Rappels d'algèbre linéaire/probabilité

Charlotte Pelletier
(Basé sur le cours de N. Courty)

22 janvier 2020

Plan

Ensemble d'outils mathématiques fonctionnant dans le domaine continu (par opposition aux mathématiques discrètes) essentiels à la compréhension des outils d'apprentissage automatique.

Nous discuterons entre autre de :

- valeurs scalaires, vectorielles matrices, tenseurs
- opérations basiques entre ces quantités (addition, produits)
- Espaces vectoriels engendrés par une base, indépendance
- diagonalisation, factorisation

Quantités

Variables scalaires : dénotées par une lettre en minuscule

- ex. $x \in \mathbb{R}$ est la pente d'une droite
- ex. $n \in \mathbb{N}$ est le nombre d'éléments dans un ensemble

Variables vectorielles : tableau de valeurs ordonnées, dénoté en minuscule **gras**

- $\mathbf{v} \in \mathbb{R}^{256}$: point dans un espace réel à 256 dimensions
- $\mathbf{v}^T = [v_1 v_2 \ldots v_i \ldots v_n]$, chaque valeur du tableau est indexé par un entier $i \in \{1, n\}$
- les valeurs v_i sont les coordonnées selon chaque axe de l'espace

Variables matricielles : tableau bi-dimensionel (2D) de valeurs ordonnées, dénoté en majuscule **gras**

- $\mathbf{A} \in \mathbb{R}^{3\times 3}$: matrice exprimant une application de $\mathbb{R}^3 \to \mathbb{R}^3$
- ces valeurs sont indexées par i (numéro de ligne) et j (numéro de colonne) : $A_{i,j}$

Tenseurs

Parfois des dimensions supplémentaires sont nécessaires pour traduire des relations complexes entre plusieurs élements

- ex. image 3D

tenseurs : tableau n-aire (n-D) de valeurs ordonnées (selon une grille régulière), dénoté en majuscule **gras**

- $\mathbf{T} \in \mathbb{R}^{28\times 28\times 28}$: tenseur de $\mathbb{R}^{28\times 28\times 28}$
- ces valeurs sont indexées par $i, j, k, \ldots$: $T_{i,j,k}$ dans le cas précédent

Addition/multiplications de matrices

Si $\mathbf{A}$ et $\mathbf{B}$ sont de même tailles (par ex. $m \times n$)

- $\mathbf{C} = \mathbf{A} + \mathbf{B} \equiv \mathbf{C}_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}$, $\forall i, j$
- ajout/multiplication par un scalaire : $\mathbf{C} = e\mathbf{A} + f \equiv \mathbf{C}_{ij} = e\mathbf{A}_{ij} + f$
- ajout d'un vecteur (notation non standard) : $\mathbf{C} = \mathbf{A} + \mathbf{v} \equiv \mathbf{C}_{ij} = \mathbf{A}_{ij} + \mathbf{v^T}_j$, aussi appelé *broadcasting* en anglais

Si $\mathbf{v}$ et $\mathbf{u}$ sont de même taille m alors

- on note $\mathbf{v^T u}$ le produit scalaire entre ces deux vecteurs
- $\mathbf{v^T u} = \sum_{k=1}^{m} \mathbf{v}_k \mathbf{u}_k$

Si $\mathbf{A}$ et $\mathbf{B}$ sont de tailles $m \times r$ et $r \times n$ alors

- $\mathbf{C} = \mathbf{AB}$ est de taille $m \times n$
- $\mathbf{C}_{ij} = \sum_{k=1}^{r} \mathbf{A}_{ik} \cdot \mathbf{B}_{kj}$
- $\mathbf{C}_{ij}$ est le produit scalaire entre la ligne i de $\mathbf{A}$ et la colonne j de $\mathbf{B}$

Propriétés du produit matriciel

- Distributivité par rapport à l'addition : $\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB}$
- Associativité : $\mathbf{C}(\mathbf{AB}) = (\mathbf{CA})\mathbf{B}$
- non-commutativité : en général $\mathbf{AB} \neq \mathbf{BA}$
- mais le produit scalaire l'est : $\mathbf{v^T u} = \mathbf{u^T v}$
- transposé d'un produit : $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- Si $\mathbf{A}$ est de taille $m \times n$, $\mathbf{x}$ et $\mathbf{b}$ de tailles n, nous avons un système à m équations et n inconnues
- cas où $m = n$. Alors la solution du système, si elle existe, est donnée par :

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

- $\mathbf{A}^{-1}$ est l'inverse de $\mathbf{A}$, i.e. telle que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$
- $\mathbf{I}_n$ matrice identité de taille $n \times n$

- Pivot de Gauss

$$\begin{cases} x + 3y - 2z = 5 \\ 3x + 5y + 6z = 7 \\ 2x + 4y + 3z = 8 \end{cases}$$

$L_2\text{-}3L_1 \rightarrow L_2 \qquad L_3\text{-}2L_1 \rightarrow L_3 \qquad \text{-}L_2/4 \rightarrow L_2$

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 3 & 5 & 6 & 7 \\ 2 & 4 & 3 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 2 & 4 & 3 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 0 & -2 & 7 & -2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & -2 & 7 & -2 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & 0 & 9 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & -15 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

- de nombreuses autres méthodes existent (LU, décomposition de Cholesky, avec des contraintes en plus sur la forme de $\mathbf{A}$)
- $\mathbf{A}^{-1}$ ne dépend pas de $\mathbf{b}$, et peut être utilisée pour résoudre plusieurs problèmes
- cependant en pratique $\mathbf{A}^{-1}$ n'est jamais réellement calculée telle quelle ($o(n^3)$ opérations) et on utilise la valeur de $\mathbf{b}$ dans la résolution (en regardant par exemple la différence entre $\mathbf{b}$ et $\mathbf{Ax}$ à une itération donnée

$$\arg\min_x ||\mathbf{Ax} - \mathbf{b}||^2$$

Espace engendré par un ensemble de vecteurs

L'espace engendré par un ensemble de vecteurs $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ est l'ensemble des points formé par toutes les combinaisons linéaires de ces vecteurs, i.e. $\mathbf{p} = \sum_i \alpha_i \mathbf{v}_i$

- savoir si $\mathbf{Ax} = \mathbf{b}$ admet une solution revient à savoir si $\mathbf{b}$ se trouve dans l'espace engendré par $\mathbf{A}$
- si $\mathbf{A}$ est de taille $n \times n$, alors $\mathbf{A}$ doit être formée de vecteurs linéairement indépendants
- rang$(\mathbf{A}) = n$

Pour des matrices de taille $m \times n$, le rang de $\mathbf{A}$ est au mieux m. D'autres méthodes d'inversion existent pour ce type de problème.

- Exemple : pseudo inverse de Moore Penrose $\mathbf{A}^+$
- si rang$(\mathbf{A}) = m$, $\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$
- si rang$(\mathbf{A}) = n$, $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$?

Normes

Une fonction f utilisée pour mesurer la 'longueur' d'un vecteur. Elle respecte les 3 conditions suivantes :

- $f(\mathbf{v}) = 0 \implies \mathbf{v} = 0$
- $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (inégalité triangulaire)
- $f(\alpha \mathbf{x}) = |\alpha| f(\mathbf{x})$

Exemples : norme L^p : $||\mathbf{x}||_p = \sqrt[p]{\sum_i |\mathbf{x}_i|^p}$

- $p = 2$ norme Euclidienne, aussi notez que $||\mathbf{x}||_2^2 = \mathbf{x}^T \mathbf{x}$
- $p = 1$ utile pour différentier des valeurs proches de 0
- $p = \infty$ norme max $||\mathbf{x}||_\infty = \max_i |\mathbf{x}_i|$

Cas particuliers de matrices

- Matrice diagonale : des entrées non-nulles seulement sur la diagonale. On note $\mathbf{V} = \text{diag}(\mathbf{v})$
- Matrice symétrique : $\mathbf{A} = \mathbf{A}^T$
- Matrice orthogonale : $\mathbf{A}^{-1} = \mathbf{A}^T$. Toutes les colonnes $\mathbf{a_{\bullet i}}$ sont orthogonales entre elles, i.e. $\forall i, j | i \neq j, \mathbf{a_{\bullet i}}^T \mathbf{a_{\bullet j}} = 0$

Décomposition de matrices

Les matrices peuvent être décomposées en facteurs (produit de matrices) pour gagner de la compréhension sur leurs structures.

- **Décomposition spectrale**, aussi appelée décomposition en valeurs/vecteurs propres
- un vecteur propre $\mathbf{v}$ d'une matrice $\mathbf{A}$ est telle qu'il existe un scalaire λ tel que $\mathbf{Av} = \lambda\mathbf{v}$
  - λ est la valeur propre associée à $\mathbf{v}$
  - tous les multiples de $\mathbf{v}$ sont des vecteurs propres de $\mathbf{A}$, i.e. les $s\mathbf{v}$ pour $s \in \mathbb{R}$
- trouver les valeurs propres est équivalent à résoudre $det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- les racines de ce polynôme sont les valeurs propres de $\mathbf{A}$
- exemple avec $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

Décomposition en vecteurs propres (eigendecomposition)

Si $\mathbf{A}$ a n vecteurs propres indépendants alors on peut l'écrire comme

$$\mathbf{A} = \mathbf{V}\text{diag}(\lambda)\mathbf{V}^{-1}$$

- $\mathbf{V}$: matrice des vecteurs propres
- λ : vecteur des valeurs propres

Cas d'une matrice symétrique :

$$\mathbf{A} = \mathbf{Q}\text{diag}(\lambda)\mathbf{Q}^{T}$$

- $\mathbf{Q}$ est une matrice orthogonale
- λ : vecteur des valeurs propres

Propriétés utiles de la décomposition en valeurs propres

- une matrice est **singulière** (non-inversible) ssi au moins une de ses valeurs propres est nulle
- le rang d'une matrice est égale au nombre de valeurs propres non nulles
- la décomposition en valeurs propres est utile pour résoudre certains problèmes d'optimisation
- ex : $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ avec $||\mathbf{x}||_2 = 1$
    - forme **quadratique**
    - si $\mathbf{x}$ est un vecteur propre, alors $f(\mathbf{x})$ est la valeur propre correspondante
    - $\min f = \min \lambda$, $\max f = \max \lambda$
- toutes les valeurs propres strictement positives : matrice **positive**
    - garantie que $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$
- toutes les valeurs propres positives ou nulles : matrice **semi-définie positive** (SDP)

Décomposition SVD

La décomposition **SVD** (**D**écomposition en **V**aleurs **S**ingulières) est une décomposition plus générale qui s'adapte aux matrices non-carrées :

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

- Si $\mathbf{A}$ est de taille $m \times n$, alors $\dim(\mathbf{U}) = m \times m$, $\dim(\mathbf{D}) = m \times n$ et $\dim(\mathbf{V}) = n \times n$
- les éléments de $\mathbf{D}$ sont appelées valeurs singulières
- les vecteurs de $\mathbf{U}$ et $\mathbf{V}$ sont respectivement vecteurs singuliers droits et gauches (resp. vecteurs propres de $\mathbf{A}\mathbf{A}^T$ et $\mathbf{A}^{\mathbf{T}}\mathbf{A}$)
- aspect pratique : calcul de pseudo-inverse avec SVD : $\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^T$, où $\mathbf{D}^+$ est formée avec la réciproque des éléments non nuls de $\mathbf{D}$.

Plan

Définition de la probabilité

Espace des épreuves Ω

Ensemble de tous les évènements possibles issus d'une expérience donnée.

Définition de $P(A)$

Soit A un ensemble d'évènements inclus dans Ω,

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n} \quad \text{si la limite existe,}$$

avec

- n le nombre d'expériences réalisées,
- $n(A)$ le nombre d'expériences où A s'est réalisé.

Exemple, dé à 6 faces

- $\Omega = \{\text{faces} : 1, 2, 3, 4, 5, 6\}$
- Si dé non pipé, alors $P(k) = 1/6, \quad \forall k \in 1, \ldots, 6$

Premier axiome

Si $A \in \Omega$ alors

$$0 \leq P(A) \leq 1$$

Deuxième axiome

$$P(\Omega) = 1 \quad P(\emptyset) = 0$$

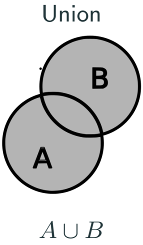
avec $\emptyset$ l'ensemble vide

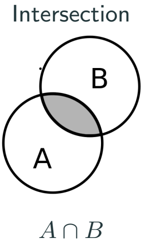
Union et intersection

Si $A \in \Omega,\ B \in \Omega$, alors

$$P(A \cup B) \quad = \quad P(A) + P(B) - P(A \cap B)$$

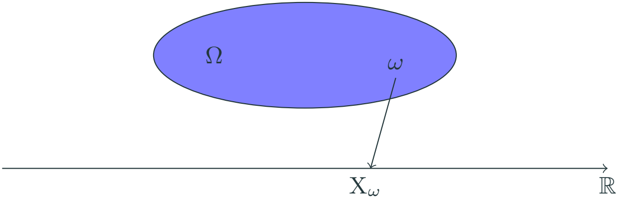
Si $A \cap B = \emptyset$ alors
$P(A \cup B) = P(A) + P(B)$.





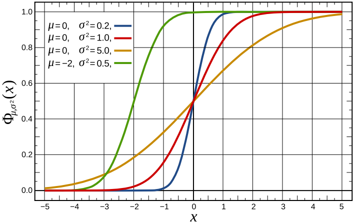
Variable Aléatoire

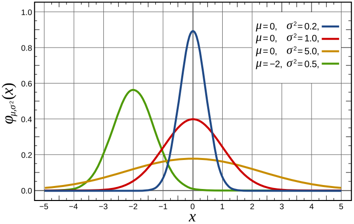
C'est un nombre (réel) X_ω dont la valeur est déterminée par le résultat ω d'une expérience aléatoire.



Exemple : dé à 6 faces

- L'évènement aléatoire (`e.a.`) est l'apparation d'une face.
- On associe un entier 1 à 6 à chaque face.





Fonction de répartition

La fonction de répartition $F_{\mathrm{X}}(x)$ d'une v.a.X est définie comme étant la probabilité que la v.a.X soit inférieur ou égale à x,

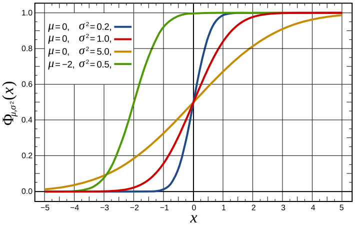
$$F_{\mathrm{X}}(x) = P(\mathrm{X} \leq x) \quad .$$

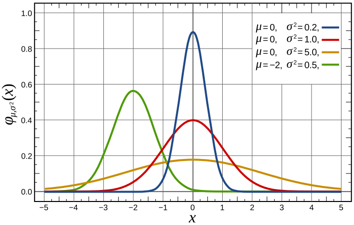
Densité de probabilité (d.d.p.)

Elle est définie comme la dérivée de la fonction de répartition,

$$p(x) = \frac{dF(x)}{dx} \quad .$$

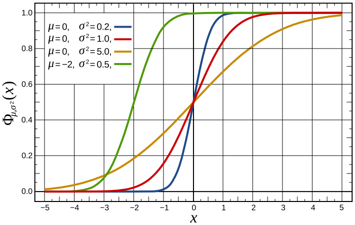
Propriétés

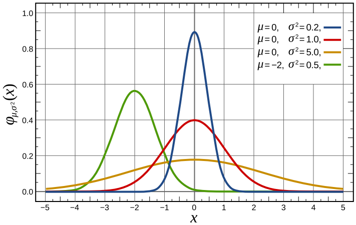




Propriétés de la fonction de répartition

Propriétés





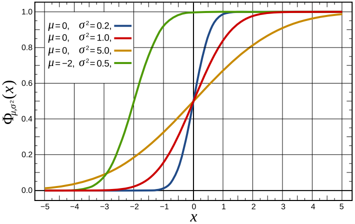
Propriétés de la fonction de répartition

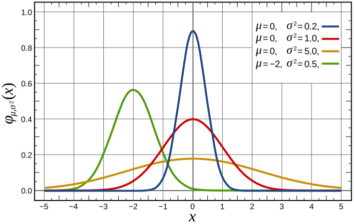
$$F_X(-\infty) = 0 \qquad\qquad F_X(\infty) = 1$$

$$0 \leq \quad F_X(x) \quad \leq 1$$

$$P(x_1 \leq x \leq x_2) \quad = \quad F_X(x_2) - F_X(x_1)$$

Propriétés de la densité de probabilité





Propriétés de la fonction de répartition

$$F_X(-\infty) = 0 \qquad F_X(\infty) = 1$$

$$0 \leq \quad F_X(x) \quad \leq 1$$

$$P(x_1 \leq x \leq x_2) \quad = \quad F_X(x_2) - F_X(x_1)$$

Propriétés de la densité de probabilité

$$p(x) \geq 0 \qquad \int_{-\infty}^{+\infty} p(x)dx = 1$$

$$P(x \leq x_1) = F_X(x_1) = \int_{-\infty}^{x_1} p(x)dx \qquad P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p(x)dx$$

Moments d'une `v.a.` (1)

Définition du moment

Le moment $g(x)$ d'une `v.a.` est donné par l'espérance,

$$E(g(x)) = \int_{-\infty}^{+\infty} g(x)p(x)dx$$

Généralement $g(x) = x^m$, on parle alors de moment d'ordre m,

$$\text{Moment d'ordre 1} \qquad m_X = E(\mathrm{X}) = \int_{-\infty}^{+\infty} xp(x)dx$$

$$\text{Moment d'ordre 2} \quad m_X^{(2)} = E(\mathrm{X}^2) = \int_{-\infty}^{+\infty} x^2p(x)dx$$

Le moment d'ordre 1 est aussi souvent appelé moyenne.

Propriété : linéarité de l'espérance

$$E(X+Y) = E(X) + E(Y), \qquad E(kX) = kE(X)$$

Pour k une constante.

Moments d'une `v.a.` (2)

Définition de la variance

La variance est l'espérance du carré des écarts par rapport à la valeur moyenne $m = E(\mathrm{X})$,

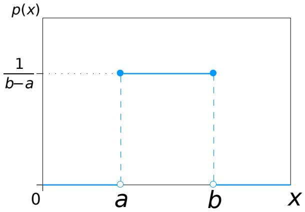
$$\begin{aligned}
\sigma_X^2 &= E\left((\mathrm{X} - m_X)^2\right) = \int_{-\infty}^{+\infty} (x - m_X)^2 p(x) dx \ , \\
\sigma_X^2 &= E\left(\mathrm{X}^2\right) - E\left(\mathrm{X}\right)^2 \ .
\end{aligned}$$

On utilise souvent aussi la notion d'écart-type σ,

$$\sigma_X = \sqrt{\sigma_X^2} \ .$$

Caractérisation incomplète

On caractérise, de manière incomplète, une `v.a.` par sa moyenne et sa variance.



Loi uniforme $\mathcal{U}(a,b)$

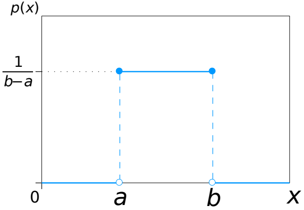
- Densité de probabilité

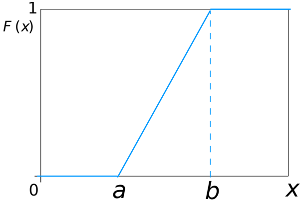
$$p(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b] \ , \\ 0 & \text{ailleurs } \ . \end{cases}$$

- Fonction de répartition

$$F(x) =$$

Exemples de lois (1)





Loi uniforme $\mathcal{U}(a, b)$

- Densité de probabilité

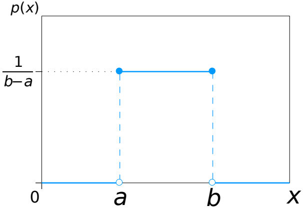
$$p(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \ , \\ 0 & \text{ailleurs} \ . \end{cases}$$

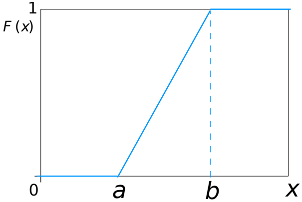
- Fonction de répartition

$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \ , \\ 1 & x > b \ . \end{cases}$$

- Espérance :

$$m_X = E(X) =$$





Loi uniforme $\mathcal{U}(a,b)$

- Densité de probabilité

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b] \ , \\ 0 & \text{ailleurs} \ . \end{cases}$$

- Fonction de répartition

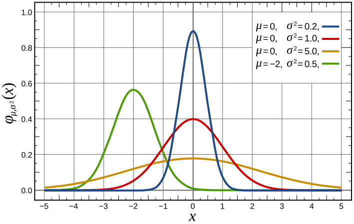
$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & \text{si } x \in [a,b] \ , \\ 1 & x > b \ . \end{cases}$$

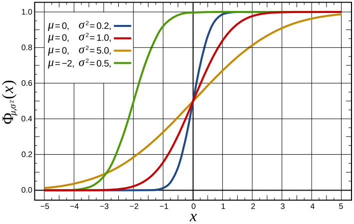
- Espérance :

$$m_X = E(X) = \frac{b+a}{2}$$

- Variance :

$$Var(X) = E((X - m_X)^2) = \frac{1}{12}(b-a)^2$$





Loi normale $\mathcal{N}(\mu, \sigma^2)$

- Densité de probabilité

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

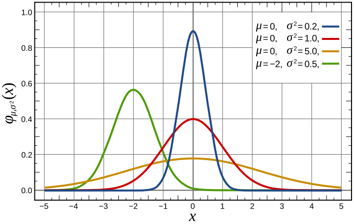
- Fonction de répartition

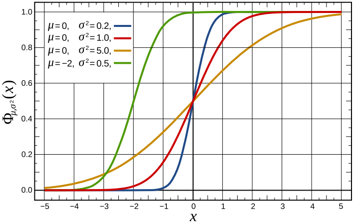
$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

- Espérance :

$$m_X = E(X) =$$

Exemples de lois (2)





Loi normale $\mathcal{N}(\mu, \sigma^2)$

- Densité de probabilité

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

- Fonction de répartition

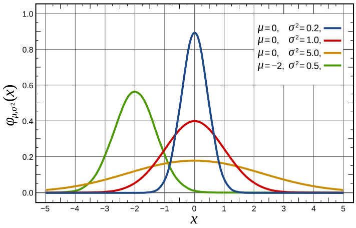
$$F(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

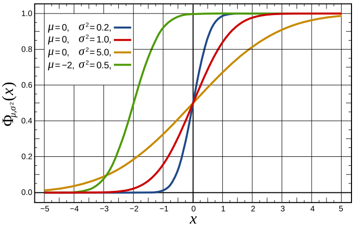
- Espérance :

$$m_X = E(X) = \mu$$

- Variance :

$$Var(X) = E((X - m_X)^2) =$$





Loi normale $\mathcal{N}(\mu, \sigma^2)$

- Densité de probabilité

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

- Fonction de répartition

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

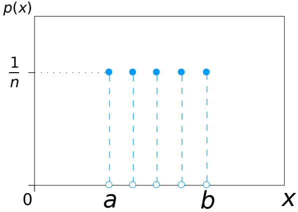
- Espérance :

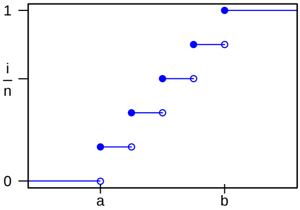
$$m_X = E(X) = \mu$$

- Variance :

$$Var(X) = E((X - m_X)^2) = \sigma^2$$

Exemples de lois (3)

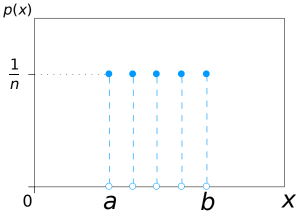


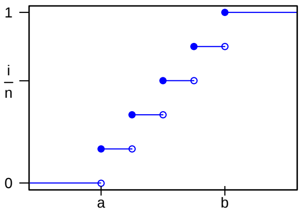


Loi uniforme discrète $\mathcal{U}(\{x_1, \ldots, x_n\})$

- $P(X = x_i) = \frac{1}{n}, \quad i \in 1, \ldots, n$
- x_i valeurs réelles.
- Densité de probabilité

$$p(x) =$$





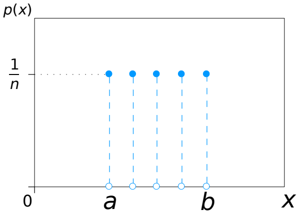
Loi uniforme discrète $\mathcal{U}(\{x_1, \ldots, x_n\})$

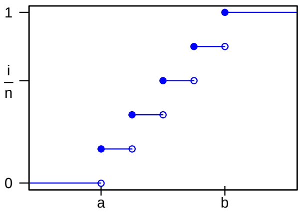
- $P(X = x_i) = \frac{1}{n}, \quad i \in 1, \ldots, n$
- x_i valeurs réelles.
- Densité de probabilité

$$p(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

- Fonction de répartition

$$F(x) =$$





Loi uniforme discrète $\mathcal{U}(\{x_1, \ldots, x_n\})$

- $P(X = x_i) = \frac{1}{n}, \quad i \in 1, \ldots, n$
- x_i valeurs réelles.
- Densité de probabilité

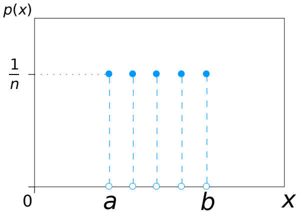
$$p(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

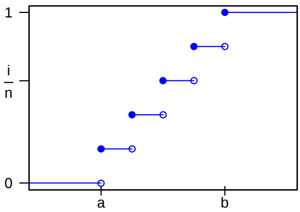
- Fonction de répartition

$$F(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x \geq x_i}$$

- Espérance :

$$m_X = E(X) =$$





Loi uniforme discrète $\mathcal{U}(\{x_1, \ldots, x_n\})$

- $P(X = x_i) = \frac{1}{n}, \quad i \in 1, \ldots, n$
- x_i valeurs réelles.
- Densité de probabilité

$$p(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

- Fonction de répartition

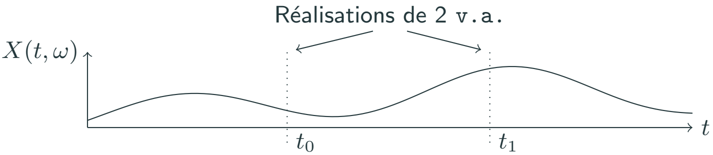
$$F(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x \geq x_i}$$

- Espérance :

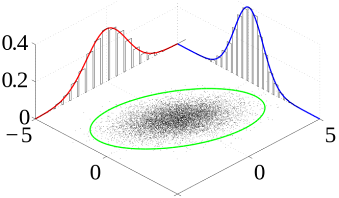
$$m_X = E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Variance :

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_X)^2$$



- On est souvent amené à considérer un ensemble de v.a.dans la mesure où à chaque instant t_i est associé une v.a..
- Modélisation jointe de ces variables.



- Lorsque l'on a plusieurs v.a. $X_1, X_2, \dots, X_d$ il est intéressant de modéliser ces v.a.par un vecteur aléatoire $\mathbf{X} \in \mathbb{R}^d$

Fonction de répartition mutuelle

Soit X et Y deux `v.a.` alors,

$$F(x, y) = P(\mathrm{X} \leq x, \mathrm{Y} \leq y)$$

Propriétés

Fonction de répartition mutuelle

Soit X et Y deux `v.a.`alors,

$$F(x,y) = P(\mathrm{X} \leq x, \mathrm{Y} \leq y)$$

Propriétés

$$0 \leq F(x,y) \leq 1$$

$$F(-\infty,-\infty) = 0$$

$$F(\infty,\infty) = 1$$

Densité de probabilité jointe

Soit X et Y deux `v.a.`alors,

$$\begin{aligned} p(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} \\ p(x) &= \int p(x,y)dy \\ p(y) &= \int p(x,y)dx \end{aligned}$$

$p(x)$ et $p(y)$ sont appelées lois marginales.

Propriétés

Densité de probabilité jointe

Fonction de répartition mutuelle

Soit X et Y deux `v.a.`alors,

$$F(x,y) = P(\mathrm{X} \leq x, \mathrm{Y} \leq y)$$

Propriétés

$$0 \leq F(x,y) \leq 1$$

$$F(-\infty,-\infty) = 0$$

$$F(\infty,\infty) = 1$$

Densité de probabilité jointe

Soit X et Y deux `v.a.`alors,

$$\begin{aligned} p(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} \\ p(x) &= \int p(x,y)dy \\ p(y) &= \int p(x,y)dx \end{aligned}$$

$p(x)$ et $p(y)$ sont appelées lois marginales.

Propriétés

$$p(x,y) \geq 0$$

$$\int\int p(x,y)dxdy = 1$$

$$\begin{aligned} p(A,B) &= P(x \in A, y \in B) \\ &= \int_A \int_B p(x,y)dxdy \end{aligned}$$

Probabilité conditionnelle

- Loi jointe $p(x, y)$.
- Probabilité d'une des variable sachant la valeur de la seconde.
- Notation : $p(x|y)$.

Théorème de Bayes

$$\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p(y)} \\
p(y|x) &= \frac{p(x,y)}{p(x)} \\
p(x,y) &= p(y|x)p(x) = p(x|y)p(y)
\end{aligned}$$

Pour caractériser l'interdépendance de deux variables, on introduit la notion de covariance.

Définitions

- Moments d'une loi jointe

$$E(g(x,y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\ p(x,y)dxdy$$

- Corrélation

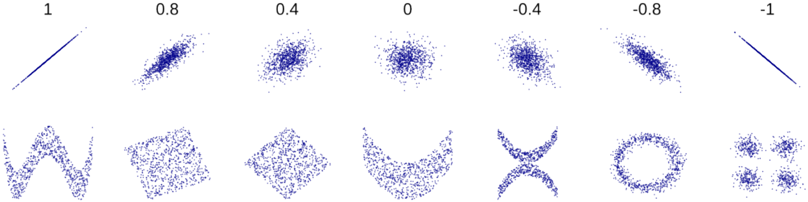
$$R_{\mathrm{XY}} = E(\mathrm{XY}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\ p(x,y)dxdy$$

- Covariance

$$\begin{aligned} C_{\mathrm{XY}} &= \sigma_{\mathrm{XY}} = E\left((\mathrm{X}-m_{\mathrm{X}})(\mathrm{Y}-m_{\mathrm{Y}})\right) \\ C_{\mathrm{XY}} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-m_{\mathrm{X}})(y-m_{\mathrm{Y}})\ p(x,y)dxdy \end{aligned}$$

- Coefficient de corrélation

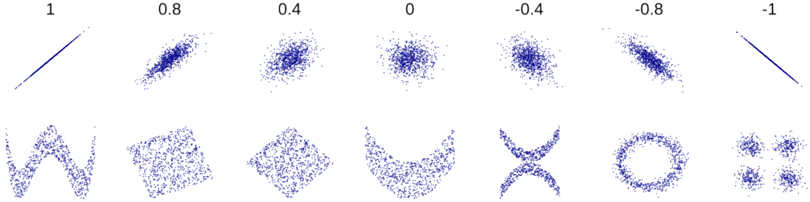
$$r_{\mathrm{XY}} = \frac{C_{\mathrm{XY}}}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}}$$



Covariance et Corrélation

$$R_{\mathrm{XY}} = E(\mathrm{XY}) =$$

Indépendance et corrélation



Covariance et Corrélation

$$R_{\mathrm{XY}} = E(\mathrm{XY}) = C_{\mathrm{XY}} + m_{\mathrm{X}} m_{\mathrm{Y}}$$

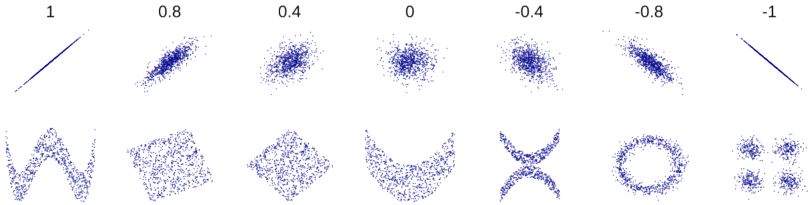
Indépendance

- Deux `v.a.` X et Y sont indépendantes si

$$p(x, y) = p(x)p(y)$$

- Si les variables sont indépendantes alors

$$R_{\mathrm{XY}} =$$



Covariance et Corrélation

$$R_{\mathrm{XY}} = E(\mathrm{XY}) = C_{\mathrm{XY}} + m_{\mathrm{X}} m_{\mathrm{Y}}$$

Indépendance

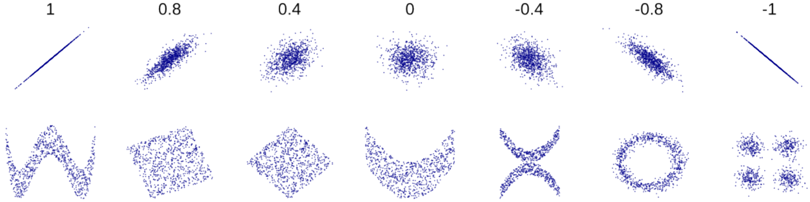
- Deux `v.a.` X et Y sont indépendantes si

$$p(x, y) = p(x)p(y)$$

- Si les variables sont indépendantes alors

$$R_{\mathrm{XY}} = m_{\mathrm{X}} m_{\mathrm{Y}} \qquad \text{et} \qquad C_{XY} =$$

Indépendance et corrélation



Covariance et Corrélation

$$R_{\mathrm{XY}} = E(\mathrm{XY}) = C_{\mathrm{XY}} + m_{\mathrm{X}} m_{\mathrm{Y}}$$

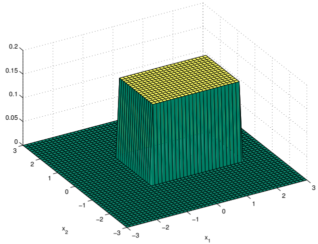
Indépendance

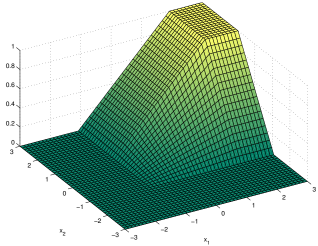
- Deux `v.a.` X et Y sont indépendantes si

$$p(x, y) = p(x)p(y)$$

- Si les variables sont indépendantes alors

$$R_{\mathrm{XY}} = m_{\mathrm{X}} m_{\mathrm{Y}} \qquad \text{et} \qquad C_{XY} = 0.$$





Loi uniforme multivariée

- $X \sim U(a_x, b_x)$ et $Y \sim U(a_y, b_y)$
- $\mathbf{X} = [X, Y]^\top$
- Densité de probabilité

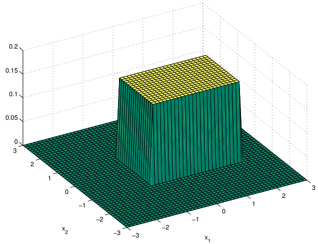
$$p(x,y) = \begin{cases} \frac{1}{S} & \text{si } x \in [a_x, b_x], \\ & \text{et } y \in [a_y, b_y] \\ 0 & \text{sinon} \end{cases}$$

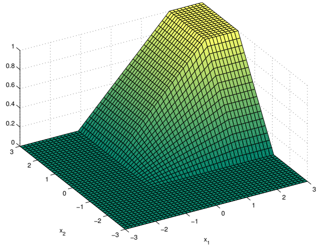
- Surface $S = (b_x - a_x)(b_y - a_y)$

- Espérance :

$$\mathbf{m}_X = E(\mathbf{X}) =$$

Exemples de système de v.a.(1)





Loi uniforme multivariée

- $X \sim U(a_x, b_x)$ et $Y \sim U(a_y, b_y)$
- $\mathbf{X} = [X, Y]^\top$
- Densité de probabilité

$$p(x, y) = \begin{cases} \frac{1}{S} & \text{si } x \in [a_x, b_x], \\ & \text{et } y \in [a_y, b_y] \\ 0 & \text{sinon} \end{cases}$$

- Surface $S = (b_x - a_x)(b_y - a_y)$

- Espérance :

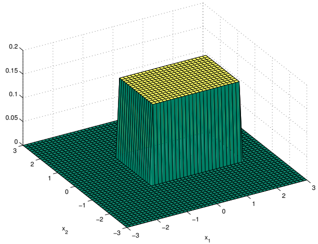
$$\mathbf{m}_X = E(\mathbf{X}) = \begin{bmatrix} \frac{b_x + a_x}{2} \\ \frac{b_y + a_y}{2} \end{bmatrix}$$

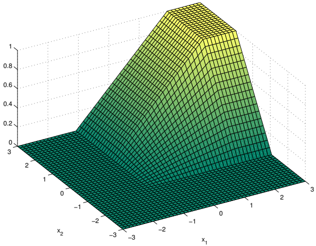
- Covariance :

$$Cov(\mathbf{X}) = E((\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top)$$

$$=$$

Exemples de système de `v.a.`(1)





Loi uniforme multivariée

- $X \sim U(a_x, b_x)$ et $Y \sim U(a_y, b_y)$
- $\mathbf{X} = [X, Y]^\top$
- Densité de probabilité

$$p(x,y) = \begin{cases} \frac{1}{S} & \text{si } x \in [a_x, b_x], \\ & \text{et } y \in [a_y, b_y] \\ 0 & \text{sinon} \end{cases}$$

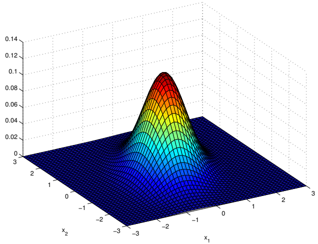
- Surface $S = (b_x - a_x)(b_y - a_y)$

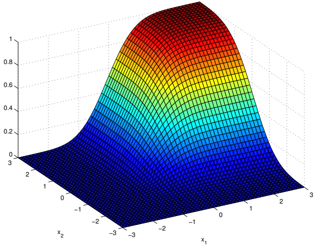
- Espérance :

$$\mathbf{m}_X = E(\mathbf{X}) = \begin{bmatrix} \frac{b_x + a_x}{2} \\ \frac{b_y + a_y}{2} \end{bmatrix}$$

- Covariance :

$$\begin{aligned} Cov(\mathbf{X}) &= E((\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top) \\ &= \begin{bmatrix} Var(X) & 0 \\ 0 & Var(Y) \end{bmatrix} \end{aligned}$$

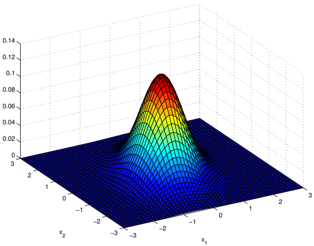


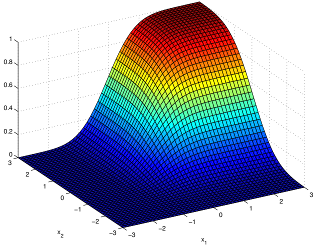


Loi gaussienne multivariée

- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Densité de probabilité

$$p(x, y) =$$





Loi gaussienne multivariée

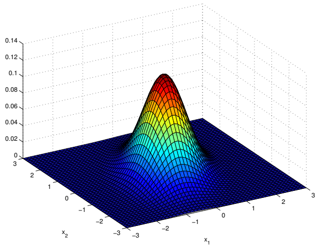
- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Densité de probabilité

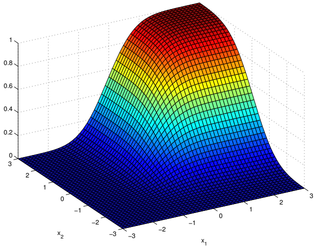
$$p(x, y) = Ke^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

- Coefficient $K = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}}$

- Espérance :

$$\mathbf{m}_X = E(\mathbf{X}) =$$





Loi gaussienne multivariée

- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Densité de probabilité

$$p(x, y) = Ke^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

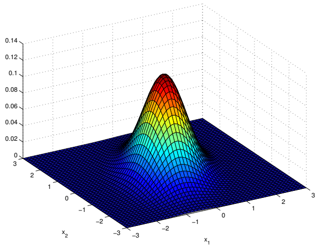
- Coefficient $K = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}}$

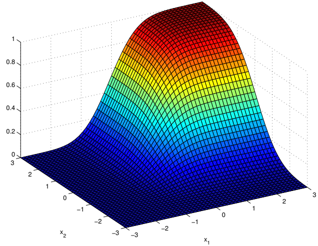
- Espérance :

$$\mathbf{m}_X = E(\mathbf{X}) = \boldsymbol{\mu}$$

- Covariance :

$$Cov(\mathbf{X}) = E((\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top)$$
$$=$$





Loi gaussienne multivariée

- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Densité de probabilité

$$p(x, y) = Ke^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

- Coefficient $K = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}}$

- Espérance :

$$\mathbf{m}_X = E(\mathbf{X}) = \boldsymbol{\mu}$$

- Covariance :

$$\begin{aligned} Cov(\mathbf{X}) &= E((\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top) \\ &= \boldsymbol{\Sigma} \end{aligned}$$