

Thérèse Phan
Jean-Pierre Rowenczyk

Exercices et problèmes de statistique et probabilités

2^e édition

DUNOD

Tout le catalogue sur
www.dunod.com



Illustration de couverture : *digitalvision*

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du

Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, Paris, 2012
ISBN 978-2-10-057501-5

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

Avertissement	vii
Chapitre 1 Probabilités	1
Rappel de cours	1
1.1 Rappels de Mathématiques	1
1.2 Axiomes du calcul des probabilités	2
1.3 Notion de variable aléatoire	3
1.4 Moments d'une variable aléatoire	4
1.5 Variables à deux dimensions	7
1.6 Indépendance de deux variables aléatoires X et Y	9
1.7 Probabilités individuelles	9
1.8 Lois de la somme de variables indépendantes connues	10
Énoncés des exercices	11
Énoncés des problèmes	13
Du mal à démarrer ?	14
Corrigés des exercices	15
Corrigés des problèmes	23
Chapitre 2 Convergences et échantillonnage	29
Rappel de cours	29
2.1 Lois statistiques	29
2.2 Propriétés	29
2.3 Échantillon gaussien	30
2.4 Convergences	30
Énoncés des exercices	32

Énoncés des problèmes	34
Du mal à démarrer ?	36
Corrigés des exercices	36
Corrigés des problèmes	41
Chapitre 3 Estimation ponctuelle	49
Rappel de cours	49
3.1 Échantillonnage	49
3.2 Estimation statistique	50
3.3 Éléments de théorie de la décision	51
Énoncés des exercices	52
Énoncés des problèmes	54
Du mal à démarrer ?	55
Corrigés des exercices	56
Corrigés des problèmes	64
Chapitre 4 Information et exhaustivité	71
Rappel de cours	71
4.1 Éléments de théorie de l'information	71
4.2 Méthode du maximum de vraisemblance	73
Énoncés des exercices	74
Énoncés des problèmes	76
Du mal à démarrer ?	78
Corrigés des exercices	79
Corrigés des problèmes	88
Chapitre 5 Estimateur sans biais de variance minimale	97
Rappel de cours	97
5.1 Théorème	97

5.2	Théorème de Rao - Blackwell	97
5.3	Théorème de Lehmann-Scheffe	97
	Énoncés des exercices	98
	Enoncés des problèmes	102
	Du mal à démarrer ?	106
	Corrigés des exercices	107
	Corrigés des problèmes	119
Chapitre 6	Intervalles de confiance	131
	Rappel de cours	131
6.1	Définition d'un intervalle de confiance	131
6.2	Intervalles de confiance pour des paramètres de lois normales	131
6.3	Intervalles de confiance pour les paramètres d'une loi inconnue	134
6.4	Intervalles de confiance pour une proportion	135
	Énoncés des exercices	135
	Énoncés des problèmes	139
	Du mal à démarrer ?	146
	Corrigés des exercices	147
	Corrigés des problèmes	160
Chapitre 7	Tests paramétriques	177
	Rappel de cours	177
7.1	Définition générale d'un problème de test	177
7.2	Théorie de la décision	178
7.3	Notion de risque	179
7.4	Théorème de Neyman et Pearson	179
	Énoncés des exercices	180
	Énoncés des problèmes	185
	Du mal à démarrer ?	188

Corrigés des exercices.....	189
Corrigés des problèmes.....	212
Chapitre 8 Tests d'adéquation et tests d'indépendance.....	223
Rappel de cours.....	223
8.1 Test d'adéquation.....	223
8.2 Test d'indépendance.....	224
Énoncés des Problèmes sur les tests non paramétriques d'adéquation....	227
Énoncés des Problèmes sur les tests non paramétriques d'indépendance .	229
Du mal à démarrer ?	229
Corrigés des problèmes.....	230
Chapitre 9 Analyse de la variance (ou ANOVA) à un seul facteur	245
Rappel de cours.....	245
9.1 Hypothèses	245
9.2 Position du test ANOVA	245
9.1 Observations réalisées.....	246
9.1 Décomposition de la variance totale.....	246
9.2 Principe de l'ANOVA.....	247
9.3 Calcul de la constante C	248
9.4 Comparaison des variances σ_i^2 de chaque population	249
9.5 Mode opératoire pour l'ANOVA.....	250
Énoncé du problème.....	251
Du mal à démarrer ?	252
Corrigé du problème	252
Index.....	255

Avertissement

Cet ouvrage est destiné aux étudiants de Licence, de première année des Grandes Écoles d'ingénieurs, de commerce et de gestion ou d'Institut Universitaires de Technologie désireux d'appréhender les concepts et les notions de base de la statistique.

Il peut être utile à tous ceux qui seraient désireux d'acquérir ou de revoir les notions opérationnelles des méthodes de base de la statistique.

Cet ouvrage comporte des rappels de cours sans démonstrations, des exercices classiques de difficultés progressives (le niveau de difficulté est repéré par un nombre d'étoiles), ainsi que des problèmes plus complexes permettant d'aborder des cas concrets d'utilisation de la statistique dans différents domaines d'application. Il est découpé en chapitres mais il comporte fondamentalement deux grandes parties :

- Une première partie concerne le calcul des probabilités

Bien que comportant des rappels de cours relativement complets, nous avons choisi, délibérément, de ne proposer dans cette partie, que des exercices abordant des notions et des calculs de probabilité qui sont utilisés en statistique : Théorème Central-Limite (ou théorème de la limite centrale), Lois de probabilités fréquemment utilisées en statistique (Loi normale, du Khi-deux, de Student, de Fisher...)

Nous avons donc évité de proposer des exercices de probabilités calculatoires classiques (exercices utilisant la combinatoire, calcul de paramètres de lois de probabilités...).

Pour cette raison, avant d'aborder les chapitres de statistique, nous conseillons vivement au lecteur, de se reporter, en cas de besoin, aux ouvrages spécialisés, afin de revoir ou de compléter leurs connaissances en matière de calcul des probabilités.

- Une deuxième partie est consacrée à l'étude des trois méthodes de base utilisées en statistique :
 - L'estimation ponctuelle
 - L'estimation par intervalle
 - Les tests d'hypothèse

Les chapitres concernant l'estimation ponctuelle permettent d'aborder les notions essentielles permettant d'étudier les estimateurs de paramètres réels de lois de probabilités.

Néanmoins, ces chapitres proposent quelques exemples d'estimation de paramètres vectoriels. Les chapitres consacrés à l'estimation par intervalle proposent un éventail large d'exercices différents, permettant d'appréhender la plupart des cas concrets rencontrés dans les différents domaines utilisant la statistique.

Les chapitres consacrés aux tests d'hypothèses sont essentiellement consacrés à l'étude des tests paramétriques dans le cas d'hypothèses simples et à l'étude de deux types de tests non paramétriques, les tests d'ajustement et les tests d'indépendance.

Les différents chapitres proposent toujours la même organisation : les énoncés, puis une rubrique « Du mal à démarrer », et enfin, les corrigés des exercices proposés.

Chaque corrigé propose, en outre, un bilan « ce qu'il faut retenir ».

Remerciements

Nous tenons, tout d'abord à exprimer toute notre gratitude à nos collègues de l'École Centrale de Paris et de l'École Spéciale des Travaux Publics, pour nous avoir incités à élaborer cet ouvrage et pour nous avoir fourni de nombreux conseils de rédaction.

En particulier, nous tenons à remercier, Alain MARRET et Michel LUCIEN, pour leur apport lors de l'élaboration du contenu de cet ouvrage.

Nos remerciements vont ensuite à Franck PHAN, pour son aide précieuse pour l'utilisation de Latex et donc de la réalisation de la maquette de cet ouvrage.

Enfin, nous tenons également à remercier vivement les Éditions DUNOD, Anne Bourguignon et Benjamin Peylet, pour leur accueil, leur compétence et leur grande compréhension au cours de la réalisation de cet ouvrage.

Thérèse PHAN et Jean-Pierre ROWENCZYK

Probabilités

RAPPEL DE COURS

1.1 Rappels de Mathématiques

a) Opérations sur les ensembles

Soit Ω un ensemble et $A, B \dots$ des parties de Ω . Si $\overline{A}$ désigne le complémentaire de A dans Ω , alors nous avons :

- $A \cup \overline{A} = \Omega$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B)$
- $A \equiv \cup_i (A \cap B_i)$
 - ▷ si les événements B_i sont incompatibles entre eux
 - ▷ et si $\Omega \equiv \cup_i B_i$

b) Analyse combinatoire

Nous rappelons ici quelques résultats :

- Nombre d'arrangements de p objets pris parmi n avec répétition

$$\mathfrak{R}_n^p = n^p$$

- Nombre d'arrangements de p objets pris parmi n sans répétition

$$A_n^p = n(n-1) \dots (n-p+1) = \frac{n!}{(n-p)!}$$

- Nombre de combinaisons de p objets pris parmi n avec répétition

$$K_n^p = C_{n+p-1}^p$$

- Nombre de combinaisons de p objets pris parmi n sans répétition

$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!}$$

- Nombre de permutations de n objets

$$Per(n) = n!$$

1.2 Axiomes du calcul des probabilités

a) Généralités

La théorie des probabilités repose sur l'étude de phénomènes aléatoires. Une expérience est dite aléatoire si on ne peut pas prévoir son résultat et si répétée dans les mêmes conditions, elle peut donner des résultats différents. Les résultats possibles de cette expérience constituent l'ensemble fondamental Ω . Un événement aléatoire est une assertion relative au résultat de l'expérience. On identifie usuellement l'événement aléatoire et la partie de Ω pour laquelle cet événement est réalisé.

Si P est une probabilité définie sur Ω , et si A et B sont deux parties de Ω , on a :

- $P(\emptyset) = 0$ et $P(\Omega) = 1$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$$P(A \cup B) = P(A) + P(B) \quad \text{si} \quad A \cap B = \emptyset$$

b) Probabilités conditionnelles

On définit la probabilité conditionnelle de l'événement A sachant que l'événement B est réalisé, par :

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

c) Formule de décomposition

Si l'ensemble des parties U_j de Ω forme un système complet d'événements, c'est-à-dire si les U_j sont indépendants et si leur réunion forme Ω tout entier, alors :

$$P(A) = \sum_{j=1}^n P(A/U_j)P(U_j)$$

d) Indépendance de deux événements

$$A \text{ et } B \text{ indépendants} \quad \Leftrightarrow \quad P(A \cap B) = P(A)P(B)$$

e) Probabilités des causes ou probabilités de BAYES

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

Si l'ensemble des parties A_i de Ω forme un système complet d'événements,

$$P(A_k/B) = \frac{P(B/A_k)P(A_k)}{\sum_i P(B/A_i)P(A_i)}$$

1.3 Notion de variable aléatoire

Lorsque l'ensemble fondamental Ω est tout ou partie de l'ensemble des réels R , le concept d'événement aléatoire est remplacé par celui de variable aléatoire. On distingue usuellement :

1. les variables aléatoires discrètes pour lesquelles l'ensemble Ω est un ensemble discret de valeurs numériques (par exemple N ensemble des entiers naturels)
2. les variables aléatoires continues pour lesquelles l'ensemble Ω est un intervalle de R ou R tout entier.

a) Fonction de répartition

On appelle « Fonction de répartition d'une variable aléatoire X » l'application F de R dans $[0, 1]$ définie par :

$$F(x) = P(X < x)$$

b) Variable aléatoire discrète

▷ On définit la probabilité attachée en un point x du domaine de définition de la variable aléatoire X discrète par :

$$P(X = x)$$

▷ Fonction de répartition de X :

$$F(x) = P(X < x) = \sum_{t < x} P(X = t)$$

c) Variable aléatoire continue

▷ On dit que la variable aléatoire X de fonction de répartition F est continue si on peut définir une fonction densité de probabilité f de X vérifiant :

$$f(x) = F'(x) \quad \text{ou} \quad F(x) = \int_{-\infty}^x f(t)dt$$

▷ La probabilité attachée au segment $[a, b]$ est alors :

$$P[a \leq X \leq b] = \int_a^b f(x)dx = F(b) - F(a)$$

d) Formule de changement de variables

- Cas discret

$$p_y = P(Y = y) = P(X \in \varphi^{-1}(y)) = \sum_{x \in \varphi^{-1}(y)} P(X = x)$$

- Cas continu

- ▷ φ est monotone croissante

$$G(y) = P(Y < y) = P(X < \varphi^{-1}(y)) = F[\varphi^{-1}(y)]$$

- ▷ φ est monotone décroissante

$$G(y) = P(Y < y) = P(X \geq \varphi^{-1}(y)) = 1 - P[X < \varphi^{-1}(y)]$$

$$G(y) = 1 - F[\varphi^{-1}(y)]$$

- ▷ φ n'est pas monotone

$$G(y) = P(Y < y) = P(X \in I_1) + \dots + P(X \in I_n)$$

où $I_1, \dots, I_n$ sont les intervalles de la variable aléatoire X qui correspondent au domaine $Y < y$.

e) Fonction caractéristique φ_X

- Cas discret

$$\varphi_X(t) = E(e^{itX}) = \sum_{D_X} \exp(itx)P(X = x)$$

- Cas continu

$$\varphi_X(t) = E(e^{itX}) = \int_{D_X} \exp(itx)f(x)dx$$

- Exemples de fonctions caractéristiques :

- ▷ Loi binomiale $B(n, p)$ $\varphi_X(t) = (pe^{it} + 1 - p)^n$

- ▷ Loi de Poisson $P(\lambda)$ $\varphi_X(t) = \exp[\lambda(e^{it} - 1)]$

- ▷ Loi de Gauss $LG(m, \sigma)$ $\varphi_X(t) = e^{itm} \times \exp\left(\frac{-t^2\sigma^2}{2}\right)$

f) Fonctions génératrice G

La fonction génératrice des moments de la variable X est définie par :

$$G_X(u) = E(e^{uX})$$

1.4 Moments d'une variable aléatoire

a) Moment d'ordre r par rapport à l'origine

► Calcul direct

- ▷ variable discrète $m_r = E(X^r) = \sum_X x_i^r P(X = x_i)$

- ▷ variable continue $m_r = E(X^r) = \int_{D_X} t^r f(t)dt$

- Utilisation de la fonction caractéristique

$$m_n = E(X^n) = \frac{\varphi_X^n(0)}{i^n}$$

- Utilisation de la fonction génératrice (pour les variables discrètes)

$$E[X(X-1)\dots(X-n+1)] = G_X^{(n)}(0)$$

où $G_X^{(n)}$ est la dérivée d'ordre n de la fonction génératrice.

b) Moments centrés d'ordre r

- ▷ variable discrète

$$\mu_r = E[(X - E[X])^r] = \sum_i (x_i - E[X])^r P(X = x_i)$$

- ▷ variable continue

$$\mu_r = E[(X - E[X])^r] = \int_{D_X} (t - E[X])^r f(t) dt$$

c) Espérance mathématique

- Calcul direct

- ▷ variable discrète

$$E(X) = \sum_X x_i P(X = x_i)$$

- ▷ variable continue

$$E(X) = \int_{D_{X_x}} t f(t) dt$$

- Espérance d'une somme

$$E(X + Y) = E(X) + E(Y)$$

- Utilisation de la fonction génératrice pour une variable aléatoire discrète

$$E(X) = G_X'(1) = \sum_{D_X} x P(X = x)$$

- Utilisation de la fonction caractéristique

$$m_1 = E(X) = \frac{\varphi_X^1(0)}{i^1}$$