TD n° 0

RAPPELS EXPRESS (et assez informels) DE PROBABILITES, avec exercices

X. Bry

Espérance

■ **Espérance**:

Soit un espace probabilisé (Ω, *P*). Pour toute variable aléatoire réelle *X* définie sur cet espace et intégrable, i.e. telle que: $\int_{\Omega} |X(\omega)| dP(\omega) < \infty$, on définit l'espérance comme:

$$E(X) = \int_{\Omega} X(\omega) dP(\omega)$$

Soit $X\ v.a. \in \mathbb{R}$. La fonction de répartition de *X* est:

$$F^X(x) = P(\{\omega ; X(\omega) \leq x\})$$

$$\forall X\ v.a. \in \mathbb{R} : \quad E(X) = \int_{\mathbb{R}} x P(\{\omega, X(\omega) \in]x ; x+dx]\})$$

$$= \int_{\mathbb{R}} x P(X \in]x ; x+dx]) = \int_{\mathbb{R}} x P^X(dx) = \int_{\mathbb{R}} x\, dP^X(x)$$

$$= \int_{\mathbb{R}} x (F^X(x+dx) - F^X(x)) = \int_{\mathbb{R}} x f^X(x) dx$$

où *f* est la densité par rapport à la mesure de Lebesgue.

Ces écritures s'étendent sans problème au cas multidimensionnel $X\ v.a. \in \mathbb{R}^p$.

■ **Théorème de transfert (très informel)**:

Soient $g : \mathbb{R}^p \to \mathbb{R}^q$, $Y = g(X)$.

On a:

$$E(Y) = \int_{\Omega} Y(\omega) dP(\omega) = \int_{\mathbb{R}} y P(\{\omega, Y(\omega) \in]y ; y+dy]\}) = \int_{\mathbb{R}} y\, dP^Y(y)$$

Mais aussi:

$$= \int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) P(\{\omega, X(\omega) \in]x ; x+dx]\})$$

$$= \int_{\mathbb{R}} g(x) P(X \in]x ; x+dx]) = \int_{\mathbb{R}} g(x) dP^X(x)$$

Autrement dit:

$$E(Y) = \int_{\mathbb{R}} y\, dP^Y(y) = \int_{\mathbb{R}} g(x) dP^X(x)$$

Intuitivement: l'espérance d'une VA est une caractéristique qui lui est intrinsèque et peut se calculer dans n'importe quel modèle.

■ **<u>Conséquences notables (exercices)</u>**:

1) Montrer que l'espérance est linéaire, i.e. A et B étant des matrices constantes, X et Y des vecteurs aléatoires:

$$E(AX+BY)=A\,E(X)+B\,E(Y)$$

2) L'espérance d'un vecteur aléatoire $X\in\mathbb{R}^p$ est le vecteur ayant pour composantes les espérances des composantes de X. Montrer que chacune peut se calculer indépendamment des autres, dès qu'on en connaît la loi.

3) Montrer que si une VAR X est intégrable et possède une distribution symétrique par rapport à μ, alors son espérance est μ. (On pourra commencer avec le cas $\mu = 0$).

■ **<u>Changement de variable</u>: technique de la fonction muette**

Soient X v.a. $\in\mathbb{R}^p, g:\mathbb{R}^p\rightarrow\mathbb{R}^q$, $Y=g(X)$. Comment trouver la loi de Y si l'on connaît celle de X? On imagine que l'on calcule l'espérance d'une fonction quelconque ϕ de Y dans les deux modèles: celui de X et celui de Y, et on écrit l'égalité entre ces deux calculs, due au théorème de transfert:

$$E(\phi(Y))=\int_{\mathbb{R}^q}\phi(y)\,dP^Y(y)=\int_{\mathbb{R}^p}\phi(g(x))\,dP^X(x)$$

Chaque élément de probabilité est ensuite exprimé en fonction de la densité et de l'élément de mesure (N.B. comme ϕ n'a aucune importance, on la remplace par un point pour alléger les notations):

$$\int_{\mathbb{R}^q}.\,f^Y(y)\,dy=\int_{\mathbb{R}^p}.\,f^X(x)\,dx \quad (1)$$

Il ne reste plus qu'à exprimer $f^X(x)$ et dx en fonction de y et dy pour en déduire, par identification dans (1), l'expression de $f^Y(y)$. Mais ATTENTION: lors du passage de x à y, dans (1), il est essentiel de faire attention aux bornes des intégrales: dès que g n'est pas injective, chaque y pouvant-être rapporté à plusieurs x, il faudra découper le domaine de X en sous-domaines où g est injective et recoller les morceaux. Il faut également s'assurer que l'élément de probabilité est positif en prenant sa valeur absolue.

En pratique:

a) Cas bijectif

$X\in\mathbb{R}^p, g:\mathbb{R}^p\rightarrow\mathbb{R}^p$ bijective. $Y=g(X)\Leftrightarrow X=g^{-1}(Y)$. On pose:

$|dx|:=|dx_1\ldots.\,dx_p|$; $|dy|:=|dy_1\ldots.\,dy_p|$. On a:

$|dx|=|H||dy|$ où H est la matrice jacobienne de g^{-1} et $|H|$ la valeur absolue de son déterminant (tout doit être positif dans cette histoire, puisqu'il s'agit d'éléments de probabilité).

On écrit alors:

$$f^Y(y)|dy|=f^X(x)|dx|=f^X(g^{-1}(y))|H||dy|$$

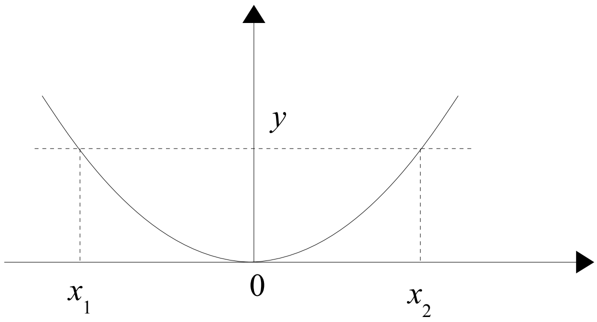
Ce qui donne par identification: $f^Y(y) = f^X(g^{-1}(y))|H|$

b) Cas non bijectif:

On découpe le domaine de X en sous-ensembles sur lesquels la restriction de g est bijective et on y applique la méthode précédente. Voyons ça sur un exemple simple et réel: $X \in \mathbb{R}$; $Y = X^2$.

Soient g_1 et g_2 les restrictions (bijectives) de g à $\mathbb{R}_-$ et $\mathbb{R}_+$ respectivement, et H_1 et H_2 les matrices jacobiennes de g_1^{-1} et g_2^{-1} respectivement.

On a $g_1^{-1} = -\sqrt{y}$ et $g_2^{-1} = \sqrt{y}$. Comme on est dans le cas réel, la matrice jacobienne n'est autre que la dérivée: $|H_1(y)| = \left| -\frac{1}{2} y^{-\frac{1}{2}} \right|$; $|H_2(y)| = \left| \frac{1}{2} y^{-\frac{1}{2}} \right|$



On écrit alors:

$$f^Y(y)|dy| = f^X(g_1^{-1}(y))|H_1||dy| + f^X(g_2^{-1}(y))|H_2||dy|$$

Ce qui donne, par identification:

$$f^Y(y) = f^X(g_1^{-1}(y))|H_1| + f^X(g_2^{-1}(y))|H_2|$$

soit, ici:

$$f^Y(y) = f^X(-\sqrt{y})\frac{1}{2}\left|y^{-\frac{1}{2}}\right| + f^X(\sqrt{y})\frac{1}{2}\left|y^{-\frac{1}{2}}\right| = \frac{1}{2}\left|y^{-\frac{1}{2}}\right|\left(f^X(-\sqrt{y}) + f^X(\sqrt{y})\right)$$

Exemples / exercices:

1) X étant une v.a.r. de loi exponentielle, i.e. de densité: $f^X(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$, trouver les densités des variables: $Y = e^X$; $Z = X^\alpha, \alpha > 0$; $T = \ln X$

2) X étant une v.a.r. de loi $U_{[0;\pi]}: f^X(x) = \frac{1}{\pi}\mathbf{1}_{[0;\pi]}(x)$, trouver les densités des variables: $Y = \sin X$; $Z = tg\, X$

3) $X \in \mathbb{R}^2$ étant un v.a. gaussien standard (i.e. d'espérance nulle et de matrice de variance = I_2), et $Y = (R, \Theta)$ sa représentation en coordonnées polaires, trouver la loi de *Y*. N.B. On rappelle la densité de *X*:

$$\forall x \in \mathbb{R}^2: \quad f^X(x) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$$

Variance et covariance

■ **Définitions**:

- ***Cas réel***:

Variance: $\forall X \in \mathbb{R}, V(X) := E(X - EX)^2$

Covariance: $\forall X, Y \in \mathbb{R}, Cov(X; Y) := E((X - EX)(Y - EY))$

N.B. La covariance est symétrique.

- ***Extension multidimensionnelle***:

Variance: $\forall X \in \mathbb{R}^p, V(X) := E(X - EX)(X - EX)'$

N.B. Donc: $(V(X))_{ij} = cov(X_i; X_j)$; $(V(X))_{ii} = V(X_i)$; et la matrice est symétrique.

Covariance: $\forall X \in \mathbb{R}^p, Y \in \mathbb{R}^q, Cov(X; Y) := E((X - EX)(Y - EY)')$

■ **Conséquences notables (exercices)**:

1) Montrer que:

$$\forall X \text{ v.a. et } a, b \text{ constantes } \in \mathbb{R} : V(aX + b) = a^2 V(X)$$

$$\forall X \text{ v.a.} \in \mathbb{R}^p,\ b \text{ constant } \in \mathbb{R}^q \text{ et } A \text{ matrice constante } (p, q) : V(AX + b) = A\, V(X) A'$$

$$\forall X, Y \text{ v.a. et } a, b, c, d \text{ constantes } \in \mathbb{R} : Cov(aX + b, cY + d) = ac\, Cov(X; Y)$$

Et enfin, que *X* et *Y* étant des vecteurs aléatoires et *A*, *C*, *b* et *d* des matrices/vecteurs constants aux dimensions convenables:

$$Cov(AX + b, CY + d) = A\, Cov(X; Y) C'$$

(… mais tout ça, bien sûr, dans l'ordre que vous voulez.)

2) Montrer que:

$$\forall X \in \mathbb{R}, V(X) = E(X^2) - (EX)^2$$

$$\forall X \in \mathbb{R}^p, V(X) = E(XX') - E(X) E(X)'$$

$$\forall X, Y \in \mathbb{R}, Cov(X; Y) = E(XY) - EX\, EY$$

$$\forall X \in \mathbb{R}^p, Y \in \mathbb{R}^q, Cov(X; Y) = E(XY') - E(X) E(Y)'$$

3) Montrer que si *X* est une v.a.r. et *g* une fonction réelle croissante:

$$Cov(X; g(X)) \geq 0$$

Et si *g* est décroissante?

Loi et moments conditionnels

■ Distribution conditionnelle:

Fonction de répartition conditionnelle:

$$\forall\, X\in\mathbb{R}^p,\, Y\in\mathbb{R}:\ F^Y_{X=x}(y) := \lim_{dx\to 0} P_{(x<X\le x+dx)}(Y\le y)$$

$$=\lim_{dx\to 0}\frac{P(x<X\le x+dx\ \wedge\ Y\le y)}{P(x<X\le x+dx)}=\lim_{dx\to 0}\frac{\frac{1}{dx}\left(F^{(X,Y)}(x+dx,y)-F^{(X,Y)}(x,y)\right)}{\frac{1}{dx}\left(F^X(x+dx)-F^X(x)\right)}$$

$$\Leftrightarrow\quad F^Y_{X=x}(y)=\frac{\frac{\partial}{\partial x}F^{(X,Y)}(x,y)}{\frac{d}{dx}F^X(x)}$$

Densité conditionnelle:

$$f^Y_{X=x}(y)=\frac{d}{dy}F^Y_{X=x}(y)=\frac{\frac{\partial^2}{\partial x\partial y}F^{(X,Y)}(x,y)}{\frac{d}{dx}F^X(x)}=\frac{f^{(X,Y)}(x,y)}{f^X(x)}$$

Écriture utile:

$$f^{(X,Y)}(x,y)=f^Y_{X=x}(y)\,f^X(x)$$

D'où la **formule de Bayes**:

$$f^{(X,Y)}(x,y)=f^Y_{X=x}(y)\,f^X(x)=f^X_{Y=y}(x)\,f^Y(y)$$

$$\Rightarrow\quad f^Y_{X=x}(y)=\frac{f^X_{Y=y}(x)\,f^Y(y)}{f^X(x)}$$

Indépendance:

Plusieurs définitions équivalentes (l'équivalence peut être montrée en exercice)):

- ■ $X_1,\dots,X_p$ indépendantes
  $\Leftrightarrow \forall\{i_1,\dots,i_q\}\subset\{1,\dots,p\},\forall A_{i_1},\dots,A_{i_q}:\ P^{(X_{i_1},\dots,X_{i_q})}(A_{i_1},\dots,A_{i_q})=P^{X_{i_1}}(A_{i_1})\dots P^{A_{i_q}}(x_{i_q})$
- ■ $X_1,\dots,X_p$ indépendantes
  $\Leftrightarrow \forall\{i_1,\dots,i_q\}\subset\{1,\dots,p\}:\ F^{(X_{i_1},\dots,X_{i_q})}(x_{i_1},\dots,x_{i_q})=F^{X_{i_1}}(x_{i_1})\dots F^{X_{i_q}}(x_{i_q})$
- ■ $X_1,\dots,X_p$ indépendantes
  $\Leftrightarrow \forall\{i_1,\dots,i_q\}\subset\{1,\dots,p\}:\ f^{(X_{i_1},\dots,X_{i_q})}(x_{i_1},\dots,x_{i_q})=f^{X_{i_1}}(x_{i_1})\dots f^{X_{i_q}}(x_{i_q})$

- $X_1, \dots, X_p$ indépendantes $\Leftrightarrow \forall I \subset \{1, \dots, p\}$: $f_{(X_i)_{i \notin I}}^{(X_i)_{i \in I}} (x_i)_{i \in I} = f^{(X_i)_{i \in I}} (x_i)_{i \in I}$

Dans le cas de deux v.a., les formules se simplifient, regardez.

■ **Moments conditionnels**:

Ce sont les moments de la loi conditionnelle:

- Espérance conditionnelle:

$$E_{X=x} Y := \int y \, dP_{X=x}^{Y}(y) = \int y \, f_{X=x}^{Y}(y) \, dy = \Phi(x)$$

$$E_X Y := \Phi(X)$$

- Variance conditionnelle:

$$V_{X=x} Y := E_{X=x}(Y - E_{X=x} Y)^2 = \Psi(x)$$

$$V_X Y := \Psi(X)$$

- Covariance conditionnelle:

$$Cov_{X=x}(Y, Z) := E_{X=x}(Y - E_{X=x} Y)(Z - E_{X=x} Z) = \Theta(x)$$

$$Cov_X(Y, Z) := \Theta(X)$$

N.B. Les extensions multidimensionnelles sont directes.

■ **Conséquences / exercices**:

1) Montrez que:

$$X_1, \dots, X_p \text{ indépendantes} \Rightarrow E(X_1 \dots X_p) = E(X_1) \dots E(X_p)$$

En déduire que:

$$X = (X_1, \dots, X_p) \text{ où les } X_i \text{ sont indépendantes} \Rightarrow V(X) \text{ diagonale.}$$

En calculant $Cov(X ; X^2)$ où X a une distribution symétrique par rapport à 0, montrez que la réciproque est fausse.

2) Montrez que (et interprétez):

$$\forall X, Y \text{ v.a.}: \ EY = E(E_X Y)$$

3) Montrez que (et interprétez):

$$\forall X, Y \text{ v.a.}: \ VY = V(E_X Y) + E(V_X Y)$$

4) Montrez que (et interprétez):

$$\forall X, Y, Z \text{ v.a.}: \ Cov(Y ; Z) = Cov(E_X Y ; E_X Z) + E(Cov_X(Y ; Z))$$

Théorèmes asymptotiques pour un échantillon i.i.d.

$\underline{X}_n := (X_1, \ldots, X_n)$ où les X_i sont des v.a.r. indépendants et de même loi, celle de la v.a. générique $X \in \mathbb{R}^p$. On supposent qu'existent: $E(X) = \mu \,;\, V(X) = \Omega$.

1. Loi forte des grands nombres

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \underset{n \to \infty}{\overset{\text{p.s.}}{\rightarrow}} \mu$$

2. Théorème central-limite

$$\sqrt{n}(\bar{X}_n - \mu) \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;\Omega)$$

Version unidimensionnelle ($X \in \mathbb{R} \,;\, V(X) = \sigma^2$):

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;1)$$

3. Transport de convergences par des applications

3.1. Transport d'une convergence p.s. par une application continue

Soit une v.a. $Z_n \in \mathbb{R}^p$ telle que $Z_n \underset{n \to \infty}{\overset{\text{p.s.}}{\rightarrow}} m$, et $g : \mathbb{R}^p \to \mathbb{R}^q$ continue. Alors:

$$g(Z_n) \underset{n \to \infty}{\overset{\text{p.s.}}{\rightarrow}} g(m)$$

3.2. Transport d'une convergence en loi par une application différentiable: Delta-méthode

- Soit $g : \mathbb{R}^p \to \mathbb{R}^q$ différentiable de matrice jacobienne: $\Gamma(x) = \left(\left(\frac{\partial g_i}{\partial x_j}(x) \right) \right)_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}}$ tq $\Gamma(m)$ soit de plein rang en colonne.

Soit une v.a. $Z_n \in \mathbb{R}^p$. Si $\sqrt{n}(Z_n - m) \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;\Sigma)$, alors:

$$\sqrt{n}(g(Z_n) - g(m)) \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;\Gamma(m)\Sigma\Gamma(m)')$$

Moralement, cela signifie que: Z_n tend vers m de sorte que dilatée par sqrt n, leur différence converge en loi vers une loi fixe. Z_n étant pour n grand dans un voisinage de m, la fonction g y est équivalente à l'application linéaire tangente (son application différentielle). Les applications linéaires transportant le comportement gaussien, on a le résultat.

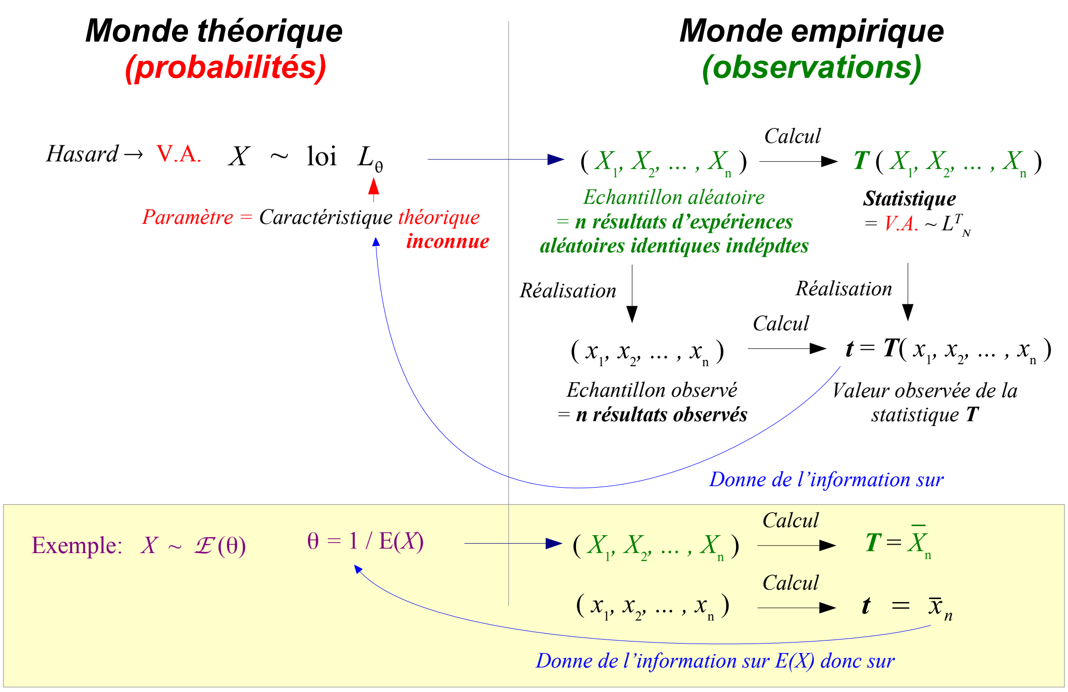
- Version unidimensionnelle ($Z_n \in \mathbb{R} \,;\, g'(m) \neq 0$):

Si $\sqrt{n}(Z_n - m) \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;\sigma^2)$, alors: $\sqrt{n}(g(Z_n) - g(m)) \underset{n \to \infty}{\overset{\text{loi}}{\rightarrow}} N(0\,;g'^2(m)\sigma^2)$

4. Utilisation

En statistique inférentielle paramétrique, on fait l'hypothèse que la loi d'une variable aléatoire appartient à une certaine famille dans laquelle elle est identifiée par la valeur du (vecteur des) paramètre(s), qui est *inconnue*. Il s'agit, à travers les observations que l'on peut faire de cette variable, d'obtenir de l'information sur ce paramètre.

Les deux mondes



Continuons avec l'exemple de la loi exponentielle $\mathcal{E}(\theta)$. La LGN nous assure que la moyenne empirique $\bar{X}_n$ converge p.s. vers $E(X)=\frac{1}{\theta}$. On dit que $\bar{X}_n$ est un estimateur convergent (p.s.) de $E(X)$. La fonction $g: x \to \frac{1}{x}$ étant continue sur $\mathbb{R}_+^*$, il s'ensuit que:

$$\frac{1}{\bar{X}_n} \underset{n\to\infty}{\overset{\text{p.s.}}{\to}} \frac{1}{E(X)} = \theta$$

On a donc obtenu un estimateur convergent (p.s.) de θ. Le TCL et la delta-méthode nous permettent même d'en obtenir la loi limite! (exercice).

Mais la question qui se pose est: "la statistique $\bar{X}_n$ est-elle celle qui nous donne le plus d'information sur θ ou y avait-il une meilleure statistique?", d'où un questionnement en chaîne: "Qu'est-ce que l'information sur θ? Comment la mesurer? En utilisant bien cette information, jusqu'à quelle précision peut-on compter pour estimer θ à partir de l'échantillon et quelle statistique nous permettra d'atteindre cette précision?"

C'est à ce questionnement que la statistique inférentielle paramétrique se propose de répondre.