

Table des matières

Préface à la deuxième édition	XIII
Avant-propos : Statistique en action	XV
Cours de statistique : synthèse et mise en perspective	1
1 Concepts fondamentaux de la statistique	3
1.1 Exemples et problématique	3
1.2 Modèle statistique	4
1.3 Bibliographie	6
2 Estimation	7
2.1 Consistance d'un estimateur	7
2.2 Constructions d'estimateurs	8
2.3 Normalité asymptotique	11
2.4 Estimation sans biais	14
2.5 Estimation optimale	15
2.6 Bibliographie	19
3 Intervalles et régions de confiance	21
3.1 Définitions	21
3.2 Premières constructions	22
3.3 Intervalles de confiance exacts obtenus par inégalités de probabilité . .	23
3.4 Intervalles de confiance asymptotiques	25
3.5 Intervalles de confiance simultanés	26
3.6 Bibliographie	27
4 Tests d'hypothèses	29
4.1 Formalisme et démarche expérimentale	29
4.2 p -valeur	34
4.3 Botanique des tests	35
4.4 Dualité entre tests et régions de confiance	37

4.5	Tests de rapport de vraisemblance	38
4.6	Bibliographie	41
5	Vecteurs gaussiens	43
5.1	Propriétés des vecteurs gaussiens et théorème de la limite centrale . .	43
5.2	Lois dérivées : lois du χ^2 et de Student	49
5.3	Conséquences statistiques : intervalles de confiance et tests	50
5.4	Bibliographie	53
6	Tests du χ^2	55
6.1	Test d'ajustement à une loi donnée	55
6.2	Test d'ajustement à une famille paramétrée de lois	59
6.3	Application : test du χ^2 d'indépendance	60
6.4	Cas particulier : test du χ^2 d'homogénéité	63
6.5	Bibliographie	64
7	Modèle linéaire gaussien	65
7.1	Définition, sens et objet du modèle linéaire gaussien	65
7.2	Estimation des paramètres du modèle	69
7.3	Régions de confiance et tests fondamentaux	71
7.4	Régression linéaire	75
7.5	Analyse de la variance	79
7.6	Discussion des hypothèses du modèle linéaire : l'analyse des résidus . .	82
7.7	Bibliographie	83
8	Fonctions de répartition	85
8.1	Consistance des fonctions de répartition empiriques	85
8.2	Tests d'ajustement à une loi ou à une famille de lois	89
8.3	Test d'homogénéité de Kolmogorov–Smirnov	94
8.4	Comparaison avec le test du χ^2	95
8.5	Estimation de quantiles	96
8.6	Bibliographie	97
9	Simulation d'échantillons de loi donnée et applications	99
9.1	Génération de nombres (pseudo-)aléatoires	99
9.2	La méthode d'inversion	100
9.3	La méthode de rejet	102
9.4	Changement de variables	104
9.5	Autres méthodes	105
9.6	Commandes <code>Matlab</code>	106
9.7	Estimation d'espérances difficiles à calculer explicitement	107
9.8	Bibliographie	109

10 Rappels de théorie de l'intégration et des probabilités	111
10.1 Résultats de théorie de l'intégration	111
10.2 Résultats de la théorie des probabilités	113
11 Compléments sur l'estimation	119
11.1 Parfois, il n'existe pas d'estimateur sans biais	119
11.2 Estimation optimale dans les modèles exponentiels	120
 Mise en action : huit thèmes de statistique	 127
12 Machines à sous	131
13 Estimation non-paramétrique pour le modèle de régression	137
14 Inférence statistique pour des modèles censurés	143
15 Étude du nombre de renouvellements	149
16 Estimation de densité de probabilité	155
17 Classification de données	161
18 Compression de données	169
19 Jeux de grattage	175
 Corrigés des thèmes de statistique en action	 181
20 Machines à sous	183
20.1 Corrigé possible du texte	183
20.1.1 Énoncé du problème et définition du modèle statistique associé	183
20.1.2 Premiers exemples de stratégies et limitations de toute stratégie	185
20.1.3 Preuve de la loi des grands nombres pour les martingales . . .	186
20.1.4 Objectif optimal et définition de la notion de bonne stratégie .	190
20.1.5 Une stratégie naïve : la décision simple selon les succès empiriques	192
20.1.6 Une stratégie plus élaborée : la décision selon les succès empi- riques avec excitation	194
20.1.7 Intervalles de confiance sur les paramètres	198
20.1.8 À propos des critères d'évaluation : uniformité, critère en espé- rance, critère d' $(1 - \epsilon)$ -optimalité	201
20.1.9 Performances pratiques de cinq stratégies et demie	206
20.2 Questions possibles, conseils aux étudiants et bibliographie	210

20.2.1	Questions éventuelles d'un enseignant	210
20.2.2	Conseils aux étudiants	211
20.2.3	Bibliographie	212
20.3	Commentaires sur les simulations <code>Matlab</code>	212
20.3.1	Architecture générale des programmes	212
20.3.2	Commentaires sur le code <code>Matlab</code>	213
20.3.3	Codes des simulations	214
21	Estimation non-paramétrique pour le modèle de régression	215
21.1	Préliminaires : questions possibles et conseils aux étudiants	215
21.1.1	Questions éventuelles d'un enseignant	215
21.1.2	Conseils aux étudiants	215
21.2	Corrigé possible du texte	216
21.2.1	Estimation non-paramétrique et fonction de perte	216
21.2.2	Les outils d'analyse nécessaires à l'étude du texte	218
21.2.3	Les espaces de Sobolev	219
21.2.4	Étude du risque par la décomposition biais-variance	221
21.2.5	Discussion sur le choix de N	227
21.2.6	Simulations	228
21.2.7	Application sur un jeu de données réelles	231
21.2.8	Construction de régions de confiance et de tests	238
21.2.9	Discussion sur la modélisation du texte	239
21.2.10	Conseils aux étudiants, suite	240
21.2.11	Bibliographie	240
21.3	Commentaires sur les simulations <code>Matlab</code>	241
21.3.1	Architecture générale des programmes	241
21.3.2	Commentaires sur le code <code>Matlab</code>	241
21.3.3	Sorties numériques	242
21.3.4	Codes des simulations	243
22	Inférence statistique pour des modèles censurés	245
22.1	Préliminaires : questions possibles et conseils aux étudiants	245
22.1.1	Questions éventuelles d'un enseignant	245
22.1.2	Conseils aux étudiants	246
22.2	Corrigé possible du texte	246
22.2.1	Modélisation statistique du problème	246
22.2.2	Notions élémentaires de fiabilité	249
22.2.3	Estimation de la fiabilité dans le cas sans censure	251
22.2.4	Retour à la problématique : construction d'un estimateur consistant de la fonction de survie dans le cas censuré	252
22.2.5	Affaiblissement des hypothèses sur les lois des durées de vie et des censures	261
22.2.6	Modèle exponentiel	261

22.2.7	Comparaisons informatiques des estimateurs de Kaplan–Meier et par plug-in	268
22.2.8	Modèle de Weibull	271
22.2.9	Conclusions et réponses aux questions introductives	272
22.2.10	Bibliographie	273
22.3	Commentaires sur les simulations <code>Matlab</code>	274
22.3.1	Architecture générale des programmes	274
22.3.2	Commentaires sur le code <code>Matlab</code>	274
22.3.3	Codes des simulations	275
23	Étude du nombre de renouvellements	277
23.1	Préliminaires : questions possibles et conseils aux étudiants	277
23.1.1	Questions éventuelles d’un enseignant	277
23.1.2	Conseils aux étudiants	277
23.2	Corrigé possible du texte	278
23.2.1	Modèle statistique	278
23.2.2	Problématique : construction d’un intervalle de prévision	279
23.2.3	Une loi classique en fiabilité : la loi exponentielle	279
23.2.4	Une généralisation de la loi exponentielle : la loi de Weibull	282
23.2.5	Étude asymptotique	283
23.2.6	Étude non-asymptotique dans un cadre général	287
23.2.7	Étude non-asymptotique dans le cadre NBU	300
23.2.8	Simulations probabilistes	303
23.2.9	Application statistique	308
23.2.10	Remarques de conclusion	310
23.2.11	Bibliographie	311
23.3	Commentaires sur les simulations <code>Matlab</code>	312
23.3.1	Architecture générale des programmes	312
23.3.2	Commentaires sur le code <code>Matlab</code>	312
23.3.3	Sorties numériques	313
23.3.4	Codes des simulations	313
24	Estimation de densité de probabilité	315
24.1	Corrigé possible du texte	315
24.1.1	Position du problème et théorème de Scheffé	315
24.1.2	Objectif : estimation consistante de f	318
24.1.3	Estimateur par fenêtres glissantes, preuve de la consistance	318
24.1.4	Estimateur par fenêtres glissantes, normalité asymptotique	321
24.1.5	Retour rapide sur les estimateurs par histogrammes	326
24.1.6	Consistance des estimateurs par noyaux	327
24.1.7	Comparaison par simulations des différentes méthodes d’estimation	336
24.1.8	L’exemple de la reconnaissance de formes	342

24.2	Questions possibles, conseils aux étudiants et bibliographie	345
24.2.1	Questions éventuelles d'un enseignant	345
24.2.2	Conseils aux étudiants	345
24.2.3	Bibliographie	347
24.3	Commentaires sur les simulations <code>Matlab</code>	348
24.3.1	Architecture générale des programmes	348
24.3.2	Commentaires sur le code <code>Matlab</code>	348
24.3.3	Codes des simulations	349
25	Classification de données	351
25.1	Corrigé possible du texte	351
25.1.1	Énoncé du problème, définition du modèle statistique associé	351
25.1.2	Erreur moyenne et fonction de classification optimales	354
25.1.3	Classification avec un expert : apprentissage supervisé	355
25.1.4	Méthode de plug-in : estimation de la fonction de régression	357
25.1.5	Estimation de la fonction de régression par histogrammes, et règle associée	358
25.1.6	Preuve du théorème général de consistance (Théorème 25.3)	361
25.1.7	Preuve du corollaire d'universalité de certaines règles par histogrammes hypercubiques (Corollaire 25.4)	365
25.1.8	Utilisation de noyaux	367
25.1.9	On n'a rien sans rien : des théorèmes « no free lunch »	369
25.1.10	Cas paramétrique : plug-in d'estimateurs des paramètres	371
25.1.11	Estimation des erreurs de classification	374
25.1.12	Comparaison numérique des performances des différentes méthodes étudiées	377
25.2	Questions possibles, conseils aux étudiants et bibliographie	380
25.2.1	Questions éventuelles d'un enseignant	380
25.2.2	Conseils aux étudiants	381
25.2.3	Bibliographie	382
25.3	Commentaires sur les simulations <code>Matlab</code>	383
25.3.1	Architecture générale des programmes	383
25.3.2	Commentaires sur le code <code>Matlab</code>	383
25.3.3	Codes des simulations	384
26	Compression de données	385
26.1	Corrigé possible du texte	385
26.1.1	Modélisation stochastique	386
26.1.2	Code optimal pour coder un mot aléatoire à loi connue	387
26.1.3	Entropie d'une loi	390
26.1.4	Codage d'une suite de mots	395
26.1.5	Code arithmétique et code de Huffman (à probabilité connue)	401

26.1.6	Guide général d'implémentation du code de Huffman sur un fichier de données	406
26.1.7	Implémentation pratique sous <code>Matlab</code> du codage de Huffman	409
26.1.8	Résultats pratiques	414
26.2	Questions possibles, conseils aux étudiants et bibliographie	415
26.2.1	Questions éventuelles d'un enseignant	415
26.2.2	Conseils aux étudiants	416
26.2.3	Bibliographie	417
26.3	Commentaires sur les simulations <code>Matlab</code>	418
26.3.1	Architecture générale des programmes	418
26.3.2	Commentaires sur le code <code>Matlab</code>	418
26.3.3	Codes des simulations	419
	Bibliographie	421
	Index	424
	Annexes et compléments techniques (disponibles en ligne) : http://www.math.ens.fr/statenaction	431
27	Jeux de grattage (corrigé)	433
27.1	Corrigé possible du texte	433
27.1.1	Construction d'un modèle statistique général	433
27.1.2	Construction et rejet d'un modèle probabiliste de référence : répartition totalement aléatoire des lots significatifs	437
27.1.3	Inférence sur ν	444
27.1.4	Étude de la redistribution des sommes mises en jeu	448
27.1.5	Sur l'existence et la pertinence des stratégies de joueur informé	452
27.1.6	Conclusions ?	456
27.2	Questions possibles, conseils aux étudiants et bibliographie	457
27.2.1	Questions éventuelles d'un enseignant	457
27.2.2	Conseils aux étudiants	458
27.2.3	Bibliographie	458
27.3	Commentaires sur les simulations <code>Matlab</code>	458
27.3.1	Architecture générale des programmes	458
27.3.2	Commentaires sur le code <code>Matlab</code>	459
27.3.3	Codes des simulations	460
28	Machines à sous (compléments)	461
28.1	Compléments mathématiques	461
28.1.1	Inégalité de Hoeffding–Azuma	461
28.1.2	Inégalité de Bernstein–Freedman pour les martingales	462
28.2	Codes des simulations	464

29 Estimation non-paramétrique pour le modèle de régression (compléments)	473
29.1 Codes des simulations	473
30 Inférence statistique pour des modèles censurés (compléments)	475
30.1 Codes des simulations	475
31 Étude du nombre de renouvellements (compléments)	481
31.1 Codes des simulations	481
32 Estimation de densité de probabilité (compléments)	487
32.1 Compléments mathématiques	487
32.1.1 Lemme de Scheffé et théorème de Glick	487
32.1.2 Un autre schéma de preuve pour le Théorème 24.7	489
32.1.3 À propos des autres applications de l'estimation de densité	490
32.1.4 Quelques commentaires sur l'optimalité des procédures d'estimation	491
32.2 Codes des simulations	493
33 Classification de données (compléments)	503
33.1 Compléments mathématiques	503
33.1.1 Il est plus facile de classer que d'estimer la fonction de régression	503
33.1.2 Preuve du Théorème 25.7	505
33.1.3 Éléments de preuve pour le Théorème 25.8	508
33.2 Codes des simulations	511
34 Compression de données (compléments)	523
34.1 Compléments mathématiques	523
34.1.1 Sous-additivité de l'entropie, taux d'entropie d'un processus stationnaire	523
34.1.2 Codes universels (un aperçu de la recherche en codage)	526
34.2 Codes des simulations	527
35 Jeux de grattage (compléments)	535
35.1 Codes des simulations	535