# Probabilité et statistique

Alexandre Allauzen
allauzen@limsi.fr

Université Paris Sud — LIMSI

Septembre 2017

# Variable aléatoire (VA) ?

- variable aléatoire = le résultat d'une expérience ; sa valeur change chaque fois qu'elle est regardée
- variable aléatoire =fonction qui permet d'associer un événement à un nombre (plus précisément c'est une fonction de l'espace des événements vers un espace mesurable)

# Les 3 étages de la fusée

## Distribution de probabilité

- Fonction qui associe une probabilité à la réalisation d'une V.A (pour toutes les réalisations possibles).
- Caractérise une VA (discrète ou continue)
- Elle définit par un ensemble de paramètres (approche paramètrique)

## Les paramètres

Ils sont à estimer à partir d'un échantillon :

- représentatif ? grand ? fiable ?

$\rightarrow$ les données d'estimation

## Données d'estimation

- Plus il y en a, mieux c'est ! Mais ...
- L'estimation des paramètres n'est qu'une version compressée des données d'estimation

# Exemple 1 : données d'apprentissage

| Refund $X_1$ | Status $X_2$ | Tax.inc. $X_3$ | Age $X_4$ | Cheat $Y$ |
|---:|---:|---:|---:|---:|
| Yes | Single | 125,6 | 25 | No |
| No | Married | 100,9 | 45 | No |
| No | Single | 70,0 | 33 | No |
| Yes | Married | 120,2 | 78 | No |
| No | Divorced | 95,5 | 72 | Yes |
| No | Married | 60,1 | 55 | No |
| Yes | Divorced | 220,7 | 41 | No |
| No | Single | 85,5 | 49 | Yes |
| No | Married | 75,0 | 37 | No |
| No | Single | 90,8 | 42 | Yes |

5 V.As : $X_1, X_2, X_3, X_4, Y$ :

- discrètes et binaires : $X_1, Y$
- discrète : $X_2$
- continues : $X_3, X_4$

Espace de réalisation :

- $\mathcal{A}_{X_1} = \mathcal{A}_Y = \{No, Yes\}$
- $\mathcal{A}_{X_2} = \{S., M., D.\}$
- $\mathcal{A}_{X_3} = \mathcal{A}_{X_3} = \mathbb{R}$

# Plan

# Variable aléatoire (VA) discrète

## Définition

Le triplet $X = (x, \mathcal{A}_X, \mathcal{P}_X)$ représente une variable aléatoire.

- $x$ est la réalisation de la VA
- $\mathcal{A}_X = \{x_1, ...x_K\}$ : le domaine de réalisation de $X$
- les probabilités associées sont définis par $\mathcal{P}_X = \{\beta_1, ...\beta_K\}$

$$\begin{aligned} P(X = x_i) &= \beta_i \\ \sum_{x_i \in \mathcal{A}_X} P(X = x_i) &= 1 \\ 0 \leq \beta_i &\leq 1 \end{aligned}$$

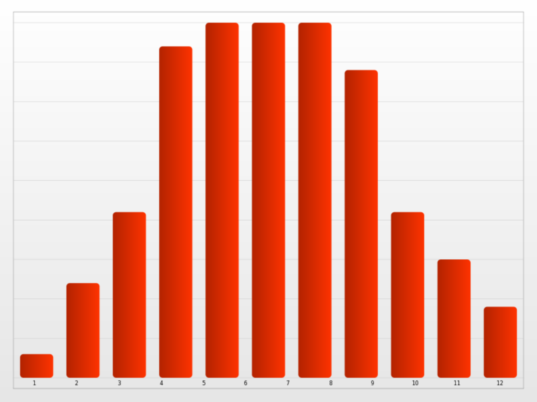
Ce type de distribution est souvent nommée : distribution catégorielle, à ne pas confondre avec la distribution multinomiale qui en est une extention.

### Exemple

Nombre de pages visitées $i$ par visite

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n(i)$ | 7 | 33 | 58 | 116 | 125 | 126 | 121 | 107 | 56 | 37 | 25 | 4 | 815 |
| fréquence | 0,01 | 0,04 | 0,07 | 0,14 | 0,15 | 0,15 | 0,15 | 0,13 | 0,07 | 0,05 | 0,03 | 0,1 | 1 |

# Graphiquement




distribution de probabilité = base de données

# Exemple 1 : 1 variable aléatoire

| Refund | Status | Tax.inc. | Age | Cheat |
|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y$ |
| Yes | Single | 125,6 | 25 | No |
| No | Married | 100,9 | 45 | No |
| No | Single | 70,0 | 33 | No |
| Yes | Married | 120,2 | 78 | No |
| No | Divorced | 95,5 | 72 | Yes |
| No | Married | 60,1 | 55 | No |
| Yes | Divorced | 220,7 | 41 | No |
| No | Single | 85,5 | 49 | Yes |
| No | Married | 75,0 | 37 | No |
| No | Single | 90,8 | 42 | Yes |

- $X_2$ : Marital status
- $\mathcal{A}_{X_2} = \{M., S., D.\}$
- $\mathcal{P}_{X_2} = (\beta_i)_{i=1}^{|\mathcal{A}_{X_2}|} = (\beta_{M.}, \beta_{S.}, \beta_{D.})$
- $\beta_{M.} + \beta_{S.} + \beta_{D.} = 1$
- Estimation :

$$\begin{aligned} P(X_2 = M.) &= \beta_{M.} \\ &= \frac{n(X_2 = M.)}{n(X_2 = *)} \\ &= \frac{n(X_2 = M.)}{\sum_{x \in \mathcal{A}_{X_2}} n(X_2 = x)} \end{aligned}$$

$P(X_2 = x)$ : $\boxed{\phantom{x}}\boxed{\phantom{x}}\boxed{\phantom{x}}$

# Exemple 1 : 1 (autre) variable aléatoire

| Refund $X_1$ | Status $X_2$ | Tax.inc. $X_3$ | Age $X_4$ | Cheat $Y$ |
|---|---|---|---|---|
| Yes | Single | 125,6 | 25 | No |
| No | Married | 100,9 | 45 | No |
| No | Single | 70,0 | 33 | No |
| Yes | Married | 120,2 | 78 | No |
| No | Divorced | 95,5 | 72 | Yes |
| No | Married | 60,1 | 55 | No |
| Yes | Divorced | 220,7 | 41 | No |
| No | Single | 85,5 | 49 | Yes |
| No | Married | 75,0 | 37 | No |
| No | Single | 90,8 | 42 | Yes |

- $X_1$ : Refund
- $\mathcal{A}_{X_1} = \{No, Yes\}$
- $\mathcal{P}_{X_1} = (\beta_i)_{i=1}^{|\mathcal{A}_{X_1}|} = (\beta_{No}, \beta_{Yes})$
- $\beta_{No} + \beta_{Yes} = 1$
- $\beta_{No} = 1 - \beta_{Yes}$
- Estimation :

$$\begin{aligned} P(X_1 = Yes) = \beta_{Yes} &= 1 - \beta_{No} \\ &= \frac{n(X_1 = Yes)}{n(X_1 = *)} \end{aligned}$$

- De même pour $Y$

# Exemple 1 : $X_2$ et $Y$, probabilité jointe - 1

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

- $X_2$ : Refund, et $Y$ : la classe
- $\mathcal{A}_Y = \{No, Yes\}$
- $\mathcal{A}_{X_2} = \{M., S., D.\}$
- $\mathcal{P}_{X_2,Y} = (\beta_{x_2,y})\forall x_2 \in \mathcal{A}_{X_2}$ et $y \in \mathcal{A}_Y$
- $|\mathcal{A}_Y| \times |\mathcal{A}_{X_2}|$ paramètres
- Estimation :

$$P(X_2 = x_2, Y = y) = \beta_{x_2,y} = \frac{n(X_2 = x_2, Y = y)}{n(X_2 = *, Y = *)}$$
$$\frac{n(X_2 = x_2, Y = y)}{\sum_{x\in\mathcal{A}_{X_2}, y\in\mathcal{A}_Y} n(X_2 = x, Y = y)}$$

# Exemple 1 : $X_2$ et $Y$, probabilité jointe - 2

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

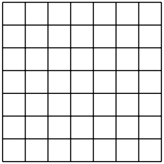
| | M. | S. | D. |
|---|---|---|---|
| No | | | |
| Yes | | | |

$P(X_2 = S., Y = Yes)$

# Probabilité jointe en 2D

- plusieurs variables aléatoires peuvent interagir
- $P(X = x, Y = y) = P(X = x \text{ et } Y = y) = \text{“}P(x, y)\text{”}$

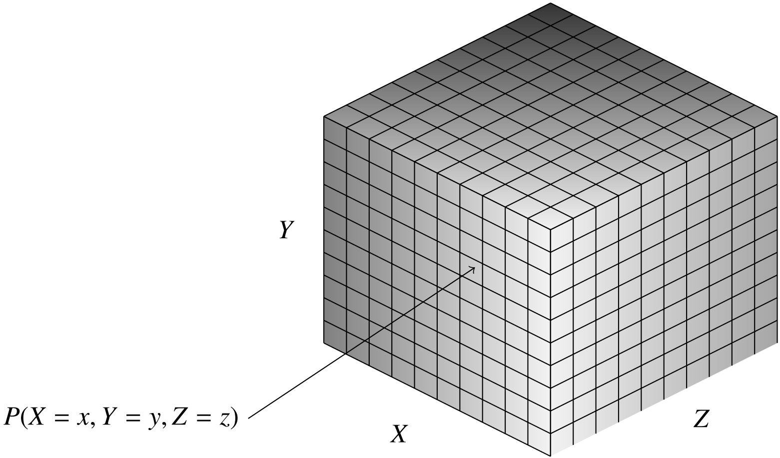
$P(X, Y)$



$$\sum_{x,y} P(X = x, Y = y) = 1$$

# Probabilité jointe en 3D

- plusieurs variables aléatoires peuvent interagir
- $P(X = x, Y = y, Z = z) = P(X = x \text{ et } Y = y \text{ et } Z = z)$



# Exemple 1 : $X_2$ et $Y$, probabilité conditionnelle - 1

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

- $X_2$ : Refund, et $Y$ : la classe
- $\mathcal{A}_Y = \{N, Y\}$
- $\mathcal{A}_{X_2} = \{M., S., D\}$
- Conditionnelle : une variable est fixée (connue)
- Fixons $Y = Yes$ :

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Divorced | Yes |
| Single | Yes |
| Single | Yes |

- Une distribution sur $X_2$ à $Y = y$ fixé

$$\sum_{x_2} P(X_2 = x_2 | Y = Yes) = 1$$

# Exemple 1 : $X_2$ et $Y$, probabilité conditionnelle - 2

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Divorced | Yes |
| Single | Yes |
| Single | Yes |

- Pour chaque réalisation de $Y$ : une distribution sur $X_2$
- $|\mathcal{A}_Y| \times |\mathcal{A}_{X_2}|$ paramètres

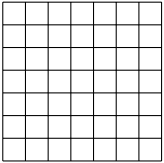
| | M. | S. | D. |
|---|---|---|---|
| No | | | |
| Yes | | | |

$$P(X_2 = S.|Y = Yes)$$

# Probabilité conditionnelle en 2D

- une des variables est connue
- revient à prendre une « tranche »

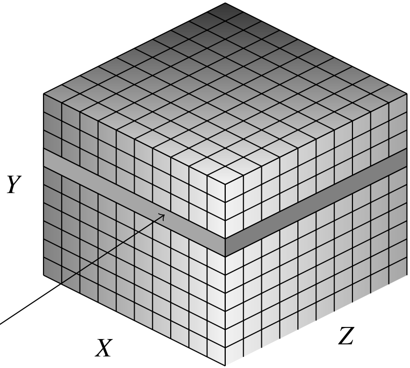
$P(X|Y)$



Quelle différence avec la distribution jointe ?

$$\sum_{y} P(X = x|Y = y) = 1$$

# Probabilité conditionnelle en 3D




$P(X = x, Z = z|Y = y)$

# Probabilité jointe et conditionnelle

$$P(X_2 = x_2, Y = y) = P(Y = y) \times P(X_2 = x_2 | Y = y)$$
$$P(X_2 = S., Y = Yes) = P(Y = Yes) \times P(X_2 = S. | Y = Yes)$$

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Divorced | Yes |
| Single | Yes |
| Single | Yes |

$$P(X_2 = S., Y = Yes) = \frac{3}{10} \times \frac{2}{3} = \frac{2}{10}$$

# Probabilité jointe et conditionnelle - 2

$$P(X_2 = x_2, Y = y) = P(X_2 = x_2) \times P(X_2 = x_2)P(Y = y|X_2 = x_2)$$
$$P(X_2 = S., Y = Yes) = P(X_2 = S.) \times P(Y = Yes|X_2 = x_2)$$

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Married | No |
| Single | No |
| Married | No |
| Divorced | Yes |
| Married | No |
| Divorced | No |
| Single | Yes |
| Married | No |
| Single | Yes |

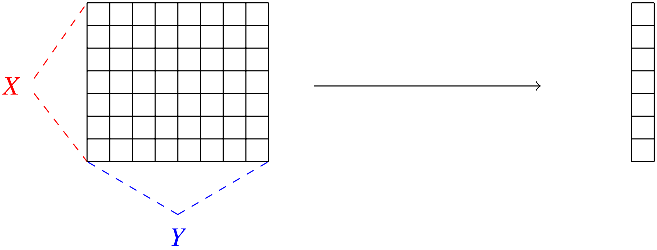
| Status $X_2$ | Cheat $Y$ |
|---|---|
| Single | No |
| Single | No |
| Single | Yes |
| Single | Yes |

$$P(X_2 = S., Y = Yes) = \frac{4}{10} \times \frac{1}{2} = \frac{2}{10}$$

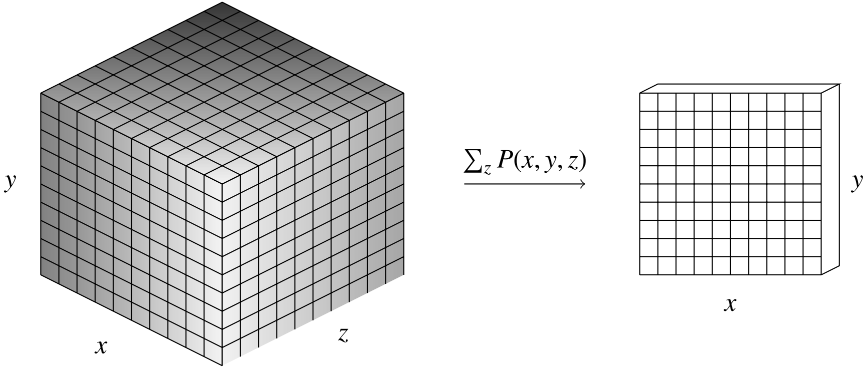
# Probabilité marginale en 1D

$P(X = x, Y = y)$

$P(X = x) = \sum_y P(X = x, Y = y)$



# Probabilité marginale




$$P(x, y) = \sum_z P(x, y, z)$$

# Distributions jointe, conditionnelles et marginales

La distribution jointe *contient* les distributions conditionnelles et marginales.

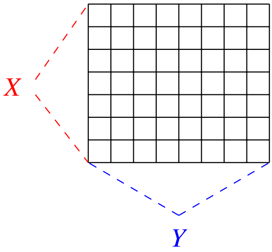
- Marginalisation : $P(X,Y) \rightarrow P(X)$ et $P(Y)$
- Passage au conditionnel :

$$P(Y|X) = \frac{P(X,Y)}{P(X)}$$
$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

On peut retrouver la distribution jointe à partir de la distribution conditionnelle **et** marginale.

# Estimation des probabilités jointe et conditionnelle

$M(i,j) = compte(X = x_i \text{ et } Y = y_j)$

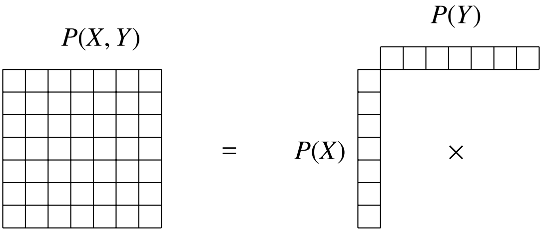



- Remplir la matrice avec les comptes des tirages
- La différence se fait à la normalisation.

# (in)dépendance statistique

- deux variables sont indépendantes si et seulement si

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$




- deux variables sont conditionnellement indépendantes si :

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z) \times P(Y = y|Z = z)$$

# Exemple 1 : Vraisemblance des données pour $X_2$

| Refund $X_1$ | Status $X_2$ | Tax.inc. $X_3$ |
|---|---|---|
| Yes | Single | 125,6 |
| No | Married | 100,9 |
| No | Single | 70,0 |
| Yes | Married | 120,2 |
| No | Divorced | 95,5 |
| No | Married | 60,1 |
| Yes | Divorced | 220,7 |
| No | Single | 85,5 |
| No | Married | 75,0 |
| No | Single | 90,8 |

Connaissant la distribution de $X_2$ (ses paramètres)
Soit les observations :

$$\mathcal{D} = (x_{2,1}, x_{2,2}, ..., x_{2,N}) = (x_{2,i})_{i=1}^{N}$$

Hypothèse i.i.d : indépendament et identiquement distribuées

$$\begin{aligned} P(\mathcal{D}) &= \prod_{i=1}^{N} P(X_2 = x_{2,i}) = \prod_{i=1}^{N} \beta_{x_{2,i}} \\ &= \beta_{S.} \times \beta_{M.} \times \beta_{S.} \times \beta_{M.} \times \beta_{D.} \times \cdots \\ &= \beta_{S.}^{c(S.)} \times \beta_{M.}^{c(M.)} \times \beta_{D.}^{c(D.)} \end{aligned}$$

# Exemple 1 : Vraisemblance des données pour $X_1$

| Refund $X_1$ | Status $X_2$ |
|---|---|
| Yes | Single |
| No | Married |
| No | Single |
| Yes | Married |
| No | Divorced |
| No | Married |
| Yes | Divorced |
| No | Single |
| No | Married |
| No | Single |

Connaissant la distribution de $X_1$ (son paramètre)
Soit les observations :

$$\mathcal{D} = (x_{1,1}, x_{1,2}, ..., x_{1,N}) = (x_{1,i})_{i=1}^{N}$$

Hypothèse i.i.d : indépendament et identiquement distribuées

$$\begin{aligned} P(\mathcal{D}) &= \prod_{i=1}^{N} P(X_1 = x_{1,i}) = \prod_{i=1}^{N} \beta_{x_{1,i}} \\ &= \beta_{Yes} \times \beta_{No} \times \beta_{No} \times \beta_{Yes} \times \beta_{No} \times \cdots \\ &= \beta_{Yes} \times (1-\beta_{Yes}) \times (1-\beta_{Yes}) \times \beta_{Yes} \times (1-\beta_{Yes}) \times \cdots \\ &= \beta_{Yes}^{c(Yes)} \times (1-\beta_{Yes})^{c(No)} \end{aligned}$$

# Théorème de Bayes

$$
\begin{aligned}
P(Y = y_j | X = x_i) &= \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{P(X = x_i, Y = y_j)}{\sum_{y_j \in \mathcal{A}_Y} P(X = x_i, Y = y_j)} \\
&= \frac{P(X = x_i | Y = y_j) P(Y = y_j)}{P(X = x_i)} \\
&= \frac{P(X = x_i | Y = y_j) P(Y = y_j)}{\sum_{y_j \in \mathcal{A}_Y} P(x = X_i | Y = y_j) P(Y = y_j)}
\end{aligned}
$$

## Interprétation

- Supposons que $Y$ représente la classe du modèle et $X$ l'observation.
- Inversion des dépendances statistiques
- Réécriture de l'inférence statistique

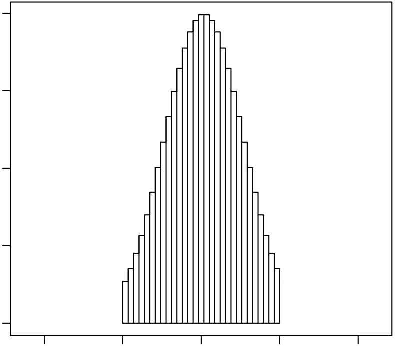
# Que peut-on faire avec une distribution de probabilité ?

- générer des données *(sampling)*
- calcul d'une probabilité jointe : déterminer la probabilité d'une configuration donnée
- inférence certaines v.a. sont connues, quelle est la valeur des autres v.a. ?
- estimation : on observe un ensemble de réalisations d'une distribution ; comment retrouver les paramètres de celle-ci ?

# Plan

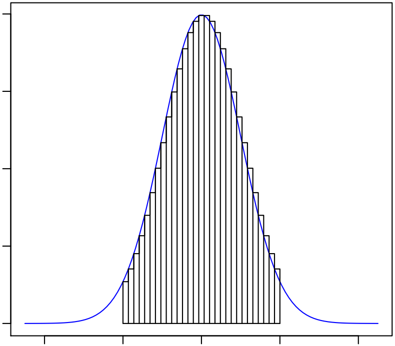
# Variable aléatoire continue

- $X$ = taille d'un homme donné
- solution la plus simple : on discrétise les tailles possibles : $P(X \in [\![1,90,1,95]\!])$
- que se passe-t-il si on diminue le pas de discrétisation ?



# Variable aléatoire continue

- $X$ = taille d'un homme donné
- solution la plus simple : on discrétise les tailles possibles : $P(X \in [\![1,90,1,95]\!])$
- que se passe-t-il si on diminue le pas de discrétisation ?



# Variable aléatoire continue

Si $X$ est une variable aléatoire continue :

- $P(X = x) = 0$ : la probabilité que la variable prenne exactement une valeur donnée est toujours nulle.
- on ne peut connaître que la probabilité que $X$ soit dans un intervalle donné : $P(a \leq X \leq b)$
- la distribution de la masse de probabilité est caractérisé par densité de probabilité $f(x)$ :

$$P(a < X \leq b) = \int_a^b f(x) \cdot dx$$

- $f(x) \cdot dx$ aire d'un intervalle de taille infinitésimal $d(x)$

# Rappel : loi normale

## En dimension 1

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \times e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

avec : $\begin{cases} \mu = \mathbb{E}[x] & \text{moyenne} \\ \sigma^2 = \mathbb{E}[(x-\mu)^2] & \text{variance} \end{cases}$

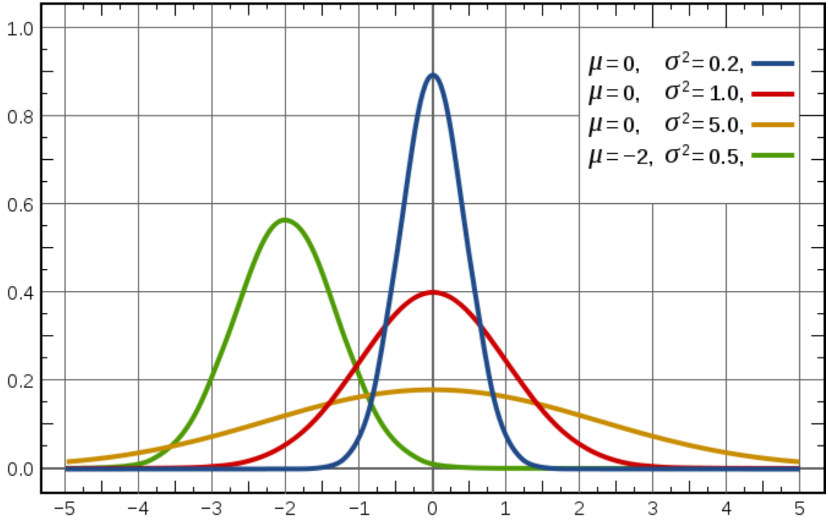
## En dimension $d$

$$f(\mathbf{x}) = \frac{1}{(2 \cdot \pi)^{\frac{d}{2}} \cdot ||\Sigma||^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^t \Sigma^{-1} (\mathbf{x}-\mu)}$$

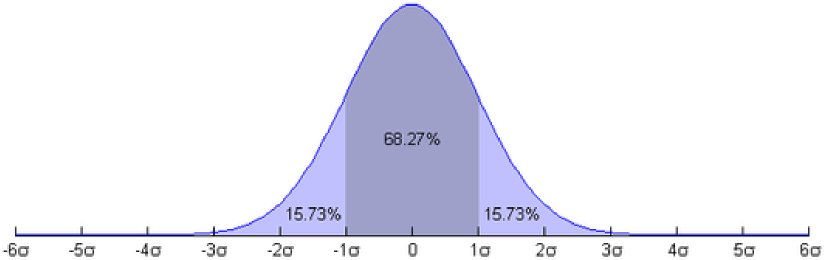
avec :

$$\begin{cases} \mu = \mathbb{E}[\mathbf{x}] & \text{vecteur moyenne} \\ \Sigma = \mathbb{E}\left[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^t\right] & \text{matrice de covariance (matrice carré définie positive)} \end{cases}$$

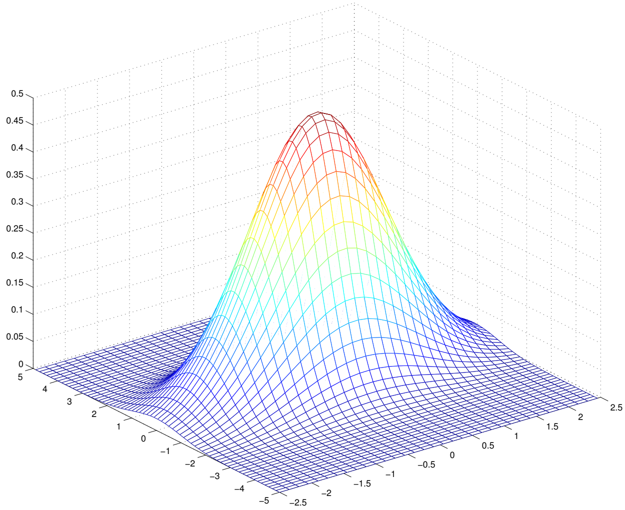
# Graphiquement (en 1D)



# Interprétation des paramètres



# Graphiquement (en 2D)



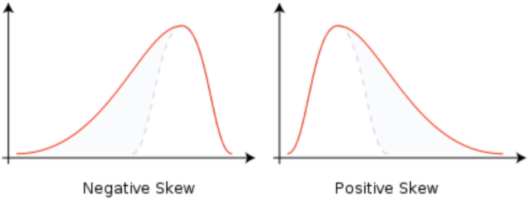
# Plan

# Moyenne et variance

Soit $x_1, x_2, ..., x_n$ un ensemble de valeurs générées par une distribution de probabilité inconnue

On peut caractériser cette distribution par :

- moyenne $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ qui caractérise le centre de la distribution
- variance $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ qui mesure la dispersion de la distribution
- (a)symétrie de la distribution *(skewness)*



# Variance, Covariance et corrélation

## Variance

$$\mathrm{var}(x) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})^2$$

## Covariance : les variations de deux variables sont-elles liées :

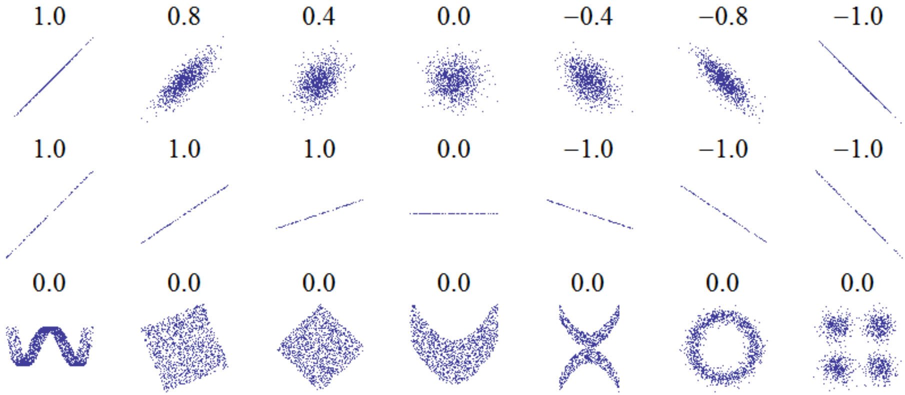
$$\mathrm{cov}(x, y) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x})(y_n - \overline{y})$$

## Corrélation, une covariance normalisée

$$cor(x, y) = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)\mathrm{var}(y)} = \frac{\mathrm{cov}(x, y)}{\mathrm{cov}(x, x)\mathrm{cov}(y, y)}$$

elle quantifie la qualité de l'approximation linéaire de $x$ par $y$ (et recipr.)

# Corrélation illustrée



# Espérance (d'une VA)

## Définition

L'espérance d'une VA discrète $X$ est définie par :

$$\mathbb{E}(X) = \sum_{x \in \mathcal{A}_X} x \, \mathrm{P}(x) = \sum_{i=1}^{m} x_i p_i$$

Avec $f(X)$ une fonction quelconque de $X$

$$\mathbb{E}(f(X)) = \sum_{x \in \mathcal{A}_X} f(x) \, \mathrm{P}(x)$$

# Variance et écart type

$$var[X] = \sigma^2 = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x \in \mathcal{A}_X} (x - \mathbb{E}(X))^2 \, \mathrm{P}(x)$$

La racine carrée de la variance est l'écart-type (ou *standard deviation*)

## Propriétés et interprétation

- l'écart-type est toujours positif,
- il est nul *ssi* toute la masse de probabilité est concentrée en un point (distribution de Dirac).
- L'espérance peut être interprétée comme le "centre" de la VA, autour de laquelle se dispersent les autres valeurs.
- L'écart-type rend compte de la dispertion autour de l'espérance.

# Corrélations et covariances

## Mesure du lien statistique entre 2 VA

### Définitions

- Corrélation

$$\mathbb{E}(XY) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} xy \, \mathrm{P}(x, y)$$

- Covariance

$$\sigma_{XY}^2 = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \tag{1}$$

$$= \sum_{x \in \mathcal{A}_X} \sum_{y \in \mathcal{A}_Y} (x - \mathbb{E}(X))(y - \mathbb{E}(Y)) P(x, y) \tag{2}$$

### Interprétation

- La covariance est une mesure du degré de dépendance entre deux VA.
- $X$ et $Y$ indépendantes $\Rightarrow \sigma_{XY} = 0$ (pas équivalence).

# Vecteurs de VA

Généralisation aux VA multidimensionnelles

## Notation vectorielle

- $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$
- $\mathcal{A}_{\boldsymbol{X}}$ produit cartésien $\mathcal{A}_{X_1} \times \mathcal{A}_{X_2} \ldots \mathcal{A}_{X_d}$
- $P(\boldsymbol{X}) = P(X_1, X_2, \ldots, X_d)$

## $\boldsymbol{X}$ continu

- Vecteur moyenne

$$\mu_{\boldsymbol{X}} = \mathbb{E}(\boldsymbol{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots \mathbb{E}(X_n))$$

- Matrice de covariance

$$\begin{aligned} \Sigma_{\boldsymbol{X}} &= \mathbb{E}((\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}))(X - \mathbb{E}(\boldsymbol{X}))^T) = (\sigma_{i,j}) \\ \sigma_{i,j} &= \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))^T) \end{aligned}$$