2 Reasons for the usefulness of Normal Distribution

1. It serves as a true population model in some instances
2. B/C of CLT, the sampling distributions of many multivariate statistics are approximately normal.

4.2 The Multivariate Normal Density and Its Properties

- The multivariate normal density is a generalization of the univariate normal density to $p \geq 2$ dimensions

$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \sim N(\mu, \sigma^2)$, and the power term of the exponent measures the square of the distance from x to μ in standard deviation unit.

$$\frac{(x-\mu)^2}{\sigma^2} = (x-\mu)(\sigma^2)^{-1}(x-\mu) = \underbrace{(x-\mu)\Sigma^{-1}(x-\mu)}_{\text{square of the generalized distance}}$$

Σ is a positive definite

- The multivariate normal density is obtained by replacing the univariate distance in (4-2) by the multivariate generalized distance.
- Probabilities in multivariate distributions are represented by volumes under the surface over regions defined by intervals of the X_i values.

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}, \quad -\infty < x < \infty, \quad i=1,2,\dots,p \sim N_p(\mu,\Sigma)$$

- Consider $p=2$ case.

$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \Rightarrow \Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{21} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$, because $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$,

$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$, so the squared distance $(x-\mu)'\Sigma^{-1}(x-\mu)$ becomes

$$(x-\mu)'\Sigma^{-1}(x-\mu) = [x_1-\mu_1, x_2-\mu_2]\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}\begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}$$

$$= \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]$$

이것을 위한 빌드업

Input the $|\Sigma|$ and Σ^{-1} in $f(x_1, x_2)$

$$f(x_1,x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

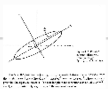
Constant Probability Density Contour $= \{\text{all } x \text{ such that } (x-\mu)'\Sigma^{-1}(x-\mu) = c^2\}$

$=$ surface of an ellipsoid centered at μ.

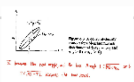
points of the tip of the axes

have axes $\pm c\sqrt{\lambda_i}\, e_i$, where $\Sigma e_i = \lambda_i e_i$

The axes of each ellipsoid of constant density are in the direction of the eigenvectors of Σ^{-1}, and their lengths are proportional to the reciprocals of the square roots of the eigenvalues of Σ^{-1}



origin의 차이



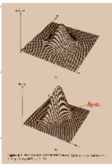
Theorem :

If Σ is positive definite, so that Σ^{-1} exists, then

$\Sigma e = \lambda e$ implies $\Sigma^{-1} e = \frac{1}{\lambda} e$

- We need to know that the constant $c^2 = \chi^2_p(\alpha)$, where $\chi^2_p(\alpha)$ is the upper $(100\alpha)^{th}$ percentile of a chi-square distribution with p degrees of freedom, which means contours contain $(1-\alpha)\times 100\%$ of the probability

$$(x-\mu)'\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)$$



$P(X \le \chi^2_p(\alpha)) = 0.95$, at $\alpha = 0.05$

- Remember the joint probability density function of p variables. This joint function has the largest p-value when the squared distance is 0, $x = \mu$, specifically. Thus, μ is the point of maximum density, or mode, as well as the expected value of X, or mean.

true for any random vector X having a multivariate normal distribution

Additional Properties of the Multivariate Normal Distribution

1. Linear combinations of the components of X are normally distributed.
2. All subsets of the components of X have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.

Properties of Multivariate Normal Distributions (Extended Explanations)

- If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $a'X = a_1X_1 + a_2X_2 + \cdots a_pX_p$ is distributed as $N(a'\mu, a'\Sigma a)$. Also, if $a'X$ is distributed as $N(a'\mu, a'\Sigma a)$ for every a, then $X \sim N_p(\mu, \Sigma)$

- If X is distributed as $N_p(\mu, \Sigma)$, the q linear combinations

$$\underset{(q\times p)(p\times 1)}{A\,X} = \begin{bmatrix} a_{11}X_1 + \cdots\cdots a_{1p}X_p \\ a_{21}X_1 + \cdots\cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots\cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_p(A\mu, A\Sigma A')$. Also, $\underset{(p\times 1)}{X} + \underset{(p\times 1)}{d}$, where d is a vector of constants, is distributed as $N_p(\mu + d, \Sigma)$

- $\underset{(p\times1)}{X} = \begin{bmatrix} \underset{(q\times1)}{X_1} \\ \hline \underset{(p-q)\times1}{X_2} \end{bmatrix}$, $\underset{(p\times1)}{\mu} = \begin{bmatrix} \underset{(q\times1)}{\mu_1} \\ \hline \underset{(p-q)\times1}{\mu_2} \end{bmatrix}$, $\underset{(p\times p)}{\Sigma} = \left[\begin{array}{c|c} \underset{(q\times q)}{\Sigma_{11}} & \underset{q\times(p-q)}{\Sigma_{12}} \\ \hline \underset{(p\times q)\times q}{\Sigma_{21}} & \underset{(p\times q)\times(p\times q)}{\Sigma_{22}} \end{array}\right]$

then X_1 is distributed as $N_p(\mu_1, \Sigma_{11})$

- zero correlation = statistical independence

1. If $q_1\times 1$ matrix X_1 and $q_2\times 1$ matrix X_2 are independent, then $Cov(X_1, X_2) = 0$, a $q_1\times q_2$ matrix of zeros

2. If $\begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$, then X_1 and X_2 are independent iff $\Sigma_{12} = 0$

3. If X_1 and X_2 are independent and are distributed as $N_{q_1}(\mu_1, \Sigma_{11})$ and $N_{q_2}(\mu_2, \Sigma_{22})$, then $\begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix}$ has the multivariate normal distribution $N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & 0 \\ \hline 0 & \Sigma_{22} \end{array}\right]\right)$

- Let $X = \begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix}$ be distributed as $N_p(\mu, \Sigma)$ with $\mu = \begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}$, $\Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$, and $|\Sigma_{22}| > 0$.

Then the conditional distribution of X_1, given that $X_2 = x_2$, is normal and has

Mean $= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ and Covariance $= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Note that the covariance does not depend on the value x_2 of the conditioning variable.

Theorem 2

Suppose that

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{(1)} \\ \hline \boldsymbol{X}_{(2)} \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_{(1)} \\ \hline \boldsymbol{\mu}_{(2)} \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{(11)} & \Sigma_{(12)} \\ \hline \Sigma_{(21)} & \Sigma_{(22)} \end{array}\right]\right).$$

Then, the following properties hold:

1. marginally normal
$\boldsymbol{X}_{(1)}$ *and* $\boldsymbol{X}_{(2)}$ *are independent if and only if* $\Sigma_{(12)} = \boldsymbol{0}_{q_1\times q_2}$.
2. $\boldsymbol{X}_{(j)} \sim N_{q_j}(\boldsymbol{\mu}_{(j)}, \Sigma_{(jj)})$ *for* $j = 1, 2$.
3. $\boldsymbol{X}_{(1)} | \boldsymbol{X}_{(2)} = \boldsymbol{x}_{(2)} \sim N_{q_1}(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2})$
where $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_{(1)} + \Sigma_{(12)}\Sigma_{(22)}^{-1}(\boldsymbol{x}_{(2)} - \boldsymbol{\mu}_{(2)})$ *and*
$\Sigma_{1|2} = \Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}$.

확대하면 보이니까 Example 4.7 못 알아듣겠음 ....

$$(\text{statistical distance})^2 = x A x = \sum_{i=1}^{k} \lambda_i e_i e_i'$$

let $V = \operatorname{diag}(\sigma_{11}, \sigma_{22}, \cdots \sigma_{kk})$

Σ = variance-covariance matrix

P = correlation matrix

$$V^{\frac{1}{2}} P V^{\frac{1}{2}} = \Sigma$$

$$V^{-\frac{1}{2}} \Sigma V^{-\frac{1}{2}} = P$$

$$\frac{d_i' d_k}{L_{d_i} L_{d_k}} = \cos(\theta_{ik}) = r_{ik} = cor(i,k)$$

$$= \frac{x'y}{L_x L_y} = \cos(\theta)$$

$$\begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix} \approx (25-\lambda)\{(4-\lambda)(9-\lambda)-1\} + 2\{-2(9-\lambda)-4\} + 4\{4(4-\lambda)+2\}$$

$$= (25-\lambda)\{35-13\lambda+\lambda^2\} + 2\{-22+2\lambda\} + 4\{18-4\lambda\}$$

$$= ($$

For the multivariate normal situation, it is worth emphasizing the following:

1. All conditional distributions are (multivariate) normal.
2. The conditional mean is of the form

$$\mu_1 + \beta_{1,q+1}(X_{q+1} - \mu_{q+1}) + \cdots + \beta_{1,p}(X_p - \mu_p)$$

$$\vdots$$

$$\mu_q + \beta_{q,q+1}(X_{q+1} - \mu_{q+1}) + \cdots + \beta_{q,p}(X_p - \mu_p)$$

where the β's are defined by $\Sigma_{12}\Sigma_{22}^{-1} = \begin{bmatrix} \beta_{1,q+1} & \beta_{1,q+2} & \cdots & \beta_{1,p} \\ \beta_{2,q+1} & \beta_{2,q+2} & \cdots & \beta_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q,q+1} & \beta_{q,q+2} & \cdots & \beta_{q,p} \end{bmatrix}$

3. The conditional covariance, $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, does not depend upon the value(s) of the conditioning variable(s).

- Let X be distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| > 0$. Then,
  1. $(X-\mu)'\Sigma^{-1}(X-\mu)$ is distributed as χ^2_p, where χ^2 denotes the chi-square distribution with p degrees of freedom.
  2. The $N_p(\mu, \Sigma)$ distribution assigns probability $1-\alpha$ to the solid ellipsoid $\{X : (X-\mu)'\Sigma^{-1}(X-\mu) \le \chi^2_p(\alpha)\}$, where $\chi^2_p(\alpha)$ denotes the upper $(100\alpha)^{th}$ percentile of the χ^2_p distribution

- Let $X_1, X_2, \ldots, X_n$ be mutually independent with X_j distributed as $N_p(\mu_j, \Sigma)$. Then $V_1 = c_1X_1 + c_2X_2 + \cdots + c_nX_n$ is distributed as $N_p\left(\sum_{j=1}^{n} c_j\mu_j, \left(\sum_{j=1}^{n} c_j^2\right)\Sigma\right)$. Moreover, V_1 and $V_2 = b_1X_1 + b_2X_2 + \cdots + b_nX_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} \left(\sum_{j=1}^{n} c_j^2\right)\Sigma & (b'c)\Sigma \\ (b'c)\Sigma & \left(\sum_{j=1}^{n} b_j^2\right)\Sigma \end{bmatrix}$$

Consequently, V_1 and V_2 are independent if $b'c = \sum_{j=1}^{n} c_jb_j = 0$

4.3 Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation

The Multivariate Normal Likelihood

Maximum Likelihood Estimates

- selecting the parameter values that maximize the joint density evaluated at the observations

* Let A be a $k\times k$ symmetric and X be a $k\times 1$ vector. Then

$x'Ax = tr(x'Ax) = tr(Axx')$

$tr(A) = \sum_{i=1}^{k} \lambda_i$

- Consider the joint density function

$$\left\{\begin{matrix}\text{Joint density} \\ \text{of } \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\end{matrix}\right\} = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} e^{-(\mathbf{x}_j-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu})/2} \right\}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\mathbf{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu})/2}$$

, and now,

the exponent in there can be simplified.

$$(X_j-\mu)'\Sigma^{-1}(X_j-\mu) = tr[(X_j-\mu)'\Sigma^{-1}(X_j-\mu)] = tr[\Sigma^{-1}(X_j-\mu)(X_j-\mu)'], \text{ so}$$

$$\sum_{j=1}^{n}(X_j-\mu)'\Sigma^{-1}(X_j-\mu) = \sum_{j=1}^{n} tr[(X_j-\mu)'\Sigma^{-1}(X_j-\mu)] = \sum_{j=1}^{n} tr[\Sigma^{-1}(X_j-\mu)(X_j-\mu)']$$

$$= tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\mu)(X_j-\mu)'\right)\right]$$

Let's break this one down. Add and subtract $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} X_j$ in each $(X_j-\mu)$, we get

$$\sum_{j=1}^{n}(X_j-\bar{x}+\bar{x}-\mu)(X_j-\bar{x}+\bar{x}-\mu)' = \sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})' + \sum_{j=1}^{n}(\bar{x}-\mu)(\bar{x}-\mu)'$$

$$= \sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)', \text{ so the joint density}$$

can be written as

$$= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}} \cdot \exp\left\{-tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]/2\right\}$$

$$= L(\mu, \Sigma)$$

※ We can also express the exponential term as,

$$tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]$$

$$= tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})'\right)\right] + n \cdot tr\left[\Sigma^{-1}(\bar{x}-\mu)(\bar{x}-\mu)'\right]$$

$$= tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\bar{x})(X_j-\bar{x})'\right)\right] + n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

Maximum Likelihood Estimation of μ and Σ

- Given a $p\times p$ symmetric positive definite matrix B and a scalar $b>0$, it follows that,

$$\frac{1}{|\Sigma|^b} e^{-tr(\Sigma^{-1}B)/2} \le \frac{1}{|B|^b}(2b)^{pb}e^{-bp},$$

for all positive definite Σ, with equality holding only for $\Sigma = (1/2b)B$.

Proof. Let $\mathbf{B}^{1/2}$ be the symmetric square root of $\mathbf{B}$ [see Equation (2-22)], so $\mathbf{B}^{1/2}\mathbf{B}^{1/2} = \mathbf{B}$, $\mathbf{B}^{1/2}\mathbf{B}^{-1/2} = \mathbf{I}$, and $\mathbf{B}^{-1/2}\mathbf{B}^{-1/2} = \mathbf{B}^{-1}$. Then $\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{B}) = \text{tr}[(\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2})\mathbf{B}^{1/2}] = \text{tr}[\mathbf{B}^{1/2}(\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2})]$. Let η be an eigenvalue of $\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}$. This matrix is positive definite because $\mathbf{y}'\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}\mathbf{y} = (\mathbf{B}^{1/2}\mathbf{y})'\mathbf{\Sigma}^{-1}(\mathbf{B}^{1/2}\mathbf{y}) > 0$ if $\mathbf{B}^{1/2}\mathbf{y} \ne 0$ or, equivalently, $\mathbf{y} \ne \mathbf{0}$. Thus, the eigenvalues η_i of $\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}$ are positive by Exercise 2.17. Result 4.9(b) then gives

$$\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{B}) = \text{tr}(\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}) = \sum_{i=1}^{p}\eta_i$$

and $|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}| = \prod_{i=1}^{p}\eta_i$ by Exercise 2.12. From the properties of determinants in Result 2A.11, we can write

$$|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}| = |\mathbf{B}^{1/2}||\mathbf{\Sigma}^{-1}||\mathbf{B}^{1/2}| = |\mathbf{\Sigma}^{-1}||\mathbf{B}^{1/2}||\mathbf{B}^{1/2}|$$
$$= |\mathbf{\Sigma}^{-1}||\mathbf{B}| = \frac{1}{|\mathbf{\Sigma}|}|\mathbf{B}|$$

or

$$\frac{1}{|\mathbf{\Sigma}|} = \frac{|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}|}{|\mathbf{B}|} = \frac{\prod_{i=1}^{p}\eta_i}{|\mathbf{B}|}$$

Combining the results for the trace and the determinant yields

$$\frac{1}{|\mathbf{\Sigma}|^b}e^{-\text{tr}[\mathbf{\Sigma}^{-1}\mathbf{B}]/2} = \frac{\left(\prod_{i=1}^{p}\eta_i\right)^b}{|\mathbf{B}|^b}e^{-\sum_{i=1}^{p}\eta_i/2} = \frac{1}{|\mathbf{B}|^b}\prod_{i=1}^{p}\eta_i^b e^{-\eta_i/2}$$

But the function $\eta^b e^{-\eta/2}$ has a maximum, with respect to η, of $(2b)^b e^{-b}$, occurring at $\eta = 2b$. The choice $\eta_i = 2b$, for each i, therefore gives

$$\frac{1}{|\mathbf{\Sigma}|^b}e^{-\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{B})/2} \le \frac{1}{|\mathbf{B}|^b}(2b)^{pb}e^{-bp}$$

The upper bound is uniquely attained when $\mathbf{\Sigma} = (1/2b)\mathbf{B}$, since, for this choice,

$$\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2} = \mathbf{B}^{1/2}(2b)\mathbf{B}^{-1}\mathbf{B}^{1/2} = (2b)\underset{(p\times p)}{\mathbf{I}}$$

and

$$\text{tr}[\mathbf{\Sigma}^{-1}\mathbf{B}] = \text{tr}[\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}] = \text{tr}[(2b)\mathbf{I}] = 2bp$$

Moreover,

$$\frac{1}{|\mathbf{\Sigma}|} = \frac{|\mathbf{B}^{1/2}\mathbf{\Sigma}^{-1}\mathbf{B}^{1/2}|}{|\mathbf{B}|} = \frac{|(2b)\mathbf{I}|}{|\mathbf{B}|} = \frac{(2b)^p}{|\mathbf{B}|}$$

Straightforward substitution for $\text{tr}[\mathbf{\Sigma}^{-1}\mathbf{B}]$ and $1/|\mathbf{\Sigma}|^b$ yields the bound asserted. ∎

확대해서 한번씩 체크
(이어지는거임)

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\boldsymbol{\Sigma}}|^{n/2}} \tag{4-18}$$

or, since $|\hat{\boldsymbol{\Sigma}}| = [(n-1)/n]^p |\mathbf{S}|$,

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \text{constant} \times (\text{generalized variance})^{-n/2} \tag{4-19}$$

The generalized variance determines the "peakedness" of the likelihood function and, consequently, is a natural measure of variability when the parent population is multivariate normal.

Maximum likelihood estimators possess an *invariance property*. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$, and consider estimating the parameter $h(\boldsymbol{\theta})$, which is a function of $\boldsymbol{\theta}$. Then the *maximum likelihood estimate* of

$$\underset{\text{(a function of }\boldsymbol{\theta})}{h(\boldsymbol{\theta})} \quad \text{is given by} \quad \underset{\text{(same function of }\hat{\boldsymbol{\theta}})}{h(\hat{\boldsymbol{\theta}})} \tag{4-20}$$

Sufficient Statistics

- $\bar{X}$ and S are sufficient statistics, meaning they have all of the information about μ and Σ in the data matrix X

4.4 The Sampling Distribution of $\bar{X}$ and S

Wishart Distribution:

- the sampling distribution of the sample covariance matrix

defined as the sum of independent products of multivariate normal random vectors

$W_m(\cdot \mid \Sigma)$ = Wishart Distribution with m d.f.

= distribution of $\sum_{j=1}^{m} Z_j Z_j'$

1. $\bar{X}$ is distributed as $N_p(\mu, \frac{1}{n}\Sigma)$
2. $(n-1)S$ is distributed as a Wishart random matrix with $n-1$ df.
3. $\bar{X}$ and S are independent

Properties of the Wishart Distribution

1. If A_1 is distributed as $W_{m_1}(A_1 \mid \Sigma)$ independently of A_2, which is distributed as $W_{m_2}(A_2 \mid \Sigma)$, then $A_1 + A_2$ is distributed as $W_{m_1+m_2}(A_1 + A_2 \mid \Sigma)$.
2. If A is distributed as $W_m(A \mid \Sigma)$, then CAC' is distributed as $W_m(CAC' \mid C\Sigma C')$.

As long as $n > p$, the pdf value of a Wishart distribution at the positive definite matrix A is,

$$W_{n-1}(A \mid \Sigma) = \frac{|A|^{(n-p-2)/2} \cdot e^{-tr[A\Sigma^{-1}]/2}}{2^{p(n-1)/2} \cdot \pi^{p(p-1)/4} \cdot |\Sigma|^{(n-1)/2} \cdot \prod_{i=1}^{p} \Gamma(\frac{1}{2}(n-i))}$$

4.5 Large - Sample Behavior of $\bar{X}$ and S

뭐야... 뭘 말하고 싶은거야...

Q-Q plots

4.6 Assessing the Assumption of Normality

Evaluating the Normality of the Univariate Marginal Distributions

$$|\hat{p}_{i1} - .683| > 3\sqrt{\frac{(.683)(.317)}{n}} = \frac{1.396}{\sqrt{n}}$$

$$|\hat{p}_{i2} - .954| > 3\sqrt{\frac{(.954)(.046)}{n}} = \frac{.628}{\sqrt{n}}$$

Evaluating Bivariate Normality

- Remember,

$(x-\mu)'\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)$, we consider the case where $p=2$, $\alpha = 0.5$,

Calculate the above formula and count the number of pairs that satisfies the inequality.

If the aggregated number is less than the α, this means that, if the observations were normally distributed, we would expect about $100 \cdot (\alpha)$ of them to be within this contour.

but ! different procedure is recommended.

Squared Generalized Distances

$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x})$, $j = 1, 2, \dots, n$

When the parent population is multivariate normal and both n and $n-p$ are greater than 25 or 30, each of the squared distances should behave like a χ^2 random variable.

Drawing χ^2 plot

1. order the squared distances from smallest to largest.
2. graph the pairs $(\underbrace{q_{c,p}((j-\frac{1}{2})/n)}, d_j^2)$

100 $(j-\frac{1}{2})/n$ quantile of the χ^2 distribution with p d.f.

$q_{c,p}((j-\frac{1}{2})/n) = \chi^2_p((n-j+\frac{1}{2})/n)$

4.7 Detecting Outliers and Cleaning Data