

Collection



Statistique
et probabilités
appliquées

Michel Lejeune

Statistique

La théorie et ses applications

Deuxième édition

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}$$

$$= \frac{\sum_{i=1}^{n_1} (X_i - \bar{X}_1)^2}{n_1 - 1}$$

$$S_1 = \sqrt{S_1^2}$$

$$\bar{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$$



Plus de
150 exercices
corrigés

 Springer

$$S_1 = \sqrt{S_1^2} \quad \bar{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1} \quad \text{et} \quad \bar{X}_2 = \frac{\sum_{j=1}^{n_2} X_j}{n_2}$$

Springer

Paris

Berlin

Heidelberg

New York

Hong Kong

Londres

Milan

Tokyo

Michel Lejeune

Statistique

La théorie et ses applications

Deuxième édition
avec exercices corrigés

 Springer

Michel Lejeune
Professeur émérite
Université de Grenoble 2
IUT 2 département statistique
BP 47
38040 Grenoble cedex 9

ISBN : 978-2-8178-0156-8 Springer Paris Berlin Heidelberg New York

© Springer-Verlag France, Paris, 2010
Imprimé en France

Springer-Verlag France est membre du groupe Springer Science + Business
Media

Cet ouvrage est soumis au copyright. Tous droits réservés, notamment la reproduction et la représentation, la traduction, la réimpression, l'exposé, la reproduction des illustrations et des tableaux, la transmission par voie d'enregistrement sonore ou visuel, la reproduction par microfilm ou tout autre moyen ainsi que la conservation des banques de données. La loi française sur le copyright du 9 septembre 1965 dans la version en vigueur n'autorise une reproduction intégrale ou partielle que dans certains cas, et en principe moyennant le paiement de droits. Toute représentation, reproduction, contrefaçon ou conservation dans une banque de données par quelque procédé que ce soit est sanctionnée par la loi pénale sur le copyright.

L'utilisation dans cet ouvrage de désignations, dénominations commerciales, marques de fabrique, etc. même sans spécification ne signifie pas que ces termes soient libres de la législation sur les marques de fabrique et la protection des marques et qu'ils puissent être utilisés par chacun.

La maison d'édition décline toute responsabilité quant à l'exactitude des indications de dosage et des modes d'emploi. Dans chaque cas, il incombe à l'utilisateur de vérifier les informations données par comparaison à la littérature existante.

Maquette de couverture : Jean-François Montmarché
© détail du tableau " Chrysler impressions " de Philippe Lejeune (1990)



Collection
Statistique et probabilités appliquées
dirigée par Yadolah Dodge

Professeur Honoraire
Université de Neuchâtel
Suisse
yadolah.dodge@unine.ch

Comité éditorial :

Christian Genest

Département de Mathématiques
et de statistique
Université Laval
Québec G1K 7P4
Canada

Stephan Morgenthaler

École Polytechnique Fédérale
de Lausanne
Département des Mathématiques
1015 Lausanne
Suisse

Marc Hallin

Université libre de Bruxelles
Campus de la Plaine CP 210
1050 Bruxelles
Belgique

Gilbert Saporta

Conservatoire national
des arts et métiers
292, rue Saint-Martin
75141 Paris Cedex 3
France

Ludovic Lebart

Télécom-ParisTech
46, rue Barrault
75634 Paris Cedex 13
France

Aurore Delaigle

Department of Mathematics and statistics
Richard Berry Building
The university of Melbourne
VIC, 3010
Australia

Christian Mazza

Département de mathématiques
Université de Fribourg
Chemin du Musée 23
CH-1700 Fribourg
Suisse

Louis-Paul Rivest

Département de mathématiques et de statistique
Université Laval
1045, rue de la Médecine
Québec G1V 0A6
Canada

Dans la même collection :

- *Statistique. La théorie et ses applications*
Michel Lejeune, avril 2004
- *Optimisation appliquée*
Yadolah Dodge, octobre 2004
- *Le choix bayésien. Principes et pratique*
Christian P. Robert, novembre 2005
- *Maîtriser l'aléatoire. Exercices résolus de probabilités et statistique*
Eva Cantoni, Philippe Huber, Elvezio Ronchetti, novembre 2006

- *Régression. Théorie et applications*
Pierre-André Cornillon, Éric Matzner-Løber, janvier 2007
- *Le raisonnement bayésien. Modélisation et inférence*
Éric Parent, Jacques Bernier, juillet 2007
- *Premiers pas en simulation*
Yadolah Dodge, Giuseppe Melfi, juin 2008
- *Génétique statistique*
Stephan Morgenthaler, juillet 2008
- *Pratique du calcul bayésien*
Jean-Jacques Boreux, Éric Parent, 2009
- *Maîtriser l'aléatoire*
Eva Cantoni, Philippe Huber, Elvezio Ronchetti, septembre 2009

À paraître :

- *Le logiciel R*
Pierre Lafaye de Micheaux, Rémi Drouilhet, Benoit Liqueur, 2010

AVANT-PROPOS

L'objectif de cet ouvrage est de rendre accessibles les fondements théoriques de la statistique à un public de niveau mathématique moyen : étudiants du premier cycle des filières scientifiques, élèves ingénieurs, chercheurs dans les domaines appliqués (économie, gestion, biologie, médecine, géographie, sciences de la vie, psychologie. . .) et, plus généralement, tous les chercheurs désireux d'approfondir leur compréhension des résultats utilisés dans la pratique. Pour ces derniers un minimum de connaissance de l'arrière-plan théorique apportera une vision plus claire et plus critique des méthodes qu'ils emploient et permettra d'éviter bien des écueils.

Les prérequis principaux sont la maîtrise de la dérivation, de l'intégration et de bases minimales du calcul des probabilités. Sur le plan purement mathématique, nous pensons que l'essentiel de l'exposé est accessible à quiconque aurait parfaitement assimilé le programme d'un bac scientifique. Il reste cependant quelques notions qui ne sont abordées qu'en premier cycle supérieur, notamment les approximations par développement de Taylor, les développements en série entière, les fonctions de plusieurs variables (dérivation et intégration) et, très marginalement, le calcul matriciel. Mais ces notions n'interviennent le plus souvent que dans les aspects techniques de démonstration, ce qui ne devrait pas nuire à la compréhension des concepts. Pour satisfaire la curiosité de mathématiciens qui voudraient, par la lecture de cet ouvrage, s'initier sans peine à la science statistique, mention sera faite ici ou là de résultats ou démonstrations exigeant des connaissances plus approfondies d'analyse. Ces éléments seront consignés en petits caractères, généralement dans des «notes» détachées que l'on pourra ignorer totalement. Quelques exercices plus difficiles, repérés par un astérisque, leur sont également proposés.

Notons que les premiers chapitres concernent la théorie des probabilités qui, toutefois, est abordée non comme une fin en soi mais de façon simplifiée dans la perspective de ce qui est nécessaire pour la théorie statistique de l'estimation et des tests.

Pour atteindre l'objectif fixé nous avons pris le parti de toujours privilégier la facilité de compréhension au détriment éventuel de la pureté formelle (si tant est qu'elle existe). Nous sommes d'avis que trop de formalisme nuit à l'assimilation des concepts et qu'il faut s'efforcer sans cesse de s'en tenir à un niveau compatible avec celui des connaissances du public visé. Ceci a été un souci constant dans la rédaction. Cela ne signifie pas que nous ayons renoncé à la rigueur du propos, c'est-à-dire à la cohérence des éléments apportés tout au long de l'ouvrage.

Par ailleurs, nous faisons partie de ceux qui pensent que la statistique ne relève pas uniquement de la mathématique qui n'est qu'un instrument. Sa

raison d'être consiste à appréhender le monde réel à partir des observations que l'on en fait. C'est pourquoi la discipline est rangée dans le domaine des mathématiques appliquées, ce terme ne devant pas, à notre sens, rester un vain mot. Fidèle à cette vision nous avons tenté de commenter le plus largement possible les concepts et résultats de façon concrète pour montrer leur utilité dans l'approche du réel. Dans les chapitres débouchant immédiatement sur des méthodes usuelles nous avons également introduit des exercices «appliqués» pour illustrer l'intérêt et la mise en oeuvre des principes théoriques. L'ouvrage n'est donc pas uniquement un traité mathématique. Cela a motivé le choix de son sous-titre « La théorie et ses applications » pour marquer la distinction, même si son objectif premier reste l'exposé de la théorie.

L'essentiel de l'apport de cette nouvelle édition est constitué des corrigés détaillés des exercices proposés. Cette demande m'a été faite de façon récurrente et il est vrai que ces corrigés doivent permettre d'améliorer nettement l'assimilation de la matière.

Je remercie mes collègues Alain Latour et Pierre Lafaye de Micheaux pour leur aide technique précieuse ainsi qu'Alain Catalano, Yves-Alain Gerber, Jérôme Hennet, Alexandre Junod, Julien Junod, Vincent Voirol et Mathieu Vuilleumier pour leurs appréciations.

J'adresse des remerciements particuliers à Yadolah Dodge, directeur de cette collection « Statistique et probabilités appliquées », sans les encouragements duquel cet ouvrage n'aurait sans doute pas abouti.

Michel Lejeune

Grenoble, juin 2010

Table des matières

1	Variables aléatoires	1
1.1	Notion de variable aléatoire	1
1.2	Fonction de répartition	4
1.3	Cas des variables aléatoires discrètes	6
1.4	Cas des variables aléatoires continues	6
1.5	Notion essentielle de quantile	9
1.6	Fonction d'une variable aléatoire	11
1.7	Exercices	12
2	Espérance mathématique et moments	15
2.1	Introduction et définition	15
2.2	Espérance d'une fonction d'une variable aléatoire	16
2.3	Linéarité de l'opérateur $E(\cdot)$, moments, variance	18
2.4	Tirage aléatoire dans une population finie : distribution empirique et distribution probabiliste	21
2.5	Fonction génératrice des moments	21
2.6	Formules d'approximation de l'espérance et de la variance d'une fonction d'une v.a.	24
2.7	Exercices	25
3	Couples et n-uplets de variables aléatoires	27
3.1	Introduction	27
3.2	Couples de v.a.	28
3.3	Indépendance de deux variables aléatoires	31
3.4	Espérance mathématique, covariance, corrélation	32
3.5	Somme de deux v.a.	36
3.6	Les n -uplets de v.a. ; somme de n v.a.	37
3.7	Sondage aléatoire dans une population et v.a. i.i.d.	38
3.8	Notation matricielle des vecteurs aléatoires	39
3.9	Loi de Gauss multivariée	40
3.10	Exercices	43

4	Les lois de probabilités usuelles	45
4.1	Les lois discrètes	45
4.1.1	La loi uniforme discrète	45
4.1.2	Loi de Bernoulli $\mathcal{B}(p)$	46
4.1.3	Le processus de Bernoulli et la loi binomiale $\mathcal{B}(n, p)$	47
4.1.4	Les lois géométrique $\mathcal{G}(p)$ et binomiale négative $\mathcal{BN}(r, p)$	49
4.1.5	La loi hypergéométrique $\mathcal{H}(N, M, n)$	50
4.1.6	La loi multinomiale	51
4.1.7	Le processus et la loi de Poisson $\mathcal{P}(\lambda)$	51
4.2	Les lois continues	54
4.2.1	La loi continue uniforme $\mathcal{U}[a, b]$	54
4.2.2	La loi exponentielle $\mathcal{E}(\lambda)$	55
4.2.3	La loi gamma $\Gamma(r, \lambda)$	56
4.2.4	La loi de Gauss ou loi normale $\mathcal{N}(\mu, \sigma^2)$	57
4.2.5	La loi lognormale $L\mathcal{N}(\mu, \sigma^2)$	60
4.2.6	La loi de Pareto	61
4.2.7	La loi de Weibull $W(\lambda, \alpha)$	61
4.2.8	La loi de Gumbel	62
4.2.9	La loi bêta $Beta(\alpha, \beta)$	63
4.3	Génération de nombres issus d'une loi donnée	63
4.4	Exercices	64
5	Lois fondamentales de l'échantillonnage	67
5.1	Phénomènes et échantillons aléatoires	67
5.2	Moyenne, variance, moments empiriques	69
5.3	Loi du Khi-deux	72
5.4	Loi de Student	74
5.5	Loi de Fisher-Snedecor	76
5.6	Statistiques d'ordre	77
5.7	Fonction de répartition empirique	78
5.8	Convergence, approximations gaussiennes, grands échantillons	79
5.8.1	Les modes de convergence aléatoires	79
5.8.2	Lois des grands nombres	81
5.8.3	Le théorème central limite	82
5.9	Exercices	86
6	Théorie de l'estimation paramétrique ponctuelle	91
6.1	Cadre général de l'estimation	91
6.2	Cadre de l'estimation paramétrique	92
6.3	La classe exponentielle de lois	94
6.4	Une approche intuitive de l'estimation : la méthode des moments	96
6.5	Qualités des estimateurs	98
6.5.1	Biais d'un estimateur	99
6.5.2	Variance et erreur quadratique moyenne d'un estimateur	100
6.5.3	Convergence d'un estimateur	103

6.5.4	Exhaustivité d'un estimateur	105
6.6	Recherche des meilleurs estimateurs sans biais	110
6.6.1	Estimateurs UMVUE	110
6.6.2	Estimation d'une fonction de θ et reparamétrisation	114
6.6.3	Borne de Cramer-Rao et estimateurs efficaces	114
6.6.4	Extension à un paramètre de dimension $k > 1$	118
6.7	L'estimation par la méthode du maximum de vraisemblance	121
6.7.1	Définitions	122
6.7.2	Exemples et propriétés	123
6.7.3	Reparamétrisation et fonctions du paramètre	126
6.7.4	Comportement asymptotique de l'EMV	127
6.8	Les estimateurs bayésiens	128
6.9	Exercices	131
7	Estimation paramétrique par intervalle de confiance	135
7.1	Définitions	135
7.2	Méthode de la fonction pivot	138
7.3	Méthode asymptotique	140
7.4	Construction des IC classiques	144
7.4.1	IC pour la moyenne d'une loi $\mathcal{N}(\mu, \sigma^2)$	144
7.4.2	IC pour la variance σ^2 d'une loi de Gauss	146
7.4.3	IC sur la différence des moyennes de deux lois de Gauss	147
7.4.4	IC sur le rapport des variances de deux lois de Gauss	149
7.4.5	IC sur le paramètre p d'une loi de Bernoulli	150
7.4.6	IC sur la différence des paramètres de deux lois de Bernoulli	152
7.5	IC par la méthode des quantiles	153
7.6	Approche bayésienne	157
7.7	Notions d'optimalité des IC	158
7.8	Région de confiance pour un paramètre de dimension $k > 1$	159
7.9	Intervalles de confiance et tests	163
7.10	Exercices	163
8	Estimation non paramétrique et estimation fonctionnelle	167
8.1	Introduction	167
8.2	Estimation de la moyenne et de la variance de la loi	168
8.2.1	Estimation de la moyenne μ	168
8.2.2	Estimation de la variance σ^2	169
8.3	Estimation d'un quantile	170
8.4	Les méthodes de rééchantillonnage	172
8.4.1	Introduction	172
8.4.2	La méthode du jackknife	173
8.4.3	La méthode du bootstrap	177
8.5	Estimation fonctionnelle	181
8.5.1	Introduction	181
8.5.2	L'estimation de la densité	182

8.5.3	L'estimation de la fonction de répartition	192
8.6	Exercices	198
9	Tests d'hypothèses paramétriques	201
9.1	Introduction	201
9.2	Test d'une hypothèse simple avec alternative simple	202
9.3	Test du rapport de vraisemblance simple	208
9.3.1	Propriété d'optimalité	208
9.3.2	Cas d'un paramètre de dimension 1	212
9.4	Tests d'hypothèses multiples	213
9.4.1	Risques, puissance et optimalité	213
9.4.2	Tests d'hypothèses multiples unilatérales	214
9.4.3	Tests d'hypothèses bilatérales	219
9.5	Test du rapport de vraisemblance généralisé	220
9.6	Remarques diverses	226
9.7	Les tests paramétriques usuels	228
9.7.1	Tests sur la moyenne d'une loi $\mathcal{N}(\mu, \sigma^2)$	229
9.7.2	Test sur la variance σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$	231
9.7.3	Tests de comparaison des moyennes de deux lois de Gauss	232
9.7.4	Tests de comparaison des variances de deux lois de Gauss	235
9.7.5	Tests sur le paramètre p d'une loi de Bernoulli (ou test sur une proportion)	235
9.7.6	Tests de comparaison des paramètres de deux lois de Bernoulli (comparaison de deux proportions)	237
9.7.7	Test sur la corrélation dans un couple gaussien	240
9.8	Dualité entre tests et intervalles de confiance	242
9.9	Exercices	244
10	Tests pour variables catégorielles et tests non paramétriques	251
10.1	Test sur les paramètres d'une loi multinomiale	252
10.1.1	Test du rapport de vraisemblance généralisé	252
10.1.2	Test du khi-deux de Pearson	254
10.1.3	Équivalence asymptotique des deux tests	255
10.1.4	Cas particulier de la loi binomiale	256
10.2	Test de comparaison de plusieurs lois multinomiales	257
10.3	Test d'indépendance de deux variables catégorielles	259
10.3.1	Test du RVG et test du khi-deux	259
10.3.2	Test exact de Fisher (tableau 2×2)	262
10.4	Tests d'ajustement à un modèle de loi	264
10.4.1	Ajustement à une loi parfaitement spécifiée	265
10.4.2	Ajustement dans une famille paramétrique donnée	267
10.5	Tests non paramétriques sur des caractéristiques de lois	272
10.5.1	Introduction	272
10.5.2	Les statistiques de rang	272
10.5.3	Tests sur moyenne, médiane et quantiles	273

10.5.4	Tests de localisation de deux lois	274
10.5.5	Test pour la corrélation de Spearman	281
10.6	Exercices	283
11	Régressions linéaire, logistique et non paramétrique	289
11.1	Introduction à la régression	289
11.2	La régression linéaire	292
11.2.1	Le modèle	292
11.2.2	Les estimateurs du maximum de vraisemblance	293
11.2.3	Intervalles de confiance	296
11.2.4	Test $H_0 : \beta_1 = 0$	297
11.2.5	Cas non gaussien	299
11.2.6	Régression et corrélation linéaires	300
11.2.7	Extension à la régression multiple	303
11.3	La régression logistique	305
11.3.1	Le modèle	305
11.3.2	Estimation de la fonction $p(x)$	306
11.3.3	Matrice des variances-covariances de $\hat{\beta}$	308
11.3.4	Test $H_0 : \beta_1 = 0$	309
11.3.5	Intervalles de confiance	310
11.3.6	Remarques diverses	312
11.4	La régression non paramétrique	314
11.4.1	Introduction	314
11.4.2	Définition des estimateurs à noyaux	314
11.4.3	Biais et variance	315
11.4.4	La régression polynomiale locale	318
11.5	Exercices	320
	Corrigés des exercices	323
	Tables	415
	Bibliographie	421
	Index	425



Collection
Statistique
et probabilités
appliquées

Dirigée par
Yadolah Dodge

COMITÉ ÉDITORIAL :

Aurore Delaigle
Université de Melbourne, Australie

Christian Genest
Université Laval, Québec

Marc Hallin
Université libre de Bruxelles, Belgique

Ludovic Lebart
Télécom-ParisTech, Paris

Christian Mazza
Université de Fribourg, Suisse

Stephan Morgenthaler
EPFL, Lausanne

Louis-Paul Rivest
Université Laval, Québec

Gilbert Saporta
CNAM, Paris

Michel Lejeune

Statistique

La théorie et ses applications

Deuxième édition

Plus de
150 exercices
corrigés

Cette collection met à la disposition du public intéressé par la statistique (étudiants, enseignants, chercheurs) des ouvrages qui concilient effort pédagogique et travail permanent de mise à jour.

Cette démarche implique de prendre en compte de façon sélective et critique les renouvellements des concepts, des champs d'application et des outils de traitement. Seules une compréhension profonde et une appropriation des connaissances permettront de s'adapter aux évolutions qui n'ont pas fini de bouleverser cette discipline.

Cet ouvrage expose les fondements théoriques des méthodes classiques de la statistique (estimation et tests) ainsi que des approches introduites plus récemment.

Les premiers chapitres sont consacrés aux notions de la théorie des probabilités, nécessaires à la statistique. Puis sont développés les tests et méthodes d'estimation dans les situations paramétriques et non paramétriques. Les modèles de base de la régression sont traités en fin d'ouvrage.

Chaque chapitre est accompagné d'exemples concrets, mais aussi d'exercices – plus de 150 au total – dont les corrigés ont été intégrés dans cette deuxième édition.

La présentation témoigne d'un réel souci pédagogique de l'auteur qui bénéficie d'une vaste expérience d'enseignement auprès de publics très variés. Les résultats exposés sont, autant que possible, replacés dans la perspective de leur utilité pratique.

Le niveau mathématique requis rend ce livre accessible aux étudiants de premier cycle universitaire et aux chercheurs dans les divers domaines des sciences appliquées. Il sera donc utile aux étudiants devant aborder les aspects théoriques de la statistique ou aux utilisateurs, pour les assurer du choix judicieux des méthodes qu'ils emploient.

45 € TTC

ISBN : 978-2-8178-0156-8



9 782817 801568

springer.com