Rapport séance 4

Probabilités

Mohsine AABID
Etienne BOST

Enseignant : M.Thomas Shatz

Septembre 2022

Contents

1 Modèle graphique **2**
- 1.1 Qu'est ce qu'un modèle graphique ? . . . 2
- 1.2 Définition du modèle graphique (Réseaux bayésiens) . . . 2
- 1.3 Exemple d'un réseaux bayésien . . . 2

2 Propriétés sur l'espérance **3**
- 2.1 Règle d'espérance totale . . . 3
- 2.2 Application de la règle . . . 3

3 Covariance **4**
- 3.1 Qu'est-ce la covariance? . . . 4
- 3.2 Définition . . . 4
- 3.3 Propriétés . . . 4
- 3.4 Correlation de Pearson . . . 4
- 3.5 Variance d'une somme de variables aléatoires . . . 5

4 Espérance et variance de la moyenne empirique **5**
- 4.1 L'espérance de la moyenne empirique . . . 5
- 4.2 Variance d'une moyenne empirique . . . 6

5 Théorème de la variance totale **6**
- 5.1 Théorème . . . 6
- 5.2 Exemple . . . 7

6 Quelques exemples de distributions **7**
- 6.1 Bernoulli . . . 7
- 6.2 Binomiale . . . 8
- 6.3 Géométrique . . . 8
- 6.4 Poisson . . . 8
- 6.5 Uniforme . . . 8
- 6.6 Loi normale (Gaussiènne) . . . 8
- 6.7 Exponentielle . . . 9

7 Vecteurs aléatoires **9**
- 7.1 définition . . . 9
- 7.2 Espérance d'un vecteur aléatoire . . . 9

8 Matrices des covariances **9**
- 8.1 définition . . . 9
- 8.2 Propriétés: . . . 10

9 Loi normale multidimensionnelle **10**
- 9.1 Définition . . . 10
- 9.2 Cas d'utilisation . . . 11
- 9.3 Exemples de relations entre la matrice de covariance et la distribution gaussienne en deux dimensions . . . 11

1 Modèle graphique

1.1 Qu'est ce qu'un modèle graphique ?

Un modèle graphique est un graphe qui peut être orienté ou non orienté dont les sommets représentent des variables aléatoires et dont les arcs (les arrêtes dans le cas non orienté) représentent les dépendances de ces variables.

Le modèle graphique est une représentation d'objets probabilistes qui est notamment utilisée en apprentissage automatique.

Il existe deux familles de modèles graphiques, les réseaux bayésiens qu'on a vu dans le cours (les graphes orientés acycliques), et les champs aléatoires de Markov qui sont non orientés.

1.2 Définition du modèle graphique (Réseaux bayésiens)

- Soit $\{X_i\}_{i\in[\![1,n]\!]}$ une famille de variables aléatoires.
- Soit $G = (V, A)$ un graphe orienté acyclique.
- Soit la fonction f définit comme ci-dessous:

$$\begin{array}{rcl} f \quad : \{X_i\}_{i\in[\![1,n]\!]} & \rightarrow & V \\ X_i & \mapsto & P(X_i|(X_{\pi i})) \end{array}$$

Où $(X_{\pi i})$ sont les parents de X_i dans le graphe G.

Alors la probabilité jointe des variables $\{X_i\}_{i\in[\![1,n]\!]}$ est égale:

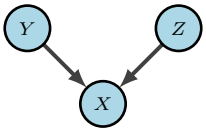
$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|(X_{\pi i}))$$

En d'autre termes chaque sommet est conditionnellement indépendant de ses non-descendants.

On remarque que cette formule de la probabilité jointe est en réalité la chain rule $P(X_1, ..., X_n) = P(X_1)P(X_2|X_1)...P(X_n|X_1, ..., X_{n-1})$ où les variables indépendantes n'apparaissent pas. Notons que pour n=3, toute distribution de probabilité se factorise dans un modèle graphique. Ce n'est pas le cas en général mais si c'est possible, un modèle graphique permet de simplifier la loi jointe et de trouver les indépendances.

1.3 Exemple d'un réseaux bayésien

Voici un exemple d'un modèle graphique:



On remarque que le graphe est orienté est acyclique. Nous remarquons que Y et Z sont indépendants des autres variables aléatoires car (Y et Z n'ont pas de parents) et que X dépend à la fois de Y et Z Nous pouvons donc à partir de ce graphe exprimer la loi de probabilité jointe comme ci-dessous:

$$p(X, Y, Z) = P_1(X|Y, Z)P_2(Y)P_3(Z)$$

Supposons que nos variables aléatoires soient binaires et prennent donc les valeurs 0 ou 1 et que les distributions suivantes soient données:

- $P_1(X|Y, Z) = \mathbb{1}_{XOR(Y,Z)}\mathcal{N}(\mu, 1) + \mathbb{1}_{XOR(Y,Z)}\mathcal{N}(0, 1)$
- $P_2(Y = 1) = a$; avec: $0 \leq a \leq 1$
- $P_3(Z = 1) = b$; avec: $0 \leq b \leq 1$

La question est de calculer la probabilité $p(Y, Z|X)$. Pour cela, on peut se reporter à un schéma de résolution assez général qui consiste à exprimer la quantité que l'on cherche en fonction de la distribution de probabilité jointe.

Dans notre exemple, utilisons la règle de Bayes

- On a $p(Y, Z|x) = \frac{p(X,Y,Z)}{P_1(X)}$
- On connaît la probabilité jointe $p(X, Y, Z)$ (elle est donné par le modèle graphique)
- Il nous reste donc de calculer la probabilité $P_1(X)$

Remarque 1. *La manière de calculer la probabilité d'une variable aléatoire peut être différente selon le fait d'être dans un espace discret ou continu. Voici comment calculer $P_1(X)$ dans les deux cas:*

- ***Cas discret:*** $\sum_y \sum_z p(X|Y, Z)$
- ***Cas continu:*** $\int_z \int_y p(X|Y, Z)dydz$

(On peut visualiser une intégrale comme une somme continue).

Dans notre cas les variables aléatoire Y et Z prennent leurs valeurs sur $0, 1$, nous sommes donc dans le cas discret.

$$P_1(X) = \sum_{Z\in\{0,1\}} \sum_{Y\in\{0,1\}} p(X|Y, Z)$$

Finalement on trouve:

$$\begin{aligned} P_1(X) = &(1-a)(1-b)P_1(X|Y=0, Z=0) \\ &+ a(1-b)P_1(X|Y=1, Z=0) \\ &+ (1-a)bP_1(X|Y=0, Z=1) \\ &+ abP_1(X|Y=1, Z=1) \end{aligned}$$

2 Propriétés sur l'espérance

2.1 Règle d'espérance totale

Soient X, Y deux variables aléatoires on a alors:

$$E_{P_y}[E_{P_{X|Y}}[X|Y]] = E_{P_X}[X]$$

Remarque 2. *On a $E[X|Y] = \sum_{x\in Val(x)} x.p_{x|y}(x|y)$ est une fonction qui dépend de Y. On pourrait aussi noter $E[X|Y] = E_{P_{X|Y}}[X]$.*

2.2 Application de la règle

El Goog se procure deux batteries, A et B, pour son téléphone. Un téléphone avec la batterie A fonctionne en moyenne 12 heures sur une seule charge, mais seulement 8 heures en moyenne avec la batterie B. El Goog met la batterie A en 80% de ses téléphones et la batterie B dans le reste. Si vous achetez un téléphone d'El Goog, combien d'heures pensez-vous qu'il fonctionnera sur une seule charge?

On commence par définir les variables aléatoires de notre problème.

- X: Le type de la batterie
- T: Nombre d'heure que le téléphone va tenir en une seule charge.

D'après l' énoncé on deux types de batteries. Dans 80% des cas le téléphone a une batterie A et 20% des cas a une batterie B.

On a donc:

$$\text{p(X = A)} = 0.8 \text{ et p(X=B)} = 0.2$$

On sait qu'un téléphone avec une batterie A tiens en moyenne 12 heures et un téléphone avec une batterie B peut tenir 8 heures.

$$E[T|X = A] = 12 \text{ et } E[T|X = B] = 8$$

On cherche à calculer l'espérance du temps de fonctionnement d'un téléphone El Goog. D'après la règle de l'espérance totale on a:

$$E[T] = E_X[E_{T|X}[T]] = E[E[T|X]$$

Voici le calcul:
$E[T] = \sum_{X\in\{A,B\}} E[T|X].p_X(X) = E[T|A]p_X(A) + E[T|B]p_X(B) = 12\text{x}0.8 + 8\text{x}0.2 = 11,2$

3 Covariance

3.1 Qu'est-ce la covariance?

la covariance entre deux variables aléatoires est un nombre permettant de quantifier leurs écarts conjoints par rapport à leurs espérances respectives. Elle est utiliser pour déterminer la direction d'une relation linéaire entre ces deux variables

3.2 Définition

Soient X, Y deux variables aléatoires on définit la covariance comme ci-dessous:

$$\text{Cov}(X, Y) \equiv \text{E}[(X - \text{E}[X])\,(Y - \text{E}[Y])]$$

- Si $\text{Cov}(X, Y) = 0$ alors X et Y ne sont pas corrélés.
- Si $\text{Cov}(X, Y) > 0$ alors X et Y ont une corrélation positive.
- Si $\text{Cov}(X, Y) < 0$ alors X et Y ont une corrélation négative.

3.3 Propriétés

Voici quelques propriétés de la covariance qui découlent de sa définition.

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(cX, Y) = c\,\text{Cov}(X, Y)$ où c est une constante.
- $\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$ où c est une constante
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ où X, Y et Z sont trois variables.
- Si $X \perp\!\!\!\perp Y$ alors $\text{Cov}(X, Y) = E[XY] = E[X].E[Y] = 0$ La réciproque n'est pas forcément vrai.
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$

3.4 Correlation de Pearson

Il est possible de définir à partir de la covariance de deux variable aléatoire la corrélation de Pearson:
$Corr(X, Y) = \frac{Cov(X,Y)}{Var(X)Var(Y)}$
Cette quantité prend ses valeurs dans [-1,1]. Si elle est proche de 1, les variables X et Y sont fortement corrélées, si elle est proche de -1, elles sont anti-corrélées et si $Corr(X, Y) = 0$ alors X et Y sont indépendantes.

3.5 Variance d'une somme de variables aléatoires

Soient $\{X_i\}_{i\in[\![1,n]\!]}$ une familles de variables aléatoires. La variances de la somme de ces variables aléatoires est définie comme ci-dessous:

$$\mathrm{Var}(\textstyle\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j)$$

Proof. Montrons par récurrence que $\forall n \in \mathbb{N}^*$:

$$\mathrm{Var}(\textstyle\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j)$$

Pour n= 1: $Var(X_1) = Var(X_1) + 0$

Soit $n \in \mathbb{N}^*$ supposons que:

$$\mathrm{Var}(\textstyle\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j)$$

Montrons pour n+1:

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^{n+1} X_i) &= \mathrm{Var}(\sum_{i=1}^{n} X_i + X_{n+1}) \\
&= E[((\sum_{i=1}^{n} X_i + X_{n+1}) - E[\sum_{i=1}^{n} X_i + X_{n+1})^2] \\
&= E[((\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]) + (X_{n+1} - E[X_{n+1}]))^2] \\
&= E[(\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i])^2] + E[(X_{n+1} - E[X_{n+1}])^2] + 2E[(\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i])(X_{n+1} - E[X_{n+1}])] \\
&= \mathrm{Var}(\sum_{i=1}^{n} X_i) + \mathrm{Var}(X_{n+1}) + 2\,\mathrm{Cov}(\sum_{i=1}^{n} X_i, X_{n+1}) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j) + \mathrm{Var}(X_{n+1}) + 2\sum_{i=1}^{n} \mathrm{Cov}(X_i, X_{n+1}) \\
&= \sum_{i=1}^{n+1} \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n+1} \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

On déduit que: $\forall n \in \mathbb{N} : \mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j)$

□

4 Espérance et variance de la moyenne empirique

On définit la moyenne empirique comme suit: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$

Dans cette section on cherche à savoir si la moyenne empirique est proche de la moyenne réelle.

4.1 L'espérance de la moyenne empirique

Proof. On va calculer l'espérance de cette moyenne empirique:

$$
\begin{aligned}
E[\hat{\mu}] &= E[\frac{1}{n}\sum_{i=1}^{n} X_i] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\
&= \frac{1}{n} n\mu \\
&= \mu
\end{aligned}
$$

□

4.2 Variance d'une moyenne empirique

Proof. Avant de commencer la preuve, il est important de remarquer que les variables aléatoires sont iid (indépendantes et identiquement distribuer), donc leurs covariances vont être nulles.

$$
\begin{aligned}
Var(\hat{\mu}) &= Var(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}(\sum_{i=1}^{n} Var(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

□

Remarque 3. *On remarque que quand n tends vers $+\infty$ alors la variance de $\hat{\mu}$ tends vers 0.*

5 Théorème de la variance totale

5.1 Théorème

Aussi connu sous le nom de formule de décomposition de la variance. Si X et Y sont deux variables aléatoires sur un même espace de probabilité, et si l'espérance de Y est finie alors:

$$\mathrm{Var}(Y) = E_X[Var(Y|X)] + Var_X(E[Y|X])$$

Proof. On écrit la variance de Y comme ci-dessous:

$Var(Y) = E[Y^2] - (E[Y])^2$

On applique ensuite la formule des espérances totales sur chaque terme.
$Var(Y) = E[E[Y^2|X]] - (E[E[Y|X]])^2$

On ajoute et on retranche $E[Y|X]^2$ de la formule pour faire apparaître la variance.

$Var(Y) = E[E[Y^2|X] - E[Y|X]^2 + E[Y|X]^2] - (E[E[Y|X]])^2$

$Var(Y) = E[Var(Y|X) + E[Y|X]^2] - (E[E[Y|X]])^2$

En utilisant la linéarité de l'espérance on trouve que:

$Var(Y) = E[Var(Y|X)] - (E[E[Y|X]^2] - E[E[Y|X]]^2)$

On remarque que:

$Var(E[Y|X]) = E[E[Y|X]^2] - E[E[Y|X]]^2$

On remplace et on trouve:

$Var(Y) = E[Var(Y|X)] - Var(E[Y|X])$

□

5.2 Exemple

On a une procédure aléatoire pour entraîner un classificateur binaire, dont on obtient n échantillons (n classifieurs). Pour tester la qualité de la procédure d'entraînement, on a une procédure aléatoire de test qui produit une erreur de classification et qu'on applique m fois sur chacun des n classificateurs entraîné . Comment mesurer la variance de l'erreur de classification (par exemple pour savoir si elle est significativement en dessous du hasard) à partir des erreurs de classifications $(e_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$?

Solution:

On va utiliser le fait que qu'ils existe des estimateurs de l'espérance et de la variance qui sont non biaisés. On remarque que les erreurs de classification ne sont pas forcément indépendantes deux à deux. Pour calculer la variance de l'erreur de classification voici comment on procède.

$Var(erreur) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(e_{i,j} - \hat{\mu}_i)^2 + \frac{1}{m}\sum_{i=1}^{n}[\frac{1}{m}\sum_{j=1}^{m} e_{i,j} - \frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} e_{i,j}]$

Avec $\hat{\mu} = \sum_{k=1}^{m} e_{i,k}$

On peut aussi utiliser une notation plus condensée avec $e_{i.}$ la moyenne des erreurs de classification du i ème classifieur:
$Var(erreur) = \frac{1}{n}\sum_{i=1}^{n}[\sum_{j=1}^{m}[e_{ij} - e_{i.}]^2] + \frac{1}{n}\sum_{i=1}^{n}[e_{i.} - e_{..}]^2$

6 Quelques exemples de distributions

6.1 Bernoulli

la loi de Bernoulli, désigne la loi de probabilité d'une variable aléatoire discrète qui prend la valeur 1 avec la probabilité p et 0 avec la probabilité q = 1 − p.

- **Fonction de masse:** $\mathbb{P}(X = x) = \begin{cases} p & \text{si } x = 1, \\ 1-p & \text{si } x = 0, \\ 0 & \text{sinon.} \end{cases}$
- **Espérance:** p
- **Variance:** $p(1-p)$

6.2 Binomiale

La loi binomiale modélise la fréquence du nombre de succès obtenus lors de la répétition de plusieurs expériences aléatoires identiques et indépendantes.

- **Fonction de masse:**$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- **Espérance:** np
- **Variance:** $np(1-p)$

6.3 Géométrique

la loi géométrique désigne a loi du nombre X d'épreuves de Bernoulli indépendantes de probabilité de succès $p \in]0, 1[$(ou q = 1 − p d'échec) nécessaire pour obtenir le premier succès. X est la variable aléatoire donnant le rang du premier succès.

- **Fonction de masse:**$\mathbb{P}(X = k) = q^{k-1}p$.
- **Espérance:** $\frac{1}{p}$
- **Variance:** $\frac{1-p}{q^2}$

6.4 Poisson

La loi de Poisson est une loi de probabilité discrète qui décrit le comportement du nombre d'événements se produisant dans un intervalle de temps fixé, si ces événements se produisent avec une fréquence moyenne ou espérance connue, et indépendamment du temps écoulé depuis l'événement précédent.

- **Fonction de masse:**$p(k) = \mathbb{P}(X = k) = \dfrac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$
- **Espérance:** λ
- **Variance:** λ

En règle général, la loi de poisson est principalement utilisée pour modéliser des évènements rares.

6.5 Uniforme

une loi discrète uniforme est une loi de probabilité discrète indiquant une probabilité de se réaliser identique (équiprobabilité) à chaque valeur d'un ensemble fini de valeurs possibles.

- **densité:**$f(x) = \begin{cases} \frac{1}{b-a} & \text{pour } a \leq x \leq b, \\ 0 & \text{sinon.} \end{cases}$
- **Espérance:** $\dfrac{a+b}{2} \frac{a+b}{2}$
- **Variance:** λ

6.6 Loi normale (Gaussiènne)

Les lois normales sont parmi les lois de probabilité les plus utilisées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires. C'est une loi de probabilité absolument continue qui dépend de deux paramètres : son espérance, un nombre réel noté μ, et son écart type, un nombre réel positif noté σ.

- **densité:** $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- **Espérance:** μ
- **Variance:** σ^2

6.7 Exponentielle

La loi exponentielle est beaucoup utilisée pour décrire des évènements qui évoluent dans le temps.

- **densité:** $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$
- **Espérance:** $\frac{1}{\lambda}$
- **Variance:** $\frac{1}{\lambda^2}$

7 Vecteurs aléatoires

7.1 définition

Soient $X_1, X_2, ..., X_n$ n variables aléatoires, on peut définir un vecteur aléatoire comme un vecteur dont les composantes sont ces variables.

$$X = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{pmatrix}$$

En ayant une fonction:

$$\begin{array}{cccc} g & : & \mathbb{R}^n & \rightarrow & \mathbb{R}^m \\ & & X & \mapsto & g(X) \end{array}$$

$$g(X) = \begin{pmatrix} g(X_1) \\ g(X_2) \\ . \\ . \\ . \\ g(X_m) \end{pmatrix}$$

7.2 Espérance d'un vecteur aléatoire

On peut définir l'espérance de ce vecteur aléatoire comme ci-dessous:

$$\mathbb{E}[g(X)] = \begin{pmatrix} \mathbb{E}[g(X_1)] \\ \mathbb{E}[g(X_2)] \\ . \\ . \\ . \\ \mathbb{E}[g(X_m)] \end{pmatrix}$$

8 Matrices des covariances

8.1 définition

On défini la covariance de deux variables aléatoires de la manière suivante:

$$Cov[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}(X_i)(X_j)_{\mathbb{E}}(X_j)]$$

On a donc que:

$$Cov[X_i, X_i] = Var[X_i]$$

On peut donc écrire la matrice de covariance du vecteur aléatoire:

$$\Sigma = \begin{pmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] & ... & Cov[X_1, X_n] \\ Cov[X_2, X_1] & Cov[X_2, X_2] & ... & Cov[X_2, X_n] \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ Cov[X_n, X_1] & Cov[X_n, X_2] & ... & Cov[X_n, X_n] \end{pmatrix}$$

8.2 Propriétés:

- Si Σ est diagonale, cela signifie qu'il n'y a pas de corrélation entre les variables.
- Σ est une matrice symétrique définie positive.
- L'inverse de la matrice de covariance est parfois désignée matrice de précision.

Remarque 4. *L'espérance d'un vecteur aléatoire nous permet d'obtenir la matrice de covariance de ces variables.*

$$\sigma = \mathbb{E}[(X - \mathbb{E}(X)(X)_{\mathbb{E}}(X)^T]$$

Remarque 5. *Dans un cadre plus concret, On peut imaginer des variables aléatoires qui définissent la couleur des pixels d'une image. Si on calcul leur matrice de covariance, on va observer de fortes corrélations entre les pixels qui sont géographiquement proches dans l'image.*

9 Loi normale multidimensionnelle

9.1 Définition

Il existe une généralisation de la loi normale à un vecteur aléatoire (d'où son nom loi normale multidimensionnelle ou gaussienne multivariée). Sa fonction de densité peut s'exprimer de la manière suivante:
$X \sim \mathbf{N}(\mu, \Sigma)$ avec $X \in \mathbb{R}^n$ et Σ la matrice de covariance associée à X.

$$p(x; \mu, \Sigma) = \frac{1}{det(\Sigma)^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$$

Par identification, si on prend n=1 (càd on a une seule VA x) alors:

$$\Sigma = Cov[x, x] = Var(x) = \sigma^2$$
$$\sqrt{(det(\Sigma))} = \sigma$$
$$\Sigma^{-1} = \frac{1}{\sigma^2}$$

Et ainsi, pour n=1:

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

Ce qui correspond à la loi de densité pour la loi normale à une variable ce qui valide cette généralisation à plusieurs variables.

9.2 Cas d'utilisation

Cette généralisation est très utile pour modéliser des évènements aléatoires et est facile à manipuler notamment pour faire de l'inférence dessus. De plus la loi gaussienne multivariée possède certaines propriétés remarquables. Parmi celles-ci on retrouve:

- Les lois marginales et conditionnelles d'une gaussienne multivariée sont gaussiennes.
- Une gaussienne multivariée de dimension d dont la matrice de covariance est diagonale est une collection de d lois normales indépendantes deux à deux.
- Les isocontours d'une gaussienne multivariée sont généralement des ellipsoïdes (on peut retrouver l'équation d'une ellipse dans la formule de la densité) dont les axes principaux sont les vecteurs propres de la matrice de covariance (exemples ci-dessous)

9.3 Exemples de relations entre la matrice de covariance et la distribution gaussienne en deux dimensions

Les 4 figures présentées ci-dessus sont des représentation graphiques d'une distribution de probabilité suivant une loi gaussienne à deux variables centrée en (0,0) pour lesquelles on fait varier la matrice de covariance afin de se représenter graphiquement ce que ces modifications entraînent.

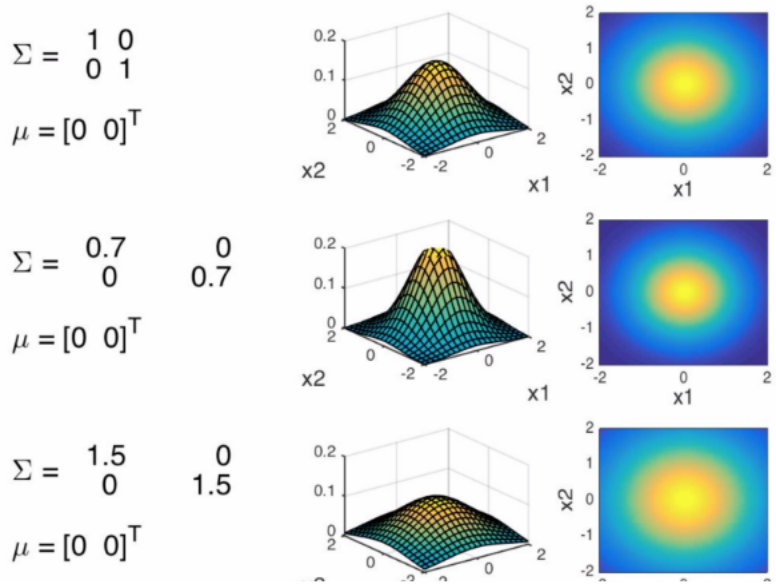


Figure 1: Variables indépendantes avec la même variance
Dans le cas où la matrice de covariance est diagonale, on constate que des variables aléatoires indépendantes ayant la même variance font apparaître une distribution centrée sur μ où la densité maximale est atteinte. Autour de ce point la distribution est identique selon chaque direction. La valeur de densité maximale dépend de la variance des variables aléatoires de manière analogue à ce que l'on observe pour une loi normale sur une variable.

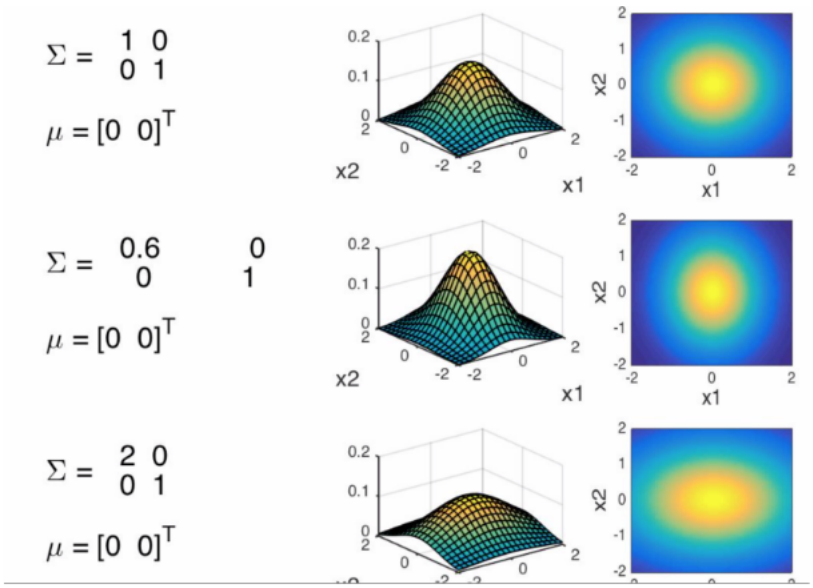


Figure 2: Variables indépendantes avec une variance différentes

Si les variables aléatoires sont indépendantes mais que leur variances respectives sont différentes, cela va étirer la distribution selon l'axe correspondant à la variables avec la variance la plus importante. Plus les variables aléatoires sont corrélées, plus la distribution est étirée selon un axe qui correspond au vecteur propre de la matrice de covariance.

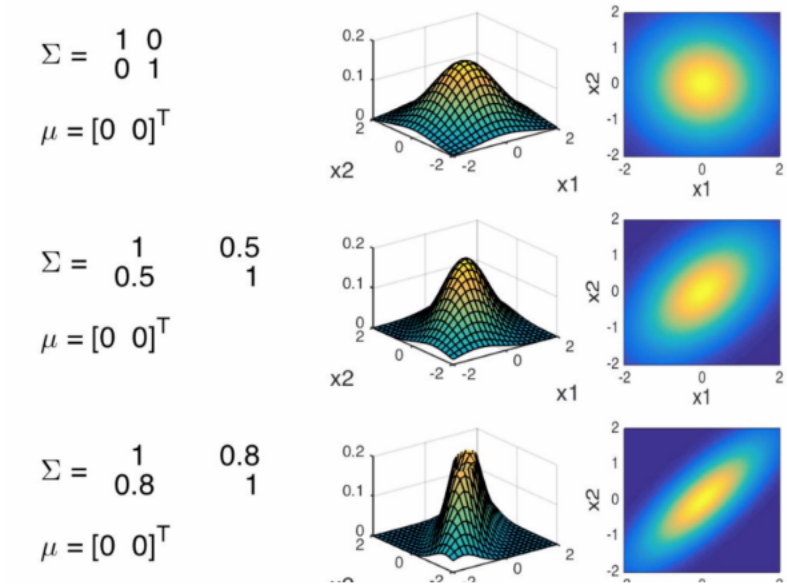


Figure 3: Variables corrélées

Les variables aléatoires sont anti-corrélées et la distribution est étirée selon l'axe orthogonal au précédant.

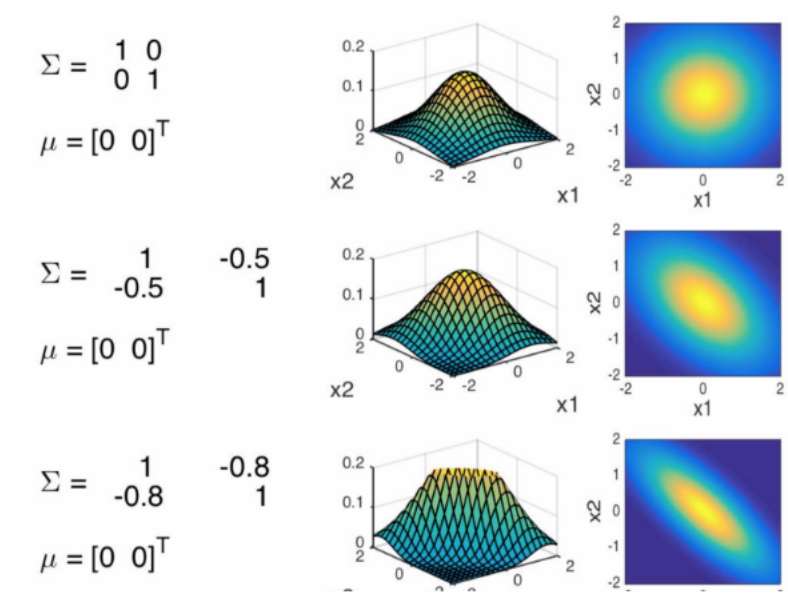


Figure 4: Variables anti-corrélées

La matrice de covariance est un outil essentiel pour l'analyse multivariée :

- L'analyse en composantes principales exploite la diagonalisation de cette matrice .
- L'analyse discriminante se fonde sur l'examen des coefficients de cette matrice.