Chapitre X : notions de bases des probabilités

Stéphane Le Roux leroux@lsv.fr

ENS Paris-Saclay

2023-2024

Incertitude et intuition

Deux élèves d'une même classe ont été jugées pour sorcellerie, car elles étaient nées le même jour. Est-ce bien raisonnable ?

Une enveloppe contient de l'argent, une autre le double. Vous pouvez en ouvrir une seule avant de décider laquelle prendre. Que faites-vous ?

On tire à pile ou face. Pile, vous gagnez 1 € et le jeu s'arrête ; face, on tire encore. Puis, pile, vous gagnez 2 € et le jeu s'arrête ; face, on continue... 2^n €. Combien seriez-vous prêts à payer pour jouer à ce jeu ?

Si les étudiants qui mangent au Krouce tombent malades trois fois plus que les autres, que peut-on en conclure ?

Bart et Lisa essayent, chacun de leur côté, de finir le plus de niveaux possibles dans les jeux A et B : ils peuvent choisir n'importe quel jeu le lundi et l'autre le mardi. Le lundi, Bart finit son jeu à 80% et Lisa à 40% ; le mardi, Bart finit à 30% et Lisa à 15%. Que peut-on en conclure ?

Formaliser l'incertitude

Motivation

- On veut prendre une décision sans connaître toutes les informations pertinentes.
- On formalise l'incertitude pour raisonner/calculer, puis prendre une bonne décision, si possible optimale.
- On doit d'abord formaliser les objets sur lesquels porte l'incertitude : l'état ou l'évolution du monde ou d'une partie du monde.

Les évènements

- Si on tire une fois à pile ou face, la partie du monde qui nous concerne a deux états possibles. On pose $\Omega := \{pile, face\}$.
- Si on jette un dé un nombre dénombrable de fois, la partie du monde qui nous concerne a beaucoup d'états possibles. On pose $\Omega := \{1, 2, 3, 4, 5, 6\}^{\mathbb{N}}$.

Ω est appelé univers ou ensemble fondamental. Ses éléments sont appelés les évènements élémentaires. Lors d'une expérience, exactement un évènement élémentaire se réalise/se produit.

Si on s'intéresse seulement à la parité de la face d'un dé (par exemple pour simuler un tirage à pile ou face), on est amené à considérer des parties de Ω, e.g. $\{1, 3, 5\}$ et $\{2, 4, 6\}$.

Toute partie E de l'univers Ω est appelée évènement. On dit que E se réalise si un de ses évènements élémentaires se réalise.

Exprimer la vraisemblance

- On appelle donc $\emptyset$ l'évènement impossible et Ω l'évènement certain (son troisième nom).
- Notre décision en environnement incertain dépendra de la vraisemblance des évènements. On cherche donc un moyen pratique d'attribuer des niveaux de vraisemblance aux parties de Ω.
    1. Si $A \subseteq B \subseteq \Omega$, on voudrait que B soit au moins aussi vraisemblable que A.
    2. Si $A \sqcup B \sqcup C \subseteq \Omega$ (union disjointe), on voudrait aussi combiner leurs vraisemblances de manière associative : combiner celle de $A \sqcup B$ et celle de C devrait revenir à combiner celle de A et celle de $B \sqcup C$.
- Pour cela on utilise des niveaux de vraisemblance dans $[0, 1]$, qu'on appelle alors probabilités. On associe 0 à $\emptyset$ et 1 à Ω, et on combine les probabilités en les additionnant.

Quantifier la vraisemblance

Définition

Soit Ω un univers dénombrable. Une fonction $\delta : \Omega \to [0, 1]$ telle que $\sum_{x \in \Omega} \delta(x) = 1$ est appelée distribution de probabilité. La fonction

$$\begin{aligned} \mathrm{P}_\delta : \quad & \mathcal{P}(\Omega) \to [0, 1] \\ & A \mapsto \sum_{x \in A} \delta(x) \end{aligned}$$

est appelée mesure de probabilité (engendrée par δ).

- La distribution δ sur $\{1, 2, 3, 4, 5, 6\}$ constante égale à $\frac{1}{6}$ est appelée distribution uniforme. Dans ce cas $\mathrm{P}(\{2, 5\}) = \frac{1}{3}$.
- Si on tire à pile ou face tant qu'on observe pile, l'univers peut être défini par $\{pile\}^*\{face\} \cup \{pile\}^\omega$. (cf un exemple du début.)

On tire à pile ou face ω fois et on gagne si la différence entre les piles et les faces reste bornée. On considère alors peut-être plutôt $\Omega := \{pile, face\}^\omega$.

Univers indénombrables

Proposition

- Soit Ω un ensemble et $\delta : \Omega \to [0, +\infty[$ tel que $\sum_{x \in \Omega} \delta(x) = 1$. Alors l'ensemble $S := \{x \in \Omega \mid 0 < \delta(x)\}$ est dénombrable.
- Soit la mesure de probabilité P construite par extension de δ. Alors $\mathrm{P}(A) = 0$ pour tout $A \subseteq \Omega \setminus S$ et $\mathrm{P}(B) = \mathrm{P}(B \cap S)$ pour tout $B \subseteq \Omega$.

Preuve

- Soit $S_n := \{\omega \in \Omega \mid \frac{1}{2^n} < \delta(\omega)\}$. D'une part S_n est fini (de cardinal inférieur à 2^n), et d'autre part $S = \cup_{n \in \mathbb{N}} S_n$, car pour tout $\epsilon > 0$, il existe $n \in \mathbb{N}$ tel que $\frac{1}{2^n} < \epsilon$. Ainsi, S est dénombrable.
- Pour tout $A \subseteq \Omega \setminus S$ on a $\mathrm{P}(A) = \sum_{\omega \in A} 0 = 0$.
  Et $\mathrm{P}(B) = \mathrm{P}(B \cap S) + \mathrm{P}(B \setminus S) = \mathrm{P}(B \cap S) + 0$.

Univers indénombrables

Proposition

- Soit Ω un ensemble et $\delta : \Omega \to [0, +\infty[$ tel que $\sum_{x \in \Omega} \delta(x) = 1$. Alors l'ensemble $S := \{x \in \Omega \mid 0 < \delta(x)\}$ est dénombrable.
- Soit une mesure de probabilité P construite par extension de δ. Alors $\mathrm{P}(A) = 0$ pour tout $A \subseteq \Omega \setminus S$ et $\mathrm{P}(B) = \mathrm{P}(B \cap S)$ pour tout $B \subseteq \Omega$.

Conséquences et exemple

- Si on construit une mesure de probabilité (sur les évènements) en étendant une distribution de probabilité (sur les évènements élémentaires), on se ramène donc au cas d'un univers dénombrable.
- Pour définir un cadre plus riche, on va distribuer les poids (de somme 1) de manière plus équitable sur l'univers. Pour cela, on va utiliser des évènements de base différents des évènements élémentaires.
- Soient $a, b \in \mathbb{R}$ tels que $a < b$. La mesure uniforme sur $[a, b]$ associe une proba $\mathrm{P}([c, d]) := \frac{d-c}{b-a}$ à tout intervalle $[c, d] \subseteq [a, b]$.

Les tribus

Définition

Soit Ω un ensemble et $\mathcal{T}$ une collection de parties de Ω. On dit que $\mathcal{T}$ est une tribu (ou σ-algèbre), si les trois conditions suivantes sont vérifiées.

(T1) $\Omega \in \mathcal{T}$.

(T2) $A \in \mathcal{T} \Rightarrow \Omega \setminus A \in \mathcal{T}$

(T3) Si $A_n \in \mathcal{T}$ pour tout $n \in \mathbb{N}$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{T}$

$\{\emptyset, \Omega\}$ est la tribu grossière et $\mathcal{P}(\Omega)$ la tribu discrète.

Si $\mathcal{T}$ est une tribu de Ω, on dit que $(\Omega, \mathcal{T})$ et un espace mesurable.

Les tribus (II)

Soit Ω un ensemble et $\mathcal{T}$ une collection de parties de Ω. On dit que $\mathcal{T}$ est une tribu (ou σ-algèbre), si les trois conditions suivantes sont vérifiées.

(T1) $\Omega \in \mathcal{T}$.

(T2) $A \in \mathcal{T} \Rightarrow \Omega \setminus A \in \mathcal{T}$

(T3) Si $A_n \in \mathcal{T}$ pour tout $n \in \mathbb{N}$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{T}$

Proposition

Toute intersection non-vide de tribus est une tribu.

Soit $(\mathcal{T}_i)_{i\in I}$ une famille de tribus sur un univers Ω. Soit $\mathcal{T} = \cap_{i\in I}\mathcal{T}_i$.

- $\Omega \in \mathcal{T}_i$ pour tout $i \in I$, donc $\Omega \in \cap_{i\in I}\mathcal{T}_i = \mathcal{T}$.
- Soit $A \in \mathcal{T}$. Alors $A \in \mathcal{T}_i$ pour tout $i \in I$, donc $\Omega \setminus A \in \mathcal{T}_i$ pour tout $i \in I$, donc $\Omega \setminus A \in \cap_{i\in I}\mathcal{T}_i = \mathcal{T}$.
- Si $A_n \in \mathcal{T}$ pour tout $n \in \mathbb{N}$, alors $A_n \in \mathcal{T}_i$ pour tout $n \in \mathbb{N}$ et $i \in I$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{T}_i$ pour tout $i \in I$, donc $\cup_{n\in\mathbb{N}} A_n \in \cap_{i\in I}\mathcal{T}_i = \mathcal{T}$

Les tribus (III)

Proposition

- Pour toute collection $\mathcal{C}$ de parties de Ω, il existe une plus petite tribu contenant $\mathcal{C}$. On note $\mathcal{T}(\mathcal{C})$ cette tribu engendrée par $\mathcal{C}$.
- Pour toutes collections $\mathcal{C}$ on a $\mathcal{C} \subseteq \mathcal{T}(\mathcal{C}) = \mathcal{T}(\mathcal{T}(\mathcal{C}))$.
- Pour toutes collections $\mathcal{C}$ et $\mathcal{C}'$ telles que $\mathcal{C} \subseteq \mathcal{C}'$ on a $\mathcal{T}(\mathcal{C}) \subseteq \mathcal{T}(\mathcal{C}')$

Preuve

Les tribus de Ω forment un système de clôture et $\mathcal{T}$ est l'opérateur de clôture associé.

$\{\emptyset, A, \Omega \setminus A, \Omega\}$ est la tribu $\mathcal{T}(\{A\})$ engendrée par $\{A\}$.
$\{\emptyset, A, B, A \cup B, A \cap B, \Omega \setminus A, \Omega \setminus B, \Omega \setminus (A \cup B), A \cup \Omega \setminus B, B \cup \Omega \setminus A, \ldots, \Omega\}$ est la tribu $\mathcal{T}(\{A, B\})$ engendrée par $\{A, B\}$.

Tribus non triviales

Définition

La tribu borélienne de la droite réelle est la tribu engendrée par $\{]a, b[\mid a, b \in \mathbb{R}\}$.

Proposition

La tribu borélienne est aussi engendrée par $\{]-\infty, b] \mid b \in \mathbb{R}\}$ ou $\{[a, b] \mid a, b \in \mathbb{R}\}$ ou les ouverts de $\mathbb{R}$ (topologie usuelle).

Preuve pour les ouverts

Les $]x, y[$ sont des ouverts, donc $\mathcal{T}(I_O) \subseteq \mathcal{T}(\text{ouverts})$. Réciproquement, montrons que $\mathcal{T}(I_O)$ contient tous les ouverts. Soit O un ouvert de $\mathbb{R}$. Alors pour tout $x \in O$ il existe $a_x, b_x \in \mathbb{Q}$ tels que $x \in]a_x, b_x[\subseteq O$. Donc $\cup_{x \in O}]a_x, b_x[\subseteq O$ et $O \subseteq \cup_{x \in O}]a_x, b_x[$, d'où égalité. Or les $]a_x, b_x[$ sont en nombre dénombrable, donc tout ouvert est dans $\mathcal{T}(I_O)$, donc $\mathcal{T}(\text{ouverts}) \subseteq \mathcal{T}(\mathcal{T}(I_O)) = \mathcal{T}(I_O)$.

Tribus non triviales (II)

Définition

Soit $\Sigma := \{0, 1\}$. Pour tout $u \in \Sigma^*$ on note $u\Sigma^\omega$ l'ensemble des éléments de Σ^ω de préfixe u. La tribu borélienne de Σ^ω est la tribu engendrée par $\{u\Sigma^\omega \mid u \in \Sigma^*\}$. (Comparer avec la topologie usuelle sur Σ^ω.)

Définition : tribu produit

Soient $(E, \mathcal{T}_E)$ et $(F, \mathcal{T}_F)$ deux espaces mesurables. On note $\mathcal{T}_E \times \mathcal{T}_F := \{A \times B \mid A \in \mathcal{T}_1 \wedge B \in \mathcal{T}_2\}$. Alors $\mathcal{T}(\mathcal{T}_E \times \mathcal{T}_F)$ est appelée la tribu produit sur $E \times F$.

Lemma

Soient E et F deux ensembles. Soient $\mathcal{C}_E \subseteq \mathcal{P}(E)$ et $\mathcal{C}_F \subseteq \mathcal{P}(F)$. Alors $\mathcal{T}(\mathcal{T}(\mathcal{C}_E) \times \mathcal{T}(\mathcal{C}_F)) = \mathcal{T}(\mathcal{C}_E \times \mathcal{C}_F)$

Proof.

Il suffit de montrer $\mathcal{T}(C_E \times \mathcal{T}(\mathcal{C}_F)) = \mathcal{T}(\mathcal{C}_E \times \mathcal{C}_F)$ puis d'invoquer une symmétrie. □

Mesures de probabilité

Avant de construire des mesures de probabilité, on précise deux axiomes qu'elles devront vérifier.

Définition

Soit $(\Omega, \mathcal{T})$ un univers muni d'une tribu, on appelle mesure de probabilité toute application $\mathrm{P} : \mathcal{T} \to [0,1]$ telle que

(P1) $\mathrm{P}(\Omega) = 1$,

(P2) (σ-additivité) Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'évènements de $\mathcal{T}$ disjoints deux à deux, alors

$$\mathrm{P}\left(\sqcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathrm{P}(A_n)$$

Le triplet $(\Omega, \mathcal{T}, \mathrm{P})$ est appelé espace probabilisé.

Les mesures de probabilité définies précédemment sur des univers Ω dénombrables sont les mesures de probabilité sur la tribu discrète $\mathcal{P}(\Omega)$.

Construction de measures

- Rappel : dans un univers dénombrable, on étend une distribution de probabilité sur les évènements élémentaires.
- Dans un univers indénombrable non-trivial, les événements de base ne sont pas les événements élémentaires.

Définion : anneau d'ensembles

Soit E un ensemble. Une collection $A \subseteq \mathcal{P}(E)$ est un anneau d'ensembles si elle vérifie les axiomes suivants :

1. $A \neq \emptyset$.
2. Pour tout $X, Y \in A$ on a $X \setminus Y \in A$
3. Pour tout $X, Y \in A$ on a $X \cup Y \in A$

Toute tribu est un anneau d'ensembles.

Corollaire des théorèmes d'extension de Carathéodory

Soient R un anneau d'ensembles et $P : R \to [0, 1]$ une pré-mesure, i.e. vérifiant $\mathrm{P}(\emptyset) = 0$ et la σ-additivité. Alors il existe une unique extension de P sur $\mathcal{T}(R)$ qui soit une mesure de probabilité.

Mesure produit

- Soit $(\Omega_1, \mathcal{T}_1, \mathrm{P}_1)$ et $(\Omega_2, \mathcal{T}_2, \mathrm{P}_2)$ deux espaces probabilisés.
- Soit $\mathrm{P} : \mathcal{T}_1 \times \mathcal{T}_2 \to [0, 1]$ telle que $\mathrm{P}(A \times B) := \mathrm{P}_1(A)\mathrm{P}_2(B)$

Lemma

Le P ci-dessus s'étend en une unique mesure de probabilité.

Proof.

Soit R l'ensemble des unions disjointes finies de produits cartésiens $A \times B \in \mathcal{T}_1 \times \mathcal{T}_2$. Alors R est un anneau d'ensembles. (Faire des dessins.) P s'étend naturellement sur R en une pré-mesure. On invoque alors les théorèmes d'extension de Carathéodory. $\square$

Conséquences des axiomes

Proposition

Soit un espace probabilisé $(\Omega, \mathcal{T}, \mathrm{P})$.

- $\mathrm{P}(\emptyset) = 0$
- Si $A \cap B = \emptyset$ alors $\mathrm{P}(A \sqcup B) = \mathrm{P}(A) + \mathrm{P}(B)$.
- $\mathrm{P}(A) + \mathrm{P}(A^C) = 1$, où $A^C := \Omega \setminus A$ (quand Ω est non ambigü)
- $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$. (cf formule de Poincaré)
- Si $A \subseteq B$ alors $\mathrm{P}(A) \leq \mathrm{P}(B)$.
- Si $\mathrm{P}(B) = 0$, alors $\mathrm{P}(A \cap B) = 0$.
- Si $\mathrm{P}(B) = 1$, alors $\mathrm{P}(A \cap B) = \mathrm{P}(A)$.

Inégalité de Boole

Inégalité de Boole

Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événement on a

$$\mathrm{P}(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \mathrm{P}(A_n)$$

où la somme de droite peut diverger. (Peut être utile si $\sum_{n\in\mathbb{N}} \mathrm{P}(A_n) < 1$.)

Preuve

- Pour tout $n \in \mathbb{N}$, soit $B_n := A_n \setminus \cup_{k=0}^{n-1} A_k$.
- On montre par récurrence sur n que $\cup_{k=0}^{n} B_k = \cup_{k=0}^{n} A_k$.
- On en déduit que $B_n \cap (\cup_{k=0}^{n-1} B_k) = \emptyset$.
- Donc $\mathrm{P}(\cup_{n\in\mathbb{N}} A_n)) = \mathrm{P}(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathrm{P}(B_n) \leq \sum_{n\in\mathbb{N}} \mathrm{P}(A_n)$, car $B_n \subseteq A_n$.

Continuité

Proposition

Soit un espace probabilisé $(\Omega, \mathcal{T}, \mathrm{P})$. Pour toute suite décroissante (pour l'inclusion) d'événements $(A_n)_{n\in\mathbb{N}}$, la suite des $\mathrm{P}(A_n)$ converge et

$$\lim_{k\to+\infty} \mathrm{P}(A_k) = \mathrm{P}(\cap_{n\in\mathbb{N}} A_n)$$

De même, pour toute suite croissante d'événements $(A_n)_{n\in\mathbb{N}}$, la suite des $\mathrm{P}(A_n)$ converge et

$$\lim_{k\to+\infty} \mathrm{P}(A_k) = \mathrm{P}(\cup_{n\in\mathbb{N}} A_n)$$

Cas décroissant : $A_k = (\cap_{n\in\mathbb{N}} A_n) \sqcup (\sqcup_{k\leq n} A_n \setminus A_{n+1})$,
donc $\mathrm{P}(A_k) = \mathrm{P}(\cap_{n\in\mathbb{N}} A_n) + \sum_{k\leq n} \mathrm{P}(A_n \setminus A_{n+1})$
donc $\sum_{n=k}^{+\infty} \mathrm{P}(A_n \setminus A_{n+1}) \to_{k\to+\infty} 0$
donc $\mathrm{P}(A_k) \to_{k\to+\infty} \mathrm{P}(\cap_{n\in\mathbb{N}} A_n)$
Cas croissant : alors la suite des $\Omega \setminus A_n$ est décroissante.

Probabilité conditionnelle

Définition et Proposition

Soient un espace probabilisé $(\Omega, \mathcal{T}, \mathrm{P})$ et $B \in \mathcal{T}$ tels que $\mathrm{P}(B) > 0$. Pour tout $A \in \mathcal{T}$, on pose alors

$$P(A \mid B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$$

qu'on appelle la probabilité (conditionnelle) de A sachant B. De plus,

$$\begin{aligned} P(\cdot \mid B) : \quad & \mathcal{T} \to [0,1] \\ & A \mapsto P(A \mid B) \end{aligned}$$

est aussi une mesure de probabilité sur $(\Omega, \mathcal{T})$.

Proof.

- $\mathrm{P}(\Omega \mid B) = \frac{\mathrm{P}(\Omega \cap B)}{\mathrm{P}(B)} = 1$
- $\mathrm{P}(\sqcup_n A_n \mid B) = \frac{\mathrm{P}(\sqcup_n A_n \cap B)}{\mathrm{P}(B)} = \frac{\sum_n \mathrm{P}(A_n \cap B)}{\mathrm{P}(B)} = \sum_n \mathrm{P}(A_n \mid B).$

Probabilité conditionnelle (II)

Example

Une urne contient deux boules rouges et deux boules vertes. On tire une boule au hasard et on la met de côté, puis une deuxième. Soit A l'évènement "la deuxième boule est rouge" et B l'évènement "la première boule est verte". Alors $\mathrm{P}(A) = \frac{1}{2}$, mais $\mathrm{P}(A \mid B) = \frac{2}{3}$.

Formule des probabilités totales

Formule des probabilités totales

Soient $B_1, \ldots, B_n$ des parties disjointes de Ω telles que $\mathrm{P}(\sqcup_i B_i) = 1$ et pour tout i, $\mathrm{P}(B_i) > 0$. Alors pour tout évènement A on a

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)$$

Preuve

$$\begin{aligned}\mathrm{P}(A) &= \mathrm{P}(A \cap \sqcup_i B_i) = \mathrm{P}(\sqcup_i (A \cap B_i)) \\ &= \textstyle\sum_i \mathrm{P}(A \cap B_i) = \sum_i \mathrm{P}(A \mid B_i)\mathrm{P}(B_i).\end{aligned}$$

Exemple

Une urne contient n boules noires, b boules blanches et r boules rouges. On tire deux boules sans remise. Quelle est la probabilité de l'évènement "la deuxième boule tirée est noire" ?

Formule des probabilités totales (II)

Exemple

Une urne contient n boules noires, b boules blanches et r boules rouges. On tire deux boules sans remise. Quelle est la probabilité de l'évènement "la deuxième boule tirée est noire" ?

Soit $S = n + b + r$, soit DN l'événement "la deuxième boule tirée est noire", et N, B, R pour les premières boules.

$$\begin{aligned}
\mathrm{P}(DN) &= \mathrm{P}(DN \mid N) \cdot \mathrm{P}(N) + \mathrm{P}(DN \mid B) \cdot \mathrm{P}(B) + \mathrm{P}(DN \mid R) \cdot \mathrm{P}(R) \\
&= \frac{n-1}{S-1} \cdot \frac{n}{S} + \frac{n}{S-1} \cdot \frac{b}{S} + \frac{n}{S-1} \cdot \frac{r}{S} \\
&= \frac{n(n-1+b+r)}{S(S-1)} = \frac{n(S-1)}{S(S-1)} = \frac{n}{S}
\end{aligned}$$

Formule de Bayes

Formule de Bayes généralisée

Soient $B_1, \ldots, B_n$ des parties disjointes de Ω telles que $\mathrm{P}(\sqcup_i B_i) = 1$ et pour tout i, $\mathrm{P}(B_i) > 0$. Alors pour tout évènement A tel que $\mathrm{P}(A) > 0$ et tout k on a

$$\mathrm{P}(B_k \mid A) = \frac{\mathrm{P}(A \mid B_k)\mathrm{P}(B_k)}{\sum_{i=1}^{n} \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)}$$

Preuve

$\mathrm{P}(B_k \mid A)\mathrm{P}(A) = \mathrm{P}(A \cap B_k) = \mathrm{P}(A \mid B_k)\mathrm{P}(B_k)$ par définition des probabilités conditionnelles. Donc

$$\mathrm{P}(B_k \mid A) = \frac{\mathrm{P}(A \mid B_k)\mathrm{P}(B_k)}{\mathrm{P}(A)}$$

Or, d'après la formules des probabilités totales, on a $\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)$.

Formule de Bayes (II)

Formule de Bayes généralisée

Soient $B_1, \ldots, B_n$ des parties disjointes de Ω telles que $\mathrm{P}(\sqcup_i B_i) = 1$ et pour tout i, $\mathrm{P}(B_i) > 0$. Alors pour tout évènement A tel que $\mathrm{P}(A) > 0$ et tout k on a

$$\mathrm{P}(B_k \mid A) = \frac{\mathrm{P}(A \mid B_k)\mathrm{P}(B_k)}{\sum_{i=1}^{n} \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)}$$

Le problème du tricheur

Il existe des pièces normales (resp. biaisées) qui donnent pile avec une chance sur deux (resp. sur trois). On choisit une pièce dans un coffre, elle est biaisée avec probabilité x. On la lance et elle tombe sur face. Quelle est la probabilité qu'elle soit biaisée ?

Normale n, biaisée b, pile p, face f.

$$\mathrm{P}(b \mid f) = \frac{\mathrm{P}(f \mid b)\mathrm{P}(b)}{\mathrm{P}(f \mid b)\mathrm{P}(b) + \mathrm{P}(f \mid n)\mathrm{P}(n)} = \frac{\frac{2}{3}x}{\frac{2}{3}x + \frac{1}{2}(1-x)} = \frac{4x}{3+x}$$

Formule du double conditionnement

Proposition

Soit A, B, C trois évènements tels que $\mathrm{P}(A \cap B) > 0$. Alors

$$\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B \mid A)\mathrm{P}(C \mid A \cap B)$$

- $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A \cap B)\mathrm{P}(C \mid A \cap B)$
- $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B \mid A)$ car $0 < \mathrm{P}(A \cap B) \leq \mathrm{P}(A)$.

Proposition (Généralisation)

Soit $A_1, \ldots A_n$ des évènements tels que $\mathrm{P}(\cap_{i=1}^{n-1} A_i) > 0$. Alors

$$\mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{k=1}^{n} \mathrm{P}\left(A_k \mid \bigcap_{i=1}^{k-1} A_i\right)$$

Preuve par récurrence sur n.

Indépendance de deux évènements

Définition

Soient A et B deux évènements. Si $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$, on dit que A et B sont (mutuellement) indépendants.

Proposition

Deux évènements disjoints et de probabilités non nulles ne sont pas indépendants.

Preuve : $\mathrm{P}(A \cap B) = \mathrm{P}(\emptyset) = 0 \neq \mathrm{P}(A)\mathrm{P}(B)$

Proposition

Tout évènement est indépendant de tout évènement de probabilité 0 ou 1.

- Si $\mathrm{P}(B) = 0$ alors $\mathrm{P}(A \cap B) \leq \mathrm{P}(B) = 0 = \mathrm{P}(B)\mathrm{P}(A)$.
- Si $\mathrm{P}(B) = 1$ alors $\mathrm{P}(A)\mathrm{P}(B) = \mathrm{P}(A) = \mathrm{P}(A \cap B)$.

Indépendance de deux évènements (II)

Proposition

A et B sont indépendants ssi A et B^C le sont.

Preuve : Supposons $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$. Alors $\mathrm{P}(A) = \mathrm{P}((A \cap B) \sqcup (A \cap B^C)) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^C) = \mathrm{P}(A)\mathrm{P}(B) + \mathrm{P}(A \cap B^C)$. Donc $\mathrm{P}(A \cap B^C) = \mathrm{P}(A)(1 - \mathrm{P}(B)) = \mathrm{P}(A)\mathrm{P}(B^C)$.

Proposition

Soient A et B deux évènements tels que $\mathrm{P}(B) > 0$. Alors A et B sont indépendants ssi $\mathrm{P}(A \mid B) = \mathrm{P}(A)$.

Proof.

- Si A et B sont indépendants, alors $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A)$
- Si $\mathrm{P}(A \mid B) = \mathrm{P}(A)$, alors $\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B)\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B)$.

$\square$

Indépendance de plusieurs évènements

Définition

Soient $\mathcal{A}$ une famille d'évènements. On dit que ces évènements sont (mutuellement) indépendants si pour tout $A_1, \ldots, A_n \in \mathcal{A}$ distincts on a $\mathrm{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{P}(A_i)$.

Proposition

Soient $\mathcal{A}$ un famille d'évènements tels que $\mathrm{P}(A) > 0$ pour tout $A \in \mathcal{A}$. Alors ils sont mutuellement indépendants ssi pour tout $A_1, \ldots, A_n \in \mathcal{A}$ distincts on a $\mathrm{P}(\cap_{i=1}^{n-1} A_i) > 0$ et $\mathrm{P}(A_n \mid \cap_{i=1}^{n-1} A_i) = \mathrm{P}(A_n)$.

Preuve : si les évènements sont indépendants, alors
$\mathrm{P}(A_n \mid \cap_{i=1}^{n-1} A_i) = \frac{\mathrm{P}(\cap_{i=1}^n A_i)}{\mathrm{P}(\cap_{i=1}^{n-1} A_i)} = \frac{\prod_{i=1}^n \mathrm{P}(A_i)}{\prod_{i=1}^{n-1} \mathrm{P}(A_i)} = \mathrm{P}(A_n)$.
Réciproquement, montrons par récurrence sur n que pour tout $A_1, \ldots, A_n \in \mathcal{A}$ distincts on a $\mathrm{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{P}(A_i)$. Pour le cas inductif, $\mathrm{P}(\cap_{i=1}^{n+1} A_i) = \mathrm{P}(A_{n+1} \mid \cap_{i=1}^n A_i)\mathrm{P}(\cap_{i=1}^n A_i) = \mathrm{P}(A_{n+1})\mathrm{P}(\cap_{i=1}^n A_i)$ par hypothèse, égal à $\mathrm{P}(A_{n+1}) \prod_{i=1}^n \mathrm{P}(A_i)$ par HR, donc à $\prod_{i=1}^{n+1} \mathrm{P}(A_i)$.

Indépendance de plusieurs évènements (II)

Définition

Soient $\mathcal{A}$ une famille d'évènements. On dit que ces évènements sont (mutuellement) indépendants si pour tout $A_1, \ldots, A_n \in \mathcal{A}$ distincts on a $\mathrm{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{P}(A_i)$.

L'indépendance (mutuelle) n'équivaut pas à l'indépendance deux à deux.

On tire deux fois à pile ou face. Soient les événements suivants : A, le premier tirage est pile ; B, le deuxième tirage est face ; C, les deux tirages produisent le même résultat. Alors

- $\mathrm{P}(A) = \frac{1}{2}$ et $\mathrm{P}(B) = \frac{1}{2}$ et $\mathrm{P}(C) = \frac{1}{2}$.
- $\mathrm{P}(A \cap B) = \frac{1}{4}$ et $\mathrm{P}(B \cap C) = \frac{1}{4}$ et $\mathrm{P}(A \cap C) = \frac{1}{4}$.
- Donc $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$, etc.
- Mais $\mathrm{P}(A \cap B \cap C) = 0$.

Application mesurable

Définition

Soit $(E, \mathcal{T}_E)$ et $(F, \mathcal{T}_F)$ deux espaces mesurables. Une application $f : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$ est mesurable si pour tout $B \in \mathcal{T}_F$ on a $f^{-1}[B] \in \mathcal{T}_E$.

Proposition

La composée de deux applications mesurables est mesurable.

Preuve

Soient $f : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$ et $g : (F, \mathcal{T}_F) \to (G, \mathcal{T}_G)$ deux applications mesurables. Alors pour tout $B \in \mathcal{T}_G$ on a $g^{-1}[B] \in \mathcal{T}_F$, donc $(g \circ f)^{-1}[B] = f^{-1}[g^{-1}[B]] \in \mathcal{T}_E$.

Plus petite tribu rendant une application mesurable

$g : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$ est mesurable si $\forall B \in \mathcal{T}_F$ on a $g^{-1}[B] \in \mathcal{T}_E$.

Proposition

Soit $g : E \to F$.

1. $g^{-1}[F] = E$ et $g^{-1}[\emptyset] = \emptyset$.
2. $g^{-1}[F \setminus B] = E \setminus g^{-1}[B]$ pour tout $B \subseteq F$.
3. $g^{-1}[\cup_n B_n] = \cup_n g^{-1}[B_n]$ et $g^{-1}[\cap_n B_n] = \cap_n g^{-1}[B_n]$
4. Soient $g : E \to F$ et $\mathcal{T}_F$ tribu de F. Alors $g^{-1}[\mathcal{T}_F]$ est une tribu de E.

- $F \in \mathcal{T}_F$ donc $E = g^{-1}[F] \in g^{-1}[\mathcal{T}_F]$.
- Si $A \in g^{-1}[\mathcal{T}_F]$, soit $B \in \mathcal{T}_F$ tel que $A = g^{-1}[B]$. Ainsi $E \setminus A = E \setminus g^{-1}[B] = g^{-1}[F \setminus B]$, avec $F \setminus B \in \mathcal{T}_F$ (complément).
- Soit $(An)_n$ une suite de $g^{-1}[\mathcal{T}_F]$. Pour tout $n \in \mathbb{N}$, soit $B_n \in \mathcal{T}_F$ tel que $A_n = g^{-1}[B_n]$. Ainsi $\cup_n A_n = \cup_n g^{-1}[B_n] = g^{-1}[\cup_n B_n]$, avec $\cup_n B_n \in \mathcal{T}_F$ (union dénombrable).

$g^{-1}[\mathcal{T}_F]$ est la + petite tribu $\mathcal{T}_E$ tq $g : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$ soit mesurable.

Plus grande tribu rendant une application mesurable

Proposition

Soient $g : E \to F$ et $\mathcal{T}_E$ une tribu de E, et $\mathcal{D} := \{B \subseteq F \mid g^{-1}[B] \in \mathcal{T}_E\}$.
Alors $\mathcal{D}$ est une tribu de F.

Proof.

- $g^{-1}[F] = E \in \mathcal{T}_E$, donc $F \in \mathcal{D}$.
- Soit $B \in \mathcal{D}$, alors $g^{-1}[B] \in \mathcal{T}_E$, et $g^{-1}[F \setminus B] = E \setminus g^{-1}[B] \in \mathcal{T}_E$.
  Ainsi $F \setminus B \in \mathcal{D}$.
- Soit $(B_n)_n$ une suite de $\mathcal{D}$. Alors $g^{-1}[\cup_n B_n] = \cup_n g^{-1}[B_n] \in \mathcal{T}_E$.
  Ainsi $\cup_n B_n \in \mathcal{D}$.

$\square$

$\mathcal{D}$ est la plus grande tribu telle que $g : (E, \mathcal{T}_E) \to (F, \mathcal{D})$ soit mesurable.
Et toute tribu incluse dans $\mathcal{D}$ rend g mesurable.

Caractérisation de la mesurabilité d'une application

Proposition

Soit $g : (E, \mathcal{T}_E) \to (F, \mathcal{T}_F)$. Si $\mathcal{T}_F = \mathcal{T}(\mathcal{C})$, alors g est mesurable ssi pour tout $B \in \mathcal{C}$ on a $g^{-1}[B] \in \mathcal{T}_E$.

Proof.

- Si g est mesurable, alors on note que $\mathcal{C} \subseteq \mathcal{T}(\mathcal{C})$.
- Supposons que pour tout $B \in \mathcal{C}$ on a $g^{-1}[B] \in \mathcal{T}_E$. Alors $\mathcal{C}$ est inclus dans la tribu $\mathcal{D} := \{B \subseteq F \mid g^{-1}[B] \in \mathcal{T}_E\}$, donc $\mathcal{T}(\mathcal{C})$ aussi.

$\square$

Mesurabilité et continuité

Une topologie sur un ensemble E est une collection $\mathcal{O} \subseteq \mathcal{P}(E)$ telle que

- $E, \emptyset \in \mathcal{O}$,
- Pour tout $A, B \in \mathcal{O}$, $A \cap B \in \mathcal{O}$,
- Pour toute famille $(O_i)_{i \in I}$ de $\mathcal{O}$, $\cup_{i \in I} O_i \in \mathcal{O}$.

Les éléments de $\mathcal{O}$ sont appelés les ouverts de E.

- Soit $\mathcal{O}$ une topologie sur E. Alors $\mathcal{T}(\mathcal{O})$ est appelé la tribu borélienne de $(E, \mathcal{O})$.
- Soient $g : (E, \mathcal{O}_E) \to (F, \mathcal{O}_F)$. On dit que g est continue si $g^{-1}[O] \in \mathcal{O}_E$ pour tout $O \in \mathcal{O}_F$.

Proposition

Si $g : (E, \mathcal{O}_E) \to (F, \mathcal{O}_F)$ est continue, alors $g : (E, \mathcal{T}(\mathcal{O}_E)) \to (F, \mathcal{T}(\mathcal{O}_F))$ est mesurable.

Variable aléatoire

Définitions, notations, exemples

- Soit $(\Omega, \mathcal{T}, \mathrm{P})$ un espace probablisé et $n \in \mathbb{N}$. Alors toute application mesurable $X : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}^n, \mathcal{B}^n)$ est appelée variable aléatoire réelle si $n = 1$, à n dimensions sinon. ($\mathcal{B}^n = \mathcal{T}(\mathit{ouverts})$, tribu borélienne.)
- Informellement, une variable aléatoire transforme une observation fine d'une expérience aléatoire en une observation grossière.
  - On jette un dé, on observe la parité.
    $X_p : \{1, 2, 3, 4, 5, 6\} \to \{\mathit{pair}, \mathit{impair}\}$ avec les tribus discrètes.
  - On jette deux dés, on observe la somme. $X_s : [6]^2 \to [2..12]$.
- Si $B \in \mathcal{B}^n$ alors $X^{-1}[B] = \{\omega \in \Omega \mid X(\omega) \in B\}$ est dans $\mathcal{T}$ et est noté $\{X \in B\}$.
  - $X_p^{-1}[\{\mathit{pair}\}] = \{X_p = \mathit{pair}\} = \{2, 4, 6\}$.
  - $X_s^{-1}[\{2, 4\}] = \{X_s \in \{2, 4\}\} = \{(1, 1), (1, 3), (3, 1), (2, 2)\}$.

Variable aléatoire (II)

Définitions, notations, exemples

- Soit P une mesure de probabilité sur $(\Omega, \mathcal{T})$. Alors $\mathrm{P}(X^{-1}[B])$ est noté $\mathrm{P}\{X \in B\}$.
  - $\mathrm{P}\{X_p = pair\} = \mathrm{P}(\{2,4,6\})$.
  - $\mathrm{P}\{X_s \in \{2,4\}\} = \mathrm{P}(\{(1,1),(1,3),(3,1),(2,2)\})$
- Si $b \in \mathbb{R}^n$, alors $\mathrm{P}(X^{-1}[\{b\}]) = \mathrm{P}\{X \in \{b\}\}$ est noté $\mathrm{P}\{X = b\}$.
- Si $n = 1$ et $b \in \mathbb{R}$, alors $\mathrm{P}(X^{-1}\left[]-\infty, b]\right]) = \mathrm{P}\{X \in]-\infty, b]\}$ est noté $\mathrm{P}\{X \leq b\}$.
  - $\mathrm{P}\{X_s \leq 3\} = \mathrm{P}(\{(1,1),(1,2),(2,1)\})$
- Pour $A_1, \ldots A_n \in \mathcal{T}$, on note $\mathrm{P}(A_1, \ldots, A_n)$ au lieu de $\mathrm{P}(A_1 \cap \cdots \cap A_n)$.
- Si X_1 et X_2 sont deux variables aléatoires $(\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B}^n)$, alors $\mathrm{P}\{X_1^{-1}[B_1] \cap X_2^{-1}[B_2]\}$ est noté $\mathrm{P}\{X_1 \in B_1, X_2 \in B_2\}$.
  - $\mathrm{P}\{X_p^1 = pair, X_s \leq 3\} = \mathrm{P}(\{(2,1)\})$

Loi de probabilité d'une variable aléatoire

Définition

Soit $X : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}^n, \mathcal{B}^n)$ une variable aléatoire. On appelle loi de probabilité de X l'application $\mathrm{P}_X : \mathcal{B}^n \to [0, 1]$ telle que
$\mathrm{P}_X(B) := \mathrm{P}(X^{-1}[B])$.

Proposition

P_X est une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}^n)$.

Premièrement, $X^{-1}[B] \in \mathcal{T}$ pour tout $B \in \mathcal{B}^n$ car X est mesurable, donc P_X est bien définie. D'une part, $\mathrm{P}_X(\mathbb{R}^n) = \mathrm{P}(X^{-1}(\mathbb{R}^n)) = \mathrm{P}(\Omega) = 1$.
D'autre part, si $(B_k)_{k\in\mathbb{N}}$ est une suite de boréliens disjoints de $\mathcal{B}^n$, on a
$\mathrm{P}_X(\sqcup_k B_k) = \mathrm{P}(X^{-1}[\sqcup_k B_k]) = \mathrm{P}(\sqcup_k X^{-1}[B_k]) = \sum_k \mathrm{P}(X^{-1}(B_k)) = \sum_k \mathrm{P}_X(B_k)$.

Multiplication des notations : on a $\mathrm{P}_X(B) = \mathrm{P}(X^{-1}[B]) = \mathrm{P}\{X \in B\}$.

Fonction de répartition d'une variable aléatoire réelle

Définition

Soit $X : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B}^1)$ une variable aléatoire réelle. On appelle fonction de répartition de X l'application

$$\begin{aligned} F : &\mathbb{R} \to \mathbb{R} \\ &x \mapsto \mathrm{P}\{X \leq x\} = \mathrm{P}(X^{-1}[]-\infty, x]]) = \mathrm{P}_X(]-\infty, x]) \end{aligned}$$

Proposition

La fonction de répartition F d'une variable aléatoire réelle X a les propriétés suivantes.

1. $0 \leq F \leq 1$ et F est croissante,
2. F est continue à droite en tout $x \in \mathbb{R}$,
3. $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.

Fonction de répartition d'une variable aléatoire réelle (II)

Proposition

F est continue à droite en tout $x \in \mathbb{R}$. (Rappel $F(x) = \mathrm{P}\{X \leq x\}$)

Preuve

Soient $x \in \mathbb{R}$ et une suite $(x_n)_{\in \mathbb{N}}$ telle que $(x_n) \downarrow x$. Alors $\cap_{n \in \mathbb{N}}] - \infty, x_n] =] - \infty, x]$. Donc
$F(x) = \mathrm{P}_X(] - \infty, x]) = \mathrm{P}_X(\cap_{n \in \mathbb{N}}] - \infty, x_n])$,
Donc $F(x) = \lim_{n \to +\infty} \mathrm{P}_X(] - \infty, x_n])$ par continuité de P_X.
Donc $F(x) = \lim_{n \to +\infty} F(x_n)$.

Pourquoi la preuve ci-dessus ne peut-elle pas être adaptée pour prouver que F est continue à gauche ? Parce que pour tout $(x_n) \uparrow x$ on a $\cup_{n \in \mathbb{N}}] - \infty, x_n] =] - \infty, x[\neq] - \infty, x]$. Et aussi parce que F peut ne pas être continue à gauche : si $\mathrm{P}_X(\{x\}) = 1$ alors $F(x) = 1$ mais $F(x_n) = 0$ pour tout $x_n < x$.

Fonction de répartition d'une variable aléatoire réelle (III)

Proposition

$\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$. (Rappel $F(x) = \mathrm{P}\{X \leq x\}$)

Preuve

Par croisance et "bornitude", F admet des limites en $-\infty$ et $+\infty$.
De plus $\lim_{-\infty} F = \lim_{n \to +\infty} F(-n)$.
Donc $\lim_{-\infty} F = \lim_{n \to +\infty} \mathrm{P}_X(] - \infty, -n])$.
Donc $\lim_{-\infty} F = \mathrm{P}_X(\cap_{n \in \mathbb{N}}] - \infty, -n])$.
Donc $\lim_{-\infty} F = \mathrm{P}_X(\emptyset) = 0$.
De même $\lim_{+\infty} F = \lim_{n \to +\infty} F(n) = \lim_{n \to +\infty} \mathrm{P}_X(] - \infty, n]) = \mathrm{P}_X(\cup_{n \in \mathbb{N}}] - \infty, n]) = \mathrm{P}_X(\mathbb{R}) = 1$

Fonctions de répartition de variables aléatoires à valeurs non réelles.

Pour $X, Y : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B})$, le vecteur (X, Y) est une variable aléatoire. On pose $F(x, y) = \mathrm{P}(X \leq x, Y \leq y)$.

Tribu engendrée par une variable aléatoire

Définition

On appelle tribu engendrée par X, notée $\mathcal{T}(X)$ la plus petite tribu rendant X mesurable.

Proposition

$\mathcal{T}(X) = X^{-1}[\mathcal{B}^n]$

Indépendance de classes d'évènements

Définition

Soient $(\Omega, \mathcal{T})$ un espace mesurable. On dit que deux classes d'évènements $\mathcal{C}_1$ et $\mathcal{C}_2$ sont indépendantes si A_1 et A_2 sont indépendants pour tout $A_1 \in \mathcal{C}_1$ et $A_2 \in \mathcal{C}_2$. (Cela se généralise à un ensemble de classes.)

Lemme

Soient $(\Omega, \mathcal{T}, \mathrm{P})$ un espace probabilisé, $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements disjoints de $\mathcal{T}$, et $B \in \mathcal{T}$ indépendant de chaque A_n. Alors B est aussi indépendant de $\sqcup_{n\in\mathbb{N}} A_n$.

Preuve

$\mathrm{P}((\sqcup_n A_n) \cap B) = \mathrm{P}(\sqcup_n (A_n \cap B)) = \sum_n \mathrm{P}(A_n \cap B) = \sum_n \mathrm{P}(A_n)\mathrm{P}(B) = (\sum_n \mathrm{P}(A_n))\mathrm{P}(B) = \mathrm{P}(\sqcup_n A_n)\mathrm{P}(B)$.

Indépendance de classes d'évènements (II)

Proposition

Soient $(\Omega, \mathcal{T})$ un espace mesurable et deux classes d'évènements $\mathcal{C}_1$ et $\mathcal{C}_2$ indépendantes et stables par intersection finie. Alors $\mathcal{T}(\mathcal{C}_1)$ et $\mathcal{T}(\mathcal{C}_2)$ sont aussi indépendantes.

- Il suffit de montrer l'indépendance de $\mathcal{C}_2$ et $\mathcal{T}(\mathcal{C}_1)$, qui est stable par intersection finie, puis de leur appliquer ce résultat partiel.
- Soit $C \in \mathcal{C}_2$. Il suffit de mq C est indépendant de tout $X \in \mathcal{T}(\mathcal{C}_1)$.
- Soit $S := \{X \in \mathcal{T}(\mathcal{C}_1) \mid X \text{ et } C \text{ indép}\} \supseteq \mathcal{C}_1$. Il suffit de mq S est une tribu.
  - $\Omega \in S$ car $\mathrm{P}(\Omega \cap C) = \mathrm{P}(\Omega)\mathrm{P}(C)$.
  - Soit $X \in S$ alors $\Omega \setminus X \in \mathcal{T}(\mathcal{C}_1)$ et C indépendant de $\Omega \setminus X \in S$.
  - Soit $(A_n)_n$ une suite de S. Soit $B_n := A_n \setminus \cup_{i=0}^{n-1} A_i = A_n \setminus \cup_{i=0}^{n-1} B_i$. Alors $B_n \in$ par récurrence sur n. Alors $\mathrm{P}(\cup_n A_n \cap C) = \mathrm{P}(\cup_n B_n \cap C) = \sum_n \mathrm{P}(B_n \cap C) = \sum_n \mathrm{P}(B_n)\mathrm{P}(C) = \mathrm{P}(\cup_n A_n)\mathrm{P}(C)$.

Variables aléatoires indépendantes

Définition

Soient deux variables aléatoires X, Y définies sur le même espace probabilisé (de départ). On dit qu'elles sont indépendantes si $\mathcal{T}(X)$ et $\mathcal{T}(Y)$ le sont. Cela se généralise à un ensemble de variables aléatoires.

Reformulation

Soient deux variables aléatoires $X : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}^n, \mathcal{B}^n)$ et $Y : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}^m, \mathcal{B}^m)$. Elles sont indépendantes, si pour tout $A \in \mathcal{B}^n$ et $B \in B^m$, on a $\mathrm{P}\{X \in A, Y \in B\} = \mathrm{P}\{X \in A\}\mathrm{P}\{Y \in B\}$.

Proposition

Soient $X, Y : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B}^1)$ deux variables aléatoires réelles ; F_X, F_Y et $F_{(X,Y)}$ les fonctions de répartition de X, Y et (X, Y). Les variables sont indépendantes ssi $F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$ pour tout $x, y \in \mathbb{R}$.

Variables aléatoires indépendantes (II)

Proposition

Soient $X, Y : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B}^1)$ deux variables aléatoires réelles ; F_X, F_Y et $F_{(X,Y)}$ les fonctions de répartition de X, Y et (X, Y). Les variables sont indépendantes ssi $F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$ pour tout $x, y \in \mathbb{R}$.

Si X et Y sont indépendantes alors pour tout $x, y \in \mathbb{R}$ on a
$F_{(X,Y)}(x, y) = \mathrm{P}\{X \leq x, Y \leq y\} = \mathrm{P}\{X \leq x\}\mathrm{P}\{Y \leq y\} = F_X(x)F_Y(y)$.

Réciproquement, soient $\mathcal{C}_X$ la classe des évènements $\{X \leq x\} = X^{-1}[]-\infty, x]]$ pour $x \in \mathbb{R}$. Cette classe est stable par intersection finie, car la classe des $]-\infty, x]$ l'est et par propriété de X^{-1} par rapport à l'intersection. De même pour $\mathcal{C}_Y$.
L'hypothèse $F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$ pour tout $x, y \in \mathbb{R}$ montre que $\mathcal{C}_X$ et $\mathcal{C}_Y$ sont indépendantes. Par un théorème ci-dessus, $\mathcal{T}(\mathcal{C}_X) = X^{-1}[\mathcal{B}^1] = \mathcal{T}(X)$ et $\mathcal{T}(\mathcal{C}_Y) = Y^{-1}[\mathcal{B}^1] = \mathcal{T}(Y)$ sont aussi indépendantes, donc X et Y sont indépendantes.

Indépendance de variables aléatoires discrètes

Définition

Une variable aléatoire $X : \Omega \to \mathbb{R}^n$ est dite discrète si P_X est discrète, i.e. s'il existe $(x_k)_{k\in\mathbb{N}}$ dans $\mathbb{R}^n$ tel que $\mathrm{P}_X(\{x_k \mid k \in \mathbb{N}\}) = 1$.

Proposition

Soit X, Y deux variables aléatoires réelles discrètes Elles sont indépendantes ssi pour tout i, j on a

$$\mathrm{P}\{X = x_i, Y = y_j\} = \mathrm{P}\{X = x_i\}\mathrm{P}\{Y = y_j\}$$

- Si X et Y sont indép., alors chaque $X^{-1}[\{x_i\}] = \{X = x_i\}$ est indép. de chaque $Y^{-1}[\{y_j\}] = \{Y = y_j\}$, car les singletons sont boréliens.
- Réciproquement, soit $A, B \in \mathcal{B}^n$. Alors $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(\cup_{i,j|x_i\in A\wedge y_j\in B}\{X = x_i, Y = y_j\}) = \sum_{i,j|x_i\in A\wedge y_j\in B} \mathrm{P}\{X = x_i, Y = y_j\} = \sum_{i,j|x_i\in A\wedge y_j\in B} \mathrm{P}\{X = x_i\}\mathrm{P}\{Y = y_j\} = (\sum_{i|x_i\in A} \mathrm{P}\{X = x_i\})(\sum_{j|y_j\in B} P\{Y = y_j\} = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$

Transformation de variables aléatoires

Proposition

Soient X une variable discrète, de loi $\mathrm{P}_X = \sum_{k\in\mathbb{N}} \alpha_k 1_{x_k}$, où $1_x(B) := 1$ si $x \in B$, et 0 sinon, et $\alpha_k = \mathrm{P}\{X = x_k\}$. Soit $g : \mathbb{R}^n \to \mathbb{R}^p$ mesurable. Alors $g \circ X : \Omega \to \mathbb{R}^p$ est une variable aléatoire discrète de loi

$$\mathrm{P}_{g\circ X} = \sum_{k\in\mathbb{N}} \alpha_k 1_{g(x_k)}$$

De plus pour tout $z \in \mathbb{R}^p$ on a

$$\mathrm{P}_{g\circ X}(\{z\}) = \mathrm{P}_X\{g = z\} = \mathrm{P}\{g \circ X = z\}$$

Preuve

$\mathrm{P}_{g\circ X}(\{z\}) = \mathrm{P}\{g \circ X = z\} = \mathrm{P}\{X \in g^{-1}(z)\} = \mathrm{P}_X(g^{-1}(z)) = \sum_{k\in\mathbb{N}} \alpha_k 1_{x_k}(g^{-1}(z))$. Or $1_{x_k}(g^{-1}(z)) = 1_{g(x_k)}(z)$, d'où le résultat.

Espérance mathématique

Définition

L'espérance mathématique d'une variable aléatoire réelle discrète X de loi $\mathrm{P}_X = \sum_k \alpha_k 1_{x_k}$ (avec $\alpha_k = \mathrm{P}\{X = x_k\}$) est
$\mathbb{E}[X] := \sum_k \alpha_k x_k = \sum_k \mathrm{P}\{X = x_k\} x_k$. N'est pas toujours bien définie.

L'espérance mathématique est une moyenne arithmétique pondérée.

Théorème de transfert

Soit $X : (\Omega, \mathcal{T}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B})$ tel que Ω est dénombrable et $X[\Omega] = \{x_n\}_{n\in\mathbb{N}}$. Alors

$$\sum_{\omega\in\Omega} X(\omega)\mathrm{P}(\{\omega\}) = \sum_{n\in\mathbb{N}} x_n \mathrm{P}_X(\{x_n\}) = \mathbb{E}[X]$$

Ci-dessus, la somme de droite “regroupe” les ω de même image x_n par X.

Espérance mathématique (II)

Propositions

Soient X et Y deux variables aléatoires réelles sur le même espace probabilisé.

- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- $X = a \Rightarrow \mathbb{E}[X] = a$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$
- $X \leq Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$
- Si X et Y sont indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Preuve (cas discret)

$\mathbb{E}[XY] = \sum_{i,j} \mathrm{P}\{X = x_i, Y = y_j\} x_i y_j = \sum_{i,j} \mathrm{P}\{X = x_i\}\mathrm{P}\{Y = y_j\} x_i y_j = (\sum_i \mathrm{P}\{X = x_i\} x_i) \sum_j (\mathrm{P}\{Y = y_j\} y_j) = \mathbb{E}[X]\mathbb{E}[Y]$.

Variance

Définition

La variance d'une variable aléatoire X est $V[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Propositions

- $V[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- Pour tout $a \in \mathbb{R}$, on a $V[X] \leq \mathbb{E}[(X - a)^2]$
- $V[aX] = a^2 V[X]$
- Si $X_1, \ldots, X_n$ sont indépendantes alors $V[\sum_k X_k] = \sum_k V[X_k]$.

- $V[X] = \mathbb{E}[X^2 - 2\mathbb{E}(X)X + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2$
- $\mathbb{E}[(X - \mathbb{E}[X] + b)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] + 2b\mathbb{E}[X - \mathbb{E}[X]] + \mathbb{E}[b^2]$. Or $\mathbb{E}[b^2] = b^2 \geq 0$ et $\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$.
- $V[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) = V[X] + V[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$. Le dernier terme est nul si X et Y sont indépendantes.

Inégalités de Markov et de Tchebychev

Proposition (Inégalité de Markov)

Si $0 \leq X$, alors pour tout $t > 0$ on a $\mathrm{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$.

$$\mathbb{E}[X] = \sum_{x_k < t} \alpha_k x_k + \sum_{x_k \geq t} \alpha_k x_k \geq t \sum_{x_k \geq t} \alpha_k = t\mathrm{P}(X \geq t).$$

Au plus 20% des gens ont un revenu au moins 5 fois supérieur à la moyenne.

Proposition (Inégalité de Tchebychev)

Pour tout $t > 0$ on a $\mathrm{P}\{|X - \mathbb{E}[X]| \geq t\} \leq \frac{V[X]}{t^2}$

$\mathrm{P}\{|X - \mathbb{E}[X]| \geq t\} = \mathrm{P}\{(X - \mathbb{E}[X])^2 \geq t^2\} \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2}$ par l'inégalité de Markov.

Lois de probabilité usuelles

- Variable aléatoire de Bernoulli: X telle que $\mathrm{P}(X \in \{0,1\}) = 1$. On dit alors que $p := \mathrm{P}\{X = 1\}$ est le paramètre de X.
- Variable aléatoire géométrique de paramètre $p \in [0,1]$: pour tout $n \in \mathbb{N}$ on a $\mathrm{P}\{X = n\} = (1-p)^n p$, la probabilité que n échecs indépendants d'une épreuve de Bernoulli soient suivis d'un succès. On vérifie que $\sum_{n \in \mathbb{N}} (1-p)^n p = 1$. (On "décale" parfois la définition sur $\mathbb{N} \setminus \{0\}$. Alors $\mathrm{P}\{X = n\} := (1-p)^{n-1} p$.)
- Variable aléatoire binomiale: X est de paramètres $n \in \mathbb{N} \setminus \{0\}$ et $p \in [0,1]$, si pour tout $k \in [\![0, n]\!]$ on a $\mathrm{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$. On note que $\sum_{k=0}^{n} \mathrm{P}\{X = k\} = 1$ et que $\mathrm{P}\{X = k\}$ est la probabilité que la répétition de n épreuves de Bernoulli indépendantes produisent k succès et $n-k$ échecs.