Inference for sampling from MVN (multivariate normal) population:

→ Let $\underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $N_p(\underline{\mu}, \Sigma)$

Parameter Space

$$\theta := (\mu, \Sigma) \in \mathbb{R}^p \times \mathbb{R}^{\frac{p(p+1)}{2}}, \quad \Sigma > 0$$

The joint pdf is

$$f_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{\mu, \Sigma}(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\underline{x}_i - \underline{\mu})^T \Sigma^{-1} (x_i - \mu) \right\}$$

$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\underline{x}_i - \underline{\mu})^T \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right\}$$

Observe

$$\sum_{i=1}^{n} (\underline{x}_i - \underline{\mu})^T \Sigma^{-1} (\underline{x}_i - \underline{\mu})$$

$$= \sum_{i=1}^{n} (\underline{x}_i - \underline{x} + \underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x}_i - \underline{x} + \underline{x} - \underline{\mu})$$

$$= \sum_{i=1}^{n} (\underline{x}_i - \underline{x})^T \Sigma^{-1} (\underline{x}_i - \underline{x}) + n (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})$$

$$+ 2 \sum_{i=1}^{n} (\underline{x}_i - \underline{x})^T \Sigma^{-1} (\underline{x} - \underline{\mu})$$ (this term is zero)

$$\therefore f_{X_1 \ldots X_n}(x_1, \ldots, x_n) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{ \frac{-1}{2} \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^T \Sigma^{-1} (\underline{x}_i - \underline{\bar{x}}) - \frac{n}{2} (\underline{\bar{x}} - \underline{\mu})^T \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}) \right\}$$

Sufficient Statistic:

Neyman factorization theorem

$$T := T(X_1 \ldots X_n)$$

$$\text{If} \quad f_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \overbrace{g(x_1 \ldots x_n)}^{\text{parameter independent}} \, h(T, \theta)$$

Then T is a sufficient statistic

[when μ & Σ are both unknown]

$$f_{X_1 \ldots X_n}(x_1 \ldots x_n) = g(x_1 \ldots x_n) \; h(\underline{\bar{x}}, S_n, \theta)$$

$$f_{X_1 \ldots X_n}(x_1 \ldots x_n) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^T \Sigma^{-1} (x_i - \bar{x}) - \frac{n}{2} (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu) \right]$$

<u>Observe</u>

$$\sum_{i=1}^{n} (x_i - \bar{x})^T \Sigma^{-1} (x_i - \bar{x})$$

$$= tr\left(\Sigma^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \right)$$

$$= tr\left(\Sigma^{-1} S_n \right)$$

$$\therefore \quad f_{X_1 \ldots X_n}(x_1 \ldots x_n) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{ -\frac{n}{2} tr(\Sigma^{-1} S_n) - \frac{n}{2} (\bar{x}-\mu)^T \Sigma^{-1} (\bar{x}-\mu) \right\}$$

$\Rightarrow (\bar{x}, S_n)$ is sufficient for (μ, Σ)

When Σ is known

$$f_{X_1 \ldots X_n}(x_1 \ldots x_n) = g(x_1 \ldots x_n)\, h(\bar{x}, \mu)$$

$\Rightarrow \bar{x}$ is sufficient for μ

When μ is known

H/w

Minimal Sufficient Statistic (MSS)

Def^n T is MSS if it is function of every other sufficient statistic

Characterization Suppose T is a sufficient statistic

Let $(X_1 \ldots X_n)$ & $(Y_1 \ldots Y_n)$ be two samples for the pop.

$\dfrac{f_{X_1 \ldots X_n}(x_1 \ldots x_n)}{f_{Y_1 \ldots Y_n}(y_1 \ldots y_n)}$ is independent of the parameter iff $T(x_1 \ldots x_n) = T(y_1 \ldots y_n)$

then T is a MSS

Unbiasedness

$\rightarrow$ $E(\bar{x}) = \mu$

$E(S_{n-1}) = \Sigma$

Maximum Likelihood Estimation

$$\mathcal{L}(\mu, \Sigma) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{ -\frac{n}{2} tr(\Sigma^{-1} S_n) - \frac{n}{2} (\bar{\underline{x}} - \underline{\mu})^T \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) \right\}$$

we want to maximize $\mathcal{L}(\mu, \Sigma)$ for (μ, Σ)

Suppose $\Sigma > 0$ is fixed

$\mathcal{L} := \mathcal{L}(\mu, \Sigma)$ is maximum iff $(\bar{\underline{x}} - \underline{\mu})^T \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu})$ is minimum

$\Leftrightarrow (\bar{\underline{x}} - \underline{\mu}) = 0 \Leftrightarrow \underline{\mu} = \bar{\underline{x}}$

$\Rightarrow \hat{\mu} = \hat{\mu}_{MLE} = \bar{\underline{x}}$

Likelihood function at $\mu = \hat{\mu}$

$$\mathcal{L}(\bar{\underline{x}}, \Sigma) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{n}{2} tr(\Sigma^{-1} S_n) \right]$$

$$LL(\Sigma) = \log \mathcal{L}(\bar{\underline{x}}, \Sigma) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log|\Sigma| - \frac{n}{2} tr(\Sigma^{-1} S_n)$$

we have to maximize $LL(\Sigma)$ wrt Σ

$\Leftrightarrow$ $-\frac{n}{2} \log|\Sigma| - \frac{n}{2} tr(\Sigma^{-1} S_n)$ is maximized wrt Σ

$\Leftrightarrow$ $\frac{n}{2} \log|\Sigma^{-1}| - \frac{n}{2} tr(\Sigma^{-1} S_n)$

$\Leftrightarrow$ $\frac{n}{2} \log|\Sigma^{-1} S_n| - \frac{n}{2} tr(\Sigma^{-1} S_n) - \frac{n}{2} \log|S_n|$

$\Leftrightarrow \quad \log|\Sigma^{-1}S_n| - tr(\Sigma^{-1}S_n)$ maximized wrt Σ

→ $\Sigma^{-1}S_n > 0$

Let $\lambda_1, \ldots, \lambda_p$ are eigen values of $\Sigma^{-1}S_n$

$$|\Sigma^{-1}S_n| = \lambda_1 \ldots \lambda_p$$
$$tr(\Sigma^{-1}S_n) = \sum_{i=1}^{p} \lambda_i$$

$$\psi(\lambda_1 \ldots \lambda_p) = \log|\Sigma^{-1}S_n| - tr(\Sigma^{-1}S_n)$$

$$= \sum_{i=1}^{p} \log \lambda_i - \sum_{i=1}^{p} \lambda_i$$

$$\frac{\partial \psi}{\partial \lambda_i} = \frac{1}{\lambda_i} - 1 = 0 \Rightarrow \lambda_i = 1$$

$$\left.\frac{\partial^2 \psi}{\partial \lambda_i^2}\right|_{\lambda_i = 1} = -1 < 0$$

$\therefore \psi$ is maximized at $\lambda_i = 1$, $i = 1, 2, \ldots, p$

$$\Sigma^{-1}S_n = PIP^T = I$$
$$\Rightarrow \hat{\Sigma} = S_n$$

(Read about Jensen's Inequality, AM-GM)

→ MLE of μ is unbiased

→ MLE of Σ is biased