Modélisation statistiques

aurore.lavigne@univ-lille.fr

Plan du cours

Notations et définitions

Soit $(\Omega, A, \mathbb{P})$ un espace de probabilité et p un entier positif.

Définition

On appelle vecteur aléatoire de dimension p (ou variable aléatoire p-dimensionnelle) le vecteur

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}, \text{ où les } X_i \text{ sont des v.a réelles.}$$

Les v.a . $X_i, i = 1, \ldots, p$ sont appelées les marginales de $\mathbf{X}$.

- $\mathbf{X}$ est donc une application définie par :

$$\begin{aligned} \mathbf{X} : \quad \Omega \quad & \longrightarrow \mathbb{R}^p \\ \omega \quad & \longrightarrow \mathbf{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix} \end{aligned}$$

- La loi de $\mathbf{X}$ est caractérisée par sa fonction de répartition (appelée aussi fonction de distribution) jointe définie par

$$F(\mathbf{x}) = \mathbb{P}\left(X_1 \leq x_1, \ldots, X_p \leq x_p\right),$$

pour tout $\mathbf{x} = (x_1, ..., x_p)^T$.

- On distingue essentiellement deux grands types vecteurs aléatoires :
  - **Vecteur aléatoire discret.** Si pour tout $i = 1, \ldots, p$, $\Omega(X_i)$ est un sous-ensemble fini ou dénombrable de $\mathbb{R}$ alors $\Omega(\mathbf{X})$ est un sous-ensemble fini ou dénombrable de $\mathbb{R}^p$. On dit alors que X est un vecteur aléatoire discret. Sa distribution est dite discrète.
  - **Vecteur aléatoire continu.** Si pour tout $i = 1, \ldots, p$, $\Omega(X_i)$ n'est pas un sous-ensemble fini ou dénombrable de $\mathbb{R}$. On dit alors que $\mathbf{X}$ est un vecteur aléatoire continu. Sa distribution est dite absolûment continue par rapport à la mesure de Lebesgue dans $\mathbb{R}^d$.

Le théorème de Radon-Nikodym

Théorème

Si $\mathbf{X}$ est un vecteur aléatoire continu, alors il existe une fonction $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ mesurable telle que

$$\forall B \in \mathcal{B},\ \mathbb{P}(\mathbf{X} \in B) = \int_B f(\mathbf{x})d\mathbf{x}.$$

On dit que f est la **fonction de densité de probabilité** (pdf) de $\mathbf{X}$ (ou densité jointe des composantes de $\mathbf{X}$).

Propriété.

$$f(\mathbf{x}) \geq 0 \text{ pour tout } \mathbf{x} \in \mathbb{R}^p \text{ et } \int_{\mathbb{R}^p} f(\mathbf{x})d\mathbf{x} = 1.$$

Dans le cas d'un vecteur aléatoire discret, les équivalents de la pdf sont les probabilités

$$p_i = \mathbb{P}(\mathbf{X} = \mathbf{x}_i)$$

où les $\mathbf{x}_i$ sont les valeurs possibles de $\mathbf{X}$, attention il s'agit de vecteurs.

	cas continu	cas discret
Valeurs possibles	$\mathbf{x}$	$\mathbf{x}_i$
loi de $\mathbf{X}$(pdf/probabilités)	$f(\mathbf{x})$	p_i

Espérance et moments d'un vecteur aléatoire

- L'espérance d'un vecteur aléatoire X est définie par

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_p) \end{pmatrix}.$$

- De manière général si $g : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ est une fonction mesurable, on a

$$\mathbb{E}(g(\mathbf{X})) = \begin{pmatrix} \mathbb{E}(g_1(\mathbf{X})) \\ \vdots \\ \mathbb{E}(g_q(\mathbf{X})) \end{pmatrix}.$$

- On suivra la même règle si g est à valeurs dans un espace de matrices.

Variance d'un vecteur aléatoire

La notion de variance peut s'étendre aux vecteurs aléatoires de la manière suivante :

Définition

Si $\mathbf{X}$ est un vecteur aléatoire de dimension p alors :

$$\mathbb{V}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^T\right] = (\sigma_{ij})_{1 \leq i,j \leq p},$$

avec

$$\begin{aligned} \sigma_{ij} = & \ \mathbb{C}ov(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right] \\ = & \ \mathbb{E}\left(X_i X_j\right) - \mathbb{E}(X_i)\mathbb{E}(X_j) \end{aligned}.$$

$\mathbb{V}(\mathbf{X})$ est connue sous le nom de matrice de variance-covariance de $\mathbf{X}$. Elle est souvent notée Σ_X ; nous adoptons cette notation.

Propriétés

1. Σ_X est une matrice symétrique définie positive.
2. Si A est une matrice $q \times p$ déterministe, alors

$$\mathbb{E}(A\mathbf{X}) = A\mathbb{E}(\mathbf{X}) \text{ et } \Sigma_{A\mathbf{X}} = A\Sigma_{\mathbf{X}}A^T.$$

Rappel

- On dit qu'une matrice M est définie positive si pour tout vecteur u, $u^T Mu > 0$.
- Ou bien de manière équivalente toutes les valeurs propres de M sont positives.
- D'autre part, une combinaison linéaire de matrices définies positives de même dimension est définie positive.
- Enfin, le théorème spectral entraîne que si M est symétrique (définie positive), alors il existe une matrice orthogonale P et une matrice diagonale D telle que $M = PDP^T$.
- On en déduit qu'il existe une matrice A, telle que $M = AA^T$; par exemple $A = PD^{1/2}$.

Preuve.

1. La matrice Σ_X est symétrique par construction. En effet, elle s'écrit comme l'espérance d'une matrice symétrique [1].
   Montrons que Σ_X est définie positive. Soit le vecteur $u = (u_1, \ldots, u_p)^T$. Alors, on a par définition de l'espérance :

$$\begin{aligned} u^T \Sigma_X u = & \ \mathbb{E}\left\{\left[u^T X - \mathbb{E}\left(u^T X\right)\right]\left[u^T X - \mathbb{E}\left(u^T X\right)\right]^T\right\} \\ = & \ \mathbb{V}\left(u^T X\right) = \mathbb{V}\left(\sum_{j=1}^p u_j X_j\right) \geq 0. \end{aligned}$$

2. Evident pour l'espérance, par définition. Pour la variance, on a :

$$\mathbb{V}(AX) = \mathbb{E}\left\{[AX - \mathbb{E}(AX)][AX - \mathbb{E}(AX)]^T\right\} = A\Sigma_X A^T.$$

1. le produit d'un vecteur et de sa transposée est toujours symétrique.

Remarque

Plus généralement, on peut définir la covariance de deux vecteurs aléatoires X et Y, de dimension p par

$$\mathbb{C}ov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^T\right]$$

Attention au fait que contrairement au cas de la dimension 1, si X et Y sont des vecteurs aléatoires on a en général :

$$\mathbb{C}ov(X,Y) \neq \mathbb{C}ov(Y,X).$$

Propriété

Si A et B sont des matrices déterministes alors

$$\mathbb{C}ov\left(AX, BX\right) = A\Sigma_X B^T.$$

- Dans ce cours, on traite uniquement le cas des vecteurs aléatoires continus.
- Ainsi, on considère le vecteur aléatoire

$$\mathbf{X} = (X_1, \ldots, X_p)^T$$

où les X_i sont des v.a. unidimensionnelles continues, avec des densités de probabilités respectives

$$f_1(x_1), \ldots, f_p(x_p),$$

et de fonctions de répartition respectives

$$F_1(x_1), \ldots, F_p(x_p).$$

et la densité de probabilité jointe des composantes de $\mathbf{X}$ est

$$f(x_1, \ldots, x_p)$$

.

- La loi de X est caractérisée par sa fonction de répartition (appelée aussi fonction de distribution) jointe

$$F(x_1, \ldots, x_p) = \mathbb{P}\left(X_1 \leq x_1, \ldots, X_p \leq x_p\right).$$

- On a

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_p} f(u_1, \ldots, u_p) du_1 \ldots du_p,$$

où $f(x_1, \ldots, x_p)$ est la fonction densité de probabilité jointe de $\mathbf{X}$.

- Si les $X_i, i = 1, \ldots, p$ sont des **v.a. indépendantes** alors

$$f(x_1, \ldots, x_p) = \prod_{i=1}^{p} f_i(x_i) \text{ et } F(x_1, \ldots, x_p) = \prod_{i=1}^{p} F_i(x_i).$$

- Réciproquement, si les factorisations précédentes sont vraies, alors les $X_i, i = 1, \ldots, p$ sont des v.a indépendantes.

- L'hypothèse d'indépendance des X_i permet donc de simplifier considérablement les calculs.
- Pour cette raison, plusieurs méthodes statistiques supposent toujours l'indépendance des variables aléatoires étudiées.
- Mais il faut faire attention au fait que l'hypothèse d'indépendance est parfois difficile à vérifier en pratique.

Définition

Soit X une variable aléatoire p-dimensionnelle. La densité jointe g d'une partie des composantes de X est obtenue en intégrant la densité jointe de X dans le domaine des variables qui ne sont pas dans la partie considérée. Ainsi, si l'on re-numérote les composantes de X par

$$X_1 \ldots, X_q, X_{q+1}, \ldots, X_p$$

alors on peut écrire

$$g\left(x_1, \ldots, x_q\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(x_1, \ldots, x_p\right) dx_{q+1} \ldots, dx_p.$$

- On obtient également la fonction de distribution jointe G associée par

$$\begin{aligned} G\left(x_1, \ldots, x_q\right) &= \mathbb{P}\left(X_1 \leq x_1, \ldots, X_q \leq x_q\right) \\ &= F\left(x_1, \ldots, x_q, +\infty, \ldots, +\infty\right). \end{aligned}$$

- En particulier la densité marginale d'une variable X_i s'écrit :

$$f_i(x_i) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_p) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p.$$

Exemple.

Soient Y_1 et Y_2 deux variables aléatoires réelles indépendantes et de même loi. On note par g et G la fonction de densité et la fonction de répartition de ces variables. On définit

$$X_1 = \min\{Y_1, Y_2\} \text{ et } X_2 = \max\{Y_1, Y_2\}.$$

1. La densité de $X = (X_1, X_2)$ est donnée par

$$f(x_1, x_2) = \begin{cases} 2g(x_1)g(x_2) & \text{si } x_1 \leq x_2 \\ 0 & \text{sinon} \end{cases}.$$

2. Les densités marginales de X_1 et X_2

Distribution conditionnelle

- En analyse multi-variée, il est parfois nécessaire de connaître la distribution d'un vecteur aléatoire conditionnellement à un autre vecteur aléatoire. On parle alors de loi conditionnelle.

Définition

La fonction de densité de probabilité conditionnelle de $X_1, \ldots, X_q$ sachant $X_{q+1} = x_{q+1}, \ldots, X_p = x_p$ est définie par

$$h\left(x_1, \ldots, x_q | x_{q+1} \ldots x_p\right) = \frac{f\left(x_1, \ldots, x_p\right)}{g\left(x_{q+1}, \ldots, x_p\right)},$$

où $f\left(x_1, \ldots, x_p\right)$ est la densité jointe de $\left(X_1, \ldots, X_p\right)$ et $g\left(x_{q+1}, \ldots, x_p\right)$ la densité jointe de $\left(X_{q+1}, \ldots, X_p\right)$.

- Lorsque les variables aléatoires $X_1, \ldots, X_p$ sont indépendantes alors la densité conditionnelle coïncide avec la densité jointe de $X_1, \ldots X_q$.
- A partir de la densité conditionnelle on, définit la fonction de répartition conditionnelle par

$$F\left(x_1, \ldots, x_q | x_{q+1} \ldots x_p\right) = \frac{\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_q} f\left(u_1, \ldots, u_q, x_{q+1}, \ldots, x_p\right) du_1 \ldots, du_q}{g\left(x_{q+1}, \ldots, x_p\right)}.$$

Vecteurs gaussiens

Définition

$\mathbf{X}$ est un vecteur Gaussien à p dimensions si toutes combinaisons linéaires de ses composantes suit une loi normale à une dimension.

Remarque

La normalité de chacune des composantes de $\mathbf{X}$ ne suffit pas à définir un vecteur gaussien.

Exemple :

$$\begin{pmatrix} X \\ -X \end{pmatrix} \text{ avec } X \sim \mathcal{N}(0, \sigma^2)$$

n'est pas un vecteur gaussien.

Densité de probabilité de $\mathbf{X}$

Théorème

Si Σ est régulière (déterminant non nul), $\mathbf{X}$ admet pour densité :

$$f(x_1, x_2, \cdots, x_p) = \frac{1}{(2\pi)^p (det(\Sigma))^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Remarque 1

La loi de $\mathbf{X}$ est entièrement déterminée par la connaissance de son espérance $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)^T$ et de sa matrice de variance covariance Σ. On notera

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Remarque 2 : linéarité

Si $\mathbf{X}$ est un vecteur gaussien de dimension p, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ et A une matrice à k lignes et p colonnes, et $\mathbf{b}$ est un vecteur de dimension k, le vecteur $A\mathbf{X} + b$ est gaussien de dimension k et

$$A\mathbf{X} + \mathbf{b} \sim \mathcal{N}(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$$

Remarque 3

Si les composantes de $\mathbf{X}$ sont décorrélées alors elles sont indépendantes. En effet, on remarque que si $\Sigma = \sigma^2 \mathbf{I}_p$ alors

$$f(x_1, x_2, \cdots, x_p) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_i)^2\right)$$

Propriété de l'espérance conditionnelle

L'espérance conditionnelle de $(Y, X_1, \ldots, X_p)^T$ sachant $X_1 = x_1, \ldots, X_p = x_p$ est une fonction affine de $x_1, \cdots, x_p$.
En particulier

$$E(Y|X_1, \cdots, X_p) = \sum_{i=1}^{p} a_i(X_i - E(X_i)) + E(Y)$$

avec

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = V(X)^{-1} \begin{pmatrix} cov(Y, X_1) \\ cov(Y, X_2) \\ \vdots \\ cov(Y, X_p) \end{pmatrix}$$

Loi du χ^2

Soit $X_1, \cdots, X_n$, n v. a. gaussiennes indépendantes, $X_i \sim \mathcal{N}(0,1)$ pour tout i, alors la variable aléatoire V donnée par $V = \sum_{i=1}^{n} X_i^2$ suit une loi du χ^2 à n degrés de libertés.

$$V = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

$$E(V) = n \qquad V(V) = 2n$$

Espaces euclidiens

On considère des sous espaces vectoriels de $\mathbb{R}^n$ munis du produit scalaire usuel. On rappelle que

- pour tout $u = (u_1, \cdots, u_n)$ et $v = (v_1, \cdots, v_n)$ de $\mathbb{R}^n$, $< u, v >= \sum_{i=1}^n u_i v_i$
- $||u|| = \sqrt{\sum_{i=1}^n u_i^2} =< u, u >$
- si u et v sont deux vecteur colonnes, $< u, v >= u^T v$,
- on dit que u est orthogonal à v si $< u, v >= 0$.

Sous espaces d'un espace vectoriel euclien

Deux sous-espaces orthogonaux

Soit E un espace vectoriel de $\mathbb{R}^n$ et F et G deux sous-espaces de E. F et G sont deux sous espaces orthogonaux ssi

$$\forall u \in F, \quad \forall v \in G, \quad < u, v >= 0$$

Supplémentaire orthogonal

Soit E un espace vectoriel de $\mathbb{R}^n$ et F un sous-espaces de E. Il existe un unique sous espace G de E tel que

- F et G sont orthogonaux
- F et G sont en somme directe
- $F \oplus G = E$

On dit que G est le supplémentaire orthogonal de F dans E, on le note $F^\perp$. C'est l'ensemble des éléments de E orthogonaux à tous les éléments de F.

Base orthormée de $\mathbb{R}^n$

Soit E un sous-espace vectoriel de $\mathbb{R}^n$. La famille de vecteur $(v_1, v_2, \cdots, v_n)$ est une base orthonormée de E ssi

- $(v_1, v_2, \cdots, v_n)$ est une base de E (famille libre et génératrice),
- pour tout couple (i, j) avec $i \neq j$, v_i et v_j sont orthogonaux,
- pour tout i, $||v_i|| = 1$.

- Il est toujours possible de trouver une base orthonormée d'un sous espace vectoriel de $\mathbb{R}^n$. On pourra par exemple, utiliser la méthode d'orthonormalisation de Gram-Schmidt.
- Soit u un vecteur de E, la décomposition de u dans la base orthonormée $(v_1, v_2, \cdots, v_n)$ est

$$u =< u, v_1 > v_1+ < u, v_2 > v_2 + \cdots + < u, v_n > v_n$$

Matrice orthogonale

Une matrice M est orthogonal si c'est la matrice de passage d'une base orthonormée à une autre base orthonormée.
Si M est une matrice orthogonale alors $M = M^T$.

Projecteur

Déf. 1

Soient E_1 et E_2 deux sous-espaces supplémentaires de $\mathbb{R}^n$. Soit $u \in \mathbb{R}^n$, la décomposition $u = u_1 + u_2$, avec $u_1 \in E_1$ et $u_2 \in E_2$ est unique. On appelle projecteur sur E_1 parallèlement à E_2, l'application linéaire p de $\mathbb{R}^n$ dans $\mathbb{R}^n$, qui à tout $u \in \mathbb{R}^n$ associe $p(u) = u_1$.

Déf. 2

On appelle projecteur, une application linéaire p de $\mathbb{R}^n$ dans $\mathbb{R}^n$ telle que $p \circ p = p$.

Les deux définitions sont équivalentes.

Projecteur orthogonal

Définition

Un projecteur p est un projecteur orthogonal, si $Im(p) \perp Ker(p)$.

Remarques

Soit u un vecteur de $\mathbb{R}^n$, alors $u - p(u) \in Ker(p)$ et est orthogonal à tout vecteur de $Im(p)$.

Proposition

Un projecteur, de matrice P dans une base orthonormée de $\mathbb{R}^n$, est un projecteur orthogonal si et seulement si sa matrice vérifie $P^2 = P$ et $P^T = P$

Propriété

La projection orthogonale sur $Im(p)$ minimise les distances : soient u et v deux vecteurs de $\mathbb{R}^n$, u fixé et $v \in Im(p)$, $||u - v||$ est minimale pour $v = p(u)$.

Théorème de Cochran

Soient $V_1, V_2, \cdots, V_k$, k sous-espaces vectoriels orthogonaux et supplémentaires de $\mathbb{R}^n$. On note $p_i = dim(V_i)$ et P_{V_i} une matrice de projection orthogonale sur V_i.
Soit $Z \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$.

1. $P_{V_i} Z \sim \mathcal{N}(0, \sigma^2 P_{V_i})$
2. $P_{V_1} Z, P_{V_2} Z, \cdots, P_{V_k} Z$ sont indépendants.
3. $\frac{||P_{V_i} Z||^2}{\sigma^2} \sim \chi^2_{p_i}$.

Application : loi de la variance empirique

Soient $X_1, X_2, \cdots, X_n$, n V.A. i.i.d., où $X_i \sim \mathcal{N}(0, \sigma^2)$. On note X le vecteur $(X_1, X_2, \cdots, X_n)^T$. Soit E la droite vectorielle engendrée par le vecteur $(1, 1, \cdots, 1)^T$.

1. Donner le projeté orthogonal de X dans E. On le note $P_E(X)$.
2. Donner le projeté orthogonal de X dans le suplémentaire orthogonal de E dans $\mathbb{R}^4$. On le note $P_{E^\perp}(X)$.
3. D'après le théorème de Cochran que peut on dire de $P_E(X)$ et $P_{E^\perp}(X)$.
4. Quelle relation existe-t-il entre la variance empirique $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ et $||P_{E^\perp}(X)||^2$?
5. En déduire la loi de S_n^2/σ^2.

Rappel sur la loi de Student

Soient $U \sim \mathcal{N}(0,1)$ et $V \sim \chi^2_n$, U et V indépendantes, alors la variable T définie par

$$T = \frac{U}{\sqrt{V/n}} \sim \mathcal{S}_n.$$

Rappel sur la loi de Fisher

Soient $V_1 \sim \chi^2_{n_1}$ et $V_2 \sim \chi^2_{n_2}$, V_1 et V_2 indépendantes, alors la variable F définie par

$$F = \frac{V_1/n_1}{V_2/n_2} \sim \mathcal{F}_{n_1,n_2}.$$