M2 Statistics & Data Science

Advanced Statistics & Machine Learning

Faicel Chamroukhi
Professeur
https://chamroukhi.com/

Overview

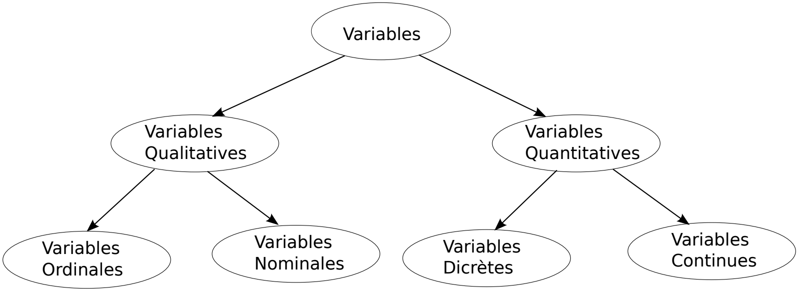
Rappels

1. Rappels
   - Variables aléatoires
   - Couple de variables aléatoires
   - Vecteurs Aléatoires
   - Variables et vecteurs aléatoires gaussiens

Variables aléatoires

- Variables aléatoires

Variables aléatoires : récapitulatif



Fonction de densité de probabilité

Definition

On appelle densité de probabilité toute fonction continue par morceaux (intégrable) :

$$\begin{aligned} f:\mathbb{R} &\rightarrow \mathbb{R}^{+} \\ x &\rightarrow f(x) \end{aligned} \tag{1}$$

telle que :

$$\begin{array}{ll} (1) & \forall x \in \mathbb{R} \quad f(x) \geq 0 \\ (2) & \displaystyle\int_{\mathbb{R}} f(x)dx = 1 \end{array}$$

Espérance mathématique d'une v.a

L'espérance mathématique d'une variable aléatoire réelle représente la valeur moyenne prise par cette variable aléatoire.

Espérance mathématique d'une variable aléatoire discrète

Soit X une variable aléatoire discrète prenant ses valeurs dans l'ensemble $\{x_1, \ldots, x_n\}$ avec des probabilités respectives $p_1, \ldots, p_n$ ç.à.d $P(X = x_k) = p_k$ et $\sum_{k=1}^{n} p_k = 1$. L'espérance de X est donnée par

$$\mathbb{E}[X] = \sum_k x_k P(X = x_k) = \sum_k x_k p_k \tag{2}$$

⚠ Remarque : On peut remarquer que l'espérance d'une variable aléatoire binaire est égale à la sa probabilité de valoir 1, autrement dit si X est une v.a binaire (prenant ses valeurs dans $\{0, 1\}$, on a alors $\mathbb{E}[X] = P(X = 1)$.

Espérance mathématique d'une v.a

Espérance mathématique d'une variable aléatoire continue

Soit X une variable aléatoire continue à valeurs réelles ayant comme fonction de densité de probabilité f. L'espérance de X est donnée par

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx. \tag{3}$$

Plus généralement, soit $g : \mathbb{R} \to \mathbb{R}$, l'espérance de $g(X)$ est donnée par

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx \tag{4}$$

Variance d'une variable aléatoire

Definition

On appelle variance d'une variable aléatoire X la quantité définie par :

$$var[X] = \mathbb{E}[(X - E[X])^2] = \mathbb{E}[X^2] - (E[X])^2 \quad (5)$$

Ainsi, la variance notée σ_X^2 et sa racine carrée l'écart-type (noté σ_X) mesurent la dispersion de la variable aléatoire X autour de son espérance $E[X]$.

⚠ Remarque : D'après (4) et (5), la variance est donc aussi donnée par

$$var[X] = \int_{\mathbb{R}} (x - E[X])^2 f(x) dx. \quad (6)$$

Quelques propriétés de l'espérance et de la variance

Soit X une v.a réelle, deux fonctions $g, h : \mathbb{R} \to \mathbb{R}$ et deux réels a et b, alors

$$\mathbb{E}[a.g(X) + b.h(X)] = a.\mathbb{E}[g(X)] + b.\mathbb{E}[h(X)]$$

ainsi le cas particulier affine consiste en la propriété suivante :

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$\Rightarrow$ On dit que l'espérance est **linéaire**
De même, pour la variance, on a la propriété suivante :

$$Var(aX + b) = a^2 Var(X)$$

$\Rightarrow$ On dit que la variance est **quadratique**

Quelques statistique sur variables aléatoires

Definition

Soit $(X_1, \ldots, X_n)$ un échantillon donné d'une population ayant comme fonction de densité de probabilité la fonction f de paramètre θ (une distribution de probabilités $P(.;\theta)$ pour le cas discret). Une *statistique* est toute fonction de l'échantillon donné $(X_1, \ldots, X_n)$ qui ne dépend pas de θ.

moyenne empirique

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7}$$

variance empirique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2 \,. \tag{8}$$

Couple de variables aléatoires

- Couple de variables aléatoires

Couple de variables aléatoires I

Lois jointes, lois conditionnelles et théorème de Bayes

Soient X et Y deux variables aléatoires de densités respectives $f(x)$ et $f(y)$ (ou de lois respectives $P(X)$ et $P(Y)$ pour le cas discret).

Le comportement du couple (X, Y) est entièrement décrit par leur fonction de **densité de probabilité jointe** $f_{XY}(x, y)$ (ou **distribution jointe** de probabilité dans le cas discret $P_{XY}(X = x, Y = y)$) .

Il se peut que lors du tirage d'une réalisation de (X, Y), que la valeur observée (réalisation) x de X fournisse une information sur la valeur possible de Y.

$\Rightarrow$ Cette information est représentée par la distribution de Y conditionnellement à $X = x$ (distribution de Y étant donné à $X = x$) soit $f_{Y|X}(y|x)$.

Théorème de Bayes

Ceci est explicité par le théorème de suivant, appelé **théorème de Bayes** (ou aussi règle de Bayes ou formule de Bayes) comme suit

Theorem

$$f(x, y) = f(y|x)f(x) \tag{9}$$

et par symétrie on a également

$$f(x, y) = f(x|y)f(y) \; . \tag{10}$$

Pour les VA discrètes cela revient à

Theorem

$$\begin{aligned} P(Y = y, X = x) &= P(Y = y|X = x)P(X = x) & (11) \\ &= P(X = x|Y = y)P(Y = y) & (12) \end{aligned}$$

Loi marginale de v.a discrètes

La loi de chaque variable est donc donnée par la **loi marginale** comme suit :

$$P(X = x) = \sum_{y \in B} P(X = x, Y = y) \tag{13}$$

et

$$P(Y = y) = \sum_{x \in A} P(X = x, Y = y) \tag{14}$$

⚠ Remarque : On peut remarquer que ces lois correspondent respectivement à la somme des lignes et la somme des colonnes dans le tableau précédent (??).

Densités marginales

Les variables X et Y sont des variables continues de densité respectives

$$f(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)dy \tag{15}$$

et

$$f(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)dx. \tag{16}$$

Covariance de deux variables aléatoires

Pour une variable aléatoire, on parle de variance
pour un couple de deux variables aléatoires, on parle de **covariance**
La covariance mesure la dépendance entre deux v.a. (Si les deux v.a sont indépendantes, leur covariance est donc nulle.)

Definition

On définit la covariance de deux variables aléatoires X et Y comme le terme

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{17}$$

et on la propriété suivante

Definition

$$cov(X, Y) = \sigma_{XY} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{18}$$

Covariance de deux variables aléatoires I

⚠ Remarque : L'espérance du produit $E(XY)$ se calcule à partir de la loi jointe du couple (X, Y).
Ainsi, dans le cas discret, si X prend ses valeurs dans A et Y prends ses valeurs dans B, on a

$$\mathbb{E}[XY] = \sum_{x \in A} \sum_{y \in B} xyP(X = x, Y = y) \tag{19}$$

Dans le cas de v.a réelles continu, on a

$$\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_{X,Y}(x, y) \,\mathrm{d}\, x \,\mathrm{d}\, y. \tag{20}$$

Indépendance de deux variables aléatoires

Soient X et Y deux variables aléatoires de densités respectives $f(x)$ et $f(y)$ (ou de lois respectives $P(X)$ et $P(Y)$ pour le cas discret).

Definition

X et Y sont dites indépendantes si et seulement si leur densité de probabilité jointe est égale au produit de leurs densités marginales. Plus spécifiquement

$$f_{XY}(x, y) = f(x)f(y) \tag{21}$$

Pour les VA discrètes cela revient à

$$P(Y = y, X = x) = P(X = x)P(Y = y) \tag{22}$$

Indépendance et espérance

Theorem

X et Y sont indépendantes si et seulement si pour toutes fonctions $f(X)$ et $g(Y)$

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

si les espérances existent.

Le théorème direct est facile, la réciproque un peu plus difficile. Un cas particulier est bien celui où Si X et Y sont indépendantes, alors

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Vecteurs Aléatoires

- Vecteurs Aléatoires

Vecteurs Aléatoires

Definition

Un vecteur aléatoire réel $\boldsymbol{X}$ de dimension d est un vecteur $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ dont les composantes X_j, $j = 1, \ldots, d$ sont des variables aléatoires réelles.

Espérance et matrice de covariance d'un vecteur aléatoire

Soit $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ un vecteur aléatoire réel.

Espérance

L'espérance du vecteur aléatoire $\mathbf{X}$ est donnée par le vecteur déterministe

$$\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])^T = (\mu_1, \ldots, \mu_d)^T \text{ avec } \mu_j = \mathbb{E}[X_j], j = 1, \ldots, d$$

Matrice de covariance

La matrice de variance-covariance (appelée aussi matrice de covariance) d'un vecteur aléatoire $\boldsymbol{X}$ est la matrice carrée parfois notée $\boldsymbol{\Sigma}$ dont le terme générique est donné par : $\boldsymbol{\Sigma}_{i,j} = \mathrm{cov}\left(X_i, X_j\right)$ Elle est définie comme :

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathrm{cov}(\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\right] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \mathbb{E}[\boldsymbol{X}](\mathbb{E}[\boldsymbol{X}])^T, \tag{23}$$

$\mathbb{E}[\boldsymbol{X}]$ étant l'espérance mathématique de $\boldsymbol{X}$.

Espérance et matrice de covariance d'un vecteur aléatoire

En développant les termes on obtient la forme suivante :

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}$$

$$= \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1X_2) & \cdots & \mathrm{cov}(X_1X_p) \\ \mathrm{cov}(X_2X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_PX_1) & \cdots & \cdots & \mathrm{var}(X_p) \end{pmatrix} = \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_p} \\ \sigma_{X_2X_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_pX_1} & \cdots & \cdots & \sigma^2_{X_p} \end{pmatrix}$$

où les $\sigma^2_{X_j}, j = 1, \ldots, p$ sont les variances respectives des v.a X_j et les $\sigma^2_{X_i,X_j}, i \neq j$ sont les covariances respectives des couples de v.a (X_i, X_j) : $\sigma^2_{X_i,X_j} = cov(X_i, X_j)$.

Quelques propriétés de la matrice de covariance

La matrice de covariance possède les propriétés suivantes :

1. La matrice est **symétrique** car on a $cov(X, Y) = cov(Y, X)$
2. Elle est **semi-définie positive** (ses valeurs propres sont positives ou nulles).
3. Les éléments de sa diagonale ($\mathbf{\Sigma}_{i,i}$) représentent la variance de chaque variable, étant donné la propriété $cov(X, X) = var(X)$
4. Les éléments en dehors de la diagonale ($\mathbf{\Sigma}_{i,j}, i \neq j$) représentent la covariance entre les variables i et j.
5. la matrice de covariance d'un vecteur décorrélé ou indépendant est diagonale ($\mathbf{\Sigma}_{i,j} = 0 \; \forall i \neq j$)

Indépendance de vecteurs aléatoires

notion de VA i.i.d

Cas des variables discrètes

Soit $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ une suite de variables aléatoires discrètes, et soit $(S_1, S_2, ..., S_n)$ une suite d'ensembles finis ou dénombrables tels que, pour tout $i \leq n$, $\mathbb{P}(X_i \in S_i) = 1$.

Definition

$(X_1, X_2, ..., X_n)$ est une suite de variables aléatoires indépendantes si et seulement si, pour tout $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \prod_{i=1}^{n} S_i$,

$$\mathbb{P}(\boldsymbol{X} = \mathbf{x}) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

Indépendance de vecteurs aléatoires

Cas des variables aléatoires à densité

Soit une suite $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ de variables aléatoires réelles définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et de densités de probabilité respectives f_i.

Definition

Les variables aléatoires réelles $(X_1, X_2, \ldots, X_n)$ sont dites indépendantes si et seulement si le vecteur $\boldsymbol{X}$ a une densité de probabilité f qui se définie par

$$\forall \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n, \qquad f(x_1, \ldots, x_n) \; = \; \prod_{i=1}^{n} f_i(x_i).$$

Quelques statistique sur vecteurs aléatoires I

Soit un échantillon de valeurs de vecteurs aléatoires continues $(\mathbf{x}_1, \ldots, \mathbf{x}_n), \mathbf{x}_i \in \mathbb{R}^d$, on a le vecteur moyen empirique et la matrice de covariance empirique sont respectivement données par

Vecteur moyen empirique

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i \tag{24}$$

Matrice de covariance empirique

$$\mathbf{S} = \frac{1}{n}\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{25}$$

Variables et vecteurs aléatoires gaussiens

- Variables et vecteurs aléatoires gaussiens

Variables et vecteurs aléatoires gaussiens

Definition

On appelle loi ou densité normale (ou gaussienne) univariée de paramètres (μ, σ^2) (où $\sigma \geq 0$) la loi de probabilité définie par la densité $f : \mathbb{R} \to \mathbb{R}^+$, telle que pour tout $x \in \mathbb{R}$:

$$f(x) = \frac{1}{\sigma\sqrt{2\,\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{26}$$

Soit X une variable aléatoire réelle admettant pour densité de probabilité la loi normale (26), appelée variable aléatoire gaussienne, son espérance est μ et son écart type est σ (sa variance est σ^2).

On note

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

et on lit "X suit la loi normale d'espérance μ et de variance σ^2". Sa densité de probabilité dessine une courbe dite courbe en cloche

Loi normale centrée réduite

La loi normale centrée réduite correspond à un cas particulier de la loi normale générale (26).

Definition

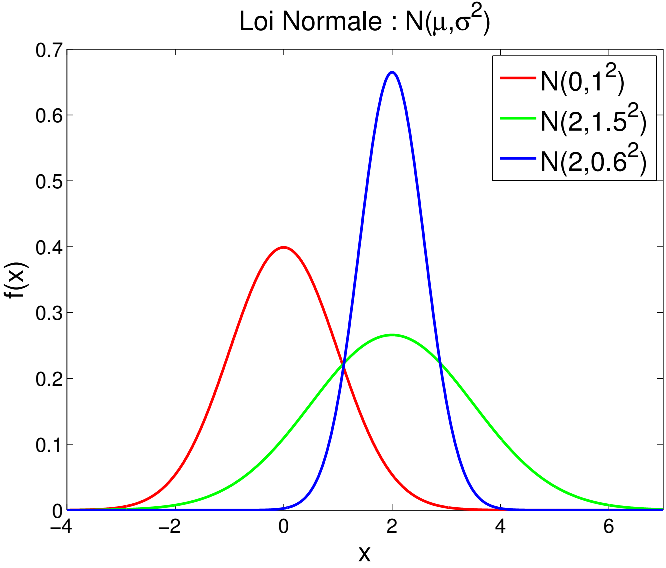
On appelle loi normale (ou gaussienne) centrée réduite la loi définie par la densité de probabilité $\varphi : \mathbb{R} \to \mathbb{R}^+$ définie par :

$$\varphi(x) = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2} \tag{27}$$

qui se note $\mathcal{N}(0, 1)$. Elle est dite centrée de part son espérance nulle et réduite de part sa variance unité.

Représentations graphiques

Loi normale monodimensionnelle



Quelques propriétés I

Soit une variable aléatoire X qui suit la loi normale $\mathcal{N}(\mu, \sigma^2)$. Alors :

- son espérance et sa variance existent et $\mathbb{E}(X) = \mu, \ \mathrm{Var}(X) = \sigma^2 \geq 0$,
- la variable aléatoire $X^\star = \frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}}$, c'est-à-dire $X^\star = \frac{X-\mu}{\sigma}$, suit la loi normale centrée réduite $\mathcal{N}(0, 1)$,
- si a, b sont deux réels ($a \neq 0$), alors la variable aléatoire $a\,X + b$ suit la loi normale $\mathcal{N}(a\,\mu + b, \ a^2\,\sigma^2)$

Vecteurs aléatoires gaussiens

Definition

Un vecteur aléatoire $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ est dit gaussien si, pour tout $\mathbf{u} \in \mathbb{R}^d$, la variable aléatoire réelle $\mathbf{u}^T\boldsymbol{X}$ est une variable aléatoire gaussienne. C'est à dire si toute combinaison linéaire de ses composantes est une variable gaussienne.

Loi normale multidimensionnelle

La loi normale multidimensionnelle représente la distribution d'un vect. a. gaussien. Soit $\boldsymbol{X}$ un vect. a. gaussien de dimension d. On appelle loi normale multidimensionnelle d'espérance $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$ la densité de probabilité $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$ définie par :

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\,|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad (28)$$

$|\boldsymbol{\Sigma}|$ étant le déterminant de $\boldsymbol{\Sigma}$.

Vecteurs aléatoires gaussiens

La loi normale multidimensionnelle se note habituellement $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

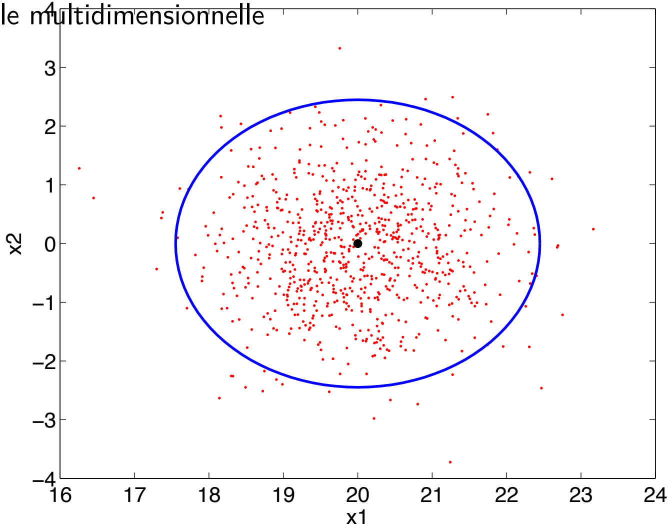
Distance de Mahalanobis

Le terme présent dans l'exponentielle dans (28) $(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$ est le carré de la **distance de Mahalanobis**. Cette dernière est en effet donnée par

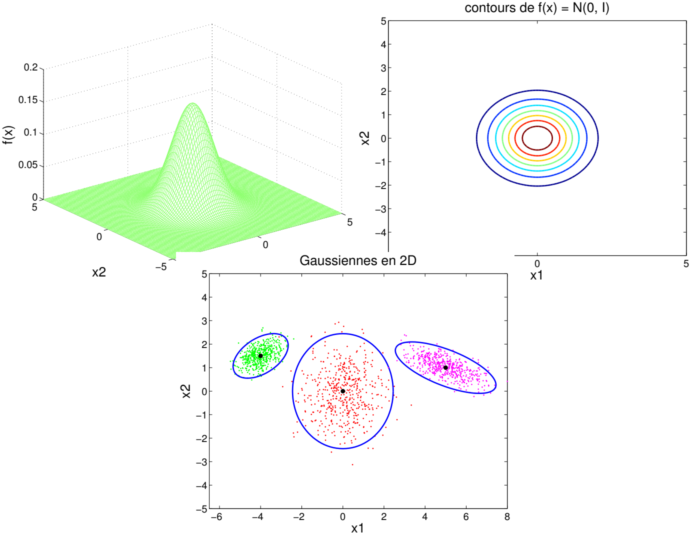
$$\sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}$$

Vecteurs aléatoires gaussiens

Loi normale multidimensionnelle



Repr. graphiques : Loi normale multidimensionnelle



Vecteurs aléatoires gaussiens

Soit $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ un vecteur aléatoire gaussien de dimension d suivant la loi $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, on a :

Queqlues Propriétés

- l'espérance et la matrice de covariance de $\boldsymbol{X}$ sont respectivement $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ et $\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$,
- les variables aléatoires $X_1, \ldots, X_d$ sont indépendantes si et seulement si la matrice de covariance est diagonale,
- Tout sous-vecteur d'un vecteur aléatoire gaussien suit une loi normale. En particulier, ses composantes sont toutes gaussiennes,
- Si $\mathbf{A}$ est un matrice constante de dimensions $[n \times d]$ et $\mathbf{b}$ un vecteur constant dans $\mathbb{R}^d$, alors la densité du vecteur aléatoire $\boldsymbol{Y} = \mathbf{A}\,\boldsymbol{X} + \mathbf{b}$ (de dimension $[n \times d]$ est la loi normale de n dimensions suivante $\mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}\,, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.