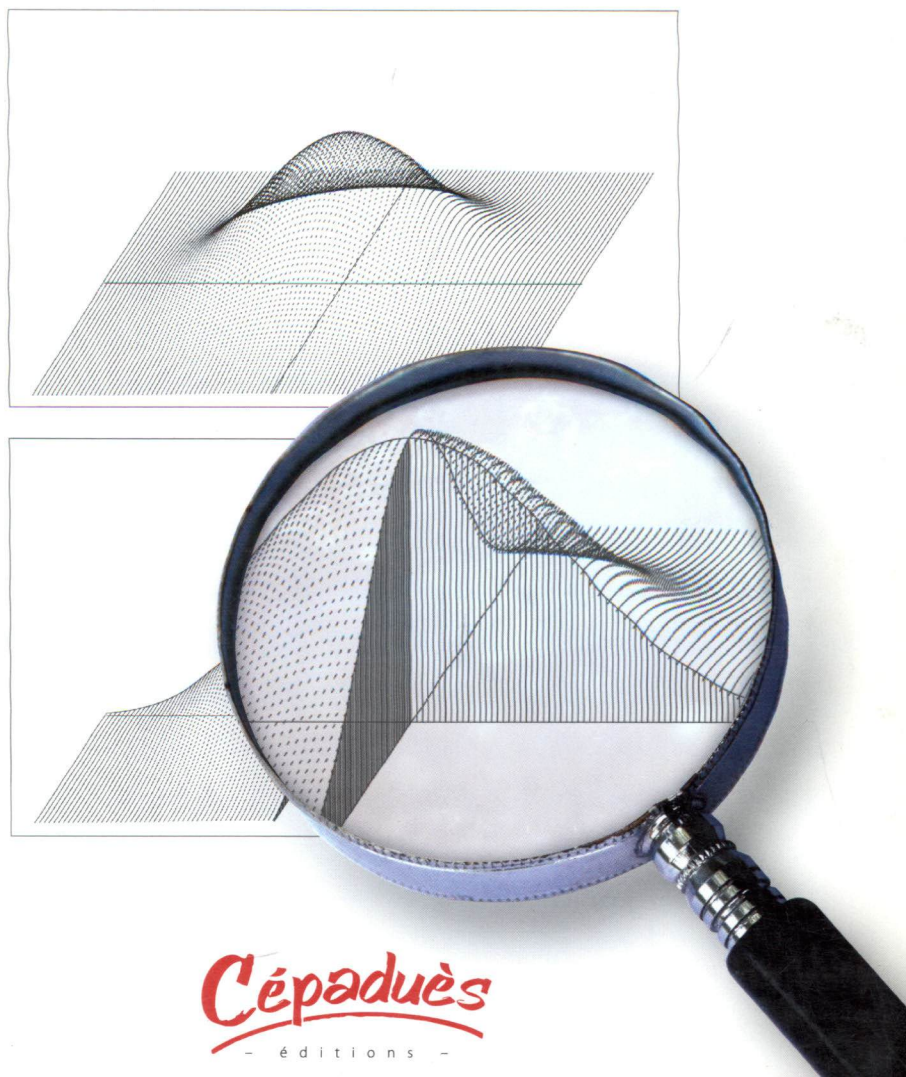


Bernard Prum

La démarche statistique

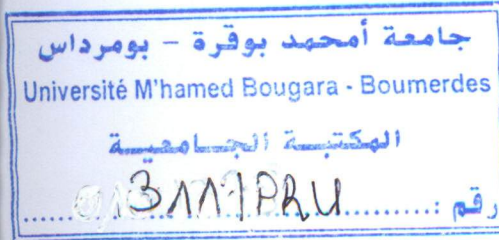


Cépaduès
- éditions -

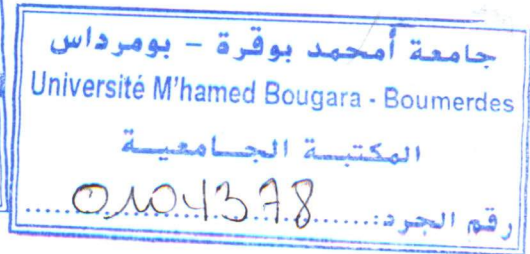


La démarche statistique

Bernard PRUM



01 exp.



CÉPADUÈS-ÉDITIONS

111, rue Nicolas Vauquelin

31100 Toulouse – France

Tél. : 05 61 40 57 36 – Fax : 05 61 41 79 89

www.cepades.com

Courriel : cepades@cepades.com

Coordonnées GPS en WGS 84

N 43° 34'43,2"

E 001° 24'21,5"

Table des matières

1	Modèles statistiques	9
1.1	L'aléatoire	9
1.1.1	Inférence	10
1.1.2	Lois empiriques, lois théoriques	12
1.2	Modèles statistiques	13
1.2.1	La vraisemblance	15
1.2.2	... et la log-vraisemblance	17
1.2.3	Un exemple à suivre	18
1.2.4	Les trois avatars de θ	21
1.3	Bases probabilistes	22
2	Premiers éléments d'inférence statistique	25
2.1	Le modèle binomial	25
2.2	Estimation (cas de la binomiale)	26
2.3	Notion de test	28
2.3.1	Avant de rompre la symétrie	29
2.3.2	Où l'on rompt la symétrie	33
2.3.3	La démarche du test	36
2.3.4	La démarche par l'absurde	45
2.3.5	La relation entre niveau et puissance	48
2.4	Tests unilatères	50
2.4.1	Un exemple	50
2.4.2	Tests unilatères (contre $p > p_0$)	52
2.4.3	Influence de la taille d'échantillon	54
2.5	Tests bilatères	56
3	Rapport de vraisemblance, Neyman-Pearson	61
3.1	Tests d'hypothèses simples	61
3.1.1	Rapport continu	64
3.1.2	Rapport non nécessairement continu	69
3.2	Rapport de vraisemblance monotone	77
3.2.1	Exemple gaussien	77
3.2.2	Cas général	80
3.2.3	Tests de $H_0 : \theta^* \leq \theta_0$	82

3.2.4	Familles exponentielles	83
3.2.5	Discussion	89
4	Vraisemblance, Information	91
4.1	La vraisemblance et ses dérivées	91
4.1.1	Échantillon	91
4.1.2	Le score $L'_X(\theta)$	93
4.2	L'information de Fisher	95
4.2.1	L'information empirique	99
4.2.2	Changement de paramètre	99
4.2.3	La matrice d'information de Fisher	100
4.3	Exhaustivité	101
4.4	Exemples	105
4.4.1	Lois de Poisson	105
4.4.2	Lois gaussiennes	105
4.4.3	Lois Gamma	108
4.5	Perte d'information par image	110
5	Estimation	115
5.1	Estimation	115
5.1.1	Estimateur, biais, écart quadratique	115
5.1.2	Consistance	117
5.1.3	La borne de Cramér-Rao	118
5.1.4	Deux autres qualités asymptotiques	119
5.2	Maximum de vraisemblance	120
5.2.1	Définitions	120
5.2.2	Propriétés	121
5.3	Contrastes	125
5.4	Méthode des moments	131
5.5	Le théorème de Blackwell-Rao	133
5.5.1	Exemples	135
5.6	Paramètre θ multidimensionnel	138
5.6.1	Image réciproque d'une hypothèse	140
6	Les trois tests	141
6.0.2	L'obligation de manipuler deux θ	141
6.0.3	Le problème de ce chapitre	142
6.0.4	Un intervalle, deux points de vue	143

TABLE DES MATIÈRES

v

6.4.1	Un exemple "non standard"	158
7	Modèles multiparamétriques	163
7.1	Quelques propriétés probabilistes	163
7.2	Les trois tests de H_0 contre H_a	165
7.2.1	Les théorèmes	165
7.2.2	Les tests	167
7.3	Les trois tests de H_c contre H_d	172
7.3.1	Eléments théoriques pour le test de H_1 contre H_2	173
7.3.2	En dimensions quelconques	176
7.4	Le cas du modèle linéaire	182
7.4.1	Le modèle	182
7.4.2	Un théorème, des projections	185
7.4.3	Lois de Student et lois de Fisher-Snedecor	188
7.4.4	Les tests	189
7.4.5	Un exemple, la droite de régression	190
8	Test d'une hypothèse composée	195
8.1	Définitions	195
8.1.1	Niveau, puissance	195
8.1.2	Modèle de tests	196
8.2	θ réel, tests unilatères	197
8.2.1	Test fondé sur $T(X)$	199
8.2.2	Test fondé sur $T_\theta(X)$	202
8.2.3	Lien avec le rapport de vraisemblance monotone	204
8.3	θ réel, tests bilatères	205
8.4	Paramètre multidimensionnel	206
9	Régions de confiance	207
9.1	Intervalles : pari, non rejet, confiance	207
9.2	Une démarche incontrôlée	209
9.3	Mise en œuvre : $\theta \in \mathbb{R}$	211
9.3.1	Intervalle de confiance unilatère	211
9.3.2	Intervalle de confiance bilatère	212
9.4	Région de prédiction	214
10	Résumé : l'équation centrale	217
11	Tests du χ^2	219
11.1	Bases probabilistes	219
11.2	Tests d'ajustement	220
11.2.1	Ajustement à une loi multinomiale fixe	220
11.2.2	Ajustement à une loi quelconque	225
11.2.3	Test avec estimation de paramètres	227
11.3	Tests d'indépendance	229

12 Tests non paramétriques	231
12.1 Définitions	231
12.2 Tests de rangs sur <i>un</i> échantillon	232
12.2.1 Statistiques d'ordre et de rang	233
12.2.2 Test "signe et rang"	234
12.3 Tests de rangs sur <i>deux</i> échantillons	237
12.4 Tests de Kolmogorov-Smirnov	241
13 Taille d'échantillon	245
13.1 Niveau, puissance, taille d'échantillon	245
13.2 Tests séquentiels	249
13.2.1 Test séquentiel du rapport de vraisemblances	250
14 Choix de modèle	257
14.1 Modèles emboîtés et non emboîtés	257
14.1.1 Choix de la pénalisation : prédiction optimale	261
14.1.2 Choix de la pénalisation : consistance de l'ordre estimé	264
14.2 Sur-ajustement, la validation croisée	268
15 Multi-tests	273
16 Approches stochastiques	281
16.1 Ré-échantillonnage	284
16.1.1 Le bootstrap	285
16.1.2 Autres ré-échantillonnage, le Jackknife	287
16.2 Algorithmes EM	288
17 L'approche bayésienne	293
17.1 Lois de probabilité sur θ	293
17.1.1 Bayésien généralisé	297
17.2 Estimateur bayésien, admissibilité	298
17.2.1 Lois conjuguées	300
17.3 Le bayésien séquentiel	302
17.4 Loi non informative	307
Envoi	311
A Rappels de probabilités	315

12 Tests non paramétriques	231
12.1 Définitions	231
12.2 Tests de rangs sur <i>un</i> échantillon	232
12.2.1 Statistiques d'ordre et de rang	233
12.2.2 Test "signe et rang"	234
12.3 Tests de rangs sur <i>deux</i> échantillons	237
12.4 Tests de Kolmogorov-Smirnov	241
13 Taille d'échantillon	245
13.1 Niveau, puissance, taille d'échantillon	245
13.2 Tests séquentiels	249
13.2.1 Test séquentiel du rapport de vraisemblances	250
14 Choix de modèle	257
14.1 Modèles emboîtés et non emboîtés	257
14.1.1 Choix de la pénalisation : prédiction optimale	261
14.1.2 Choix de la pénalisation : consistance de l'ordre estimé	264
14.2 Sur-ajustement, la validation croisée	268
15 Multi-tests	273
16 Approches stochastiques	281
16.1 Ré-échantillonnage	284
16.1.1 Le bootstrap	285
16.1.2 Autres ré-échantillonnage, le Jackknife	287
16.2 Algorithmes EM	288
17 L'approche bayésienne	293
17.1 Loïs de probabilité sur θ	293
17.1.1 Bayésien généralisé	297
17.2 Estimateur bayésien, admissibilité	298
17.2.1 Loïs conjuguées	300
17.3 Le bayésien séquentiel	302
17.4 Loi non informative	307
Envoi	311
A Rappels de probabilités	315
A.1 Événements, probabilités, densités	315
A.2 Variables aléatoires	319
A.3 Fonctions de variables aléatoires	324
A.4 Loïs multinomiales, Equilibre de Hardy-Weinberg	329
A.5 Les trois piliers des Probabilités	331
A.6 Transformée de Laplace	332
B Quelques programmes en R	335



Cet ouvrage ne suppose aucune connaissance préalable, si ce n'est quelques lois élémentaires de probabilité (les bases des probabilités sont rappelées en annexe). S'appuyant sur un exemple simple, il introduit les concepts de base et la démarche de test.

Peu à peu il rassemble l'essentiel de ce qu'un étudiant – à l'Université de niveau M, ou en Grande École – doit connaître pour « faire des statistiques » : traitement de plusieurs paramètres, tests non paramétriques, approche bayésienne – sans oublier les incontournables tests du khi-2.

S'appuyant sans cesse sur des exemples illustrés, il défend les tests exacts (par opposition aux approximations, gaussiennes ou autres) et donc l'emploi permanent d'un logiciel de calcul simple (en l'occurrence R).

Cet ouvrage s'appuie sur les enseignements dispensés par l'auteur, essentiellement face à des médecins ou des biologistes, soucieux de comprendre comment les statistiques peuvent les aider dans leur démarche disciplinaire.

Chez le même éditeur

- Simulation & Algorithmes Stochastiques
BARTOLI N. / DEL MORAL P.
- Modélisation Probabiliste et Statistique
GAREL B.
- Probabilités et Statistiques pour Ingénieurs et Commerciaux
PELLAUMAIL J. / PERRET / BASLÉ
- Prévision statistique de la résistance
PLUVINAGE G. / SAPUNOV V.T.
- RNTI E16 Analyse statistique implicative (2 volumes)
RÉGIS GRAS / COLLECTIF RNTI

www.cepadues.com

Réf. 920

ISBN : 978.2.85428.920.6



9 782854 289206