1. Given

$$f(x) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

where $x, \mu \in \mathbb{R}^k$, Σ is a k-by-k positive definite matrix and $|\Sigma|$ is its determinant.
Show that $\int_{\mathbb{R}^k} f(x)\, dx = 1$.

let $A = \Sigma^{-1}$, since A is real and symmetric, we can decompose A into $A = T\Lambda T^{-1}$ where T is an orthonormal matrix, Λ is a diagonal matrix

let $y = T^T(x-\mu)$,

so, $\int e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}dx = \int e^{-\frac{1}{2}y^T\Lambda y}|J|dy$ where $|J|$ is the determinant of the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

since $x = Ty + \mu \Rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n t_{1i} y_i \\ \vdots \\ \sum_{i=1}^n t_{ni} y_i \end{bmatrix} + \mu$

Thus, $J = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & & \vdots \\ t_{n1} & & t_{nn} \end{bmatrix} = T$

Since T is orthonormal, $|J| = |T| = 1$

Calculate $\int e^{-\frac{1}{2}y^T\Lambda y}dy$:

$$\int e^{\frac{1}{2}y^T\Lambda y} = \prod_{k=1}^n \int e^{-\frac{1}{2}\lambda_k y_k^2} dy_k$$
$$= \prod_{k=1}^n \left(\frac{2\pi}{\lambda_k}\right)^{\frac{1}{2}}$$
$$= \left(\frac{(2\pi)^n}{\prod_{k=1}^n \lambda_k}\right)^{\frac{1}{2}}$$
$$= \left(\frac{(2\pi)^n}{|\Lambda|}\right)^{\frac{1}{2}}$$

since $|A| = |T^T\Lambda T| = |\Lambda|$,

$$\int e^{\frac{-1}{2}(x-\mu)^T A(x-\mu)}dx = \int e^{-\frac{1}{2}y^T\Lambda y}|J|dy$$
$$= \int e^{-\frac{1}{2}y^T\Lambda y}dy$$
$$= \left(\frac{(2\pi)^n}{|\Lambda|}\right)^{\frac{1}{2}}$$
$$= \left(\frac{(2\pi)^n}{|A|}\right)^{\frac{1}{2}}$$
$$= \left(\frac{(2\pi)^n}{|\Sigma^{-1}|}\right)^{\frac{1}{2}} = \sqrt{(2\pi)^n|\Sigma|}$$

$\Rightarrow$

$$\int_{\mathbb{R}^k} f(x)dx = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \int e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}dx$$
$$= \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \cdot \sqrt{(2\pi)^n|\Sigma|}$$
$$= 1$$

2. Let A, B be n-by-n matrices and x be a n-by-1 vector.

(a) Show that $\frac{\partial}{\partial A}\operatorname{trace}(AB) = B^T$.

(b) Show that $x^T A x = \operatorname{trace}(x x^T A)$.

(b) Derive the maximum likelihood estimators for a multivariate Gaussian.

(a) $\operatorname{trace}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} b_{ji}$

$$\Rightarrow \frac{\partial \operatorname{trace}(AB)}{\partial A} = \begin{bmatrix} \frac{\partial \operatorname{trace}(AB)}{\partial a_{11}} & \cdots & \frac{\partial \operatorname{trace}(AB)}{\partial a_{1n}} \\ \vdots & & \vdots \\ \frac{\partial \operatorname{trace}(AB)}{\partial a_{n1}} & & \frac{\partial \operatorname{trace}(AB)}{\partial a_{nn}} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{n1} \\ b_{12} & & & \vdots \\ \vdots & & & b_{n\,n-1} \\ b_{1n} & \cdots & b_{n-1\,n} & b_{nn} \end{bmatrix} = B^T$$

(b) $$x^T A x = [x_1 \cdots x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \left[\sum_{i=1}^{n} x_i a_{i1} \cdots \sum_{i=1}^{n} x_i a_{in} \right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} x_i a_{i1} x_1 + \cdots + \sum_{i=1}^{n} x_i a_{in} x_n$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{n} x_i a_{ik} x_k$$

$$x x^T A = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} [x_1 \cdots x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 x_2 \cdots & x_1 x_n \\ x_2 x_1 & x_2^2 & x_2 x_n \\ \vdots & \ddots & \vdots \\ x_n x_1 & \cdots & x_n^2 \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} x_1 \sum_{i=1}^{n} x_i a_{i1} & \cdots & x_1 \sum_{i=1}^{n} x_i a_{in} \\ \vdots & & \\ x_n \sum_{i=1}^{n} x_i a_{i1} & \cdots & x_n \sum_{i=1}^{n} x_i a_{in} \end{bmatrix}$$

$$\Rightarrow \operatorname{trace}(x x^T A) = x_1 \sum_{i=1}^{n} x_i a_{i1} + \cdots + x_n \sum_{i=1}^{n} x_i a_{in}$$

$$= \sum_{k=1}^{n} x_k \sum_{i=1}^{n} x_i a_{ik}$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{n} x_i a_{ik} x_k = x^T A x \quad \square$$

(C)

PDF : $\frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$ for a sample $\{x_1 \cdots x_m\}$, $x_i \in \mathbb{R}^n$

$$\Rightarrow L(\mu,\Sigma) = \sum_{i=1}^{m} \ln\left(\frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)} \right) = \sum_{i=1}^{m} -\ln\sqrt{(2\pi)^n |\Sigma|} - \frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$= -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

MLE for μ:

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu}\left(-\frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\frac{\partial}{\partial \mu}\operatorname{trace}\left((x_i-\mu)(x_i-\mu)^T \Sigma^{-1}\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\frac{\partial}{\partial (x_i-\mu)(x_i-\mu)^T}\operatorname{trace}\left((x_i-\mu)(x_i-\mu)^T \Sigma^{-1}\right) \cdot \frac{d(x_i-\mu)(x_i-\mu)^T}{d\mu}$$

$\left((x_i-\mu)(x_i-\mu)^T = \operatorname{trace}\left((x_i-\mu)^T(x_i-\mu)\right)\right.$
$= \sum_{j=1}^{n}(x_{ij}-\mu_j)^2$

$$= \Sigma^{-1} \cdot \sum_{i=1}^{m}(x_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{m} x_i = m\mu \quad \Rightarrow \mu = \frac{1}{m}\sum_{i=1}^{m} x_i$$

MLE for Σ:

$$\frac{\partial L}{\partial \Sigma} = -\frac{m}{2} \cdot \frac{1}{|\Sigma|} \cdot \frac{d|\Sigma|}{d\Sigma} - \frac{1}{2}\sum_{i=1}^{m}\frac{\partial}{\partial \Sigma}\operatorname{trace}\left((x_i-\mu)(x_i-\mu)^T \Sigma^{-1}\right)$$

$$= -\frac{m}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\left(\sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1} = 0, \quad S = \sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T$$

$$\Rightarrow m\Sigma = \sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T$$

$$\Rightarrow \Sigma = \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T$$