# INTRODUCTION AU CALCUL DES PROBABILITES

Cours de Licence- Université Paris VI

**L.Mazliak**

# 1 PROBABILITES SUR LES ESPACES FINIS

## 1.1 Qu'est-ce que la probabilité?

Par essence, on ne peut pas avoir une connaissance complète du hasard: il est donc illusoire de vouloir déterminer "exactement" le comportement d'un phénomène aléatoire. Ce qu'on peut faire, par contre, c'est essayer de dégager d'un phénomène une certaine conduite en moyenne, en fréquence et affirmer que tel événement a *tant de chances* de se produire. C'est ici la notion de probabilité.

Deux points de vue sont envisageables:

- ou bien on ne connait rien et le but de l'étude est de donner, au vu des résultats d'une expérience aléatoire, les probabilités qui gouvernent l'expérience aléatoire. Par exemple, en jetant un grand nombre de fois un dé et en inscrivant le nombre de fois qu'on a obtenu chaque chiffre, on constate que ces nombres sont à peu près égaux et donc que la probabilité (fréquence) de chaque face est de 1/6. Ce problème est celui de la Statistique

- ou bien on connait *a priori* la loi qui gouverne le hasard et le but de l'étude est de déterminer le comportement de certains phénomènes (comportement asymptotique par exemple, mais aussi interaction d'un phénomène sur un autre...). C'est là l'objet du calcul des probabilités.

La donnée fondamentale d'un problème de calcul des probabilités est donc la donnée:

- d'un ensemble $\Omega$ représentant les différentes "versions" du hasard.

- d'un objet noté $P$, appelé probabilité, et qui donnera la notion de fréquence des événements citée plus haut.

Nous commencerons par le cas le plus aisé.

## 1.2 Probabilité sur un espace fini

Nous supposerons jusqu'à la fin de ce chapitre que l'espace $\Omega$ est fini, c'est à dire qu'il a un nombre d'éléments (cardinal) fini.

Voici quelques exemples standard:

- jeté d'un dé: on choisit de représenter le hasard par les résultats du jeté, autrement dit par un chiffre de 1 à 6:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- jeté de deux dés: de même ici,

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2$$

- cinq lancers d'une pièce

$$\Omega = \{0, 1\}^5$$

On a alors

**Définition 1.1** *On appelera événement tout sous ensemble de* $\Omega$

En quelque sorte, un événement est un lien entre les différentes versions du hasard, une sélection parmi elles.

<u>Exemple</u>: lancer de deux dés

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2$$

L'événement "la somme des lancers est paire" est

$$S = \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),$$
$$(4,2),(4,4),(4,6),(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\}$$

On va définir la probabilité naturellement de la façon suivante:

**Définition 1.2** *On appelle probabilité sur* $\Omega$ *toute application*

$$P : \mathcal{P}(\Omega) \longrightarrow [0, 1]$$

*où* $\mathcal{P}(\Omega)$ *est l'ensemble des parties de* $\Omega$ *telle que*

*(i)* $P(\Omega) = 1$ *(normalisation)*

*(ii) si* $(A, B) \in \mathcal{P}(\Omega)^2, A \cap B = \emptyset$,

$$P(A \cup B) = P(A) + P(B)$$

<u>Remarques</u>: 1) Si $A_1, A_2, \ldots, A_k$ sont disjoints,

$$P(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i).$$

2) Si $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, on peut poser $p_i = P(\{\omega_i\})$ et on remarque:

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} P(\{\omega_i\}) = P(\bigcup_{i=1}^{n}\{\omega_i\}) = P(\Omega) = 1.$$

Ceci permet de donner une définition équivalente à la précédente

**Définition 1.3** *Une probabilité sur l'ensemble* $\Omega = \{\omega_1, \ldots, \omega_n\}$ *est la donnée de* $(p_1, p_2, \ldots, p_n) \in [0,1]^n$ *tel que* $\sum_{i=1}^{n} p_i = 1$

La probabilité d'un événement est alors définie comme la somme des $p_i$ correspondants aux $\omega_i$ dans cet événement.

Un exemple fondamental est celui de la <u>probabilité uniforme</u>

$$\Omega = \{\omega_1, \ldots, \omega_n\}, p_i = \frac{1}{n}$$

Soit $A$ un événement; sa probabilité est donnée par $P(A) = \frac{\mid A \mid}{n}$

Dans le cas d'un dé pipé, on peut avoir

$$p_1 = p_2 = p_3 = p_4 = p_5 = \frac{1}{10}, p_6 = \frac{1}{2}$$

## 1.3 Variables aléatoires

Au lieu de s'intéresser directement à l'événement aléatoire, on va s'occuper de sa lecture à travers un effet qu'il produit. C'est un peu la même démarche que suit quelqu'un qui se levant le matin et voulant connaître la température extérieure, lit son thermomètre plutôt que de sortir.Il s'agit là de la notion fondamentale du calcul des probabilités: celle du transport du hasard.

**Définition 1.4** *On appelle variable aléatoire (v.a. en abrégé) à valeurs dans l'ensemble* $E$*, toute application de* $\Omega$ *dans* $E$*. Si* $E = \mathbf{R}$*, la variable aléatoire est dite réelle (v.a.r.)*

<u>Exemples</u>: - jeté de deux dés

$X$ ="somme des points" est définie par

$$X((x,y)) = x + y$$

-cinq lancers d'une pièce (0 représente pile et 1 face)

$X$ ="nombre de "face" obtenus"

$$X(x_1, x_2, x_3, x_4, x_5) = \sum_{i=1}^{5} x_i$$

## 1.4 Loi d'une variable aléatoire

C'est ici la véritable notion de transport.

Soit $(\Omega, P)$ donné et $X : \Omega \to E$ une variable aléatoire.

Notons $F(\subset E)$ l'ensemble des valeurs prises par $X$.

Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ et $F = \{x_1, \dots x_k\}$

Posons $(X = x_i) = \{\omega \in \Omega, X(\omega) = x_i\} = X^{-1}(\{x_i\})$ et $\alpha_i = P(X = x_i)$.

On remarque

$$\sum_{i=1}^{k} \alpha_i = \sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} P(\{\omega \in \Omega, X(\omega) = x_i\})$$

$$= P(\bigcup_{i=1}^{k} \{\omega \in \Omega, X(\omega) = x_i\})$$

(puisque ce sont des ensembles disjoints)

$$= P(\Omega) = 1$$

Donc $(\alpha_1, \dots, \alpha_k)$ définit une probabilité sur $F$.

**Définition 1.5** *On appelle loi de la variable aléatoire $X$ la probabilité $\mu_X$ sur $F$ définie ci-dessus*

Exemple: Considérons une fois de plus le jeté de deux dés. Soit $X$ la variable "somme". Sa loi sera une probabilité sur l'ensemble $\{2, 3, 4, \dots, 10, 11, 12\}$. Par exemple,

$$\mu_X(\{3\}) = P(X^{-1}(3)) = P(\{(1,2), (2,1)\}) = \frac{2}{36} = \frac{1}{18}$$

## 1.5 Espérance et variance d'une v.a.r

Placons nous dans le cas précédent mais avec $E = \mathbf{R}$. On va définir une notion de moyenne stochastique de la variable aléatoire. Cette moyenne sera un barycentre des valeurs de la v.a.r. pondérées de poids égaux aux probabilités qu'elle prenne ces valeurs.

On arrive donc à la définition suivante:

**Définition 1.6** *On appelle espérance de la variable réelle $X$ à valeurs dans un ensemble fini de réels le nombre réel*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

**Propriété 1.7** *(i)* $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$ *(linéarité)*

*(ii) Si* $X \geq 0, E(X) \geq 0$

*(iii)Si* $X \geq 0$ *et* $E(X) = 0$, *alors* $P(X \neq 0) = 0$

*(iv)* $E(\mid X \mid) \geq \mid E(X) \mid$

La variance de $X$ est définie comme la moyenne quadratique des écarts à la moyenne:

$$\mathrm{Var}(X) = E((X - E(X))^2)$$

L'écart-type est $\sigma(X) = \sqrt{\mathrm{Var}(X)}$

**Propriété 1.8** *(i)* $\mathrm{Var}(X) \geq 0$

*(ii)* $\mathrm{Var}(X) = E(X^2) - E(X)^2$

*(iii)* $\mathrm{Var}(X) = 0 \iff P(X \neq E(X)) = 0$

Exercice: Calculer espérance et variance des v.a.r évoquées plus haut.

La propriété suivante relie le calcul de l'espérance et la loi d'une variable aléatoire.

**Propriété 1.9** *Soit $X$ une variable réelle à valeurs dans $F = \{x_1, \ldots, x_k\}$ et soit $f$ : $X \to \mathbf{R}$ une fonction. On a alors*

$$E(f(X)) = \sum_{x \in F} f(x) P(= x)$$

démonstration: On a $\Omega = \bigcup_{j=1}^{k} (X = x_j)$ et donc

$$E(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}) =$$

$=$

$$= \sum_{j=1}^{k} f(x_j) \sum_{\omega \in \Omega, X(\omega) = x_j} P(\{\omega\}) = \sum_{j=1}^{k} f(x_j) P(X = x_j) \quad \square$$

## 1.6 Indépendance

On considère toujours donné l'espace $\Omega$ et $P$ une probabilité sur lui.

L'indépendance de deux événements $A$ et $B$ va correspondre à une "démultiplication" des cas favorables.

Précisément:

**Définition 1.10** *On dit que deux événements $A$ et $B$ sont indépendants si $P(A \cap B) = P(A).P(B)$*

<u>Exemple</u>: Dans le lancer de deux dés, on vérifie que les événements $A$ = "le premier dé donne un résultat impair" et $B$ = "la somme des dés est paire" sont indépendants.

**Définition 1.11** *On dit que $X$ et $Y$ à valeurs $\{x_1, \ldots, x_n\}$ (resp. $\{y_1, \ldots, y_p\}$) sont deux variables aléatoires indépendantes si et seulement si les événements $(X = x_i)$ et $(Y = y_j)$ sont indépendants quels que soient $i$ et $j$*

Plus généralement,

**Définition 1.12** *$A_1, \ldots, A_n$ sont indépendants si $\forall F \subset \{1, \ldots, n\}$,*

$$P(\bigcap_{j \in F} A_j) = \prod_{j \in F} P(A_j)$$

**Propriété 1.13** *(i) Si $X$ et $Y$ sont deux variables aléatoires réelles indépendantes,*

$$E(XY) = E(X)E(Y)$$

*(ii) Si $X$ et $Y$ sont deux variables indépendantes,*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Démonstration:

(i) Si $X$ et $Y$ sont deux indicatrices $\mathbf{1}_A$ et $\mathbf{1}_B$, le résultat n'est autre que la définition de l'indépendance. Il suffit ensuite de remarquer que $X = \sum_j x_j \mathbf{1}_{X=x_j}$ et $Y = \sum_j y_j \mathbf{1}_{X=y_j}$ et d'appliquer la linéarité de l'espérance.

(ii) On a $\mathrm{Var}(X+Y) = E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)^2 - 2E(X)E(Y)$ et ceci est égal à $\mathrm{Var}(X) + \mathrm{Var}(Y)$ en utilisant (i). $\square$

## 1.7 Conditionnement

Si on connait une information sur un événement $A$, comment peut-on l'utiliser pour dire quelque chose de l'événement $B$? Plus précisément, si on sait que $A$ se réalise, comment évaluer les chances de $B$?

L'idée va être de limiter notre espace à $A$ (i.e. prendre $A$ comme notre nouvel espace de probabilités) en le munissant de la probabilité

$$Q(C) = \frac{P(A \cap C)}{P(A)}$$

**Définition 1.14** *Soit $A$ un événement tel que $P(A) > 0$. On appelle probabilité conditionnelle de $B$ sachant $A$ la quantité*

$$Q_A(B) = P(B/A) = \frac{P(A \cap B)}{P(A)}$$

**Propriété 1.15** *Si $A$ et $B$ sont indépendants, $P(B/A) = P(B)$*

Exercice: Dans le jeté de deux dés, calculer $P(A/B)$ avec $A =$"le premier dé est inférieur ou égal à 3" et $B =$"les deux dés sont impairs"

## 1.8 Deux exemples importants de lois

a- Loi de Bernoulli de paramètre $0 \leq p \leq 1$

C'est la loi d'une variable aléatoire $X$ prenant deux valeurs 0 et 1 avec

$$P(X = 1) = p, P(X = 0) = 1 - p = q$$

b- Loi binômiale

**Définition 1.16** *Soient $n \in \mathbf{N}^*$ et $0 \leq p \leq 1$. La loi binômiale $\mathcal{B}(n, p)$ est la loi sur $\{0, 1, \ldots, n\}$ définie par*

$$0 \leq i \leq n, p_i = C_n^i p^i (1 - p)^{n-i}$$

La propriété importante est alors la suivante.

**Proposition 1.17** *Soient* $(X_1, \ldots, X_n)$ $n$ *variables de Bernoulli de paramètre* $p$ *indépendantes. Alors* $Y = \sum_{i=1}^{n} X_i$ *suit une loi* $\mathcal{B}(n,p)$

Démonstration: Soit $0 \leq k \leq n$. Pour que $Y = k$, il faut que parmi les $X_i$, $k$ vaillent 1 et $n-k$, 0. Si les indices valant 1 sont fixés, on a un événement de probabilité $p^k(1-p)^{n-k}$. Il y a $C_k^n$ possibilités de choisir $k$ indices parmi $n$, et l'événement résultant (à savoir $Y = k$) est donc de probabilité $C_n^k p^k (1-p)^{n-k}$. $\square$

# PROBABILITES DISCRETES

On va commencer ici à généraliser les notions précédentes au cas dénombrable. Pour commencer, nous allons donner une définition générale d'un espace de probabilité.

## 1- Espace de probabilités.

Soit $\Omega$ un ensemble quelconque. Pour des raisons techniques, mais fondamentales!, qui dépassent pour l'instant le cadre de notre propos, on ne va s'interesser, comme événements, qu'à des sous-ensembles de $\Omega$ qui appartiennent à une certaine structure, appelée tribu.

> **Définition**: *Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{A}$ est une tribu sur $\Omega$ si*
> *(i)* $\emptyset \in \mathcal{A}$
> *(ii)* $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
> *(iii)* $(A_n) \in \mathcal{A} \Rightarrow \bigcup_{n \in I\!N} A_n \in \mathcal{A}$

<u>Remarque</u>: Cette notion n'interviendra plus désormais et jusqu'à nouvel avis qu'implicitement. On supposera toujours notre espace $\Omega$ muni d'une tribu suffisement riche pour contenir tous les événements que l'on aura besoin de considérer.

La notion de tribu va intervenir dans la définition générale d'un espace de probabilités.

> **Définition**: *On dit que $(\Omega, \mathcal{A}, P)$ est un espace de probabilités si $P$ est une application de $\mathcal{A}$ dans $[0,1]$ telle que*
> *(i)*$P(\Omega) = 1$ *(ii) si* $(A_n) \in \mathcal{A}^{I\!N}, A_i \cap A_j = \emptyset, \forall i \neq j$,
>
> $$P(\bigcup_{n \in I\!N} A_n) = \sum_{n \in I\!N} P(A_n)$$

<u>Remarques</u>: 1) Notez qu'il s'agit bien d'une généralisation de la notion précédente avec $\mathcal{A} = \mathcal{P}(\Omega)$ et $P$ définie sur la tribu $\mathcal{P}(\Omega)$
2) Quand $\Omega$ est dénombrable, i.e. $\Omega = \{x_1, \ldots, x_n\}$, on supposera toujours implicitement que $\mathcal{A} = \mathcal{P}(\Omega)$. Dans ce cas, une probabilité sur $\Omega$ est donnée par une suite $(p_n)$ de réels positifs telle que $\sum_n p_n = 1$ avec $p_n = P(\{x_n\}$.

## 2- Variables aléatoires discrètes

Soit $E$ un ensemble dénombrable, soit $E = \{x_n\}_{n \in I\!N}$

> **Définition**: *Une variable aléatoire à valeurs dans $E$ est une application $X : \Omega \to E$ telle que*
>
> $$\forall n \in I\!N, (X = x_n) = \{\omega \in \Omega, X(\omega) = x_n\} \in \mathcal{A}$$

<u>Exemple-Exercice</u>: Dans un jeu (infini) de pile ou face, on peut considérer la variable $X =$"numéro du coup où face sort pour la première fois" (le lecteur trouvera en exercice $\Omega$ et une tribu sur $\Omega$ de telle sorte que $X$ soit bien une variable aléatoire).

La loi de la variable aléatoire $X$ se définira sans problème comme dans le cas fini.

**Définition**: *Soit $X$ une variable aléatoire à valeurs dans $E$. Posant*

$$p_n = P(X = x_n) = P(\{\omega, X(\omega) = x_n\}),$$

*on définit une probabilité $\mu_X$ sur $E$ dite la loi de $X$.*

3- Espérance

Nous allons généraliser ici la notion de moyenne introduite dans le cas fini. Cependant, comme ici les sommes intervenant sont *a priori* infinies, il faut prendre des précautions supplémentaires.

**Définition**: *Soit $X$ une variable aléatoire à valeurs dans $E \subset \mathbb{R}$, de loi $p_n = P(X = x_n)$.*
*On dira que $X$ admet une espérance (ou un moment d'ordre 1) si la série*

$$\sum_{n\in\mathbb{N}} \mid x_n \mid p_n$$

*est convergente. Dans ce cas on pose*

$$E(X) = \sum_{n\in\mathbb{N}} x_n . p_n$$

<u>Convention</u>: <u>**Si** $X \geq 0$</u>, et si la somme précédente n'est pas finie, on posera $E(X) = +\infty$ (ceci permet de définir dans tous les cas l'espérance d'une variable positive).

Voici une première propriété essentielle de l'espérance.

**Proposition**: *Soit $X$ une variable aléatoire à valeurs dans $E$ et soit $\varphi$ une application de $E$ dans $\mathbb{R}$ positive ou bornée. Alors on a*

$$E(\varphi(X)) = \sum_{k=0}^{\infty} \varphi(x_k) P(X = x_k)$$

<u>démonstration</u>: Notons $(y_i)_{i\geq 0}$ les différentes valeurs prises par $\varphi$ sur $E$ et

$$\varphi^{-1}(y_i) = \{x_{\alpha^i(k)}\}_{k\geq 0}$$

On a

$$E(\varphi(X)) = E(Y) = \sum_{i=0}^{\infty} y_i P(Y = y_i) =$$

$$= \sum_{i=0}^{\infty} y_i P(\varphi(X) = y_i) = \sum_{i=0}^{\infty} y_i P(X \in \varphi^{-1}(y_i)) =$$

$$= \sum_{i=0}^{\infty} y_i \sum_{k=0}^{\infty} P(X = x_{\alpha^i(k)}) = \sum_{k=0}^{\infty} \sum_{i=0} y_i P(X = x^i(k)) =$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \varphi(x^i(k)) P(X = x^i(k)) =$$

$$= \sum_{l=0}^{\infty} \varphi(x_l) P(X = x_l) \quad \square$$

Remarque: Noter que la dernière égalité est licite par sommation par paquets justifiée par la convergence absolue de la série. Cette convergence absolue est donc une hypothèse *essentielle* pour bien définir l'espérance.

Plus généralement, on dira que la variable aléatoire réelle $X$ admet un moment d'ordre $n$ si

$$\sum_{k \geq 0} | x_k |^n p_k < \infty$$

On a évidemment

$$E(X^n) = \sum_{k \geq 0} x_k^n p_k$$

La variance de $X$ est définie si $X$ admet un moment d'ordre 2

$$\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

Remarque: Ceci demande une petite démonstration laissée en exercice. Il faut prouver que

$$E(X^2) < \infty \Rightarrow E(| X |) < \infty$$

4- Indépendance

La notion d'indépendance de deux événements ou de deux variables aléatoires reste identique que précédemment. Enonçons simplement une proposition fondamentale

**Proposition**: *Soient $X_1, \ldots, X_n$ des variables aléatoires à valeurs dans les espaces dénombrables $E_1, E_2, \ldots, E_n$. Alors on a l'équivalence entre*
*(i) $(X_1, \ldots X_n)$ indépendantes*
*(ii) $\forall \varphi_i : E_I \to I\!R$ positive bornée,*

$$E(\prod_{i=1}^{n} \varphi_i(X_i)) = \prod_{i=1}^{n} E(\varphi_i(X_i))$$

démonstration: Nous nous limitons au cas $n = 2$, la généralisation résultant d'une récurrence immédiate. Notons $E_i = \{x_1^i, x_2^i, \dots, x_{k_i}^i\}$
(i)⇒(ii): Soient $\varphi_1, \varphi_2$. On a

$$\varphi_i = \sum_k \varphi_i(x_k^i) \mathbb{1}_{x_k^i}$$

Alors

$$E(\varphi_1(X_i))\varphi_2(X_i)) = \sum_{j,k} \varphi_1(x_j^i)\varphi_2(x_k^i) E(\mathbb{1}_{X_1=x_j^1} \mathbb{1}_{X_2=x_k^2})$$

Or, par l'indépendance,

$$E(\mathbb{1}_{X_1=x_j^1} \mathbb{1}_{X_2=x_k^2}) = P(X_1 = x_j^1, X_2 = x_k^2) = P(X_1 = x_j^1)P(X_2 = x_k^2) = E(\mathbb{1}_{X_1=x_j^1})E(\mathbb{1}_{X_2=x_k^2})$$

et donc

$$E(\varphi_1(X_i))\varphi_2(X_i)) = \sum_{j,k} \varphi_1(x_j^i)\varphi_2(x_k^i) E(\mathbb{1}_{X_1=x_j^1})E(\mathbb{1}_{X_2=x_k^2}) =$$

$$= E(\sum_j \varphi_1(x_j^i)\mathbb{1}_{X_1=x_j^1})E(\sum_k \varphi_2(x_k^i)\mathbb{1}_{X_2=x_k^2}) =$$

$$= E(\varphi_1(X_i))E(\varphi_2(X_i))$$

(ii)⇒(i): On applique simplement la propriété (ii) à

$$\varphi_1 = \mathbb{1}_{x_j^1}, \varphi_2 = \mathbb{1}_{x_k^2}$$

pour obtenir

$$P(X_1 = x_j^1, X_2 = x_k^2) = P(X_1 = x_j^1)P(X_2 = x_k^2)$$

5- Exemples importants de lois

Soit $X$ une variable aléatoire.

a- Loi géométrique

> **Définition**: *On dit que $X$ suit la loi géométrique de paramètre $0 < a < 1$ si $X$ est à valeurs dans $\mathbb{N}$ et*
>
> $$P(X = n) = (1 - a)a^n$$

b- Loi de Poisson

> **Définition**: *Une variable aléatoire à valeurs dans $\mathbb{N}$ suit la loi de Poisson de paramètre $\lambda$ si*
>
> $$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

c- Somme de deux variables aléatoires indépendantes

Soient $X$ et $Y$ deux variables aléatoires réelles indépendantes. Notons, avec des notations claires,

$$P(X = x_i) = p_i, P(Y = y_j) = q_j.$$

La loi de $X + Y$ est le produit de convolution des suites $(p_i)$ et $(q_j)$ défini par

$$P(X + Y = z_k) = r_k = (p_i) * (q_j)(k) = \sum_{z_k = x_i + y_j} p_i q_j$$

Dans le cas de deux variables $X$ et $Y$ à valeurs dans $\mathbb{N}$, l'expression devient plus sympathique

$$P(X + Y = k) = \sum_{i+j=k} p_i q_j = \sum_{i=0}^{k} p_i q_{k-i}$$

6- Fonctions génératrices

Soit $X$ une variables aléatoire à valeurs dans $\mathbb{N}$. on note $p_k = P(X = k)$.

**Définition**: *On appelle fonction génératrice de $X$ la fonction $f_X : [-1, 1] \to \mathbb{R}$ définie par*

$$f_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k p_k$$

**Propriétés**: *(i) La fonction $f_X$ est $C^\infty$ dans l'intervalle $]-1, 1[$*
*(ii) On a $\forall k \geq 0$,*

$$p_k = \frac{f_X^{(k)}(0)}{k!}$$

*En particulier, $f_X$ caractérise la loi de $X$*

Pour la suite, nous aurons besoin d'un lemme technique dont la démonstration est laissée en exercice.

**Lemme**: *Soient $(a_n^k)_{n \geq 0}$ des suites positives telles que pour tout $n$ fixé, la suite $(a_n^k)_{k \geq 0}$ croit vers $a_n$. Alors*

$$\lim_{k \to \infty} \sum_{n \geq 0} a_n^k = \sum_{n \geq 0} a_n$$

On a alors

**Proposition**: *La fonction génératrice $f_X$ permet de déterminer les moments de $X$*

Moyen mnémotechnique: $f_X(s) = E(s^X)$
$f'_X(s) = E(X s^{X-1})$ donc $f'_X(1) = E(X)$ ...
$f_X^{(k)}(s) = E(X(X-1)\dots(X-k+1)s^{X-k})$ donc $f_X^{(k)}(1) = E(X(X-1)\dots(X-k+1))$
d'où se déduit $E(X^k)$ de proche en proche.

Vérifions simplement que $E(X) = \lim_{s\to 1} f'_X(s)$. Soit $(s_k)$ une suite dans $]-1,1[$ qui tend vers 1. Avec les notations précédentes, $f'_X(s_k) = \sum_n p_n s_k^n$ et on peut alors appliquer le lemme précédent à $a_n^k = p_n s_k^n$ $\square$

Exemples de fonctions génératrices:

- Loi de Poisson

**Proposition**: *Soit $X$ une variable de loi de Poisson de paramère $\lambda$. Alors*

$$f_X(s) = c^{-\lambda} \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} = e^{\lambda(s-1)}$$

- Loi géométrique de paramètre $a$

**Proposition**: *Soit $X$ tel que $P(X=n) = (1-a)a^n$ La fonction génératrice est donnée par*

$$f_X(s) = \sum_{k=0}^{\infty} (1-a)a^k s^k = (1-a)\sum_{k=0}^{\infty} (sa)^k = \frac{1-a}{1-sa}$$

- Loi binômiale

**Proposition**: *Soit $X$ tel que $P(X=k) = C_n^k p^k (1-p)^{n-k}, 0 \le k \le n$, on a*

$$f_X(s) = (ps + (1-p))^n$$

- Loi de Bernoulli

**Proposition**: *Soit $X$ une variable de loi de Bernoulli de paramètre $p$. On a*

$$f_X(s) = ps + (1-p)$$

Application au calcul de la loi de la somme de deux variables aléatoires indépendantes.

**Proposition**: *Soient $X$ et $Y$ deux variables aléatoires réelles indépendantes à valeurs dans $\mathbb{N}$. Alors $f_{X+Y}(s) = f_X(s) f_Y(s)$*

démonstration:

$$f_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) =$$
$$= E(s^X)E(s^Y) = f_X(s) f_Y(s) \quad \square$$

vskip 2mm Exemple: Soient $X_1, \ldots, X_n$ $n$ variables de Bernoulli indépendantes, de paramètre $p$. La fonction génératrice de $X_1 + \ldots X_n$ est

$$f_{X_1+\ldots+X_n}(s) = \prod_{i=1}^{n} f_{X_i}(s) = (ps + (1-p))^n$$

donc $X_1 + \ldots + X_n$ suit une loi binômiale de paramètres $n$ et $p$.

7- Convergence des suites de variables aléatoires

La définition générale de la convergence restant identique quel que soit l'ensemble où les variables aléatoires prennent leurs valeurs, nous l'énoncerons dans le cas général avant de nous intéresser plus explicitement au cas où $E$ est dénombrable.

**Définition**: *Soit $(X_n)$ une suite de variables aléatoires à valeurs dans un espace métrique $(E,d)$ séparable et complet.*
*1) On dit que $(X_n)$ converge en probabilités vers la variable aléatoire $X_\infty$ si*

$$\forall a > 0, P(d(X_n, X_\infty) > a) \to_{n\to\infty} 0$$

*2) On dit que $(X_n)$ converge en loi vers $X_\infty$ si pour toute fonction $\varphi : E \to \mathbb{R}$, continue et bornée,*

$$E(\varphi(X_n)) \to_{n\to\infty} E(\varphi(X_\infty))$$

On a alors le lien suivant entre les deux convergences

**Proposition**: *Si $(X_n)$ converge en probabilités vers $X_\infty$, alors $(X_n)$ converge en loi vers $X_\infty$*

démonstration: Soit $\varphi$ continue et bornée de $E$ dans $\mathbb{R}$. Notons qu'alors $\varphi$ est uniformément continue sur $E$ (pourquoi?).
Soit $\varepsilon > 0$.
On peut trouver $\alpha > 0$ tel que $\forall x, y \in E$, $d(x,y) \leq \alpha \Rightarrow \mid \varphi(x) - \varphi(y) \mid < \varepsilon$. Puisque on a la convergence en probabilités, il existe $N$ tel que $\forall n \geq N$, $P(d(X_n, X_\infty) > \alpha) < \varepsilon$. On peut alors écrire

$$\mid E(\varphi(X_\infty) - \varphi(X_n)) \mid \leq E(\mid \varphi(X_\infty) - \varphi(X_n) \mid) \leq$$

$$\leq E(\mid \varphi(X_\infty) - \varphi(X_n) \mid \mathbb{1}_{d(X_n,X_\infty)\leq\alpha}) + E(\mid \varphi(X_\infty) - \varphi(X_n) \mid \mathbb{1}_{d(X_n,X_\infty)>\alpha}) \leq$$

$$\varepsilon + 2MP(d(X_n, X_\infty) > \alpha) \leq$$

$$\varepsilon + 2M\varepsilon$$

où $M$ désigne un majorant de $\varphi$. $\square$

Nous supposons désormais que $X_n$ et $X_\infty$ sont des variables à valeurs dans $\mathbb{N}$, i.e. $E = \mathbb{N}$. Notez qu'alors toute fonction définie sur $E = \mathbb{N}$ est continue.

**Proposition**: *$X_n \to X_\infty$ en loi si et seulement si*

$$\forall k \in \mathbb{N}, \lim_n P(X_n = k) = P(X_\infty = k)$$

démonstration: Supposons d'abord que $X_n \to X_\infty$ en loi. On considére $\varphi = \mathbb{I}_{\{k\}}$ et on obtient $\lim_n P(X_n = k) = P(X_\infty = k$. Réciproquement, soit $\varphi$ une fonction bornée sur $\mathbb{N}$. On a

$$E(\varphi(X_n)) = \sum_k \varphi(k) P(X_n = k)$$

et de même

$$E(\varphi(X_\infty)) = \sum_k \varphi(k) P(X_\infty = k)$$

Par convergence de la série, on peut trouver $K$ tel que

$$\sum_{k \geq K} P(X_\infty = k) < \varepsilon$$

. Par hypothèse, on peut trouver $N$ tel que $\forall n \geq N, \forall k \leq K$,

$$\mid P(X_n = k) - P(X_\infty = k) \mid \leq \varepsilon$$

Alors $\forall n \geq N$,

$$\sum_{k \geq K} P(X_n = k) < (K+1)\varepsilon$$

. Alors,

$$\mid E(\varphi(X_n) - \varphi(X_\infty)) \mid \leq$$

$$\sum_k \varphi(k) \mid P(X_n = k) - P(X_\infty = k) \mid \leq$$

$$\sum_{k \leq K} \varphi(k) \mid P(X_n = k) - P(X_\infty = k) \mid + \sum_{k > K} \varphi(k) \mid P(X_n = k) - P(X_\infty = k) \mid \leq$$

$$MK\varepsilon + M(K+1)\varepsilon \quad \square$$

On a également l'importante caractérisation suivante

**Théorème**: *On a l'équivalence entre les deux assertions suivantes*
*(i)* $X_n \to X_\infty$ *en loi*
*(ii)* $\forall s \in ]0,1[, f_{X_n}(s) \to f_{X_\infty}(s)$

démonstration:
(i)⇒(ii) est une simple application de la définition avec $\varphi(k) = s^k, s \in ]0,1[$.
(ii)⇒(ii): Supposons (ii). On va montrer qu'alors $P(X_n = 0) = f_{X_n}(0) \to P(X_\infty = 0) = f_{X_\infty}(0)$ et on procèdera ensuite par une récurrence laissée en exercice.
Soit $\varepsilon > 0$. Soit $0 < s < a < 1$. On a

$$f_{X_n}(s) - f_{X_n}(0) \leq \sum_{k \geq 1} \frac{s}{a} P(X_n = k) a^k \leq \frac{s}{a} f_{X_n}(a)$$

. Or, $f_{X_n}(a) \to f_{X_\infty}(a)$. Donc, $\exists \sigma > 0$,

$$0 \leq s \leq \sigma \Rightarrow f_{X_n}(s) - f_{X_n}(0) < \varepsilon, \forall n$$

et $f_{X_\infty}(s) - f_{X_\infty}(0) < \varepsilon$. Alors

$$| f_{X_n}(0) - f_{X_\infty}(0) | \leq | f_{X_n}(0) - f_{X_n}(s) | + | f_{X_n}(s) - f_{X_\infty}(s) | + | f_{X_\infty}(s) - f_{X_\infty}(0) | <$$

$$< 2\varepsilon + | f_{X_n}(s) - f_{X_\infty}(s) |, \forall 0 \leq s \leq \sigma$$

et pour $s$ fixé, $\exists N, \forall n \geq N, | f_{X_n}(s) - f_{X_\infty}(s) | < \varepsilon$ $\square$

On va maintenant s'intéresser à une suite particulière de variables aléatoires: la suite des moyennes arithmétiques d'une suite de variables indépendantes et de même loi.

**Théorème (Loi faible des Grands Nombres**: *Soit* $(X_k)_{k \geq 0}$ *une suite de variables aléatoires entières indépendantes et de même loi. On suppose que* $E(X_1^2) < \infty$. *Alors, si* $S_n = X_1 + \ldots X_n$, *on a* $\dfrac{S_n}{n} \to E(X_1)$ *en probabilités*

démonstration: Notons $m = E(X_1)$ et $\sigma^2 = \mathrm{Var}(X_1)$. Notons également $Y_n = \dfrac{S_n}{n}$. On a $m = E(Y_n)$ et, par indépendance, $\dfrac{1}{n}\sigma^2 = \mathrm{Var}(Y_n)$. Utilisant l'inégalité de Bienaymé-Tchebitcheff, on a pour tout $\varepsilon > 0$,

$$P(| Y_n - m | > \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2$$

et cette dernière quantité tend vers 0 avec $n$ $\square$.

<u>Application</u>: Soit $A$ une partie de $\mathbb{N}$ et soit $(X_k)$ une suite de variables aléatoires de mêmes lois et indépendantes. Notons

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_A(X_k)$$

la fréquence de réalisation de $A$ dans la suite avant l'instant $n$. Alors on a $Y_n \to P(X_1 \in A)$ et on retrouve ainsi la définition intuitive d'une probabilités en terme de fréquence d'un événement.

En fait, on a un résultat beaucoup plus fort, dit loi forte des grands nombres.

**Théorème**: *Si* $X_1, X_2, \ldots$ *sont indépendantes et de même loi et d'espérance finie, alors* $\dfrac{S_n}{n} \to E(X_1)$ *avec probabilité 1*

Nous ne démontrerons ce résultat difficile que dans le cas élémentaire où les variables $X_i$ sont bornées.

démonstration: On suppose donc qu'il exite $M$ tel que $| X_i | \leq M, \forall i$. Remplaçant $X_i$ par $Y_i = X_i - m$, on se ramène au cas $E(X_1) = 0$. On a

$$E(S_n^4) = nE(X_1^4) + 3n(n-1)E(X_1^2) \leq K.n^2$$

d'où

$$P(| S_n | \geq n\varepsilon) = P(| S_n |^4 \geq n^4\varepsilon^4) \leq \frac{E(| S_n |^4)}{n^4\varepsilon^4} \leq \frac{K}{n^2\varepsilon^4}$$

d'où $\sum_{n\geq 1} P(| S_n | \geq n\varepsilon) < \infty$ et par le lemme de Borel-Cantelli (voir exercices), $P(\exists N, \forall n \geq N, \frac{| S_n |}{n} < \varepsilon) = 1$. On prend alors l'intersection de ces ensembles sur les $\varepsilon$ rationnels pour conclure. $\square$

# CHAINES DE MARKOV DISCRETES

## 1- Présentation

Quand on regarde l'évolution d'un phénomè ne aléatoire, une propriété importante qu'on peut chercher à traduire, c'est que la dépendance entre le futur de l'instant $n$ où l'on se trouve et le passé de cet instant, ne se fait qu'à travers la valeur observée à l'instant $n$. On résume celà en la superbe formule: " le futur ne dépend du passé que par le présent". Notez qu'il s'agit d'une propriété déjà rencontrée dans d'autres domaines mathématiques (par exemple les équations différentielles). D'un autre côté, il ne faut pas perdre de vue qu'il s'agit d'une hypothè se trè s "concrè te" (on en verra des exemples, entre autres celui de l'évolution de la fortune d'un joueur à travers le temps).
Evidemment, comme on va s'interesser à un phénomè ne aléatoire, la valeur $X_n$ prise à l'instant $n$ ne décidera pas de $X_{n+1}$ mais de la *loi* de $X_{n+1}$.
Ceci mè ne à la définition suivante, où nous nous limitons bien sûr au cas dénombrable.

**Définition**: *Soit* $(X_n)$ *une suite de variables aléatoires à valeurs dans un ensemble* $E = \{x_k\}_{k\geq 0}$. *On dit que* $X_n$ *est une chaîne de Markov de transition* $(p_{i,j})_{i,j\geq 0}$ *si* $\forall n, \forall i_0, i_1, \dots, i_{n+1}$,

$$P(X_{n+1} = i_{n+1}/X_n = i_n, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1}/X_n = i_n)$$

*Cette quantité sera notée* $p_{i_n, i_{n+1}}$

Une remarque importante est que la loi de $X_0$ n'est pas fixée par la définition. Cette loi sera dite <u>loi initiale</u>.

Commençons par remarquer que la loi initiale et les probabilités de transition caractérisent les lois fini-dimensionnelles. En effet, on a (vérification immédiate)

$$P(X_n = i_n, \dots, X_0 = i_0) =$$

$$= P(X_n = i_n/X_{n-1} = i_{n-1}, \dots, X_0 = i_0)P(X_{n-1} = i_{n-1}/X_{n-2} = i_{n-2}, \dots, X_0 = i_0)\dots$$

$$\dots P(X_1 = i_1/X_0 = i_0)P(X_0 = i_0) =$$

$$p_{i_n, i_{n-1}} p_{i_{n-1}, i_{n-2}} \cdots p_{i_1, i_0} \pi_{i_0}$$

si $\pi_k = P(X_0 = x_k)$.

**Corollaire**: *Les probabilités de transition de rang supérieur sont données par*

$$p_{i,j}^{(n)} = P(X_n = x_j/X_0 = x_i) = \sum_{i_1, i_2, \dots, i_{n-1}} p_{j, i_{n-1}} p_{i_{n-1}, i_{n-2}} \cdots p_{i_1, i}$$

Dans le cas particulier où $E$ est fini, i.e. $E = \{x_0, x_1, \ldots, x_K\}$, si on note $P$ la matrice $(K+1) \times (K+1)$ d'élément générique $p_{i,j}$, $p_{i,j}^{(n)}$ devient l'élément générique de la matrice $P^n$.

> **Remarque**: *Dans ce cours, nous ne nous préoccuperons pas de l'existence d'une telle chaîne. Disons en deux mots que pour toute suite* $(p_{i,j})_{0 \leq i,j}$ *telle que* $\forall i, \sum_j p_{i,j} = 1$ *et pour toute loi initiale* $(\pi_i)_{i \geq 0}$, *il existe un espace et une suite de variables aléatoires vérifiant les bonnes propriétés.*

2- Transience et récurrence

Une question naturelle qu'on se pose immédiatement est de savoir si un état de la chaîne revient ou non. Une question annexe est le comportement asymptotique de tels retours. Commençons par introduire une notion technique

> **Définition**: *Une chaîne de Markov* $(X_n)$ *est dite* <u>*irréductible*</u> *si* $\forall i, j, \exists n, p_{i,j}^{(n)} > 0$

A Partir de maintenant, on suppose fixée une chaîne de Markov $(X_n)_{n \geq 0}$ de transition $(p_{i,j})_{i,j \geq 0}$ et de loi initiale $\pi_i > 0, \forall i$. On notera $P_i(A) = P(A / X_0 = x_i)$.
Reprenons le cours de notre récit.

> **Définition**: *Un état* $x_i$ *est dit récurrent si*
>
> $$f_{ii} = P_i(\exists n \geq 0, X_n = x_i) = 1$$
>
> *Dans le cas contraire, il est dit transient*

<u>Notation</u>: On introduit plus généralement

$$f_{ij}^{(n)} = P_i(X_1 \neq x_j, \ldots, X_{n-1} \neq x_j, X_n = x_j)$$

et

$$f_{ij} = \sum_{n \geq 1} f_{ij}^{(n)} = P_i(\exists n > 0, X_n = x_j)$$

Un premier résultat - trè s intuitif - va nous dire que si l'état $x_i$ est récurrent la chaîne y revient une infinité de fois.

> **Proposition**: *Si* $f_{ii} = 1$, *alors* $P_i(X_n = x_i$ *une infinité de fois* $) = 1$.
> *Si* $f_{ii} < 1$, *alors* $P_i(X_n = x_i$ *une infinité de fois* $) = 0$

démonstration: En effet, si $f_{ii} = 1$ (resp. $< 1$) $P_i(X_n = x_i$ une infinité de fois $) =$

$$= \sum_{n_1 < n_2 < \ldots} P_i(X_1 \neq x_i, \ldots, X_{n_1 - 1} \neq x_i, X_{n_1} = x_i, X_{n_1 + 1} \neq x_i, \ldots, X_{n_2 - 1} \neq x_i, X_{n_2} = x_i, \ldots) =$$

$$\sum_{n_1 < n_2 < \ldots} f_{ii}^{(n_1)} f_{ii}^{(n_2 - n_1)} \ldots =$$

$$= \sum_{0=n_0<n_1<n_2<\dots} \prod_{k\geq 0} f_{ii}^{(n_{k+1}-n_k)} = \prod_{k\geq 0} \sum_{l} f_{ii}^{(l)} =$$

$= 1$ (resp. $0$) $\square$

Le théorè me suivant donne un critè re de transience ou de recurrence.

**Théorè me**: *(i) $x_i$ est transient si et seulement si $P_i(X_n = i$ une infinité de fois $) = 0$ et si et seulement si $\sum_n p_{ii}^{(n)} < \infty$*

*(ii) $x_i$ est récurrent si et seulement si $P_i(X_n = i$ une infinité de fois $) = 1$ et si et seulement si $\sum_n p_{ii}^{(n)} = \infty$*

démonstration: Bien entendu, (i) $\Longleftrightarrow$ (ii) et donc on se contente de démontrer (i). Par ailleurs, un certain nombre de points sont déjà prouvés.

- Supposons $\sum_n p_{ii}^{(n)} < \infty$. Alors, par le lemme de Borel-Cantelli, $P_i(X_n = i$ une infinité de fois $) = 0$ et donc $f_{ii} < 1$

- Supposons que $f_{ii} < 1$. On utilise un argument de "premier passage".

$$p_{ii}^{(n)} = P_i[X_n = x_i] = \sum_{k=0}^{n-1} P_i[X_1 \neq x_i, \dots, X_{n-k} \neq x_i, X_n = x_i] =$$

$$= \sum_{k=0}^{n-1} P_i(X_1 \neq x_i, \dots, X_{n-k-1} \neq x_i, X_{n-k} = x_i) P_i(X_k = x_i) =$$

$$= \sum_{k=0}^{n-1} f_{ii}^{(n-k)} p_{ii}^{(k)}$$

Donc

$$\sum_{n=1}^{t} p_{ii}^{(n)} = \sum_{n=1}^{t} \sum_{k=0}^{n-1} f_{ii}^{(n-k)} p_{ii}^{(k)} =$$

$$= \sum_{k=0}^{t-1} p_{ii}^{(k)} \sum_{k+1}^{t} f_{ii}^{(n-k)} \leq$$

$$\leq \sum_{k=0}^{t-1} p_{ii}^{(k)} f_{ii} \leq \sum_{k=0}^{t} p_{ii}^{(k)} f_{ii}$$

donc

$$(1 - f_{ii}) \sum_{n=1}^{t} p_{ii}^{(n)} \leq f_{ii}$$

et si $f_{ii} < 1$, on obtient $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ $\square$

Dans le cas où la chaîne de Markov est irréductible, la classification des états est donnée apr le théorè me suivant.

**Théorè me**: *Si $(X_n)$ est une chaîne irréductible alors on est dans l'un des deux cas suivants*
*(i) Tous les états sont transients*
*(ii) Tous les états sont récurrents*

démonstration: - Supposons que $x_i$ soit un état récurrent. Alors on a $\sum_n p_{ii}^{(n)} = \infty$. Soit $x_j$ un autre état. $\exists n, n'$ tels que $p_{ij}^{(n)} > 0$ et $p_{ji}^{(n')} > 0$. Pour $k \geq n + n'$, on peut écrire

$$p_{jj}^{(k)} \geq p_{ji}^{(n')} p_{ii}^{k-(n+n')} p_{ij}^{(n)} > 0$$

donc $x_j$ est récurrent.
- Si, maintenant, $x_j$ est transient, on en déduit de même que $\sum_n p_{ii}^{(n)} < \infty$ donc $x_i$ est récurrent $\square$

En fait, on peut même préciser ce résultat; supposons que $x_j$ soit transient. On a donc $\sum_n p_{jj}^{(n)} < \infty$. Alors,

$$\sum_{n\geq 1} p_{ij}^{(n)} = \sum_{n\geq 1} \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} =$$

$$(\sum_{k=1}^{\infty} f_{ij}^{(k)}) \sum_{l=0}^{\infty} p_{jj}^{(l)} = f_{ij} \sum_{l=0}^{\infty} p_{jj}^{(l)} \quad \square$$

On en déduit en particulier le

**Corollaire**: *Si $E$ est un ensemble fini, la chaîne irréductible $(X_n)$ est récurrente (i.e. tous les états sont récurrents)*

démonstration: c'est clair car $\sum_j p_{i,j}^{(n)} = 1$ donc $\sum_j \sum_n p_{i,j}^{(n)} = \infty \quad \square$

3- Mesures stationnaires

Au lieu de démarrer avec un état fixé, on va supposer que $X_0$ est aléatoire. Une question naturelle qui se pose est de savoir si on peut trouver une loi pour $X_0$ telle que $\forall n, X_n$ aura la même loi que $X_0$.

Soit $\pi$ une telle loi i.e. $P(X_0) = \pi_i$. Elle doit vérifier

$$P(X_1 = x_i) = \pi_i = \sum_j P(X_1 = x_i / X_0 = x_j) P(X_0 = x_j) =$$

$$= \sum_j p_{ij}\pi_j$$

Ceci amè ne la définition suivante:

**Définition**: *La loi de probabilités $(\pi_j)$ sur $S$ est dite mesure stationnaire si et seulement si $\forall j$,*

$$\pi_j = \sum_i p_{j,i}\pi_i$$

L'étude de cette notion nécessite l'introduction de la périodicité.

**Définition**: *L'état $x_j$ est dit de période $t$ si*

$$t = \text{pgcd}\{n \geq 1, p_{jj}^{(n)} > 0\}$$

*Ce nombre est dit la période de $x_j$*

Pour une chaîne irréductible $(X_n)$, on a la propriété suivante:

**Proposition**: *Soit $(X_n)$ une chaîne de Markov irréductible. Alors tous les états ont la même période*

démonstration: Supposons que $x_i$ ait une période $t_i$ et $x_j$ une période $t_j$. On a un $n$ et un $n'$ tels que $p_{ij}^{(n)} > 0$ et $p_{ji}^{(n')} > 0$. Notons que $t_i \mid n + n'$ et $t_j \mid n + n'$. De plus, $p_{jj}^{(n+n'+k)} \geq p_{ji}^{(n')}p_{ii}^{(k)}p_{ij}^{(n)}$ donc $p_{ii}^{(k)} > 0 \Longrightarrow t_j \mid k$ d'où $t_j \mid t_i$. Comme on peut interchanger $i$ et $j$, on obtient $t_i = t_j$ $\square$

**Définition**: *Quand la période de la chaîne de Markov est 1, la chaîne est dite apériodique*

Nous utiliserons le résultat suivant dont la démonstration est laissée en exercice.

**Lemme**: *Une chaîne irréductible et apériodique est telle que*

$$\forall i, j, \exists n_0(i,j), \forall n \geq n_0(i,j), p_{i,j}^{(n)} > 0$$

Nous concluerons alors ce chapitre par le théorè me suivant qui donne un résultat d'existence et de convergence dans le cas fini.

**Théorè me**: *Supposons que l'espace d'états $E$ soit fini et que la chaîne $(X_n)$ soit irréductible et apériodique. Alors il existe une loi stationnaire $\{\pi_i\}$ , $A \geq 0$ et $0 \leq \rho < 1$ tels que $\mid p_{ij}^{(n)} - \pi_j \mid \leq A\rho^n$*

Remarque: L'observation essentielle est que $\pi_j$ **ne dépend plus de** $i$

démonstration: Notons $E = \{x_1, x_2, \ldots, x_r\}$
Posons $m_j^{(n)} = \min_i p_{ij}^{(n)}$ et $M_j^{(n)} = \max_i p_{ij}^{(n)}$. On a

$$m_j^{(n+1)} = \min_i \sum_k p_{ik}p_{kj}^{(n)} \geq m_j^{(n)} \min_i \sum_k p_{ik} = m_j^{(n)}$$

$$M_j^{(n+1)} = \max_i \sum_k p_{ik} p_{kj}^{(n)} \leq M_j^{(n)} \sum_k p_{ik} = M_j^{(n)} \geq m_j^{(n)}$$

On va montrer qu'il s'agit en fait de deux suites adjacentes.

<u>1er cas</u>: $\forall i, j, p_{ij} > 0$
Posons $\delta = \min_{i,j} p_{i,j}$. On a alors $0 < \delta \leq r^{-1}$. Fixons $u$ et $v$ dans $\{1, 2, \ldots, r\}$. Posons $S' = \{j, p_{uj} \geq p_{vj}\}$ et $S'' = \{j, p_{uj} < p_{vj}\}$ Bien entendu, $S' \cup S'' = S$ On a alors

$$\sum_{S'} (p_{uj} - p_{vj}) + \sum_{S''} (p_{uj} - p_{vj}) = 0$$

et

$$\sum_{S'} (p_{uj} - p_{vj}) = 1 - \sum_{S''} p_{uj} - \sum_{S'} p_{vj} \leq 1 - r\delta$$

Alors,

$$p_{uk}^{(n+1)} - p_{vk}^{(n+1)} = \sum_j (p_{uj} - p_{vj}) p_{jk}^{(n)} \leq$$

$$\leq \sum_{S'} (p_{uj} - p_{vj}) M_k^{(n)} + \sum_{S''} (p_{uj} - p_{vj}) m_k^{(n)} \leq$$

$$\leq \sum_{S'} (p_{uj} - p_{vj}) (M_k^{(n)} - m_k^{(n)}) \leq$$

$$\leq (1 - r\delta) [M_k^{(n)} - m_k^{(n)}]$$

d'où

$$M_k^{(n+1)} - m_k^{(n+1)} \leq (1 - r\delta) [M_k^{(n)} - m_k^{(n)}]$$

et donc

$$M_k^{(n)} - m_k^{(n)} \leq (1 - r\delta)^n$$

D'où ,

$$\lim_n m_k^{(n)} = \lim_n M_k^{(n)} = \lim_n p_{ik}^{(n)} = \pi_k, \forall i$$

et

$$\mid p_{ik}^{(n)} - \pi_k \mid \leq (1 - r\delta)^n$$

<u>2è me cas</u>: S'il existe des $i, j$ tels que $p_{ij}^{(n)} = 0$, de toute façon d'aprè s le lemme précédent, $\exists m$ tel que $\forall k \geq m$, $\forall i, j,$, $p_{ij}^{(k)} > 0$. Alors, on peut refaire la même démonstration que précedemment avec $m_j^{(nm)}$ et $M_j^{(nm)}$ et obtenir

$$M_j^{(nm)} - m_j^{(nm)} \leq (1 - r\delta)^n$$

et donc

$$M_j^{(n)} - m_j^{(n)} \leq (1 - r\delta)^{[\frac{n}{m}]} \leq (1 - r\delta)^{-1} (1 - r\delta)^{\frac{n}{m}} \quad \square$$

quit end

# CHAPITRE 4: ELEMENTS DE THEORIE DE LA MESURE

Le cadre général du calcul des probabilités est la théorie de la mesure. Cette théorie, assez complexe, recquiert un certain nombre de techniques de base dont nous allons essayer de donner une idée ici. Certains résultats faisant appel à des notions -topologiques en particulier- dépassant le niveau de ce cours seront admis sans démonstration.
Comme nous l'avons déjà dit précédemment, la notion fondamentale va être celle de tribu.
A partir d'ici, $E$ désigne un ensemble quelconque.

## 1- Tribus

Redonnons la définition d'une tribu.

**Définition**: *Soit $\mathcal{A} \subset \mathcal{P}(E)$. On dit que $\mathcal{A}$ est une tribu sur $E$ si*
*(i)*$\emptyset \in \mathcal{A}$
*(ii)*$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
*(iii)* $\forall (A_n) \in \mathcal{A}, \bigcup_n A_n \in \mathcal{A}$

Remarque: On a déjà noté que (i) et (ii) entrainent $E \in \mathcal{A}$

On dit que $(E, \mathcal{A})$ est un espace mesurable.

Exemples: - $\mathcal{A} = \mathcal{P}(E)$ (tribu discrète)
-$\mathcal{A} = \{\emptyset, E\}$ (tribu grossière)
- Si $A \subset E$, $\mathcal{A} = \{\emptyset, A, A^c, E\}$

En fait, un résultat essentiel pour la formation de tribus est le suivant

**Proposition**: *Soit $\mathcal{A}_{i \in I}$ une famille de tribus sur $E$. Alors, $\widetilde{\mathcal{A}} = \bigcap_{i \in I} \mathcal{A}_i$ est une tribu sur $E$*

démonstration: évident et laissé en exercice $\square$

Ceci permet d'introduire la définition suivante

**Définition**: *Soit $\mathcal{B} \subset \mathcal{P}(E)$, une famille de parties de $E$. On appelle tribu engendrée par $\mathcal{B}$ la plus petite tribu sur $E$ contenant $\mathcal{B}$, soit encore l'intersection de toutes les tribus contenant $\mathcal{B}$. On la note $\sigma(B)$*

Exemples: 1- $\mathcal{A} = \{\emptyset, A, A^c, E\}$ est la tribu engendrée par $\{A\}$
2- Exemple fondamental: la tribu $\mathcal{B}(I\!\!R)\}$ engendrée par les intervalles ouverts est dite tribu borélienne. Plus généralement, si $E$ est un espace métrique (topologique), $\mathcal{B}(E)$ est la tribu engendrée par les ouverts de $E$.

La notion de tribu sert à introduire le concept, central pour toute la théorie, de fonction mesurable.

2- Fonctions mesurables

Soit $(F, \mathcal{F})$ un espace mesurable. $E$ est muni d'une tribu $\mathcal{A}$.

> **Définition**: *On dit que $f : (E, \mathcal{A}) \to (F, \mathcal{F})$ est mesurable si $\forall B \in \mathcal{F}$, $f^{-1}(B) = \{x \in E, f(x) \in B\} \in \mathcal{A}$*

<u>Remarque</u>: Notez bien que cette notion fait intervenir <u>deux tribus</u> (celle d'arrivée et celle de départ).

<u>Exemple</u>: 1- Si $\mathcal{A} = \mathcal{P}(E)$, toute application est mesurable.
2- Si $\mathcal{A} = \{\emptyset, E\}$, et si $\mathcal{F}$ contient les singletons, $f$ est mesurable si et seulement si elle est constante
3- Si $f : I\!R \to I\!R$ est continue, alors $f$ est mesurable pour $\mathcal{B}(I\!R)$.

Un parallèle naturel avec la notion de tribu engendrée est:

> **Théorème-Définition**: *Soit $f : E \to (F, \mathcal{F})$. $\mathcal{A} = \{f^{-1}(B), B \in \mathcal{F}\}$ est une tribu sur $E$ dite <u>tribu engendrée</u> par $f$. On la note $\sigma(f)$*

Un autre exemple d'application mesurable est très utile:

> **Lemme**: *Soit $A \in \mathcal{A}$. L'application $1\!\!I_A : E \to I\!R$ est mesurable de $\mathcal{A}$ dans $\mathcal{B}(I\!R)$*

3- Mesures

> **Définition**: *Soit $(E, \mathcal{A})$ un espace mesurable. On appelle mesure sur $(E, \mathcal{A})$ toute application $\mu : \mathcal{A} \to \overline{I\!R^+}$ telle que $\forall (A_n) \in \mathcal{A}, A_n \cap A_p = \emptyset, n \neq p$,*
>
> $$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

<u>Remarque</u>: Notez bien ici qu'on n'a plus de restriction sur $\mu(E)$. On peut avoir $\mu(E) = \infty$.

$(E, \mathcal{A}, \mu)$ est dit un espace mesuré.

<u>Exemples</u>:
(i) Si $x \in E$, $\delta_x$ est une mesure sur $E$ dite mesure de Dirac en $x$
(ii) Si $\alpha \in I\!R$, $\alpha\mu$ est une mesure. Si $\mu$ et $\nu$ sont deux mesures, $\mu + \nu$ est une mesure.

Une propriété essentielle est la "continuité" d'une mesure.

> **Théorème**: *Soit $(A_n)$ une suite d'événements et soit $\mu$ une mesure. Alors*
>
> $$\mu(\bigcup_n A_n) = \lim_k (\bigcup_{n=1}^k A_n)$$

L'exemple le plus important non trivial est donné par le théorème suivant.

**Théorème**: *Il existe une unique mesure $\lambda$ sur $\mathcal{B}(\mathbb{R})$, dite mesure de Lebesgue, telle que $\lambda(]a,b[) = b - a, \forall a < b$*

Nous admettrons ce résultat.

Le résultat suivant va nous permettre de le prolonger à $\mathbb{R}^n$.

**Définition**: *Soit $(E, \mathcal{A})$ un espace mesurable et $(F, \mathcal{F})$ un autre espace mesurable. Alors $E \times F$ est muni naturellement de $\mathcal{A} \otimes \mathcal{F} = \sigma(A \times B, A \in \mathcal{A}, B \in \mathcal{F})$ dite tribu-produit*

Ainsi $\mathbb{R}^n$ est muni de $\mathcal{B}(\mathbb{R})^{\otimes n}$, dont il est aisé de voir qu'elle coincide avec $\mathcal{B}(\mathbb{R}^n)$. On énonce alors

**Théorème-Définition**: *Il existe unique mesure sur $E \times F$, dite mesure produit, telle que $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$*

La théorie de la mesure a pour but de définir une intégrale qui dans un certain nombre de cas prolonge l'intégrale de Riemann et pour laquelle des fonctions plus générales seront intégrables. (R)appellons en effet que si $f$ est intégrable au sens de Riemann, elle est continue "presque partout" (ce terme est volontairement laissé vague pour l'instant).

4- L'intégrale de Lebesgue

Avant toute chose, commençons par énoncer un résultat qui nous sera bientôt utile et dont le résultat sera démontré en exercice.

**Proposition**: *Soit $(f_n)$ une suite de fonctions de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose que $f_n$ converg simplement vers une fonction $f$, i.e. $\forall x \in E, f_n(x)$ tend vers $f(x)$. Alors $f$ est mesurable.*

A partir d'ici, $(E, \mathcal{A}, \mu)$ est un espace mesuré fixé. Comme pour l'intégrale de Riemann, on va commencer par définir l'intégrale sur des fonctions simples.

**Définition**: *Une fonction $f$ de $E$ dans $\mathbb{R}$ est dite élémentaire si elle est de la forme*

*(*)* $f(x) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(x)$ *où* $A_i \in \mathcal{A}, \alpha_i \in \mathbb{R}, \forall i$ *et* $\mu(A_i) < \infty, \forall i$

Remarques: (i) Notez bien que les $A_i$ peuvent ne pas être disjoints 2 à 2.
(ii) C'est le prolongement naturel de la notion de fonction en escalier. En particulier l'enemble $\mathcal{E}$ des fonctions élémentaires est un espace vectoriel.

**Définition**: *L'intégrale d'une fonction $f \in \mathcal{E}$ de la forme (*) est donnée par*

$$I(f) = \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

On définit $\| f \|_1 = \sum_{i=1}^{n} | \alpha_i | \mu(A_i)$

<u>Remarques</u>: (i) Une petite vérification s'impose pour voir que la définition ne dépend pas de la représentation (*) qui n'est pas unique. Elle est laissée en exercice
(ii) $|| \, . \, ||_1$ est clairement une norme "aux ensembles de mesure nulle près" (pour l'inégalité triangulaire on se ramène à deux fonctions dont les supports sont identiques)

Le théorème fondamental suivant sera admis. Il fait appel à la notion de complété d'un espace.

> **Théorème-Définition**: *L'ensemble $\mathcal{E}$ des fonctions élémentaires peut être complété dans l'ensemble des fonctions mesurables de $(E, \mathcal{A})$ dans $\mathbb{R}$ pour la norme $|| \, . \, ||_1$. Le complété $\hat{\mathcal{E}}$ de $\mathcal{E}$ s'appelle l'espace $L^1$ des fonctions intégrables.*
>
> *L'intégrale de $f \in L^1$ est définie comme limite des intégrales des fonctions élémentaires. On la note $\displaystyle\int_E f d\mu$*

Comme la définition précédente est très abstraite, nous allons donner de suite une liste de propriétés permettant des calculs.

5- Propriétés fondamentales de l'intégrale.

> **Théorème**:
>
> *(i) L'intégrale est une forme linéaire sur $L^1$*
>
> *(ii) Si $f \in L^1$, $\displaystyle\int | f | d\mu < \infty$ i.e. $| f | \in L^1$. D'ailleurs, $|| f ||_1 = \displaystyle\int | f | d\mu$*
>
> *(iii) Si $f \leq g$, $\displaystyle\int f d\mu \leq \int g d\mu$ et si $f \geq 0$, $\displaystyle\int f d\mu \geq 0$*
>
> *(iv) Si $f \geq 0$ et $\displaystyle\int d d\mu = 0$ alors $f = 0$ presque partout (i.e. $\mu\{x, f(x) \neq 0\} = 0$)*

<u>Exemples</u>:
(i) $\displaystyle\int_E f(z)\delta_x(dz) = f(x)$, $\forall f$ mesurable
(ii) Si $f : [0, 1] \to \mathbb{R}$ est continue

$$\int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x) f(x) d\lambda(x) = \int_0^1 f(x) dx$$

(ceci montre comme annoncé la coincidence des deux intégrales pour les fonctions continues sur les intervalles bornés).

# CHAPITRE 5: MODELE GENERAL DE PROBABILITES

Nous allons maintenant aborder la théorie des Probabilités dans toute sa généralité. Bien entendu, l'ensemble des résultats que nous avons déjà obtenu restera valable dans ce nouveau contexte. C'est d'ailleurs un excellent exercice de vérifier ce parallèlisme.
Recommençons par donner la définition définitive d'une probabilité.

## 1- Espace de probabilités; variables aléatoires.

**Définition**: *On dit qu'un triplet* $(\Omega, \mathcal{A}, P)$ *est un espace de probabilités si* $(\Omega, \mathcal{A}, P)$ *est un espace mesuré et si* $P(\Omega) = 1$

La définition d'une variable aléatoire est alors naturelle.

**Définition**: *Soit* $(E, \mathcal{B})$ *un espace mesurable. On dit que* $X : \Omega \to E$ *est une variable aléatoire à valeurs* $E$ *si* $X$ *est mesurable de* $(\Omega, \mathcal{A})$ *dans* $(E, \mathcal{B})$

Nous pouvons aussi donner une définition générale de la loi

**Théorème-Définition**: *Soit* $X$ *une variable aléatoire à valeurs* $(E, \mathcal{B})$. *Alors, si* $B \in \mathcal{B}$,

$$\mu_X(B) = P(X \in B) = P(X^{-1}(B))$$

*définit une mesure de probabilités sur* $E$ *dite loi de* $X$

Exemple: Si $E = I\!N$, ,otant $p_n = P(X = n)$, on a

$$\mu_X = \sum_{n \geq 0} p_n \delta_n$$

**Définition**: *Soit* $X$ *une variable aléatoire à valeurs dans* $I\!R$ *et soit* $f : I\!R^n \to I\!R$ *mesurable. On dit que* $X$ *admet la densité* $f$ *sur* $I\!R^n$ *si* $\forall A \in \mathcal{B}(I\!R^n)$,

$$P(X \in A) = \mu_X(A) = \int_{I\!R^n} 1\!\!I_A(x) f(x) dx$$

Remarques: (i) ici $dx$ désigne bien sûr la mesure de Lebesgue sur $I\!R^n$.
(ii) Dans la définition précédente, il est sous-entendu que $\int_{I\!R^n} f(x)dx = 1$

Exemples fondamentaux de densités sur $I\!R$

- Loi exponentielle

$$f(x) = \lambda e^{-\lambda x} 1\!\!I_{x \geq 0}$$

- Loi Gaussienne centrée réduite

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathbb{1}_{\mathbb{R}}(x)$$

- Loi uniforme sur $[a, b], a < b$

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$

-Loi de Cauchy

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \mathbb{1}_{\mathbb{R}}(x)$$

2- Moments

Commençonc par définir l'espérance

**Définition**: *Soit $X$ une variable aléatoire à valeurs réelles. On dit que $X$ admet un moment d'ordre 1 ou une espérance si $X$ est intégrable sur $\Omega$. On pose alors*

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

Le théorème essentiel qui relie l'espérance et la loi d'une variable aléatoire est le suivant, qu'on démontrera ensuite plus généralement.

**Théorème**: *Soit $X$ une variable aléatoire réelle. On a, si cette quantité existe,*

$$E(X) = \int_{\mathbb{R}} x \mu_X(dx)$$

Dans le cas particulier d'une variable admettant la densité $f$, l'expression précédente prend la forme

$$E(X) = \int_{\mathbb{R}} x f(x) dx$$

<u>Exercice</u>: calculer les espérances des diverses lois données précédemment.

Plus généralement, on définiera comme d'habitude les moments d'ordres supérieurs.

Le théorème suivant fournit une caractérisation de la loi

**Théorème**: *Soit $X : \Omega \to E$ une variable aléatoire. Soit $\mu$ une mesure sur $E$. Alors, les assertions suivantes sont équivalentes*

*(i) $\mu$ est la loi de $X$*

*(ii) $\forall f$ mesurable bornée de $E$ dans $\mathbb{R}^+$,*

$$E(f(X)) = \int_E f(x) \mu(dx)$$

<u>Application</u>: Ce théorème s'utilise comme outil théorique pour la recherche d'une loi:

Méthodologie pour le problème suivant: *Si $X$ est une variable de loi $\mu_X$ donnée, trouver la loi de $\varphi(X)$ où $\varphi : E \to F$.*
On prend une fonction $f : F \to I\!\!R$ mesurable et bornée, et on écrit

$$E(f(\varphi(X))) = \int_E f(\varphi(x))\mu_X(dx) = \int_F f(y)\mu_Y(dy)$$

par le changement de variable "théorique" $y = \varphi(x)$.

<u>Exemple</u>: Soit $X$ de loi normale centrée réduite. Calculer la loi de $X^2$.
On prend donc $f : I\!\!R^+ \to I\!\!R$ mesurable et bornée. On a

$$E(f(X^2)) = \int_{I\!\!R} f(x^2)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

Faisant séparémént sur $I\!\!R^+$ et $I\!\!R^-$ le changement de variable $y = x^2$, on écrit

$$E(f(X^2)) = 2.\int_{I\!\!R^+} f(y)\frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}\frac{dy}{2\sqrt{y}}$$

et on en déduit que $X^2$ est une variable à densité donnée par

$$1\!\!1_{I\!\!R^+}(y)\frac{2}{\sqrt{2\pi}}e^{-\frac{y}{2}}\frac{1}{2\sqrt{y}}dy$$

3- Indépendance

$(\Omega, \mathcal{A}, P)$ est un espace de probabilités fixé. Rappelons la définition de l'indépendance de deux événements.

> **Définition**: *Soient $A$ et $B$ dans $\mathcal{A}$.$A$ et $B$ sont dits indépendants si $P(A \cap B) = P(A)P(B)$*

On généralise cette définition aux tribus.

> **Définition**: *Soient $\mathcal{B}_1$ et $\mathcal{B}_2$ deux sous-tribus de $\mathcal{A}$. $\mathcal{B}_1$ et $\mathcal{B}_2$ sont dites indépendantes si*
> $$\forall B_1 \in \mathcal{B}_1, \forall B \in \mathcal{B}_2, P(B_1 \cap B_2) = P(B_1)P(B_2)$$

Ceci permet de donner un sens à l'indépendance de deux variables aléatoires

> **Définition**: *Deux variables aléatoires $X$ et $Y$ sont dites indépendantes si et seulement si $\sigma(X)$ et $\sigma(Y)$ sont indépendantes*

Le théorème important est le suivant:

**Théorème**: *Soit $X$ et $Y$ deux variables aléatoires à valeurs dans $(E, \mathcal{E})$. Alors sont équivalentes*
*(i) $X$ et $Y$ sont indépendantes*
*(ii) $\forall f : E \to \mathbb{R}^+$, $\forall g : E \to \mathbb{R}^+$, mesurables bornées,*

$$E(f(X)g(Y)) = E(f(X))E(g(Y))$$

*(iii) $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$*

4- Convergence des variables aléatoires

Rappelons les divers types de convergence d'une suite de variables aléatoires réelles. $(X_n)$ désigne donc une telle suite et $X$ une autre variable qui jouera le rôle de limite.

a) Convergence $L^2$

Supposons que $\forall n, E(X_n^2) < \infty$ et $E(X^2) < \infty$.

**Définition**: *On dit que $(X_n)$ converge vers $X$ dans $L^2$ si*

$$\lim_n E((X - X_n)^2) = 0$$

b) Convergence $L^1$

Supposons que $\forall n, E(\mid X_n \mid) < \infty$ et $E(\mid X \mid) < \infty$.

**Définition**: *On dit que $(X_n)$ converge vers $X$ dans $L^1$ si*

$$\lim_n E(\mid X - X_n \mid) = 0$$

c) Convergence en probabilités

**Définition**: *$X_n$ converge en probabilités vers $X$ si*

$$\forall \varepsilon > 0, P(\mid X_n - X \mid > \varepsilon) \to 0$$

d) Convergence presque-sûre

**Définition**: *$X_n$ converge presque sûrement vers $X$ si*

$$P(\{\omega, X_n(\omega) \to X(\omega)\}) = 1$$

e) Convergence en loi

**Définition**: *$X_n$ converge en loi vers $X$ si $\forall h$ continue et bornée de $I\!R$ dans $I\!R$*

$$\lim_n E(h(X_n)) = E(h(X))$$

On a le résultat suivant qui relie ces diverses notions

**Théorème**: *On a les implications:*
*(i) (convergence $L^2$)$\Rightarrow$ (convergence $L^1$) $\Rightarrow$ (convergence en probabilités)*
*(ii) (convergence en probabilités)$\Rightarrow$ (convergence en loi)*
*(iii) (convergence presque sûre )$\Rightarrow$ (convergence en probabilités)*

De plus, comme conséquence du Lemme de Borel-Cantelli, on a le résultat partiel suivant

**Théorème**: *Si $X_n$ converge vers $X$ dans $L^1$ alors il existe une sous-suite $(X_{n_p})_{p\geq 0}$ de $(X_n)$ telle que $X_{n_p}$ converge preque sûrement vers $X$*

Prolongement: La convergence en loi se prolonge naturellement aux variables aléatoires à valeurs dans un espace plus général que $I\!R$. Redonnons sa définition

**Définition**: *Soit $(X_n)$ une suite de variables aléatoires à valeusr dans un espace métrique $(E,d)$. Soit $X$ une variable aléatoire à valeurs dans $E$. On dit que $(X_n)$ converge en loi vers $X$ si $\forall h$ continue et bornée de $E$ dans $I\!R$*

$$\lim_n E(h(X_n)) = E(h(X))$$

5- Fonctions caractéristiques

Dans le cas des variables aléatoires à valeurs entières, un outil efficace dans l'étude de la convergence en loi étaient les fonctions génératrices. Dans le cas des variables plus générales, c'est la transformée de Fourier qui va jouer ce rôle.

**Définition**: *Soit $X$ une variable aléatoire à valeurs dans $I\!R^n$. On appelle fonction caractéristique de $X$ la fonction définie sur $I\!R^n$ par*

$$\varphi_X(a) = E(e^{i<a,X>}) = \int_{I\!R^n} e^{i<a,x>} \mu_X(dx)$$

Remarque: Noter que $\varphi_X$ est bien définie pour toute loi puisque $\mid e^{i<a,x>} \mid = 1$.

Les principales propriétés de $\varphi_X$ sont regroupées dans le théorème suivant:

**Proposition**:
*(i) $\varphi_X$ est continue sur $I\!R^n$ ( et même uniformément continue)*
*(ii) $\lim_\infty \varphi_X = 0$ si $\mu_X$ est à densité*
*(iii) Si $X$ et $Y$ sont indépendantes, alors $\varphi_{(X,Y)} = \varphi_X.\varphi_Y$*

La démonstration, aisée, est laissée en exercice.

Le résultat essentiel à connaître est alors

> **Théorème**: *Soit $(X_n)$ une suite de variables aléatoires à valeurs $\mathbb{R}^n$. Alors, on a l'équivalence entre*
> *(i) $X_n$ converge en loi vers $X$*
> *(ii) $\varphi_{X_n}(a)$ converge vers $\varphi_X(a)$ pour tout $a \in \mathbb{R}^n$*

Nous admettrons ce théorème qui recquiert un peu de technique.

Jusqu'à présent, nous n'avons pas vu que la fonction caractéristique caractérisait (!) la loi. C'est l'objet du théorème suivant qui fournit une "formule d'inversion".

> **Théorème**: *Soit $\mu$ une mesure de probabilités sur $\mathbb{R}$, de fonction caractéristique $\varphi$. Soient $a$ et $b$ tels que $\mu(\{a\}) = \mu(\{b\}) = 0$.*
> *Alors,*
> $$\mu(]a,b]) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt$$
> *En particulier,*
> $$\mu = \nu \iff \varphi_\mu = \varphi_\nu$$

démonstration: On pose

$$I_T = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt$$

Par le théorème de Fubini, on obtient

$$I_T = \int_{-\infty}^{\infty} \mu(da) [\frac{\operatorname{sgn}(x-a)}{\pi} S(T. \mid x-a \mid) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T. \mid x-b \mid)]$$

où $S(t) = \displaystyle\int_0^t \frac{\sin u}{u} du$ □

> **Corollaire**: *Supposons $\displaystyle\int_{-\infty}^{\infty} \mid \varphi(t) \mid dt < \infty$. Alors, $\mu$ admet une densité continue.*

démonstration: On a, si $F(x) = \displaystyle\int_{-\infty}^{x} \mu dt$

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-it(x+h)}}{ith} \varphi(t) dt$$

Comme

$$\mid \frac{e^{-itb} - e^{-ita}}{it} \mid \leq \mid b - a \mid$$

, on peut appliquer le théorème de convergence dominée quand $h \to 0$ et obtenir que $F$ est dérivable (les détails sont laissés en exercice) $\square$

Nous allons maintenant donner quelques exemples fondamentaux de fonctions caractéristiques de lois à densité

| Loi | Fonction Caractéristique |
| --- | --- |
| $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ | $e^{-\frac{t^2}{2}}$ |
| $\mathbb{1}_{[0,1]}$ | |
| $\frac{e^{it}-1}{it}$ | |
| $e^{-x}\mathbb{1}_{x>0}$ | $\frac{1}{1-it}$ |
| $\frac{1}{2}e^{-\|x\|}$ | $\frac{1}{1+t^2}$ |
| $\frac{1}{\pi}\frac{1}{1+x^2}$ | $e^{-\|t\|}$ |
| $\mathbb{1}_{]-1,1[}(1- \mid x \mid)$ | $2.\frac{1-\cos t}{t^2}$ |
| $\frac{1}{\pi}\frac{1-\cos x}{x^2}$ | $(1- \mid t \mid)\mathbb{1}_{[-1,1]}(t)$ |

6- Le théorème central-limit

Le théorème qui fait l'objet de ce paragraphe est sans doute le plus important de toutes les probabilités. C'est Laplace qui l'a découvert le premier mais le mérite de la première démonstration revient à Gauss. Pour cette raison, il est aussi appelé théorème de Laplace-Gauss.

L'idée fondamentale est la suivante: si on a un très grand nombre de variables aléatoires indépendantes et de même loi, leur somme, convenablement normalisée va être distribuée suivant une loi normale.

Nous n'énoncerons qu'une seule version de ce théorème. Il est bon de savoir qu'il en existe de nombreuses variantes plus fortes.

**Théorème**: *Soit $(X_n)$ une suite de variables aléatoires indépendantes et de mêmes lois, de moyenne $c$ et de variance $\sigma > 0$ finies. Alors la suite de variables $\dfrac{S_n - nc}{\sigma\sqrt{n}}$ converge en loi vers la loi normale centrée réduite.*

démonstration: Elle se fait via les fonctions caractéristiques. Notons $\varphi(t) = E(e^{it(X_1-c)})$. On a

$$\varphi_{\frac{S_n-nc}{\sigma\sqrt{n}}}(t) = E(e^{it\frac{S_n-nc}{\sigma\sqrt{n}}}) = \varphi(\frac{t}{\sigma\sqrt{n}})^n$$

Par ailleurs, la formule de Taylor-Young s'écrit

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + t^2 \varepsilon(t)$$

avec $\lim_{t \to 0} \varepsilon(t) = 0$ d'où

$$\varphi(\frac{t}{\sigma\sqrt{n}}) = 1 - \frac{1}{2n}t^2 + \frac{t^2}{\sigma^2 n}\varepsilon(\frac{t}{\sigma\sqrt{n}})$$

Restant au voisinage de 1, on peut utiliser un Logarithme complexe et obtenir

$$n \ln \varphi(\frac{t}{\sigma\sqrt{n}}) = -\frac{t^2}{2} + t^2 o(1)$$

d'où

$$\lim_n \varphi(\frac{t}{\sigma\sqrt{n}})^n = e^{-\frac{t^2}{2}} \quad \square$$

******************************

# 1 Probabilités élémentaires

1- Le Chevalier de Méré s'étonnait qu'en lançant deux dés, il obtienne plus souvent 11 que 12 alors que l'un et l'autre de ces nombres n'était obtenu que par une combinaison (5+6 et 6+6). Qu'en pensez-vous?

2- a) On fait rouler quatre dés. Quelle est la probabilité d'obtenir au moins un "six"?

b) On fait rouler deux dés vingt-quatre fois. Quelle est la probabilité d'obtenir au moins une fois deux "cinq"?

3- $n$ personnes sont réunies dans une pièce. Calculez la probabilité pour que deux d'entre elles au moins aient la même date d'anniversaire.

4- On suppose que dans une course, il y a $n$ chevaux au départ.

a) Calculez le nombre de tiercés possibles

b) Calculez la probabilité de gagner, avec un ticket, le tiercé

1-dans l'ordre

2-dans l'ordre ou le désordre

3-dans le désordre

c) Application numérique avec $n = 14$.

5- Dans les $p$ boîtes à lettres d'un immeuble, un facteur est chargé de distribuer $n$ lettres dont $r_1$ sont pour la boîte 1, $\ldots$,$r_p$ pour la boîte $p$. Peu consciencieux, il les distribue au hasard.

a) Quelle est la probabilité pour que la distribution soit correcte?

b) Quelle est la probabilité pour que la boîte 1 soit correctement remplie?

c) Quelle est la probabilité pour que dans la boîte 1 il n'y ait aucune lettre destinée à un voisin?

d) Quelle est la probabilité pour qu'il y ait dans chaque boîte exactement le nombre de lettres qui lui était destiné?

6- On lance deux dés au hasard et on considère les événements suivants

$A =$ le premier dé tombe sur une face impaire

$B =$ le deuxième dé amène une face impaire

$C =$ la somme des valeurs des faces des deux dés est impaire

Montrer que $A, B, C$ sont deux à deux indépendants mais pas indépendants.

7- Soient $n$ événements indépendants $A_1, \ldots, A_n$ dans $(\Omega, P)$. Calculer en fonction de $P(A_i)$ la probabilité $p = P(A_1 \cup \ldots \cup A_n)$ et montrer que

$$1 - p \leq \exp\{-\sum_i P(A_i)\}$$

8- On tire au hasard,selon une loi uniforme, un entier compris entre 1 et $n$

a) Si $q$ divise $n$, quelle est la probabilité de tirer un multile de $q$

b) On suppose que la décomposition en facteurs irréductibles de $n$ soit

$$n = q_1^{\alpha_1} \ldots q_p^{\alpha_p}$$

On note $A_i$ l'événement: "on tire un multiple de $q_i$".

Montrer que les $A_i$ sont indépendants.

# 2 Espérances et Variances. Probabilités conditionnelles

1- En utilisant la loi de $(X, Y)$, démontrer que $E(X+Y) = E(X) + E(Y)$

2- Construire un exemple de variable aléatoire non nulle pour laquelle $\mathrm{Var}(X) = 0$.

3- Une population comporte 60ont les cheveux longs et que 40

Une personne se présente avec les cheveux longs. Quelle est la probabilité pour que ce soit une femme?

4- Soit $A$ un événement. On note $\mathbf{1}_A$ la variable aléatoire qui vaut 1 sur $A$ et 0 ailleurs (fonction caractéristique de $A$). Montrer que $E(\mathbf{1}_A) = P(A)$.

5- Soit $X$ une cariable aléatoire réelle.

a) Montrer que pour tout $\alpha > 0$, si $X \geq 0$,

$$P(X \geq \alpha) \leq \frac{1}{\alpha} E(X) \text{ (Inégalité de Markov)}$$

b) Montrer que pour tout $\varepsilon > 0$,

$$P(\mid X - E(X) \mid > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \text{ (Inégalité de Bienaymé-Tchebitcheff)}$$

6- Soient $X_1$ et $X_2$ deux variables indépendantes, de lois $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$. Quelle est la loi de $X_1 + X_2$?

7- Calculer la probabilité pour qu'en répartissant $r$ boules dans $n$ cellules, toutes les cellules soient occupées.

8- Un joueur joue à la roulette à 37 cases 10 francs sur le 19 et 100 francs sur "pair". Quelle est l'espérance de son gain?

9- a) Mon voisin a deux enfants dont une fille. Quelle est la probabilité pour que l'autre soit un garçon?

b) Un autre voisin a deux enfants dont le plus jeune est une fille. Quelle est la probabilité pour que l'autre soit un garçon?

10- Ce matin là, Monsieur Martin, philosophe à ses heures, avait entrepris, compte tenu des prévisions météorologiques pessimistes, de se rendre à son travail en voiture et avait eu la bonne idée de proposer à son voisin, l'ingénieur Félicien Optimal de l'emmener.

Hélas, bientôt pris dans des encombrements désespérants, ils durent se résoudre à engager les plaisirs de la conversation ce qui leur donna l'occasion de mieux se connaître.

M.Martin expliqua qu'il avait trois enfants dont un prénommé Jacques et un autre Paul.

"N'avez-vous pas aussi une fille?" demanda Félicien. "D'ailleurs, je n'ai qu'une chance sur quatre de me tromper".

M.Martin continua son propos qui fit apparaître que l'aîné des enfants était justement Jacques.

"Je pense encore que vous avez une fille, reprit Félicien, mais j'ai maintenant une chance sur trois de me tromper. - Puisque celà vous intéresse, je puis vous donner une autre indication, dit Monsieur Martin: mon benjamin est Paul.

- Alors, répondit Félicien, je ne sais plus du tout si vous avez une fille ou non!"

Cette démonstration de rationnalisme laissa notre philosophe un peu perplexe: ilne lui apparaissait pas clair en effet que les informations successives qu'il avait données avaient pu augmenter l'incertitude de son voisin Félicien. Ces informations étaient-elles des connaissances ou des anti-connaissances? Il s'engagea alors dans une méditation sur le réel et aboutit à la conclusion que puisque effectivement le cadet de ses enfants était une fille, la première impression de Félicien avait été la bonne.

*(d'après N.Bouleau: Probabilités pour l'ingénieur, Hermann*

# 3 Probabilités dénombrables

1- Soit $(A_n)$ une suite d'événements

a) Montrer que

$$\lim_{n\to\infty} P(\bigcup_{k=1}^{n} A_k) = P(\bigcup_{k=1}^{\infty} A_k)$$

b) Montrer que

$$\lim_{n\to\infty} P(\bigcap_{k=1}^{n} A_k) = P(\bigcap_{k=1}^{\infty} A_k)$$

2- Soit $(A_n)$ une suite d'événements telle que $\sum_n P(A_n) < \infty$

Montrer que $P($"une infinité de $A_i$ se produisent simultanément"$) = 0$ (Lemme de Borel-Cantelli)

3- Soient $X$ et $Y$ deux variables aléatoires indépendantes telles que $X$ suit une loi de Poisson de paramètre 1 et $Y$ une loi de Poisson de paramètre $m$. Déterminer la loi de $X + Y$.

4- On s'intéresse à l'effet biologique produit par des électrons à l'issue d'une cathode. Précisément, on suppose que chaque électron émis a une probabilité $p$ d'être actif. On suppose que tous les électrons ont un comportement indépendant les uns des autres. On note $Z$ le nombre d'électrons émis et $Y$ le nombre d'entre eux qui sont actifs.

On suppose que $Z$ suit une loi de Poisson de paramètre 1.

Déterminer la loi de $Y$.

5- Un joueur va au casino avec une fortune $a \in \mathbf{N}$. A chaque partie, il peut gagner 1 franc avec une probabilité $p$ et perdre 1 franc avec une probabilité $q = 1 - p$.

Son but est de jouer jusqu'à l'obtention de la fortune $c \geq a, c \in \mathbf{N}$ mais il doit s'arrêter s'il est ruiné. On note $s_c(a)$ sa probabilité de succès (atteindre $c$ avant la ruine).

a) Calculer $s_c(0)$ et $s_c(c)$

b) Montrer, en raisonnant sur ce qui s'est passé au premier coup, la relation

$$s_c(a) = ps_c(a+1) + qs_c(a-1)$$

c) Déduire la valeur de $s_c(a)$

d) Application numérique: $a = 100, c = 200$; $a = 100, c = 20000$

6- Soient $A_1, A_2, \ldots, A_n$ des événements.

Montrer que

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2/A_1)\ldots P(A_n/A_1 \cap \ldots \cap A_{n-1})$$

7- Soient $X$ et $Y$ deux variables aléatoires de lois de Poisson de paramètres $\alpha$ et $\beta$. Montrer que pour tout $i$ suffisamment grand, $\mid P(X = i) - P(Y = i) \mid \leq \mid a - b \mid$.

8- Soit $X$ une variables aléatoire à valeurs entières. On pose $p_k = P(X = k)$ et $q_k = \sum\limits_{j \geq k+1} p_j$

Montrer que $E(X) = \sum\limits_{j \geq 0} q_j$

9- Trouver la loi du minimum de deux variables aléatoires exponentielles indépendantes

# 4 Convergence des suites de variables aléatoires discrètes

1- Démontrer que la convergence en probabilités entraîne la convergence en loi.

2- On dit que la suite $(X_n)$ converge dans $L^1$ vers $X_\infty$ si pour tout $n$, $E(\mid X_n \mid) < \infty$ et $\lim\limits_n E(\mid X_n - X_\infty \mid) = 0$.

a) Montrer que la convergence dans $L^1$ entraîne la convergence en probabilités.

b) Plus généralement, si $p \geq 1$, on dit que la suite $(X_n)$ converge dans $L^p$ si pour tout $n, E(\mid X_n \mid^p) < \infty$ et $\lim\limits_n E(\mid X_n - X_\infty \mid^p) = 0$. Montrer que la convergence dans $L^p$ entraîne la convergence dans $L^r, \forall 1 \leq r \leq p$.

Indication: On pourra démontrer d'abord l'inégalité de Hölder: si $X$ et $Y$ sont des variables aléatoires réelles telles que $E(\mid X \mid^p) < \infty$ et $E(\mid Y \mid^q) < \infty$ avec $\frac{1}{p} + \frac{1}{q} = 1$, alors

$$E(\mid XY \mid) \leq E(\mid X \mid^p)^{\frac{1}{p}} E(\mid X \mid^q)^{\frac{1}{q}}$$

3- On dit que la suite $(X_n)$ converge presque sûrement (p.s.) vers $X_\infty$ si

$$P(\{\omega, \lim_n X_n(\omega) = X_\infty(\omega)\}) = 1$$

a) Soit $(Y_n)$ une suite de variables aléatoires qui converge p.s. vers la variable nulle et telles qu'il existe $K \geq 0$ tel que $\forall n, \mid Y_n \mid \leq K$.

Montrer que $\lim\limits_n E(Y_n) = 0$.

b)Déduire que la convergence presque sûre entraîne la convergence en probabilités.

4- Soit $(X_n)$ une suite de variables indépendantes centrées et telles que $\sup_n E(X_n^4) < \infty$.

Montrer que $\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$ converge vers 0 p.s.

5- Soit $f : [0,1] \to \mathbf{R}$, continue.

On pose $B_n(x) = \sum_{k=1}^{n} f(\frac{k}{n}) C_n^k x^k (1-x)^{n-k}$.

a) Montrer que si $M = \sum_x \mid f(x) \mid$ et $\delta(\varepsilon) = \sum_{|x-y| \leq \varepsilon} \mid f(x) - f(y) \mid$, on a pour tout $n$

$$\sup_x \mid f(x) - B_n(x) \mid \leq \delta(\varepsilon) + \frac{2M}{n\varepsilon^2}$$

b) Conclure

6- Soit $(X_n)$ une suite de variabls aléatoires à valeurs entières telles que pour tout $\omega$, $X_n(\omega)$ croît vers $X_\infty(\omega)$.

Montrer que $E(X_n)$ tend vers $E(X_\infty)$.

# 5 Fonctions génératrices

1- Soit $(S_n)$ une suite de variables aléatoires telle que $S_n$ suit une loi binômiale de paramètres $n$ et $p_n$, $0 \leq p_n \leq 1$. On suppose que $\lim_n np_n = \lambda \in \mathbf{R}_*^+$.

Etudier la convergence en loi de la suite $(S_n)$.

2- Soit $X$ une variables léatoire à valeurs entières, de fonction génératrice $P(s)$. On suppose qu'il existe $s_0 > 1$ tel que $P(s_0) < \infty$.

a) Montrer que $m_r = E(X^r) < \infty$ pour tout $r \geq 0$.

b) On pose

$$F(s) = \sum_{r \geq 0} \frac{m_r}{r!} s^r$$

Montrer que $F$ converge pour $\mid s \mid < \ln s_0$.

3- Dans le problème de la ruine du joueur, on s'intéresse au temps d'attente du premier gain.

a) Posons

$\phi_n = P($"au $n$ ième coup, pour la première fois, le joueur réalise un gain"$)$.

Par convention, $\phi_0 = 0$.

Calculer $\phi_1$.

b) On pose $\Phi(s) = \sum_{n\geq 0} \phi_n s^n$. Montrer que pour $n > 1$,

$$\phi_n = q(\phi_1\phi_{n-2} + \ldots + \phi_{n-2}\phi_1)$$

c) Déduire que $\Phi(s) - ps = qs\Phi^2(s)$.

d) Résoudre l'équation et en déduire $\Phi$.

e) Donner la valeur de $\phi_k$.

f) Calculer $\sum_{n\geq 0} \phi_n$

g) Soit $N$ le numéro du coup où le joueur réalise un gain pour la première fois.

Calculer $E(N)$.

# 6 Chaînes de Markov discrètes

1- $r$ boules noires et $r$ boules blanches sont réparties dans deux boîtes avec la contrainte que chaque boîte contienne $r$ boules.

Soit $X_n$ le nombre de boules blanches dans la première boîte au temps $n$. A chaque étape, une boule est choisie dans chaque boîte et elles sont échangées.

Montrer que $X_n$ est une chaîne de Markov et donnez sa matrice de transition.

2- Montrer que la fortune du joueur est une chaîne de Markov dont on précisera les transitions.

3- Montrer sur un exemple que l'image d'une chaîne de Markov par une application $f$ n'et pas nécessairement une chaîne de Markov.

Voyez-vous un cas évident qui va conserver la structure markovienne?

4- Montrer que si $(X_n)$ est irréductible et apériodique, pour tous $i$ et $j$ il existe $n_0(i,j)$ tel que pour tout $n \geq n_0(i,j)$, $p_{ij}^{(n)} > 0$.

Indication: on commencera par démontrer le résultat suivants sur les entiers.

Soit $N$ un ensemble d'entiers positifs, stable par l'addition, et contenant dont entiers premiers entre eux. Montrer qu'il existe $m_0$ tel que $\forall n \geq m_0, n \in N$.

En déduire le résulat cherché pour $i = j$ puis pour $i$ et $j$ quelconques.

5- Un penseur qui possède $r$ parapluies va de sa maison à son bureau en portant un parapluie s'il y en a un à sa disposition en cas de pluie (probabilité $p$) mais pas en cas de beau temps (probabilité $q$).

Soit $X_n$ le nombre de parapluies disponibles à l'endroit où se trouve le penseur.

a) Montrer que $(X_j)$ est une chaîne de Markov.

b) Trouver les probabilités stationnaires.

c) Quelle est la probabilité qu'il soit mouillé?

6- Une suite $(X_n)$ est une chaîne de Markov du deuxième ordre si

$$P(X_{n+1} = j/X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = j/X_{n-1} = i_{n-1}, X_n = i_n) = p_{i_{n-1}, i_n; j}$$

Montrer qu'il n'y a rien de nouveau sous le soleil!...

7- Préliminaire: soit $(\mu_i)$ une suite de nombres positifs telle que $\sum_i \mu_i < \infty$. Soient $K > 0$, $g$ un fonction de $\mathbf{N}$ dans $\mathbf{N}$ et $(g_n)$ une suite de fonctions de $\mathbf{N}$ dans $\mathbf{N}$ telle que $\forall n, \mid g_n \mid \leq K$ et $\forall i, g_n(i)$ converge vers $g(i)$. Montrer que

$$\sum_{i \geq 0} \mu_i g_n(i) \to \sum_{i \geq 0} \mu_i g(i)$$

Soit $Q = (q_{ij})_{i,j \in I}$ une matrice sous-stochastique i.e. $q_{ij} \geq 0$ et $\sum_j q_{ij} \leq 1$ pour tout $i$.

On pose $Q^n = (q_{ij}^{(n)}$, puissance $n$ ième de $Q$ et $s_i^{(n)} = \sum_j q_{ij}^{(n)}$.

a) Montrer que la suite $(s_i^{(n)})$ est décroissante. On pose $s_i = \lim_n s_i^{(n)}$.

b) Montrer que $s_i$ est la solution maximale de

$$x_i = \sum_{j \in I} q_{ij} x_j \text{ et } 0 \leq x_i \leq 1 (*)$$

c) Soit $(X_n)$ une chaîne de Markov d'espace d'états $E$ et de transition $P$. Soit $I$ fini inclus dans $E$. On notera $q_{ij} = p_{i,j}$ pour $i, j \in I$.

On pose $s_i = P_i(X_t \in I, \forall t \geq 1)$.

Montrer que $s_i$ est une solution maximale de (*).

d) Soit $x^*$ dans $E$. Considérons $I = E - \{x^*\}$. On suppose $(X_n)$ irréductible. Montrer que $(X_n)$ est transiente si et seulement si (*) admet une solution non triviale.

8- Soit $(X_n)$ une chaîne de Markov à espace d'états $E = \{0, 1, \ldots, n, \ldots\}$.

On suppose que $p_{00} = 1$ et que $f_{i,0}$, la probabilité de retour en 0, est strictement positive pour tout $i$.

a) Montrer que pour tout $i$,

$$P_i\left(\bigcup_{j=1}^{\infty}\{X_n = j \text{ pour une infinité de } n\}\right) = 0$$

b) Interpréter $(X_n)$ comme la taille d'une population.

9- Montrer qu'une chaîne de Markov dont un élément diagonal n'est pas nul n'est pas périodique.

10- Soit $(X_n)$ une chaîne de Markov à états $\{0, 1\}$.

a) Montrer que $p_{0,0}^{(n)} = p_{10} + (p_{00} - p_{10})p_{00}^{(n-1)}$

b) Déduire une formule explicite pour $p_{00}^{(n)}$ et montrer que si les $p_{ij} > 0$, $p_{00}^{(n)} \to \dfrac{p_{10}}{p_{01} + p_{10}}$

Vérifier que la convergence est exponentielle.

# 7 Boréliens

1- Montrer que les singletons sont mesurables dans $\mathcal{B}(\mathbf{R})$

2- Montrer que $\mathcal{B}(\mathbf{R})$ est engendrée par les intervalles $]a, +\infty[$, $a \in \mathbf{R}$

3- Sit $(f_n)$ une suite de fonctions mesurables de $\mathbf{R}$ dans $\mathbf{R}$ telle que $\lim_n f_n(x) = f(x)$ pour tout $x \in \mathbf{R}$. Montrer que $f$ est mesurable.

4- Montrer que tout ensemble dénombrable dans $\mathbf{R}$ est de mesure de Lebesgue nulle.

5- La fonction $\frac{\sin(x)}{x}$ est-elle dans $L^1(\mathbf{R}^+)$?

6- Démontrer que toute fonction mesurable est limite simple de fonctions élémentaires.

7- Montrer que l'ensemble de Cantor est de mesure nulle mais non dénombrable dans $\mathbf{R}$.

8- (Théorème de Classe monotone)

Soit $\mathcal{F}$ une *algèbre de Boole* c'est à dire un ensemble de parties d'un ensemble $\Omega$, contenant $\Omega$, stable par complémentaire et par union *finie*.

Soit $\mathcal{M}$ une *classe monotone* c'est à dire un ensemble de parties de $\Omega$ stable par union dénombrable croissante (c'est à dire que si $(A_n)$ est une suite telle que $A_n \subset A_{n+1}$, alors $\bigcup_n A_n \in \mathcal{M}$) et par intersection dénombrable décroissante.

On suppose que $\mathcal{F} \subset \mathcal{M}$.

Montrer que $\sigma(\mathcal{F}) \subset \mathcal{M}$.

# 8 Variables aléatoires continues

1- Soit $X$ une variable aléatoire réelle. On pose $F(t) = P(X \leq t)$ (fonction de répartition).

a) Montrer que $F$ est croissante et continue à droite.

b) Montrer que $F$ est continue si $X$ est une variable à densité.

c) Si $X$ est à densité, relier $F$ et la densité de $X$.

2- On suppose que $X$ est à valeurs dans $\mathbf{R}^+$.

Montrer que $E(X) = \int_0^\infty (1 - F(t))dt$

3- Si $X$ est une variable aléatoire réelle positive de densité $f$, déterminer la densité de $X^{-1}$.

4- Soit $(X_n)$ une suite de variables aléatoires indépendantes. Soit $Y$ une variable aléatoire. On suppose que pour tout $n$, $Y$ est $\sigma(X_n, X_{n+1}, \ldots,)$-mesurable. Montrer qu'il existe $a \in \mathbf{R}$ tel que $P(Y = a) = 1$.

Indication: Appliquer le théorème de classe monotone à une classe bien choisie.

5- Soient $X_1, \ldots, X_n$ indépendantes et de loi de Cauchy. Déterminer la loi de $\frac{(X_1 + \ldots + X_n)}{n}$.

6- Supposons que $X$ et $Y$ soient indépendantes et $f(x, y) \geq 0$. On pose $g(x) = E(f(x, Y))$.

Montrer que $E(g(X)) = E(f(X, Y))$.

7- Soient 3 nombres $X, Y, Z$ choisis indépendamment et uniformément dans [0,1]. Quell est la probabilité pour que l'on puisse former un triangle à l'aide de segments de longueurs $X, Y, Z$?

8- Soit $(X, Y)$ une variable aléatoire à valeurs dans $\mathbf{R}^2$ de densité

$$f(x,y) = \frac{1}{2\sqrt{x}} e^{-y} \mathbf{1}_D(x,y)$$

avec $D = \{x > 0, y > 0, y^2 > x\}$.

a) Déterminer les lois de $X$ et $Y$.

b) Les variables aléatoires $X$ et $Y$ sont-elles indépendantes?

c) Les variables $X$ et $Y - X^{\frac{1}{2}}$ sont-elles indépendantes?

d) Les variables aléatoires $\frac{X}{Y^2}$ et $Y$ sont-elles indépendantes?

9- Soient $X_1$ et $X_2$ deux variables aléatoires indépendantes, de mêmes lois de densité $f(x) = 2x, 0 < x < 1$

Déterminer la loi de $\frac{X_1}{X_2}$.

10- (Aiguille de Buffon).

Le plan est strié de droites parallèles équidistantes de $2a$. Une aiguille de longueur $2l, l < a$ est jetée au hasard sur le plan au sens où la distance du milieu de l'aiguille à la droite la plus proche est une variable aléatoire $X$ uniforme sur $[0, a]$ et où l'angle que fait l'aiguille avec cette droite est une variable aléatoire $\varphi$, indépendante de $X$, uniforme sur $[0, \pi]$. Quelle est la probabilité que l'aiguille coupe l'une des parallèles?

# 9 Convergences et fonctions caractéristiques

1- Démontrer que si la fonction caractéristique $\varphi$ vérifie

$$\int_{-\infty}^{\infty} \mid \varphi(t) \mid dt < \infty$$

alors la variable aléatoire est à densité.

2- Montrer que les implications entre les diverses convergences énoncées dans le cours sont les seules.

3- a) Monter que des variables à densité peuvent converger vers des variables sans densité.

b) Et réciproquement.

4- Montrer que si $X_n$ converge en loi vers $X$, alors $\mu_{X_n}(I) \to \mu_X(I)$ pour tout borélien $I$ tel que $\mu(\partial I) = 0$.

5- Soit $X$ une v.a. de loi $\mu$ et de fonction caractéristique $\varphi$.

a) Montrer que

$$\mu(\{a\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita}\varphi(t)dt$$

b) Soit $\{x_k\}$ la suite des points de masse non nulle pour $\mu$. Montrer que

$$\lim_{t\to\infty} \frac{1}{2T} \int_{-T}^{T} \mid \varphi(t) \mid^2 dt = \sum_k \mu(\{x_k\})^2$$

Indications: Commencer par considérer deux variables indépendantes $Z_1$ et $Z_2$ de loi $\mu$. Montrer que

$$\lim_{t\to\infty} \frac{1}{2T} \int_{-T}^{T} \mid \varphi(t) \mid^2 dt = P(Z_1 - Z_2 = 0)$$

Montrer alors que

$$P(Z_1 - Z_2 = 0) = \int_{-\infty}^{\infty} P(X = y)\mu(dy) = \sum_k \mu(\{x_k\})^2$$

c) Montrer que $\mu$ n'a pas de point de masse si $\varphi$ est dans $L^2(dt)$.

# 10 Processus de Poisson

1- Soit une variable aléatoire telle que

$$P(X > x + y / X > x) = P(X > y), x, y \geq 0$$

Déterminer la loi de $X$ et interpréter.

2- On considère une suite de variables aléatoires $(X_n)$ indépendantes, de même loi exponentielle de paramètre $\lambda$.

On pose $T_0 = 0$ , $T_n = X_1 + \ldots + X_n$ et pour tout $t \geq 0$, $N_t = \inf\{n, T_{n+1} > t\}$.

On dit que la collection de variables aléatoires $(N_t)_{t\geq 0}$ est un processus de Poisson de paramètre $\lambda$.

a) Montrer que $N_t < \infty$ avec probabilité 1 pour tout $t \geq 0$

b) Montrer que $t \mapsto N_t$ est continue à droite, ne saute que de 1 et est croissante

c) Déterminer la loi de $N_t$ et montrer que pour $d > 0$, $N_{t+d} - N_t$ et $N_t$ sont indépendantes

3- On reprend les notations de l'exercice précédent

a) Soient $A_t = t - T_{N_t}$ et $B_t = T_{N_{t+1}} - t$

Montrer que $A_t$ et $B_t$ sont indépendantes et trouver leur lois.

b) On pose $L_t = A_t + B_t$. Déterminer la loi de $L_t$

c) Déterminer la limite de $E(L_t)$ quand $t$ tend vers l'infini

4- Soient $N_t$ et $N'_t$ deux processus de Poisson indépendants de paramètres $\lambda$ et $\lambda'$. Montrer que $N''_t = N_t + N'_t$ est un processus de Poisson dont on déterminera l'intensité.