1 Lemma 4 (Lemma 3.2.1)

Let $x_1, \ldots, x_N$ be N (p-component) vectors, and let $\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha$. Then for any vector b, we have

$$\sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)^T = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})^T + N(\bar{x} - b)(\bar{x} - b)^T.$$

Proof.

$$\begin{aligned}
&\sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)^T \\
&= \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x}) + (\bar{x} - b)][(x_\alpha - \bar{x}) + (\bar{x} - b)]^T \\
&= \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x})(x_\alpha - \bar{x})^T + (x_\alpha - \bar{x})\boxed{(\bar{x} - b)}^T + (\bar{x} - b)(x_\alpha - \bar{x})^T + (\bar{x} - b)(\bar{x} - b)^T] \\
&= \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})^T + \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(\bar{x} - b)^T + (\bar{x} - b) \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})^T + N\boxed{(\bar{x} - b)(\bar{x} - b)}^T.
\end{aligned}$$

Note that $\boxed{\sum_{\alpha=1}^{N} (x_\alpha - \bar{x}) = \sum_{\alpha=1}^{N} x_\alpha - N\bar{x} = 0,}$
the second and third terms on the right hand side are 0. □

Let $b = 0$, we have

$$\sum_{\alpha=1}^{N} x_\alpha x_\alpha^T = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})^T + N\bar{x}\bar{x}^T.$$

That is

$$\sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})^T = \sum_{\alpha=1}^{N} x_\alpha x_\alpha^T - N\bar{x}\bar{x}^T.$$

2 Assertion 5 (Lemma 3.2.3 and Corollary 3.2.1)

Assertion 5 The maximum likelihood estimators of functions of the parameters are those functions of the maximum likelihood estimators of the parameters.
Corollary 3.2.1 If on the basis of a given sample $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are maximum likelihood estimators of the parameters $\theta_1, \ldots, \theta_m$ of a distribution, and if the transformation from $\theta_1, \ldots, \theta_m$ to $\phi_1, \ldots, \phi_m$ is one to one, then $\phi_1(\hat{\theta}_1, \ldots, \hat{\theta}_m), \ldots, \phi_m(\hat{\theta}_1, \ldots, \hat{\theta}_m)$ are maximum likelihood estimators of $\phi_1(\theta_1, \ldots, \theta_m), \ldots, \phi_m(\theta_1, \ldots, \theta_m)$.
If the estimators of $\theta_1, \ldots, \theta_m$ are unique, then the estimators of $\phi_1, \ldots, \phi_m$ are unique.
Lemma 3.2.3 Let $f(\theta)$ be a real-valued function defined on a set S, and let ϕ be a single-valued function, with a single-valued inverse, on S to a set S^*; that is, to each $\theta \in S$ there

corresponds a unique $\theta^* \in S^*$, and, conversely, to each $\theta^* \in S^*$ there corresponds a unique $\theta \in S$. Let $g(\theta^*) = f[\phi^{-1}(\theta^*)]$. Then,
(1) If $f(\theta)$ attains a maximum at $\theta = \theta_0$, $g(\theta^*)$ attains a maximum at $\theta^* = \theta_0^* = \phi(\theta_0)$.
(2) If the maximum of $f(\theta)$ at θ_0 is unique, so is the maximum of $g(\theta^*)$ at θ_0^*.

Proof. (1) $f(\theta)$ attains a maximum at $\theta = \theta_0$, that is, $f(\theta_0) \geq f(\theta)$ for all $\theta \in S$. For any $\theta^* \in S^*$,

$$g(\theta^*) = f[\phi^{-1}(\theta^*)] = f(\theta) \leq f(\theta_0) = f(\phi^{-1}(\theta_0^*)) = g(\theta_0^*).$$

That is, $g(\theta^*)$ attains a maximum at θ_0^*.
(2) If the maximum of $f(\theta)$ at θ_0 is unique, there is strict inequality above for $\theta \neq \theta_0$, and the maximum of $g(\theta^*)$ is unique. □

Theorem 3 (Theorem 3.2.1)

If $x_1, x_2, \ldots, x_N$ constitute a sample from $N(\mu, \Sigma)$ with $p < N$, the maximum likelihood estimator of μ and Σ are

$$\boxed{\hat{\mu}} = \bar{x} = \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha, \tag{4}$$

and

$$\boxed{\hat{\Sigma}} = \frac{1}{N}\sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})^T, \tag{5}$$

respectively.

Lemma 6 (Corollary 3.2.2)

If $x_1, x_2, \ldots, x_N$ constitutes a sample from $N(\mu, \Sigma)$, where $\sigma_{ij} = \sigma_i\sigma_j\rho_{ij}$, and $\rho_{ii} = 1$, then the maximum likelihood estimator of μ, σ_i^2, and ρ_{ij} are:

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha,$$

$$\hat{\sigma}_i^2 = \frac{1}{N}\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2,$$

and

$$\hat{\rho}_{ij} = \frac{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2}\sqrt{\sum_{\alpha=1}^{N}(x_{j\alpha} - \bar{x}_j)^2},$$

respectively.

3 Lemma 8 (Lemma 3.3.1)

Lemma 8 If $C = (c_{\alpha\beta})$ is orthogonal (that is $C^T C = CC^T = \boxed{I_N}$), then $\boxed{\sum_{\alpha=1}^{N} x_\alpha x_\alpha^T = \sum_{\alpha=1}^{N} y_\alpha y_\alpha^T,}$ where $y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} x_\beta$, for $\alpha = 1, \ldots, N$.

Proof. Let $X = (x_1, \ldots, x_N)$ and $Y = (y_1, \ldots, y_N)$. Note that $X, Y \in \mathbb{R}^{p \times N}$.

$$Y = (y_1, \ldots, y_N) = (\sum_{\beta=1}^{N} c_{1\beta} x_\beta, \ldots, \sum_{\beta=1}^{N} c_{N\beta} x_\beta) = (x_1, \ldots, x_N) \begin{pmatrix} c_{11} & c_{21} & \ldots & c_{N1} \\ c_{12} & c_{22} & \ldots & c_{N2} \\ \ldots & \ldots & \ldots & \ldots \\ c_{1N} & c_{2N} & \ldots & c_{NN} \end{pmatrix}$$

That is $Y = XC^T$.

$$\sum_{\alpha=1}^{N} y_\alpha y_\alpha^T = YY^T = XC^T(XC^T)^T = XC^T CX^T = XX^T = \sum_{\alpha=1}^{N} x_\alpha x_\alpha^T.$$

□

4 Unbiasedness

Definition 10 (Definition 3.3.1)

An estimator t of a parameter vector θ, is unbiased if and only if $\mathbb{E}_\theta(t) = \theta$.

- We further claim that $\bar{X}$ in (4) is an unbiased estimator of the true mean (i.e., population mean),

$$\mathbb{E}(\bar{X}) = \frac{1}{N}\mathbb{E}(\sum_{\alpha=1}^{N} X_\alpha) = \mu.$$

- $\hat{\Sigma}$ in (5) is a biased estimator of Σ,

$$\mathbb{E}(\hat{\Sigma}) = \frac{1}{N}\mathbb{E}(\sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha^T) = \frac{N-1}{N}\Sigma.$$

- S (defined in (2) and (3)) is an unbiased estimator of Σ,

$$S = \frac{1}{N-1}A = \frac{1}{N-1}\sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})^T.$$

5 Sufficiency of maximum likelihood estimators

Sufficiency A statistic T is sufficient for a family of distributions of X or for a parameter θ if the conditional distribution of X given $T = t$ does not depend on θ.
The sufficient statistic T gives as much information about θ as the entire sample X.

Factorization Theorem A statistic $t(y)$ is sufficient for θ if and only if the density $f(y \mid \theta)$ can be factored as

$$f(y \mid \theta) = g[t(y), \theta] h(y),$$

where $g[t(y), \theta]$ and $h(y)$ are nonnegative and $h(y)$ does not depend on θ.

Theorem 3.4.1

(1) If $x_1, \ldots, x_N$ are observations from $N(\mu, \Sigma)$, then $\bar{x}$ and S are sufficient for μ and Σ.

(2) If μ is given, $\sum_{\alpha=1}^{N} (x_\alpha - \mu)(x_\alpha - \mu)^T$ is sufficient for Σ.

(3) If Σ is given, $\bar{x}$ is sufficient for μ.

(4) Note that if μ is given, S is not sufficient for Σ.

Proof. The density of $\mathbf{X}_1, \ldots, \mathbf{X}_N$ is

$$(2) \quad \prod_{\alpha=1}^{N} n(\mathbf{x}_\alpha | \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$= (2\pi)^{-\frac{1}{2}Np} |\boldsymbol{\Sigma}|^{-\frac{1}{2}N} \exp\left[-\boxed{\tfrac{1}{2}\mathrm{tr}}\, \boldsymbol{\Sigma}^{-1} \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \boldsymbol{\mu})(\mathbf{x}_\alpha - \boldsymbol{\mu})' \right]$$

$$= (2\pi)^{-\frac{1}{2}Np} |\boldsymbol{\Sigma}|^{-\frac{1}{2}N} \exp\left\{ -\tfrac{1}{2}\left[N(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) + \boxed{(N-1)\,\mathrm{tr}\, \boldsymbol{\Sigma}^{-1} \mathbf{S} \big]} \right\}.$$

The right-hand side of (2) is in the form of (1) for $\bar{\mathbf{x}}$, $\mathbf{S}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the middle is in the form of (1) for $\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \boldsymbol{\mu})(\mathbf{x}_\alpha - \boldsymbol{\mu})'$, $\boldsymbol{\Sigma}$; in each case $h(\mathbf{x}_1, \ldots, \mathbf{x}_N) = 1$. The right-hand side is in the form of (1) for $\bar{\mathbf{x}}, \boldsymbol{\mu}$ with $\boxed{h(\mathbf{x}_1, \ldots, \mathbf{x}_N)} = \exp\{-\frac{1}{2}(N-1)\,\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\mathbf{S}\}$. ■