Lecture 5: Gaussian Vectors and applications to linear models

Sébastien Gadat

TSE

November 19, 2023

1. Gaussian Distributions
   - Definitions and basic properties
   - Sampling a (multivariate) Gaussian distribution
   - Fourier Transform / Characteristic Function
   - Moment Generating Function

2. Important results on Gaussian Vectors for Statisticians

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 μ ? How to estimate Σ^2 ?
- Mean Square Error (M.S.E. for short) in the linear model situation ?

$$R(\theta_0, \hat{\theta}_n) = \mathbb{E}_{\theta_0}[\|\theta_0 - \hat{\theta}_n\|^2]$$

Definition

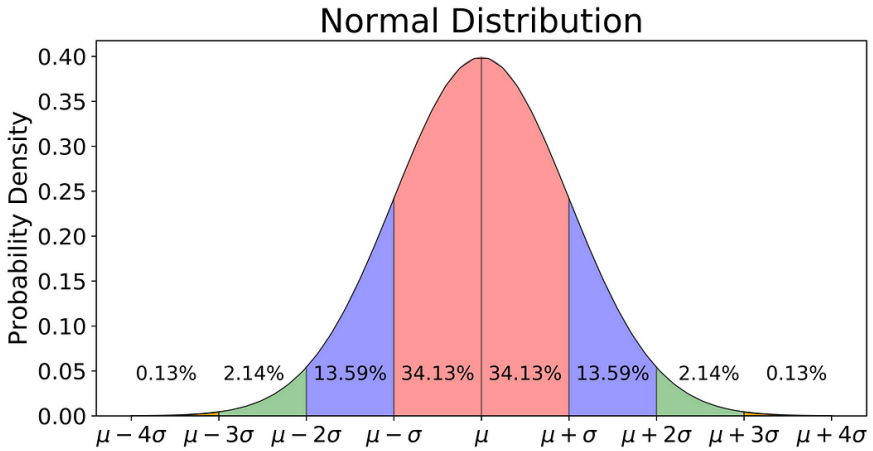
Univariate Gaussian distribution : $\mathcal{N}(\mu, \sigma^2)$ defined through its p.d.f. :

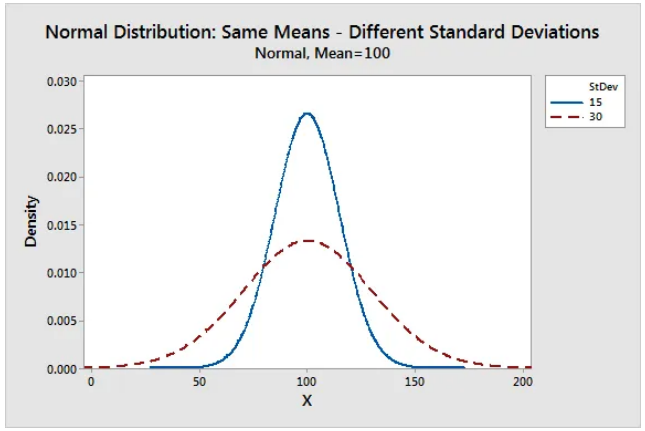
$$\gamma_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Proposition

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

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





Definition

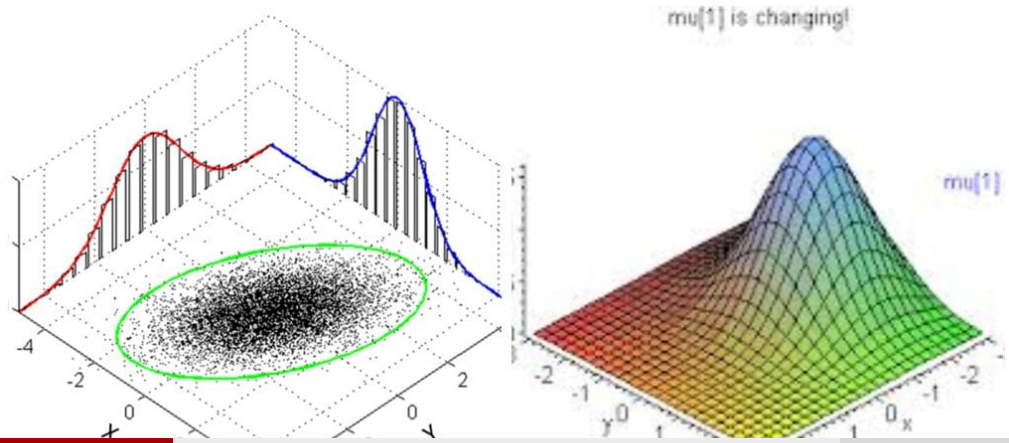
Multivariate Gaussian distribution : $\mathcal{N}(\mu, \Sigma^2)$ defined through its p.d.f. (Σ^2 invertible) :

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

Proposition

μ is the mean and Σ^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 $X \sim \mathcal{N}(\mu, I_d) \Longrightarrow X - \mu \sim \mathcal{N}(0, I_d)$
- 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 Σ^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 $\langle v, X \rangle \sim \mathcal{N}(\langle \mu, v \rangle, |\Sigma v|^2)$.*

Sampling a univariate Gaussian distribution

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

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

is untractable !
Fortunately, the Box-Muller algorithm works !

- Sample $R > 0$ with

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

- Sample $\theta \sim \mathcal{U}(0, 2\pi)$
- $$X = R\cos(\theta) \qquad \text{and} \qquad Y = R\sin(\theta)$$
- $(X, Y) \sim \mathcal{N}(0, I_2)$.

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 $X = \mu + AY$.

Characteristic function of a Gaussian distribution

Standard normal case

Theorem

Consider $\xi \in \mathbb{R}^d$ *and* $X \sim \mathcal{N}(0, I_d)n$ *then :*

$$\varphi_X(\xi) = \mathbb{E}[e^{\mathrm{i}\langle \xi, X\rangle}] = e^{-\frac{1}{2}\|\xi\|^2}$$

Multivariate normal case

Theorem

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

$$\varphi_X(\xi) = \mathbb{E}[e^{\mathrm{i}\langle \xi, X\rangle}] = e^{\mathrm{i}\langle \mu, \xi\rangle - \frac{1}{2}\xi^T \Sigma^2 \xi}$$

Moment Generating Function of a Gaussian distribution

Theorem

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 consequence (either with M.G.F. of Fourier transform)

Theorem

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

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

Chi-square distribution

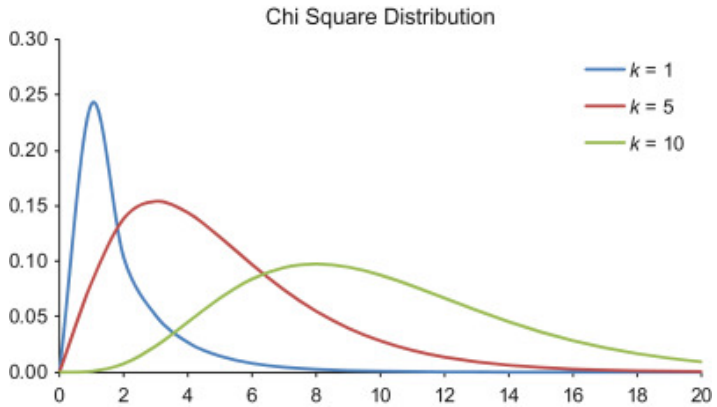
Consider $d \geq 1$ and $X_1, \ldots, X_d$ i.i.d. $\mathcal{N}(0,1)$, then :

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

Proposition

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

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



Independence of Gaussian vectors

Proposition

Assume that X, Y are two Gaussian vectors, then X and Y are independent if and only if $Cov(X, Y) = 0$.

This result is straightforward using the characteristic function of the random variable (X, Y) !

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

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

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 :

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