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

prove $\int_{\mathbb{R}^k} f(x)\, dx = 1$

令 $y = x - \mu \Leftrightarrow dx = dy$

$$\Rightarrow \int_{\mathbb{R}^k} f(x)\, dx = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \int_{\mathbb{R}^k} e^{-\frac{1}{2} y^T \Sigma^{-1} y}\, dy$$

$\Sigma^{-1} = Q \Lambda Q^T$, where $Q^T Q = I$;

$z = Q^T y \Rightarrow y = Qz \Leftrightarrow dy = dz$

$$\Rightarrow y^T \Sigma^{-1} y = (Qz)^T (Q \Lambda Q^T)(Qz) = z^T Q^T Q \Lambda Q^T Q z$$

$$= z^T I \Lambda I z = z^T \Lambda z = \sum_{i=1}^{k} \lambda_i z_i^2$$

$$\Rightarrow \int_{\mathbb{R}^k} \cancel{f(x)\, dx} = \int_{\mathbb{R}^k} e^{-\frac{1}{2}\sum_{i=1}^{k} \lambda_i z_i^2}\, dz = \prod_{i=1}^{k} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\lambda_i z_i^2}\, dz_i$$

and we have $\int_{-\infty}^{\infty} e^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}} \Rightarrow \prod_{i=1}^{k} \sqrt{\frac{2\pi}{\lambda_i}} = \sqrt{\frac{(2\pi)^k}{\prod_{i=1}^{k} \lambda_i}}$

$$|\Sigma^{-1}| = \prod_{i=1}^{k} \lambda_i = \frac{1}{|\Sigma|} \Rightarrow \int_{\mathbb{R}^k} f(x)\, dx = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \sqrt{\frac{(2\pi)^k}{\prod_{i=1}^{k} \lambda_i}}$$

$$= \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \sqrt{(2\pi)^k |\Sigma|} = 1 \quad \#$$

2.

(a.)

$$\mathrm{trace}(AB) = \sum_i (AB)_{ii} = \sum_{i,j} A_{ij} B_{ji}$$

$$\frac{\partial}{\partial A_{kl}} \left(\sum_i \sum_j A_{ij} B_{ji} \right) = \begin{cases} 0, & i \neq k \text{ or } j \neq l \\ B_{lk}, & i = k \ \& \ j = l \end{cases}$$

$$(B^T)_{kl} = B_{lk} \Rightarrow \frac{\partial}{\partial A} \mathrm{tra}(AB) = B^T$$

(b.) $\underbrace{x^T A x}_{\mathbb{R}} = \mathrm{tra}(x^T A x)$

$$\Rightarrow \mathrm{tra}(x^T A x) = \mathrm{tra}(x x^T A)$$

$$\Rightarrow x^T A x = \mathrm{tra}(x x^T A) \ \#$$

(c.)

$$L(\mu, \Sigma) = \ln \prod_{i=1}^{N} f(x_i) = \sum_{i=1}^{N} \ln f(x_i)$$

$$= \sum_{i=1}^{N} \left[-\frac{k}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right]$$

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

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu}\left(-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\right) = 0$$

by $\frac{\partial}{\partial z} z^T A z = (A+A^T)z \Rightarrow -\frac{1}{2}\sum_{i=1}^{N} 2\Sigma^{-1}(x_i-\mu)(-1) = 0$

$$\Rightarrow \mu_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$\hat{=}\ S = \Sigma^{-1}$

$$L = C + \frac{N}{2}\ln|S| - \frac{1}{2}\sum_{i=1}^{N} \mathrm{tra}\left((x_i-\mu)(x_i-\mu)^T S\right)$$

$$= C + \frac{N}{2}\ln|S| - \frac{1}{2}\mathrm{tra}\left(\left(\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T\right)S\right)$$

$\hat{=}\ M = \sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T$

$$\Rightarrow L = C + \frac{N}{2}\ln|S| - \frac{1}{2}\mathrm{tra}(MS)$$

$$\frac{\partial L}{\partial S} = \frac{N}{2}(S^{-1})^T - \frac{1}{2}M^T = 0 \Rightarrow \frac{N}{2}\Sigma - \frac{1}{2}M = 0$$

$$\Rightarrow \Sigma = \frac{1}{N}M \Rightarrow \Sigma_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T$$