Statistique Mathématique 2 (MATH-F-309, Chapitre #1)

Thomas Verdebout

Université Libre de Bruxelles

2015/2016

Plan du cours

1. Vecteurs aléatoires.
2. Loi normale multivariée.
3. Inférence dans les modèles gaussiens.
4. Méthodes classiques de l'analyse multivariée.
5. Données directionelles.
6. Modèle linéaire.

Chapitre 1

Chapitre 1

$\boxed{\text{1. Vecteurs aléatoires.}}$

1.1. Définition, types de distributions.

1.2. Espérance et moments.

1.3. Indépendance.

1.4. Estimateurs usuels de μ et Σ.

1.5. Modes de convergence, résultats limites.

1.6. Fonction caractéristique.

1.7. Jacobiens.

Définition

Soit $(\Omega, \mathcal{A}, P)$ un espace de mesure.

Définition: *un p-vecteur aléatoire (ou p-v.a.) X est une fonction*

$$\begin{array}{rccl} X: & \Omega & \to & \mathbb{R}^p \\ & \omega & \mapsto & X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix} = (X_1(\omega), \ldots, X_p(\omega))' \end{array}$$

qui est mesurable, c'est-à-dire qui est telle que pour tout $B \in \mathcal{B}^p$, $X^{-1}(B) = \{\omega \in \Omega \,|\, X(\omega) \in B\} \in \mathcal{A}$.

Terminologie: les composantes X_i d'un v.a. $X = (X_1, \ldots, X_p)'$ sont appelées les marginales de X.

Les marginales sont elles-mêmes des v.a..

Définition

La condition de mesurabilité permet de considérer la distribution P^X de X, qui est la mesure de probabilité sur $(\mathbb{R}^p, \mathcal{B}^p)$ définie par

$$P^X[B] := P[X \in B] := P[X^{-1}(B)], \quad \forall B \in \mathcal{B}^p.$$

Le p-v.a. X induit donc une correspondance entre les espaces de mesure $(\Omega, \mathcal{A}, P)$ et $(\mathbb{R}^p, \mathcal{B}^p, P^X)$.

La fonction de répartition de X est définie par

$$F^X(x) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p],$$

pour tout $x = (x_1, \ldots, x_p)'$.

Types de distributions

On distingue essentiellement deux grands types de distributions P^X:

- les distributions absolument continues (par rapport à la mesure de Lebesgue m_p) et
- les distributions singulières.

Types de distributions

Définition: *P^X est absolument continue (par rapport à m_p) ssi $\forall B \in \mathcal{B}_p$ tel que $m_p(B) = 0$, on a $P^X[B] = 0$.*

Le théorème de Radon-Nikodym assure que, pour une telle distribution, il existe une fonction $f : \mathbb{R}^p \to \mathbb{R}$ mesurable (appelée fonction de densité de probabilité (pdf) de X) telle que

$$P^X[B] = \int_B f(x)\, dx, \quad \forall B \in \mathcal{B}^p.$$

Propriété d'une pdf: $f(x) \geq 0\ \forall x \in \mathbb{R}^p$ et $\int_{\mathbb{R}^p} f(x)\, dx = 1$.

distribution de X	
valeurs possibles	x
pdf	$f(x)$

Types de distributions

Définition: *P^X est singulière (par rapport à m_p) ssi $\exists N \in \mathcal{B}_p$ tel que $m_p(N) = 0$ et $P^X[N] = 1$.*

Une classe importante de distributions singulières est celle des distributions discrètes:

Définition: *P_X est discrète ssi il existe une collection au plus dénombrable $\{x_i, i \in \mathcal{I}\}$ de p-vecteurs telle que*

$$P^X[B] = \sum_{i \,|\, x_i \in B} P\left[X = x_i\right], \quad \forall B \in \mathcal{B}^p.$$

Une telle distribution est complétement déterminée par la donnée des x_i et des probabilités $p_i = P[X = x_i]$.

distribution de X			
valeurs possibles	x_1	x_2	$\dots$
probabilités	p_1	p_2	$\dots$

Plan du cours

$\boxed{\text{1. Vecteurs aléatoires.}}$

1.1. Définition, types de distributions.

$\boxed{\text{1.2. Espérance et moments.}}$

1.3. Indépendance.

1.4. Estimateurs usuels de μ et Σ.

1.5. Modes de convergence, résultats limites.

1.6. Fonction caractéristique.

1.7. Jacobiens.

Espérance et moments

Soit X un p-v.a.
Soit $g : \mathbb{R}^p \to \mathbb{R}$ une fonction mesurable.

Définition: *l'espérance de $g(X)$ est définie par*

$$\mathrm{E}[g(X)] = \int_\Omega g(X(\omega))\, dP(\omega) = \int_{\mathbb{R}^p} g(x)\, dP^X(x),$$

pour peu que la valeur commune de ces intégrales soit finie.

En pratique, on calculera ces espérances par

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}^p} g(x)\, f(x)\, dx \quad \text{et} \quad \mathrm{E}[g(X)] = \sum_{i \in \mathcal{I}} g(x_i) p_i$$

dans le cas absolument continu et le cas discret, respectivement.

Espérance et moments

Si $g : \mathbb{R}^p \to \mathbb{R}^q$ ($q \geq 2$), on prendra l'espérance composante par composante, c'est-à-dire

$$\mathrm{E}[g(X)] = \begin{pmatrix} \mathrm{E}[g_1(X)] \\ \vdots \\ \mathrm{E}[g_q(X)] \end{pmatrix}.$$

En particulier,

$$\mathrm{E}[X] = \begin{pmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_p] \end{pmatrix}.$$

Remarque: on suivra la même règle si g est à valeurs dans un espace de matrices.

Espérance et moments

Définition: *X a des moments finis d'ordre r ($r > 0$) ssi* $\mathrm{E}[\|X\|^r] < \infty$.

Proposition: *(i)* $\mathrm{E}[\|X\|^r] < \infty \Leftrightarrow$ *(ii)* $\mathrm{E}[|X_i|^r] < \infty$ *pour tout* $1 \leq i \leq p \Leftrightarrow$ *(iii)* $\mathrm{E}[|v'X|^r] < \infty$ *pour tout* $v \in \mathbb{R}^p$.

Preuve:

(i) $\Rightarrow$ (ii): cela suit de $|X_i|^r \leq \|X\|^r$.

(ii) $\Rightarrow$ (i): cela suit de $\|X\|^r \leq \left(\sum_{i=1}^p |X_i|\right)^r \leq C_{p,r}\left(\sum_{i=1}^p |X_i|^r\right)$.

(i) $\Rightarrow$ (iii): cela suit de $|v'X|^r \leq \|v\|^r \|X\|^r$.

(iii) $\Rightarrow$ (ii): il suffit de prendre $v = e_i$ (le $i$ème vecteur de la base canonique de $\mathbb{R}^p$).
$\square$

Remarque: en particulier, si X a des moments finis d'ordre 1, on peut parler de son espérance $\mu := \mathrm{E}[X] = (\mathrm{E}[X_1], \ldots, \mathrm{E}[X_p])'$, puisqu'alors $|\mathrm{E}[X_i]| \leq \mathrm{E}[|X_i|] < \infty$ pour

On veillera toujours dans la suite à imposer les hypothèses les moins fortes possibles en termes de moments.

Dans cette optique, le résultat suivant permet "d'ordonner" les hypothèses de moments finis.

Proposition: *(i) Si* $\mathrm{E}[\|X\|^r] < \infty$ *pour un certain* $r > 0$*, alors* $\mathrm{E}[\|X\|^s] < \infty$ *pour tout* $0 < s \leq r$*.*

Preuve: fixons $s \in (0, r)$. Comme la fonction

$$\begin{array}{rccc} f: & \mathbb{R}^+ & \to & \mathbb{R} \\ & t & \mapsto & t^s/(1+t^r) \end{array}$$

est bornée, on a $\mathrm{E}[\|X\|^s] \leq \mathrm{E}[C_{s,r}(1 + \|X\|^r)] = C_{s,r}(1 + \mathrm{E}[\|X\|^r]) < \infty$. $\square$

Espérance et moments

Soit X un p-v.a. avec des moments finis d'ordre 2.

Définition: *la matrice de variance-covariance de X est*

$$\Sigma = \mathrm{Var}[X] = \mathrm{E}[(X-\mu)(X-\mu)'].$$

Propriétés:

- $(\Sigma_{ij}) = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{Cov}[X_i, X_j]$ et $(\Sigma_{ii}) = \mathrm{Var}[X_i]$, ce qui justifie la terminologie.
- Σ est bien définie. En effet, $|\Sigma_{ij}| \leq |\Sigma_{ii}|^{1/2}|\Sigma_{jj}|^{1/2}$ (CS), où $|\Sigma_{ii}| = \mathrm{E}[(X_i)^2] + \mu_i^2 < \infty$.
- Σ est symétrique et semi-définie positive (pour tout $v \in \mathbb{R}^p$, $v'\mathrm{E}[(X-\mu)(X-\mu)']v = \mathrm{E}[v'(X-\mu)(X-\mu)'v] = \mathrm{E}[|v'(X-\mu)|^2] \geq 0$).
- $\Sigma = \mathrm{E}[XX'] - \mu\mathrm{E}[X'] - \mathrm{E}[X]\mu' + \mu\mu' = \mathrm{E}[XX'] - \mu\mu'$, qui est une expression plus commode pour le calcul de Σ.

Espérance et moments

Pour un X qui a des moments finis d'ordre 2, on peut donc considérer

$$\mu = \mathrm{E}[X] = \begin{pmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_p] \end{pmatrix}$$

et

$$\begin{aligned} \Sigma &= \mathrm{Var}[X] = \Big(\mathrm{Cov}[X_i, X_j]\Big)_{i,j=1,\ldots,p} \\ &= \begin{pmatrix} \mathrm{Var}[X_1] & \ldots & \mathrm{Cov}[X_1, X_p] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_1, X_p] & \ldots & \mathrm{Var}[X_p] \end{pmatrix}. \end{aligned}$$

Soient X un p-v.a. et Y un q-v.a., qui ont tous deux des moments finis d'ordre 2.

Définition: *la covariance entre X et Y est*

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)'].$$

Propriétés:

- $(\mathrm{Cov}[X, Y])_{ij} = \mathrm{Cov}[X_i, Y_j]$.
- Comme pour Σ, on vérifie facilement que $\mathrm{Cov}[X, Y]$ est bien définie.
- $\mathrm{Cov}[\alpha_1 X_1 + \alpha_2 X_2, Y] = \alpha_1 \mathrm{Cov}[X_1, Y] + \alpha_2 \mathrm{Cov}[X_2, Y]$ et $\mathrm{Cov}[X, \alpha_1 Y_1 + \alpha_2 Y_2] = \alpha_1 \mathrm{Cov}[X, Y_1] + \alpha_2 \mathrm{Cov}[X, Y_2]$.
- $\mathrm{Cov}[X, Y] = \mathrm{E}[XY'] - \mu_X \mu_Y'$.
- Si $p = q$, $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + \mathrm{Cov}[X, Y] + \mathrm{Cov}[Y, X]$.

Espérance et moments

Si on pose $Z = (X', Y')'$, on a

$$\mathrm{E}[Z] = \begin{pmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \end{pmatrix}$$

et

$$\mathrm{Var}[Z] = \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}[X, Y] \\ \mathrm{Cov}[Y, X] & \mathrm{Var}[Y] \end{pmatrix}.$$

Plan du cours

Indépendance

Soient X un p-v.a. et Y un q-v.a.

Définition: *X et Y sont indépendants ($X \perp\!\!\!\perp Y$) ssi*
$P[X \in B_1, Y \in B_2] = P[X \in B_1]P[Y \in B_2]$, $\forall B_1 \in \mathcal{B}^p$, $\forall B_2 \in \mathcal{B}^q$.

Remarque: si $X \perp\!\!\!\perp Y$, $g(X) \perp\!\!\!\perp h(Y)$ pour tout g, h. En particulier, si $X \perp\!\!\!\perp Y$, $X_i \perp\!\!\!\perp Y_j$ pour tout i, j.

Indépendance

Proposition: $X \perp\!\!\!\perp Y \Leftrightarrow$ *(i)* $F^{(X',Y')'}(x,y) = F^X(x)F^Y(y)\ \forall x,y$
$\Leftrightarrow$ *(ii)* $f^{(X',Y')'}(x,y) = f^X(x)f^Y(y)\ \forall x,y$ *(dans le cas abs^t continu)*
$\Leftrightarrow$ *(iii)* $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]\ \forall g,h$ *à valeurs réelles.*

Remarque: si $X \perp\!\!\!\perp Y$, on a $X - \mu_X \perp\!\!\!\perp Y - \mu_Y$, et donc $(X - \mu_X)_i \perp\!\!\!\perp (Y - \mu_Y)_j$ pour tout i, j. La proposition ci-dessus implique donc que
$(\mathrm{Cov}[X, Y])_{ij} = \mathrm{E}[(X - \mu_X)_i(Y - \mu_Y)_j] = \mathrm{E}[(X - \mu_X)_i]\mathrm{E}[(Y - \mu_Y)_j] = 0$, de sorte que $\mathrm{Cov}[X, Y] = 0$.

Espérance et moments

En particulier, si on pose $Z = (X', Y')'$,

$$\mathrm{Var}[Z] = \begin{pmatrix} \mathrm{Var}[X] & 0 \\ 0 & \mathrm{Var}[Y] \end{pmatrix},$$

et si $p = q$, $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Par contre, $\mathrm{Cov}[X, Y] = 0$ n'implique pas que $X \perp\!\!\!\perp Y$ (exemple aux TP).

Extension à plus de deux v.a.:

Définitions: *(i)* $X_1, \ldots, X_n \perp\!\!\!\perp$ *ssi* $\forall B_1 \in \mathcal{B}^{p_1}, \ldots, \forall B_n \in \mathcal{B}^{p_n}$, $P[X_1 \in B_1, \ldots, X_n \in B_n] = P[X_1 \in B_1] \ldots P[X_n \in B_n]$.

(ii) $X_1, X_2, \ldots \perp\!\!\!\perp$ *ssi* $\forall k \; \forall i_1 < i_2 < \ldots < i_k$, $X_{i_1}, \ldots, X_{i_k} \perp\!\!\!\perp$.

Plan du cours

Estimateurs

Soient $X_1, \ldots, X_n$ des p-v.a. i.i.d., avec des moments finis d'ordre 1.
$\rightsquigarrow$ on peut estimer $\mu = \mathrm{E}[X]$ par

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Quels sont les propriétés de $\bar{X}$?

Proposition: *si les X_i ont des moments finis d'ordre 2,*
(i) $\mathrm{E}[\bar{X}] = \mu$ *et (ii)* $\mathrm{Var}[\bar{X}] = \frac{1}{n}\Sigma$.

Preuve:

(i) $\mathrm{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[X_i] = \mu$.
(ii) $\mathrm{Var}[\bar{X}] = \frac{1}{n^2}\mathrm{Var}[\sum_{i=1}^{n} X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n^2} n\Sigma = \frac{1}{n}\Sigma$. $\square$

Estimateurs

Soient $X_1, \ldots, X_n$ des p-v.a. i.i.d., avec des moments finis d'ordre 2.
$\rightsquigarrow$ on peut estimer $\Sigma = \mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])']$ par

$$\hat{\Sigma} = S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'.$$

Comme $\bar{X}$, cet estimateur est sans biais:

Proposition: *si les X_i ont des moments finis d'ordre 2, $\mathrm{E}[S] = \Sigma$.*

Preuve: Soit $W = (n-1)S$. On a $W = \left(\sum_i X_i X_i'\right) - n\bar{X}\bar{X}'$, de sorte que
$\mathrm{E}[W] = \left(\sum_i \mathrm{E}[X_i X_i']\right) - n\mathrm{E}[\bar{X}\bar{X}'] = n\mathrm{E}[X_1 X_1'] - n\mathrm{E}[\bar{X}\bar{X}']$.
Or $\Sigma = \mathrm{E}[X_1 X_1'] - \mu\mu'$ et $\frac{1}{n}\Sigma = \mathrm{Var}[\bar{X}] = \mathrm{E}[\bar{X}\bar{X}'] - \mu\mu'$.
Donc $\mathrm{E}[W] = n(\Sigma + \mu\mu') - n(\frac{1}{n}\Sigma + \mu\mu') = (n-1)\Sigma$. $\square$

Estimateurs

Soit X un p-v.a.
Soient A une matrice constante $m \times p$ et $b \in \mathbb{R}^m$.

Clairement, $\mathrm{E}[AX + b] = A\mathrm{E}[X] + b \quad \text{et} \quad \mathrm{Var}[AX + b] = A\mathrm{Var}[X]A'$.

Définitions:
(i) Un estimateur $T(X_1, \ldots, X_n)$ de $\mu = \mathrm{E}[X]$ est affine-équivariant ssi $T(AX_1 + b, \ldots, AX_n + b) = AT(X_1, \ldots, X_n) + b$, $\forall A, b$.
(ii) Un estimateur $S(X_1, \ldots, X_n)$ de $\Sigma = \mathrm{Var}[X]$ est affine-équivariant

ssi $S(AX_1 + b, \ldots, AX_n + b) = AS(X_1, \ldots, X_n)A'$, $\forall A, b$.

Proposition: *$\bar{X}$ et S sont des estimateurs affine-équivariants de μ et Σ, respectivement.*

Preuve: exercice. $\square$

Estimateurs

Ce qu'on attend avant tout d'un estimateur $\hat{\theta}_n$ de θ est qu'il soit proche de θ.

Il est en fait désirable qu'il soit arbitrairement proche de θ si la taille d'échantillon n est suffisamment grande, c'est-à-dire que

$$\hat{\theta}_n \to \theta,$$

si $n \to \infty$.

Il nous faut d'abord donner un sens précis à cette convergence...

Plan du cours

Modes de convergence, résultats limites

Soit $(\Omega, \mathcal{A}, P)$ un espace de mesure.
Soit (X_n) une suite de p-v.a. et X un p-v.a. définis sur $(\Omega, \mathcal{A}, P)$.

Pour rappel, dans le cas univarié ($p = 1$), on peut définir de diverses manières la convergence de (X_n) vers X:

$X_n \overset{p.s.}{\to} X$ (presque sûrement) $\Leftrightarrow P[\{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega)\}] = 1$.

$X_n \overset{P}{\to} X$ (en probabilité) $\Leftrightarrow$ Pour tout $\varepsilon > 0$, $P[|X_n - X| > \varepsilon] \to 0$.

$X_n \overset{L^r}{\to} X$ (en norme L^r, $r > 0$) $\Leftrightarrow \mathrm{E}[|X_n - X|^r] \to 0$.

$X_n \overset{\mathcal{D}}{\to} X$ en loi $\Leftrightarrow F^{X_n}(x) \to F^X(x)$ pour tout $x \in \mathbb{R}$ en lequel F^X est continue.

Ces concepts s'étendent aisément au cas multivarié:

$X_n \stackrel{p.s.}{\to} X$ (presque sûrement) $\Leftrightarrow P[\{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega)\}] = 1$.

$X_n \stackrel{P}{\to} X$ (en probabilité) $\Leftrightarrow$ Pour tout $\varepsilon > 0$, $P[\|X_n - X\| > \varepsilon] \to 0$.

$X_n \stackrel{L^r}{\to} X$ (en norme L^r, $r > 0$) $\Leftrightarrow \mathrm{E}[\|X_n - X\|^r] \to 0$.

$X_n \stackrel{\mathcal{D}}{\to} X$ en loi $\Leftrightarrow F^{X_n}(x) \to F^X(x)$ pour tout $x \in \mathbb{R}^p$ en lequel F^X est continue.

Modes de convergence, résultats limites

Les divers liens entre ces concepts de convergence sont maintenus dans le cas général:

$$\begin{array}{ccccc} X_n \stackrel{p.s.}{\to} X & \Rightarrow & X_n \stackrel{P}{\to} X & \Leftarrow & X_n \stackrel{L^r}{\to} X \\ & & \Downarrow & & \\ & & X_n \stackrel{\mathcal{D}}{\to} X & & \end{array}$$

Remarques:

- $X_n \stackrel{P}{\to} X \Rightarrow$ il existe une sous-suite (X_{n_k}) telle que $X_{n_k} \stackrel{p.s.}{\to} X$.
- $X_n \stackrel{\mathcal{D}}{\to} a$, a constant $\Rightarrow X_n \stackrel{P}{\to} a$.
- $X_n \stackrel{P}{\to} X$ et $\lim_{K\to\infty} \sup_{n\geq 1} E|X_n| I\{|X_n| \geq K\} = 0 \Leftrightarrow X_n \stackrel{L^1}{\to} X$.

Le résultat suivant montre que la convergence p.s. composante par composante implique la convergence p.s. multivariée (et qu'il en va de même pour la convergence en probabilité)...

Proposition: *soient* (X_n) *une suite de p-v.a. et* X *un p-v.a. Alors*
(i) $(X_n)_i \stackrel{p.s.}{\to} X_i \ \forall i = 1, \ldots, p \Rightarrow (X_n) \stackrel{p.s.}{\to} X$;
(ii) $(X_n)_i \stackrel{P}{\to} X_i \ \forall i = 1, \ldots, p \Rightarrow (X_n) \stackrel{P}{\to} X$.

Preuve:

(i) Soient $A := \{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega)\}$ et $A_i := \{\omega \in \Omega \,|\, \big(X_n(\omega)\big)_i \to X_i(\omega)\}$, $i = 1, \ldots, p$. Alors $P[A^c] = P[\cup_{i=1}^p A_i^c] \leq \sum_{i=1}^p P[A_i^c] = 0$, de sorte que $P[A] = 1 - P[A^c] = 1$.

(ii) $P[\|X_n - X\| > \varepsilon] = P[\sum_{i=1}^p |(X_n)_i - X_i|^2 > \varepsilon^2] \leq P[\cup_{i=1}^p [|(X_n)_i - X_i|^2 > \varepsilon^2/p]] \leq \sum_{i=1}^p P[|(X_n)_i - X_i|^2 > \varepsilon^2/p] = \sum_{i=1}^p P[|(X_n)_i - X_i| > \varepsilon/\sqrt{p}] \to 0$, si $n \to \infty$. $\square$

Modes de convergence, résultats limites

Il en découle directement que la loi des grands nombres est également valide dans le cas multivarié. Plus précisément:

Corollaire: *soient* $X_1, X_2, \ldots$ *des p-v.a. i.i.d., avec des moments finis d'ordre 1. Notons* $\mu = \mathrm{E}[X_1]$ *et* $\bar{X}^{(n)} := \frac{1}{n}\sum_{i=1}^n X_i$*. Alors*
(i) $(\bar{X}^{(n)}) \stackrel{p.s.}{\to} \mu$ *(loi forte);*
(ii) $(\bar{X}^{(n)}) \stackrel{P}{\to} \mu$ *(loi faible).*

Remarques:

- La moyenne empirique est donc un estimateur fortement convergent pour μ.
- On montre de la même manière que, si $X_1, X_2, \ldots$ sont i.i.d., avec des moments finis d'ordre 2, $S = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ est un estimateur fortement convergent pour Σ.

Modes de convergence, résultats limites

A l'inverse des convergences p.s. et en proba, il n'est pas suffisant de considérer la convergence en loi des marginales.

Plus précisément, il ne suffit pas d'établir que $X_n \xrightarrow{\mathcal{D}} X$ et que $Y_n \xrightarrow{\mathcal{D}} Y$ pour en déduire que

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Par contre, on a le fameux lemme de Slutzky :

Lemme: *Si* $X_n \xrightarrow{\mathcal{D}} X$ *et* $Y_n \xrightarrow{\mathcal{D}} a$ *(a constant), alors* $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} X \\ a \end{pmatrix}$.

Quelques résultats supplémentaires sur la convergence en loi:

Proposition: *(i)* $X_n \stackrel{\mathcal{D}}{\to} X$ *et* $X_n \stackrel{\mathcal{D}}{\to} Y \Rightarrow X \stackrel{\mathcal{D}}{=} Y$ *(c'est-à-dire* $P^X = P^Y$*). (ii) Soit* $g : \mathbb{R}^p \to \mathbb{R}^q$ *une fonction telle que* $P[X \in C_g] = 1$*, où* $C_g = \{\textit{points de continuité de } g\}$*. Alors* $X_n \stackrel{\mathcal{D}}{\to} X \Rightarrow g(X_n) \stackrel{\mathcal{D}}{\to} g(X)$*.*

Il découle directement de cette proposition et du lemme de Slutzky le résultat suivant:

Corollaire: *Si* $X_n \stackrel{\mathcal{D}}{\to} X$ *et* $Y_n \stackrel{\mathcal{D}}{\to} a$ *(*$a \in \mathbb{R}^s$ *constant) et si* $g : \mathbb{R}^{p+s} \to \mathbb{R}^q$ *est continue, alors* $g(X_n, Y_n) \stackrel{\mathcal{D}}{\to} g(X, a)$*.*

En particulier, si les dimensions coïncident, $X_n \stackrel{\mathcal{D}}{\to} X$ et $Y_n \stackrel{\mathcal{D}}{\to} a$ impliquent que $X_n + Y_n \stackrel{\mathcal{D}}{\to} X + a$, $X_n' Y_n \stackrel{\mathcal{D}}{\to} X' a$, etc.

Le résultat suivant est connu sous le nom de “méthode delta":

Proposition: *Soit* (X_n) *une suite de p-v.a. telle que* $\nu_n(X_n - a) \stackrel{\mathcal{D}}{\to} Z$*, où* $\nu_n \to \infty$*.*
Soit $g : \mathbb{R}^p \to \mathbb{R}^q$ *une fonction différentiable en a. Alors*
$\nu_n(g(X_n) - g(a)) \stackrel{\mathcal{D}}{\to} Dg(a)\, Z$*, où* $Dg(a) = (\frac{\partial g_i}{\partial x_j}(a))$*.*

Preuve: le lemme de Slutzky implique que

$$X_n - a = \nu_n^{-1}[\nu_n(X_n - a)] \stackrel{\mathcal{D}}{\to} 0.$$

D'autre part, l'hypothèse de différentiabilité permet d'écrire

$$g(t) - g(a) = Dg(a)(t-a) + k(t)\,\|t-a\|,$$

où la fonction $k(.)$ définie par

$$k(t) = [g(t) - g(a) - Dg(a)(t-a)]/\|t-a\|\; I_{[t\neq a]}$$

est continue en $t = a$. On a donc que $k(X_n) \stackrel{\mathcal{D}}{\to} k(a) = 0$. Une nouvelle application du lemme de Slutzky permet de conclure que

$$\begin{aligned} \nu_n(g(X_n) - g(a)) &= Dg(a)\nu_n(X_n - a) + k(X_n)\,\|\nu_n(X_n - a)\| \\ &\stackrel{\mathcal{D}}{\to} Dg(a)Z + 0 \times \|Z\| = Dg(a)Z. \end{aligned}$$

$\square$

Plan du cours

Fonction caractéristique

Nous définissons maintenant un outil qui, comme nous le verrons, est très puissant pour établir des résultats distributionnels.

Définition: la fonction caractéristique du p-v.a. X est la fonction

$$\begin{array}{rccl} \varphi_X : & \mathbb{R}^p & \to & \mathbb{C} \\ & t & \mapsto & \mathrm{E}[e^{it'X}]. \end{array}$$

Remarques:

- φ_X est à valeurs dans le disque unité de $\mathbb{C}$ (puisque $|\varphi_X(t)| \leq \mathrm{E}[|e^{it'X}|] = \mathrm{E}[1] = 1$).
- $\varphi_X(0) = 1$; $\overline{\varphi_X(t)} = \varphi_X(-t)$.
- φ_X est uniformément continue (ce qui découle de l'inégalité $|\varphi_X(s) - \varphi_X(t)| \leq \mathrm{E}[|e^{i(s-t)'X} - 1|]$).

Fonction caractéristique

Si la loi de X est absolument continue (de densité f^X),

$$\varphi_X(t) = \mathrm{E}[e^{it'X}] = \int_{\mathbb{R}^p} e^{it'x} f^X(x)\, dx,$$

de sorte que φ_X est la transformée de Fourier de f^X.

L'existence de la célèbre formule d'inversion

$$f^X(x) = \left(\frac{1}{2\pi}\right)^p \int_{\mathbb{R}^p} e^{-ix'y} \varphi_X(y)\, dy$$

justifie le qualificatif "caractéristique":

Proposition:
(i) $X \stackrel{\mathcal{D}}{=} Y \Leftrightarrow \varphi_X(t) = \varphi_Y(t)\ \forall t \in \mathbb{R}^p$.
(ii) $(X_n) \stackrel{\mathcal{D}}{\to} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi_X(t)\ \forall t \in \mathbb{R}^p$.

Fonction caractéristique

Exemple: si $X \sim \mathcal{N}(0,1)$,

$$\varphi_X(t) = \mathrm{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} \left[\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right] dx = \ldots = e^{-t^2/2}.$$

Proposition: *Soient $A \in \mathbb{R}^{q \times p}$, $b \in \mathbb{R}^q$. Alors $\varphi_{AX+b}(t) = e^{it'b} \varphi_X(A't)$ $\forall t \in \mathbb{R}^q$.*

Preuve: $\varphi_{AX+b}(t) = \mathrm{E}[e^{it'(AX+b)}] = \mathrm{E}[e^{it'b} e^{i(A't)'X}] = e^{it'b} \varphi_X(A't)$. $\square$

On en déduit que, si $X \sim \mathcal{N}(\mu, \sigma^2)$ ($X \stackrel{\mathcal{D}}{=} \sigma Z + \mu$, où $Z \sim \mathcal{N}(0,1)$),
$\varphi_X(t) = e^{it\mu} \varphi_Z(\sigma t) = e^{it\mu - (\sigma t)^2/2}$.

Remarque: si $\mu = 0$, φ_X est à valeurs réelles. Ceci est en fait une illustration du résultat suivant:

Proposition: $\varphi_X : \mathbb{R}^p \to \mathbb{R} \Leftrightarrow X \stackrel{\mathcal{D}}{=} -X$ *(exercice).*

Fonction caractéristique

Proposition: *soient X, Y des p-v.a. indépendants. Alors $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ $\forall t \in \mathbb{R}^p$.*

Preuve: $\varphi_{X+Y}(t) = \mathrm{E}[e^{it'(X+Y)}] = \mathrm{E}[e^{it'X}e^{it'Y}] = \mathrm{E}[e^{it'X}]\mathrm{E}[e^{it'Y}] = \varphi_X(t)\varphi_Y(t)$. $\square$

Proposition: *pour autant que les espérances ci-dessous existent,*

$$\mathrm{E}[X_1^{k_1} \dots X_p^{k_p}] = \frac{1}{i^{k_1+\dots+k_p}} \frac{\partial^{k_1+\dots+k_p}}{\partial_{t_1}^{k_1} \dots \partial_{t_p}^{k_p}} \varphi_X(t)\Big|_{t=0}$$

Preuve: exercice.

Fonction caractéristique

Théorème (Cramer-Wold): $X \stackrel{\mathcal{D}}{=} Y \Leftrightarrow u'X \stackrel{\mathcal{D}}{=} u'Y \ \forall u \in \mathcal{S}^{p-1}$.

Preuve:

($\Rightarrow$) Pour tout $t \in \mathbb{R}$, $\varphi_{u'X}(t) = \varphi_X(tu) = \varphi_Y(tu) = \varphi_{u'Y}(t)$.

($\Leftarrow$) Soit $t \in \mathbb{R}^p \backslash \{0\}$. Posons $u = t/\|t\|$. Alors $\varphi_X(t) = \varphi_X(\|t\|u) = \varphi_{u'X}(\|t\|) = \varphi_{u'Y}(\|t\|) = \varphi_Y(\|t\|u) = \varphi_Y(t)$. Ceci permet de conclure, puisque $\varphi_X(0) = 1 = \varphi_Y(0)$.

$\square$

Plan du cours

$\boxed{\text{1. Vecteurs aléatoires.}}$

1.1. Définition, types de distributions.

1.2. Espérance et moments.

1.3. Indépendance.

1.4. Estimateurs usuels de μ et Σ.

1.5. Modes de convergence, résultats limites.

1.6. Fonction caractéristique.

$\boxed{\text{1.7. Jacobiens.}}$

Jacobiens

Le résultat suivant découle facilement de la formule de changement de variables dans les intégrales multiples.

Proposition: *soit X un p-v.a. de loi absolument continue (de densité f^X). Soit $Y = \phi(X)$, où ϕ est un difféomorphisme. Alors Y est également de loi absolument continue et sa densité est donnée par $f^Y(y) = f^X(\phi^{-1}(y)) \left|D\phi^{-1}(y)\right|, \forall y \in \mathbb{R}^p$ (où $|Dg(y)|$ désigne la valeur absolue du jacobien de g en y).*

Preuve: soit $B \in \mathcal{B}^p$.

$$P[Y \in B] = P[\phi(X) \in B] = P[X \in \phi^{-1}(B)] = \int_{\phi^{-1}(B)} f^X(x)\, dx,$$

ce qui, en posant $x = \phi^{-1}(y)$, livre

$$P[Y \in B] = \int_B f^X(\phi^{-1}(y)) \left|D\phi^{-1}(y)\right| dy. \qquad \square$$