Ouvertures Économiques

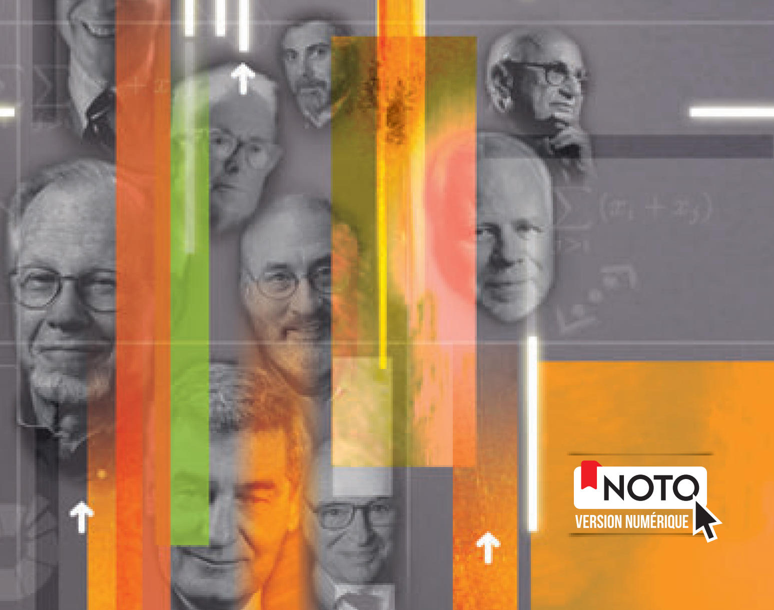
Inférence statistique et probabilités

Stéphane **Mussard**
Françoise **Seyte**

Préface de Michel **Terraza**

Exercices pédagogiques et fiches de synthèses

Version numérique enrichie de QCM : noto.deboeck.com

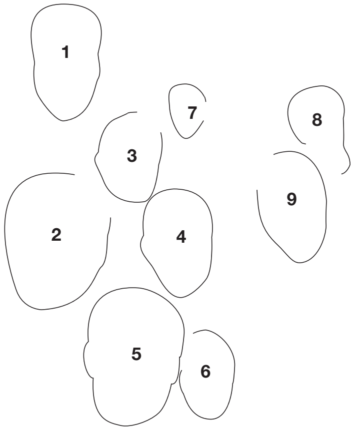


L
M
D

de boeck

NOTO
VERSION NUMÉRIQUE

Inférence statistique et probabilités



1 | **Georges AKERLOF** (1940-). *Né dans le Connecticut, Georges Akerlof est docteur en sciences économiques du Massachusetts Institute of Technology (MIT). Professeur à Berkeley, le prix Nobel d'économie lui a été décerné en 2001, en compagnie de Joseph Stiglitz et Michael Spence pour ses travaux sur l'asymétrie d'information et la « sélection adverse ».*

2 | **Oliver E. WILLIAMSON** (1932-). *Né dans le Wisconsin, Oliver E. Williamson est docteur de l'Université Carnegie-Mellon. Professeur à Berkeley, il est le fondateur de la « nouvelle économie institutionnelle », où un rôle central est attribué au concept de coût de transaction, développé dans un article célèbre du prix Nobel 1991, Ronald Coase.*
Photo : © http://groups.haas.berkeley.edu/bpp/oew/

3 | **Maurice ALLAIS** (1911-). *Né à Paris, Maurice Allais est sorti major de l'École polytechnique en 1933. Il a obtenu le prix Nobel d'économie en 1988. Ses travaux ont eu une influence déterminante après-guerre sur les ingénieurs-économistes français (*L'Économie pure *(1943) et* Économie et intérêt *(1947)) mais une part significative de sa réputation internationale est due aussi au « paradoxe d'Allais », remise en cause de la théorie face au risque de von Neumann et Morgenstern.*

4 | **Joseph STIGLITZ** (1943-). *Né dans l'Indiana, Joseph Stiglitz est, à 26 ans, professeur à l'Université de Yale. La thèse de cet ancien étudiant du Massachusetts Institue of Technology (MIT), portant sur le rationnement du crédit, est célèbre dans le monde universitaire. J. Stiglitz développera par la suite ses analyses sur l'imperfection de l'information et ses conséquences sur le fonctionnement des marchés. Chef de file des nouveaux keynésiens, il a obtenu le prix Nobel d'économie en 2001 (en même temps que G. Akerlof et M. Spence).*

5 | **Robert LUCAS** (1937-). *Né dans l'État de Washington, Robert Lucas enseigne depuis 1965 à l'Université de Chicago. Principal représentant de la «nouvelle macroéconomie classique», le prix Nobel d'économie lui a été décerné en 1995 pour ses travaux sur les anticipations rationnelles et leurs conséquences quant à la stabilité des modèles économétriques* (Lucas's critique) *et aux limites des interventions publiques* (impotence result).
Photo : © Université de Chicago

6 | **Kenneth Joseph ARROW** (1921-). *Né à New-York, Kenneth J. Arrow s'oriente en 1941 vers l'économie à l'Université de Columbia. Il est connu pour sa démonstration de l'existence d'un équilibre général de concurrence, ses travaux sur le risque et son « théorème d'impossibilité » (agrégation 'impossible' des préférences individuelles en une fonction satisfaisante de choix collectif). Il a obtenu le prix Nobel d'économie en 1972, avec John Hicks.*

7 | **Paul KRUGMAN** (1953-). *Né à New-York, Paul Krugman est diplômé du Massachusetts Institue of Technology (MIT), université où il enseigne ainsi qu'à Yale, Stanford et Princeton. Ce nouveau keynésien, défenseur du libre-échange tempéré et spécialiste de l'économie internationale, s'appuie sur l'analyse de la concurrence imparfaite pour rectifier certaines des conclusions de l'analyse néoclassique.*

8 | **Milton FRIEDMAN** (1912 – 2006). *Né à Brooklyn, Milton Friedman a enseigné à l'Université de Chicago, de 1946 à 1977. Il a été le pape du retour au libre marché, de la déréglementation et de l'abandon de la politique budgétaire au profit de la politique monétaire. Chef de file d'une véritable contre-révolution keynésienne dès les années 50, il a vu ses idées triompher dans les années 70 et a reçu le prix Nobel en 1976.*

9 | **Barry EICHENGREEN** (1952-). *Né en Californie, Barry Eichengreen a fait des études d'économie et d'histoire à l'Université de Yale et enseigne aujourd'hui à l'Université de Berkeley. Il a notamment fait des propositions pour construire une architecture financière internationale et une architecture financière européenne.*
Photo : © 2008 Robert Houser

Source : « L'essentiel de l'économie », in Alternatives économiques*, Hors série pratique n° 21, novembre 2005.*

Ouvertures Économiques

Inférence statistique et probabilités

Stéphane **Mussard**
Françoise **Seyte**

Préface de Michel **Terraza**

de boeck

Crédits photos de couverture :
Si malgré nos soins attentifs, certaines demandes n'étaient pas parvenues aux auteurs ou à leurs ayants droits, qu'ils veuillent bien nous en tenir informés.

1re édition
Fond Jean Pâques, 4 – 1348 Louvain-la-Neuve

Imprimé en Belgique

Dépôt légal :
Bibliothèque nationale, Paris : mai 2014
Bibliothèque royale de Belgique, Bruxelles : 2014/0074/022

ISSN 2030-2061
ISBN 978-2-8041-8338-7

à Gladys, Cassandra et Michèle

à Auriane, Bertrand et Pierre[1]

1. En bons statisticiens les auteurs ont décidé de tirer à pile ou face l'ordre des remerciements.

PRÉFACE

L'enseignement de la Statistique à la Faculté d'Économie de Montpellier a débuté en 1969 à l'initiative du Professeur Jean-Pierre VIGNAU. Son objectif était de préparer les étudiants de l'époque aux méthodes statistiques indispensables pour développer une filière d'Économétrie au sein de la Faculté. Le cours de statistique mathématique, dénommé ainsi, pour le distinguer de la statistique non probabiliste, a été repris par la suite par la professeur Gilberte VIGNAU qui en a développé un programme scientifique cohérent et durable.

Cet enseignement perdure aujourd'hui avec Madame Françoise SEYTE, Maître de Conférences qui, à son tour, a su initier, à travers de multiples exercices pédagogiques de travaux dirigés, les étudiants de la Faculté à la Statistique Inférentielle. C'est le fruit de son travail qu'elle présente dans cet ouvrage, accompagnée de son collègue Stéphane MUSSARD, Maître de Conférences, qui a, lui aussi, œuvré dans ce projet pédagogique.

Je suis persuadé que les étudiants de nos Facultés sauront tirer profit des exercices corrigés qui sont proposés par mes deux collègues dans ce livre pour une meilleure compréhension de leur cours de statistique mais aussi, pour une bonne utilisation de cet outil lors de la pratique économétrique qu'ils exerceront dans la suite de leurs études.

Michel TERRAZA
Professeur d'Économie
LAMETA
Faculté d'Économie de Montpellier

AVANT-PROPOS

La Statistique Inférentielle consiste à déduire à partir d'un échantillon issu d'une population, les caractéristiques de cette population. La loi de la population peut être connue. Si elle ne l'est pas, des tests permettent de la trouver.

L'objet de cet ouvrage est de s'exercer à la Statistique Inférentielle à l'aide de problèmes corrigés et expliqués. Les fiches de synthèse en début de chapitre permettent de retenir les concepts clés nécessaires à la résolution des exercices. Chaque fin de chapitre est consacrée à une mise au point concernant les difficultés rencontrées lors des exercices corrigés. A la fin des chapitres 5 et 11 deux synthèses sont proposées. Il s'agit de vérifier les connaissances sur des exercices qui rassemblent plusieurs problèmes.

Le livre s'articule de la façon suivante :

Le Chapitre 1 porte sur le calcul de probabilités, rappelant les propriétés de base des opérateurs (provenant des Algèbres de Boole).

Le Chapitre 2 introduit les variables aléatoires à une dimension afin de revoir des opérateurs tels que l'espérance, la variance, la covariance, etc. Les lois de comptage usuelles : la loi de Bernoulli, la loi Binomiale, la loi Hypergéométrique et la loi de Poisson y sont aussi développées.

Le Chapitre 3 traite des variables aléatoires continues à une dimension permettant ainsi le calcul de probabilités avec densités et fonctions de répartition (rappel sur la fonction Gamma et les intégrations par parties).

Le Chapitre 4 insiste sur les variables aléatoires discrètes à deux dimensions. Les tableaux de contingence sont étudiés à travers la notion d'indépendance notamment ; il en est de même pour les courbes de régression, les probabilités conditionnelles et espérances conditionnelles.

Le Chapitre 5 s'intéresse aux variables aléatoires continues à deux dimensions et permet de s'entrainer au calcul de probabilités jointes ou conditionnelles avec intégrales doubles.

Le Chapitre 6 est consacré aux lectures des tables statistiques relatives aux lois continues : lois Normale, Khi-Deux, Student, Fisher-Snedecor. La question de la construction des lois continues est abordée.

Le Chapitre 7 sur la convergence en loi explique les conditions pour lesquelles certaines variables aléatoires (Binomiale ou Poisson) convergent vers la loi Normale.

Le Chapitre 8 sur les distributions d'échantillonnage montre comment on peut associer à certaines statistiques issues d'échantillons empiriques une loi de probabilité afin de résoudre des problèmes concrets portant sur les moyennes, les variances, les écart-types et les proportions.

Le Chapitre 9 s'intéresse à la méthode du maximum de vraisemblance. L'utilisation de certaines lois de probabilité nécessite l'estimation d'un ou plusieurs paramètres. La méthode du maximum de vraisemblance permet de trouver des estimateurs pour ces paramètres inconnus et de vérifier leur qualité.

Le Chapitre 10 développe les tests du khi-deux d'homogénéité et d'indépendance. Ces tests montrent la liaison existante ou non entre deux échantillons ou deux variables aléatoires. Il porte également sur les tests du khi-deux d'adéquation qui permettent de vérifier que les données observées s'adaptent bien à une loi initialement définie.

Le Chapitre 11 explique enfin la construction des intervalles de confiance et des tests d'hypothèses (signification et comparaison).

Ce livre s'adresse aux étudiants des facultés d'économie, de gestion, de sciences, d'écoles de commerce et d'ingénieurs qui veulent approfondir la Statistique Inférentielle à partir de cas pratiques. Mais n'oublions pas aussi les chercheurs, les économistes d'entreprise qui, confrontés à des problèmes d'échantillonnage et de tests statistiques, peuvent trouver des réponses pratiques aux questions qu'ils se posent.

1

Algèbre de Boole et rappels sur le calcul des probabilités

SOMMAIRE

Ce chapitre définit le concept d'Algèbre de Boole et rappelle les principaux résultats à connaître dans le domaine des probabilités. Nous montrons comment la plus petite Algèbre de Boole peut être construite à partir d'une famille de parties donnée. Cette notion est importante à comprendre car en statistique inférentielle, nous travaillons dans le contexte d'une Algèbre de Boole probabilisée. La méthode de calcul est d'abord expliquée. Puis les principaux axiomes du calcul des probabilités sont abordés.

FICHE DE SYNTHÈSE

Notations

→ F : famille de parties

→ Ω : univers

→ A et B : deux parties (sous-ensembles) de Ω

Algèbre de Boole

→ Propriété 1 (P1) : F est fermée (stable) par rapport à la complémentation relativement à Ω :

$\forall\ A \subseteq \Omega$, si $A \in F$, alors $\overline{A} \in F$.

→ Propriété 2 (P2) : F est fermée (stable) par rapport à l'union finie :

$\forall\ A, B \subseteq \Omega$, si $A \in F$ et si $B \in F$ alors $A \cup B \in F$.

Probabilité

→ Il s'agit d'une fonction d'ensemble c'est-à-dire une application qui a un évènement associe un réel compris entre 0 et 1.

→ Notation : P ou Prob tel que, $A \overset{P}{\longmapsto} P(A) \in [0\ ;\ 1]$.

→ $P(\varnothing) = 0$

→ $P(\Omega) = 1$ (axiome de normalisation)

→ P : fonction croissante

→ $0 \leqslant P(A) \leqslant 1$

→ $P(\overline{A}) = 1 - P(A)$

Théorèmes

➢ Théorème des probabilités totales :

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

→ A et B évènements indépendants : $P(A \cap B) = P(A) \times P(B)$

→ A et B évènements incompatibles : $A \cap B = \varnothing$ soit $P(A \cap B) = 0$

➢ Axiome des probabilités composées (Théorème de Bayes) :

$$P(B/A) = P_A(B) = \frac{P(A \cap B)}{P(A)}, \text{ avec } P(A) \neq 0$$

→ A et B évènements indépendants : $P(B/A) = P(B)$

→ Généralisation : $P\left(\bigcup_{i=1}^{n} B_i/A\right) = \sum_{i=1}^{n} P(B_i/A)$

➢ Calcul d'une probabilité dans le cas où les issues sont en nombre fini et équiprobables :

$$P(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}.$$

EXERCICE 1 : DÉTERMINATION D'UNE ALGÈBRE DE BOOLE MINIMALE

Soit Ω un ensemble de cinq éléments (a), (b), (c), (d), (e). Une partie (un sous-ensemble) de Ω est désignée par ses éléments écrits entre parenthèses. Construire la plus petite algèbre de Boole notée F contenant les parties (a), (a, b) et (b, c, d).

Rappelons tout d'abord les deux seules propriétés qui nous permettent de résoudre ce type d'exercice. F est une algèbre de Boole si :

Propriété 1 (P1) : F est fermée par rapport à la complémentation relativement à Ω :

$\forall A \subseteq \Omega$, si $A \in F$, alors $\overline{A} \in F$;

Propriété 2 (P2) : F est fermée par rapport à l'union finie :

$\forall A, B \subseteq \Omega$, si $A \in F$ et si $B \in F$, alors $A \cup B \in F$.

- Le but de l'exercice est d'obtenir un ensemble F contenant au moins les parties *initiales* (a), (a, b) et (b, c, d). Si F est une algèbre de Boole alors les unions des parties initiales vont nécessairement donner naissance à de nouvelles parties.

Ce qui nous incite à dire que les parties *initiales* imposées dans l'énoncé (a), (a, b) et (b, c, d) ne sont pas suffisantes pour créer une algèbre de Boole.

- La technique consiste à alterner les propriétés P1 et P2, à la manière d'un algorithme, permettant de caractériser une algèbre de Boole. On peut commencer par l'une ou l'autre. Après avoir alterné plusieurs fois ces deux propriétés, l'algorithme va s'arrêter : aucune nouvelle partie supplémentaire ne pourra être créée. Il ne manquera plus qu'à recenser toutes les parties qui ont été trouvées au cours de l'alternance des deux propriétés. Toutes ces nouvelles parties, avec les parties *initiales* (a), (a, b) et (b, c, d) donneront la plus petite algèbre de Boole.
- Afin de ne pas comptabiliser plusieurs fois les parties trouvées à l'aide de P1 et P2, nous noterons les nouvelles parties créées de la manière suivante ✓ et celles que nous avons déjà trouvées (ou même les parties initiales) avec $\times$.
- Remarque 1 : Une algèbre de Boole doit toujours contenir Ω et $\varnothing$ (conséquence de la définition d'une algèbre de Boole).
- Remarque 2 : Lorsqu'une nouvelle partie a été créée, veiller à bien écrire les éléments de cette nouvelle partie en respectant l'ordre alphabétique pour éviter d'obtenir plusieurs fois la même partie [car en effet (a, b) = (b, a)].
- Remarque 3 : Afin d'éviter toute erreur, avant de commencer l'algorithme, il est nécessaire de bien identifier Ω. L'énoncé indique que Ω est constitué de cinq éléments (a), (b), (c), (d), (e). Il s'écrit donc Ω = (a, b, c, d, e). Ceci est crucial notamment pour la propriété P1. Lorsque des parties complémentaires doivent être déterminées, celles-ci le sont par rapport à Ω. Par exemple $(\bar{a})$ est le complémentaire de (a) relativement à Ω. Par conséquent $(\bar{a})$ est la partie de Ω qui ne contient pas (a). Il suffit donc d'enlever (a) à Ω = (a, b, c, d, e) : $(\bar{a})$ = (b, c, d, e). Si on avait eu Ω = (a, b, c, d, e, f), alors $(\bar{a})$ = (b, c, d, e, f).
- Commençons maintenant la résolution du problème avec P2.

Étape 1 : *application de P2*

La propriété P2 donnera une seule nouvelle partie : (a, b, c, d). Pour le voir, il faut appliquer P2 : c'est-à-dire faire l'union des parties initiales (a), (a, b) et (b, c, d) deux à deux :

(a) $\cup$ (a, b) = (a, b) $\times$: nous mettons une croix car (a, b) est une partie que nous avons déjà.

(a) $\cup$ (b, c, d) = (a, b, c, d) ✓ : nous cochons cette partie car nous ne l'avions pas.

(a, b) $\cup$ (b, c, d) = (a, b, c, d) $\times$: nous mettons une croix car nous venons de la trouver.

Conclusion : l'algèbre de Boole contiendra au final la partie (a, b, c, d) et les parties *initiales.*

Étape 2 : *application de P1*

Cette étape consiste à chercher les complémentaires des parties *initiales* **et** des parties créées à l'étape précédente !

$(\overline{a}) = (b, c, d, e)$ ✓

$\overline{(a, b)} = (c, d, e)$ ✓

$\overline{(b, c, d)} = (a, e)$ ✓

$\overline{(a, b, c, d)} = (e)$ ✓

Conclusion : Ces 4 nouvelles parties seront incluses dans l'algèbre de Boole (avec les parties *initiales* et la partie trouvée à l'étape 1).

Étape 3 : *application de P2*

Cette étape consiste à unir les nouvelles parties que nous avons trouvées. <u>Attention</u> : il faut établir les unions entre les parties créées précédemment mais aussi avec les parties initiales !

- Partons justement des parties initiales et formons les unions avec les parties trouvées aux étapes 1 et 2 :

  $(a) \cup (a, b, c, d) = (a, b, c, d)$ ×
  $(a) \cup (b, c, d, e) = (a, b, c, d, e) = \Omega$ ✓
  $(a) \cup (c, d, e) = (a, c, d, e)$ ✓
  $(a) \cup (a, e) = (a, e)$ ×
  $(a) \cup (e) = (a, e)$ ×

  $(b, c, d) \cup (a, b, c, d) = (a, b, c, d)$ ×
  $(b, c, d) \cup (b, c, d, e) = (b, c, d, e)$ ×
  $(b, c, d) \cup (c, d, e) = (b, c, d, e)$ ×
  $(b, c, d) \cup (a, e) = \Omega$ ×
  $(b, c, d) \cup (e) = (b, c, d, e)$ ×

  $(a, b) \cup (a, b, c, d) = (a, b, c, d)$ ×
  $(a, b) \cup (b, c, d, e) = (a, b, c, d, e) = \Omega$ ×
  $(a, b) \cup (c, d, e) = \Omega$ ×
  $(a, b) \cup (a, e) = (a, b, e)$ ✓
  $(a, b) \cup (e) = (a, b, e)$ ×

- Ensuite, avec les parties créées à l'étape 1, formons les unions avec les parties trouvées à l'étape 2 :
  $(a, b, c, d) \cup (b, c, d, e) = \Omega$ ×
  $(a, b, c, d) \cup (c, d, e) = \Omega$ ×
  $(a, b, c, d) \cup (a, e) = \Omega$ ×
  $(a, b, c, d) \cup (e) = \Omega$ ×

  Lorsqu'une partie contient beaucoup d'éléments comme (a, b, c, d), il n'est pas nécessaire de faire toutes les unions possibles puisque (a, b, c, d) peut aboutir à Ω uniquement.

- Finalement, avec les parties créées à l'étape 2, formons les unions avec ces mêmes parties trouvées à l'étape 2 (inutile de faire des unions avec les parties des étapes précédentes, nous referions les mêmes calculs que ceux effectués précédemment) :

(b, c, d, e) ∪ (c, d, e) = (b, c, d, e) ×
(b, c, d, e) ∪ (a, e) = **Ω** ×
(b, c, d, e) ∪ (e) = (b, c, d, e) ×
(c, d, e) ∪ (a, e) = (a, c, d, e) ×
(c, d, e) ∪ (e) = (c, d, e) ×
(a, e) ∪ (e) = (a, e) ×

Étape 4 : *application de P1*

On utilise P1 sur les parties trouvées à l'étape précédente (étape 3 : ne pas revenir aux parties de l'étape 2, car en appliquant P1 sur celles-ci on retrouverait automatiquement des parties existantes) :

$\overline{\Omega} = \varnothing$ ✓
$\overline{(a, c, d, e)} = (b)$ ✓
$\overline{(a, b, e)} = (c, d)$ ✓

Étape 5 : *application de P2*

On va construire les unions entre les nouvelles parties créées à l'étape 4 précédente et toutes les autres parties créées avant cette étape **sans oublier** les parties *initiales.*

∅ ∪ (a) = (a) × : unir l'ensemble vide avec d'autres parties ne donnera aucune nouvelle partie. Ainsi les calculs de P2 liés à ∅ seront toujours omis (de même que ceux liés à **Ω**).

(b) ∪ (a) = (a, b) ×
(b) ∪ (a, b) = (a, b) ×
(b) ∪ (b, c, d) = (b, c, d) ×
(b) ∪ (a, b, c, d) = (a, b, c, d) ×
(b) ∪ (b, c, d, e) = (b, c, d, e) ×
(b) ∪ (c, d, e) = (b, c, d, e) ×
(b) ∪ (a, e) = (a, b, e) ×
(b) ∪ (e) = (b, e) ×
(b) ∪ **Ω** = **Ω** ×
(b) ∪ (a, c, d, e) = **Ω** ×
(b) ∪ (a, b, e) = (a, b, e) ×
(b) ∪ ∅ = (b) ×
(b) ∪ (c, d) = (b, c, d) ×

(c, d) ∪ (a) = (a, c, d) ✓
(c, d) ∪ (a, b) = (a, b, c, d) ×
(c, d) ∪ (b, c, d) = (b, c, d) ×
(c, d) ∪ (a, b, c, d) = (a, b, c, d) ×
(c, d) ∪ (b, c, d, e) = (b, c, d, e) ×
(c, d) ∪ (c, d, e) = (c, d, e) ×
(c, d) ∪ (a, e) = (a, c, d, e) ×
(c, d) ∪ (e) = (c, d, e) ×
(c, d) ∪ **Ω** = **Ω** ×
(c, d) ∪ (a, c, d, e) = (a, c, d, e) ×
(c, d) ∪ (a, b, e) = **Ω** ×
(c, d) ∪ ∅ = (c, d) ×

Étape 6 : *application de P1*

On applique P1 sur la seule partie créée à l'étape précédente :

$\overline{(a, c, d)} = (b, e)$ ✓

Étape 7 : *application de P2*

On va unir la partie (b, e) trouvée à l'étape 6 avec toutes les parties créées sans oublier les parties *initiales* :

$(b, e) \cup (a) = (a, b, e) \times$

$(b, e) \cup (a, b) = (a, b, e) \times$

$(b, e) \cup (b, c, d) = (b, c, d, e) \times$

$(b, e) \cup (a, b, c, d) = \Omega \times$

$(b, e) \cup (b, c, d, e) = (b, c, d, e) \times$

$(b, e) \cup (c, d, e) = (b, c, d, e) \times$

$(b, e) \cup (a, e) = (a, b, e) \times$

$(b, e) \cup (e) = (b, e) \times$

$(b, e) \cup \Omega = \Omega \times$

$(b, e) \cup (a, c, d, e) = \Omega \times$

$(b, e) \cup (a, b, e) = (a, b, e) \times$

$(b, e) \cup \varnothing = (b, e) \times$

$(b, e) \cup (c, d) = (b, c, d, e) \times$

$(b, e) \cup (a, c, d) = \Omega \times$

Étape 8 : *Conclusion générale*

L'étape précédente n'a donné aucune nouvelle partie. L'algorithme a donc convergé. On peut ainsi clore l'exercice en rangeant toutes les parties trouvées avec les parties *initiales* par ordre alphabétique, soit au total 16 parties :

$F = \{\varnothing, (a), (b), (e), (a, b), (a, e), (b, e), (c, d), (a, b, e), (a, c, d), (b, c, d), (c, d, e), (a, b, c, d), (a, c, d, e), (b, c, d, e), \Omega\}$.

EXERCICE 2 : DÉTERMINATION D'UNE ALGÈBRE DE BOOLE MINIMALE

Soit Ω un ensemble de cinq éléments (a), (b), (c), (d), (e). Une partie de Ω est désignée par ses éléments écrits entre parenthèses. Quelle est la plus petite algèbre de Boole notée F contenant les parties (c), (e), (b, e), (a, d, e) ?

Pour cet exercice, nous allons commencer par l'étape P1.

Étape 1 : application de P1

Cette étape consiste à chercher les complémentaires des parties *initiales* (c), (e), (b, e) et (a, d, e) :

$\overline{(c)}$ = (a, b, d, e) ✓

$\overline{(e)}$ = (a, b, c, d) ✓

$\overline{(b, e)}$ = (a, c, d) ✓

$\overline{(a, d, e)}$ = (b, c) ✓

Étape 2 : application de P2

Cette étape consiste à unir les parties *initiales* entre elles, les nouvelles parties trouvées à l'étape 1 entre elles, et à unir deux à deux les parties initiales avec les nouvelles parties de l'étape 1.

(c) ∪ (e) = (c, e) ✓
(c) ∪ (b, e) = (b, c, e) ✓
(c) ∪ (a, d, e) = (a, c, d, e) ✓
(e) ∪ (b, e) = (b, e) ×
(e) ∪ (a, d, e) = (a, d, e) ×
(b, e) ∪ (a, d, e) = (a, b, d, e) ×

(a, b, d, e) ∪ (a, b, c, d) = Ω ✓
(a, b, d, e) ∪ (a, c, d) = Ω ×
(a, b, d, e) ∪ (b, c) = Ω ×
(a, b, c, d) ∪ (a, c, d) = (a, b, c, d) ×
(a, b, c, d) ∪ (b, c) = (a, b, c, d) ×
(a, c, d) ∪ (b, c) = (a, b, c, d) ×

(c) ∪ (a, b, d, e) = Ω ×
(c) ∪ (a, b, c, d) = (a, b, c, d) ×
(c) ∪ (a, c, d) = (a, c, d) ×
(c) ∪ (b, c) = (b, c) ×
(e) ∪ (a, b, d, e) = (a, b, d, e) ×
(e) ∪ (a, b, c, d) = Ω ×
(e) ∪ (a, c, d) = (a, c, d, e) ×
(e) ∪ (b, c) = (b, c, e) ×

(b, e) ∪ (a, b, d, e) = (a, b, d, e) ×
(b, e) ∪ (a, b, c, d) = Ω ×
(b, e) ∪ (a, c, d) = Ω ×
(b, e) ∪ (b, c) = (b, c, e) ×
(a, d, e) ∪ (a, b, d, e) = (a, b, d, e) ×
(a, d, e) ∪ (a, b, c, d) = Ω ×
(a, d, e) ∪ (a, c, d) = (a, c, d, e) ×
(a, d, e) ∪ (b, c) = Ω ×

Étape 3 : application de P1

On utilise P1 sur les parties trouvées à l'étape 2.

$\overline{(c, e)}$ = (a, b, d) ✓

$\overline{(b, c, e)}$ = (a, d) ✓

$\overline{(a, c, d, e)}$ = (b) ✓

$\overline{\Omega} = \varnothing$ ✓

Étape 4 : *application de P2*

Prenons toutes les unions entre les nouvelles parties déterminées à l'étape 3 et toutes les autres parties créées avant cette étape **sans oublier** les parties *initiales* (là encore on omettra les unions entre ∅ et les autres parties).

(a, b, d) ∪ (c) = (a, b, c, d) ×
(a, b, d) ∪ (e) = (a, b, d, e) ×
(a, b, d) ∪ (b, e) = (a, b, d, e) ×
(a, b, d) ∪ (a, d, e) = (a, b, d, e) ×
(a, b, d) ∪ (a, b, d, e) = (a, b, d, e) ×
(a, b, d) ∪ (a, b, c, d) = (a, b, c, d) ×
(a, b, d) ∪ (a, c, d) = (a, b, c, d) ×
(a, b, d) ∪ (b, c) = (a, b, c, d) ×
(a, b, d) ∪ (c, e) = **Ω** ×
(a, b, d) ∪ (b, c, e) = **Ω** ×
(a, b, d) ∪ (a, c, d, e) = **Ω** ×
(a, b, d) ∪ **Ω** = **Ω** ×
(a, b, d) ∪ (a, d) = (a, b, d) ×
(a, b, d) ∪ (b) = (a, b, d) ×

(a, d) ∪ (c) = (a, c, d) ×
(a, d) ∪ (e) = (a, d, e) ×
(a, d) ∪ (b, e) = (a, b, d, e) ×
(a, d) ∪ (a, d, e) = (a, d, e) ×
(a, d) ∪ (a, b, d, e) = (a, b, d, e) ×
(a, d) ∪ (a, b, c, d) = (a, b, c, d) ×
(a, d) ∪ (a, c, d) = (a, c, d) ×
(a, d) ∪ (b, c) = (a, b, c, d) ×
(a, d) ∪ (c, e) = (a, c, d, e) ×
(a, d) ∪ (b, c, e) = **Ω** ×
(a, d) ∪ (a, c, d, e) = (a, c, d, e) ×
(a, d) ∪ **Ω** = **Ω** ×
(a, d) ∪ (b) = (a, b, d) ×

(b) ∪ (c) = (b, c) ×
(b) ∪ (e) = (b, e) ×
(b) ∪ (b, e) = (b, e) ×
(b) ∪ (a, d, e) = (a, b, d, e) ×
(b) ∪ (a, b, d, e) = (a, b, d, e) ×
(b) ∪ (a, b, c, d) = (a, b, c, d) ×
(b) ∪ (a, c, d) = (a, b, c, d) ×
(b) ∪ (b, c) = (b, c) ×
(b) ∪ (c, e) = (b, c, e) ×
(b) ∪ (b, c, e) = (b, c, e) ×
(b) ∪ (a, c, d, e) = **Ω** ×
(b) ∪ **Ω** = **Ω** ×

Étape 5 : *Conclusion générale*

L'étape précédente n'a donné aucune nouvelle partie. L'algorithme s'arrête. En rangeant toutes les parties trouvées avec les parties *initiales* par ordre alphabétique, on a :

F = {∅, (b), (c), (e), (a, d), (b, c), (b, e), (c, e), (a, b, d), (a, c, d), (a, d, e), (b, c, e), (a, b, c, d), (a, b, d, e), (a, c, d, e), Ω}.

<u>Remarque</u> : On constate que la résolution de cet exercice compte 5 étapes contre 8 pour le précédent. Dans la pratique, il est en effet préférable de commencer par l'étape 1 (cela permettra d'extraire le maximum de parties à la deuxième étape).

EXERCICE 3 : DÉDUIRE LA BONNE ALGÈBRE DE BOOLE

Soit Ω un ensemble de cinq éléments a, b, c, d et e. Une partie de Ω est désignée par ses éléments écrits entre parenthèses. Quelle est la plus petite algèbre de Boole F contenant les parties (b), (d), (a, e) et (a, d, e) ?

Réponses : F =

$\boxed{\text{a}}$ {∅, (b), (d), (b, d), (a, e), (b, c), (a, b, e), (b, c, d), (a, d, e), (a, b, d, e), (a, c, d, e), (a, b, c, e), Ω}.

$\boxed{\text{b}}$ {∅, (b), (c), (d), (a, e), (c, d), (a, b, e), (a, c, e), (a, d, e), (b, c, d), (a, b, c, e), (a, b, d, e), (a, c, d, e), (a, b, c, d, e)}.

$\boxed{\text{c}}$ {∅, (b), (c), (d), (a, d), (b, c), (b, d), (a, b, e), (a, c, d), (a, d, e), (b, c, e), (a, b, , d), (a, b, d, e), (a, c, d, e)}.

$\boxed{\text{d}}$ autre :

La réponse a) n'est pas une algèbre de Boole car le nombre de parties est impair. La seconde réponse b) n'est pas une algèbre de Boole car les parties (b), (c) sont incluses dans F mais pas leur union (b, c). La réponse c) est fausse car F ne contient pas Ω. La réponse est donc d).

EXERCICE 4 : CALCULS DE PROBABILITÉS ÉLÉMENTAIRES

Cassandra est à l'aéroport Pierrefonds. Seulement deux cabines téléphoniques sont présentes dans le hall de l'aéroport. On note C_1 et C_2 les deux seules cabines téléphoniques.

Soit A l'évènement « la Cabine C_1 est occupée ».

Soit B l'évènement « la Cabine C_2 est occupée ».

On donne les probabilités suivantes : P(A) = 0,7 ; P(B) = 0,5. Les évènements A et B sont indépendants. Calculez les probabilités des évènements E, F, G, H, I suivants.

- E : « la Cabine C_1 est libre ».
- F : « une Cabine au moins est occupée ».
- G : « les deux Cabines sont libres ».
- H : « une Cabine seulement est occupée ».
- I : « une Cabine au moins est libre ».

Calculer P(E)

Notons respectivement $\overline{A}$ et $\overline{B}$ les évènements « la Cabine C_1 est libre » et « la Cabine C_2 est libre ».

Rappelons que la recherche d'une probabilité liée à un évènement complémentaire nécessite la connaissance de Ω qui est l'univers avec $P(\Omega) = 1$. D'où :

$$P(E) = P(\overline{A}) = 1 - P(A) = 1 - 0{,}7 = 0{,}3.$$

Calculer P(F)

On pourrait réécrire l'évènement F : « C_1 est occupée **'ou'** C_2 est occupée ». En écrivant les évènements de manière à faire explicitement intervenir les opérateurs unions **'ou'** [$\cup$] ou intersection **'et'** [$\cap$], il est plus facile de répondre à la question posée. Ainsi, l'évènement F s'exprime de la manière suivante :

$$P(F) = P(A \cup B)$$

<u>Remarque</u> : Théorème des probabilités totales : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Donc : $P(F) = P(A) + P(B) - P(A \cap B)$.

On constate que $P(A \cap B)$ est inconnu. On utilise une information contenue dans l'énoncé. Les évènements A et B sont indépendants : ceci implique que $P(A \cap B) = P(A).P(B)$. D'où

$$P(A \cap B) = P(A).P(B) = 0{,}7 \times 0{,}5 = 0{,}35.$$

On en déduit alors :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}7 + 0{,}5 - 0{,}35 = 0{,}85.$$

Calculer P(G)

Là aussi, nous pouvons exprimer G de la manière suivante : « C_1 est libre **et** C_2 est libre ». On trouve alors :

$$P(G) = P(\overline{A} \cap \overline{B}).$$

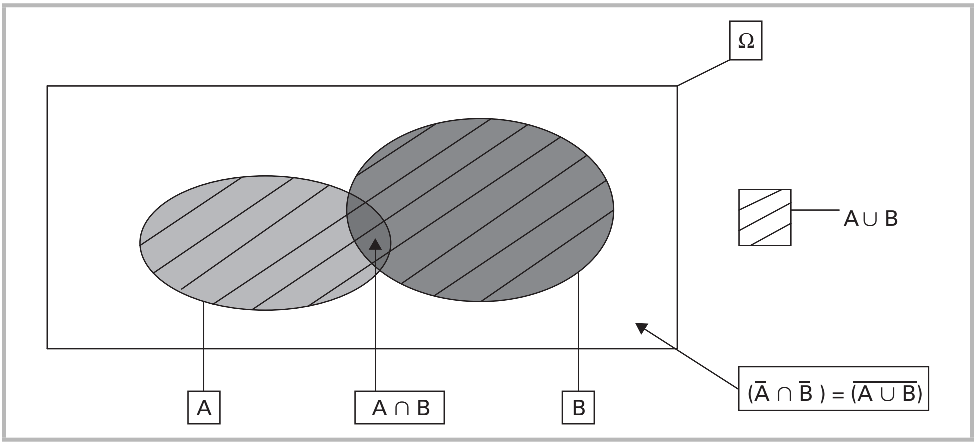
Pour ce type d'évènement légèrement plus complexe, deux types de résolution sont possibles.

→ On passe par la loi (théorème) de Morgan : $\overline{\cup} = \cap$ et $\overline{\cap} = \cup$. Donc en réécrivant l'évènement G, on a :

$$P(G) = P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0{,}85 = 0{,}15.$$

<u>Remarque</u> : G est l'évènement complémentaire de I.

→ On peut aussi résoudre le problème avec un graphique (diagramme de Venn).



Calculer P(H)

H : « une Cabine seulement est occupée » *i.e.* « soit C_1 est occupée soit C_2 est occupée ».

Remarque : Pour l'évènement F, l'intersection était permise : les deux cabines pouvaient être occupées simultanément ('ou' non exclusif). Pour H, l'intersection est non permise (il s'agit d'un 'ou' exclusif). Il existe plusieurs manières de l'écrire :

→ L'une au moins est occupée $(A \cup B)$ mais les deux cabines ne peuvent pas être simultanément occupées $(A \cap B)$: donc on mesure $P(H) = P(A \cup B) - P(A \cap B) = 0{,}85 - 0{,}35 = 0{,}5$.

→ C_1 est occupée mais pas (C_1 **'et'** C_2) **'ou'** C_2 est occupée mais pas (C_1 **'et'** C_2) :

$$P(H) = P\{[A \backslash (A \cap B)] \cup [B \backslash (A \cap B)]\}.$$

Le signe \ permet de soustraire à un ensemble un autre ensemble. D'où :

$$P(H) = P\{[A \backslash (A \cap B)] \cup [B \backslash (A \cap B)]\} = P(A) - P(A \cap B) + P(B) - P(A \cap B)$$
$$= 0{,}7 - 0{,}35 + 0{,}5 - 0{,}35 = 0{,}5.$$

Remarque : Si on note l'évènement $C = [A \backslash (A \cap B)]$ et l'évènement $D = [B \backslash (A \cap B)]$, on devrait avoir en appliquant le théorème des probabilités totales :

$$P(H) = P(C) + P(D) - P(C \cap D).$$

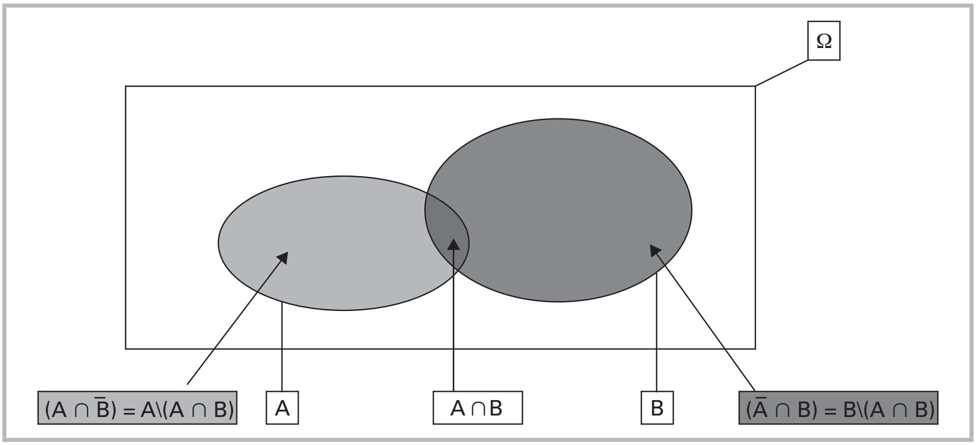
Or on voit sur le graphique précédent que $(C \cap D) = \varnothing$. Le résultat $P(H) = 0{,}5$ est donc correct.

→ Une autre manière d'écrire le 'ou' exclusif est :

$$P(H) = P[(\overline{A} \cap B) \cup (A \cap \overline{B})].$$

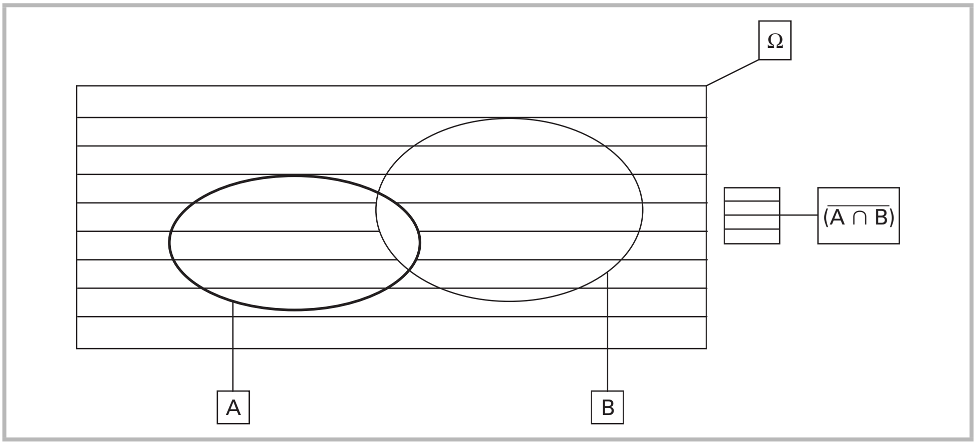
Là encore on voit que $(\overline{A} \cap B) \cap (A \cap \overline{B}) = \varnothing$: ces évènements sont disjoints. On retrouve alors le résultat précédent puisque :

$$(\overline{A} \cap B) = B \backslash (A \cap B) \text{ et } (A \cap \overline{B}) = A \backslash (A \cap B).$$



Calculer P(I)

$P(I) = P(« C_1 \text{ est libre } \textbf{'ou'} \ C_2 \text{ est libre »}) = P(\overline{A} \cup \overline{B}) = P(\overline{A \cap B})$
$= 1 - P(A \cap B) = 1 - 0{,}35 = 0{,}65.$



EXERCICE 5 : CALCULS DE PROBABILITÉS ÉLÉMENTAIRES

Auriane participe à deux jeux organisés par la kermesse de son école primaire. Soit A l'évènement : « Auriane gagne au premier jeu », et l'évènement B : « Auriane gagne au deuxième jeu », tel que $P(A) = 0{,}3$; $P(B) = 0{,}4$ et $P(A \cap B) = 0{,}15$. Calculer les probabilités suivantes :

- Auriane gagne au moins à un des jeux ;
- Auriane ne gagne qu'à un seul jeu ;

- Auriane gagne le deuxième jeu, sachant qu'Auriane a perdu au premier jeu ;
- Auriane perd aux deux jeux.

Auriane gagne au moins à un des jeux ('ou' non exclusif) :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}3 + 0{,}4 - 0{,}15 = 0{,}55.$$

Remarque : on utilise ici $(A \cap B)$ tel que celui-ci est donné dans l'énoncé sans utiliser la notion d'indépendance (celle-ci n'étant pas donnée dans l'énoncé, de plus, A et B ne semblent pas forcément indépendants).

Auriane ne gagne qu'à un seul jeu ('ou' exclusif) :

$$P(A \cup B) - P(A \cap B) = 0{,}55 - 0{,}15 = 0{,}4.$$

Auriane gagne le deuxième jeu, sachant qu'Auriane a perdu au premier jeu :

On applique le théorème de Bayes :

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

On peut déduire des graphiques précédents :

$$P(B \cap \overline{A}) = P[B\backslash(A \cap B)] = P(B) - P(A \cap B) = 0{,}4 - 0{,}15 = 0{,}25.$$

Comme $P(\overline{A}) = 1 - P(A) = 0{,}7$, alors :

$$P(B/\overline{A}) = 0{,}25/0{,}7 = 0{,}357.$$

<u>Remarque</u> : on note une probabilité conditionnelle $P(B/\overline{A})$ ou de manière équivalente $P_{\overline{A}}(B)$.

Auriane perd aux deux jeux :

$$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 0{,}45.$$

EXERCICE 6 : REMARQUE SUR L'INTERSECTION

Soient deux évènements A et B, tels que $P(A) = 0{,}5$; $P(B) = 0{,}2$ et $P(A \cap B) = 0{,}3$. Calculer $P(A \cup B)$.

Calculer P(A ∪ B)

Imaginons que les probabilités données dans l'énoncé proviennent d'observations (fréquences). Supposons que nous analysons le chiffre d'affaire d'un magasin tel que P(A) = 50 % du chiffre d'affaire provient des ventes de chaussures ; P(B) = 20 % du chiffre d'affaire provient des articles de sport. P(A ∩ B) = 0,3 signifie que 30 % du chiffre d'affaire provient des articles de sport qui sont des chaussures. Ceci est impossible puisque le chiffre d'affaire provenant des articles de sport ne peut dépasser 20 %.

Remarque : On doit toujours avoir $P(A \cap B) \leqslant P(A)$ et $P(A \cap B) \leqslant P(B)$.

Par conséquent, il est inutile de répondre à la question, le calcul des probabilités n'aura aucun sens.

EXERCICE 7 : CALCULS DE PROBABILITÉS CONDITIONNELLES

Soient deux évènements A et B, tels que P(A) = 0,5 ; P(B) = 0,3 et P(A ∩ B) = 0,2. Calculer :

1) $P(\overline{A} \cap \overline{B})$
2) $P(\overline{A}/\overline{B})$
3) $P(A/\overline{B})$
4) $P(A/A \cup B)$
5) $P(A/A \cap B)$
6) $P(A \cap B/A \cup B)$

1) $P(\overline{A} \cap \overline{B})$:

$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] = 0{,}4.$

2) $P(\overline{A}/\overline{B})$:

$P(\overline{A}/\overline{B}) = P(\overline{A} \cap \overline{B})/P(\overline{B}) = 0{,}4/0{,}7 = 4/7.$

3) $P(A/\overline{B})$:

$P(A/\overline{B}) = 1 - P(\overline{A}/\overline{B}) = 1 - 4/7 = 3/7.$

4) $P(A/A \cup B)$:

$P(A/A \cup B) = P[A \cap (A \cup B)] / P(A \cup B)$

On sait que $A \cap (A \cup B) = A$. En effet, la partie commune à A et à $(A \cup B)$ est bien A (*Cf.* les graphiques précédents). D'où :

$P(A/A \cup B) = P[A] / P(A \cup B) = 0{,}5 / 0{,}6 = 5/6$.

5) $P(A/A \cap B)$

$P(A/A \cap B) = P[A \cap (A \cap B)] / P(A \cap B)$

L'intersection entre (ou les parties communes à) A et $(A \cap B)$ est : $(A \cap B)$. D'où

$P(A/A \cap B) = P[(A \cap B)] / P(A \cap B) = 1$.

6) $P(A \cap B/A \cup B)$

$P(A \cap B/A \cup B) = P[(A \cap B) \cap (A \cup B)] / P(A \cup B)$

Graphiquement on peut voir que l'intersection entre $(A \cap B)$ et $(A \cup B)$ est $(A \cap B)$. Donc :

$P(A \cap B/A \cup B) = P[(A \cap B)] / P(A \cup B) = 0{,}2 / 0{,}6 = 1/3$.

EXERCICE 8 : CALCULS DE PROBABILITÉS CONDITIONNELLES

I. Pour se rendre de son domicile à son travail, Bertrand peut utiliser deux types de train soit le TGV soit le TER. Les TER effectuent 30 % des trajets entre la ville de Bertrand et son lieu de travail, tandis que les TGV en effectuent 70 %. S'il s'agit d'un TER, la probabilité que le TER soit en retard est de 5 % contre 7 % pour les TGV. Bertrand prend un train au hasard dans sa gare.

On notera les évènements TGV : « prendre un TGV » ; TER : « prendre un TER » et R : « le train est en retard ».

1) Calculer la probabilité pour qu'un train soit en retard.

2) Si un train est en retard, quelle est la probabilité que Bertrand ait pris un TER (arrondir à 10^{-2}) ?

3) Sachant qu'un train est à l'heure quelle est la probabilité que ce soit un TER ?

1) Calculer la probabilité pour qu'un train soit en retard

$$P(R) = P(R \cap TGV) + P(R \cap TER)$$

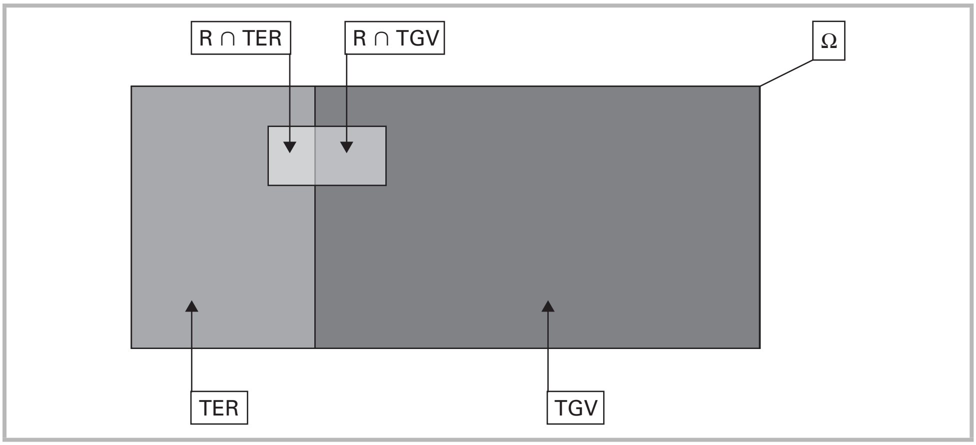
D'après l'énoncé, des conditions sont émises tout d'abord sur les TER puis sur les TGV. En effet : il est mentionné **« s'il s'agit d'un TER ».** La condition porte sur l'évènement TER, on déduit de l'énoncé la probabilité conditionnelle suivante : $P(R/TER) = 5\ \%$. De même on aura : $P(R/TGV) = 7\ \%$. On trouve alors avec le théorème de Bayes :

$P(R \cap TER) = P(R/TER).P(TER) = 0{,}05 \times 0{,}3 = 0{,}015$;

$P(R \cap TGV) = P(R/TGV).P(TGV) = 0{,}07 \times 0{,}7 = 0{,}049$. D'où :

$P(R) = P(R \cap TGV) + P(R \cap TER) = 0{,}015 + 0{,}049 = 0{,}064$.

Le graphique suivant illustre la situation :



2) Si un train est en retard, quelle est la probabilité que Bertrand ait pris un TER (arrondir à 10^{-2}) ?

On remarque ici que la condition porte sur l'évènement R (« **si un train est en retard** »). On en déduit que la question posée s'écrit : P(TER/R). Or :

$P(TER/R) = P(TER \cap R)/P(R) = 0{,}015/0{,}064 = 0{,}23$.

3) Sachant qu'un train est à l'heure quelle est la probabilité que ce soit un TER ?

La condition est ici « **sachant qu'un train est à l'heure** », donc on cherche :

$P(TER/\overline{R}) = P(TER \cap \overline{R})/P(\overline{R})$.

On cherche tout d'abord $P(TER \cap \overline{R})$:

$P(TER \cap \overline{R}) = P(\overline{R}/TER).P(TER) = [1-P(R/TER)].P(TER) = 0{,}95.0{,}3 = 0{,}285$.

D'où :

$P(TER/\overline{R}) = P(TER \cap \overline{R})/P(\overline{R}) = 0{,}285 / [1-P(R)] = 0{,}285/0{,}939 = 0{,}3035$

II. Une fois à son bureau, Bertrand décide de prendre un café. Il a le choix entre deux machines à café, M_1 et M_2. Soit A l'évènement : « la machine M_1 est en panne » et B l'évènement « la machine M_2 est en panne », tel que $P(A) = 0{,}2$; $P(B) = 0{,}3$ et $P(A \cap B) = 0{,}15$.

Calculer les probabilités suivantes :

- la machine M_1 fonctionne ;
- une machine au moins est en panne ;
- les deux machines fonctionnent.

<u>La machine M_1 fonctionne</u> : $P(\overline{A}) = 0{,}8$.

<u>Une machine au moins est en panne</u> : $P(A \cup B) = 0{,}35$.

<u>Les deux machines fonctionnent</u> : $P(\overline{A} \cap \overline{B}) = 0{,}65$.

CONSEILS PRATIQUES

1) Ne pas confondre « évènements indépendants » et « évènements disjoints » ou mutuellement exclusifs.
Les évènements disjoints ont une intersection vide :
$P(A \cap B) = \varnothing$. Par exemple A : « être un humain » et B : « être un animal ». En principe l'intersection est vide (si on ne tient pas compte des personnages fictifs comme ceux de twilight !). Ou encore A : « être un homme » et B : « être une femme ». Là encore l'intersection est vide.
Les évènements indépendants donnent $P(A \cap B) = P(A).P(B)$. Exemple :
A : « être une femme européenne » et B : « avoir plus de 20 ans ». Ces évènements sont indépendants car le fait d'être une femme européenne n'implique pas le fait d'avoir plus de 20 ans et inversement le fait d'avoir plus de 20 ans n'implique pas le fait d'être une femme de type européen.
2) Le **'ou'** en probabilité est non exclusif : l'évènement 'A ou B' implique que l'union $(A \cup B)$ contient l'intersection $(A \cap B)$. Il est parfois trompeur car l'emploi du 'ou bien' dans la vie de tous les jours est plutôt exclusif même si les évènements peuvent se réaliser séquentiellement :
A : « aller au restaurant » 'ou bien' B : « aller au cinéma ».
3) Attention $P(A/B) = 1 - P(\overline{A}/B)$ et non $1 - P(A/\overline{B})$.
4) Les algèbres de Boole contiennent toujours l'ensemble vide, l'univers et un nombre pair de parties.
5) Pourquoi dit-on qu'une algèbre de Boole est minimale ? La plus petite algèbre de Boole est celle qui contient les parties nécessaires pour définir une algèbre de Boole à partir des parties données dans l'énoncé.

2

VARIABLES ALÉATOIRES DISCRÈTES À UNE DIMENSION

SOMMAIRE

Ce chapitre, consacré aux variables aléatoires discrètes à une dimension, permet de faire le point sur :

- les caractéristiques de tendance centrale comme l'espérance mathématique, les quantiles, (médianes, quartiles, etc.) et le mode ;
- les caractéristiques de dispersion comme les moments centrés, non centrés et factoriels ;
- les propriétés des opérateurs espérance mathématique, variance et covariance ;
- les inégalités utilisées en statistique comme l'inégalité de Bienaymé-Tchebychev ;
- les fonctions génératrices et ses caractéristiques.

Le chapitre traite aussi des lois de probabilités élémentaires comme la loi de Bernoulli, la loi Binomiale, la loi de Poisson ou encore la loi hypergéométrique. Elles sont très utilisées en statistique car elles sont des lois de comptage. Elles sont facilement exploitables avec des enquêtes. Elles permettent par exemple de calculer des probabilités associées à un nombre de pannes, au nombre de clients achetant un bien de consommation spécifique, aux arrivées de voitures, au nombre de clients à un guichet, etc.

FICHE DE SYNTHÈSE

Variable aléatoire, fonction de répartition et loi de probabilité

➢ Variable aléatoire X

Il s'agit d'une application qui à un évènement (e) lui associe un réel :

$\{X(e)\}$: ensemble des images de X

$X(e) = x$: image de X, c'est une valeur discrète.

➢ Fonction de répartition

$F(x) = P[X \leqslant x], \ \ x \in \mathbb{Z} \ \mapsto F(x) \in [0\,;\,1]$ ($\mathbb{Z}$ étant l'ensemble des entiers naturels positifs et négatifs)

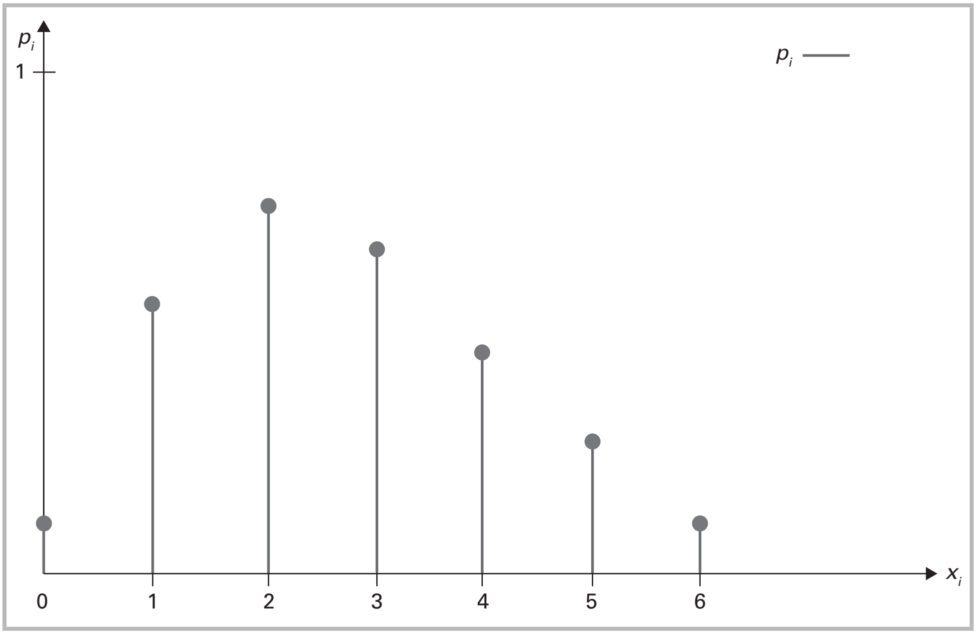
$0 \leqslant F(x) \leqslant 1, \ \ \forall x \in \mathbb{Z}$

$P(a < X \leqslant b) = F(b) - F(a)$

➢ Loi de probabilité

$\{x, p_x\}$ ou $\{x_i, p_i\}$

Représentation graphique : diagramme en bâton.



Caractéristiques de tendance centrale

➢ L'espérance mathématique

➔ *Définition :*

- $\mathrm{E}[\mathrm{X}] = \mathrm{E}(x)^* = \sum\limits_{x \in \chi} x p_x = \sum\limits_{i=1}^{n} x_i p_i$

➔ *Propriétés :*

- $\mathrm{E}[a] = a$ si $a =$ constante
- $\mathrm{E}[\mathrm{X}+\mathrm{Y}] = \mathrm{E}[\mathrm{X}] + \mathrm{E}[\mathrm{Y}]$
- $\mathrm{E}[\mathrm{X}-\mathrm{Y}] = \mathrm{E}[\mathrm{X}] - \mathrm{E}[\mathrm{Y}]$
- $\mathrm{E}[a\mathrm{X}] = a\mathrm{E}[\mathrm{X}]$
- E[XY] = E[X].E[Y] pour X, Y indépendants
- À partir d'une variable aléatoire X quelconque, avec E[X], on définit la variable centrée par :

  X – E[X] telle que $\mathrm{E}\left[\mathrm{X} - \mathrm{E}[\mathrm{X}]\right] = 0$.

* On écrira indifféremment les opérateurs espérance, variance, covariance entre crochets ou entre parenthèses.

- Moyenne de variables aléatoires :

  Soit $X_1, X_2, \ldots, X_n$, n variables aléatoires suivant une loi de probabilité quelconque et ayant chacune une espérance mathématique égale à m :

  $E\left[\bar{X}\right] = m.$

➢ Les quantiles

On appelle quantile d'ordre p ou fractile d'une variable aléatoire X dont la fonction de répartition est F, l'image x_p telle que $F(x_p) = p$.

★ **Médiane :** notée x_{me}

- $P(X \leqslant x_{\text{me}}) = \frac{1}{2} \Leftrightarrow F(x_{1/2}) = \frac{1}{2}$

★ **Quartiles** : notés Q_i

$$*\, Q_1 = x_{1/4} \textit{ tel que } F(x_{1/4}) = \frac{1}{4} \;\Leftrightarrow\; P(X \leqslant x_{1/4}) = 0{,}25$$

$$*\, Q_2 = x_{2/4} = x_{1/2} = x_{\text{me}} \textit{ tel que } F(Q_2) = \frac{1}{2} \;\Leftrightarrow\; P(X \leqslant Q_2) = 0{,}5$$

$$*\, Q_3 = x_{3/4} \textit{ tel que } F(x_{3/4}) = \frac{3}{4} \;\Leftrightarrow\; P(X \leqslant x_{3/4}) = 0{,}75$$

La valeur de l'intervalle interquartile $[Q_1\ ;\ Q_3]$ renseigne sur la dispersion de la distribution.

$[Q_1\ ;\ Q_3]$ est tel que $P(Q_1 \leqslant X \leqslant Q_3) = 0{,}5$

★ **Déciles** : notés d_i

Il existe 10 déciles, qui partagent la distribution en 10 parts égales à 1/10 :

- $F(d_i) = \frac{i}{10}$, $i = 1, \ldots, 10.$

➢ Le Mode

Dans le cas d'une variable aléatoire discrète : le mode noté x_{mo} est le (ou les) point(s) de l'ensemble de définition tel(s) que la probabilité attachée à ce(s) point(s) est supérieure aux deux probabilités adjacentes.

$$x_{\text{mo}} = x_k :\ P(X = x_k) > P(X = x_{k+1}) \text{ et } P(X = x_k) > P(X = x_{k-1})$$

Deux types de mode :

– mode local, existence d'un ou de plusieurs modes ;

– mode global, mode unique.

Caractéristiques de dispersion

➤ Moments non centrés

- $m_k(\mathrm{X}) = m_k = \mathrm{E}\left[\mathrm{X}^k\right]$

  $\mathrm{E}\left[\mathrm{X}^k\right]$ peut être exprimée pour différentes valeurs de k :

  $k = 0 \rightarrow m_0 = \mathrm{E}\left[\mathrm{X}^0\right] = 1$

  $k = 1 \rightarrow m_1 = \mathrm{E}[\mathrm{X}]$

  $k = 2 \rightarrow m_2 = \mathrm{E}\left[\mathrm{X}^2\right]$

- $m_k = \sum\limits_{i=1}^{n} x_i^k p_i = \sum\limits_{x \in X} x^k p_x$

➤ Moments centrés

- $\mu_k[\mathrm{X}] = \mu_k = \mathrm{E}[\mathrm{X} - \mathrm{E}[\mathrm{X}]]^k$

- $\mu_k = \sum\limits_{i=1}^{n} (x_i - \mathrm{E}(\mathrm{X}))^k p_i = \sum\limits_{i=1}^{n} (x_i - m)^k p_i$

  $k = 0 \rightarrow \mu_0 = \mathrm{E}[\mathrm{X} - \mathrm{E}(\mathrm{X})]^0 = 1$

  $k = 1 \rightarrow \mu_1 = \mathrm{E}[\mathrm{X} - \mathrm{E}(\mathrm{X})]^1 = 0$ (variable centrée : $\mathrm{X} - \mathrm{E}[\mathrm{X}]$)

  $k = 2 \rightarrow \mu_2 = \mathrm{E}[\mathrm{X} - \mathrm{E}(\mathrm{X})]^2 = \mathrm{E}[\mathrm{X} - m]^2$

➤ Relations entre les moments centrés et non centrés :

- $\mu_2 = m_2 - m_1^2$
- $\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$

➤ Moments factoriels

Soit X une variable aléatoire et k un entier positif, on appelle moment factoriel d'ordre k, l'expression :

- $\mu_{[k]}[\mathrm{X}] = \mu_{[k]} = \mathrm{E}\big[\underbrace{\mathrm{X}(\mathrm{X}-1)(\mathrm{X}-2)\ldots(\mathrm{X}-k+1)}_{k\ facteurs}\big]$

  $\mu_{[k]} = \sum\limits_{i=1}^{n} x_i(x_i - 1)\ldots(x_i - k + 1)p_i$

  $\mu_{[1]} = \mathrm{E}[\mathrm{X}] = m_1$

  $\mu_{[2]} = \mathrm{E}[\mathrm{X}(\mathrm{X}-1)] = m_2 - m_1$

  $\mu_{[3]} = \mathrm{E}[\mathrm{X}(\mathrm{X}-1)(\mathrm{X}-2)] = m_3 - 3m_2 + 2m_1$

Ainsi $\mu_{[k]}$ s'exprime en fonction des différents moments non centrés $m_k, m_{k-1}, \ldots, m_1$.

- À l'inverse les moments non centrés s'expriment en fonction de $\mu_{[k]}, \mu_{[k-1]}, ..., \mu_{[1]}$
  $m_1 = \mu_{[1]}$
  $m_2 = \mu_{[2]} + \mu_{[1]}$
  $m_3 = \mu_{[3]} + 3\mu_{[2]} + \mu_{[1]}$

➢ Variance : propriétés

- $V(X) = \mu_2 = E[X - E(X)]^2 = \sigma^2 = m_2 - m_1^2$
  Elle mesure la dispersion (en unités ^{2}) de la distribution autour de la moyenne.
- $V(aX + b) = a^2V(X)$ ($V(b) = 0$).
- Si $Y = \dfrac{X - m}{\sigma}$: $E(Y) = 0$ et $V(Y) = 1$ (Y : variable centrée réduite)
- $V(X \pm Y) = V(X) + V(Y) \pm 2\,\text{cov}(X,Y)$, ou 'cov' est l'opérateur covariance.
- $V(\bar{X}) = \dfrac{\sigma^2}{n}$

➢ Covariance

Elle mesure le sens de fluctuation (variation) entre deux variables :

- $\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$
- $\text{cov}(X,Y) = E[YX] - E[X]E[Y]$
- $\text{cov}(aX, bY) = ab\,\text{cov}(X,Y)$
- $\text{cov}(X_1 + X_2, X_1) = \text{cov}(X_1, X_1) + \text{cov}(X_2, X_1) = V(X_1) + \text{cov}(X_2, X_1)$
- Si X et Y sont indépendantes $\text{cov}(X,Y) = 0$ donc
  $V(X + Y) = V(X) + V(Y)$; $V(X - Y) = V(X) + V(Y)$

Inégalité de Bienaymé-Tchebychev

Pour tout $\varepsilon > 0$:

$P[|X - E(X)| \geqslant \varepsilon] \leqslant \dfrac{V(X)}{\varepsilon^2}$ pour déterminer une borne supérieure

$P[|X - E(X)| < \varepsilon] \geqslant 1 - \dfrac{V(X)}{\varepsilon^2}$ pour déterminer une borne inférieure

Fonction génératrice des moments

Soit X une variable aléatoire discrète :

- $g_X(U) = E[U^X] = \sum_{x \in \chi} u^x p_x = \sum_{i=1}^{n} u^{x_i} p_i$

- $g'_X(U) = \sum_{x\in\chi} p_x x u^{x-1} \leqslant \sum_{x\in\chi} p_x x = E[X]$

  $g''_X(U) = \sum_{x\in\chi} p_x x(x-1)u^{x-2} \leqslant \sum_{x\in\chi} p_x x(x-1) = \mu_{[2]}$

  —

  $g_X^{(k)}(U) = \sum_{x\in\chi} p_x x(x-1)...(x-k+1)u^{x-k} \leqslant \mu_{[k]}$

  $\boxed{g_X^{(k)}(1) = \mu_{[k]}}$

- Obtention des moments factoriels soit $U = V + 1$ avec U voisin de 1, V voisin de 0 :

  $$g_X(1+V) = 1 + \mu_{[1]}\frac{V}{1!} + \mu_{[2]}\frac{V^2}{2!} + \cdots + \mu_{[K]}\frac{V^k}{k!} + \cdots$$

Fonction caractéristique des moments

- $\varphi_X(t) = E\left[e^{itX}\right]$ avec i partie imaginaire d'un nombre complexe : $i = \sqrt{-1}$; $i^2 = -1$.
- La fonction caractéristique est une généralisation de la fonction génératrice :

  $\varphi_X(t) = E\left[\left(e^{it}\right)^X\right] = g_X(e^{it})$
- Si X et Y sont indépendantes :

  $\varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t)$
- Obtention des moments non centrés (par développement limité) :

$$E\left[e^{itX}\right] = E\left[1 + \frac{itX}{1!} + \frac{(itX)^2}{2!} + \cdots + \frac{(itX)^k}{k!} + \cdots\right]$$

$$= 1 + \frac{it\ E(X)}{1!} + \frac{(it)^2 E(X^2)}{2!} + \cdots + \frac{(it)^k E(X^k)}{k!} + \cdots$$

$$\varphi_X(t) = 1 + m_1\frac{it}{1!} + m_2\frac{(it)^2}{2!} + \cdots + m_k\frac{(it)^k}{k!} + \cdots$$

Loi de BERNOULLI

➢ Notation

X ~ $B(1\ ;\ p)$ avec p la probabilité de réalisation de l'évènement étudié lors du tirage.

➢ Loi de probabilité

Ensemble des images : $X(\Omega) = \{0\ ;\ 1\}$

Probabilités : $P(X = x) = p_x = p^x(1-p)^{1-x}$

x_i	0	1
p_i	$q = 1 - p$	p

➢ Caractéristiques

$E(X) = p$

$V(X) = pq$

Fonction génératrice : $g_X(U) = up + q$

Fonction caractéristique : $\varphi_X(t) = pe^{it} + q$

Loi BINOMIALE

➢ Notation

$X \sim B(n\,;p)$ avec n le nombre d'épreuves (ou tirages) identiques indépendantes dans une population où on a 2 caractères et p la probabilité de réalisation de l'évènement étudié lors de chacun des tirages indépendants.

➢ Loi de probabilité

Ensemble des images : $X(\Omega) = \{0\,;1\,;2\,;\dots;n\}$

Probabilités : $P(X = x) = p_x = C_n^x p^x q^{n-x}$

➢ Caractéristiques

$E(X) = np$

$V(X) = npq$

Fonction génératrice : $g_X(U) = (pu + q)^n$

Fonction caractéristique : $\varphi_X(t) = (pe^{it} + q)^n$

Mode x_{mo} : $np - q < x_{mo} < np + p$

➢ Théorèmes

★ $B(n\,; p) \equiv \sum_{i=1}^{n} B_i(1;\ p)$ indépendantes

★ Si $X_i \sim B(n_i\,; p)$ $\forall i \neq j$ X_i, X_j indépendantes, alors $\sum_{i=1}^{n} X_i \sim B(\sum_{i=1}^{n} n_i\,; p)$.

Loi de POISSON

➢ Notation

X ~ $P(\lambda)$, $\lambda > 0$

➢ Loi de probabilité

Ensemble des images : $X(\Omega) = \{0 ; 1 ; 2 ; \ldots ; n\}$

Probabilités : $p_x = e^{-\lambda} \lambda^x / x\,!$

➢ Caractéristiques

$E(X) = V(X) = \lambda$

Fonction génératrice : $g_X(U) = e^{\lambda(u-1)}$

Fonction caractéristique : $\varphi_X(t) = e^{\lambda\left(e^{it}-1\right)}$

Mode x_{mo} : $\lambda - 1 < x_{mo} < \lambda$

Moment factoriel d'ordre k : $\mu_{[k]} = \lambda^k$

➢ Théorèmes

★ Si $n > 50$ et $p < 10\ \%$ alors X~$B(n ; p)$ converge vers une loi de Poisson : $X \overset{C}{\approx} P(\lambda = np)$

★ Si $X_i \sim P(\lambda_i)$, $\forall i \neq j$ X_i, X_j indépendantes, alors $\sum_{i=1}^{n} X_i \sim P(\sum_{i=1}^{n} \lambda_i)$.

Loi Hypergéométrique

➢ Notation

X ~ $H(N ; n ; p)$ avec N le nombre total d'individus dans l'échantillon, n le nombre d'individus relatifs à la variable aléatoire X, p la probabilité d'un individu interrogé relative à la variable aléatoire X.

➢ Loi de probabilité

Ensemble des images : $X(\Omega) = \{0 ; 1 ; \ldots ; a\}$

Probabilités : $P(X = k) = \dfrac{C_n^k C_{N-n}^{a-k}}{C_N^a}$; a : nombre de tirages

➢ Caractéristiques

$E(X) = np$; $V(X) = npq\,(N-n)/(N-1)$

EXERCICE 1 : UTILISATION DES PROPRIÉTÉS DE L'OPÉRATEUR ESPÉRANCE MATHÉMATIQUE ET VARIANCE

Soient deux variables aléatoires indépendantes X et Y sur un ensemble fini Ω.

$E(X) = 5$; $V(X) = 25$

$E(Y) = 2$; $V(Y) = 20$

On définit les variables aléatoires suivantes :

$A = X - 5Y$, $B = 3X - 3Y$.

Calculez :

1 – E(A), E(B), V(A) et V(B)

2 – cov(A, B) de deux façons différentes

3 – V(B + Y)

4 – V(B) + V(Y)

5 – cov(B, Y)

6 – Le coefficient de corrélation linéaire entre A et B.

1 – E(A), E(B), V(A) et V(B)

- $\underline{E(A)} = E(X - 5Y)$
- L'opérateur est linéaire d'où :
- $E(A) = E(X - 5Y) = E(X) + E(-5Y) = E(X) - 5E(Y) = 5 - 5 \times 2 = -5.$
- $\underline{E(B)} = E(3X - 3Y) = E(3X) + E(-3Y) = 3E(X) - 3E(Y) = 3 \times 5 - 3 \times 2 = 9.$
- $\underline{V(A)} = E(A^2) - [E(A)]^2 = E[(X - 5Y)^2] - [E(A)]^2 = E[X^2 + 25Y^2 - 10XY] - [-5]^2$
  $= E(X^2) + 25E(Y^2) - 10E(XY) - 25$

→ À cette étape, des calculs intermédiaires sont nécessaires : $E(X^2)$, $E(Y^2)$, $E(XY)$.

- Pour trouver $E(X^2)$ il n'est pas possible d'appliquer la formule du moment non centré d'ordre 2 à savoir : $E(X^2) = \sum_i p_i x_i^2$ avec p_i la probabilité rattachée à la réalisation de l'évènement $X = x_i$. La distribution des probabilités de X est inconnue. Il n'est donc pas possible de passer par la formule précédente. On peut cependant retrouver $E(X^2)$ avec la variance de X :

  $V(X) = E(X^2) - E(X)^2$ d'où

  $E(X^2) = V(X) + E(X)^2 = 25 + 5^2 = 50.$
- De même pour $E(Y^2)$:

  $E(Y^2)= V(Y) + E(Y)^2 = 20 + 2^2 = 24.$

- Enfin il nous reste à trouver E(XY). Là encore, il n'est pas possible de passer par la formule directe $E(XY) = \sum_i \sum_j p_{ij} x_i y_j$. Il faudrait en effet connaître pour cela la distribution de probabilité du couple (X, Y) [ceci sera analysé en détail au Chapitre 4]. D'après l'énoncé les variables aléatoires X et Y sont indépendantes, donc :

E(XY) = E(X).E(Y). Ceci provient du fait que lorsque les variables sont indépendantes la variation de l'une n'implique aucune variation de l'autre. Cela se traduit par une covariance nulle :

cov(X, Y) = E(XY) – E(X).E(Y). Si cov(X, Y) = 0 alors → E(XY) = E(X).E(Y). D'où : E(XY) = E(X).E(Y) = 5 × 2 = 10.

On peut donc revenir au calcul de V(A) :

$$V(A) = E(X^2) + 25E(Y^2) - 10E(XY) - 25$$
$$= 50 + 25 \times 24 - 10 \times 10 - 25 = 525.$$

- <u>V(B)</u> $= E(B^2) - [E(B)]^2$

$$= E[(3X - 3Y)^2] - [E(B)]^2$$
$$= 9E(X^2) + 9E(Y^2) - 18E(XY) - [E(B)]^2$$
$$= 9 \times 50 + 9 \times 24 - 18 \times 10 - 9^2 = 405.$$

2 – cov(A, B) de deux façons différentes

Premièrement utilisons la formule de la covariance :

- <u>cov(A, B)</u> $= E(AB) - E(A) \times E(B) = E[(X - 5Y) \times (3X - 3Y)] - (-5) \times 9$

$$= E[3X^2 - 3XY - 15XY + 15Y^2)] + 45$$
$$= 3E(X^2) - 3E(XY) - 15E(XY) + 15E(Y^2) + 45$$
$$= 3 \times 50 - 3 \times 10 - 15 \times 10 + 15 \times 24 + 45$$
$$= 375.$$

Deuxièmement, utilisons la formule la variance :

- V(A + B) = V(A) + V(B) + 2cov(A, B) . On en déduit alors cov(A, B) :

  cov(A, B) = [V(A + B) – V(A) – V(B)] / 2. Un calcul intermédiaire est ici nécessaire puisque V(A + B) est inconnue :

- V(A + B) = V[(X – 5Y) + (3X – 3Y)] = V[4X – 8Y]

$$= 4^2V(X) + 8^2V(Y) - cov(4X, -8Y).$$

Puisque X et Y sont indépendantes alors les variables U = 4X et V = –8Y le sont aussi, donc cov(U, V) = cov(4X, –8Y) = 0. On en déduit alors :

$V(A + B) = 4^2V(X) + 8^2V(Y) – 0 = 16 \times 25 + 64 \times 20 = 1680$. On peut maintenant revenir au calcul de cov(A, B) :

- $cov(A, B) = [V(A + B) – V(A) – V(B)] / 2 = [1680 – 525 – 405] / 2 = 375$.

<u>Remarque</u> : La covariance est positive : les variables aléatoires A et B fluctuent (varient) dans le même sens.

3 – V(B + Y). Nous pouvons là aussi utiliser deux méthodes :

- $V(B + Y) = V(B) + V(Y) + 2cov(B, Y)$. Calculons au préalable :
- $cov(B, Y) = E(BY) – E(B) \times E(Y)$
- $E(BY) = E[(3X – 3Y) \times Y] = E[3XY – 3Y^2]$

  $= 3E(XY) – 3E(Y^2) = 3 \times 10 – 3 \times 24 = – 42$. D'où

  $cov(B, Y) = E(BY) – E(B).E(Y) = – 42 – 9 \times 2 = – 60$. Donc :

  $V(B + Y) = V(B) + V(Y) + 2cov(B, Y) = 405 + 20 + 2 \times (–60) = 305$.
- $V(B + Y) = V[(3X – 3Y) + Y] = V[3X –2Y] = 3^2V(X) + 2^2V(Y) – 2cov(3X, –2Y)$.

Comme X et Y sont indépendantes : on a $cov(X, Y) = cov(3X, –2Y) = 0$, d'où

$V(B + Y) = 3^2V(X) + 2^2V(Y) – 0 = 9 \times 25 + 4 \times 20 = 305$.

4 – V(B) + V(Y).

- $V(B) + V(Y) = 405 + 20 = 425$
- Ou en utilisant la question précédente :

  $V(B) + V(Y) = V(B + Y) – 2cov(B, Y) = 305 – 2(–60) = 305 + 120 = 425$.

5 – cov(B, Y).

- $cov(B, Y) = E(BY) – E(B).E(Y) = – 42 – 9 \times 2 = – 60$ (déjà calculé : *Cf.* 3).
- Ou en utilisant encore la question précédente :

  $V(B + Y) = V(B) + V(Y) + 2cov(B, Y)$. Donc :

  $cov(B, Y) = [V(B + Y) – V(B) – V(Y)] / 2 = [305 – 405 – 20] /2 = – 60$.
- Enfin en remarquant que :

  $cov(B, Y) = cov(3X – 3Y, Y) = 3cov(X, Y) – 3cov(Y, Y) = –3V(Y) – 0 = – 60$.

6 – Le coefficient de corrélation linéaire entre A et B.

Par définition $r_{A, B} = \dfrac{cov(A, B)}{\sqrt{V(A) \times V(B)}} = \dfrac{375}{\sqrt{525 \times 405}} = 0{,}8124$.

Remarque : le coefficient de corrélation linéaire r $\in$ [–1 ; 1]. La valeur 0,8124 indique que les variables aléatoires A et B sont fortement et positivement corrélées de manière linéaire.

EXERCICE 2 : CALCUL DE PROBABILITÉ ET DE LA VARIANCE D'UNE VARIABLE ALÉATOIRE

Le nombre de pauses prises par un travailleur lors d'une journée habituelle de travail est une variable aléatoire X qui suit la loi de probabilité suivante :

x_i	0	1	2	3	4	5
p_i	0,1	0,15	0,15	0,3	0,2	0,1

1) Calculer la variance de X.

2) Quelle est la probabilité que le travailleur fasse au maximum 2 pauses ?

1) Calculer la variance de X.

$V(X) = E(X^2) - E(X)^2$

- $E(X) = \sum_{i=1}^{5} p_i x_i = 0,1 \times 0 + 0,15 \times 1 + 0,15 \times 2 + 0,3 \times 3 + 0,2 \times 4 + 0,1 \times 5 = 2,65.$
- $E(X^2) = \sum_{i=1}^{5} p_i x_i^2 = 0,1 \times 0^2 + 0,15 \times 1^2 + 0,15 \times 2^2 + 0,3 \times 3^2 + 0,2 \times 4^2 + 0,1 \times 5^2$
  $= 9,15.$

D'où $V(X) = 9,15 - 2,65^2 = 2,1275.$

2) Quelle est la probabilité que le travailleur fasse au maximum 2 pauses ?

$P(X \leqslant 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,1 + 0,15 + 0,15 = 0,4.$

EXERCICE 3 : CARACTÉRISTIQUES DE TENDANCES CENTRALES, DE DISPERSION, ET CALCULS DE PROBABILITÉS.

Le nombre de sangliers observés par semaine par un chasseur du Parc du Cup est une variable aléatoire X dont la loi de probabilité est la suivante :

x_i	0	1	2	3	4	5	6
p_i	0,1	0,2	0,2	0,25	0,1	0,1	0,05

1) Calculer E(X).

2) Calculer V(X).

3) Déterminer la fonction de répartition : faire son graphique.

4) Déterminer la médiane.

5) Calculer la probabilité que le chasseur observe au moins 4 sangliers.

6) Calculer $P(2 \leqslant X < 5)$.

7) Calculer la probabilité que le chasseur observe au moins 4 sangliers par semaine sachant que le lundi un sanglier a été observé.

8) Calculer le moment centré d'ordre 3.

1) Calculer E(X).

$$\begin{aligned}E(X) &= \sum_{i=1}^{7} p_i x_i\\ &= 0{,}1 \times 0 + 0{,}2 \times 1 + 0{,}2 \times 2 + 0{,}25 \times 3 + 0{,}1 \times 4 + 0{,}1 \times 5 + 0{,}05 \times 6\\ &= 2{,}55.\end{aligned}$$

2) Calculer V(X).

$$\begin{aligned}E(X^2) &= \sum_{i=1}^{7} p_i x_i^2\\ &= 0{,}1 \times 0^2 + 0{,}2 \times 1^2 + 0{,}2 \times 2^2 + 0{,}25 \times 3^2 + 0{,}1 \times 4^2 + 0{,}1 \times 5^2 + 0{,}05 \times 6^2\\ &= 9{,}15.\end{aligned}$$

$$V(X) = E(X^2) - E(X)^2 = 9{,}15 - 2{,}55^2 = 2{,}6475.$$

3) Déterminer la fonction de répartition : faire son graphique.

La fonction de répartition permet de cumuler les probabilités individuelles :

$F(x) = P(X \leqslant x)$. Il s'agit d'une fonction dont la représentation graphique est en escalier. La première valeur est 0, la plus forte valeur est 1. Pour la première probabilité égale à 0, le premier évènement ($x_1 = 0$) ne s'est pas réalisé :

Si $x < x_1$: $F(x) = 0$.

Si le premier évènement $[x_1 = 0]$ s'est réalisé mais pas le second $[x_2 = 1]$, on a :

Si $x \in [x_1 ; x_2[$: $F(x) = P(X \leqslant x_1) = P(X = x_1) = 0{,}1$.

Si le deuxième évènement $[x_2 = 1]$ s'est réalisé mais pas le troisième $[x_3 = 2]$, on a :

Si $x \in [x_2\,; x_3[$: $F(x) = P(X \leqslant x_2) = P(X = x_1) + P(X = x_2) = 0{,}1 + 0{,}2 = 0{,}3$. On remarque à ce stade que lorsque le deuxième évènement s'est réalisé les évènements qui le précèdent sont comptabilisés (puisque par définition $F(x)$ cumule les probabilités individuelles).

Si $x \in [x_3\,; x_4[$: $F(x) = P(X \leqslant x_3) = P(X = x_1) + P(X = x_2) + P(X = x_3) = 0{,}1 + 0{,}2 + 0{,}2 = 0{,}5$.

Si $x \in [x_4\,; x_5[$: $F(x) = P(X \leqslant x_4) = P(X = x_1) + \cdots + P(X = x_4) = 0{,}1 + \cdots + 0{,}25 = 0{,}75$.

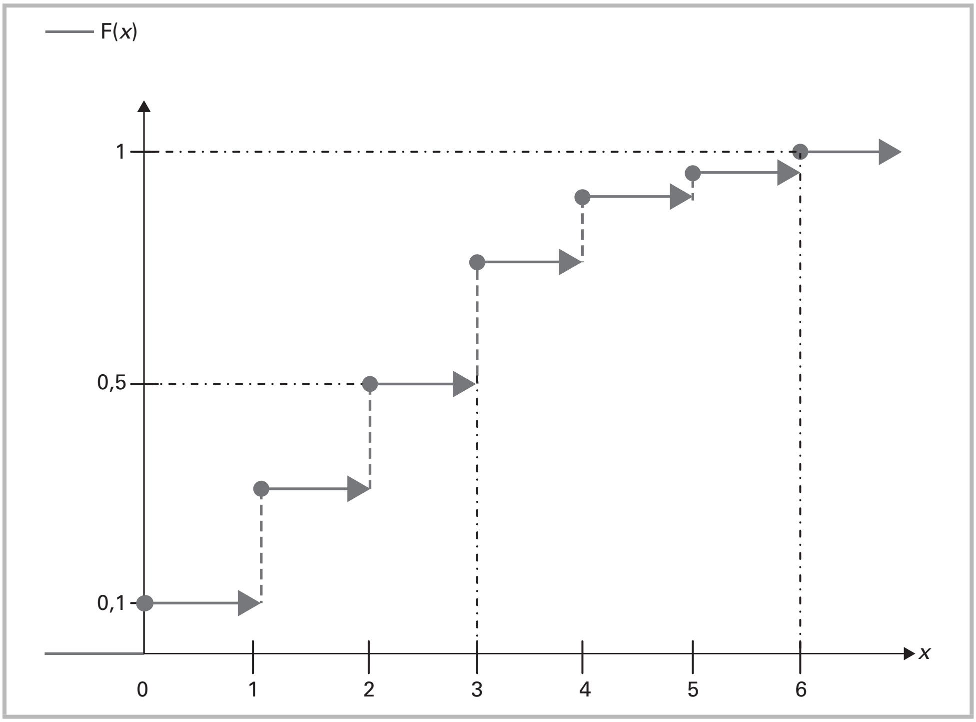
Si $x \in [x_5\,; x_6[$: $F(x) = P(X \leqslant x_5) = P(X = x_1) + \cdots + P(X = x_5) = 0{,}1 + \cdots + 0{,}1 = 0{,}85$.

Si $x \in [x_6\,; x_7[$: $F(x) = P(X \leqslant x_6) = P(X = x_1) + \cdots + P(X = x_6) = 0{,}1 + \cdots + 0{,}1 = 0{,}95$.

Enfin, pour le dernier évènement, on aura :

Si $x \geqslant x_7$: $F(x) = P(X = x_1) + \cdots + P(X = x_7) = 0{,}1 + \cdots + 0{,}05 = 1$.

Concernant la représentation graphique, on notera que la fonction de répartition est continue à gauche (vers moins l'infini pour la probabilité nulle). La continuité est représentée par un point. Par contre elle est ouverte à droite (représentation par la pointe de la flèche qui indique que l'image associée à l'évènement n'est pas atteinte). Par exemple pour $x = 6$, la valeur de $F(x)$ est 1 (on tombe en effet sur le point qui indique que l'évènement 6 est atteint. Au contraire, on tombe sur la pointe de la flèche en $F(x) = 0{,}95$ indiquant que l'image de l'évènement 5 a déjà été prise en compte.



4) Déterminer la médiane

La médiane peut être trouvée graphiquement. Il s'agit de l'image qui partage la distribution de probabilités en deux parties égales. Autrement dit, nous cherchons l'image x_{me} telle que : $F(x_{me}) = 0{,}5$.

Dans notre cas, la valeur 0,5 correspond à $x \in [x_3 ; x_4[= [2 ; 3[$: on dit qu'il s'agit d'un intervalle médian.

Remarque : Lorsque la valeur $F(x_{me}) = 0{,}5$ tombe sur une contre marche, on descend le long de celle-ci pour trouver l'image x_{me} correspondant à la médiane.

5) Calculer la probabilité que le chasseur observe au moins 4 sangliers

$$P(X \geqslant 4) = 0{,}1 + 0{,}1 + 0{,}05 = 0{,}25.$$

6) Calculer $P(2 \leqslant X < 5)$

Il est possible de procéder de plusieurs manières :

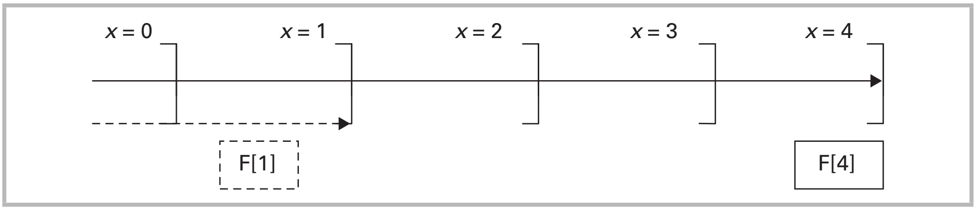
- $P(2 \leqslant X < 5) = P(X = 2) + P(X = 3) + P(X = 4) = 0{,}2 + 0{,}25 + 0{,}1 = 0{,}55.$
- En passant par la fonction de répartition, on a :
  $P(2 \leqslant X < 5) = P(2 \leqslant X \leqslant 4) = P(1 < X \leqslant 4) = F(4) - F(1).$

Ceci sera particulièrement intéressant lorsque $F(x) = P(X \leqslant x)$ pourra se lire directement sur une table statistique. Ce qui n'est pas le cas ici. Il faut calculer :

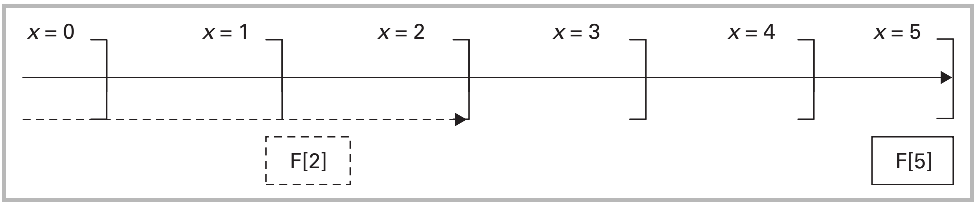
- $F(4) = P(X \leqslant 4) = 0{,}1 + 0{,}2 + 0{,}2 + 0{,}25 + 0{,}1 = 0{,}85.$
- $F(1) = P(X \leqslant 1) = 0{,}1 + 0{,}2 = 0{,}3$. D'où :
- $P(2 \leqslant X < 5) = F(4) - F(1) = 0{,}85 - 0{,}3 = 0{,}55.$

Remarque : on a toujours $P(a < X \leqslant b) = F(b) - F(a)$.

Attention ! $P(a \leqslant X < b) \neq F(b) - F(a)$.



En observant la différence entre les deux flèches (la pleine qui donne F(4) et celle en pointillé qui donne F(1)) : on constate que sont prises en compte les images $x = 2$, $x = 3$, et $x = 4$. On répond donc bien à la question posée. En revanche, en faisant $P(2 \leqslant X < 5) = F(5) - F(2)$, on constate facilement l'erreur commise :



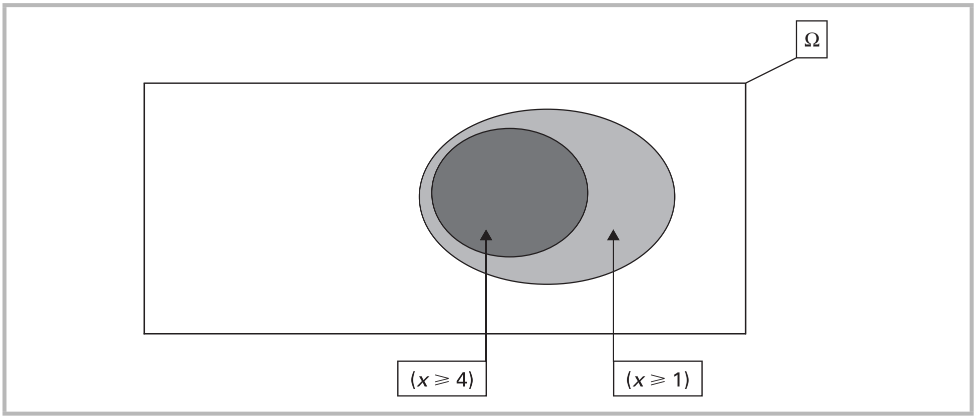
La différence entre les deux flèches montre que sont prises en compte les images $x = 3$, $x = 4$ et $x = 5$, alors qu'il était demandé $P(2 \leqslant X < 5)$, c'est-à-dire de prendre en compte les images $x = 2$, $x = 3$, et $x = 4$.

7) Calculer la probabilité que le chasseur observe au moins 4 sangliers par semaine sachant que le lundi un sanglier a été observé

- La variable aléatoire X est définie sur la semaine. L'évènement « Si le lundi 1 sanglier a été observé » doit être ramené à un évènement relatif à la semaine. Si le lundi 1 sanglier a été observé, alors cela implique que pour cette semaine-là, 1 sanglier au moins sera observé : $P(X \geqslant 1)$. Étant donné que la condition porte sur l'évènement $(X \geqslant 1)$: le théorème de Bayes donne :

$$P[(X \geqslant 4)/(X \geqslant 1)] = \frac{P\left[(X \geqslant 4) \cap (X \geqslant 1)\right]}{P(X \geqslant 1)}.$$

La difficulté est ici de trouver l'intersection $P\left[(X \geqslant 4) \cap (X \geqslant 1)\right]$. Remarquons que l'ensemble $(X \geqslant 1)$ est plus grand que $(X \geqslant 4)$, car il contient plus d'éléments. En effet, le chasseur peut observer 6 sangliers par semaine. Or, pour $(X \geqslant 4)$ il n'y a que trois éléments (4 ; 5 et 6) [ceux-ci étant inclus dans $(X \geqslant 1)$]. L'évènement $(X \geqslant 4)$ est donc inclus dans l'évènement $(X \geqslant 1)$:



On en déduit alors que $(X \geqslant 4) \cap (X \geqslant 1) = (X \geqslant 4)$. Par conséquent :

$$P[(X \geqslant 4)/(X \geqslant 1)] = \frac{P\left[(X \geqslant 4) \cap (X \geqslant 1)\right]}{P(X \geqslant 1)} = \frac{P(X \geqslant 4)}{P(X \geqslant 1)} = 0{,}25/0{,}9 = 0{,}2778.$$

8) Calculer le moment centré d'ordre 3

Le moment centré d'ordre 3 nous indique si la distribution de probabilités est centrée en sa moyenne (le coefficient est de 0), s'il y a plus de probabilité à droite de l'espérance (le coefficient est positif), ou encore si les probabilités sont concentrées à gauche de l'espérance (le coefficient est négatif). On le mesure de la manière suivante :

$$\mu_3 = E[X - E(X)]^3.$$

En utilisant l'identité remarquable $(a-b)^3 = a^3 - b^3 - 3a^2b + 3ab^2$, on a :

$$\mu_3 = E[X^3 - E(X)^3 - 3X^2\, E(X) + 3XE(X)^2].$$

Remarquons que l'opérateur espérance appliqué à lui-même donne l'espérance :

$$\mu_3 = E[X^3] - E(X)^3 - 3E[X^2]\, E(X) + 3E[X]E(X)^2.$$

Les moments non centrés d'ordre k, notés m_k sont donnés par $m_k = E[X^k]$, d'où

$$\mu_3 = m_3 - m_1^{\,3} - 3m_2\, m_1 + 3m_1m_1^{\,2} = m_3 - 3m_2\, m_1 + 2\, m_1^{\,3}.$$

On calcule les moments non centrés :

$$\begin{aligned} m_2 = E(X^2) &= \sum_{i=1}^{7} p_i x_i^2 \\ &= 0{,}1 \times 0^2 + 0{,}2 \times 1^2 + 0{,}2 \times 2^2 + 0{,}25 \times 3^2 + 0{,}1 \times 4^2 + 0{,}1 \times 5^2 + 0{,}05 \times 6^2 \\ &= 9{,}15. \end{aligned}$$

$$\begin{aligned} m_3 = E(X^3) &= \sum_{i=1}^{7} p_i x_i^3 \\ &= 0{,}1 \times 0^3 + 0{,}2 \times 1^3 + 0{,}2 \times 2^3 + 0{,}25 \times 3^3 + 0{,}1 \times 4^3 + 0{,}1 \times 5^3 + 0{,}05 \times 6^3 \\ &= 38{,}25. \end{aligned}$$

On en déduit :

$$\mu_3 = m_3 - 3m_2\, m_1 + 2\, m_1^{\,3} = 38{,}25 - 3 \times 9{,}15 \times 2{,}55 + 2 \times 2{,}55^3 = 1{,}41525.$$

Le coefficient est positif, il y a donc plus de probabilités à droite de l'espérance (autrement dit pour les images dont la valeur est supérieure à la valeur de l'espérance qui est de 2,55). On peut le vérifier sur le graphique de la fonction de répartition.

<u>Remarques</u> :

- Il est possible d'appliquer la formule $\mu_3 = E[X - E(X)]^3$ sans passer par son développement :

$$\begin{aligned} \mu_3 &= E[X - E(X)]^3 \\ &= \sum_{i=1}^{7} p_i\left(x_i - 2{,}55\right)^3 = 0{,}1\left(0 - 2{,}55\right)^3 + 0{,}2\left(1 - 2{,}55\right)^3 + \cdots + 0{,}05\left(6 - 2{,}55\right)^3. \end{aligned}$$

- Dans certains cas, le calcul n'est pas nécessaire. Imaginons la distribution suivante :

x_i	1	2	3	4	5	6
p_i	1/12	1/6	1/4	1/4	1/6	1/12

L'espérance est $E(X) = m_1 = 3{,}5$. On constate que de part et d'autre de cette valeur, les probabilités sont symétriquement distribuées. Il n'y a pas d'asymétrie : $\mu_3 = 0$.

EXERCICE 4 : LOI BINOMIALE ET SES PROPRIÉTÉS

On réalise une enquête sur un échantillon E d'une population. Afin de savoir si les individus sont aptes à répondre aux questions, le sondeur demande aux individus interrogés s'ils sont majeurs. Seulement 15 % des personnes souhaitent ne pas révéler leur âge.

1) Le sondeur interroge 25 personnes de l'échantillon E. Après avoir identifié la loi, déterminer la probabilité pour que les 25 personnes révèlent leur âge.

2) Déterminer le mode, la fonction caractéristique, la fonction génératrice, l'espérance mathématique, la variance.

1) Identifier la loi.

Contrairement à l'exercice précédent, où la distribution de probabilité est donnée (le tableau qui reporte les probabilités aux évènements), il est ici nécessaire d'identifier la loi de probabilité afin d'en déduire toutes les probabilités possibles. Il faut pour cela spécifier la variable aléatoire.

- ▫ **Dans un premier temps, déterminons la variable aléatoire**. L'énoncé indique que l'on souhaite calculer des probabilités **« pour que les 25 personnes révèlent leur âge ».** On constate que les probabilités sont liées à un nombre : ici c'est le nombre de personnes révélant leur âge. On est dans le cadre des lois de comptages : celles-ci sont habituellement des lois *Binomiales*, des lois de *Poisson*, ou des lois *hypergéométriques*. Une autre difficulté est de bien distinguer les types de lois. C'est ce que nous nous efforcerons de faire dans les exercices qui suivent. Rappelons que ces lois sont discrètes : autrement dit le comptage concerne un nombre entier naturel positif : 0, 1, 2, 3, …. On pourra avoir par exemple un nombre de pannes d'un véhicule, le nombre de Tee-shirts vendus, etc.
- o Soit **X la variable aléatoire** : « le nombre de personnes ne souhaitant pas révéler leur âge ». On peut être gêné par ce choix. On aurait pu choisir une variable notée Y telle que :
- o Soit **Y la variable aléatoire** : « le nombre de personnes souhaitant révéler leur âge ».

TABLE DES MATIÈRES

Ouvertures Économiques

Abdelmalki Lahsen, Sandretto René, *Politiques commerciales des grandes puissances*. La tentation néoprotectionniste.

Barbier-Gauchard Amélie, *Intégration budgétaire européenne*. Enjeux et perspectives pour les finances publiques européennes.

Beffy Pierre-Olivier, *Initiation à l'économie*. Compléments en ligne : QCM - exercices.

Burgenmeier Beat, *Politiques économiques du développement durable*.

Cayatte Jean-Louis, *Microéconomie de l'incertitude*.

Dumas André, *Économie mondiale*. Les règles du jeu commercial, monétaire et financier.

Faucheux Sylvie, Hue Christelle, Nicolaï Isabelle, *T.I.C. et développement durable*. Les conditions du succès.

Lupton Sylvie, *Économie des déchets*. Une approche institutionnaliste.

Mussard Stéphane, Seyte Françoise, *Inférence statistique et probabilités*

Szpiro Daniel, *Économie monétaire et financière*.

Tazdaït Tarik, *L'analyse économique de la confiance*.

Terraza Virginie, Toque Carole, *Analyse statistique pour la gestion bancaire et financière*. Applications avec R.

Le livre de référence en statistique inférentielle : fiches de synthèses avec exercices corrigés !

- Ce livre a pour objet **de guider et d'aider les étudiants dans la résolution d'exercices concernant l'inférence statistique**. Les **fiches de synthèse** en début de chapitre permettent de retenir les concepts clés nécessaires à la **résolution des exercices**. Chaque fin de chapitre est consacrée à une mise au point concernant les difficultés rencontrées.
- Le livre s'intéresse au **calcul des probabilités** dans un contexte d'algèbre de Boole, aux **variables aléatoires discrètes à une dimension** (lois inconnues ou connues comme les lois Binomiale, Bernoulli ou Poisson) ainsi qu'aux **variables aléatoires continues à une dimension**. Le problème du calcul de probabilités s'oriente ensuite vers les **variables aléatoires bidimensionnelles discrètes et continues**. Après avoir rappelé la manière dont les **tables statistiques** doivent être lues, les auteurs montrent l'utilité du recours à la **loi normale**, notamment pour certaines statistiques d'échantillonnage. Les **techniques d'estimations (ponctuelles et par intervalles de confiance)** sont étudiées, afin de s'orienter progressivement vers une des parties les plus utiles en statistique : **la théorie des tests** (indépendance, homogénéité, adéquation, signification, comparaison).
- **Cet ouvrage intéressera non seulement les étudiants en Faculté d'économie et de gestion, de niveau L2 et L3, les étudiants en écoles d'ingénieurs, les écoles de commerce, les étudiants préparant certains concours administratifs, les étudiants en MIAGE, en AES (L2-L3), sociologie (L2-L3) et psychologie (L3), mais aussi les professionnels de la statistique (économiste-statisticien de l'État, de l'entreprise, chargé de mission CRM, les professionnels du contrôle de qualité des produits, les entreprises ayant un département statistique).**

Stéphane Mussard *est Maître de conférences HDR à la Faculté d'Économie de l'Université Montpellier I.*

Françoise Seyte *est Maître de conférences HDR à la Faculté d'Économie de l'Université de Montpellier I et responsable du Master Finance de Marché et Analyse des Risques ainsi que du parcours Data Mining Relation Client à l'ISEM. Elle enseigne également à l'École des Mines d'Alès et à l'ENAC.*

Ils sont tous deux membres du LAMETA et enseignent la statistique inférentielle ainsi que l'économétrie. Leurs thèmes de recherche concernent l'économétrie appliquée dans le domaine de la finance et de la répartition personnelle du revenu : lois de distribution du revenu, mesures d'inégalités et de concentration, méthodes de décomposition.

INFSTA
ISBN 978-2-8041-8338-7
ISSN 2030-2061

www.deboeck.com

9 782804 183387

Dans le cadre du nouveau Système Européen de Transfert de Crédits (E.C.T.S.), ce manuel couvre le niveau : Licence et Baccalauréat.

En Belgique Baccalauréat
En Suisse Baccalauréat
Au Canada Licence

L
M
D