

# Réseaux bayésiens avec R

Jean-Baptiste Denis  
Marco Scutari

# Réseaux bayésiens avec R

Vj ku'r ci g'kpvgpvkqpcmf 'ighv'dncpm

Jean-Baptiste Denis et Marco Scutari

# Réseaux bayésiens avec R



**ISBN : 978-2-7598-1198-4**

© **2014, EDP Sciences**, 17, avenue du Hoggar, BP 112, Parc d'activités de Courtabœuf,  
91944 Les Ulis Cedex A

Imprimé en France

Tous droits de traduction, d'adaptation et de reproduction par tous procédés réservés pour tous pays. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (art. L. 122-4, L. 122-5 et L. 335-2 du Code de la propriété intellectuelle). Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre français d'exploitation du droit de copie, 3, rue Hautefeuille, 75006 Paris. Tél. : 01 43 26 95 35.

**Collection Pratique R**  
**dirigée par Pierre-André Cornillon**  
**et Eric Matzner-Løber**

Département MASS  
Université Rennes-2-Haute-Bretagne  
France

**Comité éditorial**

**Eva Cantoni**

Institut de recherche en statistique  
& Département d'économétrie  
Université de Genève  
Suisse

**Pierre Lafaye de Micheaux**

Département de Mathématiques  
et Statistique  
Université de Montréal  
Canada

**François Husson**

Département Sciences de l'ingénieur  
Agrocampus Ouest  
France

**Sébastien Marque**

Directeur Département Biométrie  
Danone Research, Palaiseau  
France

**Déjà paru dans la même collection :**

*Analyse factorielle multiple avec R*

Jérôme Pagès, 2013

**ISBN : 978-2-7598-0963-9** – EDP Sciences

*Psychologie statistique avec R*

Yvonnick Noël, 2013

**ISBN : 978-2-8178-0425-5** – Springer

*Séries temporelles avec R*

Yves Aragon, 2011

**ISBN : 978-2-8178-0208-4** – Springer

*Régression avec R*

Pierre-André Cornillon, Eric Matzner-Løber, 2011

**ISBN : 978-2-8178-0184-1** – Springer

*Méthodes de Monte-Carlo avec R*

Christian P. Robert, George Casella, 2011

**ISBN : 978-2-8178-0181-0** – Springer

Vj ku'r ci g'kpvgpvkqpcmf 'ighv'dncpm

À mes enfants, leurs conjoints et mes petit-enfants.

*Jean-Baptiste*

Pour mon amie Julie Bertrand, à l'occasion de son mariage.

*Marco*



Vj ku'r ci g'kpvgpvkqpcmf 'ighv'dncpm

## PRÉFACE

L'établissement de liens entre les chercheurs semble procéder d'un mécanisme difficilement prédictible. Il fait fi de toute différence de statut : alors que je n'étais que tout jeune doctorant, un professeur émérite de l'université d'Arizona, croisé au hasard d'une conférence, a mis à mal pendant trois ans le budget courrier et photocopie de son département pour me faire parvenir, régulièrement et par-delà l'Atlantique, quantité de reproductions d'articles scientifiques. Cet étrange mécanisme d'association chez les chercheurs ne doit rien non plus à la langue : l'un français, l'autre italien, voilà que les auteurs de cet ouvrage se découvrent par Internet il y a trois ans, publient ensemble d'abord un premier ouvrage en langue anglaise (Scutari & Denis, 2014), puis le reprennent pour le mettre à jour, l'améliorer et produire la version française que nous tenons entre les mains. Cette alchimie ignore enfin superbement le temps et la géographie : alors que nous avons en commun une même cotutelle administrative et quoique seulement une petite dizaine de kilomètres séparent nos laboratoires, le trop prochain départ en retraite du premier auteur me fait réaliser qu'il a fallu finalement beaucoup de temps pour que mon chemin croise le sien et que nous nous attelions à la conduite de quelques chantiers partagés, pour notre plus grand bonheur. Mais grâce au fil de la passion scientifique, des affinités et du hasard, les liens se tissent et le réseau se met en place...

Et c'est aussi de réseau que traite cet ouvrage. Le réseau bayésien réalise un mariage heureux entre la théorie des graphes, un remarquable outil pour communiquer simplement des cartes de relations structurées, et la théorie des probabilités, un outil non moins remarquable de mesure des incertitudes. Comment raisonner de façon quantitative en avenir incertain ? Un petit dessin valant mieux qu'un long discours, voici donc qu'aujourd'hui, un modèle statistique (paramétrique) sera avantageusement représenté par un assemblage de noeuds reliés par des flèches indiquant des relations de cause à effet. Les probabilités conditionnelles sont le ciment de ces assemblages. Une fois le réseau construit, l'inférence bayésienne précise la distribution de probabilité des paramètres (i.e. les causes) à partir de deux sources d'information : les observations (i.e. les effets) et l'expertise. Le cas échéant, sous réserve de disposer d'une base de données suffisante, l'utilisateur pourra également s'appuyer sur des algorithmes de choix de modèles pour apprendre la structure même du réseau. Associés aux techniques de Monte-Carlo, les réseaux bayésiens favorisent le dialogue interdisciplinaire et, par là, des modèles innovants et utiles. Ils interviennent aux interfaces de toutes les disciplines scientifiques où se mêlent savoir qualitatif et données quantitatives.

Nul doute qu'en tant qu'intellectuel non conformiste passionné de philosophie, de théologie, de sciences naturelles, de mécanique et de mathématiques, le révérend Thomas Bayes lui-même n'aurait renié cette union. Associée aux avantages de la représentation graphique, sa célèbre formule, qui permet d'évaluer la pertinence

de ce que l'on croit savoir à l'aune de l'information apportée par une observation, montre ici toute sa puissance opérationnelle. La formule de Bayes – l'équation de l'apprentissage statistique – n'a vraiment pas fini de révolutionner la boîte à outils du chercheur : friande de méthodes de simulation, elle surfe sur les vagues des raz de marée informatiques successifs et porte la promesse d'aider efficacement à désenchevêtrer les réseaux complexes de causes qui apparaissent dans les défis scientifiques en ce début du troisième millénaire. D'ailleurs, nos neurones ne formeraient-ils pas eux-mêmes des réseaux bayésiens ?

Sous l'influence des recherches de Judea Pearl entreprises dans les années 1980 à UCLA, les disciplines issues des technologies de l'information (data mining, systèmes experts, intelligence artificielle, recherche opérationnelle...) ont été nettement plus promptes à percevoir le potentiel d'applications des réseaux bayésiens que la communauté statistique elle-même. Sans doute cette dernière était-elle alors moins soucieuse d'application que de la controverse philosophique qui semble parfois encore sévir quant à l'interprétation concrète de la probabilité, propriété de l'objet ou du sujet. Tant il est vrai que les réseaux bayésiens peuvent être à la fois perçus comme une contribution à l'ingénierie – ce qui nous apprend à faire –, qu'une contribution à la science – ce qui nous apprend à connaître. A la décharge des statisticiens, inquiétante engeance qui a fait du doute sa profession, l'enthousiasme des promoteurs des réseaux bayésiens et la commodité d'emploi de ceux-là ont parfois besoin du garde-fou du sceptique, à la fois sur le concept (Dawid, 2010) et sur les outils de recherche automatique de structure de réseau (Raskutti & Uhler, 2013). L'intelligence artificielle, même sous sa forme la plus subtile, le réseau bayésien, ne garantit en rien que l'on se soit affranchi de toute trace de bêtise très humaine, surtout quand les enjeux touchent au Graal du moderne : la quête de la Connaissance. Peut-être cette zone de friction irritante entre la théorie mathématique et notre besoin d'explications causales n'est-elle rien d'autre qu'un nouvel avatar des contemporains et successeurs de Bayes cherchant à s'appuyer sur son théorème pour prouver l'existence de Dieu.

La plupart des étudiants redoutent les cours de statistique, notamment à cause de la difficulté des concepts probabilistes et du formalisme intimidant du langage mathématique qui les sous-tendent. Une initiation aux réseaux bayésiens, à laquelle cet ouvrage pourrait immédiatement servir de support d'enseignement très motivant, leur permettrait de s'apercevoir qu'il s'agit simplement de se doter de moyens pour prendre en compte l'incertitude dans le raisonnement ! Dans un souci de pédagogie, les auteurs ont ici balisé un chemin très accessible pour progresser dans le maniement des réseaux bayésiens et la découverte de leurs applications les plus intéressantes. Les premiers chapitres, fort didactiques, sont faciles à suivre : s'appuyant à chaque fois sur un exemple qu'ils développent sous le logiciel R, les auteurs mettent en évidence de façon très intuitive et progressive les concepts théoriques utiles, tout en rappelant les briques de base, modèles multinomiaux

et normaux multivariés. Les chapitres suivants présentent des applications de cas d'études réels et de structures plus complexes construites à partir des briques élémentaires et où le caractère opérationnel des réseaux bayésiens apparaît dans toute son élégante efficacité. La lecture terminée, le praticien sera en mesure de comprendre et de se réappropriier le modèle probabiliste : le réseau bayésien lui donne la clé pour ériger, sur des bases rationnelles solides, un lien entre la connaissance du problème et les données expérimentales.

En conclusion, j'adresse mes félicitations les plus vives aux deux auteurs pour cette excellente contribution à la statistique appliquée en général et au raisonnement bayésien en particulier. Gageons qu'étudiants et enseignants, ingénieurs et chercheurs, néophytes ou spécialistes s'approprieront facilement les idées-forces de ce beau livre et y trouveront inspiration et matière à progresser avec plus d'assurance sur leur chemin.

Cambridge, Grande-Bretagne  
mai, 2014

*Eric Parent*  
Professeur à AgroParisTech

Vj ku'r ci g'kpvgpvkqpcmf 'ighv'dncpm

## AVANT-PROPOS

Les applications basées sur les *réseaux bayésiens* se sont multipliées ces dernières années. On les retrouve dans des champs aussi différents que la biologie des systèmes, l'économie, les sciences sociales et l'informatique médicale. Différents aspects et propriétés de cette classe de modèles mathématiques sont cruciaux dans la pratique de ces applications : la possibilité d'apprendre les effets causaux à partir de données observées en sciences sociales alors que le recueil de données expérimentales est le plus souvent impossible ; la représentation graphique simple des dépendances entre variables qui permet une compréhension intuitive et qualitative des réseaux de réactions en sciences biologiques ; la possibilité de construire des modèles hiérarchiques complexes pour les phénomènes qui mettent en jeu de très nombreuses composantes, en associant à chacune d'entre elles les distributions de probabilité appropriées... Et pourtant, toutes ces possibilités sont basées sur un ensemble très réduit de définitions et de propriétés que nous détaillons tout au long de cet ouvrage. Cependant, la manipulation de données multidimensionnelles comportant des valeurs manquantes, les détails fins concernant le raisonnement causal, l'apprentissage sous contraintes spécifiques à des champs d'application particuliers et autres sujets très avancés sont au-delà de la portée de ce livre. Le lecteur trouvera des références utiles dans les sections *Lectures complémentaires*, mais nous pouvons déjà citer Nagarajan *et al.* (2013), Pourret *et al.* (2008) et Pearl (2009) comme étant les principales.

De la même manière, le choix du langage R est d'abord motivé par sa croissante popularité dans de très nombreuses disciplines. Le principal reproche que l'on pourrait lui adresser serait qu'il ne dispose que d'une interface en lignes commande, et semble donc d'un abord rébarbatif et d'un apprentissage ardu pour les experts dont les activités habituelles sont éloignées de la programmation. Cependant, une fois passée la phase d'initiation, R met à disposition de ses utilisateurs un environnement extrêmement polyvalent, à la fois pour l'interprétation de données et pour la mise au point d'approches statistiques originales. La disponibilité de plusieurs paquets couvrant les différents aspects des réseaux bayésiens offerts à la communauté par différents contributeurs permet au lecteur d'explorer le contenu de ce livre sans avoir à réécrire les approches standards décrites dans la littérature scientifique. Parmi ces paquets, nous mettrons en exergue les possibilités étendues de **bnlearn** (réalisé par le second auteur, qui en est déjà à sa version 3.5 au moment où nous écrivons ces lignes) ; il permet au lecteur de se concentrer sur l'étude des réseaux bayésiens sans avoir à entrer dans les particularités de chacun des autres paquets, grâce à des fonctions d'interfaçage. Pour le lecteur intéressé, une présentation plus approfondie des capacités des différents paquets se trouve dans Højsgaard *et al.* (2012) et dans les documentations qui les accompagnent, en particulier les vignettes et articles de référence.

*Réseaux bayésiens avec R : élaboration, manipulation et utilisation en modélisa-*

*tion appliquée* introduit le lecteur aux réseaux bayésiens de manière pragmatique, au travers d'exemples simples mais porteurs de sens, construits et interprétés avec le logiciel de calcul statistique R. De fait, le caractère *pratique* est un point clef de ce livre, et les explications et exemples qui émaillent le texte s'ingénient à détailler chaque étape simplement en proposant toutes les commandes R nécessaires pour pouvoir reproduire la démarche. D'autres excellents ouvrages existent aussi, indiqués dans les sections « Lectures complémentaire » placées en fin de chaque chapitre. Cependant, nous pensons que la manière dont nous présentons les choses est différente, basée sur l'interprétation et l'intuition. Ce livre constitue d'abord une revue introductive sur les réseaux bayésiens et fournit, à celui qui veut mettre en œuvre les réseaux bayésiens, des exemples de l'utilisation de R à adapter à ses propres données, pour des classes très variées de modèles.

Nous espérons aussi que ce livre sera utile à des non-statisticiens travaillant sur des domaines très différents. Évidemment, il n'est pas possible de donner des exemples pour tous les domaines où les réseaux bayésiens sont utiles. C'est pourquoi nous avons tenté d'offrir la meilleure compréhension d'approches génériques et des différentes étapes qui les composent. Pour cela, nous avons limité le nombre d'exemples pour les traiter en profondeur, certains que les experts seront capables de les transposer dans leurs applications respectives. Nous partons toujours des notions les plus simples et augmentons progressivement la complexité des concepts ; la partie théorique est ainsi concentrée dans le 4<sup>e</sup> chapitre et n'est pas nécessaire pour bien suivre les autres chapitres. Nous avons aussi pris garde de bien distinguer la définition des modèles probabilistes et leur estimation à partir de jeux de données, faute de quoi nombre de confusions s'introduisent lors des applications.

*Réseaux bayésiens avec R : élaboration, manipulation et utilisation en modélisation appliquée* peut servir de base pour un enseignement d'un semestre (ou demi-semestre), éventuellement en y ajoutant quelques livres parmi ceux que nous recommandons. Le matériel plus théorique et l'analyse détaillée de deux applications réelles que nous avons menées se trouvent en seconde partie de l'ouvrage pour compléter une bonne compréhension de ce que sont les réseaux bayésiens, d'un point de vue à la fois théorique et appliqué. Le niveau du livre se place entre master et doctorat selon les études suivies par les étudiants. Pour les disciplines comme les mathématiques, les statistiques ou l'informatique, le niveau est plutôt master 1 ; pour les sciences de la vie ou les sciences sociales, le manque de fondement en théorie des probabilités le rend plus adapté au niveau du doctorat dans une optique d'application. Dans le premier cas, c'est sans doute la seconde partie du livre qui sera la plus profitable, alors que dans le second cas, les étudiants auront intérêt à bien approfondir les trois premiers, voire les deux premiers, chapitres avant d'aborder les bases méthodologiques sous-jacentes. Ces derniers pourront aussi bénéficier des annexes dédiées à la théorie des graphes et aux probabilités. Enfin, les solutions des exercices proposés à la fin des chapitres sont données pour la commodité du lecteur qui s'y penchera. Nous avons tenté de varier les contextes

d'application dans les premiers chapitres ; les exemples réels du dernier chapitre devraient motiver les étudiants en leur montrant des applications qui ne sont pas que pédagogiques.

Les compétences requises pour comprendre en profondeur le contenu du livre sont pour la plupart de niveau licence. Cependant, quelques parties plus spécialisées reposent sur des concepts dont l'illustration ne pouvait être entreprise dans l'ouvrage. Les bases de la programmation R ne sont pas rappelées, mais il existe de très bons livres introductifs ; nous suggérons entre beaucoup d'autres Venables & Ripley (2002), Spector (2009), Cornillon *et al.* (2012) et Crawley (2013). Des rappels simples sur la théorie des graphes et les distributions de probabilité sont en annexes, comme références faciles. Des indications de la littérature que nous croyons utiles se trouvent en fin de chaque chapitre et du matériel de soutien sera mis en ligne sur le site [www.bnlearn.com](http://www.bnlearn.com), en particulier les codes R utilisés. Dans ce livre, une commande à lancer sera précédée de `>` et sa typographie, comme celle du résultat obtenu, sera en chasse fixe :

```
> 1 + 2 + 3  
[1] 6
```

L'organisation du livre est la suivante. Les réseaux bayésiens discrets sont les premiers à être décrits (chapitre 1), suivis par les réseaux bayésiens gaussiens (chapitre 2) ; il s'agit des réseaux bayésiens les plus couramment utilisés. Les réseaux hybrides (qui comprennent tout type de variables aléatoires suivant tout type de distributions de probabilité, mélangeant typiquement variables discrètes et continues) sont présentés au chapitre 3. Ces trois premiers chapitres détaillent le processus complet de modélisation à l'aide des réseaux bayésiens, depuis l'apprentissage de structure jusqu'à l'estimation des paramètres. Et toutes ces étapes sont illustrées par de nombreuses lignes de code R. Un traitement concis mais rigoureux des fondements des réseaux bayésiens est donné en chapitre 4 ; il inclut une rapide introduction aux réseaux bayésiens causaux. Pour être plus complet, nous proposons dans le chapitre 5 un tour d'horizon des logiciels disponibles, tant liés à R qu'indépendants de R. Finalement, deux exemples réels sont analysés au chapitre 6. Le premier examine diverses approches de modélisation graphique pour prédire la proportion de muscles, graisse et os de différents segments corporels à partir d'une petite base de données. Le second reprend l'étude de référence du réseau causal de protéines-signal publiée dans la revue *Science* par Sachs *et al.* (2005).

Pour finir, nous voudrions signaler notre reconnaissance aux collègues et aux amis qui nous ont aidés : en nous suggérant la démarche, et en nous soutenant dans la conception et l'écriture de ce livre et de sa version anglaise *Bayesian Networks: with Examples in R*. Plus particulièrement, nous voudrions remercier les éditeurs scientifiques d'EDP Sciences, Pierre-André Cornillon, Eric Matzner-Løber et François Husson. Initialement, c'est ce dernier qui a sollicité le premier auteur pour



rédigé un livre en français sur les réseaux bayésiens. Ne se sentant pas suffisamment assuré sur tous les aspects, celui-là a cherché (et trouvé) un bon co-auteur pour mener à bien le projet. Ce dernier ne maîtrisant pas suffisamment la langue de Molière, une version anglaise a d'abord été rédigée. Elle a finalement été proposée et acceptée en parallèle (Scutari & Denis, 2014). Il existe cependant quelques petits ajouts pour répondre aux critiques des éditeurs scientifiques de EDP Sciences, en particulier un bref chapitre de conclusion et l'indexation de la quasi-totalité des fonctions R utilisées dans les portions de code illustratives. Merci à Eric Parent qui nous a rédigé une fort intéressante préface. Cette version a aussi bénéficié de la relecture attentive de François Husson, Pierre-André Cornillon, Eric Matzner-Løber et Jeanie Denis qui ont grandement diminué le nombre de coquilles présentes dans la version initiale ; nous les en remercions.

Jouy-en-Josas, France  
Londres, Grande-Bretagne  
mai 2014

*Jean-Baptiste Denis*  
*Marco Scutari*

# Table des matières

<b>Préface</b>	<b>xi</b>
<b>Avant-propos</b>	<b>xiii</b>
<b>1 Cas discret : les réseaux bayésiens multinomiaux</b>	<b>1</b>
1.1 Exemple introductif . . . . .	1
1.2 Représentation graphique . . . . .	2
1.3 Représentation probabiliste . . . . .	7
1.4 Estimation des tables de probabilités . . . . .	12
1.5 Apprentissage de la structure du graphe . . . . .	15
1.5.1 Tests d'indépendance conditionnelle . . . . .	16
1.5.2 Scores de réseaux . . . . .	19
1.6 Utilisation d'un RB discret . . . . .	22
1.6.1 Utilisation du DAG . . . . .	22
1.6.2 Utilisation des distributions de probabilités . . . . .	25
1.7 Diagrammes . . . . .	31
1.7.1 Tracer le DAG d'un RB . . . . .	31
1.7.2 Représenter des probabilités conditionnelles discrètes . . . . .	33
1.8 Lectures complémentaires . . . . .	35
<b>2 Cas continu : les réseaux bayésiens gaussiens</b>	<b>37</b>
2.1 Exemple introductif . . . . .	37
2.2 Représentation graphique . . . . .	39
2.3 Représentation probabiliste . . . . .	42
2.4 Estimation des paramètres . . . . .	46
2.5 Apprentissage de la structure du DAG . . . . .	50
2.5.1 Tests d'indépendance conditionnelle . . . . .	50
2.5.2 Scores de réseaux . . . . .	53
2.6 Utilisation d'un RBG . . . . .	54
2.7 Diagrammes . . . . .	59
2.7.1 Tracé du DAG d'un RB . . . . .	59
2.7.2 Représenter des distributions conditionnelles . . . . .	62

2.8	Propriétés supplémentaires . . . . .	65
2.9	Lectures complémentaires . . . . .	66
<b>3</b>	<b>Réseaux bayésiens hybrides</b>	<b>67</b>
3.1	Exemple introductif . . . . .	68
3.1.1	Mélange de variables discrètes et continues . . . . .	69
3.1.2	Discrétisation de variables continues . . . . .	72
3.1.3	Utilisation d'autres distributions de probabilité . . . . .	74
3.2	Exemple d'une récolte parasitée . . . . .	75
3.2.1	Modélisation . . . . .	77
3.2.2	Exploration . . . . .	78
3.3	A propos des logiciels BUGS . . . . .	84
3.4	Lectures complémentaires . . . . .	85
<b>4</b>	<b>Théorie des réseaux bayésiens et algorithmes associés</b>	<b>87</b>
4.1	Indépendance conditionnelle et séparation graphique . . . . .	87
4.2	Les réseaux bayésiens . . . . .	89
4.3	Couvertures de Markov . . . . .	94
4.4	Graphes moralisés . . . . .	97
4.5	Apprentissage d'un réseau bayésien . . . . .	99
4.5.1	Apprentissage de la structure . . . . .	103
4.5.2	Estimation des paramètres . . . . .	115
4.6	Inférence par réseaux bayésiens . . . . .	116
4.6.1	Raisonnement probabiliste et évidence . . . . .	116
4.6.2	Mise à jour de la connaissance . . . . .	118
4.7	Réseaux bayésiens causaux . . . . .	124
4.8	Lectures complémentaires . . . . .	128
<b>5</b>	<b>Logiciels pour réseaux bayésiens</b>	<b>131</b>
5.1	Revue des paquets R . . . . .	131
5.1.1	Le paquet <b>deal</b> . . . . .	133
5.1.2	Le paquet <b>catnet</b> . . . . .	135
5.1.3	Le paquet <b>pcalg</b> . . . . .	137
5.2	Les logiciels BUGS . . . . .	139
5.2.1	Distributions de probabilités . . . . .	139
5.2.2	Dépendances complexes . . . . .	140
5.2.3	Inférence par échantillonnage MCMC . . . . .	140
5.3	Autres logiciels . . . . .	142
5.3.1	BayesiaLab . . . . .	142
5.3.2	Hugin . . . . .	142
5.3.3	GeNIe . . . . .	143
<b>6</b>	<b>Réseaux bayésiens en grandeur réelle</b>	<b>145</b>
6.1	Prédiction de la Composition Corporelle . . . . .	145

6.1.1	Approche par régression et utilité des RB . . . . .	147
6.1.2	Approche prédictive . . . . .	148
6.1.3	Recherche de RB candidats . . . . .	152
6.2	Apprentissage d'un réseau de protéines-signal . . . . .	159
6.2.1	Un réseau bayésien gaussien . . . . .	161
6.2.2	Discrétisation de l'expression des gènes . . . . .	164
6.2.3	Moyennage de modèles . . . . .	165
6.2.4	Choix du seuil de significativité . . . . .	171
6.2.5	Traitement de données d'intervention . . . . .	173
6.2.6	Requêtes sur le réseau . . . . .	177
6.3	Lectures complémentaires . . . . .	180
<b>Conclusion</b>		<b>181</b>
<b>A Théorie des graphes</b>		<b>187</b>
A.1	Graphes, nœuds et arcs . . . . .	187
A.2	La structure d'un graphe . . . . .	189
A.3	Lectures complémentaires . . . . .	190
<b>B Distributions de probabilités</b>		<b>191</b>
B.1	Généralités . . . . .	191
B.2	Distributions marginales et conditionnelles . . . . .	193
B.3	Distributions discrètes . . . . .	194
B.3.1	Distribution binomiale . . . . .	194
B.3.2	Distribution multinomiale . . . . .	195
B.3.3	Autres distributions courantes . . . . .	195
B.4	Distributions continues . . . . .	196
B.4.1	Distribution normale . . . . .	196
B.4.2	Distribution normale multivariable . . . . .	197
B.4.3	Autres distributions courantes . . . . .	198
B.5	Distributions conjuguées . . . . .	200
B.6	Lectures complémentaires . . . . .	200
<b>C Glossaire</b>		<b>201</b>
<b>D Solutions</b>		<b>207</b>
D.1	Exercices du chapitre 1 . . . . .	207
D.2	Exercices du chapitre 2 . . . . .	212
D.3	Exercices du chapitre 3 . . . . .	217
D.4	Exercices du chapitre 4 . . . . .	221
D.5	Exercices du chapitre 5 . . . . .	225
<b>Bibliographie</b>		<b>229</b>
<b>Index</b>		<b>236</b>

- densité, 193
- distribution cumulée, 174, 194
- fonction R, 139
  - all.equal, 6, 52, 110, 153, 170, 173
  - anova, 151, 160
  - arc.set, 6
  - arc.strength, 18, 21
  - arcs, 6, 11
  - as.graphNEL, 60, 155
  - autosearch, 136
  - averaged.network, 168–171, 173, 175, 178
  - barchart, 34
  - bn.fit, 13, 15, 47, 62, 154, 179, 180
  - bn.fit.barchart, 34
  - bn.fit.qqplot, 62
  - bnfit2nbn, 45
  - boot.strength, 167
  - capture.output, 136
  - cbind, 153–155
  - cextend, 172
  - children, 98
  - chisq.test, 28
  - ci.test, 17, 18, 51, 111
  - cnFindBIC, 138, 171
  - cnMatEdges, 139, 171
  - cnSamples, 139
  - cnSearchOrder, 171
  - cnSearchSA, 138
  - coda.samples, 73, 80
  - coef, 11, 49
  - colMeans, 151, 154
  - colnames, 151, 154
  - colSums, 151, 154
  - compare, 136, 164, 175, 178
  - compile, 26, 179, 180
  - condi4joint, 63, 154
  - confint, 49
  - cor, 51
  - cor2pcor, 51
  - cpdag, 94, 114, 136, 172, 175
  - cpdist, 30, 58, 63
  - cpquery, 29, 30, 59, 124, 125
  - custom.fit, 9, 44, 123
  - custom.strength, 169, 171, 178
  - directed.arcs, 163
  - discretize, 128, 138, 166
  - dnorm, 73
  - drop.arc, 180
  - dsep, 23, 24, 28, 40, 41, 63, 96
  - E, 60, 155
  - edge.render, 32
  - empty.graph, 3, 139
  - factor, 128, 135
  - fci, 140
  - gema2mn, 45, 154
  - graphviz.plot, 31, 32, 39, 78
  - gs, 52, 53, 110, 111
  - hc, 21, 113, 114, 116, 128, 152, 153, 169
  - iamb, 52, 110
  - igraph.formula, 59
  - igraph.from.graphNEL, 60, 155
  - igraph.options, 59
  - inter.iamb, 163
  - jags.model, 72, 80
  - jointprior, 136
  - lapply, 169, 171, 178
  - learn, 136
  - levels, 179
  - list, 176
  - lm, 47, 49, 151
  - load, 170
  - logLik, 125
  - mb, 97–99
  - mmhc, 117
  - model2network, 9, 39, 78, 93, 110, 114, 123, 136, 164
  - modelstring, 5, 20, 21, 110, 113, 114, 128, 136, 153
  - moral, 100
  - mutilated, 124
  - names, 169, 171, 178, 179, 181
  - narcs, 163, 172
  - nbn2gema, 45, 154
  - network, 135
  - nnorm, 57

- nodeRender, 32
  - nodes, 6, 9, 40
  - nparams, 11, 20
  - nrow, 154, 171
  - outer, 151
  - parents, 98
  - paste, 152, 153, 164
  - path, 23, 42
  - pc, 140
  - penalized, 48
  - plot, 60, 63, 155
  - print8gema, 55
  - print8mn, 45, 55
  - print8nbn, 54
  - prop.table, 13, 30
  - querygrain, 26, 27, 180, 181
  - random.graph, 20, 169, 178
  - rbn, 46, 57, 123, 125
  - read.table, 163, 166, 175
  - renderGraph, 32
  - rsmax2, 117
  - sapply, 178
  - score, 19–21, 54, 113, 170, 172
  - set.arc, 4, 95
  - set.edge, 100
  - set.seed, 166
  - setEvidence, 26, 27, 179, 180
  - skeleton, 93, 101, 168, 170
  - substr, 152, 153
  - tabu, 176, 178
  - tiers2blacklist, 176
  - unlist, 136, 154, 164, 175, 178
  - update, 73, 80
  - V, 60
  - vector, 171
  - vstructs, 93, 100
  - which, 155
  - which.max, 181
- graphe, 189
- acyclique, 191
  - acyclique orienté, 2, 89
  - moralisé, 99–101
  - non orienté, 189
  - ordre, 191
  - orienté, 189
  - partiellement orienté, 189
  - vide, 3
- inférence, 118–120
- approchée, 28–31, 56–59, 80–86, 122–126
  - exacte, 25–28, 54–56, 121–122, 179–182
- logiciel
- BayesiaLab, 144
  - BUGS, 86–87, 141–144
  - GeNIe, 145
  - Hugin, 144–145
  - JAGS, 77–87, 141–144
  - paquets R, *voir* paquets R
- maps, 90
- matrice de covariance, *voir* covariance
- morphisme, 90
- moyenne, *voir* espérance
- nœud, 2, 189
- ancêtre, 192
  - descendant, 192
  - enfant, 2, 192
  - parent, 2, 192
  - voisin, 192
- ordre des nœuds, *voir* graphe ordre
- ordre topologique, *voir* graphe ordre
- paquets R, 133
- bnlearn**, 3–25, 39–45, 47–54, 56–59, 62–65, 93–95, 97–101, 110–111, 113–117, 123–130, 133, 152–153, 163–164, 166–170, 172–173, 175–179
  - BRugs**, 86
  - catnet**, 134, 137–139, 170–172
  - corpcor**, 50, 51
  - deal**, 134–137
  - gRain**, 25–28, 134, 179–182