# Master 1 Informatique
# Éléments de statistique inférentielle


Faicel Chamroukhi
Maître de Conférences
UTLN, LSIS UMR CNRS 7296

email: chamroukhi@univ-tln.fr
web: chamroukhi.univ-tln.fr


2014/2015

# Plan I

1 Introductions et Rappels

# Notions de population, échantillon, individu

Une **population** statistique (qu'on note $\Omega$) est un ensemble concernée par une étude statistique.

Un **individu** est un élément individuel considéré dans l'analyse statistique.

Un **Échantillon** est un sous-ensemble effectivement observé de la population. La taille $n$ de l'échantillon est le cardinal de ce sous-ensemble considérée dans l'étude statistique.

⚠ Remarque : Dans la communauté du traitement de signal, on appelle aussi échantillon un point d'un signal, ce qui correspond donc, dans certains cas, à un individu en analyse statistique.

Statistique inférentielle :

**Definition**

L'inférence statistique consiste à induire (inférer) les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population

Statistique descriptive :

**Definition**

L'objectif de la statistique descriptive est de décrire, résumer ou représenter, par des statistiques, les données disponibles dans un échantillon

Wikipedia : Descriptive statistics are distinguished from inferential statistics (or inductive statistics), in that descriptive statistics aim to summarize a sample, rather than use the data to learn about the population that the sample of data is thought to represent.

# Expérience aléatoire

On appelle **expérience aléatoire** une expérience dont on connaît l'ensemble des résultats possibles mais dont on ne peut prédire le résultat effectif avec certitude et qui, répétée plusieurs fois dans des conditions opératoires identiques, produit des résultats qui peuvent être différents. L'ensemble de tous les résultats possibles est appelé *univers*. On appelle *événement* un ensemble de résultats de l'expérience aléatoire (un sous-ensemble de l'univers).

La notion d'expérience aléatoire $\mathcal{E}$ se formalise mathématiquement en définissant :

1. l'ensemble fondamental $\Omega$ (appelé *l'univers*) définissant l'ensemble des résultats possibles de $\mathcal{E}$, appelés événements élémentaires ;
2. un ensemble $\mathcal{A}$ de parties de $\Omega$, appelées événements.
3. une fonction $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, appelée mesure ou distribution de probabilité, qui à tout événement $A$ associe un nombre $\mathbb{P}(A)$ appelé probabilité de cet événement.

# Variables aléatoires

- Variables aléatoires
- Couple de variables aléatoires
- Vecteurs Aléatoires
- Variables et vecteurs aléatoires gaussiens

# Variables aléatoires

Une **Variable statistique** $X$ est une application de $\Omega$ dans un ensemble $E$
$X : \Omega \longrightarrow E$

## Types de variables aléatoires

$X$ est dite une **variable quantitative** si elle prend des valeurs dans $\mathbb{R}$. Il s'agit d'une variable représentée par une quantité (une valeur) telle que l'âge, le poids et la taille, etc.

Une **variable qualitative** est quant à elle une variable representée par une qualité, telles que le sexe, encore l'état civil, le degré de satisfaction d'un service quelconque, etc. La variable qualitative est donc représenté par une modalité plutôt que d'une valeur.

Il existe deux types de variables de variables quantitatives : **variables discrètes** et **variables continues**.

# Variables aléatoires

## Definition

Une variable aléatoire est dite **discrète** si elle ne prend que des valeurs discontinues dans un intervalle fini ou dénombrable.

Exemple : le résultat du jet d'un dé, le nombre d'enfants dans une famille, sont des variables aléatoires discrètes.

## Definition

Une variable aléatoire est dite **continue** si elle ne prend ses valeurs dans $\mathbb{R}$ ou une partie ou un ensemble de parties de $\mathbb{R}$.

Exemple : la moyenne des étudiants, la taille, etc sont des variables aléatoires continues.

# Variables aléatoires

Pour les variable aléatoires qualitatives, il existe également deux type de variables : **variables nominales** et **variables ordinales**.

## Definition

Une variable aléatoire qualitative **nominale** est une variable qui correspondent à des noms, il n'y a aucun ordre précis sur les modalités. Ce sont seulement des mots dans le désordre.

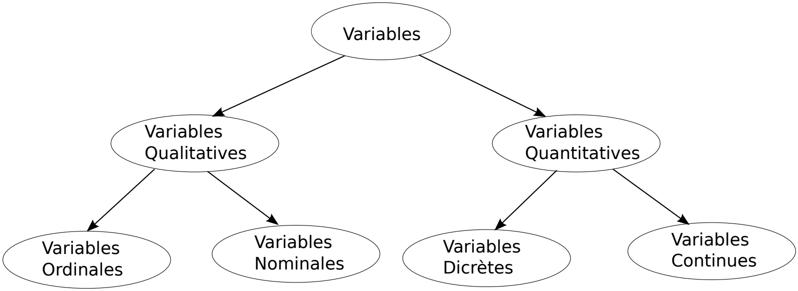
Par exemple, la variable sexe est une variable qualitative nominale qui a deux modalités possibles : féminin ou masculin et dont l'ordre n'importe pas. Un autre exemple est la catégorie socioprofessionnelle

## Definition

Une variable aléatoire qualitative **ordinale** est une variable qui correspondent à un ordre. Il y a un ordre sur les modalités.

Par exemple, le degré de satisfaction par rapport à un service : très satisfait, satisfait, insatisfait, etc

# Variables aléatoires : récapitulatif



# Fonction de densité de probabilité

## Definition

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

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

telle que :

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

# Fonction de répartition

La fonction de répartition d'une variable aléatoire réelle $X$, notée $F_X(x)$, caractérise la loi de probabilité de cette variable. Elle représente la probabilité que la variable aléatoire réelle $X$ prenne une valeur inférieure ou égale à $x$ :

$$F_X(x) = \mathbb{P}(X \leq x) \tag{2}$$

et, pour tout nombre réel $x$, est donnée par :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, \mathrm{d}u. \tag{3}$$

La fonction de répartition $F_X(x)$ est donc l'une des primitives de la fonction de densité de probabilités.

⚠ Remarque : La probabilité que $X$ se trouve dans l'intervalle $]a, b]$ est donc, si $a < b$,

$$\mathbb{P}(a < X \leq b) \;=\; F_X(b) - F_X(a)$$

# 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

$$\mathbb{E}[X] = \sum_k x_k P(X = x_k) = \sum_k x_k p_k \quad (4)$$

⚠ 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 discrète

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

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx. \quad (5)$$

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

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx \quad (6)$$

# Variance d'une variable aléatoire

## Definition

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

$$var[X] = \mathbb{E}[(X - E[X])^2] = \mathbb{E}[X^2] - (E[X])^2 \quad (7)$$

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

⚠ Remarque : D'après (6) et (7), la variance est donc aussi donnée par

$$var[X] = \int_{\mathbb{R}} (x - E[X])^2 f_X(x) dx. \quad (8)$$

# 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 :

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

$\Rightarrow$ Propriété de **linéarité** de l'espérance

Soient $n$ variables aléatoires réelles $X_1, \ldots, X_n$. Alors $\mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i]$

# Quelques propriétés de l'espérance et de la variance

De même, pour la variance, on a la propriété suivante :

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

$\Rightarrow$ Propriété **quadratique** de la variance

⚠ Remarque : Plus généralement on parle de moment d'ordre $r$ ($r > 0$) de la v.a $X$ la quantité donnée par

$$\mathbb{E}[X^r] = \int_{\mathbb{R}} x^r f_X(x) dx$$

et respectivement de moment centré d'ordre $r$ ($r > 0$) la quantité donnée par

$$\mathbb{E}[(X - \mathbb{E}[X])^r] = \int_{\mathbb{R}} (x - \mathbb{E}(X))^r f_X(x) dx.$$

# 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 $f_X$ de paramètre $\theta$ (une distribution de probabilités $P_X(.;\theta)$ pour le cas discret). Une *statistique* est toute fonction de l'échantillon donné $(X_1, \ldots, X_n)$ qui ne dépend pas de $\theta$.

## moyenne empirique

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{9}$$

## variance empirique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2\,. \tag{10}$$

## Quelques statistique sur variables aléatoires

La moyenne est la variance sont dit le moment d'ordre 1 et le moment d'ordre 2, respectivement.

Plus généralement, on parle du *moment empirique d'ordre k* donné par :

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \,. \tag{11}$$

# Couple de variables aléatoires

- Variables aléatoires
- Couple de variables aléatoires
- Vecteurs Aléatoires
- Variables et vecteurs aléatoires gaussiens

# Couple de variables aléatoires, lois jointes, loi conditionnelles, loi marginales, notion d'indépendance I

Lois jointes, lois conditionnelles et théorème de Bayes :

Soient $X$ et $Y$ deux variables aléatoires de densités respectives $f_X(x)$ et $f_Y(y)$ (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

$$f_{XY}(x, y) = f_{Y|X}(y|x)f_X(x) \tag{12}$$

*et par symétrie on a également*

$$f_{XY}(x, y) = f_{X|Y}(x|y)f_Y(y)\ . \tag{13}$$

Pour les VA discrètes cela revient à

## Theorem

$$\begin{aligned} P(Y = y, X = x) &= P(Y = y|X = x)P(X = x) && (14) \\ &= P(X = x|Y = y)P(Y = y) && (15) \end{aligned}$$

# Loi jointe de couple de v.a discrètes

Soient $X$ et $Y$ deux variables discrètes telle que $X$ prend ses valeurs dans l'ensemble $A$ et $Y$ prend ses valeurs dans l'ensemble $B$.

## Definition

La loi du couple $(X, Y)$ est définie par l'ensemble des probabilités :

$$P(X = x, Y = y) \quad \text{avec } x \in A \text{ et } y \in B.$$

⚠ Remarque : Cette loi peut être représenté sous la forme d'un tableau dans le cas où les v.a prennent un nombre petit de valeurs

| $Y/X$ | ... | $x$ | ... | Somme des colonnes |
|---|---|---|---|---|
| $\vdots$ | | | | |
| $y$ | | $P(X = x, Y = y)$ | | $P(Y = y)$ |
| $\vdots$ | | | | |
| Somme des lignes | | $P(X = x)$ | | |

# Loi marginale de v.a discrètes

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

$$P(X = x) = \sum_{y \in B} P(X = x, Y = y) \qquad (16)$$

et

$$P(Y = y) = \sum_{x \in A} P(X = x, Y = y) \qquad (17)$$

⚠ Remarque : On peut remarquer que ces lois correspondent respectivement à la somme des lignes et la somme des colonnes dans le tableau précédent (1).

# Loi jointe de couple de v.a continues (à densité)

## Definition

Un couple de v.a. réelles $(X, Y)$ est dit à densité s'il existe une fonction $f_{(X,Y)}$ telle que la fonction de répartition du couple s'écrit

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)dudv \tag{18}$$

satisfaisant les conditions suivantes :

1. $f_{X,Y}(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$,
2. $\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y)dxdy = 1$.

# Densités marginales

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

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy \tag{19}$$

et

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx. \tag{20}$$

## 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

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{21}$$

et on la propriété suivante

### Definition

$$cov(X, Y) = \sigma_{XY} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{22}$$

# 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

$$\mathbb{E}[XY] = \sum_{x \in A} \sum_{y \in B} xyP(X = x, Y = y) \tag{23}$$

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

$$\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_{X,Y}(x, y) \, \mathrm{d}\, x \, \mathrm{d}\, y. \tag{24}$$

# Corrélation entre deux variables aléatoires I

La corrélation entre deux variables aléatoires se mesure par le coefficient de corrélation linaire noté $\rho$.

## Definition

Le coefficient de corrélation linéaire entre deux variables aléatoires $X$ et $Y$ est défini par :

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}. \tag{25}$$

On a toujours $\rho_{X,Y} \in [-1,1]$. Plus $|\rho_{X,Y}|$ est proche de 1 plus la corrélation entre les variables $X$ et $Y$ est forte. *Si $X$ et $Y$ sont indépendantes, alors $Cov(X,Y) = 0$ et donc $\rho_{X,Y} = 0$. On a par conséquent, $Var(X+Y) = Var(X) + Var(Y)$.*

# Indépendance de deux variables aléatoires

Soient $X$ et $Y$ deux variables aléatoires de densités respectives $f_X(x)$ et $f_Y(y)$ (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

$$f_{XY}(x, y) = f_X(x) f_Y(y) \tag{26}$$

Pour les VA discrètes cela revient à

$$P(Y = y, X = x) = P(X = x) P(Y = y) \tag{27}$$

# Indépendance de deux variables aléatoires

Cela vient du fait que, d'une part le comportement du couple $(X, Y)$ est entièrement décrit par leur fonction de densité de probabilité jointe (ou distribution jointe de probabilité dans le cas discret) $f_{XY}(x, y)$

d'autre part, après le tirage d'une réalisation de $(X, Y)$, la valeur observée (réalisation) $x$ de $X$ fournit une certaine quantité d'information sur la valeur de $Y$ $\Rightarrow$ Cette information est représentée par la distribution de $Y$ conditionnellement à $X = x$ soit $f_{Y|X}(y|x)$.

$X$ et $Y$ sont dites indépendantes si observer une réalisation $x$ de $X$ n'a aucun effet et n'apporte aucune information sur la réalisation possible de $Y$ étant donné $x$. Autrement dit, la distribution de $Y$ conditionnellement à $X$ ne dépend pas de $x$.

Or, la densité de probabilité jointe $f_{XY}(x, y)$ est :

$$f_{XY}(x, y) = f_{Y|X}(y|x) f_X(x) \tag{28}$$

par le théorème de Bayes.

# Indépendance de deux variables aléatoires

Ensuite la densité marginale de $Y$ est obtenue en intégrant cette densité jointe par rapport à $x$ :

$$f_Y(y) = \int_{\mathcal{X}} f_{XY}(x,y)dx = \int_{\mathcal{X}} f_{Y|X}(y|x)f_X(x) \tag{29}$$

Mais si la densité conditionnelle de $Y$ ne dépend pas de $x$ (cas de VA indépendantes), on peut la sortir de l'intégrale et nous avons alors

$$f_Y(y) = f_{Y|X}(y|x)\int_{\mathcal{X}} f_X(x) = f_{Y|X}(y|x) \tag{30}$$

donc

$$f_{Y|X}(y|x) = f_Y(y) \tag{31}$$

et par conséquent

$$f_{XY}(x,y) = f_X(x)f_Y(y) \tag{32}$$

⚠ Remarque : Pour le cas des VA discrètes, le raisonnement est le même en remplaçant les intégrales par des sommes.

# 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.*

Un cas particulier est bien celui où $f(X) = X$ et $g(Y) = Y$. On peut alors énoncer le thérome ainsi : Si X et Y sont indépendantes, alors

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

# Indépendance et corrélation

Une conséquence de ce théorème est que deux variables indépendantes sont décorrélées (leur covariance est nulle).

⚠ Remarque : La réciproque est fausse : la décorrélation de deux variables n'implique pas indépendance, sauf dans le cas v.a à densité normale comme on le verra plus tard dans le chapitre dédié aux variables aléatoires gaussiennes.

# Loi des grands nombres I

## Convergence en probabilité : Définition

On dit qu'une suite $(X_n)$ de v.a. converge en probabilité vers une v.a. $X$ si

$$\forall \epsilon > 0, \qquad \lim_{n \to \infty} \mathbb{P}\left(|X_n - X| \geq \epsilon\right) = 0. \tag{33}$$

## Theorem (Loi des grands nombres)

*Soit $X_1, \ldots, X_n$ un échantillon indépendant d'une suite de v.a. indépendantes et de même loi d'espérance $\mathbb{E}[X] = \mu$ et de variance* $\mathrm{var}(X) = \sigma^2$ *. Alors quel on a :*

$$\forall \epsilon > 0, \quad \lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mathbb{E}(X)\right| \geqslant \epsilon\right) = 0 \tag{34}$$

$\Rightarrow$ La moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge en probabilité vers l'espérance $\mathbb{E}[X]$ quand $n \to +\infty$ (asymptotiquement)

# Loi des grands nombres II

⚠ Remarque : Ici on parle donc de convergence en probabilité.

$\Rightarrow$ La L.G.N nous dit que, pour tout réel $\epsilon$ strictement positif, la probabilité que la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ s'éloigne de l'espérance $\mathbb{E}[X] = \mu$ d'au moins $\epsilon$ tend vers 0 quand la taille de l'échantillon $n$ tend vers l'infini.

# Vecteurs Aléatoires

- Variables aléatoires
- Couple de variables aléatoires
- Vecteurs Aléatoires
- Variables et vecteurs aléatoires gaussiens

# Vecteurs Aléatoires

## Definition

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

## Fonction de répartition

La fonction de répartition d'un vecteur aléatoire $X$ décrit la loi de probabilité d'un vecteur aléatoire et est donnée par

$$F_X(\mathbf{x}) = F_X(x_1, \ldots, x_d) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d) \qquad (35)$$

# Vecteurs Aléatoires

## f.d.p : cas continu

Dans le cas de vecteurs aléatoires réels continus (ayant des composantes continues), la fonction de densité de probabilité $f : \mathbb{R}^d \to \mathbb{R}^+$ de $X$ est définie par

$$f_X(\mathbf{x}) = f_X(x_1, \ldots, x_p) = \frac{\partial^d F_X(\mathbf{x})}{\partial x_1 \partial x_2 \ldots \partial x_d}. \tag{36}$$

# Vecteurs Aléatoires

## f.d.p : cas continu

Dans le cas de vecteurs aléatoires réels continus (ayant des composantes continues), la fonction de densité de probabilité $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$ de $X$ est définie par

$$f_X(\mathbf{x}) = f_X(x_1, \ldots, x_p) = \frac{\partial^d F_X(\mathbf{x})}{\partial x_1 \partial x_2 \ldots \partial x_d} \cdot \qquad (36)$$

Si cette f.d.p existe, on alors

$$\mathbb{P}(X \in A) = \int_A f_X(\mathbf{x}) \, \mathrm{d}\,\mathbf{x} \, , \qquad (37)$$

pour tout $A \subseteq \mathbb{R}^d$ pour lequel cette intégrale est définie, et comme $f$ est une f.d.p, on a :

$$\int_{\mathbb{R}^d} f_X(\mathbf{x}) \, \mathrm{d}\,\mathbf{x} = 1 \cdot \qquad (38)$$

# Vecteurs Aléatoires

## Loi de probabilité : cas discret

Lorsque $X$ est un vecteur aléatoire discret, la loi de probabilité (fonction de masse de probabilité) se définit par $P_X(\mathbf{x}) = \mathbb{P}(X = \mathbf{x})$. On a aussi

$$\mathbb{P}(X \in A) = \sum_{\mathbf{x} \in A} P_X(\mathbf{x}) \,, \tag{39}$$

et

$$\sum_{\mathbf{x} \in \mathcal{X}} P_X(\mathbf{x}) = 1 \,, \tag{40}$$

$\mathcal{X}$ étant l'ensemble de valeurs que prend $\mathbf{x}$.

# Espérance et matrice de covariance d'un vecteur aléatoire

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

## Espérance

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

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

## Matrice de covariance

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

$$\mathbf{\Sigma}_X = \mathrm{cov}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right] = \mathbb{E}[XX^T] - \mathbb{E}[X](\mathbb{E}[X])^T, \quad (41)$$

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

# 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 ($\boldsymbol{\Sigma}_{i,i}$) représentent la variance de chaque variable, étant donné la propriété $cov(X, X) = var(X)$
4. Les éléments en dehors de la diagonale ($\boldsymbol{\Sigma}_{i,j}, i \neq j$) représentent la covariance entre les variables $i$ et $j$.
5. la matrice de covariance d'un vecteur décorrélé ou indépendant est diagonale ($\boldsymbol{\Sigma}_{i,j} = 0 \ \forall i \neq j$)
6. etc

# Matrice de corrélation

## Definition

La matrice de corrélation (notée $R$) du vecteur aléatoire $X$, définie de manière analogue à la matrice de covariance, est la matrice dont le terme général est le cœfficient de corrélation linéaire $\rho_{i,j}$ donné par l'équation (42).

## Cœfficient de corrélation

Chaque terme $\rho_{i,j}$ représente la corrélation entre le couple de variables $(X_i, X_j)$ du vecteur $X$ et pour rappel est donné par

$$\rho_{i,j} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j}. \tag{42}$$

⚠ Remarque : Toutes les valeurs de la matrice de corrélation sont donc comprises entre -1 et +1 et les termes de la diagonale sont égaux à 1.

# Indépendance de vecteurs aléatoires

notion de VA i.i.d

## Cas des variables discrètes

Soit $X = (X_1, X_2, ..., X_n)$ 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 $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \prod_{i=1}^{n} S_i$,

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

# Indépendance de vecteurs aléatoires

## Cas des variables aléatoires à densité

Soit une suite $X = (X_1, X_2, \ldots, X_n)$ 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 $X$ 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

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i \tag{43}$$

### Matrice de covariance empirique

$$\mathbf{S} = \frac{1}{n}\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{44}$$

# Variables et vecteurs aléatoires gaussiens

- Variables aléatoires
- Couple de variables aléatoires
- Vecteurs Aléatoires
- Variables et vecteurs aléatoires gaussiens

**Introduction**

L'une des densités les plus répandues en probabilité et statistique est la *densité normale* dite aussi *densité gaussienne* par référence à celui qui l'a proposé C. F. Gauss.

Elle sert à modéliser de nombreux phénomènes, et comme on le verra plus tard, la distribution de la moyenne d'un échantillon de variables aléatoires issues d'une densité quelconque tend vers une loi normale quand la taille de l'échantillon augmente.

Ceci montre donc la grande importance de la loi normale.

# Variables et vecteurs aléatoires gaussiens

## Definition

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

$$f(x) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{45}$$

Soit $X$ une variable aléatoire réelle admettant pour densité de probabilité la loi normale (45), appelée variable aléatoire gaussienne, son espérance est $\mu$ et son écart type est $\sigma$ (sa variance est $\sigma^2$).

On note

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

et on lit "$X$ suit la loi normale d'espérance $\mu$ et de variance $\sigma^2$". 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 (45).

## Definition

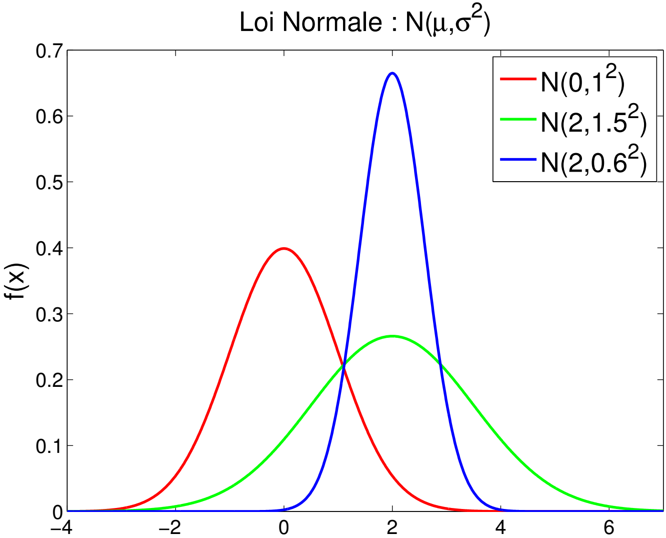
On appelle loi normale (ou gaussienne) centrée réduite la loi définie par la densité de probabilité $\varphi : \mathbb{R} \to \mathbb{R}^+$ définie par :

$$\varphi(x) = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2} \tag{46}$$

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



# Fonction de répartition de la loi normale

La fonction de répartition de la loi normale générale s'exprime communément à l'aide de la fonction caractéristique de la loi normale centrée réduite.

## Fonction de répartition de la loi normale centrée réduite

On note $\Phi$ la fonction de répartition de la loi normale centrée réduite. Elle est définie, pour tout réel $x$, par :

$$\Phi(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \varphi(u)\,\mathrm{d}u = \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u\,\cdot \tag{47}$$

Cette primitive ne s'exprime pas à l'aide des fonctions usuelles (exponentielle, etc.) mais devient elle-même une fonction usuelle.

La fonction $\Phi$ s'exprime à l'aide de la fonction d'erreur notée erf (error function) (appelée aussi fonction d'erreur de Gauss)

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-v^2} \quad (48)$$

par

$$\Phi(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \cdot \quad (49)$$

*ou a encore*

$$\mathrm{erf}(x) = 2\Phi\left(x\sqrt{2}\right) - 1 \cdot \quad (50)$$

## Fonction de répartition de la loi normale générale

La fonction de répartition $F$ pour la loi normale générale est donc donnée par, pour tout $x \in \mathbb{R}$,

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \cdot \quad (51)$$

# 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 $\mathcal{N}(a\,\mu + b,\ a^2\,\sigma^2)$

# Théorème central limite

La grande importance pratique associée à la distribution normale découle du théorème central limite présenté ci-dessous.

## Theorem (Théorème Central Limite)

*Soit $X_1, X_2, \cdots, X_n$ une suite de variables aléatoires réelles mutuellement indépendantes et suivant la meme densité $f$ (variables i.i.d) d'espérance $\mu$ et d'écart-type $\sigma$. Soit la variable aléatoire somme*

$$S = X_1 + X_2 + \cdots + X_n$$

*Alors la densité de probabilité de la somme $S$ converges vers la loi normale $\mathcal{N}(n\mu, n\sigma^2)$ quand $n$ tend vers l'infini :*

$$\lim_{n\to\infty} f(s) = \mathcal{N}(n\mu, n\sigma^2)$$

⚠ Remarque : : Ici on parle de convergence en probabilité.

# Théorème central limite

$\Rightarrow$ Ceci montre donc l'importance que joue la loi normale pour approximer la densité de données issues de l'accumulation de plusieurs phénomènes notamment physiques.

Ce théorème peut aussi se formuler ainsi. Soit la variable aléatoire moyenne

$$\bar{X} = \frac{S}{n} = \frac{X_1 + \cdots + X_n}{n},$$

et la variable aléatoire centrée réduite

$$Z = \frac{S - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

On a donc la densité de $Z$ converges vers la loi normale centrée réduite quand $n$ tend vers l'infini.

# Couple de variables aléatoires gaussiennes I

Considérons deux variables aléatoires réelles $X$ et $Y$. La densité jointe $f_{X,Y}$ du couple $(X, Y)$ est dite normale (gaussienne) si elle prend la forme suivante :

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} \times \\ & \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right\}. \end{aligned}$$

avec $\sigma_1 et \sigma_2 \geq 0$ et $\rho \in [-1, 1]$. On montre que $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ et que $\rho$ est le coefficient de corrélation entre $X$ et $Y$.

*Une corrélation nulle (décorrélation) implique l'indépendance quand les variables aléatoires sont gaussiennes.*

# Couple de variables aléatoires gaussiennes II

## Démonstration.

Soit $\rho = 0$ dans l'équation (52). On a la densité jointe de $(X, Y$ est donnée par

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right\}. \\ &= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \times \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\} \\ &= f_X(x)f_Y(y) \end{aligned}$$

$\square$

qui est lé résultat recherché.

# Vecteurs aléatoires gaussiens

## Definition

Un vecteur aléatoire $X = (X_1, \ldots, X_d)^T$ est dit gaussien si, pour tout $\mathbf{u} \in \mathbb{R}^d$, la variable aléatoire réelle $\mathbf{u}^T X$ 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 $X$ 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 $f : \mathbb{R}^d \to \mathbb{R}^+$ définie par :

$$f(x) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})\right), \qquad (52)$$

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

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

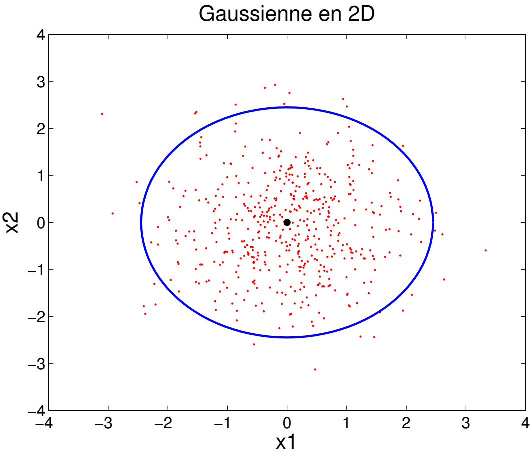
### Distance de Mahalanobis

Le terme présent dans l'exponentielle dans (52) $(x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})$ est le carré de la **distance de Mahalanobis**. Cette dernière est en effet donnée par

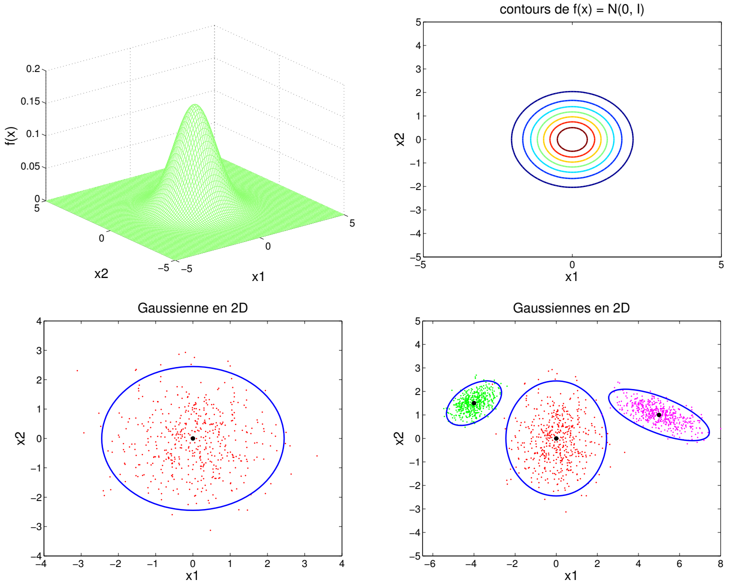
$$\sqrt{(x - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})}$$

# Représentations graphiques

## Loi normale multidimensionnelle



# Repr. graphiques : Loi normale multidimensionnelle



# Vecteurs aléatoires gaussiens

Soit $X = (X_1, \ldots, X_d)^T$ 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 $X$ sont respectivement $\mathbb{E}[X] = \boldsymbol{\mu}$ et $\mathrm{Cov}(X) = \boldsymbol{\Sigma}$,
- 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 $Y = \mathbf{A}\,X + \mathbf{b}$ (de dimension $[n \times d]$ est la loi normale de $n$ dimensions suivante $\mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}\,, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.