Anne-Marie CHARBONNEL

ELEMENTS DE PROBABILITÉS

UNIVERSITÉ DE NANTES
IREM des Pays de Loire Janvier 2001

Ces notes sont destinées aux étudiants à l'Agrégation Externe de Mathématiques. Elles ont servi de support aux cours intensifs faits à Nantes en Septembre 99 et en Septembre 2000. Elles constituent une synthèse des notions au programme du concours si on ne présente pas l'option "probabilités et statistiques" à l'oral.

Elles doivent beaucoup à Didier Robert, avec qui j'ai enseigné en Licence plusieurs années, et à qui j'ai emprunté nombre de ses présentations. Merci aussi à Xavier Saint Raymond pour sa relecture attentive.

L'auteur tient également à remercier l'IREM des Pays de Loire, qui a assuré l'impression et la diffusion de ce polycopié.

Chapitre 1 : Espaces probabilisés, variables aléatoires.

I - Intégrale (abstraite) de Lebesgue

I-1 - Espaces probabilisés

On considère un espace probabilisé $(\Omega, \mathcal{A}, P)$, c'est à dire un ensemble Ω muni d'une tribu $\mathcal{A}$, sur lequel on définit une mesure de masse totale égale à 1.

Rappels :

* Une tribu sur un ensemble est une famille de parties stable par passage au complémentaire, comprenant l'ensemble lui-même (donc l'ensemble vide), stable par unions dénombrables (donc par intersections dénombrables).

* Une mesure (positive) P est une application de $\mathcal{A}$ dans $\mathbf{R}_+$ telle que $P(\emptyset) = 0$, et qui est σ-additive, c'est à dire qu'elle vérifie : $P(\cup_{n\in\mathbf{N}} A_n) = \sum_{n\in\mathbf{N}} P(A_n)$, pour toute suite d'éléments de $\mathcal{A}$ deux à deux disjoints.

Une mesure de probabilités vérifie en outre $P(\Omega) = 1$. Elle a donc en particulier toutes les propriétés des mesures finies.

Un tel espace permet de modéliser une expérience probabiliste concrète :

Ω est l'ensemble des tous les résultats possibles. Un élément ω de Ω s'appelle un résultat, ou évènement élémentaire. $\mathcal{A}$ est la famille des évènements. Ce sont les parties mesurables de Ω, celles auxquelles on pourra attribuer une probabilité. Un évènement est dit réalisé s'il contient le résultat de l'expérience. Ω est l'évènement certain, puisqu'il est toujours réalisé.

Exemples

1er Exemple : On joue à pile ou face.

a) **Première expérience : on lance une fois une pièce de monnaie.**

Cette expérience n'a que deux résultats possibles (on dit encore deux issues). On la modélise donc par un ensemble à deux éléments : $\Omega = \{P, F\}$, ou $\Omega = \{0, 1\}$. Sur un ensemble de ce genre, qui est discret fini, il suffit de calculer la probabilité de chaque élément pour connaître la probabilité de chaque partie de Ω. Si on désigne par p, $0 < p < 1$ la probabilité d'avoir "pile", on a : $P(P) = p$, $P(F) = 1 - p$, et évidemment $P(\Omega) = 1$ et $P(\emptyset) = 0$. P est ainsi définie sur toutes les parties de Ω : $\mathcal{A} = \mathcal{P}(\Omega)$.

b) **Deuxième expérience : on lance n fois de suite la même pièce de monnaie.**

On peut décrire l'expérience de plusieurs façons possibles. D'ailleurs on n'a jamais unicité du modèle! On peut ou bien décrire un résultat en disant ce qu'on a eu à chaque lancer, ou bien se contenter de dire combien de fois on a obtenu "pile".

Dans le premier cas l'ensemble Ω_1 de tous les résultats possibles est celui des n-uplets dont les composantes sont égales à P ou F (ou encore 0 ou 1). Comme c'est un ensemble discret fini, on calcule la probabilité de chaque élément, et on en déduit l'espace probabilisé $(\Omega_1,\ \mathcal{P}(\Omega_1),\ P_1)$ suivant :

$$(1)\quad \Omega_1 \;=\; \{\omega = (\omega_1, \ldots, \omega_n)\ ;\ \omega_i = P \text{ ou } F \quad \text{pour} \quad i = 1, \ldots, n\ \}$$

$$\text{avec} \quad P(\omega) = p^k(1-p)^{n-k}$$

k étant le nombre des composantes de ω égales à P.

. **Cas particulier :** $p = 1/2$.
Dans ce cas tous les éléments ont même probabilité : $P(\omega) = (1/2)^n$. Ce modèle est appelé "espace probabilisé de Laplace". C'est un ensemble discret fini, dont tous les éléments sont équiprobables. La probabilité d'un évènement se calcule alors aisément : $P(A) = \frac{cardA}{card\Omega}$.

Dans le second cas, un résultat est déterminé par deux nombres, le nombre des "pile" et celui des "face", dont la somme est évidemment n. On ne tient plus compte de l'ordre dans lequel ils sont obtenus. On obtient ainsi l'espace probabilisé $(\Omega_2,\ \mathcal{P}(\Omega_2),\ P_2)$ suivant :

$$(2)\quad \Omega_2 \ = \ \{\omega = k, l\ ;\ k, l \in \{0, 1, \ldots, n\} \text{ avec } l = n - k\}$$
$$\text{avec} \quad P(\omega) = C_n^k\ p^k (1-p)^{n-k}$$

k étant le nombre des piles obtenus.

c) **Troisième expérience : on joue à pile ou face jusqu'à obtenir face.**
Un résultat est alors une suite formée de P et de F. La probabilité d'un résultat ω est égale à :

$$(3)\quad P(\omega) = p^k(1-p) \quad ,$$

k étant le nombre de jets avant celui qui a donné face.
L'ensemble des résultats possibles est cette fois un ensemble discret infini.
Exercice 1.1 Vérifier qu'on a bien construit ainsi un espace probabilisé.
Quelle est la probabilité que le jeu ne s'arrête pas ?

Deuxième exemple Expérience : On place devant soi une feuille de papier, et on suppose être muni d'un crayon dont la pointe est un point matériel.
L'expérience consiste à piquer au hasard la pointe du crayon sur la feuille. On suppose que deux parties isométriques de la feuille ont la même chance d'être atteintes.
Exercice 1.2
1) Quelle est la probabilité de tomber :
a) dans la moitié supérieure de la feuille ?
b) dans une partie dessinée à l'avance ?
2) Modéliser cette expérience.
3) Quelle est la probabilité de tomber en un point particulier de la feuille ?
Ce dernier évènement est improbable, mais non pas impossible !

I-2 - Intégrales de variables aléatoires

Définition 1.1 : *On appelle "variable aléatoire" sur l'espace probabilisable $(\Omega, \mathcal{A})$, à valeurs dans $\mathbf{R}$, ou $\mathbf{R}^n$, ou $\mathbf{C}$, toute fonction X mesurable de $(\Omega, \mathcal{A})$ dans $\mathbf{R}$, ou $\mathbf{R}^n$, ou $\mathbf{C}$, ie vérifiant $X^{-1}(B) \in \mathcal{A}$ pour tout B borélien de $\mathbf{R}$, $\mathbf{R}^n$, ou $\mathbf{C}$.*
On peut donc parler de l'intégrabilité d'une variable aléatoire sur l'espace Ω par rapport à la mesure P. La construction de l'intégrale se fait selon le schéma suivant

Définition 1.2 : *On appelle "fonction étagée"(ou fonction simple) sur $(\Omega, \mathcal{A})$ toute combinaison linéaire de fonctions caractéristiques d'éléments de $\mathcal{A}$, c'est à dire toute fonction de Ω dans $\mathbf{C}$ s'écrivant sous la forme :*

$$(4)\quad f \ = \ \sum_{i=1}^{N} \lambda_i \mathbf{1}_{A_i} \quad \text{avec } \lambda_i \in \mathbf{C} \quad \text{et} \quad A_i \in \mathcal{A} \quad .$$

Remarques :

1 f ne peut prendre qu'un nombre fini de valeurs.

2 L'écriture de f comme dans (4) n'est pas unique. On a une écriture canonique si on suppose que tous les λ_i sont distincts, et que les A_i forment une partition de Ω.

Définition 1.3 : *Si f est une fonction étagée sur Ω, on appelle "intégrale de f sur Ω par rapport à la probabilité P", ou moyenne de f, ou "espérance" de f, le nombre complexe :*

$$(5) \quad \int_\Omega f \, dP = \sum_{i=1}^{N} \lambda_i P(A_i) \quad .$$

Remarque : ce nombre est défini sans ambiguïté puisque, P étant une mesure finie, on a une somme algébrique de quantités finies. Par ailleurs, il ne dépend pas de l'écriture de f sous la forme (4) (à vérifier).

Proposition 1.1 *Toute variable aléatoire positive est limite partout d'une suite croissante de fonctions étagées positives.*

On peut alors définir l'intégrale d'une variable aléatoire X à valeurs dans $[0, +\infty]$:

Définition 1.4 : *Soit X une variable aléatoire sur Ω, à valeurs dans $[0, +\infty]$; on définit l'intégrale de X sur Ω de la façon suivante :*

$$(6) \quad \int_\Omega X \, dP = \sup \{ \int_\Omega f \, dP \} \quad ,$$

où f décrit l'ensemble des fonctions étagées positives, telles que l'on ait : $f \leq X$.

Le nombre défini par (6) est fini ou infini.

Pour pouvoir travailler avec des intégrales de fonctions à valeurs dans $[0, +\infty]$, on fait les conventions suivantes:

$a + \infty = \infty + a = +\infty \quad \forall\, a \ : 0 \leq a \leq +\infty$.

$a \times \infty = \infty \times a = +\infty \quad \forall\, a \ : 0 < a \leq +\infty$.

$0 \times \infty = \infty \times 0 = 0$.

$a + b = a + c \Rightarrow b = c \quad \forall\, a \ : 0 \leq a < +\infty$.

$a \times b = a \times c \Rightarrow b = c \quad \forall\, a \ : 0 < a < +\infty$.

Définition 1.5 : *Si X est une variable aléatoire sur Ω, à valeurs dans $\mathbf{C}$, on dit que X est intégrable sur Ω, ou possède un moment d'ordre 1, si l'on a : $\int_\Omega |X| \, dP < +\infty$.*

On peut alors définir son intégrale sur Ω comme l'unique forme linéaire vérifiant ce qui précède, c'est à dire:

$$(7) \quad \int_\Omega X \, dP = \int_\Omega u^+ \, dP - \int_\Omega u^- \, dP + i \int_\Omega v^+ \, dP - i \int_\Omega v^- \, dP \quad ,$$

où u et v sont respectivement les parties réelle et imaginaire de X, et u^+ (resp. u^-), désigne la partie positive (resp. négative) de u.

Proposition 1.2 *Soit X une variable aléatoire. Les 2 assertions suivantes sont équivalentes*

1) $E(|X|) < +\infty$.

2) X est limite simple d'une suite de fonctions étagées f_n, de Cauchy pour la norme $\mathcal{L}^1$.

On a alors $E(X) = \lim_{n\to\infty} E(f_n)$, la limite étant indépendante de la suite f_n.

Exemples

1) Intégrale d'une variable aléatoire par rapport à la mesure de Dirac en un point

2) Intégrale d'une variable aléatoire par rapport à une mesure du type $p_1\delta_{\omega_1} + p_2\delta_{\omega_2}$

3) Intégrale d'une variable aléatoire par rapport à la mesure uniforme sur un pavé de $\mathbf{R}^n$

I-3 - Lois et moments de variables aléatoires

Dans la suite on notera souvent v. a. pour variable aléatoire.

Définition 1.6 *Si X est une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$ et à valeurs dans $\mathbf{R}$ (resp $\mathbf{R}^n$ ou $\mathbf{C}$), on appelle loi de X la mesure image de P par X, c'est à dire la mesure ν_X sur $(\mathbf{R}, \mathcal{B}_\mathbf{R})$ (resp $(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n})$ ou $(\mathbf{C}, \mathcal{B}_\mathbf{C})$) définie par : $\nu_X(B) = P(X^{-1}(B))$, pour $B \in \mathcal{B}_\mathbf{R}$ (resp $\mathcal{B}_{\mathbf{R}^n}$ ou $\mathcal{B}_\mathbf{C}$).*

On a alors la proposition suivante :

Proposition 1.3 *f est une fonction ν_X - intégrable si, et seulement si, $f \circ X$ est P-intégrable, et on a alors:*

$$\int_Y f d\nu_X = \int_\Omega (f \circ X)\, dP$$

où Y désigne $\mathbf{R}$ (resp $\mathbf{R}^n$ ou $\mathbf{C}$).

Exemples

Dans les expériences du premier exemple, on considère la v. a. X égale au nombre de "pile" obtenu. Cette variable prend un nombre discret de valeurs (fini dans les cas a) et b), infini dans le cas c)). On dit que c'est une v. a. discrète.

Exercice 1.3

Vérifier que, dans les 3 cas, la loi de X s'écrit sous la forme $\nu_X = \sum_{k\in\mathbf{N}} a_k\, \delta_k$ où δ_k est la mesure de Dirac au point $k \in \mathbf{N}$ et où les a_k sont des coefficients que l'on calculera. Calculer leur somme.

(L'écriture $\mu = \sum_{k\in\mathbf{N}} a_k\, \mu_k$ signifie simplement que, pour toute fonction f μ-intégrable, on a : $\int f d\mu = \sum_{k\in\mathbf{N}} a_k \int f d\mu_k$, écriture formelle que l'on pourra traduire plus tard en termes de limite faible).

La v. a. de l'exercice 1.3 est dite suivre la loi de Bernoulli dans le cas a), la loi binomiale dans le cas b), et dans le cas c) la loi géométrique.

Définition 1.7 : Moments d'une variable aléatoire réelle

Soit p un entier supérieur ou égal à 1, et X une variable aléatoire à valeurs dans $\mathbf{R}$ et appartenant à $L^p(\Omega)$. On appelle moment d'ordre p de X sa moyenne d'ordre p sur Ω.

Le moment d'ordre 1 s'appelle l'espérance de X (ou sa moyenne), et on le note $E(X)$.

On appelle variance de la v. a. X la moyenne quadratique des écarts de X par rapport à sa moyenne : $V(X) = E(X - E(X))^2$.

On vérifiera que l'on a $V(X) = E(X^2) - (E(X))^2$.

Pour des v. a. X et Y appartenant à $L^2(\Omega)$, on appelle covariance de X et de Y la quantité

$$cov(X,Y) \;=\; E\big((X-E(X)).(Y-E(Y))\big) \;=\; E(XY)-E(X)E(Y) \quad .$$

Cas des vecteurs aléatoires Si X est à valeurs dans $\mathbf{R}^n$, c'est à dire si c'est un vecteur aléatoire, on dit qu'elle est dans $L^p(\Omega)$ si chacune de ses composantes s'y trouve. Sa moyenne d'ordre p est alors le vecteur des moyennes d'ordre p de chaque composante.

Si X appartient à $L^2(\Omega)$, sa (matrice carrée de) covariance est égale à :

$$cov(X) \;=\; [E\big((X_i-E(X_i)).(X_j-E(X_j))\big)]_{1\leq i,j\leq n}$$

Remarque Une variable aléatoire est très souvent donnée par l'ensemble de ses valeurs et sa loi. L'espace probabilisé sur lequel elle vit n'a pas besoin d'être précisé, mais on y fait référence, si besoin est, pour exprimer certaines intégrales.

Exemples et exercices

Exercice 1.4 : Inégalité de Bienaymé Tchebichev

Soit X une v. a. sur un espace $(\Omega,\mathcal{A},P)$ à valeurs dans $\mathbf{R}$. Vérifier que la probabilité qu'elle s'écarte de sa moyenne de plus de ϵ est majorée par $\frac{V(X)}{\epsilon^2}$.

Remarque Cette inégalité, qui s'obtient sans effort, n'est pas très précise, mais elle permet tout de même d'obtenir la loi (faible) des grands nombres.

Exercice 1.5

Calculer l'espérance et la variance des variables de Bernoulli, binomiale et géométrique après avoir vérifié leur existence.

Exercice 1.6

On appelle v. a. de Poisson de paramètre $\lambda>0$ une v. a. à valeurs entières dont la loi est donnée par : $\nu \;=\; \sum_{k\in\mathbf{N}} p_k\delta_k$ avec $p_k=e^{-\lambda}\frac{\lambda^k}{k!}$.

Montrer que cette v. a. est de carré sommable et calculer son espérance et sa variance.

Exercice 1.7

a) Loi exponentielle

Soit X une variable aléatoire suivant la loi exponentielle de paramètre $\lambda>0$, c'est à dire que la loi de X admet par rapport à la mesure de Lebesgue sur $\mathbf{R}$ la densité de probabilité

$$f(x)=\lambda \mathrm{e}^{-\lambda x}\mathbf{1}_{]0,+\infty[}$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de l'ensemble A.

Déterminer la moyenne et la variance de X.

b) Loi de Gauss (ou normale) sur $\mathbf{R}$

Soit X une variable aléatoire admettant par rapport à la mesure de Lebesgue sur $\mathbf{R}$ la densité de probabilité

$$g(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad .$$

où σ est un réel strictement positif et μ un réel.

Déterminer la moyenne et la variance de X.

Exercice 1.8 : une démonstration du théorème de Weierstrass

Référence : l'ouvrage de Loeve par exemple, mais aussi Foata et Fuchs p 234.

On considère une suite de variables aléatoires X_n telle que

$$P[X_n = k] = \binom{n}{k} x^k (1-x)^{n-k} = b_{k,n}(x) \quad \text{avec} \quad x \in [0,1] \quad .$$

1) Déterminer l'espérance mathématique $E(X_n)$ et la variance $V(X_n)$.

2) Soit f une fonction continue, réelle, définie sur $[0,1]$.

Montrer que

$$E[f(\frac{X_n}{n})] = \sum_{0 \leq k \leq n} f(\frac{k}{n})\ b_{k,n}(x)$$

3) En utilisant l'inégalité de Bienaymé - Tchebichev, montrer que pour $\delta > 0$ on a

$$\sum_{|k-nx|>n\delta} b_{k,n}(x) \leq \frac{1}{n^2\delta^2} \sum_{0 \leq k \leq n} (k-nx)^2\ b_{k,n}(x)$$

4) On pose

$$P_n(x) = \sum_{0 \leq k \leq n} f(\frac{k}{n})\ b_{k,n}(x)$$

En utilisant ce qui précède montrer que la suite de polynômes P_n converge uniformément vers f sur $[0,1]$.

Exercice 1.9 : fonctions génératrices

On considère une variable aléatoire X à valeurs dans $\mathbf{N}$, de loi définie par la suite $p_k = P([X = k])$.

1) Montrer que l'égalité $G_X(z) = \sum_{k\in\mathbf{N}} p_k z^k$ définit dans $\mathbf{C}$ une série entière de rayon de convergence ≥ 1, appelée fonction génératrice de la loi de probabilité $\{p_k\}$.

Montrer que $G_X(z) = E(z^X)$ où $E(Y)$ désigne l'espérance mathématique de la v.a Y.

2) A l'aide de la formule de Cauchy exprimer p_k en fonction de G_X.

3) Déterminer les fonctions génératrices d'une loi binomiale de paramètres (n,p) et d'une loi de Poisson de paramètre λ.

4) On suppose que le rayon de convergence de la série définissant G_X est strictement plus grand que 1. Montrer alors que X admet une espérance mathématique et une variance que l'on exprimera en fonction de $G'_X(1)$ et $G''_X(1)$.

5) Soit $\{X_n\}_{n\in\mathbf{N}}$ une suite de variables aléatoires à valeurs dans $\mathbf{N}$, de loi de probabilité $P[X_n = k] = p_{n,k}$ et X de loi $P[X = k] = p_k$.

Montrer que $\lim_{n\to+\infty} p_{n,k} = p_k, \quad \forall k \in \mathbf{N}$ si, et seulement si, les fonctions génératrices vérifient

$\lim_{n\to+\infty} G_{X_n}(z) = G_X(z)\ \forall z \in \mathbf{C}\ |z| \leq 1$.

6) Utiliser ce qui précède pour montrer que si X_n suit la loi binomiale $(n, \frac{\lambda}{n})$ alors

$$\lim_{n\nearrow+\infty} P[X_n = k] = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}, \quad \forall k \in \mathbf{N}$$

N.B. Ce dernier résultat peut s'obtenir directement sans l'aide des fonctions génératrices!

I-4 - Fonctions caractéristiques

On vient de voir dans l'exercice 1.9 que, pour une variable discrète, la donnée de la fonction génératrice contient toutes les informations. Les fonctions caractéristiques jouent le même rôle, pour tout type de variables aléatoires.

Définition 1.8 : Fonction caractéristique d'une mesure positive bornée

Soit μ une mesure positive bornée sur $\mathbf{R}$ (resp $\mathbf{R}^n$). Alors pour tout t appartenant à $\mathbf{R}$ (resp $\mathbf{R}^n$), la fonction $x \longrightarrow \mathrm{e}^{itx}$ (resp $x \longrightarrow \mathrm{e}^{i\langle t,x\rangle}$) est μ-intégrable sur $\mathbf{R}$ (resp $\mathbf{R}^n$), et on appelle fonction caractéristique de μ la fonction Φ_μ définie par

$$\Phi_\mu(t) = \int_{\mathbf{R}} \mathrm{e}^{itx}\, d\mu(x) \quad \big(\text{resp } \Phi_\mu(t) = \int_{\mathbf{R}^n} \mathrm{e}^{i\langle t,x\rangle}\, d\mu(x)\big).$$

Définition 1.9 : Fonction caractéristique d'une variable aléatoire

On appelle fonction caractéristique de la variable aléatoire X à valeurs dans $\mathbf{R}$ (resp $\mathbf{R}^n$) la fonction caractéristique de sa loi, c'est à dire la fonction définie sur $\mathbf{R}$ (resp $\mathbf{R}^n$) par $t \longrightarrow \Phi_X(t) = E(\mathrm{e}^{i\langle t,X\rangle})$, où $\langle t, X\rangle$ désigne la fonction qui à $\omega \in \Omega$ associe le produit scalaire de t et de $X(\omega)$ (produit ordinaire si $n = 1$).

Remarque : Le coefficient de $\langle t, X\rangle$ dans l'exponentielle est variable suivant les auteurs, comme c'est le cas pour la définition de la transformée de Fourier. D'ailleurs, si X est une v. a. à densité, sa fonction caractéristique au point t est égale, avec les conventions nantaises, à la transformée de Fourier de cette densité prise au point $-t$.

Exercice 1.10

Calculer $\Phi_X(t)$ pour les variables aléatoires rencontrées dans les exercices 1.5, 1.6 et 1.7.

Proposition 1.4 Premières propriétés des fonctions caractéristiques

Soit X une variable aléatoire à valeurs dans $\mathbf{R}^n$. Sa fonction caractéristique vérifie :

1) $|\Phi_X(t)| \leq 1 \quad \forall t \in \mathbf{R}^n$, $\Phi_X(0) = 1$ et $\Phi_X(-t) = \overline{\Phi_X(t)}$ ($Re\Phi_X$ est paire, et $Im\Phi_X$ impaire).

2) Φ_X est uniformément continue sur $\mathbf{R}^n$.

Preuve :

1) provient des propriétés de l'intégrale ; on obtient immédiatement aussi que Φ_X est continue comme intégrale dépendant d'un paramètre.

2) Démonstration de la continuité uniforme :

$$|\Phi_X(t_1) - \Phi_X(t_2)| \leq \int_{\mathbf{R}^n} |\mathrm{e}^{i\langle t_1,x\rangle} - \mathrm{e}^{i\langle t_2,x\rangle}|\, dP_X(x)\ ,$$

où P_X désigne la loi de la variable aléatoire X.

$$|\Phi_X(t_1) - \Phi_X(t_2)| \leq \int_{|x|\leq R} + \int_{|x|>R}\ .$$

Or l'inégalité des accroissements finis permet d'écrire $|e^{i\langle t_1,x\rangle} - e^{i\langle t_2,x\rangle}| \leq |t_1-t_2||x|$, ce qui conduit à :

$$|\Phi_X(t_1) - \Phi_X(t_2)| \leq R|t_1-t_2| + 2\int_{|x|>R} dP_X(x) \quad .$$

Soit maintenant $\epsilon > 0$. Il existe $R_\epsilon > 0$ tel que $2\int_{|x|>R_\epsilon} dP_X(x) \leq \frac{\epsilon}{2}$, puis $\eta \leq \frac{\epsilon}{2R_\epsilon}$ tel que $|t_1-t_2| \leq \eta \Rightarrow |\Phi_X(t_1) - \Phi_X(t_2)| \leq \epsilon$.

Proposition 1.5 Liens entre fonctions caractéristiques et moments
Soit X une variable aléatoire à valeurs dans $\mathbf{R}$.

1) On suppose que X admet des moments d'ordre p pour $p \leq m$, où $m \in \mathbf{N}$. Alors la fonction Φ_X est m fois continûment dérivable sur $\mathbf{R}$ et on a :

$$\Phi_X^{(k)}(t) = i^k\, E(X^k e^{itX}) \quad \text{pour } k \leq m \quad .$$

En particulier on a :

$$\Phi_X^{(k)}(0) = i^k\, E(X^k) \quad \text{pour } k \leq m \quad .$$

Pour une v. a. à valeurs dans $\mathbf{R}^n$, on a une propriété analogue, que l'on prouve en regardant les dérivées partielles.

2) Réciproquement, supposons que Φ_X soit m fois dérivable en 0. Alors X admet des moments jusqu'à l'ordre $2k$, avec $2k \leq m$.

Preuve :

1) est une conséquence immédiate du théorème de dérivation des intégrales.

2) Supposons d'abord $m = 2$. Donc $\Phi_X^{(2)}(0)$ existe, et on peut remarquer qu'elle est égale à $Re\Phi_X^{(2)}(0)$. (ou Re désigne la partie réelle).

On peut écrire

$$1 - Re\Phi_X(t) = \int (1-\cos tx)\, dP_X(x) \quad .$$

On démontre aisément la relation

$$|tx| \leq 1 \Rightarrow 1-\cos tx \geq \frac{(tx)^2}{3} \quad .$$

Il suffit d'écrire $1-\cos u \geq \frac{u^2}{2} - \frac{u^4}{24}$ pour $|u| \leq 1$ et de minorer. Alors il vient

$$1 - Re\Phi_X(t) \geq \int_{-\frac{1}{t}}^{\frac{1}{t}} (1-\cos tx)\, dP_X(x) \geq \int_{-\frac{1}{t}}^{\frac{1}{t}} \frac{(tx)^2}{3}\, dP_X(x) \quad ,$$

et donc

$$\int_{-\frac{1}{t}}^{\frac{1}{t}} x^2\, dP_X(x) \leq 3\frac{1-Re\Phi_X(t)}{t^2} \quad .$$

Comme $Re\Phi'_X(0)$ est nul, $\frac{1-Re\Phi_X(t)}{t^2}$ converge vers $-\frac{Re\Phi_X^{(2)}(0)}{2}$, et en invoquant par exemple le théorème de Beppo-Levi, on en déduit que X est de carré sommable.

Pour m quelconque, on fait la démonstration par récurrence sur m, en passant de l'existence du moment d'ordre 2k à celui d'ordre (2k+2) pour tout k tel que $2k+2 \leq m$: Il s'agit en effet de montrer que X admet un moment d'ordre 2p+2, où p est le plus grand entier tel que $2p+2 \leq m$.
Comme pour le cas $m = 2$ on écrit :

$$Re[\Psi(0) - \Psi(t)] = \int (1 - \cos tx)\, x^{2k}\, dP_X(x) \quad \text{avec} \quad \Psi(t) = \int e^{itx}\, x^{2k}\, dP_X(x) \quad .$$

En effet, d'après l'hypothèse de récurrence, l'intégrale $\int (1 - \cos tx)\, x^{2k}\, dP_X(x)$ existe bien.
On en déduit alors, comme pour le cas $m = 2$, que l'on a :

$$\int_{-\frac{1}{t}}^{\frac{1}{t}} x^{2k+2}\, dP_X(x) \leq 3\frac{Re[\Psi(0) - \Psi(t)]}{t^2} \quad .$$

Comme $E(X^{2k})$ existe, Φ admet des dérivées partout jusqu'à l'ordre 2k, et on a $\Phi^{(2k)}(t) = (-1)^k\Psi(t)$, de sorte que l'existence de la dérivée (2k+2)ième en 0 assure l'existence de la dérivée seconde de Ψ en 0. Comme $Re\Psi'(0) = 0$, on termine exactement comme pour le cas $m = 2$.

Proposition 1.6 Inversion des fonctions caractéristiques

On se contente dans ce qui suit d'énoncer le résultat pour des v. a. à valeurs réelles.

Soit X une v. a. dont on note P_X la loi, et F_X la fonction de répartition, c'est à dire la fonction définie sur $\mathbf{R}$ par $F_X(t) = P(\{X < t\}) = \int_{\{x<t\}} dP_X$. F_X ainsi définie est continue à gauche. La donnée de F_X détermine complètement la mesure P_X, c'est à dire la loi de X. Soit Φ_X la fonction caractéristique de X. Notons $\widetilde{F_X}(t) = \frac{F_X(t+0)+F_X(t-0)}{2} = \frac{F_X(t+0)+F_X(t)}{2}$.

Alors, quels que soient a et b dans $\mathbf{R}$, on a la formule d'inversion :

$$\widetilde{F_X}(b) - \widetilde{F_X}(a) = \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{-it}\, \Phi_X(t)\, dt \quad .$$

Autrement dit, si 2 v. a. ont même fonction caractéristique, elles ont la même loi.

Remarque On a une formule analogue dans le cas de la dimension n.

Preuve :

Pour simplifier on suppose $a = 0$ et $b > 0$. Le cas général se traite de la même manière.

On peut alors écrire :

$$\int_{-T}^{T} \frac{e^{-itb} - 1}{-it}\, \Phi_X(t)\, dt = \int_{-T}^{T} \frac{e^{-itb} - 1}{-it}\, [\int_{\mathbf{R}} e^{itx}\, dP_X(x)]\, dt \quad .$$

On peut utiliser le théorème de Fubini, ce qui donne :

$$\int_{-T}^{T} \frac{e^{-itb}-1}{-it} \Phi_X(t)\, dt = \int_{\mathbf{R}} [\int_{-T}^{T} \frac{e^{-itb}-1}{-it} e^{itx}\, dt]\, dP_X(x)$$

$$= 2\int_{\mathbf{R}} [\int_{T(x-b)}^{Tx} \frac{\sin u}{u}\, du]\, dP_X(x) \quad .$$

En effet, l'expression $\frac{e^{-itb}-1}{-it}$ e^{itx} peut s'écrire comme intégrale de e^{ity} entre $x-b$ et x, et via le théorème de Fubini, cela conduit à l'expression ci-dessus.

Comme $Sup_T|\int_{T(x-b)}^{Tx} \frac{\sin u}{u}\, du|$ est borné indépendamment de x, on peut utiliser le théorème de convergence dominée.

L'expression $\int_{T(x-b)}^{Tx} \frac{\sin u}{u}\, du$ admet une limite, quand T tend vers l'infini, pour tous x et b appartenant à $\mathbf{R}$, qui est égale à : $l(x,b) = \pi$ pour $0 < x < b$, $\frac{\pi}{2}$ pour $x = 0$ ou $x = b$, 0 sinon.

On en déduit alors :

$$\lim_{T\to\infty} \frac{1}{\pi} \int_{\mathbf{R}} [\int_{T(x-b)}^{Tx} \frac{\sin u}{u} du] dP_X(x) = F(b) - F(0+0) + \frac{1}{2}\big(F(b+0) - F(b) + F(0+0) - F(0)\big)$$

$$= \tilde{F}(b) - \tilde{F}(0) ,$$

où on a noté F au lieu de F_X.

Cas particuliers

1) Lois à densité de probabilité :

Si P_X admet une densité f par rapport à la mesure de Lebesgue sur $\mathbf{R}$ ou $\mathbf{R}^n$, et si Φ_X appartient à $L^1(\mathbf{R})$ ou à $L^1(\mathbf{R}^n)$, f s'obtient par la formule d'inversion de Fourier.

Qui plus est, si X est une variable aléatoire dont la fonction caractéristique Φ_X appartient à $L^1(\mathbf{R}^n)$, la loi de X admet une densité f continue sur $\mathbf{R}^n$ donnée par la formule :

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \Phi_X(t)\, e^{-i\langle t,x\rangle}\, dt \quad \text{pour} \quad x \in \mathbf{R}^n \quad .$$

2) Lois discrètes à valeurs entières (dans $\mathbf{Z}$) :

Si une v. a. X prend les valeurs $n \in \mathbf{Z}$ avec la probabilité p_n, alors sa fonction caractéristique Φ_X est égale à :

$$\Phi_X(t) = \sum_{n\in\mathbf{Z}} e^{itn}\, p_n$$

et on a

$$p_n = \frac{1}{2\pi} \int_0^{2\pi} e^{-itn}\, \Phi_X(t)\, dt \quad .$$

Exercice 1.11

Trouver la densité d'une v. a. admettant pour fonction caractéristique $\Phi_X(t) = e^{-|t|}$.

Exercice 1.12

Démontrer le résultat suivant :

Soit X une v. a. qui prend des valeurs discrètes x_k , $k \in \mathbf{N}$, avec les probabilités p_k. On suppose avoir $x_j \neq x_k$ pour $j \neq k$. Montrer que, si Φ_X est la fonction caractéristique de X, on a :

$$p_n = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \Phi_X(t)\, e^{-itx_n}\, dt \quad \forall\, n \in \mathbf{N} \quad .$$

Les fonctions caractéristiques sont extrêmement utiles, en particulier pour étudier la loi d'une somme de variables aléatoires indépendantes, comme nous allons le voir au chapitre suivant ; elles joueront un grand rôle aussi pour l'étude de la convergence d'une suite de v. a.

Chapitre 2 : Indépendance

I - Notion d'indépendance

I-1 - Indépendance d'évènements

Définition 2.1 *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Deux évènements A et B appartenant à $\mathcal{A}$ sont dits indépendants s'ils vérifient :*

$$P(A \cap B) = P(A) \, . \, P(B) \quad .$$

Définition 2.2 Indépendance mutuelle *Des évènements $A_1, \ldots, A_n$ sont dits mutuellement indépendants s'ils vérifient :*

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j) \quad , \forall \, J \subseteq \{1, \ldots, n\} \quad .$$

Exercice 2.1

1) Vérifier que, si deux évènements sont indépendants, il en est de même de leurs complémentaires.

2) On dit que deux évènements sont incompatibles s'ils ne peuvent se réaliser simultanément, c'est à dire s'ils sont disjoints. Des évènements incompatibles peuvent-ils être indépendants ?

3) Comparer l'indépendance 2 à 2 et l'indépendance mutuelle d'une famille finie d'évènements. (Indication : on pourra considérer un modèle de l'expérience consistant à lancer 2 dés non truqués, et regarder les évènements suivants : "le 1er chiffre est pair", "le 2ème chiffre est pair", "la somme des chiffres est paire".)

Définition 2.3 Indépendance de tribus d'évènements *Soient $\mathcal{A}_1, \ldots, \mathcal{A}_n$ n tribus incluses dans $\mathcal{A}$. On dit qu'elles sont indépendantes si on a :*

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j) \quad , \forall \, A_j \in \mathcal{A}_j \ , \ J \subset \{1, \ldots, n\} \quad .$$

Exercice 2.2

On considère l'espace probabilisé produit de $(\Omega, \mathcal{A}, P)$ par lui-même. Vérifier que les évènements $A \times \Omega$ et $\Omega \times B$ sont indépendants.

Exercice 2.3

On considère l'espace probabilisé $(\Omega, \mathcal{A}, P)$, où Ω a 11 éléments, et où P est la probabilité de Laplace. Trouver tous les couples (A, B) d'évènements tels que A et B soient indépendants.

Définition 2.4 *Une famille infinie d'évènements est dite indépendante, mutuellement ou deux à deux, si toute sous-famille finie l'est.*

I-2 - Indépendance de variables aléatoires

Définition 2.5 *Soit X une variable aléatoire, c'est à dire une application mesurable d'un espace mesurable $(\Omega, \mathcal{A})$ dans un espace mesurable $(\Omega', \mathcal{A}')$. On appelle tribu engendrée par X la sous-tribu de $\mathcal{A}$ définie par :*

$$X^{-1}(\mathcal{A}') = \{X^{-1}(A) \ ; \ A \in \mathcal{A}'\} \quad .$$

Définition 2.6 *Soient n variables aléatoires $X_j \ : \ (\Omega, \mathcal{A}) \to (\Omega_j, \mathcal{A}_j)$, avec $j = 1, \ldots, n$. On dit qu'elles sont indépendantes si les tribus qu'elles engendrent : $X_j^{-1}(\mathcal{A}_j)$ pour $j = 1, \ldots, n$, le sont.*

Définition 2.7 *Une famille infinie de variables aléatoires est dite indépendante si toute sous-famille finie l'est.*

Exercice 2.4

Soient $X_1, \ldots, X_n$ n v. a. indépendantes à valeurs réelles. Soient $f_1, \ldots, f_n$ n fonctions boréliennes définies sur $\mathbf{R}$. Montrer que les v. a. $f_1(X_1), \ldots, f_n(X_n)$ sont également indépendantes.

Remarque : Par la suite, quand on parlera d'indépendance d'évènements ou de v. a. sans plus de précision, il s'agira de l'indépendance mutuelle.

II - Caractérisation de l'indépendance de v. a.

Proposition 2.1 *Soient n variables aléatoires réelles $X_1, \ldots, X_n$ définies sur $(\Omega, \mathcal{A}, P)$.*
Elles sont indépendantes si, et seulement si, la loi du vecteur aléatoire $X = (X_1, \ldots, X_n)$ est égale au produit des lois des variables X_j, c'est à dire ssi :

$$\nu_X = \nu_{X_1} \otimes \ldots \otimes \nu_{X_n} \quad .$$

On dit que la loi conjointe est égale au produit des lois marginales.

Preuve : laissée au lecteur en exercice.

Corollaire 2.2 Propriétés de v. a. indépendantes

Soient n variables aléatoires réelles indépendantes $X_1, \ldots, X_n$, définies sur $(\Omega, \mathcal{A}, P)$. Alors on a :

1) $E(X_1.X_2.\ldots.X_n) = E(X_1).E(X_2).\ldots.E(X_n)$ si les X_j admettent une espérance.

2) $V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$, si les X_j admettent une variance.

En particulier, si X_1 et X_2 sont des v. a. indépendantes, leur covariance est nulle.

Preuve :

Pour 1), il suffit d'appliquer le théorème de Fubini, après avoir écrit l'espérance de $X_1.X_2.\ldots.X_n$ comme une intégrale sur Ω, puis comme une intégrale sur $\mathbf{R}^n$ par rapport à la loi conjointe du n-uplet $(X_1, X_2, \ldots, X_n)$.

Pour 2), on écrit la variance comme la moyenne quadratique des écarts à la moyenne, on développe la forme quadratique obtenue et on utilise 1).

Remarque Si deux v. a. X_1 et X_2 sont indépendantes, leur covariance $cov(X_1, X_2)$ est nulle, mais la réciproque est fausse.

Proposition 2.3 *Soient n variables aléatoires réelles $X_1, \ldots, X_n$ définies sur $(\Omega, \mathcal{A}, P)$. Elles sont indépendantes si, et seulement si, la fonction caractéristique de la v. a. $X = (X_1, \ldots, X_n)$ est égale au produit tensoriel des fonctions caractéristiques des v.a. X_j, ie*

$$\Phi_X(t) \ = \ \prod_{j=1,\ldots,n} \Phi_{X_j}(t_j) \ \text{ avec } \ t = (t_1, \ldots, t_n) \quad .$$

Preuve :

Si $X_1, \ldots, X_n$ sont des v. a. indépendantes, la fonction caractéristique du n-uplet $X = (X_1, \ldots, X_n)$ s'écrit :

$$\Phi_X(t) \ = \ E(\exp^{i(t_1x_1+t_2x_2+\ldots+t_nx_n)}) \ = \ E(\prod_{j=1,\ldots,n} \exp^{it_jx_j}) \ = \ \prod_{j=1,\ldots,n} E(\exp^{it_jx_j}) \quad .$$

Réciproquement, si X_1 et X_2 sont des v. a. de fonctions caractéristiques respectives Φ_{X_1} et Φ_{X_2}, et si le couple a pour fonction caractéristique $\Phi_{(X_1,X_2)} \ = \ \Phi_{X_1} \otimes \Phi_{X_2}$, alors la loi du couple coïncide avec le produit des lois marginales, d'après l'injectivité de l'application qui à une mesure associe sa fonction caractéristique.

III - Loi d'une somme de v. a. indépendantes

III-1 - Convolution de mesures boréliennes

Définition 2.8 *Soient $\nu_1, \ldots, \nu_p$ p mesures boréliennes sur $\mathbf{R}^n$. On appelle produit de convolution et on note $\nu_1 * \ldots * \nu_p$ la mesure image du produit $\nu_1 \otimes \ldots \otimes \nu_p$ par l'application*

$$(x_1, \ldots, x_p) \to x_1 + \ldots + x_p \quad .$$

On peut donc écrire, si B est un Borélien de $\mathbf{R}^n$:

$$(\nu_1 * \ldots * \nu_p)(B) \ = \ (\nu_1 \otimes \ldots \otimes \nu_p)(\{ \ (x_1, \ldots, x_p) \ ; \ x_1 + \ldots + x_p \in B \ \}) \quad .$$

Exercice 2.5

Si ν_1 et ν_2 sont des mesures de densités respectives f_1 et f_2 par rapport à la mesure de Lebesgue sur $\mathbf{R}^n$, montrer que $\nu_1 * \nu_2$ est la mesure de densité $f_1 * f_2$.

Exercice 2.6

Déterminer le produit de convolution $\nu_1 * \nu_2$ si ν_1 et ν_2 sont respectivement les mesures de Dirac en a_1 et a_2 sur $\mathbf{R}$.

Proposition 2.4 : Propriétés du produit de convolution

1) Il est associatif et commutatif.

2) Il a un élément neutre : la mesure de Dirac en 0.

3) Si $\nu_1, \ldots, \nu_p$ sont des mesures de probabilité, il en est de même de $\nu_1 * \ldots * \nu_p$.

III-2 - Sommes de v. a. indépendantes

Proposition 2.5 *Soient n variables aléatoires réelles indépendantes $X_1, \ldots, X_n$. La loi de la somme est égale au produit de convolution des lois :*

$$\nu_{X_1+X_2+\ldots+X_n} = \nu_{X_1} * \ldots * \nu_{X_n} \quad .$$

Preuve : Exercice sur les mesures images!

Proposition 2.6 *Soient n variables aléatoires réelles indépendantes $X_1, \ldots, X_n$. La fonction caractéristique de la somme est égale au produit des fonctions caractéristiques :*

$$\Phi_{X_1+X_2+\ldots+X_n}(t) = \prod_{j=1,\ldots,n} \Phi_{X_j}(t) \quad .$$

Preuve : C'est encore un exercice pour le lecteur.

Exercice 2.7

Trouver la loi d'une somme de v. a. dans les cas suivants :

1) $X_1, \ldots, X_p$ sont p v. a. de Bernoulli indépendantes et de même paramètre.

2) $X_1, \ldots, X_p$ sont p v. a. binomiales indépendantes de tailles différentes et de même paramètre.

3) $X_1, \ldots, X_p$ sont p v. a. indépendantes suivant des lois de Poisson.

4) $X_1, \ldots, X_p$ sont p v. a. indépendantes suivant des lois gaussiennes.

5) $X_1, \ldots, X_p$ sont p v. a. indépendantes suivant la loi uniforme sur $[-1, 1]$.

Exercice 2.8

Soit X une variable aléatoire de Cauchy, ie suivant la loi de densité $c_\alpha = \frac{\alpha}{\pi(\alpha^2+x^2)}$. α est un paramètre positif.

1) Calculer sa fonction caractéristique.

2) On se place dans le cas $\alpha = 1$. Soit Y une v. a. égale à X.

Calculer ϕ_{X+Y} et $\phi_X \phi_Y$. Conclusion ?

Exercice 2.9

On considère 3 variables aléatoires indépendantes X, Y et Z.

On note ϕ_X , ϕ_Y et ϕ_Z leurs fonctions caractéristiques.

1) Quelle est la fonction caractéristique du couple (X+Y , Y+Z) ?

2) Trouver une CNS sur les fonctions caractéristiques pour que X+Y et Y+Z soient indépendantes.

Exercice 2.10

1)

a) X, Y étant 2 variables indépendantes, de lois respectives ν_X et ν_Y et f une fonction $(\nu_X \otimes \nu_Y)$-intégrable sur $\mathbf{R}^2$, donner l'expression de $E(f(X,Y))$ en fonction de f, ν_X, ν_Y.

b) Rappeler, sans la recalculer, la transformée de Fourier de la fonction de Gauss $g(x) = \exp(-\frac{x^2}{2})$, et en déduire la fonction caractéristique de la loi de Gauss centrée réduite de densité $g_0(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.

2) On considère 2 variables aléatoires X et Y de même densité g_0 et indépendantes. On note $Z = XY$ la variable aléatoire égale à leur produit. Montrer que la fonction caractéristique de Z est égale à :

$$\phi_Z(t) = \frac{1}{\sqrt{1+t^2}} \quad .$$

3) On considère maintenant 4 variables aléatoires, X, Y, L, et T, indépendantes et suivant toutes la même loi gaussienne centrée réduite. On note $\Delta = XY - LT$. Calculer ϕ_Δ et montrer que Δ suit la loi de densité h égale à $h(x) = \frac{1}{2}\exp(-|x|)$.

Chapitre 3 : Autour du jeu de "Pile ou Face"

Lois des grands nombres dans le schéma de Bernoulli. Lemme de Borel-Cantelli.

Le jeu de "Pile ou Face" est un modèle avec lequel on peut illustrer plusieurs jolis théorèmes de la théorie des Probabilités.
Nous avons vu au premier chapitre des modèles associés au jeu de 'Pile ou Face", où l'espace des résultats était fini ou dénombrable. On va considérer ici un jeu d'une durée infinie, et, dans ce cas, l'existence d'un espace probabilisé permettant de le modéliser n'est pas du tout évidente.

I - Le modèle de Bernoulli

Pour modéliser un jeu de "Pile ou Face" de durée infinie, on introduit l'espace Ω des suites infinies $\omega = (x_m)_{m\geq 1}$ formées de 0 et de 1 qui sont effectivement les résultats possibles du jeu, ainsi que les v. a. égales aux coordonnées de cet espace, X_m $m \geq 1$, qui donnent les résultats de chaque lancer : $X_m(\omega) = x_m$ si $\omega = (x_1, x_2, \ldots, x_m, \ldots)$.
On suppose que $p : 0 < p < 1$ est la probabilité du 1, et que la suite de v. a. $(X_m)_{m\geq 1}$ est une famille de v. a. indépendantes.
La probabilité P dont est muni Ω doit donc être telle que l'on ait pour $n \geq 1$:

$$P(X_1 = x_1, \ldots, X_n = x_n) \ = \ p(x_1)\ldots p(x_n)$$

$$\text{avec } x_m = 0 \text{ ou } 1 \text{ , et } p(1) = p, \ p(0) = 1 - p \ .$$

Par ailleurs la tribu $\mathcal{A}$ sur Ω doit inclure tous les ensembles cylindriques $\{ X_1 = x_1, \ldots, X_n = x_n \}$ de Ω.
On peut admettre l'existence d'un tel espace probabilisé. On peut aussi savoir démontrer son existence en montrant que l'existence de la probabilité P sur Ω équivaut en fait à celle de la mesure de Lebesgue sur $[0, 1[$. cf par exemple le polycopié de J. Neveu.
Exercice 3.1 : Construction du modèle de Bernoulli dans le cas $p = \frac{1}{2}$.

Soit λ la mesure de Lebesgue sur $[0, 1[$, c'est à dire l'unique probabilité sur la tribu borélienne de $[0, 1[$ telle que $\lambda[(a, b)] \ = \ b - a$ pour tout intervalle (a, b) tel que $0 \leq a < b \leq 1$. Soit f l'application de $[0, 1[$ dans l'espace $\Omega' \ = \ \{0, 1\}^{\mathbf{N}^*}$ des suites infinies de 0 et de 1 définies par le développement dyadique

$$f(x) \ = \ (x_1, x_2, \ldots) \quad \text{si } x \ = \ \sum_1^{\infty} x_i\, 2^{-i} \quad (x_i = 0 \text{ ou } 1) \quad .$$

(Pour les nombres dyadiques on convient de considérer le développement qui n'a pas une infinité de 1).
On munit Ω' de la tribu engendrée par les cylindres

$$C_{y_1,\ldots,y_k} = \{ \ (x_1, x_2, \ldots) \ ; \ x_i = y_i \text{ pour } 1 \leq i \leq k\} \quad \text{avec } k \geq 1 \text{ et } y_1, \ldots, y_k \ = 0 \text{ ou } 1 \quad .$$

1) Montrer que l'application f est alors mesurable.
2) Montrer que la probabilité image de λ par f s'identifie à la probabilité P du schéma de Bernoulli dans le cas $p = \frac{1}{2}$.

Exercice 3.2

Montrer que chaque singleton inclus dans Ω est un évènement de probabilité nulle.

Le modèle de Bernoulli est donc celui qui traduit un jeu de "Pile ou Face" de durée infinie ; d'une manière générale, c'est celui qui permet de traduire une expérience composée d'une suite infinie d'épreuves identiques et indépendantes, à deux issues possibles : pile ou face, oui ou non... C'est celui auquel on fait référence quand on dit : "on considère une suite infinie de v. a. de Bernoulli, indépendantes et de même paramètre".

Le cas d'une suite de v. a. indépendantes et équidistribuées est très important en Probabilités, et c'est dans ce cadre que nous allons démontrer la loi des grands nombres sous les formes "faible" et "forte".

Pour les paragraphes II et III, on considère donc une suite de variables aléatoires indépendantes et équidistribuées.

II - Loi faible des grands nombres. Cas de v. a. équidistribuées.

Définition 3.1 : Convergence stochastique *On dit qu'une suite de v. a. X_n converge stochastiquement (ou en probabilités) vers une v. a. X_0 si on peut écrire :*

$$\lim_{n\to\infty} P(|X_n - X_0| > \epsilon) \;=\; 0 \;\; \forall\, \epsilon > 0 \quad .$$

La convergence presque sûre notée p. s., qui est la convergence presque partout, implique la convergence stochastique, mais la réciproque est fausse. Pour le démontrer je vous propose les deux exercices suivants :

Exercice 3.3

On considère l'ensemble $[X_n \to 0] = \{\omega \in \Omega,\ \lim_{n\nearrow+\infty} X_n(\omega) = 0\}$. Pour tout entier $m \geq 1$ on définit

$$A_m = \cap_{n\geq 1} \cup_{k\geq 1} [|X_{n+k}| \geq \frac{1}{m}]$$

1) Montrer que $A_m \in \mathcal{A}$, et en déduire que $[X_n \to 0] \in \mathcal{A}$.

2) Montrer que, si la suite X_n converge p.s. vers 0, elle converge en probabilités vers 0.

Exercice 3.4

On considère, sur $\Omega \;=\; [0,1]$ muni de la mesure de Lebesgue, la suite de variables aléatoires X_n définies par

$$\text{pour } n = 2^h + k \ ,\ 0 \leq k < 2^h \ :\ X_n \;=\; \mathcal{X}_{[\frac{k}{2^h}, \frac{k+1}{2^h}[} \quad .$$

Etudier la convergence stochastique et la convergence presque sûre de cette suite.

Proposition 3.1 Loi faible des grands nombres pour des v. a. équidistribuées.

Soient X_n une suite de v. a. indépendantes et équidistribuées, de seconds moments finis. On note S_n la somme $S_n = X_1 + X_2 + \ldots + X_n$. La suite des moyennes $M_n = \frac{S_n}{n}$ converge en probabilités vers la moyenne commune des X_n quand n tend vers l'infini, ie :

$$\lim_{n \to \infty} P(|\frac{S_n}{n} - E(X_1)| > \epsilon) = 0 \ \forall \epsilon > 0 \quad .$$

Preuve : Conséquence immédiate de l'inégalité de Bienaymé-Tchebichev !

Exercice 3.5 : Application aux sondages.

On désire procéder à un sondage afin d'estimer la probabilité p qu'a le candidat Durand d'être élu. Pour cela on interroge un grand nombre d'électeurs, et on comptabilise les "oui" obtenus. On note N le nombre de personnes interrogées. On traduit par la v. a. X_i la réponse de la i-ième personne interrogée. On suppose les X_i indépendantes.

Remarque : on appelle échantillon de taille N un N-uplet de v. a. indépendantes et de même loi.

1) Donner la loi des X_i.

2) Montrer que la v. a. égale au nombre de "oui" obéit à une loi binomiale de taille N et de paramètre égal à p.

3) Montrer que, quand on estime la probabilité p par la fréquence du "oui", on applique la loi des grands nombres.

Le modèle de Bernoulli, avec la démonstration complète de la loi faible (y compris l'inégalité de Bienaymé-Tchebichev), puis l'application aux sondages, peut faire un point de développement à l'oral.

La loi faible signifie que la v. a. $\frac{S_n}{n}$ pour n grand, a peu de chances de différer sensiblement de $E(X_1)$. C'est un résultat facile à prouver, mais il existe un résultat plus fort, à savoir la loi forte des grands nombres.

En réalité, quand on fait un sondage, on approche la probabilité par la fréquence observée, et c'est la forme forte qu'on utilise.

III - Loi forte des grands nombres. Cas de v. a. équidistribuées.

Proposition 3.2

Soient X_n une suite de v. a. indépendantes et équidistribuées, de seconds moments finis. On note S_n la somme $S_n = X_1 + X_2 + \ldots + X_n$. La suite des moyennes $M_n = \frac{S_n}{n}$ converge presque sûrement vers la moyenne commune des X_n quand n tend vers l'infini, ie :

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to E(X_1) \ \text{p.s.} \quad .$$

On peut en fait remplacer l'hypothèse "de seconds moments finis" par l'existence seulement de $E(X_1)$. Le résultat est à retenir, mais la démonstration est plus difficile.

Preuve :

Il existe plusieurs façons de prouver cette proposition. On peut s'appuyer sur le lemme de Borel-Cantelli, qui fait l'objet du paragraphe suivant. On peut aussi, et c'est ce qu'on va faire ici, faire une démonstration "self-contained" en utilisant un argument analogue à celui qui permet de prouver le lemme. La démonstration qui suit est due à E. Borel, et nous la présentons sous forme d'exercice.

Exercice 3.6 : Preuve de la loi forte, contexte de Borel

On reprend les données et les notations de l'exercice 3.3 . Soient un espace probabilisé $(\Omega, \mathcal{A}, P)$ et $\{X_n\}$ une suite de variables aléatoires réelles. On pose $[X_n \to 0] = \{\omega \in \Omega,\ \lim_{n \nearrow +\infty} X_n(\omega) = 0\}$. Pour tout entier $m \geq 1$ on définit

$$A_m = \cap_{n\geq 1} \cup_{k\geq 1} [|X_{n+k}| \geq \frac{1}{m}]$$

1) Montrer que $A_m \in \mathcal{A}$. En déduire que $[X_n \to 0] \in \mathcal{A}$.

2) On suppose que pour tout $m \geq 1$ on a

$$\sum_{n\geq 0} P[|X_n| \geq \frac{1}{m}] < +\infty$$

Montrer que $P(A_m) = 0$ pour tout $m \geq 1$ et en déduire que $P[X_n \to 0] = 1$.

On suppose maintenant que les variables $\{X_n\}$ sont mutuellement indépendantes, de même espérance mathématique μ et de même variance σ^2. On pose $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$.

3) Montrer que

$$P[|S_{n^2} - \mu| \geq \frac{1}{m}] \leq \frac{m^2\sigma^2}{n^2}$$

et en déduire avec l'aide de la question précédente que la suite $\{S_{n^2}\}$ converge presque surement vers μ.

4) On suppose de plus qu'il existe $M > 0$ telle que

$$|X_n(\omega)| \leq M, \quad \forall \omega \in \Omega, \quad \forall n \geq 1$$

Montrer que pour $k^2 \leq n < (k+1)^2$ on a $|S_n - S_{k^2}| \leq \frac{4M}{k}$.

En déduire que la suite $\{S_n\}$ converge presque surement vers μ (loi forte des grands nombres, cas particulier dû à E.Borel, 1909)).

5) Donner l'interprétation probabiliste du résultat précédent sur le jeu de pile ou face en introduisant l'espace probabilisé $\{0, 1\}^{\mathbf{N}}$.

IV - Lemme de Borel-Cantelli.

Remarque : Ni la loi forte des grands nombres, ni le lemme de Borel-Cantelli ne sont au programme de l'Agrégation si vous ne choisissez pas les probabilités à l'épreuve de modélisation. Mais il me paraît important de les connaître si on veut avoir une culture minimale en Probabilités.

On continue à jouer à "Pile ou Face".

Exercice 3.7

Répondre intuitivement aux deux questions suivantes :

1) On joue une infinité de fois à "Pile ou Face". Quelle est la probabilité d'obtenir 2 fois de suite "Pile"? k fois de suite ?

2) On met un singe immortel devant une machine à écrire. Il tape au hasard sur les touches. Quelle est la probabilité qu'il tape les oeuvres complètes de Proust, sans une seule erreur ?

Si vous avez répondu de manière différente à ces deux questions, vous avez perdu. Il est facile de voir qu'il s'agit du même problème dans les deux cas, à savoir : "déterminer la probabilité d'obtenir une séquence finie imposée de résultats lors d'une suite infinie d'épreuves identiques et indépendantes, n'ayant chacune qu'un nombre fini d'issues possibles".

Le lemme de Borel-Cantelli va vous permettre de prouver que dans les deux cas le résultat est 1. Ce qui est intuitif pour la première question, mais beaucoup moins pour la seconde !

Définition 3.2 : Evènements terminaux. *Soit* $(A_n)_{n\in\mathbf{N}}$ *une suite d'évènements d'un espace probabilisé* $(\Omega, \mathcal{A}, P)$. *On appelle limite supérieure des* A_n*, et on note* $\limsup_{n\to\infty} A_n$ *l'évènement :*

$$\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbf{N}} \bigcup_{j\geq n} A_j \quad .$$

Ces évènements sont appelés évènements terminaux.

Proposition 3.3 : Lemme de Borel-Cantelli.

Soit $(A_n)_{n\in\mathbf{N}}$ *une suite d'évènements d'un espace probabilisé* $(\Omega, \mathcal{A}, P)$. *Alors :*

1) Si la série de terme général $P(A_n)$ *est convergente, l'évènement* $\limsup_{n\to\infty} A_n$ *a une probabilité nulle.*

2) Si la série diverge, et si les évènements sont indépendants, alors l'évènement $\limsup_{n\to\infty} A_n$ *a une probabilité égale à 1.*

Remarque : On peut remplacer dans 2) l'hypothèse d'indépendance par l'existence d'une sous-suite d'évènements indépendants, et même l'indépendance, ie l'indépendance mutuelle, par l'indépendance 2 à 2.

Preuve :

1) Soit $B_n = \bigcup_{j\geq n} A_j$. Alors l'inégalité $P(B_n) \leq \sum_{j\geq n} P(A_j)$ prouve que $P(B_n)$ tend vers 0 quand n tend vers ∞. Or B_n est une suite décroissante d'évènements dont l'intersection est $\limsup_{n\to\infty} A_n$. On a donc prouvé 1).

2) On va montrer que l'on a $P(B_n) = 1$ pour tout n. Pour cela considérons le complémentaire de B_n :

$$B_n^C = \bigcap_{j\geq n} A_j^C = \bigcap_{N\in\mathbf{N}} \bigcap_{N\geq j\geq n} A_j^C \quad .$$

Or on peut écrire :

$$P(\bigcap_{N\geq j\geq n} A_j^C) = \prod_{N\geq j\geq n} P(A_j^C) = \prod_{N\geq j\geq n} (1 - P(A_j)) \quad ,$$

puisque les A_j sont indépendants.

Par ailleurs, en utilisant l'inégalité $0 \leq x < 1 \Rightarrow \ln(1-x) \leq -x$, on obtient

$$\sum_{N \geq j \geq n} \ln(1 - P(A_j)) \leq - \sum_{N \geq j \geq n} P(A_j) \to -\infty \text{ quand } N \to \infty \quad .$$

On en déduit alors que la probabilité de B_n^C est nulle et donc que l'on a : $P(B_n) = 1$.
Le résultat $P(B_n) = 1$ signifie que, pour tout n, il existe $j \geq n$ tel que A_j se réalise presque sûrement. Une infinité des A_n se réalise donc presque sûrement.

Exercice 3.8

Faire l'exercice 3.7 en remplaçant "intuitivement" par "en justifiant votre réponse".

Pour clore ce chapitre, on pourrait parler du théorème de la limite centrale, qui dans le cas de la loi binomiale peut se démontrer directement. Il permet de répondre simplement à des questions du type "on lance 100 fois une pièce de monnaie, on obtient 80 fois "Pile". Peut-on en conclure que la pièce est fausse ?" On verra ce théorème au chapitre suivant, comme conséquence du théorème de Lévy, et dans un contexte plus général.

Chapitre 4 : Convergence en loi.

Dans ce chapitre on considère une suite de v. a. $(X_j)_{j\in\mathbf{N}}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbf{R}^k$, et on s'intéresse au comportement asymptotique de cette suite.

I - Les modes de convergence déjà rencontrés

Pour le moment on s'est intéressé au comportement asymptotique d'une suite de v. a. $(X_j)_{j\in\mathbf{N}}$ en les considérant comme fonctions mesurables sur $(\Omega, \mathcal{A}, P)$. On a rencontré les types suivants : convergence presque sûre, convergence en moyenne, convergence en probabilités. On les compare dans la proposition suivante :

Proposition 4.1

Soit $(X_j)_{j\in\mathbf{N}}$ et X des v. a. définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On a les assertions suivantes :

1) La convergence presque sûre de X_j vers X implique la convergence en probabilités. La réciproque est fausse en général.

2) Soit $p \geq 1$. La convergence en moyenne d'ordre p de X_j vers X implique la convergence en probabilités. La réciproque est fausse en général.

3) La convergence presque sûre de X_j vers X implique la convergence en moyenne d'ordre p si la suite de v. a. X_j est uniformément bornée.

4) La convergence en moyenne d'ordre p de X_j vers X implique la convergence presque sûre pour une sous-suite.

5) Si $q \geq p \geq 1$ la convergence en moyenne d'ordre q de X_j vers X implique la convergence en moyenne d'ordre p.

Preuve :

1) a été fait au chapitre précédent.

3) résulte du théorème de convergence dominée.

4) est un résultat d'intégration.

Démontrons 2) et 5).

2) Si la suite X_j converge vers X en moyenne d'ordre p, on peut écrire :

$$\lim_{j\to+\infty} \int_\Omega \mid X_j(\omega) \; - \; X(\omega) \mid^p dP(\omega) \; = \; 0 \quad .$$

Or on a, pour tout $\delta > 0$:

$$\int_\Omega |X_j(\omega) - X(\omega)|^p dP(\omega) = \int_{\{|X_j-X|>\delta\}} |X_j(\omega) - X(\omega)|^p dP(\omega) + \int_{\{|X_j-X|\leq\delta\}} |X_j(\omega) - X(\omega$$

$$\geq \; \delta^p \; P(\{|X_j - X| > \delta\}) \quad .$$

On obtient ainsi la convergence en probabilités de X_j vers X.

5)

a) La convergence en moyenne d'ordre p implique la convergence en moyenne (ie d'ordre 1).

En effet grâce à l'inégalité de Hölder on obtient :

$$\int_\Omega \mid Y \mid dP \leq [\int_\Omega \mid Y \mid^p dP\,]^{\frac{1}{p}} \quad \forall\, p \geq 1 \text{ , pour toute v. a. } Y \quad .$$

b) Supposons $q > p \geq 1$. Alors $r = \frac{q}{p}$ admet un conjugué s, et, en utilisant l'inégalité de Hölder pour r et s, on obtient :

$$\int_\Omega \mid Y \mid^p dP \leq [\int_\Omega \mid Y \mid^q dP\,]^{\frac{p}{q}} \text{ , pour toute v. a. } Y \quad .$$

Exercice 4.1 : contre-exemple à la réciproque de 2) :

On considère la suite de v. a. $(X_n)_{n\in N}$ de lois égales à $\nu_{X_n} = \frac{1}{n}\delta_{n^2} + (1-\frac{1}{n})\delta_0$.

Montrer que la suite X_n converge vers 0 en probabilités, mais pas en moyenne.

La proposition 4.1 donne les résultats les plus importants. On peut raffiner, en particulier regarder des contextes où les réciproques sont exactes. Vous pouvez pour cela consulter Foata et Fuchs, où vous trouverez des compléments que vous pourrez faire sous forme d'exercices. (Les maths se font le crayon à la main!).

II - La convergence en loi

On va s'intéresser dans ce paragraphe au comportement asymptotique des lois des v. a. X_j considérées comme mesures bornées sur $\mathbf{R}^k$. Ceux qui veulent approfondir ce qui suit peuvent consulter Dacunha-Castelle - Duflo.

Notons C_K, (resp. C_0, resp. C_b), l'espace des fonctions continues sur $\mathbf{R}^k$, à support compact (resp. tendant vers 0 à l'infini, resp. bornées). Soit μ une mesure bornée sur $\mathbf{R}^k$. On peut considérer μ comme une forme linéaire continue sur chacun de ces 3 espaces muni de la topologie de la convergence uniforme ; en effet on peut écrire :

$$\mid \int_{\mathbf{R}^k} f\, d\mu \mid \leq \|f\|_\infty \times \mu\, (\mathbf{R}^k) \quad \forall\, f \in C_K \quad (\text{resp. } \mathrm{C}_0 \text{ , } \mathrm{C}_\mathrm{b} \text{)} \quad .$$

On regarde donc les mesures bornées sur $\mathbf{R}^k$ comme des éléments du dual topologique de C_K , (resp. C_0, resp. C_b), et les convergences que l'on va considérer sont les topologies faibles sur C'_K , (resp. C'_0, resp. C'_b). Ainsi, pour μ_n et μ mesures bornées sur $\mathbf{R}^k$:

Définitions 4.1

1) La suite μ_n converge vers μ "vaguement" si l'on a $\int_{\mathbf{R}^k} f\, d\mu_n \rightarrow \int_{\mathbf{R}^k} f\, d\mu$ pour toute fonction f appartenant à C_K.

2) La suite μ_n converge vers μ "faiblement" si l'on a $\int_{\mathbf{R}^k} f\, d\mu_n \rightarrow \int_{\mathbf{R}^k} f\, d\mu$ pour toute fonction f appartenant à C_0.

3) La suite μ_n converge vers μ "étroitement" si l'on a $\int_{\mathbf{R}^k} f\, d\mu_n \rightarrow \int_{\mathbf{R}^k} f\, d\mu$ pour toute fonction f appartenant à C_b.

On a évidemment 3) $\Rightarrow$ 2) $\Rightarrow$ 1). Pour la réciproque, on peut énoncer le résultat suivant :

Proposition 4.2

Soient $(\mu_n)_{n\in N}$ et μ des mesures bornées sur $\mathbf{R}^k$. Sous l'hypothèse $\mu_n(\mathbf{R}^k) \rightarrow \mu(\mathbf{R}^k)$, si μ_n converge vers μ "vaguement", alors μ_n converge vers μ "étroitement".

Autrement dit, sous cette hypothèse, les 3 types de convergence sont équivalents. Par contre c'est faux sans cette hypothèse, comme on peut s'en convaincre en considérant $\mu_n = \delta_n$ sur $\mathbf{R}$.

C'est le cas en particulier si les mesures μ_n sont des mesures de probabilité sur $\mathbf{R}^k$, et si μ est aussi une mesure de probabilité sur $\mathbf{R}^k$.

Preuve :

Pour simplifier, on la fait dans le cas particulier $\mu_n(\mathbf{R}^k) = \mu(\mathbf{R}^k) = 1$ pour tout n. Il est très facile de l'adapter au cas général !

Soit f appartenant à C_b. On désire prouver l'assertion suivante :

$$\forall\, \epsilon > 0 \;\exists\, j_0 \;:\; \forall\, j \geq j_0 \quad | \int_{\mathbf{R}^k} f \, d\mu_j \;-\; \int_{\mathbf{R}^k} f \, d\mu \;| \;<\; \epsilon \quad ,$$

sachant qu'elle est vraie pour f appartenant à C_K. Tout d'abord, comme μ est une mesure bornée, on peut écrire :

$$\forall\, \epsilon > 0 \;\exists\, T \;:\; \mu\, (\{|x| \geq T\}) \;<\; \epsilon/4M \quad \text{avec} \quad M = \sup(||f||_\infty \;;\; 1) \quad .$$

Soit alors χ appartenant à C_0^∞ $(\mathbf{R}^k)$, telle que χ soit égale à 1 sur l'ensemble $\{|x| \leq T\}$, partout comprise entre 0 et 1, et à support dans la boule $\{|x| \leq T+1\}$.

χ appartient à C_K et donc on a :

$$\forall\, \alpha > 0 \;\exists\, j_1 = j_1(\alpha, \epsilon) \;:\; \forall\, j \geq j_1 \quad | \int_{\mathbf{R}^k} \chi \, d\mu_j \;-\; \int_{\mathbf{R}^k} \chi \, d\mu \;| \;<\; \alpha \quad .$$

On choisit $\alpha = \epsilon/4M$. Alors on peut écrire :

$$| \int_{\mathbf{R}^k} f \, d\mu_j \;-\; \int_{\mathbf{R}^k} f \, d\mu \;| \;\leq\; | \int_{\mathbf{R}^k} f \, d\mu_j \;-\; \int_{\mathbf{R}^k} f\chi \, d\mu_j \;| \;+$$

$$+\; | \int_{\mathbf{R}^k} f\chi \, d\mu_j \;-\; \int_{\mathbf{R}^k} f\chi \, d\mu \;| \;+\; | \int_{\mathbf{R}^k} f\chi \, d\mu \;-\; \int_{\mathbf{R}^k} f \, d\mu \;| \quad .$$

Notons respectivement A, B et C les 3 termes du second membre. Pour C on a :

$$|C| \;\leq\; M \int (1-\chi)\, d\mu \;\leq\; M\, \mu\, (\{|x| \geq T\}) \;<\; \epsilon/4 \quad .$$

Il existe j_2 tel que l'on ait $|B| \;\leq\; \epsilon/4$ pour $j \geq j_2$, car $f\chi$ appartient à C_K.

Quant à A, l'hypothèse $\mu_j(\mathbf{R}^k) = \mu(\mathbf{R}^k)$ pour tout j permet d'écrire, grâce au fait que l'on a

$| \int_{\mathbf{R}^k} \chi \, d\mu_j \;-\; \int_{\mathbf{R}^k} \chi \, d\mu| \;<\; \epsilon/4M$ pour tout $j \geq j_1$:

$$| \int_{\mathbf{R}^k} (1-\chi)\, d\mu_j \;-\; \int_{\mathbf{R}^k} (1-\chi)\, d\mu| \;=\; | \int_{\mathbf{R}^k} \chi \, d\mu_j \;-\; \int_{\mathbf{R}^k} \chi \, d\mu \;| \;<\; \frac{\epsilon}{4M} \;\; \forall\, j \geq j_1 \quad ,$$

et donc $|A|$ est majoré par $M \times [\;| \int_{\mathbf{R}^k} (1-\chi)\, d\mu| \;+\; \epsilon/4M] \;\leq \epsilon/2$.

Pour ϵ donné, on a donc bien construit un indice $j_3 = \sup(j_1, j_2)$, tel que pour tout $j \geq j_3$, on ait $| \int_{\mathbf{R}^k} f \, d\mu_j - \int_{\mathbf{R}^k} f \, d\mu| < \epsilon$.

Définition 4.2

Soit $(X_j)_{j\in\mathbf{N}}$ une suite de variables aléatoires à valeurs dans $\mathbf{R}^k$. On dit que la suite $(X_j)_{j\in\mathbf{N}}$ **converge en loi** *vers une variable aléatoire X si la suite de leurs lois $(\nu_{X_j})_{j\in\mathbf{N}}$ converge vers la loi ν_X de X, au sens des assertions de la proposition 4.2.*

Comme les mesures ν_{X_j} et ν_X sont des lois de probabilité, les 3 notions de convergence sont en effet équivalentes.

Il est alors naturel de se poser alors la question suivante :

Q Si B est un borélien de $\mathbf{R}^k$, et si les v. a. X_j convergent en loi vers X, peut-on dire que $\nu_{X_j}(B)$ converge vers $\nu_X(B)$?

Pour répondre on peut traiter l'exercice suivant :

Exercice 4.2

Soit μ_σ la mesure de densité égale à la fonction de Gauss g_σ : $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$ par rapport à la mesure de Lebesgue sur $\mathbf{R}$ ($\sigma > 0$).

1) Trouver la limite de μ_σ pour la convergence étroite quand σ tend vers 0.

2) En considérant l'intervalle $[0, +\infty[$, répondre à la question **Q**.

Proposition 4.3

On considère une suite $(X_j)_{j\in\mathbf{N}}$ de v. a. réelles, et X une v. a. réelle. Les assertions suivantes sont équivalentes :

1) X_j converge vers X en loi.

2) $\nu_{X_j}(I)$ converge vers $\nu_X(I)$ pour tout intervalle I borné dont ν_X ne charge pas les extrémités (ie si $I = (a, b)$, alors $\nu_X(a) = \nu_X(b) = 0$).

Remarque : On peut prouver une proposition équivalente dans $\mathbf{R}^k$ (cf Neveu).

Avant de démontrer la proposition 4.3, nous allons énoncer une autre forme de convergence équivalente à la convergence en loi.

Proposition 4.4

On reprend le contexte de la proposition 4.3. Les 2 assertions de cette proposition sont équivalentes à la suivante :

3) La suite F_{X_j} des fonctions de répartition des v. a. X_j converge vers la fonction de répartition F_X de X en tout point où F_X est continue.

Remarque C'est cette caractérisation que Foata - Fuchs donne comme définition de la convergence en loi.

Nous n'utiliserons pas les fonctions de répartition dans la suite, contrairement à ce que certains d'entre vous ont fait en L7 et à ce que fait l'ouvrage de Foata - Fuchs. C'est pourquoi nous ne démontrerons pas cette proposition. Pour se convaincre que 3) est une caractérisation du même type que 2), on peut faire l'exercice suivant :

Exercice 4.3

Soit X une variable aléatoire réelle. Alors la fonction de répartition F_X de X est continue en $x \in \mathbf{R}$ si, et seulement si, la mesure $\nu_X(\{x\})$ du point x est nulle.

Remarque Si on définit F_X par $F_X(x) = P(\{X < x\})$ on obtient une fonction continue à gauche, et si on prend comme définition $F_X(x) = P(\{X \leq x\})$, alors F_X est continue à droite. Par ailleurs F_X est toujours une fonction croissante. Ces remarques peuvent vous faciliter la résolution de l'exercice qui précède.

Preuve de la proposition 4.3

Dans ce qui suit nous allons noter ν_j au lieu de ν_{X_j}, et ν au lieu de ν_X.

Démonstration de 1) $\Rightarrow$ 2)

Soit $I = (a, b)$ un intervalle dont ν ne charge pas les extrémités, et χ sa fonction caractéristique. On encadre χ entre deux suites de fonctions f_n et g_n, continues à support compact,comprises entre 0 et 1, telles que l'on ait : $0 \leq f_n \leq \chi \leq g_n \leq 1$. Prenons par exemple f_n nulle en dehors de (a, b), égale à 1 sur $[a + \frac{1}{n}, b - \frac{1}{n}]$, et affine ailleurs, et g_n nulle en dehors de $[a - \frac{1}{n}, b + \frac{1}{n}]$, égale à 1 sur (a, b) et affine ailleurs (Faire un graphe).
Alors $\nu(\partial I) = 0 \Rightarrow \lim_{n\to\infty} \int (g_n - f_n)\, d\nu = 0$
En effet, on peut écrire :

$$\int (g_n - f_n)\, d\nu = \int_{a-\frac{1}{n}}^{a+\frac{1}{n}} (g_n - f_n)\, d\nu + \int_{b-\frac{1}{n}}^{b+\frac{1}{n}} (g_n - f_n)\, d\nu$$

$$\leq \nu[a - \frac{1}{n}, a + \frac{1}{n}] + \nu[b - \frac{1}{n}, b + \frac{1}{n}] \quad .$$

$A_n = [a - \frac{1}{n}, a + \frac{1}{n}]$ est une suite décroissante et donc $\nu(A_n)$ tend vers $\nu(\{a\}) = 0$. De même pour l'autre extrémité. Pour ϵ donné, on peut donc choisir N tel que $\int (g_N - f_N)\, d\nu \leq \epsilon/2$. Fixons ce N.

Alors

$$\exists\, j_0 : \forall\, j \geq j_0 \; |\int f_N d\nu_j - \int f_N \, d\nu| \leq \epsilon/2 \text{ et } |\int g_N \, d\nu_j - \int g_N \, d\nu| \leq \epsilon/2$$

Considérons :

$$|\nu_j(I) - \nu(I)| = |\int \chi \, d\nu_j - \int \chi \, d\nu|$$

En encadrant χ entre f_N et g_N on obtient :

$$|\nu_j(I) - \nu(I)| = \sup[\int \chi \, d\nu_j - \int \chi \, d\nu \, , \int \chi \, d\nu - \int \chi \, d\nu_j \,]$$

$$\leq \sup[\int g_N \, d\nu_j - \int f_N \, d\nu \, ; \int g_N \, d\nu - \int f_N \, d\nu_j \,] \quad .$$

Ce qui permet d'écrire :

$$|\nu_j(I) - \nu(I)| \leq$$

$$\sup[\,|\int g_N d\nu_j - \int g_N d\nu| + |\int (g_N - f_N) d\nu| \, ; \, |\int f_N d\nu_j - \int f_N d\nu| + |\int (g_N - f_N) d\nu|\,].$$

On obtient ainsi $|\nu_j(I) - \nu(I)| \leq \epsilon$ pour tout $j \geq j_0$.

Démonstration de 2) $\Rightarrow$ 1)

On sait que $\nu_j(I)$ converge vers $\nu(I)$ pour tout intervalle I dont ν ne charge pas les extrémités.

Soit $f \in C_K(\mathbf{R})$. On veut prouver l'assertion suivante :

$$\forall\, \epsilon > 0 \; \exists\, j_0 : \forall\, j \geq j_0 \; |\int f d\nu_j - \int f\, d\nu| \leq \epsilon \quad .$$

Une fonction continue à support compact est limite uniforme d'une suite de fonctions en escalier.
Soir $\epsilon > 0$. Il existe une fonction en escalier ϕ_ϵ telle que $||f - \phi_\epsilon||_\infty < \epsilon/4$.
On peut écrire :

$$\phi_\epsilon = \sum_{k=1}^{p_\epsilon} \alpha_{k,\epsilon}\, \chi_{(a_{k,\epsilon},b_{k,\epsilon})} = \sum_{k=1}^{p_\epsilon} \alpha_{k,\epsilon}\, \chi_{I_{k,\epsilon}} \quad .$$

On obtient alors :

$$\int \phi_\epsilon\, d\nu = \sum_{k=1}^{p_\epsilon} \alpha_{k,\epsilon}\, \nu(I_{k,\epsilon}) \quad ,$$

de sorte que, si la mesure $\nu(\{a_{k,\epsilon}, b_{k,\epsilon}\}_{k=1,\ldots,p_\epsilon})$ est nulle, alors $\nu_j(I_{k,\epsilon})$ converge vers $\nu(I_{k,\epsilon})$ quand j tend vers l'infini, pour tout $k = 1,\ldots,p_\epsilon$, et donc :

$$\forall\, \beta > 0 \; \exists\, j_0 : \forall\, j \geq j_0 \; |\, \nu_j(I_{k,\epsilon}) - \nu(I_{k,\epsilon})\,| < \frac{\beta}{\sum_{k=1}^{p_\epsilon} |\alpha_{k,\epsilon}|} \quad \forall\, k = 1,\ldots,p_\epsilon \quad .$$

On prend $\beta = \epsilon/2$ et on obtient alors $|\int \phi_\epsilon\, d\nu_j - \int \phi_\epsilon\, d\nu| < \epsilon/2$ pour tout $j \geq j_0$, ce qui permet d'écrire :

$$\begin{aligned}|\int f\, d\nu_j - \int f\, d\nu| &\leq |\int (f - \phi_\epsilon)\, d\nu_j| + |\int \phi_\epsilon\, d\nu_j - \int \phi_\epsilon\, d\nu| + |\int (\phi_\epsilon - f)\, d\nu| \\ &\leq 2||f - \phi_\epsilon||_\infty + \epsilon/2 \leq \epsilon \text{ pour } j \geq j_0 \quad .\end{aligned}$$

On a utilisé le fait que ν ne chargeait pas les extrémités des intervalles $I_{k,\epsilon}$. On peut montrer en effet qu'il est toujours possible de choisir ainsi ϕ_ϵ, puisque l'ensemble D des points où $\nu(\{x\})$ est non nul est au plus dénombrable, sinon la famille $\{\nu(\{x\})\}_{x \in D}$ ne serait pas sommable, or sa somme est inférieure ou égale à 1.

Proposition 4.5

Soit $(X_j)_{j \in \mathbf{N}}$ et X des v. a. définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbf{R}^k$. Si la suite X_j converge en probabilités vers X, alors elle converge en loi vers X.

Preuve :

On suppose donc avoir, pour tout $\delta > 0$: $\lim_{n\to\infty} P(\{|X_n - X|\} > \delta) = 0$.
Considérons, pour $f \in C_K$, la quantité $\Delta = |\int f\, d\nu_{X_n} - \int f\, d\nu_X|$.
Δ s'écrit $\Delta = |\int_\Omega (f \circ X_n - f \circ X)\, dP|$.
f étant continue à support compact est uniformément continue.
Soit $\epsilon > 0$. Il existe $\eta = \eta_\epsilon > 0$ tel que $\forall\, x, y\; |x - y| \leq \eta \Rightarrow |f(x) - f(y)| < \epsilon/2$.
On obtient alors :

$$\Delta = \int_{\{|X_n - X| \leq \eta\}} [f(X_n) - f(X)]\, dP + \int_{\{|X_n - X| > \eta\}} [f(X_n) - f(X)]\, dP$$

$$\leq \epsilon/2 \int_\Omega dP \; + \; 2||f||_\infty \; P(\{|X_n - X| > \eta\}) \quad .$$

La convergence en probabilités permet alors de rendre $\Delta \leq \epsilon$ pour tout n suffisamment grand.

Remarque 1 : Une suite de v. a. peut converger en loi sans converger en probabilités.

Exercice 4.4 : exemple

Soient X et Y deux v.a. telles que $\nu_X = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ et $Y = 1 - X$. On considère alors X_n définie par $X_n = Y$ pour tout $n \geq 1$.

Montrer que X_n converge en loi vers X, mais pas en probabilités.

Remarque 2 : La convergence en loi d'une suite $(X_j)_{j\in\mathbf{N}}$ vers une v. a. X n'implique pas la convergence de la suite des espérances vers $E(X)$, même si les espérances existent ! Ce n'est pas étonnant d'ailleurs, puisque la convergence en probabilités n'implique déjà pas celle en moyenne. Pour s'en convaincre, de traiter l'exercice suivant :

Exercice 4.5 : une remarque

Etudier la convergence en loi et la convergence de la suite des espérances, pour les v. a. suivantes :

1) $\nu_{X_n} = \frac{1}{n}\delta_{n^2} + (1 - \frac{1}{n})\delta_0$.

2) $\nu_{X_n} = \frac{1}{n}\delta_n + (1 - \frac{1}{n})\delta_0$.

3) $\nu_{X_n} = \frac{1}{n}\delta_{n(2+(-1)^n)} + (1 - \frac{1}{n})\delta_0$.

Pour terminer ce paragraphe, donnons quelques exercices pour familiariser le lecteur avec la convergence en loi :

Exercice 4.6

On considère la suite de mesures μ_n définies par

$$\mu_n = 2^{-n} \sum_{k=1}^{2^n} \delta_{\frac{k}{2^n}} \quad .$$

Montrer que la suite μ_n converge étroitement vers la mesure de Lebesgue sur $[0,1]$.

Exercice 4.7

Soit μ_n une suite de mesures de densité h_n par rapport à la mesure de Lebesgue sur $\mathbf{R}$. On suppose que h_n converge partout sur $\mathbf{R}$ vers une fonction h, et qu'il existe une fonction $q \in L^1(\mathbf{R})$ telle que l'on ait $|h_n| \leq q$ pour tout $n \in \mathbf{N}$.

1) Montrer que la suite μ_n converge étroitement vers la mesure de densité h par rapport à la mesure de Lebesgue sur $\mathbf{R}$.

Vérifier que h est bien une densité de probabilité.

2) En déduire que la mesure μ_σ de densité égale à la fonction de Gauss g_σ : $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$ converge vers μ_{σ_0} de densité g_{σ_0} quand σ converge vers σ_0 dans $]0,\infty[$.

III - Le théorème de Lévy

Les fonctions caractéristiques vont donner un critère très important de convergence en loi. Nous allons d'abord donner un énoncé général du théorème de Lévy sur $\mathbf{R}^k$.

Théorème 4.6 : 1er énoncé du théorème de Lévy.

Soit $(\nu_j)_{j\in\mathbf{N}}$ une suite de lois de probabilités sur $\mathbf{R}^k$, de fonctions caractéristiques ϕ_j.

Alors les assertions suivantes sont équivalentes :

1) $\exists\ \nu$, loi de probabilité sur $\mathbf{R}^k$, telle que $\int f d\nu_j$ converge vers $\int f d\nu$ pour toute fonction f continue et bornée sur $\mathbf{R}^k$. (ie la suite ν_j converge étroitement vers ν sur $\mathbf{R}^k$)

2) $\exists\ \phi\ : \mathbf{R}^k \ \rightarrow\ \mathbf{C}$, telle que la suite ϕ_j converge vers ϕ uniformément sur tout pavé borné de $\mathbf{R}^k$.

3) $\exists\ \phi\ : \mathbf{R}^k \ \rightarrow\ \mathbf{C}$ continue en 0, telle que la suite ϕ_j converge vers ϕ simplement sur $\mathbf{R}^k$.

Remarque : Ce théorème est fin, car dans l'implication 3) $\Rightarrow$ 1), par exemple, il faut démontrer l'existence d'une mesure, et c'est toujours délicat. Par contre, en pratique, il est fréquent que l'on reconnaisse ϕ comme la fonction caractéristique d'une mesure connue, et c'est en fait un résultat moins fin que l'on utilise. C'est cette version que nous allons énoncer et démontrer.

Théorème 4.7 : 2ème énoncé du théorème de Lévy.

Soit $(\nu_j)_{j\in\mathbf{N}}$ une suite de lois de probabilité sur $\mathbf{R}^k$, de fonctions caractéristiques ϕ_j, et soit ν une mesure de probabilité sur $\mathbf{R}^k$, de fonction caractéristique ϕ.

Alors les assertions suivantes sont équivalentes :

1) La suite ν_j converge étroitement vers ν sur $\mathbf{R}^k$.

2) $\phi_j(t)\ \rightarrow\ \phi(t)\ \forall\ t\in\mathbf{R}^k$.

Autrement dit : Soit $(X_j)_{j\in\mathbf{N}}$ et X des v. a. définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et à valeurs dans $\mathbf{R}^k$. Pour que la suite X_j converge en loi vers X, il est nécessaire et suffisant que la suite des fonctions caractéristiques des X_j converge simplement vers la fonction caractéristique de X.

Preuve :

1) $\Rightarrow$ 2)

Si la suite ν_j converge étroitement vers ν, alors, comme la fonction $x \rightarrow e^{it.x}$ est continue et bornée sur $\mathbf{R}^k$ pour tout $t\in\mathbf{R}^k$, la suite des fonctions caractéristiques $\phi_j(t)\ =\ \int e^{it.x}\ d\nu_j$ converge bien partout sur $\mathbf{R}^k$ vers $\phi(t)\ =\ \int e^{it.x}\ d\nu$.

2) $\Rightarrow$ 1)

Cette assertion est un joli résultat d'analyse de Fourier, que nous allons démontrer en deux étapes : d'abord en prouvant la convergence des intégrales pour une fonction appartenant à $C_0^\infty(\mathbf{R}^k)$, puis en l'étendant par régularisation aux fonctions continues à support compact. Supposons donc avoir $\phi_j(t)\ \rightarrow\ \phi(t)\ \forall\ t\in\mathbf{R}^k$.

Il s'agit de montrer que l'on a $\lim_{j\to\infty}\ \int f\ d\nu_j\ =\ \int f\ d\nu$ pour toute fonction f continue à support compact sur $\mathbf{R}^k$.

A) 1ère étape : $f\in C_0^\infty(\mathbf{R}^k)$.

Supposons f dans $C_0^\infty(\mathbf{R}^k)$. Alors la formule d'inversion de Fourier permet d'écrire :

$$f(x)\ =\ (\frac{1}{2\pi})^k\ \int e^{ix.\xi}\ \hat{f}(\xi)\ d\xi\ \ ,$$

de sorte que $\int f\ d\nu_j$ est égale à :

$$\int f(x)\ d\nu_j(x)\ =\ (\frac{1}{2\pi})^k\ \int[\int e^{ix.\xi}\ \hat{f}(\xi)\ d\xi]\ d\nu_j(x)\ \ .$$

Comme la fonction de 2 variables $(x,\xi) \quad \rightarrow \quad e^{ix.\xi}\ \hat{f}(\xi)$ appartient à $L^1[(\mathbf{R}_\xi{}^k\ ,\ d\xi) \otimes (\mathbf{R_x^k}\ ,\ d\nu_j(x))]$, on peut utiliser le théorème de Fubini pour écrire

$$\int f(x)\ d\nu_j(x) \ = \ (\frac{1}{2\pi})^k \ \int [\int e^{ix.\xi}\ d\nu_j(x)]\hat{f}(\xi)\ d\xi \ = \ \int \phi_j(\xi)\ \hat{f}(\xi)\ d\xi \quad .$$

La convergence ponctuelle des fonctions caractéristiques vers ϕ permet d'appliquer le théorème de convergence dominée pour passer à la limite sous l'intégrale, puisque, grâce à l'inégalité $|\phi_j(\xi)| \leq 1\ \forall \xi \in \mathbf{R}^k$, la fonction sous l'intégrale est dominée par $|\hat{f}|$ qui est intégrable.
Il suffit alors de faire pour ν la démarche inverse de celle qu'on a faite pour ν_j, l'utilisation du théorème de Fubini étant encore licite, justifiée par les mêmes arguments.

B) 2ème étape : $f \in C_K(\mathbf{R}^k)$.

On va procéder par régularisation. On rappelle la construction d'une suite régularisante :
Pour $j \in \mathbf{N}\ :\ g_j(x) \ = \ j^k\ g_1(jx)$ avec $g_1 \in C_0^\infty(\mathbf{R}^k), suppg_1 \subset B(0,1),\ g_1(x) \geq 0$ $\forall x \in \mathbf{R}^k$, $\int g_1(x)\ dx = 1$.
Alors g_j appartient à $C_0^\infty(\mathbf{R}^k)$, est également positive d'intégrale 1, et est à support dans la boule de centre 0 et de rayon $\frac{1}{j}$.

Lemme 4.8

Soit $f \in C_K(\mathbf{R}^k)$. La suite $f * g_j$ appartient à $C_0^\infty(\mathbf{R}^k)$ et converge vers f uniformément sur $\mathbf{R}^k$.

Preuve :

On peut écrire :

$$f * g_j\ (x) - f(x) \ = \ \int [f(x-y) - f(x)]\ g_j(y)\ dy \ = \ j^k \int [f(x-y) - f(x)]\ g_1(jy)\ dy \quad .$$

On effectue le changement de variables défini par $u = jy$ et on obtient :

$$f * g_j\ (x) - f(x) \ = \ \int [f(x - \frac{u}{j}) - f(x)]\ g_1(u)\ du \ = \ \int_{B(0,1)} [f(x - \frac{u}{j}) - f(x)]\ g_1(u)\ du \quad .$$

f étant continue à support compact est uniformément continue, et on a :

$$\forall \epsilon > 0\ \exists \eta\ :\ \forall\ z\ |z| \leq \eta\ \Rightarrow\ |f(x-z) - f(x)| < \epsilon \quad .$$

Alors :

$$\exists\ j_0\ :\ \forall\ j \geq j_0\ \forall u \in\ B(0,1)\ |\frac{u}{j}|\ \leq\ \eta \quad .$$

Et on obtient :

$$|f * g_j\ (x) - f(x)|\ \leq\ \epsilon \int g_1(u)\ du\ \leq\ \epsilon\ \forall\ j \geq j_0 \quad .$$

On a donc prouvé :

$$\forall \epsilon > 0\ \exists\ j_0\ :\ \forall\ j \geq j_0\ ||f * g_j - f||_\infty\ <\ \epsilon \quad .$$

Terminons maintenant la démonstration du théorème de Lévy.
f étant donnée dans $C_K(\mathbf{R}^k)$, on l'approche à $\epsilon/3$ près par $f * g_{j_0}$, j_0 étant l'indice de la dernière assertion dans la preuve du lemme, mais pour $\epsilon/3$ au lieu de ϵ.
$f * g_{j_0} \ = \ f_\epsilon$ appartient à $C_0^\infty(\mathbf{R}^k)$. Alors, pour tout $\alpha > 0$ il existe $N(\alpha, \epsilon)$ tel que

$$|\int f_\epsilon \, d\nu_n \ - \ \int f_\epsilon \, d\nu| \ < \ \alpha \ \forall \, n \ \geq \ N \ .$$

On choisit $\alpha = \epsilon/3$, et on obtient :

$$|\int f \, d\nu_j \ - \ \int f \, d\nu| \ \leq \ |\int (f - f_\epsilon) \, d\nu_j| \ + \ |\int f_\epsilon \, d\nu_j - \int f_\epsilon \, d\nu| \ + \ |\int (f - f_\epsilon) \, d\nu| \ .$$

Le premier terme et le dernier sont majorés chacun par $\epsilon/3$, puisque $\nu_j(\mathbf{R}^k) = \nu(\mathbf{R}^k) = 1$.
Quant au second, il est également majoré par $\epsilon/3$ pour $n \geq N$.
On a donc construit N tel que l'on ait $|\int f \, d\nu_n \ - \ \int f \, d\nu| \ < \epsilon$ pour tout $n \geq N$.

IV - Une application du théorème de Lévy : le théorème "de la limite centrale".

Théorème 4.8 : Théorème "de la limite centrale".

Soient $(X_n)_{n\geq 1}$ une suite de v. a. réelles, indépendantes et équidistribuées (c'est à dire de même loi). On suppose que les v. a. X_n possédent une moyenne m et une variance σ^2. Alors, pour tous a et $b \in \mathbf{R}$, avec $a < b$, on a :

$$\lim_{n\to\infty} \ P[a \leq \frac{X_1 + X_2 + \ldots + X_n - nm}{\sigma\sqrt{n}} \leq b] \ = \ \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \, dx \quad .$$

Preuve :

Pour simplifier on va supposer les v. a. centrées (ie $m = 0$) et réduites (ie $\sigma = 1$). Sinon il suffit de considérer les v. a. centrées réduites associées, à savoir $\frac{X_j - m}{\sigma}$.
Soit $Y_n \ = \ \frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}$. Notons ϕ la fonction caractéristique commune des v. a. X_n.
On a alors :

$$\phi_{Y_n}(u) \ = \ E[e^{iu(\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}})}] \ = \ [E(e^{i\frac{uX_j}{\sqrt{n}}})]^n \ = \ [\phi(\frac{u}{\sqrt{n}})]^n \quad .$$

Les v. a. sont centrées réduites ; on en déduit :

$$\phi(0) = 1 \ \ \phi'(0) = 0 \ \ \phi''(0) = -1 \quad .$$

On peut donc, pour tout u fixé, écrire un développement de Taylor de $\phi(\frac{u}{\sqrt{n}})$ pour n grand :

$$\phi(\frac{u}{\sqrt{n}}) \ = \ 1 \ - \ \frac{u^2}{2n} \ + \ o(\frac{u^2}{n})$$

et

$$\phi_{Y_n}(u) = [1 - \frac{u^2}{2n} + o(\frac{u^2}{n})]^n$$

a pour limite $e^{-\frac{u^2}{2}}$ quand n tend vers l'infini, pour tout $u \in \mathbf{R}$. Pour justifier ce résultat, comme le petit o dans la formule de Taylor est complexe, il suffit de remarquer que l'on peut utiliser la détermination principale du logarithme complexe qui est holomorphe dans un disque de centre 1 et de rayon strictement inférieur à 1. On a alors un développement de Taylor pour le logarithme.
La fonction caractéristique de la v. a. Y_n converge donc simplement vers celle d'une v. a. suivant la loi de Gauss centrée réduite.
Comme cette loi est à densité, elle ne charge aucun point, et la mesure d'un intervalle $[a, b]$ pour la loi de Y_n converge vers celle de ce même intervalle pour la loi de Gauss. C'est ce qu'affirme l'énoncé.

Pour terminer, on va donner une application de ce dernier théorème au jeu de "Pile ou face".
Exercice 4.8 : un test.
On joue 100 fois à "Pile ou face ". On obtient 75 fois "Pile". Peut-on en conclure que la pièce est truquée?
Résolution
On fait "l'hypothèse nulle" (H_0) "La pièce est normale".
On va calculer la probabilité d'avoir autant de "Pile" sous cette hypothèse que lors de l'expérience. Si cette probabilité est "trop faible", c'est à dire inférieure à un seuil α que l'on s'est fixé, on déclarera, quitte à se tromper avec une probabilité inférieure à α, que la pièce est truquée. Si cette probabilité n'est pas "trop petite", c'est à dire est supérieure au seuil de risque α qu'on ne veut pas dépasser, on déclarera que les résultats ne sont pas significatifs, et que l'on ne peut pas dire que la pièce n'est pas normale.
Ce seuil de risque, α, majore la probabilité de se tromper en rejetant une hypothèse qui était vraie.
Soit X_i la v. a. qui au i-ème lancer associe 1 si l'on obtient "Pile", et 0 si on obtient "Face".
Sous l'hypothèse (H_0), les v. a. X_i sont des v. a. de Bernoulli, indépendantes, et de paramètre $\frac{1}{2}$.
La somme des 100 premières suit une loi binomiale de taille 100 et de paramètre $\frac{1}{2}$, et la v. a. centrée réduite associée, Y_{100} est égale à :

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{npq}} \quad \text{soit} \quad Y_{100} = \frac{X_1 + X_2 + \ldots + X_{100} - 50}{\sqrt{25}} \quad .$$

Alors :

$$P(X_1 + X_2 + \ldots + X_{100} \geq 75) = P(Y_{100} \geq 5) \quad .$$

Si on considère que Y_{100} suit approximativement la loi normale centrée réduite, alors la probabilité de cet évènement est donnée par la table de la loi de Gauss :

$$P(Y_{100} \geq 5) \simeq 29.10^{-8}$$

Si on choisit un seuil de risque du type $0,01$, ou $0,005$, ... on décide sans hésiter que la pièce est truquée!
Pour être précis, il faudrait connaître les limites de validité de l'approximation que nous venons de faire. Ce sont des statistiques fines, et cela sort du cadre de ce cours... Mais en pratique les statisticiens donnent comme domaine de validité "$n \geq 50$ et p voisin de $\frac{1}{2}$".
L'approximation est alors excellente.

Nous allons arrêter là ces éléments de Probabilités.
Pour compléter vos connaissances, il reste à voir la démonstration directe de l'approximation "loi binomiale - loi de Poisson" que nous avons rencontrée dans l'exercice sur les fonctions génératrices, et aussi les probabilités conditionnelles qui, si elles ne sont pas explicitement au programme de l'Agrégation, font partie du bagage culturel de tout futur professeur de Mathématiques...

BIBLIOGRAPHIE

P. Barbe et M. Ledoux : Probabilité (Belin)
D. Dacunha-Castelle - M. Duflo : Probabilités et Statistiques Tomes 1 et 2. (Masson)
D. Foata et A. Fuchs : Calcul des probabilités (Masson)
M. Loeve : Probability Theory I (Springer)
M. Métivier : Notions fondamentales de la théorie des probabilités (Dunod)
J. Neveu : Probabilités (Cours à l'Ecole Polytechnique. Edition 95)
J-Y Ouvrard : Probabilités 1 (Cassini)
A. Renyi : Calcul des Probabilités (Dunod)

Table des Matières

Ce polycopié peut être commandé à

l'IREM des Pays de Loire
Université de Nantes — Faculté des Sciences
2 rue de la Houssinière
BP 92208
44322 Nantes cedex 3

TITRE: Eléments de probabilités.

I.R.E.M.: des Pays de Loire (Nantes).

AUTEUR: Anne-Marie CHARBONNEL, Maître de Conférences
à l'Université de Nantes.

DATE: Janvier 2001.

NIVEAU: de la licence à l'agrégation de mathématiques.

PUBLIC CONCERNÉ: Étudiants de deuxième cycle universitaire.

MOTS-CLÉS: Mesures de probabilités,
Variables aléatoires,
Convergence en loi,
Indépendance.

RÉSUMÉ: Après des rappels sur la théorie de l'intégrale de Lebesgue, on présente la notion de variable aléatoire, et ausitôt après celle de fonction caractéristique .
On caractérise ensuite la notion d'indépendance dans le cas de variables aléatoires et d'évènements.
Dans le contexte du schéma de Bernoulli on énonce les lois des grands nombres faible et forte.
Enfin on étudie les différents concepts de convergence permettant de prouver la convergence en loi.

FORMAT	NOMBRE DE PAGES	PRIX	TIRAGE
$21 \times 29{,}7$	41	35,00 F	100 ex.

N° I.S.B.N.: 2 – 86300 – 030 – 6