

Université Claude Bernard Lyon 1
Année universitaire 2007-2008

**Préparation au CAPES de
Mathématiques
Probabilités**

F. Bienvenue-Duheille

Chapitre 1

Probabilités

1 Mesure

1.1 Définitions

On se place sur un ensemble Ω appelé **espace de probabilité** que l'on muni d'une tribu Σ .

Définition 1.1 Une tribu Σ est un sous-ensemble de $\mathcal{P}(\Omega)$ tel que

- $\emptyset \in \Sigma$,
- Si $A \in \Sigma$, $\Omega \setminus A \in \Sigma$,
- Si les $(A_n)_{n \geq 1}$ sont des éléments de Σ , alors $\cup_{n \geq 1} A_n \in \Sigma$.

En théorie de la mesure, un sous-ensemble $A \in \Sigma$ est dit mesurable.

En probabilité, un élément ω de Ω est appelé une **expérience** et un sous-ensemble A de Ω appartenant à Σ , un **événement**.

En pratique, lorsque Ω est fini ou dénombrable la tribu utilisée sera (presque) toujours la tribu $\mathcal{P}(\Omega)$. Si Ω est \mathbf{R} (ou un intervalle de \mathbf{R}), la tribu utilisée sera le plus souvent la tribu borélienne $\mathcal{B}(\mathbf{R})$, c'est-à-dire la plus petite tribu (au sens de l'inclusion des tribus) contenant (au choix) : les intervalles, les ouverts ou les fermés.

Une fonction $h : \Omega \rightarrow \mathbf{R}$ sera dite $(\Sigma, \mathcal{B}(\mathbf{R}))$ -mesurable si, pour tout $B \in \mathcal{B}(\mathbf{R})$, on a

$$h^{-1}(B) = \{\omega \in \Omega, h(\omega) \in B\} = \{h \in B\} \in \Sigma.$$

Une fonction borélienne $h : \mathbf{R} \rightarrow \mathbf{R}$ est une fonction $(\mathcal{B}(\mathbf{R}), \mathcal{B}(\mathbf{R}))$ -mesurable.

Définition 1.2 Une **mesure de probabilité** \mathbf{P} est une fonction définie sur la tribu Σ et à valeurs dans $[0, 1]$ vérifiant les propriétés suivantes :

1. $\mathbf{P}(\Omega) = 1$.
2. Si A et B sont deux sous-ensembles disjoints de Ω et appartenant à Σ $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.
3. Si $(A_n)_{n \geq 1}$ est une famille dénombrables de sous-ensembles de Ω deux à deux disjoints, appartenant à Σ , on a

$$\mathbf{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_n \mathbf{P}(A_n).$$

On déduit de la définition d'une mesure de probabilité que

- $\mathbf{P}(\emptyset) = 0$,
- si A est un événement, $\mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A)$,
- si $A \subset B$ sont deux événements, $\mathbf{P}(A) \leq \mathbf{P}(B)$,
- et si A et B sont deux événements, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

On montre facilement par récurrence le résultat suivant appelé formule de Poincaré ou formule du crible :

Proposition 1.3 Soit $(A_k)_{1 \leq k \leq n}$ n événements quelconques de Ω . On a

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbf{P}(A_i) \\ &\quad - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) + \\ &\quad + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i \cap A_j \cap A_k) \\ &\quad + \dots + (-1)^n \mathbf{P}\left(\bigcap_{i=1}^n A_i\right). \end{aligned}$$

1.2 Probabilités discrètes

La mesure de probabilité \mathbf{P} est dite **discrète** dès que l'espace Ω est fini ou dénombrable ou plus généralement, dès qu'il existe un sous-ensemble Ω_0 de Ω fini ou dénombrable et tel que $\mathbf{P}(\Omega_0) = 1$. Une probabilité sur un ensemble dénombrable sera toujours discrète.

Une probabilité sur un ensemble dénombrable est complètement déterminée par les $\mathbf{P}(\{\omega\})$ pour tout $\omega \in \Omega$. En effet, pour $A \subset \Omega$, on a $\mathbf{P}(A) = \sum_{\omega \in (A \cap \Omega_0)} \mathbf{P}(\omega)$ où Ω_0 est un événement dénombrable et de probabilité 1.

Remarques :

- Les poids d'une probabilité discrète vérifient $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$.
- Une mesure de probabilité ne permet d'évaluer a priori que la taille de sous-ensembles de Ω .

Des exemples

- *Lancer d'une pièce équilibrée* : on souhaite modéliser le résultat du lancer d'une pièce sans tricherie. Pour cela, on choisit $\Omega_1 = \{\text{pile}, \text{face}\}$, et donc $\mathbf{card} \Omega_1 = 2$. L'ensemble des parties de Ω_1 comporte quatre éléments et on définit la mesure de probabilité \mathbf{P} par $\mathbf{P}\{\text{pile}\} = \mathbf{P}\{\text{face}\} = 1/2$ puisque les deux événements sont équiprobables (c'est-à-dire de même probabilité)..

Remarque : On aurait très bien pu choisir $\Omega_1 = \{\text{pile}, \text{face}, \text{rouge}, \text{vert}\}$, et comme mesure de probabilité $\mathbf{P}\{\text{pile}\} = \mathbf{P}\{\text{face}\} = 1/2$ et $\mathbf{P}\{\text{rouge}\} = \mathbf{P}\{\text{vert}\} = 0$, mais tant qu'à faire, on choisit le plus simple...

- *Lancer de k pièces, $k \geq 2$* : on prend cette fois-ci $\Omega_k = (\Omega_1)^k$, c'est-à-dire l'ensemble des k -uplets de pile ou face. On a $\mathbf{card} \Omega_k = 2^k$ et $\mathbf{card} \mathcal{P}(\Omega_k) = 2^{2^k}$. Les différents k -uplets sont tous équiprobables donc $\mathbf{P}(\omega) = 2^{-k}$, pour tout $\omega \in \Omega_k$.

• *Probabilité uniforme discrète* : sur un ensemble **fini** $\Omega = \{\omega_1, \dots, \omega_n\}$, on définit la probabilité uniforme par $\mathbf{P}(\omega_i) = 1/n$ pour tout i entre 1 et n . Dans ce cas, tous les ω_i ont la même probabilité de se produire (i.e. sont **équiprobables**), et pour une partie A de Ω , on a

$$\mathbf{P}(A) = \frac{\text{card } A}{n} = \frac{\text{nb cas favorables}}{\text{nb cas possibles}}.$$

Par exemple, lors du lancer d'un dé régulier à six faces, chaque face est obtenue avec la même probabilité $1/6$.

Remarque : Il ne peut bien sûr pas y avoir de loi uniforme sur \mathbf{N} .

• *Exemple de mesure de probabilité sur \mathbf{N}^** . On lance un dé de façon répétée jusqu'à obtenir un 6, et on note le numéro du tirage du premier 6. On a évidemment $\mathbf{P}(1) = 1/6$.

On a également

$$\begin{aligned} \mathbf{P}(2) &= \mathbf{P}(\text{au premier tirage, on n'a pas eu de 6 ; au deuxième tirage, on a eu un 6}) \\ &= \frac{5}{36} \end{aligned}$$

car sur les 36 tirages possibles équiprobables, seuls 5 permettent d'obtenir le premier 6 au deuxième tirage.

De même, pour tout $k \geq 2$,

$$\mathbf{P}(k) = \mathbf{P}(k-1 \text{ échecs puis une réussite}) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

Cela constitue bien une mesure de probabilité discrète sur \mathbf{N}^* puisque $\sum_{k \geq 1} \mathbf{P}(k) = 1$.

Attention : Ne pas confondre cette probabilité avec la probabilité de tirer un 6 exactement parmi les k premiers lancers.

1.3 Probabilité à densité

On se place sur \mathbf{R} et on note dx l'élément d'intégration de la mesure de Lebesgue. Soit $f : \mathbf{R} \rightarrow \mathbf{R}$ une fonction positive, borélienne et d'intégrale sur \mathbf{R} égale à 1. Il est facile de vérifier que l'on définit une mesure μ en posant, pour tout borélien A :

$$\mu(A) = \int_{\mathbf{R}} \mathbf{1}_A(x) f(x) dx.$$

Une telle mesure est dite à densité (par rapport à la mesure de Lebesgue sur \mathbf{R}). On dit également que c'est une probabilité continue. La tribu utilisée sur \mathbf{R} sera (presque) toujours la tribu borélienne.

Des exemples

• La mesure uniforme sur l'intervalle $[a, b]$, où $a < b$: On définit

$$\mu(A) = \int_{\mathbf{R}} \mathbf{1}_{A \cap [a, b]}(x) \frac{dx}{b-a}.$$

• La mesure de Gauss sur \mathbf{R} . On utilise ici la fonction

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

où $m \in \mathbf{R}$ et $\sigma \in \mathbf{R}^{+*}$ sont deux paramètres fixés.

1.4 Probabilité conditionnelle, indépendance

Définition 1.4 On se donne deux événements A et B de Ω , avec $\mathbf{P}(B) > 0$. On définit la probabilité conditionnelle de A sachant B , notée $\mathbf{P}(A|B)$ par

$$\mathbf{P}(A|B) = \mathbf{P}(A \cap B) / \mathbf{P}(B).$$

Par exemple, si l'on dispose d'un échantillon de maïs comportant des grains lisses (L) ou frippés (F) et de couleur jaune (J) ou bleue (B), on peut calculer la proportion de jaunes parmi les frippés dans l'échantillon considéré c'est-à-dire $\mathbf{P}(J \text{ et } F) / \mathbf{P}(F)$.

La probabilité conditionnelle vérifie les mêmes propriétés qu'une probabilité : on a ainsi $\mathbf{P}(\Omega|B) = 1$, $\mathbf{P}(\emptyset|B) = 0$, si A_1 et A_2 sont disjoints, $\mathbf{P}(A_1 \cup A_2|B) = \mathbf{P}(A_1|B) + \mathbf{P}(A_2|B)$, $\mathbf{P}(\Omega \setminus A|B) = 1 - \mathbf{P}(A|B)$...

Les probabilités conditionnelles permettent de décomposer un événement suivant des sous-ensembles de Ω sur lesquels on maîtrise mieux ce qui se passe. Pour cela introduisons la notion de système complet d'événements :

Définition 1.5 Un système complet d'événements est une famille dénombrable ou finie (B_n) d'événements deux à deux disjoints et vérifiant $\sum_n \mathbf{P}(B_n) = 1$.

Formule des probabilités totales : Soit B_n un système complet d'événements et A un événement. On a $\mathbf{P}(A) = \sum_n \mathbf{P}(A|B_n)\mathbf{P}(B_n)$.

Remarque : Si par exemple $\mathbf{P}(B_1) = 0$, on peut poser $\mathbf{P}(A|B_1) = 0$, ou 1, ou 1/2 pour tout A , cela n'interviendra pas dans la formule ci-dessus.

Preuve : Par définition, $\mathbf{P}(A|B_n)\mathbf{P}(B_n) = \mathbf{P}(A \cap B_n)$ et les événements $A \cap B_n$ sont deux à deux disjoints car les B_n le sont. On en déduit donc que

$$\sum_n \mathbf{P}(A|B_n)\mathbf{P}(B_n) = \mathbf{P}(\cup(A \cap B_n)) = \mathbf{P}(A \cap (\cup_n B_n)).$$

Notons $\Omega_0 = \cup B_n$.

Comme $A = (A \cap \Omega_0) \cup (A \cap (\Omega \setminus \Omega_0))$, on a $\mathbf{P}(A) = \mathbf{P}(A \cap \Omega_0) + \mathbf{P}(A \cap (\Omega \setminus \Omega_0))$.

Or $A \cap (\Omega \setminus \Omega_0) \subset (\Omega \setminus \Omega_0)$ et $\mathbf{P}(\Omega \setminus \Omega_0) = 1 - \mathbf{P}(\Omega_0) = 0$ donc $\mathbf{P}(A \cap (\Omega \setminus \Omega_0)) = 0$.

On a donc bien $\sum_n \mathbf{P}(A|B_n)\mathbf{P}(B_n) = \mathbf{P}(\cup(A \cap B_n)) = \mathbf{P}(A)$.

Un problème courant est de déterminer $\mathbf{P}(A|B)$ à partir de $\mathbf{P}(B|A)$. La seule donnée de $\mathbf{P}(B|A)$ n'y suffit pas. Il faut par exemple connaître aussi $\mathbf{P}(A)$ et $\mathbf{P}(B)$: on a alors $\mathbf{P}(A|B) = \mathbf{P}(B|A)\mathbf{P}(A) / \mathbf{P}(B)$. Une autre possibilité est de connaître $\mathbf{P}(A)$ et $\mathbf{P}(B|\Omega \setminus A)$:

Formule de Bayes : Soient A et B deux événements de probabilité strictement positive. On vérifie que

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|\Omega \setminus A)\mathbf{P}(\Omega \setminus A)}.$$

Preuve : Le dénominateur du membre de droite vaut en fait $\mathbf{P}(B)$, alors que le numérateur vaut $\mathbf{P}(A \cap B)$, d'où le résultat.

Définition 1.6 Deux événements A et B sont dits **indépendants** si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On a alors $\mathbf{P}(A|B) = \mathbf{P}(A)$ et $\mathbf{P}(B|A) = \mathbf{P}(B)$ si $\mathbf{P}(A) > 0$ et $\mathbf{P}(B) > 0$.

Exercice :

1. Montrer qu'un événement de probabilité nulle est indépendant de tout événement.
2. Montrer que si A et B sont indépendants, alors $\Omega \setminus A$ et B le sont.
3. Montrer qu'un événement de probabilité 1 est indépendant de tout événement.

Exemples :

- Lors d'un lancer de *pile ou face*, les événements « tomber sur pile au premier tirage » et « tomber sur pile au deuxième tirage » sont généralement indépendants (sauf en cas de tricherie...)

- *Tirage avec remise.* On dispose d'une urne contenant N boutons noirs et J boutons jaunes. À chaque tirage, on prend un bouton au hasard, on note la couleur du bouton obtenu et on le remet dans l'urne. Les événements $A = \{\text{tirer un bouton noir au premier tirage}\}$ et $B = \{\text{tirer un bouton jaune au deuxième tirage}\}$ sont-ils indépendants ?

- *Urne de Polya.* On dispose toujours d'une urne contenant N boutons noirs et J boutons jaunes. À chaque tirage, on note la couleur du bouton obtenu et on le remet dans l'urne accompagné d'un bouton de la même couleur. Même question que précédemment.

Définition 1.7 n événements A_1, \dots, A_n sont (**mutuellement ou n à n**) **indépendants** si pour tout choix d'indices i_1, \dots, i_k deux à deux distincts, on a

$$\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \times \dots \times \mathbf{P}(A_{i_k}).$$

Des événements n à n indépendants le sont bien évidemment 2 à 2 mais la réciproque est fausse.

Exercice :

- On choisit $\Omega = \{1, 2, 3, 4\}$ et on le munit de la probabilité uniforme. Trouver trois événements deux à deux indépendants mais pas trois à trois.

- Sur $\Omega = \{1, \dots, 8\}$ muni de la probabilité uniforme, trouver trois événements A , B et C tels que $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ mais tels que A , B et C ne sont pas indépendants.

Chapitre 2

Variables aléatoires réelles

1 La loi

1.1 Définition

Une variable aléatoire $X : \Omega \rightarrow \mathbf{R}$ est une fonction $(\Sigma, \mathcal{B}(\mathbf{R}))$ -mesurable. Sa loi est donnée par l'ensemble $X(\Omega)$ ainsi que par la mesure $\mu(B) = \mu(X \in B)$ pour tout borélien B . μ est appelée la mesure image de \mathbf{P} par X . On note parfois $\mu = X(\mathbf{P})$ ou $\mu = \mathbf{P}_X$.

La loi est la principale information dont on disposera sur une variable aléatoire : souvent l'ensemble Ω sera inconnu ou implicite, on n'aura donc pas d'information sur $X(\omega)$.

La variable aléatoire X sera discrète si elle prend ses valeurs dans un ensemble discret (et sa mesure-image est alors une mesure discrète). Sa loi sera caractérisée par l'ensemble $X(\Omega)$ (ou par un ensemble dénombrable contenant $X(\Omega)$) et par les probabilités $\mathbf{P}(X = x)$ pour tout $x \in X(\Omega)$.

X sera à densité (on dit aussi que X est continue) si sa mesure image admet une densité, que l'on notera f . On aura alors pour tout intervalle $[a, b]$ de \mathbf{R} (avec $a < b$),

$$\mathbf{P}(X \in [a, b]) = \int_a^b f(x) dx.$$

En particulier en prenant $a = b$ dans l'égalité ci-dessus, on remarque $\mathbf{P}(X = a) = 0$ pour tout $a \in \mathbf{R}$.

Remarque : Si Ω est un ensemble fini ou dénombrable, toute variable aléatoire définie sur Ω sera discrète.

Attention : Deux variables aléatoires peuvent suivre la même loi sans être égales : par exemple deux tirages successifs de pile ou face.

Nous allons maintenant étudier quelques exemples de variables aléatoires discrètes ou à densité, mais il faut garder à l'esprit que cela ne recouvre pas tous les types de variables aléatoires.

1.2 Exemples de variables aléatoires discrètes

Définition 2.1 La loi d'une variable aléatoire discrète est donnée par
– l'ensemble (dénombrable) $X(\Omega)$,

– pour tout $x \in X(\Omega)$, la quantité

$$\mathbf{P}(\{\omega \in \Omega \text{ tels que } X(\omega) = x\}) = \mathbf{P}(X^{-1}\{x\}) = \mathbf{P}(X = x)$$

Remarque : On doit avoir $\sum_x \mathbf{P}(X = x) = 1$, où la somme est prise sur $x \in X(\Omega)$.

Pour construire une variable aléatoire discrète, on peut aussi commencer par définir une mesure de probabilité sur \mathbf{N} en se donnant le poids p_n de chaque entier n (avec $\sum p_n = 1$) puis considérer une variable aléatoire X d'un certain espace Ω dans \mathbf{N} dont la loi est donnée par $\mathbf{P}(X = n) = p_n$.

Exercice : On se donne une variable aléatoire $X : \Omega \rightarrow \mathbf{N}$. Montrer que la famille $A_n = \{\omega, X(\omega) = n\}$ pour tout $n \geq 0$ forme un système complet d'événements.

Des exemples

- Pour un événement $A \subset \Omega$, on note $\mathbf{1}_A$ la fonction suivante : $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon. Cette fonction, appelée **l'indicatrice de l'événement** A , est une variable aléatoire discrète très utile.

- Le nombre de piles obtenus lors des 8 premiers tirages d'un jeu de pile ou face est aussi une variable aléatoire discrète.

- *Loi de Dirac* en $a \in \mathbf{R}$. On fixe un nombre réel a . La loi de Dirac en a , généralement notée δ_a , est la loi de la variable aléatoire suivante : $X(\Omega) = \{a\}$ et $\mathbf{P}(X = a) = 1$. On dit que X vaut « presque-sûrement » a .

Exercice : Montrer que, pour $A \subset \mathbf{R}$, $\mathbf{P}(X \in A) = \mathbf{1}_A(a)$ si X suit la loi de Dirac en a .

- *Loi de Bernoulli*. La loi de Bernoulli $\mathcal{B}(p)$ de paramètre $p \in [0, 1]$ est donnée par $X(\Omega) = \{0, 1\}$ et $\mathbf{P}(X = 1) = p = 1 - \mathbf{P}(X = 0)$. Lors d'un tirage de pile ou face d'une pièce équilibrée, si on note $X = 1$ si la pièce tombe sur pile et 0 sinon, on obtient une variable aléatoire de loi de Bernoulli $\mathcal{B}(\frac{1}{2})$. Plus généralement, pour un événement A quelconque, la variable aléatoire $\mathbf{1}_A$ suit une loi de Bernoulli de paramètre $\mathbf{P}(A)$.

- *Loi binomiale*. La loi binomiale $\text{Bin}(n, p)$, pour $n \in \mathbf{N}^*$ et $p \in [0, 1]$ est donnée par $X(\Omega) = \{0, \dots, n\}$ et $\mathbf{P}(X = k) = C_n^k p^k (1 - p)^{n-k}$, pour tout $k \in \{0, \dots, n\}$. On retrouve ici la probabilité d'obtenir k fois au cours de n tentatives la réalisation d'un événement dont la probabilité est p . Par exemple, la probabilité de tirer exactement k 6 lors des n premiers lancers d'un dé est $C_n^k 5^{n-k} 6^{-n}$.

- *Loi uniforme* sur $\{1, \dots, n\}$. On a ici $X(\Omega) = \{1, \dots, n\}$ et cette loi affecte le même poids à chacun des éléments. On a donc $\mathbf{P}(X = k) = 1/n$, pour tout $k \in \{1, \dots, n\}$.

- *Loi géométrique* $\mathcal{G}(p)$, $p \in]0, 1[$: Cette loi est donnée par $X(\Omega) = \mathbf{N}^*$ et $\mathbf{P}(X = k) = p(1 - p)^{k-1}$ pour tout $k \in \mathbf{N}^*$. On a vu plus haut que c'est la loi du numéro du tirage où un événement se réalise pour la première fois.

- *Loi de Poisson* $\mathcal{P}(\lambda)$, $\lambda > 0$. C'est la loi de la variable aléatoire X vérifiant $X(\Omega) = \mathbf{N}$ et $\mathbf{P}(X = k) = e^{-\lambda} \lambda^k / k!$. Elle est généralement utilisée pour modéliser le nombre d'appels reçus par un serveur au cours d'un laps de temps donné.

1.3 Exemples de loi à densité

Loi uniforme sur $[a, b]$: c'est la loi de la variable aléatoire X de densité $\mathbf{1}_{[a,b]}/(b - a)$. La

probabilité qu'une variable aléatoire de loi uniforme sur $[a, b]$ appartienne à un sous-intervalle de $[a, b]$ est proportionnelle à la longueur de ce sous-intervalle. On a en particulier $\mathbf{P}(X \in [a, b]) = 1$.

Loi exponentielle de paramètre λ . Il s'agit de la loi de densité $f_\lambda(x) = \lambda \exp(-\lambda x) \mathbf{1}_{x>0}$. Si X suit cette loi, on a $\mathbf{P}(X \geq 0) = 1$. La loi exponentielle est dite **sans mémoire** au sens où pour tous réels positifs s et t , on a $\mathbf{P}(X > t + s | X > s) = \mathbf{P}(X > t)$. C'est pour cette raison qu'elle est utilisée généralement pour modéliser des temps d'attente entre deux événements : par exemple entre deux pannes successives d'une machine, ou entre deux requêtes reçues par un serveur informatique.

Loi normale, ou loi de Gauss centrée réduite. Il s'agit de la loi de la variable aléatoire X de densité $f(x) = e^{-x^2/2}/\sqrt{2\pi}$. C'est une loi très utilisée en statistique et en modélisation. Nous allons commencer par vérifier que c'est bien la densité d'une probabilité : f est une fonction positive, il reste à voir que $I = \int_{\mathbf{R}} f(t) dt = 1$. Pour cela, calculons I^2 . On a

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{+\infty} f(t) dt \right)^2 \\ &= \left(\int_{-\infty}^{+\infty} f(t) dt \right) \times \left(\int_{-\infty}^{+\infty} f(s) ds \right) \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(s) ds \right) f(t) dt \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(s)f(t) ds dt \\ &= \int_{\mathbf{R}^2} e^{-(s^2+t^2)/2} \frac{ds dt}{2\pi}. \end{aligned}$$

Procédons à un changement de variables en coordonnées polaires en posant $s = r \cos \theta$ et $t = r \sin \theta$. Il vient

$$\begin{aligned} I^2 &= \int_0^\infty \int_{-\pi}^\pi e^{-r^2/2} \frac{r dr d\theta}{2\pi} \\ &= \int_0^\infty r e^{-r^2/2} dr \\ &= 1. \end{aligned}$$

La loi normale $\mathcal{N}(m, \sigma^2)$ est de densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \left(\frac{(x - m)^2}{2\sigma^2} \right)$$

2 Espérance, variance, fonctions génératrice et caractéristique

2.1 Espérance et variance

On se donne une variable aléatoire $X : \Omega \rightarrow \mathbf{R}$.

Définition 2.2 Une variable aléatoire X est dite **intégrable** si la quantité

$$\int_{\Omega} |X| d\mathbf{P} = \int_{\mathbf{R}} |x| d\mu$$

est finie. On définit alors son **espérance** par

$$\mathbf{E}(X) = \int_{\Omega} X d\mathbf{P} = \int_{\mathbf{R}} x d\mu.$$

Plus généralement, pour toute fonction borélienne $h : \mathbf{R} \rightarrow \mathbf{R}$, on a

$$\mathbf{E}(h(X)) = \int_{\Omega} h(X) d\mathbf{P} = \int_{\mathbf{R}} h(x) d\mu,$$

lorsque la quantité

$$\int_{\Omega} |h(X)| d\mathbf{P} = \int_{\mathbf{R}} |h(x)| d\mu < \infty.$$

Dans le langage courant (et aussi probabiliste), l'espérance est appelée **moyenne**.

Si X est une variable aléatoire **discrète**, l'intégrabilité se traduit par

$$\sum_{x \in X(\Omega)} |x| \mathbf{P}(X = x) < \infty$$

et on a alors :

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

Plus généralement, pour toute fonction $h : X(\Omega) \rightarrow \mathbf{R}$

$$\mathbf{E}(h(X)) = \sum_{x \in X(\Omega)} h(x) \mathbf{P}(X = x)$$

si $h(X)$ est intégrable, c'est-à-dire, si

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| \mathbf{P}(X = x) < \infty.$$

Si X est une variable aléatoire à valeurs réelle et de **densité** $f : \mathbf{R} \rightarrow \mathbf{R}^+$, elle est intégrable si

$$\int_{\mathbf{R}} |x| f(x) dx < \infty$$

et on a dans ce cas

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

Plus généralement, pour toute fonction borélienne $h : \mathbf{R} \rightarrow \mathbf{R}$, la variable aléatoire $h(X)$ est intégrable, si

$$\mathbf{E}|h(X)| = \int_{\mathbf{R}} |h(x)| f(x) dx < \infty$$

et on a alors

$$\mathbf{E}(h(X)) = \int_{\mathbf{R}} h(x)f(x) dx.$$

Pourquoi l'intégrabilité ? Les séries (et les intégrales) qui ne convergent pas absolument ne sont pas facilement manipulables. Par exemple, la série

$$\sum_{n \in \mathbf{Z}^*} \frac{1}{n}$$

est mal définie. En effet, les sommes partielles de $-N$ à $+N$ sont nulles, et donc on a plutôt envie de dire que la série converge vers 0. Mais considérons la somme S_N de $-N$ à $+N^2$, on a

$$\begin{aligned} S_N &= \sum_{n=N+1}^{n=N^2} \frac{1}{n} \\ &\geq \int_{N+1}^{N^2+1} \frac{1}{x} dx \\ &\geq \ln(N^2 + 1) - \ln(N + 1) \end{aligned}$$

et donc S_N tend vers l'infini...

En pratique, on vérifie donc **toujours** qu'une variable aléatoire est intégrable avant de l'intégrer, sauf si $\mathbf{P}(X \in \mathbf{R}^+) = 1$.

Remarque :

Si $X(\Omega) \subset \mathbf{R}^+$, on pourra écrire $\mathbf{E}(X)$ même lorsque la série ne converge pas, mais on ne pourra évidemment pas dire que X est intégrable avant d'avoir vérifié si $\mathbf{E}(X) < \infty$.

Exemple : Calcul de l'espérance de la loi géométrique de paramètre p . On se donne une variable aléatoire X de loi $\mathcal{G}(p)$. On a vu que $X(\Omega) = \mathbf{N}^*$: il est inutile de vérifier l'intégrabilité de X puisque $X(\Omega) \subset \mathbf{R}^+$.

On a

$$\begin{aligned} \mathbf{E}(X) &= \sum_{n \geq 1} n \mathbf{P}(X = n) \\ &= \sum_{n \geq 1} np(1-p)^{n-1} \\ &= p \sum_{n \geq 0} n(1-p)^{n-1} \end{aligned}$$

Or $n(1-p)^{n-1}$ est la dérivée de $-(1-p)^n$ et $\sum_{n \geq 0} (1-p)^n = 1/p$. On obtient :

$$\sum_{n \geq 0} n(1-p)^{n-1} = -\frac{d}{dp} \frac{1}{p} = \frac{1}{p^2}.$$

Finalement, $\mathbf{E}(X) = 1/p$.

Exemple : Calcul de l'espérance d'une variable aléatoire Y de loi exponentielle, c'est-à-dire de densité $f(y) = \lambda \exp(-\lambda y) \mathbf{1}_{\mathbf{R}^+}(y)$ où $\lambda > 0$ est une constante. On constate que Y est une

variable aléatoire positive (l'intégrale de sa densité sur \mathbf{R}^- est nulle). On peut donc se passer de vérifier l'intégrabilité, et passer directement au calcul de l'espérance.

On a par définition

$$\mathbf{E}(Y) = \int_{\mathbf{R}} yf(y) dy = \int_{\mathbf{R}^+} \lambda y e^{-\lambda y} dy.$$

Cette intégrale s'intègre par parties :

$$\mathbf{E}(Y) = \left[y \times (-\exp(-\lambda y)) \right]_0^{+\infty} - \int_0^{+\infty} (-\exp(-\lambda y)) dy = \frac{1}{\lambda}.$$

Exercice : Calculer l'espérance des lois des paragraphes 1.2 et 1.3.

Proposition 2.3 • Soient a une constante et X une variable aléatoire intégrable. Les variables aléatoires $X + a$ et aX sont intégrables et on a $\mathbf{E}(X + a) = a + \mathbf{E}(X)$ et $\mathbf{E}(aX) = a\mathbf{E}(X)$.

• Si X est une variable aléatoire positive et intégrable, alors $\mathbf{E}(X) \geq 0$.

• Si X et Y sont deux variables aléatoires intégrables vérifiant $X \leq Y$ alors $\mathbf{E}(X) \leq \mathbf{E}(Y)$.

On déduit de cette propriété que toute variable aléatoire bornée par une constante (ou plus généralement par une variable aléatoire intégrable) est intégrable.

• L'espérance est une opération linéaire : si X et Y sont deux variables aléatoires intégrables et a et b deux nombres réels, $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$

Exercice :

• On se donne un événement A . Montrer que $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$.

• On considère une variable aléatoire X et un réel x . Calculer $\mathbf{E}(\mathbf{1}_{X \leq x})$.

Définition 2.4 La variance d'une variable aléatoire de carré intégrable X est égale à

$$\text{var } X = \mathbf{E}((X - \mathbf{E}(X))^2).$$

On appelle écart-type la quantité $\sigma = \sqrt{\text{var}(X)}$.

Proposition 2.5 – Pour toute variable aléatoire de carré intégrable X , on a $\text{var } X = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$.

– Si X est une variable aléatoire intégrable et si a est une constante, on a $\text{var}(X + a) = \text{var}(X)$ et $\text{var}(aX) = a^2 \text{var}(X)$.

– La variance d'une variable aléatoire de carré intégrable est toujours une quantité positive. Elle n'est nulle que si la variable aléatoire suit une loi de Dirac.

Preuve :

1) Notons $m = \mathbf{E}(X)$. On a $(X - m)^2 = X^2 - 2mX + m^2$. Donc

$$\begin{aligned} \mathbf{E}(X - m)^2 &= \mathbf{E}(X^2) - \mathbf{E}(2mX) + \mathbf{E}(m^2) \\ &= \mathbf{E}(X^2) - 2m\mathbf{E}(X) + m^2 \\ &= \mathbf{E}(X^2) - m^2. \end{aligned}$$

2) On pose $Y = X + a$. On a $\mathbf{E}(Y) = a + \mathbf{E}(X)$ et $\text{var}(Y) = \mathbf{E}((Y - \mathbf{E}(Y))^2)$.

D'où $\text{var}(Y) = \mathbf{E}((X - \mathbf{E}(X))^2) = \text{var}(X)$.

Posons aussi $Z = aX$. On a $\mathbf{E}(Z) = a\mathbf{E}(X)$ et $\text{var}(Z) = \mathbf{E}((aX - \mathbf{E}(Z))^2)$.

D'où $\text{var}(Y) = a^2\mathbf{E}((X - \mathbf{E}(X))^2) = a^2\text{var}(X)$.

3) La variance est l'espérance d'une variable aléatoire positive : elle est donc positive. Elle ne peut être nulle que si $X = \mathbf{E}(X)$ p.s.

Exemple : Calcul de la variance d'une variable aléatoire X de loi $\mathcal{G}(p)$. On a déjà vu que $\mathbf{E}(X) = 1/p$. Calculons maintenant $\mathbf{E}(X^2)$. On a :

$$\begin{aligned}\mathbf{E}(X^2) &= \sum_{n \geq 1} n^2 \mathbf{P}(X = n) \\ &= \sum_{n \geq 1} n^2 p (1-p)^{n-1}\end{aligned}$$

On va essayer de faire apparaître une dérivée seconde en écrivant $n^2 = n(n-1) + n$:

$$\begin{aligned}\mathbf{E}(X^2) &= p(1-p) \sum_{n \geq 1} n(n-1)(1-p)^{n-2} + p \sum_{n \geq 1} n(1-p)^{n-1} \\ &= p(1-p) \frac{d^2}{dp^2} \left(\sum_{n \geq 0} (1-p)^n \right) + \frac{1}{p} \\ &= p(1-p) \frac{d^2}{dp^2} \frac{1}{p} + \frac{1}{p} \\ &= \frac{2p(1-p)}{p^3} + \frac{1}{p} \\ &= \frac{2-p}{p^2}\end{aligned}$$

On en déduit maintenant la variance de X : $\text{var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = (1-p)/p^2$.

Exemple : Calcul de la variance d'une variable aléatoire Y de loi exponentielle de paramètre λ . On a vu (cf p.10) que l'espérance de Y est égale à $1/\lambda$. Calculons maintenant $\mathbf{E}(Y^2)$:

$$\begin{aligned}\mathbf{E}(Y^2) &= \int_{\mathbf{R}} x^2 \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0} dx \\ &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx \\ &= [-x^2 e^{-\lambda x}]_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= 2/\lambda^2\end{aligned}$$

On a donc $\text{var}(Y) = 1/\lambda^2$.

Exercice : Calculer les variances des lois des paragraphes 1.2 et 1.3.

2.2 Fonction de répartition de la loi d'une variable aléatoire

Définition 2.6 On considère une variable aléatoire $X : \Omega \rightarrow \mathbf{R}$. La fonction de répartition de la loi de X est la fonction $F_X : \mathbf{R} \rightarrow [0, 1]$ définie pour tout $x \in \mathbf{R}$ par

$$F_X(x) = \mathbf{P}(X \leq x).$$

La fonction de répartition caractérise la loi d'une variable aléatoire.

Si X suit une loi discrète, on a

$$F_X(x) = \sum_{t \leq x} \mathbf{P}(X = t),$$

où la somme est prise sur tous les $t \in X(\Omega)$ inférieurs ou égaux à x . On obtient une fonction constante par morceaux.

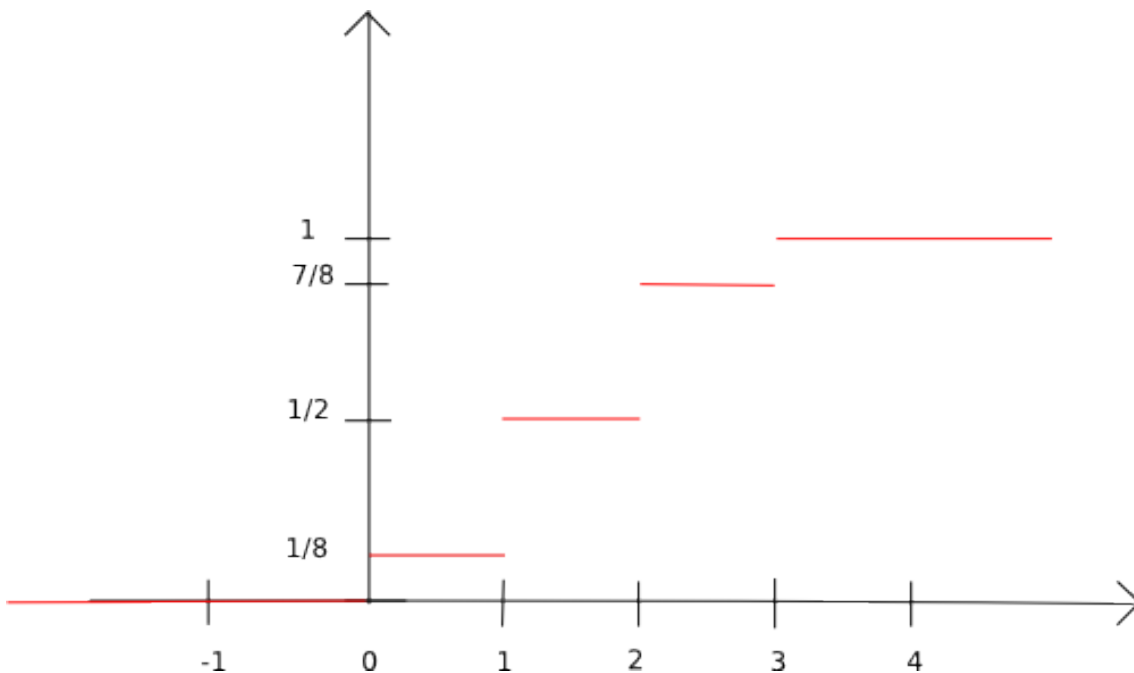
Si X suit la loi de densité f ,

$$F_X(x) = \int_{-\infty}^x f(t) dt.$$

C'est alors une fonction continue.

Proposition 2.7 La fonction de répartition d'une variable aléatoire discrète est toujours croissante et continue à droite. On a $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Exemple : Fonction de répartition de la loi binômiale $\mathcal{B}(3, 1/2)$. On a $\mathbf{P}(0) = 1/8$, $\mathbf{P}(1) = 3/8$, $\mathbf{P}(2) = 3/8$, $\mathbf{P}(3) = 1/8$, et $\mathbf{P}(k) = 0$ pour tout $k \geq 4$.



2.3 Fonctions génératrice, caractéristique et transformée de Laplace

Définition 2.8 La fonction génératrice de la loi de la variable aléatoire $X : \Omega \rightarrow \mathbf{R}^+$ est la fonction G_X définie pour tout $s \in [0, 1]$ par

$$G_X(s) = \mathbf{E}(s^X) = \sum_{x \in X(\Omega)} s^x \mathbf{P}(X = x).$$

Cette fonction est définie (et finie) pour $s \in [0, 1]$ puisque pour tout $s \in [0, 1]$, la variable aléatoire s^X est positive et majorée par 1 : c'est donc une variable aléatoire intégrable.

La fonction génératrice est surtout utilisée pour les variables aléatoires à valeurs dans \mathbf{N} . Dans ce cas, on peut également la définir pour $s \in [-1, 0]$ et on a

$$G_X(s) = \mathbf{E}(s^X) = \sum_{n \in \mathbf{N}} s^n \mathbf{P}(X = n).$$

Remarque : La fonction G_X est polynômiale si et seulement si $X(\Omega) \subset \{0, \dots, n\}$. Plus généralement, si $X(\Omega) \in \mathbf{N}$, G_X est une série entière.

Proposition 2.9

- Soit X une variable aléatoire discrète positive et intégrable. On a $\mathbf{E}(X) = G'_X(1^-)$.
- Plus généralement, si $X : \Omega \rightarrow \mathbf{R}^+$ est une variable aléatoire vérifiant $\mathbf{E}(X^n) < \infty$, la dérivée $n^{\text{ième}}$ de G_X en 1^- est égale à $\mathbf{E}(X(X-1)\dots(X-n+1))$.
- Si deux variables aléatoires positives ont la même fonction génératrice, alors elles suivent la même loi. On dit que la fonction génératrice caractérise la loi des variables aléatoires positives.

On utilise également la **transformée de Laplace** $\mathcal{L}_X(\lambda) = \mathbf{E}(\exp(-\lambda X)) = G_X(\exp(-\lambda))$. Si X est une variable aléatoire positive, sa transformée de Laplace est définie sur \mathbf{R}^+ (au moins) et caractérise sa loi. Si X est intégrable, on a $\mathbf{E}(X) = -\mathcal{L}'_X(0+)$, et si X est suffisamment intégrable, $\mathbf{E}(X^k) = (-1)^k \mathcal{L}_X^{(k)}(0+)$.

Pour des variables aléatoires $X : \Omega \rightarrow \mathbf{R}$ dont le signe n'est pas constant, on utilisera plutôt la **transformée de Fourier** ou **fonction caractéristique** de $X : \phi_X(\alpha) = \mathbf{E}(\exp(i\alpha X))$. Cette fonction est définie pour tout α dans \mathbf{R} et caractérise la loi de la variable aléatoire de la même façon que la fonction génératrice ou la transformée de Laplace. Tout comme pour la fonction génératrice ou pour la transformée de Laplace, il existe un lien entre les dérivées de ϕ et l'espérance de X : si X est intégrable, on a $\mathbf{E}(X) = i\phi'_X(0)$ et si X est suffisamment intégrable, $\mathbf{E}(X^k) = (-i)^k \phi_X^{(k)}(0)$.

Remarque : En pratique, il est parfois rentable de calculer la fonction génératrice d'une variable aléatoire (ou transformée de Laplace ou transformée de Fourier), et de la dériver pour en déduire l'espérance et la variance.

Aide-mémoire : La dérivée de $s \rightarrow s^X$ est égale à Xs^{X-1} . En prenant l'espérance pour $s = 1$, on obtient donc $G'_X(1) = \mathbf{E}(X)$.

De même, la dérivée seconde de $s \rightarrow s^X$ est égale à $X(X-1)s^{X-2}$, et après la même opération que précédemment, $G''_X(1) = \mathbf{E}(X(X-1))$.

En faisant le même genre d'opération à partir des fonctions $\lambda \rightarrow \exp(-\lambda X)$ ou $\alpha \rightarrow \exp(i\alpha X)$, on retrouve les liens entre espérance et dérivées des transformées de Laplace et de Fourier.

Exemple : Calcul de l'espérance et de la variance d'une loi de Poisson à partir de la fonction génératrice. Soit X une variable aléatoire de loi de Poisson de paramètre $m > 0$. Calculons sa fonction génératrice G_X .

Comme X est à valeurs dans \mathbf{N} , on a

$$\begin{aligned} G_X(s) &= \mathbf{E}(s^X) \\ &= \sum_{n \geq 0} s^n \mathbf{P}(X = n) \\ &= \sum_{n \geq 0} s^n e^{-m} \frac{m^n}{n!} \\ &= e^{-m} e^{ms} \end{aligned}$$

En dérivant, on obtient, $\mathbf{E}(X) = G'_X(1) = e^{-m} m e^{m \times 1} = m$.

En dérivant à nouveau : $\mathbf{E}(X(X-1)) = G''_X(1) = e^{-m} m^2 e^{m \times 1} = m^2$.

On en déduit alors la variance de X : par définition, $\text{var } X = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$ et en écrivant $\mathbf{E}(X^2) = \mathbf{E}(X(X-1)) + \mathbf{E}(X)$, on obtient $\text{var}(X) = m$.

2.4 Comment calculer la loi

Le problème se pose fréquemment de calculer la loi d'une variable aléatoire Y définie par exemple comme fonction d'une autre variable aléatoire X : $Y = g(X)$, la fonction g étant borélienne.

Si Y est une variable discrète, il suffit de déterminer l'ensemble des valeurs prises par Y , puis pour tout $y \in Y(\Omega)$, on aura $\mathbf{P}(Y = y) = \mathbf{P}(X \in g^{-1}(y))$, ce qui se calcule donc à l'aide de la loi de X .

Si on pense que Y va avoir une densité, on peut imaginer calculer sa fonction de répartition : $\mathbf{P}(Y \leq y) = \mathbf{P}(g(X) \leq y)$. Si la fonction g est monotone, on imagine facilement la suite... mais sinon ? La méthode habituellement utilisée consiste à utiliser une fonction test $h : \mathbf{R} \rightarrow \mathbf{R}$ borélienne et bornée. Si on réussit alors à écrire $\mathbf{E}(h(Y))$ sous la forme

$$\mathbf{E}(h(Y)) = \int_{\mathbf{R}} h(y) d\mu(y),$$

on aura gagné : la mesure de probabilité μ obtenue sera la mesure image de Y (de la forme $d\mu = f_Y(y) dy$ si Y est à densité). En effet, en prenant h de la forme $h(y) = \mathbf{1}_{y \leq x}$, on retrouve ainsi la fonction de répartition de la loi de Y . L'avantage est que le changement de variable auquel il faut procéder apparaît clairement.

Exemple : Soit X une variable aléatoire de densité f_X . Déterminons les lois de $Y = aX + b$ et de $Z = (1 + X)/(1 - X)$. Prenons donc une fonction test h borélienne et bornée. On a

$$\mathbf{E}(h(Y)) = \mathbf{E}(h(aX + b)) = \int_{\mathbf{R}} h(ax + b) f_X(x) dx.$$

On effectue alors le changement de variables $y = ax + b$. Ce changement de variables est bien un bi-difféomorphisme de \mathbf{R} et son jacobien vaut $dx = dy/|a|$. Il vient :

$$\mathbf{E}(h(Y)) = \int_{\mathbf{R}} h(y) f_X\left(\frac{y-b}{a}\right) \frac{dy}{|a|}.$$

On en déduit donc que Y admet pour densité la fonction f_Y définie sur \mathbf{R} par

$$f_Y(Y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}.$$

Procédons de même pour Z :

$$\begin{aligned}\mathbf{E}(h(Z)) &= \int_{\mathbf{R}} h\left(\frac{1+x}{1-x}\right) f_X(x) dx \\ &= \int_{\mathbf{R}} h(z) f_X\left(\frac{z-1}{z+1}\right) \frac{2}{(z+1)^2} dz\end{aligned}$$

On a en effet posé $z = (1+x)/(1-x)$, soit $x = (z-1)/(z+1) = 1 - 2/(z+1)$. Ce changement de variable est un bi-difféomorphisme de $\mathbf{R} \setminus \{-1\}$ vers $\mathbf{R} \setminus \{1\}$. Remarquons qu'il est souvent de raisonner un terme de jacobien plutôt qu'en terme de dérivée : il n'y a pas à s'occuper du sens des bornes des intégrales, mais il suffit de vérifier qu'un domaine s'envoie bien sur un autre domaine de \mathbf{R} (et de ne pas oublier de mettre la valeur absolue de la dérivée et non la dérivée elle-même). On traite alors ce changement de variable comme un changement de variable dans une intégrale multiple.

Chapitre 3

Couples et vecteurs de variables aléatoires

1 Couples et vecteurs aléatoires discrets

1.1 Loi conjointe

On se donne X et Y deux variables aléatoires discrètes avec $X(\Omega) = \{x_i, i \in \mathbf{N}\}$ et $Y(\Omega) = \{y_i, i \in \mathbf{N}\}$. La **loi conjointe** du couple (X, Y) est donnée par $(X, Y)(\Omega)$ (ou par $X(\Omega)$ et $Y(\Omega)$) ainsi que par les probabilités

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}\{\omega, X(\omega) = x \text{ et } Y(\omega) = y\}, \text{ pour tout couple } (x, y) \in (X, Y)(\Omega).$$

Remarque : On doit bien entendu avoir $\sum_{x,y} \mathbf{P}(X = x, Y = y) = 1$.

Plus généralement, si X_1, \dots, X_n sont n variables aléatoires discrètes à valeurs dans \mathbf{N} , la loi conjointe du vecteur (X_1, \dots, X_n) est donnée par $(X_1, \dots, X_n)(\Omega) \subset \mathbf{N}^n$ ainsi que par les probabilités $\mathbf{P}(X_1 = i_1, \dots, X_n = i_n)$, pour tout n -uplet $(i_1, \dots, i_n) \in \mathbf{N}^n$.

Exemple 1 : Fixons $p \in]0, 1[$ et $\lambda > 0$ et considérons le couple de variables aléatoires (X, Y) à valeurs dans $\{0, 1\} \times \mathbf{N}$ dont la loi est donnée par :

$$\begin{aligned} \mathbf{P}(X = 0, Y = 0) &= 1 - p \\ \mathbf{P}(X = 1, Y = k) &= pe^{-\lambda} \lambda^k / k!, \text{ pour tout } k \in \mathbf{N} \\ \mathbf{P}(X = j, Y = k) &= 0 \text{ sinon.} \end{aligned}$$

On a bien $\sum_{i,j} \mathbf{P}(X = i, Y = j) = 1$: on a donc bien écrit la loi d'un couple aléatoire discret.

Exemple 2 : On dispose d'une urne contenant quatre jetons numérotés de 1 à 4, et on tire au sort successivement deux jetons sans remise. On note (X, Y) les résultats des deux tirages.

On a : $\mathbf{P}(X = i, Y = i) = 0$ pour tout i entre 1 et 4 et $\mathbf{P}(X = i, Y = j) = 1/12$ si $1 \leq i, j \leq 4$ et $i \neq j$.

On peut écrire les probabilités sous la forme du tableau suivant (où par exemple dans la deuxième case de la première ligne, on lit $\mathbf{P}(X = 1, Y = 2)$) :

	1	2	3	4
1	0	1/12	1/12	1/12
2	1/12	0	1/12	1/12
3	1/12	1/12	0	1/12
4	1/12	1/12	1/12	0

Exemple 3 : Loi trinomiale. On se fixe un nombre entier n strictement positif et deux paramètres réels positifs p_x et p_y tels que $p_x + p_y \leq 1$. La loi trinomiale (n, p_x, p_y) est la loi du couple (X, Y) tel que $(X, Y)(\Omega) \in \mathbf{N}^2$ et donnée pour tout $(i, j) \in \mathbf{N}^2$ tels que $i + j \leq n$ par :

$$\mathbf{P}(X = i, Y = j) = \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1 - p_x - p_y)^{n-i-j},$$

et $\mathbf{P}(X = i, Y = j) = 0$ sinon.

1.2 Lois marginales

Définition 3.1 Les (deux) lois marginales du couple (X, Y) sont les lois des variables aléatoires X et Y . On les obtient de la façon suivante :

$$\begin{aligned} \mathbf{P}(X = x) &= \sum_{y \in Y(\Omega)} \mathbf{P}(X = x, Y = y), \\ \mathbf{P}(Y = y) &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x, Y = y). \end{aligned}$$

Preuve : On a

$$\{X = x\} = \{X = x, Y \in Y(\Omega)\} = \bigcup_{y \in Y(\Omega)} \{X = x, Y = y\}.$$

Comme la réunion est dénombrable et disjointe, il vient :

$$\mathbf{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbf{P}(X = x, Y = y).$$

Plus généralement, un vecteur (X_1, \dots, X_n) à valeurs dans \mathbf{Z}^n possède n lois marginales unidimensionnelles, mais également $n(n-1)$ lois marginales bidimensionnelles, et ainsi de suite. On a par exemple

$$\mathbf{P}(X_1 = x) = \sum_{(x_2, \dots, x_n) \in \mathbf{Z}^{n-1}} \mathbf{P}(X_1 = x, X_2 = x_2, \dots, X_n = x_n).$$

Reprenons les exemples précédents :

Exemple 1.

Déterminons la loi de X : X est à valeurs dans $\{0, 1\}$ et on a :

$$\mathbf{P}(X = 0) = \sum_{j \in \mathbf{N}} \mathbf{P}(X = 0, Y = j) = 1 - p.$$

De la même façon :

$$\mathbf{P}(X = 1) = \sum_{j \in \mathbf{N}} \mathbf{P}(X = 1, Y = j) = \sum_{j \geq 0} p e^{-\lambda} \lambda^j / j! = p.$$

La variable aléatoire X suit donc une loi de Bernoulli de paramètre p . Calculons aussi la loi de Y :

$$\mathbf{P}(Y = 0) = \mathbf{P}(X = 0, Y = 0) + \mathbf{P}(X = 1, Y = 0) = 1 - p + p e^{-\lambda}.$$

Et pour tout $j \geq 1$,

$$\mathbf{P}(Y = j) = \mathbf{P}(X = 0, Y = j) + \mathbf{P}(X = 1, Y = j) = p e^{-\lambda} \lambda^j / j!.$$

Exemple 2 : Il suffit de sommer en colonne pour avoir la loi de X , et en ligne pour obtenir celle de Y . En pratique, on peut ajouter une colonne et une ligne au tableau pour y écrire les lois de X et Y . Et avant de conclure, on prend le soin de vérifier que la somme de cette colonne (et de cette ligne) vaut 1.

On trouve ici que X et Y suivent une loi uniforme sur $\{1, 2, 3, 4\}$.

Exemple 3 : On considère le couple (X, Y) de loi trinomiale (n, p_x, p_y) . Déterminons la loi marginale de X : fixons $j \in \{0, \dots, n\}$ et évaluons $\mathbf{P}(X = j)$.

On a

$$\begin{aligned} \mathbf{P}(X = j) &= \sum_{k=0}^n \mathbf{P}(X = j, Y = k) \\ &= \sum_{k=0}^{n-j} \mathbf{P}(X = j, Y = k) + \sum_{k=n-j+1}^n \mathbf{P}(X = j, Y = k) \\ &= \sum_{k=0}^{n-j} \frac{n!}{j! k! (n-j-k)!} p_x^j p_y^k (1-p_x-p_y)^{n-j-k} + 0 \\ &= \frac{n!}{j! (n-j)!} p_x^j \sum_{k=0}^{n-j} \frac{(n-j)!}{k! (n-j-k)!} p_y^k (1-p_x-p_y)^{n-j-k} \\ &= C_n^j p_x^j (1-p_x)^{n-j} \end{aligned}$$

La variable aléatoire X suit donc une loi $\text{Bin}(n, p_x)$. Un calcul similaire montre que Y suit une loi binomiale $\text{Bin}(n, p_y)$.

1.3 Loi de $f(X, Y)$

Problème : On dispose d'un couple de variables aléatoires discrètes (X, Y) dont on connaît la loi conjointe et on voudrait connaître la loi de la variable aléatoire $Z = f(X, Y)$, où $f : X(\Omega) \times Y(\Omega) \rightarrow \mathbf{R}$ est une fonction donnée. Par exemple, on a souvent besoin de connaître la loi de $X + Y$, ou celle de $X - Y$, ou de XY . Et déterminer la loi de X à partir de celle de (X, Y) revient à considérer la fonction $f(x, y) = x$.

Proposition 3.2 On a $Z(\Omega) = f((X, Y)(\Omega))$ et pour tout $z \in f((X, Y)(\Omega))$, on a

$$\mathbf{P}(Z = z) = \sum_{(x,y) \in (X,Y)(\Omega), f(x,y)=z} \mathbf{P}(X = x, Y = y).$$

Exemple : Reprenons une nouvelle fois l'exemple 1 et considérons la fonction $f(x, y) = xy$. La variable aléatoire XY est à valeurs dans \mathbf{N} et on a

$$\mathbf{P}(XY = 0) = \mathbf{P}(X = 0, Y = 0) + \mathbf{P}(X = 1, Y = 0) = 1 - p + pe^{-\lambda}$$

et, pour tout $k \in \mathbf{N}^*$,

$$\mathbf{P}(XY = k) = \mathbf{P}(X = 1, Y = k) = pe^{-\lambda} \frac{\lambda^k}{k!}.$$

Un cas particulier. Nous considérons ici la fonction $f(x, y) = x + y$. On obtient :

$$\begin{aligned} \mathbf{P}(X + Y = z) &= \sum_{(x,y) \in (X,Y)(\Omega), x+y=z} \mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x, Y = z - x) \\ &= \sum_{y \in Y(\Omega)} \mathbf{P}(X = z - y, Y = y) \end{aligned}$$

Plus généralement, si $\bar{X} = (X_1, \dots, X_n)$ est un vecteur aléatoire discret et $f : \bar{X}(\Omega) \rightarrow \mathbf{R}$ est une fonction donnée, on a :

$$\mathbf{P}(f(X_1, \dots, X_n) = z) = \sum_{x_1, \dots, x_n, f(x_1, \dots, x_n) = z} \mathbf{P}(X_1 = x_1, \dots, X_n = x_n).$$

Exemple : Supposons que (X, Y) suit une loi trinomiale (n, p_x, p_y) et calculons la loi de $X + Y$. Cette variable aléatoire est à valeurs dans \mathbf{N} et on a, pour tout entier k :

$$\mathbf{P}(X + Y = k) = \sum_{j=0}^n \mathbf{P}(X = j, Y = k - j).$$

Pour tout $k > n$, chacun des termes de cette somme est nul donc $\mathbf{P}(X + Y = k) = 0$.

Fixons maintenant un entier $k \in \{0, \dots, n\}$. On a :

$$\begin{aligned} \mathbf{P}(X + Y = k) &= \sum_{j=0}^k \mathbf{P}(X = j, Y = k - j) \\ &= \sum_{j=0}^k \frac{n!}{j!(k-j)!(n-k)!} p_x^j p_y^{k-j} (1 - p_x - p_y)^{n-k} \\ &= \frac{n!}{(n-k)!} (1 - p_x - p_y)^{n-k} \sum_{j=0}^k \frac{1}{j!(k-j)!} p_x^j p_y^{k-j} \\ &= \frac{n!}{k!(n-k)!} (1 - p_x - p_y)^{n-k} (p_x + p_y)^k. \end{aligned}$$

La variable aléatoire $X + Y$ suit donc une loi binomiale $\text{Bin}(n, p_x + p_y)$.

1.4 Loi conditionnelle

Considérons un couple (X, Y) de variables aléatoires discrètes, dont on connaît la loi jointe et fixons y tel que $\mathbf{P}(Y = y) > 0$.

La **loi conditionnelle** de X sachant l'événement $\{Y = y\}$ est donnée par le fait que c'est une loi sur $X(\Omega)$ ainsi que par les probabilités conditionnelles $\mathbf{P}(X = x|Y = y)$ pour tout $x \in X(\Omega)$.

On vérifie aisément que

$$\sum_{x \in X(\Omega)} \mathbf{P}(X = x|Y = y) = 1$$

ce qui implique que la loi conditionnelle de X sachant $\{Y = y\}$ est la loi d'une variable aléatoire.

De plus, la formule des probabilités totales implique que

$$\mathbf{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbf{P}(X = x|Y = y)\mathbf{P}(Y = y).$$

Cette définition de la loi conditionnelle s'étend à des vecteurs aléatoires : par exemple pour un triplet aléatoire (X, Y, Z) , on peut étudier la loi conditionnelle de X sachant $\{Y = y\}$, la loi conditionnelle de X sachant $\{Y = y$ et $Z = z\}$, la loi conditionnelle du couple (X, Y) sachant $\{Z = z\}$...

Exemple 1 : La loi de Y sachant $\{X = 1\}$ est donnée par $Y(\Omega) \in \mathbf{N}$) et, pour tout entier positif k , on a

$$\mathbf{P}(Y = k|X = 1) = \frac{\mathbf{P}(Y = k \text{ et } X = 1)}{\mathbf{P}(X = 1)} = pe^{-\lambda} \frac{\lambda^k}{k!} \times \frac{1}{p} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

La loi conditionnelle de Y sachant que $\{X = 1\}$ est donc une loi de Poisson de paramètre λ .

Exemple 2 : La loi de X sachant $\{Y = 1\}$ est la loi uniforme sur $\{2, 3, 4\}$.

Exemple 3 : Supposons que (X, Y) suit une loi trinomiale (n, p_x, p_y) et calculons la loi conditionnelle de X sachant $\{Y = k\}$, pour un entier $k \in \{0, \dots, n\}$. Remarquons tout d'abord que si $j < 0$ ou $j > n - k$, $\mathbf{P}(X = j|Y = k) = 0$. Fixons maintenant un entier $j \in \{0, \dots, n - k\}$. On a

$$\begin{aligned} \mathbf{P}(X = j|Y = k) &= \frac{\mathbf{P}(X = j \text{ et } Y = k)}{\mathbf{P}(Y = k)} \\ &= \frac{n! p_x^j p_y^k (1 - p_x - p_y)^{n-j-k}}{j! k! (n - j - k)!} \frac{k! (n - k)!}{n! p_y^k (1 - p_y)^{n-k}} \\ &= \frac{(n - k)!}{j! (n - k - j)!} \left(\frac{p_x}{1 - p_y} \right)^j \left(\frac{1 - p_x - p_y}{1 - p_y} \right)^{n-k-j}. \end{aligned}$$

La loi conditionnelle de X sachant $\{Y = k\}$ est donc une loi binomiale $(n - k, p_x/(1 - p_y))$.

1.5 Indépendance de variables aléatoires discrètes

Définition 3.3 Deux variables aléatoires discrètes X et Y sont dites **indépendantes** si pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants, c'est-à-dire : $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$.

Si X et Y sont deux variables aléatoires discrètes indépendantes, on aura donc, pour tout $y \in Y(\Omega)$ et tout $x \in X(\Omega)$ tels que $\mathbf{P}(Y = y) > 0$ et $\mathbf{P}(X = x) > 0$, $\mathbf{P}(X = x|Y = y) = \mathbf{P}(X = x)$ et $\mathbf{P}(Y = y|X = x) = \mathbf{P}(Y = y)$.

Plus généralement, les n variables aléatoires discrètes X_1, \dots, X_n sont (**mutuellement ou n à n**) **indépendantes** si, pour tout choix de $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$, on a $\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbf{P}(X_1 = x_1) \dots \mathbf{P}(X_n = x_n)$.

Remarques :

- L'indépendance n à n entraîne l'indépendance 2 à 2 (mais la réciproque est fautive). Écrire la preuve de ce résultat pour $n = 3$.

- Des événements A_1, \dots, A_n sont indépendants si et seulement si les variables aléatoires $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ le sont.

Proposition 3.4 Si les n variables aléatoires discrètes X_1, \dots, X_n sont indépendantes alors on a $\overline{X}(\Omega) = X_1(\Omega) \times \dots \times X_n(\Omega)$.

Remarque : Lorsque les variables aléatoires X et Y sont indépendantes, connaître les lois marginales permet donc de connaître la loi jointe du couple (X, Y) , alors que pour des variables aléatoires quelconques, cela ne suffit pas.

1.6 Espérance, matrice de covariance

Dans un souci de clarté, tous les résultats de ce paragraphe et du suivant sont énoncés pour les couples de variables aléatoires discrètes, et ils s'étendent sans peine aux vecteurs aléatoires discrets.

On considère un couple aléatoire discret (X, Y) .

Définition 3.5 • L'**espérance du couple** (X, Y) est définie si X et Y sont intégrables et on a alors : $\mathbf{E}(X, Y) = (\mathbf{E}(X), \mathbf{E}(Y))$.

Remarque : Cette opération ne fait intervenir que les lois marginales.

- Si X et Y sont deux variables aléatoires discrètes de carré intégrable, la covariance de X et de Y , ou covariance du couple (X, Y) , est donnée par

$$\text{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))].$$

- Si X et Y sont deux variables aléatoires de carré intégrable, la matrice de covariance du couple (X, Y) est la matrice

$$C = \begin{pmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \text{var}(Y) \end{pmatrix}.$$

Plus généralement, la matrice de covariance d'un vecteur (X_1, \dots, X_n) , dont chacune des composantes est de carré intégrable, est une matrice $n \times n$ dont les termes diagonaux sont les variances des X_i et dont le terme (i, j) est la covariance $\text{cov}(X_i, X_j)$ pour tout $i \neq j$.

Proposition 3.6 1. Une matrice de covariance C est toujours une matrice symétrique et positive (i.e., pour tout $v \in \mathbf{R}^n$, $\langle v, Cv \rangle \geq 0$).

2. Si X et Y sont deux variables aléatoires indépendantes et intégrables, on a $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ et donc $\text{cov}(X, Y) = 0$. La réciproque de ce résultat est fausse.
3. Si X et Y sont deux variables aléatoires indépendantes et f et g deux fonctions telles que les variables aléatoires $f(X)$ et $g(Y)$ sont intégrables, on a $\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y))$. La réciproque de ce résultat est fausse.
4. Si les variables aléatoires X_1, \dots, X_n sont indépendantes et de carré intégrable, alors la matrice de covariance de (X_1, \dots, X_n) est diagonale. La réciproque de ce résultat est fausse.

Preuve : 1. Soit $\bar{X} = (X_1, \dots, X_n)$ un vecteur aléatoire dont chaque composante est de carré intégrable, notons C sa matrice de covariance et fixons $\bar{v} = (v_1, \dots, v_n)$ un vecteur de \mathbf{R}^n .

On a

$$\begin{aligned}
 \langle v, Cv \rangle &= \sum_{i,j} C_{ij} v_i v_j \\
 &= \sum_{i,j} \mathbf{E}(X_i X_j) v_i v_j \\
 &= \mathbf{E} \left(\sum_{i,j} v_i X_i v_j X_j \right) \\
 &= \mathbf{E} \left(\left(\sum_i v_i X_i \right) \times \left(\sum_j v_j X_j \right) \right) \\
 &= \mathbf{E} \left(\left(\sum_i v_i X_i \right)^2 \right)
 \end{aligned}$$

On a donc bien $\langle v, Cv \rangle \geq 0$.

2. Considérons deux variables aléatoires discrètes X et Y indépendantes et intégrables. Montrons que XY est intégrable et calculons son espérance. On a

$$\mathbf{E}(|XY|) = \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| \mathbf{P}(X = x, Y = y)$$

Comme X et Y sont indépendantes, $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$ donc

$$\begin{aligned} \mathbf{E}(|XY|) &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| \mathbf{P}(X = x) \mathbf{P}(Y = y) \\ &= \sum_{x \in X(\Omega)} \left(|x| \mathbf{P}(X = x) \sum_{y \in Y(\Omega)} |y| \mathbf{P}(Y = y) \right) \\ &= \left(\sum_{x \in X(\Omega)} |x| \mathbf{P}(X = x) \right) \left(\sum_{y \in Y(\Omega)} |y| \mathbf{P}(Y = y) \right) \\ &= \mathbf{E}(|X|) \mathbf{E}(|Y|). \end{aligned}$$

On montre alors par un calcul similaire que $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$, puis on en déduit que $\text{cov}(X, Y) = 0$.

3. et **4.** se déduisent aisément de **2.**

La proposition suivante, bien que très simple à prouver, est fort utile :

Proposition 3.7 *Si X et Y sont deux variables aléatoires discrètes et de carré intégrable, on a*

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

Si de plus, X et Y sont indépendantes, on a

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

De plus, pour deux variables aléatoires de carré intégrable, il est possible d'obtenir une majoration de $\text{cov}(X, Y)$ à partir des variances de X et Y :

Proposition 3.8 (Inégalité de Cauchy Schwarz) *Soient X et Y deux variables aléatoires de carré intégrable. On a*

$$\begin{aligned} |\mathbf{E}(XY)| &\leq \mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)} \\ |\text{cov}(X, Y)| &\leq \sqrt{\text{var}(X)\text{var}(Y)} \end{aligned}$$

Preuve : Le deuxième point s'obtient en appliquant le premier à $X - \mathbf{E}(X)$ et $Y - \mathbf{E}(Y)$.

Le premier point se montre de la même façon que dans le cas classique (produit scalaire de deux vecteurs) : on étudie le polynôme

$$R(\lambda) = \mathbf{E}(X^2)\lambda^2 - 2\mathbf{E}|XY|\lambda + \mathbf{E}(Y^2).$$

Ce polynôme se factorise en

$$R(\lambda) = \mathbf{E}[(\lambda|X| - |Y|)^2].$$

Il n'admet donc pas deux racines réelles distinctes, ce qui signifie que son discriminant (réduit) est négatif :

$$(\mathbf{E}|XY|)^2 - \mathbf{E}(X^2)\mathbf{E}(Y^2) \leq 0.$$

On étudie de la même façon le cas d'égalité : le discriminant n'est nul que si le polynôme admet une racine réelle double. Dans ce cas, il existe donc λ_0 tel que

$$\mathbf{E}[(\lambda_0|X| + |Y|)^2] = 0;$$

l'espérance d'une variable aléatoire positive ne pouvant être nulle que si la variable aléatoire est (presque sûrement) nulle, on aura alors : $|Y| = -\lambda_0|X|$, presque sûrement.

1.7 Fonction génératrice d'un couple

Définition 3.9 La fonction génératrice d'un couple de variables aléatoires discrètes positives est la fonction définie sur $[0, 1]^2$ par

$$G_{(X,Y)}(s, t) = \mathbf{E}(s^X t^Y) = \sum_{x \in X(\Omega), y \in Y(\Omega)} s^x t^y \mathbf{P}(X = x, Y = y).$$

Proposition 3.10 La fonction génératrice détermine la loi du couple (X, Y) au sens où si deux couples de variables aléatoires positives ont la même fonction génératrice, alors ils suivent la même loi.

La fonction génératrice de (X, Y) permet de retrouver par exemple :

- la fonction génératrice de X : $G_X(s) = G_{(X,Y)}(s, 1)$ et de Y ,
- l'espérance de X si X est intégrable : $\mathbf{E}(X) = \frac{\partial}{\partial s} G_{(X,Y)}(1, 1)$,
- l'espérance de X^2 si X est de carré intégrable : $E(X^2) = \frac{\partial^2}{\partial s^2} G_{(X,Y)}(1, 1) + \frac{\partial}{\partial s} G_{(X,Y)}(1, 1)$,
- l'espérance de XY si X et Y sont de carré intégrable : $\mathbf{E}(XY) = \frac{\partial^2}{\partial s \partial t} G_{(X,Y)}(1, 1)$.

Une autre utilité importante de la fonction génératrice est de permettre de calculer simplement la loi de somme de variables aléatoires, à partir du moment où on connaît leur loi jointe. En effet, on a le résultat suivant :

Proposition 3.11 Soient (X, Y) un couple de variables aléatoires positives dont on connaît la loi jointe. Notons $G_{X,Y}$ la fonction génératrice du couple. On a alors $G_{X+Y}(s) = G_{(X,Y)}(s, s)$.

Preuve : On a $G_{(X,Y)}(s, s) = \mathbf{E}(s^X s^Y) = \mathbf{E}(s^{X+Y}) = G_{X+Y}(s)$.

Si X et Y sont à valeurs entières, il ne reste plus qu'à décomposer la fonction G_{X+Y} en série entière pour obtenir les probabilités $\mathbf{P}(X + Y = k)$.

Il est facile de voir que si X et Y sont deux variables aléatoires positives indépendantes, alors on a $G_{(X,Y)}(s, t) = G_X(s)G_Y(t)$. La réciproque de ce résultat est également vraie :

Proposition 3.12 Les variables aléatoires positives X et Y sont indépendantes si et seulement si $G_{(X,Y)}(s, t) = G_X(s)G_Y(t)$.

Remarque : La fonction génératrice du vecteur aléatoire $\bar{X} = (X_1, \dots, X_n)$ vérifiant $\bar{X}(\Omega) \subset \mathbf{N}^n$ est la fonction $G : [0, 1]^n \rightarrow \mathbf{R}^+$ définie par

$$G(s_1, \dots, s_n) = \mathbf{E}(s_1^{X_1} \times \dots \times s_n^{X_n}).$$

Cette fonction de n variables caractérise la loi du n -uplet, permet de retrouver les lois marginales ...

1.8 Exemple d'utilisation de la fonction génératrice

Calculons la fonction génératrice de la loi trinominale (n, p_x, p_y) .

Soit (X, Y) un couple de variable aléatoire de loi trinominale (n, p_x, p_y) .

On a

$$\begin{aligned}
 G_{(X,Y)}(s, t) &= \mathbf{E}(s^X t^Y) \\
 &= \sum_{i,j \in \{0, \dots, n\}, i+j \leq n} s^i t^j \mathbf{P}(X = i, Y = j) \\
 &= \sum_{i=0}^n \sum_{j=0}^{n-i} s^i t^j \frac{n!}{i! j! (n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\
 &= \sum_{i=0}^n \frac{n!}{i! (n-i)!} (s p_x)^i \sum_{j=0}^{n-i} \frac{(n-i)!}{j! (n-i-j)!} (t p_y)^j (1-p_x-p_y)^{n-i-j} \\
 &= \sum_{i=0}^n \frac{n!}{i! (n-i)!} (s p_x)^i (1-p_x - (1-t)p_y)^{n-i} \\
 &= (1 - (1-s)p_x - (1-t)p_y)^n.
 \end{aligned}$$

En prenant $t = 1$ dans le résultat ci-dessus, on trouve : $G_{(X,Y)}(s, 1) = (1 - (1-s)p_x)^n$: on reconnaît la fonction caractéristique d'une loi binomiale (n, p_x) ce qui prouve que X suit une loi binomiale (n, p_x) .

La fonction génératrice de (X, Y) permet également de calculer celle de $X + Y$. On a en effet :

$$\mathbf{E}(s^{X+Y}) = \mathbf{E}(s^X s^Y) = G_{(X,Y)}(s, s).$$

On trouve ici : $G_{X+Y}(s) = (1 - (1-s)(p_x + p_y))^n$: on retrouve donc que $X + Y$ suit une loi binomiale $(n, p_x + p_y)$.

1.9 Somme de deux variables aléatoires de loi de Poisson et indépendantes

Considérons deux variables aléatoires X et Y indépendantes et de loi de Poisson de paramètre respectivement $\mu_1 > 0$ et $\mu_2 > 0$, et calculons la loi de $X + Y$.

Première méthode

Commençons par calculer la fonction génératrice de la loi de X . On a

$$\begin{aligned}
 G_X(s) &= \mathbf{E}(s^X) \\
 &= \sum_{n \geq 0} s^n \mathbf{P}(X = n) \\
 &= \sum_{n \geq 0} s^n e^{-\mu_1} \frac{\mu_1^n}{n!} \\
 &= e^{-\mu_1} e^{\mu_1 s} \\
 &= e^{-\mu_1(1-s)}
 \end{aligned}$$

De la même façon, on a $G_Y(t) = e^{-\mu_2(1-t)}$. Comme X et Y sont des variables aléatoires indépendantes, on en déduit la fonction génératrice du couple (X, Y) :

$$G_{(X,Y)}(s, t) = e^{-\mu_1(1-s)} e^{-\mu_2(1-t)}.$$

La fonction génératrice de $X + Y$ est donc

$$G_{X+Y}(s) = G_{(X,Y)}(s, s) = e^{-(\mu_1+\mu_2)(1-s)}.$$

On reconnaît ici la fonction génératrice d'une loi de Poisson de paramètre $\mu_1 + \mu_2$ et on peut conclure que $X + Y$ suit une loi de Poisson de paramètre $\mu_1 + \mu_2$.

Attention : La variable aléatoire $X - Y$ ne suit pas une loi de Poisson : cette variable aléatoire prend dans \mathbf{Z} et non dans \mathbf{N} et, en conséquence, on ne peut pas calculer sa fonction génératrice.

Deuxième méthode

On a évidemment $(X + Y)(\Omega) \subset \mathbf{N}$. Fixons donc $n \in \mathbf{N}$ et calculons directement la probabilité $\mathbf{P}(X + Y = n)$.

On a

$$\begin{aligned} \mathbf{P}(X + Y = n) &= \sum_{k \geq 0} \mathbf{P}(X = k, X + Y = n) \\ &= \sum_{k \geq 0} \mathbf{P}(X = k, Y = n - k) \\ &= \sum_{k \geq 0} \mathbf{P}(X = k) \mathbf{P}(Y = n - k) \\ &= \sum_{k=0}^n e^{-\mu_1} \frac{\mu_1^k}{k!} e^{-\mu_2} \frac{\mu_2^{n-k}}{(n-k)!} \\ &= e^{-(\mu_1+\mu_2)} \frac{1}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \mu_1^k \mu_2^{n-k} \\ &= e^{-(\mu_1+\mu_2)} \frac{(\mu_1 + \mu_2)^n}{n!} \end{aligned}$$

On retrouve donc que $X + Y$ suit une loi de Poisson de paramètre $\mu_1 + \mu_2$.

2 Couples aléatoires à densité

2.1 Densité d'un couple

Définition 3.13 *Un couple (X, Y) de variables aléatoires réelles sera dit « à densité par rapport à la mesure de Lebesgue de \mathbf{R}^2 » s'il existe une fonction borélienne $f : \mathbf{R}^2 \rightarrow \mathbf{R}^+$ telle que, pour tous boréliens (ou intervalles) I et J et pour toute fonction borélienne bornée ou positive h on ait :*

$$\mathbf{P}(X \in I \text{ et } Y \in J) = \int_{I \times J} f(x, y) \, dx \, dy \quad \text{et} \quad \mathbf{E}(h(X, Y)) = \int_{\mathbf{R}^2} h(x, y) f(x, y) \, dx \, dy$$

Remarque : La densité f est toujours une fonction positive, mesurable et dont l'intégrale sur \mathbf{R}^2 par rapport à la mesure de Lebesgue vaut 1.

Exemples :

1. Soit $D = \{(x, y) \in \mathbf{R}^2, x^2 + y^2 \leq 1\}$ le disque unité de \mathbf{R}^2 et f la fonction définie sur \mathbf{R}^2 par : $f(x, y) = \mathbf{1}_D(x, y)/\pi$. f est bien une densité de probabilité sur \mathbf{R}^2 : c'est la densité de la loi uniforme sur D , c'est-à-dire la loi que l'on obtient en jetant un point au hasard et uniformément sur D .
2. Soit X une variable aléatoire de loi exponentielle. La loi du couple (X, X) n'admet pas de densité par rapport à la mesure de Lebesgue sur \mathbf{R}^2 . En effet, si une telle densité f existait, elle devrait être (presque partout) nulle en dehors de la droite $\{x = y\}$ puisque $\mathbf{P}(X \neq Y) = 0$. On aurait alors

$$1 = \int_{\mathbf{R}^2} f(x, y) \, dx \, dy = \int_{\{x=y\} \subset \mathbf{R}^2} f(x, y) \, dx \, dy.$$

Or une droite est de mesure nulle pour la mesure de Lebesgue sur \mathbf{R}^2 , donc l'intégrale, toujours par rapport à la mesure de Lebesgue sur \mathbf{R}^2 , de n'importe quelle fonction sur ce domaine est nulle. On aboutit donc à une contradiction.

On définit, comme pour les couples discrets l'espérance d'un couple aléatoires intégrable comme étant le couple des espérances, et la matrice de covariance si les variables aléatoires sont de carré intégrable. Le terme (1,2) (et (2,1)) de cette matrice sera $\text{cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ où

$$\mathbf{E}(XY) = \int_{\mathbf{R}^2} xyf(x, y) \, dx \, dy.$$

La matrice de covariance est comme dans le cas discret une matrice symétrique et positive (au sens des formes bilinéaires).

Ces définitions s'étendent naturellement au cas des vecteurs aléatoires à densité, en remplaçant les intégrales doubles par des intégrales multiples.

Pour identifier la densité d'un couple, on utilise habituellement une fonction test $h : \mathbf{R}^2 \rightarrow \mathbf{R}$ borélienne et bornée et on essaie d'écrire $\mathbf{E}(h(X, Y))$ sous la forme

$$\mathbf{E}(h(X, Y)) = \int_{\mathbf{R}^2} h(x, y)f(x, y) \, dx \, dy.$$

La fonction f ci-dessus, si elle existe, sera la densité du couple (X, Y) .

2.2 Loi marginale

Les lois marginales du couple (X, Y) sont les lois des variables aléatoires X et Y . On peut vérifier aisément la proposition suivante :

Proposition 3.14 *Soit (X, Y) un couple de variables aléatoires de densité (conjointe) $f : \mathbf{R}^2 \rightarrow \mathbf{R}^+$. Alors les deux variables aléatoires X et Y ont chacune une densité, notée respectivement f_X et f_Y données par :*

$$f_X(x) = \int_{y \in \mathbf{R}} f(x, y) \, dy \quad \text{et} \quad f_Y(y) = \int_{x \in \mathbf{R}} f(x, y) \, dx.$$

Preuve : En effet, soit $t \in \mathbf{R}$ et calculons la fonction de répartition de X . On a

$$\begin{aligned} \mathbf{P}(X \leq t) &= \mathbf{P}(X \in]-\infty, t] \text{ et } Y \in \mathbf{R}) \\ &= \int_{]-\infty, t] \times \mathbf{R}} f(x, y) \, dx \, dy \\ &= \int_{-\infty}^t \left(\int_{y \in \mathbf{R}} f(x, y) \, dy \right) dx \end{aligned}$$

Ce principe s'applique aux vecteurs à densité, à ce près que les lois marginales d'un triplet sont les lois de chacune des variables qui le constituent ainsi que les lois des couples de variables.

Exercice : Calculer les lois marginales du couple (X, Y) de loi uniforme sur le disque D de centre 0 et de rayon 1.

2.3 Indépendance

Rappelons la définition de l'indépendance de variables aléatoires :

Définition 3.15 Les variables aléatoires X et Y sont **indépendantes** si, pour tous boréliens (ou intervalles) I et J , on a

$$\mathbf{P}(X \in I \text{ et } Y \in J) = \mathbf{P}(X \in I)\mathbf{P}(Y \in J).$$

Plus généralement, n variables aléatoires X_1, \dots, X_n sont **(mutuellement) indépendantes** si, pour tous boréliens (ou intervalles) I_1, \dots, I_n , on a

$$\mathbf{P}\left(\bigcap_{k \leq n} \{X_k \in I_k\}\right) = \prod_{k \leq n} \mathbf{P}\{X_k \in I_k\}.$$

Le lien entre densités et indépendance est clair :

Proposition 3.16 Deux variables aléatoires (X, Y) de densité respectivement f et g sont indépendantes si et seulement si la loi du couple admet une densité et que cette densité est la fonction $(x, y) \rightarrow f(x)g(y)$.

En particulier, on peut en déduire que si X et Y sont deux variables aléatoires indépendantes et intégrables, on aura $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$, **mais ce n'est pas une équivalence**. En effet, en reprenant l'exemple de la loi uniforme sur le disque de centre 0 et de rayon 1, on peut vérifier que $\mathbf{E}(XY) = 0 = \mathbf{E}(X)\mathbf{E}(Y)$ mais que ces variables aléatoires ne sont pas indépendantes.

2.4 Fonction caractéristique

La fonction caractéristique (appelée aussi transformée de Fourier) est utilisée à la place de la fonction génératrice pour étudier la loi d'un couple à densité. Elle est définie ainsi :

Définition 3.17 Soit (X, Y) un couple de variables aléatoires de densité conjointe $f : \mathbf{R}^2 \rightarrow \mathbf{R}^+$. La fonction caractéristique du couple (X, Y) est la fonction $\Phi : \mathbf{R}^2 \rightarrow \mathbf{R}$ suivante :

$$\Phi(\alpha, \beta) = \mathbf{E}(e^{i\alpha X + i\beta Y}) = \int_{\mathbf{R}^2} e^{i\alpha x + i\beta y} f(x, y) \, dx \, dy.$$

On retrouve la fonction caractéristique de X ou de Y en posant $\beta = 0$ ou $\alpha = 0$, et celle de la somme $X + Y$ en posant $\alpha = \beta$.

Remarque : La fonction caractéristique d'un vecteur aléatoire (X_1, \dots, X_n) à valeurs dans \mathbf{R}^n est la fonction Φ définie sur \mathbf{R}^n par $\Phi(\alpha_1, \dots, \alpha_n) = \mathbf{E}(\exp(i \sum \alpha_k X_k))$.

2.5 Loi de la somme

Il est facile de calculer la fonction caractéristique de la somme de deux variables aléatoires à partir de celle du couple formé par ces variables. On peut également vérifier que si le couple (X, Y) a une densité, la variable $X + Y$ aura également une densité, et on peut expliciter cette densité.

Proposition 3.18 – *Si le couple (X, Y) admet pour densité la fonction f , alors la densité de la variable aléatoire $X + Y$ est la fonction g définie par*

$$g(z) = \int_{\mathbf{R}} f(x, z - x) dx = \int_{\mathbf{R}} f(z - y, y) dy.$$

– *Si X et Y sont deux variables aléatoires indépendantes de densité f_X et f_Y , la densité, notée g , de $X + Y$ est le produit de convolution de f_X et f_Y :*

$$g(z) = \int_{\mathbf{R}} f_X(x) f_Y(z - x) dx = \int_{\mathbf{R}} f_X(z - y) f_Y(y) dy.$$

Preuve : Soit $h : \mathbf{R} \rightarrow \mathbf{R}$ une fonction borélienne et bornée. On a

$$\begin{aligned} \mathbf{E}h(X + Y) &= \int_{\mathbf{R}^2} h(x + y) f(x, y) dx, dy \\ &= \int_{\mathbf{R}} h(z) \int_{\mathbf{R}} f(x, z - x) dx dz \end{aligned}$$

en posant $x = x$ et $z = x + y$. Si on avait effectué le changement de variables $z = x + y$ et y , on aurait obtenu l'autre expression annoncée.

Chapitre 4

Théorèmes limites et tests

Le début de ce chapitre s'intéresse à la façon dont une suite de variables aléatoires peut converger. Plus précisément, on se donne une suite de variables aléatoires (X_n) indépendantes et identiquement distribuées et on désire étudier le comportement de la suite $S_n = \sum_{k \leq n} X_k$ lorsque n tend vers $+\infty$.

On abordera ensuite la question des tests statistiques, qui peuvent reposer sur des résultats exacts ou asymptotiques.

1 Lois des grands nombres

1.1 Loi faible des grands nombres

Théorème 4.1 *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, identiquement distribuées et de carré intégrable. On note $S_n = \sum_{k \leq n} X_k$. On a alors*

$$\lim_{n \rightarrow +\infty} \mathbf{E} \left(\left(\frac{S_n}{n} - \mathbf{E}(X_1) \right)^2 \right) = 0.$$

Définition 4.2 *Lorsqu'une suite de variables aléatoires Y_n de carrés intégrables et une variable aléatoires Y , elle aussi de carré intégrable, vérifient*

$$\lim_n \mathbf{E} \left((Y_n - Y)^2 \right) = 0$$

on dit que Y_n tend vers Y en moyenne quadratique (ou dans \mathcal{L}^2).

Lorsque l'on a plutôt $\lim_n \mathbf{E}|Y_n - Y| = 0$, on dit que Y_n tend vers Y en moyenne (ou dans \mathcal{L}^1).

Sous les hypothèses de la loi faible des grands nombres, on peut donc dire que S_n/n tend vers $\mathbf{E}(X_1)$ en moyenne quadratique.

La proposition suivante se déduit immédiatement de l'inégalité de Cauchy-Schwarz :

Proposition 4.3 *Si Y_n tend vers Y en moyenne quadratique, alors Y_n tend vers Y en moyenne.*

Preuve : de la loi faible des grands nombres. Soit X_n une suite de variables aléatoires indépendantes et de carré intégrable. Quitte à remplacer X_n par $X_n - \mathbf{E}(X_n)$, on voit que l'on peut supposer que les X_n sont d'espérance nulle. Notons donc $S_n = \sum_{k \leq n} X_k$ et montrons que $\mathbf{E}(S_n^2/n^2)$ tend vers 0. On a

$$\begin{aligned} \mathbf{E}(S_n^2) &= \mathbf{E} \left(\left(\sum_{k=1}^n X_k \right)^2 \right) \\ &= \mathbf{E} \left(\sum_{k=1}^n X_k^2 + 2 \sum_{1 \leq k < l \leq n} X_k X_l \right) \\ &= \sum_{k=1}^n \mathbf{E}(X_k^2) + 2 \sum_{1 \leq k < l \leq n} \mathbf{E}(X_k X_l) \\ &= n\mathbf{E}(X_1^2) + 2 \sum_{1 \leq k < l \leq n} \mathbf{E}(X_k)\mathbf{E}(X_l) \end{aligned}$$

Comme on a supposé que $\mathbf{E}(X_k) = 0$, on obtient que $\mathbf{E}(S_n^2) = n\mathbf{E}(X_1^2)$. En particulier,

$$\lim_n \mathbf{E} \left(\frac{S_n^2}{n^2} \right) = 0.$$

1.2 Loi forte des grands nombres

On part à nouveau d'une suite de variables aléatoires indépendantes et identiquement distribuées. On a alors :

Théorème 4.4 *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, identiquement distribuées et intégrables. On pose $S_n = \sum_{\{k \leq n\}} X_k$. Alors :*

$$\mathbf{P} \left(\omega, \lim_n \frac{S_n(\omega)}{n} \text{ existe et vaut } \mathbf{E}(X_1) \right) = 1.$$

Ce résultat est plus fin que le précédent car on suppose les variables aléatoires intégrables (et non nécessairement de carré intégrable) et la limite obtenue est plus précise : on a une information sur le comportement de $S_n(\omega)/n$ pour presque tout ω , au lieu d'une information sur une mesure.

2 Théorème central limite

Théorème 4.5 *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées, de carré intégrable. On a alors pour tout $x \in \mathbf{R}$*

$$\mathbf{P} \left(\frac{X_1 + \dots + X_n - n\mathbf{E}(X_1)}{\sqrt{n\text{var}(X_1)}} \leq x \right) \longrightarrow \int_{-\infty}^x e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

On dit que $(X_1 + \dots + X_n - n\mathbf{E}(X_1))/\sqrt{n\text{var}(X_1)}$ tend en loi vers la loi normale centrée réduite.

Exercice : Montrer que si les (X_n) sont indépendantes, identiquement distribuées et de carré intégrable, $(X_1 + \dots + X_n - n\mathbf{E}(X_1))/\sqrt{n\text{var}(X_1)}$ est une variable aléatoire de moyenne 0 et de variance 1.

La loi forte des grands nombres permet de dire que, si (X_n) est une suite de variables aléatoires intégrables et identiquement distribuées, $(X_1 + \dots + X_n)/n$ tend vers $\mathbf{E}(X_1)$. Le théorème central limite permet de préciser ce résultat au sens où il donne un équivalent de $(X_1 + \dots + X_n)/n - \mathbf{E}(X_1)$. La différence essentielle entre ces deux théorèmes est dans la nature de la convergence : pour la loi forte des grands nombres, c'est une convergence « presque sûre », c'est un mode exigeant de convergence. Dans le théorème central limite, c'est une convergence en loi. C'est le plus faible des modes de convergences que nous rencontrerons car il n'apporte aucune information sur le comportement presque sûr de la suite. Si $(Y_n)_n$ converge en loi, en général, la suite de réels $(Y_n(\omega))_n$ n'est pas convergente. Seule la suite des fonctions de répartition des Y_n converge.

Le théorème central limite explique le rôle fondamental joué par la loi normale en probabilité et en statistiques. Il est également à la source d'un certain nombre de tests statistiques.

3 Statistique descriptive

3.1 Échantillons

Un **échantillon aléatoire** $(x_k)_{1 \leq k \leq n}$ est une suite de résultats numériques obtenus soit lors de la répétition d'une expérience, soit au cours d'une simulation informatique. Il faut les conditions de l'expérience soient rigoureusement reproduites : ce sera le cas par exemple pour un tirage avec remise dans une urne, mais pas pour un tirage sans remise. Dans le cas d'une simulation informatique, il faudra supposer ou vérifier que l'échantillon fourni par l'ordinateur est conforme aux attentes ; dans ce cas, on parle d'un générateur de nombres pseudoaléatoires.

Du point de vue mathématique, un échantillon est une suite finie $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ des valeurs prises par une suite de variables aléatoires indépendantes $(X_k)_{k \geq 1}$ lors de l'expérience ω . Dire qu'un échantillon relève d'une loi donnée signifie que les variables aléatoires X_k suivent la loi en question, et que l'on a $x_k = X_k(\omega)$. Cela signifie aussi que si l'on trace un histogramme de l'échantillon, on s'attend à ce que l'histogramme se rapproche du tracé de la densité de la loi de X .

Un échantillon est donc une suite de réels : il ne s'agit en aucun cas de variables aléatoires (qui sont, rappelons-le, des fonctions).

Les tests statistiques porteront notamment sur la moyenne empirique, la variance empirique, la médiane d'un échantillon. Nous étudierons également des tests dits « d'ajustement » : ces tests permettront de dire si oui ou non un échantillon est susceptible de relever de la loi envisagée, pour un seuil de risque donné. Mais attention, les tests sont des outils qui permettent de quantifier des résultats qualitatifs (l'histogramme ressemble à la densité...), ils fournissent donc un critère d'évaluation ; mais ils ne permettent pas de dire avec certitude ce qui va se passer si on renouvelle l'expérience.

Notation : Nous réserverons l'usage des lettres majuscules aux noms des variables aléatoires, et celui des minuscules aux échantillons. Ainsi, par exemple, un échantillon $(x_k)_{1 \leq k \leq 100}$ sera susceptible de relever de la loi de la variable aléatoire X .

3.2 Paramètres d'un échantillon

Les principaux paramètres $(x_k)_{1 \leq k \leq n}$ d'un échantillon sont

- sa **taille**, c'est-à-dire le nombre n d'expériences réalisées. La taille peut être soit choisie par le statisticien, qui demandera la réalisation du nombre d'un nombre d'expériences déterminé, soit imposée par l'expérience elle-même (nombres de personnes atteintes d'une maladie donnée par exemple).
- sa **moyenne empirique**, c'est-à-dire $\bar{x} = (x_1 + \dots + x_n)/n$. Il ne faut pas confondre cette moyenne empirique avec la moyenne (dite *théorique*) de la loi dont relève l'échantillon. Deux échantillons relevant de la même loi ont toutes les chances d'avoir des moyennes empiriques différentes.
- la **variance empirique**, c'est-à-dire

$$\overline{\sigma^2} = \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}.$$

Attention à nouveau à bien différencier variances empirique et théorique. On utilisera également l'écart-type empirique.

- la **médiane**. Si $n = 2r + 1$ et si les valeurs de l'échantillon sont **classées** par ordre croissant, la médiane est x_{r+1} . Si $n = 2r$, la médiane est $(x_r + x_{r+1})/2$, toujours pour un échantillon classé. On la définit parfois dans ce cas comme étant l'intervalle $]x_r, x_{r+1}[$.

Il faudra bien prendre soin à différencier les quantités théoriques des quantités empiriques. La notion de « théorique » se rapporte à la loi dont est supposé relever l'échantillon, alors que les quantités empiriques sont déduites directement de l'échantillon. Les tests et les intervalles de confiance peuvent être utilisés dans les deux sens :

- soit on part de l'échantillon, et on estime le paramètre théorique (par un intervalle de confiance par exemple) à partir du paramètre empirique,
- soit on connaît a priori la loi dont doit relever l'échantillon et on souhaite donner un intervalle de fluctuation des paramètres empiriques, ou vérifier que l'échantillon relève bien de cette loi.

Quantiles, quartiles et pourcentiles

Nous venons de voir la médiane : c'est la valeur qui permet de découper l'échantillon en deux intervalles de même effectif (ie avec le même nombre de valeurs). Les $k - 1$ quantiles (ou fractiles) d'ordre k permettent de façon similaire de découper un échantillon en k intervalles de même effectif. On les numérote par ordre croissant.

Les quartiles sont les quantiles d'ordre 4 : il y en a donc trois. Le premier quartile est la plus petite valeur de l'échantillon à laquelle sont inférieures au moins un quart des valeurs de l'échantillon ; le deuxième quartile est la plus petite valeur de l'échantillon à laquelle sont inférieures au moins la moitié des valeurs de l'échantillon ; et le troisième quartile est la valeur de l'échantillon à laquelle sont inférieures au moins trois quarts des valeurs de l'échantillon.

Ils sont généralement notés Q_1 , Q_2 et Q_3 . Si l'échantillon comporte un nombre impair de valeurs, le deuxième quartile est égal à la médiane. L'intervalle inter-quartile est $[Q_1, Q_3]$ et l'écart inter-quartile est $Q_3 - Q_1$.

Par exemple, si un échantillon comporte 14 valeurs, Q_1 est la 4^{ième} valeur ($14 \times 0.25 = 3,5$), Q_2 est la 7^{ième}, et Q_3 est la 11^{ième} ($14 \times 0.75 = 10,5$).

Les pourcentiles (ou, suivant la terminologie anglo-saxonne, percentiles) sont les quantiles d'ordre 100.

On parle également de l'**étendue** d'un échantillon : c'est la différence de la plus grande et de la plus petite des valeurs.

Le **mode** d'un échantillon est la valeur la plus fréquente : cette donnée est pertinente uniquement pour les échantillons relevant d'une loi discrète.

Exemple : Les chiffres suivants représentent le salaire mensuel (en euros) des 30 salariés d'une entreprise.

1	2	3	4	5	6	7	8	9	10
1000	1200	1400	1800	1800	2000	2100	2400	2400	2400
11	12	13	14	15	16	17	18	19	20
2500	2500	2550	2600	2700	2800	2900	3000	3000	3200
21	22	23	24	25	26	27	28	29	30
3200	3300	3400	3500	3800	4200	4600	4600	4700	5000

La médiane de cet échantillon est la moitié de sa 15^{ième} et de sa 16^{ième} valeur, soit 2750.

Q_1 est la 8^{ième} valeur : $Q_1 = 2400$; Q_2 est la 15^{ième} valeur : $Q_2 = 2700$; Q_3 est la 23^{ième} valeur : $Q_3 = 3400$.

La moyenne empirique de l'échantillon vaut 2885 euros et la variance empirique 1028192 euros², soit un écart-type de 1014 euros.

3.3 Représentation graphique

Lorsque l'on est face à un échantillon, on commence généralement par calculer les paramètres ci-dessus, et on représente graphiquement l'échantillon, par exemple par une « boîte à moustaches » ou un histogramme ou encore un « camembert ».

Boîte à moustaches

Ces diagrammes sont également appelés diagrammes en boîte ou box-plots. Sur une échelle généralement verticale, on indique :

- le premier quartile Q_1 ,
- la médiane M ,
- le troisième quartile Q_3 ,
- on construit une « boîte » (un rectangle) enserrant ces trois valeurs,
- on ajoute deux traits horizontaux, l'un situé au niveau de la plus petite valeur de l'échantillon supérieure à $Q_1 - 1,5(Q_3 - Q_1)$ et l'autre à celui de la plus grande valeur de l'échantillon inférieure à $Q_3 + 1,5(Q_3 - Q_1)$; ce sont les moustaches.
- on indique les valeurs extrêmes, c'est-à-dire celles non comprises entre les moustaches par des *,
- on peut ajouter la moyenne (une croix).

La quantité $Q_3 - Q_1$ est appelée l'écart inter-quartile.

Au lieu de $Q_1 - 1,5(Q_3 - Q_1)$ et $Q_3 + 1,5(Q_3 - Q_1)$, on peut indiquer les percentiles d'ordre 10 et 90 par exemple (parfois 1 et 99).

Lorsque plusieurs échantillons sont à comparer, il peut être judicieux de tracer leurs box-plots en parallèle.

Histogramme

On regroupe l'échantillon en classes, c'est-à-dire en intervalles, et on comptabilise le nombre de valeurs de l'échantillon dans chacune des classes : ce sont les effectifs empiriques des classes. Il reste alors à tracer l'histogramme : sur l'axe des abscisses, on indique les bornes des classes, et on trace des rectangles dont l'aire est proportionnelle à l'effectif empirique de la classe. Reste à savoir comment on construit les classes. Pour cela, il y a plusieurs solutions. Les deux suivantes figurent les plus courantes :

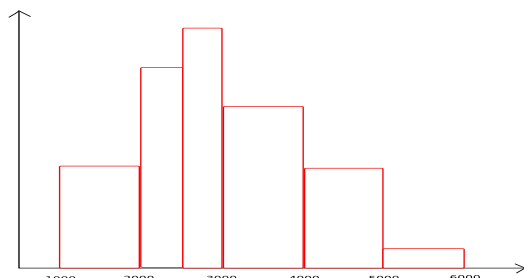
- on peut choisir des classes de largeur constante : dans ce cas (et uniquement dans celui-ci!), la hauteur est directement liée à l'effectif ;
- ou préférer des classes d'aire constante (c'est-à-dire contenant le même nombre de valeurs).

L'histogramme peut être complété par l'indication de l'effectif de chaque classe (ou des effectifs cumulés), soit en valeurs, soit en pourcentage.

Attention : Dans un histogramme, c'est l'aire des rectangles, et non la hauteur, qui est proportionnelle à l'effectif de la classe.

Exemple : Histogramme de l'échantillon de salaires donnés ci-dessus, en prenant pour classes $[1000, 2000[$, $[2000, 2500[$, $[2500, 3000[$, $[3000, 4000[$, $[4000, 5000[$ et $[5000, 6000[$. Les effectifs de chacun des intervalles sont : 5, 5, 7, 8, 5 et 1.

On obtient l'histogramme suivant :



Camembert

Ce mode de représentation sera utilisé essentiellement pour des échantillons prenant des valeurs discrètes (par exemple, un tirage de dé, ou lorsqu'il s'agit de données qualitatives). C'est le mode de représentation privilégié des résultats des sondages politiques, et aussi de la répartition effective des sièges suite à un scrutin de listes.

La réalisation d'un camembert débute, comme celle d'un histogramme par une répartition en classes. L'amplitude des secteurs angulaires du camembert sera proportionnelle à l'effectif de la classe considérée.

Attention à ce que la représentation graphique ne trouble pas l'esprit : les camemberts en perspective, qui sont sûrement très jolis et très à la mode dans les rapports divers et variés, sont également très difficiles à lire, et on peut leur faire dire à peu près n'importe quoi !

3.4 Estimateur

Le plus souvent, on étudie un échantillon pour obtenir des renseignements sur la loi ou les paramètres de la loi théorique dont relève cet échantillon : on dit que l'on **estime**, ou que l'on utilise un **estimateur** de ces paramètres.

Considérons une suite de variables aléatoires $(X_n)_{n \geq 1}$ indépendantes et de même loi $\mathcal{L}(\theta)$, θ étant un paramètre inconnu et une fonction T_N de (X_1, \dots, X_N) , définie pour tout $N \geq 1$, destinée à estimer θ . T_N sera appelé un **estimateur** de θ si lorsque N tend vers $+\infty$, T_N tend vers θ . Cet estimateur sera dit **sans biais** si $\mathbf{E}(T_N) = \theta$.

Par exemple, si le paramètre θ à estimer est l'espérance de la loi des X_n , on utilisera le plus souvent comme estimateur

$$T_N = \frac{X_1 + \dots + X_N}{N}.$$

On peut vérifier que cet estimateur est sans biais.

Mis à part le fait d'être sans biais, on demandera à un « bon » estimateur d'être robuste, c'est-à-dire peu sensible aux variations du type de loi et aux valeurs aberrantes de l'échantillon. L'estimateur de la moyenne donné ci-dessus est relativement sensible aux grandes valeurs de l'échantillon mais il est valable pour un grand nombre de lois.

4 Tests Statistiques

4.1 Principe des tests statistiques

Un test statistique est un outil d'aide à la décision : il permet de comparer différents échantillons soit entre eux, soit avec une loi théorique fixée. Les tests portent soit sur un paramètre de l'échantillon, soit sur la loi dont est supposé relever l'échantillon.

Un test part toujours d'une hypothèse H_0 que l'on suppose vraie. Par exemple : « La loi de l'échantillon est une loi de Gauss » ou « la moyenne (théorique) de la loi dont relève l'échantillon est égale à 3 ». L'hypothèse H_1 est le contraire de H_0 : on pense donc qu'elle est fautive. On se fixe alors un risque α (en général entre 1% et 10%) : α est le risque de rejeter l'hypothèse H_0 alors qu'elle est vérifiée.

Par exemple, si on fait un tirage de pile ou face avec une pièce parfaitement équilibrée, on s'attend à ce que la proportion de piles soit proche de 1/2. Néanmoins, si on fait 100 tirages, il y a une probabilité très faible mais non nulle pour que cette proportion soit égale à 1 : si on fait un test sur la proportion de piles pour cet échantillon (avec comme hypothèse H_0 « La pièce est équilibrée »), on va rejeter H_0 , alors qu'on a créé l'échantillon avec une pièce équilibrée !

Pour pratiquer un test statistique, il faut aussi disposer d'un résultat théorique (en loi) sur les variables aléatoires : par exemple si les (X_n) sont indépendantes identiquement distribuées et de loi de Gauss $N(m, \sigma^2)$, on sait que $(X_1 + \dots + X_n - nm) / \sqrt{n\sigma^2}$ suit une loi de Gauss centrée réduite. Pour un risque α fixé (ici, 5%), on détermine alors a (par exemple avec une table statistique) de telle sorte que, si Y suit une loi normale centrée réduite, on a

$$\mathbf{P}(|Y| \geq a) = \alpha = 0,05.$$

On trouve $a = 1,96$. On en déduit que

$$\mathbf{P}\left(\left|\frac{X_1 + \dots + X_n - nm}{\sqrt{n\sigma^2}}\right| \geq a\right) = 0,05.$$

Si on dispose d'un échantillon $(x_k)_{k \leq n}$ et si l'on désire tester l'hypothèse H_0 « La moyenne théorique de la loi dont relève cet échantillon est égale à 3 », on calcule la moyenne (empirique) $\bar{m} = (x_1 + \dots + x_n)/n$. Si \bar{m} vérifie $|\bar{m} - 3| > a\sigma/\sqrt{n}$, on rejette, l'hypothèse H_0 , avec un risque de 5%. Sinon, on conclut que le test pratiqué ne permet pas d'écarter l'hypothèse H_0 pour un risque de 5%, ou plus simplement : on accepte cette hypothèse pour un risque de 5%.

On dit encore que ce test est basé sur un résultat exact car on connaît la loi exacte de $\sqrt{n}(\bar{m}_n - m)/\sigma$ et pas simplement sa loi asymptotique.

4.2 Test du χ^2 d'adéquation

On dit test du khi 2.

Ce test permet de vérifier si un échantillon relève d'une loi fixée ou non : plus précisément, on teste si l'histogramme ressemble au tracé de la densité (si elle existe). L'hypothèse à tester est H_0 « l'échantillon relève de la loi de X . »

Pour cela, on répartit l'échantillon $(x_i)_{i \leq n}$ en un certain nombre de classes (intervalles) $(I_j)_{j \leq k}$. Ces intervalles doivent former une partition de $X(\Omega)$.

On commence par calculer les **effectifs empiriques**, c'est-à-dire, pour tout $j \leq k$,

$$n_j = \text{card} \{i \leq n, x_i \in I_j\}$$

et les **effectifs théoriques**

$$N_j = n\mathbf{P}(X \in I_j).$$

Remarque : Les effectifs empiriques sont toujours des entiers, les effectifs théoriques sont simplement des réels positifs.

Le test permet de comparer l'écart entre les n_j et les N_j ; on calcule la distance de Pearson D_n :

$$D_n = \sum_{j=1}^k \frac{(n_j - N_j)^2}{N_j}.$$

ou encore, en développant cette expression :

$$D_n = \sum_{j=1}^k \frac{n_j^2}{N_j} - n.$$

Lorsque n tend vers $+\infty$, la loi de D_n tend vers une loi du χ^2 à $k - 1$ degrés de liberté notée χ_{k-1}^2 (où k est le nombre de classes). En pratique, on considère que, à partir de $n = 50$, la loi de D_n est suffisamment proche de la loi du χ^2 pour pouvoir appliquer le test. Pour un risque $\alpha > 0$ fixé, on lit donc dans la table de la loi du χ^2 du bon nombre de degré de liberté la valeur x_c telle que $\mathbf{P}(\chi_{k-1}^2 \geq x_c) = \alpha$. Pour l'échantillon observé, si $D_n \geq x_c$, on rejettera l'hypothèse H_0 , pour le risque α . Sinon, on conclura que ce test ne permet pas de rejeter cette hypothèse, pour le risque α .

On choisit les classes de sorte que les effectifs théoriques N_j soient tous supérieurs ou égaux à 5 : on estime que, ainsi, la loi de D_n est plus proche d'une loi de χ_{k-1}^2 .

La loi à tester n'est pas toujours complètement déterminée : on peut tester par exemple si un échantillon relève d'une loi normale, sans préciser la moyenne et la variance. Pour créer les classes et calculer les effectifs théoriques, on estime les paramètres nécessaires (moyenne et/ou

variance en général). cela provoque une diminution de nombre de degrés de liberté du χ^2 limite : il devient $k - p - 1$ si on a estimé p paramètres.

Exemple : Tester avec un test du χ^2 d'ajustement si le dé ayant donné les tirages suivants est pipé, pour un niveau de risque de 5%.

6 6 4 4 2 3 1 1 2 3 6 2 1 6 2 1 4 6 6 3 3 3 1 5 5 6 1 1 2 4 3 3 5 1 6 1 3 6 3 3 5 2 6 2 5 6 5 6 4 5

Les effectifs empiriques sont :

$$n_1 = 9 \quad n_2 = 7 \quad n_3 = 10 \quad n_4 = 5 \quad n_5 = 7 \quad n_6 = 12,$$

et on les regroupe sous la forme d'un tableau contenant aussi les effectifs théoriques, qui sont tous égaux à $50/6$ si le dé n'est pas pipé.

	1	2	3	4	5	6
N_j	50/6	50/6	50/6	50/6	50/6	50/6
n_j	9	7	10	5	7	12

On calcule

$$D_{50} = \frac{(9 - 50/6)^2 + (7 - 50/6)^2 + (10 - 50/6)^2 + (5 - 50/6)^2 + (7 - 50/6)^2 + (12 - 50/6)^2}{50/6}.$$

On obtient $D_{50} = 3,76$.

On lit ensuite dans une table du χ^2 à 5 degrés de liberté : $\mathbf{P}(\chi_5^2 \geq 11,1) = 0,05$. On accepte donc l'hypothèse que le dé n'est pas pipé, pour un risque de 5%.

4.3 Intervalle de confiance

Un test classique porte sur la moyenne d'un échantillon : on connaît \bar{m} et on souhaite obtenir un « intervalle de confiance » de la moyenne théorique m . On commence par se fixer un risque α ou un niveau de confiance $1 - \alpha$.

Le test est basé sur le résultat suivant :

Théorème 4.6 *Si les variables X_1, \dots, X_n sont indépendantes et de loi de Gauss de moyenne m et de variance σ^2 , la variable*

$$T = \frac{\bar{X} - m}{\sqrt{\bar{V}/(n-1)}}$$

suit une loi de Student à $n - 1$ degrés de liberté, avec

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad \bar{V} = \frac{X_1^2 + \dots + X_n^2}{n} - \bar{X}^2.$$

Pour réaliser l'intervalle de confiance de la moyenne de l'échantillon supposé gaussien $(x_k)_{k \leq n}$, on calcule donc la moyenne empirique \bar{m} et la variance empirique \bar{v} de cet échantillon. On lit alors dans une table de la loi de Student T_{n-1} le réel t tel que

$$\mathbf{P}(|T_{n-1}| \geq t) = \alpha,$$

où α est le niveau de risque choisi. L'intervalle de confiance obtenu est

$$m \in \left[\bar{m} - t \sqrt{\frac{\bar{v}}{n-1}}, \bar{m} + t \sqrt{\frac{\bar{v}}{n-1}} \right] \quad \text{pour un risque } \alpha.$$

Cet intervalle de confiance est basé sur un résultat exact (et non asymptotique), valable uniquement pour les échantillons gaussiens, mais lorsque l'échantillon est de taille suffisamment grande ($n \geq 30$), on peut le mettre en pratique pour un échantillon non gaussien, en remplaçant la loi de Student par une loi de Gauss : le résultat devient asymptotique, et pour n suffisamment grand (supérieur à 30 en pratique), on a avec une confiance α ,

$$m \in \left[\bar{m} - a \sqrt{\frac{\bar{v}}{n-1}}, \bar{m} + a \sqrt{\frac{\bar{v}}{n-1}} \right]$$

où a est choisi de sorte que, si X est une variable aléatoire de loi normale centrée réduite,

$$\mathbf{P}(|X| \geq a) = \alpha.$$

5 Séries statistiques à deux variables

Il s'agit ici d'étudier les échantillons statistiques $(x_i, y_i)_{i \leq n}$, par exemple, le poids et la taille d'enfants du même âge, ou le chiffre d'affaire d'une entreprise en fonction de l'année. En pratique, on commencera toujours par tracer le nuage de points associé. Si ce nuage est relativement rectiligne, on cherchera alors la droite D passant « le plus près » de ces points. Dans cette étude, le rôle des x_i et des y_i n'est pas interchangeable : les x_i seront les variables explicatives et les y_i les variables à expliquer. Un cas courant est celui où les x_i représentent le temps (heure, année), et où les y_i sont une certaine quantité variant au cours du temps : il sera alors possible d'obtenir une estimation de la valeur y correspondant à un certain x n'appartenant pas aux données initiales. Il existe des tests et des intervalles de confiance, portant sur les coefficients de la régression ou sur les valeurs y estimées, mais ce n'est pas l'objet de ce cours !

Écrivons l'équation catésienne de la droite D :

$$D = \{(x, y), y = ax + b\}.$$

Les paramètres à déterminer sont a et b . On les cherche de telle sorte que la quantité

$$\Delta(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

soit minimale. C'est la somme des carrés des résidus. On peut remarquer qu'il s'agit d'un polynôme de degré 2 en a et b . Étant toujours positif, ce polynôme admet un minimum. Pour déterminer le couple (a, b) réalisant ce minimum, on cherche à annuler

$$\frac{\partial \Delta}{\partial a} \quad \text{et} \quad \frac{\partial \Delta}{\partial b}.$$

On a donc :

$$\sum_{i=1}^n (x_i(y_i - ax_i - b)) = 0 \quad \text{et} \quad \sum_{i=1}^n (y_i - ax_i - b) = 0.$$

En notant

$$\bar{x} = \frac{\sum x_i}{n}, \quad \bar{y} = \frac{\sum y_i}{n}, \quad \overline{v_x} = \frac{\sum x_i^2}{n} - \bar{x}^2, \quad \overline{c_{xy}} = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y}$$

respectivement les moyennes empiriques des (x_i) et des (y_i) , la variance empirique des (x_i) et la covariance empirique des (x_i, y_i) on obtient :

$$a = \frac{\overline{c_{xy}}}{\overline{v_x}} \quad \text{et} \quad b = \bar{y} - a\bar{x}.$$

On constate que cette droite, appelée droite des moindres carrés ordinaires, passe par (\bar{x}, \bar{y}) , le barycentre du nuage de points : c'est souvent plus utile de connaître ce résultat pour tracer la droite des moindres carrés, que de connaître la valeur de b : si 0 est « loin » des x_i , il n'apparaîtra en effet pas sur l'axe des abscisses !

La régression linéaire sera d'autant meilleure que la valeur absolue du coefficient de corrélation

$$r = \frac{\overline{c_{xy}}}{\sqrt{\overline{v_x} \overline{v_y}}}$$

est proche de 1. En effet, la formule de Cauchy-Schwarz permet de montrer que l'on a toujours $r \in [-1, 1]$, et que r ne vaut 1 ou -1 que si les points sont colinéaires, auquel cas les (x_i, y_i) forment une droite parfaite.

Exemple : Les données suivantes représentent le nombre de retraités en France, en millions, en fonction de l'année :

année	1990	1991	1992	1993	1994	1995	1996	1997
retraités	7.3	7.6	7.9	8.2	8.5	8.8	9.0	9.2
année	1998	1999	2000	2001	2002	2003	2004	2005
retraités	9.4	9.6	9.7	9.8	10.0	10.2	10.4	10.7

La variable explicative est l'année x et la variable à expliquer est le nombre de retraités y . On calcule :

$$\begin{aligned} \bar{x} &= 1997.5 & \bar{y} &= 9.15 & \overline{v_x} &= \frac{1}{16} \sum x_i^2 - \bar{x}^2 = 21.25 \\ \overline{v_y} &= \frac{1}{16} \sum y_i^2 - \bar{y}^2 = 0.98 & \overline{c_{xy}} &= \frac{1}{16} \sum x_i y_i - \bar{x}\bar{y} = 4.54. \end{aligned}$$

L'équation de la droite des moindres carrés ordinaire est donc : $y = 0,2135(x - \bar{x}) + \bar{y}$.

Le graphique ci-dessous représente le nuage de points ainsi que la droite de la régression linéaire. Le point (\bar{x}, \bar{y}) est représenté par une étoile *.

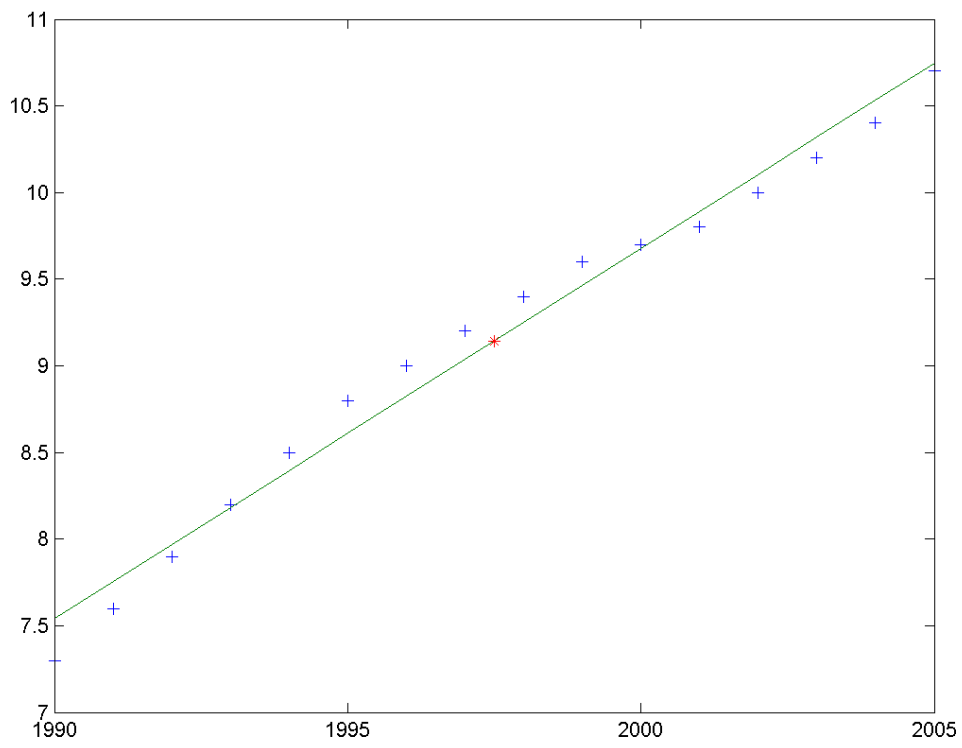


Table des matières

1	Probabilités	1
1	Mesure	1
1.1	Définitions	1
1.2	Probabilités discrètes	2
1.3	Probabilité à densité	3
1.4	Probabilité conditionnelle, indépendance	4
2	Variables aléatoires réelles	6
1	La loi	6
1.1	Définition	6
1.2	Exemples de variables aléatoires discrètes	6
1.3	Exemples de loi à densité	7
2	Espérance, variance, fonctions génératrice et caractéristique	8
2.1	Espérance et variance	8
2.2	Fonction de répartition de la loi d'une variable aléatoire	12
2.3	Fonctions génératrice, caractéristique et transformée de Laplace	13
2.4	Comment calculer la loi	15
3	Couples et vecteurs de variables aléatoires	17
1	Couples et vecteurs aléatoires discrets	17
1.1	Loi conjointe	17
1.2	Lois marginales	18
1.3	Loi de $f(X, Y)$	19
1.4	Loi conditionnelle	21
1.5	Indépendance de variables aléatoires discrètes	22
1.6	Espérance, matrice de covariance	22
1.7	Fonction génératrice d'un couple	25
1.8	Exemple d'utilisation de la fonction génératrice	26
1.9	Somme de deux variables aléatoires de loi de Poisson et indépendantes	26
2	Couples aléatoires à densité	27
2.1	Densité d'un couple	27
2.2	Loi marginale	28
2.3	Indépendance	29
2.4	Fonction caractéristique	29
2.5	Loi de la somme	30

4	Théorèmes limites et tests	31
1	Lois des grands nombres	31
	1.1 Loi faible des grands nombres	31
	1.2 Loi forte des grands nombres	32
2	Théorème central limite	32
3	Statistique descriptive	33
	3.1 Échantillons	33
	3.2 Paramètres d'un échantillon	34
	3.3 Représentation graphique	35
	3.4 Estimateur	37
4	Tests Statistiques	37
	4.1 Principe des tests statistiques	37
	4.2 Test du χ^2 d'adéquation	38
	4.3 Intervalle de confiance	39
5	Séries statistiques à deux variables	40