Multivariate Gaussian.

$$Z \sim N(\mu, \Sigma), \quad \mu \in \mathbb{R}^p, \quad \Sigma \in \mathbb{R}^{p \times p}.$$

density function.
$$f_Z(z) = (2\pi)^{-\frac{p}{2}} \left[\det(\Sigma)\right]^{-\frac{1}{2}} e^{-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)}.$$

X is not correlated with $Y \Rightarrow X \perp\!\!\!\perp Y$.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix} \right). \quad \Sigma_{xy} = \mathrm{Cov}(X, Y) = 0.$$

$\mu_x \in \mathbb{R}^{p_x}$
$\mu_y \in \mathbb{R}^{p_y}$

$$\Sigma = \begin{pmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{pmatrix}. \quad \det(\Sigma) = \det(\Sigma_x) \cdot \det(\Sigma_y).$$

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_x^{-1} & 0 \\ 0 & \Sigma_y^{-1} \end{pmatrix}.$$

thus $$\left((X-\mu_x)^T, (y-\mu_y)^T\right) \begin{pmatrix} \Sigma_x^{-1} & 0 \\ 0 & \Sigma_y^{-1} \end{pmatrix} \begin{pmatrix} X-\mu_x \\ Y-\mu_y \end{pmatrix}.$$

$$= (X-\mu_x)^T \Sigma_x^{-1} (X-\mu_x) + (y-\mu_y)^T \Sigma_y^{-1} (y-\mu_y).$$

$f_{X,Y}(x,y)$

$$= (2\pi)^{-\frac{p_x+p_y}{2}} \left(\det(\Sigma_x)\det(\Sigma_y)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_x)^T\Sigma_x^{-1}(x-\mu_x)-\frac{1}{2}(y-\mu_y)^T\Sigma_y^{-1}(y-\mu_y)}$$

$$= f_X(x)\cdot f_Y(y).$$

maximum likelihood estimator.

$$Z_1, \cdots, Z_n \overset{iid}{\sim} N(\mu, \Sigma).$$

the MLE $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n Z_i$, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T$.

basic matrix algebra.

for $M \in \mathbb{R}^{n\times n}$. $Tr(M) = M_{11} + M_{22} + \cdots + M_{nn}$.

for $A, B \in \mathbb{R}^{n\times n}$, $Tr(AB^T)$

$$= Tr\left(\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} b_{11} & b_{21} & \cdots & b_{n1} \\ b_{12} & b_{22} & \cdots & b_{n2} \\ \vdots & \vdots & & \vdots \\ b_{1n} & b_{2n} & \cdots & b_{nn} \end{pmatrix}\right).$$

$$= Tr\begin{pmatrix} a_{11}b_{11} + a_{12}b_{12} + \cdots + a_{1n}b_{1n} & & \\ & a_{21}b_{21} + \cdots & \\ & & \\ & & a_{n1}b_{n1} + \cdots \end{pmatrix} = \sum_{ij} A_{ij}B_{ij}.$$

therefore $Tr(AB^T) = Tr(B^TA)$, moreover $\sum_{i=1}^n a_i Tr(A_i) = Tr\left(\sum_{i=1}^n a_i A_i\right)$.

for a matrix $X \in \mathbb{R}^{n\times m}$, $f(X) \in \mathbb{R}$.

$\frac{\partial f(X)}{\partial X}$ is a matrix with (i,j)-th entry $\frac{\partial f(X)}{\partial X_{ij}}$

important identity 1. $\dfrac{\partial \operatorname{Tr}(AX)}{\partial X} = A$.

2. for symmetric invertible X,

$$\frac{\partial \log\det(X)}{\partial X} = X^{-1}.$$

likelihood function

$$L_n(\mu, \Sigma) = \prod_{i=1}^{n} f(z_i) = (2\pi)^{-\frac{pn}{2}} \left(\det(\Sigma)\right)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu)}$$

$$\log L_n(\mu,\Sigma) = -\frac{n}{2}\log\det(\Sigma) - \frac{n}{2}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu) - \frac{pn}{2}\log(2\pi).$$

equivalent to minimize.

$$g_n(\mu,\Sigma) = \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu).$$

$$= \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z}+\bar{z}-\mu)^T\Sigma^{-1}(z_i-\bar{z}+\bar{z}-\mu)$$

$$= \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z}) + (\bar{z}-\mu)^T\Sigma^{-1}(\bar{z}-\mu)$$

(cross product is $\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(\bar{z}-\mu)$.
$= (\bar{z}-\bar{z})^T\Sigma^{-1}(\bar{z}-\mu) = 0$).

$(\bar{z}-\mu)^T\Sigma^{-1}(\bar{z}-\mu) \geq 0$. $(=0)$ iff $\mu = \bar{z}$.

thus $\hat{\mu} = \bar{z}$.

$$g_n(\hat{\mu},\Sigma) = \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z}).$$

$$= \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}\operatorname{Tr}\left((z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z})\right).$$

$$= \log\det(\Sigma) + \frac{1}{n}\sum_{i=1}^{n}\operatorname{Tr}\left((z_i-\bar{z})(z_i-\bar{z})^T\Sigma^{-1}\right)$$

$$= -\log\det(\Sigma^{-1}) + \mathrm{Tr}\left(\left(\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T\right)\Sigma^{-1}\right).$$ Pge 8

$$\frac{\partial g_n(\hat{\mu},\Sigma)}{\partial \Sigma^{-1}} = -\Sigma + \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T = 0.$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T.$$