

# Table des matières

Avant-propos

XVII

<b>COURS DE STATISTIQUE : SYNTHÈSE ET MISE EN PERSPECTIVE</b>	<b>1</b>
<b>1 Concepts fondamentaux de la statistique</b>	<b>3</b>
1.1 Exemples et problématique . . . . .	3
1.2 Modèle statistique . . . . .	4
1.3 Bibliographie . . . . .	5
<b>2 Estimation</b>	<b>7</b>
2.1 Consistance d'un estimateur . . . . .	7
2.2 Constructions d'estimateurs . . . . .	8
2.3 Normalité asymptotique . . . . .	10
2.4 Estimation sans biais . . . . .	13
2.5 Estimation optimale . . . . .	14
2.6 Bibliographie . . . . .	19
<b>3 Intervalles et régions de confiance</b>	<b>21</b>
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 . . . . .	24
3.5 Intervalles de confiance simultanés . . . . .	26
3.6 Bibliographie . . . . .	27
<b>4 Tests d'hypothèses</b>	<b>29</b>
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 . . . . .	40

<b>5</b>	<b>Vecteurs gaussiens</b>	<b>41</b>
5.1	Propriétés des vecteurs gaussiens et théorème de la limite centrale . . . . .	41
5.2	Lois dérivées : lois du $\chi^2$ et de Student . . . . .	47
5.3	Conséquences statistiques : intervalles de confiance et tests . . . . .	48
5.4	Bibliographie . . . . .	51
<b>6</b>	<b>Tests du <math>\chi^2</math></b>	<b>53</b>
6.1	Test d'ajustement à une loi donnée . . . . .	53
6.2	Test d'ajustement à une famille paramétrée de lois . . . . .	57
6.3	Application : test du $\chi^2$ d'indépendance . . . . .	58
6.4	Cas particulier : test du $\chi^2$ d'homogénéité . . . . .	61
6.5	Bibliographie . . . . .	62
<b>7</b>	<b>Modèle linéaire gaussien</b>	<b>63</b>
7.1	Définition, sens et objet du modèle linéaire gaussien . . . . .	63
7.2	Estimation des paramètres du modèle . . . . .	67
7.3	Régions de confiance et tests fondamentaux . . . . .	69
7.4	Régression linéaire . . . . .	73
7.5	Analyse de la variance . . . . .	76
7.6	Discussion des hypothèses du modèle linéaire : l'analyse des résidus . . . . .	79
7.7	Bibliographie . . . . .	81
<b>8</b>	<b>Fonctions de répartition</b>	<b>83</b>
8.1	Consistance des fonctions de répartition empiriques . . . . .	83
8.2	Tests d'ajustement à une loi ou à une famille de lois . . . . .	87
8.3	Test d'homogénéité de Kolmogorov–Smirnov . . . . .	92
8.4	Comparaison avec le test du $\chi^2$ . . . . .	92
8.5	Estimation de quantiles . . . . .	93
8.6	Bibliographie . . . . .	95
<b>9</b>	<b>Simulation d'échantillons de loi donnée et applications</b>	<b>97</b>
9.1	Génération de nombres (pseudo-)aléatoires . . . . .	97
9.2	La méthode d'inversion . . . . .	98
9.3	La méthode de rejet . . . . .	100
9.4	Changement de variables . . . . .	101
9.5	Autres méthodes . . . . .	103
9.6	Commandes Matlab . . . . .	104
9.7	Estimation d'espérances difficiles à calculer explicitement . . . . .	104
9.8	Bibliographie . . . . .	107
	<b>MISE EN ACTION : HUIT THÈMES DE STATISTIQUE</b>	<b>109</b>
<b>10</b>	<b>Machines à sous</b>	<b>113</b>

<b>11 Estimation non-paramétrique pour le modèle de régression</b>	<b>119</b>
<b>12 Inférence statistique pour des modèles censurés</b>	<b>125</b>
<b>13 Étude du nombre de renouvellements</b>	<b>131</b>
<b>14 Estimation de densité de probabilité</b>	<b>137</b>
<b>15 Classification de données</b>	<b>143</b>
<b>16 Compression de données</b>	<b>149</b>
<b>17 Jeux de grattage</b>	<b>155</b>
<b>CORRIGÉS DES THÈMES DE STATISTIQUE EN ACTION</b>	<b>161</b>
<b>18 Estimation non-paramétrique pour le modèle de régression</b>	<b>163</b>
18.1 Préliminaires : questions possibles et conseils aux étudiants . . . . .	163
18.1.1 Questions éventuelles d'un enseignant . . . . .	163
18.1.2 Conseils aux étudiants . . . . .	163
18.2 Corrigé possible du texte . . . . .	164
18.2.1 Estimation non-paramétrique et fonction de perte . . . . .	164
18.2.2 Les outils d'analyse nécessaires à l'étude du texte . . . . .	165
18.2.3 Les espaces de Sobolev . . . . .	166
18.2.4 Etude du risque par la décomposition biais-variance . . . . .	168
18.2.5 Discussion sur le choix de $N$ . . . . .	174
18.2.6 Simulations . . . . .	176
18.2.7 Application sur un jeu de données réelles . . . . .	179
18.2.8 Construction de régions de confiance et de tests . . . . .	185
18.2.9 Discussion sur la modélisation du texte . . . . .	186
18.2.10 Conseils aux étudiants, suite . . . . .	187
18.2.11 Bibliographie . . . . .	187
18.3 Commentaires sur les simulations <code>Matlab</code> . . . . .	188
18.3.1 Architecture générale des programmes . . . . .	188
18.3.2 Commentaires sur le code <code>Matlab</code> . . . . .	188
18.3.3 Sorties numériques . . . . .	189
18.3.4 Codes des simulations . . . . .	190
<b>19 Inférence statistique pour des modèles censurés</b>	<b>191</b>
19.1 Préliminaires : questions possibles et conseils aux étudiants . . . . .	191
19.1.1 Questions éventuelles d'un enseignant . . . . .	191
19.1.2 Conseils aux étudiants . . . . .	192
19.2 Corrigé possible du texte . . . . .	192

19.2.1	Modélisation statistique du problème . . . . .	192
19.2.2	Notions élémentaires de fiabilité . . . . .	196
19.2.3	Estimation de la fiabilité dans le cas sans censure . . . . .	197
19.2.4	Retour à la problématique : construction d'un estimateur consistant de la fonction de survie dans le cas censuré . . . . .	198
19.2.5	Affaiblissement des hypothèses sur les lois des durées de vie et des censures . . . . .	206
19.2.6	Modèle exponentiel . . . . .	207
19.2.7	Comparaisons informatiques des estimateurs de Kaplan–Meier et par plug-in . . . . .	214
19.2.8	Modèle de Weibull . . . . .	216
19.2.9	Conclusions et réponses aux questions introductives . . . . .	217
19.2.10	Bibliographie . . . . .	218
19.3	Commentaires sur les simulations MatLab . . . . .	219
19.3.1	Architecture générale des programmes . . . . .	219
19.3.2	Commentaires sur le code MatLab . . . . .	219
19.3.3	Codes des simulations . . . . .	220
<b>20</b>	<b>Classification de données</b>	<b>221</b>
20.1	Corrigé possible du texte . . . . .	221
20.1.1	Enoncé du problème, définition du modèle statistique associé . . . . .	221
20.1.2	Erreur moyenne et fonction de classification optimales . . . . .	224
20.1.3	Classification avec un expert : apprentissage supervisé . . . . .	225
20.1.4	Méthode de plug-in : estimation de la fonction de régression . . . . .	227
20.1.5	Estimation de la fonction de régression par histogrammes, et règle associée . . . . .	228
20.1.6	Preuve du théorème général de consistance (Théorème 20.3) . . . . .	231
20.1.7	Preuve du corollaire d'universalité de certaines règles par histogrammes hypercubiques (Corollaire 20.4) . . . . .	235
20.1.8	Utilisation de noyaux . . . . .	237
20.1.9	On n'a rien sans rien : des théorèmes « no free lunch » . . . . .	238
20.1.10	Cas paramétrique : plug-in d'estimateurs des paramètres . . . . .	240
20.1.11	Estimation des erreurs de classification . . . . .	243
20.1.12	Comparaison numérique des performances des différentes méthodes étudiées . . . . .	246
20.2	Questions possibles, conseils aux étudiants et bibliographie . . . . .	249
20.2.1	Questions éventuelles d'un enseignant . . . . .	249
20.2.2	Conseils aux étudiants . . . . .	250
20.2.3	Bibliographie . . . . .	251
20.3	Commentaires sur les simulations MatLab . . . . .	252
20.3.1	Architecture générale des programmes . . . . .	252
20.3.2	Commentaires sur le code MatLab . . . . .	252
20.3.3	Codes des simulations . . . . .	253

<b>21 Compression de données</b>	<b>255</b>
21.1 Corrigé possible du texte . . . . .	255
21.1.1 Modélisation stochastique . . . . .	256
21.1.2 Code optimal pour coder un mot aléatoire à loi connue . . . . .	257
21.1.3 Entropie d'une loi . . . . .	259
21.1.4 Codage d'une suite de mots . . . . .	264
21.1.5 Code arithmétique et code de Huffman (à probabilité connue) . . . .	270
21.1.6 Guide général d'implémentation du code de Huffman sur un fichier de données . . . . .	275
21.1.7 Implémentation pratique sous Matlab du codage de Huffman . . . .	277
21.1.8 Résultats pratiques . . . . .	282
21.2 Questions possibles, conseils aux étudiants et bibliographie . . . . .	283
21.2.1 Questions éventuelles d'un enseignant . . . . .	283
21.2.2 Conseils aux étudiants . . . . .	284
21.2.3 Bibliographie . . . . .	285
21.3 Commentaires sur les simulations Matlab . . . . .	285
21.3.1 Architecture générale des programmes . . . . .	285
21.3.2 Commentaires sur le code Matlab . . . . .	286
21.3.3 Codes des simulations . . . . .	287
 <b>Bibliographie</b>	 <b>289</b>
 <b>Index</b>	 <b>293</b>
 <b>ANNEXES (DISPONIBLES EN LIGNE) :</b> <b><a href="http://www.dma.ens.fr/statenaction">http://www.dma.ens.fr/statenaction</a></b>	 <b>299</b>
 <b>22 Machines à sous</b>	 <b>301</b>
22.1 Corrigé possible du texte . . . . .	301
22.1.1 Enoncé du problème et définition du modèle statistique associé . . . .	301
22.1.2 Premiers exemples de stratégies et limitations de toute stratégie . . . .	303
22.1.3 Preuve de la loi des grands nombres pour les martingales . . . . .	304
22.1.4 Objectif optimal et définition de la notion de bonne stratégie . . . . .	308
22.1.5 Une stratégie naïve : la décision simple selon les succès empiriques . .	310
22.1.6 Une stratégie plus élaborée : la décision selon les succès empiriques avec excitation . . . . .	311
22.1.7 Intervalles de confiance sur les paramètres . . . . .	315
22.1.8 A propos des critères d'évaluation : uniformité, critère en espérance, critère d' $(1 - \varepsilon)$ -optimalité . . . . .	318
22.1.9 Performances pratiques de cinq stratégies et demie . . . . .	322
22.2 Questions possibles, conseils aux étudiants et bibliographie . . . . .	327
22.2.1 Questions éventuelles d'un enseignant . . . . .	327

22.2.2	Conseils aux étudiants . . . . .	328
22.2.3	Bibliographie . . . . .	329
22.3	Commentaires sur les simulations Matlab . . . . .	329
22.3.1	Architecture générale des programmes . . . . .	329
22.3.2	Commentaires sur le code Matlab . . . . .	329
22.3.3	Codes des simulations . . . . .	330
<b>23</b>	<b>Étude du nombre de renouvellements</b>	<b>331</b>
23.1	Préliminaires : questions possibles et conseils aux étudiants . . . . .	331
23.1.1	Questions éventuelles d'un enseignant . . . . .	331
23.1.2	Conseils aux étudiants . . . . .	331
23.2	Corrigé possible du texte . . . . .	332
23.2.1	Modèle statistique . . . . .	332
23.2.2	Problématique : construction d'un intervalle de prévision . . . . .	333
23.2.3	Une loi classique en fiabilité : la loi exponentielle . . . . .	333
23.2.4	Une généralisation de la loi exponentielle : la loi de Weibull . . . . .	335
23.2.5	Etude asymptotique . . . . .	336
23.2.6	Etude non-asymptotique dans un cadre général . . . . .	341
23.2.7	Etude non-asymptotique dans le cadre NBU . . . . .	354
23.2.8	Simulations probabilistes . . . . .	356
23.2.9	Application statistique . . . . .	361
23.2.10	Remarques de conclusion . . . . .	364
23.2.11	Bibliographie . . . . .	365
23.3	Commentaires sur les simulations Matlab . . . . .	365
23.3.1	Architecture générale des programmes . . . . .	365
23.3.2	Commentaires sur le code Matlab . . . . .	366
23.3.3	Sorties numériques . . . . .	366
23.3.4	Codes des simulations . . . . .	366
<b>24</b>	<b>Estimation de densité de probabilité</b>	<b>367</b>
24.1	Corrigé possible du texte . . . . .	367
24.1.1	Position du problème et théorème de Scheffé . . . . .	367
24.1.2	Objectif : estimation consistante de $f$ . . . . .	369
24.1.3	Estimateur par fenêtres glissantes, preuve de la consistance . . . . .	370
24.1.4	Estimateur par fenêtres glissantes, normalité asymptotique . . . . .	373
24.1.5	Retour rapide sur les estimateurs par histogrammes . . . . .	376
24.1.6	Consistance des estimateurs par noyaux . . . . .	379
24.1.7	Comparaison par simulations des différentes méthodes d'estimation . . . . .	387
24.1.8	L'exemple de la reconnaissance de formes . . . . .	393
24.2	Questions possibles, conseils aux étudiants et bibliographie . . . . .	396
24.2.1	Questions éventuelles d'un enseignant . . . . .	396
24.2.2	Conseils aux étudiants . . . . .	396
24.2.3	Bibliographie . . . . .	397

24.3	Commentaires sur les simulations Mat1ab . . . . .	398
24.3.1	Architecture générale des programmes . . . . .	398
24.3.2	Commentaires sur le code Mat1ab . . . . .	399
24.3.3	Codes des simulations . . . . .	399
<b>25</b>	<b>Jeux de grattage</b>	<b>401</b>
25.1	Corrigé possible du texte . . . . .	401
25.1.1	Construction d'un modèle statistique général . . . . .	401
25.1.2	Construction et rejet d'un modèle probabiliste de référence : répartition totalement aléatoire des lots significatifs . . . . .	405
25.1.3	Inférence sur $\nu$ . . . . .	411
25.1.4	Etude de la redistribution des sommes mises en jeu . . . . .	415
25.1.5	Sur l'existence et la pertinence des stratégies de joueur informé . . .	419
25.1.6	Conclusions? . . . . .	423
25.2	Questions possibles, conseils aux étudiants et bibliographie . . . . .	424
25.2.1	Questions éventuelles d'un enseignant . . . . .	424
25.2.2	Conseils aux étudiants . . . . .	424
25.2.3	Bibliographie . . . . .	425
25.3	Commentaires sur les simulations Mat1ab . . . . .	425
25.3.1	Architecture générale des programmes . . . . .	425
25.3.2	Commentaires sur le code Mat1ab . . . . .	425
25.3.3	Codes des simulations . . . . .	426
	<b>COMPLÉMENTS TECHNIQUES (DISPONIBLES EN LIGNE) :</b>	
	<b><a href="http://www.dma.ens.fr/statenaction">http://www.dma.ens.fr/statenaction</a></b>	<b>427</b>
<b>26</b>	<b>Rappels de théorie de l'intégration et des probabilités</b>	<b>429</b>
26.1	Résultats de théorie de l'intégration . . . . .	429
26.1.1	Théorème de dérivation des intégrales à paramètre . . . . .	429
26.1.2	Le théorème fondamental du calcul différentiel et intégral . . . . .	429
26.1.3	Le théorème de changement de variables . . . . .	431
26.2	Résultats de la théorie des probabilités . . . . .	431
26.2.1	Convergences . . . . .	431
26.2.2	Inégalités utiles . . . . .	434
<b>27</b>	<b>Compléments sur l'estimation</b>	<b>437</b>
27.1	Preuve des assertions de l'exemple 2.16 . . . . .	437
27.2	Estimation optimale dans les modèles exponentiels . . . . .	438

<b>28 Machines à sous (compléments)</b>	<b>443</b>
28.1 Compléments mathématiques . . . . .	443
28.1.1 Inégalité de Hoeffding–Azuma . . . . .	443
28.1.2 Inégalité de Bernstein–Freedman pour les martingales . . . . .	444
28.2 Codes des simulations . . . . .	445
<b>29 Estimation non-paramétrique pour le modèle de régression (compléments)</b>	<b>455</b>
29.1 Codes des simulations . . . . .	455
<b>30 Inférence statistique pour des modèles censurés (compléments)</b>	<b>461</b>
30.1 Codes des simulations . . . . .	461
<b>31 Étude du nombre de renouvellements (compléments)</b>	<b>467</b>
31.1 Codes des simulations . . . . .	467
<b>32 Estimation de densité de probabilité (compléments)</b>	<b>473</b>
32.1 Compléments mathématiques . . . . .	473
32.1.1 Lemme de Scheffé et théorème de Glick . . . . .	473
32.1.2 Un autre schéma de preuve pour le Théorème 24.7 . . . . .	475
32.1.3 A propos des autres applications de l’estimation de densité . . . . .	475
32.1.4 Quelques commentaires sur l’optimalité des procédures d’estimation	477
32.2 Codes des simulations . . . . .	479
<b>33 Classification de données (compléments)</b>	<b>489</b>
33.1 Compléments mathématiques . . . . .	489
33.1.1 Il est plus facile de classer que d’estimer la fonction de régression . . .	489
33.1.2 Preuve du Théorème 20.7 . . . . .	491
33.1.3 Éléments de preuve pour le Théorème 20.8 . . . . .	494
33.2 Codes des simulations . . . . .	497
<b>34 Compression de données (compléments)</b>	<b>507</b>
34.1 Compléments mathématiques . . . . .	507
34.1.1 Sous-additivité de l’entropie, taux d’entropie d’un processus stationnaire	507
34.1.2 Codes universels (un aperçu de la recherche en codage) . . . . .	510
34.2 Codes des simulations . . . . .	511
<b>35 Jeux de grattage (compléments)</b>	<b>519</b>
35.1 Codes des simulations . . . . .	519