

**Ph. Capéraà
B. Van Cutsem**

**méthodes et modèles
en statistique
non paramétrique
exposé fondamental**

**Presses de l'Université
Laval**

Dunod

METHODES ET MODELES

EN STATISTIQUE NON PARAMETRIQUE

Tome 1 : Exposé fondamental

Table des Matières

Chapitre I. INTRODUCTION	1
I. Quelques modèles non paramétriques	2
II. Méthodes non paramétriques	4
III. Qualités des méthodes non paramétriques	6
IV. Plan de l'exposé et cadre de l'étude	9
V. Quelques conventions	10
Chapitre II. STATISTIQUES D'ORDRE	12
I. Introduction	12
II. Définitions et propriétés générales des statistiques d'ordre et de rang	13
II.1. Définitions	13
II.2. Statistiques d'ordre et fonctions symétriques des observations	15
II.3. Indépendance des statistiques d'ordre et des rangs	16
II.4. Statistiques usuelles fonctions des statistiques d'ordre	18
II.4.a. Les p-quantiles empiriques	18
II.4.b. Les (r,s)-étendues	20
II.4.c. Les L-statistiques	20
III. Lois de probabilité associées aux statistiques d'ordre	21
III.1. Lois des composées des statistiques d'ordre et de la fonction de répartition	21
III.2. Lois marginales des statistiques d'ordre	24
III.2.a. Loi de la r ^{ème} statistique d'ordre	24
III.2.b. Loi du couple $(X_{(r)}, X_{(s)})$	26
III.3. Lois asymptotiques de quelques fonctions des statistiques d'ordre	27
III.3.a. Introduction	27
III.3.b. Loi asymptotique d'un sous-vecteur central du vecteur des statistiques d'ordre	28

TABLE DES MATIERES

III.3.c. Lois asymptotiques des extrêmes	30
III.3.d. Lois asymptotiques des L-statistiques	32
III.4. Lois de probabilité des p-quantiles empiriques	34
III.4.a. Lois exactes des p-quantiles empiriques	34
III.4.b. Lois asymptotiques des p-quantiles empiriques	35
III.5. Lois de probabilité des étendues	37
IV. Moments des statistiques d'ordre	38
IV.1. Existence des moments des statistiques d'ordre	38
IV.2. Les moments et leurs approximations	40
V. Préordres sur les lois de probabilité sur \mathbb{R}	43
V.1. Introduction à la notion de poids des queues d'une loi	44
V.2. Comment évaluer le poids des queues ?	46
V.3. Les préordres de van Zwet et Lawrence	47
V.4. Comparaison des préordres de van Zwet et Lawrence	48
V.5. Lien avec les statistiques d'ordre	50
Exercices proposés	51
Chapitre III. EXEMPLES D'UTILISATION DES STATISTIQUES D'ORDRE	53
I. Introduction	53
I.1. Les modèles étudiés	53
I.2. Efficacité relative asymptotique	54
I.3. Qualités des méthodes	55
II. Moyenne et médiane d'un échantillon	55
II.1. Comparaison des variances	55
II.2. Comportement de l'efficacité relative asymptotique	57
II.3. Influence de la taille de l'échantillon	58
II.4. Tolérance aux valeurs extrêmes	59
II.5. Sensibilité aux nouvelles observations	60
III. Moyennes empiriques α -tronquées	62
III.1. Définition	62
III.2. Comparaison des efficacités relatives asymptotiques	63
III.3. Tolérance aux valeurs extrêmes	66
III.4. Sensibilité aux nouvelles observations	66
IV. L-estimateurs	67
IV.1. Quelques propriétés asymptotiques	69
IV.2. Tolérance aux valeurs extrêmes	77
IV.3. Sensibilité à une nouvelle observation, courbe d'influence	78
V. Intervalles de confiance	81
V.1. Intervalles de confiance pour un p-quantile	82
V.2. Intervalles de confiance pour un paramètre de localisation	83
VI. Tests de symétrie	87

TABLE DES MATIERES

VII. Autres utilisations des statistiques d'ordre : estimations d'un paramètre de localisation et d'un paramètre d'échelle	89
VII.1. L-estimateurs des moindres carrés	89
VII.1.a. Le modèle linéaire	89
VII.1.b. Les estimateurs	90
VII.2. Application à des données de durée de vie censurées	93
VII.2.a. Les modèles usuels	93
VII.2.b. Estimation des paramètres	94
VII.2.b.1. Estimateurs du maximum de vraisemblance	94
VII.2.b.2. L-estimateurs des moindres carrés	94
VII.2.b.3. Un exemple	95
VII.3. Méthodes graphiques et L-estimateurs des moindres carrés	96
Exercices proposés	99
Chapitre IV. PROBLEMES A DEUX ECHANTILLONS	101
I. Introduction	101
I.1. Les modèles	101
I.2. Statistiques de rang	105
I.3. Statistiques linéaires de rang	106
II. Comparaison générale de deux échantillons	108
II.1. Le test de Kolmogorov-Smirnov	108
II.2. Loi de la statistique $D_{m,n}$	110
III. Statistiques linéaires de rang	112
III.1. Loi des statistiques linéaires de rang sous l'hypothèse H_0	112
III.2. Loi du vecteur de rang $R(\cdot)$ sous l'hypothèse alternative	115
III.3. Tests de rang localement les plus puissants	116
III.3.a. Définition et caractérisation	117
III.3.b. Modèles de localisation et d'échelle pour deux échantillons	118
III.3.c. Tests de rang avec fonction score croissante	121
IV. Modèle de localisation pour deux échantillons : test de la médiane, test de Wilcoxon	123
IV.1. Test de la médiane	123
IV.2. Tests de Wilcoxon et de Mann-Whitney	125
IV.2.a. Définition	125
IV.2.b. Loix exactes des statistiques de Wilcoxon et de Mann-Whitney sous l'hypothèse H_0	126
IV.2.c. Loix asymptotiques des statistiques de Wilcoxon et de Mann-Whitney sous l'hypothèse H_0	127
IV.3. Comparaison de quelques tests	130

TABLE DES MATIERES

IV.4. Traitement des données avec des ex-æquo	132
IV.4.a. Méthode des rangs au hasard	132
IV.4.b. Méthode des rangs moyens	133
V. Lois et comparaisons asymptotiques	135
V.1. Lois asymptotiques des statistiques linéaires de rang	135
V.1.a. Les théorèmes généraux	135
V.1.b. Cas du test de Wilcoxon	138
V.1.c. Quelques conditions suffisantes de normalité asymptotique sous l'hypothèse H_0	139
V.2. Efficacité relative asymptotique de Pitman	141
V.2.a. Comment comparer deux tests de rang ?	141
V.2.b. Efficacité relative de Pitman	143
V.2.c. Calcul de l'efficacité relative asymptotique	144
VI. Comparaisons de tests de rang pour un paramètre de translation dans un modèle de localisation pour deux échantillons	147
VI.1. Introduction	147
VI.2. Description et propriétés de quelques tests de rang	149
VI.2.a. Test de la médiane	149
VI.2.b. Test de Wilcoxon ou de Mann-Whitney	150
VI.2.c. Test de Fisher-Yates	150
VI.2.d. Test de van der Waerden	151
VI.2.e. Test de Gastwirth	151
VI.3. Comparaison des tests de rang pour un paramètre de translation	152
VII. Estimateurs d'un paramètre de translation dans un modèle de localisation pour deux échantillons	158
VII.1. Introduction	158
VII.2. Estimateurs de Hodges-Lehmann	159
VII.2.a. Définition et exemples	159
VII.2.b. Propriétés des estimateurs	163
VII.2.c. Comparaison des estimateurs	164
VII.3. Intervalles de confiance	165
VIII. Tests de rang dans un modèle d'échelle	169
VIII.1. Introduction	169
VIII.2. Description et propriétés de quelques tests de rang	170
VIII.2.a. Test de Ansari-Bradley	170
VIII.2.b. Test de Mood	171
VIII.2.c. Test de Klotz	172
VIII.2.d. Test de Gastwirth	172
VIII.2.e. Test de Siegel-Tukey	173
VIII.2.f. Test de Savage	173
VIII.3. Comparaison de quelques tests de rang	174

TABLE DES MATIERES

IX. Tests dans un modèle de localisation-échelle pour deux échantillons	176
IX.1. Tests de $\mu_1 = \mu_2$	176
IX.2. Tests de $\sigma_1 = \sigma_2$	179
IX.3. Tests de $\mu_1 = \mu_2$ et $\sigma_1 = \sigma_2$	179
Exercices proposés	179
Chapitre V. PROBLEMES A UN ECHANTILLON	183
I. Introduction	183
II. Test du signe	185
III. Test de rang signé de Wilcoxon	187
III.1. La statistique W_n^+	187
III.2. Loi de la statistique W_n^+ sous l'hypothèse H_0	189
III.3. Loi de la statistique W_n^+ sous l'hypothèse H_1	191
III.4. Propriétés du test de rang signé de Wilcoxon	194
III.4.a. Propriétés non asymptotiques et comparaisons	194
III.4.b. Propriétés asymptotiques, approximation de la puissance	196
IV. Tests de rang signé généraux	199
IV.1. Définition des tests de rang signé	199
IV.2. Lois asymptotiques, comparaisons asymptotiques	200
V. Estimation d'un paramètre de localisation	203
V.1. Estimation ponctuelle	203
V.1.a. Définition et exemples	203
V.1.b. Propriétés des estimateurs	206
V.1.c. Comparaison des L-estimateurs et des R-estimateurs	211
V.2. Intervalles de confiance	213
VI. Tests de symétrie	214
Exercices proposés	216
Chapitre VI. COMPARAISON DE k TRAITEMENTS	218
I. Introduction	218
II. Comparaison de traitements à partir de k échantillons indépendants	221
II.1. Test de Kruskal-Wallis	221
II.1.a. Définition	221
II.1.b. Loi asymptotique de la statistique de Kruskal-Wallis sous l'hypothèse H_0	223
II.1.c. Propriétés et utilisation du test de Kruskal-Wallis	226
II.2. Hypothèse alternative spécifiant un ordre sur les paramètres	228

TABLE DES MATIERES

II.3. Comparaison de $k-1$ traitements avec un traitement de contrôle	230
II.3.a. Tests déduits du test de Wilcoxon	231
II.3.b. Tests déduits du test de Kruskal-Wallis	235
II.4. Comparaisons multiples	236
II.4.a. Tests déduits du test de Wilcoxon	236
II.4.b. Tests déduits du test de Kruskal-Wallis	239
III. Comparaison de k traitements à partir de blocs aléatoires complets	241
III.1. Test de Friedman	241
III.1.a. Définition	241
III.1.b. Loi asymptotique de la statistique de Friedman sous l'hypothèse H_0	243
III.2. Hypothèse alternative spécifiant un ordre sur les paramètres	245
III.3. Comparaisons multiples	247
III.4. Alignement des observations	248
Exercices proposés	250
Chapitre VII. TESTS D'INDEPENDANCE. MESURES D'ASSOCIATION	254
I. Introduction	254
II. Les modèles de dépendance	256
II.1. Les modèles de dépendance entre deux variables aléatoires	256
II.1.a. Le modèle général	256
II.1.b. Le modèle de dépendance par quadrant	256
II.1.c. Le modèle de dépendance par régression	257
II.1.d. Le modèle de dépendance par couplage	258
II.2. Les modèles de dépendance pour des observations multivariées	260
III. Tests d'indépendance	261
III.1. Lois des rangs sous l'hypothèse H_0	261
III.2. Test du quadrant	262
III.3. Le coefficient de corrélation des rangs de Spearman	264
III.3.a. Définition	264
III.3.b. Loi de probabilité sous l'hypothèse H_0	266
III.3.c. Tests d'indépendance associés à ρ_n	268
III.4. Le coefficient de Kendall	269
III.4.a. Définition	269
III.4.b. Loi de probabilité sous l'hypothèse H_0	271
III.4.c. Tests d'indépendance associés à τ_n	274
III.5. Tests d'indépendance généraux	275
III.5.a. Tests de rang généraux	275
III.5.b. Tests d'indépendance utilisant les fonctions de répartition empiriques	278
III.6. Liens entre les coefficients de Spearman et de Kendall	279

TABLE DES MATIERES

IV. Test d'indépendance de n vecteurs aléatoires	281
IV.1. Coefficient de concordance de Kendall	281
IV.2. Loi de probabilité sous l'hypothèse H_0	282
IV.3. Test d'indépendance associé à C_n	283
V. Mesures d'association, indices de dépendance	283
V.1. Indice de dépendance associé à la statistique Q_n	284
V.2. Indice de dépendance associé à la statistique τ_n	285
V.3. Indices de dépendance associés à la statistique ρ_n	286
V.4. Liens entre ces indices de dépendance	288
V.5. Expressions de ces indices à l'aide des fonctions de répartition	288
V.6. Qualités requises d'un indice de dépendance	288
Exercices proposés	289
Chapitre VIII. TESTS D'ADEQUATION	291
I. Introduction	291
II. Tests d'adéquation à un modèle paramétrique	292
II.1 Le test du khi-deux	292
II.2 Tests construits à partir de la fonction de répartition empirique	295
III. Vérification de la normalité des observations	297
III.1. Quelques tests de normalité	297
III.1.a. Le test du khi-deux avec des classes aléatoires	297
III.1.b. Tests utilisant la fonction de répartition empirique	298
III.1.c. Tests utilisant les moments empiriques	299
III.1.d. Tests utilisant des L-statistiques	300
III.2. Comparaison des tests de normalité	302
III.3. Méthodes graphiques	302
III.4. Remarques sur la vérification de la normalité des erreurs dans une régression linéaire	306
Exercices proposés	307
RAPPELS	310
I. Notations générales	310
II. Le modèle probabiliste et le modèle statistique	311
III. Les notions de probabilité	312
III.1. Notions de base	312
III.2. Lois usuelles	313
III.3. Théorèmes de convergence	314
III.4. Ordres stochastiques	318
IV. Les notions de statistique	320
Tables des lois usuelles	323

TABLE DES MATIERES

BIBLIOGRAPHIE	327
TABLES STATISTIQUES	339
INDEX	355

Écrit à l'intention des étudiants du 2^e cycle universitaire qui suivent un enseignement de Mathématiques appliquées ou de Statistique, ce livre présente les principaux thèmes de la statistique non paramétrique : les statistiques d'ordre et leurs utilisations, les tests de rangs et leurs applications, les tests d'adéquation...

Pour chacun de ces domaines, les auteurs proposent une synthèse de l'état actuel des connaissances et mettent en évidence les techniques les plus utilisées (test de Wilcoxon pour un et deux échantillons; tests de Kruskal-Wallis et de Friedman pour plusieurs traitements; tests de Spearman et de Kendall, etc.).

Ce livre consacré à l'exposé fondamental de la statistique non paramétrique est complété par un volume *d'exercices corrigés* (réalisé en collaboration avec Alain Baille).

Philippe Capéraà est professeur à l'Université Laval à Québec. Bernard Van Cutsem, docteur ès-sciences, est professeur à l'Université Joseph Fourier (Grenoble I).