Bayesian Learning and Spatio-temporal Statistics

On some properties of the multivariate Gaussian distribution

François Septier

http://www.univ-ubs.fr/septier/

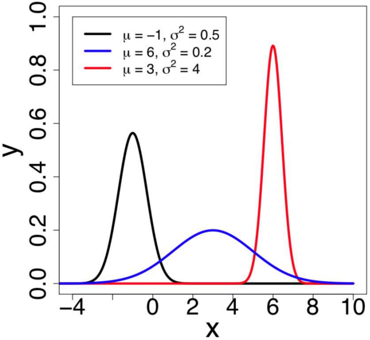
francois.septier@univ-ubs.fr

Scalar Gaussian (Normal) distribution

For a a scalar variable x, the Gaussian distribution can be written on the form

$$\mathcal{N}(x;\mu,\sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{Z} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

- μ is the mean (expected value of the distribution)
- σ is the standard deviation
- σ^2 is the variance
- Z is the normalization constant



What if $\boldsymbol{x}$ is a vector $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}^T$?

Multivariate Gaussian

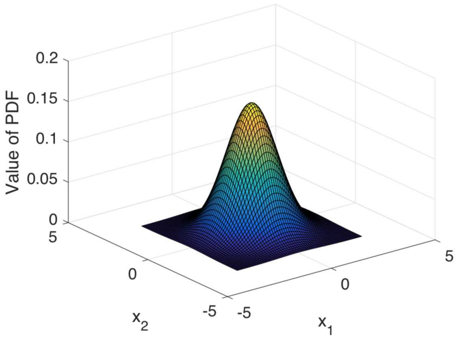
For a D-dimensional vector $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}^T$, the **multivariate** Gaussian distribution can be written on the form

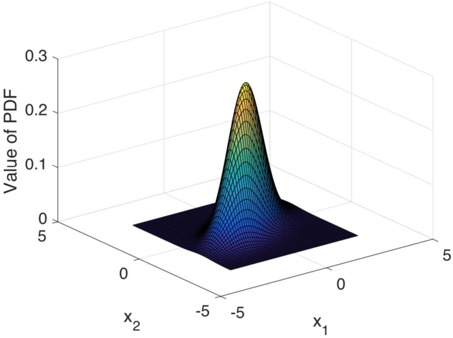
$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \underbrace{\frac{1}{(2\pi)^{D/2} \det(\boldsymbol{\Sigma})^{1/2}}}_{Z} \exp\left(-\frac{1}{2} \underbrace{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}_{\text{quadratic form}} \right)$$

- $\boldsymbol{\mu}$ is the mean vector
- $\boldsymbol{\Sigma}$ is the covariance matrix
- Z is the normalization constant

Remark: Gaussian $\propto \exp($quadratic form on $\boldsymbol{x})$

Multivariate Gaussian





$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}$$

Partitioned Gaussian - marginalization

Partition the Gaussian random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} \in \mathbb{R}^n$ into two sets of random vectors, $\boldsymbol{x}_a \in \mathbb{R}^{n_a}$ and $\boldsymbol{x}_b \in \mathbb{R}^{n_b}$, such that

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}, \quad .$$

Task: Compute the marginal distribution of the random vector $\boldsymbol{x}_a$,

$$p(\boldsymbol{x}_a) = \int p(\boldsymbol{x}_a, \boldsymbol{x}_b) d\boldsymbol{x}_b$$

Partitioned Gaussian - marginalization

Theorem 1 (Marginalization)

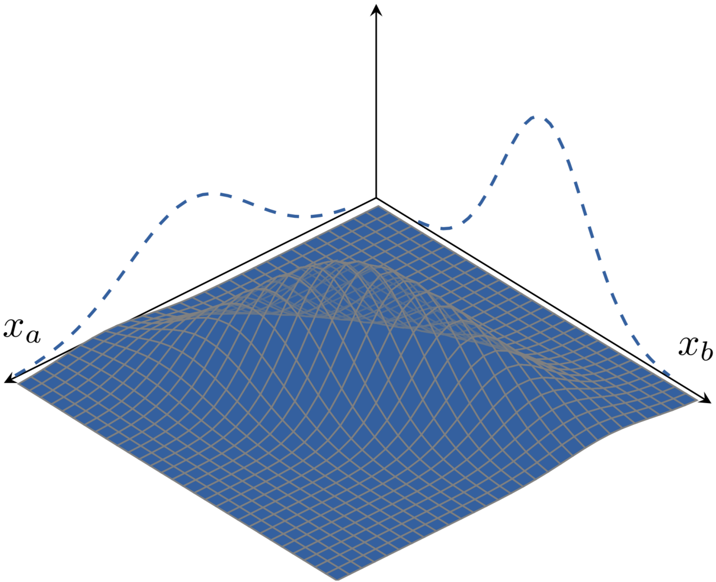
Partition the Gaussian random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, according to*

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}, \quad .$$

The marginal distribution $p(\boldsymbol{x}_a)$ *is then given by*

$$p(\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_a; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$$

Partitioned Gaussian - marginalization



Partitioned Gaussian - Theorems

$$p(\mathbf{x}_a, \mathbf{x}_b)$$

Thm 1

$$p(\mathbf{x}_a)$$

$$p(\mathbf{x}_b)$$

Partitioned Gaussian - conditioning

Theorem 2 (Conditioning)

Partition the Gaussian random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *according to*

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}, \quad .$$

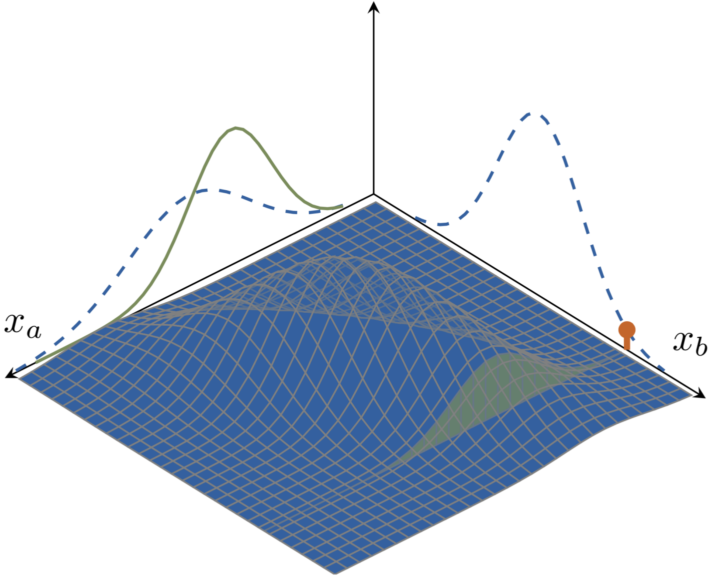
The conditional distribution $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$ *is then given by*

$$p(\boldsymbol{x}_a|\boldsymbol{x}_b) = \mathcal{N}(\boldsymbol{x}_a; \boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b})$$

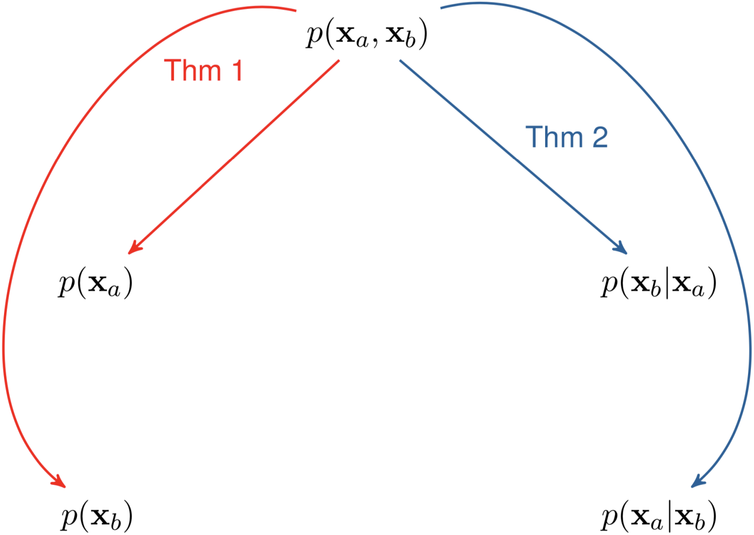
with

$$\begin{aligned} \boldsymbol{\mu}_{a|b} &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b), \\ \boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}. \end{aligned}$$

Partitioned Gaussian - conditioning



Partitioned Gaussian - Theorems



Affine transformation of multivariate Gaussian r.v.

We can also be interested to do the opposite $\mapsto$ compute $p(\boldsymbol{x}_a, \boldsymbol{x}_b)$ based on $p(\boldsymbol{x}_b|\boldsymbol{x}_a)$ and $p(\boldsymbol{x}_a)$

Theorem 3 (Affine transformation)

Assume that $\boldsymbol{x}_a$, as well as $\boldsymbol{x}_b$ conditioned on $\boldsymbol{x}_a$, are normally distributed according to

$$p(\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_a; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a),$$
$$p(\boldsymbol{x}_b|\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_b; \boldsymbol{M}\boldsymbol{x}_a, \boldsymbol{\Sigma}_{b|a}).$$

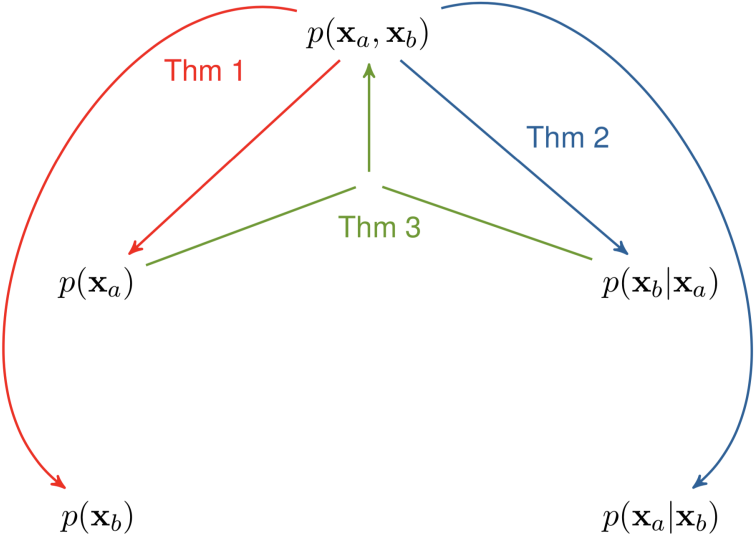
Then the joint distribution of $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is

$$p(\boldsymbol{x}_a, \boldsymbol{x}_b) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}_a \\ \boldsymbol{x}_b\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_a \\ \boldsymbol{M}\boldsymbol{\mu}_a\end{bmatrix}, \boldsymbol{R}\right)$$

with

$$\boldsymbol{R} = \begin{bmatrix}\boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_a \boldsymbol{M}^T \\ \boldsymbol{M}\boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_{b|a} + \boldsymbol{M}\boldsymbol{\Sigma}_a\boldsymbol{M}^T\end{bmatrix}$$

Partitioned Gaussian - Theorems



Affine transformation of multivariate Gaussian r.v.

We can also be interested [e.g. **Bayesian**] to compute directly $p(\boldsymbol{x}_a|\boldsymbol{x}_b)$ (posterior) based on $p(\boldsymbol{x}_b|\boldsymbol{x}_a)$ (likelihood) and $p(\boldsymbol{x}_a)$ (prior) [Use Theorem 1, 2 and 3]

Corollary 1 (Affine transformation + Conditional)

Assume that $\boldsymbol{x}_a$, as well as $\boldsymbol{x}_b$ conditioned on $\boldsymbol{x}_a$, are normally distributed according to

$$p(\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_a; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a),$$
$$p(\boldsymbol{x}_b|\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_b; \boldsymbol{M}\boldsymbol{x}_a + \mathbf{b}, \boldsymbol{\Sigma}_{b|a}).$$

Then the marginal density of $\boldsymbol{x}_b$ is given by

$$p(\boldsymbol{x}_b) = \mathcal{N}\left(\boldsymbol{x}_b; \boldsymbol{\mu}_b; \boldsymbol{\Sigma}_b\right)$$

with

$$\boldsymbol{\mu}_b = \boldsymbol{M}\boldsymbol{\mu}_a + \boldsymbol{b},$$
$$\boldsymbol{\Sigma}_b = \boldsymbol{\Sigma}_{b|a} + \boldsymbol{M}\boldsymbol{\Sigma}_a\boldsymbol{M}^T.$$

Then the joint distribution of $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is

$$p(\boldsymbol{x}_a|\boldsymbol{x}_b) = \mathcal{N}\left(\boldsymbol{x}_a; \boldsymbol{\mu}_{a|b}; \boldsymbol{\Sigma}_{a|b}\right) \quad \textit{with}$$

$$\begin{aligned}\boldsymbol{\mu}_{a|b} &= \boldsymbol{\Sigma}_{a|b}\left(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a + \boldsymbol{M}^T\boldsymbol{\Sigma}_{b|a}^{-1}(\boldsymbol{x}_b - \boldsymbol{b})\right)\\ &= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_a\boldsymbol{M}^T\boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}_b - \boldsymbol{b} - \boldsymbol{M}\boldsymbol{\mu}_a),\\ \boldsymbol{\Sigma}_{a|b} &= \left(\boldsymbol{\Sigma}_a^{-1} + \boldsymbol{M}^T\boldsymbol{\Sigma}_{b|a}^{-1}\boldsymbol{M}\right)^{-1}\\ &= \boldsymbol{\Sigma}_a - \boldsymbol{\Sigma}_a\boldsymbol{M}^T\Sigma_b^{-1}\boldsymbol{M}\boldsymbol{\Sigma}_a.\end{aligned}$$

** last line obtained by using the Woodbury matrix identity*

Partitioned Gaussian - Theorems