Processus stochastiques
Introduction

Les processus aléatoires (où stochastiques) permettent de modéliser des systèmes dont le comportement n'est que partiellement prévisible. La théorie est fondée sur le calcul des probabilités et les statistiques.

Les domaines d'applications sont très nombreux : de nombreuses questions de télécommunications, la modélisation et la gestion du trafic dans les réseaux à accès multiples, la commande adaptative, le traitement du signal et le filtrage .... Ce cours a pour objectif d'introduire les méthodes à la base de l'étude de tels systèmes, en faisant largement appel aux exemples rencontrés par les ingénieurs.

Contenu de la matière :

- Chap 1 : Définitions et généralités
- Chap 2 : Processus stationnaires.
- Chap 3 : Processus à accroissements indépendants et stationnaires (PAIS).
- Chap 4 : Processus de Poisson
- Chap 5 : Processus de Naissance et de Mort
- Chap 6 : Processus markoviens.
- Chap 7 : Chaînes de Markov

Le stochastique en général

On fait appel aux méthodes stochastiques lorsqu'on est en présence de phénomènes qu'il n'est pas possible ou peu pratique d'étudier de façon détaillée et déterministe. C'est notamment le cas lorsque les systèmes étudiés présentent une très grande complexité, ou lorsqu'on ne dispose que d'une connaissance partielle de leurs caractéristiques. Les méthodes stochastiques permettent alors d'étudier les comportements en moyenne, en modélisant de façon probabiliste les parties d'un système qu'on ne souhaite pas ou qu'on ne peut pas décrire en détails. Les méthodes stochastiques fournissent les outils nécessaires pour déduire les distributions de probabilités des grandeurs de sortie importantes, en fonction de celles des entrées et du modèle (déterministe ou non) du système. Elles permettent ensuite d'utiliser au mieux ces informations pour prendre des décisions appropriées.

Pourquoi enseigner les méthodes stochastiques ?

Le besoin des méthodes stochastiques est reconnu depuis assez longtemps (enseignement du calcul des probabilités et des statistiques en candidature ingénieur, depuis quelques décennies). Cependant, cet enseignement a très peu évolué au cours du temps, et surtout le volume associé est resté constant. Il apparaît que la partie théorique relative au calcul des probabilités est mieux assimilée par les étudiants, alors que la statistique est assez vite oubliée, ce qui tient au fait qu'on ne l'utilise pas assez par après.

Si on fait un inventaire des travaux de fin d'études des ingénieurs électriciens (électronique, informatique, génie électronique) on se rend compte qu'une proportion importante fait appel de façon directe ou indirecte aux méthodes stochastiques.

Enfin, si on s'interroge sur le rôle des futurs ingénieurs, on se rend compte qu'on leur demande un esprit critique et une capacité d'innovation de plus grande. Or, il est reconnu que l'enseignement actuel des ingénieurs, qui est surtout basé sur le raisonnement déductif (appliquer des "lois" générales aux cas particuliers), tend à étouffer l'imagination et la créativité. L'enseignement du stochastique a pour premier objectif d'engendrer une plus grande ouverture d'esprit et de renforcer la capacité de raisonnement inductif (c'est-à-dire à tirer des conclusions intéressantes à partir de cas particuliers). Il apparaît donc comme souhaitable, sinon nécessaire, pour renforcer la capacité d'innovation des ingénieurs.

Domaine de l'ingénieur électricien faisant appel au stochastique

Les télécommunications reposent sur les méthodes stochastiques en ce qui concerne l'optimisation des performances, le codage, et le filtrage du bruit. En particulier, les télécommunications font largement appel au traitement du signal, aux techniques d'optimisation des performances de systèmes informatiques distribués, à la compression et au codage de données. Par exemple, dans un réseau ATM chaque connexion se présente sous la forme d'une suite virtuellement unique de cellules transmises, qui peut être représentée comme la réalisation d'un processus stochastique. Les méthodes stochastiques permettent alors d'étudier les performances du système lorsqu'il est soumis à différents types de trafic (communications téléphoniques, transferts de données numériques, trafic multimédia...). Elles permettent aussi l'optimisation du codage des données dans le but de minimiser les pertes d'informations suite au bruitage en cours de transmission.

Le traitement du signal (filtrage, traitement de la parole et de signaux physiologiques, traitement d'images) repose en grande partie sur des méthodes stochastiques. En imagerie spatiale, par exemple, ces techniques sont utilisées pour éliminer le bruit des informations brutes captées par les télescopes et ainsi identifier automatiquement les corps stellaires. En médecine, elles permettent l'interprétation automatique de signaux physiologiques (électrocardiogrammes, électroencéphalogrammes) et facilitent ainsi la surveillance des malades.

La théorie des systèmes est une discipline générale qui est utilisée pour l'étude et la conception d'une très grande diversité de systèmes, que ce soit en mécanique, en électricité, ou encore en informatique. Une partie de la théorie des systèmes s'intéresse aux systèmes stochastiques qui sont notamment utilisées en estimation d'état et pour la conception de systèmes auto-adaptatifs. L'estimation d'état, permet de tirer le meilleur profit des informations fournies par divers capteurs, notamment en filtrant les erreurs de mesures et en permettant la détection de fonctionnement anormaux de certains capteurs. Les systèmes auto-adaptatifs sont capables d'adapter leur stratégie de commande en fonction de changement des caractéristiques de l'environnement, du système piloté, et/ou ses objectifs de réglage.

En informatique de nombreuses questions font directement appel aux méthodes stochastiques : l'optimisation des performances des systèmes, la compression de données, l'intelligence artificielle, l'analyse de données, les réseaux de neurones.... Par exemple, l'analyse de données est utilisée par certains constructeurs automobiles afin de détecter les raisons pour lesquelles certains pannes répétitives sont observées, les méthodes stochastiques permettent dans ce contexte d'identifier les facteurs qui provoquent effectivement des pannes des nombreuses autres informations disponibles dans les bases de données. La compression des données est basée sur le codage de suites de symboles en fonction de leur probabilité d'apparition, les suites les plus fréquentes recevant les codes les plus courts, elles permet de réduire coûts de stockage et délais de transmission dans de nombreuses applications (stockage de masse, réseaux informatiques, disques compacts, multimédia, télévision digitale...).

La gestion des risques industriels et technologiques fait appel aux méthodes stochastiques pour l'analyse et la maîtrise de la fiabilité et de la sécurité des systèmes. Les méthodes stochastiques sont notamment utilisées pour la conception des centrales nucléaires, la planification des réseaux d'énergie électrique, l'évaluation des risques des moyens de transport en commun (aéronautique, trains à grande vitesse), la maîtrise de la fiabilité des lanceurs spatiaux... Ces techniques permettent notamment d'identifier les sources de pannes les plus probables et de déterminer des parades à la fois efficaces et aussi économiques que possibles. Notons que l'étude des performances des logiciels informatiques fait partie de ce domaine.

1

Définitions et généralités

1.1 Rappels de probabilité

1.1.1 Espace de probabilité

Un espace de probabilité est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où :

— Ω est un ensemble.

— $\mathcal{F}$ est une tribu (où σ-algèbre) sur Ω.

— $\mathbb{P}$ est une (mesure de) probabilité sur $(\Omega, \mathcal{F})$

Définition 1.1.1. *Une tribu (ou σ-algèbre) sur Ω est une famille $\mathcal{F}$ de sous-ensemble de Ω (appelés "évènements") tels que*

$$\begin{cases} i)\ \emptyset \in \mathcal{F}, \\ ii)\ A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}, \\ iii)\ (A_n)_{n=1}^{\infty} \subset \mathcal{F} \Rightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}. \end{cases}$$

En particulier : A, B $\in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

De même A, B $\in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

Exemple 1.1.1. *-Soit $\Omega = \{1, ..., 6\}$. On peut définir plusieurs tribus sur Ω :*

$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, ..., \{1, 2\}, ..., \Omega\}$= *tribu complète (la plus grande),*

$\mathcal{F}_0 = \{\emptyset, \Omega\}$= *tribu triviale (la plus petite) ($\emptyset$ = événement impossible, Ω = événement arbitraire),*

$\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, ..., 6\}, \Omega\}$

$\mathcal{F}_2 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$, *etc.*

- Soit $\Omega = [0,1]$ et $I_1, ..., I_n$ une famille d'intervalles formant une partition de Ω. La famille de sous-ensembles définie par

$$\mathcal{G} = \{\emptyset, I_1, I_2, ..., I_1 \cup I_2, ..., I_1 \cup I_2 \cup I_3, ..., \Omega\}$$

est une tribu sur Ω.

Définition 1.1.2. *Soit $\mathcal{A} = \{A_i, i \in I\}$ une famille de sous-ensembles de Ω. Alors la tribu engendrée par $\mathcal{A}$ est la plus petite tribu sur Ω qui contient tous les sous-ensembles $A_i, i \in I$. Elle est notée $\sigma(\mathcal{A})$. (NB : l'ensemble I n'est pas forcément dénombrable).*

Exemple 1.1.2. *Reprenons l'exemple 1*
-Soit $\Omega = \{1, ..., 6\}$. Si $\mathcal{A}_1 = \{\{1\}\}$, alors $\sigma(A_1) = \{\emptyset, \{2,3,4,5,6\}, \{1\}, \Omega\} = \mathcal{F}_1$.
Si $\mathcal{A}_2 = \{\{1,3,5\}\}$, alors $\sigma(A_2) = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\} = \mathcal{F}_2$.
Et si $\mathcal{A} = \{\{1,3,5\}, \{1,2,3\}\}$, alors $\sigma(\mathcal{A}) = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \{1,2,3\}, \{4,5,6\}, \{1,2,3,5\}, \{4,6\}, \{1,3,4,5,6\}, \{2\}, \{1,2,3,4,6\}, \{5\}, \Omega\}$.
-Soit $\Omega = [0,1]$. Si $\mathcal{A} = \{I_1, ..., I_n\}$, alors $\sigma(\mathcal{A}) = \mathcal{G}$.

Définition 1.1.3. *Soit $\Omega = [0,1]$. La tribu borélienne sur [0,1] est la tribu engendrée par la famille de sous-ensembles $\mathcal{A} = \{]a,b[: 0 \leq a < b \leq 1\}$ = {intervalles ouverts dans [0,1]}. Elle est notée $\mathcal{B}([0,1])$. Elle contient un très grande nombre de sous-ensembles de [0,1], mais pas tous.*

Remarque 1.1.1. *-Pour Ω ensemble fini, on choisit souvent $\mathcal{F} = \mathcal{P}(\Omega)$.*
-Pour $\Omega \subset \mathbb{R}$ ou $\Omega \subset \mathbb{R}^n$, on choisit souvent $\mathcal{F} = \mathcal{B}(\Omega)$

Définition 1.1.4. *Une sous-tribu de $\mathcal{F}$ est une tribu $\mathcal{G}$ telle que si $A \in \mathcal{G}$ alors $A \in \mathcal{F}$. On note $\mathcal{G} \subset \mathcal{F}$*

Exemple 1.1.3. *Reprenons l'exemple 1. On a $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$, $\mathcal{F}_0 \subset \mathcal{F}_2 \subset \mathcal{F}$, mais pas $\mathcal{F}_1 \subset \mathcal{F}_2$, ni $\mathcal{F}_2 \subset \mathcal{F}_1$.*

Remarque importante

∘ Il est toujours vrai que : $A \in \mathcal{G}$ et $\mathcal{G} \subset\in \mathcal{F} \Rightarrow A \in \mathcal{F}$.

∘ Mais il est faux de dire que $A \subset B$ et $B \in \mathcal{F} \Rightarrow A \in \mathcal{F}$.

Contre exemple

$\{1\} \subset \{1,3,5\}, \{1,3,5\} \in \mathcal{F}_2$, mais $\{1\} \notin \mathcal{F}_2$

Définition 1.1.5. *Soit $\mathcal{F}$ une tribu sur Ω. Une (mesure de) probabilité sur $(\Omega, \mathcal{F})$ est une application $\mathbb{P} : \mathcal{F} \to [0,1]$ telle que*

$$\begin{cases} i)\ \mathbb{P}(\emptyset) = 0 \text{ et } \mathbb{P}(\Omega) = 1, \\ ii)\ (A_n)_{n=1}^{\infty} \subset \mathcal{F} \text{ disjoints } (i.e\ A_n \cap A_m = \emptyset, \forall n \neq m) \Rightarrow \mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \end{cases}$$

En particulier : $A, B \in \mathcal{F}$ et $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
De plus :

$$\begin{cases} i)\ Si\ (A_n)_{n=1}^{\infty} \subset \mathcal{F}, A_n \subset A_{n+1} \text{ et } \cup_{n=1}^{\infty} A_n = A, alors\ \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A) \\ ii)\ Si\ (A_n)_{n=1}^{\infty} \subset \mathcal{F}, A_{n+1} \subset A_n \text{ et } \cap_{n=1}^{\infty} A_n = A, alors\ \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A) \end{cases}$$

Exemple 1.1.4. *Soient $\Omega = \{1, ..., 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. On définit :*
1- $\mathbb{P}_1(\{i\}) = \frac{1}{6}, \forall i$ (mesure de probabilité associée à un dé-équilibre).
Dans ce cas, on voit par exemple que :

$$\mathbb{P}_1(\{1,3,5\}) = \mathbb{P}_1(\{1\}) + \mathbb{P}_1(\{3\}) + \mathbb{P}_1(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

2- $\mathbb{P}_2(\{i\}) = 0, \forall i = \overline{1,5}$ et $\mathbb{P}_2(\{6\}) = 1$ (mesure de probabilité associée à un dé-pipé) .

Définition 1.1.6. *Soient $\Omega = [0,1]$ et $\mathcal{F} = \mathcal{B}([0,1])$. On appelle mesure de Lebesgue sur [0,1] la mesure de probabilité définie par*

$$\mathbb{P}(]a,b[) = b - a, \quad \forall\ 0 \leq a < b \leq 1.$$

$\mathbb{P}$ n'est définie a priori que sur les intervalles, mais est uniquement extensible à tout ensemble borélien $B \in \mathcal{B}([0,1])$. Elle est notée $\mathbb{P}(B) = |B|, B \in \mathcal{B}([0,1])$.

En utilisant la propriété (ii) ci-dessus, on déduit que $\forall x \in [0,1]$:

$$\mathbb{P}(\{x\}) = |\{x\}| = \lim_{n\to\infty} \mathbb{P}(]x - \frac{1}{n}, x + \frac{1}{n}[) = \lim_{n\to\infty} \frac{2}{n} = 0.$$

Généralisation à n dimensions : Soit $\Omega = [0,1]^n$.
-Tribu borélienne : $\mathcal{B}(\Omega) = \sigma(\mathcal{A})$ où $\mathcal{A} = \{]a_1, b_1[\times]a_2, b_2[\times ... \times]a_n, b_n[,\ 0 \leq a_i < b_i \leq 1\}$. $\mathcal{A}$ est la famille des "rectangles" dans Ω.
-Mesure de Lebesgue : $\mathbb{P}(]a_1, b_1[\times]a_2, b_2[\times ... \times]a_n, b_n[) = (b_1 - a_1)(b_2 - a_2)...(b_n - a_n)$.
Comme dans le cas uni-dimensionnel, $\mathbb{P}$ n'est pas définie a priori que sur certains ensembles (les rectangles), mais est uniquement extensible à tout $B \in \mathcal{B}(\Omega)$ (par exemple B = disque, ovale...).

1.2 Variable aléatoire

Définition 1.2.1. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. Une variable aléatoire (souvent abrégé v.a par la suite) est une application* $X : \Omega \to \mathbb{R}$ *telle que*

$$\{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} \in \mathcal{F}, \ \forall B \in \mathcal{B}(\mathbb{R})$$

Proposition 1.2.1. X *est une variable aléatoire ssi* $\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{F}, \ \forall t \in \mathbb{R}$

Remarque 1.2.1. *-X est aussi dite une fonction (ou v.a)* $\mathcal{F}$*-mesurable.*
-Si $\mathcal{F} = \mathcal{P}(\Omega)$*, alors* X *est toujours* $\mathcal{F}$*-mesurable.*
-Si $\Omega = \mathbb{R}$ *et* $\mathcal{F} = \mathcal{B}(\mathbb{R})$*, alors* X *est dite une fonction borélienne.*

Définition 1.2.2. *Pour* $A \subset \Omega$*, on pose* $\mathbb{1}_A(\omega) = \begin{cases} 1 & si\ \omega \in A \\ 0 & sinon \end{cases}$

On vérifie que la v.a $\mathbb{1}_A$ est $\mathcal{F}$-mesurable ssi $A \in \mathcal{F}$.

Exemple 1.2.1. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *l'espace de probabilité du dé-équilibre*
$X_1(\omega) = \omega : \mathbb{P}(\{\omega \in \Omega : X_1(\omega) = i\}) = \mathbb{P}(\{i\}) = \frac{1}{6}$
$X_2(\omega) = \mathbb{1}_{\{1,3,5\}}(\omega) : \mathbb{P}(\{\omega \in \Omega : X_2(\omega) = 1\}) = \mathbb{P}(\{1,3,5\}) = \frac{1}{2}$

Soit $\mathcal{F} = \mathcal{P}(\Omega)$*.* X_1 *et* X_2 *sont toutes deux* $\mathcal{F}$*-mesurables.*
Soit $\mathcal{F}_2 = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$*. Seule* X_2 *est* $\mathcal{F}_2$*-mesurable ;* X_1 *ne l'est pas.*
En effet :
$\{\omega \in \Omega : X_2(\omega) = 1\} = \{1,3,5\} \in \mathcal{F}_2$ *et* $\{\omega \in \Omega : X_2(\omega) = 0\} = \{2,4,6\} \in \mathcal{F}_2$*, tandis que* $\{\omega \in \Omega : X_1(\omega) = 1\} = \{1\} \notin \mathcal{F}_2$*.*

Définition 1.2.3. *La tribu engendrée par une famille de v.a* $\{X_i, i \in I\}$ *sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *est définie par*

$$\sigma(X_i, i \in I) = \sigma(\{X_i \in B\}, i \in I, B \in \mathcal{B}(\mathbb{R})) = \sigma(\{X_i \leq t\}, i \in I, t \in \mathbb{R})$$

Exemple 1.2.2. *Reprenons l'exemple précédent :* $\sigma(X_1) = \mathcal{F} = \mathcal{P}(\Omega)$*,* $\sigma(X_2) = \mathcal{F}_2 \neq \mathcal{P}(\Omega)$*.*

Proposition 1.2.2. *Si* $g : \mathbb{R} \to \mathbb{R}$ *est borélienne et* $X : \Omega \to \mathbb{R}$ *est une v.a, alors g(X) est une v.a.*

Démonstration.
Soit $B \in \mathcal{B}(\mathbb{R})$. On a

$$\{\omega \in \Omega : g(X(\omega)) \in B\} = \{\omega \in \Omega : X(\omega) \in g^{-1}(B)\}$$

Or $g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\} \in \mathcal{B}(\mathbb{R})$, car g est une fonction borélienne.
Comme X est une v.a, on en déduit que $\{\omega \in \Omega : X(\omega) \in g^{-1}(B)\} \in \mathcal{F}$, et donc finalement que $g(X)$ est une v.a. $\square$

Proposition 1.2.3. *Toute fonction continue est borélienne (et pratiquement toute fonction discontinue l'est aussi).*

1.3 Loi d'une variable aléatoire

Définition 1.3.1. *La loi d'une variable aléatoire X est l'application $\mu_X(B) : \mathcal{B}(\mathbb{R}) \to [0,1]$ définie par*

$$\mu_X(B) = \mathbb{P}(\{X \in B\}), \quad B \in \mathcal{B}(\mathbb{R}).$$

N.B : $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ forme un nouveau espace de probabilité.

Exemple 1.3.1. *-Soit $\Omega = \{1, ..., 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{i\}) = \frac{1}{6}, \forall\, i$*
$X_1(\omega) = \omega, \mu_X(i) = \mathbb{P}(\{X = i\}) = \mathbb{P}(\{i\}) = \frac{1}{6}$.
-Soit $\Omega = \{1, ..., 6\} \times \{1, ..., 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{(i,j)\}) = \frac{1}{36}, \forall\, (i,j) \in \Omega$
$X(\omega) = X(\omega_1, \omega_2) = \omega_1 + \omega_2$
On a alors : $X(\Omega) = \{2, 3, 4, ..., 12\}$
Par exemple :

$$\begin{aligned}
\mu_X(\{7\}) &= \mathbb{P}(\{X \in \{7\}\}) \\
&= \mathbb{P}(\{X = 7\}) \\
&= \mathbb{P}(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) \\
&= \frac{1}{6}
\end{aligned}$$

1.3.1 Fonction de répartition d'une variable aléatoire

Définition 1.3.2. *La fonction de répartition d'une v.a X est l'application $F_X : \mathbb{R} \to [0,1]$ définie par*

$$F_X(t) = \mathbb{P}(\{X \leq t\}) = \mu_X(]-\infty, t[), \quad t \in \mathbb{R}.$$

Proposition 1.3.1. *La donnée de F_X équivaut à celle de μ_X.*

1.3.2 Deux types particulières de variables aléatoires

A) Variable aléatoire discrète :
X prend ses valeurs dans un ensemble D dénombrable ($X(\omega) \in D, \forall \omega \in \Omega$). Dans ce cas, on a : $\sigma(X) = \sigma(\{X = x\}, x \in D)$, $p(x) = \mathbb{P}(\{X = x\}) \geq 0$ et $\sum_{x \in D} p(x) = \mathbb{P}(\{x \in D\}) = 1$.
De plus, $F_X(t) = \sum_{x \in D : x \leq t} p(x)$.
B) Variable aléatoire continue :
$\mathbb{P}(\{X \in B\}) = 0$ si |B| = 0 (en part $\mathbb{P}(\{X = x\}) = 0, \ \forall x$). Sous cette condition, le théorème de Randon-Nikodym assure l'existence d'une fonction borélienne $f_X : \mathbb{R} \to \mathbb{R}$ (appelée densité) telle que

$$f_X(x) \geq 0, \, \forall x \in \mathbb{R}, \qquad \int_{\mathbb{R}} f_X(x)dx = 1 \quad et \quad \mathbb{P}(\{X \in B\}) = \int_B f_X(x)dx.$$

De plus, $F_X(t) = \int_{-\infty}^{t} f_X(x)dx$ et $F_X^{'}(t) = f_X(t)$

Exemple 1.3.2. *A) Loi binomiale B(n,p), $n \geq 1$, $p \in [0,1]$:*

$$p(k) = \mathbb{P}(\{X = k\}) = \begin{pmatrix} n \\ p \end{pmatrix} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n,$$

où $\begin{pmatrix} n \\ p \end{pmatrix} = \frac{n!}{k!(n-k)!}$

B) La loi gaussienne $\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \quad x \in \mathbb{R}$$

Remarque 1.3.1. *Si X et Y suivant une même loi, on dit que X et Y sont identiquement distribuées (i.d) et on note $X \sim Y$ (ne pa confondre avec $X \simeq Y$, qui veut dire que X est à peu près égale à Y).*

1.4 Espérance d'une variable aléatoire

1.4.1 Construction de l'espérance (= intégrale de Lebesgue !)

On procède en trois étapes :
1. Soit $X(\omega) = \sum_{i=0}^{\infty} x_i \mathbb{1}_{A_i}(x), x_i \geq 0, A_i \in \mathcal{F}$. On définit l'espérance de telles v.a (dites simples) comme suit :

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} x_i \mathbb{P}(A_i) \in [0, +\infty]$$

Exemple 1.4.1. *-Si $X = \mathbb{1}_A$, $\mathbb{P}(A) = p$, alors $\mathbb{E}(X) = \mathbb{P}(A) = p$.*
-Si $X = c\mathbb{1}_\Omega = cte$ sur Ω, alors $\mathbb{E}(X) = c$.

2. Soit X une v.a $\mathcal{F}$-mesurable telle que $X(\omega) \geq 0, \forall \omega \in \Omega$. On pose

$$X_n(\omega) = \sum_{i=0}^{\infty} \frac{i}{2^n} \mathbb{1}_{\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\}}(\omega).$$

Alors (X_n) est une suite croissante de v.a qui tend vers X. On définit

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} \sum_{i=0}^{\infty} \frac{i}{2^n}\mathbb{P}(\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\}) \in [0, +\infty]$$

3. Soit X une v.a $\mathcal{F}$-mesurable quelconque. On pose

$$X(\omega) = X^+(\omega) - X^-(\omega) \quad avec \begin{cases} X^+(\omega) = \max(0, X(\omega)) \geq 0, \\ X^-(\omega) = \max(0, -X(\omega)) \geq 0. \end{cases}$$

On a alors $|X(\omega)| = X^+(\omega) + X^-(\omega) \geq 0$

— Si $\mathbb{E}(|X|) < \infty$, alors on définit $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$

— Si $\mathbb{E}(|X|) = \infty$, alors on dit que $\mathbb{E}(X)$ n'est pas définie.

Remarque 1.4.1. *-Si $\mathbb{E}(X) = 0$, alors on dit que X est une <u>v.a centrée</u>.*
-Si $\mathbb{E}(|X|) < \infty$, alors on dit que X est une <u>v.a intégrable</u>.
-Si $\mathbb{E}(X^2) < \infty$, alors on dit que X est une <u>v.a de carré intégrable</u>.
-On dit que X une <u>v.a de carré bornée</u> s'il existe une cte $K > 0$ telle que $|X(\omega)| \leq K, \forall \omega \in \Omega$.

Remarque 1.4.2. *X bornée $\Rightarrow \mathbb{E}(X^2) < \infty \Rightarrow \mathbb{E}(|X|) < \infty$.*

Proposition 1.4.1. *Soient X une v.a et $g : \mathbb{R} \to \mathbb{R}$ une fonction borélienne telle que $\mathbb{E}(|g(X)|) < \infty$. Alors*

A) Si X est une v.a discrète (à valeurs dans D dénombrable), alors

$$\mathbb{E}(g(X)) = \sum_{x\in D} g(x)\mathbb{P}(\{X = x\})$$

B) Si X est une v.a continue (avec densité f_X), alors

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

Ceci s'applique en particulier si $g(x) = x$.

1.4.2 Variance et covariance de variables aléatoires

Définition 1.4.1. *Soient X et Y deux v.a de carré intégrable. On pose*

$$Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geq 0$$
$$Cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Remarque 1.4.3. *-Un événement $A \in \mathcal{F}$ est dit négligeable si $\mathbb{P}(A) = 0$.*
-Un événement $A \in \mathcal{F}$ est dit presque sûr (souvent abrégé p.s) si $\mathbb{P}(A) = 1$, i.e si $\bar{A}$ est négligeable.

Exemple 1.4.2. *Soit X une v.a telle que $\mathbb{P}(\{X = c\}) = 1$. Alors on dit que $X = c$ presque sûrement ("$X = c$ p.s")*

Propriété 1.4.1. ***(Propriétés de l'espérance)***
Soient X, Y deux v.a intégrables.
-Linéarité : $\mathbb{E}(cX + Y) = c\,\mathbb{E}(X) + \mathbb{E}(Y), c \in \mathbb{R}$ et X, Y v.a intégrables.
-Positivité : si $X \geq 0$ p.s, alors $\mathbb{E}(X) \geq 0$.
-Positivité stricte : si $X \geq 0$ p.s, alors $\mathbb{E}(X) = 0$, alors $X = 0$ p.s.

1.5 Indépendance

1.5.1 Indépendance des événements

Définition 1.5.1. *Deux événements A et B ($\in \mathcal{F}$) sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Notation : Si A est indépendant de B, on note $A \perp B$
Conséquence

$$\begin{aligned}
\mathbb{P}(\bar{A} \cap B) &= \mathbb{P}(B \backslash (A \cap B)) \\
&= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(B)(1 - \mathbb{P}(A)) \\
&= \mathbb{P}(\bar{A})\mathbb{P}(B)
\end{aligned}$$

De même, on a $\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A)\mathbb{P}(\bar{B})$ et $\mathbb{P}(\bar{A} \cap \bar{B}) = \mathbb{P}(\bar{A})\mathbb{P}(\bar{B})$

Définition 1.5.2. *n événements $A_1, ..., A_n \in \mathcal{F}$ sont indépendants si*

$$\mathbb{P}(A_1^* \cap ... \cap A_n^*) = \prod_{i=1}^{n} \mathbb{P}(A_i^*), \quad telleque\ A_i^* = A_i \vee \bar{A}_i$$

Proposition 1.5.1. *n événements $A_1, ..., A_n \in \mathcal{F}$ sont indépendants ssi*

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i), \quad \forall I \subset \{1, ..., n\}.$$

Remarque 1.5.1. *-Pour $n > 2$, la condition $\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$ ne suffit pas.*
-L'indépendance de n événements telle que définie ci-dessus est plus forte que i'indépendance deux à deux $(A_i \perp A_j, \forall i \neq j)$

1.5.2 Indépendance de tribus

Définition 1.5.3. *Une famille $(\mathcal{F}_1, ..., \mathcal{F}_n)$ de sous-tribus de $\mathcal{F}$ est indépendante si*

$$\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1)...\mathbb{P}(A_n), \quad \forall A_1 \in \mathcal{F}_1, ..., A_n \in \mathcal{F}_n.$$

Proposition 1.5.2. *$(\sigma(A_1), ..., \sigma(A_n))$ est une famille de sous-tribus indépendantes ssi les événements $(A_1, ..., A_n)$ sont indépendants.*

1.5.3 Indépendance de variables aléatoires

Définition 1.5.4. *Une famille $(X_1, ..., X_n)$ de v.a ($\mathcal{F}$-mesurables) est indépendante si $(\sigma(A_1), ..., \sigma(A_n))$ est indépendante.*

Proposition 1.5.3. *$(X_1, ..., X_n)$ est une famille de v.a indépendantes ssi les événements $\{X_1 \leq t_1\}, ..., \{X_n \leq t_n\}$ sont indépendants $\forall t_1, ..., t_n \in \mathbb{R}$.*
En particulier : $X \perp Y$ ssi $\mathbb{P}(\{X \leq t, Y \leq s\}) = \mathbb{P}(\{X \leq t\})\mathbb{P}(\{Y \leq s\}), \quad \forall t, s \in \mathbb{R}$.

Exemple 1.5.1. *Soit $\Omega = \{1, ..., 6\} \times \{1, ..., 6\}, \mathcal{F} = \mathcal{P}(\Omega), \mathbb{P}(\{i, j\}) = \frac{1}{36}$*
On pose $X(\omega) = X(\omega_1, \omega_2) = \omega_1, \quad Y(\omega) = Y(\omega_1, \omega_2) = \omega_2$.
Calculons

$$\begin{aligned}
\mathbb{P}(\{X = i\}) &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = i\}) \\
&= \mathbb{P}(\{(i, 1), ..., (i, 6)\}) \\
&= \frac{1}{6}
\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{Y=j\}) &= \mathbb{P}(\{\omega\in\Omega : Y(\omega)=j\})\\ &= \mathbb{P}(\{(1,j),...,(6,j)\})\\ &= \frac{1}{6}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{X=i,Y=j\}) &= \mathbb{P}(\{\omega\in\Omega : X(\omega)=i\wedge Y(\omega)=j\})\\ &= \mathbb{P}(\{(i,j)\})\\ &= \frac{1}{36}=\frac{1}{6}\times\frac{1}{6}\\ &= \mathbb{P}(\{X=i\}).\mathbb{P}(\{Y=j\}),\quad \forall (i,j)\in\Omega\end{aligned}$$

donc X et Y sont indépendantes.

Exemple 1.5.2. *Si $X(\omega)=c, \forall\omega\in\Omega$, alors $X\perp Y$, $\forall\, Y$ (une v.a constante est indépendante de toute autre v.a définie sur le même espace).*

Proposition 1.5.4. *Soient $c\in\mathbb{R}$ et X, Y deux v.a de carré intégrable. Si $X\perp Y$, alors*

$$\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)\quad ...\,(1)$$
$$Var(cX+Y)=c^2Var(X)+Var(Y)\quad ...\,(2)$$

La première égalité signifie que deux v.a sont indépendants alors elles sont décorrélées ; la réciproque n'est pas vraie.

Proposition 1.5.5. *Si $X\perp Y$ et X, Y sont deux v.a continues possèdent une densité conjointe $f_{X,Y}$ (c-à-d $\mathbb{P}((X,Y)\in B)=\int\int_B f_{X,Y}(x,y)dxdy$, $\forall B\in\mathcal{B}(\mathbb{R}^2)$), alors*

$$f_{X,Y}(x,y)=f_X(x)f_Y(y),\quad \forall x,y\in\mathbb{R}$$

1.6 Probabilités conditionnelles et lois conditionnelles

1.6.1 Conditions des événements

Soit $(\Omega,\mathcal{F},\mathbb{P})$ un espace de probabilité.

Définition 1.6.1. *Étant donné deux événements A et $B\in\mathcal{F}$, la probabilité conditionnelle de B sachant A (ou de B conditionné par A) où $\mathbb{P}(A)\neq 0$ est définie par :*

$$\mathbb{P}(B\backslash A)=\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)}\overset{notée}{=}\mathbb{P}^A(B)$$

Théorème 1.6.1. *Deux événements quelconque A et B telle que : $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$ vérifiant :*

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(B \backslash A)\mathbb{P}(A)\\ &= \mathbb{P}(A \backslash B)\mathbb{P}(B)\end{aligned}$$

Théorème 1.6.2. ***(Probabilité conditionnelle)***
Pour un événement A telle que $\mathbb{P}(A) \neq 0$, la mesure $\mathbb{P}(.\backslash A) = \frac{\mathbb{P}(.\cap A)}{\mathbb{P}(A)}$ définit sur $(\Omega, \mathcal{F})$ est une probabilité dite conditionnelle par rapport à l'événement A pour tous $B \in \mathcal{F}$:

$$\begin{aligned}\mathbb{P}(.\backslash A) &: (\Omega, \mathcal{F}) \to [0,1]\\ \mathbb{P}(B \backslash A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}\end{aligned}$$

1.6.2 Lois conditionnelles

Dans ce paragraphe, il s'agit d'explicité la probabilité conditionnelle dans deux cas caractérisé par le caractère discret et continu des variables conditionnantes et les variables conditionnées.

I. Loi d'une variable aléatoire discrète

Théorème 1.6.3. *Etant donné un vecteur aléatoire (X, Y) défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans le sous-ensemble au plus dénombrable $D \in \mathbb{R}^2$ de loi conjointe $\mathbb{P}_{X,Y}$ connue. La loi conditionnelle de Y sachant X est définie par :*
$\forall (x_i, y_i) \in D$ *telleque* $\mathbb{P}(X = x_i) \neq 0$:

$$\begin{aligned}\mathbb{P}_{Y\backslash X}(y_i \backslash X = x_i) &= \mathbb{P}_{Y\backslash X=x_i}(y_i)\\ &= \frac{\mathbb{P}_{X,Y}(X = x_i, Y = y_i)}{\mathbb{P}_X(X = x_i)}\end{aligned}$$

Exemple 1.6.1. *On lance une pièce parfaite 4 fois de suite.*
Soit X le nombre aléatoire de face obtenus et Y le nombre de séries de face obtenus par exemple :
$(FFPP) \Rightarrow X = 2 \wedge Y = 1$
1) trouver la loi conjointe du couple (X, Y)
2) les lois marginales de X et de Y
3) la matrice de dispersion (variance-covariance)

Solution

1) La loi conjointe du couple (X,Y) :
$\Omega = \{(PPPP), (PPPF), (PPFP), (PPFF), (PFPP), (PFPF), (PFFP), (PFFF), (FPPP),$

$(FPPF), (FPFP), (FPFF), (FFPP), (FFPF), (FFFP), (FFFF)\}$

$D = X(\Omega) \times Y(\Omega)$ telle que $X(\Omega) = \{0, 1, 2, 3, 4\} \wedge Y(\Omega) = \{0, 1, 2\}$

$\Rightarrow D = \{0, 1, 2, 3, 4\} \times \{0, 1, 2\}$

La loi conjointe de (X , Y) : $\mathbb{P}_{X,Y}(X = x_i, Y = y_i);\ x_i = \overline{0,4},\ y_i = \overline{0,2}$

$$\begin{aligned}\mathbb{P}(\{X = 0, Y = 0\}) &= \mathbb{P}(\{\omega \in \Omega / X(\omega) = 0 \wedge Y(\omega) = 0\}) \\ &= \mathbb{P}\{(PPPP)\} = \tfrac{1}{16}\end{aligned}$$

$\mathbb{P}(\{X = 0, Y = 1\}) = \mathbb{P}(\emptyset) = 0 = \mathbb{P}(\{X = 0, Y = y_i\})$

$$\begin{aligned}\mathbb{P}(\{X = 1, Y = 1\}) &= \mathbb{P}\{(FPPP), (PFPP), (PPFP), (PPPF)\} \\ &= \tfrac{4}{16}\end{aligned}$$

Y \ X	0	1	2	3	4	$\mathbb{P}_Y(y_i)$
0	$\frac{1}{16}$	0	0	0	0	$\frac{1}{16}$
1	0	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{10}{16}$
2	0	0	$\frac{3}{16}$	$\frac{2}{16}$	0	$\frac{5}{16}$
$\mathbb{P}_X(x_i)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	$\frac{16}{16} = 1$

2) Loi marginale de X :

$\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = \sum_{y_i=0}^{2} \mathbb{P}_{X,Y}(X = x_i, Y = y_i)$

Loi marginale de Y :

$\mathbb{P}_Y(y_i) = \mathbb{P}(Y = y_i) = \sum_{x_i=0}^{4} \mathbb{P}_{X,Y}(X = x_i, Y = y_i)$

3) La matrice de dispersion de (X,Y) : :

$$\Sigma_{X,Y} = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

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

Alors

$$\begin{cases} Var(X) = 1 \\ Var(Y) = \frac{5}{16} \end{cases}$$

$$\begin{aligned}Cov(X,Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

avec $\mathbb{E}(XY) = \sum_{x_i} \sum_{y_i} x_i y_i \mathbb{P}_{X,Y}(X = x_i, Y = y_i)$

$= \frac{7}{2}$

Donc

$$\Sigma_{X,Y} = \begin{pmatrix} 1 & 1 \\ 1 & \frac{5}{16} \end{pmatrix}$$

Remarque 1.6.1. $X \rightsquigarrow B(4, \frac{1}{2}) \Rightarrow \forall k = \overline{0,4}, \quad \mathbb{P}(X = k) = C_4^k (\frac{1}{2})^k (\frac{1}{2})^{4-k}$

II. Loi d'une variable aléatoire continue

Théorème 1.6.4. *Etant donnée un vecteur aléatoire (X,Y) de densité conjointe $f_{X,Y}(x, y)$ telle que : $\forall x, f_X(x) \neq 0$, la loi conditionnelle de Y sachant $X = x$ est absolument continue de densité :*

$$f_{Y \backslash X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Exemple 1.6.2. *Soit $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda y} 1_D(x, y)$ telle que : $D = \{(x, y) \in \mathbb{R}^2$ telleque : $0 \leq x \leq y\}$*

- *Trouvons $f_{Y \backslash X=x}(y)$*

Il faut trouver la loi marginale de X et de Y telle que :

$$\begin{aligned} f_X(x) &= \int_D f_{X,Y}(x, y)dy \\ &= \int_x^\infty \lambda^2 e^{-\lambda y} dy \\ &= -\lambda e^{-\lambda y}|_x^{+\infty} \\ &= \lambda e^{-\lambda x}, \ \forall x \in [0, +\infty[\\ f_Y(y) &= \int_D f_{X,Y}(x, y)dx \\ &= \int_0^y \lambda^2 e^{-\lambda y} dx \\ &= \lambda^2 y e^{-\lambda y}, \ \forall y \in [0, +\infty[\end{aligned}$$

Alors :

$$\begin{aligned} f_{Y \backslash X=x}(y) &= \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\lambda^2 e^{-\lambda y} 1_D(x, y)}{\lambda e^{-\lambda x} 1_{[0,+\infty[}(x)} \\ &= \lambda e^{\lambda(x-y)} 1_{[0,+\infty[}(y) \end{aligned}$$

► *Exercice 1*

Faisons l'hypothèse (optimiste) que deux politiciens mentent parfois indépendamment l'un de l'autre, l'un disant la vérité 3 fois sur 4, l'autre 4 sur 5. S'ils énoncent la même affirmation, quelle la probabilité pour qu'elle soit vraie ?

Soit E :"les deux politiciens énoncent la même assertion ", et pour i=$\overline{1,2}$: A_i : "le i$^{\text{ème}}$ ne ment pas"

Solution

$$\mathbb{P}(A_1 \cap A_2 \backslash E) = \frac{\mathbb{P}(A_1 \cap A_2 \cap E)}{\mathbb{P}(E)}$$

avec : $E = (A_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2)$

$$\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(\bar{A}_1 \cap \bar{A}_2) \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2) + (1 - \mathbb{P}(A_1))(1 - \mathbb{P}(A_2)) \\
&= \frac{13}{20} \\
\mathbb{P}(A_1 \cap A_2 \cap E) &= \mathbb{P}(A_1 \cap A_2) = \frac{3}{5}
\end{aligned}$$

Par conséquent ;

$$\mathbb{P}(A_1 \cap A_2 \backslash E) = \frac{12}{13}$$

► *Exercice 2*

Propriétés des système "sans mémoire", décrits par des v.a géométrique et exponentielle.

Soit X une v.a géométrique de loi $\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \forall k \in \mathbb{N}^*$

1) Démontrer que $\mathbb{P}(X > m + n \backslash X > n) = \mathbb{P}(X > m)$.

Interpréter ce résultat si X est la durée de vie d'un système.

2) Cette propriété caractérisé aussi les v.a exponentielle, et ce sur prenant ?

Solution

1) Démontrons que :

$$\begin{aligned}
\circ\, \mathbb{P}(X > m + n \backslash X > n) &= \frac{\mathbb{P}(\{X > m + n\} \cap \{X > n\})}{\mathbb{P}(X > n)} \\
&= \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > n)} \\
&= \frac{\sum_{k=m+n+1}^{\infty}(1-p)^{k-1}p}{\sum_{k=n+1}^{\infty}(1-p)^{k-1}p} \\
&= \frac{\sum_{k=m+n+1}^{\infty}(1-p)^{k-1}}{\sum_{k=n+1}^{\infty}(1-p)^{k-1}}
\end{aligned}$$

Soit $j = k - m \Rightarrow j = n+1 \to +\infty$

$$= \frac{\sum_{j=n+1}^{\infty}(1-p)^{j+m-1}}{\sum_{k=n+1}^{\infty}(1-p)^{k-1}}$$

$$= (1-p)^m$$

$$\circ\, \mathbb{P}(X > m) = \sum_{k=m+1}^{\infty}(1-p)^{k-1}p$$

$$= p\sum_{k=m+1}^{\infty}(1-p)^{k-1}$$

$$= p(1-p)^m \frac{1}{1-(1-p)}$$

$$= (1-p)^m$$

Ça veut dire que la probabilité que la durée de vie de ce système dépasse un certain tempe "m+n" sachant que ce système à dépasse le temps "n" égale à la probabilité que la durée de vie de ce système dépassera le temps "m", c-à-d : qu'il n'est pas obligatoire de savoir le temps passé de son fonctionnement.

2) Soit $X \rightsquigarrow \xi(\lambda), \quad \lambda > 0 \Leftrightarrow f(x) = \lambda e^{-\lambda x} 1_{[0,+\infty[(x)}$

Trouvons la fonction de répartition de X :

$$F_X(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(X \in]-\infty, x]) = \int_{-\infty}^{x} f(t)dt$$

$$= \begin{cases} 0 & si\ x < 0 \\ 1 - e^{-\lambda x} & si\ x \geq 0 \end{cases}$$

$$\mathbb{P}(X > t+s \backslash X > s) = \frac{\mathbb{P}(\{X > t+s\} \cap \{X > s\})}{\mathbb{P}(X > s)}$$

$$= \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)}$$

$$= \frac{1 - \mathbb{P}(X \leq t+s)}{1 - \mathbb{P}(X \leq s)}$$

$$= \frac{1 - (1 - e^{-\lambda(t+s)})}{1 - (1 - e^{-\lambda s})}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) = 1 - \mathbb{P}(\{X \leq t\}) = \mathbb{P}(\{X > t\})$$

Ce n'est sur prenant car la loi exponentielle étude aussi la durée de vie des systèmes sans mémoire dans le cas continue.

1.7 Espérance conditionnelle

Connaissant la notion de probabilité conditionnelle de l'événement B sachant l'événement A, il est naturel de défini la loi de probabilité conditionnelle de v.a Y sachant la v.a X et définir par conséquence son espérance mathématique lorsque $X = \mathbb{1}_A$ et $Y = \mathbb{1}_B$.

1.7.1 Espérance conditionnelle pour les v.a

I. Cas discret : Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \to E, Y : \Omega \to F$ deux v.a discrètes sur cet espace. Notons $\mathbb{P}_{X,Y}(x, y)$ la loi du couple (X,Y) et $\mathbb{P}_{Y\setminus X=x}(y)$ la loi conditionnelle de $Y\setminus X = x$ telle que : $\mathbb{P}_X(x) = \mathbb{P}(X = x) \neq 0, \quad \forall x \in E$.

Définition 1.7.1. *Lorsque $F \subset \mathbb{R}^2$ et sous réserve d'intégrabilité, l'espérance de Y pour la loi conditionnelle $\mathbb{P}_{Y\setminus X=x}(y)$ vaut :*

$$\begin{aligned}\mathbb{E}(Y\setminus X = x) &= \sum_{y\in F} y\, \mathbb{P}_{Y\setminus X=x}(y) \\ &= \sum_{y\in F} y\, \mathbb{P}(Y = y\setminus X = x)\end{aligned}$$

*et on appelle **l'espérance conditionnelle** de Y sachant X=x.*

Remarque 1.7.1. *Il s'agit d'une fonction de x (la valeur prise par X) $\psi(x) = \mathbb{E}(Y\setminus X = x)$, mais il est très utile de la considérer comme fonction $\psi(x)$.*

Définition 1.7.2. *La variable aléatoire $\psi(x)$ est l'espérance conditionnelle de Y sachant X et on la note $\psi(x) = \mathbb{E}(Y\setminus X)$.*

Attention : $\mathbb{E}(Y\setminus X)$ est une v.a fonction de X.

Exemple 1.7.1. *Reprenons l'exemple de jet 4 fois d'une pièce de manie et trouvons $\mathbb{E}(Y\setminus X)$ et $\mathbb{E}(X\setminus Y)$.*
$\triangleright$ *Posons $\psi(x) = \mathbb{E}(Y\setminus X)$.*
Trouvons $\psi(x) = \mathbb{E}(Y\setminus X = x), \quad \forall x = \overline{0, 4}$.
Pour X = 0 : $\psi(0) = \mathbb{E}(Y\setminus X = 0)$
$= \sum_{y=0}^{2} y\, \mathbb{P}(Y = y\setminus X = 0)$
Trouvons $\mathbb{P}(Y = y\setminus X = 0), \quad \forall i = \overline{0, 2}$.
$\mathbb{P}(Y = 0\setminus X = 0) = 1$
$\mathbb{P}(Y = 1\setminus X = 0) = \mathbb{P}(Y = 2\setminus X = 0) = 0$ *telle que :* $(\sum_{y=0}^{2} y\, \mathbb{P}(Y = y\setminus X = 0) = 1)$

Alors : $\psi(0) = 0$, $\psi(1) = 1$, $\psi(2) = \frac{3}{2}$, $\psi(3) = \frac{3}{2}$, $\psi(4) = 1$

$\Rightarrow \psi(x) = \mathbb{E}(Y\backslash X) = 0.\mathbb{1}_{\{X=0\}} + \mathbb{1}_{\{X=1\}\cup\{X=4\}} + \frac{3}{2}\mathbb{1}_{\{X=2\}\cup\{X=3\}}$

Donc

$$\psi(x) = \begin{cases} 0 & si\ X = 0 \\ 1 & si\ X = 1 \vee X = 4 \\ \frac{3}{2} & si\ X = 2 \vee X = 3 \end{cases}$$

$\triangleright$ $\mathbb{E}(X\backslash Y)$ *(exercice).*

II. Cas continue

Définition 1.7.3. *Soit (X,Y) un couple aléatoire à densité sur* $\mathbb{R}^2$ *et* $x \in \mathbb{R}$ *telle que :* $\mathbb{P}_X(x) = \mathbb{P}(X = x) > 0$.

On définit l'espérance conditionnelle de Y sachant X = x par :

$$\mathbb{E}(Y\backslash X) = \psi(x) \quad avec \quad \psi(x) = \int_{\mathbb{R}} y\ f_{Y\backslash X=x}(y\backslash x)dy$$

Exemple 1.7.2. *Reprenons l'exemple de* $f_{X,Y}(x,y) = \lambda^2 e^{-\lambda y} 1_D(x,y)$ *tq :* $D = \{(x,y) \in \mathbb{R}^2\ tq : 0 \leq x \leq y\}$

Trouvons $\mathbb{E}(X\backslash Y)$ *:*

Soit $\psi(y) = \mathbb{E}(X\backslash Y = y) = \int_{\mathbb{R}} x f_{X\backslash Y=y}(x\backslash y)dx$

Trouvons la densité conjointe de $X\backslash Y = y$ *:*

$$\begin{aligned} Ona : f_{X\backslash Y=y}(x\backslash y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad f_Y(y) > 0 \\ &= \frac{\lambda^2 e^{-\lambda y}\mathbb{1}_D(x,y)}{\lambda^2 y e^{-\lambda y}\mathbb{1}_{[0,+\infty[}(y)} \\ &= \frac{\mathbb{1}_{[0,+\infty[}(y)\mathbb{1}_{[0,y]}(x)}{y\mathbb{1}_{[0,+\infty[}(y)} \\ &= \frac{1}{y}\mathbb{1}_{[0,y]}(x) \\ \Rightarrow \psi(y) &= \int_{\mathbb{R}} x\frac{1}{y}\mathbb{1}_{[0,y]}(x)dx \\ &= \int_0^y \frac{x}{y}dx \\ &= \frac{1}{2}y \end{aligned}$$

Alors : $\psi(Y) = \mathbb{E}(X\backslash Y) = \frac{1}{2}Y$

Exemple 1.7.3. *Considérons le couple aléatoire (X,Y) de densité conjointe* $f_{X,Y}(x,y) = \frac{1}{x}\mathbb{1}_{\{0<y<x<1\}}$.

Calculons $\psi(X) = \mathbb{E}(Y\backslash X)$ *:*

Soit $\psi(x) = \mathbb{E}(Y\backslash X = x) = \int_{\mathbb{R}} y\, f_{Y\backslash X=x}(y\backslash x)dy$

La densité conjointe de $Y\backslash X = x$ *:*

$$f_{Y\backslash X=x}(y\backslash x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_X(x) > 0$$

$f_X(x) = 1, \quad \forall x \in]0,1[$

$$\begin{aligned} Alors : f_{Y\backslash X}(y\backslash x) &= \frac{\frac{1}{x}\mathbb{1}_{]0,1[}(x)\mathbb{1}_{]0,x[}(y)}{\mathbb{1}_{]0,1[}(x)} \\ &= \frac{1}{x}\mathbb{1}_{]0,x[}(y) \end{aligned}$$

$\psi(x) = \int_{\mathbb{R}} y\frac{1}{x}\mathbb{1}_{]0,x[}(y)dy = \frac{x}{2}$

$\Rightarrow \psi(X) = \mathbb{E}(Y\backslash X) = \frac{X}{2}$

• Propriétés de l'espérance conditionnelle

La variable aléatoire conditionnelle $\mathbb{E}(X\backslash Y)$ vérifie les propriétés suivantes :

1. **Positivité** : Si g(.) est une fonction positive $\Rightarrow$ $\mathbb{E}(g(X)\backslash Y) \geq 0$.
2. **Constante** : $\mathbb{E}(c\backslash X) = c, \quad \forall c = cte \in \mathbb{R}$.
3. **Linéarité** : $\mathbb{E}(\alpha g(X) + \beta h(Y)\backslash Z) = \alpha\, \mathbb{E}(g(X)\backslash Z) + \beta\, \mathbb{E}(h(Y)\backslash Z)$
4. **Projection** : $\mathbb{E}(X\backslash X) = X$ (pour la stabilité).

   Démonstration.

   Il convient de démontrer que $\forall \omega \in \Omega : X(\omega) = \mathbb{E}(X\backslash X)(\omega)$.

   Soit $\omega \in \Omega$ *tq* : $X(\omega) = x$ et on a à démontrer que : $x = \mathbb{E}(X\backslash X = x)$.

   Supposons que X est une v.a discrète, alors :

$$\begin{aligned} \mathbb{E}(X\backslash X = x) &= \sum_{x_i \in X(\Omega)} x_i\, \mathbb{P}(X = x_i\backslash X = x) \\ &= x\, \underbrace{\mathbb{P}(X = x\backslash X = x)}_{=1} = x \end{aligned}$$

$\square$

5. $\mathbb{E}(h(X)\backslash X) = h(X)$
6. $\mathbb{E}(\mathbb{E}(X\backslash Y)\backslash Y) = \mathbb{E}(X\backslash Y)$

Démonstration.

On a : $\mathbb{E}(X\backslash Y) = \psi(Y)$

$$\begin{aligned}\overset{P.5}{\Longrightarrow} \mathbb{E}(\mathbb{E}(X\backslash Y)\backslash Y) &= \mathbb{E}(\psi(Y)\backslash Y) = \psi(Y) \\ &= \mathbb{E}(X\backslash Y)\end{aligned}$$

□

7. $\mathbb{E}(h(X).g(Y)\backslash X) = h(X).\mathbb{E}(g(Y)\backslash X)$
8. La loi de l'espérance totale : $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X\backslash Y))$

Démonstration.

Supposons que X est discrète (cas continue à titre exercice).

On a :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x\in X(\Omega)} x\,\mathbb{P}(X = x) \\ &= \sum_{x\in X(\Omega)} x \sum_{y\in Y(\Omega)} \mathbb{P}(X = x, Y = y) \\ &= \sum_{x\in X(\Omega)} x \sum_{y\in Y(\Omega)} \mathbb{P}(Y = y)\mathbb{P}(X = x\backslash Y = y) \\ &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} x\,\mathbb{P}(X = x\backslash Y = y)\mathbb{P}(Y = y) \\ &= \sum_{y\in Y(\Omega)} [\sum_{x\in X(\Omega)} x\,\mathbb{P}(X = x\backslash Y = y)]\mathbb{P}(Y = y) \\ &= \sum_{y\in Y(\Omega)} \mathbb{E}(X\backslash Y = y)\mathbb{P}(Y = y) \\ &= \mathbb{E}(\mathbb{E}(X\backslash Y))\end{aligned}$$

□

1.7.2 Cas générale

I. Espérance conditionnelle par rapport à un événement

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

Définition 1.7.4. *Soit X une v.a intégrable, $B \in \mathcal{F}$ tq : $\mathbb{P}(B) > 0$. On appelle espérance conditionnelle de X sachant B :*

$$\begin{aligned}\mathbb{E}(X\backslash B) &= \int X d\mathbb{P}(.\backslash B) = \frac{1}{\mathbb{P}(B)}\int_B X d\mathbb{P} \\ &= \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}\end{aligned}$$

Remarque 1.7.2.

a) Si X est intégrable, alors : $\mathbb{E}(|X|) = \int |X| d\mathbb{P} < \infty \Rightarrow \int |X| d\mathbb{P}(.\backslash B) < \infty$

b) Si $X = \mathbb{1}_A \Rightarrow \mathbb{E}(\mathbb{1}_A\backslash B) = \mathbb{P}(A\backslash B)$.

II. Espérance conditionnelle par rapport à une tribu

Supposons que $\mathbb{P}(B) > 0$ et $\mathbb{P}(\bar{B}) > 0$, on définit alors $\mathcal{G} = \sigma(B) = \{\emptyset, B, \bar{B}, \Omega\}$.

Définition 1.7.5. *On appelle* ***espérance conditionnelle*** *de X sachant $\mathcal{G}$, la v.a aléatoire $\mathcal{G}$-mesurable :*

$$\begin{aligned}\mathbb{E}(X\backslash \mathcal{G}) &= \mathbb{E}(X\backslash B)\mathbb{1}_B + \mathbb{E}(X\backslash \bar{B})\mathbb{1}_{\bar{B}} \\ &= \mathbb{E}(X\backslash \mathcal{G})(\omega) = \begin{cases} \mathbb{E}(X\backslash B)(\omega) & \omega \in B \\ \mathbb{E}(X\backslash \bar{B})(\omega) & \omega \in \bar{B}\end{cases}\end{aligned}$$

Ainsi, si $(B_k)_k$ est une partition de Ω par des événements telle que : $\forall k \in \mathbb{N} : \mathbb{P}(B_k) > 0$ et $\mathcal{G} = \sigma(B_k,\ \forall k \in \mathbb{N})$, alors :

$$\mathbb{E}(X\backslash \mathcal{G}) = \sum_{k\in\mathbb{N}} \mathbb{E}(X\backslash B_k)\mathbb{1}_{B_k}$$

Remarque 1.7.3. *Il n'y a pas de problème de convergence de la densité car $(B_k)_k$ est une partition de Ω, donc pour tout $\omega \in \Omega$, seule B_{k_0} compte, c-à-d :*

$$\mathbb{E}(X\backslash \mathcal{G})(\omega) = \mathbb{E}(X\backslash B_{k_0})\mathbb{1}_{B_k}$$

Exemple 1.7.4. *Soient (X, Y) deux jets de dés indépendants.*

Trouvons $\mathbb{E}(X + Y\backslash Y)$, en utilisant les événements.

Soit :

$$\begin{aligned}\psi(y) &= \mathbb{E}(X + Y\backslash Y = y) \\ &= \frac{\mathbb{E}[(X+Y)\mathbb{1}_{\{Y=y\}}]}{\mathbb{P}(\{Y = y\})} = \frac{\mathbb{E}(X+Y)}{1} \\ &= \mathbb{E}(X) + y \\ &= \frac{7}{2} + y\end{aligned}$$

$$\Rightarrow \mathbb{E}(X + Y\backslash Y) = \mathbb{E}(X) + Y, \quad Y(\Omega) = \{1, ..., 6\}$$

Exemple 1.7.5. *On lance une pièce de manie deux fois de suite et soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *l'espace de probabilité correspondant :*

$$\Omega = \{(i,j), i = P \vee F \, et \, j = P \vee F\} \Rightarrow card\,\Omega = 4$$
$$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, (P,P), (P,F), (F,P), (F,F), (P,P), (P,F), (F,P), (F,F), ..., \Omega\}$$

$$\Rightarrow card\,\mathcal{F} = 2^{card\,\Omega} = 16 < +\infty$$

La mesure $\mathbb{P}$ *est la probabilité uniforme sur* Ω *c-à-d :*

$$\mathbb{P}(\omega) = \frac{1}{card\,\Omega} = \frac{1}{4}$$

1 *Soient* X, X_1 *et* X_2 *les v.a définies sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *par :*

$X : \Omega \to \{0,1,2\} = X(\Omega)$

$\omega \mapsto X(\omega) = nombre \; de \; piles \; dans \, \omega = (\omega_1, \omega_2)$

$X_1 : \Omega \to \{0,1\} = X_1(\Omega)$

$\omega \mapsto X_1(\omega) = nombre \; de \; piles \; dans \, \omega_1$

$X_2 : \Omega \to \{0,1\} = X_2(\Omega)$

$\omega \mapsto X_2(\omega) = nombre \; de \; piles \; dans \, \omega_2$

i)

$$\begin{aligned}\mathcal{F}_{X_1} = \sigma(X_1) &= \sigma(\{\omega \in \Omega, X_1(\omega) = x, x \in X_1(\Omega)\}) \\ &= \sigma(\{\omega \in \Omega, X_1(\omega) = 0, X_1(\omega) = 1, \omega \in \Omega\}) \\ &= \sigma(\{(F,P),(F,F)\}, \{(P,P),(P,F)\}) \\ &= \{\emptyset, \{(F,P),(F,F)\}, \{(P,P),(P,F)\}, \Omega\}\end{aligned}$$

ii)

$$\begin{aligned}\mathcal{F}_{X_2} = \sigma(X_2) &= \sigma(\{\omega \in \Omega, X_2(\omega) = x, x \in X_2(\Omega)\}) \\ &= \sigma(\{\omega \in \Omega, X_2(\omega) = 0, X_2(\omega) = 1, \omega \in \Omega\}) \\ &= \sigma(\{(F,F),(P,F)\}, \{(P,P),(F,P)\}) \\ &= \{\emptyset, \{(F,F),(P,F)\}, \{(P,P),(F,P)\}, \Omega\}\end{aligned}$$

Enfin : $\mathcal{F}_X = \sigma(X) = \{\emptyset, \{(F,F)\}, \{(P,F),(F,P)\}, \{(P,P)\}, \{(F,F),(P,F),(F,P)\}, \{(F,F),(P,P)\},$
$\{(P,F),(F,P),(P,P)\}, \Omega\}$

2 *Calculons* $\mathbb{E}(X \backslash \mathcal{F}_{X_1})$

On remarque $\mathcal{F}_{X_1} = \sigma(\{B, \bar{B}\})$ *telle que :* B = *{(F,P), (F,F)}*

$\Rightarrow \mathbb{E}(X \backslash \mathcal{F}_{X_1}) = \mathbb{E}(X \backslash B)\mathbb{1}_B + \mathbb{E}(X \backslash \bar{B})\mathbb{1}_{\bar{B}}$

$*$ *Trouvons* $\mathbb{E}(X\backslash B), tq : B = \{(F,P),(F,F)\}$ *et* $\mathbb{E}(X\backslash B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)},\ \mathbb{P}(B) > 0$
telle que :

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(\{(F,P),(F,F)\}) = \frac{1}{2}\\
\mathbb{E}(X\mathbb{1}_B) &= \sum_{x=0}^{2} x\,\mathbb{P}(X=x,B)\\
&= \sum_{x=0}^{2} x\,\mathbb{P}(\{\omega\in\Omega, X(\omega)=x\}\cap B) = \frac{1}{4}
\end{aligned}$$

$\Rightarrow \mathbb{E}(X\backslash B) = \frac{1}{2}$ *et* $\mathbb{E}(X\backslash \bar{B}) = \frac{3}{2}$
Alors :

$$\begin{aligned}
\mathbb{E}(X\backslash \mathcal{F}_{X_1}) &= \frac{1}{2}\mathbb{1}_B + \frac{3}{2}\mathbb{1}_{\bar{B}}\\
&= \frac{1}{2}\mathbb{1}_{\{(F,F),(F,P)\}}(\omega) + \frac{3}{2}\mathbb{1}_{\{(P,P),(P,F)\}}(\omega)\\
&= \begin{cases} \frac{1}{2} & \omega\in B\\ \frac{3}{2} & \omega\in\bar{B}\end{cases}
\end{aligned}$$

- ***Exercice***

$*$ *Calculons* $\mathbb{E}(X\backslash X_1)$ *et on déduire la relation entre* $\mathbb{E}(X\backslash X_1)$ *et* $\mathbb{E}(X\backslash \mathcal{F}_{X_1})$.
$*$ *Calculons* $\mathbb{E}(X\backslash \mathcal{F}_{X_2})$ *et comparer avec* $\mathbb{E}(X\backslash X_2)$.

Théorème 1.7.1. *Soit* X *une v.a intégrable (ou positive) sur* $(\Omega,\mathcal{F},\mathbb{P})$. *Soit* $\mathcal{G}$ *une sous-tribu de* $\mathcal{F}$ *alors il existe une v.a* Y $\mathcal{G}$*-mesurable et* $\mathbb{P}$*-intégrable (ou positive) vérifiant :*

$$\forall A\in\mathcal{G}, \int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$$

Cette v.a est définie $\mathbb{P}_{\mathcal{G}}$*-presque sûrement de manière unique, elle est appelée* ***espérance conditionnelle*** *de* X *sachant* $\mathcal{G}$ *et se note* $\mathbb{E}(X\backslash\mathcal{G})$.

Remarque 1.7.4. *Dans le cas où* X^2 *est intégrable, on établit facilement cette existence.*

En effet, on notons L^2 l'espace des v.a réelles X où : $\mathbb{E}(X^2) < +\infty$.
L^2 muni du produit scalaire $< X,Y >= \|\mathbb{E}(XY)\|$ et de la norme associée $\|X\|_2 = \sqrt{< X,X >} = \sqrt{\mathbb{E}(X^2)}$ est un espace de Hilbert.
Rappelons qu'on définit L^2 en introduisant sur :

$$\mathscr{L}^2 = \{X \ v.a\ tq : \mathbb{E}(X^2) < \infty\}$$

La relation d'équivalence : $X\sim Y \Leftrightarrow (\mathbb{P}(X=Y)=1 \ et\ L^2 = \mathscr{L}^2/\sim)$.
Soit $L^2(\mathcal{G}) \subset L^2$ le sous espace constituée des v.a.r $\mathcal{G}$-mesurables. D'après le théorème de

projection : $L^2 = L^2(\mathcal{G}) \oplus L^2(\mathcal{G})^\perp$ et tout élément $X \in L^2$ s'écrit d'une façon unique $X = Y + (X - Y)$ où Y est la projection orthogonale de X sur $L^2(\mathcal{G})$ c-à-d caractérisée par :

i) Y est $\mathcal{G}$-mesurable.

ii) $(X - Y) \perp Z$ i.e $\mathbb{E}(XZ) = \mathbb{E}(YZ), \forall Z \in L^2(\mathcal{G})$.

On a alors : $\forall A \in \mathcal{G}, \int \mathbb{1}_A X d\mathbb{P} = \int \mathbb{1}_A Y\, d\mathbb{P}$

Ceci prouve l'existence de l'espérance conditionnelle pour des v.a dans L^2 et donne une interprétation géométrique concrète, pour $X \in L^2$, $Y = \mathbb{E}(X \backslash \mathcal{G})$ est la meilleure approximation de X au sens L^2 pour les v.a de $L^2(\mathcal{G})$. Il s'agit donc de l'unique élément de $L^2(\mathcal{G})$ réalisant l'égalité :

$$\|X - Y\|_2 = \inf_{Z \in L^2(\mathcal{G})} \|X - Z\|_2$$

Démonstration.

1) L'unicité

Supposons qu'il existe Y_1, Y_2 vérifiant la définition, on a alors : $\{Y_1 < Y_2\} \in \mathcal{G}$ et $\int_{\{Y_1<Y_2\}} X d\mathbb{P} = \int_{\{Y_1<Y_2\}} Y_1\, d\mathbb{P} = \int_{\{Y_1<Y_2\}} Y_2\, d\mathbb{P}$.

D'où : $\int_{\{Y_1<Y_2\}} (Y_1 - Y_2) d\mathbb{P}$ et donc : $(Y_1 - Y_2)\mathbb{1}_{\{Y_1<Y_2\}} = 0\ p.s$

$\Rightarrow \mathbb{P}(Y_1 < Y_2) = 0$.

De même : $\mathbb{P}(Y_1 > Y_2) = 0 \Rightarrow \mathbb{P}(Y_1 = Y_2) = 1$

$\Rightarrow Y_1 = Y_2\ p.s.$

2) L'existence

On a vu l'existence de Y si $X \in L^2$.

En décomposant $X = X^+ - X^-$.

On voit qu'il suffit de traiter le cas $X \geq 0$, $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Posons : $X_n = \inf(X, n)$, on voit que $X_n \in L^2$.

De plus, il s'agit d'une suite croissante vers X et pour toute $n \in \mathbb{N}, \exists\, Y_n = \mathbb{E}(X_n \backslash \mathcal{G}) \geq 0$ p.s.

En effet ;

$\int_{\{Y_n<0\}} X_n\, d\mathbb{P} = \int_{\{Y_n<0\}} Y_n\, d\mathbb{P} \leq 0$

Par construction, on a : $X_n \geq 0$ et $X_n \mathbb{1}_{\{Y<0\}} \geq 0$ donc : $\mathbb{P}(Y_n < 0) = 0$

Comme $X_{n+1} - X_n \geq 0$, on a : $\mathbb{E}(X_{n+1} \backslash \mathcal{G}) - \mathbb{E}(X_n \backslash \mathcal{G}) = \mathbb{E}(X_{n+1} - X_n \backslash \mathcal{G}) \geq 0$

On obtient alors par la convergence monotone :

$$\forall A \in \mathcal{G}, \int_A X d\mathbb{P} = \lim \int_A X_n d\mathbb{P} = \int_A Y_n d\mathbb{P} = \lim \int_A Y d\mathbb{P}$$

Donc : il existe une v.a $Y = \lim \mathbb{E}(X_n \backslash \mathcal{G})$. □

Remarque 1.7.5. *1) $\mathcal{G}$ représente l'information disponible et $\mathbb{E}(X\backslash\mathcal{G})$ est la meilleure prévision qu'on peut faire sur X en connaissant $\mathcal{G}$.*
2) On ne définit l'espérance conditionnelle qu'à un presque sûr près.

1.8 Propriétés de l'espérance conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité. Considérons $\mathcal{G}$ une sous tribu de $\mathcal{F}$. Soient X, Y des v.a, $(a, b) \in \mathbb{R}^2$ et $A \in \mathcal{G}$:

1. **Linéarité** : $\mathbb{E}(aX + bY\backslash\mathcal{G}) = a\,\mathbb{E}(X\backslash\mathcal{G}) + b\,\mathbb{E}(Y\backslash\mathcal{G})$.
2. **Positivité** : Si $X \geq 0 \Rightarrow \mathbb{E}(X\backslash\mathcal{G}) \geq 0$.
3. **Information minimale** : Soit $\tau \subset \mathcal{G} \Rightarrow \mathbb{E}(X\backslash\tau) = \mathbb{E}(\mathbb{E}(X\backslash\mathcal{G})\backslash\tau)$
4. $\mathbb{E}(X\backslash\{\emptyset, \Omega\}) = \mathbb{E}(X)$ p.s
5. Si X est $\mathcal{G}$-mesurable $\Rightarrow \mathbb{E}(X\backslash\mathcal{G}) = X$ p.s
6. Si X et $\mathcal{G}$ sont indépendants $\Rightarrow \mathbb{E}(X\backslash\mathcal{G}) = \mathbb{E}(X)$ p.s
   En effet ;
   Il suffit d'écrire :
$$\begin{aligned}\int_A X\,d\mathbb{P} &= \int \mathbb{1}_A X\,d\mathbb{P} = \mathbb{P}(A)\int X\,d\mathbb{P}\\ &= \mathbb{P}(A)\,\mathbb{E}(X)\\ &= \int_A \mathbb{E}(X)\,d\mathbb{P}\end{aligned}$$
7. Si X, Y $\in L^1$ et $X \leq Y$ p.s $\Rightarrow \mathbb{E}(X\backslash\mathcal{G}) \leq \mathbb{E}(Y\backslash\mathcal{G})$ p.s.
   En effet ;
   Soit $A = \{\omega \in \Omega,\ \mathbb{E}(X\backslash\mathcal{G})(\omega) > \mathbb{E}(Y\backslash\mathcal{G})(\omega)\}$
   On a : $\int_A X\,d\mathbb{P} = \int_A \mathbb{E}(X\backslash\mathcal{G})\,d\mathbb{P} \leq \int_A Y\,d\mathbb{P} = \int_A \mathbb{E}(Y\backslash\mathcal{G})\,d\mathbb{P}$
   D'où : $\int \mathbb{1}_A\mathbb{E}(X\backslash\mathcal{G})\,d\mathbb{P} - \int \mathbb{1}_A\mathbb{E}(Y\backslash\mathcal{G})\,d\mathbb{P} \leq 0$
   Et donc : $\mathbb{P}(\mathbb{E}(X\backslash\mathcal{G})) > \mathbb{E}(Y\backslash\mathcal{G}) = 0$
   $\Leftrightarrow \mathbb{P}(A) = 0.$
8.

   Théorème 1.8.1. ***(Convergence monotone conditionnelle)***
   Si $X_n \geq 0$ et X_n croit p.s vers $X \in L^1$, alors :
$$\lim_{n\to\infty} \mathbb{E}(X_n\backslash\mathcal{G}) = \mathbb{E}(X\backslash\mathcal{G})\ p.s$$

Démonstration. (Exercice) □

9.

Théorème 1.8.2. ***(Convergence dominée conditionnelle)***
Soit $|X_n| \leq M$, $X_n \in L^1$ convergeant p.s vers X, alors la suite $(\mathbb{E}(X_n \backslash \mathcal{G}))_n$ converge p.s vers $\mathbb{E}(X \backslash \mathcal{G})$

10.

Théorème 1.8.3. ***(Lemme de Fatou conditionnelle)***
Si $X_n \geq 0$, alors : $\mathbb{E}(\liminf X_n \backslash \mathcal{G}) \leq \liminf \mathbb{E}(X_n \backslash \mathcal{G})$

11. Soient X, Y deux v.a tq X, Y et XY sont intégrables, si X est $\mathcal{G}$-mesurable, alors :

$$\mathbb{E}(XY \backslash \mathcal{G}) = X\mathbb{E}(Y \backslash \mathcal{G})$$

Démonstration. On suppose $X, Y \geq 0$, il existe une suite $X_n \nearrow X$, où les X_n sont étagées et $\mathcal{G}$-mesurables (car X l'est).
On a : $0 \leq X_n Y$ et $X_n Y \nearrow XY$.
Par le théorème de convergence monotone conditionnelle : $\mathbb{E}(X_n Y \backslash \mathcal{G}) \nearrow \mathbb{E}(XY \backslash \mathcal{G})$ p.s
Soit $C \in \mathcal{G}, \forall A \in \mathcal{G}$ on a : $C \cap A \in \mathcal{G}$ et :

$$\begin{aligned} \int_A \mathbb{1}_C Y \, d\mathbb{P} &= \int_{A \cap C} Y \, d\mathbb{P} = \int_{A \cap C} \mathbb{E}(Y \backslash \mathcal{G}) \, d\mathbb{P} \\ &= \int_A \mathbb{1}_C \mathbb{E}(Y \backslash \mathcal{G}) \, d\mathbb{P} \end{aligned}$$

Par l'unicité de l'espérance conditionnelle, on trouve : $\mathbb{E}(Y \backslash \mathcal{G})\mathbb{1}_C = \mathbb{E}(\mathbb{1}_C Y \backslash \mathcal{G})$
Donc : $\mathbb{E}(X_n Y \backslash \mathcal{G}) = X_n \, \mathbb{E}(Y \backslash \mathcal{G}) \nearrow X \, \mathbb{E}(Y \backslash \mathcal{G})$ □

12.

Théorème 1.8.4. ***(Inégalité de Jensen conditionnelle)***
Soit X une v.a intégrable à valeurs dans I, $f : I \to \mathbb{R}$ une fonction convexe, alors $\mathbb{E}(X \backslash \mathcal{G}) \in I$ p.s et si $\mathbb{E}(f) < +\infty$, alors :

$$f(\mathbb{E}(X \backslash \mathcal{G})) \leq \mathbb{E}(f(X) \backslash \mathcal{G}) \; p.s$$

Proposition 1.8.1. *Soit X une v.a.r, alors X est indépendant de la tribu $\mathcal{G}$ ssi $\forall\, U \in \mathbb{R}$, $\mathbb{E}(e^{iUX} \backslash \mathcal{G}) = \mathbb{E}(e^{iUX})$ p.s*

Démonstration.
Compte tenu de la proposition 6, on a la 1[er] implication, il suffit de démontrer que si :

$\forall\, U \in \mathbb{R} : \mathbb{E}(e^{iUX}\backslash\mathcal{G}) = \mathbb{E}(e^{iUX}) \overset{??}{\Rightarrow} X\ et\ \mathcal{G}$ sont indépendants.

• Si $\mathbb{E}(e^{iUX}\backslash\mathcal{G}) = \mathbb{E}(e^{iUX})$ p.s

On a pour tout $G \in \mathcal{G} : \mathbb{E}(e^{iUX}\mathbb{1}_G) = \mathbb{E}(e^{iUX})\mathbb{1}_G$

Par définition de l'espérance conditionnelle si $\mathbb{P}(G) \neq 0$, on peut écrire :

$$\mathbb{E}(e^{iUX}\frac{\mathbb{1}_G}{\mathbb{P}(G)}) = \mathbb{E}(e^{iUX})$$

Cela signifie que la fonction caractéristique de X est la même sous la probabilité $\mathbb{P}$ et sous la probabilité $\frac{\mathbb{1}_G}{\mathbb{P}(G)}$.

L'égalité des fonctions caractéristiques entraine l'égalité des lois et par conséquent, pour toute fonction borélienne bornée

$$\mathbb{E}(f(X)\frac{\mathbb{1}_G}{\mathbb{P}(G)}) = \mathbb{E}(f(X))$$

Ce qui implique **L'indépendance**. □

13. Si $\mathcal{G} = \sigma(Y)$ alors : $\mathbb{E}(X\backslash\mathcal{G}) = \mathbb{E}(X\backslash Y)$.
    Si $\mathcal{G} = \mathcal{F}$ la tribu de l'espace alors $\mathbb{E}(X\backslash\mathcal{F}) = X$.

1.9 Distribution conditionnelle par rapport à une tribu

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité sur laquelle est définie une v.a X et soit $B \in \mathcal{B}(\mathbb{R})$, on veut définir $\mathbb{P}(X \in B\backslash\mathcal{G})$ telle que $\mathcal{F}$ est une sous-tribu de $\mathcal{F}$.

Définition 1.9.1.
1) La distribution conditionnelle par rapport à la tribu $\mathcal{G}$ est la mesure $\mathbb{P}_X(.\backslash\mathcal{G})$ telle que :

$$\begin{aligned}\mathbb{P}_X(.\backslash\mathcal{G}) : \mathcal{B}(\mathbb{R}) &\longrightarrow [0,1]\\ B &\longmapsto \mathbb{P}_X(B\backslash\mathcal{G}) = \mathbb{E}(\mathbb{1}_B\backslash\mathcal{G})\end{aligned}$$

2) La fonction de répartition conditionnelle $F_X(.\backslash\mathcal{G})$ de X par rapport à la tribu $\mathcal{G}$ est :

$$\begin{aligned}F_X(.\backslash\mathcal{G}) : \mathbb{R} &\longrightarrow [0,1]\\ x &\longmapsto F_X(x\backslash\mathcal{G}) = \mathbb{P}_X(X \leq x\backslash\mathcal{G}) = \mathbb{E}(\mathbb{1}_{\{X\leq x\}}\backslash\mathcal{G})\end{aligned}$$

3) Si $F_X(.\backslash\mathcal{G})$ est absolument continue par rapport à la mesure de Lebesgue, alors il existe une fonction $f_X(.\backslash\mathcal{G})$ dite densité de ***Randon-Nikodym*** *définie par :*

$$f_X(x\backslash\mathcal{G}) = \frac{dF_X(x\backslash\mathcal{G})}{dx}$$

4) Si X est discrète alors : on définit la fonction de masse conditionnelle par rapport à $\mathcal{G}$ par :

$$\begin{aligned} \mathbb{P}_X(.\backslash\mathcal{G}) : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto \mathbb{P}_X(x\backslash\mathcal{G}) = \mathbb{P}_X(X = x\backslash\mathcal{G}) = \mathbb{E}(\mathbb{1}_{\{X=x\}}\backslash\mathcal{G}) \end{aligned}$$

5) La variance conditionnelle de X par rapport à $\mathcal{G}$ est :

$$\begin{aligned} Var(X\backslash\mathcal{G}) &= \mathbb{E}([X - \mathbb{E}(X\backslash\mathcal{G})]^2\backslash\mathcal{G}) \\ &= \mathbb{E}(X^2\backslash\mathcal{G}) - \mathbb{E}(X\backslash\mathcal{G})^2 \end{aligned}$$

Théorème 1.9.1. ***(Théorème de Transfert)***
Soit φ une fonction mesurable réelle définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
$\varphi(x)$ est $\mathbb{P}$-intégrable ssi φ est $\mathbb{P}_X$-intégrable et $\int_\Omega \varphi(X(\omega))\, d\mathbb{P}(\omega) = \int_\mathbb{R} \varphi(X)\, d\mathbb{P}_X(x)$.
Espérance :

$$\begin{aligned} \mathbb{E}(X) &= \int_\Omega X\, d\mathbb{P} = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) \\ &= \int_\mathbb{R} x\, d\mathbb{P}_X(x) \end{aligned}$$