

# Table des matières

<b>I</b>	<b>Bases théoriques</b>	<b>1</b>
<b>1</b>	<b>Bases probabilistes utiles</b>	<b>3</b>
1.1	Introduction . . . . .	3
1.2	L'incertitude . . . . .	5
1.2.1	Aléa et variabilité naturelle . . . . .	5
1.2.2	Incertitude par ignorance . . . . .	7
1.2.3	Compléments . . . . .	8
1.2.4	Événements incertains . . . . .	10
1.3	Qu'est-ce qu'une probabilité? . . . . .	13
1.3.1	La probabilité par l'exemple . . . . .	13
1.3.2	Définition axiomatique et propriétés . . . . .	15
1.3.3	Extension au cas où l'univers est infini dénombrable ou continu	16
1.3.4	La probabilité : une propriété de l'objet . . . . .	17
1.3.5	La probabilité : une propriété du sujet . . . . .	18
1.4	Probabilité conditionnelle et indépendance . . . . .	19
1.4.1	Probabilité conditionnelle et statistique bayésienne . . . . .	19
1.4.2	La probabilité conditionnelle par l'exemple . . . . .	20
1.4.3	Définition, propriétés et applications . . . . .	21
1.4.4	Indépendance stochastique . . . . .	23
1.4.5	La clé de voûte de la statistique bayésienne : la formule de Bayes	25
1.4.6	Casse-têtes résolus par les probabilités conditionnelles . . . . .	29
1.5	Variable aléatoire et loi de probabilité . . . . .	35
1.5.1	Qu'est-ce qu'une variable aléatoire? . . . . .	35
1.5.2	Variabes aléatoires discrètes . . . . .	38
1.5.3	Variabes aléatoires continues . . . . .	46
1.5.4	Comment simuler une variable aléatoire? . . . . .	55
1.6	Couple de variables aléatoires . . . . .	55
1.6.1	Qu'est-ce qu'un couple de v.a.? Définition et exemples . . . . .	56
1.6.2	Couple de v.a. discrètes . . . . .	58
1.6.3	Couple de v.a. continues . . . . .	65
1.6.4	Modélisation d'un couple de v.a. . . . .	67

<b>2</b>	<b>Modélisation probabiliste</b>	<b>71</b>
2.1	Introduction . . . . .	71
2.1.1	La construction d'un modèle . . . . .	72
2.1.2	Les différents ingrédients d'un modèle probabiliste . . . . .	73
2.2	Lois de probabilité usuelles . . . . .	77
2.2.1	Loi de Bernoulli (et modèle d'urne) . . . . .	77
2.2.2	Loi binomiale . . . . .	78
2.2.3	Loi de Poisson . . . . .	79
2.2.4	Lois géométrique et exponentielle . . . . .	80
2.2.5	Loi normale . . . . .	82
2.2.6	Loi bêta . . . . .	84
2.2.7	Loi gamma . . . . .	85
2.2.8	Loi binormale . . . . .	87
<b>3</b>	<b>Du classique au bayésien</b>	<b>89</b>
3.1	Introduction . . . . .	89
3.1.1	Propos . . . . .	89
3.1.2	Exemple discret : un modèle binomial . . . . .	90
3.1.3	Exemple continu : un modèle normal . . . . .	90
3.1.4	Modèle probabiliste <i>versus</i> inférence statistique . . . . .	91
3.2	Inférence statistique classique . . . . .	92
3.2.1	Estimation . . . . .	92
3.2.2	Propriétés des estimateurs . . . . .	95
3.2.3	Tests et intervalles de confiance . . . . .	95
3.3	Inférence statistique bayésienne . . . . .	102
3.3.1	Généralités . . . . .	102
3.3.2	Lois <i>a priori</i> et <i>a posteriori</i> . . . . .	103
3.3.3	Démarche . . . . .	110
3.3.4	Influence de la loi <i>a priori</i> et des données sur la loi <i>a posteriori</i> . . . . .	111
3.3.5	Estimation et intervalle de confiance en bayésien ? . . . . .	113
3.4	Conclusion . . . . .	114
3.4.1	Récapitulatif de la démarche bayésienne . . . . .	115
3.4.2	Recommandations . . . . .	115
3.4.3	Pourquoi préférer le bayésien ? . . . . .	116
<b>4</b>	<b>La modélisation graphique</b>	<b>121</b>
4.1	Introduction . . . . .	121
4.2	Graphes et dépendances . . . . .	121
4.2.1	Graphes et graphes orientés . . . . .	121
4.2.2	Graphes acycliques orientés ou DAG . . . . .	123
4.2.3	Indépendance conditionnelle . . . . .	125
4.2.4	DAGs équivalents . . . . .	127
4.2.5	Réseau bayésien . . . . .	129
4.3	Corrélation n'est pas causalité . . . . .	130
4.4	Vers les structures hiérarchiques . . . . .	131
4.4.1	Un premier exemple simple . . . . .	131

4.4.2	Un second exemple . . . . .	133
4.4.3	Généralisation : la modélisation hiérarchique bayésienne . . . . .	137
4.5	Propriétés mathématiques des DAGs . . . . .	138
4.5.1	La factorisation de la loi jointe . . . . .	138
4.5.2	L'inférence bayésienne . . . . .	138
4.5.3	Loi conditionnelle complète dans un DAG . . . . .	140
4.6	Modèles multidimensionnels traditionnels . . . . .	142
4.6.1	Lois normales multidimensionnelles et GMRF . . . . .	142
4.6.2	Loi <i>a posteriori</i> de la moyenne . . . . .	143
4.6.3	Modèles normaux-Wishart . . . . .	144
4.6.4	Modèles Dirichlet-multinomiaux ou Polya . . . . .	146
4.7	Les logiciels . . . . .	148
4.8	Conclusions . . . . .	149
<b>5</b>	<b>Méthodes numériques</b>	<b>151</b>
5.1	Introduction : pourquoi utiliser des méthodes numériques? . . . . .	151
5.2	Exemple d'illustration en épidémiologie spatiale . . . . .	153
5.2.1	Modèle dynamique . . . . .	154
5.2.2	Données et évaluation des lois <i>a posteriori</i> . . . . .	156
5.2.3	Lois <i>a priori</i> . . . . .	157
5.3	Méthode du ré-échantillonnage par importance . . . . .	158
5.3.1	Principe . . . . .	158
5.3.2	Application . . . . .	159
5.3.3	Limites . . . . .	160
5.3.4	Pour aller plus loin : échantillonnage séquentiel par importance . . . . .	161
5.4	Algorithme de Monte Carlo par chaînes de Markov . . . . .	163
5.4.1	Principe . . . . .	163
5.4.2	Échantillonneur de Gibbs . . . . .	164
5.4.3	Échantillonneur de Metropolis-Hastings . . . . .	165
5.4.4	Application 1 : échantillonnage complet . . . . .	165
5.4.5	Application 2 : échantillonnage incomplet temporellement . . . . .	168
5.4.6	Application 3 : échantillonnage incomplet spatialement et temporellement . . . . .	169
5.4.7	Limites . . . . .	171
5.4.8	Pour aller plus loin : algorithmes MCMC en interaction . . . . .	171
5.5	Calcul bayésien approché ou ABC . . . . .	173
5.5.1	Principe . . . . .	173
5.5.2	Limites . . . . .	174
5.6	Conclusion . . . . .	175
5.7	Annexe . . . . .	176
<b>6</b>	<b>Évaluation de modèles</b>	<b>179</b>
6.1	Introduction . . . . .	179
6.2	Prédiction de la teneur en protéines du blé . . . . .	180
6.3	Analyse de sensibilité . . . . .	181
6.4	Qualité des prédictions <i>a posteriori</i> . . . . .	182

6.5	Validation croisée . . . . .	185
6.6	Facteur de Bayes . . . . .	187
6.7	Paradoxe de Lindley-Bartlett . . . . .	189
6.7.1	Test classique d'une hypothèse nulle . . . . .	189
6.7.2	Paradoxe de Lindley . . . . .	190
6.7.3	Paradoxe de Bartlett . . . . .	191
6.7.4	Mises en garde . . . . .	191
6.8	Critères de vraisemblance pénalisée . . . . .	192
6.9	Combinaison de modèles . . . . .	193
6.10	Conclusion . . . . .	195
<b>7</b>	<b>Distributions <i>a priori</i> et élicitation</b>	<b>197</b>
7.1	Introduction . . . . .	197
7.2	Lois <i>a priori</i> conjuguées . . . . .	198
7.3	Lois <i>a priori</i> non informatives . . . . .	201
7.3.1	Lois <i>a priori</i> dispersées . . . . .	204
7.3.2	Lois <i>a priori</i> de Jeffreys . . . . .	207
7.3.3	Lois <i>a priori</i> de référence . . . . .	210
7.4	Élicitation de dires d'expert . . . . .	213
7.4.1	Introduction . . . . .	213
7.4.2	Le questionnement de l'expert . . . . .	214
7.4.3	La traduction des réponses en loi <i>a priori</i> . . . . .	223
<b>II</b>	<b>Cas d'études</b>	<b>229</b>
<b>8</b>	<b>Modèle binomial et <i>Listeria</i></b>	<b>231</b>
8.1	Études complètement comparables . . . . .	232
8.2	Variabilité inter-études . . . . .	237
8.3	Deux sous-populations? . . . . .	240
<b>9</b>	<b>Régression pour la processionnaire du pin</b>	<b>247</b>
9.1	Dispositif expérimental . . . . .	247
9.2	Modèle de régression linéaire multiple . . . . .	250
9.2.1	Modèle . . . . .	250
9.2.2	Écriture matricielle et vraisemblance . . . . .	251
9.2.3	Lois <i>a priori</i> classiques pour l'inférence bayésienne . . . . .	252
9.2.4	Prendre des lois <i>a priori</i> conjuguées simplifie l'inférence bayésienne . . . . .	252
9.2.5	Loi <i>a priori</i> de Zellner pour le modèle linéaire . . . . .	254
9.3	Mise en pratique de la régression linéaire multiple . . . . .	256
9.3.1	Régression linéaire multiple sur la variable sans transformation . . . . .	257
9.3.2	Régression linéaire multiple sur la variable $\log(NbNids)$ . . . . .	258
9.3.3	Vérifier les hypothèses du modèle linéaire . . . . .	258
9.3.4	Estimation des paramètres et tests sur les paramètres . . . . .	263
9.4	Sélection de variables explicatives . . . . .	264
9.4.1	Contexte . . . . .	264

9.4.2	Principe de l'algorithme . . . . .	264
9.4.3	Lois <i>a priori</i> . . . . .	265
9.4.4	Résultats . . . . .	267
<b>10</b>	<b>Modèle hiérarchique normal</b>	<b>271</b>
10.1	Introduction . . . . .	271
10.2	Les données de Potthoff et Roy . . . . .	272
10.2.1	Des mesures longitudinales entachées d'erreurs . . . . .	272
10.2.2	Objectifs de l'analyse statistique des données . . . . .	274
10.3	ANCOVA . . . . .	274
10.3.1	Modèle normal de vraisemblance . . . . .	274
10.3.2	Lois <i>a priori</i> . . . . .	275
10.3.3	Inférence . . . . .	276
10.3.4	Mise en œuvre sous WinBUGS . . . . .	277
10.4	Modèle à effet aléatoire . . . . .	279
10.4.1	Construction . . . . .	279
10.5	Discussion . . . . .	283
10.5.1	Mise en évidence le dimorphisme sexuel . . . . .	283
10.5.2	Traiter facilement les données manquantes . . . . .	283
10.5.3	<i>Borrowing strength</i> . . . . .	284
10.5.4	Conclusions générales . . . . .	284
<b>11</b>	<b>ABC et histoire démographique de bovins</b>	<b>287</b>
11.1	Contexte et objectif . . . . .	287
11.2	Les données microsatellites . . . . .	288
11.3	Implémentation du calcul bayésien approché . . . . .	289
11.3.1	Simulation de jeux de données . . . . .	289
11.3.2	Lois <i>a priori</i> pour les paramètres $\theta$ . . . . .	292
11.3.3	Choix des statistiques descriptives . . . . .	292
11.4	Pré-évaluation des lois <i>a priori</i> . . . . .	292
11.5	Estimation des lois <i>a posteriori</i> approchées . . . . .	293
11.6	Évaluation du modèle . . . . .	296
11.7	Conclusions . . . . .	296
<b>12</b>	<b>Cas d'étude à explorer</b>	<b>299</b>
12.1	La catastrophe de Montroc . . . . .	299
12.1.1	Rappel des faits . . . . .	299
12.1.2	Le problème décisionnel . . . . .	299
12.1.3	Objectif . . . . .	300
12.1.4	Questions . . . . .	300
12.1.5	Solution . . . . .	302
12.2	Invertébrés épibenthiques du Golfe du St Laurent . . . . .	302
12.2.1	Contexte . . . . .	302
12.2.2	Objectif . . . . .	302
12.2.3	Données . . . . .	303
12.2.4	Questions . . . . .	304

12.3	Courbes de croissance de poulets . . . . .	306
12.3.1	Contexte . . . . .	306
12.3.2	Objectif . . . . .	306
12.3.3	Données . . . . .	306
12.3.4	Questions . . . . .	306
12.4	Risques de pollution de l'eau par les nitrates . . . . .	308
12.4.1	Contexte . . . . .	308
12.4.2	Objectif . . . . .	309
12.4.3	Données . . . . .	309
12.4.4	Questions . . . . .	309
12.5	Nombre de copies d'ADN . . . . .	311
12.5.1	Introduction . . . . .	311
12.5.2	Données . . . . .	311
12.5.3	Questions . . . . .	313