4. Sample Distribution

4.1. Distribution of Statistics

Definition:

random sample : random variables that are independent and follow the same distribution.

population parameter : 모형의 설정에 사용되는 미지의 변수; numerical value that describes a characteristic of an entire population, representing a true, fixed value for the whole group.

Ex. 4.1.1 Examples of models in real life and their parameter/parameter space.

(a) Use case : Statistical research on the process defect rate.
Model : Bernoulli(p)
parameter : p (defect rate)
parameter space : $\Omega = \{p : 0 \le p \le 1\}$

(b) Use case : Research on water pollution level (# of planktons in unit volume of water)
Model : Poisson
parameter : λ (mean # of planktons)
parameter space : $\Omega = \{\lambda : \lambda \ge 0\}$

(c) Use case : model for distribution of error in a scientific research
Model : $N(\mu, \sigma^2)$
parameter : μ, σ^2 (mean, variance)
parameter space : $\Omega = \{(\mu, \sigma^2)^t : -\infty < \mu < +\infty, \sigma^2 > 0\}$

θ : parameter
Ω : parameter space
$f(x;\theta)$: pdf of the population distribution.
Statistic : function of observable random variables that are used to make assumption on unknown parameters.

Definition: Random sample and statistics.

Random samples $X_1, \cdots, X_n$ sampled from a population with pdf $f(x;\theta), \theta \in \Omega$:

$$X_1, X_2, \cdots, X_n \overset{i.i.d}{\sim} f(x;\theta),\ \theta \in \Omega$$

Statistic $u(X_1, \cdots, X_n)$: function of random samples. The value of a statistic is determined when the values of random samples are given.

Ex. 4.1.2.
Examples of statistics from Ex.4.1.1

(a) $\hat{p} = (X_1 + \cdots + X_n)/n$ $\quad$ $\hat{p}$: sample proportion

(b) $\bar{x} = \frac{1}{n}(X_1 + \cdots + X_n)$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $\bar{x}$: sample mean, σ^2 : sample variance.

(c) $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$: order statistics

$\underset{1 \le i \le n}{\text{med}}\, X_i = \begin{cases} X_{(m+1)} & \text{if } n = 2m+1 \\ (X_{(m)} + X_{(m+1)})/2 & \text{if } n = 2m \end{cases}$ $\quad$ $\underset{1 \le i \le n}{\text{med}}\, X_i$: sample median

the statistic $u(X_1, \cdots, X_n)$ is a random variable and has its own distribution that is determined by the population distribution. The distribution of $u(X_1, \cdots, X_n)$ is called the "sampling distribution".

Ex.4.1.3.

(a) Let $X_1, \cdots, X_n$ denote the random samples from a Bernoulli(p) distribution, where p denotes the defect rate.

The total number of defects, $X_1 + \cdots + X_n$, follows the binomial distribution $B(n,p)$. Thus, the sample proportion $\hat{p} = (X_1 + \cdots + X_n)/n$ has the distribution

$$P(X_1 + \cdots + X_n = k) = P(n\hat{p} = k) = P(\hat{p} = k/n)$$

$$\therefore P(\hat{p} = k/n) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \cdots, n$$

The mean and variance of $\hat{p}$ are $E[\hat{p}] = p$, $Var(\hat{p}) = p(1-p)/n$ respectively. Meaning, the distribution of $\hat{p}$ gets closer to the population parameter p as the sample size $n \to \infty$

(b) Let $X_1, \cdots, X_n$ denote the random samples from a Poisson(λ) distribution, where λ is the mean number of planktons in unit volume of water. From Theorem 3.4.1, $X_1 + \cdots + X_n \sim$ Poisson($n\lambda$). Thus, the distribution of sample mean $\bar{x}$ is :

$$P(X_1 + \cdots + X_n = k) = P(n\bar{x} = k) = P(\bar{x} = k/n) = e^{-n\lambda}(n\lambda)^k / k!,\ k = 0, 1, 2, \cdots$$

$E[\bar{x}] = \lambda$, $Var(\bar{x}) = \lambda/n$ $\quad$ let $X = X_1 + \cdots + X_n$. $E[X] = n\lambda$, $Var(X) = n\lambda$
$E[\bar{x}] = E[X/n] = n\lambda/n = \lambda$, $Var(\bar{x}) = Var(X/n) = \lambda/n$

Thus, the distribution of $\bar{x}$ gets closers to the population parameter λ as $n \to \infty$

Definition: Transformation of variables and pdfs.

(a) (discrete case) $P(Y = y) = P(u(x) = y) = \sum_{x : u(x) = y} P(X = x)$

$$\therefore pdf_Y(y) = \sum_{x : u(x) = y} pdf_X(x)$$

(b) (continuous case) $P(Y \in y \pm |\Delta y|) = P(u(x) \in y \pm |\Delta y|)$

$$\approx \sum_{x : u(x) = y} P(X \in x \pm |\Delta x|)$$

$$\therefore pdf_Y(y)|\Delta y| \approx \sum_{x : u(x) = y} pdf_X(x)|\Delta x|,\ \text{meaning}\ pdf_Y(y) \approx \sum_{x : u(x) = y} pdf_X(x)|\Delta y/\Delta x|^{-1}$$

Ex. 4.1.4

(a) Let $X_1 \sim Bin(n_1, p)$ and $X_2 \sim Bin(n_2, p)$; X_1 and X_2 are independent.
We can find the pdf of $Y = X_1 + X_2$ as follows:

$$pdf_Y(y) = \sum_{x:u(x)=y} P(x=x) = \sum_{x:\, x_1+x_2=y} \binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1} \binom{n_2}{x_2} p^{x_2}(1-p)^{n_2-x_2}$$

$$= \sum_{x_1=0}^{y} \binom{n_1}{x_1}\binom{n_2}{y-x_1} p^{x_1+x_2}(1-p)^{n_1+n_2-x_1-x_2}$$

$$= \sum_{x_1=0}^{y} \binom{n_1}{x_1}\binom{n_2}{y-x_1} p^{y}(1-p)^{n_1+n_2-y}$$

$$= \binom{n_1+n_2}{y} p^y (1-p)^{n_1+n_2-y} \quad : \text{pdf of } Bin(n_1+n_2, y)$$

(b) Let $X_1 \sim Poisson(\lambda_1)$ and $X_2 \sim Poisson(\lambda_2)$; X_1 and X_2 are independent.
If $Y = X_1 + X_2$, the distribution of Y is:

$$P(Y=y) = \sum_{x:u(x)=y} P(u(x)=y) = \sum_{x:\, x_1+x_2=y} \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!} = \sum_{x_1=0}^{y} \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{y-x_1}}{x_1!(y-x_1)!}$$

$$= e^{-(\lambda_1+\lambda_2)}\frac{1}{y!}\sum_{x_1=0}^{y}\frac{y!}{x_1!(y-x_1)!}\lambda_1^{x_1}\lambda_2^{y-x_1} = e^{-(\lambda_1+\lambda_2)}\frac{1}{y!}\sum_{x_1=0}^{y}\binom{y}{x_1}\lambda_1^{x_1}\lambda_2^{y-x_1}$$

$$= e^{-(\lambda_1+\lambda_2)}\frac{1}{y!}(\lambda_1+\lambda_2)^y = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^y}{y!}$$

: pdf $Poisson(\lambda_1+\lambda_2)$

(c) Let $X \sim N(0,1)$. The pdf of $Y = X^2$ can be calculated by finding the derivative of $cdf_Y(y)$.

$$P(Y<y) = P(X^2<y) = P(-\sqrt{y} < X < \sqrt{y}) = \Phi(\sqrt{y}) - (1-\Phi(\sqrt{y}))$$

$$= 2\Phi(\sqrt{y}) - 1$$

$$P(Y=y) = \frac{\partial}{\partial y}P(Y<y) = \frac{\partial}{\partial y}2\cdot\Phi(\sqrt{y}) = \frac{1}{2}y^{-1/2}\cdot 2\,\phi(\sqrt{y})$$

$$= y^{-1/2}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}y\right\}, \; y>0 \quad \checkmark\checkmark$$

[Textbook Proof]

$$\frac{d}{dy}cdf_Y(y) = \frac{d}{dy}\{P(x \le \sqrt{y}) - P(x \le -\sqrt{y})\}$$

$$= \frac{1}{2}y^{-1/2}\{pdf_X(\sqrt{y}) + pdf_X(-\sqrt{y})\}$$

$$\therefore pdf_Y(y) = \frac{d}{dy}cdf_Y(y) = \sum_{x:\,x^2=y} pdf_X(x)\left|\frac{dy}{dx}\right|^{-1} = \frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}, \; y>0$$

〈Theorem 4.1.1. Continuous random variables one-to-one transformation and pdf〉

Suppose that the following holds for K dimension random variable $X = (X_1, \cdots, X_K)^t$ and function $u = (u_1, \cdots, u_K)^t : \mathcal{X} \to \mathcal{Y}$

(a) $P(X \in \mathcal{X}) = 1$

(b) The vector value function (벡터값 함수) $u = (u_1, \cdots, u_K)^t : \mathcal{X} \to \mathcal{Y}$ has the domain $\mathcal{X}$, range $\mathcal{Y}$ and is one-to-one.

(c) The domain $\mathcal{X}$ of the function $u = (u_1, \cdots, u_K)^t$ is an open set in $\mathbb{R}^K$, function u has a derivative (u가 미분가능하며), the first derivate is a continuous function and has a non-zero Jacobian matrix (1차 편도함수가 연속함수로서 0이 아닌 야코비안 행렬식을 갖는다).

i.e.,

$$J_u(x) = \det\left(\frac{\partial}{\partial x_i}u_j(x)\right) \neq 0 \quad \forall x \in \mathcal{X}$$

Now, the pdf of $Y = u(X) = (Y_1, \cdots, Y_K)^t = (u_1(X), \cdots, u_K(X))^t$ is:

$$pdf_Y(y) = pdf_X(u^{-1}(y))\,|J_{u^{-1}}(y)|, \; y \in \mathcal{Y}$$

i.e., $$pdf_Y(y) = pdf_X(x)\left|\det\left(\frac{\partial y}{\partial x}\right)\right|^{-1}, \; y = u(x) \in \mathcal{Y}$$

(proof) From the integration by substitution in Appendix I, the following holds for an arbitrary set $B \subseteq \mathcal{Y}$. Thus, the pdf of $Y = u(X)$ is given as above.

$$P(Y \in B) = P(u(X) \in B)$$

$$= \int_{x \in u^{-1}(B)} pdf_X(x)\,dx$$

$$= \int_{y \in B} pdf_X(u^{-1}(y))\,|J_{u^{-1}}(y)|\,dy \quad (\because x = u^{-1}(y) \text{ 로 치환})$$

Ex. 4.1.5 pdf using location & scale parameter

If the continuous type random variable Z has the pdf $f(z)$, let

$$X = \sigma Z + \mu, \; \sigma > 0, \; \mu \in \mathbb{R}$$

$$pdf_X(x) = pdf_Z(z)\left|\frac{dx}{dz}\right|^{-1}, \; x = \sigma z + \mu$$

$$= \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$$

Here, we call μ and Sigma the location parameter and scale parameter respectively.

(a) Normal Distribution $N(\mu, \sigma^2)$, $N(\mu, \sigma^2) \overset{d}{=} \sigma N(0,1) + \mu$

$$pdf_X(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \; \phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}I_{(-\infty<z<+\infty)}$$

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

(b) Logistic Distribution $L(\mu, \sigma)$, $L(\mu, \sigma) \overset{d}{=} \sigma L(0,1) + \mu$

$$pdf_X(x) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right), \; f(z) = \frac{e^z}{(1+e^z)^2}I_{(-\infty<z<+\infty)}$$

$$X \sim L(\mu, \sigma) \Leftrightarrow X \overset{d}{=} \sigma Z + \mu, \; Z \sim L(0,1)$$

(c) Double Exponential Distribution $DE(\mu, \sigma)$, $DE(\mu, \sigma) = \sigma DE(0,1) + \mu$

$$pdf_X(x) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right), \; f(z) = \frac{1}{2}e^{-|z|}I_{(-\infty<z<+\infty)}$$

$$X \sim DE(\mu, \sigma) \Leftrightarrow X \overset{d}{=} \sigma Z + \mu, \; Z \sim DE(0,1)$$

(d) Cauchy Distribution $C(\mu, \sigma)$, $C(\mu, \sigma) \overset{d}{=} \sigma C(0,1) + \mu$

$$pdf_X(x) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right), \; f(z) = \frac{1}{\pi}\frac{1}{1+z^2}I_{(-\infty<z<+\infty)}$$

$$X \sim C(\mu, \sigma) \Leftrightarrow X \overset{d}{=} \sigma Z + \mu, \; Z \sim C(0,1)$$

(e) Exponential Distribution $Exp(\sigma)$, $Exp(\sigma) \overset{d}{\equiv} \sigma\, Exp(1)$

$$pdf_X(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right),\quad f(z) = e^{-z} I_{[0\le z<+\infty)}$$

$$X \sim Exp(\sigma) \Leftrightarrow X \overset{d}{\equiv} \sigma Z,\quad Z \sim Exp(1)$$

(f) Gamma Distribution $Gamma(\alpha,\beta)$, $Gamma(\alpha,\beta) \overset{d}{\equiv} \beta\, Gamma(\alpha,1)$

$$pdf_X(x) = \frac{1}{\beta} f\left(\frac{x}{\beta}\right),\quad f(z) = \frac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-z} I_{[0\le z<+\infty)}$$

$$X \sim Gamma(\alpha,\beta) \Leftrightarrow X \overset{d}{\equiv} \beta Z,\quad Z \sim Gamma(\alpha,1)$$

Ex. 4.1.6

Let $X_1 \sim Gamma(\alpha_1,\beta)$, $X_2 \sim Gamma(\alpha_2,\beta)$, X_1 and X_2 are independent.

If $Y_1 = \frac{X_1}{X_1+X_2}$, $Y_2 = X_1+X_2$ find the joint pdf and marginal pdf of Y_1 and Y_2

$$f_{X_1}(x_1) = \frac{1}{\Gamma(\alpha_1)\beta^{\alpha_1}} x_1^{\alpha_1-1} \exp\left[-x_1/\beta\right],\quad f_{X_2}(x_2) = \frac{1}{\Gamma(\alpha_2)\beta^{\alpha_2}} x^{\alpha_2-1} e^{-x_2/\beta}$$

$$f_{X_1,X_2}(x_1=x_1, x_2=x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-(x_1+x_2)/\beta}$$

$$Y_1 = Y_1 Y_2 \qquad \partial x_1/\partial y_1 = y_2 \qquad \partial x_1/\partial y_2 = y_1$$
$$X_2 = Y_2 - Y_1 Y_2 = Y_2(1-Y_1) \qquad \partial x_2/\partial y_1 = -y_2 \qquad \partial x_2/\partial y_2 = 1-y_1$$

$$J_{u^{-1}}(x) = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{vmatrix} = y_2(1-y_1) - (y_1 \cdot -y_2) = y_2 - y_1 y_2 + y_1 y_2 = y_2$$

$$\therefore f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} (y_1 y_2)^{\alpha_1-1} (y_2(1-y_1))^{\alpha_2-1} e^{-y_2/\beta} \cdot |J_{u^{-1}}|$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} y_2^{\alpha_1+\alpha_2-2} e^{-y_2/\beta} \cdot y_2$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} I_{(0,1)}(y_1) I_{(0,\infty)}(y_2)$$

$$= \left[\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} I_{(0,1)}(y_1)\right]\left[\frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} I_{(0,\infty)}(y_2)\right]$$

$f_{Y_1,Y_2}(y_1,y_2) = g_1(y_1)\, g_2(y_2)$ $\therefore$ Y_1 and Y_2 are independent.

$$f_{Y_1}(y_1) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} I_{(0,1)}(y_1),\quad f_{Y_2}(y_2) = \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} I_{(0,\infty)}(y_2)$$

[Textbook proof]

The joint pdf of X_1 and X_2 is:

$$pdf_{X_1,X_2}(x_1,x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-(x_1+x_2)/\beta} I_{(x_1>0,\, x_2>0)}$$

The domain $\mathcal{X} = \{(x_1,x_2)^t : pdf_{X_1,X_2}(x_1,x_2) > 0\} = \{(x_1,x_2)^t : x_1>0,\, x_2>0\}$

Check if the function $u(x) = (x_1/(x_1+x_2),\ x_1+x_2)^t$ satisfies the conditions in Theorem 4.1.1

Let $(y_1,y_2)^t$ denote the values of function u.

$$u: \begin{cases} y_1 = \frac{x_1}{x_1+x_2} \\ y_2 = x_1+x_2 \end{cases} \qquad u^{-1}: \begin{cases} x_1 = y_1 y_2 \\ x_2 = y_2(1-y_1) \end{cases}$$

$$\mathcal{Y} = \{(y_1,y_2)^t : y_1 y_2 > 0,\ y_2(1-y_1) > 0\} = \{(y_1,y_2)^t : 0 < y_1 < 1,\ y_2 > 0\}$$

From the precedent range of the function u, we know that the function $u: \mathcal{X} \to \mathcal{Y}$ is one-to-one, satisfies the conditions in Theorem 4.1.1 and the Jacobian matrix of its inverse function $u^{-1}(x)$ is given as:

$$J_{u^{-1}} = \det\begin{pmatrix} y_2 & -y_2 \\ y_1 & 1-y_1 \end{pmatrix} = y_2$$

Thus, from Theorem 4.1.1 the joint pdf of Y_1, Y_2 is:

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} I_{(0,1)}(y_1) I_{(0,\infty)}(y_2)$$

which is the product of two functions: $g_1(y_1)$ and $g_2(y_2)$.

Thus, Y_1 and Y_2 are independent, and their marginal pdfs are given as:

$$pdf_{Y_1}(y_1) = \int_{-\infty}^{+\infty} pdf_{Y_1,Y_2}(y_1,y_2)\, dy_2$$

$$= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} I_{(0,1)}(y_1)$$

$$pdf_{Y_2}(y_2) = \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} I_{(0,\infty)}(y_2)$$

The distribution of Y_1 in Ex. 4.1.6 is called the Beta distribution [56] and is denoted as $Beta(\alpha_1,\alpha_2)$.

Definition: Beta Distribution

$$Beta(\alpha_1,\alpha_2) \overset{d}{\equiv} Gamma(\alpha_1,\beta) / (Gamma(\alpha_1,\beta) \oplus Gamma(\alpha_2,\beta))$$

$$X \sim Beta(\alpha_1,\alpha_2) \Leftrightarrow X \overset{d}{\equiv} Z_1/(Z_1+Z_2),\quad Z_i \sim Gamma(\alpha_i,\beta)\ (i=1,2)$$
Z_i and Z_j are independent $i \ne j$

$$\Leftrightarrow pdf_X(x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1} I_{(0,1)}(x)$$

* 56) In Ex. 4.1.6 the integral of Y_1 over $(-\infty,+\infty)$ is 1. Thus, for $\alpha_1>0$ and $\alpha_2>0$,

$$\int_0^1 x^{\alpha_1-1}(1-x)^{\alpha_2-1} dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$$

A function whose definite integral is $\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$ is called a Beta function and is denoted as $B(\alpha_1,\alpha_2)$

Ex. 4.1.7.

If X_1 and X_2 are mutually independent and follows the distribution

$$f(x) = I_{(0,1)}(x)$$

find the pdf of $Y = (X_1 + X_2)/2$

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = I_{(0,1)}(x_1) \cdot I_{(0,1)}(x_2)$$

$$\begin{aligned} Y_1 &= (X_1+X_2)/2 & X_1 &= Y_1 + Y_2 \\ Y_2 &= (X_1-X_2)/2 & X_2 &= Y_1 - Y_2 \end{aligned} \qquad |J_{u^{-1}}| = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = |-1-1| = 2$$

$u(x): X \to Y$ is a one-to-one transformation.

$\mathcal{X}$: open set in $\mathbb{R}^2$

Thus, we can apply theorem 4.1.1

$$f_{X_1,X_2}(x_1,x_2) = I_{(0,1)}(x_1) I_{(0,1)}(x_2) = I_{(0,1)}^{(y_1+y_2)} I_{(0,1)}^{(y_1-y_2)} |J|$$

$$I_{(-y_1,1-y_1)}^{(y_2)} I_{(-y_1,1-y_1)}^{(-y_2)} = I_{(-y_1,1-y_1)}^{(y_2)} I_{(y_1-1,y_1)}^{(y_2)} = I\{-y_1 \vee (y_1-1) < y_2 < (1-y_1) \wedge y_1\}$$

if $0 \le y_1 \le 1/2$, $-y_1 \vee (y_1-1) = -y_1$ and $(1-y_1) \wedge y_1 = y_1$

if $1/2 < y_1 \le 1$, $-y_1 \vee (y_1-1) = y_1 - 1$ and $(1-y_1) \wedge y_1 = 1-y_1$

$\therefore$ if $0 < y_1 \le 1/2$, $f(y_1) = 2\int_{-y_1}^{y_1} 1 \cdot dy_2 = 2(y_1+y_1) = 4y_1$

if $1/2 < y_2 \le 1$, $f(y_1) = 2\int_{y_1-1}^{1-y_1} 1 \cdot dy_2 = 4 - 4y_1$

$\therefore f(y_1) = 4y_1 I_{[0,1/2]}(y_1) + (4-4y_1) I_{(1/2,1)}(y_1)$

[Textbook Solution]

$Y = (X_1+X_2)/2 \quad Z = (X_1-X_2)/2$

$$pdf_{Y,Z}(y,z) = I_{\{0<y+z<1,\ 0<y-z<1\}} |-2|$$

이로부터 Y의 주변확률밀도함수를 나타내기 위하여, 지표함수(indicator function)의 연립부등식을 Z에 대한 부등식으로 나타내면

$$-y < z < 1-y, \quad y-1 < z < y$$

즉 $a \vee b = \max(a,b)$, $a \wedge b = \min(a,b)$ 이라고 하면

$$(-y) \vee (y-1) < z < (1-y) \wedge y$$

그러므로 $y \le 1/2$ 인 경우와 $y > 1/2$ 인 경우로 나누어 Y의 주변확률밀도함수를 구하면 다음과 같다.

(i) $y \le 1/2$ 인 경우에, $(-y) \vee (y-1) = -y$, $(1-y) \wedge y = y$ 이고 $-y < y \Leftrightarrow y > 0$ 이므로

$$pdf_Y(y) = \int_{-\infty}^{+\infty} pdf_{Y,Z}(y,z)\,dz = \int_{-y}^{y} 2\,dz\, I_{(0,1/2]}(y) = 4y I_{(0,1/2]}(y)$$

$y > 1/2$ 인 경우에, $(-y) \vee (y-1) = y-1$, $(1-y) \wedge y = 1-y$ 이고 $y-1 < 1-y \Leftrightarrow y < 1$ 이므로

$$pdf_Y(y) = \int_{-\infty}^{+\infty} pdf_{Y,Z}(y,z)\,dz = \int_{y-1}^{1-y} 2\,dz\, I_{(1/2,1)}(y) = 4(1-y) I_{(1/2,1)}(y)$$

따라서 Y의 주변밀도함수를 하나의 식으로 나타내면 〈예 1.4.3〉에서와 같이

$$pdf_Y(y) = (2 - 4|y - 1/2|) I_{(0,1)}(y)$$

임을 알 수 있다.

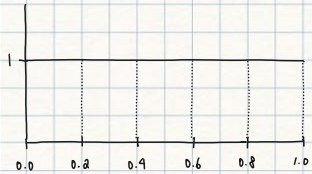
X_1, X_2 distributions are called uniform distribution and the distribution of $Y = (X_1+X_2)/2$ is called the triangular distribution.

Definition: Uniform Distribution.

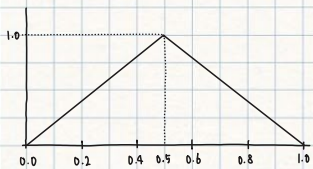
$$U(a,b) \overset{d}{=} (b-a)U(0,1) + a$$

$$X \sim U(a,b) \Leftrightarrow pdf_X(x) = \frac{1}{b-a} I_{(a,b)}(x)$$

$$\Leftrightarrow X \overset{d}{=} (b-a)Z + a,\ Z \sim U(0,1)$$



: uniform distribution



: triangular distribution

★ Ex 4.1.8 If X and Y are independent and follow the $N(0,1)$ distribution, what is the joint pdf of R and Θ which satisfies the polar coordinate conversion

$$\begin{cases} X = R\cos\Theta \\ Y = R\sin\Theta \end{cases}$$

$$\partial X/\partial R = \cos\Theta \qquad \partial X/\partial \Theta = -R\sin\Theta$$

$$\partial Y/\partial R = \sin\Theta \qquad \partial Y/\partial \Theta = R\cos\Theta$$

$(X,Y) \to (R,\Theta)$ is one-to-one

$(X,Y) \in \mathbb{R}^2$ and are open

$|J| = R\cos^2\Theta + R\sin^2\Theta = R \neq 0$

Thus we can apply Theorem 4.1.4.

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}x^2\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}y^2\right\} = \frac{1}{2\pi} \exp\left\{-\tfrac{1}{2}(x^2+y^2)\right\}$$

$$x^2 + y^2 = R^2\cos^2\Theta + R^2\sin^2\Theta = R^2$$

$$\therefore f_{R,\Theta}(r,\theta) = \frac{1}{2\pi} \exp\left\{-\tfrac{1}{2}r^2\right\} \cdot r\, I_{[0,+\infty)}^{(r)} I_{(0,2\pi]}^{(\theta)}$$

[Textbook Solution] $pdf_{R,\Theta}(r,\theta) = \frac{1}{2\pi} e^{-\frac{1}{2}r^2(\cos^2\theta+\sin^2\theta)} I_{[0,+\infty)}^{(r)} I_{[0,2\pi]}^{(\theta)} |r|$

$$= r e^{-\frac{1}{2}r^2} I_{[0,+\infty)}^{(r)} \frac{1}{2\pi} I_{[0,2\pi]}^{(\theta)}$$

* 정의역과 치역을 뚜렷히 정의하여 $u(x)$가 1-1 함수라는 것을 보여주었다.

즉 R, Θ는 서로 독립이고

$R^2/2 \sim Exp(1)$ $\quad \Theta \sim U[0, 2\pi)$

☆ Ex. 4.1.9.

Let the random variables $X_i\ (i=1,\cdots,K,K+1)$ be mutually independent and follow the Gamma(a_i, β) distribution.

Let $\quad Y_1 = \dfrac{X_1}{X_1+\cdots+X_{K+1}}, \cdots, Y_K = \dfrac{X_K}{X_1+\cdots+X_{K+1}}$

Find the joint pdf of $Y_1,\cdots,Y_K$

$$Y_1+\cdots+Y_K = \frac{X_1+\cdots+X_K}{X_1+\cdots+X_{K+1}}$$

$$1-(Y_1+\cdots+Y_K) = \frac{X_{K+1}}{X_1+\cdots+X_{K+1}}$$

$$(Y_1+\cdots+Y_K)^{-1} = \frac{X_1+\cdots+X_{K+1}}{X_1+\cdots+X_K} = 1+\frac{X_{K+1}}{X_1+\cdots+X_K}$$

Let $\quad y_{K+1} = X_1+\cdots+X_K$

$$u: \begin{cases} y_1 = x_1 \\ \vdots \\ y_K = x_K/(x_1+\cdots+x_{K+1}) \\ y_{K+1} = x_1+\cdots+x_{K+1} \end{cases} \qquad u^{-1}: \begin{cases} x_1 = y_1 y_{K+1} \\ \vdots \\ x_K = y_K y_{K+1} \\ x_{K+1} = y_{K+1}(1-y_1-\cdots-y_K) \end{cases}$$

$$\mathcal{Y} = \left\{ (y_1,\cdots,y_K,y_{K+1})^t : y_i>0\ (i=1,\cdots,K),\ y_1+\cdots+y_K<1,\ y_{K+1}>0 \right\}$$

이로부터 함수 u는 $\mathcal{X}$에서 $\mathcal{Y}$로의 일대일 함수로서 〈정리 4.1.1〉의 조건을 만족시키고 그 역함수 u^{-1}의 야코비안 행렬식이 다음과 같이 주어지는 것을 알 수 있다.

$$J_{u^{-1}} = \begin{pmatrix} y_{K+1} & \cdots & 0 & -y_{K+1} \\ \vdots & & \vdots & -y_{K+1} \\ 0 & \cdots & y_{K+1} & -y_{K+1} \\ y_1 & \cdots & y_K & (1-y_1-\cdots-y_K) \end{pmatrix} = y_{K+1}^K$$

column: fix $g_i(y)$ ↓ ← row: fix y_i * Textbook is wrong.

$$X_1 = g_1(Y) \\ X_2 = g_2(Y) \\ X_3 = g_3(Y) \qquad |J_{u^{-1}}| = \begin{vmatrix} \partial g_1(y)/\partial y_1 & \partial g_2(y)/\partial y_1 & \partial g_3(y)/\partial y_1 \\ \partial g_1(y)/\partial y_2 & \partial g_2(y)/\partial y_2 & \partial g_3(y)/\partial y_2 \\ \partial g_1(y)/\partial y_3 & \partial g_2(y)/\partial y_3 & \partial g_3(y)/\partial y_3 \end{vmatrix}$$

Theorem 4.3.2 *If $X \sim$ Poisson(θ) and $Y \sim$ Poisson(λ) and X and Y are independent, then $X+Y \sim$ Poisson$(\theta+\lambda)$.*

If (X, Y) is a continuous random vector with joint pdf $f_{X,Y}(x, y)$, then the joint pdf of (U, V) can be expressed in terms of $f_{X,Y}(x, y)$ in a manner analogous to (2.1.8). As before, $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$ and $\mathcal{B} = \{(u, v) : u = g_1(x, y) \text{ and } v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}$. The joint pdf $f_{U,V}(u, v)$ will be positive on the set $\mathcal{B}$. For the simplest version of this result we assume that the transformation $u = g_1(x, y)$ and $v = g_2(x, y)$ defines a one-to-one transformation of $\mathcal{A}$ onto $\mathcal{B}$. The transformation is onto because of the definition of $\mathcal{B}$. We are assuming that for each $(u, v) \in \mathcal{B}$ there is only one $(x, y) \in \mathcal{A}$ such that $(u, v) = (g_1(x, y), g_2(x, y))$. For such a one-to-one, onto transformation, we can solve the equations $u = g_1(x, y)$ and $v = g_2(x, y)$ for x and y in terms of u and v. We will denote this inverse transformation by $x = h_1(u, v)$ and $y = h_2(u, v)$. The role played by a derivative in the univariate case is now played by a quantity called the *Jacobian of the transformation*. This function of (u, v), denoted by J, is the determinant of a matrix of partial derivatives. It is defined by

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}.$$

따라서 $Y_1,\cdots,Y_K, Y_{K+1}$의 결합확률밀도함수는

The joint pdf of $Y_1,\cdots,Y_K, Y_{K+1}$ is:

$$f_{X_1,\cdots,X_{K+1}}(x_1,\cdots,x_{K+1}) = \frac{1}{\Gamma(a_1)\beta^{a_1}} x_1^{a_1-1} e^{-x_1/\beta} \cdots \frac{1}{\Gamma(a_{K+1})\beta^{a_{K+1}}} x_{K+1}^{a_{K+1}-1} e^{-x_{K+1}/\beta}$$

$$= \left\{ \prod_{i=1}^{K+1} \frac{1}{\Gamma(a_i)\beta^{a_i}} \right\} x \left\{ \prod_{i=1}^{K} (y_i y_{K+1})^{a_i-1} \right\} x \left\{ y_{K+1}(1-y_1-\cdots-y_K) \right\}^{a_{K+1}-1}$$

$$x\, e^{-y_{K+1}/\beta} \left| y_{K+1}^K \right|, \quad (y_1,\cdots,y_{K+1})^t \in \mathcal{Y}$$

$$\therefore \text{pdf}_{Y_1,\cdots,Y_K,Y_{K+1}} = \left\{ \prod_{i=1}^{K} \frac{1}{\Gamma(a_i)} y_i^{a_i-1} \right\} (1-y_1-\cdots-y_K)^{a_{K+1}-1} I_{(y_1>0,\cdots,y_K>0,\ y_1+\cdots+y_K<1)}$$

$$x \left\{ \frac{1}{\Gamma(a_{K+1})\beta^{a_1+\cdots+a_{K+1}}} y_{K+1}^{a_1+\cdots+a_{K+1}-1} e^{-y_{K+1}/\beta} I_{(y_{K+1}>0)} \right\}$$

그런데

$$\int_0^\infty \frac{1}{\beta^{a_1+\cdots+a_{K+1}}} y_{K+1}^{a_1+\cdots+a_{K+1}-1} e^{-y_{K+1}/\beta} dy_{K+1} = \Gamma(a_1+\cdots+a_{K+1})$$

이므로 $Y_1,\cdots,Y_K, Y_{K+1}$의 결합확률밀도함수를 y_{K+1}에 대하여 적분하면

$$\text{pdf}_{Y_1,\cdots,Y_K}(y_1,\cdots,y_K) = \int_{-\infty}^{+\infty} \text{pdf}_{Y_1,\cdots,Y_K,Y_{K+1}}(y_1,\cdots,y_K,y_{K+1})\, dy_{K+1}$$

$$= \frac{\Gamma(a_1+\cdots+a_{K+1})}{\Gamma(a_1)\cdots\Gamma(a_{K+1})} \left\{ \prod_{i=1}^{K} y_i^{a_i-1} \right\} (1-y_1-\cdots-y_K)^{a_{K+1}-1} I_{(y_1>0,\cdots,y_K>0,\ y_1+\cdots+y_K<1)}$$

The joint distribution of $Y_1,\cdots,Y_K$ is the generalized Beta distribution in upper dimensions, and is called the Dirichlet distribution.
The distribution is denoted as: $(Y_1,\cdots,Y_K)^t \sim \text{Dirichlet}(a_1,\cdots,a_K,a_{K+1})$

☆ Definition: Dirichlet Distribution

$$\text{Dirichlet}(a_1,\cdots,a_K,a_{K+1}) \overset{d}{=} \left(\frac{\text{Gamma}(a_i)}{\text{Gamma}(a_1,\beta)\oplus\cdots\oplus\text{Gamma}(a_{K+1},\beta)} \right)_{1\le i\le K}$$

$$Y = (Y_1,\cdots,Y_K)^t \sim \text{Dirichlet}(a_1,\cdots,a_K,a_{K+1})$$

$$\Leftrightarrow Y \overset{d}{=} \left(\frac{X_1}{X_1+\cdots+X_{K+1}}, \cdots, \frac{X_K}{X_1+\cdots+X_{K+1}} \right), \quad X_i \sim \text{Gamma}(a_i,\beta)$$
& X_i and X_j are independent $(i\neq j)$

$$\Leftrightarrow \text{pdf}_{Y_1,\cdots,Y_K}(y_1,\cdots,y_K)$$

$$= \frac{\Gamma(a_1+\cdots+a_{K+1})}{\Gamma(a_1)\cdots\Gamma(a_{K+1})} \left\{ \prod_{i=1}^{K} y_i^{a_i-1} \right\} (1-y_1-\cdots-y_K)^{a_{K+1}-1} I_{(y_1>0,\cdots,y_K>0,\ y_1+\cdots+y_K<1)}$$

* $X \sim \text{Dirichlet}(\underbrace{a_1,\cdots,a_{K+1}}_{K+1\text{개}})$, $X \in \mathbb{R}^K$ (K r.v.s)

〈Theorem 4.1.2. Many-to-one transformation of Continuous type random variables and their pdfs〉

Suppose that the following are true for continuous type random variable $X = (X_1,\cdots,X_K)^t$ and function $u = (u_1,\cdots,u_K)^t : \mathcal{X} \to \mathcal{Y}$

(a) $P(X \in \mathcal{X}) = 1$

(b) The vector value function $u = (u_1,\cdots,u_K)^t$, whose domain is $\mathcal{X}$ and range is $\mathcal{Y}$, is an m-to-1 function.

(c) The domain $\mathcal{X}$ of function $u = (u_1,\cdots,u_K)^t$ can be expressed into sum of non-overlapping open sets $\mathcal{X}_1,\cdots,\mathcal{X}_m$. For the function u whose domain has been restricted to $\mathcal{X}_r\ (r=1,\cdots,m)$, namely,

$$u^r(x) = u(x), \quad x \in \mathcal{X}_r\ (r=1,2,\cdots,m)$$

is a one-to-one function from $\mathcal{X}$ to $\mathcal{Y}$, has a derivative and its first derivative is continuous and has a non-zero Jacobian matrix. Meaning, let $u^r(x) = (u_1^r(x),\cdots,u_K^r(x))^t$, then

$$J_{u^r}(x) = \det\left\{ \frac{\partial}{\partial x_j} u_j^r(x) \right\} \neq 0 \quad \forall x \in \mathcal{X}_r$$

under these conditions, the K dimension random variable $Y = u(X)$, $Y = (Y_1, \cdots, Y_K)^t = (u_1(X), \cdots, u_K(X))^t$ has the pdf:

$$pdf_Y(y) = \sum_{r=1}^{m} pdf_X(|(u^r)^{-1}(y)|)\, |J_{(u^r)^{-1}}(y)|, \quad y \in \mathcal{Y}$$

i.e., $$\sum_{x:u(x)=y} pdf_X(x) \left| \det\left(\frac{\partial u(x)}{\partial x} \right) \right|^{-1}, \quad y \in \mathcal{Y}$$

(proof)

〈정리 4.1.1〉의 증명에서와 같이 분할된 각각의 정의역 $\mathcal{X}_r$ $(r = 1, \cdots, m)$에서의 일대일 함수 $y = u^r(x)$ $(r = 1, \cdots, m)$의 역함수 $x = (u^r)^{-1}(y)$를 이용한 치환적분법을 적용하면 임의의 부분집합 B에 대하여 다음이 성립함을 알 수 있다.

$$P(Y \in B) = P(u(X) \in B) = \sum_{r=1}^{m} P(u^r(X) \in B, X \in \mathcal{X}_r)$$

$$= \sum_{r=1}^{m} \int_{x \in (u^r)^{-1}(B)} pdf_X(x)\, dx$$

$$= \int_{y \in B} \sum_{r=1}^{m} pdf_X(|(u^r)^{-1}(y)|)\, |J_{(u^r)^{-1}}(y)|\, dy \quad (\because x = (u^r)^{-1}(y))$$

$$\therefore pdf_Y(y) = \sum_{r=1}^{m} pdf_X(|(u^r)^{-1}(y)|)\, |J_{(u^r)^{-1}}(y)|, \quad y \in \mathcal{Y}$$

따라서 $$pdf_Y(y) = \sum_{x:u(x)=y} pdf_X(x) \left| \det\left(\frac{\partial u(x)}{\partial x} \right) \right|^{-1}, \quad y \in \mathcal{Y}$$

Ex. 4.1.10

Let $X \sim U(-1, 1)$. Find the pdf of $Y = X^2$

$u(x) = x^2$ $\quad pdf_Y(y) = \sum_{x:u(x)=y} pdf_X(x)\, |J_{u^{-1}}|$

$$f_X(x) = \frac{1}{2} I_{[-1,1]}(x), \quad |J| = \left| \frac{dx}{dy} \right| = \frac{1}{2\sqrt{y}}$$

$$\therefore pdf_Y(y) = \sum_{x:u(x)=y} \frac{1}{2} I_{[0,1)}(y) \cdot \frac{1}{2\sqrt{y}}$$

$$= 2 \cdot \frac{1}{2} I_{[0,1)}(x) \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2\sqrt{y}} I_{[0,1)}(y)$$

$$\int_0^1 \frac{1}{2\sqrt{y}}\, dy = \left[y^{1/2} \right]_0^1 = 1 - 0 = 1$$

$$\therefore pdf_Y(y) = \frac{1}{2\sqrt{y}} I_{[0,1)}(y) \quad \checkmark$$

★★ Ex. 4.1.11

If the joint pdf of $X = (X_1, X_2)^t$ is: $pdf_X(x) = \frac{1}{\pi} I_{(0 < x_1^2 + x_2^2 < 1)}$

Find the joint pdf of $Y = (Y_1, Y_2)^t$

$$Y_1 = X_1^2 + X_2^2, \quad Y_2 = \frac{X_1^2}{X_1^2 + X_2^2}$$

$$X_1^2 = Y_1 Y_2 \qquad X_2^2 = Y_1 - Y_1 Y_2 = Y_1(1 - Y_2)$$

$$X_1 = \pm (Y_1 Y_2)^{1/2} \qquad X_2 = \pm \left\{ Y_1(1 - Y_2) \right\}^{1/2}$$

$$\therefore \begin{cases} x_1 = \pm \sqrt{y_1 y_2} \\ x_2 = \pm \sqrt{y_1(1 - y_2)} \end{cases} \quad \therefore \text{The function is 4-to-1}$$

$$Y_1 = X_1^2 + X_2^2, \quad Y_2 = \frac{X_1^2}{X_1^2 + X_2^2}$$

$$\frac{\partial y_1}{\partial x_1} = 2x_1 \qquad \frac{\partial y_2}{\partial x_1} = 2x_1 / (x_1^2 + x_2^2)^2 - 2x_1^3 / (x_1^2 + x_2^2)^2$$

$$\frac{\partial y_1}{\partial x_2} = 2x_2 \qquad \frac{\partial y_2}{\partial x_2} = -2x_2 x_1^2 / (x_1^2 + x_2^2)^2$$

$$\frac{2x_1 x_2^2}{(x_1^2 + x_2^2)^2} \quad \frac{\{2x_1(x_1^2 + x_2^2) - 2x_1^3\}}{(x_1^2 + x_2^2)^2}$$

$$|J| = \left| \frac{dy}{dx} \right|^{-1} = \begin{vmatrix} 2x_1 & 2x_2 \\ 2x_1 x_2^2/(x_1^2 + x_2^2)^2 & -2x_1 x_2^2/(x_1^2 + x_2^2)^2 \end{vmatrix}^{-1}$$

$$= \left| \left(-4x_1^3 x_2 - 4x_1 x_2^3 \right) / (x_1^2 + x_2^2)^2 \right|$$

$$= 4 \left| x_1 x_2 (x_1^2 + x_2^2) / (x_1^2 + x_2^2)^2 \right|$$

$$= 4 \left| x_1 x_2 / (x_1^2 + x_2^2) \right|$$

$$= 4\sqrt{y_2(1 - y_2)} \quad \cdots (*)$$

$$x_1^2 = Y_1 Y_2, \quad x_2^2 = Y_1 - Y_1 Y_2 = Y_1(1 - Y_2)$$

$$x_1 x_2 = (Y_1 Y_2 \cdot Y_1(1 - Y_2))^{1/2} = Y_1 \sqrt{Y_2(1 - Y_2)}$$

$$\therefore x_1 x_2 / (x_1^2 + x_2^2) = Y_1 \sqrt{Y_2(1 - Y_2)} / Y_1 = \sqrt{Y_2(1 - Y_2)}$$

$$pdf_X(x_1, x_2) = \frac{1}{\pi} I_{(0 < x_1^2 + x_2^2 < 1)}$$

$$\mathcal{X} := \left\{ (x_1, x_2)^t : 0 < x_1^2 + x_2^2 < 1,\ x_1 \neq 0,\ x_2 \neq 0 \right\}$$

$$\mathcal{Y} := \left\{ (y_1, y_2)^t : 0 < y_1 < 1,\ 0 < y_2 < 1 \right\}$$

$$\Rightarrow pdf_Y(y_1, y_2) = \sum_{\substack{x_1, x_2 : \\ x_1^2 + x_2^2 = y_1}} \pi^{-1} I_{(0 < x_1^2 + x_2^2 < 1)}$$

$$x_1, x_2 : x_1^2 / (x_1^2 + x_2^2) = y_2$$

$$= \frac{4}{\pi} \left| \frac{\partial y}{\partial x} \right|^{-1} = \frac{4}{\pi} \cdot \frac{1}{4\sqrt{y_2(1 - y_2)}} I_{(0,1)}(y_1) \cdot I_{(0,1)}(y_2)$$

$$= I_{(0,1)}(y_1) \cdot \frac{1}{\pi} y_2^{-1/2} (1 - y_2)^{-1/2} I_{(0,1)}(y_2)$$

$$= g_1(y_1) \cdot g_2(y_2)$$

$\therefore Y_1$ and Y_2 are independent.

*notice that $\Gamma(1/2) = \sqrt{\pi}$ $\quad \therefore \frac{1}{\pi} y_2^{-1/2} (1 - y_2)^{-1/2} = \frac{\Gamma(1/2 + 1/2)}{\Gamma(1/2)\Gamma(1/2)} y_2^{-1/2} (1 - y_2)^{-1/2}$

$\sim Beta(1/2, 1/2)$

$\therefore Y_1 \sim U(0, 1)$ and $Y_2 \sim Beta(1/2, 1/2)$ □

〈Random Variable Transformation key Takeaways〉

1. define domain space $\mathcal{X}$ and range space $\mathcal{Y}$ first. This lets you know if function $u(x)$ is m-to-1.
2. ★ Find the Jacobian matrix and its determinant. $|J_{u^{-1}}| = \left| \frac{\partial y}{\partial x} \right|^{-1} = \left| \frac{\partial x}{\partial y} \right|$ (w.r.t y or x ← very important. In this exercise you had to find $|J|$ w.r.t y)
3. Express $f_X(x) = \sum_{x:u(x)=y} f_X(y)$ and proceed.

* Try deriving the pdf of the Dirichlet distribution.

★ : finding Jacobian w.r.t y first then change $x \to y$ and any ± signs are canceled out by the absolute operation(?) ! !

⟨4.2. TYPICAL SAMPLE DISTRIBUTIONS⟩

Ex. 4.2.1.

Let $X_1, \cdots, X_r \sim N(0,1)$. Find the distribution of $Y = X_1^2 + \cdots + X_r^2$.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$$

$$Y = X^2 \iff X = \pm\sqrt{Y}$$

$\therefore u(x) = x^2$ is a 2-to-1 function.

$$|J_{u^{-1}}| = \left|\frac{dY}{dX}\right|^{-1} = |2x|^{-1} = |2\sqrt{Y}|^{-1} = \frac{1}{2\sqrt{y}}$$

$$f_X(x) = \sum_{x:\, x^2 = Y} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y\right\} \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{\sqrt{2}\sqrt{\pi}} \exp\left\{-\frac{1}{2}y\right\} \cdot y^{-1/2}$$

$$= \frac{1}{\Gamma(1/2)\, 2^{1/2}} y^{1/2 - 1} \exp\left\{-y/2\right\} \sim \text{Gamma}(1/2, 2)$$

$X_1^2, \cdots, X_r^2 \sim \text{Gamma}(1/2, 2)$ and by Gamma distribution's property,

$$X_1^2 + \cdots + X_r^2 \sim \text{Gamma}\left(\frac{r}{2}, 2\right)$$

The distribution in Ex. 4.2.1, $\text{Gamma}(r/2, 1)$, is called the chi-square distribution with degrees of freedom r and is denoted as: $Y \sim \chi^2(r)$

⟨Definition: Chi-square Distribution⟩

$$\chi^2(r) \overset{d}{\equiv} N_1(0,1)^2 \oplus \cdots \oplus N_r(0,1)^2 \overset{d}{\equiv} \text{Gamma}(r/2, 2)$$

$$Y \sim \chi^2(r) \ (r > 0) \iff Y \sim \text{Gamma}(r/2, 2)$$

$$\iff \text{pdf}_Y(y) = \frac{1}{\Gamma(r/2)\, 2^{r/2}} y^{r/2 - 1} e^{-y/2} I_{(0,\infty)}(y)$$

$$\iff Y \overset{d}{\equiv} X_1^2 + \cdots + X_r^2, \ X_i \overset{i.i.d}{\sim} N(0,1) \ \ (i = 1, \cdots, r) \ \ r \in \mathbb{N}$$

The value $\chi^2_\alpha(r)$ satisfying $P(Y > \chi^2_\alpha(r)) = \alpha$ is called the "upper α quartile of χ^2 distribution with degrees of freedom r."

$$X \sim \text{Gamma}(v, \beta) \iff X/\beta \sim \text{Gamma}(v, 1) \iff 2X/\beta \sim \text{Gamma}(v, 2) = \chi^2(2v)$$

Thus, you can find the upper quartile number of a $\text{Gamma}(v, \beta)$ distribution using the χ^2 upper quartile number. Meaning:

$$P(X > \beta\chi^2_\alpha(2v)/2) = P(2X/\beta > \chi^2_\alpha(2v)) = \alpha$$

Hence, the upper α quartile number of a $\text{Gamma}(v, \beta)$ distribution is $\beta \cdot \chi^2_\alpha(2v)/2$.

Ex. 4.2.3

If two random variables Z, V are mutually independent and $Z \sim N(0,1)$, $V \sim \chi^2(r)$ find the pdf of:

$$X = \frac{Z}{\sqrt{V/r}}$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}$$

$$V \sim \chi^2(r) \overset{d}{\equiv} \text{Gamma}\left(\frac{r}{2}, 2\right) \quad f_V(v) = \frac{1}{\Gamma(r/2)\, 2^{r/2}} v^{r/2 - 1} \exp\left\{-v \cdot 2/r\right\}$$

$$f_{Z,V}(z,v) = f_Z(z) \cdot f_V(v) = \left(\sqrt{2\pi} \cdot \Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}\right)^{-1} v^{r/2 - 1} \exp\left\{-\frac{1}{2}z^2 - \frac{2v}{r}\right\}$$

Let $Y = V$

$$Z = X \cdot \sqrt{Y/r} \qquad X = Z/\sqrt{V/r}$$

$$V = Y \qquad Y = V$$

$$\mathcal{X} := \{(z,v) : -\infty < z < +\infty, \ 0 < v < +\infty\}$$

$$\mathcal{Y} := \{(x,y) : -\infty < x < +\infty, \ 0 < v < +\infty\}$$

$\therefore u(z,v) \to (x,y)$ is a one-to-one function.

$$\partial z/\partial x = (y/r)^{1/2}$$

$$\partial z/\partial y = \partial/\partial y \, x \cdot (y/r)^{1/2} = 1/2 \, y \cdot 1/r \cdot (y/r)^{-1/2} = y/2r \cdot (y/r)^{-1/2} = \frac{y}{2r} \cdot \frac{r^{1/2}}{y^{1/2}} = \left(\frac{y}{2r}\right)^{1/2}$$

$$\partial v/\partial x = 0, \quad \partial v/\partial y = 1$$

$$\therefore |J| = \left|\det\left(\frac{\partial u(x)}{\partial x}\right)\right|^{-1} = \begin{vmatrix} \partial z/\partial x & \partial v/\partial x \\ \partial z/\partial y & \partial v/\partial y \end{vmatrix}^{-1} = \begin{vmatrix} (y/r)^{1/2} & 0 \\ (y/2r)^{1/2} & 1 \end{vmatrix} = (y/r)^{1/2}$$

$$\text{pdf}_{X,Y}(x,y) = \text{pdf}_{Z,V}(z,v) \left|\det\left(\frac{\partial(z,v)}{\partial(x,y)}\right)\right|$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \frac{1}{\Gamma(r/2) \cdot 2^{r/2}} v^{r/2 - 1} e^{-v/2} I_{(v>0)} \sqrt{\frac{y}{r}}$$

$$= \frac{1}{\Gamma(1/2)\Gamma(r/2)\, 2^{(r+1)/2} \sqrt{r}} y^{(r+1)/2 - 1} e^{-(1 + x^2/r)y/2} I_{(y>0)}$$

$$\text{pdf}_X(x) = \int_0^{+\infty} \frac{1}{\Gamma(1/2)\Gamma(r/2)\, 2^{(r+1)/2} \sqrt{r}} y^{(r+1)/2 - 1} e^{-(1 + x^2/r)y/2} \, dy$$

Let $t = (1 + x^2/r)y/2$, $y = 2t(1 + x^2/r)^{-1}$ $\qquad dt/dy = \frac{(1 + x^2/r)}{2}$

$$= \int_0^{\infty} \frac{1}{\Gamma(1/2)\Gamma(r/2)\, 2^{(r+1)/2} \sqrt{r}} \left\{2t(1 + x^2/r)^{-1}\right\}^{(r+1)/2 - 1} e^{-t} \left\{2(1 + x^2/r)^{-1}\right\} dt$$

$$= \int_0^{\infty} \frac{1}{\Gamma(1/2)\Gamma(r/2)\, 2^{(r+1)/2} \sqrt{r}} \left\{2t(1 + x^2/r)^{-1}\right\}^{(r+1)/2 - 1} e^{-t} \, dt$$

$$= \frac{1}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}} (1 + x^2/r)^{-(r+1)/2} \int_0^{\infty} t^{(r+1)/2 - 1} e^{-t} \, dt$$

$$= \frac{1}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}} (1 + x^2/r)^{-(r+1)/2} \Gamma\left(\frac{r+1}{2}\right)$$

$$= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}} (1 + x^2/r)^{-(r+1)/2}$$

The distribution in Ex.4.2.3 is called the t-distribution with degrees of freedom r and is denoted as $X \sim t(r)$

Definition: t-distribution [57]

$t(r) \overset{d}{\equiv} N(0,1) / \sqrt{\chi^2(r)/r}$, $N(0,1) \perp \chi^2(r)$

$X \sim t(r)\ (r>0) \Leftrightarrow X \overset{d}{\equiv} \dfrac{Z}{\sqrt{V/r}}$, $Z \sim N(0,1)$, $V \sim \chi^2(r)$, Z and V are independent.

$\Leftrightarrow pdf_X(x) = \dfrac{\Gamma((r+1)/2)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}} (1+x^2/r)^{-(r+1)/2}$

The distribution of t is symmetric about 0 and similar to the standard normal distribution, when $X \sim t(r)$, we call the value $t_\alpha(r)$ the "upper α quartile number of t-distribution with degrees of freedom r", if $t_\alpha(r)$ satisfies the following:

$$P(X > t_\alpha(r)) = \alpha \quad (0<\alpha<1)$$

*57) $X \perp Y$: X and Y are independent.

〈Theorem 4.3.3. Fundamental Theorems of sample distributions in which the population is Normal〉

Let $X_1, \cdots, X_n$ denote the random samples from the normal distribution $N(\mu, \sigma^2)$. The following holds true:

(a) Sample mean $\bar{X} = (X_1 + \cdots + X_n)/n \sim N(\mu, \sigma^2/n)$

(b) Sample variance $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$ and $\bar{X}$ are independent

(c) $(n-1)S^2/\sigma^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2 \sim \chi^2(n-1)$

(proof)

(a) $M_{(X_1+\cdots+X_n)/n}(t) = E[e^{tX_1/n}] \cdots E[e^{tX_n/n}] = M_X(t/n)^n = \exp\{\mu(t/n) + \frac{1}{2}(\frac{t}{n})^2\sigma^2\}^n$
$= \exp\{(\mu \cdot \frac{t}{n} + \frac{1}{2}\frac{t^2}{n^2}\cdot\sigma^2)n\} = \exp\{\mu t + \frac{1}{2}t^2(\sigma^2/n)\}$
: MGF of $N(\mu, \sigma^2/n)$

$\therefore \bar{X} = (X_1 + \cdots + X_n)/n \sim N(\mu, \sigma^2/n)$

[Textbook]: We can repeatedly apply the properties of the normal distribution shown in 〈Thm. 3.6.1〉 and 〈3.6.2〉

〈Theorem 3.6.1. Properties of the Normal Distribution〉

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

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

(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,

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

〈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) $X \sim N(\mu, \sigma^2) \Leftrightarrow \dfrac{X-\mu}{\sigma} \sim N(0,1) \Leftrightarrow X \overset{d}{\equiv} \sigma Z + \mu$, $Z \sim N(0,1)$

(proof continued)

(b) $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$ and $\bar{X}$ are independent.

① $E[S^2 \cdot \bar{X}] = E[S^2] \cdot E[\bar{X}]$

② $M_{S^2,\bar{X}}(t) = M_{S^2,\bar{X}}(t_1, t_2 = 0) \cdot M_{S^2,\bar{X}}(t_1 = 0, t_2)$

③ $pdf_{S^2,\bar{X}}(s^2, \bar{x}) = g_1(s^2)\, g_2(\bar{x})$

S^2 is a function of $(X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$. Hence, from theorem 2.5.10, if we show that $\bar{X}$ and $(X_1 - \bar{X}, \cdots, X_n - \bar{X})$ are independent, $\bar{X}$ and S^2 are independent. The joint mgf of $\bar{X}$ and $Y = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$ is:

$$mgf_{\bar{X},Y}(s, t_1, \cdots, t_n) = E[e^{s\bar{X} + t_1(X_1 - \bar{X}) + \cdots + t_n(X_n - \bar{X})}]$$

$= E[\exp\{s\bar{X} + t_1(X_1 - \bar{X}) + \cdots + t_n(X_n - \bar{X})\}]$

$= E[\exp\{s\frac{1}{n}(X_1 + \cdots + X_n) + t_1X_1 + \cdots + t_nX_n - (t_1 + \cdots + t_n)\bar{X}\}]$

$= E[\exp\{s\cdot\frac{1}{n}(X_1 + \cdots + X_n) + t_1X_1 + \cdots + t_nX_n - \bar{t}(X_1 + \cdots + X_n)\}]$

$= E[\exp\{X_1(s/n + t_1 - \bar{t}) + X_2(s/n + t_2 - \bar{t}) + \cdots + X_n(s/n - t_n - \bar{t})\}]$

$= E[\exp\{X_1(s/n + t_1 - \bar{t})\}] \times \cdots \times E[\exp\{X_n(s/n - t_n - \bar{t})\}]$

$= \prod_{i=1}^{n} \exp\{\mu(s/n + t_i - \bar{t}) + \frac{1}{2}\sigma^2(s/n + t_i - \bar{t})^2\}$

$= \exp\{\sum_{i=1}^{n} \mu(\frac{s}{n} + t_i - \bar{t}) + \frac{1}{2}\sigma^2(\frac{s}{n} + t_i - \bar{t})^2\}$

$= \exp\{\sum_{i=1}^{n} \mu(\frac{s}{n} + t_i - \bar{t}) + \frac{1}{2}\sigma^2(\frac{s^2}{n^2} + 2\cdot\frac{s}{n}(t_i - \bar{t}) + (t_i - \bar{t})^2)\}$

(Cancel out) (Cancel out)

$= \exp\{\mu s + \frac{1}{2}\sigma^2\} \exp\{\frac{1}{2}\sigma^2 \sum_{i=1}^{n} (t_i - \bar{t})^2\}$

$\therefore \bar{X}$ and $Y = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$ are independent

Thus, from Theorem 2.5.9., $\bar{X}$ and $Y = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$ are independent.

" 이로부터 $\bar{X}$ 와 $Y = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$ 의 결합적률생성함수는 각각의 적률적률생성함수의 곱인 s의 함수와 $t = (t_1, \cdots, t_n)^t$ 의 함수의 곱으로 나타내어진다.
따라서 〈정리 2.5.9〉로부터 $\bar{X}$와 $Y = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^t$는 서로 독립이다.

(c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1) \overset{d}{\equiv} Gamma(\frac{2}{2}(n-1), 2)$

$(n-1)S^2/\sigma^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2$

$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \{(X_i - \bar{X}) + (\bar{X} - \mu)\}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$

$X \sim N(\mu, \sigma^2) \therefore \dfrac{X-\mu}{\sigma} \sim N(0,1) \Rightarrow \left(\dfrac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1) \Rightarrow \sum_{i=1}^{n} \left(\dfrac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$

$\Rightarrow \underbrace{\frac{1}{\sigma^2}\Sigma(X_i - \mu)^2}_{\chi^2(n)} = \frac{1}{\sigma^2}\Sigma(X_i - \bar{X})^2 + \underbrace{\frac{1}{\sigma^2}\Sigma(\bar{X} - \mu)^2}_{\chi^2(1)}$

$\therefore \frac{1}{\sigma^2}\Sigma(X_i - \bar{X})^2 \sim \chi^2(n-1)$

[Textbook proof]

Let $U=\sum_{i=1}^{n}(X_i-\mu)^2/\sigma^2$, $V=\sum_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2=(n-1)S^2/\sigma^2$, $W=n(\bar{X}-\mu)^2/\sigma^2$

$U=V+W$ and from (b), V and W are independent.

Hence, $mgf_U(t)=mgf_V(t)\,mgf_W(t)$

Meanwhile, $(X_i-\mu)/\sigma$ $(i=1,2,\cdots,n)$ are mutually independent and follow the standard normal distribution $N(0,1)$.

Thus,

$$U=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n),\quad \therefore\ mgf_U(t)=(1-2t)^{-n/2}I_{(t<1/2)}$$

Also from (a), $(\bar{X}-\mu)/(\sigma/\sqrt{n})\sim N(0,1)$

$$W=\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2\sim\chi^2(1),\ \text{i.e.,}\ mgf_W(t)=(1-2t)^{-1/2}I_{(t<1/2)}$$

$$\therefore\ mgf_V(t)=mgf_U(t)/mgf_W(t)=(1-2t)^{-(n-1)/2}$$

$$\Updownarrow$$

$$V=(n-1)S^2/\sigma^2\sim\chi^2(n-1)\quad\square$$

〈Theorem 4.2.3. Estimation of population mean in a normal population〉

Let random samples $X_1,\cdots,X_n\sim N(\mu,\sigma^2)$. The following is true:

$$(\bar{X}-\mu)/(S/\sqrt{n})\sim t(n-1)$$

$$P\left\{\bar{X}-t_{\alpha/2}(n-1)S/\sqrt{n}\le\mu\le\bar{X}+t_{\alpha/2}(n-1)S/\sqrt{n}\right\}=1-\alpha$$

(proof)

$\bar{X}\sim N(\mu,\sigma^2/n)\Rightarrow\bar{X}-\mu=N(0,\sigma^2)\Rightarrow\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)\sim N(0,1)$ let $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

from 〈Theorem 4.2.2〉-(c), $(n-1)S^2/\sigma^2\sim\chi^2(n-1)$. let $V=(n-1)S^2/\sigma^2$

$$\frac{Z}{\sqrt{V/(n-1)}}=\underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{Z}\cdot\sqrt{\frac{\sigma^2}{(n-1)S^2}}(n-1)=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)\quad\square$$

$$P\left\{\bar{X}-t_{\alpha/2}(n-1)S/\sqrt{n}\le\mu\le\bar{X}+t_{\alpha/2}(n-1)S/\sqrt{n}\right\}$$

$$=P\left\{-t_{\alpha/2}(n-1)S/\sqrt{n}\le\mu-\bar{X}\le t_{\alpha/2}(n-1)S/\sqrt{n}\right\}$$

$$=P\left\{-t_{\alpha/2}(n-1)\le\frac{\mu-\bar{X}}{S/\sqrt{n}}\le t_{\alpha/2}(n-1)\right\}=P\left\{-t_{\alpha/2}(n-1)\le\frac{\bar{X}-\mu}{S/\sqrt{n}}\le t_{\alpha/2}(n-1)\right\}$$

$$=P\left\{X^*\le t_{\alpha/2}(n-1)\right\}-P\left\{X^*\le-t_{\alpha/2}(n-1)\right\},\quad X^*\sim t(n-1)$$

$$=P\left\{X^*\le t_{\alpha/2}(n-1)\right\}-\left[1-P\left\{X^*\le t_{\alpha/2}(n-1)\right\}\right]$$

$$=2P\left\{X^*\le t_{\alpha/2}(n-1)\right\}-1$$

$$=2(1-\alpha/2)-1$$

$$=1-\alpha\quad\square$$

[Textbook proof]

Let $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $V=(n-1)S^2/\sigma^2$. From Theorem 4.2.2,

$Z\sim N(0,1)$, $V\sim\chi^2(n-1)$ and Z,V are independent.

Thus, $\frac{\bar{X}-\mu}{S/\sqrt{n}}=\frac{Z}{\sqrt{V/(n-1)}}\sim t(n-1)$

and by the definition of upper quartile number,

$$P\left\{\left|\frac{\bar{X}-\mu}{S/\sqrt{n}}\right|\le t_{\alpha/2}(n-1)\right\}=1-\alpha$$

The interval for μ given in theorem 4.2.3 is called the "Confidence interval of μ with confidence level $(1-\alpha)$"

Confidence level indicates that $100\cdot(1-\alpha)\%$ of confidence intervals generated from random samples will contain the true population parameter μ.

〈Theorem 4.2.4. Estimation of population variance in a normal population〉

If $X_1,\cdots,X_n\sim N(\mu,\sigma^2)$, the following is true:

$$P\left\{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}S^2\le\sigma^2\le\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}S^2\right\}=1-\alpha$$

(proof) $(n-1)S^2/\sigma^2\sim\chi^2(n-1)$

$$\therefore P\left\{\chi^2_{1-\alpha/2}(n-1)\le\frac{(n-1)S^2}{\sigma^2}\le\chi^2_{\alpha/2}(n-1)\right\}=1-\alpha$$

$$=P\left\{\chi^2_{1-\alpha/2}(n-1)\frac{1}{(n-1)S^2}\le\frac{1}{\sigma^2}\le\chi^2_{\alpha/2}(n-1)\frac{1}{(n-1)S^2}\right\}$$

$$=P\left\{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\ge\sigma^2\ge\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}\right\}=1-\alpha$$

Ex. 4.2.4.

Random variables V_1,V_2 are independent and $V_1\sim\chi^2(r_1)$, $V_2\sim\chi^2(r_2)$

Find the pdf of $X=\frac{V_1/r_1}{V_2/r_2}$

(proof)

let $\begin{cases}V_1=r_1XY/r_2\\ V_2=Y\end{cases}$ $\quad X=\frac{V_1/r_1}{V_2/r_2}$, $Y=V_2$

$$\mathcal{X}:=\{(v_1,v_2):0<v_1<+\infty,\ 0<v_2<+\infty\}$$

$$\mathcal{Y}:=\{(x,y):0<x<+\infty,\ 0<y<+\infty\}$$

$$u(v_1,v_2)=(x,y)\Leftrightarrow u^{-1}(x,y)=(v_1,v_2)$$

(a) $u(\cdot)$ is a one to one function

(b) $\mathcal{X},\mathcal{Y}$ are open intervals in $\mathbb{R}^2$

(c) $|J|=\left|\det\left(\frac{\partial(v_1,v_2)}{\partial(x,y)}\right)\right|=\begin{vmatrix}r_1y/r_2 & 0\\ r_1x/r_2 & 1\end{vmatrix}=r_1y/r_2\neq0$

$$f_{V_1,V_2}(v_1,v_2) = \frac{1}{\Gamma(\frac{r_1}{2})2^{r_1/2}} v_1^{\frac{r_1}{2}-1}\exp\left\{-\frac{v_1}{2}\right\}\frac{1}{\Gamma(\frac{r_2}{2})2^{r_2/2}} v_2^{\frac{r_2}{2}-1}\exp\left\{-\frac{v_2}{2}\right\}$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})2^{(r_1+r_2)/2}} v_1^{\frac{r_1}{2}-1} v_2^{\frac{r_2}{2}-1}\exp\left\{-\frac{v_1}{2}-\frac{v_2}{2}\right\}$$

$$\begin{cases} v_1 = r_1 X Y / r_2 & X = \dfrac{v_1/r_1}{v_2/r_2} \\ v_2 = Y & Y = v_2 \end{cases}$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})2^{(r_1+r_2)/2}} \left(r_1 x y / r_2\right)^{\frac{r_1}{2}-1} y^{\frac{r_2}{2}-1}\exp\left\{-\frac{1}{2}\left(\frac{r_1 x}{r_2}+y\right)\right\}\frac{r_1 y}{r_2}$$ ✓✓

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1} y^{\frac{(r_1+r_2)}{2}-1}\exp\left\{-\frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)y\right\}$$

Let $t = \frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)y$, $dt = \frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)dy \Rightarrow dy = \left[\frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)\right]^{-1}dt$

$$f_X(x) = \int_0^\infty f_{X,Y}(x,y)\,dy$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}\int_0^\infty \left[\frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)\right]^{-\frac{(r_1+r_2)}{2}+1} t^{\frac{(r_1+r_2)}{2}-1}\exp\{-t\}\left[\frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)^{-1}\right]dt$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}\left[\frac{1}{2}\left(\frac{r_1}{r_2}x+1\right)\right]^{-\frac{r_1+r_2}{2}}\int_0^\infty t^{\frac{r_1+r_2}{2}-1}\exp\{-t\}\,dt$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}\left[\frac{r_1}{r_2}x+1\right]^{-\frac{r_1+r_2}{2}}\Gamma\left(\frac{r_1+r_2}{2}\right)$$

$$= \frac{\Gamma(\frac{r_1+r_2}{2})}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}\left[\frac{r_1}{r_2}x+1\right]^{-\frac{r_1+r_2}{2}} I_{(x>0)}\ \square$$ ✓✓

Definition: F-distribution

$$F(r_1,r_2) \overset{d}{\equiv} \frac{\chi^2(r_1)/r_1}{\chi^2(r_2)/r_2}, \quad \chi^2(r_1) \perp \chi^2(r_2)$$

$$X \sim F(r_1,r_2)\ (r_i>0,\ i=1,2)$$

$$\Leftrightarrow X \overset{d}{\equiv} \frac{V_1/r_1}{V_2/r_2},\ V_i \sim \chi^2(r_i)\ (i=1,2),\ V_1 \text{ and } V_2 \text{ are independent.}$$

$$\Leftrightarrow pdf\ f_X(x) = \frac{\Gamma((r_1+r_2)/2)}{\Gamma(r_1/2)\Gamma(r_2/2)}\left(\frac{r_1}{r_2}\right)^{r_1/2} x^{r_1/2-1}\left(1+\frac{r_1}{r_2}x\right)^{-(r_1+r_2)/2} I_{(x>0)}$$

Similar to the χ^2_α, the value $F_\alpha(r_1,r_2)$ is called the upper α quartile number if it satisfies the following:

$$P(X > F_\alpha(r_1,r_2)) = \alpha \quad (0<\alpha<1)$$

〈Theorem 4.2.5. Properties of the F-distribution〉

(a) If $X \sim F(r_1,r_2)$, $\frac{1}{X} \sim F(r_2,r_1)$. Hence,

$$F_{1-\alpha}(r_1,r_2) = 1/F_\alpha(r_2,r_1)$$

(b) If $X \sim t(r)$, $X^2 \sim F(1,r)$. Hence,

$$t^2_{\alpha/2}(r) = F_\alpha(1,r)$$

(proof) If $X \sim F(r_1,r_2)$, $X \overset{d}{\equiv} \frac{V_1/r_1}{V_2/r_2}$, $V_1 \sim \chi^2(r_1)$, $V_2 \sim \chi^2(r_2)$

$\therefore \frac{1}{X} \equiv \frac{V_2/r_2}{V_1/r_1} \sim F(r_2,r_1)$ $\therefore \frac{1}{X} \sim F(r_2,r_1)$

$P(F_{1-\alpha}(r_1,r_2) < X) = 1-\alpha$

$\Rightarrow P(F_{1-\alpha}(r_1,r_2) > X) = \alpha$

$P(F_\alpha(r_2,r_1) < X^{-1}) = \alpha$

$\Rightarrow P(F_\alpha(r_2,r_1)^{-1} > X) = \alpha$

$\therefore P(F_{1-\alpha}(r_1,r_2) > x) = P(F_\alpha(r_2,r_1)^{-1} > x) = \alpha$

$\therefore F_{1-\alpha}(r_1,r_2) = F_\alpha(r_2,r_1)^{-1}$ ✓✓

[Textbook proof] $1-\alpha = P\{X \geq F_{1-\alpha}(r_1,r_2)\}$, $\alpha = P\{1/X \geq F_\alpha(r_2,r_1)\}$

$\therefore \alpha = P\{X \leq F_{1-\alpha}(r_1,r_2)\} = P\{X \leq 1/F_\alpha(r_2,r_1)\}$

(b) $X \sim t(r) \Leftrightarrow X \overset{d}{\equiv} \frac{Z}{\sqrt{V/r}}$, $Z \sim N(0,1)$, $V \sim \chi^2(r)$, $Z \perp V$.

$$X^2 = \left(\frac{Z}{\sqrt{V/r}}\right)^2 = \frac{Z^2}{V/r} = \frac{Z^2/1}{V/r},\quad Z^2 \sim \chi^2(1),\ V \sim \chi^2(r) \text{ with } r_1=1 \text{ and } r_2=r$$

$\therefore X^2 \sim F(r_1=1, r_2=r) = F(1,r)$

$$P(X > t_{\alpha/2}) = \alpha/2 \Rightarrow P(X^2 > t^2_{\alpha/2}) = P(\pm X > t_{\alpha/2})$$
$$= P(X > t_{\alpha/2}) + P(-X > t_{\alpha/2})$$
$$= P(X > t_{\alpha/2}) + P(X < -t_{\alpha/2})$$
$$= \alpha/2 + \alpha/2$$
$$= \alpha$$

$X^2 \sim F(1,r)$ $\therefore P(X^2 > F_\alpha(1,r)) = \alpha$

$\therefore t^2_{\alpha/2} = F_\alpha(1,r)$

[Textbook proof] $\alpha = P\{X^2 \geq F_\alpha(1,r)\}$, $\alpha = P\{|X| \geq t_{\alpha/2}(r)\} = P\{X^2 \geq t_{\alpha/2}(r)^2\}$

Ex. 4.2.5.

From Appendix Ⅳ, $F_{0.975}(5,10) = 1/F_{0.025}(10,5) = 1/6.625 = 0.151$

$t^2_{0.025}(5) = 2.571^2 = 6.610$, $F_{0.05}(1,5) = 6.610$

〈Theorem 4.2.6. Comparison of population variance in two normal distributions〉

Let $X_{11}, X_{12}, \cdots, X_{1n_1} \sim N(\mu_1, \sigma_1^2)$ and $X_{21}, X_{22}, \cdots, X_{2n_2} \sim N(\mu_2, \sigma_2^2)$.

If the two random samples are mutually exclusive, and

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i,\quad S_i^2 = \sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2/(n_i-1) \quad (i=1,2)$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

$$P\left\{\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{\alpha/2}(n_2-1,n_1-1)\right\} = 1-\alpha$$

(proof)

$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n)$ and $(n_i-1)S_i^2/\sigma_i^2 \sim \chi^2(n_i-1)$

Let $V = (n_1-1)S_1^2/\sigma_1^2$, $W = (n_2-1)S_2^2/\sigma_2^2$

$V \sim \chi^2(n_1-1)$, $W \sim \chi^2(n_2-1)$

$$\frac{V/(n_1-1)}{W/(n_2-1)} \sim F(n_1-1, n_2-1)$$

$$\frac{V/(n_1-1)}{W/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad \therefore \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

$$P\left\{ F_{1-\alpha/2}(n_1-1, n_2-1) \le \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \le F_{\alpha/2}(n_1-1, n_2-1) \right\} = 1-\alpha$$

$$= P\left\{ \frac{S_2^2}{S_1^2} F_{1-\alpha/2}(n_1-1, n_2-1) \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2} F_{\alpha/2}(n_1-1, n_2-1) \right\} = 1-\alpha$$

$$= P\left\{ \frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2}(n_1-1, n_2-1)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)} \right\} = 1-\alpha$$

$$= P\left\{ \frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2}(n_1-1, n_2-1)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} F_{\alpha/2}(n_2-1, n_1-1) \right\} = 1-\alpha$$

D

When we compare multiple populations, we often assume/suppose that the variances of each population are equal and compare the mean of each population. The model that is most often used in the precedent test is the one way classification model (이원배치 분산분석). It means that we observe i.i.d random samples $X_{i1}, \cdots, X_{in_i}$ from $N(\mu_i, \sigma^2)$ distributions. ($X_{i1}, \cdots, X_{in} \sim N(\mu_i, \sigma^2)$)

Definition: One-way Classification Model

$$\begin{cases} X_{ij} = \mu_i + e_{ij} \\ e_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2), \ i = 1, 2, \cdots, K; \ j = 1, \cdots, n_i \end{cases}$$

〈Theorem 4.2.7 Comparison of population means from multiple populations.〉

In the one-way classification model : $\begin{cases} X_{ij} = \mu_i + e_{ij} \\ e_{ij} \sim N(0, \sigma^2), \ i = 1, \cdots, K, \ j = 1, \cdots, n_i \end{cases}$

Let $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$, $\bar{X} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} X_{ij}/n$, $n = \sum_{i=1}^{K} n_i$, $\bar{\mu} = \sum_{i=1}^{K} n_i\mu_i/n$

Then the following are true :

(a) $\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\sigma^2 \sim \chi^2(K-1)$

(b) $\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/\sigma^2 \sim \chi^2(n-K)$

(c) $\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2$ and $\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$ are independent.

Hence, if $\hat{\sigma}^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/(n-K)$,

$$\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\hat{\sigma}^2 \sim F(K-1, n-K)$$

$$P\left\{ \sum_{i=1}^{K} n_i(\mu_i - \bar{\mu} - (\bar{X}_i - \bar{X}))^2 \le (K-1)\hat{\sigma}^2 F_\alpha(K-1, n-K) \right\} = 1-\alpha.$$

(a) $\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\sigma^2 \sim \chi^2(K-1)$

(b) $\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/\sigma^2 \sim \chi^2(n-K)$

(c) $\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2$ and $\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$ are independent.

(proof)

$\bar{X}_i \sim N(\mu_i, \sigma^2/n)$, $\bar{X} = N(\bar{\mu}, \sigma^2/n)$

$S_i^2 = \frac{1}{(n_i-1)}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$, $(n_i-1)S_i^2/\sigma^2 \sim \chi^2(n_i-1)$, $\bar{X}_i \perp S_i^2$

(c) $\sum_{i=1}^{n} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2 = g_1(\bar{X}_1, \cdots, \bar{X}_K)$

$\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = g_2(S_1^2, \cdots, S_K^2)$

$S_i^2 \perp \bar{X}_j \quad \forall\, i,j \in [1, 2, \cdots, K]$

$\therefore g_1(\bar{X}_1, \cdots, \bar{X}_K)$ and $g_2(S_1^2, \cdots, S_K^2)$ are independent also. ✓✓

(b) $\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/\sigma^2 = (n_1-1)S_1^2/\sigma^2 \oplus \cdots + (n_K-1)S_K^2/\sigma^2$

$= V_1 \oplus \cdots \oplus V_K$

$V_1 \sim \chi^2(n_1-1), \cdots, V_K \sim \chi^2(n_K-1)$

$\therefore V_1 \oplus \cdots \oplus V_K \sim \chi^2(n_1 + \cdots + n_K - K) = \chi^2(n-K)$ ✓✓

(a) proof for the general case is in Ch.4.4.

Here, we will prove (a) for $n_1 = n_2 = \cdots = n_K = \bar{n}$

$\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\sigma^2 \sim \chi^2(K-1)$

If $n_1 = \cdots = n_K = \bar{n}$

$\Rightarrow \bar{n}\sum_{i=1}^{K} (\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\sigma^2$

$= \bar{n}\sum_{i=1}^{K} (\bar{X}_i - \mu_i - (\bar{X} - \bar{\mu}))^2/\sigma^2 = \sum_{i=1}^{K} (Z_i - \bar{Z})^2 \sim \chi^2(K-1)$

$Z_i = (\bar{X}_i - \mu_i)/(\sigma/\sqrt{n_i}) = \sqrt{\bar{n}}\,(\bar{X}_i - \mu_i)/\sigma \quad (i = 1, 2, \cdots, K)$

$Z_i \sim N(0,1)$

Lastly, $\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\hat{\sigma}^2$

$= \frac{\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/(K-1)\sigma^2}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/(n-K)\sigma^2}$

$\therefore$ from (a), (b), (c) and the definition of the F distribution,

$$\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\hat{\sigma}^2 \sim F(K-1, n-K)$$

Thus the following is true :

$$P\left\{ \frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2/\hat{\sigma}^2 \le F_\alpha(K-1, n-K) \right\} = 1-\alpha$$

If we change the inequality into an equation inside an ellipsoid (타원체 내부의 방정식), we get the following:

$$P\left\{ \sum_{i=1}^{k} n_i [\mu_i - \bar{\mu} - (\bar{X}_i - \bar{X})]^2 \le (k-1)\hat{\sigma}^2 F_\alpha(k-1, n-k) \right\} = 1-\alpha$$

★ The set of $\mu_i - \bar{\mu}$ $(i=1,2,\cdots,k)$ given in theorem 4.2.7 is called the cofidence set with confidence level $100(1-\alpha)\%$. Confidence set is the generalized confidence interval. For better understanding, consider the case $k=2$.

$$\begin{cases} \mu_1 - \bar{\mu} = \dfrac{n_2}{n_1+n_2}(\mu_1 - \mu_2), & \mu_2 - \bar{\mu} = \dfrac{n_1}{n_1+n_2}(\mu_2 - \mu_1) \\ \bar{X}_1 - \bar{X} = \dfrac{n_2}{n_1+n_2}(\bar{X}_1 - \bar{X}_2), & \bar{X}_2 - \bar{X} = \dfrac{n_1}{n_1+n_2}(\bar{X}_2 - \bar{X}_1) \end{cases}$$

Thus, $$\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2 = \frac{n_1 n_2}{n_1+n_2}(\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1))^2$$

$$\therefore \frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2 / \hat{\sigma}^2 = \left\{ \frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{n_1^{-1}+n_2^{-1}}\,\hat{\sigma}} \right\}^2$$

Also from (Theorem 4.2.5), $F_\alpha(k-1, n-k) = F_\alpha(1, n_1+n_2-2) = t^2_{\alpha/2}(n_1+n_2-2)$

For $k=2$, the probability expression (확률 표현) can be shown as following:

$$P\left\{ \left| \frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{n_1^{-1}+n_2^{-1}}\,\hat{\sigma}} \right| \le t_{\alpha/2}(n_1+n_2-2) \right\} = 1-\alpha$$

Thus, the confidence interval of $\mu_2 - \mu_1$ is:

$$\left[\bar{X}_2 - \bar{X}_1 - t_{\alpha/2}(n_1+n_2-2)\sqrt{n_1^{-1}+n_2^{-1}}\,\hat{\sigma},\right.$$

$$\left. \bar{X}_2 - \bar{X}_1 + t_{\alpha/2}(n_1+n_2-2)\sqrt{n_1^{-1}+n_2^{-1}}\,\hat{\sigma} \right]$$

We can estimate the equivalence of population distributions by checking whether the confidence interval/set contains 0 or not.

〈4.3. DISTRIBUTION OF ORDER STATISTICS

Ex. 4.3.1.

Let X_1, X_2, X_3 be i.i.d ramdom samples from Expl(1) and let $X_{(1)} < X_{(2)} < X_{(3)}$ be the order statistics. Find the joint pdf of $Y = (X_{(1)}, X_{(2)}, X_{(3)})^t$.

Let $u(x_1,x_2,x_3) = (x_{(1)}, x_{(2)}, x_{(3)})^t$

$u(\cdot)$ is a 3!-to-1 function.

$$\therefore f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3) = 3!\, e^{-x_1-x_2-x_3} I_{(0<x_1)} I_{(0<x_2)} I_{(0<x_3)}$$

[Textbook]

Let $u(x_1,x_2,x_3)$ be the function that outputs the order statistics.

$$\mathcal{X} := \{ (x_1,x_2,x_3)^t : x_i > 0 \ (i=1,2,3),\ x_1 \neq x_2,\ x_2 \neq x_3,\ x_1 \neq x_3 \}$$

$$\mathcal{Y} := \{ (y_1,y_2,y_3)^t : 0 < y_1 < y_2 < y_3 \}$$

$u : \mathcal{X} \mapsto \mathcal{Y}$ is a 3!-to-1 function that satisfies the condition in Theorem 4.1.2.

Specifically, let π denote the permutation in $\{1,2,3\}$ and let

$$\mathcal{X}_\pi : \{ (x_1,x_2,x_3)^t : 0 < x_{\pi 1} < x_{\pi 2} < x_{\pi 3} \}$$

The function u^π - function u whose domain is restricted to $\mathcal{X}_\pi$ - is a one-to-one function; $u^\pi : \mathcal{X}_\pi \mapsto \mathcal{Y}$, and is differentiable and

$$J_{u^\pi}(x) = \det\left(\frac{\partial x_{\pi_j}}{\partial x_i} \right) = \pm 1 \quad \forall x \in \mathcal{X}_\pi.$$

For example if $(\pi 1, \pi 2, \pi 3) = (3,2,1)$

$$J_{u^\pi}(x) = \det \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = -1$$

Thus, from theorem 4.1.2, the joint pdf of $Y = (X_{(1)}, X_{(2)}, X_{(3)})^t$ is

$$pdf_Y(y) = \sum_\pi pdf_x((u^\pi)^{-1}(y)) \left| J_{(u^\pi)^{-1}}(y) \right|, \quad y \in \mathcal{Y}$$

$$= \sum_\pi pdf_x(y_{\pi^{-1}1}, y_{\pi^{-1}2}, y_{\pi^{-1}3}) \cdot |\pm 1|, \quad y \in \mathcal{Y}$$

Meanwhile, the joint pdf of $X = (X_1, X_2, X_3)^t$ is:

$$pdf_X(x_1,x_2,x_3) = e^{-(x_1+x_2+x_3)} I_{(x_1>0,\ x_2>0,\ x_3>0)}$$

The number of permutation $\pi = 3! = 6$

$$pdf_Y(y_1,y_2,y_3) = \sum_\pi e^{-(y_{\pi^{-1}1} + y_{\pi^{-1}2} + y_{\pi^{-1}3})}, \quad y \in \mathcal{Y}$$

$$= 6\, e^{-(y_1+y_2+y_3)} I_{(0<y_1<y_2<y_3)}$$

〈Theorem 4.3.1 : Joint pdf of order statistics〉

If the population distribution is continuous and its pdf is $f(x)$, the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$'s joint pdf is:

$$pdf_Y(y_1,\cdots,y_n) = n!\, f(y_1)\cdots f(y_n)\, I_{(y_1<\cdots<y_n)}$$

(proof)

let $u(x_1,\cdots,x_n) = (x_{(1)},\cdots,x_{(n)})^t$. Function u is

$$\mathcal{X} := \{(x_1,\cdots,x_n)^t : f(x_i)>0,\ x_i \neq x_j\ (i\neq j)\}$$

$$\mathcal{Y} := \{(y_1,\cdots,y_n)^t : f(y_i)>0,\ y_1<y_2<\cdots<y_{n-1}<y_n\}$$

$u: \mathcal{X} \mapsto \mathcal{Y}$ is a $n!$-to-1 function and satisfies the conditions in theorem 4.1.2.

specifically, let π denote a permutation of $\{1,2,\cdots,n\}$

$$\mathcal{X}_\pi = \{(x_1,\cdots,x_n)^t : f(x_i)>0\ (i=1,\cdots,n),\ x_{\pi_1}<\cdots<x_{\pi_n}\}$$

$u^\pi(x_1,\cdots,x_n) = (x_{\pi_1},\cdots,x_{\pi_n})^t$.

u^π: function u whose domain is restricted to $\mathcal{X}_\pi$

u^π is a one-to-one function that maps to $\mathcal{Y}$, is differentiable, and its first derivative matrix (Jacobian Matrix) $\left|\frac{\partial x_{\pi_i}}{\partial x_i}\right|$ is equivalent to an identity matrix whose column and rows have been changed.

Thus, $J_{u^\pi}(x) = \det\left(\frac{\partial x_{\pi_i}}{\partial x_i}\right) = \pm 1 \quad \forall x \in \mathcal{X}_\pi$.

from theorem 4.1.a, the joint pdf of $Y=(X_{(1)},\cdots,X_{(n)})^t$ is:

$$pdf_Y(y) = \sum_\pi pdf_X\left((u^\pi)^{-1}(y)\right)\left|J_{(u^\pi)^{-1}}(y)\right|,\ y\in\mathcal{Y}$$

$$= \sum_\pi pdf_X(y_{\pi^{-1}_1},\cdots,y_{\pi^{-1}_n})\,|\pm 1|,\ y\in\mathcal{Y}$$

Meanwhile, the pdf of $X=(X_1,\cdots,X_n)^t$ is:

$$pdf_X(x_1,\cdots,x_n) = f(x_1)\cdots f(x_n)$$

The number of permutations is $n!$.

$$\therefore pdf_Y(y) = \sum_\pi f(y_{\pi^{-1}_1})\cdots f(y_{\pi^{-1}_n}),\ y\in\mathcal{Y}$$

$$= n!\, f(y_1)\cdots f(y_n)\, I_{(y_1<\cdots<y_n)}$$

〈Theorem 4.3.2. Marginal pdf of order statistics〉

If the population distribution is continuous, its pdf is $f(x)$ and its cdf is $F(x)$, the order statistics of $X_1,\cdots,X_n$ = $X_{(1)},\cdots,X_{(n)}$ its one- or two- dimensional pdf is:

(a) pdf of $X_{(r)}$ $(1\le r\le n)$

$$pdf_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}\left[F(x)\right]^{r-1} f(x)\left[1-F(x)\right]^{n-r}$$

(b) pdf of $(X_{(r)},X_{(s)})^t$ $(1\le r\le s\le n)$

$$pdf_{X_{(r)},X_{(s)}}(x,y) = \frac{n!}{(r-1)!(s-1-r)!(n-s)!}\left[F(x)\right]^{r-1} f(x)\left[F(y)-F(x)\right]^{s-1-r} f(y)\left[1-F(y)\right]^{n-s}$$

(proof)

Let $f_{1,\cdots,n}(y_1,\cdots,y_n)$ denote the joint pdf of $(X_{(1)},\cdots,X_{(n)})^t$ and let $f_{(r)}(x)$ and $f_{r,s}(x,y)$ denote the joint pdf of $X_{(r)}$ and $(X_{(r)},X_{(s)})^t$ respectively.

$$f_r(x) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f_{1,\cdots,n}(y_1,\cdots,x,\cdots,y_n)\,dy_1\cdots dy_{r-1}\,dy_{r+1}\cdots dy_n$$

$$f_{r,s}(x,y) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f_{1,\cdots,n}(y_1,\cdots,x,\cdots,y,\cdots,y_n)\,dy_1\cdots dy_{r-1}dy_{r+1}\cdots dy_{s-1}dy_{s+1}\cdots dy_n$$

(a) In this case, within the integrand $f_{1,\cdots,n}(y_1,\cdots,x,\cdots,y_n)$, the term $f(y_1)\cdots f(y_{r-1})$ is invariant under permutations of $y_1,\cdots,y_{r-1}$

$f_r(x) =$

① $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(y_1)\cdots f(y_{r-1})\, I_{(y_1<\cdots<y_{r-1}<x)}\,dy_1\cdots dy_{r-1}$

$= \frac{1}{(r-1)!}\int_{-\infty}^{x}\int_{-\infty}^{x} f(y_1)\cdots f(y_{r-1})\,dy_1\cdots dy_{r-1}$

$= \frac{1}{(r-1)!}\left[F(x)\right]^{r-1}$

② $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(y_{r+1})\cdots f(y_n)\, I_{(x<y_{r+1}<\cdots<y_n)}\,dy_{r+1}\cdots dy_n$

$= \frac{1}{(n-r)!}\int_{x}^{+\infty}\cdots\int_{x}^{+\infty} f(y_{r+1})\cdots f(y_n)\,dy_{r+1}\cdots dy_n$

$= \frac{1}{(n-r)!}\left[1-F(x)\right]^{n-r}$

$$\therefore f_r(x) = n!\, f(x)\frac{1}{(r-1)!}\left[F(x)\right]^{r-1}\frac{1}{(n-r)!}\left[1-F(x)\right]^{n-r}$$

(b) In this case, within the integrand $f_{1,\cdots,n}(y_1,\cdots,x,\cdots,y,\cdots,y_n)$, the term $f(y_1),\cdots,f(y_{r-1})$ is invariant under the permutations of $y_1,\cdots,y_{r-1}$ and $f(y_{s+1}),\cdots,f(y_n)$ is invariant under the permutations of $y_{s+1},\cdots,y_n$. Hence

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(y_{r+1})\cdots f(y_{s-1})\, I_{(x<y_{r+1}<\cdots<y_{s-1}<y)}\,dy_{r+1}\cdots dy_{s-1}$$

$$= \frac{1}{(s-1-r)!}\int_{x}^{y}\cdots\int_{x}^{y} f(y_{r+1})\cdots f(y_{s-1})\,dy_{r+1}\cdots dy_{s-1}$$

$$= \frac{1}{(s-1-r)!}\left[F(y)-F(x)\right]^{s-1-r}$$

$$\therefore f_{r,s}(x,y) = n!\, f(x)f(y)\frac{1}{(r-1)!}\left[F(x)\right]^{r-1}\cdot\frac{1}{(s-1-r)!}\left[F(y)-F(x)\right]^{s-1-r}\frac{1}{(n-s)!}\left[1-F(y)\right]^{n-s}$$

★ Ex. 4.3.2 Order Statistics in Uniform Distribution

Let $X_{(1)} < \cdots < X_{(n)}$ denote the order statistic of random sample of size n, from $U(0,1)$ distribution. Prove the following:

(a) $X_{(r)} \sim \text{Beta}(r, n-r+1)$, i.e.,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r} I_{(0,1)}(x)$$

(b) $Z = (X_{(1)}, X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)})^t \sim \text{Dirichlet}(\underbrace{1, \cdots, 1}_{n+1 \text{ 개}})$. i.e.,

$$pdf_Z(z_1, \cdots, z_n) = \frac{\Gamma(n+1)}{\Gamma(1)\cdots\Gamma(1)} I_{(z_1>0, \cdots, z_n>0,\ z_1+\cdots+z_n<1)}$$

(proof)

$$\begin{aligned} pdf_{X_{(r)}}(x) &= \frac{n!}{(r-1)!(n-r)!}[F_X(x)]^{r-1} f_X(x)[1-F_X(x)]^{n-r} \\ &= \frac{n!}{(r-1)!(n-r)!} x^{r-1} \cdot 1 \cdot (1-x)^{n-r} \\ &= \frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r} \\ &= \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} x^{r-1}(1-x)^{n-r} \end{aligned}$$

$\therefore X_{(r)} \sim \text{Beta}(r, n-r+1)$

(b) Let

$$\begin{aligned} Y_1 &= X_{(1)} & X_{(1)} &= Y_1 \\ Y_2 &= X_{(2)} - X_{(1)} & X_{(2)} &= Y_1 + Y_2 \\ &\vdots & &\vdots \\ Y_n &= X_{(n)} - X_{(n-1)} & X_{(n)} &= Y_1 + \cdots + Y_n \end{aligned}$$

$u : (X_{(1)}, \cdots, X_{(n)}) \mapsto (Y_1, \cdots, Y_n)$ is one-to-one

$$|J| = \left| \det\left(\frac{\partial Y}{\partial x}\right) \right| = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = 1$$

$$\begin{aligned} f_{X_{(1)}\cdots X_{(n)}}(x_1, \cdots, x_n) &= n! I_{(0<x_1<\cdots<x_n<1)} \\ &= n! I_{(0<y_1<1)} \cdots I_{(0<y_n<1)} I_{(y_1+\cdots+y_n<1)} \\ &= \frac{\Gamma(n+1)}{\Gamma(1)\cdots\Gamma(1)} I_{(0<y_1<1)} \cdots I_{(0<y_n<1)} I_{(y_1+\cdots+y_n<1)} \quad \checkmark\checkmark \end{aligned}$$

[Textbook proof].

For $Z = (X_{(1)}, X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)})^t$ and $Y = (X_{(1)}, \cdots, X_{(n)})^t$

$$Z = (Z_1, \cdots, Z_n)^t = u((Y_1, \cdots, Y_n)^t) = u(Y)$$

$$u : \begin{cases} z_1 = y_1 \\ z_2 = y_2 - y_1 \\ \vdots \\ z_n = y_n - y_{n-1} \end{cases} \qquad u^{-1} : \begin{cases} y_1 = z_1 \\ y_2 = z_1 + z_2 \\ \vdots \\ y_n = z_1 + z_2 + \cdots + z_n \end{cases}$$

$$\mathcal{Y} = \{(y_1, \cdots, y_n)^t : 0 < y_1 < \cdots < y_n < 1\}$$
$$\mathcal{Z} = \{(z_1, \cdots, z_n)^t : z_i > 0\ (i = 1, 2, \cdots, n),\ z_1 + \cdots + z_n < 1\}$$

functions $u : \mathcal{Y} \mapsto \mathcal{Z}$ is one-to-one and satisfies the conditions in Theorem 4.1.1 and its Jacobian matrix of the inverse function u^{-1} is 1 hence, the pdf of $Z = (X_{(1)}, X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)})^t$ is:

$$pdf_Z(z_1, \cdots, z_n) = n! I_{(z_1>0, \cdots, z_n>0,\ z_1+\cdots+z_n<1)}$$

Ex 4.3.3. Order Statistics of Exponential Distribution * Exercise 7.13

Let $X_{(1)} < \cdots < X_{(n)}$ denote the order statistics of random samples $X_1, \cdots, X_n \overset{i.i.d}{\sim} Exp(1)$

Let $Z_1 = nX_{(1)}$, $Z_2 = (n-1)(X_{(2)} - X_{(1)})$, $\cdots$, $Z_n = X_{(n)} - X_{(n-1)}$

Show that $Z_1, Z_2, \cdots, Z_n$ are independent and follow the $Exp(1)$ distribution.

(Solution)

$$\mathcal{X} := \{(x_1, \cdots, x_n) : f(x_i) > 0\ (i = 1, \cdots, n),\ x_1 < \cdots < x_n\}$$
$$\mathcal{Y} := \{(z_1, \cdots, z_n) : 0 < z_i < +\infty\ (i = 1, \cdots, n)\}$$

Let $u : \mathcal{X} \mapsto \mathcal{Y}$

$$u \begin{cases} Z_1 = nX_{(1)} \\ Z_2 = (n-1)(X_{(2)} - X_{(1)}) \\ \vdots \\ Z_n = X_{(n)} - X_{(n-1)} \end{cases} \qquad u^{-1} \begin{cases} X_{(1)} = \frac{1}{n} Z_1 \\ X_{(2)} = \frac{1}{n} Z_1 + \frac{1}{n-1} Z_2 \\ \vdots \\ X_{(n-1)} = \frac{1}{n} Z_1 + \cdots + \frac{1}{2} Z_{n-1} \\ X_{(n)} = \frac{1}{n} Z_1 + \cdots + \frac{1}{2} Z_{n-1} + Z_n \end{cases}$$

$$|J_{u^{-1}}| = \left| \det\left(\frac{\partial y^{-1}}{\partial x}\right) \right| = \begin{array}{c} \\ X_{(1)} \to \\ X_{(2)} \to \\ \\ \\ X_{(n)} \to \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & & \end{array} \\ \begin{bmatrix} 1/n & 0 & \cdots & 0 & 0 \\ 1/n & 1/n-1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1/n & 1/n-1 & \cdots & 1/2 & 1 \end{bmatrix} \end{array}$$

$$= (n!)^{-1}$$

$$\begin{aligned} f_{X_{(1)}, \cdots X_{(n)}}(x_1, \cdots, x_n) &= n! e^{-(x_1 + \cdots + x_n)} I_{(0<x_1<\cdots<x_n<+\infty)} \\ &= n! e^{-(z_1 + (z_1+z_2) + \cdots + (z_1+\cdots+z_n))} I_{(0<z_1)} \cdots I_{(0<z_n)} \cdot n! \cdot |J| \\ &= \left[e^{-z_1} I_{(0<z_1)}\right]\left[e^{-z_2} I_{(0<z_2)}\right] \cdots \left[e^{-z_n} I_{(0<z_n)}\right] \\ &= f_{Z_1}(z_1) \cdots f_{Z_n}(z_n) \end{aligned}$$

$\therefore Z_1, \cdots, Z_n$ are mutually independent and $Z_i \overset{i.i.d}{\sim} Exp(1)$ $(i = 1, 2, \cdots)$

Definition: Representational Definition of Order Statistics in Exp(1)

Let $X_{(1)} < \cdots < X_{(n)}$ denote the order statistics from Exp(1) distribution.

Then, $$(X_{(r)})_{1\le r\le n} \overset{d}{\equiv} \left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)_{1\le r\le n}, \quad Z_r \overset{i.i.d}{\sim} \text{Exp}(1) \ (r=1,\cdots,n)$$

〈Theorem 4.3.3. Probability Integral Transformation〉 [58]

If the random variable X is continuous and it cdf is strictly increasing, i.e.,

if $x_1 < x_2$, $F(x_1) < F(x_2)$

and the inverse of cdf is F^{-1} the following are true.

(a) $F(X) \sim U(0,1)$ i.e., $F(X)$ follows the $U(0,1)$ distribution.

(b) For a random variable $U \sim U(0,1)$, $F^{-1}(U) \overset{d}{\equiv} X$
meaning, the cdf of $F^{-1}(U)$ is F and thus $F^{-1}(U)$ and X have the same distribution.

(proof) since F is strictly increasing, F and F^{-1} are one-to-one transformations.

* use cdf to find the distribution!

We also know the following:

(a) $F(x) \le u \Leftrightarrow x \le F^{-1}(u)$, $F^{-1}(u) \le x \Leftrightarrow u \le F(x)$, $F(F^{-1}(u)) = u$

$P(F(X) \le u) = P[x \le F^{-1}(u)] = F(F^{-1}(u)) = u$

$\because P(F(x) \le u) = u \Leftrightarrow \frac{\partial}{\partial u} P[F(x) \le u) = 1 \quad \therefore \text{pdf}_{F(X)}(u) = 1$

Since the domain of $F(x)$ is $(0,1)$ and its pdf is 1, we can conclude that $F(X) \sim U(0,1)$

(b) Let $F^{-1}(U) = Y$

$P(Y \le x) = P[F(Y) \le F(x)] = P(U \le F(x)) = F(x) \quad \forall x : -\infty < x < +\infty$
(cdf of $U(0,1)$: $P(U \le x) = x$)

$\therefore$ cdf of $F^{-1}(U) = Y$ is $F(x)$. Hence $F^{-1}(U) \overset{d}{\equiv} X$.

*58) If the cdf is not strictly increasing but simply increasing, define the inverse function $F^{-1}(u) = \inf\{x : F(x) \ge u\}$ $(0 \le u \le 1)$ and the precedent theorem-(b) is true for general random variables.

Ex. 4.3.4. Order Statistics of Uniform and Exponential Distribution.

Let $U_{(1)} < \cdots < U_{(n)}$ denote the order statistics from $U(0,1)$ and let $Y_{(1)} < \cdots < Y_{(n)}$ denote the order statistics from Exp(1).

$$(U_{(r)})_{1\le r\le n} \overset{d}{\equiv} (1 - e^{-Y_{(r)}})_{1\le r\le n}$$

(Solution)

Let $Y_1, \cdots, Y_n \overset{i.i.d}{\sim} \text{Exp}(1)$

The cdf of Exp(1) is: $G(y) = \begin{cases} 1 - e^{-y}, & y \ge 0 \\ 0, & y < 0 \end{cases}$

From Theorem 4.3.3. $G(Y_i) \overset{i.i.d}{\sim} U(0,1)$
and since cdf is an increasing function

$$(G(Y_{(r)}))_{1\le r\le n} \overset{d}{\equiv} (U_{(r)})_{1\le r\le n}$$

〈Theorem 4.3.4. The Representational Definition of Order Statistics〉

If the population distribution is continuous and its cdf is $F(x)$, for the order statistics $X_{(1)} < \cdots < X_{(n)}$ from n random samples, the following is true:

let $h(y) = F^{-1}(1 - e^{-y}) I_{(0,+\infty)}(y)$

$$(X_{(r)})_{1\le r\le n} \overset{d}{\equiv} \left(h\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)\right)_{1\le r\le n}, \quad Z_r \overset{i.i.d}{\sim} \text{Exp}(1) \ (r=1,\cdots,n)$$

(proof) from previous theorems & exercises, we know the following

(1) $F(X) \overset{d}{\equiv} U$ $U \sim U(0,1)$ and X is an on arbitrary random variable.

(2) Since F is increasing, $F(X_{(r)}) \overset{d}{\equiv} U_{(r)}$ $(r=1,\cdots,n)$ Thm 4.3.3

(3) $U_{(r)} \overset{d}{\equiv} 1 - e^{-y_{(r)}}$ Ex. 4.3.4

(4) $Y_{(r)} \overset{d}{\equiv} \frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r$, $Z_i \overset{i.i.d}{\sim} \text{Exp}(1)$ $(i=1,\cdots,r)$ Ex. 4.3.3.

(5) combine (3) and (4) and we get $U_{(r)} \overset{d}{\equiv} 1 - \exp\left\{-\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)\right\}$

(6) combine (2) and (5): $X_{(r)} \overset{d}{\equiv} F^{-1}(U_{(r)}) \overset{d}{\equiv} F^{-1}\left(1 - \exp\left\{-\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)\right\}\right)$

(7) Let $h(x) = F^{-1}(1 - e^{-x}) I_{(0,+\infty)}(x)$

Hence, $X_{(r)} \overset{d}{\equiv} h\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)$, $Z_i \overset{i.i.d}{\sim} \text{Exp}(1)$ $(i=1,\cdots,r)$

4.3. MULTIVARIATE NORMAL DISTRIBUTIONS

Let $Z_1, Z_2, \cdots, Z_n \overset{i.i.d}{\sim} N(0,1)$, $A = (a_{ij})_{\substack{1\le i\le n \\ 1\le j\le n}}$, $\mu = (\mu_1, \cdots, \mu_n)^t$

$$\begin{cases} X_1 = a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1 \\ X_2 = a_{21}Z_1 + \cdots + a_{2n}Z_n + \mu_2 \\ \vdots \\ X_n = a_{n1}Z_1 + \cdots + a_{nn}Z_n + \mu_n \end{cases}, \text{ i.e., } X = AZ + \mu, \ X = (X_1, \cdots, X_n)^t, Z = (Z_1, \cdots, Z_n)^t$$

The following are true

(a) if the matrix A is a nonsingular matrix, the pdf of $X = AZ + \mu$ is

$$pdf_X(x) = (|\det(2\pi\Sigma)|)^{-1/2} \exp\left\{-(x-\mu)^t \Sigma^{-1}(x-\mu)/2\right\}, \ x \in \mathbb{R}^n \ (\Sigma = AA^t)$$

(b) The mgf of $X = AZ + \mu$ is:

$$mgf_X(t) = \exp\left\{\mu^t t + \frac{1}{2} t^t \Sigma t\right\}, \ t \in \mathbb{R}^n.$$

(proof)

(a) Let $Z = (Z_1, \cdots, Z_n)^t$

$$pdf_Z(z_1, \cdots, z_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}(z_1^2 + \cdots + z_n^2)\right\}$$
$$= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2} z^t z\right\}$$

$$X = AZ + \mu \ \therefore \ A^{-1}(X-\mu) = Z \Rightarrow z^t z = (X-\mu)^t (A^{-1})^t A^{-1}(X-\mu)$$
$$= (X-\mu)^t (A^t)^{-1} A^{-1}(X-\mu)$$
$$= (X-\mu)^t (AA^t)^{-1}(X-\mu)$$
$$= (X-\mu)^t \Sigma^{-1}(X-\mu)$$

$$Z = A^{-1}(X-\mu) \ \therefore \ |J| = \left|\det\left(\frac{\partial Z}{\partial X}\right)\right| = |\det(A^{-1})|$$
$$= |\det(A)^{-1}|$$

$$\det(\Sigma) = \det(AA^t) = \det(A)\det(A^t) = \det(A)\det(A) = \det(A)^2$$

$$\therefore |J| = |\det(A)^{-1}| = |\det(\Sigma)|^{-1/2}$$

$$\therefore pdf_X(x) = pdf_Z(u^{-1}(z))|J| = (2\pi)^{-n/2} \cdot |\det(\Sigma)|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\}$$
$$= (|\det(2\pi\Sigma)|)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\}$$
$$= (|\det(2\pi\Sigma)|)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\} \ \square.$$

* multiplicative properties of determinant:

(1) $\det(AB) = \det(A)\det(B)$
(2) $\det(A^{-1}) = \det(A)^{-1}$
(3) $\det(A^t) = \det(A)$

[Textbook proof]

(a) Let $X = AZ + \mu$. If A is a nonsingular matrix, the function $u: \mathbb{R}^n \mapsto \mathbb{R}^n$ is a one-to-one function. The conditions in Theorem 4.1.1 are satisfied and the Jacobian of its inverse function $u^{-1}(x) = A^{-1}(x-\mu)$ is:

$$J_{u^{-1}} = \det(A^{-1}) = \det(A)^{-1}$$

Hence, $pdf_X(x) = pdf_Z(u^{-1}(x))|J_{u^{-1}}(x)|$

$$= (2\pi)^{-n/2} \cdot \exp\left\{-\frac{1}{2}(A^{-1}(x-\mu))^t A^{-1}(x-\mu)\right\} |\det(A)|^{-1}$$

If $\Sigma = AA^t$, $\det(2\pi\Sigma) = (2\pi)^n \det(A)^2$

$$\therefore pdf_X(x) = (|\det(2\pi\Sigma)|)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\}$$

(b) $mgf_X(t) = E\left[\exp\left\{t^t X\right\}\right] = E\left[\exp\left\{t^t(AZ+\mu)\right\}\right]$

$$= e^{t^t\mu} \cdot E\left[\exp\left\{t^t AZ\right\}\right]$$

* $M_Z(t) = \exp\left\{\frac{1}{2}t_1^2 + \cdots + \frac{1}{2}t_n^2\right\} = \exp\left\{+\frac{1}{2}t^t t\right\}$

$$\therefore e^{t^t\mu} \cdot E\left[\exp\left\{t^t AZ\right\}\right] = e^{t^t\mu} \cdot E\left[\exp\left\{+\frac{1}{2}(A^t t)^t Z\right\}\right]$$
$$= \exp\left\{t^t\mu + \frac{1}{2}(A^t t)^t (A^t t)\right\}$$
$$= \exp\left\{t^t\mu + \frac{1}{2} t^t AA^t t\right\}$$
$$= \exp\left\{\mu^t t + \frac{1}{2} t^t \Sigma t\right\} \ \square$$

The distribution of X from theorem 4.4.1 is called the multivariate normal distribution and is denoted as $X \sim N_n(\mu, \Sigma)$ or $X \sim N(\mu, \Sigma)$ omitting the number of random variables. Especially, if A is a nonsingular matrix with an inverse, as in the case of (a), the distribution of X is called the "nonsingular multivariate normal distribution".

〈Theorem 4.4.2〉

(a) If $X \sim N(\mu, \Sigma)$, $E[X] = \mu$, $Var(X) = \Sigma$

(b) For matrix Σ, if $AA^t = \Sigma$, $A = A^t$, $A = \Sigma^{1/2}$,

$$X \sim N_n(\mu, \Sigma) \Leftrightarrow X \overset{d}{=} \Sigma^{1/2} Z + \mu, \ Z \sim N_n(0, I)$$

(proof)

(a) As shown in Chapter 2 Section 5, the multivariate random variable's mean and variance can be calculated using the derivative of the cgf. The cgf of $N(\mu, \Sigma)$ is

$$C(t) = \mu^t t + \frac{1}{2} t^t \Sigma t$$

$$\therefore E[X] = \dot{C}(t=0) = \mu, \ Var(X) = \ddot{C}(t=0) = \Sigma$$

(b) Similar to Theorem 4.4.1 (b) let $\Sigma^{1/2}\Sigma^{1/2}=\Sigma$ and $\Sigma^{1/2}=(\Sigma^{1/2})^t$

$$mgf_{\Sigma^{1/2}Z+\mu}(t) = \exp\left\{\mu^t t + \frac{1}{2}t^t\Sigma t\right\}$$

Thus $X \sim N_n(\mu,\Sigma)$

Definition: Multivariate Normal Distribution.

$Z \sim N_n(0,I) \Leftrightarrow Z=(Z_1 \cdots Z_n)^t,\ Z_i \overset{i.i.d}{\sim} N(0,1)\quad (i=1,\cdots,n)$

$X \sim N(\mu,\Sigma) \Leftrightarrow$ (1) $^{60)}$ $X \overset{d}{=} AZ+\mu$, $Z \sim N(0,I)$, $AA^t=\Sigma$

$\Leftrightarrow$ (2) $X \overset{d}{=} \Sigma^{1/2}Z+\mu$, $Z \sim N(0,I)$, $\Sigma^{1/2}\Sigma^{1/2}=\Sigma$, $\Sigma^{1/2}=(\Sigma^{1/2})^t$

$\Leftrightarrow$ (3) $mgf_X(t) = \exp\left\{\mu^t t + \frac{1}{2}t^t\Sigma t\right\}$, $t \in \mathbb{R}^n$

$\Leftrightarrow$ (4) (If Σ is a non-Singular matrix)

$$pdf_X(x) = |\det(2\pi\Sigma)|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}$$
$$x \in \mathbb{R}^n.$$

*60) in (1) A is not necessarily a nxn non-Singular matrix.
$A \in \mathbb{R}^{n\times m}$ s.t. $AA^t=\Sigma$ is sufficient.

Ex. 4.4.1. Bivariate Normal Distribution $N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$
or $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho)$

For the variance-covariance matrix Σ, $\det(\Sigma)=\sigma_1^2\sigma_2^2(1-\rho^2)$.
If $\sigma_1>0$, $\sigma_2>0$, $-1<\rho<1$,

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix} \quad * \Sigma=\begin{pmatrix}a & b\\ c & d\end{pmatrix} \Rightarrow \Sigma^{-1}=\frac{1}{ad-bc}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$$

Thus, the pdf is given as: $f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}Q}$

$$Q = \frac{1}{1-\rho^2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

However, if $\sigma_1>0$, $\sigma_2>0$, $\rho=\pm1$, inverse of the covariance matrix does not exist, and by ⟨Theorem 2.2.3⟩ the following linear relationship (선형 관계) holds between two random variables.

$$\rho=1 \Leftrightarrow P\left(\frac{x_2-\mu_2}{\sigma_2}=\frac{x_1-\mu_1}{\sigma_1}\right)=1,\quad \rho=-1 \Leftrightarrow P\left(\frac{x_2-\mu_2}{\sigma_2}=-\frac{x_1-\mu_1}{\sigma_1}\right)=1$$

⟨Theorem 4.4.3. Properties of the Multivariate Normal Distribution⟩

(a) If $X \sim N(\mu,\Sigma)$, for constant matrix and vector A, b

$$AX+b \sim N(A\mu, A\Sigma A^t)$$

(b) If $\begin{pmatrix}X_1\\ X_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$ and $cov(X_1,X_2)=\Sigma_{12}=\Sigma_{21}^t=0$,
X_1 and X_2 are independent.

(c) If $X \sim N(\mu,\Sigma)$, A and B are constant matrices,
$Cov(AX,BX)=A\Sigma B=0$, AX and BX are independent.

(proof)

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

$$\begin{aligned}M_{AX+b}(t) &= E[\exp\{t^t(AX+b)\}] = E[\exp\{t^t b\}\exp\{t^t AX\}]\\ &= E[\exp\{b^t t\}\exp\{(A^t t)^t X\}]\\ &= \exp\{b^t t\}\cdot\exp\left\{\mu^t(A^t t)+\frac{1}{2}(A^t t)^t\Sigma(A^t t)\right\}\\ &= \exp\left\{(A\mu+b)^t t + \frac{1}{2}t^t A\Sigma A^t t\right\}\end{aligned}$$

$\therefore$ mgf of $N(A\mu+b, A\Sigma A^t)$

$\therefore AX+b \sim N(A\mu+b, A\Sigma A^t)$

(b) If $\Sigma_{12}=\Sigma_{21}^t=0$, $\begin{pmatrix}X_1\\ X_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}\right)$

Let $X=\begin{pmatrix}X_1\\ X_2\end{pmatrix}$, $M_X(t)=E[\exp\{t_1^t X_1 + t_2^t X_2\}]$

$$\begin{aligned}&= \exp\left\{\begin{pmatrix}\mu_1^t & \mu_2^t\end{pmatrix}\begin{pmatrix}t_1\\ t_2\end{pmatrix} + \frac{1}{2}\begin{pmatrix}t_1^t & t_2^t\end{pmatrix}\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}\begin{pmatrix}t_1\\ t_2\end{pmatrix}\right\}\\ &= \exp\left\{\mu_1^t t_1 + \mu_2^t t_2 + \frac{1}{2}(t_1^t\Sigma_{11}t_1 + t_2^t\Sigma_{22}t_2)\right\}\\ &= \exp\left\{\mu_1^t t_1 + \frac{1}{2}t_1^t\Sigma_{11}t_1\right\}\exp\left\{\mu_2^t t_2 + \frac{1}{2}t_2^t\Sigma_{22}t_2\right\}\\ &= M_{X_1}(t_1)\cdot M_{X_2}(t_2)\end{aligned}$$

$\therefore X_1$ and X_2 are independent.

(c) $X \sim N(\mu,\Sigma)$

Let $C=\begin{pmatrix}A\\ B\end{pmatrix}$, $CX=\begin{pmatrix}AX\\ BX\end{pmatrix} \sim N\left(\begin{pmatrix}A\mu\\ B\mu\end{pmatrix}, \begin{pmatrix}A\\ B\end{pmatrix}\Sigma\begin{pmatrix}A^t & B^t\end{pmatrix}\right)$

$$\begin{pmatrix}AX\\ BX\end{pmatrix} \sim N\left(\begin{pmatrix}A\mu\\ B\mu\end{pmatrix}, \begin{pmatrix}A\Sigma A^t & A\Sigma B^t\\ B\Sigma A^t & B\Sigma B^t\end{pmatrix}\right)$$
$$\sim N\left(\begin{pmatrix}A\mu\\ B\mu\end{pmatrix}, \begin{pmatrix}A\Sigma B^t & 0\\ 0 & B\Sigma B^t\end{pmatrix}\right)$$

$\therefore$ from (b) AX and BX are independent.

Ex. 4.4.2 Independence in bivoriate normal distributions.

From theorem 2.4.4 and Ex. 2.4.3 we know that if random variables X_1 and X_2 are independent, $Cov(X_1,X_2)=0$. However, the reverse is not always true. For $(X_1,X_2)^t \sim$ bivariate normal distribution, if $Cov(X_1,X_2)=0$, X_1 and X_2 are independent.

〈Theorem 4.4.4 Marginal/Conditional Distribution of Multivariate Normal Distributions〉:

(a) (Marginal Distribution) If $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$, $X_1 \sim N(\mu_1, \Sigma_{11})$

(b) (Conditional Distribution) If $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$ and its Covariance matrix is non-Singular,

$$X_2|X_1=x_1 \sim N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

(proof)

(a) $X_1=(I,0)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, thus $X_1 \sim N\left[(I\ 0)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, (I\ 0)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I \\ 0 \end{pmatrix}\right]$

$\sim N(\mu_1, \Sigma_{11})$

(b)

① It is known that: [61]

$$\det\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \det(\Sigma_{11})\det(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

Let $\Sigma_{22\cdot1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_{11}^{-1}+\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1} \\ -\Sigma_{22\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22\cdot1}^{-1} \end{pmatrix}$$

use this to find the Conditional pdf $pdf_{X_2|X_1=x_1}(x_2|x_1) = pdf_{X_1,X_2}(x_1,x_2)/pdf_{X_1}(x_1)$

② $(X_2|X_1=x_1) \sim N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$

$\Rightarrow (X_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1)\,|\,X_1=x_1) \sim N(0, \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$

$$\Rightarrow \begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1) \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}$$

$$\begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}\begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}^t = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}$$

$$\therefore \begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}\right)$$

Thus, from Theorem 4.4.3-(b) $X_1-\mu_1$ and $X_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1)$ are independent and

$$X_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1) \sim N(0, \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$ □

Ex. 4.4.3 Conditional Distribution of the bivariate normal distribution.

$$X_{2|1} \sim N(\mu_2+\rho\sigma_2(x_1-\mu_1)/\sigma_1, \sigma_2^2(1-\rho^2))$$

〈Theorem 4.4.5 Distribution of Quadratic Form〉

(a) If $X \sim N_K(\mu,\Sigma)$ and Σ is non-Singular,

$$(X-\mu)^t\Sigma^{-1}(X-\mu) \sim \chi^2(K)$$

(b) If $Z \sim N(0,I)$ and $A^2=A$, $Z^tAZ \sim \chi^2(r)$ and degrees of freedom is $r=\text{trace}(A)$ [62]

* [62] 일반적으로 $x^tAx = x^t\left(\frac{A+A^t}{2}\right)x^t$ 이므로 이차형식 z^tAz 에서 행렬 A는 언제나 대칭행렬인 것으로 가정되어 있으며, (b)에서의 조건 $A^2=A$는 $z^tAz \sim \chi^2(A)$ 이기 위한 충분조건일 뿐만 아니라 필요조건인 것도 알려져 있다. 여기에서 trace(A)는 행렬 A의 모든 대각원소의 합이다.

(proof)

(a) $\Sigma^{-1/2}(X-\mu) \overset{d}{\equiv} Z \sim N(0, I_K)$

Distribution of z^tz : $M_{z^tz}(t) = E\left[\exp\left\{t\,(z^tz)\right\}\right]$

$= E\left[\exp\left\{t\cdot z_1^2+\cdots+tz_K^2\right\}\right]$

$= E\left[\exp\left\{tz_1^2\right\}\right]^K$

We know that $z^2 \overset{d}{\equiv} V \sim \chi^2(1)$, $M_V(t) = (1-2t)^{-1/2}$

$\therefore E\left[\exp\left\{t\cdot z_1^2\right\}\right]^K = (1-2t)^{-K/2}$: MGF of $\chi^2(K)$ ✓

[Textbook proof]

From Theorem 4.4.3 (a) $\Sigma^{-1/2}(X-\mu) \sim N_K(0,I)$

Thus, let $Z=(z_1,\cdots,z_K)^t = \Sigma^{-1/2}(X-\mu)$ and we have:

$$(X-\mu)^t\Sigma^{-1}(X-\mu) = z^tz = z_1^2+\cdots+z_K^2, \quad z_i \overset{i.i.d}{\sim} N(0,1)\ (i=1,\cdots,K)$$

From the representational definition of the Chi-square distribution,

$$(X-\mu)^t\Sigma^{-1}(X-\mu) \sim \chi^2(K)$$

(b)

In *[62] it is mentioned that $A^T=A$ (symmetric).

$\therefore$ A can be decomposed to: $A=P\Lambda P^T$

$\because PP^t=I$, $P^tP=I$ and $\Lambda = \text{diag}(\lambda_1,\cdots,\lambda_r,0,\cdots,0)$, $\lambda_i=1\ (i=1,\cdots,r)$

$z^tAz = z^tP\Lambda P^tz = (P^tz)\Lambda(P^tz)$ $\quad P^tz \sim N(P^t0, P^tP) = N(0,I)$

$\therefore (P^tz)\Lambda(P^tz) = z_1^2+\cdots+z_r^2$, $z_i^2 \sim \chi^2(1)$, $(i=1,\cdots,r)$

$z_1^2+\cdots+z_r^2 \sim \chi^2(r)$, $r=tr(A)$ $\therefore z^tAz \sim \chi^2(r)$

[Textbook proof]

Matrix A is a symmetric matrix thus the following diagonalization is possible.

$$A = P^t \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} P, \quad P^tP = PP^t = I$$

Here $\lambda_1, \cdots, \lambda_n$ satisfies $\det(A-\lambda_i I)=0$ and since $A^2=A$,
i.e, $\lambda_i^2=\lambda_i$, $\lambda_i = 0$ or 1. say, the first r eigenvalues $\lambda_1, \cdots, \lambda_r = 1$ and the others $\lambda_{r+1}, \cdots, \lambda_n = 0$ we can argue the following:

$$z^tAz = (Pz)^t \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} (Pz)$$

Hence, by letting $X = (X_1, \cdots, X_n)^t = PZ$

$$Z^tAZ = X^t \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} X = X_1^2 + \cdots + X_r^2, \quad X \sim N_n(0, PP^t) = N_n(0, I)$$

$$\therefore Z^tAZ = X_1^2 + \cdots + X_r^2, \quad X_i \overset{i.i.d}{\sim} N(0,1) \ (i=1, \cdots, r)$$

Thus from the representational definition of Chi-square distribution,

$$z^tAz \sim \chi^2(r)$$

* (63) 헌성가 실수인 대칭행렬의 대각화에 대한 정리는 부록 II에 추가적 있으며, $\det(A-\lambda I)=0$ 을 만족하는 λ 값을 행렬 A의 고유값 (eigen value)이라고 한다.

Ex 4.4.4.

In (theorem 4.2.7)-(a)'s one-way classification model, for $\bar{X}_i \sim N(\mu_i, \sigma^2/n_i)$ $(i=1, \cdots, K)$ the following sample distribution is introduced.

$$\sum_{i=1}^{K} n_i (\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2 / \sigma^2 \sim \chi^2(K-1)$$

Now, we wish to prove this using (Theorem 4.4.5).

Let $Y_i = \bar{X}_i - \bar{X} - (\mu_i - \bar{\mu})$ $(i=1, \cdots, K)$, $Y = (Y_1, \cdots, Y_r)^t$, $r = K-1$

$$\bar{X} = \sum_{i=1}^{K} n_i \bar{X}_i / n, \quad \bar{\mu} = \sum_{i=1}^{K} n_i \mu_i / n, \quad n = n_1 + \cdots + n_K$$

from the precedent $\bar{X}$ and $\bar{\mu}$ we can find the mean, variance and covariance of Y_i:

$$E[Y_i] = 0, \quad \mathrm{Var}(Y_i) = (n_i^{-1} - n^{-1})\sigma^2,$$

$$\mathrm{Cov}(Y_i, Y_j) = -n^{-1}\sigma^2 \ (i,j = 1, \cdots, K) \ (i \neq j)$$

* $E[Y_i] = E[\bar{X}_i] - E\left[\frac{1}{n}\sum_{i=1}^{n} n_i X_i\right] - E[\mu_i] + E\left[\frac{1}{n}\sum_{i=1}^{n} n_i \mu_i\right]$

$= \mu_i - \underbrace{\frac{1}{n}\sum_{i=1}^{n} n_i E[X_i]}_{\text{cancel out}} - \mu_i + \underbrace{\frac{1}{n}\sum_{i=1}^{n} n_i \mu_i}_{\text{cancel out}}$

$= 0$

(63)

$$\mathrm{Var}(Y_i) = E[Y_i^2] - E[Y_i]^2 = E[Y_i^2] = E[(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}))^2]$$
$$= E[(\bar{X}_i - \mu_i)^2 - 2(\bar{X}_i - \mu_i)(\bar{X} - \bar{\mu}) + (\bar{X} - \bar{\mu})^2]$$
$$(\text{or}) = E[(\bar{X}_i - \bar{X})^2 - 2(\bar{X}_i - \bar{X})(\mu_i - \bar{\mu}) + (\mu_i - \bar{\mu})^2]$$

$$\bar{X}_i - \bar{X} = \bar{X}_i - \frac{1}{n}\sum_{j=1}^{K} n_j \bar{X}_j = (1 - \frac{n_i}{n})\bar{X}_i - \frac{1}{n}\sum_{j \neq i}^{K} n_j \bar{X}_j$$

$$\mathrm{Var}(\bar{X}_i - \bar{X}) = \mathrm{Var}\left((1 - \frac{n_i}{n})\bar{X}_i - \frac{1}{n}\sum_{j \neq i}^{K} n_j \bar{X}_j\right)$$

$$= (1 - \frac{n_i}{n})^2 \mathrm{Var}(\bar{X}_i) + \sum_{j \neq i}^{K} \left(\frac{n_j}{n}\right)^2 \mathrm{Var}(\bar{X}_j)$$

$$= (1 - \frac{n_i}{n})^2 \cdot \frac{\sigma^2}{n_i} + \sum_{j \neq i}^{K} \frac{n_j^2}{n^2} \cdot \frac{\sigma^2}{n_j}$$

$$= (1 - \frac{2n_i}{n} + \frac{n_i^2}{n^2})\frac{\sigma^2}{n_i} + \sum_{j \neq i}^{K} \frac{n_j}{n^2}\sigma^2$$

$$= \frac{\sigma^2}{n_i} - \frac{2}{n}\sigma^2 + \sum_{j=1}^{K} \frac{n_j}{n^2}\sigma^2$$

$$= \frac{\sigma^2}{n_i} - \frac{2}{n}\sigma^2 + \frac{1}{n}\sigma^2 = (n_i^{-1} - n^{-1})\sigma^2$$

$$\mathrm{Var}(Y_i - Y_j) = \mathrm{Var}(Y_i) + \mathrm{Var}(Y_j) - 2\mathrm{Cov}(Y_i, Y_j)$$

$$\therefore \mathrm{Cov}(Y_i, Y_j) = \frac{1}{2}\left\{\mathrm{Var}(Y_i) + \mathrm{Var}(Y_j) - \mathrm{Var}(Y_i - Y_j)\right\}$$

$$\mathrm{Var}(Y_i - Y_j) = \mathrm{Var}(\bar{X}_i - \bar{X} - (\mu_i - \bar{\mu}) - \bar{X}_j + \bar{X} + (\mu_j - \bar{\mu}))$$
$$= \mathrm{Var}(\bar{X}_i - \bar{X}_j)$$
$$= \mathrm{Var}(\bar{X}_i) + \mathrm{Var}(\bar{X}_j)$$
$$= \sigma^2/n_i + \sigma^2/n_j$$

$$\mathrm{Var}(Y_i) = (n_i^{-1} + n^{-1})\sigma^2 \qquad \mathrm{Var}(Y_j) = (n_j^{-1} + n^{-1})\sigma^2$$

$$\therefore \mathrm{Var}(Y_i) + \mathrm{Var}(Y_j) - \mathrm{Var}(Y_i - Y_j) = (n_i^{-1} - n^{-1})\sigma^2 + (n_j^{-1} - n^{-1})\sigma^2 - (\sigma^2/n_i + \sigma^2/n_j)$$
$$= -2n^{-1}\sigma^2$$

$$\frac{1}{2}\left\{\mathrm{Var}(Y_i) + \mathrm{Var}(Y_j) - \mathrm{Var}(Y_i - Y_j)\right\} = -n^{-1}\sigma^2$$

$$\therefore \mathrm{Cov}(Y_i, Y_j) = -n^{-1}\sigma^2 \ \checkmark\checkmark$$

$$Y = (Y_1, \cdots, Y_r)^t \sim N(0, \Sigma), \quad \Sigma = \left[D(n_i^{-1}) - n^{-1}\mathbb{1}\mathbb{1}^t\right]\sigma^2$$

$$D(n_i^{-1}) = \begin{bmatrix} n_1^{-1} & 0 & \cdots & 0 \\ 0 & n_2^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_r^{-1} \end{bmatrix}, \quad \mathbb{1} = (1, \cdots, 1)^t$$

Meanwhile, (64)

$$\left[D(n_i^{-1}) - n^{-1}\mathbb{1}\mathbb{1}^t\right]^{-1} = \left[I - n^{-1}D^{-1}(n_i^{-1})\mathbb{1}\mathbb{1}^t\right]^{-1} D^{-1}(n_i^{-1})$$
$$= D^{-1}(n_i^{-1}) + n_K^{-1}(n_1, \cdots, n_r)^t(n_1, \cdots, n_r)$$

$$D(n_i) \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_r \end{bmatrix}$$

$$\therefore Y^t\Sigma^{-1}Y = \sum_{i=1}^{r} n_i Y_i^2/\sigma^2 + n_K^{-1}\left[\sum_{i=1}^{r} n_i Y_i\right]^2/\sigma^2 = \sum_{i=1}^{K} n_i Y_i^2/\sigma^2$$

Also from 〈Theorem 4.4.5〉 (a) $Y^t\Sigma^{-1}Y \sim \chi^2(r)$

$$\therefore \sum_{i=1}^{K} n_i(\bar{x}_i-\bar{x}-(\mu_i-\bar{\mu}))^2/\sigma^2 \sim \chi^2(K-1)$$

* 64) for vectors a,b if $1+b^ta\neq 0$, $(I+ab^t)^{-1}=I+cab^t$, $c=-1/(1+b^ta)$

Ex. 4.4.5 Sample Distribution of sample variance.

We know from Theorem 4.2.a that for $X_1,\cdots,X_n \overset{i.i.d}{\sim} N(\mu,\sigma^2)$, $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and $S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$ are mutually independent and $(n-1)S^2/\sigma^2\sim\chi^2(n-1)$.

Now, let's use the theorems from the multivariate normal distribution to prove it. Let $X=(X_1,\cdots,X_n)^t$ then,

$$X\sim N(\mu\mathbb{1},\sigma^2\mathbb{1}),\quad \mathbb{1}=(1,\cdots,1)^t$$

$$\bar{X}=n^{-1}\mathbb{1}^tX,\quad (n-1)S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2=X^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)X$$

Meanwhile, $I-n^{-1}\mathbb{1}\mathbb{1}^t$ is a symmetric matrix and $\mathbb{1}^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)=0$. Hence,

$$\mathrm{Cov}(\mathbb{1}^tX,(I-n^{-1}\mathbb{1}\mathbb{1}^t)X)=\mathbb{1}^t(\sigma^2I)(I-n^{-1}\mathbb{1}\mathbb{1}^t)^t=0$$

Therefore from 〈Theorem 4.4.3〉-(c), $\bar{X}=n^{-1}\mathbb{1}^tX$ and $(I-n^{-1}\mathbb{1}\mathbb{1}^t)X$ are independent. However,

$$(I-n^{-1}\mathbb{1}\mathbb{1}^t)^2=I-n^{-1}\mathbb{1}\mathbb{1}^t \quad \text{(idempotent)}$$

$$\therefore\ X^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)X=[(I-n^{-1}\mathbb{1}\mathbb{1}^t)X]^t[(I-n^{-1}\mathbb{1}\mathbb{1}^t)X]$$

Hence, $\bar{X}$ and $(n-1)S^2=X^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)X$ are independent. In addition,

$$X^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)X/\sigma^2=(X-\mu\mathbb{1})^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)(X-\mu\mathbb{1})/\sigma^2$$

$(X-\mu\mathbb{1})/\sigma\sim N_n(0,I_n)$, $(I-n^{-1}\mathbb{1}\mathbb{1}^t)^2=I-n^{-1}\mathbb{1}\mathbb{1}^t$, $\mathrm{trace}(I-n^{-1}\mathbb{1}\mathbb{1}^t)=n-1$

Thus, from 〈Theorem 4.4.5 (b)〉

$$(n-1)S^2/\sigma^2=X^t(I-n^{-1}\mathbb{1}\mathbb{1}^t)X/\sigma^2\sim\chi^2(n-1)\ \square$$

Definition: Linear Regression Model (error is normal)

$$\begin{cases} Y_i=x_{i0}\beta_0+x_{i1}\beta_1+\cdots+x_{ip}\beta_p+e_i \\ e_i\overset{i.i.d}{\sim}N(0,\sigma^2),\quad i=1,2,\cdots,n \end{cases}$$

To express the linear regression model using vectors and matrices, let:

$$Y=\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix},\quad X=\begin{pmatrix}x_{10} & x_{11} & \cdots & x_{1p}\\ \vdots & & & \\ x_{n0} & x_{n1} & \cdots & x_{np}\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{pmatrix},\quad e=\begin{pmatrix}e_1\\ \vdots\\ e_n\end{pmatrix}$$

we can simplify the model as:

$$\begin{cases} Y=X\beta+e \\ e\sim N_n(0,\sigma^2I) \end{cases}$$

Here, the matrix X is the explanatory variable matrix (설명변수의 행렬 X) and is rank $p+1$.[65] $\beta=(\beta_0,\beta_1,\cdots,\beta_p)^t$ which shows the specific shape of linear relation (선형관계의 구체적 형태) is called the regression coefficient and its estimation is:

$$\hat{\beta}=(X^TX)^{-1}X^TY$$

$\hat{\beta}$ is called the "sample regression coefficient (표본회귀계수)."

* 65) Definition and details of "rank" is in appendix Ⅱ.

Ex. 4.4.6. Sample Regression Coefficient in Simple Linear Regression.

In a linear regression model if $p=1$ and $x_0=1$, i.e.,

$$\begin{cases} Y_i=\beta_0+x_{i1}\beta_1+e_i \\ e_i\overset{iid}{\sim}N(0,\sigma^2),\quad i=1,2,\cdots,n \end{cases}$$

the model is called a "simple" linear regression model. In this case,

$$X=\begin{pmatrix}1 & x_{11}\\ \vdots & \vdots\\ 1 & x_{n1}\end{pmatrix},\quad X^tX=\begin{pmatrix}n & \sum x_{i1}\\ \sum x_{i1} & \sum x_{i1}^2\end{pmatrix},\quad X^tY=\begin{pmatrix}\sum Y_i\\ \sum x_{i1}Y_i\end{pmatrix}$$

$$(X^tX)^{-1}=\frac{1}{n\sum x_{i1}^2-(\sum x_{i1})^2}\begin{pmatrix}\sum x_{i1}^2 & -\sum x_{i1}\\ -\sum x_{i1} & n\end{pmatrix}$$

Hence, the sample regression coefficient $\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1)^t$ is given as:

$$\hat{\beta}_1=\frac{S_{xY}}{S_{xx}},\quad S_{xY}=\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)Y_i,\quad S_{xx}=\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2,\quad \bar{x}_1=\frac{1}{n}\sum_{i=1}^{n}x_{i1}$$

$$\hat{\beta}_0=\bar{Y}-\bar{x}_1\hat{\beta}_1,\ \text{i.e.,}\ \hat{\beta}_0+x_{i1}\hat{\beta}_1=\bar{Y}+(x_{i1}-\bar{x}_1)\hat{\beta}_1.$$

In the linear regression model, the estimation of σ^2 (variance of error) is:

$$\begin{aligned}\hat{\sigma}^2&=(Y-X\hat{\beta})^t(Y-X\hat{\beta})/(n-p-1)\\ &=\sum_{i=1}^{n}(Y_i-\hat{\beta}_0x_{i0}-\hat{\beta}_1x_{i1}-\cdots-\hat{\beta}_px_{ip})^2/(n-p-1)\end{aligned}$$

$\hat{\sigma}^2$ is called the "mean squared error sum of squares".
(Mean Squared ESS, 평균오차제곱합)

Ex. 4.4.7. Mean Squared error sum of squares in a Simple linear regression model

The mean squared error sum of squares of the simple linear regression model in Ex. 4.4.6 is:

$$\hat{\sigma}^2=\sum_{i=1}^{n}(Y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1})^2/(n-2)=\sum_{i=1}^{n}(Y_i-\bar{Y}-\hat{\beta}_1(x_{i1}-\bar{x}_1))^2/(n-2)$$

〈Theorem 4.4.6. Basic theorems on the sample distribution in Simple Linear regression models〉

Suppose that in the simple linear regression model:

$$\begin{cases} Y = X\beta + e \\ e \sim N_n(0, \sigma^2 I) \end{cases}, \quad X \in \mathbb{R}^{n\times(p+1)}$$

the following holds:

(a) $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$

(b) $\hat{\beta}$ and $\hat{\sigma}^2$ are independent

(c) $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1)$

(proof)

(a) $\hat{\beta} = (X^TX)^{-1}X^TY$, $Y \sim N(X\beta, \sigma^2 I)$

$\therefore \hat{\beta} \sim N((X^TX)^{-1}X^TX\beta, (X^TX)^{-1}X^T\sigma^2 I \cdot X(X^TX)^{-1})$
$\sim N(\beta, \sigma^2(X^TX)^{-1})$

(b) $\hat{\beta} = (X^TX)^{-1}X^TY$, $\hat{\sigma}^2 = (Y-X\hat{\beta})^t(Y-X\hat{\beta})/(n-p-1)$

$Y - X\hat{\beta}$ is also normal

$$\begin{aligned} Cov(\hat{\beta}, Y-X\hat{\beta}) &= (X^TX)^{-1}X^T Cov(Y, Y-X\hat{\beta}) \\ &= (X^TX)^{-1}X^T Cov(Y,Y) - (X^TX)^{-1}X^T Cov(Y, X(X^TX)^{-1}X^TY) \\ &= \sigma^2(X^TX)^{-1}X^T - \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \\ &= \sigma^2(X^TX)^{-1}X^T - \sigma^2(X^TX)^{-1}X^T \\ &= 0 \end{aligned}$$

$\therefore$ by 〈Theorem 4.4.3〉, $\hat{\beta}$ and $Y-X\hat{\beta}$ are independent.

We also know that if X and Y are independent, so is $g_1(X)$ and $g_2(Y)$.
$g_1(\cdot)$ and $g_2(\cdot)$ are arbitrary functions defined in $\mathbb{R}$.
let $g(X) = X^tX/(n-p-1)$. $g(Y-X\hat{\beta}) = (Y-X\hat{\beta})^t(Y-X\hat{\beta})/(n-p-1) = \hat{\sigma}^2$.
Thus, $\hat{\beta}$ and $\hat{\sigma}^2$ are independent. ✓

[Textbook]
Let $\Pi = X(X^tX)^{-1}X^t$. Π is symmetric and idempotent.
$\therefore \Pi^2 = \Pi$, $\Pi X = X$

Meanwhile, $Y - X\hat{\beta} = Y - \Pi Y = (I-\Pi)Y$

$\therefore Cov((X^tX)^{-1}X^tY, (I-\Pi)Y) = (X^tX)^{-1}X^t(\sigma^2 I)\cdot(I-\Pi)^t = 0$.

Thus from theorem 4.4.3 (c), $\hat{\beta} = (X^tX)^{-1}X^tY$ and $Y-X\hat{\beta} = (I-\Pi)Y$ are independent and from this we know that $\hat{\beta}$ and $\hat{\sigma}^2 = (Y-X\hat{\beta})^t(Y-X\hat{\beta})/(n-p-1)$ are independent.

(c) $$\begin{aligned} (n-p-1)\hat{\sigma}^2/\sigma^2 &= (Y-X\hat{\beta})^t(Y-X\hat{\beta})/\sigma^2 \\ &= [(I-\Pi)Y]^t[(I-\Pi)Y]/\sigma^2 \\ &= Y^t(I-\Pi)^t(I-\Pi)Y/\sigma^2 \\ &= Y^t(I-\Pi)^tY/\sigma^2 \\ &= (Y-X\beta)^t(I-\Pi)^t(Y-X\beta)/\sigma^2 \quad (\because (I-\Pi)X = 0) \\ &= \left(\frac{Y-X\beta}{\sigma}\right)^t(I-\Pi)^t\left(\frac{Y-X\beta}{\sigma}\right) \\ &= Z^t(I-\Pi)^tZ, \quad Z \sim N(0, I_n) \\ &\sim \chi^2(n-p-1) \quad \because \text{Theorem 4.4.5 (b)} \end{aligned}$$

$$\begin{aligned} rank(I-\Pi) &= trace(I-\Pi) \\ &= trace(I) - trace(\Pi) \\ &= n-(p+1) = n-p-1 \end{aligned}$$

□

Ex. 4.4.8. Sample Distribution of Simple Linear Regression Model.

From 〈Ex.4.4.6〉, 〈Ex.4.4.7〉 and 〈Theorem 4.4.6〉,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \frac{\sigma^2}{S_{xx}}\begin{pmatrix} \sum_{i=1}^n x_{i1}^2/n & -\bar{x}_1 \\ -\bar{x}_1 & 1 \end{pmatrix}\right)$$

$$(n-2)\hat{\sigma}^2/\sigma^2 = \sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta}_1(x_{i1}-\bar{x}_1))^2/\sigma^2 \sim \chi^2(n-2)$$

and $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Thus, from the representational definition of the t-distribution,

$$(\hat{\beta}_1 - \beta_1)/\sqrt{\hat{\sigma}^2/S_{xx}} = \frac{(\hat{\beta}_1-\beta_1)/\sqrt{\sigma^2/S_{xx}}}{\sqrt{(n-2)(\hat{\sigma}^2/\sigma^2)/(n-2)}} \sim t(n-2)$$

we can find mean squared error sum of square (ESS) using the following:

$$\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta}_1(x_{i1}-\bar{x}_1))^2/(n-2) = \left\{S_{YY} - (S_{xY})^2/S_{xx}\right\}/(n-2)$$

$$S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

* test for $H_0: \beta = \beta^*$ vs. $H_1: \beta \neq \beta^*$

다변량 정규분포의 성질

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

(2) $AN(\mu,\Sigma) + b \overset{d}{\equiv} N(A\mu+b, A\Sigma A^t)$

(3) $N(\mu_1,\Sigma) \oplus N(\mu_2,\Sigma_2) \overset{d}{\equiv} N(\mu_1+\mu_2, \Sigma_1+\Sigma_2)$

(4) if $X \sim N(\mu,\Sigma)$, $AX \perp BX \Leftrightarrow Cov(AX,BX) = A\Sigma B^t = 0$

(5) if $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$

$\Leftrightarrow$ (a) $X_1 \sim N(\mu_1,\Sigma_{11})$ and

(b) $X_2|X_1=x_1 \sim N(\mu_1 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1),\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$

(6) $N(0,I)^t A N(0,I) \sim \chi^2(r) \Leftrightarrow A^2 = A$, $r = trace(A)$