More on Probability
Random Vectors

ROB 501

Necmiye Ozay

- Random Vectors
  - Covariance matrices
- Multi-variate Normal Distribution
- BLUE: Best Linear Unbiased Estimator
- (if time) QR decomposition

$Q = E\left((X-m)(X-m)^T\right)$

If we have two random variables X_1, X_2 s.t. $E(X_1 X_2) = 0$, we call X_1, X_2 to be uncorrelated.

Covariance matrix examples:

1) $Q = \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix} \Rightarrow Q^{-1} = \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & 10 \end{bmatrix}$

2) $Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \Theta^T \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}}_{\Sigma} \Theta$

$\lambda_1 + \lambda_2 = 4 \quad \lambda_1 = 3$

$\lambda_1 \lambda_2 = 3 \quad \lambda_2 = 1$

$\Theta = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Define a new random vector by

$\underline{Y = \Theta \cdot X} = \frac{1}{\sqrt{2}} \begin{bmatrix} X_1 + X_2 \\ X_1 - X_2 \end{bmatrix}$ $\quad Y_1 = \frac{1}{\sqrt{2}}(X_1 + X_2)$, $\quad Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$

Assume $m = E(\underline{X}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$E(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} E(X_1) + E(X_2) \\ E(X_1) - E(X_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$Cov(Y) = E\left((Y - m_y)(Y - m_y)^T\right)$

*Expected value operator is a linear operator.
$E(Y_1) = \frac{1}{\sqrt{2}}(E(X_1) + E(X_2))$
$= 0$
$E(Y_2) = 0$

$$= E(yy^T)$$

$$= E(\Theta \cancel{X} \cancel{X}^T \Theta^T) = \Theta \underbrace{E(XX^T)}_{Q} \Theta^T$$

$$\boxed{m_y = \begin{bmatrix} E(Y_1) \\ E(Y_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q = \Theta^T \Sigma \Theta$$

$$\Theta Q = \Sigma \Theta$$

$$\Theta Q \Theta^T = \Sigma$$

lower uncertainty in Y_2 than Y_1

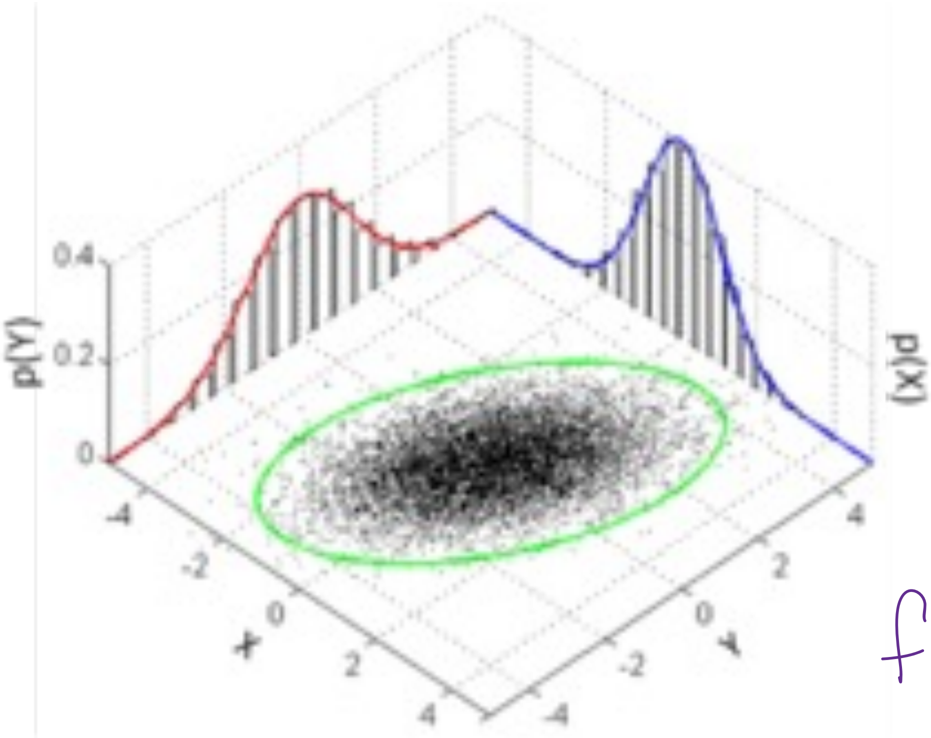
With slight abuse of terminology, let's define the variance of a random vector X

$$\mathrm{var}(X) = E\left((X - m_x)^T (X - m_x)\right) = E\left(\|X - m_x\|_2^2\right)$$

$$= E\left(\sum_{i=1}^{n} (X_i - m_{x_i})(X_i - m_{x_i})\right)$$

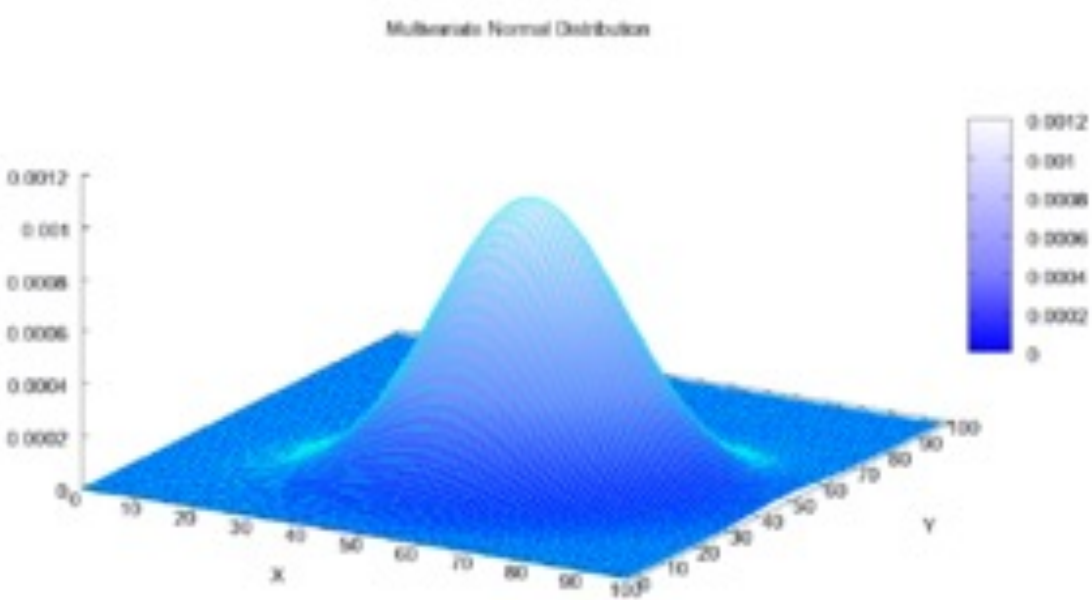
$$\boxed{Q = E\left((X-m)(X-m)^T\right)}$$

$$= \mathrm{tr}(Q)$$



k-dimensional Gaussian

$$f(X) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} e^{\left(\frac{-1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right)}$$



From: https://proofwiki.org/wiki/Variance_of_Gaussian_Distribution/Proof_1

From the definition of the Gaussian distribution, X has probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

From Variance as Expectation of Square minus Square of Expectation:

$$\operatorname{var}(X) = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx - (\operatorname{E}(X))^2$$

So:

$\operatorname{var}(X) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx - \mu^2$	Expectation of Gaussian Distribution
$= \dfrac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} \left(\sqrt{2}\sigma t + \mu\right)^2 \exp\left(-t^2\right) dt - \mu^2$	substituting $t = \dfrac{x-\mu}{\sqrt{2}\sigma}$
$= \dfrac{1}{\sqrt{\pi}} \left(2\sigma^2 \displaystyle\int_{-\infty}^{\infty} t^2 \exp\left(-t^2\right) dt + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} t \exp\left(-t^2\right) dt + \mu^2 \int_{-\infty}^{\infty} \exp\left(-t^2\right) dt\right) - \mu^2$	
$= \dfrac{1}{\sqrt{\pi}} \left(2\sigma^2 \displaystyle\int_{-\infty}^{\infty} t^2 \exp\left(-t^2\right) dt + 2\sqrt{2}\sigma\mu \left[-\frac{1}{2}\exp\left(-t^2\right)\right]_{-\infty}^{\infty} + \mu^2\sqrt{\pi}\right) - \mu^2$	Fundamental Theorem of Calculus, Gaussian Integral
$= \dfrac{1}{\sqrt{\pi}} \left(2\sigma^2 \displaystyle\int_{-\infty}^{\infty} t^2 \exp\left(-t^2\right) dt + 2\sqrt{2}\sigma\mu \cdot 0\right) + \mu^2 - \mu^2$	Exponential Tends to Zero and Infinity
$= \dfrac{2\sigma^2}{\sqrt{\pi}} \displaystyle\int_{-\infty}^{\infty} t^2 \exp\left(-t^2\right) dt$	
$= \dfrac{2\sigma^2}{\sqrt{\pi}} \left(\left[-\dfrac{t}{2}\exp\left(-t^2\right)\right]_{-\infty}^{\infty} + \dfrac{1}{2} \displaystyle\int_{-\infty}^{\infty} \exp\left(-t^2\right) dt\right)$	Integration by Parts
$= \dfrac{2\sigma^2}{\sqrt{\pi}} \cdot \dfrac{1}{2} \displaystyle\int_{-\infty}^{\infty} \exp\left(-t^2\right) dt$	Exponential Tends to Zero and Infinity
$= \dfrac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}}$	Gaussian Integral
$= \sigma^2$	

BLUE: Best Linear Unbiased Estimator

Recap: on non-stochastic setting:

Underdetermined case: $y = Cx \quad x \in \mathbb{R}^n, \; y \in \mathbb{R}^m$

rows are linearly indep. $n > m$.

Given weight matrix $S > 0$, $\|x\|_S^2 = x^T S x$

$$\left(\hat{x} := \underset{x \in \mathbb{R}^n}{\arg\min} \; \|x\|_S^2 \quad \text{s.t.} \quad y = Cx \right) = S^{-1} C^T (C S^{-1} C^T)^{-1} y$$

Overdetermined case: $y = Cx + e$ (cols are lin. indep)

$x \in \mathbb{R}^n, \; y \in \mathbb{R}^m$

$m > n$

$S > 0$

$$\left(\hat{x} := \underset{e \in \mathbb{R}^m, \, x \in \mathbb{R}^n}{\arg\min} \; \|e\|_S^2 \quad \text{s.t.} \quad y = Cx + e \right) = \underset{x \in \mathbb{R}^n}{\arg\min} \; \|y - Cx\|_S^2$$

$$= (C^T S C)^{-1} C^T S y$$

Back to BLUE:

Goal: Given a stochastic noise model, we will find out how to choose the weight

matrix in an overdetermined setting $(m>n)$
s.t. the resulting estimator has some
nice properties.

$$y = Cx + \varepsilon, \quad y \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad \varepsilon \in \mathbb{R}^m$$

Noise model: Assume ε is a R.V. with
$m = E(\varepsilon) = 0$, $Q = \text{cov}(\varepsilon) > 0$.
x is unknown constant, no probabilistic
model for x.

Want to find a linear estimator: $\hat{x} = Ky$

$K \in \mathbb{R}^{n \times m}$

Unbiased estimator: $E(\hat{x}) = x$

Best in terms of minimizing the
variance

$$\text{var}(\hat{x}) = E\left((\hat{x} - m_{\hat{x}})^T (\hat{x} - m_{\hat{x}}) \right)$$

since, we want $\underbrace{E(\hat{x})}_{m_{\hat{x}}} = x$

$$var(\hat{x}) = E\left((\hat{x}-x)^T(\hat{x}-x)\right)$$

we want linear estimator: $\hat{x} = Ky = K(Cx+\varepsilon)$

$*$ $\hat{x} - x = KCx + K\varepsilon - x$

To have $E(\hat{x}) = x \Rightarrow E(\hat{x}) = E(KCx + K\varepsilon)$

$= KCx + \cancel{KE(\varepsilon)}^{0}$

$= KCx$

To have $E(\hat{x}) = x \Rightarrow \underline{KC = I}$

guarantees unbiasedness

$*$ becomes

for $KC = I$, $\hat{x} - x = K\varepsilon$

$\varepsilon^T K^T K\varepsilon = tr(K\varepsilon(K\varepsilon)^T)$

$\boxed{x^T x = tr(xx^T)}$

$$var(\hat{x}) = E((K\varepsilon)^T K\varepsilon)$$

$$= E\left(tr\left((K\varepsilon)(K\varepsilon)^T\right)\right)$$

$$= E\left(tr(K\varepsilon\varepsilon^T K^T)\right)$$

trace is a linear operator, expectation

$$= tr(K\underbrace{E(\varepsilon\varepsilon^T)}K^T)$$

$$= \operatorname{tr}(K \; Q \; K^T)$$

$$\boxed{\text{find } \quad \hat{K} = \underset{K}{\operatorname{argmin}} \; \operatorname{tr}(KQK^T) \quad \text{s.t.} \quad KC = I}$$

<u>Theorem:</u> Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $y = Cx + \varepsilon$, $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon^T) = Q > 0$, and $\operatorname{rank}(C) = n$.

Then, the BLUE is $x = \hat{K} y$ where

$$\hat{K} = (C^T Q^{-1} C)^{-1} C^T Q^{-1}.$$

Moreover, the covariance of the error is

$$E\left((\hat{x} - x)(\hat{x} - x)^T\right) = (C^T Q^{-1} C)^{-1} = \hat{K} Q \hat{K}^T$$

<u>Key observation:</u>

BLUE $\equiv$ Weighted least squares w/ (weight matrix) $S = Q^{-1}$ (information matrix)

<u>Proof:</u>

k_i is ith row of K

$$K = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$$

$$\boxed{\hat{K} = \underset{K}{\operatorname{argmin}} \ tr(KQK^T) \quad \text{s.t.} \quad KC = I}$$

$$K^T = [k_1^T \mid k_2^T \mid \dots \mid k_n^T]$$

$$tr(KQK^T) = tr\left(\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} Q [k_1^T \mid k_2^T \mid \dots \mid k_n^T]\right)$$

$$= tr\left(\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} [Qk_1^T \mid Qk_2^T \mid \dots \mid Qk_n^T]\right)$$

$$= \sum_{i=1}^{n} k_i Q k_i^T = \underbrace{\sum_{i=1}^{n} \|k_i\|_Q^2}$$

objective function can be separated into a sum of functions of each k_i

$$KC = I$$
$$\Updownarrow$$
$$C^T K^T = I \iff C^T [k_1^T | k_2^T | \dots k_n^T] = I$$

$$\iff C^T k_i^T = e^i \quad \text{for all } i = 1, \dots, n$$

e^i: i^{th} col'n of the identity matrix

→ Since both the objective function and constraints are decomposed, we can solve n independent smaller optimization problems of the form:

$$\hat{k}_i^T = \operatorname{argmin} \|k_i^T\|_Q$$
$$\text{s.t.} \quad C^T k_i^T = e^i$$

Minimum norm solution of an undetermined set of equations:

$$\Longleftrightarrow \quad \hat{k}_i^T = Q^{-1} C (C^T Q^{-1} C)^{-1} e^i$$