# Lecture 3: Gaussian Random Vectors and applications

Sébastien Gadat

APM 3F004 Ecole Polytechnique

September 19, 2025

# Gaussian Distributions

Let $(\Omega, \mathcal{P})$ be a parametric model with

$$\mathcal{P} = \{\mathbb{P}_\theta;\ \theta \in \Theta\}.$$

- **Objective 1 :** $\theta = (\mu, \sigma^2)$ for univariate random variables
- **Objective 2 :** $\theta = (\mu, \Sigma^2)$ for multivariate random variables
- Beyond the simple definitions : key properties ?
- Point estimation problem : How to estimate $\mu$ ? How to estimate $\Sigma^2$ ?

# Definition

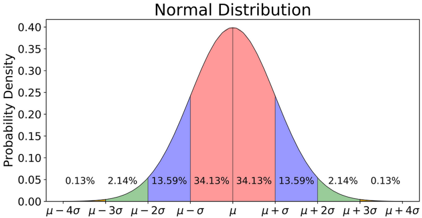
**Univariate Gaussian distribution** : $\mathcal{N}(\mu, \sigma^2)$ defined through its p.d.f. :

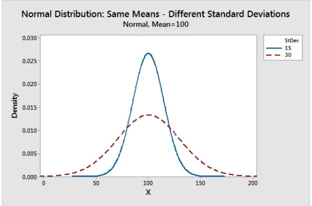
$$\forall x \in \mathbb{R} \qquad \gamma_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## Proposition

*$\mu$ is the mean and $\sigma^2$ the variance of $\mathcal{N}(\mu, \sigma^2)$ :*

$$\mu = \mathbb{E}_\theta[X] \qquad \sigma^2 = \mathbb{E}_\theta[(X-\mu)^2.]$$



# Sampling a univariate Gaussian distribution

Major issue : The C.D.F. is not explicit... The following method :

$$F_{\mu,\sigma^2}^{-1}(\mathcal{U}) \sim \mathcal{N}(\mu, \sigma^2)$$

is untractable !

Fortunately, the Box-Muller algorithm works !

- Sample $R > 0$ with

$$R = \sqrt{-2\log(\mathcal{U}(0,1))}$$

- Sample $\theta \sim \mathcal{U}(0, 2\pi)$
-
$$X = R\cos(\theta) \qquad \text{and} \qquad Y = R\sin(\theta)$$
- Then $X$ and $Y$ are *independent* and $\mathcal{N}(0,1)$.

# Sampling a univariate Gaussian distribution - Proof :

# Moment Generating Transform : a key object

**Definition (Laplace / Moment Generating Transform)**

- *Univariate case* For any real random variable $X$, we define $m_X$ the function defined as :

$$\forall t \in \mathcal{D}_X \subset \mathbb{R} \qquad m_X(t) = \mathbb{E}[e^{tX}].$$

- *Multivariate case* For any multivariate random variable $X$, we define $m_X$ the function defined as :

$$\forall t \in \mathcal{D}_X \subset \mathbb{R}^d \qquad m_X(t) = \mathbb{E}[e^{\langle t, X \rangle}].$$

$\mathcal{D}_X$ is the domain where the Laplace transform is defined. For certain distributions, $\mathcal{D}_X = \mathbb{R}$ or $\mathbb{R}^d$, for other ones $\mathcal{D}_X$ is only defined on intervals or rectangles.

# Moment Generating Transform : two key properties

### Theorem (One to one association with M.G.T.)

*For any random variables/vectors* $(X_1, X_2)$ *: the two properties are* **equivalent**

- *i) The c.d.f. of* $X_1$ *and* $X_2$ *are identical*
- *ii)* $m_{X_1} = m_{X_2}$ *in an open rectangle that includes* $0$.

In practice : M.G.T., identify $m_{X_1} = m_{X_2}$, deduce $\mathcal{L}(X_1) = \mathcal{L}(X_2)$.

### Theorem (Independence)

*For any r.v.* $X = (X_1, \ldots, X_d)$ *: the two properties are* **equivalent**

- $X_1, \ldots, X_d$ *are mutually independent*
- $$\forall t = (t_1, \ldots, t_d) \qquad m_X(t) = m_{X_1}(t_1) \times \ldots \times m_{X_d}(t_d),$$
  *in an open rectangle that includes* $0$.

# Moment Generating Transform

**Theorem (Laplace / Moment Generating Transform)**

- *Consider $t \in \mathbb{R}$ and $X \sim \mathcal{N}(0,1)$ then :*

$$\mathbb{E}[e^{tX}] = e^{t^2/2}$$

- *If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then*

$$\mathbb{E}[e^{tY}] = e^{t\mu + \sigma^2 t^2/2}$$

**Important conclusion !**

**Theorem**

- *If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $\sigma^{-1}(Y - \mu) \sim \mathcal{N}(0,1)$*
- *If $X \sim \mathcal{N}(0,1)$, then $Y = \mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$.*

# Sum of Gaussian random variables

## Important consequence

### Theorem

*If* $(X_1, \ldots, X_n)$ *are* $n$ **independent** *Gaussian r.v.* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$*, then*

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Proof :

# Definition : Multivariate Standard Gaussian

**Multivariate Standard Gaussian** : $\mathcal{N}(0, I_d)$ defined through its p.d.f. :

$$\gamma_{0,I_d}(x) = \frac{1}{\sqrt{|2\pi I_d|}} e^{-\frac{\|x\|^2}{2}} = (2\pi)^{-d/2} e^{-\frac{\|x\|^2}{2}}$$

## Proposition

- *0 is the mean and $I_d$ is the covariance of $\mathcal{N}(0, I_d)$ :*

$$0 = \mathbb{E}_\theta[X] \qquad I_d = \mathbb{E}_\theta[XX^T].$$

- *The M.G.T. is*

$$\forall t \in \mathbb{R}^d \qquad m_X(t) = e^{\|t\|^2/2} = e^{\frac{t_1^2+\ldots+t_d^2}{2}}$$

- *$X_1, \ldots, X_d$ are independent.*

# Definition : Multivariate Gaussian

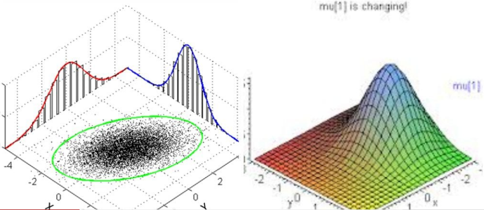
**Multivariate Gaussian distribution** : $\mathcal{N}(\mu, \Sigma^2)$ defined through its p.d.f. ($\Sigma^2$ invertible) :

$$\gamma_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{|2\pi\Sigma^2|}} e^{-\frac{(x-\mu)^T \Sigma^{-2}(x-\mu)}{2}}$$

## Proposition

*$\mu$ is the mean and $\Sigma^2$ is the covariance of $\mathcal{N}(\mu, \Sigma^2)$ :*

$$\mu = \mathbb{E}_\theta[X] \qquad \Sigma^2 = \mathbb{E}_\theta[(X-\mu)(X-\mu)^T].$$



# Proof of these results ?

- Starting point : study the case $\mathcal{N}(0, I_d)$ :

$$\mathbb{E}[X] = 0 \quad \text{and} \quad \mathbb{E}[XX^T] = I_d.$$

- Prove that $X \sim \mathcal{N}(\mu, I_d) \Longleftrightarrow X - \mu \sim \mathcal{N}(0, I_d)$
- Prove that $X \sim \mathcal{N}(0, I_d) \Longrightarrow AX \sim \mathcal{N}(0, AA^T)$
- Prove that $X \sim \mathcal{N}(\mu, \Sigma^2) \Longrightarrow AX \sim \mathcal{N}(A\mu, A\Sigma^2 A^T)$

### Theorem

*If $X \sim \mathcal{N}(\mu, \Sigma^2)$ and $\Sigma^2$ invertible, then :*

- $Y = \Sigma^{-1}(X - \mu) \sim \mathcal{N}(0, I_d)$.
- $X =^{\mathcal{L}} \Sigma Y + \mu$ *where* $Y \sim \mathcal{N}(0, I_d)$
- *If $v$ is any vector of $\mathbb{R}^d$, then* $\langle v, X \rangle \sim \mathcal{N}(\langle \mu, v \rangle, |\Sigma v|^2)$.

# Definition : Multivariate Gaussian

**Multivariate Gaussian distribution** : We shall define indeed $\mathcal{N}(\mu, \Sigma^2)$ through :

### Definition

If $Z \sim \mathcal{N}(0, I_d)$, $X$ is a r.v. $\mathcal{N}(\mu, \Sigma^2)$ if and only if

$$X = \mu + \sqrt{\Sigma^2} Z = \mu + \Sigma Z$$

Below, we oftenly use $A$ a symetric positive definite matrix such that

$$A = \sqrt{\Sigma^2} \Longleftrightarrow AA^T = AA = \Sigma^2.$$

$A$ is the so-called square root matrix of $\Sigma^2$.
Another important implicit hypothesis : $\Sigma^2$ is assumed to be invertible

# M.G.T. of Multivariate Gaussian

Theorem

*For any $X \sim \mathcal{N}(\mu, \Sigma^2)$, we have :*

$$\forall t \in \mathbb{R}^d \qquad m_X(t) = e^{\langle t,\mu\rangle + \frac{t^T \Sigma^2 t}{2}}$$

Key consequence : If $\Sigma^2$ is a **diagonal matrix**, then

$$m_X(t) = m_{X_1}(t_1) \times \ldots \times m_{X_d}(t_d),$$

which implies that $X_1, \ldots, X_d$ are mutually independent !

**For Gaussian random vectors : "independence $\Longleftrightarrow$ no correlation"**

# Sampling a multivariate Gaussian distribution

Starting point : $X \sim \mathcal{N}(\mu, \Sigma^2) \Longrightarrow X =^{\mathcal{L}} \mu + \Sigma Y$ where $Y \sim \mathcal{N}(0, I_d)$.

- Compute $A$ such that
$$AA^T = \Sigma^2$$
- Sample $Y \sim \mathcal{N}(0, I_d)$ with the Box-Muller algorithm.
- Compute $X = \mu + AY$.

# Operations on multivariate Gaussian vectors

Two important manipulations :
**Linear combinations :** $\langle X, a \rangle$

### Proposition

*Assume* $X \sim \mathcal{N}(\mu, \Sigma^2)$ *and* $a \in \mathbb{R}^d$, *then*

$$\langle X, a \rangle \sim \mathcal{N}(\langle \mu, a \rangle, a^T \Sigma^2 a).$$

A linear combination of a normal r.v. is still a normal r.v.
**Affine transformation :** $Y = a + BX$

### Proposition

*Assume* $X \sim \mathcal{N}(\mu, \Sigma^2)$ *and* $Y = a + BX$, *then*

$$Y \sim \mathcal{N}(B\mu + a, B\Sigma^2 B^T).$$

# Chi-square distribution

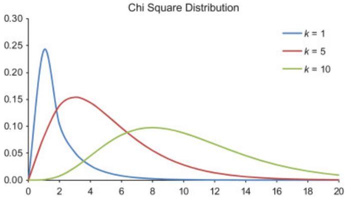
Consider $d \geq 1$ and $X_1, \ldots, X_d$ i.i.d. $\mathcal{N}(0,1)$, then :

$$Z = \sum_{i=1}^{d} X_i^2 \sim \chi^2(d) \quad \text{and} \quad \mathbb{E}[Z] = d \quad \text{and} \quad Var(Z) = 2d.$$

### Proposition

*The density of $Z \sim \chi^2(d)$ is :*

$$\forall x \in \mathbb{R} \qquad f_d(x) = \frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} e^{-x/2} \mathbb{1}_{x \geq 0}.$$



# Mean and Variance of a Gaussian sample

Assume that $X_1, \ldots X_n$ are i.i.d. sample of $\mathcal{N}(\mu, \sigma^2)$.
Then define

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n} \quad \text{and} \quad \widehat{\sigma_n^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

**Theorem (Gosset's Theorem)**

- $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$
- *$\bar{X}_n$ and $\widehat{\sigma_n^2}$ are independent*
- $$\frac{(n-1)\widehat{\sigma_n^2}}{\sigma^2} \sim \chi^2(n-1)$$

# Projections of Gaussian vectors

We define :

$$\forall d \geq r \geq 1 \qquad J_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Consider $P$ an orthogonal projection matrix of rank $r$ :

$$P^2 = P \quad \text{and} \quad \exists Q \quad P = Q^T J_r Q \quad \text{with} Q^{-1} = Q^T.$$

Theorem (Cochran's theorem)

*Assume that $X \sim \mathcal{N}(0, I_d)$, then $\|PX\|^2 \sim \chi^2(r)$ and*

$$PX \quad \textit{and} \quad X - PX \quad \textit{are independent random variables}$$

# Linear model

We observe $(X_i, Y_i)_{1\leq i\leq n}$ i.i.d. and assume a Gaussian linear model $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}^d}$ :

$$Y = \langle X, \theta\rangle + \varepsilon \qquad \text{with} \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

Notation : $\mathbb{X} = (X_1, \ldots, X_n)^T$ and $\mathbb{Y} = (Y_1, \ldots, Y_n)^T$.

$$\hat{\theta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y} \sim \mathcal{N}(\theta, \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}) \quad \text{and} \quad \hat{\sigma}^2 = \frac{\|\mathbb{Y} - \mathbb{X}\hat{\theta}\|^2}{n}.$$

We observe that $P = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ satisfies :

$$P^2 = P \qquad \text{and} \qquad \langle Py, y - Py\rangle = 0.$$

Consequence : $P$ is a matrix projection on $\mathbb{X}$.

# Linear model

We observe $(X_i, Y_i)_{1\leq i\leq n}$ i.i.d. and assume a Gaussian linear model $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}^d}$ :

$$Y = \langle X, \theta\rangle + \varepsilon \qquad \text{with} \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

Consequence on the variance :

$$n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-d)$$

Consequence for statistics :

$$\hat{\theta} \quad \text{and} \quad \hat{\sigma} \quad \text{are independent}$$

Consequence for statistical testing :

$$H_0 : \text{``}\theta_j = 0'' : \frac{\sqrt{n}\hat{\theta}_j}{\sqrt{\hat{\sigma}^2(\mathbb{X}^T\mathbb{X})^{-1}_{j,j}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2(n-d)}} \sim \mathcal{T}_{n-p}$$