X_1+\cdots+X_D=Bin(D,p)$

$$f(x_1+\cdots+x_N)=\binom{N}{x}p^x(1-p)^{N-x},\quad f(x_1+\cdots+x_D)=\binom{D}{y}p^y(1-p)^{D-y}$$

$$f(x_1+\cdots+x_D\mid x_1+\cdots+x_N)=f(x_1+\cdots+x_D,\ x_1+\cdots+x_n)\big/f(x_1+\cdots+x_n)$$

$$\begin{aligned}&f(x_1+\cdots+x_D=m,\ x_{D+1}+\cdots+x_N=n-m)\\&=f(x_1+\cdots+x_D=m)f(x_{D+1}+\cdots+x_N=n-m)\\&=\binom{D}{m}p^m(1-p)^{D-m}\binom{N-D}{n-m}p^{n-m}(1-p)^{N-D-(n-m)}\\&=\binom{D}{m}\binom{N-D}{n-m}p^n(1-p)^{N-n}\end{aligned}$$

$$f(x_1+\cdots+x_N=n)=\binom{N}{n}p^n(1-p)^{N-n}$$

$$\therefore f(x_1+\cdots+x_D=m\mid x_1+\cdots+x_N=n)=\binom{D}{m}\binom{N-D}{n-m}\Big/\binom{N}{n}$$

Now let's suppose that there are three or more categories in a population and their respective proportions be $p_1, \cdots, p_K$.
Let X_i $(i=1, \cdots, K)$ denote the number of items - belonging to category i - in a sample that was sampled randomly with replacement.
We can express the joint pdf of $X_1, \cdots, X_K$ using the multinomial coefficient[48] as the following:

$$f(x_1, \cdots, x_K) = \binom{n}{x_1\, x_2 \cdots x_K} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K}$$

$$x_i = 0, \cdots, n \quad (i=1, 2, \cdots, K) \qquad x_1 + \cdots + x_K = n$$

We call this joint pdf the multinomial distribution[49] and denote it as:

$$X = (X_1, \cdots, X_K)^t \sim \text{Multi}(n, (p_1, \cdots, p_K)^t)$$

Observing^ one by one from a population with three or more categorical components is called a "multinomial trial" and sampling with replacement indicates an "independent multinomial trial".
(the sample)
If a^ single observation (the result of a) of an independent multinomial trial is $Z_i = (Z_{i1}, Z_{i2}, \cdots, Z_{iK})^t$ $(i=1, \cdots, n)$, they are mutually independent and their distribution is identical since:

$$P(Z_{i1}=z_1, Z_{i2}=z_2, \cdots, Z_{iK}=z_K) = p_1^{z_1} p_2^{z_2} \cdots p_K^{z_K},$$
(lower letter case.)

$$z_i = 0, 1 \ (i=1, \cdots, K), \quad z_1 + \cdots + z_K = 1$$

Hence, the distribution of Z_i is a multinomial distribution with $n=1$:
$Z_i \sim \text{Multi}(1, (p_1, \cdots, p_K)^t)$

The number of each categorical items^ (i.e., X_i) in an independent multinomial trial is the sum of independent $Z_1, Z_2, \cdots, Z_K \sim \text{Multi}(n=1, (p_1, \cdots, p_K)^t)$
Thus, the multinomial distribution can be defined as the following:
(서로 독립인 n번의 다항시행 에서 나오는 각 유형의 개수는 서로 독립인 각 시행에서의 결과를 합한 것이므로 다음과 같이 다항분포가 정의되는 것을 알 수 있다.)

48) $\binom{n}{x_1\, x_2 \cdots x_K}$ 는 서로 다른 유형의 $x_1, x_2, \cdots, x_K$ 개를 나열하는 방법의 수인 $\frac{n!}{x_1!\, x_2! \cdots x_K!}$ 를 나타내는 다항계수 (multinomial coefficient) 이다

49) $k=2$ 인 경우의 다항분포 $\text{Multi}(n, (p, 1-p)^t)$는 이항분포 $B(n,p)$를 따르는 확률변수 X에 대하여 X와 $n-X$의 결합분포를 뜻하고 있는 것이다. 즉
$(X, n-X)^t \sim \text{Multi}(n, (p, 1-p)^t)$

Definition: Representational Definition of Multinomial Distribution

$$\text{Multi}(n, p) \overset{d}{\equiv} \text{Multi}_1(1, p) \oplus \cdots \oplus \text{Multi}_n(1, p)$$

$$X = (X_1, X_2, \cdots, X_K)^t \sim \text{Multi}(n, (p_1, p_2, \cdots, p_K)^t)$$

$$\Leftrightarrow X \overset{d}{\equiv} Z_1 + \cdots + Z_n, \ Z_i = (Z_{i1}, Z_{i2}, \cdots, Z_{iK})^t \overset{i.i.d}{\sim} \text{Multi}(1, (p_1, p_2, \cdots, p_K)^t) \quad (i=1, 2, \cdots, K)$$

〈Theorem 3.2.2: Properties of Multinomial Distribution〉

(a) If $X = (X_1, X_2, \cdots, X_K)^t \sim \text{Multi}(n, (p_1, \cdots, p_K)^t)$

$$E[X_l] = np_l \quad (l=1, \cdots, K)$$

$$\text{Var}(X_l) = np_l(1-p_l), \ \text{Cov}(X_l, X_m) = -np_lp_m \ | \ l \neq m, \ l, m = 1, 2, \cdots, K)$$

(b) If $X = (X_1, X_2, \cdots, X_K)^t \sim \text{Multi}(n, (p_1, \cdots, p_K)^t)$,

$$mgf_X(t) = (p_1 e^{t_1} + \cdots + p_K e^{t_K})^n, \ -\infty < t_l < +\infty \quad (l=1, \cdots, K)$$

(proof)

(a) $X \overset{d}{\equiv} \text{Multi}(1, p) \oplus \cdots \oplus \text{Multi}(1, p)$
$\overset{d}{\equiv} Z_1 \oplus \cdots \oplus Z_n$

Z_i : i^{th} trial of the independent multinomial trial
Z_i pdf : $f_Z(z_{i1}=z_1, z_{i2}=z_2, \cdots z_{ik}=z_K)$

$$= p_1^{z_1} p_2^{z_2} \cdots p_K^{z_K} \quad \left(z_i = 0, 1, \ \sum_{i=1}^{K} z_i = 1\right)$$

$$E[Z_i] = \sum_z z\, (p_1^{z_1} p_2^{z_2} \cdots p_K^{z_K}) = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} p_1 + \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} p_2 + \cdots + \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix} p_K = \begin{bmatrix}p_1\\\vdots\\p_K\end{bmatrix}$$

$$E[X] = E[Z_1] + \cdots + E[Z_n] = \underbrace{\begin{bmatrix}p_1\\\vdots\\p_K\end{bmatrix} + \cdots + \begin{bmatrix}p_1\\\vdots\\p_K\end{bmatrix}}_{n\text{개}} = \begin{bmatrix}np_1\\\vdots\\np_K\end{bmatrix} = np$$

$\Leftrightarrow E[X_l] = np_l \ (l=1, \cdots, K)$

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

$$\text{Var}(Z_1) = E\left[(Z_1 - E[Z_1])(Z - E[Z_1])^t\right]$$

$$E\left[\left[(Z_{11}\ Z_{12} \cdots Z_{1K})^t - (p_1\ p_2 \cdots p_K)^t\right]\left[(Z_{11}\ Z_{12} \cdots Z_{1K}) - (p_1\ p_2 \cdots p_K)\right]\right]$$

$$(Z_{1i} - p_i)^t (Z_{1j} - p_j) = Z_{1i} \cdot Z_{1j} - p_i Z_{1j} - p_j Z_{1i} + p_i p_j$$

$$E[\quad \text{〃} \quad] = \underbrace{E[Z_{1i} \cdot Z_{1j}]}_{=0\ (i \neq j),\ p_i(1-p_j)\ (i=j) \ \Rightarrow\ p_i q_j} - p_i p_j - p_j p_i + p_i p_j$$

$$= -p_i p_j \ (\text{if } i \neq j), \quad p_i q_i \ (\text{if } i = j)$$

$\therefore \text{Var}(Z_1) = (-p_i p_j) \ (i \neq j), \ p_i q_i \ (i=j)$

$$\therefore \text{Var}(X) = n \cdot \text{Var}(Z_1) = \begin{cases} -np_ip_j & \text{if } i \neq j \\ np_iq_i & \text{if } i=j, \ q_i = 1-p_i \end{cases}$$

(b) $X=(X_1,\cdots,X_k)^t \sim \text{Multi}(n,(p_1,p_2,\cdots,p_k)^t)$. mgf is:

$$mgf_X(t)=(p_1e^{t_1}+\cdots+p_ke^{t_k})^n,\ -\infty<t_i<+\infty\ \ (i=1,\cdots,k)$$

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

$$mgf_{Z_1}(t)=E\left[e^{t^tZ_1}\right]=E\left[\exp\left\{t_1Z_{11}+\cdots+t_kZ_{1k}\right\}\right]$$
$$=\sum_{Z}\exp\left\{t_1Z_{11}+\cdots+t_kZ_{1k}\right\}\cdot p_1^{Z_{11}}p_2^{Z_{12}}\cdots p_k^{Z_{1k}}$$

$$\left((Z_{11},Z_{12},\cdots,Z_{1k}):Z_{1i}=0,1,\ \sum_{i=1}^{k}Z_{1i}=1,\ i=1,\cdots,k\right)=S(Z_1)$$

$$E\left[e^{t^tZ_1}\right]=\sum_{Z_1=(Z_{11},\cdots,Z_{1k})\in S(Z_1)}\exp\left\{t_1Z_{11}+\cdots+t_kZ_{1k}\right\}\cdot p_1^{Z_{11}}p_2^{Z_{12}}\cdots p_k^{Z_{1k}}$$

$$=\exp\left\{t_1\cdot1+t_2\cdot0+\cdots+t_k0\right\}p_1+\cdots+\exp\left\{t_i\cdot0+\cdots+t_k\cdot1\right\}\cdot p_k$$
$$=e^{t_1}p_1+e^{t_2}p_2+\cdots+e^{t_k}p_k$$

$$\therefore M_X(t)=M_{Z_1}(t)\cdots M_{Z_n}(t)=E\left[e^{t^tZ_1}\right]\cdots E\left[e^{t^tZ_n}\right]$$
$$=\left(p_1e^{t_1}+\cdots+p_ke^{t_k}\right)^n,\ -\infty<t_i<+\infty\ \ (i=1,\cdots,k)$$

[Textbook proof]

다항분포의 대의적 정의와 〈정리 2.5.11〉로 부터 X의 평균벡터, 분산행렬과 적률생성함수는

$$E[X]=E[Z_1+\cdots+Z_n]=nE[Z]$$
$$Var(X)=Var(Z_1+\cdots+Z_n)=nVar(Z_1)$$
$$mgf_X(t)=mgf_{Z_1+\cdots+Z_n}(t)=\left(mgf_{Z_1}(t)\right)^n$$

한편, $Z_1=(Z_{11},Z_{12},\cdots,Z_{1k})^t \sim \text{Multi}(1,(p_1,p_2,\cdots,p_k)^t)$ 이므로

$$P(Z_{1l}=1)=p_l,\ P(Z_{1l}=0)=1-p_l,\ E[Z_{1l}Z_{1m}]=0\ \ (l\neq m,\ l,m=1,\cdots,k)$$

$$\therefore\begin{cases}E[Z_{1l}]=p_l\\ Cov(Z_{1l},Z_{1m})=E[Z_{1l}Z_{1m}]-E[Z_{1l}]E[Z_{1m}]=-p_lp_m\ \ (l\neq m,\ l,m=1,\cdots,k)\end{cases}$$

또한, $Z_1=(Z_{11},\cdots,Z_{1k})^t$ 의 적률생성함수는

$$mgf_{Z_1}(t)=E\left[e^{t_1Z_{11}+\cdots+t_kZ_{1k}}\right]=p_1e^{t_1}+\cdots+p_ke^{t_k}$$

3.3. GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTION.

Suppose that $X_1,\cdots,X_n \sim \text{Bernoulli}(p)$ and let W_1 the number of trials until the first "success" is observed.
There must be a consecutive "fails" until the first "success" thus,

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

We call this distribution the geometric distribution(기하분포)[50] and denote it as $W_1 \sim Geo(p)$.

*50) 무한급수 $\sum_{x=1}^{\infty}q^{x-1}$ 을 기하급수라고 부르는 데에서 기하분포의 이름이 유래되었다.

〈Theorem 3.3.1. Properties of the Geometric Distribution〉

(a) If $W_1 \sim \geq o(p)$*, the mgf is *"0부터 시작하는 기하분포" +ChatGPT 5.0

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

(b) If $W_1 \sim \geq o(p)$, $E[W_1]=1/p$, $Var(W_1)=q/p^2$ $(q=1-p)$

(proof)

$$f_X(x)=(1-p)^{x-1}p$$

$$E[e^{tX}]=\sum_{x=1}^{\infty}e^{tx}\cdot(1-p)^x\cdot(1-p)^{-1}\cdot p=(1-p)^{-1}p\sum_{x=1}^{\infty}(e^t-e^tp)^x$$

$$=(1-p)^{-1}p\cdot\left\{\sum_{x=0}^{\infty}(e^t-e^tp)^x-1\right\}$$

$$=(1-p)^{-1}p\cdot\left\{(1-e^t+e^tp)^{-1}-1\right\}\ ,\ |e^t-e^tp|<1$$

$$=(1-p)^{-1}p\cdot\left\{\frac{1-1+e^t-e^tp}{1-e^t-e^tp}\right\}\qquad e^t|1-p|<1\quad \because 0\leq p\leq1$$
$$e^t<(1-p)^{-1}\Leftrightarrow t<-\log q$$

$$=(1-p)^{-1}p\cdot\frac{e^t(1-p)}{(1-qe^t)}=pe^t(1-qe^t)^{-1}$$

(b) cgf of X : $\log M_X(t)=\log p+t-\log(1-qe^t)$ $\quad t<-\log q$

$$\frac{\partial}{\partial t}C_X(t)=1+\frac{qe^t}{1-qe^t}\qquad \frac{\partial}{\partial t}C_X(t=0)=1+q(1-q)^{-1}=\frac{1-q+q}{1-q}=p^{-1}$$

$$\frac{\partial}{\partial t^2}C_X(t)=qe^t(1-qe^t)^{-1}+qe^t\cdot qe^t(1-qe^t)^{-2}$$

$$\frac{\partial}{\partial t^2}C_X(t=0)=q(1-q)^{-1}+q^2(1-q)^2=\frac{q(1-q)+q^2}{(1-q)^2}=\frac{q-q^2+q^2}{(1-q)^2}=q/p^2$$

[Textbook Solution]

(a) 기하분포 $\geq o(p)$ 의 적률생성함수는 기하급수의 공식으로부터

$$E[e^{tW_1}]=\sum_{x=1}^{\infty}e^{tx}(1-p)^{x-1}p$$

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

(b) 기하분포의 누율생성함수는

$$cgf_{W_1}(t) = -\log(1-qe^t) + t + \log p$$
$$= t - \log\left\{p - q(e^t-1)\right\} + \log p$$
$$= t - \log\left\{1 - \frac{q}{p}(e^t-1)\right\}, \quad t < -\log q$$

여기에서 로그함수의 멱급수 전개식인

$$-\log(1-A) = A + A^2/2 + A^3/3 + \cdots \quad (A \simeq 0)$$

을 이용하여, 누율생성함수를 멱급수 전개하고 t의 올림차순으로 정리하면

$$cgf_{W_1}(t) = t + \frac{q}{p}(e^t-1) + \frac{1}{2}\left\{\frac{q}{p}(e^t-1)\right\}^2 + \cdots$$
$$= t + \frac{q}{p}t + \frac{1}{2}\left\{\frac{q}{p} + \left(\frac{q}{p}\right)^2\right\}t^2 + \cdots$$
$$= \frac{1}{p}t + \frac{1}{2}\frac{q}{p^2}t^2 + \cdots, \quad t < -\log q$$

$$\therefore E[x] = cgf'_{W_1}(0) = p_1^{-1}, \quad Var(W_1) = cgf''_{W_1}(0) = q/p^2$$

Meanwhile, for $X_1, \cdots, X_n \sim$ Bernoulli(p), let W_r be the number of trials until the rth success,

$$(W_r = x) = (\text{until the } (x-1)_{th} \text{ trial, } (r-1)\text{"Success"s and "Success" on the } x^{th} \text{ trial.})$$

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

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

We call this distribution the negative binomial distribution[51] and denote it as $W_r \sim Neg(r,p)$

Meanwhile, let $W_1, W_2 - W_1, \cdots, W_r - W_{r-1}$ denote the number of trials it takes next "success" after a "Success". We can observe the following:

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

$= P($ consecutive "fails" for $x_i - 1$ times and then "Success", $i = 1, \cdots, r)$

$$= \underbrace{\left\{(1-p)^{x_1-1}p\right\}}_{g_1(x_1)}\underbrace{\left\{(1-p)^{x_2-1}p\right\}}_{g_2(x_2)}\cdots\underbrace{\left\{(1-p)^{x_r-1}p\right\}}_{g_r(x_r)}, \quad x_i = 1,2,\cdots \ (i = 1,2,\cdots,r)$$

$$\therefore f(w_r) = f(w_r - w_{r-1} + w_{r-1} + \cdots + w_2 - w_1 + w_1)$$
$$= g(w_r - w_{r-1})g(w_{r-1} - w_{r-2})\cdots g(w_1)$$

Thus, $W_1, W_2 - W_1, \cdots, W_r - W_{r-1}$ are independent and follows the same geometric distribution. i.e.,

$$W_1, W_2 - W_1, \cdots, W_r - W_{r-1} \overset{i.i.d}{\sim} \simeq O(p)$$

Thus, we can define the negative binomial distribution, which is the sum of $W_1, W_2 - W_1, \cdots, W_r - W_{r-1}$ as the following:

Definition: Representational Definition of the Negative Binomial Distribution.

$$Negbin(r,p) \overset{d}{\equiv} Geo_1(p) \oplus \cdots \oplus Geo_r(p) \quad \text{meaning:}$$
$$X \sim Negbin(r,p) \Leftrightarrow X \overset{d}{\equiv} Z_1 + \cdots + Z_r, \ Z_i \overset{i.i.d}{\sim} Geo(p) \quad (i = 1, \cdots, r)$$

* 51) 음의 정수를 지수로 갖는 함수 $(1-t)^{-r}$의 다음과 같은 멱급수 전개식을 음이항 전개식 (negative binomial expansion) 이라고 부르는 데에서 음이항분포의 명칭이 유래하였다.

$$(1-t)^{-r} = \sum_{k=0}^{\infty} \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!} t^k = \sum_{k=0}^{\infty} \frac{(r+k-1)!}{k!(r-1)!}(-t)^k, \quad -1 < t < 1$$

우변의 무한합에서 $x = r+k$로 치환하고 $-t = q = 1-p$를 대입하면 다음이 성립함을 알 수 있다.

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

$$* \ (1-t)^{-r} = \sum_{k=0}^{\infty} \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!} t^k = \sum_{k=0}^{\infty} \frac{(-1)^k (r)(r+1)\cdots(r+k-1)}{k!} t^k$$
$$= \sum_{k=0}^{\infty} \frac{(r+k-1)!}{k!(r-1)!}(-t)^k \quad \text{let } -t = 1-p \text{ and } r+k = x \Leftrightarrow k = x-r$$
$$= \sum_{k=0}^{\infty} \frac{(x-1)!}{(x-r)!(r-1)!}(1-p)^k$$
$$= \sum_{x-r=0}^{\infty} \binom{x-1}{r-1}(1-p)^{x-r}$$
$$= \sum_{x=r}^{\infty} \binom{x-1}{r-1}(1-p)^{x-r} = 1$$

〈Theorem 3.3.2. Properties of Negative Binomial Distributions〉

(a) If $X \sim Negbin(r,p)$, $E[x] = r/p$, $Var(X) = rq/p^2$

(b) If $X \sim Negbin(r,p)$, $mgf_X(t) = \left\{pe^t(1-qe^t)^{-1}\right\}^r$, $t < -\log q$ $(q = 1-p)$

(c) If $X_1 \sim Negbin(r_1,p)$, $X_2 \sim Negbin(r_2,p)$ and X_1 and X_2 are independent, $X_1 + X_2 \sim Negbin(r_1 + r_2, p)$

(proof)

(a) $X = W_1 \oplus W_2 - W_1 \oplus \cdots \oplus W_r - W_{r-1}$

$$E[x] = E[W_1] + E[W_2 - W_1] + \cdots + E[W_r - W_{r-1}]$$
$$= p^{-1} + p^{-1} + \cdots + p^{-1}$$
$$= r \cdot p^{-1}$$

$$Var(X) = Var(W_1 + (W_2 - W_1) + \cdots + (W_r - W_{r-1}))$$
$$= Var(W_1) + Var(W_2 - W_1) + \cdots + Var(W_r - W_{r-1})$$
$$= q/p^2 + \cdots + q/p^2$$
$$= rq/p^2$$

(b) $M_X(t) = M_{W_1}(t_1) \cdot M_{W_2 - W_1}(t_2) \cdots M_{W_r - W_{r-1}}(t)$

$$= pe^t(1-qe^t)^{-1} \cdot pe^t(1-qe^t)^{-1} \cdots pe^t(1-qe^t)^{-1}$$
$$= \left\{pe^t(1-qe^t)^{-1}\right\}^r, \ qe^t < 1, \ e^t < q^{-1} \quad t < -\log q$$

(c) $M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = \left\{pe^t(1-qe^t)\right\}^{r_1+r_2}$

: mgf of $Negbin(r_1+r_2, p)$

$\therefore X_1 + X_2 \sim Negbin(r_1+r_2, p)$

3.4. POISSON DISTRIBUTION

If n is large in a Binomial distribution $B(n,p)$, calculation of exact probabilities become incredibly difficult. Thus, the probabilities are approximated. Especially, if n is large and p is small, we can approximate as the following:

$$\binom{n}{x} p^x (1-p)^{n-x} = n(n-1)\cdots(n-x+1)p^x(1-p)^{n-x}/x!$$
$$= \frac{n}{n}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)(np)^x\left(1-\frac{np}{n}\right)^{n-x}/x!$$
$$\simeq (np)^x e^{-np}/x!$$

This approximation assumes that n is large and p is small so that np approximates to some $\lambda>0$. This approximation can be expressed as the following and we call it the poisson approximation of the binomial probability (이항확률의 포아송 근사).

Definition : Poisson Approximation of the Binomial Probability.[52]

$$\lim_{\substack{n\to\infty \\ np_n\to\lambda}} \binom{n}{x} p_n^x (1-p_n)^{n-x} = e^{-\lambda}\lambda^x/x! \quad (\lambda>0)$$

* 52) 이 근사 계산의 과정에서는 다음과 같은 극한 값 계산 결과를 이용하고 있다.

$$\lim_{\substack{n\to\infty \\ a_n\to a}} \left(1+\frac{a_n}{n}\right)^n = e^a, \text{ 즉 } \lim_{\substack{n\to\infty \\ a_n\to a}} n\log\left(1+\frac{a_n}{n}\right) = a$$

Ex. 3.4.1
지난 시즌에서 100타석 당 홈런을 2개 친 야구 선수가 금년에 400타석에서 15개의 홈런을 칠 확률을 구하여라.

$X \sim Bin(400, 0.02)$ $P(X=15) = \binom{400}{15}(0.02)^{15}(0.98)^{385} \simeq 0.0090$ ✓✓

$X \sim Poisson(8)$ $P(X=15) = e^{-\lambda}\lambda^x/x! = e^{-8}8^{15}/15! \approx 0.009025$ ✓✓
$\lambda = np$

The function given as the approximation of the binomial distribution is $f(x) = e^{-\lambda}\lambda^x/x!$, $x=0,1,2..,(\lambda>0)$. We denote 'X follows the Poisson distribution' as $X \sim Poisson(\lambda)$.

〈 Theorem 3.4.1 Properties of the Poisson Distribution 〉

(a) If $X \sim Poisson(\lambda)$, $mgf_X(t) = \exp\{-\lambda+\lambda e^t\} = \exp\{(e^t-1)\lambda\}$, $-\infty<t<+\infty$

(b) If $X \sim Poisson(\lambda)$, $E[X]=\lambda$, $Var(X)=\lambda$

(c) If $X_1 \sim Poisson(\lambda_1)$, $X_2 \sim Poisson(\lambda_2)$ and X_1 and X_2 are independent

$$X_1+X_2 \sim Poisson(\lambda_1+\lambda_2)$$

(proof)

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

$$= \sum_{x=0}^{\infty}\frac{\exp\{-e^t\lambda+e^t\lambda-\lambda\}(e^t\lambda)^x}{x!}$$
$$= \exp\{e^t\lambda-\lambda\}\sum_{x=0}^{\infty}\frac{\exp\{-e^t\lambda\}(e^t\lambda)^x}{x!}$$
$$= \exp\{e^t\lambda-\lambda\}\sum_{x=0}^{\infty}\frac{e^{-\lambda^*}(\lambda^*)^x}{x!}, \quad \lambda^* = e^t\lambda$$
$$= \exp\{e^t\lambda-\lambda\}, \quad -\infty<t<+\infty$$

(b) $C_X(t) = \log M_X(t) = e^t\lambda-\lambda$
$C_X'(t) = e^t\lambda$, $C_X'(t=0) = E[X] = \lambda$
$C_X''(t) = e^t\lambda$, $C_X''(t=0) = Var(X) = \lambda$

(c) $M_{X_1+X_2}(t) = M_{X_1}(t)\cdot M_{X_2}(t) = \exp\{e^t\lambda_1-\lambda_1\}\exp\{e^t\lambda_2-\lambda_2\}$
$= \exp\{e^t(\lambda_1+\lambda_2)-(\lambda_1+\lambda_2)\}$
: mgf of $Poisson(\lambda_1+\lambda_2)$

by the uniqueness of mgf, we conclude that $X_1+X_2 \sim Poisson(\lambda_1+\lambda_2)$

[Textbook proof]

(a) $E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx}e^{-\lambda}\lambda^x/x = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x} = e^{-\lambda}\cdot e^{\lambda e^t} = \exp\{-\lambda+\lambda e^t\}$

(b) & (c) are identical to my proof.

〈 Definition : Poisson Process 〉

Let N_t denote the number of occurences of a phenomenon from time 0 to t. If the following conditions are satisfied, we say that $\{N_t : t\geq 0\}$ is a Poisson process with occurence rate (발생율) λ.

(a) (정상성, Stationarity) The distribution of the "number of occurences (of a phenomenon)" is independent of the starting time. i.e., distribution of N_t and $N_{s+t}-N_t$ are equivalent and $N_0 = 0$.

(b) (independent increment, 독립증분성) The number of events happening from time 0 to t is independent to the number of events happening from time t to $t+h$ $(h>0)$. i.e., N_t and $N_{t+h}-N_t$ are independent

(c) (proportionality) In a short span of time, the probability of an event occuring is directly proportional to time. Meaning,

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

here, λ is a strictly positive constant of proportionality (λ는 양수의 비례상수), and $o(h)$ means $\lim_{h\to 0} o(h)/h \to 0$
* h가 0 에 수렴하는 속도 보다 $o(h)$가 0 에 더 빨리 수렴한다는 의미

(d) (rareness, 희귀성) In a short span of time, the probability of two or more events occuring is extremely low. i.e.,

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

* $\lim_{h \to 0} o(h) = 0$

Let's think about N_t (the number of events from time 0 to t in a Poisson process). Divide the interval $(0,t]$ into smaller n intervals and let $N_{i \cdot t/n} - N_{(i-1) \cdot t/n}$ denote the number of events occured in the time interval $((i-1)\frac{t}{n}, i \cdot \frac{t}{n}]$. The Sum of events occured is N_t, meaning

$$N_t = \sum_{i=1}^{n} (N_{i \cdot t/n} - N_{(i-1) \cdot t/n}), \quad \forall n = 1, 2, \cdots$$

From conditions (a), (c) and (d),

$$P(N_{i \cdot t/n} - N_{(i-1) \cdot t/n} = 1) = \lambda \cdot \frac{t}{n} + o(\lambda \frac{t}{n}), \ n \to \infty$$

$$P(N_{i \cdot t/n} - N_{(i-1) \cdot t/n} \geq 2) = o(\lambda \frac{t}{n}) \quad , \ n \to \infty$$

Thus, if the number of intervals (n) is large, the distribution of $N_{i \cdot t/n} - N_{(i-1) \cdot t/n}$ is approximately a Bernoulli distribution with $p = \lambda \frac{t}{n}$ Also, from (b), $N_{i \cdot t/n} - N_{(i-1) \cdot t/n}$ $(i = 1, \cdots, n)$ are mutually independent. Hence, from the representational definition of the Binomial Distribution, the sum of $N_{i \cdot t/n} - N_{(i-1) \cdot t/n}$ $(i = 1, \cdots, n)$ follows the Binomial Distribution $B(n, \lambda \frac{t}{n})$

From the Poisson approximation of binomial probability, we can think that N_t follows a Poisson distribution with mean λt for large enough n.

Summary:

① N_t distribution
② $n \times (N_{i \cdot t/n} - N_{(i-1) \cdot t/n}$ distribution) ← divide $(0,t]$ into n small intervals
③ $n \times$ Bernoulli$(\lambda \frac{t}{n})$ ← use conditions of Poisson process (a), (c), (d)
④ Bin$(n, \lambda \frac{t}{n})$ distribution ← representational definition of binomial distribution
⑤ Poisson(λt) distribution ← poisson approximation of binomial probability

〈Theorem 3.4.2. Distribution of the number of events in a Poisson process〉

In a Poisson process $\{N_t : t \geq 0\}$ with occurence rate λ, the "number of events until time t" (N_t), follows a Poisson distribution with mean λt.

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

(proof)
(Continued on the right column)

$$P(N_{t+h} = x)$$

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

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

(b) independent increment, (a) stationarity

$$= P(N_t = x) P(N_h = 0) + P(N_t = x-1) P(N_h = 1) + P(N_t = x-2) P(N_h = 2) + \cdots$$

(d) rareness

$$= P(N_t = x) P(N_h = 0) + P(N_t = x-1) P(N_h = 1) + o(h) \quad \text{as } h \to 0$$

* the writer writes (b) → (a) → (d) instead of in the alphabetical order. why? Because of the order of the logic.

$\Rightarrow P(N_t = x, N_{t+h} - N_t = 0) = P(N_t = x) P(N_{t+h} - N_t)$ ⋯ (b) independent increment

$\Rightarrow P(N_t = x) P(N_{t+h} - N_t = 0) = P(N_t = x) P(N_h = 0)$ ⋯ (a) stationarity.

Let $g(x,t) = P(N_t = x)$

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

$$= 1 - \{\lambda h + o(h) + o(h)\}$$

- proportionality
- rareness

- $P(N_h = 1) = \lambda h + o(h)$ ∵ proportionality

$$g(x,t)[1 - \{\lambda h + o(h) + o(h)\}] + g(x-1, t)\{\lambda h + o(h)\} + o(h) \quad h \to 0$$

$$= g(x, t+h)$$

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

$$= \lim_{h \to 0} \frac{1}{h} \{ g(x,t)[1 - \{\lambda h + 2 \cdot o(h)\}] + g(x-1,t)\{\lambda h + o(h)\} + o(h) - g(x,t) \}$$

(1 and −g(x,t) cancel out)

$$= \lim_{h \to 0} \frac{1}{h} \{ g(x,t)[-\{\lambda h + 2 \cdot o(h)\}] + g(x-1,t)\{\lambda h + o(h)\} + o(h) \}$$

$$= \lim_{h \to 0} g(x,t)[-\{\lambda + 2 \cdot \frac{o(h)}{h}\}] + g(x-1,t)\{\lambda + \frac{o(h)}{h}\} + \frac{o(h)}{h}$$

$\to 0$ as $h \to 0$

$$= -\lambda g(x,t) + \lambda g(x-1,t)$$

$$\therefore \frac{\partial}{\partial t}\{e^{\lambda t} g(x,t)/\lambda^x\} = \lambda e^{\lambda t} g(x,t)/\lambda^x + e^{\lambda t}\left|\frac{\partial}{\partial t} g(x,t)\right|/\lambda^x$$

$$= \lambda e^{\lambda t} g(x,t)/\lambda^x + e^{\lambda t}(-\lambda g(x,t) + \lambda g(x-1,t))/\lambda^x$$

$$= \lambda e^{\lambda t} g(x,t)/\lambda^x - \lambda e^{\lambda t} g(x,t)/\lambda^x + \lambda e^{\lambda t} g(x-1,t)/\lambda^x$$

$$= e^{\lambda t} g(x-1,t)/\lambda^{x-1}$$

$$\frac{\partial^k}{\partial t^k}\{e^{\lambda t} g(x,t)/\lambda^x\} = e^{\lambda t} g(x-k,t)/\lambda^{x-k} \quad , (k \leq x)$$

Meanwhile, $N_0 = 0$, which means that

$\underbrace{g(0,0)=1}_{(1)}$, $\underbrace{g(1,0)=g(2,0)=\cdots=0}_{(2)}$ $\quad \begin{cases} g(0,0) = P(X=0, t=0), \\ g(1,0) = P(X=1, t=0) \end{cases}$

(1): when time is very short, probability of an event occuring is zero due to the "rareness" property of the Poisson process. In other words, the probability of no event occuring in a span of a very short time ($t=0$), is 1.

(2): Due to the "rareness" property, the probability of an event occuring when $t \simeq 0$, is zero, as mentioned above.

$$\frac{\partial}{\partial t} g(x,t)\Big|_{x=0} = -\lambda g(x=0,t) + \lambda g(x=-1,t)$$

Gemini 5.0 Solution. 미분방정식 활용

$g(x=-1,t)=0$ because it does not exist.

$g(x=0,t=0)=1 \quad \therefore \frac{\partial}{\partial t} g(x,t)\Big|_{\substack{x=0 \\ t=0}} = -\lambda g(x=0,t=0) = -\lambda$

The solution to $y' = -ky$ is an exponential decay:

$$g(x=0,t) = e^{-\lambda t}$$

(case: $x=1$) $\quad \frac{\partial}{\partial t} g(x,t)\Big|_{x=1} = -\lambda g(x=1,t) + \lambda g(0,t)$

$$\frac{\partial}{\partial t} g(x,t)\Big|_{x=1} = -\lambda g(x=1,t) + \lambda e^{-\lambda t}$$

$$g'(x=1,t) + \lambda g(x=1,t) = \lambda e^{-\lambda t}$$

* This is a linear first order ODE. We use the integration factor $e^{\int \lambda dt} = e^{\lambda t}$

$$\frac{\partial}{\partial t}\left[e^{\lambda t} g(x=1,t)\right] = e^{\lambda t}\left(\lambda e^{-\lambda t}\right) = \lambda$$

Integrate w.r.t t:

$$e^{\lambda t} g(x=1,t) = \lambda t + C$$

Since $N_0 = 0$, $P_1(0)=0$ which implies $C=0$. Thus:

$$P_1(t) = \lambda t e^{-\lambda t}$$

General Solution:

If we continue this process for $P_2(t)$, $P_3(t)$ etc, we find the pattern follows the terms of the Taylor expansion of $e^{\lambda t}$ weighted by $e^{-\lambda t}$. By induction, the general solution is:

$$g(x,t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

which is the pmf of a Poisson distribution with parameter $\mu = \lambda t$.

Ex. 3.4.2

단위 길이당 평균 결점수가 0.05로 알려진 전선을 생산하는 공정에서 나타나는 결점수의 모형으로서 포아송과정이 타당하다고 할 때, 100단위의 길이에 해당하는 전선에서 10개 이상의 결점이 나타날 확률을 구하여라.

$X \sim \text{Poisson}(100 \cdot 0.05) \overset{d}{=} \text{Poisson}(5)$

$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.968171\ldots = 0.031828\ldots \approx 0.032$

3.5. EXPONENTIAL AND GAMMA DISTRIBUTION

In a Poisson process $\{N_t : t \geq 0\}$ of occurence rate λ, if the first event occurs after t unit time, it means that no event has occured until t. If W_t denotes the time it took for the first event to occur, the following holds.

$$(W_t > 0) = (N_t = 0)$$

from this equality, if we calculate the cdf of W_1

$$P(W_1 > t) = P(N_t = 0) = e^{-\lambda t}, \quad t \geq 0$$

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

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

Thus, the pdf of W_1 is: $f(x) = \lambda e^{-\lambda x} I_{(x \geq 0)}$

We call this distribution the exponential distribution and denote it as $W_1 \sim \text{Exp}(1/\lambda)$ $(\lambda > 0)$

〈Theorem 3.5.1. Properties of Exponential Distribution〉

(a) If $W_1 \sim \text{Exp}(1/\lambda)$ $(\lambda > 0)$, the mgf is: $mgf_{W_1}(t) = (1 - t/\lambda)^{-1}$, $t < \lambda$

(b) If $W_1 \sim \text{Exp}(1/\lambda)$ $(\lambda > 0)$, $E[W_1] = 1/\lambda$, $\text{Var}(W_1) = 1/\lambda^2$

(proof)

(a) $E[e^{tx}] = \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x} dx = \int_0^\infty \lambda \exp\{-(\lambda - t)x\} dx$

$= \lambda(\lambda - t)^{-1} \int_0^\infty (\lambda - t) e^{-(\lambda - t)x} dx = (1 - t/\lambda)^{-1}$, $t < \lambda$

* If $X \sim \text{Exp}(\lambda)$, $M_X(t) = (1 - \lambda t)^{-1}$, $t < \lambda^{-1}$

(b) $C_X(t) = \log M_X(t) = -\log(1 - t/\lambda)$

$C_X'(t) = \lambda^{-1}/(1 - t/\lambda)$ $\quad$ $C_X''(t) = \lambda^{-2}(1 - t/\lambda)^{-2}$

$E[X] = C_X'(t=0) = \lambda^{-1}$ $\quad$ $C_X''(t=0) = \text{Var}(X) = \lambda^{-2}$

* If $X \sim \text{Exp}(\lambda)$, $E[X] = \lambda$ and $\text{Var}(X) = \lambda^2$

In a Poisson process $\{N_t : t \geq 0\}$ with occurence rate λ, if the rth event occurs after t, it means that $(r-1)$ or less events occured until t. Thus, if W_r denotes the time it takes for the rth event to occur, the following holds.

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

$$P(W_r \leq t) = 1 - P(W_r > t) = 1 - P(N_t \leq r-1) = 1 - \sum_{k=0}^{r-1} e^{-\lambda t}(\lambda t)^k / k!, \quad t \geq 0$$

derive both LHS and RHS w.r.t t to get the pdf of W_r

$$pdf_{W_r}(t) = \frac{\partial}{\partial t} cdf_{W_r}(t)$$

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

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

$$= \lambda e^{-\lambda t} \left\{ \sum_{k=0}^{r-1} (\lambda t)^k / k! - \sum_{k=0}^{r-2} (\lambda t)^k / k! \right\}$$

$$= \lambda e^{-\lambda t} (\lambda t)^{r-1} / (r-1)! = \lambda^r t^{r-1} e^{-\lambda t} / (r-1)!, \quad t > 0$$

* $t \neq 0$ because t^{r-1} may blow up and the pdf wouldn't be defined.

We call this distribution the gamma distribution and denote it as $W_r \sim \text{Gamma}(r, 1/\lambda)$

r is called the "shape parameter"(형상모수) and $\beta = 1/\lambda$ is called the "scale parameter".(척도모수).

The shape parameter may not be limited to a natural number. It can be a positive real number. In such case, we write the shape parameter as α.

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha - 1} e^{-x} \, dx, \quad \alpha > 0$$

We can use the precedent Gamma function[참] to express the Gamma(α, β) pdf.

$$f(x) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha - 1} e^{-x/\beta} I_{(x > 0)} \quad (\alpha > 0, \beta > 0)$$

* [참] 부록 I의 〈예 I 6.2〉에 소개된 감마함수 $\Gamma(\alpha)$ $(\alpha > 0)$의 성질은 다음과 같다.

(i) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, $\alpha > 1$. 특히 $\Gamma(n) = (n-1)!$, $n = 1, 2, \cdots$

(ii) $\Gamma(1/2) = \sqrt{\pi}$

$$\Gamma(1/2) = \int_0^\infty x^{1/2 - 1} e^{-x} dx, \quad \alpha > 0$$

$$= \int_0^\infty x^{-1/2} e^{-x} dx. \text{ Let } t = x^{1/2} \iff t^2 = x \quad dx = 2t \, dt$$

$$= \int_0^\infty t^{-1} e^{-t^2} \cdot 2t \, dt$$

$$= \int_0^\infty e^{-t^2} \cdot 2 \, dt = 2 \int_0^\infty e^{-t^2} dt = 2 \cdot \frac{\sqrt{\pi}}{2} = \sqrt{\pi}$$

* $\int_0^\infty e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$

〈Theorem 3.5.2 Properties of the Gamma Distribution〉

(a) If $X \sim \text{Gamma}(\alpha, \beta)$ $\quad E[X] = \alpha\beta$, $\text{Var}(X) = \alpha\beta^2$

(b) If $X \sim \text{Gamma}(\alpha, \beta)$, $M_X(t) = (1 - \beta t)^{-\alpha}$, $t < 1/\beta$

(c) If $X_1 \sim \text{Gamma}(\alpha_1, \beta)$, $X_2 \sim \text{Gamma}(\alpha_2, \beta)$ and X_1 and X_2 are independent $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$

(proof)

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

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

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

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

$$= \alpha(\alpha+1)\beta^2 \int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+1} e^{-x/\beta} dx$$

$$= \alpha(\alpha+1)\beta^2$$

$$E[X^2] - E[X]^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2$$
$$= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2$$
$$= \alpha\beta^2$$

$$\therefore \mathrm{Var}(X) = E[X^2] - E[X]^2 = \alpha\beta^2$$

(b) $$E[e^{tx}] = \int_0^\infty e^{tx} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$$

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

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

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta^*} dx, \quad \beta^* = (\beta^{-1}-t)^{-1} = \left(\frac{1}{\beta}-t\right)^{-1} = \frac{\beta}{1-\beta t} = \beta(1-\beta t)^{-1}$$

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

$$= \beta^{-\alpha}(\beta^*)^\alpha = \beta^{-\alpha} \cdot (\beta(1-t\beta)^{-1})^\alpha = \beta^{-\alpha}\beta^\alpha(1-\beta t)^{-\alpha} = (1-\beta t)^{-\alpha}, \quad t < 1/\beta$$

(c) $$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = (1-\beta t)^{-\alpha_1} \cdot (1-\beta t)^{-\alpha_2} = (1-\beta t)^{-(\alpha_1+\alpha_2)}, \quad t < 1/\beta$$

$(1-\beta t)^{-(\alpha_1+\alpha_2)}$ is the mgf of Gamma$(\alpha_1+\alpha_2, \beta)$

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

Exponential distribution is equivalent to a Gamma distribution with the shape parameter value $\alpha=1$. From the property of the Gamma distribution (Theorem 3.5.2 (c)), we can define the Gamma distribution as the following as well:

Definition: Representational Definition of the Gamma distribution with shape parameter $\alpha \in \mathbb{N}$.

If the shape parameter is $r \in \mathbb{N}$, Gamma$(r, \beta) \overset{d}{\equiv}$ Exp$_1(\beta) \oplus \cdots \oplus$ Exp$_r(\beta)$

i.e. $X \sim$ Gamma$(r, \beta) \Leftrightarrow X \overset{d}{\equiv} Z_1 + \cdots + Z_r$, $Z_i \overset{i.i.d}{\sim}$ Exp(β)

Meanwhile, we know that $W_1, W_2-W_1, W_3-W_2, \cdots, W_r-W_{r-1}$ denotes the time taken for another event to occur after an event in a Poisson process with occurence rate λ. and $W_1, W_2-W_1, W_3-W_2, \cdots, W_r-W_{r-1}$ are independent, follows the exponential distribution Exp$(1/\lambda)$, and their sum $W_r \sim$ Gamma$(r, 1/\lambda)$. i.e.,

$$W_r = W_1 \oplus W_2-W_1 \oplus \cdots \oplus (W_r - W_{r-1}) \sim \text{Gamma}(r, 1/\lambda)$$
$$W_1, \cdots, W_r - W_{r-1} \overset{i.i.d}{\sim} \text{Exp}(1/\lambda)$$

3.6. NORMAL DISTRIBUTION

The approximation of the cumulative probability of the binomial distribution is: [54]

$$\sum_{x: a \le \frac{x-np}{\sqrt{np(1-p)}} \le b} \binom{n}{x} p^x (1-p)^{n-x} \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \quad n \to \infty$$

here, the integral of the function $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $-\infty < z < \infty$, over $(-\infty, +\infty)$ is 1 and is called the pdf of the Standard normal distribution. Generally, the function

$$\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \quad -\infty < x < +\infty \ (\mu \in \mathbb{R}, \sigma > 0)$$

is called the pdf of a normal distribution and is denoted as $N(\mu, \sigma^2)$. The normal distribution is a bell-shaped curved symmetric about $x=\mu$ and σ shows the level of the distribution's dispersion. (종모양 정규)

〈Theorem 3.6.1. Properties of the Normal Distribution〉

(a) If $X \sim N(\mu, \sigma^2)$, $E[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$

(b) If $X \sim N(\mu, \sigma^2)$, $mgf_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$, $-\infty < t < +\infty$

(c) If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ and X_1 and X_2 are independent,

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

(proof)

$$E[X] = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx$$

Let $z = \frac{x-\mu}{\sigma}$, $dz = \frac{1}{\sigma}dx \Leftrightarrow dx = \sigma \cdot dz$

$$E[z] = \int_{-\infty}^{+\infty} (\sigma z + \mu) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}z^2\right\} \cdot \sigma \cdot dz$$
$$= \int_{-\infty}^{+\infty} (\sigma z + \mu) \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz$$
$$= \sigma \int_{-\infty}^{+\infty} z \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz + \mu$$
$$= \sigma \int_{-\infty}^{+\infty} z \cdot \phi(z) dz + \mu$$

let $z^2/2 = y \Leftrightarrow 2z \cdot \frac{1}{2} \cdot dz = dy \Leftrightarrow z\,dz = dy \Leftrightarrow dz = \frac{1}{z}dy$

$$= \sigma \int_0^{+\infty} z \cdot \phi(z) dz + \sigma \int_{-\infty}^0 z \cdot \phi(z) dz + \mu$$
$$= \sigma \int_0^\infty z \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}z^2\right\} dz + \sigma \int_{-\infty}^0 z \cdot \phi(z) dz + \mu$$
$$= \sigma \int_0^\infty z \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\{-y\} \cdot \frac{1}{z} \cdot dy + \sigma \int_{-\infty}^0 z \cdot \frac{1}{\sqrt{2\pi}} \exp\{-y\} \cdot \frac{1}{z} dy + \mu$$
$$= \sigma \int_0^\infty \frac{1}{\sqrt{2\pi}} \cdot \exp\{-y\} dy + \sigma \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp\{-y\} dy + \mu$$
$$= \sigma \int_0^\infty \frac{1}{\sqrt{2\pi}} \cdot \exp\{-y\} dy - \sigma \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\{-y\} dy + \mu = \mu$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx$$

Let $z = (x-\mu)/\sigma$, $dz = \frac{1}{\sigma}dx \Leftrightarrow \sigma\, dz = dx$

$$= \int_{-\infty}^{+\infty} (\sigma z+\mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \sigma\, dz$$

$$= \int_{-\infty}^{+\infty} (\sigma^2 z^2 + 2\sigma\cdot\mu\cdot z + \mu^2) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dz$$

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

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

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

$$Var(X) = E[X^2] - E[X]^2 = \sigma^2 \int_{-\infty}^{+\infty} z^2 \phi(z)\, dz$$

$$= 2\sigma^2 \int_0^{+\infty} z^2 \phi(z)\, dz = 2\sigma^2 \int_0^{+\infty} z^2 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz$$

Let $y = z^2/2$, $z = \sqrt{2y}$ $\quad dy = dz\cdot z$ $\quad \therefore dz = \frac{1}{z}dy$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{+\infty} z\cdot \exp\{-y\}\, dz = \frac{\sqrt{2}\sigma^2}{\sqrt{\pi}} \int_0^{+\infty} \sqrt{2y}\cdot \exp\{-y\}\, dy$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{+\infty} y^{1/2} e^{-y} dy = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma(3/2) = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(1+\tfrac{1}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2}\Gamma\left(\tfrac{1}{2}\right) = \frac{\sigma^2}{\sqrt{\pi}} \cdot \sqrt{\pi}$$

$$= \sigma^2 \quad \text{Q.E.D.}$$

$\Gamma(n+1) = n\Gamma(n)$

$\Gamma(n) = (n-1)! = (n-1)\cdot(n-2)! = (n-1)\Gamma(n-1)$

(b) $$M_X(t) = E[e^{tx}] = \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx$$

Let $z = (x-\mu)/\sigma \Leftrightarrow \sigma z + \mu = x$, $\quad dx = \sigma\, dz$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}z^2 + t(\sigma z+\mu)\right\} \cdot \sigma\, dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[z^2 - 2t\sigma z - 2\mu t\right]\right\} dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[z^2 - 2t\sigma z + t^2\sigma^2 - t^2\sigma^2 - 2\mu t\right]\right\} dz$$

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

$$= \exp\left\{\mu t + \frac{1}{2}t^2\sigma^2\right\} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z-t\sigma)^2\right\} dz}_{\text{pdf of } N(t\sigma,1)}$$

$$= \exp\left\{\mu t + \frac{1}{2}t^2\sigma^2\right\}$$

(c) $$M_{X_1+X_2}(t_1+t_2) = M_{X_1}(t_1)\cdot M_{X_2}(t_2) = \exp\left\{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right\} \exp\left\{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right\}$$

$$= \exp\left\{(\mu_1+\mu_2) + \frac{1}{2}(\sigma_1^2+\sigma_2^2)\right\}$$

$\because$ mgf of $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

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

〈Theorem 3.6.2 Representational Definition of Normal Distribution〉

(a) If $X \sim N(\mu, \sigma^2)$, for constants a and b,

$$aX+b \sim N(a\mu+b, a^2\sigma^2)$$

(b) $X \sim N(\mu,\sigma^2) \Leftrightarrow \frac{X-\mu}{\sigma} \sim N(0,1) \Leftrightarrow X \overset{d}{=} \sigma Z + \mu$, $Z \sim N(0,1)$

(proof)

$$M_{aX+b}(t) = E[\exp\{(aX+b)t\}] = \exp\{bt\}\, E[\exp\{X\cdot(at)\}]$$

$$= \exp(bt)\cdot \exp\left\{(at)\mu + \frac{1}{2}\sigma^2\cdot(at)^2\right\}$$

$$= \exp\left\{(a\mu+b)t + \frac{1}{2}\sigma^2 a^2 t^2\right\}$$

$\because$ mgf of $N(a\mu+b, a^2\sigma^2)$

$\therefore aX+b \sim N(a\mu+b, a^2\sigma^2)$

(b) let $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$

$$aX+b \sim N(a\mu+b, a^2\sigma^2) \overset{d}{=} N\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \left(\frac{1}{\sigma}\right)^2\cdot\sigma^2\right)$$

$$\overset{d}{=} N(0,1)$$

$\therefore aX+b = \frac{X-\mu}{\sigma} \sim N(0,1)$

$(\cdots)$

[Textbook comment] (b)는 (a)의 특별한 경우이다.
(b) is a special case of (a)

From the property of a normal distribution, we can find the cumulative probability using the standard normal distribution. Let

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

then the cumulative probability of $X \sim N(\mu,\sigma^2)$ is:

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Also, note that Φ is symmetric about 0.
Thus the cumulative probability of $x < 0$ can be calculated as follows: $\Phi(-z) = 1 - \Phi(z)$

Ex. 3.6.1

If $X \sim N(3,4)$ use the table from 부록 IV to calculate the following probabilities.

(a) $P(5 < X \le 7)$
(b) $P(1 < X \le 4)$
(c) $P(X > 0)$

(a) $P(5 < X \le 7) = P\left(\frac{5-3}{2} < \frac{X-3}{2} \le \frac{7-3}{2}\right)$
$= P(1 < Z \le 2) = \Phi(2) - \Phi(1) = 0.9772 - 0.8413$
$= 0.1359$

(b) $P(1 < X \le 4) = P\left(\frac{1-3}{2} < \frac{X-3}{2} \le \frac{4-3}{2}\right) = P(-1 < Z < 1/2)$
$= \Phi(1/2) - \Phi(-1) = \Phi(1/2) - (1 - \Phi(1))$
$= 0.6195 - (1 - 0.8413) = 0.5328$

(c) $P(X > 0) = P\left(\frac{X-3}{2} > \frac{0-3}{2}\right) = P(Z > -\frac{3}{2}) = 1 - P(Z < -\frac{3}{2})$
$= 1 - \Phi(-3/2) = 1 - (1 - \Phi(3/2) = \Phi(3/2)$

Ex. 3.6.2.

If $X \sim N(3,4)$ find $q_{0.95}$ and $q_{0.025}$ satisfying:

(a) $P(X \le q_{0.95}) = 0.95$ (b) $P(X \le q_{0.025}) = 0.025)$

(a) $P(X \le q_{0.95}) = P\left(\frac{X-3}{2} \le \frac{q_{0.95}-3}{2}\right) = \Phi((q_{0.95}-3)/2) = 0.95$

$\therefore (q_{0.95}-3)/2 = 1.645$ $\therefore q_{0.95} = 1.645 \cdot 2 + 3 = 3.29 + 3 = 6.29$ ✓✓

(b) $P(X \le q_{0.025}) = P\left(\frac{X-3}{2} \le \frac{q_{0.025}-3}{2}\right) = \Phi((q_{0.025}-3)/2) = 0.025$

$= \Phi((3 - q_{0.025})/2) = 0.975$

$\therefore (3 - q_{0.025})/2 = 1.96 \Leftrightarrow 3 - q_{0.025} = 3.92$

$\therefore q_{0.025} = -0.92$ ✓✓

〈Properties of Normal Distribution〉

(1) $N(\mu, \sigma^2) \overset{d}{\equiv} \sigma N(0,1) + \mu$

(2) $aN(\mu, \sigma^2) + b \overset{d}{\equiv} N(a\mu + b, a^2\sigma^2)$

(3) $N(\mu_1, \sigma_1^2) \oplus N(\mu_2, \sigma_2^2) \overset{d}{\equiv} N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Binomial (Σ of Bernoulli)

Poisson dist.

Negative Binomial (Σ of Geometric)

〈Definition of Various Distributions〉 p. 133, 134