1. *Maximum Likelihood Method*: consider n random samples from a multivariate normal distribution, $X_i \in \mathbb{R}^p \sim \mathcal{N}(\mu, \Sigma)$ with $i = 1, \dots, n$.

   (a) Show the log-likelihood function

$$l_n(\mu, \Sigma) = -\frac{n}{2}\mathrm{trace}(\Sigma^{-1} S_n) - \frac{n}{2}\log\det(\Sigma) + C,$$

   where $S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T$, and some constant C does not depend on μ and Σ;

   (b) Show that $f(X) = \mathrm{trace}(AX^{-1})$ with $A, X \succeq 0$ has a first-order approximation,

$$f(X + \Delta) \approx f(X) - \mathrm{trace}(X^{-1}A'X^{-1}\Delta)$$

   hence formally $df(X)/dX = -X^{-1}AX^{-1}$ (note $(I + X)^{-1} \approx I - X$);

   (c) Show that $g(X) = \log\det(X)$ with $A, X \succeq 0$ has a first-order approximation,

$$g(X + \Delta) \approx g(X) + \mathrm{trace}(X^{-1}\Delta)$$

   hence $dg(X)/dX = X^{-1}$ (note: consider eigenvalues of $X^{-1/2}\Delta X^{-1/2}$);

   (d) Use these formal derivatives with respect to positive semi-definite matrix variables to show that the maximum likelihood estimator of Σ is

$$\hat{\Sigma}_n^{MLE} = S_n.$$

(a) $f(x) = \dfrac{1}{(2\pi)^p|\Sigma|} \exp\{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\}$

$\log f(x) = \sum_{i=1}^{n}\left(-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)$

$l(x) = \mathrm{trace}(\log f(x))$

$= \mathrm{trace}(-\frac{np}{2}\log(2\pi)) - \frac{n}{2}\log[\det(\Sigma)) - \frac{1}{2}\sum_{i=1}^{n} tr((x_i-\mu)^T\Sigma^{-1}(x_i-\mu))$

$= -\frac{1}{2}\sum_{i=1}^{n} tr(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T)$

$= -\frac{n}{2} tr(\frac{1}{n}\Sigma^{-1}\cdot\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T) - \frac{n}{2}\log[\det(\Sigma)) + C$

(b) $f(x+\Delta) = tr(A(x+\Delta)^{-1})$

$= tr(A[(I+\Delta x^{-1})x]^{-1})$

$= tr(Ax^{-1}(I+\Delta x^{-1})^{-1})$

$\approx tr(AX^{-1}) - tr(Ax^{-1}\Delta x^{-1})$

$= tr(AX^{-1}) - tr((X^{-1})^T A^T (X^{-1})^T A^T)$

$= tr(AX^{-1}) - tr(X^{-1}AX^{-1}A')$

$$= tr(AX^{-1}) - tr(A'X^{-1}\Delta X^{-1})$$
$$= tr(AX^{-1}) - tr(X^{-1}A'X^{-1}\Delta)$$
$$= f(X) - tr(X^{-1}A'X^{-1}\Delta)$$

(c) $g(X+\Delta) = \log(\det(X+\Delta))$
$$= \log(\det(X^{\frac{1}{2}}(I+X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}})X^{\frac{1}{2}}))$$
$$= \log(\det X) + \log(\det(I+X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}))$$
$$= \log(\det X) + \sum_{i=1}^{n}\log(1+\lambda_i)$$

λ_i is the i-th eigenvalue of $X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}$

Δ small $\Rightarrow \lambda_i$ small

$\Rightarrow \log(1+\lambda_i) \approx \lambda_i$

Thus,

$$\log(\det(Z)) \approx \log(\det(X)) + \sum_{i=1}^{n}\lambda_i$$
$$= \log(\det(X)) + tr(X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}})$$
$$= \log(\det(X)) + tr(X^{-1}\Delta)$$

$\Rightarrow g(X+\Delta) = g(X) + tr(X^{-1}\Delta)$

(d) $l(X) = -\frac{n}{2}tr(\Sigma^{-1}S_n) - \frac{n}{2}\log(\det(\Sigma)) + C$

$$\frac{\partial l(X)}{\partial \Sigma} = -\frac{n}{2}\times(-\Sigma^{-1}S_n\Sigma^{-1}) - \frac{n}{2}\Sigma^{-1}$$

Also. $\frac{\partial l(X)}{\partial \Sigma} = 0$

Thus $\hat{\Sigma} = S_n$

□