1. 证明 $1=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$

证: 设 $I=\int_{-\infty}^{+\infty}\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$.

令 $y=\frac{x-\mu}{2}$ 则 $I=\int_{-\infty}^{+\infty}\exp(-\frac{y^2}{2\sigma^2})dy$

令 $z=\frac{y}{\sqrt{2}\sigma}$ 则 $I=\int_{-\infty}^{+\infty}\exp(-z^2)\cdot\sqrt{2}\sigma dz$

$=\sqrt{2}\sigma\int_{-\infty}^{+\infty}\exp(-z^2)dz$

$=\sqrt{2}\sigma\cdot\sqrt{\pi}$

$=\sqrt{2\pi}\sigma$

则 $\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma} f \exp(-\frac{(x-\mu)^2}{2\sigma^2})dx=\sqrt{2\pi}\sigma\cdot\frac{1}{\sqrt{2\pi}\sigma}=1$

2.5.1 假设 u,v 是同维度向量, 证 $u^Tv=tr(vu^T)$

证: 设 $u=(x_1,x_2,\cdots,x_n)^T$, $v=(y_1,y_2,\cdots,y_n)^T$

则 $u^Tv=\sum_{i=1}^{n}x_iy_i$ $\quad vu^T=\begin{bmatrix} x_1y_1 & \cdots & \cdots \\ \cdots & x_2y_2 & \\ \cdots & \ddots & x_ny_n \end{bmatrix}$

$\therefore tr(vu^T)=\sum_{i=1}^{n}x_iy_i$

$\therefore u^Tv=tr(vu^T)$

2.5.4 $x\sim N(\mu,\Sigma)$, 证明 $\cancel{\Sigma=\Sigma}$ $\mu=E(x)=\int_{-\infty}^{\infty}xp(x)dx$

证: $E(x)=\int_{-\infty}^{+\infty}\frac{x}{\sqrt{(2\pi)^N\det\Sigma}}\exp((x-\mu)^T\Sigma^{-1}(x-\mu))dx$

令 $y=x-\mu$.

则 $E(x)=\int_{-\infty}^{+\infty}\frac{y}{\sqrt{(2\pi)^N\det\Sigma}}\exp(y^T\Sigma^{-1}y)dy+\int_{-\infty}^{+\infty}\frac{\mu}{\sqrt{(2\pi)^N\det\Sigma}}\exp(y^T\Sigma^{-1}y)dy$

$=0+\mu\cdot 1$

$=\mu$

$\therefore E(x)=\mu$

2.5.5 对于高斯分布的随机变量，$x \sim N(\mu, \Sigma)$，证明.

$$\Sigma = E[(x-\mu)(x-\mu)^T] = \int_{-\infty}^{+\infty} (x-\mu)(x-\mu)^T p(x) dx.$$

证：
$$I = \int_{-\infty}^{+\infty} (x-\mu)(x-\mu)^T \frac{1}{\sqrt{(2\pi)^N \det \Sigma}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)) dx$$

令 $y = x-\mu$ $\quad \therefore dy = dx$

$$I = \int_{-\infty}^{+\infty} \frac{yy^T}{\sqrt{(2\pi)^N \det \Sigma}} \exp(-\frac{1}{2} y^T \Sigma^{-1} y) dy$$

$$= \int_{-\infty}^{+\infty} \frac{yy^T}{\sqrt{(2\pi)^N \det \Sigma}} \cdot \frac{1}{-y^T \Sigma^{-1}} \exp(-\frac{1}{2} y^T \Sigma^{-1} y) \, d(-\frac{1}{2} y^T \Sigma^{-1} y).$$

$$= \int_{-\infty}^{+\infty} \frac{-y\Sigma}{\sqrt{(2\pi)^N \det \Sigma}} \, d \exp(-\frac{1}{2} y^T \Sigma^{-1} y)$$

$$= \frac{-y\Sigma}{\sqrt{(2\pi)^N \det \Sigma}} \exp(-\frac{1}{2} y^T \Sigma^{-1} y) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \exp(-\frac{1}{2} y^T \Sigma^{-1} y) \, d(\frac{-y\Sigma}{\sqrt{(2\pi)^N \det \Sigma}})$$

$$= 0 + \Sigma \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{(2\pi)^N \det \Sigma}} \exp(-\frac{1}{2} y^T \Sigma^{-1} y) dy$$

$$= \Sigma$$

$$\therefore \Sigma = E[(x-\mu)(x-\mu)^T] = \int_{-\infty}^{+\infty} (x-\mu)(x-\mu)^T p(x) dx$$

2.5.6 对于K个相互独立的高斯变量，$x_k \sim N(\mu_k, \Sigma_k)$，请证明它们的归一化积仍是高斯分布.

证明：设随机变量 $x_k \in R^{N_k}$，服从的PDF为

$$p_k(x) = \frac{1}{\sqrt{(2\pi)^{N_k}\det\Sigma_k}} \exp\left(-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)\right)$$

定义它们的归一化积为 $p(x) = \eta\, p_1(x)p_2(x)\cdots p_K(x)$

∴指数分布的二次型为

$$-\frac{1}{2}\sum_{k=1}^{K}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)$$

$$= -\frac{1}{2}\left[x^T\left(\sum_{k=1}^{K}\Sigma_k^{-1}\right)x - 2x^T\left(\sum_{k=1}^{K}\Sigma_k^{-1}\mu_k\right) + \sum_{k=1}^{K}\mu_k^T\Sigma_k^{-1}\mu_k\right]$$

定义 列矩阵μ，使得

$$\sum_{k=1}^{K}\Sigma_k^{-1}\mu = \sum_{k=1}^{K}\Sigma_k^{-1}\mu_k$$

∴指数分布的二次型为

$$-\frac{1}{2}\left[(x-\mu)^T\left(\sum_{k=1}^{K}\Sigma_k^{-1}\right)(x-\mu) - \mu^T\left(\sum_{k=1}^{K}\Sigma_k^{-1}\right)\mu + \sum_{k=1}^{K}\mu_k^T\Sigma_k^{-1}\mu_k\right]$$

定义 $\Sigma^{-1} = \sum_{k=1}^{K}\Sigma_k^{-1}$，$D = -\mu^T\left(\sum_{k=1}^{K}\Sigma_k^{-1}\right)\mu + \sum_{k=1}^{K}\mu_k^T\Sigma_k^{-1}\mu_k$

∴指数分布的二次型为

$$-\frac{1}{2}\left[(x-\mu)^T\Sigma^{-1}(x-\mu) + D\right].$$

∴归一化积为

$$p(x) = \eta \cdot \frac{1}{\sqrt{(2\pi)^{\sum_{k=1}^{K}N_k}\prod_{k=1}^{K}\det\Sigma_k}} \cdot \exp\left(-\frac{D}{2}\right) \cdot \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$\because \cancel{\det\Sigma^{-1} = \det}$$

$$\cancel{= \eta \cdot \frac{\sqrt{(2\pi)^N\det\Sigma}}{\sqrt{(2\pi)^{\sum_{k=1}^{K}N_k}\cdot\prod_{k=1}^{K}\det\Sigma_k}} \cdot \exp\left(-\frac{D}{2}\right) \cdot \frac{1}{\sqrt{(2\pi)^N\det\Sigma}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}$$

又 $\because$ $p(x) = \lambda \cdot \frac{1}{\sqrt{(2\pi)^{\sum_{k=1}^{K} N_k} \prod_{k=1}^{K} \det \Sigma_k}} \cdot \exp \prod_{k=1}^{K} \exp\left(-\frac{1}{2}(x_k-\mu_k)^T \Sigma_k^{-1}(x_k-\mu_k)\right)$

$\therefore$ 设 $\eta = \frac{1}{\exp(-\frac{D}{2})}$，则

$$\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) = \eta \prod_{k=1}^{K} \exp\left(-\frac{1}{2}(x_k-\mu_k)^T \Sigma_k^{-1}(x_k-\mu_k)\right)$$

其中 $\Sigma^{-1} = \sum_{k=1}^{K} \Sigma_k^{-1}$，$\Sigma^{-1}\mu = \sum_{k=1}^{K} \Sigma_k^{-1}\mu_k$

归一化因子 $\eta = \frac{1}{\exp\left(-\frac{1}{2}\left(-\mu^T \sum_{k=1}^{K} \Sigma_k^{-1} \mu + \sum_{k=1}^{K} \mu_k^T \Sigma_k^{-1} \mu_k\right)\right)}$