MAITRISE de MATHEMATIQUES **Université Paul Sabatier**

Probabilités et Statistiques, B5, 2002-2003 **U.F.R. M.I.G.**

42 h cours, 48 h TD.

INTRODUCTION

Ce cours s'appuie sur deux cours de licence :
intégration et théorie de la mesure (NL 2),
probabilités (NL 7 ou 12).

1 Révisions de licence et théorèmes des classes monotones

Ce premier chapitre est essentiellement consacré à des révisions, à l'exception d'un outil important : les deux théorèmes des classes monotones.

1.1 Espace de probabilité

(i) réciter ce que sont
une expérience aléatoire $\mathcal{E}$,
les réalisations ou épreuves ω, d'ensemble Ω, $\mathcal{P}(\Omega)$ ensembles des parties de Ω,
tribu ou σ-algèbre $\mathcal{A}$ sur Ω,
tribu engendrée par une partie de $\mathcal{P}(\Omega)$,
événements particuliers : certain, impossible, contradictoires ou incompatibles, singletons,
probabilité sur une tribu $\mathcal{A}$.

(ii) Exemples sur des espaces Ω finis. Notion d'équiprobabilité.

(iii) Indépendance et conditionnement (FORMULAIRE, page 1)
Définitions de l'indépendance et de la mutuelle indépendance.
Définition de la proba conditionnelle d'un événement $A \in \mathcal{A}$ sachant $B \in \mathcal{A}$, B non impossible. les trois formules :

probas totales (sur un système complet d'événements)

probas composées (idem)

formule de BAYES (probas a priori et a posteriori)

Définition 1.1 *On dit que deux événements A et B d'une tribu $\mathcal{T}$ sont* **indépendants** *si* $\mathbb{P}(A \cap B) = \mathbb{P}(A).\mathbb{P}(B)$.

Définition 1.2 *On dit que n événements A_n d'une tribu $\mathcal{T}$ sont mutuellement indépendants si*

$$\forall k, \forall (n_1, \cdots, n_k), \mathbb{P}(A_{n_1} \cap \cdots \cap A_{n_k}) = \Pi_{j=1}^k \mathbb{P}(A_{n_j}).$$

Définition 1.3 *On appelle probabilité conditionnelle d'un événement $A \in \mathcal{A}$ sachant $B \in \mathcal{A}$, B non impossible, c'est à dire $\mathbb{P}(B) \neq 0$, la quantité notée $\mathbb{P}(A/B)$, et définie par* $\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$.

Trois formules :

probabilités totales (sur un système complet d'événements, c'est à dire une partition de Ω en événements (A_n)) :

$$\forall A \in \mathcal{T}, \mathbb{P}(A) = \sum_n \mathbb{P}(A \cap A_n).$$

probabilités composées (idem avec $\forall n, \mathbb{P}(A_n) \neq 0$)

$$\forall B \in \mathcal{T}, \mathbb{P}(B) = \sum_n \mathbb{P}(B/A_n)\mathbb{P}(A_n).$$

Formule de BAYES (probabilités a priori et a posteriori) :

$$\mathbb{P}(A_i/B) = \frac{\mathbb{P}(B/A_i)\mathbb{P}(A_i)}{\sum_n \mathbb{P}(B/A_n)\mathbb{P}(A_n)}.$$

1.2 Premier théorème de classes monotones

Définition 1.4 *Soit $\mathcal{C} \subset \mathcal{P}(\Omega)$. C'est une classe monotone si elle est stable par réunion croissante (ou intersection décroissante) dénombrable.*

Théorème 1.5 *Soit $\mathcal{J} \subset \mathcal{P}(\Omega)$ stable par intersection et $\mathcal{C}$ une classe monotone vérifiant :*

(i) $\Omega \in \mathcal{C}$,

(ii) si A *et* $B \in \mathcal{C}, B \subset A$, *alors* $A - B = A \cap B^c \in \mathcal{C}$,

(iii) $\mathcal{J} \subset \mathcal{C}$,

alors $\mathcal{C}$ contient la tribu engendrée par $\mathcal{J}, \sigma(\mathcal{J})$.

1.3 Révisions du cours d'intégration

On se place dans tout ce paragraphe sur un espace mesurable $(\Omega, \mathcal{A})$ et on se réfère au poly de Wagshall (W). On notera toujours la tribu des boréliens $\mathcal{B}$.

1.3.1

réciter ce que sont :

les fonctions étagées (W 2.1)

les fonctions mesurables (W 2.2)

Proposition 1.6 *(W 2.2.10) Toute fonction mesurable positive de* $(\Omega, \mathcal{A})$ *dans* $(\overline{\mathbb{R}_+}, \mathcal{B})$ *est limite d'une suite croissante de fonctions étagées positives.*

Corollaire 1.7 *(W 2.2.11) Toute fonction mesurable de* $(\Omega, \mathcal{A})$ *dans* $(\overline{\mathbb{R}}, \mathcal{B})$ *est limite d'une suite de fonctions étagées.*

1.4 Second théorème de classes monotones

Théorème 1.8 *Soit* $(\Omega, \mathcal{A})$ *un espace mesurable et* $\mathcal{J} \subset \mathcal{P}(\Omega)$ *stable par intersection telle que* $\mathcal{A} = \sigma(\mathcal{J})$*. Soit* $\mathcal{F}$ *un espace vectoriel de fonctions mesurables réelles sur* $(\Omega, \mathcal{A})$ *tel que :*

(a) $\mathbf{1}_\Omega \in \mathcal{F}$

(b) $\forall A \in \mathcal{J}, \mathbf{1}_A \in \mathcal{F}$,

(c) si $(f_n, n \geq 0)$ *est une suite croissante de fonctions positives de* $\mathcal{F}$ *de limite finie (resp. bornée)* f*, alors* $f \in \mathcal{F}$*,*
alors $\mathcal{F}$ *contient toutes les fonctions mesurables réelles positives finies (resp. bornées) sur* $(\Omega, \mathcal{A})$.

APPLICATIONS :

1.4.1 Construction de l'intégrale

On considère ici un espace mesuré $(\Omega, \mathcal{A}, \mu)$. On définit d'abord l'intégrale des fonctions étagées. Puis, si f est une fonction positive mesurable, elle est limite d'une une suite de fonctions étagées positives $(f_n$)et $\int f d\mu = \lim_n \int f_n d\mu$.

Enfin, on dit que f est intégrable si $\int |f| d\mu < \infty$. Et dans ce cas $\int f d\mu = \int f_+ d\mu - \int f_- d\mu$.

1.4.2 Théorèmes de convergence (W 2.6)

FORMULAIRE, page 2.

Théorème 1.9 *(convergence monotone) Soit une suite croissante de fonction mesurables positives* (f_n) *de* $(\Omega, \mathcal{A})$ *dans* $(\bar{\mathbb{R}_+}, \mathcal{B})$*, alors* $\int \lim_n f_n d\mu = \lim_n \int f_n d\mu$.

Théorème 1.10 *(Beppo-Levi, W th 2.6.4.) Soit une suite monotone de fonction mesurables* (f_n) *de* $(\Omega, \mathcal{A})$ *dans* $(\bar{\mathbb{R}}, \mathcal{B})$*, alors* $f = \lim_n f_n$ *est intégrable si et seulement si la suite* $(\int f_n d\mu)$ *converge ; dans ce cas*

$$\int \lim_n f_n d\mu = \lim_n \int f_n d\mu.$$

Théorème 1.11 *(lemme de Fatou, W th 2.6.6.) Soit une suite de fonction mesurables positives* (f_n) *de* $(\Omega, \mathcal{A})$ *dans* $(\bar{\mathbb{R}_+}, \mathcal{B})$*, alors*

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

Théorème 1.12 *(convergence dominée, W 2.6.10) Soit une suite de fonction mesurables* (f_n) *de* $(\Omega, \mathcal{A})$ *dans* $(\bar{\mathbb{R}}, \mathcal{B})$*, convergeant* μ *presque partout vers* f*, s'il existe* g *de* $(\Omega, \mathcal{A})$ *dans* $(\bar{\mathbb{R}}, \mathcal{B})$ *intégrable telle que pour tout* $n, |f_n| \leq g$ μ *presque partout, alors* f *est intégrable et*

$$\int f d\mu = \lim_n \int f_n d\mu.$$

1.4.3 Fonctions définies par une intégrale (W 2.8)

FORMULAIRE, page 2.
théorèmes de continuité et de dérivabilité sous le signe somme : W th 2.8.2 et 2.8.3.

Théorème 1.13 *Soit une fonction* $f : \Omega \times X \mapsto \overline{\mathbb{R}}$*, où* X *est un espace topologique. Si*

- *(i)* $\forall x \in X, f(., x) \in L^1(\Omega, \mathcal{T}, \mu)$,
- *(ii)* $\mathbb{P}\{\omega : x \mapsto f(\omega, x)$ *est continue en* $a\} = 1$,
- *(iii)* *il existe* $g \in L^1((\Omega, \mathcal{T}, \mu; \overline{\mathbb{R}^+})$ *et* $V(a)$ *voisinage de* a *tels que* $\forall x \in V(a), |f(\omega, x)| \leq g(\omega)$ *p. s.*

alors $x \mapsto \int_\Omega f(\omega, x) d\mu(\omega)$ *est continue au point* a*.*

Théorème 1.14 *Soit une fonction* $f : \Omega \times X \mapsto \overline{\mathbb{R}}$*, où* X *est un espace topologique. Si*

- *(i)* $\forall x \in X, f(., x) \in L^1(\Omega, \mathcal{T}, \mu)$,
- *(ii)* $\mathbb{P}\{\omega : x \mapsto f(\omega, x)$ *est dérivable (resp. de classe* C^1*) en* $a\} = 1$,
- *(iii)* *il existe* $g \in L^1((\Omega, \mathcal{T}, \mu; \overline{\mathbb{R}^+})$ *et* $V(a)$ *un voisinage de* a *tels que* $\forall x \in V(a), |D_x f(\omega, x)| \leq g(\omega)$

alors $x \mapsto \int_\Omega f(\omega, x) d\mu(\omega)$ *est dérivable (resp. de classe* C^1*) au point* a*.*

2 Variables aléatoires

Ce chapitre, aussi, contient essentiellement des révisions.

On se place sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et on considère un espace mesurable $(E, \mathcal{E})$. Le plus souvent E sera l'ensemble des réels $\mathbb{R}, \overline{\mathbb{R}}, \mathbb{R}_+, \overline{\mathbb{R}_+}$ et $\mathcal{E}$ la tribu de leurs boréliens.

2.1 Définitions générales

variable aléatoire X de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(E, \mathcal{E})$; loi d'une variable aléatoire rélle (var) ; $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ sur $(E, \mathcal{E})$ par X.
X et Y sont dites "équidistribuées" si elles ont même loi.
Tribu engendrée par une famille de variables aléatoires $(X_i, i \in I)$.

Théorème 2.1 *(lemme de Doob) Si $\sigma(X)$ est la tribu engendrée par $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, alors $Y : \Omega \to \mathbb{R}$ est $(\sigma(X), \mathcal{E})$-mesurable si et seulement si il existe une fonction borélienne sur $(E, \mathcal{E})$ telle que $Y = h \circ X$.*

cf. [1] page 168. preuve faite en exercice avec le deuxième théorème des classes monotones.
je n'ai pas eu le temps de faire le th de transport....

1er octobre

2.2 Variables aléatoires réelles

Définition 2.2 *Fonction de répartition* $F_X : x \mapsto \mathbb{P}\{X \leq x\}$.

Réciter les propriétés de la fdr.

Exemples de lois discrètes, de lois continues.

2.3 Lois images, théorème du transport ou du transfert

On se place dans tout ce paragraphe sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et on notera toujours la tribu des boréliens $\mathcal{B}$ sur $\mathbb{R}$ ou $\mathbb{R}^d$. Soit donc $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B})$ et g fonction borélienne ; $Y = g \circ X$. La loi de Y est $\mathbb{P}_Y = h_*(\mathbb{P}_X)$. (Wagshal 2.9)

Exemples importants :

lois marginales d'un vecteur

cas où g est un difféomorphisme et la loi de X admet une densité h sur un ouvert D.

Théorème 2.3 *de changement de variables : Y est de densité $g \circ h^{-1}|J_{h^{-1}}|\mathbf{1}_{h(D)}$.*

Théorème 2.4 *du transport ou du transfert : sous les mêmes hypothèses,*

$$E[g(X)] \text{ existe si et seulement si } \int |g(x)| d\mathbb{P}_X(x) < \infty,$$

et dans ce cas, $E[g(X)] = \int g(x)\ d\mathbb{P}_X(x)$.

Exercices fait en amphi Si (X, Y) est de loi *gamma* sur $\mathbb{R}^+$, trouver la loi de $U = X/X + Y$; $V = X + Y$.

Si X est de loi uniforme sur $[0, 1]$, quelle est la loi de X^2 ?
même question si X est de loi uniforme sur $[-1, 1]$?

2.3.1 Convergence de suites de variables aléatoires et lemme de Borel-Cantelli

faire réciter convergence en loi (étroite), en proba et $\mathbb{P}$-presque sûre ; liens entre les trois. Ils disent (ceux de Toulouse) qu'ils n'ont pas vu le théorème

Théorème 2.5 *Si une suite de variables aléatoires*(X_n) *sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *converge en probabilité, il existe une suite extraite* (X_{n_k}) *qui converge presque sûrement.*

Voir la preuve dans le chapitre 7.

Lemme 2.6 *Soit* $(A_n, n \geq 0)$ *une suite d'événements de* $\mathcal{A}$ *et* $A^* = \limsup_n A_n = \cap_n \cup_{k \geq n} A_k$.

a) si $\sum_n \mathbb{P}(A_n) < \infty$, *alors* $\mathbb{P}(A^*) = 0$.

b) si les événements $(A_n, n \geq 0)$ *sont deux à deux indépendants, et si* $\sum_n \mathbb{P}(A_n) = \infty$, *alors* $\mathbb{P}(A^*) = 1$.

Emile BOREL, (Saint Affrique 1871, 1956) professeur à l'ENS et député de l'Aveyron ; Francesco CANTELLI, Palerme, 1875, professeur dans une école d'économie et de commerce à Rome. *preuve de b*

2.4 Vecteurs aléatoires

On appelle **vecteur aléatoire** une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ muni de la tribu borélienne.

Son espérance existe si et seulement si $X_i = \pi_i(X)$ (π_i est la projection de $\mathbb{R}^d$ dans $\mathbb{R}$ qui à x fait correspondre sa $i$ème coordonnée) est intégrable pour tout $i = 1 \cdots d$ et c'est alors le vecteur de $\mathbb{R}^d$ de coordonnées $E(X_i)$.

Proposition 2.7 *Pour tout matrice* A *d'une application linéaire de* $\mathbb{R}^d$ *dans* $\mathbb{R}^p$,

$$E[AX] = AE[X]; V(AX) = AK_X\tilde{A}.$$

Théorème 2.8 *Soit* X *un vecteur aléatoire et* g *une fonction borélienne sur* $\mathbb{R}^d$. $g(X)$ *est inégrable si et seulement si* $\int_{\mathbb{R}^d} |g(x_1, \cdots, x_d)| d\mathbb{P}_X(x_1, \cdots, x_d) < \infty$. *Et dans ce cas,*

$$E[g(X)] = \int_{\mathbb{R}^d} g(x_1, \cdots, x_d) d\mathbb{P}_X(x_1, \cdots, x_d).$$

mettre ds le chapitre des espaces L^p

2.5 Moments des variables aléatoires réelles

Ce ne sont que des rappels de licence.

2.5.1 Espérance

Espérance ou moment d'ordre 1 notée $E[X]$ et définie par $\int_{\mathbb{R}} xd\mathbb{P}_X(x)$ si l'intégrale $\int_{\mathbb{R}} |x|d\mathbb{P}_X(x) < \infty$.

L'ensemble des var intégrables est noté $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble quotient par la relation d'équivalence "égales presque sûrement" est noté $L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Si $E|X] = 0$, on dit que X est centrée.

2.5.2 Moments d'ordre p

Le moment d'ordre p est $E[|X|^p]$, défini par $\int_{\mathbb{R}} |x|^p d\mathbb{P}_X(x)$ si cette intégrale est finie.

L'ensemble des var admettant un moment d'ordre p est noté $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble quotient par la relation d'équivalence "égales presque sûrement" est noté $L^p(\Omega, \mathcal{A}, \mathbb{P})$.

On appelle le moment centré l'espérance $E[|X - E(X)|^p]$.

Cas particulier de $p = 2$ le moment centré d'ordre 2 s'appelle la **variance**, sa racine carrée l'**écart-type** :

$$V(X) = E[|X - E(X)|^2] \ ; \ \sigma_X = \sqrt{V(X)}.$$

Lemme 2.9 *de BIENAYME (Jules)-TCHEBICHEV (Pafnouti) Soit* $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ *; pour tout* $a > 0$, $\mathbb{P}\{|X| \geq a\} \leq a^{-p}E[|X|^p]$.

Corollaire 2.10 *Soit* $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ *; pour tout* $a > 0$, $\mathbb{P}\{|X - E(X)| \geq a\sigma_X\} \leq a^{-2}$.

C'est à cause de ce résultat que l'on dit de la quantité σ_X que c'est un *écart-type*.

Proposition 2.11 *Si* $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$*,* $V(X)$ *réalise l'infimum de l'application* $a \mapsto E[|X - a|^2]$ *et cet infimum est atteint en* $a = E[X]$.

Pour tout réel a *et* b $E[aX + b] = aE[X] + b, V[aX + b] = a^2V(X)$.

Dans le cas $d = 2$, et lorsque les deux composantes X et $Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ on définit la **covariance** de X et Y : c'est l'espérance du produit $(X - E(X))(Y - E(Y))$ notée $cov(X, Y)$.
On définit également le **coefficient de corrélation**

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}.$$

Exercice :$\rho_{X,Y}$ est à valeurs dans $[-1, +1]$; si $\rho^2_{X,Y} = 1$, il existe une relation affine entre X et Y.

La matrice de covariance du vecteur aléatoire X existe si pour tout $i = 1 \cdots d, X_i \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et elle est alors définie par $K_X = E[(X - E(X))(X - \tilde{E}(X))]$, c'est à dire de terme général $cov(X_i, X_j)$.

3 Fonctions caractéristiques, transformées de Fourier et de Laplace

On se place toujours sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et on considère les var sur $(\Omega, \mathcal{A}, \mathbb{P})$.

3.1 Fonction caractéristique des vecteurs aléatoires

Définition 3.1 *Soit μ une mesure de proba sur $\mathbb{R}^d$:*

$$\phi_\mu(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} d\mu(x).$$

Ou encore, si X est un vecteur aléatoire, $\phi_X(t)$ note la fonction caractéristique de la mesure de probabilité $\mathbb{P}_X$, c'est à dire

$$\phi_X(t) = E[e^{i\langle t,X\rangle}].$$

propriétés à faire réciter
Remarque : $\phi_X(t) = (\sqrt{2\pi})^d \hat{\mathbb{P}_X}(-t)$, c'est à dire la transformée de Fourier de la mesure $\mathbb{P}_X$.

Théorème 3.2 *Si μ et μ' ont même fonction caractéristique, elles sont égales.*

C'est une conséquence du théorème d'inversion :

Théorème 3.3 *Soit F la fonction de répartition d'une var X et $F^*(x) = \frac{1}{2}(F(x)+F(x_-))$ la "régularisée" en tout point x de $\mathbb{R}$. Si $x_1 < x_2$,*

$$F^*(x_1) - F^*(x_2) = \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{itx_1} - e^{itx_2}}{it} \phi_X(t) dt.$$

Preuve avec Fubini.

Corollaire 3.4 *Si $\phi_{X_1} = \phi_{X_2}$, alors $F_1^* = F_2^*$ c'est à dire que X_1 et X_2 sont de même loi, puisque la donnée de la fdr en tous ses points de continuité caractérise la loi.*

Lorsque ϕ_X est intégrable, on a mieux :

Théorème 3.5 *([1] page 67, exo 6.9 page 152)*
Si $\phi_X \in L^1(\mathbb{R}^d, \mathcal{B}, leb)$, X admet une densité sur $\mathbb{R}^d$ donnée par :

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \int e^{-\langle t,x\rangle} \phi_X(t) dt.$$

3.2 Fonction caractéristique et moments

(cf. [1] pages 68-69)

Proposition 3.6 *Si X admet un moment d'ordre k, ϕ_X est de classe C^k. Dans ce cas,*

$$\phi_X^{(k)}(t) = i^k \int x^k e^{itx} d\mathbb{P}_X(x) = i^k E[X^k e^{itX}].$$

Corollaire 3.7 *Si X admet un moment d'ordre k, $(\phi_X)^{(k)}(0) = i^k E(X^k)$.*

Réciproquement, on a

Proposition 3.8 *Si ϕ_X est $2k$ fois dérivable en 0, X admet tous ses moment d'ordre $\leq 2k$.*

On appelle **seconde fonction caractéristique** la détermination principale ψ_X de $\log \phi_X$ au voisinage de 0. Lorsque $X \in L^2$, on a un D.L. à l'ordre 2 :

$$\psi_X(t) = i\langle E(X), t\rangle - \frac{1}{2}\tilde{t}E(X\tilde{X})t + o(\|t\|^2).$$

Si ψ_X se développe plus loin :

$$\psi_X(t) = \sum_{j \leq k} \kappa_j \frac{(it)^j}{j!} + o((\|t\|^k),$$

Les coefficients κ_j s'appellent les **cumulants**. En particulier, κ_3 s'appelle la *skewness* et κ_4 s'appelle la *kurtosis*.

3.3 Fonction caractéristique de vecteurs aléatoires

Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^n$.

Définition 3.9 *Si X est une vecteur aléatoire, $\phi_X(t)$ note pour $t \in \mathbb{R}^n$, la fonction caractéristique définie par*

$$\phi_X(t) = E[e^{i\langle t,X\rangle}]$$

où $\langle .,. \rangle$ note le produit vectoriel.

Propriétés élémentaires....

Proposition 3.10 *Si X admet tous ses moments d'ordre k, ϕ_X est ce classe C^k.*

3.4 Transformée de Laplace

Définition 3.11 *La transformée de Laplace est l'application* $L^X : \mathbb{R}^d \to \mathbb{R}^+, s \mapsto E[e^{sX}]$.

Le dommaine de définition de L^X est l'ensemble des réels s tels que $e^{sX} \in L^1$.
$L^X(O) = 1$ donc $0 \in Dom(L^X)$.
De même que ϕ_X, L^X caractérise la loi de X.

Proposition 3.12 *Soit* X *var telle que il existe un intervalle ouvert contenant* 0 *et où* $e^{sX} \in L^1$. *Alors,* L^X *est définie sur un intervalle* J *symétrique contenu dans* I *; et* l^X *est analytique sur* J *:*

$$\forall t \in J, L^X(t) = \sum_{n \geq 0} \frac{t^n}{n!} E(X^n).$$

En particulier, $(L^X)^{(n)}(0) = E(X^n)$.

Rappel de DEUG : la fonction génératrice lorsque X est à valeurs entières.

3.5 Liens avec l'indépendance

3.5.1 Tribus indépendantes

Définition 3.13 *On dit qu'un famille de sous-tribus* $(\mathcal{A}_i, i \in I)$ *de* $\mathcal{A}$ *est (mutuellement) indépendante si* $\forall A_i \in \mathcal{A}_i, i \in I$*, les événements* $(A_i, i \in I)$ *sont (mutuellement) indépendants.*
On dit qu'un famille de var $(X_i, i \in I)$ *sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *sont (mutuellement) indépendantes si la famille des sous tribus* $(\sigma(X_i), i \in I)$ *est (mutuellement) indépendante.*

La définition est la même pour une famille de classes.

Proposition 3.14 *Soit une famille de classes* $(\mathcal{C}_i, i \in I)$ *indépendante et telle que pour tout* $i, \mathcal{C}_i$ *est stable par intersection. Alors les sous tribus engendrées* $(\sigma(\mathcal{C}_i), i \in I)$ *sont indépendantes.*

Corollaire 3.15 *Soit une famille de var* $(X_i, i = 1, \cdots, n)$ *: elles sont indépendantes si et seulement si pour tout* $(a_1, \cdots, a_n)$*,*
$\mathbb{P}\{X_1 \leq a_1, \cdots, X_n \leq a_n\} = \Pi_{i=1,n} \mathbb{P}\{X_i \leq a_i\}$.

Théorème 3.16 *(loi de zéro-un) Soient* $(\mathcal{B}_n, n \in \mathbb{N})$ *des sous-tribus indépendantes et* $\mathcal{C}_k = \sigma(\mathcal{B}_n, n \geq k)$ *notée* $\vee_{n=k}^{\infty} \mathcal{B}_n$ *et* $\mathcal{B}_\infty = \cap_k \mathcal{C}_k$. *Alors* $\mathcal{B}_\infty$ *est indépendante d'elle-même c'est à dire que pour tous ses événements* $B, pr(B) = 0$ *ou* 1.

3.5.2 Variables aléatoires indépendantes

Théorème 3.17 *Soit X un veceur aléatoire de composantes X_i : les var (X_i) sont indépendantes si et seulement si la loi de X $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$.*

Proposition 3.18 *Soit X un vecteur aléatoire de composantes X_i : les var (X_i) sont indépendantes si et seulement pour tout ensembles de fonctions boréliennes $(g_1, \cdots, g_n)$, $E[\Pi_i g(X_i)] = \Pi_i E[g(X_i)]$.*

Corollaire 3.19 *Si X et Y sont deux vecteurs aléatoires indépendants, $\phi_{X+Y} = \phi_X . \phi_Y$.*

idem avec k var.

3.5.3 Somme de variables aléatoires indépendante

Rappel de licence : si μ_i est la loi de $X_i, i = 1, 2$, deux var indépendantes, la loi de $X_1 + X_2$ est donnée par le produit de convolution $\mu_1 * \mu_2$.
En géralisant à n var indépendantes, la loi de $\sum_i X_i$ est donnée par le produit de convolution $\mu_1 * \cdots * \mu_n$. Il est plus maniable de passer par les fonctions caractéristiques en utilisant le corollaire 3.19. On a d'ailleurs la caractérisation :

Proposition 3.20 *$(X_1, \cdots, X_n)$ sont des vai si et seulement si $\phi_{\sum_i X_i} = \pi_i \phi_{X_i}$.*

une notion à ne pas confondre : var non corrélées. X et Y indépendantes, alors elles sont non corrélées. la réciproque est fausse.

Proposition 3.21 *Si $(X_1, \cdots, X_n)$ sont des var deux à deux non corrélées, alors $V(X_1 + \cdots + X_n) = \sum_i V(X_i)$.*

4 Espaces $L^p, p \in [1, +\infty]$

je m'aperçois que cette année, je n'ai pas parlé de Radon-Nikodym....

suite du 9 octobre

4.1 Moments des variables aléatoires réelles

Ce ne sont que des rappels de licence : [1] II.6, V.3

les deux sous-sections suivantes pour mémoire et survolées très rapidement....

4.1.1 Espérance

Espérance ou moment d'ordre 1 notée $E[X]$ et définie par $\int_{\mathbb{R}} x d\mathbb{P}_X(x)$ si l'intégrale $\int_{\mathbb{R}} |x| d\mathbb{P}_X(x) < \infty$.

L'ensemble des var intégrables est noté $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble quotient par la relation d'équivalence "égales presque sûrement" est noté $L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Si $E|X] = 0$, on dit que X est centrée.

4.1.2 Moments d'ordre p

Le moment d'ordre p est $E[|X|^p]$, défini par $\int_{\mathbb{R}} |x|^p d\mathbb{P}_X(x)$ si cette intégrale est finie.

L'ensemble des var admettant un moment d'ordre p est noté $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble quotient par la relation d'équivalence "égales presque sûrement" est noté $L^p(\Omega, \mathcal{A}, \mathbb{P})$.

On appelle le moment centré l'espérance $E[|X - E(X)|^p]$.

Cas particulier de $p = 2$ le moment centré d'ordre 2 s'appelle la **variance**, sa racine carrée l'**écart-type** :

$$V(X) = E[|X - E(X)|^2] \ ; \ \sigma_X = \sqrt{V(X)}.$$

Lemme 4.1 *de BIENAYME (Jules)-TCHEBICHEV (Pafnouti) Soit $X \in L^p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^d)$; pour tout $a > 0$, $\mathbb{P}\{\|X - E(X)\| \geq a\} \leq a^{-p} E[\|X - E(X)\|^p]$.*

Corollaire 4.2 *Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$; pour tout $a > 0$, $\mathbb{P}\{|X - E(X)| \geq a\sigma_X\} \leq a^{-2}$.*

C'est à cause de ce résultat que l'on dit de la quantité σ_X que c'est un *écart-type.*

j'ai oublié de dire ce corollaire

Proposition 4.3 *Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $V(X)$ réalise l'infimum de l'application $a \mapsto E[|X-a|^2]$ et cet infimum est atteint en $a = E[X]$.*

Pour tout réel a et b $E[aX+b] = aE[X]+b, V[aX+b] = a^2V(X)$.

Dans le cas $d = 2$, et lorsque les deux composantes X et $Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ on définit la **covariance** de X et Y : c'est l'espérance du produit $(X - E(X))(Y - E(Y))$ notée $cov(X, Y)$.
On définit également le **coefficient de corrélation**

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

Exercice :$\rho_{X,Y}$ est à valeurs dans $[-1, +1]$; si $\rho^2_{X,Y} = 1$, il existe une relation affine entre X et Y.

La matrice de covariance du vecteur aléatoire X existe si pour tout $i = 1 \cdots d, X_i \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et elle est alors définie par $K_X = E[(X - E(X))(X - \tilde{E}(X))]$, c'est à dire de terme général $cov(X_i, X_j)$.

4.2 Définitions et premières propriétés

cf. W chapitre 4.
Définition de la norme ess sup, rappel de la définition des L^p , inégalité de Holder si $1/r = 1/p + 1/q$.
Conséquences : l'application de $L^p \times L^q$ dans L^r qui a (X, Y) fait correspondre le produit est une application bilinéaire continue.

Ici, $\mathbb{P}$ est une mesure finie, on a la suite d'inclusions

$$L^\infty(\Omega, \mathcal{A}, \mathbb{P}) \subset L^p \in (\Omega, \mathcal{A}, \mathbb{P}) \subset L^1(\Omega, \mathcal{A}, \mathbb{P}), 1 \leq p \leq \infty.$$

L'injection canonique de L^p dans $L^q, q \leq p$, est continue.

Théorème de Fisher-Riesz : les L^p sont des Banach.

L'espace vectoriel des variables aléatoires étagées est dense dans tout L^p, de même si Ω est un e.m. séparable, les fonctions continues à support compact sont denses.

Théorème 4.4 *Soit une suite de variables aléatoires (X_n) dans L^p telles que $\sum_n \|X_n\|_p < \infty$. alors il existe $X \in L^p$ telle que la suite $\sum_{k=1}^n X_k$ converge vers X dans L^p et presque sûrement.*

Preuve en TD cf. feuille 3.

Ce théorème sert a montrer

Théorème 4.5 *Soit une suite de variables aléatoires (X_n) dans L^p qui converge vers X dans L^p. Alors, il existe une suite extraite qui converge presque sûrement.*

Preuve en exo, pas faite en amphi.

4.3 Dualité

cf. l'unité B3, mais tous ne la font pas...

Définition 4.6 *Si p et $q \in [1, \infty]$ vérifient $1/p + 1/q = 1$, on dit qu'ils sont conjugués.*

Théorème 4.7 *Si p et q sont conjugués, $Y \in L^q, T : L^p \to \mathbb{R}, T(X) = E(XY)$, est une forme linéaire continue de norme $\|Y\|_q$.*

Preuve en TD

Réciproque partielle :

Théorème 4.8 *Si p et q sont conjugués, $p < \infty$, pour toute forme linéaire continue sur L^p, il existe $Y \in L^q$ tel que $T(X) = E(XY)$.*

preuve admise.

Attention ! ce n'est pas vrai si $p = \infty$!!

4.4 Espace L^2

cours du 10 octobre 2002
Cas particulier de conjugués : $p = q = 2$. On dit que L^2 est "réflexif", c'est de plus un espace de Hilbert, pour le produit scalaire

$$\langle X, Y \rangle = E(XY).$$

4.4.1 Orthogonalité

rappel du théorème de projection sur un sev fermé (cours de licence de topologie), et rappel du théorème de Pythagore. Application de ceci pour montrer

Théorème 4.9 *de RIESZ : Pour toute forme linéaire continue sur L^2, il existe $Y \in L^2$ tel que $T(X) = E(XY)$.*

Preuve en exo en amphi.

définition de l'orthogonalité ds un espace de HILBERT E, si $M \subset E$, def de $M^\perp$, $M \cap M^\perp = \{0\}$,

Théorème 4.10 *Si F sev fermé de E, Hilbert, pour tout $x \in E$, il existe un unique $y \in F$ tel que : $\|x - y\| = \inf\{\|x - z\|, z \in F\}$, et $x - y \in F^\perp$.*

Cet y s'appelle la **projection orthogonale** de x sur F.
Indications succinctes pour la preuve.
Th de Pythagore.
Application à $F = L^2$ et preuve de RIESZ ds ce cas : fait en amphi.
exercice laissé à faire seuls : *ds un Hilbert* E, *si* $M \subset E$, $M^{\perp} = \{0\}$ *si et seulement si le sev engendré par* M *est dense.*

Corollaire 4.11 *les fonctions tages sont denses dans* L^2.

4.4.2 Application : théorème de Radon-Nikodym

Théorème 4.12 *Soient deux mesures bornées μ et ν sur $(\Omega, \mathcal{A})$ telles que ν est absolument continue par rapport à μ ($\forall A \in \mathcal{A}, \mu(A) = 0 \Rightarrow \nu(A) = 0$), il existe $f \in L^1_+(\Omega, \mathcal{A}, \mu)$ telle que*

$$\forall A \in \mathcal{A}, \nu(A) = \int_A f d\mu.$$

Preuve commencée en amphi, à finir en TD.

5 Espérance conditionnelle, loi conditionnelle

fin du 9 octobre 2002

5.1 Cas discret

Soit (X,Y) un vecteur aléatoire de dim 2 à valeurs dans $E \times F \subset \mathbb{N}^2$.

Définition 5.1 *La loi conditionnelle de X sachant $Y = j, j \in F$ est la donnée, pour tout $i \in E$ de la quantité*

$$\mathbb{P}(\{X = i\}/\{Y = j\}).$$

Par extension, on a donc une application sur F

$$g_i : j \mapsto \mathbb{P}(\{X = i\}/\{Y = j\})$$

et on note $\mathbb{P}(\{X = i\}/Y) = g_i(Y), i \in E$.
Ceci est une loi sur E, on peut donc en définir l'espérance appelée
espérance conditionnelle de X sachant Y :

$$E[X/Y] = \sum_{i \in E} i\mathbb{P}(\{X = i\}/Y).$$

5.2 Cas général

On considère un vecteur aléatoire de dim 2 sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace mesurable produit $(E \times F, \mathcal{E} \otimes \mathcal{F})$ de loi notée $\mathbb{P}_{X,Y}$.

Définition 5.2 *On appelle loi conditionnelle de X sachant Y une application notée $\mathbb{P}_{X/Y}$ définie sur $\mathcal{E} \times F$ à valeurs dans $[0,1]$ telle que*

- . $\forall y \in F, B \mapsto \mathbb{P}_{X/Y}(B,y)$ *est proba sur $\mathcal{E}$.*
- . $\forall B \in \mathcal{E}, y \mapsto \mathbb{P}_{X/Y}(B,y)$ *est $\mathcal{F}$-mesurable.*
- . $\forall B \in \mathcal{E}, \forall C \in \mathcal{F}, \mathbb{P}_{X,Y}, \mathbb{P}_{X,Y}(B \times C) = \int \mathbf{1}_C \mathbb{P}_{X/Y}(B,y) d\mathbb{P}_Y(y)$.

Définition 5.3 *Si $\mathbb{P}_{X/Y}$ est la loi conditionnelle de X sachant Y, pour tout $y \in F$, l'application $B \mapsto \mathbb{P}_{X/Y}(B,y)$ s'appelle la loi conditionnelle de X sachant $Y = y$. On la note $\mathbb{P}_{X/Y=y}$.*

5.3 Exemples

X et Y indépendantes

$X = f(Y)$ avec f $(\mathcal{F}, \mathcal{E})$-mesurable

cas discret : X somme des points de deux dés, Y différence.

16 octobre 2002 autre exemple :
(X, Y) admettant une densité sur $E \times F \subset \mathbb{R}^n$.

Puisqu'on a une proba, on peut faire de l'intégration :

Proposition 5.4 *Soit $\mathbb{P}_{X/Y}$ la loi conditionnelle de X sachant Y, f mesurable sur $(E, \mathcal{E})$ et g mesurable sur $(F, \mathcal{F})$, alors*

$$E[f(X)g(Y))] = \int_F g(y) \left(\int_E f(x) d\mathbb{P}_{X/Y=y}(x) \right) d\mathbb{P}_Y(y).$$

Notation : le noyau de conditionnement s'écrit $k(y, dx)$ tel que $\int_B k(y, dx) = \mathbb{P}_{X/Y=y}(B)$ pour tout $B \in \mathcal{E}$.

Définition 5.5 *Un* **noyau de conditionnement** *est une application notée $k(.,.)$ définie sur $\mathcal{E} \times F$ à valeurs dans les probabilités sur $\mathcal{E}$ telle que*

. $\forall y \in F, B \mapsto k(y, B)$ *est proba sur $\mathcal{E}$.*

. $\forall B \in \mathcal{E}, y \mapsto k(y, B)$ *est $\mathcal{F}$-mesurable.*

Il faut penser que, de fait, $k(y, dx)$ représente une loi conditionnelle $\mathbb{P}_{X/Y=y}(y, dy)$. En effet, on remarque que si $\mathbb{P}_Y$ est la loi d'une variable aléatoire Y, il existe une variable aléatoire X telle que la loi du couple (X, Y) est le produit $k(y, dx)\mathbb{P}_Y(dy)$.

Proposition 5.6 *mesurabilité de*

$$y \mapsto \int_E h(x, y) d\mathbb{P}_{X/Y=y} = \int_E h(x, y) k(y, dx).$$

Proposition 5.7 *Soit $\mathbb{P}_{X/Y}$ la loi conditionnelle de X sachant Y, h mesurable sur $(E, \mathcal{E}) \otimes (F, \mathcal{F})$, alors*

$$E[h(X, Y))] = \int_F \left(\int_E h(x, y) d\mathbb{P}_{X/Y=y} \right) d\mathbb{P}_Y(y).$$

Preuve de ces deux théorèmes par les classes monotones à me rédiger s'ils le veulent.....

5.4 Espérance conditionnelle par rapport à une sous tribu

cf. [1] VI.2

5.4.1 Espérance conditionnelle dans L^2

Théorème 5.8 *de projection ds un Hilbert : si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{B}$ sous tribu de $\mathcal{A}$, il existe une unique var dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$ notée $E[X/\mathcal{B}]$ qui réalise l'infimum dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$ de l'application $Z \mapsto E[(X-Z)^2]$. On l'appelle l'espérance conditionnelle de X sachant $\mathcal{B}$.*

C'est la projection orthogonale de X de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$ sui est un sous espace de Hilbert.

Propriétés de l'application $L^2(\Omega, \mathcal{A}, \mathbb{P})$ dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$:$X \mapsto E[X/\mathcal{B}]$ (à montrer en exos, presque toutes corrigées en amphi) :

a) linéaire

b)$\|E[X/\mathcal{B}]\|_2 \leq \|X\|_2$

c) si $X \in L^2(\mathcal{B}), E(X/\mathcal{B}) = X$

d) pour tout $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P}), E(ZX) = E(ZE[X/\mathcal{B}])$.

e) croissante

f) si $\mathcal{C}$ est une sous tribu de $\mathcal{B}, E[E[X/\mathcal{B}]/\mathcal{C}] = E[X/\mathcal{C}]$.

g) si $\sigma(X)$ et $\mathcal{B}$ sont indépendantes, $E[X/\mathcal{B}] = E[X]$.

h)si $Z \in L^\infty(\mathcal{B}), E(ZX/\mathcal{B}) = ZE(X/\mathcal{B})$

5.4.2 Espérance conditionnelle dans L^1

Théorème 5.9 *si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{B}$ sous tribu de $\mathcal{A}$, il existe une unique var dans $L^1(\Omega, \mathcal{B}, \mathbb{P})$ notée encore $E[X/\mathcal{B}]$ qui vérifie $\forall B \in \mathcal{B}, E[1_B X] = E[1_B E(X/\mathcal{B})]$. On l'appelle encore l'espérance conditionnelle de X sachant $\mathcal{B}$.*

Preuve faite en détail

Propriétés de l'application $L^1(\Omega, \mathcal{A}, \mathbb{P})$ dans $L^1(\Omega, \mathcal{B}, \mathbb{P})$:$X \mapsto E[X/\mathcal{B}]$ *à montrer en exos à la maison pour le 22 octobre 2002* :

a) linéaire

b)$\|E[X/\mathcal{B}]\|_1 \leq \|X\|_1$

c) si $X \in L^1(\mathcal{B}), E(X/\mathcal{B}) = X$

d) pour tout $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P}), E(ZX) = E(ZE[X/\mathcal{B}])$.

e) croissante

f) si $\mathcal{C}$ est une sous tribu de $\mathcal{B}, E[E[X/\mathcal{B}]/\mathcal{C}] = E[X/\mathcal{C}]$.

g) si $\sigma(X)$ et $\mathcal{B}$ sont indépendantes, $E[X/\mathcal{B}] = E[X]$.

h) si $Z \in L^\infty(\mathcal{B}), E(ZX/\mathcal{B}) = ZE(X/\mathcal{B})$.

22 octobre 2002

Rgle pratique pour vérifier que $Y = E(X/\mathcal{B})$:

(i) $Y \in L^1(\mathcal{B})$; $\forall B \in \mathcal{B}, E[1_B X] = E[1_B Y]$.

On a une CNS :

Proposition 5.10 *$Y = E(X/\mathcal{B})$ si et seulement si*

(i) $Y \in L^1(\mathcal{B})$; $\forall Z \in L^1(\mathcal{B}), E[ZX] = E[ZY]$.

Preuve en exo tous seuls avec le th des classes monotones.

De plus, l'espérance conditionnelle est une espérance : elle vérifie les théorèmes de Beppo-Levi, de Fatou, de Lebesgue, et l'inégalité de Jensen.
preuves en exos (j'ai corrigé Beppo-Levi et Jensen)
Exercice : $\forall p \geq 1, X \in L^p, \|E(X/\mathcal{B})\|_p \leq \|X\|_p$.

5.5 Retour à $\mathcal{B} = \sigma(Y)$

On a d'une part la loi de X sachant Y et d'autre part $E[X/\mathcal{B}]$ ici $E[X/\sigma(Y)]$. Il s'agit de faire le lien entre les deux. On montre, si g est l'application mesurable $y \mapsto \int_E x\mathbb{P}_{X/Y}(dx, y)$:

$$g(Y) = E[X/\sigma(Y)]$$

avec la caractérisation (i)(ii) et le lemme de Doob.

Définition 5.11 *g s'appelle l'espérance conditionnelle de X sachant Y, notéee $E[X/Y]$.*

Remarque : $E[X/\sigma(Y)] = g(Y) = E[X/Y] \circ Y$, mais souvent on confond $E[X/\sigma(Y)]$ et $E[X/Y]$....

Proposition 5.12 *Si $P_{X/Y}$ est la loi de X sachant Y, et si $f(X, Y) \in L^1$,*

$$E[f(X,Y)/\sigma(Y)] = h(Y) \text{ où } h(y) = \int_E f(x,y)\mathbb{P}_{X/Y}(dx, y)$$

Corollaire 5.13 *$E[f(X,Y)] = E[h(Y)]$.*

5.6 Indépendance conditionnelle

Voir en complément de ce paragraphe les exos 7 9 de la feuille 5.

Définition 5.14 *Soient trois sous-tribus $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}$ de $\mathcal{A}$. On dit que $\mathcal{A}_1$ et $\mathcal{A}_2$ sont conditionnellement indépendantes sachant $\mathcal{B}$ si pour tout $Z_2 \in L^\infty(\mathcal{A}_2), E(Z_2/\mathcal{A}_1 \vee \mathcal{B}) = E(Z_2/\mathcal{B})$. (ou si pour tout $Z_1 \in L^\infty(\mathcal{A}_1), E(Z_1/\mathcal{A}_2 \vee \mathcal{B}) = E(Z_1/\mathcal{B})$)*

On a une CNS de cette définition.

Proposition 5.15 *Soient trois sous-tribus* $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}$ *de* $\mathcal{A}$*.* $\mathcal{A}_1$ *et* $\mathcal{A}_2$ *sont conditionnellement indépendantes sachant* $\mathcal{B}$ *si et seulement si* $\forall Z_1 \in L^\infty(\mathcal{A}_1), Z_2 \in L^\infty(\mathcal{A}_2), E(Z_1 Z_2/\mathcal{B}) = E(Z_1/\mathcal{B})E(Z_2/\mathcal{B})$.

Exercices

1.Si $\mathcal{B} \subset \mathcal{A}_1 \cap \mathcal{A}_2$, et si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont conditionnellement indépendantes sachant $\mathcal{B}$, pour tout $X \in L^1(\mathcal{A}), E[E(X/\mathcal{A}_1)/A_2] = E[E(X/\mathcal{A}_2)/A_1]$.

2. Chane de Markov : $X_{n+1} = f(X_n, W_{n+1})$, le passé est indépendant du futur sachant le présent.

6 Lois gaussiennes et lois associées

6.1 Lois de Gauss

6.1.1 Cas réel

révisions licence

Définition 6.1 *Soit Z une variable aléatoire réelle sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. C'est une variable gaussienne centrée réduite si sa loi admet sur $\mathbb{R}$ la densité $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$. On note cette loi $\mathcal{N}(0, 1)$.*

Une définition équivalente est de donner la fonction caractéristique de cette var : $\phi_Z(t) = e^{-\frac{1}{2}t^2}$.
Exercice : montrer que $Z \in L^p$, $\forall p \geq 0$ et calculer $E[Z^p]$ pour tout $p \geq 0$.

Définition 6.2 *Soit X une variable aléatoire réelle sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. C'est une variable gaussienne s'il existe une var $Z \sim \mathcal{N}(0, 1)$, un réel m et un réel positif σ tels que $X = \sigma Z + m$. On note cette loi $\mathcal{N}(m, \sigma^2)$.*

Remarque 6.3 *Lorsque $\sigma > 0$, on obtient la densité de X par loi image : $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(x-m)^2/\sigma^2}$.*

Exercices : montrer que $\phi_X(t) = e^{itm}e^{-\frac{1}{2}\sigma^2 t^2}$. Montrer que $X \in L^p$, $\forall p \geq 0$.

6.1.2 Vecteurs gaussiens

Définition 6.4 *Soit X un vecteur aléatoire de dimension k sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. C'est un vecteur gaussien si $\forall a \in \mathbb{R}^k$, le produit scalaire de a et X, noté $\langle a, X\rangle$ est une var gaussienne. Si $E[X] = m \in \mathbb{R}^k$ et si la matrice de variance $V(X) = \Gamma$, on note la loi de X : $\mathcal{N}(m, \Gamma)$.*

Exercice : montrer que $V(X)$ est une matrice symétrique définie positive.

Proposition 6.5 *Pour toute application linéaire de $\mathbb{R}^k$ dans $\mathbb{R}^d$ de matrice A, et tout $m \in \mathbb{R}^k$, si X est un vecteur gaussien de dimension k, alors $Y = AX + m$ est un vecteur gaussien de dimension d.*

Proposition 6.6 *Soit X un vecteur gaussien de dimension k, d'espérance $E[X] = m \in \mathbb{R}^k$ et de matrice de variance $V(X) = \Gamma$, alors pour tout $t \in \mathbb{R}^k$, la fonction caractéristique de la loi de X est donnée par*

$$\phi_X(t) = e^{i\langle t,m\rangle}e^{-\frac{1}{2}\tilde{t}\Gamma t}.$$

On a la réciproque :

Proposition 6.7 *Soit X un vecteur aléatoire de dimension k, et de fonction caractéristique*

$$\phi_X(t) = e^{i\langle t,m\rangle} e^{-\frac{1}{2}\tilde{t}\Gamma t}.$$

Alors X est un vecteur gaussien de dimension k, d'espérance $E[X] = m \in \mathbb{R}^k$ et de matrice de variance $V(X) = \Gamma$.

Corollaire 6.8 *Soit X un vecteur gaussien de dimension k, de coordonnées $(X_1, \cdots, X_n)$; les var $(X_1, \cdots, X_n)$ sont indépendantes si et seulement si la matrice de variance de X est diagonale.*

Proposition 6.9 *Soit $m \in \mathbb{R}^k$ et Γ une matrice de dimension $k \times k$ symétrique définie positive. Alors, il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et X un vecteur gaussien de dimension k, d'espérance $E[X] = m \in \mathbb{R}^k$ et de matrice de variance $V(X) = \Gamma$.*

Preuve : elle s'appuie sur le lemme suivant vu en licence

Lemme 6.10 *Soit Γ une matrice de dimension $k \times k$ symétrique définie positive de rang $r \leq k$. Alors il existe une matrice A de taille $(r \times k)$ telle que $\Gamma = \tilde{A}.A$.*

Proposition 6.11 *Soit $m \in \mathbb{R}^k$ et Γ une matrice de dimension $k \times k$ symétrique définie strictement positive. Alors la densité de la loi $\mathcal{N}(m, \Gamma)$ est donnée sur $\mathbb{R}^k$ par :*

$$\frac{1}{(2\phi)^{k/2}\sqrt{det\ \Gamma}} e^{-\frac{1}{2}(x\tilde{-}m)\Gamma^{-1}(x-m)}.$$

Preuve : en exercice.

Exercice : montrer que deux vecteurs gaussiens Y et Z sont indépendants si et seulement s'ils ne sont pas corrélés, c'et à dire que $E[(Y - E(Y))(Z - \tilde{E}(Z))] = 0$.

6.2 Loi du khi-deux

Définition 6.12 *la loi de chi-deux à n degrés de liberté, notée $\chi^2(n)$, est la loi $\Gamma(\frac{n}{2}, \frac{1}{2})$.*

Elle est due à Pearson, matḿaticien londonien (1857-1936).

Théorème 6.13 *de COCHRAN : soit un vecteur gaussien Y de dimension k et de loi $\mathcal{N}(0, I_k)$ et H un sous espace vectoriel de $\mathbb{R}^k$ de dimension $r \leq k$. Alors, le projeté orthogonal $X = P_H(Y)$ est un vecteur gaussien à valeurs dans H de loi $\mathcal{N}(0, I_r)$ et $\|X\|^2$ suit une loi $\chi^2(r)$.*

FISHER, généticien, Londres, 1890-1962.

Théorème 6.14 *de Fisher : soit un vecteur gaussien X de dimension n et de loi $\mathcal{N}(\mu\mathbf{1}, \sigma^2 I_n)$ où $\mathbf{1}$ est le vecteur de coordonnées toutes égales à 1. Alors si on pose*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ ; \ Q = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

le vecteur $(\bar{X}, \frac{1}{\sigma^2}Q)$ est de loi $\mathcal{N}(\mu, \frac{\sigma^2}{n}) \otimes \chi^2(n-1)$.

6.3 Loi de Student

William GOSSET, ingénieur à la brasserie Guinness, a publié sous le pseudonyme de "Student" sous la pression de ses employeurs (OK.... elle est mauvaise.....)

Définition 6.15 *loi de STUDENT : soit $n \geq 1$ et un couple (X, Y) de loi $\mathcal{N}(0,1) \otimes \chi^2(n)$; alors la var $T_n = \frac{X\sqrt{n}}{\sqrt{Y_n}}$ suit une loi de Student à n degrés de liberté.*

Théorème 6.16 *de Student-Fisher : si $(X_1, \cdots, X_n)$ est un $n-$échantillon de la loi $\mathcal{N}(m, \sigma^2)$, alors la statistique T_{n-1} suit une loi de Student à $n-1$ degrés de liberté avec*

$$T_{n-1} = \frac{(\bar{X} - m)\sqrt{n}}{\sqrt{\frac{Q}{n-1}}}$$

6.4 Stabilité des lois gaussiennes par convergence

Grâce au théorème d'équivalence de la convergence en loi et de celle des fonctions caractéristiques, on peut montrer

Proposition 6.17 *Une suite de vecteurs aléatoires de loi $\mathcal{N}(m_n, \Gamma_n)$ converge en loi vers un vecteur gaussien si et seulement si les limites des suites (m_n) et $\Gamma_n)$ existent. Dans ce cas, la loi de X est$\mathcal{N}(m, \Gamma)$ où m et Γ sont les limites des suites (m_n) et $\Gamma_n)$.*

6.5 Lois conditionnelles

Dans le cas de vecteurs gaussiens, la loi conditionnelle d'une composante par une autre est particulièrement simple.

Proposition 6.18 *Soit (X, Y) un vecteur gaussien de dimension 2, d'espérance (μ, ν), de matrice de variance-covariance*

$$\begin{matrix} a^2 & \rho ab \\ \rho ab & b^2 \end{matrix}$$

Alors, la loi conditionelle de X sachant Y est une loi gaussienne $\mathcal{N}(\mu + \rho\frac{a}{b}(Y - \nu), a^2(1 - \rho^2))$.

7 Convergences de mesures bornées, de variables et vecteurs aléatoires

On se place sur l'espace $\mathcal{M}_b^+$ des mesures positives bornées sur $\mathbb{R}^d$.

7.1 Convergence faible

Définition 7.1 *On dit qu'une suite de mesures* (μ_n) *dans* $\mathcal{M}_b^+$ *converge faiblement si* $\forall f \in C_0(\mathbb{R}^d), \mu_n(f) \to \mu(f)$.

7.2 Convergence étroite

Définition 7.2 *On dit qu'une suite de mesures* (μ_n) *dans* $\mathcal{M}_b^+$ *converge étroitement si elle converge faiblement et si de plus* $\mu_n(\mathbb{R}^d) \to \mu(\mathbb{R}^d)$.

Proposition 7.3 *Une suite de mesures* (μ_n) *dans* $\mathcal{M}_b^+$ *converge étroitement si et seulement si* $\forall f \in C_b(\mathbb{R}^d), \mu_n(f) \to \mu(f)$.

Exemple de convergence étroite : si (α_n) est une suite de réels tendant vers 1 et (X_n) une suite de variables aléatoires de loi géométrique de paramètre α_n, alors $(1-\alpha_n)X_n$ converge vers une loi exponentielle (1).

7.3 Application aux probabilités

C'est à dire que l'on s'intéresse aux mesures de masse 1 sur $\mathbb{R}^d$.

Proposition 7.4 *Soit une suite de variables aléatoires réelles de loi* $(\mathbb{P}_n)$ *et de fonction de répartition* F_n. *Alors*

$(\mathbb{P}_n)$ *converge étroitement vers* $\mathbb{P}$ *de fonction de répartition* F *si et seulement si la suite de fonctions* (F_n) *converge vers* F *en tout point où* F *est continue.*

Remarque : ceci dit que la convergence étroite des probas **est** la convergence en loi des var (cf def ds chap 2). De plus, cette équivalence montre que cela revient à la def donnée par Cohen en licence, l'équivalence ayant été admise. On la montre cette année. Cette preuve utilise le :

Théorème 7.5 *(de Paul Lévy, admis en licence) Soit une suite de probabilités* $(\mathbb{P}_n)$ *sur* $(\mathbb{R}^d, \mathcal{B})$ *:*

(a) si cette suite converge étroitement vers $\mathbb{P}$, alors la suite des fonctions caractéristiques converge vers la fonction caractéristique de $\mathbb{P}$ uniformément sur tout compact de $\mathbb{R}^d$.

(b) si la suite fonctions caractéristiques converge vers une fonction ϕ continue à l'origine, alors la suite de probabilités $(\mathbb{P}_n)$ converge étroitement vers $\mathbb{P}$, de fonction caractéristique ϕ.

(a) prouvé le 6

(b) prouvé le 12.

Exercice à faire pour le 13 :

Théorème 7.6 *(de Khinchine) Soit une suite de var (X_k) qui converge en loi vers X, et deux suites $(a_k) \subset \mathbb{R}_+^*$ et $(b_k) \subset \mathbb{R}$, telles que la suite de var $(a_k X_k + b_k)$ converge en loi vers Y, alors $\lim a_k = a$ existe dans $\mathbb{R}_+^*$, $\lim b_k = b$ existe dans $\mathbb{R}$, et $aX + b$ est égale en loi à Y.*

7.4 Notion de tension

cf. [1] pages 136 et sq.

On se place dans ce paragraphe sur un espace topologique muni de ses boréliens, $(E, \mathcal{E})$.

Définition 7.7 *(i) on dit que la proba $\mathbb{P}$ sur $(E, \mathcal{E})$ est* **tendue** *si $\forall \varepsilon > 0, \exists K$ compact tel que $\mathbb{P}(K) > 1 - \varepsilon$.*
(ii) on dit que la famille de probas $(\mathbb{P}_n, n \in \mathbb{N})$ sur $(E, \mathcal{E})$ est (équi- ou uniformément) **tendue** *si $\forall \varepsilon > 0, \exists K$ compact tel que $\forall n \in \mathbb{N},\ \mathbb{P}_n(K) > 1 - \varepsilon$.*
(iii) on dit que la famille de variables aléatoires $(X_n, n \in \mathbb{N})$ à valeurs dans $(E, \mathcal{E})$ est (équi- ou uniformément) **tendue** *si la famille des lois $(\mathbb{P}_{X_n}, n \in \mathbb{N})$ est tendue.*

Proposition 7.8 *S'il existe F fonction positive sur un espace topologique de dim finie, qui tend vers 0 à l'infini et vérifie $\sup_n \int_E F(x)\mathbb{P}_n(dx) \leq C < +\infty$, alors la famille $(\mathbb{P}_n)$ est tendue.*

Théorème 7.9 *(th 4.4 [1]) Si la famille $(\mathbb{P}_n)$ est tendue, elle est relativement compacte pour la topologie de convergence faible.*

Proposition 7.10 *Lorsque'une suite de probabilités $(\mathbb{P}_n)$ converge faiblement vers une mesure μ, on a les équivalences :*
(i) la convergence est étroite,
(ii) la famille $(\mathbb{P}_n)$ est tendue
(iii) la mesure μ est une proba.

7.5 Théorème de Skorohod

Théorème 7.11 *(cf. TD 1 et Billingsley) Si une suite de probas* $(\mathbb{P}_n)$ *converge étroitement vers* μ, *alors il existe un espapace de probabilité* $(\Omega, \mathcal{A}, \mathbb{P})$, *une suite de variables aléatoires* (X_n), *une variable aléatoire* X, *tels que* X_n *est de loi* $\mathbb{P}_n$, X *est de loi* μ, *et* X_n *converge presque sûrement vers* X.

7.6 Convergences des variables aéatoires

7.6.1 Rappel du chapitre 2

convergence en loi équivaut $\mathbb{P}_{X_n}$ converge étroitement vers $\mathbb{P}_X$ équivaut $\mathbb{P}\{X_n \leq x\}$ converge vers $F(x) = \mathbb{P}\{X \leq x\}$ en tout point x où F est continue équivaut la fonction caractéristique de X_n tend vers celle de X qui est continue en 0.

7.6.2 Convergence presque sûre

Définition 7.12 $\mathbb{P}\{\omega : X_n(\omega) \to X(\omega)\} = 1$.

Lemme 7.13 X_n *converge* $\mathbb{P}$ *presque sûrement vers* X *si et seulement si l'une des deux conditions suivantes :*

a) $\forall \epsilon > 0,\ \mathbb{P}[\cap_n \cup_{k \geq n} \{\|X_n - X\| > \varepsilon\} = 0$.

b) $\forall \epsilon > 0,\ \mathbb{P}[\cup_{k \geq n}\{\|X_n - X\| > \varepsilon\} \to_{n\infty} 0$.

7.6.3 Convergence en probabilité

Définition 7.14 $\forall \varepsilon > 0,\ \mathbb{P}\{\omega : \|X_n(\omega) - X(\omega)\| > \varepsilon\} \to 0$.

Proposition annoncée ds le chapitre 2 et prouvée ici :

Proposition 7.15 *a) Si* X_n *converge presque sûrement vers* X, *la convergence est aussi en proba.*

b) Si X_n *converge en proba vers* X, *il existe* $n_k \to \infty$ *telle que* X_{nk} *converge presque sûrement vers* X.

Proposition 7.16 *La convergence en proba est métrisable avec par exemple la distance définie sur les vecteurs aléatoires par* $d(X, Y) = E[\|X - Y\| \wedge]$.

Rappels : si X_n converge en proba vers X, la convergence est aussi en loi.
Réciproquement : si X_n converge en loi vers une constante a, la convergence est aussi en proba.
preuve laissée en exercice.
Pour mémoire, revoir les convergences L^p déjà redites.

7.7 Uniforme, ou équi-, intégrabilité

Barbe et Ledoux pages 125 et sq.

Définition 7.17 *Soit une famille de vecteurs aléatoires* $(X_\alpha, \alpha \in A)$ *; on dit que cette famille est équi-intégrable si*

a) $\sup_\alpha E[\|X_\alpha\|] = M < \infty$.

b) $\forall \varepsilon > 0, \exists \eta : \mathbb{P}(B) \leq \eta,$ *alors* $\sup_\alpha \int_B \|X_\alpha\| d\mathbb{P} \leq \varepsilon$.

Proposition 7.18 *Une famille de vecteurs aléatoires* $(X_\alpha, \alpha \in A)$ *est équi-intégrable si et seulement si*

$$\lim_{C \to +\infty} \sup_\alpha \int_{\{\|X_\alpha\| > C\}} \|X_\alpha\| d\mathbb{P} = 0.$$

Proposition 7.19 *Une suite de vecteurs aléatoires* $(X_\alpha, \alpha \in A)$ *converge en proba vers* X *et est équi-intégrable si et seulement si elle converge dans* $\mathcal{L}^1$.

Exemples et exercices à faire seuls :
une famille finie de va intégrables,
une famille dominée par une va intégrable,
la famille $(E[X/\mathcal{B}_\alpha])$ si $\mathcal{B}_\alpha$ est une famille de sous tribus et X une va intégrable.

8 Théorèmes limites

On considère dans tout ce chapitre une suite de vai de même loi $(X_n, n \geq 0)/$

8.1 Loi faible des grands nombres

Théorème 8.1 *Si les vai (X_n) sont intégrables et d'espérance m, alors $\bar{X}_n$ converge en proba vers m.*

admis sans preuve en licence, prouvé cette année.

8.2 Loi forte des grands nombres

Théorème 8.2 *Si les vai (X_n) sont de carré intégrable et d'espérance m, alors $\bar{X}_n$ converge presque sûrement et dans L^2 vers m.*

admis sans preuve en licence, prouvé cette année.

Barbe et Ledoux montrent, mais c'est beaucoup plus délicat, que intégrable suffit, et que la convergence ps de $\bar{X}_n$ vers une constante montre que les (X_n) sont intégrables.

8.3 Convergence vers la loi de Gauss

Théorème 8.3 *limite central : Si les vai (X_n) sont de carré intégrable de variance σ^2 et d'espérance m, alors $\bar{X}_n$ converge en loi vers la loi gaussienne centrée réduite.*

On a le même résultat ds le cas vectoriel

Proposition 8.4 *Soit une suite de vecteurs aléatoires de dimension d, de même loi, d'espérance m de matrice de variance Γ. Alors $Z_n = \frac{S_n - nm}{\sqrt{n}}$ converge en loi vers une loi gaussienne centrée de matrice matrice de variance Γ.*

Corollaire 8.5 *: formule de MOIVRE-LAPLACE (1718)*

$$\mathbb{P}\{a\sigma/\sqrt{n} \leq \bar{X}_n - m \leq b\sigma/\sqrt{n}\} \sim F(b) - F(a)$$

où F est la fdr de la loi gaussienne centrée réduite.

Exemples et exercices :

(a) On considère un lot d'ampoules dont la durée de vie suit une loi exponentielle de paramètre $1/10$ (en inverse de jours). Quelle est la proba d'user en un an plus de 50 ampoules ?

(b) Un central téléphonique comprnd 5000 abonnés, chacun a la proba de tĺéphoner de 0.2. Quel nombre de lignes faut-il ouvrir pour que la probas que tous les abonnés soient satisfaits dépasse 0.975 ?

(c) Combien faut-il sonder d'électeurs pour avoir une précision de 0.025 pour estimer un pourcentage de "oui" ?

8.4 Autres convergences de variables aléatoires

Proposition 8.6 *Soient* $(Y_i^{(n)}, i = 1, \cdots, n)$ *des vai de Bernoulli de paramètre* (p_n) *tels que* $\lim_{n\to\infty} np_n = \lambda$*. Alors,* $S_n = \sum_i Y_i^{(n)}$ *converge en loi vers la loi de Poisson de paramètre* λ*.*

Après les vacances, je fais encore un cours de proba avec des trucs pour Bakry : la tension d'une famille de probas, le théorème de Skorokhod (sans preuve), la loi du log itéré, j'espère cramer-Chernoff si j'y arrive sans référence.... plus cette limite :

Proposition 8.7 *Soit une suite de vai* T_n *de loi géométrique de paramètre* p_n *tels que* p_n *tend vers* 1*. Alors,* $(1 - p_n)T_n$ *converge en loi vers une exponentielle de paramètre* 1*.*

8.5 Loi du log itéré

cf. [1] page 149.

Théorème 8.8 *Soit une suite de vai de même loi* $(X_n), E(X_n) = m, V(X_n) = \sigma^2$, *alors, presque sûrement :*

$$\overline{\lim_n} \frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{2n \log\log n}} = 1,$$

$$\underline{\lim_n} \frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{2n \log\log n}} = -1.$$

Commentaire : ce résultat dit que la suite $\frac{X_1+\cdots+X_n-nm}{\sigma\sqrt{2n\log\log n}}$ admet pour valeurs d'adhérence l'intervalle entier $[-1, +1]$. Par ailleurs, on a là un exemple de profonde différence entre convergence en loi et convergence presque sûre : la suite $Z_n = \frac{X_1+\cdots+X_n-nm}{\sigma\sqrt{n}}$ converge en loi vers la loi $\mathcal{N}(/, \infty)$. Mais il n'existe pas de var Z de la loi $\mathcal{N}(/, \infty)$ qui soit limite presque sûre de (Z_n) : si cela était, on aurait $Z_n/\sqrt{2\log\log n}$ tend presque sûrement vers 0, ce qui contredirait la loi du log itéré : donc, Z_n ne converge pas presque sûrement.

References

[1] Ph. BARBE et M. LEDOUX : “Probabilités, de la licence à l’agrégation”, Belin, Paris, 1998.

[2] M. COTTRELL, V. GENON-CATALOT, C. DUHAMEL, T. MEYRE : “Exercices de probabilités”, Cassini, Paris, 1999.

[3] D. DACUNHA-CASTELLE et M. DUFLO : “Probabilités et statistique”, Masson, Paris, 1982 ; et le livre d’exercices qui va avec.

[4] P. JAFFARD : “Initiation aux méthodes de la statistique et du calcul des probabilités”, Masson, Paris, 1978

[5] P. JAFFARD : “Initiation aux méthodes de la statistique et du calcul des probabilités”, Masson, Paris, 1978 (réédité en 1996).

[6] J. NEVEU : “Bases mathématiques du calcul des probabilités”, Masson, Paris, 1970.

[7] J. NEVEU : “Cours de probabilités à l’école Polytechnique”.

[8] G. SAPORTA : “Probabilités, analyse des données et statistique”, ed. Technip, Paris, 1990.