Cours n°1:

Introduction
Rappels de probabilités
Variables aléatoires
Principales lois discrètes et distributions continues

LAKHEL El Hassan

Département: Informatiques et Réseaux-Télécom
Filière G.I.I.A.
ENSA, A. U. 2022-2023

Plan du module

1. Rappels de probabilités : variables aléatoires, loi des variables aléatoires, fonctions de répartition, fonctions caractéristiques, loi des grands nombres, théorème central limite.

2. Vecteurs aléatoires gaussiens. Théorème d'existence de la densité gaussienne.

3. Introduction générale aux processus stochastiques discrets et continus. Processus de Bernoulli, Marche aléatoire, Processus de Poisson, Processus de comptage, processus de renouvellement, Mouvement Brownien. Applications.

4. Chaînes de Markov : Matrice de transition d'une chaîne de Markov, dynamique markovienne, matrice stochastique. Exemples de chaîne de Markov, Equation de Chapman-Kolmogorov, Le graphe de transition, Caractérisation de la loi d'une chaîne de Markov.

5. Modélisation d'une file d'attente : Constitution d'une file d'attente. Modélisation des arrivées. Modélisation du temps de service. Modélisation de la longueur de la queue. Étude de la file en régime stationnaire. Autres modèles de files d'attente.

6. Projets : Simulation de phénomènes aléatoires et méthode de Monte Carlo.

EVALUATION : L'évaluation du cours se fera par deux notes :

- une note sur les deux devoirs surveillés (80%).
- une note sur un projet que vous ferez en binôme et pour lequel un rapport en LATEX de 10 à 15 pages devra être rédigé (20%).

1. Etude d'une file d'attente. Les files d'attente peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. On les rencontre dans les domaines d'activité les plus divers (guichet de poste, traffic routier, central téléphonique, atelier de réparation,...). L'étude mathématique des phénomènes d'attente constitue un champ d'application important des processus stochastiques.
   La théorie des files d'attente consiste en l'étude de systèmes où des clients se présentent à un dispositif de service, appelé serveur. Puisqu'un client occupe le serveur pendant un certain temps, les autres clients doivent attendre avant d'être servis, formant ainsi une file d'attente. On étudie la longueur de la file d'attente, en fonction des paramètres qui interviennent dans la modélisation, à savoir la "loi" d'arrivée des clients, "loi" de la durée des services, combien de serveurs ? comment s'organise la file ? On se demande si la file à tendance à se diminuer ou au contraire à augmenter.

   Quelques exemples d'application :

   - Réseaux informatiques : serveur = routeur, client = paquet.
   - Ateliers (job shop) : serveur = machine, client = tâche
   - Trafic aérien : serveur = trajet, client = avion.
   - Télécommunications (téléphonie, call-centers)
   - Serveurs informatiques...

Objectif : En ingénierie, on s'intéresse à des métriques de performance des files d'attente. D'innombrables questions se posent naturellement, afin d'optimiser la rentabilité de certains services et de diminuer les attentes. Par exemple,

- quel est le temps moyen d'attente ?
- quel est le nombre moyen de clients dans la file ?
- quel est le nombre de serveurs occupés ?
- quelle est la probabilité que la file soit vide / pleine ?
- quelle est la longueur de la queue à un instant donné ?
- combien de serveurs faudrait-il au minimum pour éviter la saturation de la salle d'attente ?
- ...

L'objectif de ce cours est d'étudier la structure et de calculer des valeurs caractéristiques permettant de d'ecrire les performances de tels systèmes.

Chaîne de Markov et l'algorithme PageRank de Google

On peut aussi modéliser des phénomènes aléatoires plus complexes. Donnons un exemple :

2. Le succès de google se base sur l'algorithme Pagerank. Cet algorithme permet de trier les résultats des recherches.
Que fait un moteur de recherche ?
L'utilisateur lance une requête : mots clés cherchés.
Le moteur de recherche exécute :

    - liste des pages contenant les mots clés
    - tri par ordre de pertinence
    - affichage des résultats

    Il y a donc plusieurs aspects dont

    - modélisation mathématique : comment définir/calculer la pertinence ? (personne n'est chargée de lire toutes les pages web !)
    - ressource informatique : stockage, traitement d'une quantité énorme d'informations

    L'idée : pour mesurer l'importance d'une page est de travailler sur le fait qu'il y a des liens reliant les pages les unes aux autres. **Google voit le web comme un graphe**, dont les sommets sont les pages et les arcs sont les liens allant d'une page vers l'autre. Chaque flèche est affectée d'un poids, correspondant à la probabilité qu'on a, lorsqu'on quitte une page, d'aller vers la page suivante.
    si la page P_j contient un lien vers la page P_i on matérialise cela par une flèche $P_j \to P_i$.
    on utilise les chaîne de Markov ; les marches aléatoires sur les graphes et le Théorème de Perron-Frobenius pour calculer l'importance d'une page, aussi appelée PageRank. (Voir le dernier chapitre pour plus de détails).

Prenons l'exemple suivant (où le web comporte 5 pages) :
Dans cet exemple, il est raisonnable de penser que la "Page 4" est la plus importante car c'est vers cette page que le plus d'autres pages pointent. Ensuite, la "Page 5" est plus importante que la "Page 2", car si elles ont chacune un lien entrant, celui de la "Page 5" vient de la page la plus importante, etc...
Le point de départ du PageRank est de donner des valeurs numériques à l'importance d'une page modélisant le phénomène précédent. L'idée est d'associer à chaque page la probabilité qu'on a, lorsqu'on visite le web, d'être à cette page donnée. On peut représenter la pertinence par un nombre ou un score positif avec la convention que plus le score est grand plus la page est « importante ». Ceci peut se faire à l'aide de l'algorithme suivant. On introduit la matrice de transition du graphe, qui s'écrit

$$G = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Le coefficient $G_{i,j}$ représente la probabilité, étant sur la page i, d'aller sur la page j. On introduit aussi le vecteur initial

$$\pi^{(0)} = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}),$$

on a autant de chances d'être sur n'importe quel page. Alors, le vecteur $\pi^{(n)}$, des probabilités des états du système à l'instant n, est obtenu en effectuant le produit du vecteur ligne $\pi^{(0)}$ par la matrice G^n .
Autrement dit :

$$\pi^{(n)} = \pi^{(0)} G^n.$$

Il se trouve que la suite $\pi^{(n)}$ converge (voir chapitre sur les CM). Les termes de la limite donnent le poids de chaque page, le "Pagerank".

THÉORIE DES PROBABILITÉS

- Une théorie mathématique pour quantifier le Hasard.
  - La théorie des probabilités a mis beaucoup de temps à émerger.
  - Le mot Hasard est un mot d'origine arabe : az-zahr.
  - Au cours du 17ème siècle, le calcul probabiliste commence réellement à être rigoureusement développé par Pascal,
  - 19ème siècle et début 20ème siècle : essor des probabilités grâce aux méthodes d'analyse.
    - ◇ Calcul intégral et différentiel : (Laplace, Gauss)
    - ◇ Théorie de la mesure : (Borel - Lebesgue)

  - La période moderne, caractérisée par l'étude systématique des processus aléatoires, débute vers 1930. Modéliser des phénomèmes aléatoires qui évoluent au cours du temps : Processus de Markov, mouvement Brownien, processus de Poisson,...

- Les prérequis de ce cours sont des connaissances mathématiques dispensées durant les deux premières années d'un cursus universitaire. Nous allons developper les notions dont nous avons besoin au fur et à mesure de ce module.

- L'objet de la théorie des probabilités est l'analyse mathématique de phénomènes dans lesquels le hasard intervient. Ces phénomènes sont appelés des expériences aléatoires.

- Les processus stochastiques servent à la modélisation des phénomnes aléatoires qui évoluent dans temps. Le terme modéliser signifie ici l'opération qui consiste à associer à une expérience aléatoire un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et une famille $(X_t)_{t\in T}$ de variables aléatoires définies sur le même espace de probabilité. Le plus souvent, T est un ensemble ordonné qui joue le rôle du temps.

- La notion d'une chaînes de Markov fait partie de cette branche de la théorie des probabilités.

- Les processus stochastiques trouvent des applications dans plusieurs domaines tels que :
    - ♦ Iternet : Google par exemple fonctionne avec des chaînes de Markov.
    - ♦ Résaux et télécom : la théorie des files d'attente est basée essentiellement sur la théorie des probabilités.
    - ♦ Les mathématiques financières, le traitement d'images, etc.

EXPÉRIENCE ALÉATOIRE

Certaines expériences entraînent des résultats aléatoires, c'est-à-dire qui dépendent directement du hasard. Nous les appelons des **expériences aléatoires**.
En théorie des probabilités, le but est la modélisation des expériences aléatoires, le terme **modéliser** désigne l'opération qui consiste à associer à une expérience aléatoire trois objets mathématiques, notés et appelés généralement Ω, l'univers, $\mathcal{F}$, l'ensemble des événements et $\mathbb{P}$, la probabilité.

EXPÉRIENCE ALÉATOIRE

La phrase typique dans la théorie et les exercices est la suivante : Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Essayons d'expliciter chacun des trois termes de ce triplet. C'est un peu abstrait, on verra des exemples plus usuels dans les paragraphes suivants.

Définition

Une expérience est dite aléatoire si, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prévoir le résultat par avance.

Définition

L'ensemble des résultats d'une expérience aléatoire est appelé l'univers (ou espace d'états). Il est noté Ω.

Exemple

1. *Le jet d'un dé :* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. *Le jet successif de n pièces de monnaie :* $\Omega = \{P, F\}^n$ *(pour* $n = 2$*, on a* $\Omega = \{PP, PF, FP, FF\} = \{P, F\}^2$*).*
3. *La durée de vie d'une ampoule :* $\Omega = \mathbb{R}^+$.
4. *La durée d'une communication téléphonique :* $\Omega = \mathbb{R}^+$.
5. *Cours d'un actif financier sur un intervalle de temps* $[a, b]$ *:* $\Omega = \mathcal{C}([a, b], \mathbb{R}^+)$.

ÉVÉNEMENT

Définition

Un événement est une partie A de Ω, c'est un fait lié à une expérience qui peut se produire ou non.

ÉVÉNEMENT

Définition

Un événement est une partie A de Ω, c'est un fait lié à une expérience qui peut se produire ou non.

Exemple

Dans nos situations précédentes, A pourrait, par exemple, être :

1. $A_1 = \{2, 4, 6\}$: *"obtenir un nombre pair."*
2. $A_2 = \{P\} \times \{P, F\}^{n-1}$: *" Le premier lancer est pile."*
3. $A_3 = [100, +\infty[$: *"l'ampoule fonctionne plus de cent heures."*
4. $A_4 = [3, 7]$: *"la durée de communication est entre 3 et 7 min.*

Ainsi, un événement aléatoire est représenté par l'ensemble des résultats pour lesquels il est réalisé.

Tribu ou $\sigma-$algèbre

En général, on ne peut pas prendre toutes les parties de Ω comme événements, on doit se limiter à des familles vérifiant certaines propriétés :

Définition

Soit Ω un ensemble. Une famille $\mathcal{F}$ de parties de Ω est appelée une tribu si elle vérifie les propriétés suivantes :

- $\Omega \in \mathcal{F}$
- *(stabilité par complémentaire) : si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$*
- *(stabilité par union dénombrable) : si $(A_i)_{i\in\mathbb{N}}$ sont des éléments de $\mathcal{F}$, alors $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.*

Définition

Soit Ω un ensemble muni d'une tribu $\mathcal{F}$, le couple $(\Omega, \mathcal{F})$ est appelé espace mesurable, et les éléments de $\mathcal{F}$ sont appelés des événements.

Exercice

1. *Vérifier que* $\{\emptyset, \Omega\}$ *est une tribu (la tribu grossière).*
2. *Vérifier que* $\mathcal{P}(\Omega)$ *est une tribu.*
3. *Soit* A *une partie de* Ω*. Montrer que* $\{\emptyset, A, A^c, \Omega\}$ *est une tribu sur* Ω*.*
4. *Soit* Ω *un ensemble, et* A *et* B *deux parties de* Ω*. on pose*

$$\mathcal{F} = \{\emptyset, A, B, A \cap B, A \cup B, A^c, B^c, (A \cap B)^c, (A \cup B)^c, A \cap B^c, A^c \cap B, A \cup B^c, A^c \cup B, \Omega\}$$

Montrer que $\mathcal{F}$ *est une tribu.*

Remarque

La plus petite tribu de Ω *est* $\{\emptyset, \Omega\}$ *, tandis que la plus grande est* $\mathcal{P}(\Omega)$*.*

Proposition

Soit Ω *un ensemble muni d'une tribu* $\mathcal{F}$*, alors :*

1. $\emptyset$ *est dans* $\mathcal{F}$*,*
2. *Pour tout* A *et* B *de* $\mathcal{F}$*, on a* $A \cap B$*,* $A \cup B$ *et* $A \backslash B$ *sont dans* $\mathcal{F}$*.*
3. *Si les* $(A_i)_{i \in \mathbb{N}}$*, sont des éléments de* $\mathcal{F}$*, alors* $\cap_{i \in \mathbb{N}} A_i$ *est un élément de* $\mathcal{F}$ *(stabilité par intersection dénombrable).*

Preuve :

1. $\emptyset = \overline{\Omega}$ et $\Omega \in \mathcal{F}$, donc $\emptyset \in \mathcal{F}$.

2. $A \cup B \in \mathcal{F}$ par définition.

   $A \cap B = \overline{\overline{A} \cup \overline{B}} \in \mathcal{F}$ car $\overline{A} \cup \overline{B} \in \mathcal{F}$

   $A \backslash B = A \cap \overline{B} \in \mathcal{F}$ d'après ce qui précède.

3. $\cap_{i \in \mathbb{N}} A_i = \overline{\cup_{i \in \mathbb{N}} \overline{A_i}} \in \mathcal{F}$.

Remarque

Une tribu est stable par réunion et intersection finie.

Définition d'une probabilité

Définition

Soit Ω un ensemble muni d'une tribu $\mathcal{F}$. Une application $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ est une probabilité si elle vérifie les propriétés suivantes :

i) $\mathbb{P}(\Omega) = 1$

ii) *($\sigma-$additivité) Si les $(A_i)_{i\in\mathbb{N}}$ sont des éléments de $\mathcal{F}$ deux à deux disjoints, alors*

$$\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est alors appelé un espace de probabilité.

Remarque

1. *Les événements sont donc les parties de Ω auxquels on saura attribuer une probabilité de se réaliser.*
2. *Si Ω est **fini,** on peut remplacer ii) par ii)' pour tout A, B de $\mathcal{P}(\Omega)$ tels que $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.*

Exemple

1. *Jet d'un dé :* $\Omega = \{1, 2, ..., 6\}$*, les faces sont équiprobables. On prend* $\mathcal{F} = \mathcal{P}(\Omega)$ *et on définit* $\mathbb{P}$ *par :*

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = ... = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

On a donc,

$$\mathbb{P}(\{2, 3\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = \frac{1}{3}.$$

Exemple

1. *Jet d'un dé :* $\Omega = \{1, 2, ..., 6\}$, *les faces sont équiprobables. On prend* $\mathcal{F} = \mathcal{P}(\Omega)$ *et on définit* $\mathbb{P}$ *par :*
$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = ... = \mathbb{P}(\{6\}) = \frac{1}{6}.$$
On a donc,
$$\mathbb{P}(\{2,3\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = \frac{1}{3}.$$

2. *Jet d'une pièce de monnaie* $\Omega = \{P, F\}$, *si la pièce est équilibrée, on choisit :*
$$\mathbb{P}(\{P\}) = \mathbb{P}(\{F\}) = \frac{1}{2}.$$

Attention ! Le mot probabilité désigne donc deux choses différentes : l'application et le nombre associé par cette application à un événement. Le contexte permet en général de lever toute ambiguïté.

Proposition

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

i) $\mathbb{P}(\emptyset) = 0$

ii) *si* $(A_i)_{1 \leq i \leq n}$ *sont des éléments de* $\mathcal{F}$ *deux à deux disjoints, alors*

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

iii) *si* A *est dans* $\mathcal{F}$*, alors* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

iv) *si* A *et* B *sont des éléments de* $\mathcal{F}$ *tels que* $A \subset B$*, alors* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

v) *si* A *et* B *sont deux éléments de* $\mathcal{F}$*, alors*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Preuve :...

Démonstration.

i) On applique le ii) de la définition à la famille d'événements disjoints $(\Omega, \emptyset, \emptyset, ...)$:

$$1 = \mathbb{P}(\Omega) + \sum_{i=1}^{\infty} \mathbb{P}(\emptyset).$$

La série dans le membre de droite ne converge alors que si $\mathbb{P}(\emptyset) = 0$.

ii) On applique le ii) de la définition à la famille d'événements disjoints $(A_1, A_2, \dots, A_n, \emptyset, \emptyset, \dots)$ en utilisant que $\mathbb{P}(\emptyset) = 0$:

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i=n+1}^{n} \mathbb{P}(\emptyset) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

iii) On applique ii) à la famille d'événements disjoints (A, A^c) :
$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$ d'après le i) de la définition.

Démonstration.[suite]

iv) Soit A et B deux événements tels que $A \subset B$. Comme $B = (B \cap A) \cup (B \cap A^c)$, avec $(B \cap A) \cap (B \cap A^c) \subset A \cap A^c = \emptyset$, on peut appliquer ii) :

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B \cap A) + \underbrace{\mathbb{P}(B \cap A^c)}_{\geq 0} \\ &\geq \mathbb{P}(B \cap A) = P(A) \end{aligned}$$

v) On écrit $A \cup B$ comme la réunion disjointe $A \cap B^c$, $A \cap B$ et $B \cap A^c$, et on remarque que A est la réunion disjointe $A \cap B^c$, et $A \cap B$, tandis que B est la réunion disjointe $A \cap B$ et $B \cap A^c$. On obtient donc :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) + \mathbb{P}(B \cap A^c) \\ &= (\mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B)) + (\mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c)) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned}$$

Proposition

vi) *si* $(A_i)_{i\in\mathbb{N}}$ *sont des éléments de* $\mathcal{F}$, *alors*

$$\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$$

vii) *(Limite croissante) si* $(A_i)_{i\in\mathbb{N}}$ *forment une suite croissante d'éléments de* $\mathcal{F}$, *c'est à dire s'ils vérifient* $\forall i \in \mathbb{N}\ A_i \subseteq A_{i+1}$, *alors*

$$\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = lim_{i\longrightarrow+\infty}\mathbb{P}(A_i).$$

viii) *(Limite décroissante) si* $(A_i)_{i\in\mathbb{N}}$ *forment une suite décroissante d'éléments de* $\mathcal{F}$, *c'est à dire s'ils vérifient* $\forall i \in \mathbb{N}\ A_{i+1} \subseteq A_i$, *alors*

$$\mathbb{P}(\bigcap_{i\in\mathbb{N}} A_i) = lim_{i\longrightarrow+\infty}\mathbb{P}(A_i).$$

Démonstration.
On construit à partir de la famille $(A_i)_{i\in\mathbb{N}}$ une famille $(B_i)_{i\in\mathbb{N}}$ de la façon suivante :

$$B_0 = A_0 \text{ et } \forall_i \geq 1, B_i = A_i \setminus (\bigcup_{j=0}^{i-1} A_j),$$

on vérifie alors (exercice) que

- $\forall i \in \mathbb{N}$, $B_i \subset A_i$, et donc $P(B_i) \leq P(A_i)$.
- $i \neq j \Longrightarrow B_i \cap B_j = \emptyset$,
- $\forall n \in \mathbb{N}$, $\bigcup_{i=0}^{n} B_i = \bigcup_{i=0}^{n} A_i$ (par récurrence) et donc $\bigcup_{i\in\mathbb{N}} B_i = \bigcup_{i\in\mathbb{N}} A_i$

Démonstration.[suite]

vi) $\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = \mathbb{P}(\bigcup_{i\in\mathbb{N}} B_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(B_i) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$.

vii) $\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = \mathbb{P}(\bigcup_{i\in\mathbb{N}} B_i) = lim_{n\longrightarrow\infty} \sum_{i=1}^{n} \mathbb{P}(B_i)$.
Mais
$\sum_{i=1}^{n} \mathbb{P}(B_i) = P(\bigcup_{i=1}^{n} B_i) = \mathbb{P}(\bigcup_{i=1}^{n} A_i) = \mathbb{P}(A_n)$ par croissance de la suite $(A_n)_{n\in\mathbb{N}}$.
Donc

$$\mathbb{P}(\bigcup_{i\in N} A_i) = lim_{n\longrightarrow\infty}\mathbb{P}(A_n).$$

viii) Utiliser le point précédent et passer au complémentaire (exercice).

■

Définition

Soit $(\Omega, \mathcal{A}, P)$ *un espace de probabilité et soit* A *un événement de probabilité non nulle. Pour tout événement* $B \in \mathcal{A}$, *on appelle probabilité conditionnelle de* B *sachant* A *la quantité, notée* $\mathbb{P}(B/A)$ *et définie par :*

$$\mathbb{P}_A(B) = \mathbb{P}(B/A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Remarque

1. *La Probabilité conditionnelle est une vraie probabilité. Pour cela, montrons que* $\mathbb{P}_A$ *est une application de* $\mathcal{A}$ *dans* $[0,1]$ *telle que pour toute suite* $(B_n)_{n\in\mathbb{N}}$ *d'éléments* 2 *à* 2 *disjoints de* $\mathcal{A}$ *on a :*

$$\mathbb{P}_A(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_A(B_n).$$

Remarque

1. *La Probabilité conditionnelle est une vraie probabilité. Pour cela, montrons que* $\mathbb{P}_A$ *est une application de* $\mathcal{A}$ *dans* $[0, 1]$ *telle que pour toute suite* $(B_n)_{n\in\mathbb{N}}$ *d'éléments* 2 *à* 2 *disjoints de* $\mathcal{A}$ *on a :*

$$\mathbb{P}_A(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_A(B_n).$$

- *On a* $A \cap B \in \mathcal{A}$ *donc* $\mathbb{P}(A \cap B)$ *existe et on a aussi* $A \cap B \subset A$ *donc* $\mathbb{P}_A(B) \in [0, 1]$.
- *Soit* $(B_n)_{n\in\mathbb{N}}$ *une suite d'éléments* 2 *à* 2 *disjoints. On a :*

$$\mathbb{P}_A(\cup_{n=0}^{\infty} B_n) = \frac{\mathbb{P}(A \cap (\cup_{n\geq 0} B_n))}{\mathbb{P}(A)} = \frac{\mathbb{P}(\cup_{n\geq 0}(A \cap B_n)}{P(A)} = \sum_{n=0}^{\infty} \frac{\mathbb{P}(A \cap B_n)}{\mathbb{P}(A)}.$$

Remarque

1. *La Probabilité conditionnelle est une vraie probabilité. Pour cela, montrons que* $\mathbb{P}_A$ *est une application de* $\mathcal{A}$ *dans* $[0,1]$ *telle que pour toute suite* $(B_n)_{n\in\mathbb{N}}$ *d'éléments* 2 *à* 2 *disjoints de* $\mathcal{A}$ *on a :*

$$\mathbb{P}_A(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_A(B_n).$$

- *On a* $A\cap B \in \mathcal{A}$ *donc* $\mathbb{P}(A\cap B)$ *existe et on a aussi* $A\cap B \subset A$ *donc* $\mathbb{P}_A(B) \in [0,1]$.
- *Soit* $(B_n)_{n\in\mathbb{N}}$ *une suite d'éléments* 2 *à* 2 *disjoints. On a :*

$$\mathbb{P}_A(\cup_{n=0}^{\infty} B_n) = \frac{\mathbb{P}(A\cap(\cup_{n\geq 0} B_n))}{\mathbb{P}(A)} = \frac{\mathbb{P}(\cup_{n\geq 0}(A\cap B_n)}{P(A)} = \sum_{n=0}^{\infty} \frac{\mathbb{P}(A\cap B_n)}{\mathbb{P}(A)}.$$

2. *Si* $A, B \in \mathcal{A}$ *sont tels que* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$, *on a la propriété évidente mais très utile suivante :*

$$\mathbb{P}(A\cap B) = \mathbb{P}_B(A)P(B) = \mathbb{P}_A(B)\mathbb{P}(A).$$

Théorème (Formule des probabilités composées)

Soit $(A_i)_{1\leq i\leq n}$ *une famille d'événements telle que* $P(A_1 \cap A_2 \cap ... \cap A_{n-1}) \neq 0$*. Alors,*

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P_{A_1}(A_2)...P_{A_1\cap A_2\cap...\cap A_{n-1}}(A_n).$$

Remarque

Toutes les probabilités écrites ont un sens car :

$$\forall j \in \{1, 2, ..., n-1\} \qquad A_1 \cap A_2 \cap ... \cap A_{n-1} \subset A_1 \cap A_2 \cap ... \cap A_j.$$

Donc

$$P(A_1 \cap A_2 \cap ... \cap A_j) \geq P(A_1 \cap A_2 \cap ... \cap A_{n-1}) > 0.$$

Preuve.

...

Exemple

Un sac contient 3 boules blanches et 7 noires. On tire successivement 3 boules sans remise. Quelle est la probabilité d'obtenir les trois boules blanches ?

Exemple

Un sac contient 3 boules blanches et 7 noires. On tire successivement 3 boules sans remise. Quelle est la probabilité d'obtenir les trois boules blanches ?

Soit B_k (resp. N_k) l'événement $''$ le k^{ieme} tirage donne une boule blanche $''$ (resp. noire).
Soit A l'événement $''$ on obtient 3 boules blanches $''$.

$$A = B_1 \cap B_2 \cap B_3.$$

D'après la formule des probabilités composées, on a :

$$P(A) = P(B_1)P_{B_1}(B_2)P_{B_1 \cap B_2}(B_3).$$

Puisque le sac contient 10 boules dont 3 blanches, on a :

$$P(B_1) = \frac{3}{10}.$$

À l'issue du premier tirage, le sac contient 9 boules et si la première boule tirée est blanche, il ne reste que 2 boules blanches donc,

$$P_{B_1}(B_2) = \frac{2}{9}.$$

À l'issue du deuxième tirage, le sac contient 8 boules et si les deux boules tirées sont blanches, il ne reste que 1 boule blanche donc,

$$P_{B_1 \cap B_2}(B_3) = \frac{1}{8}.$$

Finalement,

$$P(A) = \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} = \frac{1}{720}.$$

Définition

Une suite finie ou non $(B_n)_{n\in I\subseteq\mathbb{N}}$ *d'événements de* Ω *est appelée une* ***partition*** *de* Ω *si les* B_n *sont deux à deux disjoints et si leur réunion est égale à* Ω.

Définition

Une suite finie ou non $(B_n)_{n\in I\subseteq\mathbb{N}}$ *d'événements de* Ω *est appelée une* ***partition*** *de* Ω *si les* B_n *sont deux à deux disjoints et si leur réunion est égale à* Ω.

On a alors le théorème :

Théorème (Principe des probabilités totales)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_n)_{n\in I\subseteq\mathbb{N}}$ *une partition telle que* $P(B_n) > 0$ *pour tout* $n \in I$ *et soit* $A \in \mathcal{A}$.
On a :

$$P(A) = \sum_{n\in I} P_{B_n}(A)P(B_n).$$

Définition

Une suite finie ou non $(B_n)_{n\in I\subseteq\mathbb{N}}$ d'événements de Ω est appelée une ***partition*** *de Ω si les B_n sont deux à deux disjoints et si leur réunion est égale à Ω.*

On a alors le théorème :

Théorème (Principe des probabilités totales)

Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé, soit $(B_n)_{n\in I\subseteq\mathbb{N}}$ une partition telle que $P(B_n)>0$ pour tout $n\in I$ et soit $A\in\mathcal{A}$.
On a :

$$P(A)=\sum_{n\in I}P_{B_n}(A)P(B_n).$$

Preuve.

$$\begin{aligned} P(A) &= P(A\cap\Omega)=P(A\cap\cup_{n\in I}B_n)\\ &= P(\cup_{n\in I}(A\cap B_n))\\ &= \sum_{n\in I}P(A\cap B_n)=\sum_{n\in I}P_{B_n}(A)P(B_n). \end{aligned}$$

$\blacksquare$

Remarque : Quand $n=2$, on obtient en particulier :

$$P(A)=P_B(A)P(B)+P_{B^c}(A)P(B^c).$$

Exemple

On effectue des tirages sans remise dans un sac contenant 3 boules blanches et 7 boules noires. Quelle est la probabilité d'obtenir une boule noire au deuxième tirage ?

Exemple

On effectue des tirages sans remise dans un sac contenant 3 boules blanches et 7 boules noires. Quelle est la probabilité d'obtenir une boule noire au deuxième tirage ?

Le premier tirage a donné soit une boule blanche soit une boule noire, donc :

$$P(N_2) = P(B_1)P_{B_1}(N_2) + P(N_1)P_{N_1}(N_2)$$

On a

$$P(B_1) = \frac{3}{10}, \text{ et } P(N_1) = \frac{7}{10}.$$

À l'issue du premier tirage, le sac ne contient que 9 boules dont 7 noires si B_1 a été réalisé et 6 boules noires si c'est N_1 qui a été réalisé. Donc :

$$P_{B_1}(N_2) = \frac{7}{9} \qquad et \qquad P_{N_1}(N_2) = \frac{6}{9}$$

D'où :

$$P(N_2) = \frac{3}{10} \times \frac{7}{9} + \frac{7}{10} \times \frac{6}{9} = \frac{7}{10}.$$

Théorème (Formule de Bayes)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_k)_{k=1}^{n}$ *une partition de* Ω *telle que* $P(B_k) > 0$*, pour tout* $k \in \{1, ..., n\}$ *:*

$$P_A(B_k) = \frac{P_{B_k}(A)P(B_k)}{\sum_j P_{B_j}(A)P(B_j)}$$

Théorème (Formule de Bayes)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_k)_{k=1}^{n}$ *une partition de* Ω *telle que* $P(B_k) > 0$*, pour tout* $k \in \{1, \dots, n\}$ *:*

$$P_A(B_k) = \frac{P_{B_k}(A)P(B_k)}{\sum_j P_{B_j}(A)P(B_j)}$$

Preuve. On a :

$$P_{B_k}(A)P(B_k) = P_A(B_k)P(A) = P_A(B_k)\sum_{j=1}^{n} P_{B_j}(A)P(B_j).$$

■

Théorème (Formule de Bayes)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_k)_{k=1}^n$ *une partition de* Ω *telle que* $P(B_k) > 0$*, pour tout* $k \in \{1, ..., n\}$ *:*

$$P_A(B_k) = \frac{P_{B_k}(A)P(B_k)}{\sum_j P_{B_j}(A)P(B_j)}$$

Preuve. On a :

$$P_{B_k}(A)P(B_k) = P_A(B_k)P(A) = P_A(B_k)\sum_{j=1}^{n} P_{B_j}(A)P(B_j).$$

■

Interprétation

- *A chacun des événements* (B_k) *correspond une information initiale qui permet d'évaluer* ***à priori*** *les probabilités* $P(B_1)$, $P(B_2)$,...,$P(B_n)$.

Théorème (Formule de Bayes)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_k)_{k=1}^n$ *une partition de* Ω *telle que* $P(B_k) > 0$*, pour tout* $k \in \{1, ..., n\}$ *:*

$$P_A(B_k) = \frac{P_{B_k}(A)P(B_k)}{\sum_j P_{B_j}(A)P(B_j)}$$

Preuve. On a :

$$P_{B_k}(A)P(B_k) = P_A(B_k)P(A) = P_A(B_k)\sum_{j=1}^{n} P_{B_j}(A)P(B_j).$$

■

Interprétation

- *A chacun des événements* (B_k) *correspond une information initiale qui permet d'évaluer* ***à priori*** *les probabilités* $P(B_1)$, $P(B_2)$,...,$P(B_n)$.
- *Soit* A *un événement quelconque pour lequel on connaît a priori les probabilités conditionnelles* $P_{B_1}(A)$, $P_{B_2}(A)$,..., $P_{B_n}(A)$ *(elles sont appelées des* ***vraisemblances****).)*

Théorème (Formule de Bayes)

Soit $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $(B_k)_{k=1}^{n}$ *une partition de* Ω *telle que* $P(B_k) > 0$*, pour tout* $k \in \{1, ..., n\}$ *:*

$$P_A(B_k) = \frac{P_{B_k}(A)P(B_k)}{\sum_j P_{B_j}(A)P(B_j)}$$

Preuve. On a :

$$P_{B_k}(A)P(B_k) = P_A(B_k)P(A) = P_A(B_k)\sum_{j=1}^{n} P_{B_j}(A)P(B_j).$$

■

Interprétation

- *A chacun des événements* (B_k) *correspond une information initiale qui permet d'évaluer* ***à priori*** *les probabilités* $P(B_1)$, $P(B_2)$,...,$P(B_n)$.
- *Soit A un événement quelconque pour lequel on connaît a priori les probabilités conditionnelles* $P_{B_1}(A)$, $P_{B_2}(A)$,..., $P_{B_n}(A)$ *(elles sont appelées des* ***vraisemblances****).)*
- *Le théorème de Bayes permet de calculer les probabilités conditionnelles* ***a posteriori*** $P_A(B_k)$ *à partir des probabilités* ***a priori*** *les* $P(B_k)$ *et les* $P_{B_k}(A)$.

Exemple

Reprenons l'exemple précédent et effectuons deux tirages. Le second tirage ayant donné une boule blanche, quelle est la probabilité que la première boule tirée ait été blanche ?

Cherchons $P_{B_2}(B_1)$:

Exemple

Reprenons l'exemple précédent et effectuons deux tirages. Le second tirage ayant donné une boule blanche, quelle est la probabilité que la première boule tirée ait été blanche ?

Cherchons $P_{B_2}(B_1)$:

$$\begin{aligned} P_{B_2}(B_1) &= \frac{P(B_2 \cap B_1)}{P(B_2)} \\ &= \frac{P(B_1)P_{B_1}(B_2)}{P(B_1)P_{B_1}(B_2)+P(N_1)P_{N_1}(B_2)} \\ &= \frac{\frac{3}{10}\times\frac{2}{9}}{\frac{3}{10}\times\frac{2}{9}+\frac{7}{10}\times\frac{3}{9}} = \frac{2}{9}. \end{aligned}$$

Parfois, A et B sont tels que $P_B(A) = P(A)$. Autrement dit le fait de savoir que B est réalisé ne donne aucune information supplémentaire sur le fait de savoir que A l'est. Cela conduit à la :

Définition

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

1. *Deux événements A et B sont dits indépendants si :*

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

2. *Une suite $(A_i)_{i\in\{1,2,\ldots,n\}}$ d'événements est dite mutuellement indépendante si, pour toute partie non vide $I \subset \{1, 2, ..., n\}$, on a*

$$P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$$

3. *Une suite (A_i) d'événements est dite indépendante deux à deux si, on a*

$$P(A_i \cap A_j) = P(A_i)P(A_j), \qquad \forall i \neq j.$$

Remarque

1. *L'indépendance mutuelle implique évidemment l'indépendance deux à deux. Mais attention, si $n \geq 3$, la réciproque est fausse.*

Exemple

Soit $\Omega = \{1, 2, 3, 4\}$ *avec* $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$.

- *Les événements* $A = \{1, 2\}$, $B = \{1, 3\}$ *et* $C = \{1, 4\}$ *sont deux à deux indépendants. On a :* $P(A \cap B) = P(A)P(B) = \frac{1}{4}$,
  $P(A \cap C) = P(A)P(C) = \frac{1}{4}$ *et* $P(C \cap B) = P(C)P(B) = \frac{1}{4}$.

Remarque

1. *L'indépendance mutuelle implique évidemment l'indépendance deux à deux. Mais attention, si $n \geq 3$, la réciproque est fausse.*

Exemple

Soit $\Omega = \{1, 2, 3, 4\}$ *avec* $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$.

- *Les événements* $A = \{1, 2\}$, $B = \{1, 3\}$ *et* $C = \{1, 4\}$ *sont deux à deux indépendants. On a :* $P(A \cap B) = P(A)P(B) = \frac{1}{4}$, $P(A \cap C) = P(A)P(C) = \frac{1}{4}$ *et* $P(C \cap B) = P(C)P(B) = \frac{1}{4}$.
- *Mais pas mutuellement indépendants*

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

Remarque

1. *L'indépendance mutuelle implique évidemment l'indépendance deux à deux. Mais attention, si $n \geq 3$, la réciproque est fausse.*

Exemple

Soit $\Omega = \{1,2,3,4\}$ *avec* $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$.

- *Les événements* $A = \{1,2\}$, $B = \{1,3\}$ *et* $C = \{1,4\}$ *sont deux à deux indépendants. On a :* $P(A \cap B) = P(A)P(B) = \frac{1}{4}$, $P(A \cap C) = P(A)P(C) = \frac{1}{4}$ *et* $P(C \cap B) = P(C)P(B) = \frac{1}{4}$.
- *Mais pas mutuellement indépendants*

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

2. *La notion d'indépendance dépend de la probabilité considérée. On peut imaginer un espace mesurable* $(\Omega, \mathcal{A})$, *sur lequel existent deux probabilités* P *et* Q *telles que les événements* A *et* B *soient indépendants sous* P *et pas sous* Q.

Exercice

Pour montrer que l'indépendance n'est pas une propriété intrinsèque des événements et qu'elle dépend de la probabilité considérer, on choisit $\Omega = \{1, 2, 3, 4, 5, 6\}$ *muni de la tribu* $\mathcal{P}(\Omega)$. *Soit* $\mathbb{P}_1$ *et* $\mathbb{P}_2$ *les probabilités définies sur* $(\Omega, \mathcal{P}(\Omega))$ *par :*

i	*1*	*2*	*3*	*4*	*5*	*6*
$\mathbb{P}_1(\{i\})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

et

i	*1*	*2*	*3*	*4*	*5*	*6*
$\mathbb{P}_2(\{i\})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

On pose $A = \{1, 2\}$ *et* $B = \{2, 3\}$.

1. *Pour i=1, 2, calculer* $\mathbb{P}_i(A)$, $\mathbb{P}_i(B)$ *et* $\mathbb{P}_i(A \bigcap B)$.
2. *Conclure.*

♠ **Attention :** Ne confondez pas "événements indépendants et événements disjoints" ! ! Par exemple, A et A^c sont disjoints et ne sont pas indépendants : si on sait que A est réalisé, on est sûr que A^c n'est pas réalisé.

$$P(A \cap A^c) = P(\emptyset) = 0.$$

Et

$$P(A)P(A^c) = P(A)(1 - P(A)) \neq 0 \qquad \text{en général.}$$

Définition

Deux sous tribus $\mathcal{G}$ et $\mathcal{G}'$ de $\mathcal{A}$ sont indépendantes si

$$\forall A \in \mathcal{G}, \forall A' \in \mathcal{G}', \qquad P(A \cap A') = P(A)P(A').$$

i.e. tout événement de $\mathcal{G}$ est indépendant de tout événement de la tribu $\mathcal{G}'$.

Exemple

Soit $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, et $\mathcal{A} = \mathcal{P}(\Omega)$ muni de l'équiprobabilité.
On pose :

$$\mathcal{G} = \{\emptyset, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\},$$

et

$$\mathcal{G}' = \{\emptyset, \Omega, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\}.$$

Alors $\mathcal{G}$ et $\mathcal{G}'$ sont deux sous tribus de $\mathcal{A}$ indépendantes.

Exercice

Si deux tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ sur $(\Omega, \mathcal{A}, P)$ sont indépendantes et ayant un élément commun A, on a :

$$P(A) = 0 \quad ou \quad 1.$$

Définition

Deux sous tribus $\mathcal{G}$ et $\mathcal{G}'$ de $\mathcal{A}$ sont indépendantes si

$$\forall A \in \mathcal{G}, \forall A' \in \mathcal{G}', \qquad P(A \cap A') = P(A)P(A').$$

i.e. tout événement de $\mathcal{G}$ est indépendant de tout événement de la tribu $\mathcal{G}'$.

Exemple

Soit $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, et $\mathcal{A} = \mathcal{P}(\Omega)$ muni de l'équiprobabilité.
On pose :

$$\mathcal{G} = \{\emptyset, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\},$$

et

$$\mathcal{G}' = \{\emptyset, \Omega, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\}.$$

Alors $\mathcal{G}$ et $\mathcal{G}'$ sont deux sous tribus de $\mathcal{A}$ indépendantes.

Exercice

Si deux tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ sur $(\Omega, \mathcal{A}, P)$ sont indépendantes et ayant un élément commun A, on a :

$$P(A) = 0 \quad ou \quad 1.$$

On a :

$$P(A \cap A) = P(A)^2 = P(A).$$

Application :

On considère une suite infinie de lancers de pile ou face indépendants ; on suppose qu'à chaque lancer on a une probabilité $p \in]0,1[$ d'obtenir pile. Montrer qu'avec probabilité 1, pile va apparaitre dans la suite de lancers.

Application :

On considère une suite infinie de lancers de pile ou face indépendants ; on suppose qu'à chaque lancer on a une probabilité $p \in]0,1[$ d'obtenir pile. Montrer qu'avec probabilité 1, pile va apparaitre dans la suite de lancers.
On note A_n l'événement $''$ pile apparait au n-ième lancer $''$. On sait par les hypothèses que les $(A_n)_{n\in\mathbb{N}^*}$ sont indépendants et que $P(A_n) = p > 0$ pour tout n.
On note A l'événement $''$ pile apparait au mois une fois $''$.

Application :

On considère une suite infinie de lancers de pile ou face indépendants ; on suppose qu'à chaque lancer on a une probabilité $p \in]0,1[$ d'obtenir pile. Montrer qu'avec probabilité 1, pile va apparaitre dans la suite de lancers.
On note A_n l'événement $''$ pile apparait au n-ième lancer $''$. On sait par les hypothèses que les $(A_n)_{n\in\mathbb{N}^*}$ sont indépendants et que $P(A_n) = p > 0$ pour tout n.
On note A l'événement $''$ pile apparait au mois une fois $''$. Alors

$$A^c = \cap_{n\geq 1} A_n^c$$

et donc, pour tout $N \in \mathbb{N}^*$,

$$P(A^c) \leq P(\cap_{n=1}^{N} A_n^c) = \prod_{n=1}^{N} P(A_n^c) = (1-p)^N$$

qui tend vers 0 quand N tend vers l'infini.
Donc,

$$P(A^c) = 0.$$

Par suite,

$$P(A) = 1.$$

Exercice

1. *Deux évenements incompatibles peuvent-ils être indépendants ?*
2. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé et* $A \in \mathcal{F}$*. Montrer que les propositions suivantes sont équivalentes*
   a) $\mathbb{P}(A) = 0$ *ou* $\mathbb{P}(A) = 1$
   b) *Pour tout* $B \in \mathcal{F}$*,* A *et* B *sont indépendants.*
3. *Soient* A *et* B *deux événements. Montrer que si* A *et* B *sont indépendants, alors* A^c *et* B *(resp.* A *et* B^c*,* A^c *et* B^c *) sont indépendants.*
4. *On considère un espace probabilisé* $(\Omega, \mathcal{F}, \mathbb{P})$ *et deux événements* A *et* B *dans* $\mathcal{F}$*. Montrer que si* A *et* B *sont indépendants alors* $\mathcal{F}(A)$ *et* $\mathcal{F}(B)$ *sont indépendantes.*

Variable aléatoire

Modèles aléatoires sophistiqués : Point de vue fonctionnel plutôt qu'ensembliste.

Définition

Soit $(\Omega, \mathcal{A})$ *et* $(E, \mathcal{B})$ *deux espaces probabilisables. On appelle variable aléatoire toute application* $X : \Omega \longrightarrow E$ *telle que :*

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

avec

$$X^{-1}(B) = \{\omega \in \Omega / X(\omega) \in B\}.$$

On dit aussi que X *est une application* $(\mathcal{A}, \mathcal{B})-$*mesurable.*

Remarque

- *Si* $E = \mathbb{R}$ *est une v. a. ssi*
$$X^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}_{\mathbb{R}}.$$
- *Une variable aléatoire est une grandeur qui dépend du résultat de l'expérience. Mathématiquement, c'est une fonction* $X : \Omega \longrightarrow E$, $\omega \longmapsto X(\omega)$.

Exemple

1. *Si* $\mathcal{A} = \mathcal{P}(\Omega)$ *ou* $\mathcal{B} = \{\emptyset, \quad E\}$, *alors toute application :* $X : \Omega \longrightarrow E$ *est une v.a.*
2. *Fonction indicatrice d'un événement. Soit* $(\Omega, \mathcal{A})$ *un espace mesurable et* A *un élément de* $\mathcal{A}$. *On définit la fonction indicatrice de* A, *notée* 1_A *:*

$$\forall \omega \in \Omega, 1_A(\omega) = \begin{cases} 1 & si \quad \omega \in A, \\ 0 & sinon \end{cases}$$

alors 1_A *est une v. a. de* $(\Omega, \mathcal{A})$ *dans* $(\{0,1\}, \mathcal{P}(\{0,1\}))$.

Remarque

1. *L'espace* $E = X(\Omega)$ *sera connu et simple (sous-ensemble de* $\mathbb{R}$ *ou de* $\mathbb{R}^d$*), et permettra de faire des calculs.*
   (Ω *est un espace souvent difficile à décrire, voire totalement abstrait.)*
2. ***Plutôt que de travailler sur l'espace de probabilité*** $(\Omega, \mathcal{A}, \mathbb{P})$, ***on va étudier les chances de réalisation des valeurs de*** X.

loi d'une variable aléatoire

La phrase typique dans la théorie et les exercices est la suivante : Soit X une variable aléatoire de loi $\mu := \mathbb{P}_X$ sur un espace de probabilités $(\Omega, \mathcal{A}, P)$. Essayons d'expliciter chacun des termes importants. C'est un peu abstrait, on verra des exemples plus usuels dans le paragraphe suivant.

Proposition

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $(E, \mathcal{B})$ un espace probabilisable et $X : \Omega \longrightarrow E$ une v.a. A toute élément B de $\mathcal{B}$, l'application

$$\mathbb{P}_X : \quad \begin{aligned} \mathcal{B} &\longrightarrow [0,1] \\ B &\longmapsto \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \end{aligned}$$

définit une probabilité sur $(E, \mathcal{B})$, appelée loi de X ou loi image de $\mathbb{P}$ par la variable aléatoire X .

Preuve. ....

Remarque

Une v. a. permet de transporter la probabilité $\mathbb{P}$ sur Ω en une probabilité $\mathbb{P}_X$ sur l'ensemble des valeurs prises par X. $\mathbb{P}_X$ ***est beaucoup plus facile à caractériser que $\mathbb{P}$ car $X(\Omega)$ est connu dans la pratique.***

♠ **Attention :** Ne pas confondre la variable aléatoire avec sa loi. Considérons un dé bleu et un dé rouge, et notons X_b, respectivement X_r, le résultat du lancer du bleu, respectivement rouge. Alors $\mathbb{P}(X_b \neq X_r) = \frac{5}{6} > 0$

♠ **Attention :** Ne pas confondre la variable aléatoire avec sa loi. Considérons un dé bleu et un dé rouge, et notons X_b, respectivement X_r, le résultat du lancer du bleu, respectivement rouge. Alors $\mathbb{P}(X_b \neq X_r) = \frac{5}{6} > 0$

En effet ;

$$(X_b, X_r) = \{(k,k),\ k = 1, 2, ..., 6\}.$$

Donc

$$\mathbb{P}(X_b = X_r) = \mathbb{P}(\bigcup_{k=1}^{6}(k,k)) = \frac{6}{36} = \frac{1}{6}.$$

Donc $\mathbb{P}(X_b \neq X_r) = \frac{5}{6}$.
Donc les variables aléatoires X_b et X_r sont différentes.
Par contre, elles ont la même loi :

$$\mathbb{P}_{X_b}(\{k\}) = \mathbb{P}_{X_r}(\{k\}) = \frac{1}{6}, \quad k = 1, 2, ..., 6.$$

Exemple

On tire, avec remise à chaque fois, deux boules d'un sac contenant 3 boules numérotées de 1 à 3. Soit X la v.a. : "Somme des points obtenus". Déterminer la loi de X.

Exemple

On tire, avec remise à chaque fois, deux boules d'un sac contenant 3 boules numérotées de 1 à 3. Soit X la v.a. : "Somme des points obtenus". Déterminer la loi de X.

On a : $\Omega = \{1,2,3\}^2$ *et* $card(\Omega) = 3^2 = 9$. *On a aussi :*

$$X(\Omega) = \{2,3,4,5,6\}.$$

D'où :

$$\begin{array}{ll}
\{X=2\} = \{(1,1)\}, & \mathbb{P}(X=2) = \frac{1}{9}. \\
\{X=3\} = \{(1,2),(2,1)\}, & \mathbb{P}(X=3) = \frac{2}{9}. \\
\{X=4\} = \{(1,3),(3,1),(2,2)\}, & \mathbb{P}(X=4) = \frac{3}{9} = \frac{1}{3}. \\
\{X=5\} = \{(2,3),(3,2)\}, & \mathbb{P}(X=5) = \frac{2}{9} \\
\{X=6\} = \{(3,3)\}, & \mathbb{P}(X=6) = \frac{1}{9}.
\end{array}$$

On peut représenter les résultats sous forme d'un tableau :

x_i	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

On vérifie que

$$\frac{1}{9} + \frac{2}{9} + \frac{3}{9} + \frac{2}{9} + \frac{1}{9} = 1.$$

Exercice

Un joueur dispose d'un dé cubique truqué de telle sorte que la probabilité d'apparition d'un numéro est proportionnelle à ce dernier. Nous supposons que les faces sont numérotées de 1 à 6. Soit X la variable aléatoire réelle associée au lancer de ce dé.

1. *Déterminez la loi de X.*
2. *Posons $Y = \frac{1}{X}$. Déterminez la loi de Y.*

Exercice

Un joueur dispose d'un dé cubique truqué de telle sorte que la probabilité d'apparition d'un numéro est proportionnelle à ce dernier. Nous supposons que les faces sont numérotées de 1 à 6. Soit X la variable aléatoire réelle associée au lancer de ce dé.

1. *Déterminez la loi de X.*
2. *Posons $Y = \frac{1}{X}$. Déterminez la loi de Y.*

Solution

1. X prend les valeurs $k \in |[1, 6]|$. Si $p = \mathbb{P}(X = 1)$, alors nous avons $\mathbb{P}(X = k) = kp$ pour tout $k \in |[1, 6]|$.
   Nous devons avoir
   $$\sum_{k=1}^{6} \mathbb{P}(X = k) = \sum_{k=1}^{6} kp = 1.$$
   c'est-à-dire $p\frac{6\times 7}{2}$. Donc $p = \frac{1}{21}$. La loi de X est donnée par

k	1	2	3	4	5	6
$\mathbb{P}(x = k)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

2. Y prend les valeurs $y_k = \frac{1}{k}$, $k \in |[1, 6]|$ avec $\mathbb{P}(Y = y_k) = \mathbb{P}(X = k)$.

Fonction de répartition

Définition

Soit X une v. a. r. définie sur l'espace de probabilité $(\Omega, \mathbb{P})$. La fonction F_X définie sur $\mathbb{R}$ par :

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X \leq x),$$

est appelée la fonction de répartition de X.

Fonction de répartition

Définition

Soit X une v. a. r. définie sur l'espace de probabilité $(\Omega, \mathbb{P})$. La fonction F_X définie sur $\mathbb{R}$ par :

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X \leq x),$$

est appelée la fonction de répartition de X.

La fonction de répartition F_X d'une v. a. r. X vérifie les propiétés suivantes :

Fonction de répartition

Définition

Soit X une v. a. r. définie sur l'espace de probabilité $(\Omega, \mathbb{P})$. La fonction F_X définie sur $\mathbb{R}$ par :

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X \leq x),$$

est appelée la fonction de répartition de X.

La fonction de répartition F_X d'une v. a. r. X vérifie les propiétés suivantes :

Proposition

1. $0 \leq F_X \leq 1$, $\lim_{t \longrightarrow -\infty} F_X(t) = 0$ *et* $\lim_{t \longrightarrow +\infty} F_X(t) = 1$.
2. *F_X est croissante*
3. *F_X est continue à droite avec des limites à gauche (càdlàg).*
4. *pour tout $x \in \mathbb{R}$, on a : en notant $F(x_-) = \lim_{\varepsilon \longrightarrow 0} F(x - \varepsilon)$ la limite à gauche :*

$$\mathbb{P}(X = x) = F_X(x) - F(x_-).$$

Remarque

On vous rappelle la Propriété d'Archimède. $\mathbb{R}$ *est archimédien, c'est-à-dire :* $\forall x \in \mathbb{R}$, $\exists n$ $\mathbb{N}$ *tel que* $n > x$. *Pour tout réel* x, *il existe un entier naturel* n *strictement plus grand que* x.

Preuve.

1. L'encadrement $0 \leq F_X \leq 1$ vient du fait que $\mathbb{P}$ est à valeurs dans $[0, 1]$.
   On a :
   $$\emptyset = \cap_{n\in\mathbb{N}^*} \{X \leq -n\}$$
   et d'après la propriété de la continuité monotone, on a
   $$0 = \mathbb{P}(\emptyset) = \lim_{n\longrightarrow+\infty} \mathbb{P}(X \leq -n) = \lim_{n\longrightarrow+\infty} F_X(-n) = \lim_{n\longrightarrow-\infty} F_X(n).$$
   Tandis que :
   $$\Omega = \bigcup_{n\in\mathbb{N}} (X \leq n) \qquad \text{(d'après la propriété d'Archimède)},$$
   Par suite :
   $$1 = \mathbb{P}(\Omega) = \lim_{n\longrightarrow+\infty} \mathbb{P}(X \leq n) = \lim_{n\longrightarrow+\infty} F_X(n)$$

2. La croissance découle de la croissance de $\mathbb{P}$.
3. Montrons que $\lim_{n\longrightarrow+\infty} F_X(x+\frac{1}{n}) = F_X(x).$

   On a :

$$F_X(x+\frac{1}{n}) = \mathbb{P}_X(]-\infty, x+\frac{1}{n}]).$$

   La suite des événements $A_n =]-\infty, x+\frac{1}{n}]$ est $\downarrow$ et

$$]-\infty, x] = \cap_{n=1}^{+\infty}]-\infty, x+\frac{1}{n}].$$

   D'après la propriété de la croissance monotone,

$$\begin{aligned} F_X(x) &= P_X(]-\infty, x]) = P(\cap]-\infty, x+\tfrac{1}{n}]) \\ &= \lim_{n\longrightarrow+\infty} P(]-\infty, x+\tfrac{1}{n}]) \\ &= \lim_{n\longrightarrow+\infty} F_X(x+\tfrac{1}{n}). \end{aligned}$$

   La limite à gauche est également une conséquence de la croissance de F_X (Exercice.) (Ind. écrire $\cup_n]-\infty, x-\frac{1}{n}] =]-\infty, x[$ et $F_X(x_-) = P_X(]-\infty, x[))$.

4. On remarque que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}^*$:

$$\mathbb{P}_X(]x - \frac{1}{n}, x]) = \mathbb{P}_X(] - \infty, x]) - \mathbb{P}_X(] - \infty, x - \frac{1}{n}]) = F_X(x) - F_X(x - \frac{1}{n}).$$

Or,

$$\{x\} = \cap_{n\in\mathbb{N}^*}]x - \frac{1}{n}, x].$$

Ce qui donne :

$$\mathbb{P}_X(\{x\}) = \lim_{n \longrightarrow +\infty} \mathbb{P}_X(]x - \frac{1}{n}, x]).$$

Donc

$$\mathbb{P}(X = x) = F_X(x) - F(x_-).$$

Remarque

1. *La fonction de répartition caractérise la loi de la v. a. X. Il s'agit de montrer que la probabilité de tout intervalle est exprimable à l'aide de la connaissance de F.*
2. ***En pratique :*** *On retrouve la probabilité des intervalles à partir de la fonction de répartition de la façon suivante : Soit $\mathbb{P}_X$ une probabilité sur $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, et F_X sa fonction de répartition, alors pour tous $a < b$:*

$$\mathbb{P}_X(]a, b]) = F_X(b) - F_X(a)$$

$$\mathbb{P}_X([a, b]) = F_X(b) - F_X(a_-)$$

$$\mathbb{P}_X(]a, b[) = F_X(b_-) - F_X(a)$$

$$\mathbb{P}_X([a, b[) = F_X(b_-) - F_X(a_-)$$

$$\mathbb{P}_X(\{a\}) = F_X(a) - F_X(a_-)$$

Exemple

Considérons l'exemple précédent. On a :

$$X(\Omega) = \{2, 3, 4, 5, 6\}$$

On a :

x_i	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Déterminons la fonction de répartition de X :

Exemple

Considérons l'exemple précédent. On a :

$$X(\Omega) = \{2, 3, 4, 5, 6\}$$

On a :

x_i	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Déterminons la fonction de répartition de X :

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

- *Si* $x < 2$, *on a* $(X \leq x) = \emptyset \Longrightarrow F_X(x) = 0.$
- *Si* $2 \leq x < 3$, *on a* $(X \leq x) = (X = 2) \Longrightarrow F_X(x) = \mathbb{P}(X = 2) = \frac{1}{9}.$
- *Si* $3 \leq x < 4$, *on a* $(X \leq x) = (X = 2) \cup (X = 3)$, *de plus les événements* $(X = 2)$ *et* $(X = 3)$ *sont disjoints* $\Longrightarrow F_X(x) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = \frac{1}{9} + \frac{2}{9} = \frac{1}{3}.$
- *Si* $4 \leq x < 5$ *, on a :* $F_X(X) = \frac{1}{9} + \frac{2}{9} + \frac{3}{9} = \frac{2}{3}.$
- *Si* $5 \leq x < 6$, *on a* $F_x(x) = \frac{2}{3} + \frac{2}{9} = \frac{8}{9}.$
- *Si* $x \geq 6$, *on a* $F_X(x) = 1.$

Lois discrètes

Définition

On dit qu'une v.a. X est discrète si elle prend ses valeurs dans un ensembles fini ou dénombrable.

$X(\Omega) = \{x_i \in \mathbb{R} / i \in I\}$ *avec* $I = \mathbb{N}$ *ou* $\mathbb{Z}$ *ou une de leurs parties finies.*

Dans ce cas, connaitre la loi de probabilité de X, c'est connaitre les probabilités élémentaires :

$$\forall x_i \in X(\Omega) : \qquad \mathbb{P}(X = x_i) = p_i.$$

Lois discrètes

Définition

On dit qu'une v.a. X est *discrète* *si elle prend ses valeurs dans un ensembles fini ou dénombrable.*

$X(\Omega) = \{x_i \in \mathbb{R} / i \in I\}$ *avec* $I = \mathbb{N}$ *ou* $\mathbb{Z}$ *ou une de leurs parties finies.*

Dans ce cas, connaitre la loi de probabilité de X, c'est connaitre les probabilités élémentaires :

$$\forall x_i \in X(\Omega) : \qquad \mathbb{P}(X = x_i) = p_i.$$

Remarque

Il est toujours utile de vérifier que $p_i \geq 0$ *et* $\sum_{i\in I} p_i = 1$.

Loi de probabilité d'une variable aléatoire discrète

La loi de probabilité d'une variable aléatoire X discrète est définie par la donnée de :

- $x_1, x_2, ..., x_n$ les réalisations de la v.a. X.
- $p_1, p_2, ..., p_n$ les probabilités des événements $(X = x_i)$; c'est-à-dire $\mathbb{P}(X = x_i) = p_i$; pour $i = 1, 2, \ldots, n$.

Exercice

Soit X une v.a. réelle dont la loi est définie par :

x_i	-1	1	2
p_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Déterminer la loi de $Y = 2X + 1$ et de $Z = X^2$.

Solution :

On a l'univers image de Y est $\{-1, 3, 5\}$.

$$\mathbb{P}(Y = -1) = \mathbb{P}(X = -1) = \frac{1}{4}.$$

$$\mathbb{P}(Y = 3) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

$$\mathbb{P}(Y = 5) = \mathbb{P}(X = 2) = \frac{1}{4}.$$

De même , l'univers image de Z est $\{1, 4\}$. On a :

$$\mathbb{P}(Z = 1) = \mathbb{P}(X = -1) + \mathbb{P}(X = 1) = \frac{3}{4}.$$

$$\mathbb{P}(Z = 4) = \mathbb{P}(X = 2) = \frac{1}{4}.$$

Remarque

1. *Soit X une v.a. de Ω dans $\{x_1, x_2, ..., x_n\}$.*
   La fonction de répartition vérifie :

Remarque

1. *Soit X une v.a. de Ω dans $\{x_1, x_2, ..., x_n\}$.*
   La fonction de répartition vérifie :
   - $x \in]-\infty, x_1[$, $F_X(x) = P(\emptyset) = 0$,
   - $x \in [x_1, x_2[$, $F_X(x) = P(X \leq x) = F_X(x_1) = P(X = x_1) = p_1$.

$\vdots$

   - $x \in [x_i, x_{i+1}[$, $F_X(x) = p_1 + p_2 + ... + p_i$ *(car* $\{X \leq x\} = \{X = x_i \quad ou \quad X = x_{i-1} \quad ou \quad ....ou \quad X = x_i\}$*).*
   - $X \geq x_n$, $F_X(x) = 1$.

Remarque

1. *Soit X une v.a. de Ω dans $\{x_1, x_2, ..., x_n\}$.*
   La fonction de répartition vérifie :
   - $x \in]-\infty, x_1[$, $F_X(x) = P(\emptyset) = 0$,
   - $x \in [x_1, x_2[$, $F_X(x) = P(X \leq x) = F_X(x_1) = P(X = x_1) = p_1$.

$\vdots$

   - $x \in [x_i, x_{i+1}[$, $F_X(x) = p_1 + p_2 + ... + p_i$ *(car* $\{X \leq x\} = \{X = x_i \quad ou \quad X = x_{i-1} \quad ou \quad ....ou \quad X = x_i\}$*).*
   - $X \geq x_n$, $F_X(x) = 1$.
2. *Dans le cas discret, comme*

$$F_X(x_i) - F_X(x_{i-1}) = p_i = P(X = x_i).$$

Si on connait la fonction de répartition, on peut reconstruire la loi de X par différence successives.

La loi de Bernoulli

Définition

La v. a. r. X suit la loi de Bernoulli de paramètre $p \in]0,1[$ si elle ne prend que deux valeurs 0 et 1 avec :

$$\mathbb{P}(X=1)=p, \qquad \mathbb{P}(X=0)=1-p=q.$$

Dans ca cas, on note $X \sim \mathcal{B}(p)$.

La loi de Bernoulli

Définition

La v. a. r. X suit la loi de Bernoulli de paramètre $p \in]0,1[$ si elle ne prend que deux valeurs 0 et 1 avec :

$$\mathbb{P}(X=1)=p, \qquad \mathbb{P}(X=0)=1-p=q.$$

Dans ca cas, on note $X \sim \mathcal{B}(p)$.

Exemple

On jette une pièce de monnaie qui tombe sur pile avec la probabilité p et on définit X par $X=1$ si on tombe sur pile et $X=0$ sinon.

La loi de Bernoulli

Définition

La v. a. r. X suit la loi de Bernoulli de paramètre $p \in]0,1[$ si elle ne prend que deux valeurs 0 et 1 avec :

$$\mathbb{P}(X=1)=p, \qquad \mathbb{P}(X=0)=1-p=q.$$

Dans ca cas, on note $X \sim \mathcal{B}(p)$.

Exemple

On jette une pièce de monnaie qui tombe sur pile avec la probabilité p et on définit X par $X=1$ si on tombe sur pile et $X=0$ sinon.
$\longrightarrow X \sim \mathcal{B}(p)$.

La loi de Bernoulli

Définition

La v. a. r. X suit la loi de Bernoulli de paramètre $p \in]0,1[$ si elle ne prend que deux valeurs 0 et 1 avec :

$$\mathbb{P}(X=1)=p, \qquad \mathbb{P}(X=0)=1-p=q.$$

Dans ca cas, on note $X \sim \mathcal{B}(p)$.

Exemple

On jette une pièce de monnaie qui tombe sur pile avec la probabilité p et on définit X par $X=1$ si on tombe sur pile et $X=0$ sinon.
$\longrightarrow X \sim \mathcal{B}(p)$.

Remarque

Les v. a. de Bernoulli sont naturellement associées à une expérience aléatoire qui a deux issues possibles : blan ou noir, pile / face, succès / échec, pièce correcte / pièce défectueuse, oui / non,...

La loi binomiale ou loi du tirage avec remise

Définition

La v. a. r. X suit la loi binomiale de paramètre n et p ($n \in \mathbb{N}^*$ et p) si l'ensemble des valeurs possibles est $X(\Omega) = \{0, 1, 2, ..., n\}$ *et*

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \forall k = 0, 1, ..., n.$$

Dans ce cas, on note $X \sim \mathcal{B}(n, p)$.

La loi binomiale ou loi du tirage avec remise

Définition

La v. a. r. X suit la loi binomiale de paramètre n et p ($n \in \mathbb{N}^*$ et p) si l'ensemble des valeurs possibles est $X(\Omega) = \{0, 1, 2, ..., n\}$ et

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \forall k = 0, 1, ..., n.$$

Dans ce cas, on note $X \sim \mathcal{B}(n, p)$.

Remarque

1. *La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k (1-p)^{n-k}$ sont positifs et*

$$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

2. *Les v. a. binomiales sont naturellement associées à la répétition de n expériences identiques et indépendantes de Bernoulli.*

La loi géométrique de paramètre $p \in]0,1[$

Définition

Soit X une v. a. r.. On dit que X suit la loi géométrique de paramètre p si et seulement si :

- $X(\Omega) = \mathbb{N}^*$.
- $\forall k \in \mathbb{N}^*, \qquad \mathbb{P}(X = k) = p(1-p)^{k-1}$.

Remarque

Le fait que la somme des probabilités fasse 1 vient de la formule de sommation des progressions géométriques :

$$\sum_{n \in \mathbb{N}} r^n \text{ converge} \iff |r| < 1 \text{ et alors } \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

La loi de Poisson de paramètre $\lambda \in]0, \infty[$

Définition

Soit X une v. a. r.. On dit que X suit la loi Poisson de paramètre $\lambda > 0$ si et seulement si :

- $X(\Omega) = \mathbb{N}$.
- $\forall k \in \mathbb{N}, \qquad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Dans ce cas, on note $X \sim \mathcal{P}(\lambda)$.

Remarque

Le fait que la somme des probabilités fasse 1 vient de la formule de sommation de la série exponentielle :

$$\forall \lambda \in \mathbb{R}, \quad \sum_{n \in \mathbb{N}} \frac{\lambda^n}{n!} \text{ converge et } \sum_{n \in \mathbb{N}} \frac{\lambda^n}{n!} = exp(\lambda).$$

Champs d'applications : C'est la loi qu'on obtient en comptant des événements aléatoires indépendants :

- *Arrivée des particules sur un capteur,*
- *Nombre de clients entrant dans un magasin,*
- *comptage de voiture sur une route,*
- *Etude des files d'attente dans les réseaux de communication...*

La loi de Poisson s'obtient par passage à la limite :

Proposition

Soit (X_n) *une suite de v. a. r. de loi* $\mathcal{B}(n, p_n)$ *avec* (p_n) *une suite de* $[0, 1]$ *vérifiant* $\lim_{n\longrightarrow\infty} np_n = \lambda$. *On a pour tout* $k \in \mathbb{N}$ *:*

$$\lim_{n\longrightarrow\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

La loi de Poisson s'obtient par passage à la limite :

Proposition

Soit (X_n) *une suite de v. a. r. de loi* $\mathcal{B}(n, p_n)$ *avec* (p_n) *une suite de* $[0,1]$ *vérifiant* $\lim_{n\longrightarrow\infty} np_n = \lambda$. *On a pour tout* $k \in \mathbb{N}$ *:*

$$\lim_{n\longrightarrow\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Preuve. On a

$$\begin{aligned} P(X_n = k) &= C_n^k p_n^k (1-p_n)^{n-k} = \tfrac{n!}{(n-k)!k!} p_n^k (1-p_n)^{n-k} \\ &= \tfrac{1}{k!} \underbrace{(np_n)}_{\downarrow\lambda} \underbrace{((n-1)p_n)}_{\downarrow\lambda} \ldots \underbrace{((n-k+1)p_n)}_{\downarrow\lambda} (1-p_n)^{n-k}. \end{aligned}$$

Or,

$$(1-\frac{np_n}{n})^{n-k} = e^{(n-k)Ln(1-p_n)} \longrightarrow_{n\longrightarrow\infty} e^{-\lambda} \ (k \text{ fixe}),$$

car $p_n \sim_{n\infty} \frac{\lambda}{n}$ petit $\Longrightarrow Ln(1-p_n) \sim_{n\infty} -p_n$.

Lois continues

Quand la fonction de répartition d'une v. a. r. X est continue en tout point de $\mathbb{R}$, on dit que X est une variable aléatoire continue.

Définition

On dit qu'une v. a. r. X admet une densité s'il existe une fonction mesurable positive

$$f_X : \mathbb{R} \longrightarrow \mathbb{R}$$

telle que, pour tout intervalle $[a, b]$ de $\mathbb{R}$, on a

$$\mathbb{P}_X([a, b]) = \mathbb{P}(X \in [a, b]) = \int_a^b f_X(t)dt.$$

Faisons quelques remarques concernant cette notion :

Remarque

1. *Une densité f_X est positive et vérifie*

$$\int_{-\infty}^{+\infty} f_X(t)dt = 1.$$

Réciproquement toute fonction vérifiant ces deux conditions est la densité d'une v.a.r.
$\longrightarrow$ *On définit la v.a. dont la loi est donnée par*

$$\mathbb{P}_X([a,b]) = \int_a^b f_X(t)dt.$$

Remarque

1. *Une densité* f_X *est* *positive* *et vérifie*

$$\int_{-\infty}^{+\infty} f_X(t)dt = 1.$$

Réciproquement toute fonction vérifiant ces deux conditions est la densité d'une v.a.r.
$\longrightarrow$ *On définit la v.a. dont la loi est donnée par*

$$\mathbb{P}_X([a,b]) = \int_a^b f_X(t)dt.$$

2. F_X *est dérivable sur* $\mathbb{R}$*, sauf en au plus un nombre dénombrable de points de discontinuité de* f_X*, et on a*

$$(F_X)' = f_X.$$

On passe donc facilement de la fonction de répartition à la fonction densité et vice-versa.

Exercice

Soit X une v. a. continue de fonction de densité f définie par :

$$f(x) = \begin{cases} 2x & si\ x \in [0,1[, \\ 0 & sinon \end{cases}$$

1. *Vérifier que f est bien une fonction de densité et représenter-la graphiquement.*
2. *Calculer la probabilité pour que $0,5 < X < 0,7$.*
3. *Donner sa fonction de répartition et recalculer la probabilité pour que $0,5 < X < 0,7$.*

1. IL est clair que f est bien une fonction de densité de probabilité.
2. Calcul de la probabilité $\mathbb{P}(0,5 < X < 0,7)$.

$$\mathbb{P}(0,5 < X < 0,7) = \int_{0,5}^{0,7} f(t)dt = (0.7)^2 - (0.5)^2 = 0.24.$$

3. Par définition $F(x) = \mathbb{P}(X \leq x)$.

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ x^2 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

On peut maintenant recalculer la probabilité $\mathbb{P}(0,5 < X < 0,7)$ en utilisant la formule :

$$\mathbb{P}(a < X < b) = F(b) - F(a).$$

Donc

$$\mathbb{P}(0,5 < X < 0,7) = F(0.7) - F(0.5) = 0.24.$$

La loi uniforme continue sur le segment $[a, b]$:

Définition

Une v.a.r. X est dite uniforme *sur l'intervalle* $[a, b]$ *(avec $b > a$) si elle admet pour densité :*

$$f(x) = \frac{1}{b-a}\ 1_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \textit{si} \quad a \leq x \leq b, \\ 0 & \textit{sinon}. \end{cases}$$

Dans ce cas, on note $X \sim \mathcal{U}_{[a,b]}$.

La loi uniforme continue sur le segment $[a, b]$:

Définition

Une v.a.r. X est dite uniforme sur l'intervalle $[a, b]$ (avec $b > a$) si elle admet pour densité :

$$f(x) = \frac{1}{b-a}\, 1_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & si \quad a \leq x \leq b, \\ 0 & sinon. \end{cases}$$

Dans ce cas, on note $X \sim \mathcal{U}_{[a,b]}$.

Remarque

- *f est clairement mesurable (elle est C^∞ par morceaux), positive, et*

$$\int_{\mathbb{R}} f(x)dx = \int_a^b \frac{1}{b-a} dx = 1.$$

- *Soit $[\alpha, \beta] \subset [a, b]$, on a :*

$$P_X([\alpha, \beta]) = P(\alpha \leq X \leq \beta) = \int_\alpha^\beta \frac{1}{b-a} dx = \frac{\beta - \alpha}{b-a}.$$

De plus tous les intervalles de même longueur ont même probabilité.

Exercice

1. *Calculer la fonction de répartition de la loi* $\mathcal{U}_{[a,b]}$.
2. *Soit* X *une v. a. réelle dont la fonction de répartition* F *est continue strictement croissante. Montrer que la v. a.* $Y = F(X)$ *suit la loi* $\mathcal{U}_{[0,1]}$.

Exercice

1. *Calculer la fonction de répartition de la loi* $\mathcal{U}_{[a,b]}$.
2. *Soit* X *une v. a. réelle dont la fonction de répartition* F *est continue strictement croissante. Montrer que la v. a.* $Y = F(X)$ *suit la loi* $\mathcal{U}_{[0,1]}$.

Réponse :

1. Soit $x \in \mathbb{R}$. On a

$$F(x) = \begin{cases} 0 & si x < a \\ \frac{x-a}{b-a} & si\ a \leqslant x < b \\ 1 & si\ x \geqslant b. \end{cases}$$

2. On a la v.a. Y prend ses valeurs dans $[0, 1]$.
On désigne par G et g la fonction de répartition et la fonction densité de Y.

$$G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F(X) \leq y) = \mathbb{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Donc ;
si $y \in]-\infty, 0]$: $G(y) = 0$ et $g(y) = 0$
si $y \in [0, 1]$: $G(y) = y$ et $g(y) = 1$
si $y \in [1, +\infty]$: $G(y) = 1$ et $g(y) = 0$.

Remarque

$F(x) = x$ sur $[0, 1]$, avec $F(x) = 0$ pour $x \leq 0$ et $F(x) = 1$ pour $x \geq 1$ détermine une probabilité appelée probabilité uniforme sur $[0, 1]$ ou encore mesure de Lebesgue sur $[0, 1]$, qui correspond à la longueur des intervalles.

lois gaussiennes (ou lois normales)

Définition

Une v. a. r. X est dite gaussienne de moyenne m et de variance $\sigma^2 \in \mathbb{R}_+^*$, si elle admet pour densité :

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

Dans ce cas, on note $X \sim \mathcal{N}(m, \sigma^2)$.
Lorsque $m = 0$, et $\sigma = 1$ la v. a. X est dite centée réduite.
σ représente la dispersion des valeurs autour de la moyenne : plus σ est grande plus les valeurs sont dispersées.

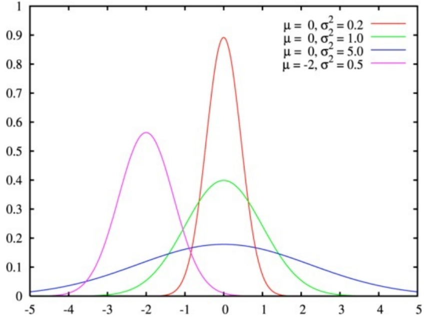
lois gaussiennes (ou lois normales)

Définition

Une v. a. r. X est dite gaussienne *de moyenne m et de variance $\sigma^2 \in \mathbb{R}_+^*$, si elle admet pour densité :*

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

Dans ce cas, on note $X \sim \mathcal{N}(m, \sigma^2)$.
Lorsque $m = 0$, et $\sigma = 1$ la v. a. X est dite centée réduite.
σ représente la dispersion des valeurs autour de la moyenne : plus σ est grande plus les valeurs sont dispersées.



Exercice

1. *Soit U une v. a. normale centré réduite $N(0,1)$ et $\Pi(.)$ la fonction de répartition de U. Montrer que*

$$\Pi(-z) = 1 - \Pi(z) \qquad \forall z \in \mathbb{R}$$

2. *Soit X la v.a. définie par*

$$X = m + \sigma U$$

où m et σ sont des réels non nuls. Soit G_X la fonction de répartition de X. Montrer que

$$G_X(x) = \Pi(\frac{x-m}{\sigma}).$$

3. *Déterminer la densité de probabilité $g(x)$ de X.*

Réponse :

1- On a

$$1 = \int_{\mathbb{R}} f_U(u)du = \int_{-\infty}^{-z} f_U(u)du + \int_{-z}^{+\infty} f_U(u)du$$

Or, par le changement de variable $u' = -u$

$$\int_{-z}^{+\infty} f_U(u)du = -\int_{z}^{-\infty} f_U(u')du' = \int_{-\infty}^{z} f_U(u)du = \Pi(z).$$

2- On a

$$G_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}) = \Pi(\frac{x-m}{\sigma})$$

3- On a

$$g(x) = G'(x) = \frac{1}{\sigma}\Pi'(\frac{x-m}{\sigma}) = \frac{1}{\sigma} f_U(\frac{x-m}{\sigma})$$

Donc

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

Utilisation de la table de la loi normale

Soit X une v. a. normale $N(3,4)$. Calculer $\mathbb{P}(X \leq 4)$, $\mathbb{P}(X \leq -1)$; $\mathbb{P}(X > 1)$. Touver a telle que $\mathbb{P}(X < a) = 0,7486$.

Réponse :

Utilisation de la table de la loi normale

Soit X une v. a. normale $N(3,4)$. Calculer $\mathbb{P}(X \leq 4)$, $\mathbb{P}(X \leq -1)$; $\mathbb{P}(X > 1)$. Touver a telle que $\mathbb{P}(X < a) = 0,7486$.

Réponse :

$$\mathbb{P}(X \leq 4) = \mathbb{P}(\frac{X-m}{\sigma} \leq \frac{4-m}{\sigma}) = \mathbb{P}(U \leq 0.5) = 0.6915.$$

$$\mathbb{P}(X \leq -1) = \mathbb{P}(\frac{X-m}{\sigma} \leq \frac{-1-m}{\sigma}) = \mathbb{P}(U \leq -2) = 0.0228$$

$$\mathbb{P}(X > 1) = \mathbb{P}(U > -1) = 0.8413.$$

$$\mathbb{P}(X \leq a) = \mathbb{P}(U \leq \frac{a-3}{2}) = 0.7486$$

Or

$$\mathbb{P}(U \leq 0.67) = 0.7486$$

Donc $\frac{a-3}{2} = 0.67$ et par suite $a = 4.35$.

La loi exponentielle

Définition

La v. a. r. X suit la loi exponentielle de paramètre $\theta > 0$ si et seulement si elle est à valeurs dans $\mathbb{R}^+$ et admet pour densité la fonction donnée par :

$$f(x) = \begin{cases} \theta \exp(-\theta x) & si \quad x > 0, \\ 0 & sinon \end{cases} = \theta \exp(-\theta x) \mathbf{1}_{]0,+\infty[}(x)$$

Exercice

Montrer que la fonction de répartition de la loi exponentielle est donnée par :

$$F(t) = (1 - e^{-\theta t}) 1_{[0,+\infty[}(t).$$

Remarque

1. *La loi exponentielle intervient en fiabilité pour modéliser la durée de fonctionnement d'un équipement technique.*
2. *Elle intervient aussi pour modéliser le temps séparant les arrivées de deux clients dans l'étude d'un phénomène d'attente.*

Remarque

1. *La loi exponentielle intervient en fiabilité pour modéliser la durée de fonctionnement d'un équipement technique.*
2. *Elle intervient aussi pour modéliser le temps séparant les arrivées de deux clients dans l'étude d'un phénomène d'attente.*

Exercice

Soit la v.a. N_t *désignant le nb de clients arrivant dans un magasin sur l'intervalle* $[0, t]$*. Si* N_t *suit la loi de Poisson de paramètre* λt*, montrer que la date d'arrivée* T *du premier client suit la loi* $exp(\lambda)$*.*

Remarque

1. *La loi exponentielle intervient en fiabilité pour modéliser la durée de fonctionnement d'un équipement technique.*
2. *Elle intervient aussi pour modéliser le temps séparant les arrivées de deux clients dans l'étude d'un phénomène d'attente.*

Exercice

Soit la v.a. N_t désignant le nb de clients arrivant dans un magasin sur l'intervalle $[0, t]$. Si N_t suit la loi de Poisson de paramètre λt, montrer que la date d'arrivée T du premier client suit la loi $exp(\lambda)$.

En effet, on a

$$P(N_t = 0) = P(T > t) = 1 - P(T \leq t)$$

d'où

$$P(T \leq t) = 1 - exp(-\lambda t).$$

Ce qui donne le résultat.

Exercice

Soit X une v. a. r. suivant la loi $\exp(\theta)$. Montrer que X vérifie la propriété d'absence de mémoire :

$$\forall s \in \mathbb{R}^+, \quad \forall t \in \mathbb{R}^+, \quad P(X > t + s / X > t) = P(X > s)$$

Exercice

Soit X une v. a. r. suivant la loi $exp(\theta)$. Montrer que X vérifie la propriété d'absence de mémoire :

$$\forall s \in \mathbb{R}^+, \quad \forall t \in \mathbb{R}^+, \quad P(X > t + s/X > t) = P(X > s)$$

Solution.

On a $X \sim exp(\theta)$, donc

$$\begin{aligned} P_{X>t}(X > t + s) &= \frac{P(X>t+s)}{P(X>t)} \\ &= \frac{\theta \int_{t+s}^{+\infty} e^{-\theta x} dx}{\theta \int_{t}^{+\infty} e^{-\theta x} dx} \\ &= \frac{e^{-\theta(t+s)}}{e^{-\theta t}} = e^{-\theta s} = P(X > s). \end{aligned}$$

■

Nous supposons ici que X est une v. a. réelle.

Définition

Soit X une v. a. discrète et

$$X(\Omega) = \{x_i \in \mathbb{R}, \quad i \in I\}, \qquad I \subseteq \mathbb{N}.$$

1. *On dit que X est intégrable si la somme $\sum_{i \in I} |x_i| \mathbb{P}(X = x_i)$ converge.*
2. *Si X est intégrable, son espérance est donnée par :*

$$\mathbb{E}(X) = \sum_{i \in I} x_i \mathbb{P}(X = x_i) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

Nous supposons ici que X est une v. a. réelle.

Définition

Soit X une v. a. discrète et

$$X(\Omega) = \{x_i \in \mathbb{R}, \quad i \in I\}, \qquad I \subseteq \mathbb{N}.$$

1. *On dit que X est intégrable si la somme $\sum_{i \in I} |x_i| \mathbb{P}(X = x_i)$ converge.*
2. *Si X est intégrable, son espérance est donnée par :*

$$\mathbb{E}(X) = \sum_{i \in I} x_i \mathbb{P}(X = x_i) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

Remarque

La somme $\sum_{i \in I} |x_i| P_X(x_i)$ est soit une somme finie (si $X(\Omega)$ est fini) et dans ce cas, l'espérance existe toujours, soit une série à terme positifs (si $X(\Omega)$ est une partie dénombrable infinie), et dans ce cas, la série peut être divergente et avoir une somme égale à $+\infty$.

Exemple

1. *Si X est une v.a. constante $X = c$, alors $\mathbb{E}(X) = c$.*
2. *Soit X une v.a. discrète dont la loi est donnée par :*

x	2	3	4	5	6
$P(x{=}x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

$$\mathbb{E}(X) =$$

Exemple

1. *Si* X *est une v.a. constante* $X = c$*, alors* $\mathbb{E}(X) = c$.
2. *Soit* X *une v.a. discrète dont la loi est donnée par :*

x	2	3	4	5	6
$P(x{=}x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

$$\mathbb{E}(X) = 2 \times \frac{1}{9} + 3 \times \frac{2}{9} + 4 \times \frac{3}{9} + 5 \times \frac{2}{9} + 6 \times \frac{1}{9} = 4.$$

3. *Soit* X *une v.a. de Bernoulli de paramètre* p.
   On a $X = \mathbf{1}_A$*, alors* $\mathbb{E}(X) = \mathbb{E}(\mathbf{1}_A) = P(A) = p$.

Remarque

Si pour tout $i \in I$, on a : $\quad a \leq x_i \leq b$, *alors :*

$$a \leq \mathbb{E}(X) \leq b.$$

Remarque

Si pour tout $i \in I$, on a : $\quad a \leq x_i \leq b$, *alors :*

$$a \leq \mathbb{E}(X) \leq b.$$

En effet,

$$\sum_{i \in I} aP(X = x_i) \leq \mathbb{E}(X) \leq \sum_{i \in I} bP(X = x_i)$$

et comme $\sum_{i \in I} P(X = x_i) = 1$, *on a le résultat.*
En particulier, si $X \geq 0$, *alors* $\mathbb{E}(X) \geq 0$.

Définition

Soit X une v. a. r. à densité.

1. *On dit que X est intégrable si l'intégrale $\int_{\mathbb{R}} |x| f_X(x)dx$ converge.*
2. *Si X est intégrable, son **espérance** est donnée par :*

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

Exemple

- *Soit $X \sim \mathcal{U}_{[a,b]}$ avec $a < b$.*
  Alors :

$$\mathbb{E}(X) =$$

Définition

Soit X une v. a. r. à densité.

1. *On dit que X est intégrable si l'intégrale $\int_{\mathbb{R}} |x| f_X(x)dx$ converge.*
2. *Si X est intégrable, son espérance est donnée par :*

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

Exemple

- *Soit $X \sim \mathcal{U}_{[a,b]}$ avec $a < b$.*
  Alors :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}[\frac{x^2}{2}]_a^b = \frac{b+a}{2}.$$

Le théorème de transfert est un théorème fondamental en théorie des probabilités qui permet d'exprimer l'espérance d'une fonction d'une variable aléatoire X en fonction d'une intégrale contre la loi de X. Sa forme générale est la suivante :

Théorème (Théorème de transfert)

1. *(Variable discrète) : Soit g une fonction définie sur $X(\Omega)$ à valeurs dans $\mathbb{R}$. Si I est fini ou si la série $\sum_{i\in I} g(x_i)P(X = x_i)$ est absolument convergente, alors la v.a. $g(X)$ admet une espérance et on a :*

$$\mathbb{E}(g(X)) = \sum_{i\in I} g(x_i)\mathbb{P}(X = x_i).$$

2. *(Variable à densité) : Soient $X : \Omega \longrightarrow \mathbb{R}$ une v. a. r. à densité et $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction mesurable avec $g \geq 0$ ou $g(X)$ est intégrable, alors, on a*

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

Exemple

Soit $X \sim \mathcal{U}_{[a,b]}$ *avec* $a < b$. *Et soit* $g(x) = exp(x)$
Alors :

$$\mathbb{E}(g(X)) \;=$$

Exemple

Soit $X \sim \mathcal{U}_{[a,b]}$ *avec* $a < b$. *Et soit* $g(x) = exp(x)$
Alors :

$$\begin{aligned} \mathbb{E}(g(X)) &= \mathbb{E}(exp(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx \\ &= \int_{-\infty}^{+\infty} \frac{exp(x)}{b-a}dx = \frac{1}{b-a}[exp(x)]_a^b = \frac{exp(b)-exp(a)}{b-a}. \end{aligned}$$

Proposition

L'espérance possède les propriétés suivantes :

1. *Linéarité : Soit X et Y deux v. a. réelles et λ unréel. On a :*

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y),$$

$$\mathbb{E}(\lambda(X)) = \lambda\mathbb{E}(X).$$

2. *Positivité : Si X est une v. a. positive, alors $\mathbb{E}(X) \geq 0$.*
3. *Croissance : Si X et Y sont deux v. a. intégrables, et si $X \leq Y$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

Autrement dit, $X \longmapsto \mathbb{E}(X)$ est une forme linéaire positive.

Preuve. La démonstration sera donnée comme exercice en T.D $N2$

Définition

Soient X une v.a.r. et $k \in \mathbb{N}^*$. On dit que X admet un moment d'ordre k si la v. a. r. X^k est intégrable. Dans ce cas, la valeur $\mathbb{E}(X^k)$ est appelée moment d'ordre k de X. On note

$$m_k(X) = \mathbb{E}(X^k) =^{\text{thm. tr.}} \begin{cases} \sum_{i \in I} x_i^k \mathbb{P}(X = x_i), & \text{cas discrèt} \\ \int_{-\infty}^{+\infty} x^k f(x)dx, & \text{cas continu.} \end{cases}$$

Remarque

1. *Le moment d'ordre 1 de X, si il existe, est l'espérance de X.*

Définition

Soient X une v.a.r. et $k \in \mathbb{N}^*$. On dit que X admet un moment d'ordre k si la v. a. r. X^k est intégrable. Dans ce cas, la valeur $\mathbb{E}(X^k)$ est appelée moment d'ordre k de X. On note

$$m_k(X) = \mathbb{E}(X^k) =^{\text{thm. tr.}} \begin{cases} \sum_{i \in I} x_i^k \mathbb{P}(X = x_i), & \text{cas discrèt} \\ \int_{-\infty}^{+\infty} x^k f(x)dx, & \text{cas continu.} \end{cases}$$

Remarque

1. *Le moment d'ordre 1 de X, si il existe, est l'espérance de X.*
2. *Si la v. a. r. X admet un moment absolu d'ordre r fini, elle a aussi un moment absolu d'ordre p fini pour tout $p \in |[0, r]|$:*

Définition

Soient X une v.a.r. et $k \in \mathbb{N}^*$. On dit que X admet un moment d'ordre k si la v. a. r. X^k est intégrable. Dans ce cas, la valeur $\mathbb{E}(X^k)$ est appelée moment d'ordre k de X. On note

$$m_k(X) = \mathbb{E}(X^k) =^{\text{thm. tr.}} \begin{cases} \sum_{i \in I} x_i^k \mathbb{P}(X = x_i), & \text{cas discrèt} \\ \int_{-\infty}^{+\infty} x^k f(x)dx, & \text{cas continu.} \end{cases}$$

Remarque

1. *Le moment d'ordre 1 de X, si il existe, est l'espérance de X.*
2. *Si la v. a. r. X admet un moment absolu d'ordre r fini, elle a aussi un moment absolu d'ordre p fini pour tout $p \in |[0, r]|$: En effet :*
   Si $0 \leq p \leq r$, on a :
   $$|X|^p \leq \mathbf{1}_{\{|X| \leq 1\}} + |X|^p 1_{\{|X|>1\}} \leq 1 + |X|^r.$$

Définition

Soit X une v. a. r. admettant un moment d'ordre 2.

1. *La variance de X est par définition :*

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

2. *L'écart-type de X est par définition égal :*

$$\sigma(X) = \sqrt{Var(X)}.$$

Définition

Soit X une v. a. r. admettant un moment d'ordre 2.

1. *La variance de X est par définition :*

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

2. *L'écart-type de X est par définition égal :*

$$\sigma(X) = \sqrt{Var(X)}.$$

Remarque

La variance d'une v. a. r. représente sa dispersion autour de sa moyenne. C'est un nombre toujours positif, ce qui est évident vu sa définition.
Nous allons montrer dans le paragraphe qui suit l'inégalité suivante : $\forall \epsilon > 0$

$$P(|X - \mathbb{E}(X)| > \epsilon) \leq \frac{Var(X)}{\epsilon^2}.$$

Une variance faible correspond à un niveau homogène, proche de la moyenne.

En pratique, on utilise la

Proposition

Soit X une v. a. r. admettant un moment d'ordre 2. On a :

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

En pratique, on utilise la

Proposition

Soit X une v. a. r. admettant un moment d'ordre 2. On a :

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

Preuve. Il suffit d'utiliser la linéarité de l'espérance :

$$\begin{aligned} Var(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned}$$

La variance vérifie également les propriétés suivantes :

Proposition

Soit X une v. a. r. admettant un moment d'ordre 2. On a :

1. $Var(aX+b) = a^2 Var(X)$, *pour tout* $a, b \in \mathbb{R}$.
2. $Var(X) = 0$ *si et seulement si* X *est une constante.*

Proposition

Soit X une v. a. r. admettant un moment d'ordre 2. On a :

1. $Var(aX + b) = a^2 Var(X)$, *pour tout* $a, b \in \mathbb{R}$.
2. $Var(X) = 0$ *si et seulement si* X *est une constante.*

Preuve.

1. Laissée en exercice.
2. Cela repose sur le résultat important suivant : si Z est une v. a. r. positive telle que $\mathbb{E}(Z) = 0$, alors $Z = 0$ p.s. En effet, on a

$$Z \geq \frac{1}{n} 1_{\{Z \geq \frac{1}{n}\}}.$$

Donc $P(Z \geq \frac{1}{n}) = 0$ pour tout $n \in \mathbb{N}^*$.
On a aussi,

$$\begin{aligned} \mathbb{P}(Z > 0) &= \mathbb{P}(\cup_n \{Z \geq \tfrac{1}{n}\}) \\ &= \lim_{n \longrightarrow \infty} P(Z \geq \tfrac{1}{n}) \\ &= \lim_{n \longrightarrow \infty} 0 = 0. \end{aligned}$$

Il suffit d'appliquer ce résultat à $Z = (X - \mathbb{E}(X))^2$.

Proposition

Si X est une v. a. r. positive, intégrable, et si $a \in \mathbb{R}_+^*$, On a :

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Preuve.
Elle découle de l'inégalité $X \geq a1_{X>a}$ et de la positivité de l'espérance.

Proposition

Si X est une v. a. r. admettant un moment d'ordre 2, et si $a \in \mathbb{R}_+^*$, On a :

$$P(|X - \mathbb{E}(X)| > a) \leq \frac{Var(X)}{a^2}.$$

Proposition

Si X est une v. a. r. admettant un moment d'ordre 2, et si $a \in \mathbb{R}_+^*$, On a :

$$P(|X - \mathbb{E}(X)| > a) \leq \frac{Var(X)}{a^2}.$$

Preuve.
Il suffit d'appliquer l'inégalité de Markov à $(X - \mathbb{E}(X))^2$ après avoir remarqué que :

$$\{|X - \mathbb{E}(X)| > a\} = \{(X - \mathbb{E}(X))^2 > a^2\}.$$

Proposition

Soient X *une v. a. r. intégrable, et* $g : \mathbb{R} \longrightarrow \mathbb{R}$ *une fonction convexe telle que la v. a. r.* $g(X)$ *soit intégrable. Alors :*

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)).$$

Proposition

Soient X une v. a. r. intégrable, et $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction convexe telle que la v. a. r. $g(X)$ soit intégrable. Alors :

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)).$$

Preuve. La convexité de g assure qu'en tout point son graphe est au-dessus de sa tangente : Pour tout $t \in \mathbb{R}$, il existe δ (on peut prendre $\delta = g'_d(t)$ ou $g'_g(t)$), tel que :

$$g(x) \geq g(t) + \delta(x - t).$$

On prend : $x = X(\omega)$ et $t = \mathbb{E}(X)$ et on prend l'espérance, on obtient :

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)).$$

Définition

Soit X une variable aléatoire réelle. On appelle fonction caractéristique de X la fonction ϕ_X définie par

$$\begin{aligned} \phi_X(t) &= \mathbb{E}(e^{itX}) \\ &= \int_{\mathbb{R}} e^{itx} f(x)dx \text{ dans le cas des v.a continues.} \\ &= \sum_{x\in X(\Omega)} e^{itx}\mathbb{P}(X = x) \text{ dans le cas discret.} \end{aligned}$$

Remarque

Si X et Y sont deux v. a. ayant la même fonction caractéristique. Alors, elles ont la même loi, i.e $\mathbb{P}_X = \mathbb{P}_Y$. Ce résultat décole du théorème d'injectivité de la transformée de Fourier sur les mesures bornées. En effet, La fonction caractéristique d'une variable aléatoire X est la transformée de Fourier de sa loi $\mathbb{P}_X$.

Exercice

1. *Vérifier que* $\phi_X(0) = 1$ *et que* $|\phi_X(t)| \leq 1 \ \forall t \in \mathbb{R}$.
2. *Déterminer la f. c.* ϕ_X *de la v. a.* X *dans les cas suivants :*

a. $X = a, \quad a \in \mathbb{R}$.

b. X *est une v. a. de Bernoulli de paramètre* p.

c. X *est une v. a. de binomiale de paramètres* n *et* p, *i.e.* $X \sim B(n, p)$.

e. X *est une v. a. de Poisson de paramètre* λ, *i.e.* $X \sim P(\lambda)$.

f. X *est une variable aléatoire de loi uniforme sur* $[-a, a]$.

Solution

a. $X = a, \quad a \in \mathbb{R}.$

$$\phi_X(t) = \sum_{x \in X(\Omega)} e^{itx}\mathbb{P}(X = x) = e^{ita}\mathbb{P}(X = a) = e^{ita}.$$

b. X est une v. a. de Bernoulli de paramètre p.

$$\phi_X(t) = \sum_{x \in X(\Omega)} e^{itx}\mathbb{P}(X = x) = e^{it}\mathbb{P}(X = 1) + e^{it0}\mathbb{P}(X = 0) = pe^{it} + 1 - p.$$

c. X est une v. a. de binomiale de paramètres n et p, i.e. $X \sim B(n, p)$.

$$\phi_X(t) = \sum_{x \in X(\Omega)} e^{itx}\mathbb{P}(X = x) = \sum_{k=0}^{n} e^{itk}\mathbb{P}(X = k) = \sum_{k=0}^{n} e^{itk}C_n^k p^k (1-p)^{n-k} = (pe^{it} + 1 - p)^n.$$

e. X est une v. a. de Poisson de paramètre λ, i.e. $X \sim P(\lambda)$.

$$\phi_X(t) = \sum_{x \in X(\Omega)} e^{itx}\mathbb{P}(X = x) = \sum_{k=0}^{+\infty} e^{itk}\mathbb{P}(X = k) = \sum_{k=0}^{+\infty} e^{itk}e^{-\lambda}\frac{\lambda^k}{k!} = e^{\lambda(e^{it}-1)}.$$

Solution

f. X est une variable aléatoire de loi uniforme sur $[-a, a]$.

$$\begin{aligned}
\phi_X(t) &= \int_{\mathbb{R}} e^{itx} f_X(x) dx \\
&= \int_{\mathbb{R}} e^{itx} \frac{1}{2a} 1_{[-a,a]}(x) dx \\
&= \frac{1}{2a} \left(\int_{-a}^{a} cos(tx) dx + i \int_{-a}^{a} sin(tx) dx \right) \\
&= \frac{1}{2a} \int_{-a}^{a} cos(tx) dx \\
&= \frac{1}{2a} \left[\frac{1}{t} sin(tx) \right]_{-a}^{a} \\
&= \frac{sin(at)}{at}.
\end{aligned}$$

Fonction caractéristique de la loi normale

Exercice

Soit X une v a normale centrée réduite.

1. *Montrer que la fonction caractéristique de X vérifie l'équation différentielle linéaire suivante :*

$$\phi_X(t)' = -t\phi_X(t).$$

2. *Montrer la fonction caractéristique de X est $\phi_X(t) = e^{\frac{-t^2}{2}}$. Déduire que si $Z = \sigma X + m$, alos :*

$$\phi_Z(t) = e^{imt-\frac{\sigma^2 t^2}{2}}$$

3. *Déduire que si $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ et X et Y sont indépendantes, alors $X + Y \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Solution

1. Montrons que la fonction caractéristique de X vérifie une équation différentielle linéaire :

$$\begin{aligned}
\phi_X(t) &= \int_{\mathbb{R}} e^{itx} f_X(x)dx = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= \int_{\mathbb{R}} cos(tx) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx + i \int_{\mathbb{R}} sin(tx) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= \int_{\mathbb{R}} cos(tx) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.
\end{aligned}$$

Comme la loi normale admet des moments de tout ordre, on peut dériver $\phi_X(t)$

$$\begin{aligned}
\phi_X(t)' &= \int_{\mathbb{R}} \frac{d}{dt}(cos(tx)) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= \int_{\mathbb{R}} -xsin(tx) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&=^{I.P.P.} - \int_{\mathbb{R}} tcos(tx) \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \\
&= -t\phi_X(t).
\end{aligned}$$

Solution

2. D'après la première question, on a établit une équation différentielle linéaire satisfaite par $\phi_X(t)$: donc

$$\forall t \in \mathbb{R} \qquad \phi_X(t) = C.e^{\frac{-t^2}{2}}.$$

Puisque $\phi_X(0) = 1$, alors $C = 1$. Donc la fonction caractéristique de X est $\phi_X(t) = e^{\frac{-t^2}{2}}$.
Si $Z = \sigma X + m$, alors :

$$\phi_Z(t) = \mathbb{E}e^{itZ} = \mathbb{E}e^{it(\sigma X+m)} = e^{itm}.\mathbb{E}e^{i(t\sigma)X)} = e^{itm}.\phi_X(t\sigma) = e^{imt-\frac{\sigma^2 t^2}{2}}$$

3. Si $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ et $X \amalg Y$, alors $X + Y \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.
On a

$$\phi_{X+Y}(t) = \mathbb{E}e^{it(X+Y)} = \mathbb{E}(e^{itX}.e^{itY}) \stackrel{\amalg}{=} \mathbb{E}(e^{itX}).\mathbb{E}(e^{itY})) = e^{i(m_1+m_2)t-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}.$$

Comme la fonction caractéristique caractérise la loi d'une v.a., alors

$$X + Y \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

Exercice

La durée de vie, exprimée en années, d'un circuit électronique est une variable aléatoire X dont la fonction de répartition F_X est définie par

$$F_X(t) = (1 - e^{\frac{-t^2}{2}})1_{t\geq 0}.$$

1. *Donner la densité de probabilité f_X de X. Calculer $\mathbb{E}(X)$.*
2. *Sachant que le circuit a déjà fonctionné durant 1 an, quelle est la probabilité qu'il continue à fonctionner encore durant au moins 2 ans ? La loi est-elle sans mémoire ?*
   Un équipement électronique est composé de 10 circuits identiques et indépendants. Au circuit i $(1 \leq i \leq 10)$ est associée la variable aléatoire X_i, avec $X_i = 1$ si la durée de vie du circuit i est inférieure à un an et $X_i = 0$ sinon.
3. *Quelle est la loi du nombre, N, de circuits de cet équipement dont la durée de vie est inférieure à 1 an ?*
4. *L'équipement est dit en série si la défaillance de l'un de ses circuits entraine sa défaillance. Quelle est alors la probabilité qu'il soit défaillant avant 1 an ?*
5. *L'équipement est dit en parallèle si sa défaillance ne peut se produire que si tous ses circuits sont défaillants. Quelle est alors la probabilité qu'il soit défaillant avant 1 an ?*

1. La densité de probabilité f_X de X est
$$f_X(t) = texp(-\frac{1}{2}t^2)1_{t>0}.$$
On a
$$\mathbb{E}(X) = \int_0^\infty t^2 exp(-\frac{1}{2}t^2)dt =^{I.P.P.} \ldots = \frac{\sqrt{2\pi}}{2}.$$
2. Sachant que le circuit a déjà fonctionné durant 1 an, la probabilité qu'il continue à fonctionner encore durant au moins 2 ans est
$$\mathbb{P}(T \geq 3 | T \geq 1) = \frac{\mathbb{P}(T \geq 3, T \geq 1)}{\mathbb{P}(T \geq 1)} = \frac{\mathbb{P}(T \geq 3)}{\mathbb{P}(T \geq 1)} = \frac{e^{-9/2}}{e^{-1/2}} = e^{-4}.$$
Puisque
$$\mathbb{P}(T \geq 3 | T \geq 1) \neq e^{-2} = \mathbb{P}(T \geq 2).$$
La loi n'est pas sans mémoire.
Un équipement électronique est composé de 10 circuits identiques et indépendants. Au circuit i ($1 \leq i \leq 10$) est associée la variable aléatoire X_i, avec $X_i = 1$ si la durée de vie du circuit i est inférieure à un an et $X_i = 0$ sinon.
3. La loi du nombre, N, de circuits de cet équipement dont la durée de vie est inférieure à 1 an est la loi binomiale $B(10, y)$ avec
$$y = \mathbb{P}(T \leq 1) = F(1) = 1 - e^{-\frac{1}{2}}.$$
4. L'équipement est dit en série si la défaillance de l'un de ses circuits entraine sa défaillance. La probabilité qu'il soit défaillant avant 1 an est
$$\mathbb{P}(N \geq 1) = 1 - \mathbb{P}(N = 0) = 1 - e^{-\frac{10}{2}} \simeq 0.99.$$
5. L'équipement est dit en parallèle si sa défaillance ne peut se produire que si tous ses circuits sont défaillants. La probabilité qu'il soit défaillant avant 1 an vaut
$$\mathbb{P}(N = 10) = (1 - e^{-\frac{1}{2}})^{10} \simeq 8.910^{-5}.$$