

Probabilités



MOHSINE AABID
ETIENNE BOST

Enseignant : M.THOMAS SHATZ

Contents

1	Modèle graphique	2
1.1	Qu'est ce qu'un modele graphique ?	2
1.2	De nition du modele graphique (Reseaux bayesiens)	2
1.3	Exemple d'un reseaux bayesien	2
2	Propriétés sur l'espérance	3
2.1	Regle d'esperance totale	3
2.2	Application de la regle	3
3	Covariance	4
3.1	Qu'est-ce la covariance?	4
3.2	De nition	4
3.3	Proprietes	4
3.4	Correlation de Pearson	4
3.5	Variance d'une somme de variables aleatoires	5
4	Espérance et variance de la moyenne empirique	5
4.1	L'esperance de la moyenne empirique	5
4.2	Variance d'une moyenne empirique	6
5	Théorème de la variance totale	6
5.1	Theoreme	6
5.2	Exemple	7
6	Quelques exemples de distributions	7
6.1	Bernoulli	7
6.2	Binomiale	8
6.3	Geometrique	8
6.4	Poisson	8
6.5	Uniforme	8
6.6	Loi normale (Gaussienne)	8
6.7	Exponentielle	9
7	Vecteurs aléatoires	9
7.1	de nition	9
7.2	Esperance d'un vecteur aleatoire	9
8	Matrices des covariances	9
8.1	de nition	9
8.2	Proprietes:	10
9	Loi normale multidimensionnelle	10
9.1	De nition	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 Modele graphique

1.1 Qu'est ce qu'un modele graphique ?

Un modele graphique est un graphe qui peut être oriente ou non oriente dont les sommets representent des variables aleatoires et dont les arcs (les arrêtes dans le cas non oriente) representent les dependances de ces variables.

Le modele graphique est une representation d'objets probabilistes qui est notamment utilisee en apprentissage automatique.

Il existe deux familles de modeles graphiques, les reseaux bayesiens qu'on a vu dans le cours (les graphes orientes acycliques), et les champs aleatoires de Markov qui sont non orientes.

1.2 Definition du modele graphique (Reseaux bayesiens)

- Soit $\{X_i\}_{i=1:n}$ une famille de variables aleatoires.
- Soit $G = (V; A)$ un graphe oriente acyclique.
- Soit la fonction f de nit comme ci-dessous:

$$f : \{X_i\}_{i=1:n} \rightarrow \prod_{X_i \in V} P(X_i | \text{Pa}(X_i))$$

Ou $\text{Pa}(X_i)$ sont les parents de X_i dans le graphe G.

Alors la probabilite jointe des variables $\{X_i\}_{i=1:n}$ est egale:

$$P(X_1; X_2; \dots; X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i))$$

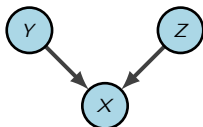
En d'autre termes chaque sommet est conditionnellement independant de ses non-descendants.

On remarque que cette formule de la probabilite jointe est en realite la chain rule

$P(X_1; \dots; X_n) = P(X_1)P(X_2|X_1) \dots P(X_n|X_1; \dots; X_{n-1})$ ou les variables independantes n'apparaissent pas. Notons que pour $n=3$, toute distribution de probabilite se factorise dans un modele graphique. Ce n'est pas le cas en general mais si c'est possible, un modele graphique permet de simplifier la loi jointe et de trouver les independances.

1.3 Exemple d'un reseaux bayesien

Voici un exemple d'un modele graphique:



On remarque que le graphe est oriente acyclique. Nous remarquons que Y et Z sont independants des autres variables aleatoires car (Y et Z n'ont pas de parents) et que X depend a la fois de Y et Z. Nous pouvons donc a partir de ce graphe exprimer la loi de probabilite jointe comme ci-dessous:

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

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

- $P_1(X|Y; Z) = \mathbb{1}_{X \text{ OR } (Y; Z)} N(1) + \mathbb{1}_{\text{NOT } (X \text{ OR } (Y; Z))} N(0; 1)$
- $P_2(Y = 1) = a$; 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 donnée 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:** $\int_z \int_y p(X; Y; Z) dy dz$

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

Dans notre cas les variables aléatoires 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 $E[X; Y] = \sum_{x \in \text{Val}(X)} x \cdot p_{X|Y}(x; y)$ 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 a 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:

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

On sait qu'un téléphone avec une batterie A tient 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:

$$E[T] = \sum_{X \in \{A, B\}} E[T|X] \cdot p_X(X) = E[T|A]p_X(A) + E[T|B]p_X(B) = 12 \times 0.8 + 8 \times 0.2 = 11.2$$

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 utilisée 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:

$$\text{Cov}(X; Y) = E[(X - E[X])(Y - E[Y])]$$

- 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.

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

3.4 Corrélation de Pearson

Il est possible de définir à partir de la covariance de deux variables aléatoires la corrélation de Pearson:

$$\text{Corr}(X; Y) = \frac{\text{Cov}(X; Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

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 $\text{Corr}(X; Y) = 0$ alors X et Y sont indépendantes.

3.5 Variance d'une somme de variables aleatoires

Soient $(X_i)_{i \in \{1, \dots, n\}}$ une famille de variables aleatoires. La variance de la somme de ces variables aleatoires est de nie comme ci-dessous:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i; X_j)$$

Proof. Montrons par recurrence que $\forall n \geq \mathbb{N}$:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i; X_j)$$

Pour $n=1$: $\text{Var}(X_1) = \text{Var}(X_1) + 0$

Soit $n \geq \mathbb{N}$ supposons que:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i; X_j)$$

Montrons pour $n+1$:

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^{n+1} X_i\right) &= \text{Var}\left(\sum_{i=1}^n X_i + X_{n+1}\right) \\ &= E\left[\left(\sum_{i=1}^n X_i + X_{n+1} - E\left[\sum_{i=1}^n X_i + X_{n+1}\right]\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right] + (X_{n+1} - E[X_{n+1}])\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right]\right)^2\right] + E\left[(X_{n+1} - E[X_{n+1}])^2\right] + 2E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right]\right)(X_{n+1} - E[X_{n+1}])\right] \\ &= \text{Var}\left(\sum_{i=1}^n X_i\right) + \text{Var}(X_{n+1}) + 2 \text{Cov}\left(\sum_{i=1}^n X_i; X_{n+1}\right) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i; X_j) + \text{Var}(X_{n+1}) + 2 \sum_{i=1}^n \text{Cov}(X_i; X_{n+1}) \\ &= \sum_{i=1}^{n+1} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n+1} \text{Cov}(X_i; X_j) \end{aligned}$$

On deduit que: $\forall n \geq \mathbb{N}$: $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i; X_j)$

□

4 Esperance et variance de la moyenne empirique

On definit la moyenne empirique comme suit: $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Dans cette section on cherche a savoir si la moyenne empirique est proche de la moyenne reelle.

4.1 L'esperance de la moyenne empirique

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

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

□

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 distribuées), donc leurs covariances vont être nulles.

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

□

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

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

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

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2$$

On applique ensuite la formule des espérances totales sur chaque terme.

$$\text{Var}(Y) = E[E[Y^2|X]] - (E[E[Y|X]])^2$$

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

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

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

En utilisant la linearite de l'esperance on trouve que:

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

On remarque que:

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

On remplace et on trouve:

$$Var(Y) = E[Var(YjX)] + Var(E[YjX])$$

□

5.2 Exemple

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

Solution:

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

$$Var(erreur) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m (e_{ij} - \hat{e}_i)^2 + \frac{1}{m} \sum_{i=1}^n \left[\frac{1}{m} \sum_{j=1}^m e_{ij} - \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n e_{kj} \right]^2$$

$$\text{Avec } \hat{e}_i = \frac{1}{m} \sum_{k=1}^m e_{i;k}$$

On peut aussi utiliser une notation plus condensee avec e_i : la moyenne des erreurs de classification du i eme classifieur:

$$Var(erreur) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{m} \sum_{j=1}^m [e_{ij} - e_i]^2 \right] + \frac{1}{n} \sum_{i=1}^n [e_i - \bar{e}]^2$$

6 Quelques exemples de distributions

6.1 Bernoulli

la loi de Bernoulli, designe la loi de probabilite d'une variable aleatoire discrete qui prend la valeur 1 avec la probabilite p et 0 avec la probabilite $q = 1 - p$.

$$\bullet \text{ Fonction de masse: } P(X = x) = \begin{cases} p & \text{si } x = 1; \\ 1 - p & \text{si } x = 0; \\ 0 & \text{sinon.} \end{cases}$$

• **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:** $P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}$;
- **Espérance:** np
- **Variance:** $np(1 - p)$

6.3 Géométrique

La loi géométrique désigne la 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:** $P(X = k) = q^{k-1} p$;
- **Espérance:** $\frac{1}{p}$
- **Variance:** $\frac{1-p}{p^2}$

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 fixe, si ces événements se produisent avec une fréquence moyenne ou espérance connue, et indépendamment du temps écoulé depuis l'événement précédent.

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

En règle générale, la loi de Poisson est principalement utilisée pour modéliser des événements 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:** $\frac{a+b}{2}$
- **Variance:**

6.6 Loi normale (Gaussienne)

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

6.7 Exponentielle

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

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

7 Vecteurs aléatoires

7.1 Définition

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

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

En ayant une fonction:

$$g : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$X \mapsto g(X)$$

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

7.2 Espérance d'un vecteur aléatoire

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

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

8 Matrices des covariances

8.1 Définition

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

$$\text{Cov}[X_i; X_j] = E[(X_i - E(X_i))(X_j - E(X_j))]$$

On a donc que:

$$\text{Cov}[X_i; X_i] = \text{Var}[X_i]$$

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

$$= \begin{pmatrix} \text{Cov}[X_1; X_1] & \text{Cov}[X_1; X_2] & \cdots & \text{Cov}[X_1; X_n] \\ \text{Cov}[X_2; X_1] & \text{Cov}[X_2; X_2] & \cdots & \text{Cov}[X_2; X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[X_n; X_1] & \text{Cov}[X_n; X_2] & \cdots & \text{Cov}[X_n; X_n] \end{pmatrix}$$

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 de rang positif.
- 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.

$$\Sigma = E[(X - E(X))(X - E(X))^T]$$

Remarque 5. Dans un cadre plus concret, on peut imaginer des variables aléatoires qui décrivent la couleur des pixels d'une image. Si on calcule 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: $X \sim \mathcal{N}(\mu; \Sigma)$ avec $X \in \mathbb{R}^n$ et Σ la matrice de covariance associée à X .

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

Par identification, si on prend $n=1$ (c'est-à-dire qu'on a une seule VA x) alors:

$$\begin{aligned} p(x; \mu; \sigma^2) &= \text{Cov}[x; x] = \text{Var}(x) = \sigma^2 \\ \frac{1}{\det(\Sigma)} &= \frac{1}{\sigma^2} \\ \Sigma &= \sigma^2 \end{aligned}$$

Et ainsi, pour $n=1$:

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

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 generalisation est tres utile pour modeliser des evenements aleatoires et est facile a manipuler notamment pour faire de l'inference dessus. De plus la loi gaussienne multivariee possede certaines proprietes remarquables. Parmi celles-ci on retrouve:

- Les lois marginales et conditionnelles d'une gaussienne multivariee sont gaussiennes.
- Une gaussienne multivariee de dimension d dont la matrice de covariance est diagonale est une collection de d lois normales independantes deux a deux.
- Les isocontours d'une gaussienne multivariee sont generalement des ellipsoïdes (on peut retrouver l'equation d'une ellipse dans la formule de la densite) 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 gures presentees ci-dessous sont des representation graphiques d'une distribution de probabilite suivant une loi gaussienne a deux variables centree en $(0,0)$ pour lesquelles on fait varier la matrice de covariance afin de se représenter graphiquement ce que ces modifications entraînent.

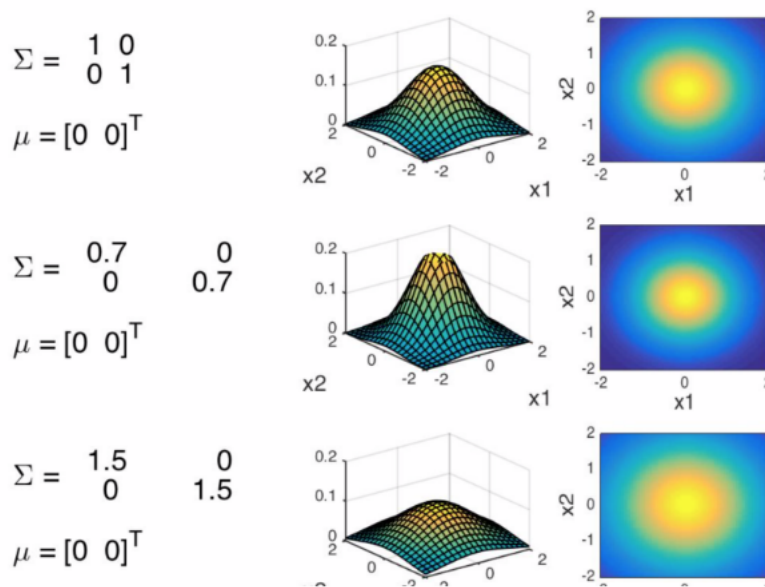


Figure 1: Variables independantes avec la même variance

Dans le cas ou la matrice de covariance est diagonale, on constate que des variables aleatoires independantes ayant la même variance font apparaître une distribution centree sur $(0,0)$ ou la densite maximale est atteinte. Autour de ce point la distribution est identique selon chaque direction. La valeur de densite maximale depend de la variance des variables aleatoires de maniere analogue a ce que l'on observe pour une loi normale sur une variable.

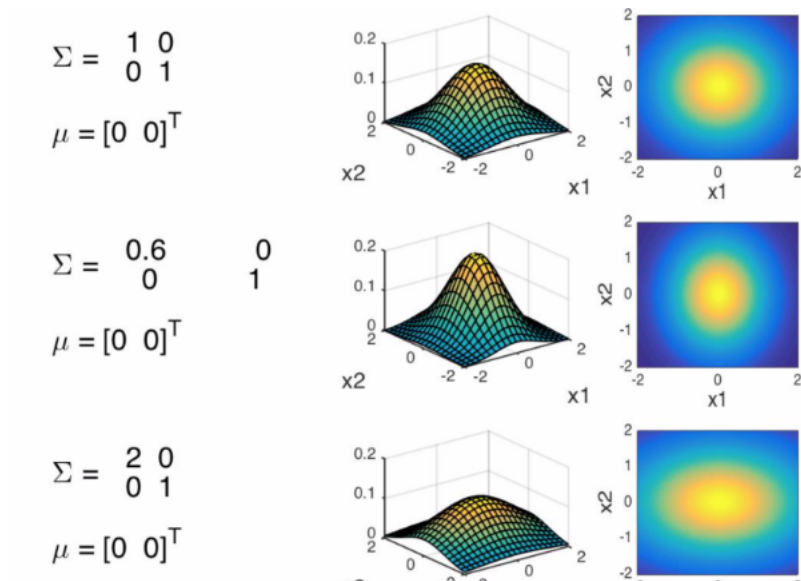


Figure 2: Variables independantes avec une variance di erentes

Si les variables aleatoires sont independantes mais que leur variances respectives sont di erentes, cela va etirer la distribution selon l'axe correspondant a la variables avec la variance la plus importante. Plus les variables aleatoires sont correlees, plus la distribution est etiree selon un axe qui correspond au vecteur propre de la matrice de covariance.

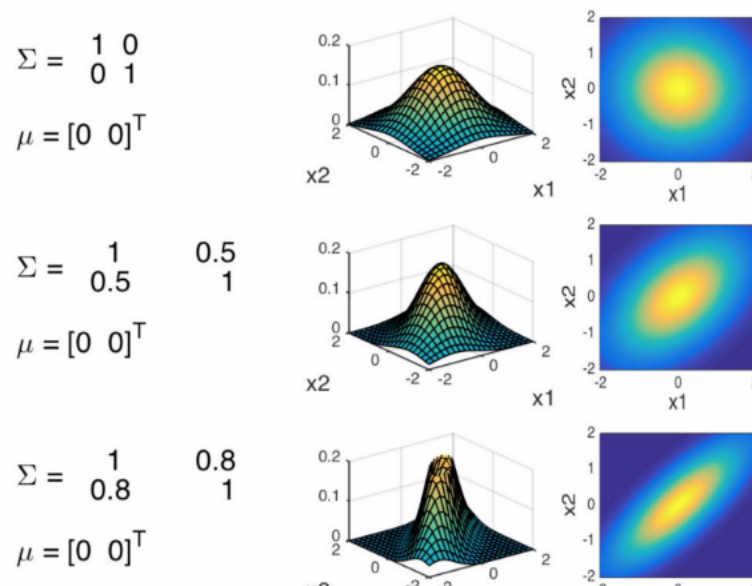


Figure 3: Variables correlees

Les variables aleatoires sont anti-correlees et la distribution est etiree selon l'axe orthogonal au precedent.

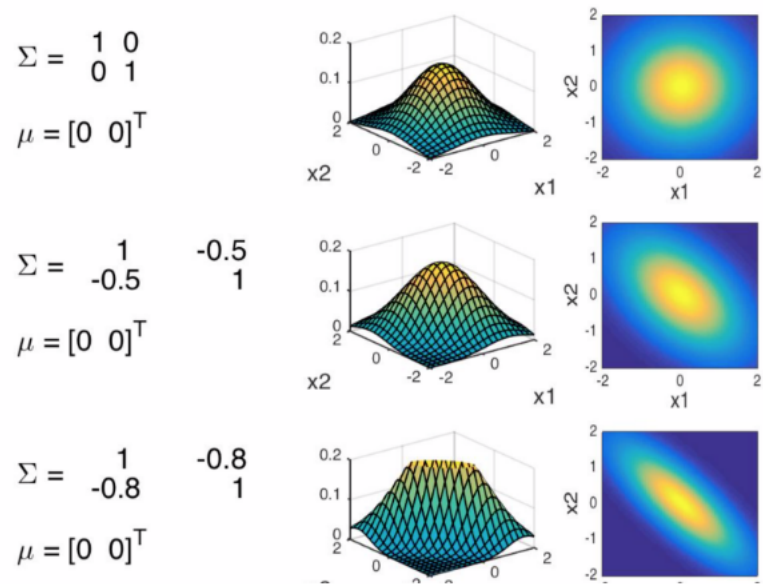


Figure 4: Variables anti-correlees

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

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