ELEC-C5310 - Introduction to Estimation, Detection and Learning: Basics of Probability

Prof. Sergiy Vorobyov

Department of Signal Processing and Acoustics
Aalto University

BASICS

Random variables

Real random variable $x(\xi)$ represents a mapping that assigns a real number x to every outcome ξ from the abstract probability space.

Probability distribution of a random variable x

$$P_x(x) = \mathrm{Probability}\{x(\xi) \leq x\}$$

Probability density function (pdf):

$$p_x(x) = \frac{\partial P_x(x)}{\partial x}$$

where

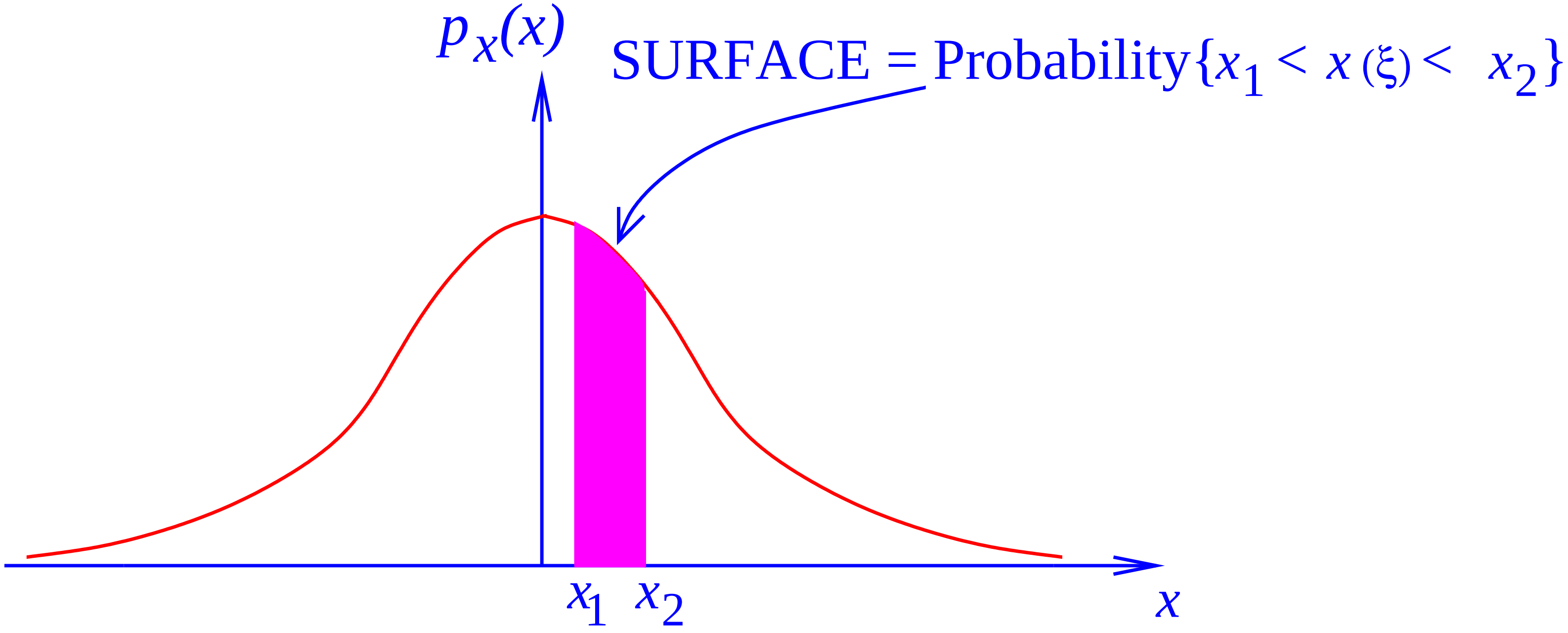
$$P_x(x_0) = \int_{-\infty}^{x_0} p_x(x)\, dx$$

Since $P_x(\infty) = 1$, we have *norming condition*

$$\int_{-\infty}^{\infty} p_x(x)\, dx = 1$$

Simple interpretation:

$$p_x(x) = \lim_{\Delta \to 0} \frac{\text{Probability}\{x - \Delta/2 \le x(\xi) \le x + \Delta/2\}}{\Delta}$$



Mathematical *expectation* of an arbitrary $f(x)$:

$$\mathrm{E}\{f(x)\} = \int_{-\infty}^{\infty} f(x)\, p_x(x)\, dx$$

Mean:

$$\mu_x = \mathrm{E}\{x\} = \int_{-\infty}^{\infty} x\, p_x(x)\, dx$$

Variance of a *real* random variable X:

$$\mathrm{var}\{x\} = \sigma_x^2 \;=\; \mathrm{E}\{(x - \mathrm{E}\{x\})^2\}$$

A *complex* random variable:

$$x(\xi) = x_R(\xi) + j x_I(\xi), \qquad \mathrm{var}\{x\} = \sigma_x^2 = \mathrm{E}\{|x - \mathrm{E}\{x\}|^2\}$$

Gaussian distribution : 1-variate case

- RV x has a **normal (Gaussian) distribution** with mean μ and variance σ^2 if its p.d.f. is of the form

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathbb{R}.$$

We denote this case by $x \sim \mathcal{N}(\mu, \sigma^2)$.

- The case $\mathcal{N}(0,1)$ is called the **standard normal distribution** :

$$\text{p.d.f.:} \quad \phi(x) = f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$\text{c.d.f.:} \quad \Phi(x) = \int_{-\infty}^{x} \phi(z)\mathrm{d}z = \frac{1}{2}\left\{1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right\}.$$

- Note that $\Phi(-x) = 1 - \Phi(x)$ as $\mathcal{N}(0,1)$ is symmetric w.r.t. the origin and that $\Phi(\cdot)$ does not have analytical expression.
- Recall that

$$\frac{x-\mu}{\sigma} \sim \mathcal{N}(0,1) \quad \text{when} \quad x \sim \mathcal{N}(\mu, \sigma^2).$$

- Thus, if $x \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(|x-\mu| \leq \sigma) = \mathbb{P}(\mu - \sigma \leq x \leq \mu + \sigma) = 2\Phi(1) - 1 \approx 0.68$$
$$\mathbb{P}(|x-\mu| \leq 2\sigma) = \mathbb{P}(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 2\Phi(2) - 1 \approx 0.95.$$

- Standard deviation σ measures how concentrated the distribution is about the mean

Some properties

1. If $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, p$ are mutually independent, then

$$\sum_{i=1}^{p} x_i \sim \mathcal{N}\Big(\sum_{i=1}^{p} \mu_i, \sum_{i=1}^{p} \sigma_i^2\Big).$$

2. If $x_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, p$ are mutually independent, then

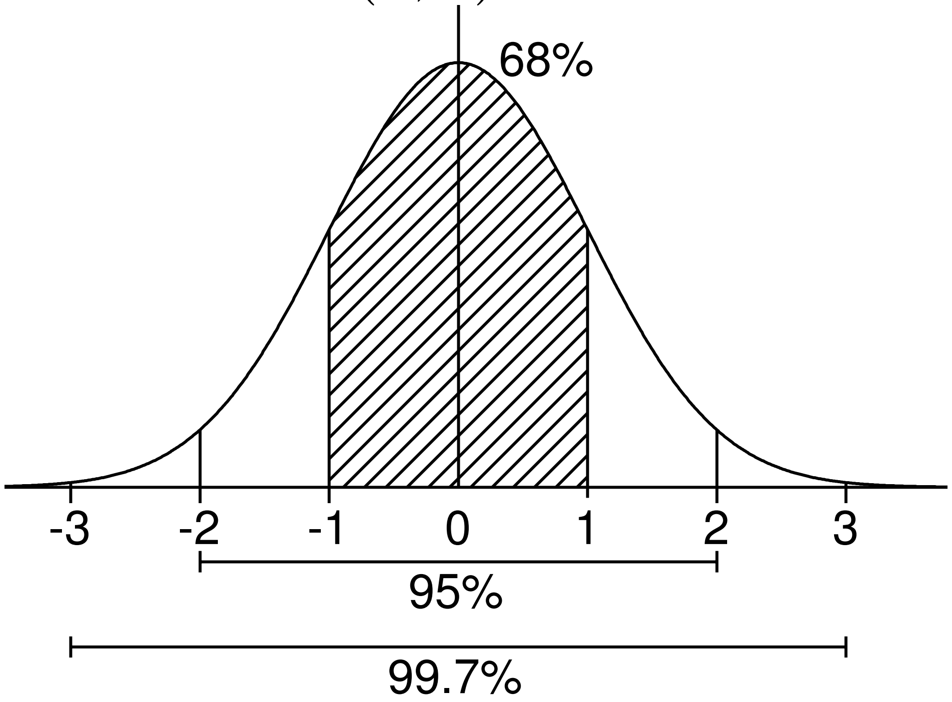
$$\sum_{i=1}^{n} x_i^2 \sim \chi_p^2$$

i.e., $\|\mathbf{x}\|^2 \sim \chi_p^2$, when $\mathbf{x} = (x_1, \ldots, x_p)^\top$.

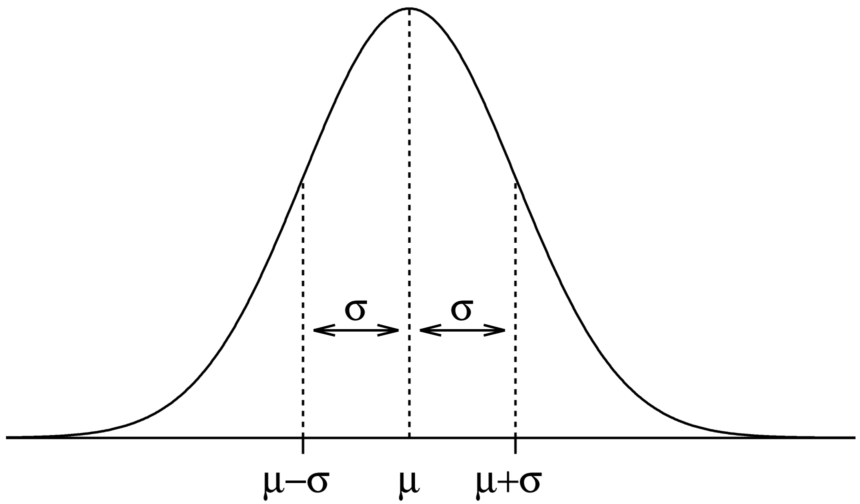
3. Let $x \sim \mathcal{N}(\mu, \sigma^2)$, then the skewness and kurtosis coefficients $\gamma_1 = \frac{\mathsf{E}[x^3]}{\sigma^3}$ and $\gamma_2 = \frac{\mathsf{E}[x^4]}{\sigma^4} - 3$ vanish (i.e., are equal to zero).

Bell shape curve

$\mathcal{N}(0,1)$ distribution



$\mathcal{N}(\mu,\sigma^2)$ distribution



Multivariate distribution

- We can go from 1-dimension to any finite dimensions:

$$\mathbb{P}(a < x_1 < b, c < x_2 < d) = ?$$

- Assume a continuous RV with (joint) p.d.f. $f(\mathbf{x})$ which thus verifies

$$f(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^p$$
$$\int_{-\infty}^{\infty} f(\mathbf{x}) \mathrm{d}\mathbf{x} \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_p) \mathrm{d}x_1 \cdots \mathrm{d}x_p = 1.$$

- The c.d.f. is

$$F(\mathbf{a}) = F(a_1, \ldots, a_p) = \mathbb{P}(x_1 \leq a_1, \ldots, x_p \leq a_p)$$
$$= \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_p} f(x_1, \ldots, x_p) \mathrm{d}x_1 \cdots \mathrm{d}x_p \equiv \int_{-\infty}^{\mathbf{a}} f(\mathbf{x}) \mathrm{d}\mathbf{x},$$

- Also the components x_i (or any multivariate components) of $\mathbf{x}$ are RV's and thus they have distributions as well. These are commonly called **marginal distributions**.

- Compose $p \times 1$ RV $\mathbf{x}$ as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix},$$

where $\mathbf{x}_1 = (x_1, \ldots, x_k)^\top$ and $\mathbf{x}_2 = (x_{k+1}, \ldots, x_p)^\top$, $k < p$.

- Then

$$F_1(\mathbf{a}_1) = F(a_1, \ldots, a_k, \infty, \ldots, \infty) = \int_{-\infty}^{\mathbf{a}_1} \int_{-\infty}^{\infty} f(\mathbf{x}_1, \mathbf{x}_2) \mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2$$

$$f_1(\mathbf{x}_1) = \int_{-\infty}^{-\infty} f(\mathbf{x}_1, \mathbf{x}_2) \mathrm{d}\mathbf{x}_2,$$

are the **marginal c.d.f.** and **marginal p.d.f.** of RV $\mathbf{x}_1$. (Similarly for RV $\mathbf{x}_2$).

- (Marginal) p.d.f. and c.d.f. of $\mathbf{x}_2$ are defined analogously.
- The p.d.f. of the **conditional distribution** of RV $\mathbf{x}_2$ given $\mathbf{x}_1$ is

$$f(\mathbf{x}_2 | \mathbf{x}_1) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f_1(\mathbf{x}_1)}$$

(Similarly one obtain the p.d.f. of the cond. distr. of $\mathbf{x}_1$ given $\mathbf{x}_2$).

- RV's $x_1, \ldots, x_p$ are (mutually) **independent** if

$$f(\mathbf{x}) = f_1(x_1) \cdots f_p(x_p) \quad \forall\, \mathbf{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$$

- RV's $\mathbf{x}_1$ ja $\mathbf{x}_2$ are (pairwise) independent if

$$f(\mathbf{x}_1, \mathbf{x}_2) = f_1(\mathbf{x}_1) f_2(\mathbf{x}_2) \quad \forall\, \mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top \in \mathbb{R}^p.$$

Multivariate normal distribution

- We assume the existence of a density (i.e., non-singular normal distribution $\Rightarrow$ full rank covariance matrix).
- Multivariate normal (MVN) distribution can be defined more generally (there can be *singular* multinormal distributions).
- Thus, we say that a RV $\mathbf{x} = (x_1, \ldots, x_p)^\top$ has a **p-variate normal distribution** if its p.d.f. is of the form

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

where $\boldsymbol{\mu} = \mathsf{E}[\mathbf{x}]$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x})$ is a positive definite symmetric covariance matrix (so $\boldsymbol{\Sigma} \succ 0$). We denote this case by $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

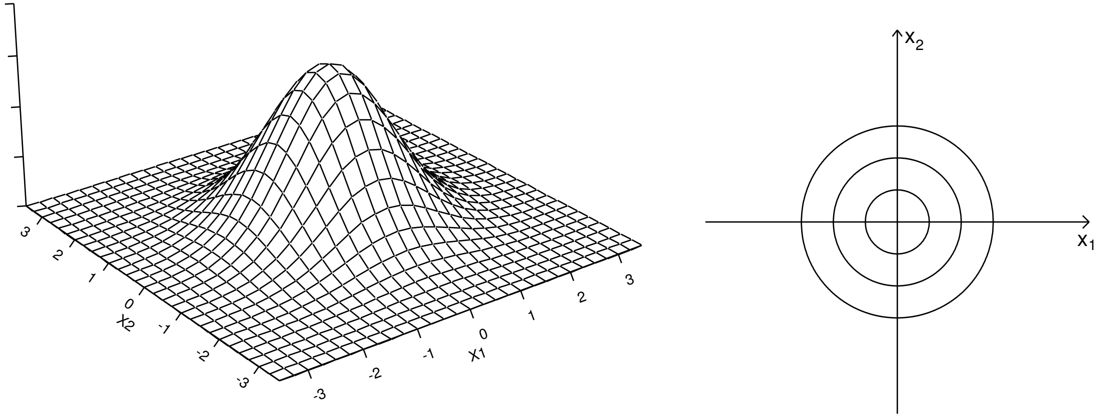
- Write $\sigma_{ij} = (\boldsymbol{\Sigma})_{ij}$ and recall that $\sigma_{ii} \equiv \sigma_i^2 = \mathrm{Var}(x_i)$ and $\sigma_{ij} = \mathrm{Cov}(x_i, x_j)$ for $i \neq j$.
- The case $\mathcal{N}_p(\mathbf{0}, \mathbf{I})$ is called the **standard (multi)normal distribution**:

$$f_{\mathbf{0},\mathbf{I}}(\mathbf{x}) = (2\pi)^{-p/2} e^{-\frac{1}{2}\|\mathbf{x}\|^2}.$$

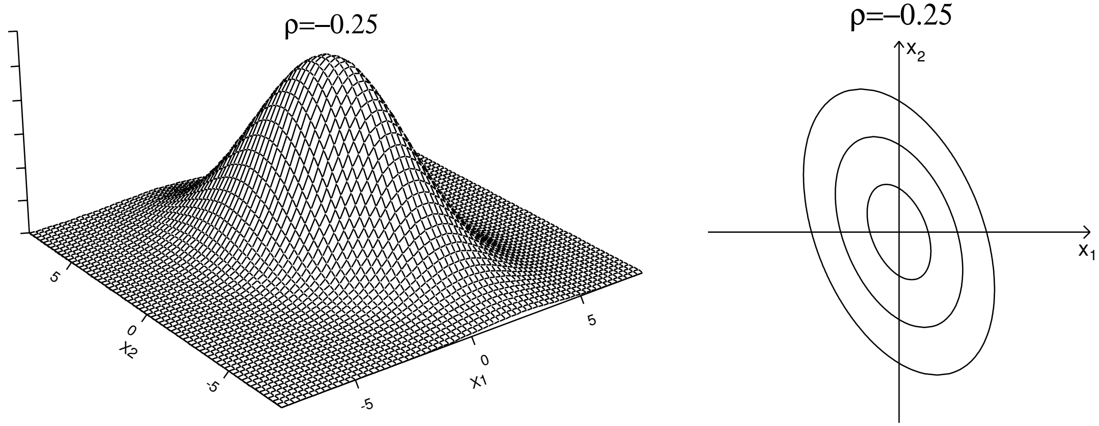
- Equidensity contours are ellipsoids in $\mathbb{R}^p$ since:

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) = \text{const.} \iff (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \text{const.}.$$

Bivariate case: equidensity contours are ellipses



$\mathcal{N}_2(\mathbf{0}, \mathbf{I})$: p.d.f. and the equidensity contours.



$\mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma})$ ($\rho = \mathrm{Corr}(x_1, x_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = -0.25$): p.d.f. and contours.

Some Properties

(N1) If $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}$ is a $q \times p$ matrix with $\mathrm{rank}(\mathbf{A}) = q \leq p$, then

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{b} \sim \mathcal{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$$

(N2) If $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $x_1, \ldots, x_p$ are mutually independent $\iff$ $\sigma_{ij} = \mathrm{Cov}(x_i, x_j) = 0 \; \forall i \neq j$.

(N3) If

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathcal{N}_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right),$$

then $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent $\iff$ $\boldsymbol{\Sigma}_{12} = 0$.

(N4) If $\mathbf{x}_1 \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{x}_2 \sim \mathcal{N}_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ are independent, then

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathcal{N}_{p+k}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right).$$

and conversely.

(N5) Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and partition $\mathbf{x}$, $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x})$ as:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

where $\mathbf{x}_1$ and $\boldsymbol{\mu}_1$ are k-vectors $(k < p)$ and $\boldsymbol{\Sigma}_{11} = \mathrm{Cov}(\mathbf{x}_1)$ a $k \times k$ matrix.

Then

$$\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \quad \text{and} \quad \mathbf{x}_2 \sim \mathcal{N}_{p-k}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

and the conditional distribution of $\mathbf{x}_2$ given $\mathbf{x}_1$ is $\mathcal{N}_{p-k}(\boldsymbol{\mu}_{2|1}, \boldsymbol{\Sigma}_{2|1})$, where

$$\begin{aligned}
\boldsymbol{\mu}_{2|1} &= \mathbb{E}[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \\
\boldsymbol{\Sigma}_{2|1} &= \mathrm{Cov}(\mathbf{x}_2|\mathbf{x}_1) = \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.
\end{aligned}$$