

Notes de Cours de Probabilités.

Thierry CHONAVEL

Télécom Bretagne

thierry.chonavel@enst-bretagne.fr

Juin 2000

Table des matières

1	Introduction.	4
2	Espace probabilisé.	4
3	Probabilité sur un ensemble fini ou dénombrable.	5
4	Probabilité sur un ensemble non dénombrable.	6
4.1	Exemple.	6
4.2	Mesures de probabilité absolument continues.	6
5	Probabilité conditionnelle.	7
6	Evènements indépendants.	7
7	Variable aléatoires réelles.	8
7.1	Définition.	8
7.2	Loi d'une variable aléatoire.	9
7.3	Fonction de répartition.	9
8	Intégration des variables aléatoires.	10
8.1	Principe.	10
8.2	Moments des variables aléatoires.	11
8.3	Fonction caractéristique.	11
9	Couples de variables aléatoires.	12
9.1	Loi conjointe.	13
9.2	Lois marginales.	14
9.3	Intégration des fonctions d'un couple de variables aléatoires.	14

9.4	Changement de variables.	15
9.5	Matrice de covariance.	15
10	Espérance conditionnelle et loi conditionnelle.	16
10.1	Théorème de projection.	16
10.2	Régression linéaire et affine.	16
10.3	Espérance conditionnelle.	17
10.4	Loi conditionnelle.	18
11	Variables aléatoires indépendantes.	19
12	Vecteurs gaussiens.	19
12.1	Loi d'un vecteur gaussien.	19
12.2	Espérance conditionnelle dans le cas gaussien.	20
13	Convergence des variables aléatoires.	21
13.1	Convergence en probabilité.	21
13.2	Convergence en moyenne d'ordre α - Loi faible des grands nombres.	21
13.3	Convergence presque sûre - Loi forte des grands nombres.	22
13.4	Convergence en loi - Théorème limite-centrale.	22
13.5	Relations entre les différents types de convergence.	23
14	Quelques lois de probabilités classiques.	23
14.1	Lois discrètes.	23
14.2	Lois à densités.	25

1 Introduction.

Dans de nombreuses expériences, le résultat n'est pas connu à l'avance. Une telle expérience est appelée expérience aléatoire. Ce caractère aléatoire peut être du à une connaissance plus ou moins incomplète des relations de cause à effet qui conduisent au résultat de l'expérience, ou de celle des valeurs de paramètres qui régissent ces relations, ou encore au fait que les relations qui permettraient de prévoir le résultat sont trop complexes pour qu'il puisse effectivement être calculé.

Dans de nombreuses expériences, on a cependant quelque idée à priori du caractère plus ou moins vraisemblable des différents résultats possibles. On peut alors convenir d'exprimer cette vraisemblance en affectant aux différents résultats des grandeurs numériques, censées représenter des mesures de leur vraisemblance. La manipulation de ces grandeurs numériques pourra alors fournir de nouvelles informations sur l'expérience. La théorie des probabilités vise à formaliser ces idées intuitives.

Il est à noter que le fait d'associer un modèle probabiliste à une expérience n'indique pas si le choix de ce modèle est pertinent, i.e. si les vraisemblances que l'on associe aux résultats sont réellement représentatives de cette expérience. Ce dernier point constitue plutôt un aspect de la théorie des statistiques, qui sera envisagée en deuxième année.

Ce polycopié reprend dans ses grandes lignes des notes de cours utilisées à l'école depuis de nombreuses années [10]. On complètera utilement le cours de probabilité par la lecture du polycopié de théorie de la mesure [11] distribué par ailleurs. La théorie des probabilités s'inscrit en effet naturellement dans le cadre de la théorie de la mesure. De plus, la théorie de la mesure est très utilisée dans l'ensemble des mathématiques de l'ingénieur.

2 Espace probabilisé.

On appelle **espace probabilisable** (ou espace mesurable) la donnée d'un couple (Ω, \mathcal{B}) , où Ω est un ensemble, et \mathcal{B} une **tribu** de parties de Ω , c'est à dire un ensemble de parties de Ω , contenant Ω , et stable pour les opérations de complémentation et de réunion dénombrable :

$$\Omega \in \mathcal{B}$$

$$A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B} \tag{1}$$

$$(A_n)_{n=1,+\infty} \in \mathcal{B} \Rightarrow \bigcup_{n=1,+\infty} A_n \in \mathcal{B}.$$

Lorsque Ω est fini ou dénombrable, on choisit souvent (mais pas nécessairement) comme tribu l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω . Lorsque $\Omega = \mathbb{R}$, on choisit généralement la tribu borélienne de \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$ qui est la plus petite tribu (au sens de l'inclusion) qui contient les intervalles de \mathbb{R} de la forme $[a, b]$ (ou de façon équivalente de la forme $[a, b[$, ou encore de la forme $] - \infty, a]$, ...).

La plus petite tribu qui contient un ensemble donné de parties de Ω est appelée la **tribu engendrée** par cet ensemble. Notons que l'intersection de plusieurs tribus est encore une

tribu. Par suite, la tribu engendrée par un ensemble est l'intersection des tribus qui le contiennent.

On appellera **évènement** associé à une expérience aléatoire une condition portant sur le résultat de cette expérience. En fait, on convient d'identifier cet évènement à l'ensemble des valeurs de Ω pour lesquelles la condition qui définit cet évènement est réalisée. Ainsi, un évènement E est identifié à l'ensemble

$$\{\omega \in \Omega; \text{ la condition } E \text{ est satisfaite pour le résultat } \omega\}.$$

La donnée d'une tribu peut donc être vue comme la spécification d'une classe d'évènements particulière. Quand les éléments de Ω définissent des singletons $\{\omega\}$ qui appartiennent à \mathcal{B} , $\{\omega\}$ est parfois appelé **évènement élémentaire**.

On verra que l'on ne peut pas toujours modéliser une expérience aléatoire en se donnant simplement une mesure de la vraisemblance de chacun des résultats possibles, mais que cela peut être fait en se donnant celle des évènements d'une tribu. La notion de mesure de probabilité permet de donner une forme mathématique à la notion intuitive de vraisemblance d'un évènement.

Un **espace probabilisé** est donné par un triplet (Ω, \mathcal{B}, P) , où (Ω, \mathcal{B}) est un espace probabilisable, et P une **mesure de probabilité** sur \mathcal{B} , i.e. une application définie sur \mathcal{B} et à valeurs dans $[0, 1]$, telle que

$$\begin{aligned} P(\Omega) &= 1, \\ \forall (A_n)_{n=1,+\infty}, \text{ avec } A_i \cap A_j &= \emptyset \text{ pour } i \neq j, P(\cup_{n=1,+\infty} A_n) = \sum_{n=1,+\infty} P(A_n). \end{aligned} \tag{2}$$

Conséquences. On vérifie que $\forall A, B \in \mathcal{B}$

$$\begin{aligned} A \subset B &\Rightarrow P(A) \leq P(B), \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B), \end{aligned} \tag{3}$$

et plus généralement, pour une famille $(A_n)_{n=1,N}$ quelconque, la formule de Poincaré indique que

$$\begin{aligned} P(\cup_{n=1,N} A_n) &= \sum_{n=1,N} P(A_n) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \\ &\quad \dots + (-1)^{N-1} P(\cap_{n=1,N} A_n). \end{aligned} \tag{4}$$

Notons enfin qu'une mesure de probabilité est entièrement caractérisée par les valeurs qu'elle prend sur une famille génératrice de la tribu sur laquelle elle est définie.

3 Probabilité sur un ensemble fini ou dénombrable.

Pour fixer les idées, supposons que $\Omega = \{1, 2, \dots, N\}$, ou $\Omega = \mathbb{N}$. Dans ce cas, on prend souvent $\mathcal{B} = \mathcal{P}(\Omega)$. La donnée d'une suite de nombres positifs $(p_k)_{k \in \Omega}$, avec $\sum_k p_k = 1$

définit entièrement une probabilité sur (Ω, \mathcal{B}) . En effet, il est alors clair que

$$\forall A \in \mathcal{B}, \quad P(A) = \sum_{k \in A} p_k. \quad (5)$$

Dans le cas particulier où les éléments de $\Omega = \{1, 2, \dots, N\}$ sont des événements élémentaires équiprobables, i.e. $\forall k \ p_k = 1/\text{card}(\Omega)$,

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}. \quad (6)$$

Dans ce cas, il apparaît clairement que le calcul de la probabilité d'un événement se ramène au problème du dénombrement des éléments contenu dans cet événement. A ce sujet rappelons les résultats utiles suivants :

- nombre de permutations de n éléments : $n! = n(n-1) \dots 2.1$
- nombre de p -uplets ordonnés parmi n éléments : $A_n^p = n!/(n-p)!$
- nombre de façons de choisir p éléments parmi n : $C_n^p = A_n^p/p!$
- formule de Stirling $n! = n^n e^{-n} \sqrt{2\pi n} (1 + \varepsilon(n))$, avec $\lim_{n \rightarrow \infty} \varepsilon(n) = 0$.

4 Probabilité sur un ensemble non dénombrable.

4.1 Exemple.

Commençons par un exemple. Prenons $\Omega = [0, 1]$, et supposons que l'on réalise une expérience aléatoire telle que toutes les valeurs de Ω aient “la même chance” de se produire. Il serait alors tentant d'affecter à chaque valeur de Ω une probabilité identique, notée p , avec $0 < p < 1$. Mais cela n'est pas possible car alors, on aurait nécessairement $P(\Omega) = +\infty$!

Par contre, on pourra caractériser une probabilité sur $[0, 1]$ pour l'expérience précédente en définissant une probabilité sur la tribu borélienne de $[0, 1]$ (i.e. la tribu engendrée par les intervalles fermés de $[0, 1]$) de la façon suivante :

$$\forall a, b \in [0, 1], \text{ avec } a \leq b, \quad P([a, b]) = \int_{[a, b]} dx = b - a.$$

Cette définition est cohérente avec l'idée d'équiprobabilité des éléments de $[0, 1]$. Cet exemple illustre l'intérêt de la notion de tribu introduite précédemment.

4.2 Mesures de probabilité absolument continues.

Dans ce cours, on va se restreindre à l'étude de deux classes essentielles de mesures de probabilités : les probabilités discrètes, définies sur des ensembles finis ou dénombrables et que l'on a déjà présentées, et les probabilités définies sur \mathbb{R} (ou plus généralement \mathbb{R}^n)

qui sont **absolument continues**. Les mesures de probabilité absolument continues sont celles pour lesquelles il existe une fonction f , appelée densité de probabilité de P , telle que

$$\forall A \in \mathcal{B}(\mathbb{R}), P(A) = \int_A f(x)dx. \quad (7)$$

5 Probabilité conditionnelle.

Soient A et B deux évènements de Ω . Lorsque $P(A) \neq 0$, on définit la **probabilité conditionnelle** de B sachant A , notée $P(B/A)$, ou $P^A(B)$, par

$$P(B/A) = \frac{P(A \cap B)}{P(A)}. \quad (8)$$

On montre alors que l'application

$$\begin{aligned} P^A : \mathcal{B} &\rightarrow [0, 1] \\ B &\rightarrow P(B/A), \end{aligned} \quad (9)$$

définit une probabilité sur (Ω, \mathcal{B}) .

Les résultats simples suivants sont d'un usage très commun :

- Formule des conditionnements successifs : si $P(A_1 \cap \dots \cap A_n) \neq 0$,

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) = \\ P(A_n/A_1 \cap \dots \cap A_{n-1})P(A_{n-1}/A_1 \cap \dots \cap A_{n-2}) \dots P(A_2/A_1)P(A_1) \end{aligned} \quad (10)$$

- Formule de la probabilité totale : si $(A_k)_{k=1,n}$, est une partition de Ω , avec $P(A_k) > 0$ pour $k = 1, n$,

$$P(A) = \sum_{k=1,n} P(A/A_k)P(A_k) \quad (11)$$

- **Formule de Bayes** : si $(A_k)_{k=1,n}$, est une partition de Ω , avec $P(A_k) > 0$ pour $k = 1, n$

$$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B/A_k)P(A_k)}{\sum_{i=1,n} P(B/A_i)P(A_i)}. \quad (12)$$

6 Evènements indépendants.

Intuitivement, deux évènements A et B sont indépendants si le fait que A soit réalisé ne modifie pas la probabilité de B . Ainsi, si $P(A) \neq 0$, on pourra caractériser l'indépendance par le fait que $P(B/A) = P(B)$. Plus généralement, on dira que deux évènements A et B sont indépendants si

$$P(A \cap B) = P(A)P(B). \quad (13)$$

On dira que les événements A_1, \dots, A_n sont indépendants si

$$\forall I \subset \{1, \dots, n\}, P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i). \quad (14)$$

Notons que cette notion est plus forte que celle de l'indépendance deux à deux des événements A_1, \dots, A_n .

Si les événements A_1, \dots, A_n sont indépendants, les événements B_1, \dots, B_n , avec $B_k = A_k$ ou $B_k = A_k^c$, sont indépendants.

On dira que deux tribus \mathcal{A} et \mathcal{B} sont indépendantes si

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{B}, P(A \cap B) = P(A)P(B). \quad (15)$$

7 Variable aléatoires réelles.

7.1 Définition.

L'ensemble Ω des résultats d'une expérience aléatoire n'est pas nécessairement constitué de grandeurs numériques, ce qui ne facilite pas leur manipulation. Par ailleurs la tribu \mathcal{B} que l'on définit sur Ω peut être très différente d'une expérience à une autre. Enfin, pour une expérience donnée, on peut s'intéresser non pas spécifiquement au résultat lui-même mais à une de ses grandeurs caractéristiques : ainsi, si l'expérience consiste à choisir de façon aléatoire un individu dans une certaine population, on pourra par exemple envisager plus particulièrement l'âge de cette personne, sa taille, L'introduction de la notion de variable aléatoire permet de prendre en compte ces diverses situations de façon satisfaisante dans le cadre d'un formalisme unique.

Rappelons tout d'abord que si $(\Omega_1, \mathcal{B}_1)$ et $(\Omega_2, \mathcal{B}_2)$ sont deux espaces probabilisables, on dit qu'une application X de Ω_1 dans Ω_2 est **mesurable** si

$$\forall A \in \mathcal{B}_2, X^{-1}(A) \in \mathcal{B}_1. \quad (16)$$

Notons que si G_2 est une famille génératrice de \mathcal{B}_2 , il suffit alors pour cela de vérifier que

$$\forall A \in G_2, X^{-1}(A) \in \mathcal{B}_1. \quad (17)$$

Une **variable aléatoire** réelle est une application mesurable X d'un espace probabilisé (Ω, \mathcal{B}, P) , dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il suffit donc par exemple de vérifier que

$$\forall x \in \mathbb{R}, X^{-1}([-\infty, x]) \in \mathcal{B} \quad (18)$$

pour que X soit une variable aléatoire. On se limitera ici aux variables aléatoires à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $\mathcal{B}(\mathbb{R}^n)$ désignant la tribu borélienne de \mathbb{R}^n , c'est à dire celle engendrée par les pavés de \mathbb{R}^n . Mais on peut définir des variables aléatoires à valeurs dans des espaces plus généraux.

Pour une variable aléatoire réelle X , l'ensemble

$$\mathcal{B}^X = \{X^{-1}(B); B \in \mathcal{B}(\mathbb{R})\} \quad (19)$$

est une tribu de Ω appelée **tribu engendrée par X** .

7.2 Loi d'une variable aléatoire.

Considérons l'application suivante

$$\begin{aligned} P_X : \mathcal{B}(\mathbb{R}) &\rightarrow [0, 1] \\ A &\rightarrow P_X(A) = P(X^{-1}(A)). \end{aligned} \quad (20)$$

Il est aisé de vérifier que P_X représente bien une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il apparaît donc que l'on a transposé la probabilité P que l'on avait initialement définie sur (Ω, \mathcal{B}) , en une probabilité P_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. P_X définit la **loi de probabilité de X** .

On utilise classiquement des notations du type $P(X \in A)$ pour désigner la probabilité $P(X^{-1}(A)) = P_X(A)$.

7.3 Fonction de répartition.

La loi P_X d'une variable aléatoire X est entièrement caractérisée par la connaissance des grandeurs $P_X([-\infty, x])$, $\forall x \in \mathbb{R}$. Il apparaît donc que la connaissance de P_X est équivalente à celle de la fonction F_X définie par

$$F_X(x) = P_X([-\infty, x]). \quad (21)$$

La fonction F_X est appelée **fonction de répartition** de la variable aléatoire réelle X . Indiquons quelques propriétés immédiates de la fonction de répartition :

- F_X est croissante
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$
- F_X est continue à droite.

On montre réciproquement que toute fonction qui satisfait à ces propriétés est la fonction de répartition d'une certaine variable aléatoire réelle.

Dans le cas de variables aléatoires discrètes, i.e. telles que $X(\Omega)$ est un ensemble fini ou dénombrable,

$$F_X(x) = \sum_{s \in X(\Omega), s \leq x} P_X(\{s\}). \quad (22)$$

Dans ce cas, la fonction de répartition est donc une fonction en escalier discontinue pour les valeurs de $X(\Omega)$, et la valeur du saut de F_X en $s \in X(\Omega)$ est donnée par

$$F_X(s) - \lim_{x \rightarrow s^-} F_X(x) = P_X(\{s\}). \quad (23)$$

Dans le cas de variables aléatoires absolument continues, i.e. telles que P_X soit absolument continues, P_X est caractérisée par une densité de probabilité f_X telle que

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad P_X(A) = \int_A f_X(x) dx. \quad (24)$$

Donc, en particulier,

$$F_X(x) = \int_{]-\infty, x]} f_X(x) dx. \quad (25)$$

La fonction de répartition est donc ici une fonction dérivable, de fonction dérivée f_X .

Notons au passage que f est la densité de probabilité d'une certaine mesure de probabilité si et seulement si

- f est positive,
- $\int_{\mathbb{R}} f(x)dx = 1$.

Remarque. La fonction de répartition est parfois définie par $F_X(x) = P_X([-\infty, x])$. Cette définition modifie légèrement les énoncés précédents. Ainsi, on aura la continuité à gauche au lieu de la continuité à droite, ...

8 Intégration des variables aléatoires.

8.1 Principe.

Pour une mesure de probabilité P sur un espace probabilisable (Ω, \mathcal{B}) , et toute fonction étagée $\sum_{i=1,n} a_i \mathbb{I}_{A_i}$, avec $A_i \in \mathcal{B}$ ($i = 1, n$), on définit la grandeur

$$\mathbb{E}\left[\sum_{i=1,n} a_i \mathbb{I}_{A_i}\right] = \sum_{i=1,n} a_i P(A_i), \quad (26)$$

où

$$\begin{aligned} \mathbb{I}_{A_i} : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow \begin{cases} 1 & \text{si } \omega \in A_i \\ 0 & \text{si } \omega \notin A_i. \end{cases} \end{aligned} \quad (27)$$

On convient de noter cette grandeur

$$\mathbb{E}\left[\sum_{i=1,n} a_i \mathbb{I}_{A_i}\right] = \int_{\Omega} \sum_{i=1,n} a_i \mathbb{I}_{A_i}(\omega) dP(\omega). \quad (28)$$

Par ailleurs, on montre que toute fonction mesurable de (Ω, \mathcal{B}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ peut s'écrire comme limite d'une suite de fonctions étagées, et qu'il est ainsi possible d'étendre la définition précédente à l'ensemble des fonctions mesurables par passage à la limite. Pour une variable aléatoire X quelconque sur (Ω, \mathcal{B}, P) , on note cette limite

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) dP(\omega). \quad (29)$$

Cette expression est appelée **espérance mathématique**, ou **moyenne**, de X . La formule suivante, appelée **théorème de transfert**, permet d'effectuer simplement le calcul de $\mathbb{E}[X]$, ou plus généralement d'expressions de la forme $\mathbb{E}[g(X)]$, où g est une fonction mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ($g(X)$ est donc une variable aléatoire) :

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) dP_X(x). \quad (30)$$

En particulier, pour une variable aléatoire discrète,

$$\mathbb{E}[g(X)] = \sum_{s \in X(\Omega)} g(s) P_X(s). \quad (31)$$

Pour une variable aléatoire absolument continue de densité f_X ,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx. \quad (32)$$

Remarques.

1) En fait, compte tenu du mode de construction de l'espérance mathématique, les expressions précédentes n'ont de sens que si on a respectivement

$$\sum_{s \in X(\Omega)} |g(s)| P_X(s) < \infty, \text{ et } \int_{\mathbb{R}} |g(x)| f_X(x) dx < \infty. \quad (33)$$

2) Pour un évènement A , il est clair que

$$P_X(A) = \mathbb{E}[\mathbb{I}_A(X)]. \quad (34)$$

8.2 Moments des variables aléatoires.

On définit le moment (resp. le moment centré) d'ordre m d'une variable aléatoire X par

$$\mathbb{E}[X^m] \text{ (resp. } \mathbb{E}[(X - \mathbb{E}[X])^m]). \quad (35)$$

Le moment centré d'ordre 2

$$\sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \quad (36)$$

est appelé **variance** de X , et σ_X est l'**écart type** de X . Les inégalités suivantes font intervenir les moments et fournissent des majorants de la probabilité de dispersion d'une variable aléatoire au delà d'un certain domaine :

— **inégalité de Markov**

$$\forall a > 0, \forall m \geq 0, \quad P(|X| \geq a) \leq \frac{\mathbb{E}[|X|^m]}{a^m}, \quad (37)$$

— **inégalité de Bienaymé-Tchebychev** (cas particulier de l'inégalité de Markov)

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\sigma_X^2}{a^2}. \quad (38)$$

8.3 Fonction caractéristique.

On a vu en traitement du signal l'intérêt de la transformée de Fourier. De même que l'on peut caractériser la loi d'une variable aléatoire par sa fonction de répartition, on pourra la décrire au moyen de sa transformée de Fourier. La fonction

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX(\omega)} dP(\omega) = \int_{\mathbb{R}} e^{itx} dP_X(x) \quad (39)$$

est appelée **fonction caractéristique** de X .

Pour une variable aléatoire discrète

$$\varphi_X(t) = \sum_{s \in X(\Omega)} e^{its} P_X(s), \quad (40)$$

et pour une variable aléatoire absolument continue,

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx. \quad (41)$$

Les fonctions caractéristiques satisfont aux propriétés suivantes

- $\varphi_X(0) = 1$, $|\varphi_X(t)| \leq 1$
- $\varphi_X(t)$ est continue (d'après le **théorème de convergence dominée de Lebesgue**)
- $\varphi_X(t) = \varphi_X^*(-t)$
- $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$
- $\varphi_{X_1} = \varphi_{X_2}$, si et seulement si X_1 et X_2 ont la même loi de probabilité.
- si P_X est absolument continue et $\int_{\mathbb{R}} |\varphi_X(t)| dt < \infty$,

$$f_X(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t) dt, \quad (42)$$

- $\varphi_X(t)$ est une fonction de type positif :

$$\forall n \in \mathbb{N}^*, \forall (z_k)_{k=1,n} \in \mathbb{C}^n, \forall (t_k)_{k=1,n} \in \mathbb{R}^n, \sum_{k=1,n} z_k z_l^* \varphi_X(t_k - t_l) \geq 0. \quad (43)$$

Réciproquement, toute fonction de type positif égale à 1 en 0 est une fonction caractéristique (théorème de Bochner).

- si $\mathbb{E}[X^n]$ existe, alors $\mathbb{E}[X^k]$ aussi pour $k = 0, n$, et

$$\varphi_X(t) = \sum_{k=0,n} \frac{(it)^k}{k!} \mathbb{E}[X^k] + t^n \varepsilon(t), \quad (44)$$

avec $\lim_{t \rightarrow 0} \varepsilon(t) = 0$. De plus, $\mathbb{E}[X^n] = (-i)^n \varphi_X^{(n)}(0)$ ($\varphi_X^{(n)}$ est la fonction dérivée d'ordre n de $\varphi_X^{(n)}$).

Fonction génératrice des moments. Pour les variables aléatoires pour lesquelles $X(\Omega) \subset \mathbb{N}$, on utilise souvent la fonction génératrice des moments de préférence à la fonction caractéristique. Celle ci est définie par

$$g_X(z) = \mathbb{E}[z^X] = \sum_{n \in X(\Omega)} z^n P_X(\{n\}), \text{ pour } z \in \mathbb{C}, \text{ et } |z| \leq 1. \quad (45)$$

On a alors $g'_X(1) = \mathbb{E}[X]$, $g''_X(1) = \mathbb{E}[X(X-1)]$, ...

9 Couples de variables aléatoires.

Les résultats que l'on va voir maintenant se généralisent immédiatement au cas de vecteurs aléatoires de taille quelconque.

Soient deux espaces probabilisables (E_1, \mathcal{B}_1) , et (E_2, \mathcal{B}_2) . On peut définir sur $E_1 \times E_2$ la tribu engendrée par les éléments $B_1 \times B_2 \in \mathcal{B}_1 \times \mathcal{B}_2$. Cette tribu sera notée $\mathcal{B}_1 \otimes \mathcal{B}_2$.

En particulier, on définit sur \mathbb{R}^2 la tribu borélienne $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, qui est la tribu engendrée par les ensembles $] -\infty, x] \times] -\infty, y]$. Si X et Y sont deux variables aléatoires réelles définies sur un même espace probabilisé (Ω, \mathcal{B}, P) , le couple (X, Y) définit une variable aléatoire sur (Ω, \mathcal{B}, P) , à valeurs dans \mathbb{R}^2 muni de la tribu $\mathcal{B}(\mathbb{R}^2)$. En effet, pour la famille génératrice de $\mathcal{B}(\mathbb{R}^2)$ constituée des ensembles de la forme $C = A \times B$, avec $A, B \in \mathcal{B}(\mathbb{R})$,

$$(X, Y)^{-1}(C) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{B}. \quad (46)$$

Plus généralement, on aura alors

$$\forall C \in \mathcal{B}(\mathbb{R}^2), \quad (X, Y)^{-1}(C) = \{\omega \in \Omega; (X(\omega), Y(\omega)) \in C\} \in \mathcal{B}. \quad (47)$$

9.1 Loi conjointe.

La loi conjointe de (X, Y) est donnée par la probabilité $P_{X,Y}$ définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}^2))$ par

$$\forall C \in \mathcal{B}(\mathbb{R}^2), \quad P_{X,Y}(C) = P((X, Y)^{-1}(C)). \quad (48)$$

$P((X, Y)^{-1}(C))$ est encore notée $P((X, Y) \in C)$. Puisque $\mathcal{B}(\mathbb{R}^2)$ est engendrée par les événements $B_1 \times B_2$, où $B_1, B_2 \in \mathcal{B}(\mathbb{R})$, la connaissance des probabilités $P_{X,Y}(B_1 \times B_2) = P(\{X \in B_1\} \cap \{Y \in B_2\})$ caractérise entièrement $P_{X,Y}$. Pour les mêmes raisons, $P_{X,Y}$ est caractérisée par la connaissance de la fonction de répartition de (X, Y) :

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \quad (x, y \in \mathbb{R}). \quad (49)$$

Dans le cas de variables aléatoires discrètes, i.e. si $X(\Omega) \times Y(\Omega)$ est fini ou dénombrable,

$$P((X, Y) \in C) = \sum_{(m,n) \in C \cap (X(\Omega) \times Y(\Omega))} P(X = m, Y = n). \quad (50)$$

On dit que $P_{X,Y}$ est absolument continue s'il existe une fonction $f_{X,Y}$ telle que $\forall x, y \in \mathbb{R}$

$$F_{X,Y}(x, y) = \int_{]-\infty, x] \times]-\infty, y]} f_{X,Y}(s, t) ds dt. \quad (51)$$

Alors, il est clair que

$$P_{X,Y}(C) = \int_C f_{X,Y}(x, y) dx dy. \quad (52)$$

Notons que dans ce cas f_{XY} s'obtient simplement par dérivation de F_{XY} :

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y). \quad (53)$$

9.2 Lois marginales.

Les lois de X et de Y peuvent être obtenues à partir de celle du couple (X, Y) (la réciproque est fausse!). P_X et P_Y sont alors appelées lois marginales du couple (X, Y) , et sont données par

$$P_X(B_1) = P(X \in B_1, Y \in \mathbb{R}), \text{ et } P_Y(B_2) = P(X \in \mathbb{R}, Y \in B_2). \quad (54)$$

De façon analogue,

$$\begin{aligned} F_X(x) &= \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) \\ \text{et } F_Y(y) &= \lim_{x \rightarrow +\infty} F_{X,Y}(x, y). \end{aligned} \quad (55)$$

En particulier, dans le cas discret,

$$P(X = n) = \sum_{m \in Y(\Omega)} P(X = n, Y = m), \quad (56)$$

et dans le cas absolument continu,

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy. \quad (57)$$

9.3 Intégration des fonctions d'un couple de variables aléatoires.

Le théorème de transfert s'écrit ici sous la forme

$$\mathbb{E}[g(X, Y)] = \int_{\mathbb{R}^2} g(x, y) dP_{X,Y}(x, y), \quad (58)$$

pour toute fonction g mesurable de $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Dans le cas discret, cette relation devient

$$\mathbb{E}[g(X, Y)] = \sum_{(m,n) \in X(\Omega) \times Y(\Omega)} g(m, n) P(X = m, Y = n), \quad (59)$$

et dans le cas absolument continu,

$$\mathbb{E}[g(X, Y)] = \int_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy. \quad (60)$$

En particulier, la fonction caractéristique du couple (X, Y) est définie par

$$\varphi_{X,Y}(u, v) = \mathbb{E}[e^{i(uX+vY)}]. \quad (61)$$

Dans les cas discret et absolument continu, on obtient alors respectivement

$$\begin{aligned} \varphi_{X,Y}(u, v) &= \sum_{(m,n) \in X(\Omega) \times Y(\Omega)} e^{i(um+vn)} P_{X,Y}(\{(m, n)\}) \\ \text{et } \varphi_{X,Y}(u, v) &= \int_{\mathbb{R}^2} e^{i(ux+vy)} f_{X,Y}(x, y) dx dy. \end{aligned} \quad (62)$$

9.4 Changement de variables.

Considérons une transformation ϕ de \mathbb{R}^2 dans \mathbb{R}^2 , et notons $(U, V) = \phi(X, Y)$. On fait les hypothèses suivantes :

- $P_{X,Y}$ est absolument continue, et $f_{X,Y} > 0$ sur un ouvert O de \mathbb{R}^2 , et nulle à l'extérieur de O .
- ϕ est bijective de O sur un ouvert O' , et $\forall (x, y) \in O$, le jacobien de ϕ , $J_\phi(x, y) = |\phi'(x, y)|$, est non nul. rappelons que

$$J_\phi(x, y) = \begin{vmatrix} \frac{\partial u(x,y)}{\partial x} & \frac{\partial u(x,y)}{\partial y} \\ \frac{\partial v(x,y)}{\partial x} & \frac{\partial v(x,y)}{\partial y} \end{vmatrix}. \quad (63)$$

$P_{U,V}$ est alors absolument continue, de densité

$$f_{U,V}(u, v) = \frac{1}{|J_\phi(x, y)|} f_{X,Y}(\phi^{-1}(u, v)) \mathbb{1}_{O'}(u, v). \quad (64)$$

Remarque. De façon plus générale, on obtient la loi de (U, V) en passant par les fonctions de répartition. Ainsi,

$$F_{U,V}(u, v) = P(U \leq u, V \leq v) = P_{X,Y}(\phi^{-1}([-\infty, u] \times [-\infty, v])). \quad (65)$$

Bien sûr, cette remarque s'applique aussi au cas de transformations du type $U = \phi(X)$: alors $F_U(u) = P_X(\phi^{-1}([-\infty, u]))$.

9.5 Matrice de covariance.

On définit la **covariance** des variables aléatoires de X et de Y par

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (66)$$

La covariance est une forme bilinéaire symétrique. $\text{cov}(X, X) = \sigma_X^2$ est la **variance** de X . Le **coefficient de corrélation** de X et de Y est donné par

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (67)$$

et est compris entre -1 et +1 car l'**inégalité de Cauchy-Schwarz** indique que

$$|\mathbb{E}[XY]| \leq \mathbb{E}[X^2]^{1/2} \mathbb{E}[Y^2]^{1/2}. \quad (68)$$

Si X et Y sont des vecteurs aléatoires $X = [X_1, \dots, X_n]^T$, et $Y = [Y_1, \dots, Y_m]^T$, la matrice de covariance de X et de Y est la matrice R_{XY} de taille $n \times m$ et de terme général $[R_{XY}]_{ij} = \text{cov}(X_i, Y_j)$, ie

$$R_{XY} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T]. \quad (69)$$

Lorsque $R_{XY} = 0$, on dit que les vecteurs X et Y sont **décorrélés**. Dans le cas où X et Y sont des variables aléatoires scalaires, la décorrélation de X et de Y se traduit donc par

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]. \quad (70)$$

Si $X = Y$, on note simplement R_X la matrice de covariance du vecteur X .

R_X est une matrice positive, i.e. pour tout vecteur $u \in \mathbb{R}^n$, $u^T R_X u \geq 0$. Notons de plus que si A et b sont une matrice et un vecteur quelconques (de tailles compatibles),

$$R_{AX+b} = AR_X A^T. \quad (71)$$

10 Espérance conditionnelle et loi conditionnelle.

On cherche ici à approcher une variable aléatoire Y à partir de la connaissance d'une variable aléatoire X .

10.1 Théorème de projection.

Notons $L^2(\Omega, \mathcal{B}, P)$ l'espace des variables aléatoires réelles définies sur (Ω, \mathcal{B}, P) pour lesquelles $\mathbb{E}[X^2] < \infty$, muni du produit scalaire défini par $\langle X, Y \rangle = \mathbb{E}[XY]$. On peut montrer que $L^2(\Omega, \mathcal{B}, P)$ est un espace de Hilbert, i.e. un espace vectoriel normé complet muni d'un produit scalaire. Dans un espace de Hilbert H , on dispose d'un résultat très intéressant, le **théorème de projection**, qui indique que pour tout sous espace vectoriel fermé F de H ,

$$\forall Y \in H, \exists ! Y^* \in F, \quad \|Y - Y^*\|_H = \min_{Z \in F} \|Y - Z\|_H, \quad (72)$$

où $\|\cdot\|_H$ représente la norme définie sur H . Y^* est la projection orthogonale de Y sur F . Une caractérisation possible de Y^* est donnée par le résultat suivant : Y^* est le seul élément de F tel que

$$\forall Z \in F, \langle Y - Y^*, Z \rangle_H = 0, \quad (73)$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire défini sur H .

10.2 Régression linéaire et affine.

Plaçons nous dans $L^2(\Omega, \mathcal{B}, P)$. Prenons pour F l'espace des combinaisons linéaires des variables aléatoires X_1, \dots, X_n , et cherchons à approximer une variable aléatoire Y de $L^2(\Omega, \mathcal{B}, P)$ comme un élément de F . L'approximation de Y qui consiste à prendre la projection de Y sur F est appelée l'approximation de Y “**au sens des moindres carrés**”. D'après la formule de caractérisation (73), la solution $Y^* = \sum_{k=1,n} a_k X_k$ de ce problème doit vérifier

$$\mathbb{E}[(Y - a^T X)X^T] = 0, \quad (74)$$

où $a = [a_1, \dots, a_n]^T$, et $X = [X_1, \dots, X_n]^T$. Donc $\mathbb{E}[XX^T]a = \mathbb{E}[XY]$.

Si les variables aléatoires $(X_k)_{k=1,n}$ sont linéairement indépendantes, i.e. si

$$\mathbb{E}\left[\left(\sum_{k=1,n} c_k X_k\right)^2\right] = 0 \Rightarrow c_1 = \dots = c_n = 0, \quad (75)$$

il est alors clair que la matrice $\mathbb{E}[XX^T]$ est inversible, auquel cas,

$$a = (\mathbb{E}[XX^T])^{-1}\mathbb{E}[XY]. \quad (76)$$

Pour ce qui concerne la régression affine, on cherche à approximer Y comme une combinaison linéaire des variables aléatoires $X_1, \dots, X_n, 1$, où 1 est la variable aléatoire qui prend toujours la valeur 1. On montre que dans ce cas l'approximation de Y "au sens des moindres carrés" est donnée par

$$Y^* = \mathbb{E}[Y] + R_{YX}R_X^{-1}(X - \mathbb{E}(X)). \quad (77)$$

Dans le cas où $X = X_1$, la droite d'équation

$$y = \mathbb{E}[Y] + \frac{\text{cov}(X, Y)}{\sigma_X^2}(x - \mathbb{E}[X]) \quad (78)$$

est appelée **droite de régression** de Y par rapport à X . En pratique, l'observation d'une réalisation x de la seule variable X permet alors d'estimer par y la réalisation correspondante de la variable Y .

10.3 Espérance conditionnelle.

On cherche maintenant à raffiner l'approximation de Y à partir de la connaissance de X , en choisissant une approximation parmi les variables aléatoires de la forme $h(X)$, où h est une fonction mesurable (pour que $h(X)$ soit une variable aléatoire) telle que $\int h(x)^2 dP_X(x) < \infty$ (pour avoir $h(X) \in L^2(\Omega, \mathcal{B}, P)$). On note $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ l'ensemble de ces fonctions h . On montre facilement que l'ensemble des variables aléatoires $h(X)$ correspondantes est $L^2(\Omega, \mathcal{B}^X, P)$, où \mathcal{B}^X est la tribu engendrée par $X : \mathcal{B}^X = \{X^{-1}(B); B \in \mathcal{B}(\mathbb{R})\}$. Notons que $L^2(\Omega, \mathcal{B}^X, P)$ est un sous espace de Hilbert de $L^2(\Omega, \mathcal{B}, P)$.

La variable aléatoire $Y^* = h^*(X)$ recherchée est donc la projection de Y sur le sous espace $L^2(\Omega, \mathcal{B}^X, P)$ de $L^2(\Omega, \mathcal{B}, P)$. On notera $Y^* = \mathbb{E}[Y/X]$ et on l'appellera espérance conditionnelle de Y sachant X . On peut caractériser la variable aléatoire $h^*(X) = \mathbb{E}[Y/X]$ en utilisant la relation (73) :

$$\forall g \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X), \mathbb{E}[(Y - h^*(X))g(X)] = 0. \quad (79)$$

En particulier, on a la relation souvent utile

$$\mathbb{E}[\mathbb{E}[Y/X]] = \mathbb{E}[Y]. \quad (80)$$

Dans le cas de variables aléatoires absolument continues, la relation (79) devient

$$\begin{aligned}\mathbb{E}[(Y - h^*(X))g(X)] &= \int_{\mathbb{R}^2} (y - h^*(x))g(x)f_{XY}(x,y)dx dy \\ &= \int_x [\int_y y \frac{f_{XY}(x,y)}{f_X(x)} dy - h^*(x)]g(x)f_X(x)dx \\ &= 0.\end{aligned}\tag{81}$$

Donc, puisque la relation est vraie $\forall g \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$,

$$h^*(x) = \int_y y \frac{f_{XY}(x,y)}{f_X(x)} dy, \tag{82}$$

sur tout ouvert où $f_X(x)$ ne s'annule pas. Notons que lorsque $f_X(x) = 0$, x ne peut pas être une réalisation de X , et $h^*(x)$ n'a donc pas besoin d'être défini en un tel point.

La courbe $y = h^*(x)$ s'appelle **courbe de régression** de Y par rapport à X .

10.4 Loi conditionnelle.

$h^*(x)$ peut être vue comme l'espérance de Y conditionnellement à l'évènement $\{X = x\}$. On note alors $h^*(x) = \mathbb{E}[Y/X = x]$. D'après (79), $h^*(x)$ est aussi l'espérance de la variable aléatoire dont la loi a pour densité

$$f_{Y/X=x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}, \tag{83}$$

que l'on note aussi $f_Y^{X=x}(y)$. Cette loi est appelée loi conditionnelle de Y sachant X .

Notons que dans le cas de densités de probabilités continues, cette définition de la loi conditionnelle est cohérente, avec le résultat que l'on obtient en calculant les grandeurs $P(Y \in B/X \in [x, x+h])$, et en faisant tendre h vers 0 :

$$\lim_{h \rightarrow 0} P(Y \in B/X \in [x, x+h]) = \int_{y \in B} f_{Y/X=x}(y) dy. \tag{84}$$

Dans le cas discret, le calcul est plus simple dans la mesure où lorsque $P(X = n) \neq 0$, on a d'après la définition de la probabilité conditionnelle d'un évènement sachant un autre

$$P(Y \in B/X = n) = \frac{P(\{Y \in B\} \cap \{X = n\})}{P(X = n)}. \tag{85}$$

On note souvent la loi conditionnelle de Y sachant $X = x$ sous la forme $P_{Y/X=x}$, ou encore $P_Y^{X=x}$.

11 Variables aléatoires indépendantes.

Deux variables aléatoires réelles X et Y définies sur (Ω, \mathcal{B}, P) sont indépendantes si

$$\forall A, B \in \mathcal{B}, P(X \in A, Y \in B) = P(X \in A) \times P(Y \in B), \quad (86)$$

où $P(X \in A, Y \in B)$ constitue une notation usuelle pour $P(\{X \in A\} \cap \{Y \in B\})$. cette relation s'écrit encore $P_{XY}(A \times B) = P_X(A)P_Y(B)$. Notons que la définition de l'indépendance de X et de Y équivaut à celle des tribus \mathcal{B}^X et \mathcal{B}^Y .

L'indépendance de X et de Y s'exprime de façon équivalente par

$$\forall x, y \in \mathbb{R}, F_{XY}(x, y) = F_X(x)F_Y(y). \quad (87)$$

Dans le cas discret, on aura donc $P(X = m, Y = n) = P(X = m)P(Y = n)$, et dans le cas absolument continu, $f_{XY}(x, y) = f_X(x)f_Y(y)$.

On peut caractériser l'indépendance de deux variables aléatoires X et Y à partir des tribus \mathcal{B}^X et \mathcal{B}^Y engendrées par X et Y respectivement. Pour cela, on définit l'**indépendance de deux tribus** : les tribus \mathcal{B}_1 et \mathcal{B}_2 sont dites indépendantes si

$$\forall B_1 \in \mathcal{B}_1, \forall B_2 \in \mathcal{B}_2, P(B_1 \cap B_2) = P(B_1)P(B_2). \quad (88)$$

L'indépendance de X et Y est équivalente à celle des tribus \mathcal{B}^X et \mathcal{B}^Y .

Autres caractérisations : les variables aléatoires X et Y sont indépendantes si et seulement si l'un des critères équivalents suivant est vérifié.

- $\forall g, h$ mesurables, $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$.
- $\forall u, v \in \mathbb{R}, \varphi_{XY}(u, v) = \varphi_X(u)\varphi_Y(v)$.
- $\forall x \in \mathbb{R}, P_{Y/X=x} = P_Y$

Il est clair que deux variables aléatoires indépendantes sont décorrélées. La réciproque est fausse en général.

Pour deux variables aléatoires indépendantes X et Y $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$, donc $f_{X+Y}(s) = f_X(s) * f_Y(s)$ ($*$: produit de convolution).

12 Vecteurs gaussiens.

12.1 Loi d'un vecteur gaussien.

Les vecteurs gaussiens sont des vecteurs aléatoires $X = [X_1, \dots, X_n]^T$ dont la loi possède une fonction caractéristique de la forme

$$\varphi_X(u) = \mathbb{E}[e^{i \sum_{k=1,p} u_k X_k}] = \exp(i.u^T m_X - (1/2)u^T R_X u). \quad (89)$$

On montre que m_X et R_X représentent alors respectivement le vecteur moyenne $\mathbb{E}[X]$ et la matrice de covariance $\mathbb{E}[XX^T] - m_X m_X^T$ de X .

Dans le cas où $|R_X| \neq 0$, la loi du vecteur X est absolument continue, de densité de probabilité

$$f_X(x) = (2\pi)^{-N/2} |R_X|^{-1/2} \exp\left[-\frac{1}{2}(x - m_X)^T R_X^{-1}(x - m_X)\right], \quad \forall x \in \mathbb{R}^n. \quad (90)$$

Dans le cas où X est un vecteur aléatoire à valeurs complexes, la densité de probabilité de X est obtenue en identifiant X à un vecteur aléatoire de \mathbb{R}^{2n} dont les composantes sont les parties réelles et imaginaires des composantes de X . Si ce vecteur aléatoire à valeurs dans \mathbb{R}^{2n} est gaussien, on dira que X est un vecteur aléatoire complexe gaussien.

La donnée d'un vecteur m et d'une matrice positive R de taille $n \times n$ caractérisent entièrement la loi d'un certain vecteur gaussien $X = [X_1, \dots, X_n]^T$ pour lequel $\mathbb{E}[X] = m$, et $\text{cov}[X_i, X_j] = R_{i,j}$.

12.2 Espérance conditionnelle dans le cas gaussien.

Notation. $X \sim \mathcal{N}(m_X, R_X)$ signifie que la loi de X est une loi gaussienne de moyenne m_X et de matrice de covariance R_X .

Pour un couple gaussien de vecteurs (X, Y) , on montre que $\mathbb{E}[X/Y]$ coïncide avec la projection de X sur l'espace vectoriel engendré par les combinaisons linéaires de composantes de Y . Plus précisément, on a le résultat suivant :

Si (X, Y) est un couple gaussien de vecteurs et $m_X, m_Y, R_X, R_Y, R_{XY}$ les moyennes, covariances et intercovariance de ces deux vecteurs,

$$\mathbb{E}[X/Y] = m_X + R_{XY} R_Y^{-1} (Y - m_Y) \quad (91)$$

et la loi conditionnelle de X sachant Y suit une loi

$$\mathcal{N}(m_X + R_{XY} R_Y^{-1} (Y - m_Y), R_X - R_{XY} R_Y^{-1} R_{XY}^T). \quad (92)$$

Démonstration. Notons

$$Z = X - m_X - R_{XY} R_Y^{-1} (Y - m_Y). \quad (93)$$

Il est facile de voir que

$$\begin{pmatrix} Z \\ Y \end{pmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ m_Y \end{bmatrix}, \begin{bmatrix} R_X - R_{XY} R_Y^{-1} R_{XY}^T & 0 \\ 0 & R_Y \end{bmatrix}\right). \quad (94)$$

Par suite,

$$\begin{aligned} \mathbb{E}[X/Y] &= \mathbb{E}[Z/Y] + m_X + R_{XY} R_Y^{-1} \mathbb{E}[(Y - m_Y)/Y] \\ &= \mathbb{E}[Z] + m_X + R_{XY} R_Y^{-1} (Y - m_Y) \\ &= m_X + R_{XY} R_Y^{-1} (Y - m_Y). \end{aligned} \quad (95)$$

Il est également clair que

$$\begin{aligned}\mathbb{E}[e^{iuX}/Y] &= \mathbb{E}[e^{iuZ}/Y] \exp[iu^T(m_X + R_{XY}R_Y^{-1}(Y - m_Y))] \\ &= \exp[iu^T(m_X + R_{XY}R_Y^{-1}(Y - m_Y)) - \frac{1}{2}u^T(R_X - R_{XY}R_Y^{-1}R_{XY}^T)u],\end{aligned}\tag{96}$$

ce qui termine la démonstration. \square

Remarques.

1) Le résultat précédent montre clairement que si deux vecteurs aléatoires gaussiens sont décorréllés, alors ils sont nécessairement indépendants. Cette propriété n'est vraie que pour des vecteurs gaussiens.

2) Si $X \sim \mathcal{N}(m_X, R_X)$, et A et b une matrice et un vecteur, $AX + b \sim \mathcal{N}(Am_X + b, AR_XA^T)$.

13 Convergence des variables aléatoires.

Malgré leur nom, les variables aléatoires sont des fonctions, et il est classique pour ce type d'objet mathématique de définir différents types de convergence. Ces différentes notions permettent de formuler plusieurs résultats remarquables concernant la convergence de suites de variables aléatoires indépendantes (dans le cas de variables dépendantes, il existe des extensions de ces résultats).

13.1 Convergence en probabilité.

La suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la variable aléatoire X , et on note $X_n \xrightarrow{P} X$ si

$$\forall \varepsilon > 0 \quad , \quad \lim_{n \rightarrow +\infty} P[|X_n - X| \geq \varepsilon] = 0.\tag{97}$$

13.2 Convergence en moyenne d'ordre α - Loi faible des grands nombres.

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en moyenne d'ordre α , avec $\alpha > 0$, vers la variable aléatoire X , et on note $X_n \xrightarrow{L^\alpha} X$, si

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^\alpha] = 0.\tag{98}$$

En pratique, α vaut souvent deux. On parle alors de convergence en moyenne quadratique.

Loi faible des grands nombres. Soit une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires. On suppose les variables aléatoires indépendantes, de même loi, et de variance finie. Alors

$(1/n) \sum_{k=1,n} X_k$ converge en moyenne quadratique vers l'espérance commune des variables, $\mathbb{E}[X_1]$:

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{L^2} \mathbb{E}[X_1]. \quad (99)$$

13.3 Convergence presque sûre - Loi forte des grands nombres.

Soit un espace probabilisé (Ω, \mathcal{B}, P) . On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement, vers la variable aléatoire X , et on note $X_n \xrightarrow{p.s.} X$, si

$$\exists N \subset \Omega, \text{ avec } P(N) = 0, \quad \omega \in \Omega - N, \Rightarrow (X_n(\omega) \rightarrow X(\omega)). \quad (100)$$

Un ensemble N tel que $P(N) = 0$ est dit **de mesure nulle** pour la mesure de probabilité P .

Critère de convergence presque sûre.¹ Le critère suivant est souvent utilisé pour démontrer la convergence presque sûre. Si on peut montrer que pour tout λ positif, la série des probabilités $P(|X_n - X| > \lambda)$ est de somme finie, alors la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X presque sûrement. En termes mathématiques,

$$(\forall \lambda > 0, \sum_{n=1}^{\infty} P(|X_n - X| > \lambda) < \infty) \Rightarrow (X_n \xrightarrow{p.s.} X). \quad (103)$$

Loi forte des grands nombres. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, de même loi et avec une espérance finie. Alors $(1/n) \sum_{k=1,n} X_k$ converge presque sûrement vers l'espérance $\mathbb{E}[X_1]$ commune aux variables aléatoires :

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p.s.} \mathbb{E}[X_1]. \quad (104)$$

13.4 Convergence en loi - Théorème limite-centrale.

C'est une propriété de convergence faible très souvent utilisée en pratique. On dit que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire X , et on note $X_n \xrightarrow{L} X$, si l'une des conditions suivantes est vérifiée

1. Ce critère est une conséquence du **lemme de Borel-Cantelli**. Pour énoncer ce lemme, rappelons tout d'abord que $\limsup A_n = \cap_{p \in \mathbb{N}} (\cup_{n \geq p} A_n)$, représente l'ensemble des éléments qui sont communs à une infinité d'ensembles A_n . On montre alors que pour un ensemble d'événements $(A_n)_{n \in \mathbb{N}}$,

$$\sum_{n \in \mathbb{N}} P(A_n) < \infty \Rightarrow P(\limsup A_n) = 0, \quad (101)$$

et si les A_n sont deux à deux indépendants,

$$\sum_{n \in \mathbb{N}} P(A_n) = \infty \Rightarrow P(\limsup A_n) = 1. \quad (102)$$

Ce résultat est également connu sous le nom de **loi du tout ou rien**.

- $F_{X_n}(x)$ converge vers $F_X(x)$ en tout point de continuité de F_X
- $\varphi_{X_n}(t)$ converge vers $\varphi_X(t)$ en tout point t de \mathbb{R} .
- pour toute fonction g continue bornée, $\mathbb{E}[g(X_n)]$ converge vers $\mathbb{E}[g(X)]$.

Lorsque les variables aléatoires sont absolument continues, la convergence en loi est équivalente à la convergence simple de $f_{X_n}(x)$ vers $f_X(x)$.

Théorème limite-centrale (en anglais central limit theorem). Soit une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes, de même loi de variance finie σ^2 . Notons $S_n = X_1 + \dots + X_n$. Alors

$$\frac{S_n - \mathbb{E}[S_n]}{\sigma_{S_n}} \xrightarrow{L} \mathcal{N}(0, 1) \quad (105)$$

D'après la linéarité de l'espérance on a évidemment $\mathbb{E}[S_n] = n\mathbb{E}[X_1]$ et $\sigma_{S_n}^2 = n\sigma_{X_1}^2$.

Remarque sur les suites gaussiennes. La limite d'une suite de variables aléatoires gaussiennes $(X_n)_{n \in \mathbb{N}}$, convergente dans l'espace $L^2(\Omega, \mathcal{B}, P)$ des variables aléatoires réelles X sur (Ω, \mathcal{B}, P) pour lesquelles $\mathbb{E}[X^2] < \infty$, est une variable aléatoire X gaussienne. Pour s'en assurer, il suffit de vérifier que $\mathbb{E}[X_n]$ et $\mathbb{E}[X_n^2]$ convergent vers $\mathbb{E}[X] = m_X$ et $\mathbb{E}[X^2] = \sigma_X^2 + m_X^2$, puis de remarquer que les fonctions caractéristiques $\varphi_{X_n}(u)$ de variables aléatoires X_n convergent vers $\exp(i.m_X u - \sigma_X^2 u^2/2)$ en tout point de $u \in \mathbb{R}$.

13.5 Relations entre les différents types de convergence.

On a les relations suivantes entre les différents types de convergence

$$\begin{array}{ccccc} X_n \xrightarrow{p.s.} X & \implies & X_n \xrightarrow{P} X & \implies & X_n \xrightarrow{L} X \\ & & \uparrow & & \\ & & X_n \xrightarrow{L^\alpha} X & & \\ & & \uparrow & & \\ & & X_n \xrightarrow{L^\beta} X & \beta > \alpha & \end{array}$$

14 Quelques lois de probabilités classiques.

14.1 Lois discrètes.

Loi discrète uniforme.

Epreuve type : choix au hasard d'un entier entre 1 et N .

$$P(X = k) = \frac{1}{N} \mathbb{I}_{\{1, \dots, N\}}(k). \quad (106)$$

$$\mathbb{E}[X] = \frac{N+1}{2}, \quad \sigma_X^2 = \frac{N^2-1}{12}, \quad g_X(z) = \frac{z}{N} \cdot \frac{z^N - 1}{z - 1}, \quad (107)$$

où $g_X(z)$ est la fonction génératrice des moments.

Loi de Bernoulli.

Epreuve type : On parle d'épreuve de Bernoulli pour une expérience qui a deux issues possibles, exprimées par $X = 0$ ("échec"), et $X = 1$ ("succès"). $P(X = 0) = 1 - p$, et $P(X = 1) = p$.

$$\mathbb{E}[X] = p, \quad \sigma_X^2 = p(1 - p), \quad g_X(z) = 1 - p + pz \quad (108)$$

Loi Binomiale : $\mathcal{B}(n, p)$.

Epreuve type : On compte le nombre de "succès" obtenus au cours de n épreuves de Bernoulli successives indépendantes. Une variable $\mathcal{B}(n, p)$ peut être vue comme la somme de n variables de Bernoulli indépendantes de même paramètre p .

$$P(X = k) = \begin{cases} C_n^k p^k (1 - p)^{n-k} & \text{si } k \in \{0, 1, \dots, n\} \\ 0 & \text{sinon} \end{cases} \quad (109)$$

$$\mathbb{E}[X] = np, \quad \sigma_X^2 = np(1 - p), \quad g_X(z) = (1 - p + pz)^n. \quad (110)$$

Loi Multinomiale.

Epreuve type : tirage de n objets **avec remise** parmi r classes d'objets. Soit p_i la probabilité de la classe i ($\sum_{i=1}^r p_i = 1$).

La probabilité de tirer n_1 objets de la classe 1, n_2 de la classe 2, ..., avec $\sum_{i=1}^r n_i = n$, est

$$P(n_1, n_2, \dots, n_r) = \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i}. \quad (111)$$

Loi de Poisson : $\mathcal{P}(\lambda)$.

Epreuve type : nombre d'événements réalisés pendant un certain intervalle de temps, ces événements se produisant de façon indépendante (nombre de clients à un guichet, d'appels à un central téléphonique, ...)

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}, \quad \lambda > 0. \quad (112)$$

$$\mathbb{E}[X] = \sigma_X^2 = \lambda \quad g_X(z) = e^{\lambda(z-1)}. \quad (113)$$

Si $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$, et X et Y indépendantes, $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Loi géométrique (ou de Pascal) : $\mathcal{G}(p)$.

Epreuve type : Dans une suite d'épreuves de Bernoulli indépendantes de même paramètre p , c'est le nombre d'épreuves à effectuer pour obtenir un premier "succès".

$$P(X = n) = pq^n \quad \text{avec } n \in \mathbb{N}, \quad 0 < p < 1 \quad \text{et } p + q = 1. \quad (114)$$

$$\mathbb{E}[X] = \frac{q}{p}, \quad \sigma_X^2 = \frac{q^2}{p^2} + \frac{q}{p}, \quad g_X(z) = \frac{p}{1 - qz} \quad (115)$$

Loi hypergéométrique : $\mathcal{H}(N, n, p)$.

Epreuve type : tirage **sans remise**. C'est la probabilité de tirer k boules noires en tirant n boules dans une urne de N boules dont Np sont noires.

$$P(X = k) = \begin{cases} \frac{C_{Np}^k \cdot C_{N(1-p)}^{n-k}}{C_N^n} & \text{si } \max(0, n - N(1-p)) \leq k \leq \min(n, Np) \\ 0 & \text{sinon} \end{cases} \quad (116)$$

Autre formulation avec $N = N_1 + N_2$, et $N_1 = Np$: $P(X = k) = \frac{C_{N_1}^k \cdot C_{N_2}^{n-k}}{C_{N_1+N_2}^n}$

$$\mathbb{E}[X] = np, \quad \sigma_X^2 = \frac{N-n}{N-1} np(1-p). \quad (117)$$

Loi binomiale négative : $\mathcal{Bn}(r, p)$.

Epreuve type : nombre d'épreuves à effectuer pour obtenir un r -ème "succès" dans une suite d'épreuves de Bernoulli.

$$P(X = k) = \begin{cases} C_{k-1}^{r-1} q^{k-r} p^r & \text{si } k \geq r \\ 0 & \text{sinon} \end{cases} \quad (118)$$

$$\mathbb{E}[X] = \frac{r}{p}, \quad \sigma_X^2 = r \frac{q}{p^2}, \quad g_X(z) = \left(\frac{pz}{1-qz} \right)^r. \quad (119)$$

14.2 Lois à densités.

Loi uniforme : $\mathcal{U}_{[a,b]}$.

Loi uniforme sur $[a, b]$: $f_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$.

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}. \quad (120)$$

Loi gaussienne : $\mathcal{N}(m, \sigma^2)$.

Pour le cas vectoriel, on pourra se référer au paragraphe 12.1. Pour le cas scalaire, $X \sim \mathcal{N}(m, \sigma^2)$ si

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (121)$$

$\mathbb{E}[X] = m$ et $\sigma_X^2 = \sigma^2$.

$$\varphi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}} \quad (122)$$

Si $X \sim \mathcal{N}(0, 1)$,

$$\mathbb{E}[X^{2p+1}] = 0, \quad \mathbb{E}[X^{2p}] = \frac{(2p!)}{p!2^p}. \quad (123)$$

Les calculs sur les lois gaussiennes amènent parfois à considérer les fonctions *erf* (error

function), $erfc$ ou Q définies par

$$\begin{aligned} erf(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du \\ erfc(x) &= \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du \\ Q(x) &= 1 - \int_{-\infty}^x \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du. \end{aligned} \quad (124)$$

Notons que $Q(x) = \frac{1}{2}erfc(\frac{x}{\sqrt{2}})$, et $erfc(x) = 2Q(x\sqrt{2})$. Ces fonctions apparaissent notamment dans les calculs de probabilités d'erreurs de transmissions en communications numériques.

Loi gamma : $\gamma(\theta, \lambda)$.

$X \sim \gamma(\theta, \lambda)$ si

$$f_X(x) = \frac{\lambda^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\lambda x} \mathbb{I}_{\mathbb{R}_+}(x), \quad (125)$$

avec $\Gamma(\theta) = \int_0^\infty x^{\theta-1} e^{-x} dx$.

$$\mathbb{E}[X] = \frac{\theta}{\lambda}, \quad \sigma_X^2 = \frac{\theta}{\lambda^2}, \quad \varphi_X(t) = \left(\frac{\lambda}{\lambda - it} \right)^\theta. \quad (126)$$

$$\mathbb{E}[X^n] = \frac{\Gamma(n+\theta)}{\Gamma(\theta)} \lambda^{-n} = (n+\theta-1)(n+\theta-2)\dots\theta \lambda^{-n} \quad (127)$$

Notons que $\lambda X \sim \gamma(\theta, 1)$.

Cas particuliers importants de la loi γ : lois d'Erlang ($\gamma(n, \lambda)$, $n \in \mathbb{N}$), loi du chi-deux ($\gamma(n/2, 1/(2\sigma^2))$), loi exponentielle ($\gamma(1, \lambda)$).

Loi du chi deux χ^2 . Soit $X = \sum_{k=1, n} X_k^2$, avec $X_k \sim \mathcal{N}(0, \sigma^2)$ et indépendantes. On dit que X suit une loi du χ^2 à n degrés de liberté.

$$f_X(x) = \frac{1}{\sigma^n 2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2\sigma^2}\right). \quad (128)$$

$$\mathbb{E}[X] = n\sigma^2, \quad \sigma_X^2 = 2n\sigma^4, \quad \varphi_X(t) = \frac{1}{(1 - 2it\sigma^2)^{n/2}}. \quad (129)$$

Loi exponentielle : $\mathcal{E}(\lambda)$. Epreuve type : temps d'attente entre deux phénomènes indépendants tels que des arrivées à un guichet, ou des appels téléphoniques.

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{\mathbb{R}_+}(x) \quad (130)$$

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \sigma_X^2 = \frac{1}{\lambda^2}, \quad \varphi_X(t) = \frac{\lambda}{\lambda - it}. \quad (131)$$

Propriété caractéristique : $P(X > b | X > a) = P(X > b - a)$.

Notons que la loi exponentielle est une loi χ^2 à deux degrés de liberté.

Loi d'Erlang. C'est une loi $\gamma(n, \lambda)$. Epreuve type : c'est la loi de l'instant où se produit le n -ème évènement pour des phénomènes indépendants qui suivent une loi $\mathcal{E}(\lambda)$ (e.g. temps du n -ème appel téléphonique à un central).

$$f_X(x) = \frac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} \mathbb{I}_{\mathbb{R}_+}(x) \quad (132)$$

$$\mathbb{E}[X] = \frac{n}{\lambda}, \quad \sigma_X^2 = \frac{n}{\lambda^2}, \quad \varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^n. \quad (133)$$

Loi de Rayleigh : $\mathcal{R}(\sigma^2)$.

Epreuve type : soit $X = \sqrt{X_1^2 + X_2^2}$, avec $X_1, X_2 \sim \mathcal{N}(0, \sigma^2)$ et indépendantes. On dit que X suit une loi de Rayleigh. C'est aussi la racine carrée d'une loi exponentielle. La loi de Rayleigh apparaît souvent pour décrire le bruit en sortie de certains récepteurs de transmissions.

$$f_X(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \mathbb{I}_{\mathbb{R}_+}(x). \quad (134)$$

$$\mathbb{E}[X] = \sigma \sqrt{\frac{\pi}{2}}, \quad \sigma_X^2 = (2 - \pi/2)\sigma^2. \quad (135)$$

$\varphi_X(t)$ n'admet pas ici d'expression analytique.

Références

- [1] J. BASS, *Eléments de Calcul des Probabilités*, Masson, 1974.
- [2] P.B. BILLINGSLEY, *Probability and measure, Wiley Series in Probability and Mathematical Statistics*, 1979.
- [3] N. BOULEAU, *Probabilités de l'ingénieur, Variables Aléatoires et Simulations*, Hermann, 1986.
- [4] P. BREMAUD, *Introduction aux Probabilités*, Springer Verlag, 1984.
- [5] G. CALOT, *Cours de Calcul de Probabilités*, Dunod, 1967.
- [6] P.Y. COCHET, *Cours de probabilités*, polycopié ENSTBr, 1996
- [7] D. DACUNHA-CASTELLE, M.DUFLO, *Probabilités et Statistiques - Tome 1 : Problèmes à temps fixe*. Masson, 1982.
- [8] J.P. DELMAS, *Introduction aux probabilités*, Ellipses, Collection Pédagogique de Télécommunication, Paris, 1993.
- [9] J.L. DOOB, *measure Theory*, Springer-Verlag, 1993.
- [10] A. HILLION, *Probabilités (résumé de cours)*, polycopié ENSTBr, 1984.
- [11] A. HILLION, *Théorie de la mesure et intégration (résumé de cours)*, polycopié ENSTBr, 1984.
- [12] N.L. JOHNSON, S. KOTZ, N. BALAKRISHNAN, *Continuous Univariate Distributions, Continuous Univariate Distributions, Univariate Discrete Distributions, Discrete Multivariate Distributions*, (4 tomes) Wiley Series in Probability & statistics, 1993-1997.
- [13] M. METIVIER, *Notions Fondamentales de la Théorie des Probabilités*, Dunod, Paris, 1972.
- [14] E. PARZEN, *Modern Probability Theory*, John Wiley and Sons, 1960.
- [15] A. PAPOULIS, *Probability and Random Variables*, Mc Graw Hill, 1991.