CM5 MIA

Baptiste Leroux, Nathan Chometton

20 September 2022

Contents

1 Revue du calcul probabiliste

1.1 Moments

En théorie des probabilités et en statistique, le moment d'ordre $\mathbf{r} \in \mathbf{N}$ d'une variable aléatoire réelle $\mathbf{X}$ est un indicateur de la dispersion de cette variable.

On définit le moment d'ordre k par $\mathbf{E}[\mathbf{X^k}]$ et le moment centré d'ordre k par $\mathbf{E}[(\mathbf{X} - \mu)^{\mathbf{k}}]$ avec $\mu = \mathbf{E}[\mathbf{X}]$.
On remarque que lorsque $\mathbf{k = 1}$ le moment centré d'ordre 1 est l'espérance et que quand $\mathbf{k = 2}$ il s'agit de la variance.

1.2 Loi Gaussienne Multivariée

Définition

On définit la loi gaussienne multivarié par $\mathbf{X} \sim \mathbf{N}(\mu, \mathbf{\Sigma}), \mathbf{X} \in \mathbf{R^n}$ par:

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

$\mu =$ la moyenne
$\Sigma =$ la variance
$\Sigma^{-}1 =$ matrice de covariance

<u>Preuve:</u>
Posons $(\mathbf{x} - \mu)\mathbf{P}(\mathbf{x} - \mu)$ avec $\mathbf{P} = \mathbf{\Sigma^{-1}}$
On sait que $\mathbf{P} = \mathbf{Q^T D Q}$ avec D:

Comme D est une matrice diagonale, on sait que:
$\mathbf{DD' = D'D = I}$ et donc que D est inversible $\implies$ D' = D^{-1} avec D':

$$\begin{bmatrix} 1/\lambda_i & 0 & 0 \\ 0 & .. & 0 \\ 0 & 0 & 1/\lambda_i \end{bmatrix} \forall i, \lambda_i > 0$$

Posons $\mathbf{P' = Q^T D^{-1} Q}$
$\mathbf{PP' = P'P = I} \implies$ P est inversible et $\mathbf{P^{-1} = P' = Q^T D^{-1} Q}$

$\mathbf{P} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \lambda_{\mathbf{i}} \mathbf{q_i q_i^T}$

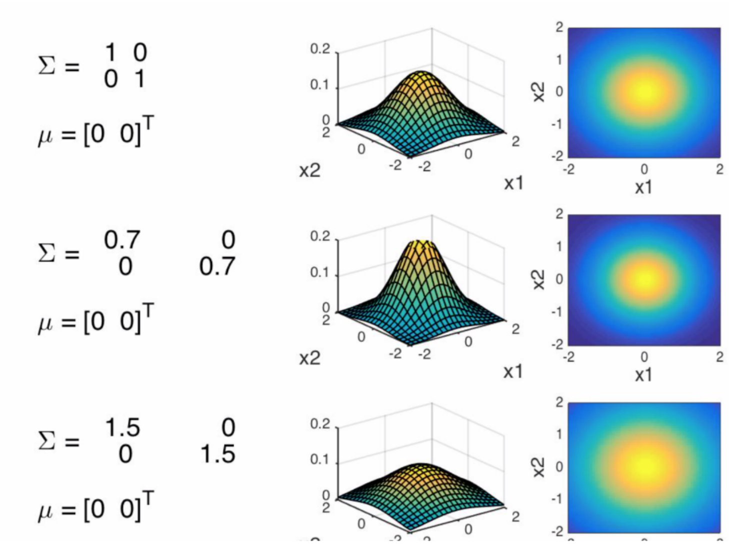
$\mathbf{v = kq_i} \implies$ Pv = k $\lambda_i q_i$
$\mathbf{v^T P v = k^2} \lambda_{\mathbf{i}}$

Si on prend $\mathbf{v} = \mu + \mathbf{x}$ alors on a:

$(\mathbf{x} - \mu)^{\mathbf{T}}\mathbf{P}(\mathbf{x} - \mu) = \lambda_{\mathbf{i}} \|\mathbf{v}\|_2^2$

$\frac{\mathbf{1}}{\mathbf{K}}\mathbf{exp}(\frac{\mathbf{-1}}{\mathbf{2}}\lambda_{\mathbf{i}}\|\mathbf{v}\|_2^2)$
Avec $\mathbf{K} = \mathbf{det}(\mathbf{\Sigma})^{\frac{1}{2}}(\mathbf{2}\pi)^{\frac{1}{2}}$

Visualisation de la variation de notre Gausienne en modifiant la variance



Définition générale

$\mathbf{X} \sim \mathbf{N}(\mu, \mathbf{\Sigma}) \Leftrightarrow \exists \mu \in \mathbf{R^k}, \mathbf{A} \in \mathbf{R^{k,l}} : \mathbf{X} = \mathbf{AZ} + \mu$ pour $\mathbf{Z_n} \sim \mathbf{N}(\mathbf{0}, \mathbf{1})$ i.d.d

Distribution conditionnelles

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N-q) \times 1 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N-q) \times 1 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & (N-q) \times (N-q) \end{bmatrix}$$

Exemple

$$\begin{array}{cc} & \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \end{array}$$

$$p(x_2, x_3) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$p(x_1, x_3) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Propriétés

- Les marginales et les conditionelles d'une Gausienne jointe sont Gausiennes.
- Une Gausienne à d-dimension $X \in N(\mu, \Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$, est équivalente à une décomposition de d Gausiennes **indépendantes** $X_i \in N(\mu_i, \sigma_i^2)$.

2 Théorie de la mesure

À quoi ça sert ? C'est un traitement rigoureux, unifié et général de la théorie des probabilités.

- Unifié: PMF et PDF deviennent deux cas particulier d'un concept général, pas de différence entre somme discrètes et intégrales continus

- Général: variable mixte (ni continues ni discrètes), variable plus compliqué (ex: signal)

Quels sont les difficultés ?

- Plus abstrait (intégral de Riemann et Lebesgue)
- Le problème de la mesurabilité des fonctions

2.1 Notation

Soit $(\mathbf{X}, \epsilon, \mu)$ un espace mesuré. Si f est une fonction mesurable, alors l'intégral de f par rapport à μ s'écrit:

$\int_X f d\mu$

2.2 Eléments de probabilité

Espace d'échantillons Ω :
{HH, HT, TH, TT}

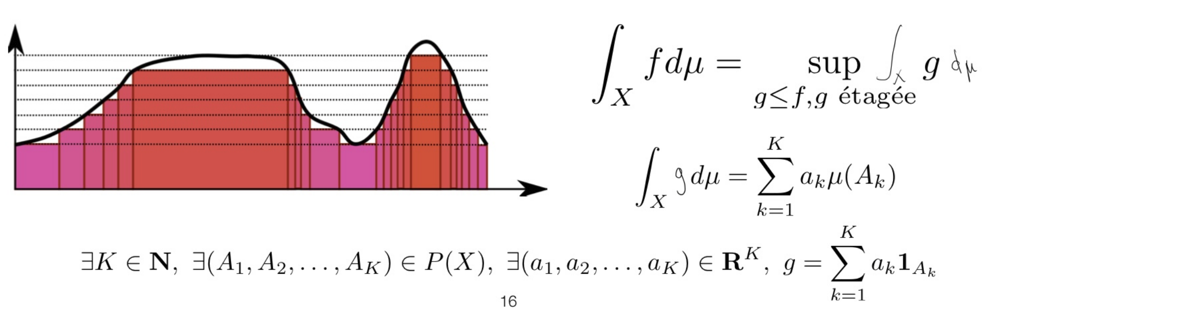
Un événement A $\sqsubseteq \Omega$:
HH, HT, Ω

Un espace d'événements F

Probabilité de la mesure P: F $\rightarrow R$
$P(A) \geq 0 \; \forall A \in F$
$P(\Omega) = 1$

Si A_1, A_2, ... sont des événements disjoints, $(A_i \cap A_j$ quand $i \neq j)$ alors
$P(\bigcup_i A_i) = \sum_i P(A_i)$

2.3 Espérance d'une variable aléatoire : intégrale de Lebesgue



3 Statistiques

3.1 Introduction

1. Problème de l'induction, inséparable de la démarche scientifique dans son ensemble

2. Bayésien VS fréquentiste

3. Modèle probabiliste VS Modèle non-probabiliste

Exemple 1
Soit $X_i = 0$ si une personne X qui ne pas mange de chocolat et $X_i = 1$ si elle en mange.
Soit $y_i = 0$ si une personne n'a pas le cancer et $y_i = 1$ si elle l'a.
$(X_1, X_2, X_3) = (0, 1, 0) \sim$ iid P
$(y_1, y_2, y_3) = (1, 1, 0) \sim$ iid Q
En réalité ici on ne peut pas tirer de conclusion de ces données. En effet ne conaissant le mode de vie des personnes de notre dataset, il y a bien d'autres facteurs qui pourraient induire un cancer. On pourrait par exemple faire entrer en jeu d'autre facteurs et pourquoi pas les pondérer.

Types de problèmes statistiques

- Modèle Statistique $P = N(\mu, \sigma^2)\ \ \mu \in R, \sigma > 0$

- Distribution P $\in P$

- Observation $O \sim P$

Estimation ponctuelle : $\hat{f}(O) \approx \hat{f}(P)$

3.2 Statistique non asymptotique: biais, variance et risque

Estimation ponctuelle

- biais: $b_P(\hat{f}) = E_{O\sim P}[\hat{f}(O)] - f(P) \qquad \left\| b_P(\hat{f}) \right\|_2$
  Estimateur non biaisé par ce modèle si le biais est nul quelque soit la distribution prise par ce modèle

- variance: $\mathrm{Var}_P(\hat{f}) = \mathrm{Var}_{O\sim P}[\hat{f}(O)]$

- risque: $\ell : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$

$$R_P(\hat{f}) = E_{O\sim P}[\ell(f(P), \hat{f}(O))]$$
$$f : P \to R^d$$
$$f : E \to R^d$$