ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΜΠΕΥΖΙΑΝΑ ΔΙΚΤΥΑ: ΜΕΘΟΔΟΙ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

Γεώργιος Ε. Κουνιός

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς

Σεπτέμβριος 2024

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. 20/13 - 7 - 2022 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Ηλιόπουλος Γεώργιος, Καθηγητής (Επιβλέπων)
- Ντζούφρας Ιωάννης, Καθηγητής
- Ρακιτζής Αθανάσιος, Επίκουρος Καθηγητής

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN

APPLIED STATISTICS

BAYESIAN NETWORKS: METHODS AND APPLICATIONS

By

George E.Kounios

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece
September 2024

Στην οικογένειά μου

Ευχαριστίες

Σε αυτή την ενότητα θα ήθελα να ευχαριστήσω τόσο τα άτομα που με βοήθησαν στην εκπόνηση της παρούσας διπλωματικής εργασίας, όσο και αυτούς που με βοήθησαν κατά τη διάρκεια των σπουδών μου στο Π.Μ.Σ. Εφαρμοσμένη Στατιστική. Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, κύριο Γεώργιο Ηλιόπουλο, για την ανάθεση του θέματος και την πολύτιμη συνεισφορά του ώστε να λάβει η διπλωματική μου εργασία την παρούσα μορφή της. Στη συνέχεια θα ήθελα να ευχαριστήσω, τα μέλη της τριμελούς συμβουλευτικής επιτροπής, τους Καθηγητή κύριο Ιωάννη Ντζούφρα και Επίκουρο Καθηγητή κύριο Αθανάσιο Ρακιντζή, για το χρόνο που αφιέρωσαν στη διόρθωση της διπλωματικής εργασίας. Ευχαριστώ επίσης, τους φίλους και συμφοιτητές μου Σωτήριο Τρούσα , Ιωάννα Χαμπεσή , Ιωάννα Μπραουδάκη, Αγγελική Μπούτση, Αγάπη Καλάρη και Δημήτρη Αναγνώστη για τις εποικοδομητικές συζητήσεις που είχαμε κατά τη διάρκεια των σπουδών μας στο Πανεπιστήμιο Πειραιώς. Τέλος, ευχαριστώ από καρδιάς την οικογένειά μου που βρίσκεται πάντα δίπλα μου και με στηρίζει σε κάθε φάση της ζωής μου και τους οφείλω πολλά από αυτά που έχω πετύχει μέχρι σήμερα.

Περίληψη

Τα Μπεϋζιανά δίκτυα (BN) προσφέρουν ένα ισχυρό πλαίσιο για τη μοντελοποίηση και την εξαγωγή συμπερασμάτων για πιθανοτικές σχέσεις, οι οποίες είναι διαδεδομένες σε διάφορες βιομηχανίες και επιστημονικούς τομείς του πραγματικού κόσμου. Η παρούσα διπλωματική εργασία εμβαθύνει στη θεωρία πίσω από τα δίκτυα Bayes, συζητώντας μεθοδολογίες τόσο για τη δομή όσο και για την εκτίμηση των παραμέτρων. Επιπλέον, διερευνά εκτενώς τις λειτουργίες του πακέτου bnlearn, ενός κύριου εργαλείου για την εξαγωγή συμπερασμάτων σε δίκτυα Bayes στην R, μέσω πρακτικών εφαρμογών.

Επιπλέον, η διπλωματική εργασία μελέτα κάποιες εφαρμογές από τις άλλες επιστήμες, τις οποίες μπορεί να βοηθήσει για την καλύτερη επεξεργασία αποτελεσμάτων με χρήση του κώδικα στην R. Διερευνά έτσι εναλλακτικές μεθόδους αξιολόγησης για να μετριάσει τον κίνδυνο επιλογής μιας δομής δικτύου με βάση αποκλειστικά τις συναρτήσεις βαθμολογίας, οι οποίες μπορεί να ευνοούν υπερβολικά πολύπλοκες δομές.

Abstract

Bayesian networks (BNs) offer a powerful framework for modelling and inferring probabilistic relationships, which are prevalent in various industries and scientific fields in the real world. This thesis delves into the theory behind Bayesian networks, discussing methodologies for both structure and parameter estimation. In addition, it extensively explores the features of the bnlearn package, a main tool for inference in Bayesian networks in R, through practical applications.

In addition, the thesis studies some applications from other sciences which can help to better process results using code in R, thus exploring alternative evaluation methods to mitigate the risk of selecting a network structure based solely on score functions, which may favour overly complex structures.

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κατάλογος Συντομογραφιών

DGMs	Κατευθυνόμενα γραφικά μοντέλα
DAGs	Directed Acyclic Graphs
BNs	Μπεϋζιανά Δίκτυα
BDe	Bayesian Dirichlet equivalent uniform
BIC	Bayesian Information Criterion
IC	Inductive Causation
GS	Grow-Shrink
IAMB	Incremental Association Markov Blanket algorithm
Fast-IAMB	Fast Incremental Association Markov Blanket algorithm
Inter-IAMB	Interleaved Incremental Association Markov Blanket algorithm
GM	Graphical Model
CI	Conditional Independence

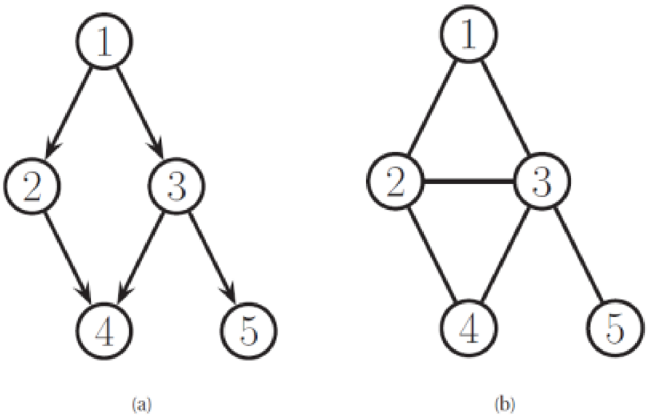
ΚΕΦΑΛΑΙΟ 1

Εισαγωγή στα Μπεϋζιανά δίκτυα

1.1 Γραφικά Μοντέλα

Η γραφική αναπαράσταση αναφέρεται στη χρήση διαισθητικών γραφημάτων για την καθαρή απεικόνιση και απλοποίηση σύνθετων συνόλων δεδομένων. Τα δεδομένα εισάγονται στη γραφική αναπαράσταση των δεδομένων και στη συνέχεια αντιπροσωπεύονται από διάφορα σύμβολα, όπως γραμμές σε ένα γραμμικό γράφημα, γραμμές σε ένα ραβδογράφημα, ή φέτες σε ένα γράφημα πίτας, από τα οποία οι χρήστες μπορούν να τα κατανοήσουν καλύτερα συγκριτικά με την αριθμητική ανάλυση. Τα αναπαραστατικά γραφικά μπορούν να απεικονίσουν πολύ γρήγορα γενικές συμπεριφορές και τονίζουν φαινόμενα, ανωμαλίες και σχέσεις μεταξύ σημείων δεδομένων που διαφορετικά μπορεί να παραλειφθούν και μπορεί να συμβάλουν σε προβλέψεις και καλύτερα δεδομένα για καθοδηγούμενες αποφάσεις. Οι τύποι των αναπαραστατικών γραφικών που χρησιμοποιούνται θα εξαρτηθούν από τον τύπο δεδομένων που διερευνώνται.

Για να ορίσουμε τη Γραφική αναπαράσταση των Μπεϋζιανών Δικτύων, πρέπει να κατανοήσουμε τη βασική έννοια της Θεωρίας Γραφημάτων. Ένα Πιθανολογικό Γραφικό Μοντέλο (GraphicalModel (GM)) παρέχει έναν μέσο για την αναπαράσταση μιας κοινής κατανομής πιθανοτήτων υποθέτοντας ανεξαρτησία υπό όρους (ConditionalIndependence (CI)). Πιο συγκεκριμένα, οι κόμβοι των γραφημάτων αντιπροσωπεύουν τυχαίες μεταβλητές και οι υποθέσεις ανεξαρτησίας υπό όρους (ConditionalIndependence(CI)) αντιπροσωπεύονται από ακμές. Υπάρχουν διάφοροι τύποι γραφικών μοντέλων που σχετίζονται με κατευθυνόμενα, μη κατευθυνόμενα γραφήματα ή συνδυασμό από τα δύο. Παρακάτω απεικονίζονται δύο κατευθυνομενοι γράφοι.



Σχήμα 1.1: (α) Ένα απλό DAG σε 5 κόμβους, αριθμημένους με τοπολογική σειρά. Ο κόμβος 1 είναι η ρίζα, οι κόμβοι 4 και 5 είναι τα φύλλα. (β) Ένα απλό μη κατευθυνόμενο γράφημα, με τα ακόλουθα μέγιστες κλίκες: {1, 2, 3}, {2, 3, 4}, {3, 5} (Murphy, K. P. (2012)[15]).

Ένα Κατευθυνόμενο Γραφικό Μοντέλο ή DGM είναι ένα Γραφικό Μοντέλο του οποίου ο γράφος είναι ένα DAG. Οι DirectedAcyclicGraphs (DAGs) είναι πολύ σημαντικοί στη σύγχρονη αιτιώδη συμπερασματολογία. Τα DAGs ορίζουν σχέσεις μεταξύ μετρήσεων που έχουν ληφθεί σε διαφορετικούς διακριτούς χρόνους. Αυτά είναι ως επί το πλείστων γνωστά ως Μπεϋζιανά Δίκτυα. Τα Μπεϋζιανά Δίκτυα (BNs) ονομάζονται επίσης δίκτυα πεποιθήσεων και μας δίνουν έναν τρόπο να αναπαραστήσουμε κατανομές πιθανοτήτων. Μοντέλα όπως αυτά ονομάζονται επίσης αιτιώδη δίκτυα, επειδή τα κατευθυνόμενα βέλη μερικές φορές ερμηνεύουν αιτιώδεις σχέσεις.

1.2 Κατευθυνόμενα γραφικά μοντέλα(DGMs)

Προκειμένου να κατανοήσουμε καλύτερα αυτά που αναφέρουμε στη συγκεκριμένη παράγραφο καλό είναι να εξηγήσουμε το εξής:
Ένα γράφημα G = (V, A) αποτελείται από ένα σύνολο κόμβων ή κορυφών,
U = {1, ... , U}, και ένα σύνολο A = (u, v) : u, v∈V τα οποία ονομάζονται τόξα, σύνδεσμοι ή ακμές. Το γράφημα μπορεί να αναπαρασταθεί από τον πίνακα γειτνίασής του, στον οποίο γράφουμε G(u, v) = 1 για να δηλώσουμε ότι (u, v) ∈A, δηλαδή αν το u → v είναι τόξο στο γράφημα. Αν G(u, v) = 1 και αν και μόνο αν G(u, v) = 1, ο γράφος ονομάζεται κατευθυνόμενος, διαφορετικά ονομάζεται μη

κατευθυνόμενος. Τις περισσότερες φορές υποθέτουμε G(u, u) = 0, πράγμα που σημαίνει ότι δεν υπάρχουν αυτοβρόχοι. Οι ακόλουθοι όροι χρησιμοποιούνται συχνά στην πράξη.

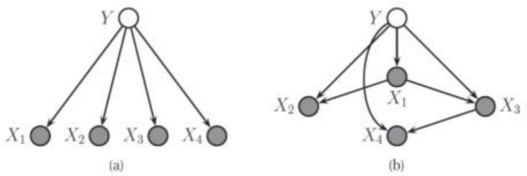
◦ Γονέας: Για κατευθυνόμενους γράφους, οι γονείς ενός κόμβου είναι το σύνολο όλων των κόμβων που τροφοδοτούνται σε αυτόν: par(v) := {u : G(u, v) = 1}.

◦ Παιδί: Για κατευθυνόμενους γράφους, τα παιδιά ενός κόμβου είναι το σύνολο όλων των κόμβων που τροφοδοτούνται από αυτόν: chi(u) := {v : G(u, v) = 1}.

Η βασική ιδιότητα των DGMs είναι ότι οι κόμβοι μπορούν να ταξινομηθούν έτσι ώστε τα παιδιά να ακολουθούν τους γονείς, πράγμα που είναι γνωστό ως τοπολογική ταξινόμηση, και μπορεί να κατασκευαστεί από οποιοδήποτε κατευθυνόμενο γραφικό μοντέλο. Εάν έχουμε μια σειρά όπως αυτή, μπορούμε να ορίσουμε την ιδιότητα του διατεταγμένου Markov (περισσότερα για την ιδιότητα Markov θα συζητηθεί αργότερα) η οποία υποθέτει ότι ένας κόμβος εξαρτάται μόνο από τους προηγούμενους γονείς του, όχι από όλους τους προηγούμενους στη διάταξη. Πιο επίσημα αυτό μπορεί να γραφτεί ως εξής:

$$x_u \perp x_{pre(u)\setminus pa(u)} \mid x_{pa(u)} \qquad (1.1)$$

όπου pa(u) είναι κόμβοι u γονείς και pre(u) είναι κόμβοι u προκάτοχοι της σειράς. Αυτή είναι μια φυσική γενίκευση της ιδιότητας Markov πρώτης τάξης από αλυσίδες στη γενική DGMs (Murphy, K. P. (2012)[15]).



Σχήμα 1.2: (α) Ένας ταξινομητής naiveBayes που αναπαρίσταται ως κατευθυνόμενο γραφικό μοντέλο. Υποθέτοντας ότι υπάρχουν F = 4 χαρακτηριστικά, για λόγους απλότητας. Οι σκιασμένοι κόμβοι είναι οι παρατηρούμενοι, οι μη σκιασμένοι κόμβοι είναι κρυφοί. (β) Ταξινομητής naiveBayes με δενδροειδείς επαυξήσεις για F = 4 χαρακτηριστικά. Γενικά, η τοπολογία του δέντρου μπορεί να αλλάξει ανάλογα με την τιμή του y (Murphy, K. P. (2012)[15])

Ο κατευθυνόμενος ακυκλικός γράφος στο Σχήμα 1.1(α), για παράδειγμα, ορίζει την επόμενη ένωση κατανομή:

$$
\begin{aligned}
p(x_{1:5}) &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, \cancel{x_2})p(x_4 \mid \cancel{x_1}, x_2, x_3)p(x_5 \mid \cancel{x_1}, \cancel{x_2}, x_3, \cancel{x_4}) \\
&= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2, x_3)p(x_5 \mid x_3)
\end{aligned} \tag{1.2}
$$

Γενικά, έχουμε

$$p(x_{1:V} \mid G) = \prod_{u=1}^{V} p(x_u \mid x_{pa(u)}) \tag{1.3}$$

όπου κάθε όρος $p(x_u \mid x_{pa(u)})$ κωδικοποιεί μια υπό συνθήκη κατανομή πιθανότητας (CPD).Γράφουμε την κατανομή ως $I(G) \subseteq I(p), p(x|G)$ για να δείξουμε ότι αυτή η εξίσωση ισχύει μόνο αν οι υποθέσεις ανεξαρτησίας υπό όρους που ορίζονται στο κατευθυνόμενο γραφικό μοντέλο G είναι δεξιά (Murphy, K. P. (2012)[15]).

Ο Μη κατευθηνόμενος γράφος στο Σχημα 1.1 (β) είναι διπλής κατεύθυνσης γραφικός σχεδιασμός. Επιτρέπει τις ακμές να συνδέουν τα ζεύγη κόμβων χωρίς συγκεκριμένη κατεύθυνση. Αυτό σημαίνει ότι, σε αντίθεση με τον κατευθυνόμενο γράφο Σχημα 1.1 (α), όπου οι ακμές έχουν συγκεκριμένη κατεύθυνση (δηλαδη από έναν κόμβο προς έναν άλλο), οι ακμές στον μη κατευθυνόμενο γράφο είναι αμφίδρομες και δεν υπάρχει προσανατολισμός.

Με άλλα λόγια, στον κατευθυνόμενο γράφο (a), η σύνδεση 1 → 2 υποδηλώνει μια μοναδική κατεύθυνση από τον κόμβο 1 προς τον κόμβο 2, ενώ στον μη κατευθυνόμενο γράφο (b), η σύνδεση μεταξύ 2 και 3 (2—3) μπορεί να θεωρηθεί αμφίδρομη, υποδηλώνοντας ότι υπάρχει αλληλεπίδραση και προς τις δύο κατευθύνσεις χωρίς συγκεκριμένη προτεραιότητα.

Ο μη κατευθυνόμενος γράφος (b) αντιπροσωπεύει ένα **Markov Random Field (MRF)**. Σε αυτό το μοντέλο, οι κόμβοι επίσης αντιπροσωπεύουν τυχαίες μεταβλητές, αλλά οι ακμές υποδεικνύουν ότι υπάρχει μια **άμεση εξάρτηση** μεταξύ των ζευγών των κόμβων.

Στο MRF, οι πιθανότητες αντιπροσωπεύονται μέσω των **κλίκων** του γράφου. Η πιθανότητα κάθε συνόλου μεταβλητών που σχετίζονται μεταξύ τους υπολογίζεται μέσω μιας **συνάρτησης πιθανοφάνειας** φ, και η συνολική κατανομή πιθανότητας δίνεται από:

$$P\left(X_{1,}X_2,\ldots,X_n\right)=\frac{1}{Z}\prod_{C\in\kappa\lambda\acute{\iota}\kappa\varepsilon\varsigma}\varphi_C\left(X_C\right)$$

όπου Z είναι ένας **κανονικοποιητικός παράγοντας** (partition function) και C οι μέγιστες κλίκες του γράφου.

1.3 Βασικά στοιχεία των Μπεϋζιανών Δικτυών

Τα Μπεϋζιανά Δίκτυα (BayesianNetworks - BNs) είναι γνωστά ως Γραφικά Μοντέλα τα οποία χρησιμοποιούνται για συμπερασματολογία υπό αβεβαιότητα και με τα οποία μπορούμε να αναπαραστήσουμε και να αιτιολογήσουμε αιτιακές σχέσεις μεταξύ μεταβλητών. Αυτές οι άμεσες συνδέσεις μπορεί μερικές φορές να είναι συνδέσεις εξάρτησης. Επιπλέον, τα Μπεϋζιανά Δίκτυα (BNs) προσδιορίζουν την ποσοτική ισχύ των συνδέσεων μεταξύ των μεταβλητών, έτσι ώστε όταν γίνονται διαθέσιμες νέες πληροφορίες να ενημερώνονται αυτόματα οι πιθανολογικές πεποιθήσεις σχετικά με αυτές. Οι κόμβοι στα Μπεϋζιανά Δίκτυα ορίζουν ένα σύνολο τυχαίων μεταβλητών από το θέμα για το οποίο θέλουμε να επιχειρηματολογήσουμε. Άμεσες εξαρτήσεις μεταξύ των μεταβλητών αναπαρίστανται με κατευθυνόμενες ακμές που συνδέουν ζεύγη κόμβων. Υποθέτοντας ότι έχουμε διακριτές μεταβλητές, η ισχύς της σχέσης μεταξύ των μεταβλητών αξιολογείται από κατανομές πιθανότητας υπό συνθήκη που συνδέεται με κάθε κόμβο.

1.4 Ένα ενδεικτικό παράδειγμα: Κόμβοι και τιμές

Το απλό παράδειγμα ιατρικής διάγνωσης που ακολουθεί θα καταδείξει τη χρησιμότητα και τη δυνατότητα εφαρμογής ενός BayesianNetwork μαζί με τα κύρια χαρακτηριστικά του BayesianNetwork.

Παράδειγμα προβλήματος: Καρκίνος του πνεύμονα (Korb and Nicolson (2010) [16]).

Ένας ασθενής υποφέρει από δύσπνοια και επισκέπτεται τον γιατρό επειδή ανησυχεί ότι μπορεί να έχει καρκίνο του πνεύμονα. Ο γιατρός γνωρίζει ότι υπάρχουν αριθμοί άλλων πιθανών αιτιών, όπως η φυματίωση και η βρογχίτιδα, καθώς και ο καρκίνος του πνεύμονα. Γνωρίζει επίσης ότι υπάρχουν και άλλες σχετικές πληροφορίες που περιλαμβάνουν το αν ο ασθενής είναι καπνιστής (αυξάνοντας τις πιθανότητες καρκίνου και βρογχίτιδας) και σε τι είδους ρύπανση του αέρα έχει εκτεθεί ο ασθενής. Μια θετική ακτινογραφία θα υποδείκνυε είτε καρκίνο του πνεύμονα είτε φυματίωση.

Αρχικά, πρέπει να προσδιοριστούν οι μεταβλητές ενδιαφέροντος. Αυτό σχετίζεται με την απάντηση στο εξής ερώτημα: «Με ποιους κόμβους μπορούμε να αναπαραστήσουμε τις μεταβλητές και ποιες τιμές μπορούν ενδεχομένως να πάρουν;» Εξετάζουμε μόνο διακριτές μεταβλητές, που σημαίνει ότι οι κόμβοι λαμβάνουν διακριτές τιμές και πρέπει να είναι αμοιβαία αποκλεισμένοι και εξαντλητικοί.

Αυτό σημαίνει, ότι η μεταβλητή πρέπει να παίρνει ακριβώς μία από αυτές τις τιμές σε μια φορά. Οι γνωστοί τύποι διακριτών κόμβων περιλαμβάνουν:

◦ Κόμβοι Boolean, οι οποίοι παίρνουν δυαδικές τιμές true (T) και false (F) και αντιπροσωπεύουν προτάσεις. Η πρόταση ότι ένας ασθενής έχει καρκίνο θα αναπαρίσταται από τον κόμβο Cancer (Korb and Nicolson (2010) [16]).

◦ Διατεταγμένες τιμές. Για παράδειγμα, ο κόμβος Pollution θα μπορούσε να αναπαριστά τη ρύπανση ενός ασθενούς και θα μπορούσε να πάρει τις ακόλουθες τιμές {χαμηλή, μεσαία, υψηλή}. (Korb and Nicolson (2010) [16]).

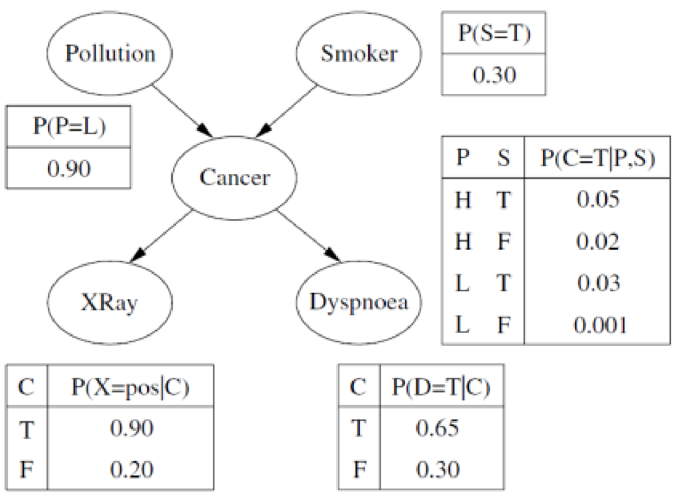
◦ Ολοκληρωτικές τιμές. Ας υποθέσουμε ότι θα θέλαμε να μοντελοποιήσουμε έναν κόμβο με το όνομα Age, ο οποίος θα αντιπροσώπευε την ηλικία ενός ασθενούς και θα είχε πιθανές τιμές από 1 έως 100 (Korb and Nicolson (2010) [16]).

Ακόμη και σε αυτό το πρώιμο επίπεδο, πρέπει να γίνουν επιλογές. Για παράδειγμα, αντί να αναπαραστήσουμε την ακριβή ηλικία ενός ασθενούς, θα μπορούσαμε να ομαδοποιήσουμε τους ασθενείς σε διάφορες ηλικιακές ομάδες , όπως {μωρό, παιδί, έφηβος, νέος, μεσήλικας, ηλικιωμένος}. Το κόλπο είναι να επιλέξουμε τιμές που αναπαριστούν αποτελεσματικά τον τομέα, παρέχοντας ταυτόχρονα αρκετές λεπτομέρειες για την επίτευξη της απαιτούμενης επιχειρηματολογίας. Θα ξεκινήσουμε με την περιορισμένη συλλογή κόμβων και τιμών που παρουσιάζεται στον πίνακα 1.1. Τα πράγματα που μπορούν να αναπαρασταθούν στο δίκτυο είναι ήδη περιορισμένα από αυτές τις επιλογές. Δεν υπάρχει αναπαράσταση άλλων ασθενειών, όπως η φυματίωση ή η βρογχίτιδα, οπότε το σύστημα δεν μπορεί να παρέχει την πιθανότητα ο ασθενής να έχει αυτές τις ασθένειες. Ένας άλλος περιορισμός είναι η ύπαρξη διαφοροποίησης, μεταξύ έντονου ή ήπιου καπνίσματος, όπου και πάλι το μοντέλο προϋποθέτει τουλάχιστον κάποια έκθεση στη ρύπανση. Αξίζει να σημειωθεί ότι όλοι αυτοί οι κόμβοι έχουν μόνο δύο τιμές, γεγονός που απλοποιεί το μοντέλο, αλλά γενικά δεν υπάρχουν περιορισμοί στον αριθμό των διακριτών τιμών που μπορούν να χρησιμοποιηθούν.

Πίνακας 1.1: (Korb and Nicolson (2010) [16]),Προκαταρκτικές επιλογές κόμβων και τιμών για το παράδειγμα του καρκίνου του πνεύμονα

Όνομα κόμβου	**Τύπος**	**Τιμές**
Ρύπανση	Δυαδικό	χαμηλό, υψηλό
Καπνιστής	Boolean	{T, F}
Καρκίνος	Boolean	{T, F}
Δύσπνοια	Boolean	{T, F}
X-ray	Δυαδικό	{pos, neg}

Στο παράδειγμα που δώσαμε παραπάνω, ο κόμβος Cancer έχει δύο γονείς, Pollution και Smoker και ο Smoker είναι πρόγονος τόσο της X-ray όσο και της Dyspnoea. Με τον ίδιο τρόπο, η ακτίνα X είναι παιδί του Καρκίνου και απόγονος του Καπνιστή και της Ρύπανσης. Parents(X) είναι το σύνολο των γονέων του κόμβου X. Μια άλλη χρήσιμη έννοια είναι ένας κόμβος Markov οι γονείς των παιδιών του. Από την αναλογία του "δέντρου" έχουμε μια πολύ συχνά χρησιμοποιούμενη ορολογία: ένας κόμβος χωρίς γονείς ονομάζεται κόμβος ρίζα και ένας κόμβος που δεν έχει παιδιά ονομάζεται κόμβος φύλλο. Κάθε μη-φύλλο και μη ρίζα ονομάζεται ενδιάμεσος κόμβος. Αυτό σημαίνει ότι οι κόμβοι ρίζας αντιπροσωπεύουν αρχικές αιτίες και ταυτόχρονα οι κόμβοι φύλλων αντιπροσωπεύουν τα τελικά αποτελέσματα. Στο παράδειγμα του καρκίνου, εμείς έχουμε δύο κόμβους ρίζα, Pollution και Smoker και δύο κόμβους φύλλο , X-ray και Dyspnoea. Κατόπιν συμφωνίας, για ευκολότερη οπτική παρατήρηση της δομής ενός BayesianNetworks συνήθως τοποθετούνται έτσι ώστε οι ακμές να δείχνουν γενικά από πάνω προς τα κάτω.



Σχήμα 1.3: (Korb and Nicolson (2010) [16]). Ένα Μπεϋζιανό Δίκτυο για το πρόβλημα του καρκίνου του πνεύμονα.

Το μοντέλο πιθανοφάνειας στο συγκεκριμένο παράδειγμα είναι (Korb and Nicolson (2010) [16]):

$$P\left(X=pos \wedge D=T \wedge C=T \wedge P=L \wedge S=F\right)=P\left(X=pos \mid D=T,C=T,P=L,S=F\right)$$
$$\times P\left(D=T \mid C=T,P=L,S=F\right)$$
$$\times P\left(C=T \mid P=L,S=F\right)P\left(P=L \mid S=F\right)P\left(S=F\right)$$
$$=P\left(X=pos \mid C=T\right)P\left(D=T \mid C=T\right)P\left(C=T \mid P=L,S=F\right)$$
$$\times P\left(P=L\right)P\left(S=F\right)$$

1.5 Παράδειγμα με χρήση Κώδικα στην R: Έρευνα για τη χρήση του τρένου

Σε αυτό το υποκεφάλαιο παραθέτουμε ένα παράδειγμα που δόθηκε αρχικά από τους (Scutari, M., & Denis, J. B. (2021)[8]). Έστω ότι έχουμε μία υποθετική έρευνα της οποίας σκοπός είναι να εξετάσει τον τρόπο χρήσης των διαφόρων μέσων μεταφοράς, εστιάζοντας στα τρένα και τα αυτοκίνητα. Αυτού του είδους οι έρευνες χρησιμοποιούνται για την εκτίμηση της ικανοποίησης των πελατών σε διάφορες κοινωνικές ομάδες, έτσι ώστε οι δημόσιες πολιτικές ή ο πολεοδομικός σχεδιασμός να μπορούν να αξιολογηθούν. Σε αυτό το παράδειγμα θέλουμε να εξετάσουμε, για κάθε άτομο, τις επερχόμενες έξι διακριτές μεταβλητές (οι ετικέτες τους που θα χρησιμοποιήσουμε στα σχήματα και τους υπολογισμούς σημειώνονται σε παρένθεση):

◦ Ηλικία (AGE): η ηλικία, η οποία μπορεί να πάρει τις τιμές young (y) για άτομα κάτω των 30 ετών , adult (a) για άτομα ηλικίας μεταξύ 30 και 60 ετών και old (ol) για άτομα άνω των 60 ετών.

◦ Φύλο (SEX): το βιολογικό φύλο του υπό εξέταση ατόμου, το οποίο μπορεί να είναι αρσενικό (M) ή γυναίκα (F).

◦ Εκπαίδευση (EDU): το υψηλότερο επίπεδο εκπαίδευσης ή κατάρτισης το οποίο έχει ολοκληρώσει ένα άτομο και μπορεί να λάβει τις αξίες λύκειο (h) ή πανεπιστημιακό πτυχίο (u).

◦ Επάγγελμα (OCC): αν το άτομο είναι μισθωτός (em) ή αυτοαπασχολούμενος (se). εργαζόμενος.

◦ Κατοικία (RES): το μέγεθος της πόλης στην οποία ζει το άτομο, που καταγράφεται είτε ως μικρή είτε ως μικρή.(s) ή μεγάλη (b).
◦ Ταξίδια (TRA): το μέσο μεταφοράς που προτιμά το άτομο, καταγεγραμμένο είτε ως αυτοκίνητο (c), τρένο (tr) ή άλλο (ot).

Σε αυτό το παράδειγμα, κάθε μεταβλητή ανήκει σε μία από τις τρεις ομάδες. Η ηλικία και το φύλο είναι δημογραφικοί δείκτες. Με απλά λόγια, είναι εγγενή χαρακτηριστικά του ατόμου που μπορεί να οδηγήσουν σε διαφορετικά πρότυπα συμπεριφοράς, αλλά δεν επηρεάζονται από το ίδιο το άτομο. Από την άλλη πλευρά, το ακριβώς αντίθετο ισχύει για τις EDU, OCC και RES. Οι μεταβλητές αυτές είναι κοινωνικοοικονομικοί δείκτες και περιγράφουν τη θέση του ατόμου στην κοινωνία. Για το λόγο αυτό, παρέχεται μια χονδρική περιγραφή του αναμενόμενου τρόπου ζωής του ατόμου, για παράδειγμα, μπορεί να χαρακτηριστούν οι συνήθειες του όσον αφορά τις δαπάνες ή το πρόγραμμα εργασίας του. Μεταβλητή, TRA, είναι ο στόχος του παραδείγματος, η ποσότητα για την οποία ενδιαφερόμαστε και της οποίας η συμπεριφορά είναι υπό έρευνα.

1.5.1 Γραφική αναπαράσταση

Η φύση των μεταβλητών που καταγράφονται στο παράδειγμα, και γενικά των τριών κατηγοριών στις οποίες ανήκουν, υποδηλώνει τη σχέση που έχουν μεταξύ τους. Ορισμένες θα είναι άμεσες, ενώ άλλες θα είναι έμμεσες, που σημαίνει ότι θα μεσολαβούν από μία ή περισσότερες μεταβλητές. Μέσα από ένα κατευθυνόμενο γράφημα, το οποίο είναι μία από τις δύο θεμελιώδεις οντότητες που χαρακτηρίζουν ένα Μπεϋζιανό Δίκτυο, μπορεί να αναπαραστήσει αποτελεσματικά και διαισθητικά και τα δύο είδη σχέσεων.

Στο παράδειγμα κάθε κόμβος του γραφήματος, αντιστοιχεί σε μία από τις μεταβλητές. Στην πραγματικότητα, συχνά λέγεται ότι είναι εναλλάξιμες στη βιβλιογραφία. Ως αποτέλεσμα, ο γράφος που δημιουργείται από αυτό το παράδειγμα θα περιέχει 6 κόμβους, με τις ακόλουθες ετικέτες: AGE, SEX, EDU, OCC, RES και TRA. Οι ακμές μεταξύ ζευγών μεταβλητών αντιπροσωπεύουν σχέσεις άμεσης εξάρτησης (δηλαδή, AGE → EDU σημαίνει ότι η EDU εξαρτάται από την AGE). Οι κόμβοι στην ουρά των ακμών ονομάζονται γονείς, ενώ αυτοί που βρίσκονται στην κεφαλή (όπου βρίσκεται το βέλος) ονομάζονται παιδιά. Δεν υπάρχει ρητή αναπαράσταση των έμμεσων σχέσεων εξάρτησης. Παρόλο που εμείς θα μπορούσαμε να τις διαβάσουμε από το γράφημα ως ακολουθίες ακμών που πηγαίνουν από μια μεταβλητή σε μια άλλη, μέσω μιας ή περισσότερων μεταβλητών που μεσολαβούν (δηλαδή, ο συνδυασμός AGE → EDU και EDU→ RES σημαίνει ότι η RES εξαρτάται από την AGE μέσω της EDU). Ακολουθίες ακμών όπως αυτές λέγεται ότι σχηματίζουν ένα μονοπάτι που πηγαίνει από μια μεταβλητή σε μια άλλη και αυτές οι δύο μεταβλητές πρέπει να πρέπει να είναι διαφορετικές. Μονοπάτια όπως AGE → . . . . → AGE, τα οποία ονομάζουμε κύκλους, δεν εγκρίνονται. Αυτός είναι ο λόγος για τον οποίο οι γράφοι που χρησιμοποιούμε στα Μπεϋζιανά Δίκτυα ονομάζονται κατευθυνόμενοι ακυκλικοί Γράφοι (DAGs). Αξίζει να σημειωθεί ότι πρέπει να είμαστε προσεκτικοί όταν ερμηνεύουμε τις άμεσες και έμμεσες εξαρτήσεις. Η ύπαρξη ακμών συνεπάγεται ότι για κάθε ακμή, μια μεταβλητή θα πρέπει να θεωρείται ως αιτία και η άλλη ως αποτέλεσμα (π.χ. AGE → EDU σημαίνει ότι AGE προκαλεί EDU). Γενικά μιλάμε για σχέσεις εξάρτησης αντί για αιτιώδεις αποτελέσματα, διότι είναι δύσκολο να δικαιολογήσουμε την ερμηνεία που ονομάζεται αιτιώδης στις περισσότερες καταστάσεις.

Για τη δημιουργία και το χειρισμό κατευθυνόμενων ακυκλικών γραφημάτων στο πλαίσιο του Bayesian Δικτύων, θα χρησιμοποιήσουμε το πακέτο bnlearn (συντομογραφία για "Bayesian network learning"). Το πρώτο βήμα είναι η δημιουργία ενός κατευθυνόμενου ακυκλικού γράφου με έναν κόμβο για κάθε μεταβλητή στο παράδειγμα και χωρίς ακμές, το οποίο συνήθως ονομάζεται κενό γράφημα, αφού έχει κενό σύνολο ακμών. Μετά από αυτό μπορούμε να αρχίσουμε να προσθέτουμε τις ακμές που κωδικοποιούν τις άμεσες εξαρτήσεις μεταξύ των μεταβλητών του παραδείγματος. Όπως είπαμε οι AGE και SEX δεν επηρεάζονται από καμία από τις άλλες μεταβλητές, οπότε δεν υπάρχουν ακμές που να φαίνονται σε καμία από τις δύο μεταβλητές. Από την άλλη πλευρά, EDU επηρεάζεται άμεσα από τις AGE και SEX. Όπως γνωρίζουμε, για παράδειγμα, ο αριθμός των ατόμων που φοιτούν σε πανεπιστήμια έχει αυξηθεί με την πάροδο των ετών. Ως αποτέλεσμα, είναι πιο πιθανό για τους νεότερους ανθρώπους να έχουν πανεπιστημιακό πτυχίο απ' ότι οι μεγαλύτεροι σε ηλικία.

Ομοίως, το SEX επηρεάζει επίσης το EDU- το χάσμα μεταξύ των δύο φύλων στις αιτήσεις για πανεπιστημιακές σπουδές γίνεται όλο και μεγαλύτερο εδώ και πολλά χρόνια, με τις γυναίκες να υπερτερούν αριθμητικά και να ξεπερνούν τους άνδρες. Επιπλέον, η Εκπαίδευση επηρεάζει έντονα το επάγγελμα και την κατοικία.

Είναι γνωστό ότι, τα υψηλότερα επίπεδα εκπαίδευσης βοηθούν στην πρόσβαση σε επαγγέλματα με μεγαλύτερο κύρος. Επιπρόσθετα, οι άνθρωποι πολύ συχνά μετακινούνται για να φοιτήσουν σε ένα συγκεκριμένο πανεπιστήμιο ή για να βρουν μια θέση εργασίας που ταιριάζει στις δεξιότητές τους.

Τελικά, τα προτιμώμενα μέσα μεταφοράς επηρεάζονται άμεσα τόσο από το OCC όσο και από το RES. Για το πρώτο, οφείλεται το γεγονός ότι λίγες θέσεις εργασίας απαιτούν τακτικά ταξίδια μεγάλων αποστάσεων, ενώ άλλες απαιτούν συχνότερα ταξίδια μικρών αποστάσεων. Το δεύτερο οφείλεται στο γεγονός ότι τόσο ο χρόνος μετακίνησης όσο και η απόσταση αποτελούν αποφασιστικούς παράγοντες για την επιλογή μεταξύ μετακινήσεων με τρένο ή με αυτοκίνητο.

Όταν προστεθούν όλες οι ακμές, ο κατευθυνόμενος ακυκλικός γράφος στο αντικείμενο dag1 κωδικοποιεί τις επιθυμητές άμεσες εξαρτήσεις. Οι άμεσες εξαρτήσεις παρατίθενται για κάθε μεταβλητή, διαχωρίζονται με άνω και κάτω τελεία (:) και συμβολίζονται με μια γραμμή (-). Για παράδειγμα, [EDU-AGE:SEX] σημαίνει ότι AGE → EDU και SEX → EDU, ενώ [AGE] σημαίνει ότι δεν υπάρχει τόξο που να δείχνει προς την AGE. Η συνάρτηση modelstring μπορεί να χρησιμοποιηθεί για την αναπαράσταση του γραφήματος δομής που έχει σχεδιαστεί για να ανακαλεί ένα γινόμενο υπό συνθήκη πιθανοτήτων.

Ο κώδικας στην R είναι (Scutari, M., & Denis, J. B. (2021)[8]):

```
install.packages("bnlearn")
library (bnlearn)
> dag1<-empty.graph(nodes=c( "AGE" , "SEX" , "EDU" , "OCC" , "RES" ,
"TRA") )
> dag1

  Random/Generated Bayesian network

  model:
   [AGE][SEX][EDU][OCC][RES][TRA]
  nodes:                                 6
  arcs:                                  0
    undirected arcs:                     0
    directed arcs:                       0
  average markov blanket size:           0.00
  average neighbourhood size:            0.00
  average branching factor:              0.00

  generation algorithm:                  Empty

> dag1 <-set.arc ( dag1 , from = "AGE" , to = "EDU" )
> dag1 <-set.arc ( dag1 , from = "SEX" , to = "EDU" )
> dag1 <-set.arc ( dag1 , from = "EDU" , to = "OCC" )
> dag1 <-set.arc ( dag1 , from = "EDU" , to = "RES" )
> dag1 <-set.arc ( dag1 , from = "OCC" , to = "TRA" )
> dag1 <-set.arc ( dag1 , from = "RES" , to = "TRA" )
> dag1

  Random/Generated Bayesian network
```

```
  model:
   [AGE][SEX][EDU|AGE:SEX][OCC|EDU][RES|EDU][TRA|OCC:RES]
  nodes:                                 6
  arcs:                                  6
    undirected arcs:                     0
    directed arcs:                       6
  average markov blanket size:           2.67
  average neighbourhood size:            2.00
  average branching factor:              1.00

  generation algorithm:                  Empty

>modelstring( dag1 )
[1] "[AGE][SEX][EDU|AGE:SEX][OCC|EDU][RES|EDU][TRA|OCC:RES]"
> nodes( dag1 )
[1] "AGE" "SEX" "EDU" "OCC" "RES" "TRA"
>all_edges<-arcs(dag1)
>all_edges
     from  to
[1,] "AGE" "EDU"
[2,] "SEX" "EDU"
[3,] "EDU" "OCC"
[4,] "EDU" "RES"
[5,] "OCC" "TRA"
[6,] "RES" "TRA"
>arc.set<-matrix( c ( "AGE" , "EDU" ,"SEX" , "EDU","EDU" ,"OCC"
,"EDU","RES","OCC","TRA","RES","TRA" ) ,byrow = TRUE, ncol = 2
,dimnames =list(NULL, c("from","to")))
> dag2 <-empty.graph ( nodes = c ( "AGE" , "SEX" , "EDU" , "OCC" ,
"RES" , "TRA" ) )
> dag1

  Random/Generated Bayesian network

  model:
   [AGE][SEX][EDU|AGE:SEX][OCC|EDU][RES|EDU][TRA|OCC:RES]
  nodes:                                 6
  arcs:                                  6
    undirected arcs:                     0
    directed arcs:                       6
  average markov blanket size:           2.67
  average neighbourhood size:            2.00
  average branching factor:              1.00

  generation algorithm:                  Empty

> dag2

  Random/Generated Bayesian network

  model:
   [AGE][SEX][EDU][OCC][RES][TRA]
  nodes:                                 6
  arcs:                                  0
    undirected arcs:                     0
    directed arcs:                       0
  average markov blanket size:           0.00
  average neighbourhood size:            0.00
  average branching factor:              0.00

  generation algorithm:                  Empty

>arc.set
     from  to
[1,] "AGE" "EDU"
[2,] "SEX" "EDU"
[3,] "EDU" "OCC"
[4,] "EDU" "RES"
```

```
[5,] "OCC" "TRA"
[6,] "RES" "TRA"
> arcs(dag2)=arc.set
> #The resulting DAG is identical to the previous one, dag.
>all.equal( dag1,dag2 )
[1] TRUE
```

Με βάση τον παραπάνω κώδικα R, φαίνεται ότι επιτυγχάνεται η δημιουργία του ορισμού ακμών και του έλεγχου ταυτότητας μεταξύ δύο Bayesian δικτύων. Εδώ είναι ένα γενικό συμπέρασμα:

1. Δημιουργία Δικτύου Bayesian (dag1):Αρχικά, δημιουργείτε ένα δίκτυο Bayesian με συγκεκριμένες ακμές μεταξύ των κόμβων "AGE," "SEX," "EDU," "OCC," "RES," και "TRA."
2. Επεξεργασία Δικτύου Bayesian (dag1):Χρησιμοποιώντας τη συνάρτηση `set.arc`, ορίζετε ακόμα περισσότερες ακμές στο δίκτυο.
3. Εκτύπωση Δικτύου Bayesian (dag1):Εκτυπώνετε τη δομή του δικτύου για επιπρόσθετο έλεγχο.
4. Λήψη Ακμών Δικτύου (dag1):Χρησιμοποιείτε τη συνάρτηση `arcs` για να πάρετε τις ακμές του δικτύου.
5. Δημιουργία Δεύτερου Δικτύου Bayesian (dag2): Δημιουργείτε ένα άλλο δίκτυο Bayesian με τις ίδιες κόμβους.
6. Ορισμός Ακμών Δικτύου (dag2) με Χρήση Πίνακα (arc.set):Χρησιμοποιώντας έναν πίνακα (`arc.set`), ορίζετε ακμές στο δεύτερο δίκτυο.
7. Έλεγχος Ταυτότητας των Δικτύων: Χρησιμοποιείτε τη συνάρτηση `all.equal` για να ελέγξετε αν τα δύο δίκτυα είναι ταυτόσημα.

Στην περίπτωση αυτή, η επιστροφή `TRUE` αποδεικνύει ότι τα δύο δίκτυα είναι πανομοιότυπα, δηλαδή έχουν την ίδια δομή ακμών.

1.5.2 Πιθανολογική αναπαράσταση

Προηγουμένως είδαμε, ότι οι αλληλεπιδράσεις μεταξύ των AGE, SEX, EDU, OCC, RES και TRA αναπαρίσταντο με τη χρήση ενός κατευθυνόμενου ακυκλικού γράφου. Μια κοινή κατανομή πιθανοτήτων σε αυτές τις μεταβλητές πρέπει να καθοριστεί προκειμένου να ολοκληρωθεί η μοντελοποίηση του BayesianNetwork. Οι εν λόγω μεταβλητές είναι διακριτές και ορίζονται σε ένα σύνολο μη ταξινομημένων καταστάσεων. Κατά συνέπεια, η λογική επιλογή για την κοινή κατανομή πιθανότητας είναι μια πολυωνυμική κατανομή, η οποία δίνει μια πιθανότητα σε κάθε συνδυασμό καταστάσεων των μεταβλητών στην έρευνα. Στα δίκτυα Bayes, η κοινή κατανομή αναφέρεται ως συνολική κατανομή. Είναι, όμως, πολύ δύσκολο να χρησιμοποιηθεί άμεσα η παγκόσμια κατανομή, ακόμη και αν μιλάμε για μικρά προβλήματα, όπως αυτό που εξετάζουμε, καθώς έχει πολλές παραμέτρους.

Σε αυτό το παράδειγμα, το σύνολο των παραμέτρων έχει 143 πιθανότητες που αντιστοιχούν στους συνδυασμούς των επιπέδων των μεταβλητών. Ευτυχώς οι κατευθυνόμενοι ακυκλικοί γράφοι κωδικοποιούν πληροφορίες και μπορούν να χρησιμοποιηθούν για να χωριστεί η συνολική κατανομή σε μία για κάθε μεταβλητή ώστε να γίνει ένα σύνολο μικρότερων τοπικών κατανομών. Οι μεταβλητές που δεν συνδέονται με μια ακμή είναι υπό όρους ανεξάρτητες, έτσι ώστε η παγκόσμια κατανομή να παραγοντοποιείται με τον ακόλουθο τρόπο:

$$\begin{aligned}&\Pr\left(AGE,SEX,EDU,OCC,RES,TRA\right)=\Pr\left(AGE\right)\Pr\left(SEX\right)\Pr\left(EDU \mid AGE,SEX\right)\\&\Pr(OCC \mid EDU)\Pr(RES \mid EDU)\Pr(TRA,OCC,RES)\end{aligned} \tag{1.4}$$

Σε έναν επίσημο ορισμό του τρόπου με τον οποίο οι εξαρτήσεις μπορούν να περιγράφουν από το DAG, δίνεται ο χάρτης στο χώρο πιθανοτήτων με σχέσεις ανεξαρτησίας υπό όρους στην εξίσωση (1.4). Η παραγοντοποίηση είναι καλά ορισμένη εάν δεν υπάρχουν κύκλοι στο DAG. Οι μεταβλητές εξαρτώνται μόνο από τους γονείς τους, η κατανομή τους είναι μονομεταβλητή και έχουν μόνο λίγες παραμέτρους. Το σύνολο όλων των τοπικών κατανομών έχει μικρότερο αριθμό παραμέτρων από την παγκόσμια κατανομή. Η τελευταία δηλώνει ένα πιο γενικό μοντέλο από το πρώτο, αφού δεν κάνει υποθέσεις σχετικά με τις εξαρτήσεις μεταξύ των μεταβλητών. Μια εμφωλευμένη στο μοντέλο ή υπομοντέλο της παγκόσμιας κατανομής ορίζεται από την παραγοντοποίηση στην Εξίσωση (1.4).

Στο παράδειγμά μας, το AGE και το SEX μοντελοποιούνται με απλές, μονοδιάστατες πιθανότητες αφού δεν έχουν γονέα. Τα OCC και RES, τα οποία εξαρτώνται από το EDU, μοντελοποιούνται από δισδιάστατους πίνακες πιθανοτήτων υπό συνθήκη. Κάθε στήλη αντιστοιχεί σε ένα επίπεδο κάθε γονέα και περιέχει την κατανομή της μεταβλητής υπό την αίρεση του συγκεκριμένου επιπέδου. Κατά συνέπεια, οι πιθανότητες αθροίζονται σε 1 σε κάθε στήλη. Τέλος, οι EDU και TRA είναι μοντελοποιούνται ως τρισδιάστατοι πίνακες, καθώς ο καθένας έχει δύο γονείς που είναι οι AGE και SEX για την EDU, OCC και RES για την TRA. Κάθε στήλη αντιστοιχεί σε έναν συνδυασμό των επιπέδων των γονέων, και περιέχει την κατανομή της μεταβλητής με την προϋπόθεση του συγκεκριμένου συνδυασμού. Σε γενικές γραμμές, οι τοπικές κατανομές που ορίζονται παραπάνω έχουν μόνο 21 παραμέτρους, σε αντίθεση με τις 143 της παγκόσμιας κατανομής. Επιπλέον, μπορούμε να χειριστούμε τις τοπικές κατανομές ανεξάρτητα τη μία από την άλλη, θέτοντας το πολύ 8 παραμέτρους στην καθεμία. Αυτή η μείωση της διάστασης είναι μία από τις θεμελιώδεις ιδιότητες των Μπεϋζιανών δικτύων και καθιστά εφικτή την εφαρμογή τους για προβλήματα υψηλών διαστάσεων. Έχοντας ορίσει και στα δύο, την τοπική κατανομή που αντιστοιχεί σε κάθε μεταβλητή και τον κατευθυνόμενο ακυκλικό γράφο οπού μπορούν να συνδυαστούν για να σχηματίσουν ένα πλήρως καθορισμένο Μπεϋζιανό Δίκτυο.

Χρησιμοποιώντας τον τύπο του μοντέλου που παρέχεται από τη συνάρτηση modelstring στην R, της οποίας η σύνταξη είναι σχεδόν πανομοιότυπη με την Εξίσωση (1.4), το κατευθυνόμενο ακυκλικό γράφημα μπορεί να αναδημιουργηθεί. Οι γονείς και οι κόμβοι κάθε κόμβου μπορούν να απαριθμηθούν με οποιαδήποτε σειρά, γεγονός που μας επιτρέπει να ακολουθήσουμε τη λογική δομή του δικτύου κατά τη συγγραφή του τύπου. Αυτό σημαίνει ότι η συνολική κατανομή μπορεί να παραγοντοποιηθεί σε μία για κάθε μεταβλητή. Μετά από αυτό ο κατευθυνόμενος ακυκλικός γράφος που αποθηκεύσαμε στο dag1 συνδυάζεται με μια λίστα που περιέχει τις τοπικές κατανομές, η οποία θα ονομάζεται cpt1, σε ένα αντικείμενο της κλάσης bn.fit που ονομάζεται bn. Μπορούμε να υπολογίσουμε τον αριθμό των παραμέτρων του Μπεϋζιανού Δικτύου με τη συνάρτηση nparams η οποία πράγματι είναι 21, όπως αναμέναμε από την σύνολα παραμέτρων των τοπικών κατανομών.

Ο κώδικας R για όλα τα παραπάνω έχει ως εξής (Scutari, M., & Denis, J. B. (2021)[8]):

```
>library (bnlearn)
>AGE.lv <-c( "y" , "a" ,"o" )
>SEX.lv <- c( "M" , "F" )
>EDU.lv <-c( "h" , "u" )
>OCC.lv <- c( "em" ,"se" )
>RES.lv <- c( "s" , "b" )
>TRA.lv <- c( "c" ,"t" ,"ot" )
>AGE.prob<- array ( c ( 0.30 , 0.50 ,0.20 ) , dim = 3 ,dimnames =
list(AGE=AGE.lv))
>AGE.prob
AGE
  y   a   o
0.3 0.5 0.2
>SEX.prob<-array(c(0.60,0.40),dim = 2 ,dimnames = list (SEX =
SEX.lv))
>SEX.prob
SEX
  M   F
0.6 0.4
>OCC.prob<-array ( c ( 0.96 , 0.04 , 0.92 , 0.08) , dim = c ( 2 , 2
) ,dimnames=list (OCC = OCC.lv , EDU = EDU.lv ) )
>OCC.prob
    EDU
OCC     h    u
```

```
em 0.96 0.92
  se 0.04 0.08
>RES.prob<-array(c(0.25,0.75,0.20,0.80),dim = c(2,2)
,dimnames=list(RES=RES.lv,EDU=EDU.lv))
>RES.prob
   EDU
RES    h   u
  s 0.25 0.2
  b 0.75 0.8
>RES.prob<-matrix(c(0.25,0.75,0.20,0.80),ncol = 2,dimnames
=list(RES= RES.lv,EDU = EDU.lv))
>RES.prob
   EDU
RES    h   u
  s 0.25 0.2
  b 0.75 0.8
>TRA.prob<-
array(c(0.48,0.42,0.10,0.56,0.36,0.08,0.58,0.24,0.18,0.70,0.21,0.09)
,dim=c(3,2 ,2) ,dimnames=list(TRA=TRA.lv,OCC =OCC.lv,RES =RES.lv))
>dag3<-
model2network("[AGE][SEX][EDU|AGE:SEX][OCC|EDU][RES|EDU][TRA|OCC:RES
]")
>dag3

  Random/Generated Bayesian network

  model:
   [AGE][SEX][EDU|AGE:SEX][OCC|EDU][RES|EDU][TRA|OCC:RES]
  nodes:                                 6
  arcs:                                  6
    undirected arcs:                     0
    directed arcs:                       6
  average markov blanket size:           2.67
  average neighbourhood size:            2.00
  average branching factor:              1.00

  generation algorithm:                  Empty

>all.equal(dag1,dag3)
[1] TRUE
>cpt1<-list(AGE=AGE.prob,SEX=SEX.prob,EDU=EDU.prob,OCC=OCC.prob,RES
=RES.prob,TRA=TRA.prob)
>bn<-custom.fit(dag1,cpt1)
>bn


Bayesian network parameters

  Parameters of node AGE (multinomial distribution)

Conditional probability table:
 AGE
  y   a   o
0.3 0.5 0.2

  Parameters of node SEX (multinomial distribution)

Conditional probability table:
 SEX
  M   F
0.6 0.4

  Parameters of node EDU (multinomial distribution)

Conditional probability table:

, , SEX = M

   AGE
```

```
EDU    y    a    o
  h 0.75 0.72 0.88
  u 0.25 0.28 0.12

, , SEX = F

   AGE
EDU    y    a    o
  h 0.64 0.70 0.90
  u 0.36 0.30 0.10


  Parameters of node OCC (multinomial distribution)

Conditional probability table:

    EDU
OCC     h    u
em 0.96 0.92
  se 0.04 0.08

  Parameters of node RES (multinomial distribution)

Conditional probability table:

   EDU
RES    h    u
  s 0.25 0.20
  b 0.75 0.80

  Parameters of node TRA (multinomial distribution)

Conditional probability table:

, , RES = s

    OCC
TRA    em   se
  c  0.48 0.56
  t  0.42 0.36
ot 0.10 0.08

, , RES = b

    OCC
TRA    em   se
c  0.58 0.70
  t  0.24 0.21
ot 0.18 0.09
```

Οι μεταβλητές της κλάσης bn.fit χρησιμοποιούνται για τον ορισμό των Μπεϋζιανών Δικτύων στο bnlearn. Μεταφέρουν πληροφορίες σχετικά με τον κατευθυνόμενο ακυκλικό γράφο (όπως οι γονείς και τα παιδιά κάθε κόμβου) και τις τοπικές κατανομές (τις παραμέτρους τους). Πρακτικά μπορούμε να τις χρησιμοποιήσουμε σαν να ήταν αντικείμενα της κλάσης bn κατά τη διερεύνηση γραφικών ιδιοτήτων. Για παράδειγμα (Scutari, M., & Denis, J. B. (2021)[8]).

```
>edges<-arcs(bn)
>edges
     from  to
[1,] "AGE" "EDU"
```

```
[2,] "SEX" "EDU"
[3,] "EDU" "OCC"
[4,] "EDU" "RES"
[5,] "OCC" "TRA"
[6,] "RES" "TRA"
```

και το ίδιο ισχύει και για άλλες λειτουργίες όπως οι κόμβοι, οι γονείς και τα παιδιά. Επιπλέον, μπορούμε να εκτυπώνουμε πίνακες υπό συνθήκη πιθανοτήτων από το αντικείμενο bn.fit (Scutari, M., & Denis, J. B. (2021)[8]).

```
>bn$R

  Parameters of node RES (multinomial distribution)

Conditional probability table:

   EDU
RES    h    u
  s 0.25 0.20
b 0.75 0.80
```

ένας άλλος τρόπος εξαγωγής τους είναι η χρήση της συνάρτησης coef ως εξής.

```
>RES.cpt<=coef(bn$RES)
   EDU
RES    h    u
  s TRUE TRUE
b TRUE TRUE
>bn
```

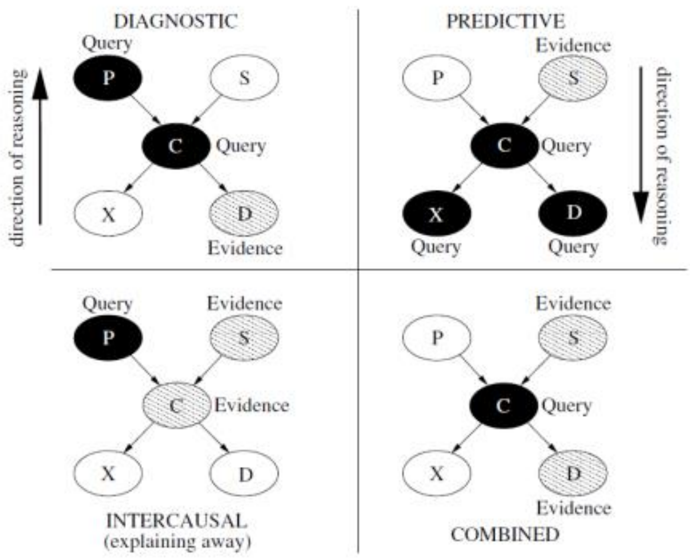
Αν απλά πληκτρολογήσουμε bn, όλοι οι πίνακες υπό συνθήκη πιθανοτήτων στο Μπεϋζιανό δίκτυο θα έχουν εκτυπωθεί.

1.6 Συλλογιστική με Μπεϋζιανά Δίκτυα

Εφόσον έχουμε ορίσει τον τρόπο αναπαράστασης ενός τομέα και της αβεβαιότητάς του σε ένα Μπεϋζιανό Δίκτυο, μπορούμε να δούμε πώς μπορούμε να χρησιμοποιήσουμε το δίκτυο για συλλογισμό σχετικά με αυτόν. Ειδικότερα, όταν παρατηρείται μια τιμή κάποιας μεταβλητής, θα θέλαμε να θέσουμε ως όρο τη νεά πληροφορία. Αυτή η διαδικασία η οποία ονομάζεται διάδοση πιθανοτήτων ή συμπέρασμα, πραγματοποιείται μέσω μιας "ροής πληροφοριών" μέσω του δικτύου. Αυτή η ροή πληροφοριών δεν περιορίζεται στις κατευθύνσεις των τόξων. Στο πιθανολογικό μας σύστημα, αυτή σημαίνει ότι πρέπει να υπολογίσουμε την εκ των υστέρων κατανομή πιθανότητας για ένα σύνολο κόμβων ερωτήματος, δεδομένων των τιμών για ορισμένα στοιχεία ή παρατηρηθέντες κόμβους.

1.7 Τύποι συλλογισμών

Τα Μπεϋζιανά δίκτυα παρέχουν μια πλήρη αναπαράσταση των κατανομών πιθανοτήτων επί των μεταβλητών. Μπορούμε να εκτελέσουμε διαγνωστικό συλλογισμό, δηλαδή συλλογισμό από τα συμπτώματα στις αιτίες, όπως όταν ένας γιατρός παρατηρεί δύσπνοια και στη συνέχεια ενημερώνει την πεποίθησή του για τον καρκίνο σε περίπτωση που ο ασθενής είναι καπνιστής. Αυτός ο συλλογισμός εμφανίζεται προς την αντίθετη κατεύθυνση από το τόξο του δικτύου. Μπορούμε επίσης να εκτελέσουμε προγνωστικό συλλογισμό ή συλλογισμό νέων πληροφοριών σχετικά με τα αίτια σε νέες πεποιθήσεις σχετικά με τα αποτελέσματα, ακολουθώντας τις κατευθύνσεις των τόξων του δικτύου. Για παράδειγμα, ο ασθενής μπορεί να πει στο γιατρό του ότι είναι καπνιστής- ακόμη και πριν από οποιαδήποτε συμπτώματα.
Ο γιατρός γνωρίζει ότι αυτό θα αυξήσει τις πιθανότητες του ασθενούς να αναπτύξει καρκίνο, καθώς και να αλλάξει τις προσδοκίες του γιατρού για άλλα συμπτώματα, όπως η δύσπνοια. αναπνοή ή ένα καλό αποτέλεσμα ακτινογραφίας.



Σχήμα 1.4:Τύποι συλλογισμών (Korb and Nicolson (2010),[16]).

Ένας άλλος τύπος συλλογισμού ονομάζεται ενδοαιτιακός συλλογισμός, ο οποίος περιλαμβάνει συλλογισμούς σχετικά με τις αμοιβαίες αιτίες ενός κοινού αποτελέσματος. Ας υποθέσουμε ότι ένα συγκεκριμένο αποτέλεσμα έχει ακριβώς δύο εναλλακτικές αιτίες, καθεμία από τις οποίες αντιπροσωπεύεται από μια δομή v στο BN. Αυτή η περίσταση εμφανίζεται στο μοντέλο μας με τα αίτια Καπνιστής και Ρύπανση που έχουν ένα κοινό αποτέλεσμα, τον Καρκίνο.

Σύμφωνα με το μοντέλο, οι δύο αυτές αιτίες είναι ανεξάρτητες μεταξύ τους, πράγμα που σημαίνει ότι είτε ένας ασθενής καπνίζει είτε όχι δεν αλλάζει την πιθανότητα ο ασθενής να υποστεί ρύπανση. Ας υποθέσουμε ότι μαθαίνουμε ότι ο ασθενής πάσχει από καρκίνο. Αυτό θα αυξήσει τις πιθανότητες και για τις δύο πιθανές αιτίες καρκίνου. Ας υποθέσουμε τώρα, ότι μαθαίνουμε ότι είναι καπνιστής. Αυτή η νέα πληροφορία εξηγεί την παρατήρηση του καρκίνου, μειώνοντας την πιθανότητα να έχει εκτεθεί σε σημαντικά επίπεδα ρύπανσης. Παρόλο που οι δύο αιτίες είναι εντελώς ανεξάρτητες, γνωρίζοντας το αποτέλεσμα, η παρουσία μιας εξηγητικής αιτίας παρέχει μια εναλλακτική αιτία λιγότερο πιθανή.

Για να το θέσουμε αλλιώς, η εναλλακτική αιτία έχει εξηγηθεί. Επειδή οποιοιδήποτε κόμβοι μπορεί να είναι κόμβοι ερωτήματος και οποιοιδήποτε μπορεί να είναι κόμβοι αποδεικτικών στοιχείων, μερικές φορές ο συλλογισμός δεν εμπίπτει πάντα σε έναν από τους τύπους που περιγράφονται παραπάνω. Πράγματι, μπορούμε να συνδυάσουμε τους παραπάνω τύπους συλλογισμών με πολλούς τρόπους. Στο Σχήμα 1.4 μπορούμε να δούμε τις διάφορες ποικιλίες της συλλογιστικής χρησιμοποιώντας το παράδειγμα του Cancer BN. Λάβετε υπόψη, ότι ο τελευταίος συνδυασμός καταδεικνύει τη χρήση διαγνωστικής και προγνωστικής συλλογιστικής ταυτόχρονα.

1.8 Τύποι αποδεικτικών στοιχείων

Όταν υπάρχουν διαθέσιμες νέες πληροφορίες τις οποίες ονομάζουμε αποδείξεις, τα Μπεϋζιανά Δίκτυα χρησιμοποιούνται για τον υπολογισμό νέων πεποιθήσεων. Θεωρήστε ως απόδειξη την οριστική διαπίστωση ότι ένας κόμβος X έχει μια ακριβή τιμή, x, την οποία συμβολίζουμε ως X = x και μερικές φορές αναφέρεται ως συγκεκριμένη απόδειξη. Για παράδειγμα, ας υποθέσουμε ότι ανακαλύπτουμε ότι ο ασθενής μας είναι καπνιστής, τότε Smoker=T ονομάζεται συγκεκριμένη απόδειξη. Παρ' όλα αυτά, μερικές φορές έχουμε στη διάθεσή μας στοιχεία τα οποία δεν είναι ξεκάθαρα. Η απόδειξη θα μπορούσε να είναι ότι το Y δεν βρίσκεται στην κατάσταση y_1,αλλά παίρνει οποιαδήποτε από τις άλλες τιμή, το οποίο ονομάζεται αρνητική απόδειξη. Στην πραγματικότητα οι νέες πληροφορίες μπορεί να είναι απλώς μια νέα υπόθεση ότι ο ακτινολόγος που ανέλυσε την ακτινογραφία στο παράδειγμα του καρκίνου δεν είναι σίγουρος. Πιστεύει ότι η ακτινογραφία φαίνεται θετική, αλλά είναι μόνο κατά 80% σίγουρος. Στα Δίκτυα Bayes ονομάζουμε και αυτή την εικονική απόδειξη ή απόδειξη πιθανότητας, δεδομένου ότι χειρίζεται μέσω της πληροφορίας πιθανότητας.

1.9 Συλλογισμοί με αριθμούς

Θα ρίξουμε μια ματιά στους πραγματικούς αριθμούς που έχουμε περιγράψει ποιοτικά τις μορφές των συλλογισμού που είναι δυνατές με τη χρήση των Μπεϋζιανών δικτύων. Ακόμη και πριν από την απόκτηση οποιουδήποτε αποδεικτικού στοιχείου, μπορεί να υπολογιστεί μια εκ των προτέρων πεποίθηση για την τιμή κάθε κόμβου, η οποία είναι η εκ των προτέρων πιθανότητα κατανομή πιθανότητας του κόμβου. Χρησιμοποιούμε την Bel(X) για να δηλώσουμε την εκ των υστέρων κατανομή πιθανότητας πάνω σε μια μεταβλητή X, για να τη διακρίνουμε από την προηγούμενη και την υπό συνθήκη κατανομή πιθανότητας (δηλαδή, P(X), P(X | Y)). Ο πίνακας 1.2 παρέχει τους συγκεκριμένους αριθμούς για τις πεποιθήσεις που έχουν ενημερωθεί για τις περιπτώσεις συλλογισμού που αναφέρθηκαν παραπάνω. Το πρώτο σύνολο είναι για τις προηγούμενες και εξαρτημένες πιθανότητες οι οποίες προσδιορίζονται αρχικά στο Σχήμα 1.6. Με το δεύτερο σύνολο είναι φαίνεται τι θα συμβεί αν το ποσοστό καπνίσματος στον πληθυσμό αυξηθεί από 30% σε 50%, μια αλλαγή που αντιπροσωπεύει το prior για τον κόμβο Smoker. Έτσι, επειδή η διαφορά μεταξύ των δύο περιπτώσεων είναι μόνο η εκ των προτέρων πιθανότητα του καπνίσματος, η οποία σημαίνει P(S = T) = 0,3 στο σε αντίθεση με την P(S = T) = 0,5, η προηγούμενη τιμή καθίσταται άνευ σημασίας και τα δύο δίκτυα καταλήγουν σε τους ίδιους αριθμούς όταν η ίδια η απόδειξη αφορά το γεγονός ότι ο ασθενής είναι καπνιστής.

Πίνακας 1.2: Επικαιροποιημένες πεποιθήσεις δεδομένων νέων πληροφοριών με ποσοστό καπνίσματος 0,3 (κορυφή σύνολο) και 0,5 (κάτω σύνολο) (Korb and Nicolson (2010),[16]).

		Συλλογιστική Υπόθεση				
Κώδικας	**ΧωρίςΠεριστατικό**	**Διάγνωση D=T**	**Πρόβλεψη S=T**	**Διασυνοριακή Σχέση C=T**	**Διασυνοριακή Σχέση C=T S=T**	**Συνδυασμός D=T S=T**
P(S)=0.3						
Bel(P=high) Bel(S=T) Bel(C=T) Bel(X=pos) Bel(D=T)	0.100 0.300 0.011 0.208 0.304	0.102 0.307 0.025 0.217 1	0.100 1 0.032 0.222 0.311	0.249 0.825 1 0.900 0.650	0.156 1 1 0.900 0.650	0.102 1 0.067 0.247 1
P(S)=0.5						
Bel(P=high) Bel(S=T) Bel(C=T) Bel(X=pos) Bel(D=T)	0.100 0.500 0.174 0.212 0.306	0.102 0.508 0.037 0.226 1	0.100 1 0.032 0.311 0.222	0.201 0.917 1 0.900 0.650	0.156 1 1 0.900 0.650	0.102 1 0.067 0.247 1

Ένας αριθμός αλγορίθμων ακριβούς και προσεγγιστικής συμπερασματολογίας μπορεί να χρησιμοποιηθεί για την ενημέρωση πεποιθήσεων. Σημειώστε ότι στα περισσότερα από τα ήδη υπάρχοντα πακέτα λογισμικού των Μπεϋζιανών Δικτύων χρησιμοποιείται ουσιαστικά ο ίδιος αλγόριθμος, και δεν είναι τόσο δύσκολο να κατασκευαστεί και να χρησιμοποιηθεί το Μπεϋζιανό Δίκτυο χωρίς να γνωρίζει κανείς τις λεπτομέρειες της επικαιροποίησης των αλγορίθμων.

1.10 Ιδιότητες ανεξαρτησίας υπό όρους των κατευθυνόμενων γραφικών μοντέλων

Ένα σύνολο υποθέσεων εξαρτημένης ανεξαρτησίας (CI) βρίσκεται στο επίκεντρο κάθε γραφικού μοντέλου. Αν το D είναι ανεξάρτητο από το E δεδομένου του F στο γράφημα G, γράφουμε $x_D \perp_G x_E \mid x_F$ χρησιμοποιώντας τη σημασιολογία που ορίζουμε παρακάτω. Υποθέτουμε ότι $I(G)$ είναι το σύνολο όλων αυτών των υπό συνθήκη δηλώσεων ανεξαρτησίας που δίνονται από το γράφημα. Ο G λέγεται ότι είναι ένας χάρτης ανεξαρτησίας (I-map) για την p, ή η p είναι Markov σε σχέση με τον G, εάν και μόνο εάν $I(G) \subseteq I(p)$ έτσι ώστε $I(p)$ είναι το σύνολο όλων των δηλώσεων υπό συνθήκη ανεξαρτησίας που είναι αληθείς για την κατανομή p. Αυτό σημαίνει ότι αν το γράφημα δεν κάνει ισχυρισμούς εξαρτημένης ανεξαρτησίας οι οποίοι δεν ισχύουν για την κατανομή, τότε ονομάζεται χάρτης ανεξαρτησίας. Με βάση αυτό, το γράφημα μπορεί να χρησιμοποιηθεί ως μια ασφαλής προσέγγιση για το p όταν σκεφτόμαστε για το $p's$ ιδιότητες ανεξαρτησίας στις δεσμευμένες πιθανότητες. Αυτό μπορεί να μας βοηθήσει να σχεδιάσουμε αλγορίθμους που λειτουργούν για μεγάλες κλάσεις κατανομών, αδιαφορώντας για τις ιδιαίτερες αριθμητικές παραμέτρους τους.

Αξίζει να σημειωθεί ότι το πλήρως συνδεδεμένο γράφημα είναι ένας χάρτης ανεξαρτησίας όλων των κατανομών, επειδή δεν γίνονται καθόλου ισχυρισμοί ανεξαρτησίας υπό όρους, αφού δεν λείπουν ακμές. Για το λόγο αυτό ο G λέγεται ότι είναι ένας ελάχιστος χάρτης ανεξαρτησίας του p, αν ο G είναι ένας χάρτης ανεξαρτησίας χάρτης του p, και αν δεν υπάρχει G′⊆G που να είναι χάρτης ανεξαρτησίας του p. Το μόνο πράγμα που απομένει και που πρέπει να διευκρινιστεί είναι πώς να προσδιοριστεί αν $X_D \perp_G X_E \mid X_F$. Αυτές οι ανεξαρτησίες για μη κατευθυνόμενες γράφους μπορούν εύκολα να εξαχθούν, αλλά η κατάσταση των DAG είναι λίγο πιο περίπλοκη, αφού είναι απαραίτητο να λαμβάνεται υπόψη ο προσανατολισμός των κατευθυνόμενων τόξων.

1.11 D-separation και αλγόριθμος BayesBall (παγκόσμιος αλγόριθμος Markov ιδιότητες)

Ο αλγόριθμος BayesBall (BB) είναι μια αλγοριθμική διαδικασία για την αξιολόγηση της D-separation ενός συνόλου κόμβων. Για να εξετάσουμε τον αλγόριθμο BB πρέπει να δώσουμε κάποιους ορισμούς.

Ορισμός 1. Δομές V: (Scutari and Denis (2021)

$K \leftarrow L \rightarrow M$ Μια συγκλίνουσα σύνδεση $A_i \rightarrow A_j \leftarrow A_k$ ονομάζεται v-structure εάν δεν υπάρχει ακμή που να συνδέει τους A_i και A_j. Επιπλέον, η A_k συνήθως ονομάζεται κόμβος σύγκρουσης και η σύνδεση ονομάζεται μη θωρακισμένη σύγκρουση, σε αντίθεση με έναν θωρακισμένο κόμβο σύγκρουσης στον οποίο είτε ισχύει $A_i \rightarrow A_j$ είτε $A_i \leftarrow A_j$.

Ορισμός 2. Μη κατευθυνόμενη διαδρομή: Οποιαδήποτε ακολουθία κόμβων μεταξύ ενός μέλους του Α και ενός μέλους του Β, έτσι ώστε κάθε γειτονικό ζεύγος κόμβων να συνδέεται με μια ακμή, χωρίς να λαμβάνεται υπόψη η κατεύθυνση και εάν κανένας κόμβος δεν εμφανίζεται στην ακολουθία δύο φορές, τότε το μονοπάτι μεταξύ των σύνολων των κόμβων Α και Β ονομάζεται μη κατευθυνόμενη διαδρομή.

Ορισμός 3. Μπλοκαρισμένη διαδρομή: αν υπάρχει ένας κόμβος C στο μονοπάτι για τον οποίο ισχύει τουλάχιστον μία από τις ακόλουθες τρεις συνθήκες, το μονοπάτι ονομάζεται μπλοκαρισμένο μονοπάτι:

◇ O C βρίσκεται στο A και ο C έχει μία ακμή στο μονοπάτι που οδηγεί προς τα μέσα και μία ακμή προς τα έξω (αλυσίδα).

◇ Η C είναι στο A και η C έχει και τις δύο ακμές του μονοπατιού που οδηγούν προς τα έξω (κοινή αιτία).

◇ Ούτε ο C ούτε κανένας απόγονος του C βρίσκεται στο A και οι δύο ακμές του μονοπατιού οδηγούν στο C (κοινό αποτέλεσμα).

Λέγεται ότι ένα μη κατευθυνόμενο μονοπάτι P διαχωρίζεται κατά D από ένα σύνολο κόμβων A το οποίο περιέχει το αποδεικτικά στοιχεία εάν και μόνο εάν ισχύει τουλάχιστον μία από τις επόμενες συνθήκες:

◇ Μια αλυσίδα περιέχεται στο P, u → s → v ή

◇ Μια σκηνή ή διακλάδωση περιέχεται στο P, u↙s↘v, όπου s∈A

◇ Μια σύγκρουση ή δομή v περιέχεται στο P, u↘s↙v, όπου s δεν είναι στο A και ούτεείναι κανένας απόγονος του s.

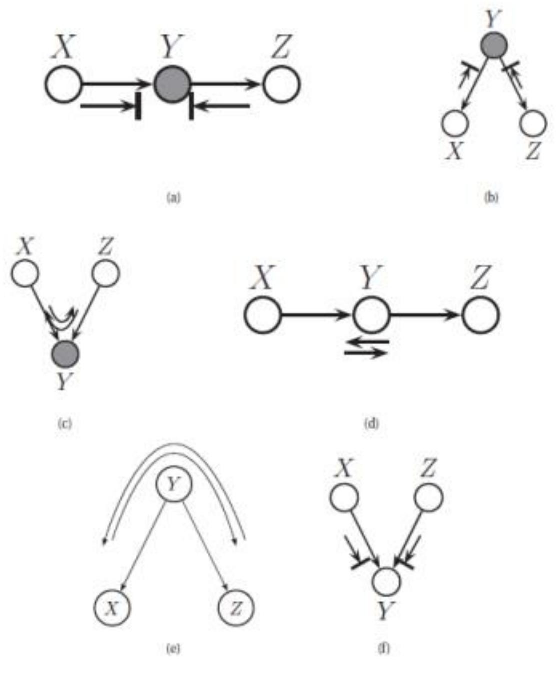
Επιπλέον, ένα σύνολο κόμβων N διαχωρίζεται κατά D από ένα διαφορετικό σύνολο κόμβων N1 δεδομένο ένα τρίτο παρατηρούμενο σύνολο N2 αν και μόνο αν κάθε μη κατευθυνόμενη διαδρομή από κάθε κόμβο n∈N σε κάθε κόμβο n1 ∈N1 είναι D-διαχωρισμένο από το N. Τέλος, οι ιδιότητες της υπό συνθήκη ανεξαρτησίας ενός DAG ορίζονται παρακάτω:

$$x_D \perp_G x_E \mid x_F \Leftrightarrow N \qquad (1.5)$$

Ένας απλός τρόπος για να δούμε αν το N διαχωρίζεται κατά D από το N1 με δεδομένο το N2 είναι ο αλγόριθμος Bayesball (Shachter 1998), ο οποίος βασίζεται στον ορισμό που δόθηκε παραπάνω. Η ιδέα της σφαίρας Bayes αλγορίθμου είναι η εξής: "σκιάζουμε" τους κόμβους του N2 για να υποδείξουμε τους παρατηρούμενους κόμβους. Στη συνέχεια τοποθετούνται "μπάλες" σε κάθε κόμβο του N, τις οποίες αφήνουμε να "αναπηδήσουν" ανάλογα με κάποιους κανόνες, και κοιτάμε αν κάποια από τις μπάλες φτάσει σε κάποιον από τους κόμβους του N1.

Στο σχήμα 1.5 φαίνονται οι τρεις βασικοί κανόνες που δόθηκαν παραπάνω. Κάθε μπάλα μπορεί να ταξιδέψει προς την αντίθετη κατεύθυνση των τόξων. Όπως βλέπουμε οι μπάλες μπορούν να περάσουν μέσα από μια αλυσίδα, αλλά όχι αν η αλυσίδα είναι σκιασμένη στη μέση. Με παρόμοιο τρόπο οι μπάλες μπορούν να περάσουν μέσα από ένα πιρούνι, αλλά όχι αν το πιρούνι είναι σκιασμένο στη μέση. Οι μπάλες δεν μπορούν να περάσουν μέσα από δομές v, εκτός εάν η δομή v είναι σκιασμένη στη μέση. Οι 3 κανόνες του αλγορίθμου μπάλας Bayes μπορούν να δικαιολογηθούν όπως φαίνεται στο παρακάτω σχήμα. Πρώτον, υποθέστε ότι έχουμε μια δομή αλυσίδας $K \to L \to M$, η οποία κωδικοποιεί

$$p(k,l,m) = p(k)p(l|m)p(m|l) \tag{1.6}$$



Σχήμα 1.5: Κανόνες Bayes-ball (Murphy, K. P. (2012),[15]) .

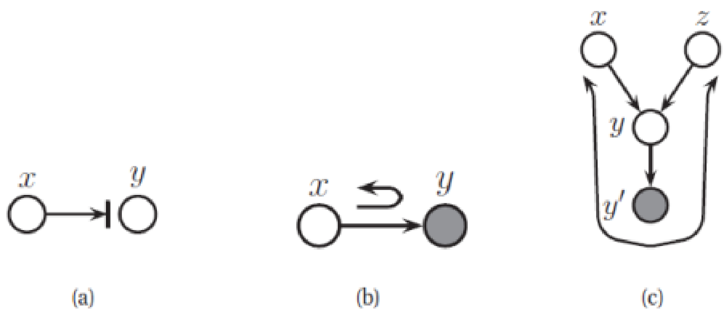
Οι σκιασμένοι κόμβοι είναι κόμβοι στους οποίους θέτουμε όρους. Η μπάλα δεν μπορεί να περάσει αν υπάρχει μια ακμή που χτυπάει μια μπάρα. Σε κάθε άλλη περίπτωση η μπάλα μπορεί να περάσει.

Υπό τον όρο l, είναι τα k και m ανεξάρτητα; Έχουμε

$$p(k,m\mid l) = p(k)p(l\mid k)p(m\mid l)p(l) = p(k,l)p(m\mid l)p(l) = p(k\mid l)p(m\mid l) \quad (1.7)$$

και αυτό δίνει $k \perp m|l$. Αυτό σημαίνει ότι η παρατήρηση του μεσαίου κόμβου μιας αλυσίδας σπάει την αλυσίδα στα δύο, όπως συμβαίνει σε μια αλυσίδα Markov. Εξετάζοντας στη συνέχεια τη δομή της σκηνής $K \leftarrow L \rightarrow M$. Η κοινή κατανομή δίνεται ως εξής

$$p(k,l,m) = p(l)p(k\mid l)p(m\mid l) \quad (1.8)$$



Σχήμα 1.6: (α-β) Οριακές συνθήκες μπάλας Bayes. (γ) Παράδειγμα του γιατί χρειάζονται οριακές συνθήκες. Το l′ είναι το παρατηρούμενο παιδί του l, καθιστώντας το l "ουσιαστικά παρατηρούμενο" και ως αποτέλεσμα, η μπάλα αναπηδά ξανά στο δρόμο της από το k στο m (Murphy, K. P. (2012),[15]) .

Είναι τα k και m ανεξάρτητα όταν εξαρτούμε το l; Έχουμε

$$p(k,m\mid l) = p(k,l,m)p(l) = p(l)p(k\mid l)p(m\mid l)p(l) = p(k\mid l)p(m\mid l) \quad (1.9)$$

ως αποτέλεσμα $k \perp m|l$. Αυτό σημαίνει ότι η παρατήρηση ενός κόμβου ρίζας διαχωρίζει τα παιδιά της ρίζας όπως συμβαίνει σε έναν ταξινομητή naiveBayes. Τέλος, υποθέτοντας ότι έχω μια v-δομή K →L ← M, η κοινή κατανομή δίνεται ως εξής

$$p\left(k,\ l,\ m\right) = p\left(k\right)p\left(m\right)p\left(l \mid k,\ m\right) \quad \left(1.10\right)$$

Είναι τα k και m ανεξάρτητα και εξαρτημενό το l; Έχουμε

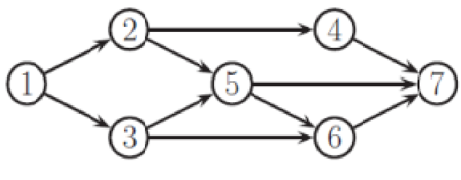
$$p(k, m \mid l) = \frac{p(k)\,p(m)\,p(l \mid k, m)}{p(l)} \tag{1.11}$$

άρα $k \not\perp m|l$. Αν και, στην από κοινού κατανομή, έχουμε

$$p(k, m) = p(k)\,p(m) \tag{1.12}$$

αυτό σημαίνει ότι τα k και m είναι οριακά ανεξάρτητα.

Ως αποτέλεσμα οι γονείς ενός κόμβου γίνονται ανεξάρτητοι αν θέσουμε ως προϋπόθεση ένα κοινό παιδί στο κάτω μέρος μιας ν-δομής. Αυτό το ονομάζουμε εξήγηση του σημαντικού αποτελέσματος, δια-αιτιακό συλλογισμό ή το παράδοξο του Berkson. Ας υποθέσουμε ότι, για παράδειγμα, ρίχνουμε δύο νομίσματα που αντιπροσωπεύουν τους δυαδικούς αριθμούς 0 και 1 το "άθροισμα" των τιμών τους παρατηρείται. Τα νομίσματα είναι ανεξάρτητα εκ των προτέρων, αλλά μόλις το άθροισμά τους παρατηρηθεί, γίνονται συνδεδεμένα (για παράδειγμα, αν το άθροισμά τους είναι 1 και η τιμή του πρώτου νομίσματος είναι 1, τότε γνωρίζουμε ότι η τιμή του δεύτερου νομίσματος είναι 0). Τέλος, η μπάλα Bayes χρειάζεται επίσης το όριο "συνθήκες", οι οποίες παρουσιάζονται στο σχήμα 1.6(α-β). Χρειάζεται να κοιτάξει κανείς το σχήμα 1.6(γ) για να καταλάβει από πού προέρχονται αυτοί οι κανόνες. Υποθέστε ότι η L' είναι ένα αντίγραφο του L χωρίς θόρυβο. Αν το Y' παρατηρείται, το L μπορεί να παρατηρηθεί επίσης, οπότε οι γονείς K και L πρέπει να ανταγωνίζονται για να το εξηγήσουν αυτό. Έτσι, στέλνοντας μια μπάλα προς τα κάτω $K \to L \to M'$, θα πρέπει να "αναπηδήσει" προς τα πάνω κατά μήκος $L' \to L \to M$. Παρόλο που, το L και όλα τα παιδιά του είναι κρυμμένα, η μπάλα δεν αναπηδά πίσω. Στα Σχήματα 1.5 και 1.6, το x συμβολίζει το k, το z το m, το y το l και το X το K, το Y το L και Z για το M.



Σχήμα 1.7: Ένα κατευθυνόμενο γραφικό μοντέλο (Murphy, K. P. (2012),[15]).

Στο Σχήμα 1.7, για παράδειγμα, όπως βλέπουμε $x_2 \perp x_6 \mid x_5$, καθώς το μονοπάτι $2 \rightarrow 5 \rightarrow 6$ έχει μπλοκαριστεί από το x_5 (το οποίο έχουμε παρατηρήσει) και το μονοπάτι $2 \rightarrow 4 \rightarrow 7 \rightarrow 6$ έχει μπλοκαριστεί από το x_7 (το οποίο είναι κρυφό) και το μονοπάτι $2 \rightarrow 1 \rightarrow 3 \rightarrow 6$ έχει μπλοκαριστεί από το x_1 (το οποίο είναι κρυμμένο). Παρόλα αυτά, μπορούμε επίσης να δούμε ότι $x_2 \not\perp x_6 \mid x_5, x_7$, καθώς το μονοπάτι $2 \rightarrow 4 \rightarrow 7 \rightarrow 6$ δεν εμποδίζεται πλέον από το x_7 (το οποίο παρατηρείται).

1.12 Markov blanket και πλήρεις συνθήκες

Κάθε σύνολο κόμβων V που καθιστά έναν κόμβο u υπό όρους ανεξάρτητο από όλους τους άλλους κόμβους σε ένα γράφημα, ονομάζεται $u's$ Markovblanket το οποίο συμβολίζεται με $mb(u)$. Μπορούμε να δείξουμε ότι το Markovblanket οποιουδήποτε κόμβου σε ένα κατευθυνόμενο γραφικό μοντέλο ισούται με τους γονείς, τα παιδιά, και τους συν-γονείς του κόμβου.

$$mb(u) = ch(u) \cup pa(u) \cup copa(u) \qquad (1.13)$$

Στο Σχήμα 1.7, για παράδειγμα, έχουμε

$$mb(5) = 6,\ 7 \cup 2,\ 3 \cup 4 = 2,\ 3,\ 4,\ 6,\ 7$$

όπου η 4 θεωρείται συν-γονέας της 5, δεδομένου ότι μοιράζονται το ίδιο παιδί, δηλαδή την 7. Σημειώνεται ότι με την εξαγωγή

$$p(x_u \mid x_{-u}) = \frac{p(x_u,\ x_{-u})}{p(x_{-u})}$$

μπορούμε να καταλάβουμε γιατί οι συν-γονείς βρίσκονται στην κουβέρτα Markov. Κάθε όρος που δεν περιλαμβάνει x_u θα ακυρωθεί μεταξύ του αριθμητή και του παρονομαστή, οπότε θα μείνουμε με ένα γινόμενο κατανομών υπό συνθήκη πιθανότητας στο πεδίο εφαρμογής του οποίου θα περιέχεται η x_u. Έτσι

$$p(x_u \mid x_{-u}) \propto p(x_u \mid x_{pa(u)}) \prod_{v \in ch(u)} p(x_v \mid x_{pa(v)})$$

Για παράδειγμα, στο Σχήμα1.7 έχουμε(τυπο)

$$p(x_5 \mid x_{-5}) \propto p(x_5 \mid x_2, x_3)\, p(x_6 \mid x_3, x_5)\, p(x_7 \mid x_4, x_5, x_6)$$

Η τελευταία έκφραση ονομάζεται $u's$ υπό όρους.

1.13 Ιδιότητες Markov των κατευθυνόμενων γραφικών μοντέλων

Από τον D-separation, μπορούμε να συμπεράνουμε ότι

$$u \perp \frac{nd(u)}{pa(u)} | pa(u) \qquad (1.14)$$

όπου οι μη απόγονοι ενός κόμβου $nd(u)$ είναι όλοι οι άλλοι κόμβοι εκτός από τους απογόνους του, $nd(u) = V \setminus u \cup desc(u)$. Η εξίσωση (1.11) ονομάζεται κατευθυνόμενη τοπική ιδιότητα Markov. Στο Σχήμα 1.7, για παράδειγμα, $nd(3) = 2,4$ και $pa(3) = 1$, οπότε έχουμε $3 \perp 2,4 | 1$. Όταν εξετάζουμε μόνο τους προκατόχους ενός κόμβου σύμφωνα με μια τοπολογική διάταξη έχουμε μια ειδική περίπτωση της παραπάνω ιδιότητας. Έτσι έχουμε(1.15)

$$u \perp \frac{pred(u)}{pa(u)} | pa(u) \qquad (1.15)$$

το οποίο προκύπτει από τη στιγμή που $pred(u) \subseteq nd(u)$ και ονομάζεται ιδιότητα OrderedMarkov. Στο Σχήμα1.7, για παράδειγμα, αν η ακόλουθη διάταξη 1, 2, ... , 7 θα διαπιστώσουμε ότι $pred(3) = \{1,2\}$ και $pa(3) = 1$ αυτό μας δίνει ότι $3 \perp 2 | 1$.

Τώρα έχουμε περιγράψει τις τρεις ιδιότητες Markov για κατευθυνόμενους ακυκλικούς γράφους: πρώτα την κατευθυνόμενη παγκόσμια ιδιότητα MarkovG στην Εξίσωση (1.7), δεύτερον την κατευθυνόμενη τοπική ιδιότητα MarkovL στην Εξίσωση (1.11), και τέλος, την ιδιότητα OrderedMarkovO στην Εξίσωση (1.12). Προφανώς $G \Rightarrow L \Rightarrow O$, αν και είναι λιγότερο προφανές ότι $O \Rightarrow L \Rightarrow G$ είναι αληθές (βλέπε π.χ. (KollerandFriedman2009)για την απόδειξη). Συνεπώς, όλες αυτές οι ιδιότητες είναι ισοδύναμες. Επιπλέον, οποιαδήποτε κατανομή p που είναι Markov ως προς την G μπορεί να παραγοντοποιηθεί όπως στην εξίσωση 1.4, την οποία ονομάζουμε ιδιότητα παραγοντοποίησηςF και προφανώς $O \Rightarrow F$, αλλά μπορούμε να δείξουμε ότι ισχύει και το αντίστροφο (βλ. π.χ., (KollerandFriedman 2009) για την απόδειξη).

Θεώρημα 1. Κλάσεις ισοδυναμίας: (Scutari and Denis (2021),[8]) Two Directed Acyclic γραφήματα τα οποία ορίζουν το ίδιο σύνολο μεταβλητών, λέγεται ότι είναι ισοδύναμα εάν και μόνο εάν έχουν τον ίδιο σκελετό. Αυτό σημαίνει ότι έχουν το ίδιο υποκείμενο μη κατευθυνόμενο γράφημα και οι v-δομές τους είναι ίδιες. Για να το θέσουμε αλλιώς, οι μόνες ακμές των οποίων οι κατευθύνσεις είναι σημαντικές είναι αυτές που αποτελούν μέρος μιας ή περισσότερων v-δομών.

Θεώρημα 2. Σύνθεση της Markov Blanket: Μια Markov Blanket είναι ένα σύνολο κόμβων A που αποτελείται από τους γονείς του A, τα παιδιά του A και όλους τους άλλους κόμβους που έχουν κοινό παιδί με τον A. Αυτό το θεώρημα προσδιορίζει τους κόμβους που πρέπει να συμπεριλάβουμε στο Markov blanket για να διαχωρίσουμε τον κόμβο-στόχο από τον υπόλοιπο κατευθυνόμενο ακυκλικό γράφο. Τα μονοπάτια που ικανοποιούν την πρώτη συνθήκη του ακόλουθου ορισμού απαιτείται να είναι μπλοκαριστούν από τους γονείς και τα παιδιά.

Ορισμός 4. Διαχωρισμός D: Αν έχουμε τρία υποσύνολα κόμβων σε ένα κατευθυνόμενο ακυκλικό γράφημα G που είναι διαχωρισμένα, D, E και F, τότε το F είναι D-διαχωρισμένο D από το E, το οποίο εμείς συμβολίζουμε ως $D \perp_G E | F$, αν σε κάθε διαδρομή μεταξύ ενός κόμβου στο D και ενός κόμβου στο E υπάρχει ένας κόμβος u που ικανοποιεί μία από τις επόμενες συνθήκες:

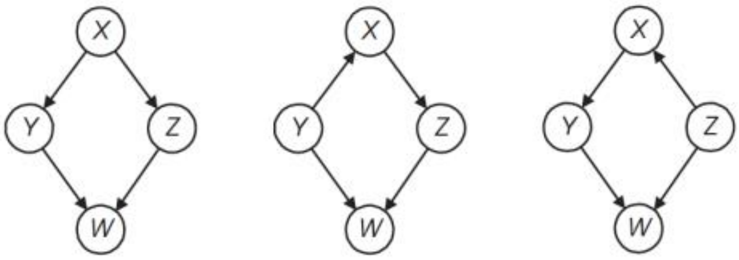
◊ ο u έχει συγκλίνουσες ακμές (δηλαδή, από τους γειτονικούς κόμβους στο μονοπάτι έχουμε δύο ακμές που δείχνουν στον u) και ο u δεν έχει ούτε κανέναν από τους απογόνους του (δηλαδή κόμβους στους οποίους μπορεί να φτάσει κανείς από τον u) είναι στο F.

◊ Το u ανήκει στο F και δεν έχει συγκλίνουσες ακμές.

Οι κόμβοι που έχουν κοινό παιδί με τον κόμβο-στόχο, συνήθως ονομάζονται σύζυγοι και απαιτείται να μπλοκάρουν τα μονοπάτια που ικανοποιούν τη δεύτερη συνθήκη. Είναι γνωστό ότι ο ίδιος ο κόμβος-στόχος δεν αποτελεί μέρος του MarkovBlanket του.

Επακόλουθο 1. Ως αποτέλεσμα του παραπάνω θεωρήματος έχουμε την ακόλουθη συμμετρική σχέση: αν ο κόμβος D βρίσκεται στο MarkovBlanket του E, τότε ο E βρίσκεται στο MarkovBlanket του D.

Ορισμός 5. Ισοδυναμία Markov: Πολλοί κατευθυνόμενοι ακυκλικοί γράφοι είναι ισοδύναμοι που σημαίνει ότι έχουν τα ίδια D-διαχωριστικά. Κάθε ένας από τους κατευθυνόμενους ακυκλικούς γράφους στο επόμενο σχήμα, για παράδειγμα, έχει τις ακόλουθες D-διαχωρισμούς $I_G(\{X\},\{W\}|\{Y,Z\})$ και $I_G(\{Y\},\{Z\}|\{X\})$,και είναι οι μόνες D-διαχωρισμοί που έχουν.



Σχήμα 1.8: Αυτοί οι κατευθυνόμενοι ακυκλικοί γράφοι είναι ισοδύναμοι κατά Markov και δεν υπάρχουν άλλοι κατευθυνόμενοι άκυκλοι γράφοι Markov ισοδύναμοι με αυτούς (Neapolitan, R. E. (2004)).

Θεωρήστε ότι $G=(V,E)$ και $G_1=(V,E_1)$ είναι δύο κατευθυνόμενοι ακυκλικοί γράφοι που περιέχουν το ίδιο σύνολο μεταβλητών V. Τότε ονομάζουμε τα G και G_1 Markov ισοδύναμα αν για κάθε τρεις αμοιβαία διαχωρισμένα υποσύνολα $K,L,M \subseteq V, K$ και L διαχωρίζονται D από το M στο G αν και μόνο αν αν τα K και L είναι D-διαχωρισμένα από το M στο G_1.

Αυτόσημαίνει

$$I_G(K,L\,|\,M) \Leftrightarrow I_{G_1}(K,L\,|\,M)$$

Ο παραπάνω ορισμός έχει να κάνει μόνο με τις ιδιότητες του γράφου, αν και η εφαρμογή του είναι σε πιθανότητες λόγω του επόμενου θεωρήματος:

Θεώρημα 3. Ισοδύναμοι κατά Markov κατευθυνόμενοι ακυκλικοί γράφοι: (Neapolitan, R. E. (2004),[17])

Λέμε ότι δύο κατευθυνόμενοι ακυκλικοί γράφοι είναι ισοδύναμοι κατά Markov εάν και μόνο εάν, απαιτούν τις ίδιες υπό συνθήκη ανεξαρτησίες με βάση τη συνθήκη Markov.

Πόρισμα 2. Έστω $G=(V,E)$ και $G_1=(V,E_1)$ δύο κατευθυνόμενοι ακυκλικοί γράφοι οι οποίοι περιέχουν το ίδιο σύνολο μεταβλητών V . Τότε έχουμε ότι τα G και G_1 αν και μόνο αν για κάθε κατανομή πιθανότητας P της V , το (G, P) ικανοποιεί τη συνθήκη Markov αν και μόνο αν (G_1, P) ικανοποιεί τη συνθήκη Markov τα G και G_1 είναι ισοδύναμα Markov.

1.14 Markov Blankets και όρια

Ένας μεγάλος αριθμός κόμβων μπορεί να εμπλέκεται σε ένα Μπεϋζιανό Δίκτυο, και η πιθανότητα ενός γνωστού κόμβου επηρεάζεται κατά τη δημιουργία ενός απομακρυσμένου κόμβου. Παρόλο που, η ενστάλαξη ενός συνόλου κοντινών κόμβων μπορεί να θωρακίσει έναν κόμβο από την επιρροή όλων των άλλων κόμβων. Μπορούμε να το δούμε αυτό από τους επόμενους ορισμούς:

Ορισμός 6. Θεωρήστε ότι V είναι ένα σύνολο τυχαίων μεταβλητών, P είναι η κοινή τους πιθανότητα και $K \in V$. Τότε το Markov BlanketMA_K του K είναι οποιοδήποτε σύνολο μεταβλητών έτσι ώστε η k να είναι υπό όρους ανεξάρτητη από όλες τις άλλες μεταβλητέςδεδομένης της MA_K. Αυτό σημαίνει ότι

$$I_P\left(\left(\{X\},\, V - \left(MA_K \cup \{K\}\right) \mid MA_K\right)\right).$$

Ορισμός 7. Έστω ότι V είναι ένα σύνολο τυχαίων μεταβλητών, P είναι η κοινή κατανομή πιθανότητας τους και $K \in V$. Οποιοδήποτε κενό Markov, έτσι ώστε καμία από τις $K's$ κατάλληλων υποσυνόλων του, είναι ένα Markov κουβέρτα του K ονομάζεται όριο Markov του K.

Κλείνοντας το πρώτο κεφάλαιο μπορούμε να πούμε ότι κάναμε μια εισαγωγή στους ορισμούς, στα διαγράμματα και στους τρόπους που χρησιμοποιούνται τα Μπεϋζιανά Δίκτυα και στα διαφορετικά εργαλεία που χρησιμοποιούν για να τα βοηθήσουν.
Στη συνέχεια θα δούμε πως μπορούμε να εξάγουμε συμπεράσματα μέσω των Μπεϋζιανών Δικτύων.

ΚΕΦΑΛΑΙΟ 2

Μπεϋζιανή Συμπερασματολογία

Όπως είδαμε, τα Γραφικά Μοντέλα παρέχουν έναν σταθερό τρόπο για τον ορισμό κοινών κατανομών πιθανοτήτων. Μια τέτοια κοινή κατανομή χρησιμοποιείται για πιθανολογική συμπερασματολογία. Αυτό σημαίνει ότι μπορούμε να εκτιμήσουμε ποσότητες που είναι άγνωστες από ποσότητες που είναι γνωστές. Γενικά, μπορούμε να σκεφτούμε ένα πρόβλημα συμπερασμάτων όπως περιγράφεται στη συνέχεια.

Ας υποθέσουμε ότι έχουμε ένα σύνολο τυχαίων μεταβλητών που συσχετίζονται και έχουν κοινή κατανομή $p(y_1 : V \mid \theta)$. Σε ό,τι ακολουθεί, υποθέτουμε ότι οι παράμετροι θ του μοντέλου είναι γνωστές και αργότερα θα συζητήσουμε τον τρόπο που οι παράμετροι μπορούν να μαθευτούν. Χωρίζουμε το διάνυσμα στις ορατές μεταβλητές y_u, οι οποίες έχουν παρατηρηθεί, και στις κρυφές μεταβλητές, y_h, οι οποίες δεν έχουν παρατηρηθεί. Η συμπερασματολογία αναφέρεται στον υπολογισμό της εκ των υστέρων κατανομής των άγνωστων μεταβλητών δεδομένων των γνωστών μεταβλητών:

$$p(y_h \mid y_u, \theta) = \frac{p(y_h, y_u \mid \theta)}{p(y_u \mid \theta)} = \frac{p(y_h, y_u \mid \theta)}{\sum_{y_u'} p(y_h', y_u \mid \theta)}$$

Στην πραγματικότητα θέτουμε όρους στα δεδομένα περιορίζοντας τις ορατές μεταβλητές στις παρατηρούμενες τιμές, y_u, και στη συνέχεια κανονικοποιούμε για να πάμε από $p(y_h, y_u)$ σε $p(y_h \mid y_u)$, $p(y_u \mid \theta)$. Είναι η σταθερά κανονικοποίησης, η οποία θεωρείται η πιθανότητα των δεδομένων και ονομάζεται επίσης πιθανότητα των στοιχείων.

Μερικές φορές δεν μας ενδιαφέρουν όλες οι κρυφές μεταβλητές. Αυτό σημαίνει ότι χωρίζουμε τις κρυφές μεταβλητές σε μεταβλητές ερωτήματος, y_q, των οποίων τις τιμές θα θέλαμε να γνωρίζουμε, και τις υπόλοιπες αδιάφορες μεταβλητές, y_n, οι οποίες δεν μας ενδιαφέρουν. Η περιθωριοποίηση των είναι ένας τρόπος για να υπολογίσουμε αυτό που μας ενδιαφέρει :

$$p(y_q \mid y_u, \theta) = \sum_{y_n} p(y_q, y_n \mid y_u, \theta)$$

2.1 Ορισμός των τυχαίων μεταβλητών και της κοινής πιθανότητας Κατανομές για Μπεϋζιανή Συμπερασματολογία

Μπορούμε να ορίσουμε μια τυχαία μεταβλητή Y ως ένα σύμβολο που αντιπροσωπεύει οποιοδήποτε σύνολο τιμών που ονομάζουμε χώρο της Y . Θεωρούμε ότι ο χώρος της Y είναι μετρήσιμος, αλλά η θεωρία μπορεί φυσικά να επεκταθεί στην περίπτωση που δεν είναι. Η τυχαία μεταβλητή LungCancer, για παράδειγμα, θα μπορούσε να αναγνωριστεί ότι έχει το χώρο {παρουσία, απουσία}. Ο ακόλουθος συμβολισμός Y = y χρησιμοποιείται ως ένα πρωτότυπο που χρησιμοποιείται σε εκφράσεις πιθανότητας. Αυτό σημαίνει ότι Y = y θα δεν θα οριστεί ως προς οτιδήποτε άλλο. Στην εφαρμογή LungCancer = present δηλώνει ότι το υπό εξέταση στοιχείο πάσχει από καρκίνο του πνεύμονα, αλλά μαθηματικά είναι απλώς ένα αρχικό στοιχείο που χρησιμοποιούμε σε εκφράσεις πιθανότητας. Δεδομένου του παραπάνω ορισμού και πρωταρχικό, έχουμε τον ακόλουθο άμεσο ορισμό μιας κοινής κατανομής πιθανοτήτων.

__Ορισμός 8__. Θεωρήστε ένα σύνολο m τυχαίων μεταβλητών $V=\{Y_1,Y_2,...,Y_m\}$ που καθορίζεται έτσι ώστε κάθε Y_i να έχει (για απλότητα) ένα μετρήσιμο άπειρο διάστημα. Ονομάζουμε μια κοινή κατανομή πιθανότητας των τυχαίων μεταβλητών στο V μια συνάρτηση που αποδίδει έναν πραγματικό αριθμό $P(Y_1=y_1,Y_2=y_2,...,Y_m=y_m)$ σε κάθε συνδυασμό τιμών των y_i έτσι ώστε η τιμή του y_i μπορεί να επιλεγεί από το χώρο των Y_i εάν ικανοποιεί τις επόμενες συνθήκες:

◊ Για κάθε συνδυασμό τιμών των $y_i' s$ έχουμε

$$0 \le P(Y_1=y_1,Y_2=y_2,...,Y_m=y_m) \le 1$$

◊Έχουμε

$$\sum_{y_1,y_2,...,y_m} P(Y_1=y_1,Y_2=y_2,...,Y_m=y_m)=1$$

Με τον συμβολισμό $\sum_{y_1,y_2,...,y_m}$ εννοούμε το άθροισμα καθώς οι μεταβλητές $y_1,y_2,...,y_m$ λαμβάνουν όλες τις δυνατές τιμές στους αντίστοιχους χώρους τους.

Ένας τρόπος για να δημιουργήσουμε μια κοινή πιθανότητα κατανομής που ικανοποιεί αυτόν τον ορισμό είναι με μια κοινή κατανομή πιθανότητας που προκύπτει από τον ορισμό τυχαίων μεταβλητών ως συναρτήσεων σε έναν δειγματικό χώρο.

Ένα κλασικό παράδειγμα Μπεϋζιανής Συμπερασματολογίας

Τα επερχόμενα παραδείγματα ερμηνεύουν πώς το θεώρημα του Bayes έχει εφαρμοστεί για τον υπολογισμό της πιθανότητας ενός γεγονότος που μας ενδιαφέρει από πιθανότητες που είναι η γνωστές.

Ας υποθέσουμε ότι η Μαίρη πρέπει να πάει για μια διαγνωστική ακτινογραφία θώρακα ρουτίνας, την οποία απαιτεί η ColonialBank. από όλους τους νέους υπαλλήλους και η ακτινογραφία δίνει θετικό αποτέλεσμα για καρκίνο του πνεύμονα. Η Μαίρη τότε είναι σίγουρη ότι έχει καρκίνο του πνεύμονα και παθαίνει κρίση πανικού. Το ερώτημα εδώ είναι "θα έπρεπε να το κάνει;" Χωρίς όμως, να γνωρίζει πόσο ακριβές είναι το τεστ, δεν υπάρχει κανένας τρόπος για τη Μαίρη να ξέρει πόσο πιθανό είναι είναι να έχει καρκίνο του πνεύμονα. Το τεστ δεν είναι απόλυτα πειστικό και όταν το ανακαλύπτει αποφασίζει να ερευνήσει πόσο ακριβές είναι πραγματικά το τεστ και τότε ανακαλύπτει ότι έχει ένα ψευδώς θετικό ποσοστό 0,02 και ένα ψευδώς αρνητικό ποσοστό 0,4. Αυτή η ακρίβεια αντιπροσωπεύεται ως εξής. Πρώτα απ' όλα πρέπει να οριστούν αυτές οι τυχαίες μεταβλητές:

Πίνακας 2.1: Ορισμός των μεταβλητών

Μεταβλητές	Τιμές	Όταν οι μεταβλητές παίρνουν αυτές τις τιμές
Τεστ	Θετικό Αρντικό	X-ray είναι θετική X-ray είναι αρντική
ΚαρκίνοςτουΠνεύμονα (LungCancer)	Παρών Απών	Lung cancer είναιπαρών Lung cancer είναιαπών

Τότε έχουμε τις ακόλουθες δεσμευμένες πιθανότητες:

$$P(\text{test} = \text{positive} \mid \text{lungCancer} = \text{present}) = 0.6$$
$$P(\text{test} = \text{positive} \mid \text{lungCancer} = \text{absent}) = 0.02$$

Δεδομένων των παραπάνω πιθανοτήτων, η Μαρία αισθάνεται κάπως καλύτερα. Αν και, στη συνέχεια, συνειδητοποιεί ότι εξακολουθεί να μην γνωρίζει πόσο πιθανό είναι να έχει καρκίνο του πνεύμονα. Έτσι, η πιθανότητα ότι η Μαίρη έχει καρκίνο του πνεύμονα είναι P(καρκίνος του πνεύμονα = παρών | εξέταση = θετική), και αυτή η πιθανότητα δεν αναφέρεται παραπάνω. Τελικά, η Μαίρη θυμάται το θεώρημα του Bayes και συνειδητοποιεί την ανάγκη της άλλης πιθανότητας για να μπορέσει να προσδιορίσει την πιθανότητα να έχει καρκίνο του πνεύμονα. Η ακόλουθη πιθανότητα P(lungCancer = present) είναι η πιθανότητα να έχει καρκίνο του πνεύμονα, πριν λάβει οποιαδήποτε πληροφορία από τα αποτελέσματα της εξέτασης. Παρόλο που η πιθανότητα αυτή έχει δεν έχει βασιστεί σε καμία πληροφορία σχετικά με τα αποτελέσματα της εξέτασης, έχει βασιστεί σε κάποιες πληροφορίες. Πιο συγκεκριμένα, βασίζεται σε όλες τις πληροφορίες που είναι σχετικές με τον καρκίνο του πνεύμονα και είναι γνωστές για τη Μαίρη πριν από τη διενέργεια της εξέτασης. Πριν η Μαίρη κάνει το τεστ, η μόνη πληροφορία γι' αυτήν ήταν ότι ήταν μία από τους νέους υπαλλήλους που έπρεπε να κάνουν το τεστ που απαιτούνταν από αυτούς. Αυτό σημαίνει ότι, όταν μαθαίνει μόνο 1 σε 1000 νέους υπαλλήλους ότι έχει καρκίνο του πνεύμονα, αποδίδει 0,001 στο P(lungCancer = παρών). Στη συνέχεια, χρησιμοποιεί το θεώρημα του Bayes, όπως δίνεται στη συνέχεια:

$$P(present \mid positive) = \frac{P(positive \mid present)P(present)}{P(positive \mid present)P(present) + P(positive \mid absent)P(absent)} = \frac{(0.6)(0.001)}{(0.6)(0.001)+(0.02)(0.999)} = 0.029$$

Η Μαίρη αισθάνεται τώρα ότι η πιθανότητα να έχει καρκίνο του πνεύμονα είναι μόλις 0,03 και χαλαρώνει κάπως, ενώ περιμένει το αποτέλεσμα των περαιτέρω εξετάσεων.

Πιθανότητες όπως η P(καρκίνος του πνεύμονα = παρών) ονομάζονται προηγούμενες πιθανότητες αφού σε ένα συγκεκριμένο μοντέλο είναι οι πιθανότητες ενός γεγονότος πριν από την ενημέρωση των πιθανοτήτων του γεγονότος αυτού, που περιλαμβάνονται στο πλαίσιο του συγκεκριμένου μοντέλου με τη χρήση νέων πληροφοριών. Πιθανότητες όπως P(lungCancer = present|test =positive) ονομάζονται εκ των υστέρων πιθανότητες, καθώς είναι οι πιθανότητες κάποιου γεγονότος μετά την ενημέρωση της προηγούμενης πιθανότητάς του, που περιλαμβάνεται στο πλαίσιο του εν λόγω μοντέλου χρησιμοποιώντας νέες πληροφορίες.

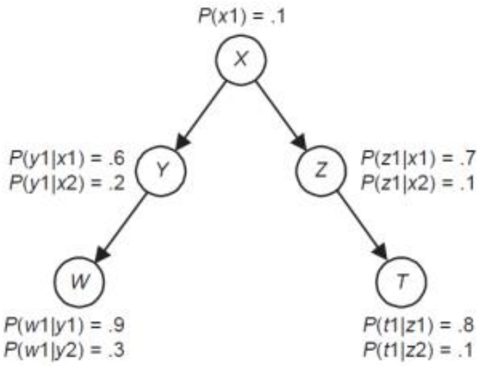
Το επόμενο παράδειγμα ερμηνεύει τον τρόπο με τον οποίο οι εκ των προτέρων πιθανότητες μπορούν να αλλάξουν και αυτό εξαρτάται για την κατάσταση που θέλουμε να μοντελοποιήσουμε.

Παράδειγμα:

Υποθέτουμε τώρα ότι η Σαμάνθα έχει την ίδια διαγνωστική ακτινογραφία θώρακα με τη Μαίρη. Έχει την ακτινογραφία δεδομένου ότι εργάζεται στα ορυχεία εδώ και 20 χρόνια και οι εργοδότες της ανησύχησαν όταν διαπίστωσαν ότι περίπου το 10% όλων των εργαζομένων αναπτύσσουν καρκίνο του πνεύμονα μετά από πολλά χρόνια στα ορυχεία. Οι εξετάσεις της Σαμάνθα δίνουν επίσης θετικά αποτελέσματα. Θα θέλαμε όμως, να γνωρίζουμε την πιθανότητα να έχει καρκίνο του πνεύμονα. Με βάση τις πληροφορίες που είχαμε για τη Σαμάνθα πριν κάνει το τεστ, είμαστε σε θέση να αποδώσουμε μια εκ των προτέρων πιθανότητα 0,1 η Σαμάνθα να έχει καρκίνο του πνεύμονα. Όπως και πριν, χρησιμοποιώντας το θεώρημα του Bayes, μπορούμε να φτάσουμε στο συμπέρασμα ότι P(καρκίνος του πνεύμονα = παρών|τεστ = θετικό) = 0,769 για τη Σαμάνθα. Έτσι, η Σαμάνθα συμπεραίνει ότι είναι πολύ πιθανό να έχει καρκίνο του πνεύμονα. Τα δύο παραπάνω παραδείγματα μας δείχνουν ότι μια τιμή πιθανότητας είναι σχετική με κάποιες πληροφορίες για ένα γεγονός και ότι δεν αποτελεί ιδιότητα του ίδιου του γεγονότος. Τόσο η Μαίρη όσο και η Σαμάνθα μπορεί να έχουν ή να μην έχουν καρκίνο του πνεύμονα. Είναι πιθανό η Μαρία να έχει και η Σαμάνθα να μην έχει. Παρόλο που, με βάση τις πληροφορίες που έχουμε, ο βαθμός πεποίθησής μας (πιθανότητα) ότι η Σαμάνθα δεν τον έχει, είναι πολύ μεγαλύτερος από τον βαθμό πεποίθησής μας ότι τον έχει η Μαίρη. Όταν αποκτήσουμε περισσότερες πληροφορίες σε σχέση με το γεγονός, για παράδειγμα αν η Μαίρη καπνίζει ή έχει ιστορικό καρκίνου στην οικογένεια, η πιθανότητα θα αλλάξει.

Αλγόριθμος μετάδοσης μηνυμάτων του Pearl:

O Pearl [1986, 1988] ανέπτυξε έναν αλγόριθμο διακίνησης μηνυμάτων για να μπορεί να βγάλει συμπέρασμα σε BayesianNetworks εκμεταλλευόμενος τις τοπικές ανεξαρτησίες. Ας υποθέσουμε ότι έχουμε ένα σύνολο v από τιμών ενός συνόλου V από ενσταντανέ μεταβλητές, ο αλγόριθμος ορίζει τις πιθανότητες P(y | v) για όλες τις τιμές y κάθε μεταβλητής Y στο δίκτυο. Αυτό επιτυγχάνεται με την έναρξη μηνυμάτων από τις ενσαρκωμένες μεταβλητές προς τους γείτονές τους. Στη συνέχεια, οι γείτονες αυτοί δίνουν μηνύματα στους δικούς τους γείτονες. Τα αποδεικτικά στοιχεία μπορούν να φτάσουν με οποιαδήποτε σειρά, δεδομένου ότι η ενημέρωση δεν επηρεάζεται από τη σειρά με την οποία ξεκινάμε αυτά τα μηνύματα. Στην αρχή ο αλγόριθμος αναπτύσσεται για BayesianNetworks, των οποίων οι κατευθυνόμενοι ακυκλικοί γράφοι είναι ριζωμένα δέντρα και στη συνέχεια ο αλγόριθμος επεκτείνεται σε δίκτυα με απλή σύνδεση.



Σχήμα 2.1: (Neapolitan, R. E. (2004),[17])Εδώ έχουμε το δίκτυο Bayes που είναι ένα δέντρο. Κάθε μεταβλητή μπορεί να πάρει μόνο δύο πιθανές τιμές και γι' αυτό μόνο η πιθανότητα της αναπαρίσταται.

2.1.1 Συμπερασματολογία σε δέντρα

Ονομάζουμε ένα δέντρο με ρίζα ένα κατευθυνόμενο ακυκλικό γράφημα όπου υπάρχει ένας μοναδικός κόμβος που ονομάζεται ρίζα, ο οποίος δεν έχει γονείς, κάθε άλλος κόμβος έχει ακριβώς έναν γονέα και κάθε κόμβος είναι απόγονος της ρίζας. Με βάση το ακόλουθο θεώρημα παίρνουμε τον αλγόριθμο.

Θεώρημα 4. Ας υποθέσουμε ότι έχουμε ένα Μπεϋζιανό Δίκτυο (G,P) του οποίου ο Κατευθυνόμενος Ακυκλικός Γράφος είναι ένα δέντρο και G = (V, A), και c είναι ένα σύνολο τιμών ενός υποσυνόλου $C \subset A$. Για κάθε μεταβλητή Y, ορίζουμε λ μηνύματα, λ τιμές, π μηνύματα και π τιμές όπως φαίνεται στη συνέχεια:

◊ λ μηνύματα: Για κάθε παιδί Z του Y, για όλες τις τιμές του y,

$$\lambda_Z(y) = \sum_z P(z \mid y)\lambda(z)$$

◊ τιμές λ: Αν $Y \in C$ και η τιμή του Yyb,

$$\lambda(y) \equiv 1$$
$$\lambda(y) \equiv 0 \; \gamma\iota\alpha \;\; y \neq y$$

Αν $Y \notin C$ και Y είναι ένα φύλλο, για κάθε τιμή του y τότε $\lambda(y) \equiv 1$.

Αν $Y \notin C$ και Y δεν είναι ένα φύλλο, για κάθε τιμή του y τότε

$$\lambda(y) \equiv \prod_{U \in CH_Y} \lambda_U(y)$$

Οπού CH_Y συμβολίζει το σύνολο των παιδιών του Y

◊ π μηνύματα:

Αν ο X είναι ο γονέας του Y, τότε για όλες τις τιμές του x,

$$\pi_Y(x) \equiv \pi(x) \prod_{U \in CH_{x} - \{Y\}} \lambda_U(x)$$

◊ τιμές π:

Εάν $Y \in C$ και η τιμή του $Y \; \varepsilon\acute{\iota}\nu\alpha\iota \; y$

$$\pi(y) \equiv 1$$
$$\pi(y) \equiv 0 \; \gamma\iota\alpha \;\; y \neq y$$

Αν $Y \notin C$ και Y είναι η ρίζα για κάθε τιμή του y τότε

$$\pi(y) \equiv P(y)$$

Αν $Y \in C$ το Y δεν είναι ρίζα και το X είναι γονέας του Y για κάθε τιμή του y τότε

$$\pi(y) \equiv \sum_x P(y \mid x) \pi_Y(x)$$

◊ Εφόσον όλα τα παραπάνω ορίζονται, για κάθε μεταβλητή Y τοτε έχουμε για κάθε τιμή του y

$$P(y \mid c) = \alpha \lambda(y) \pi(y)$$

όπου α είναι η σταθερά κανονικοποίησης.

Αλγόριθμος-Συμπεράσματα σε δέντρα:

Αλγόριθμος: Αυτός ο αλγόριθμος είναι η εφαρμογή του αλγορίθμου του Pearl σε δέντρα.

Πρόβλημα:

Υποθέτοντας ότι έχουμε ένα BayesianNetwork του οποίου το DirectedAcyclicGraph είναι ένα δέντρο. Εμείς θέλουμε να προσδιορίσουμε τις πιθανότητες των τιμών κάθε κόμβου που εξαρτώνται από τους κόμβους συγκεκριμένες τιμές σε κάποιο υποσύνολο.

Είσοδοι:

Το Μπεϋζιανό Δίκτυο (G,P) του οποίου ο Κατευθυνόμενος Άκυκλος Γράφος είναι ένα δέντρο, $G=(V,A)$, και ένα σύνολο από τιμών c ενός υποσυνόλου $C \subseteq V$.

Έξοδοι:

Το Δίκτυο Bayes (G,P) ενημερώνεται σύμφωνα με τις τιμές στο c . Οι τιμές του λ και π και των μηνυμάτων και $P(y|c)$ για κάθε $Y \in V$ θεωρούνται ότι αποτελούν μέρος του Δικτύου.

2.2 Συμπερασματολογία σε μοναδικά συνδεδεμένα δίκτυα

Ένας κατευθυνόμενος ακυκλικός γράφος ονομάζεται μοναδιαία συνδεδεμένος εάν υπάρχει κυρίως μία αλυσίδα μεταξύ δύο κόμβων. Σε οποιαδήποτε άλλη περίπτωση ονομάζεται πολλαπλά συνδεδεμένος. Ονομάζουμε το BayesianNetworksingly-connected αν έχει ένα μονό-συνδεδεμένο DirectedAcyclicGraph, ενώ σε κάθε άλλη περίπτωση ονομάζεται multiplyconnected. Η κύρια διαφορά μεταξύ ενός μονοσυνδεδεμένου Κατευθυνόμενου Κυκλικού Γραφου που δεν είναι δέντρο, και ενός δέντρου, είναι ότι στο τελευταίο είναι δυνατόν ένας κόμβος να έχει περισσότερους από έναν γονείς. Ένας μονόπλευρα συνδεδεμένος κατευθυνόμενος ακυκλικός γράφος που δεν είναι δέντρο, φαίνεται στο σχήμα 2.2.

Τώρα θα παρουσιάσουμε μια επέκταση του παραπάνω αλγορίθμου για δέντρα σε έναν αλγόριθμο για μονοσυνδεδεμένους κατευθυνόμενους ακυκλικούς γράφους.

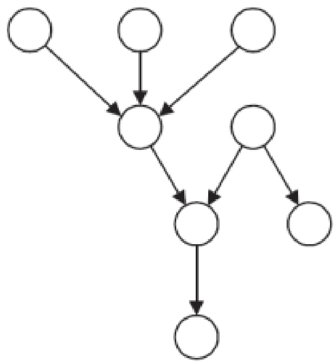
Θεώρημα 5. Έστω ότι το (G,P) είναι ένα Μπεϋζιανό Δίκτυο που είναι μονό-συνδεδεμένο, $G=(V,A)$ και c είναι ένα σύνολο τιμών ενός υποσυνόλου $C \subset V$. Θέλουμε να ορίσουμε λ μηνύματα, λ τιμές, π μηνύματα, και π τιμές για κάθε μεταβλητή Y όπως φαίνεται παρακάτω

◊ λ μηνύματα:

Για κάθε παιδί Z του Y , για όλες τις τιμές του y

$$\lambda_Z(y) \equiv \sum_z \left[\sum_{k_1,k_2,\ldots,k_j} \left(P(z \mid y, k_1, k_2, \ldots, k_j) \right) \prod_{i=1}^{j} \pi_z(k_i) \right] \lambda(z)$$

όπου $\kappa_1, \kappa_2, \ldots, \kappa_j$ είναι οι άλλοι γονείς του Z



Σχήμα 2.2: Ένα μόνο συνδεδεμένο δίκτυο που δεν είναι δέντρο (Neapolitan, R. E. (2004)[17]).

◊ τιμές λ:

Αν $Y \in C$ και η τιμή του Y είναι y,

$$\lambda(y) \equiv 1$$

$$\lambda(y) \equiv 0 \ \gamma\iota\alpha\ \ y \neq y$$

Αν $Y \notin C$ και Y είναι φύλλο τότε για κάθε τιμή του y

$$\lambda(y) \equiv 1$$

Αν $Y \notin C$ και Y δεν είναι φύλλο τότε για κάθε τιμή του y

$$\lambda(y) \equiv \prod_{U \in CH_Y} \lambda_U(y)$$

Οπού CH_Y είναι το σύνολο των παιδιών του Y.

◊ π μηνύματα:

Έστω W γονέας του Y. Τότε για όλες τις τιμές του w,

$$\pi_Y(w) \equiv \pi(w) \prod_{U \in CH_{W-\{Y\}}} \lambda_U(w)$$

◊ τιμές π:

Αν $Y \in C$ και η τιμή του Y είναι y,

$$\pi(y) \equiv 1$$
$$\pi(y) \equiv 0 \ \text{για} \ y \neq y$$

Αν $Y \notin C$ και Y είναι η ρίζα για κάθε τιμή του y τότε

$$\overline{\pi(y) \equiv P(y)}$$

Αν Y∈C, το Y δεν είναι η ρίζα, για κάθε τιμή του y,

$$\pi(y) \equiv P(y)$$

Αν $Y \notin C$ και Y είναι μη ρίζα, και $W_1, W_2, ..., W_l$ είναι οι γονείς του Y, για κάθε τιμή της

$$\pi(y) = \sum_{w_1, w_2, ..., w_l} \left(P(y \mid w_1, w_2, ..., w_l) \prod_{i=1}^{l} \pi_Y(z_i) \right)$$

◊ Δεδομένων των παραπάνω ορισμών, για κάθε μεταβλητή Y, έχουμε για κάθε τιμή του y,

$$P(y \mid c) = \alpha\lambda(y)\pi(y)$$

όπου α είναι η σταθερά κανονικοποίησης.

Αυτός ο αλγόριθμος είναι η εφαρμογή του αλγορίθμου του Pearl σε μονοσυνδεδεμένα δίκτυα.

Αλγόριθμός Inference in Trees σε μονοσυνδεδεμένα δίκτυα:

Πρόβλημα:

Υποθέτοντας ότι έχουμε ένα μόνο-συνδεδεμένο δίκτυο Bayesian, θα θέλαμε να προσδιορίσουμε τις πιθανότητες των τιμών κάθε κόμβου υπό την προϋπόθεση ότι οι κόμβοι καθορίζονται σε κάποιο υποσύνολο τιμών.

Είσοδοι:

Ένα μονό-συνδεδεμένο δίκτυο Bayesian (G, P), όπου G = (V, A), και ένα σύνολο τιμών c ενός υποσυνόλου C⊆V .

Έξοδοι:

Το δίκτυο Bayes (G, P) το οποίο ενημερώνεται σύμφωνα με τις τιμές στο C. Οι τιμές των λ και π και τα μηνύματα και P(y | c)για κάθε Y∈V θεωρούνται μέρος του Δικτύου.

2.3 Συμπερασματολογία σε πολλαπλά συνδεδεμένα δίκτυα

Προηγουμένως εξετάσαμε τον αλγόριθμο του Pearl για δίκτυα με μοναδική σύνδεση, αν και, στον πραγματικό κόσμο υπάρχουν εφαρμογές όπου αυτό δεν ισχύει. Έτσι, τώρα θα δείξουμε πώς να χειριστούμε τα πολλαπλά συνδεδεμένα δίκτυα χρησιμοποιώντας τον αλγόριθμο του Pearl για δίκτυα με μονή σύνδεση τον οποίο περιγράψαμε προηγουμένως. Αυτή η μέθοδος που θα συζητηθεί ονομάζεται conditioning. Θα εξηγήσουμε τη μέθοδο χρησιμοποιώντας ένα παράδειγμα. Ας υποθέσουμε ότι έχουμε ένα Μπεϋζιανό Δίκτυο που περιέχει μια κατανομή P και έχει κατευθυνόμενο ακυκλικό γράφημα όπως αυτό που φαίνεται στο Σχήμα 2.3, και οι τυχαίες μεταβλητές έχουν δύο πιθανές τιμές. Ο αλγόριθμος της Συμπερασματολογίας σε Δέντρα που ορίσαμε παραπάνω δεν μπορεί να εφαρμοστεί άμεσα αφού το Δίκτυο που έχουμε είναι πολλαπλά συνδεδεμένο. Παρ' όλα αυτά, αν το Y αφαιρεθεί από το Δίκτυο θα γίνει Singly-connected. Έτσι, κατασκευάζονται δύο BayesianNetworks, το ένα από τα οποία περιέχει την υπό συνθήκη κατανομή P' της P δεδομένου ότι $Y = y_1$ και το άλλο που περιέχει την υπό συνθήκη κατανομή $\boldsymbol{P''}$ της P δεδομένου ότι $Y = y_2$.

Στο Σχήμα 2.3 (β),(γ) μπορούμε να δούμε αυτά τα δύο δίκτυα αντίστοιχα. Αρχικά πρέπει να προσδιορίσουμε τις υπό συνθήκη πιθανότητες κάθε κόμβου δεδομένων των γονέων του για κάθε Δίκτυο. Εάν αυτό ισχύει, οι υπό συνθήκη πιθανότητες είναι οι ίδιες με αυτές που ορίζονται στο αρχικό Δίκτυο, εκτός από τις ρίζες Z και W. Επομένως, για αυτές τις πιθανότητες πρέπει να ψάξουμε στους τύπους

$$P'(z_1) = P(z_1 \mid z_1)P'(w_1) = P(w_1 \mid y_1)$$
$$P''(z_1) = P(z_1 \mid w_2)P'(w_1) = P(w_1 \mid y_2)$$

Τώρα είμαστε έτοιμοι να χρησιμοποιήσουμε στο αρχικό μας Δίκτυο τον Αλγόριθμο Inferencemin Trees για να συμπεράνουμε σε αυτά τα Singly-connected Networks.

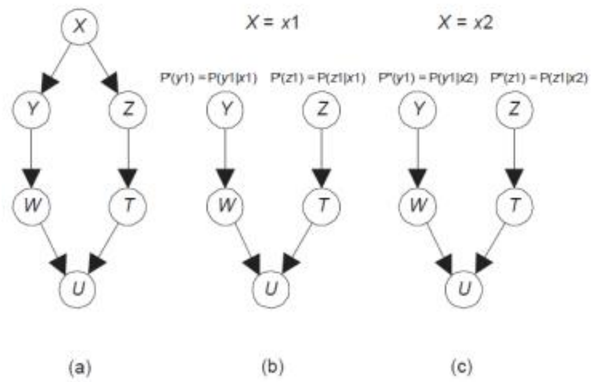
Παρακάτω παρουσιάζουμε ένα παράδειγμα για να δείξουμε το μέθοδο.

Παράδειγμα:

Έστω ότι στο Δίκτυο του Σχήματος 2.3(α) το U είναι ενσαρκωμένο για το u_1. Για την ανάγκη απεικόνισης, υποθέτουμε την υπό συνθήκη πιθανότητα του W δεδομένης αυτής της ενστάλαξης. Έτσι,

$$P(w_1 \mid u_1) = P(w_1 \mid x_1, u_1) P(x_1 \mid u_1) + P(w_1 \mid x_2, u_1) P(x_2 \mid u_1)$$

Οι τιμές $P(w_1 \mid x_1, u_1)$ και $P(w_1 \mid x_2, u_1)$ μπορούν να προκύψουν αν εφαρμόσουμε τον Αλγόριθμο Inference in Trees στα Δίκτυα που φαίνονται στο Σχήμα 2.3(β),(γ) αντίστοιχα. Η τιμή $P(x_i \mid u_1)$ δίνεται από τη σχέση $P(x_i \mid u_1) = \alpha P(u_1 \mid x_i) P(x_i)$



Σχήμα 2.3: Ένα πολυδιασυνδεδεμένο δίκτυο απεικονίζεται στο (α). Τα μονό-συνδεδεμένα δίκτυα που προκύπτουν από την ενσάρκωση του X για το x_1 και για το x_2 απεικονίζονται στα (β),(γ) αντίστοιχα. (Neapolitan, R. E. (2004)).

όπου α η σταθερά κανονικοποίησης που ισούται με $\dfrac{1}{P(u_1)}$. Αποθηκεύουμε την τιμή της $P(x_i)$ στο Δίκτυο επειδή το X είναι ρίζα και η τιμή του $P(u_1 \mid x_i)$ προκύπτει αν εφαρμόσουμε τον Αλγόριθμο Inference in Trees στα Δίκτυα που απεικονίζονται στο Σχήμα 2.3(β) και στο Σχήμα 2.3(γ). Επομένως, μπορεί να ληφθεί η τιμή του

$P(w_1 | u_1)$.Ομοίως, οι υπό συνθήκη πιθανότητες όλων των μη εξαρτημένων μεταβλητών των Δικτύων μπορούν να ληφθούν. Πρέπει να αναφέρουμε ότι με τον τρόπο αυτό έχει ήδη υπολογιστεί η υπό συνθήκη πιθανότητα, δηλαδή η $P(x_i | u_1)$της εξαρτημένης μεταβλητής.

2.4 Το μοντέλο της θορυβώδους πύλης OR-Gate

Όπως γνωρίζουμε, ένα Μπεϋζιανό Δίκτυο απαιτεί τις δεσμευμένες πιθανότητες κάθε μεταβλητής δεδομένου όλων των συνδυασμών των τιμών των γονέων μεταβλητών. Έτσι, υποθέτοντας ότι κάθε μεταβλητή έχει δύο πιθανές τιμές και κάθε μεταβλητή έχει p γονείς, 2^p δεσμευμένες πιθανότητες πρέπει να καθοριστούν για την εν λόγω μεταβλητή. Εάν έχουμε μεγάλο p, η εξαγωγή συμπερασμάτων του αλγορίθμου μας καθίσταται υπολογιστικά ανέφικτη και οι απαιτήσεις αποθήκευσης καθίστανται και οι ίδιες εξίσου ανέφικτες. Επιπλέον, εάν το p δεν είναι τόσο μεγάλο, η δεσμευμένη πιθανότητα μιας μεταβλητής δεν είναι συνήθως πολύ προσιτή δεδομένου ενός συνδυασμού τιμών των γονέων της.

Τώρα ορίζουμε ένα Μοντέλο που απαιτεί μόνο τον προσδιορισμό των τελευταίων πιθανοτήτων. Αυτές η πιθανότητες δεν είναι μόνο προσβάσιμες, αλλά υπάρχει επίσης μόνο ένας γραμμικός αριθμός τους.

Το μοντέλο:

Το μοντέλο ονομάζεταο θορυβώδης πύλη OR-Gate και η περίπτωση όπου οι σχέσεις μεταξύ των μεταβλητών συνήθως αντιπροσωπεύουν αιτιώδεις μηχανισμούς. Κάθε μεταβλητή μπορεί να λάβει μόνο δύο πιθανές τιμές και εάν η συνθήκη είναι παρούσα η μεταβλητή παίρνει την πρώτη της τιμή και τη δεύτερη τιμή της σε κάθε άλλη περίπτωση.

Για λόγους συμβολικής απλότητας οι τιμές εμφανίζονται ως 1 και 2.

Οι επόμενες τρεις υποθέσεις γίνονται σε αυτό το μοντέλο:

◊ **Αιτιώδης αναστολή:** Με την υπόθεση αυτή συνεπάγεται ότι υπάρχει ένας μηχανισμός που εμποδίζει ένα αίτιο να επιφέρει το αποτέλεσμά του, και η παρουσία αυτού του αιτίου έχει ως αποτέλεσμα στην παρουσία του αποτελέσματος, εάν και μόνο εάν ο μηχανισμός απενεργοποιηθεί.

◊ **Ανεξαρτησία εξαιρέσεων:** Με την υπόθεση αυτή συνεπάγεται ότι ο μηχανισμός που αναστέλλει ένα αίτιο είναι ανεξάρτητος από τον μηχανισμό που αναστέλλει ένα άλλο αίτιο.

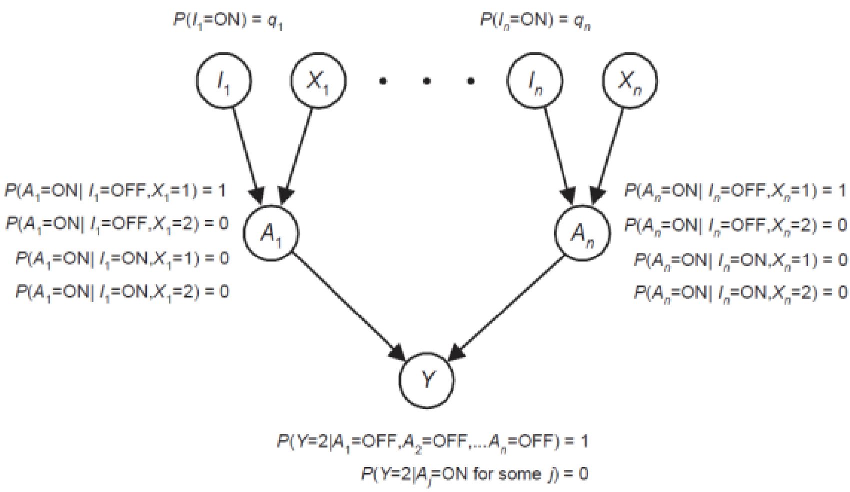
◊ **Λογοδοσία:** Με την υπόθεση αυτή συνεπάγεται ότι ένα αποτέλεσμα μπορεί να συμβεί μόνο εάν, τουλάχιστον μία από τις αιτίες του είναι παρούσα και δεν περιορίζεται. Για το λόγο αυτό όλες οι αιτίες που δεν δηλώνονται συγκεκριμένα πρέπει να ενσωματωθούν σε μία άγνωστη αιτία.

Δεδομένου των υποθέσεων των μοντέλων, μπορούμε να αναπαραστήσουμε τις σχέσεις μεταξύ των μεταβλητών με το Μπεϋζιανό Δίκτυο που παρουσιάζεται παρακάτω στο σχήμα 2.4. Σε αυτό το Σχήμα παρουσιάζουμε την κατάσταση όπου έχουμε n αιτίες $X_1, X_2, ..., X_n$ του Y και η μεταβλητή I_j δείχνει την μηχανισμό που αναστέλλει την x_j . Λόγω της παραδοχής μας για ανεξαρτησία εξαιρέσεων οι $I_j's$ είναι ανεξάρτητες. Επιπλέον, εάν και μόνο εάν X_j είναι παρούσα, που σημαίνει ότι είναι ίση με 1 και είναι δεν παρεμποδίζεται, η μεταβλητή A_j είναι ενεργοποιημένη. Αυτό σημαίνει ότι η Y πρέπει να είναι παρούσα, πράγμα που σημαίνει ίση με 1, εάν οποιαδήποτε από τις $A_j's$ είναι παρούσα, λόγω της υπόθεσής μας για αιτιώδη αναστολή. Για το λόγο αυτό, έχουμε:

$$P\left(Y = 2 \mid A_j = ON \text{ για κάποια } j\right) = 0$$

και αυτός είναι ο λόγος για τον οποίο ονομάζεται μοντέλο "OR-Gate". Έτσι, μπορεί κανείς να σκεφτεί ότι το $A_j's$ εισέρχεται σε μια πύλη OR-Gate έτσι ώστε η έξοδός της να τροφοδοτεί το Z και να ονομάζεται "θορυβώδης" λόγω της $I_j's$. Τελικά, από την υπόθεση της λογοδοσίας συνεπάγεται ότι έχουμε:

$$P\left(Y = 2 \mid A_1 = OFF, A_2 = OFF, ..., A_n = OFF\right) = 1$$



Σχήμα 2.4: Ένα δίκτυο Bayes που αναπαριστά τις υποθέσεις στη θορυβώδη πύλη OR (Neapolitan, R. E. (2004),[17]).

Έτσι έχουμε το επόμενο Θεώρημα.

Θεώρημα 6. Υποθέτοντας ότι έχουμε ένα Μπεϋζιανό Δίκτυο που αναπαριστά το Noisy-Or-Gate Μοντέλο (δηλ. Σχήμα1.12). Ας υποθέσουμε ότι

$$W = \{X_1, X_2, ..., X_n\}$$

Και

$$\{x_1, x_2, ..., x_n\}$$

είναι ένα σύνολο τιμών των μεταβλητών που έχουμε στο W. Επιπλέον, ας υποθέσουμε ότι το S είναι ένα σύνολο δεικτών έτσι ώστε $j \in S$ αν και μόνο αν $x_j = 1$. Αυτό σημαίνει ότι $S = j$ έτσι ώστε $x_j = 1$. Στη συνέχεια έχουμε

$$P(Y = 2 \,|\, W = w) = \prod_{j \in S} q_j$$

Το πραγματικό δίκτυο Bayes που έχουμε περιέχει το Y και το $X_j's$ αλλά δεν περιέχονται τα $I_j's$ ή $A_j's$. Επομένως, πρέπει να προσδιορίσουμε την υπό συνθήκη πιθανότητα του Y δεδομένων των συνδυασμών των τιμών των $X_j's$. Χρειάζεται μόνο να προσδιορίσουμε τις τιμές του q_j για όλα τα $j's$ λόγω του προηγούμενου θεωρήματος. Στη συνέχεια, χρησιμοποιώντας το Θεώρημα μπορούμε να υπολογίσουμε όλες τις απαραίτητες υπό συνθήκη πιθανότητες. Αντί αυτών συχνά προσδιορίζεται:

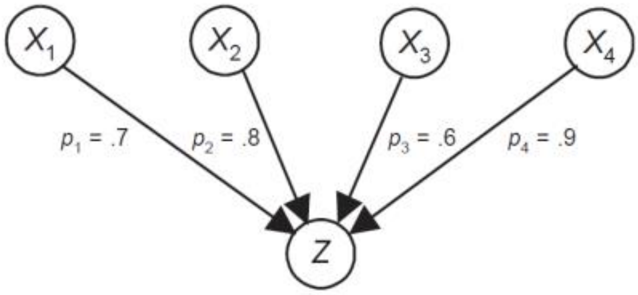
$$\boldsymbol{p_j = 1 - q_j}$$

την οποία ονομάζουμε αιτιώδη δύναμη του X για το Y . Το θεώρημα που ορίζεται παραπάνω συνεπάγεται με:

$$p_j = P\left(Y = 1 \mid X_j = 1, X_i = 2\, \gamma\iota\alpha\, i \neq j\right)$$

η τιμή της οποίας είναι κάπως προσιτή.

Για παράδειγμα, θεωρήστε ότι έχουμε μια μεγάλη βάση δεδομένων των ασθενών που έχουν μόνο καρκίνο του πνεύμονα ως ασθένεια. Θα θέλαμε να εκτιμήσουμε την αιτιώδη δύναμη του καρκίνου του πνεύμονα για την κόπωση. Για να το κάνουμε αυτό, πρέπει μόνο να προσδιορίσουμε πόσοι από τους από αυτούς τους ασθενείς είναι κουρασμένοι. Εάν θέλουμε να εκτιμήσουμε άμεσα την υπό συνθήκη πιθανότητα της κόπωσης δεδομένης της βρογχίτιδας, του καρκίνου του πνεύμονα και ορισμένων άλλων αιτιών, οι βάσεις δεδομένων περιέχουν ασθενείς με όλους τους συνδυασμούς αυτών των ασθενειών.



Σχήμα 2.5: Δίκτυο Bayes με χρήση του μοντέλου NoisyOR-Gate (Neapolitan, R. E((2004),[17])

Εξαγωγή συμπερασμάτων με το μοντέλο:

Παρόλο που το παραπάνω θεώρημα μπορεί να λύσει το πρόβλημα των προδιαγραφών μας, υπάρχει ακόμα η ανάγκη για τον υπολογισμό ενός ενδεχομένως εκθετικού αριθμού τιμών για να συμπεράνουμε χρησιμοποιώντας τον Αλγόριθμο: InferenceinTrees. Τώρα τροποποιούμε τον αλγόριθμο για να συμπεραίνουμε πιο αποτελεσματικά χρησιμοποιώντας το θορυβώδες μοντέλο OR-Gate με τις πιθανότητες που προσδιορίζονται. Ας υποθέσουμε ότι το θορυβώδες OR-Gate μοντέλο ικανοποιείται από μεταβλητές και το Y έχει m γονείς $X_1, X_2, ..., X_m$.Έστω p_i η αιτιώδης ισχύς της X_i για το Y , και $q_i = 1 - p_i$. Παρουσιάζεται η κατάσταση με m = 4 στο Σχήμα 2.5. Πριν προχωρήσουμε όμως, πρέπει να διευκρινίσουμε τον συμβολισμό. Για να δηλωθεί ότι X_i είναι παρούσα, θα χρησιμοποιούμε x_i^+ αντί για 1 και για να δηλώσουμε ότι το X_i δεν υπάρχει, θα χρησιμοποιήσουμε x_i^- αντί για 2. Εξετάζουμε πρώτα τα μηνύματα λ. Με τον παρόντα συμβολισμό μας, πρέπει να κάνουμε τον ακόλουθο υπολογισμό στον Αλγόριθμο: InferenceinTrees για να μπορέσουμε να υπολογίσουμε τα λ μήνυμα που στέλνει ο Y στον X_i :

$$\lambda_Y(x_i) = \sum_y \left[\sum_{x_i, ..., x_{i-1}, x_{i+1}, ..., x_m} \left(P(y \mid x_1, x_2, ..., x_m) \prod_{m \neq i} \pi_Y(x_m) \right) \right] \lambda(y)$$

Για να κάνουμε αυτόν τον υπολογισμό πρέπει να προσδιορίσουμε έναν εκθετικό αριθμό υπό συνθήκη πιθανοτήτων. Λαμβάνοντας τώρα υπόψη τις τιμές π και με τη χρήση του σημερινού μας συμβολισμού πρέπει να κάνουμε τον ακόλουθο υπολογισμό στον Αλγόριθμο Inference in Trees για να μπορέσουμε να υπολογίσουμε την τιμή π του Y :

$$\pi(y) = \sum_{x_1, x_2, ..., x_m} \left(P(y \mid x_1, x_2, ..., x_m) \prod_{i=1}^{m} \pi_Y(x_i) \right)$$

Κλείνοντας το κεφάλαιο αυτό είδαμε μερικούς αλγορίθμους και περιπτώσεις συμπερασματολογίας τις οποίες εφαρμόζουμε στα Μπεϋζιανά Δίκτυα ώστε να μπορέσουμε να εξάγουμε συμπεράσματα και να υπολογίσουμε τις πιθανότητες που χρειαζόμαστε.

Στη συνέχεια θα δούμε πως μπορούμε να εκπαιδεύσουμε τα Μπεϋζιανά Δίκτυα.

ΚΕΦΑΛΑΙΟ 3

Μάθηση Μπεϋζιανών Δικτύων

Η Μάθηση Μπεϋζιανών Δικτύων χρησιμοποιείται για τη διάκριση μεταξύ συμπεράσματος και μάθησης στη βιβλιογραφία των γραφικών μοντέλων. Με τον όρο «Συμπερασματολογία» εννοούμε τον υπολογισμό συναρτήσεων όπως $p(x_u \mid x_u, \theta)$, όπου u είναι οι κρυφοί κόμβοι, v είναι οι ορατοί κόμβοι και θ είναι οι παράμετροι του μοντέλου που υποθέτουμε ότι γνωρίζουμε. Συνήθως με τον όρο «Μάθηση» εννοούμε τον υπολογισμό μιας εκτίμησης MAP των παραμέτρων των δεδομένων

$$\hat{\theta} = \underbrace{argmax}_{\theta} \sum_{m=1}^{M} logp\left(x_{m,u} \mid \theta\right) + logp\left(\theta\right)$$

όπου $x_{i,u}$ είναι οι ορατές μεταβλητές στην περίπτωση i. Εάν η πρότερη εκτίμησή μας είναι ομοιόμορφη με $p(\theta) \propto 1$, τότε αυτό συνήθως ανάγεται στην MLE.

Η κύρια διαφορά μεταξύ των κρυφών μεταβλητών και των παραμέτρων είναι ότι όσο αυξάνεται ο όγκος των δεδομένων εκπαίδευσης, αφού συνήθως υπάρχει ένα σύνολο από κρυφές μεταβλητές για κάθε περίπτωση παρατηρούμενων δεδομένων, αυξάνεται και ο αριθμός των κρυφών μεταβλητών, αν και ο αριθμός των παραμέτρων είναι συνήθως σταθερός (τουλάχιστον στα παραμετρικά μοντέλα). Αυτό σημαίνει ότι οι κρυφές μεταβλητές πρέπει να ολοκληρωθούν, ώστε να αποφύγουμε την υπερπροσαρμογή, αν και θα μπορούσαμε να ξεφύγουμε με τεχνικές σημειακής εκτίμησης για τις παραμέτρους, οι οποίες είναι λιγότερες σε αριθμό.

Μάθηση Δικτύων Bayesian:

Στα Μπεϋζιανά Δίκτυα, η Μάθηση είναι γνωστή ως επιλογή και εκτίμηση μοντέλου, μια ονομασία που δανείζεται από την Τεχνητή Νοημοσύνη (ΤΝ) και τη Μηχανική Μάθηση (ΜΜ). Η Bayesian Μάθηση δικτύων πραγματοποιείται ως επί το πλείστων σαν διαδικασία δύο βημάτων:

Βήμα 1: Δομή της μάθησης, η οποία ορίζεται ως η εκμάθηση της δομής του κατευθυνόμενου ακυκλικού γραφήματος του Δικτύου.

Βήμα 2: Εκμάθηση παραμέτρων, η οποία ορίζεται ως η εκμάθηση των τοπικών κατανομών που υπονοούνται από το δομή του κατευθυνόμενου ακυκλικού γραφήματος του δικτύου που μαθαίνεται στο βήμα 1.

Μπορούμε να εκτελέσουμε και τα δύο βήματα, είτε ως μη επιβλεπόμενη μάθηση, εάν χρησιμοποιούμε τις πληροφορίες που παρέχονται από το σύνολο των δεδομένων, είτε ως μάθηση με επίβλεψη, ζητώντας από ειδικούς που είναι σχετικοί με το φαινόμενο που θέλουμε να μοντελοποιήσουμε. Είναι πολύ σύνηθες να συνδυάζουμε και τις δύο προσεγγίσεις. Ορισμένες φορές οι προηγούμενες πληροφορίες που είναι διαθέσιμες με το φαινόμενο που μοντελοποιείται, δεν είναι αρκετές για να προσδιορίσουν πλήρως ένα Μπεϋζιανό Δίκτυο. Ακόμη και ο προσδιορισμός της δομής του κατευθυνόμενου ακυκλικού γράφου είναι συχνά αδύνατος, ιδίως όταν εμπλέκεται μεγάλος αριθμός μεταβλητών.

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων D και ένα Μπεϋζιανό Δίκτυο $B=\left(G,V\right)$. Συμβολίζοντας τις παραμέτρους της παγκόσμιας κατανομής του V με Θ, χωρίς απώλεια γενικότητας, υποθέτουμε ότι το Θ προσδιορίζει μοναδικά V στην παραμετρική οικογένεια των κατανομών που επιλέγονται για τη μοντελοποίηση του D και γράφουμε $B=\left(G,\Theta\right)$. Στη συνέχεια, η μάθηση του BayesianNetwork μπορεί να τυποποιηθεί ως εξής (Scutari, M., & Denis, J. B. (2021)[8]).

$$\underbrace{Pr(B\,|\,D)=Pr(G,\Theta\,|\,D)}_{\text{μάθηση}}=\underbrace{Pr(G,D)}_{\text{δομή μάθησης}}*\underbrace{Pr(\Theta\,|\,G,D)}_{\text{παραμετρική μάθηση}} \qquad (3.1)$$

Τα δύο βήματα που περιγράψαμε παραπάνω αντικατοπτρίζονται από τη διάσπαση της $Pr\left(G,\Theta\,|\,D\right)$, η οποία διέπει τη λογική της διαδικασίας Μάθησης. Πρακτικά, μπορούμε να κάνουμε Δομή Μάθησης βρίσκοντας το κατευθυνόμενο ακυκλικό γράφημα που μεγιστοποιεί την ακόλουθη εξίσωση (Scutari, M., & Denis, J. B. (2021)[8]).

$$Pr\left(G\,|\,D\right)\propto Pr\left(G\right)Pr\left(D\,|\,G\right)=Pr\left(G\right)\int Pr\left(D\,|\,G,\Theta\right)Pr\left(\Theta\,|\,G\right)d\Theta \qquad (3.2)$$

με το θεώρημα Bayes να χρησιμοποιείται για την ανάλυση της εκ των υστέρων πιθανότητας του κατευθυνόμενου ακυκλικού γραφήματος Pr(G | D), στο γινόμενο της εκ των προτέρων κατανομής επί των πιθανών κατευθυνόμενων ακυκλικών γράφων Pr(G) και της πιθανότητας των δεδομένων Pr(D | G). Είναι σαφές ότι είναι αδύνατο να υπολογιστεί η τελευταία χωρίς να εκτιμηθούν επίσης οι παράμετροι Θ του G. Για το λόγο αυτό, λοιπόν, το Θ πρέπει να ενσωματωθεί από την εξίσωση (3.1), ώστε να γίνει Pr(G | D) ανεξάρτητη από οποιαδήποτε συγκεκριμένη επιλογή του Θ. Ένας ιδανικός τρόπος για την εισαγωγή οποιασδήποτε προηγούμενης πληροφορίας, που είναι διαθέσιμη για τις σχέσεις ανεξαρτησίας υπό όρους μεταξύ των μεταβλητών στο X, παρέχεται από την εκ των προτέρων κατανομή Pr(G). Για παράδειγμα, θα μπορούσαμε να απαιτήσουμε ότι μία ή περισσότερες ακμές θα πρέπει να υπάρχουν στο κατευθυνόμενο ακυκλικό γράφημα ή να απουσιάζουν από το Κατευθυνόμενο Άκυκλο Γράφο, για τη λογική των γνώσεων που αποκτήθηκαν σε προηγούμενες μελέτες. Θα μπορούσαμε επίσης, να απαιτήσουμε ότι εάν υπάρχουν ορισμένες ακμές στο κατευθυνόμενο ακυκλικό γράφημα, θα πρέπει να είναι προσανατολισμένες προς μία συγκεκριμένη κατεύθυνση, εφόσον αυτή είναι η μόνη που έχει νόημα υπό το πρίσμα της λογικής που διέπει το φαινόμενο που μοντελοποιείται. . Ένα μη πληροφοριακό prior πάνω στο χώρο των πιθανών κατευθυνόμενων ακυκλικών γράφων επιλέγεται συνήθως για το Pr(G) αποδίδοντας την ίδια πιθανότητα σε κάθε κατευθυνόμενο ακυκλικό γράφο.

Αν ξεκινήσουμε από τη διάσπαση σε τοπικές κατανομές, τότε η Pr(D | G) μπορεί να παραγοντοποιηθεί με παρόμοιο τρόπο:

$$
\begin{aligned}
Pr\left(D \mid G\right) &= \int \prod_{i=1}^{p} \left[Pr\left(X_i \mid \Pi_{X_i}, \Theta_{X_i}\right) Pr\left(\Theta_{X_i} \mid \Pi_{X_i}\right) \right] d\Theta_{X_i} \\
&= \prod_{i=1}^{p} \left[\int Pr\left(X_i \mid \Pi_{X_i}, \Theta_{X_i}\right) Pr\left(\Theta_{X_i} \mid \Pi_{X_i}\right) \right] d\Theta_{X_i} \\
&= \prod_{i=1}^{p} \mathrm{E}_{\Theta_{X_i}} \left[Pr\left(X_i \mid \Pi_{X_i}\right) \right]
\end{aligned}
$$

Έτσι, αν αυτή η αποσύνθεση μπορεί να γίνει, ο υπολογισμός είναι ταχύτερος.

Οι συναρτήσεις των οποίων, η παραγοντοποίηση είναι δυνατόν να γίνει με αυτόν τον τρόπο ονομάζονται αποσυνθέσιμες. Σε περίπτωση που μπορούμε να υπολογίσουμε όλες τις προσδοκίες σε στενή μορφή τότε υπάρχει η δυνατότητα να υπολογιστεί και η Pr(D|G) σε εύλογο χρόνο και για μεγάλα σύνολα δεδομένων. Αυτό μπορεί ενδεχομένως να γίνει τόσο για την πολυωνυμική κατανομή, που υποθέτουμε για τα Διακριτά Μπεϋζιανά Δίκτυα (από τη συζυγή της Dirichlet prior distribution), όσο και για την πολυμεταβλητή κατανομή Gaussian που υποθέτουμε για τα συνεχή Bayesian Networks (από τη συζυγή της προηγούμενη κατανομή InverseWishart). Στα διακριτά Μπεϋζιανά Δίκτυα, μπορούμε να εκτιμήσουμε την Pr(D|G) με το ισοδύναμο ομοιόμορφο της Bayesian Dirichlet equivalent (BDe) σκορ από τους Heckermanetal. (1995). Καθώς αποτελεί το μόνο μέλος της οικογένειας βαθμολογιών του ισοδύναμου Bayesian Dirichlet που χρησιμοποιείται συνήθως και συχνά αναφέρεται απλώς ως Bayesian Dirichlet equivalent. Όπως είναι γνωστό, οι εκ των προτέρων πιθανότητες τόσο στο χώρο των κατευθυνόμενων ακυκλικών γραφημάτων όσο και στο χώρο των παραμέτρων κάθε κόμβου, εντασσονται στο Bayesian Dirichlet equivalent:

$$Pr\left(G\right) \propto 1 \kappa\alpha\iota \; Pr\left(\Theta_{X_i} \mid \Pi_{X_i}\right) = \alpha_{ij} = \frac{a}{\left|\Theta_{X_i}\right|} \tag{3.3}$$

Η μόνη παράμετρος του φανταστικού μεγέθους του δείγματος α που συνδέεται με την Dirichlet εκ των προτέρων κατανομή, είναι η μόνη παράμετρος της Bayesian Dirichlet equivalent που ορίζει την ποσότητα του βάρους που αποδίδεται στην εκ των προτέρων ως προς το μέγεθος ενός φανταστικού δείγματος που το υποστηρίζει. Υπό τις παραδοχές που δόθηκαν ανωτέρω, το Bayesian Dirichlet equivalent λαμβάνει την ακόλουθη μορφή (Scutari, M., & Denis, J. B. (2021)[8]) :

$$BDe(G,D) = \prod_{i=1}^{p} BDe\left(X_i, \Pi_{X_i}\right) = \prod_{i=1}^{p}\prod_{j=1}^{q}\left\{\frac{\Gamma\left(\alpha_{ij}\right)}{\Gamma\left(\alpha_{ij} + \left(\alpha_{ij} + m_{ij}\right)\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(\alpha_{ijk} + m_{ij}\right)}{\Gamma\left(\alpha_{ijk}\right)}\right\} \tag{3.4}$$

Όπου:

- p μας δίνει τον αριθμό των κόμβων στο G,
- r_i αντιπροσωπεύει τις κατηγορίες για τον κόμβο X_i
- q_i είναι οι διαμορφώσεις του αριθμού των κατηγοριών των γονέων του X_i
- n_{ijk} ορίζει τον αριθμό των δειγμάτων που έχουν την j-οστή κατηγορία για τον κόμβο Xi και το k-οστή διαμόρφωση για τους γονείς του.

Οι Geiger και Heckerman (1994) ονόμασαν την Bayesian Gaussian equivalent uniform (BGeu) την αντίστοιχη εκ των υστέρων πιθανότητα για τα GaussianBayesianNetworks, η οποία επίσης αναφέρεται ομοίως ως Bayesian Gaussian equivalent (BGe). Με τον ίδιο τρόπο όπως το Bayesian Dirichlet equivalent (BDe), θεωρεί μια χωρις πληροφορία εκ των προτέρων επί του χώρου των κατευθυνόμενων ακυκλικών γράφων και του χώρου παραμέτρων κάθε κόμβου και η μόνη παράμετρος που εμπλέκεται είναι το φανταστικό μέγεθος του δείγματος α. Ο υπολογισμός των ανωτέρω είναι εξαιρετικά πολύπλοκος και δε θα αναλυθεί στη συγκεκριμένη εργασία. Εξαιτίας της περιπλοκότητας του όμως, έχουν προταθεί δύο εναλλακτικοί υπολογισμοί για τη χρήση του Pr(D | G) στη μάθηση δομής. Η πρώτη λύση, είναι η χρήση του Bayesian Information Criterion (BIC) για την προσέγγιση του Pr(D | G), όπως

$$BIC\left(G,D\right) \to \log BDe\left(G,D\right) \, \mu\varepsilon \; \mu\acute{\varepsilon}\gamma\varepsilon\theta o\varsigma \; \delta\varepsilon\acute{\iota}\gamma\mu\alpha\tau o\varsigma \; n \to \infty \tag{3.5}$$

Το BayesianInformationCriterion (BIC) μπορεί να αναλυθεί και εξαρτάται μόνο από τη συνάρτηση πιθανότητας,

$$BIG\left(G,D\right) = \sum_{i=1}^{p} \left[\log P\,\mathrm{R}\,r\left(X_i \mid \Pi_{X_i}\right) - \frac{|\Theta_{X_i}|}{2} \log n \right] \tag{3.6}$$

γεγονός που καθιστά τον υπολογισμό του πολύ εύκολο. Η άλλη εναλλακτική λύση είναι η αποφυγή της ανάγκης του ορισμού μιας μέτρησης της καλής προσαρμογής για τον κατευθυνόμενο ακυκλικό γράφο και η χρήση δοκιμών ανεξαρτησίας υπό όρους, έτσι ώστε να μαθαίνουμε την δομή από τους κατευθυνόμενους ακυκλικούς γράφους για μία ακμή κάθε φορά.

3.1 Δομή μάθησης

Στη βιβλιογραφία έχουν παρουσιαστεί μερικοί αλγόριθμοι για την επίλυση αυτού του προβλήματος, λόγω της εφαρμογής αποτελεσμάτων που προκύπτουν από τη θεωρία της πληροφορίας, των πιθανοτήτων και της βελτιστοποίησης.

Παρά την ενίοτε συγκεχυμένη ποικιλία θεωρητικών υποβάθρων και ορολογίας, όλες αυτές οι προσεγγίσεις ανάγονται σε τρεις μόνο προσεγγίσεις, οι οποίες είναι οι υβριδικές, οι βασισμένες σε περιορισμούς και οι βασισμένες σε βαθμολογία.

Οι αλγόριθμοι εκμάθησης δομής λειτουργούν με βάση ένα σύνολο κοινών παραδοχών:

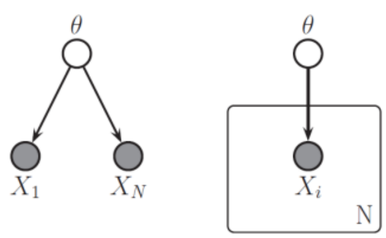
◊ Οι κόμβοι του κατευθυνόμενου ακυκλικού γράφου και οι τυχαίες μεταβλητές στο X πρέπει να έχουν μια αντιστοιχία ένα προς ένα, πράγμα που σημαίνει ότι δεν πρέπει να έχουμε πολλαπλούς κόμβους που είναι ντετερμινιστικές συναρτήσεις μιας μόνο μεταβλητής.

◊ Όλες οι σχέσεις μεταξύ των μεταβλητών στο X πρέπει να είναι υπό συνθήκη ανεξάρτητες, δεδομένου ότι ορίζονται ως το μόνο είδος σχέσεων που μπορούν να εκφραστούν από ένα Μπεϋζιανό Δίκτυο.

◊ Ένα έγκυρο, παρατηρήσιμο γεγονός πρέπει να αντιπροσωπεύεται από κάθε συνδυασμό των πιθανών τιμών στο X. Αυτό που συνεπάγεται αυτή η υπόθεση, είναι μια αυστηρά θετική συνολική κατανομή, η οποία πρέπει να έχει μοναδικά καθορισμένες MarkovBlanket και για το λόγο αυτό ένα μοναδικά προσδιορίσιμο μοντέλο. Ακόμη και όταν αυτό δεν ισχύει, οι αλγόριθμοι που βασίζονται σε περιορισμούς λειτουργούν, αφού για τη μοναδικότητα των MarkovBlanket, η ύπαρξη ενός τέλειου χάρτη είναι επίσης επαρκής συνθήκη(Pearl, 1988).

◊ Αντιμετωπίζουμε τις παρατηρήσεις ως ανεξάρτητες πραγματώσεις του συνόλου των κόμβων. Εάν υπάρχει οποιασδήποτε μορφής χρονική ή χωρική εξάρτηση, όπως στα δυναμικά Μπεϋζιανά Δίκτυα, πρέπει να ληφθεί ρητά υπόψη στον ορισμό των δικτύων.

Σημειώσεις πινακίδων:



Φ

Σχήμα 3.1: (Murphy, K. P. (2012),[15])Στα αριστερά παρουσιάζονται τα σημεία δεδομένων x_i τα οποία με δεδομένο το θ είναι υπό όρους ανεξάρτητα και στα δεξιά έχουμε τον συμβολισμό Plate. Ο συμβολισμός αυτός αντιπροσωπεύει το ίδιο μοντέλο με το μοντέλο στα αριστερά, με τη διαφορά ότι οι κόμβοι που είναι επαναλαμβανόμενοι x_i είναι μέσα στο πλαίσιο και είναι γνωστοί ως πλάκα.

Σημειώνεται ότι ο αριθμός N που βρίσκεται στην κάτω δεξιά γωνία δηλώνει τον αριθμό των επαναλήψεων των κόμβων X_i.

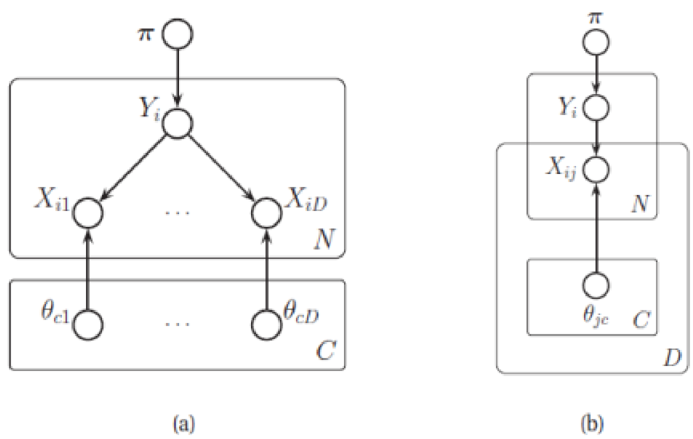
Στην περίπτωση που συμπεραίνουμε παραμέτρους από δεδομένα, συχνά υποθέτουμε ότι τα δεδομένα είναι ανεξάρτητα. και ταυτόσημα κατανεμημένα (iid), κάτι το οποίο μπορεί να αναπαρασταθεί με τη χρήση ενός γραφικού μοντέλου, όπως φαίνεται στο σχήμα 3.1(α). Η υπόθεση ότι κάθε περίπτωση δεδομένων έχει δημιουργηθεί ανεξάρτητα αλλά και από την ίδια κατανομή, απεικονίζεται με αυτό. Οι περιπτώσεις δεδομένων είναι ανεξάρτητες μόνο εάν εξαρτώνται από τις παραμέτρους θ και εδώ βλέπουμε, ότι οριακά εξαρτώνται. Ωστόσο, μπορεί να φανεί σε αυτό το παράδειγμα ότι η σειρά του τρόπου με τον οποίο τα δεδομένα περιπτώσεων που φθάνουν δε διαφέρει από τις πεποιθήσεις μας για το θ, καθώς όλες οι σειρές πρόκειται να έχουν την ίδια επαρκή στατιστική. Έτσι, λέγεται ότι τα δεδομένα είναι ανταλλάξιμα. Για την αποφυγή της οπτικής ακαταστασίας, χρησιμοποιείται συνήθως μια μορφή περιγράμματός που ονομάζεται «πλάκες», όπου απλώς σχεδιάζεται ένα μικρό πλαίσιο γύρω από τις επαναλαμβανόμενες μεταβλητές, με τη συμφωνία ότι οι κόμβοι που βρίσκονται μέσα στο πλαίσιο πρόκειται να επαναληφθούν όταν το μοντέλο ξετυλίγεται. Ο αριθμός των αντιγράφων ή των επαναλήψεων γράφεται συχνά στα πλαίσια στην κάτω δεξιά γωνία. Ένα απλό παράδειγμα παρουσιάζεται στο Σχήμα 3.1(β). Έτσι, η αντίστοιχη κοινή κατανομή έχει τη μορφή

$$p(\theta, D) = p(\theta)\left[\prod_{i=1}^{N} p(x_i \mid \theta)\right] \tag{3.7}$$

Στο σχήμα 3.2 παρουσιάζεται ένα λίγο πιο σύνθετο παράδειγμα, στο οποίο στην αριστερή πλευρά υπάρχει ένας “αφελής” Bayes που έχει "ξετυλιχτεί" για D χαρακτηριστικά, και ταυτόχρονα χρησιμοποιείται μια πλάκα για την αναπαράσταση της επανάληψης πάνω στις περιπτώσεις i = 1 : N. Στα δεξιά παρουσιάζεται η έκδοση του ίδιου μοντέλου με τη χρήση ένθετης συμβολής πλάκας. Εάν μια μεταβλητή βρίσκεται στο εσωτερικό δύο πλακών, λέμε ότι έχει δύο υποδείκτες.

Για παράδειγμα, γράφουμε θ_{ic} για την αναπαράσταση της παραμέτρου για το χαρακτηριστικό j στην πυκνότητα υπό συνθήκη κλάσης c. Μπορούμε να έχουμε πλάκες οι οποίες είναι ένθετες ή διασταυρούμενες. Συμβολισμοί για τη μοντελοποίηση περισσότερων σύνθετων προτύπων σύνδεσης παραμέτρων μπορούν να επινοηθούν (π.χ. (Heckermanetal. 2004)), αλλά δεν χρησιμοποιούνται συνήθως.

Κάτι που δεν είναι σαφές από το σχήμα είναι ότι το θjc χρησιμοποιείται για την παραγωγή του x_{ij} εάν και μόνο εάν $y_i = c$, ενώ σε κάθε άλλη περίπτωση αγνοείται. Αυτό καλείται παράδειγμα ειδικού πλαισίου ανεξαρτησίας. Επομένως, η υπό συνθήκη ανεξαρτησία σχέση $x_{ij} \perp \theta_{jc}$ ισχύει μόνο εάν $y_i \neq c$.



Σχήμα 3.2: (Murphy, K. P. (2012),[15])Ταξινομητής Naive Bayes ως DGM. (α) Με μια μόνοπλάκα. (β) Με φωλιασμένες πλάκες.

3.1.1. Εκτίμηση των παραμέτρων: Πίνακες υπό συνθήκη πιθανοτήτων

Λαμβάνοντας υπόψη την υποθετική έρευνα που περιεγράφηκε προηγουμένως στο 1.1.5, υποθέσαμε ότι γνωρίζουμε τόσο το κατευθυνόμενο ακυκλικό γράφημα όσο και τις παραμέτρους των τοπικών κατανομών που όριζαν το Μπεϋζιανό Δίκτυο. Σε αυτή την περίπτωση, τα δίκτυα Bayesian χρησιμοποιούνται ως εμπειρογνώμονα συστήματα, δεδομένου ότι τυποποιούν τη γνώση που κατέχουν ένας ή περισσότεροι εμπειρογνώμονες στους σχετικούς τομείς. Παρ' όλα αυτά, στις περισσότερες περιπτώσεις οι παράμετροι των τοπικών κατανομών εκτιμώνται ή μαθαίνονται με την παρατήρηση ενός δείγματος. Στην πραγματικότητα, όταν γίνεται χρήση της R, τα δεδομένα αποθηκεύονται σε ένα αρχείο κειμένου το οποίο μπορεί να εισαχθεί με την εντολή `read.table`, με μία μεταβλητή σε κάθε στήλη (με ετικέτα στην πρώτη γραμμή) και μία παρατήρηση σε κάθε γραμμή.

Σε αυτή την έρευνα, και γενικά στα διακριτά δίκτυα BayesianNetworks, οι παράμετροι που χρειάζονται να εκτιμηθούν είναι οι τοπικές κατανομές υπό συνθήκη πιθανοτήτων. Για παράδειγμα μπορούμε να τις εκτιμήσουμε με τις αντίστοιχες εμπειρικές συχνότητες στο σύνολο δεδομένων, οι οποίες σημαίνουν,

$$\begin{aligned}\hat{P}\left(OCC = em \mid EDU = h\right) &= \frac{\Pr\left(OCC = em, EDU = h\right)}{\Pr\left(EDU = h\right)} \\ &= \frac{\text{ο αριθμός των παρατηρήσεων που το } OCC = em \text{ και } EDU = h}{\text{ο αριθμός των παρατηρήσεων που το } EDU = h}\end{aligned} \tag{3.8}$$

Έτσι, προκύπτουν οι κλασικές συχνές εκτιμήσεις και οι εκτιμήσεις μέγιστης πιθανοφάνειας. Στην bnlearn μπορούν να υπολογιστούν με τη συνάρτηση bn.fit. Η bn.fit συμπληρώνει την custom.fit την οποία χρησιμοποιήσαμε προηγουμένως. Ένα Μπεϋζιανό Δίκτυο καθοδηγείται από την τελευταία χρησιμοποιώντας ένα σύνολο προσαρμοσμένων παραμέτρων που καθορίζουμε εμείς και εκτιμά το ίδιο από τα δεδομένα.

Με παρόμοιο τρόπο γίνεται η custom.fit, ένα αντικείμενο της κλάσης bn.fit που επιστρέφεται από την bn.fit. Ο εκτιμητής που θα χρησιμοποιηθεί καθορίζεται από το όρισμα της μεθόδου που στην περίπτωση αυτή είναι η "mle" για τον εκτιμητή μέγιστης πιθανοφάνειας. Υποθέτουμε ότι η δομή των δικτύων είναι γνωστή και πάλι και περνάει στη συνάρτηση μέσω του dag. Μπορούμε επίσης να υπολογίσουμε τις ίδιες εκτιμήσεις χειροκίνητα και να πιστοποιήσουμε ότι θα λάβουμε τα ίδια αποτελέσματα με την bn.fit (Scutari, M., & Denis, J. B. (2021)[8]).

```
survey<-read.table("survey.txt",header=TRUE)
head(survey)
bn.mle<-bn.fit(dag,data=survey,method= "mle")
prop.table(table(survey[,c("OCC","EDU")]),margin=2)
bn.mle$OCC
```

Παρουσιάζουμε τους παράμετρους του κόμβου OCC (πολυωνυμική κατανομή) στον παρακάτω πίνακα δεσμευμένων πιθανοτήτων (Scutari, M., & Denis, J. B. (2021)[8]).:

```
EDU
OCC h u
em 0.9808 0.9259
se 0.0192 0.0741
```

Εναλλακτικά, οι ίδιες υπό συνθήκη πιθανότητες σε ένα Μπεϋζιανό περιβάλλον μπορούν επίσης να εκτιμηθούν, με τη χρήση των εκ των υστέρων κατανομών τους. Στην περίπτωση αυτή, πρέπει να ορίσουμε τη μέθοδο του bn.fit σε "bayes" (Scutari, M., & Denis, J. B. (2021)[8]).

```
bn.bayes<-bn.fit(dag,data=survey,method="bayes",iss=10)
```

Από έναν ομοιόμορφο prior σε κάθε πίνακα υπό συνθήκη πιθανοτήτων μπορούμε να υπολογίσουμε τις εκτιμώμενες εκ των υστέρων πιθανότητες. Το προαιρετικό όρισμα iss, το όνομα του οποίου αντιπροσωπεύει το φανταστικό μέγεθος δείγματος που είναι επίσης γνωστό ως ισοδύναμο μέγεθος δείγματος, καθορίζει τον τρόπο με τον οποίο θα αποδοθεί το ποσό της βαρύτητας στην εκ των προτέρων κατανομή σε σύγκριση με τα δεδομένα κατά τον υπολογισμό της εκ των υστέρων κατανομής. Καθορίζουμε το βάρος ως το μέγεθος ενός φανταστικού δείγματος που υποστηρίζει την εκ των προτέρων κατανομή. Η τιμή του διαιρείται με τον αριθμό των κελιών που έχουμε στον πίνακα υπό συνθήκη πιθανοτήτων, δεδομένου ότι η εκ των προτέρων είναι επίπεδη και χρησιμοποιείται για τον υπολογισμό της εκ των υστέρων εκτίμησης ως σταθμισμένου μέσου όρου με τις εμπειρικές συχνότητες. Ας υποθέσουμε για παράδειγμα ότι έχουμε ένα δείγμα μεγέθους m, το οποίο μπορούμε να υπολογίσουμε ως εξής nrow(survey). Διατηρώντας το

$$\hat{p}_{em,h} = \frac{ο\,αριθμός\,των\,παρατηρήσεων\,οπου\,το\,OCC = em\,\,και\,\,το\,EDU = h}{m}$$

$$\hat{p}_{h} = \frac{ο\,αριθμός\,των\,παρατηρήσεων\,οπου\,το\,EDU = h}{m}$$

και στη συνέχεια συμβολίζοντας τις αντίστοιχες εκ των προτέρων πιθανότητες ως

$$\pi_{em,h} = \frac{1}{mOCC \times mEDU} \;\; και\, \pi_h = \frac{m0}{mOCC \times mEDU}$$

Όπου $mOCC = nlevels\left(bn.bayes.OCC\right)$ και $mEDU = nlevels\left(bn.bayes.EDU\right)$, έχουμε

$$\Pr\left(OCC = em, EDU = h\right) = \frac{iss}{m+iss}\pi_{em,h} + \frac{iss}{m+iss}\hat{p}_{em,h}$$

$$\Pr\left(EDU = h\right) = \frac{iss}{m+iss}\pi_{h} + \frac{iss}{m+iss}\hat{p}_{h}$$

.

Για αυτό το λόγο,

$$\Pr\left(OCC = em \mid EDU = h\right) = \frac{\Pr\left(OCC = em, EDU = h\right)}{\Pr\left(EDU = h\right)}$$

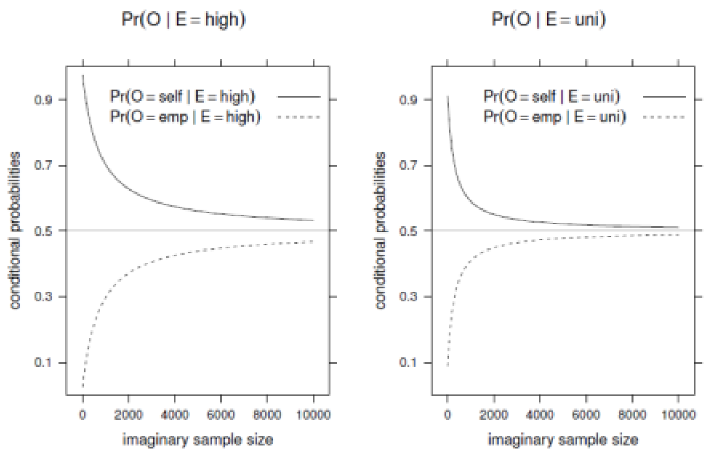
Στις περισσότερες περιπτώσεις, επιλέγουμε μια μικρή τιμή του iss, η οποία συνήθως κυμαίνεται μεταξύ 1 και 15, έτσι ώστε η εκ των προτέρων κατανομή να κυριαρχείται εύκολα από τα δεδομένα. Οι υπό όρους πιθανότητες που είναι πιο ομαλές αλλά εξακολουθούν να είναι κοντά στις εμπειρικές συχνότητες είναι το αποτέλεσμα μικρών τιμών, δηλαδή pbem,h και υπολογίζονται από

```
bn.bayes$OCC
```

Παρουσιάζουμε τους παράμετρους του κόμβου OCC (πολυωνυμική κατανομή) στον παρακάτω πίνακα δεσμευμένων πιθανοτήτων (Scutari, M., & Denis, J. B. (2021)[8]):

```
EDU
OCC h u
em 0.9743 0.9107
se 0.0257 0.0893
```

Από τον παραπάνω πίνακα των υπό συνθήκη πιθανοτήτων, φαίνεται ότι όλες οι εκ των υστέρων εκτιμήσεις απέχουν πολύ περισσότερο από το 0 και το 1 από τις αντίστοιχες εκτιμήσεις μέγιστης πιθανοφάνειας, επειδή επηρεάζονται από την εκ των προτέρων κατανομή. Αυτό μπορεί να γίνει αντιληπτό για διάφορους λόγους. Πρώτον, αυτό εξασφαλίζει ότι πληρούνται οι συνθήκες κανονικότητας των μεθόδων εκτίμησης και εξαγωγής συμπερασμάτων του μοντέλου. Ειδικότερα, είναι αδύνατο για τους αραιούς υπό συνθήκη πίνακες πιθανοτήτων (με πολλά μηδενικά κελιά) ακόμη και αν έχουμε μικρά σύνολα δεδομένων. Επιπλέον, οι εκτιμήσεις μεγίστης πιθανοφάνειας δεν είναι τόσο ισχυρές όσο οι εκ των υστέρων εκτιμήσεις και έτσι προκύπτουν τα BayesianNetworks με καλύτερη προβλεπτική ικανότητα. Μια αύξηση της τιμής του iss κάνει την εκ των υστέρων κατανομή πιο επίπεδη και την ωθεί προς την ομοιόμορφη κατανομή που θα χρησιμοποιηθεί ως εκ των προτέρων. Στο σχήμα 3.3 φαίνεται ότι για μεγάλες τιμές του iss οι υπό συνθήκη εκ των υστέρων κατανομές για τις $\Pr\left(OCC \mid EDU = h\right)$ και $\Pr\left(OCC \mid EDU = u\right)$ έχουν πιθανότητα περίπου ίση με 0,5 το οποίο αποδίδεται τόσο στην se όσο και στην em. Παρακάτω, αναφέρουμε ότι αυτή η τάση είναι ήδη εμφανής εάν οι υπό συνθήκη πιθανότητες που λαμβάνονται για iss = 10 με εκείνες για iss = 20, συγκρίνονται.



Σχήμα 3.3: (Scutari, M., & Denis, J. B. (2021)[8]). Κατανομές υπό συνθήκη πιθανοτήτων για OCC δεδομένων των πιθανών τιμών της EDU, δηλαδή Pr(OCC | EDU = h) και Pr(OCC | EDU = u), οι οποίες καθώς αυξάνεται το νοητό μέγεθος του δείγματος συγκλίνουν σε ομοιόμορφες κατανομές.

```
bn.ba,yes <- bn.fit(dag1,data=survey,method="bayes",iss=20)
bn.bayes$OCC
```

Παρουσιάζουμε τους παράμετρους του κόμβου OCC (πολυωνυμική κατανομή) στον παρακάτω πίνακα δεσμευμένων πιθανοτήτων (Scutari, M., & Denis, J. B. (2021)[8]):

```
EDU
OCC h u
em 0.968 0.897
se 0.032 0.103
```

Αλγόριθμοι βασισμένοι σε περιορισμούς

Οι αλγόριθμοι που βασίζονται σε περιορισμούς παρέχουν μια δομική ανακατασκευή και χρησιμοποιούν τεστ ανεξαρτησίας υπό όρους για να μάθουν τη δομή εξάρτησης των δεδομένων. Εντοπίζουν τους υπό συνθήκη περιορισμούς ανεξαρτησίας με στατιστικούς ελέγχους και συνδέουν κόμβους που δεν βρέθηκαν να είναι ανεξάρτητοι. Η βάση των αλγορίθμων που βασίζονται σε περιορισμούς είναι η ίδια με την εργασία των Pearl για τους χάρτες και την εφαρμογή του στα αιτιώδη γραφικά μοντέλα. Για την εκμάθηση της δομής του κατευθυνόμενου ακυκλικού γράφου των Μπεϋζιανών δικτύων, παρέχεται ένα πλαίσιο από τον αλγόριθμο InductiveCausation (IC) (VermaandPearl, 1991) με τη χρήση των υπό συνθήκη τεστ ανεξαρτησίας. Οι λεπτομέρειες του, περιγράφονται στον αλγόριθμο επαγωγικής αιτιολόγησης που θα συζητήσουμε στη συνέχεια. Πρώτα, πρέπει να προσδιορίσουμε τα ζεύγη μεταβλητών που συνδέονται με μια ακμή, αδιαφορώντας για την κατεύθυνσή της. Δεδομένου οποιουδήποτε άλλου υποσυνόλου μεταβλητών, αυτές δεν μπορούν να είναι ανεξάρτητες αφού δεν μπορούν να διαχωριστούν κατά D. Αυτό το βήμα θεωρείται επίσης ως μία προς τα πίσω διαδικασία επιλογής η οποία ξεκινά από το κορεσμένο μοντέλο με ένα πλήρες γράφημα και το «κόβει» με βάση τις στατιστικές δοκιμές για την υπό συνθήκη ανεξαρτησία.

3.1.2 Αλγόριθμος επαγωγικής αιτιολόγησης

Για των αλγόριθμο της επαγωγικής αιτιολόγισης εχουμε τα παρακάτω βήματα.

◊ Για κάθε ζεύγος κόμβων A και B στο σύνολο V πρέπει να αναζητήσουμε το σύνολο $S_{AB} \subset V$ έτσι ώστε οι A και B να είναι ανεξάρτητοι δεδομένου του S_{AB} και των $A, B \notin S_{AB}$. Τοποθετούμε μια μη κατευθυνόμενη ακμή μεταξύ των A και B εάν δεν υπάρχει τέτοιο σύνολο.

◊ Για κάθε ζεύγος μη γειτονικών κόμβων A και B που μοιράζονται τον ίδιο γείτονα C, πρέπει να ελέγξουμε αν ο $C \in S_{AB}$. Θέτουμε την κατεύθυνση των ακμών A - C και C - B σε A → C και C, εάν αυτό δεν ισχύει.

◊ Ορίζουμε την κατεύθυνση των ακμών που εξακολουθούν να είναι μη κατευθυνόμενες ακολουθώντας αναδρομικά ως εξής τους επόμενους δύο κανόνες:

(1) αν το A είναι γειτονικό με το B και έχουμε αυστηρά κατευθυνόμενη διαδρομή από το A στο B, τότε θέτουμε κατεύθυνση του A - B σε A → B,

(2) αν τα A και B δεν είναι γειτονικά αλλά A → Γ και Γ - B, τότε αλλάζουμε την τελευταία σε Γ → B.

◊ Επιστρέφει τα αποτελέσματα που προκύπτουν στο κατευθυνόμενο ακυκλικό γράφο τις δεσμευμένης πιθανότητας.

Στο επόμενο βήμα πρέπει να προσδιορίσουμε τις v-δομές μεταξύ όλων των ζευγών των μη γειτονικών κόμβων A και B που έχουν τον ίδιο γείτονα C. Οι δομές v είναι εξ ορισμού η μόνη θεμελιώδης σύνδεση στην οποία, υπό την προϋπόθεση του τρίτου κόμβου, οι δύο μη γειτονικοί κόμβοι δεν είναι ανεξάρτητοι. Για το λόγο αυτό, αν έχουμε ένα υποσύνολο κόμβων στο οποίο περιέχεται ο C και το D διαχωρίζει τους A και B. Αυτοί οι τρεις κόμβοι θεωρούνται ότι είναι μέρος μιας v-δομής που έχει κέντρο τον C. Μπορούμε να επαληθεύσουμε αυτήν τη συνθήκη αν εκτελέσουμε έναν έλεγχο ανεξαρτησίας υπό όρους για τους A και B, σε αντίθεση με κάθε πιθανό υποσύνολο των κοινών γειτόνων στο οποίο περιλαμβάνεται ο C. Σε αυτό το τέλος των βημάτων ο σκελετός και οι v-δομές του δικτύου θεωρούνται γνωστά και ως αποτέλεσμα ή κλάση ισοδυναμίας στην οποία ανήκει το BayesianNetwork που προσδιορίζεται μοναδικά. Το τελευταίο βήμα του επαγωγικού Causation (IC) αλγορίθμου είναι ο προσδιορισμός των αναγκαστικών ακμών και του προσανατολισμού τους αναδρομικά, για να προκύψει ο κατευθυνόμενος ακυκλικός γράφος, της υπό συνθήκη πιθανότητας που περιγράφει την κλάση ισοδυναμίας που προσδιορίστηκε από τα προηγούμενα βήματα. Ένα μεγάλο πρόβλημα του αλγορίθμου InductiveCausation (IC) είναι ότι δεν μπορούμε να εφαρμόσουμε τα δύο πρώτα βήματα με τη μορφή η οποία περιγράφεται στον αλγόριθμο επαγωγικής αιτιολόγησης (IC) σε οποιοδήποτε πρόβλημα του πραγματικού κόσμου λόγω του εκθετικού αριθμού των πιθανών σχέσεων ανεξαρτησίας υπό όρους.

Για το λόγο αυτό έχουν αναπτυχθεί βελτιωμένοι αλγόριθμοι. Η πρώτη πρακτική εφαρμογή του αλγορίθμου της επαγωγικής αιτιολόγησης (IC) ήταν το PC (Spirtesetal., 2000). Στη συνέχεια, ο αλγόριθμος Grow-Shrink (GS) ο οποίος βασίζεται στον αλγόριθμο Grow-Shrink Markov Blanket (Margaritis, (2003),[18]) και αποτελεί μια προσέγγιση ανίχνευσης της απλής επιλογής προς τα εμπρός Markov Blanket. Επιπροσθέτως ο αλγόριθμος Incremental Association (IAMB) ο οποίος βασίζεται στον αλγόριθμο Incremental Association Markov Blanket (Tsamardinos etal.,(2003),[19]) και διαθέτει ένα σχήμα επιλογής δύο φάσεων. Μετέπειτα, έχουμε τον αλγόριθμο Fast Incremental Association (Fast-IAMB) ο οποίος είναι μια παραλλαγή του IAMB που χρησιμοποιεί κερδοσκοπική βηματική προς τα εμπρός επιλογή, έτσι ώστε να μειώνεται ο αριθμός των δοκιμών ανεξαρτησίας υπό όρους (Yaramakala and Margaritis, (2005),[20]). Τέλος έχουμε τον Interleaved Incremental Association (Inter-IAMB) αλγόριθμο που είναι μια άλλη παραλλαγή του IAMB στην οποία η βηματική εμπρόσθια επιλογή χρησιμοποιείται (Tsamardinos etal., (2003),[19]), έτσι ώστε να αποφεύγονται τα ψευδώς θετικά αποτελέσματα στh φαση ανίχνευσης του Markov Blanket. Σε όλους τους αλγορίθμους που δόθηκαν παραπάνω, εκτός από τον αλγόριθμο PC, εμείς πρέπει πρώτα να μάθουμε κάθε κόμβο Markov Blanket.

Το προηγούμενο βήμα απλοποιεί εξαιρετικά τη διαδικασία αναγνώρισης των γειτόνων. Έτσι αυτό έχει ως αποτέλεσμα, τη σημαντική μείωση του αριθμού των δοκιμών ανεξαρτησίας υπό όρους και συνεπώς τη συνολική υπολογιστική πολυπλοκότητα του αλγορίθμου εκμάθησης. Επίσης, κάποιες βελτιώσεις είναι δυνατές αν αξιοποιήσουμε τη συμμετρία των MarkovBlankets. Επιπλέον, αφορά την ποιότητα των μαθημένων κατευθυνόμενων ακυκλικών γραφημάτων υπό συνθήκη πιθανότητας. Κατά μέσο όρο το Inter-IAMB παράγει λιγότερα ψευδώς θετικά αποτελέσματα από το Grow-Shrink, το IAMB ή το FastIAMB, ενώ έχει συγκρίσιμο αριθμό ψευδώς αρνητικών αποτελεσμάτων. Ο αλγόριθμος PC όπως επεκτάθηκε στους (Kalisch και Buhlmann (2007),[21]), (Kalisch και Buhlmann (2008),[22]) και (Buhlmannetal. (2010) [23])είναι επίσης ανταγωνιστικός. Σε σύνολα δεδομένων υψηλών διαστάσεων, ο ημιενδοδιαμετρικός Hiton-PC από τους Aliferisetal. (2010) είναι ίσως η καλύτερη επιλογή που μπορούμε να κάνουμε, η οποία κλιμακώνεται καλα μέχρι και χιλιάδες μεταβλητές.

3.1.3 Εκμάθηση της δομής DAG

Προηγουμένως εξετάσαμε τον κατευθυνόμενο ακυκλικό γράφο που διέπει το Μπεϋζιανό Δίκτυο. Με πιο απλά λόγια, βασιστήκαμε στην προηγούμενη γνώση που είχαμε για το φαινόμενο που μοντελοποιήσαμε, έτσι ώστε να αποφασίσουμε ποιες ακμές θα υπάρχουν στο γράφημα και ποιες δε θα είναι παρούσες. Παρ' όλα αυτά, μερικές φορές δεν είναι δυνατό ή επιθυμητό η δομή του κατευθυνόμενου ακυκλικού γράφου να είναι η ίδια με το αντικείμενο της έρευνας. Στο πλαίσιο των κοινωνικών επιστημών, η δομή του κατευθυνόμενου ακυκλικού γραφήματος μπορεί να προσδιορίζει τους κόμβους που σχετίζονται άμεσα με τον στόχο της ανάλυσης και για το λόγο αυτό μπορεί να χρησιμοποιηθεί για να βελτιώσει της διαδικασίας χάραξης πολιτικής. Για παράδειγμα, ο κατευθυνόμενος ακυκλικός γράφος της έρευνας που χρησιμοποιήσαμε ως παράδειγμα, υποδεικνύει ότι οι τιμές των εισιτηρίων των τρένων θα πρέπει να προσαρμοστούν έτσι ώστε το κέρδος να μεγιστοποιείται μόνο με βάση την κατοικία και την απασχόληση.

Η εκμάθηση ενός κατευθυνόμενου ακυκλικού γράφου των Μπεϋζιανών Δικτύων είναι μια πραγματικά πολύπλοκη εργασία, για δύο λόγους. Ο πρώτος, είναι ότι ο χώρος των πιθανών κατευθυνόμενων ακυκλικών γράφων είναι πολύ μεγάλος, πράγμα που σημαίνει ότι καθώς αυξάνεται ο αριθμός των κόμβων, ο αριθμός των κατευθυνόμενων ακυκλικών γραφημάτων αυξάνεται υπερεκθετικά. Το αποτέλεσμα αυτού, είναι ότι μπορούμε να διερευνήσουμε μόνο ένα μικρό κλάσμα των στοιχείων του σε εύλογο χρονικό διάστημα. Επιπλέον, ο χώρος αυτός διαφέρει πολύ από πραγματικούς χώρους (π.χ. $R, R^2, R^3, ...$) αφού δεν είναι συνεχής και έχει πεπερασμένο αριθμό στοιχείων. Αυτός είναι ο λόγος για τον οποίο χρειαζόμαστε ad-hoc αλγορίθμους για την εξερεύνησή του. Προς το παρόν η προσοχή μας θα περιοριστεί στις δύο κατηγορίες στατιστικών κριτηρίων που οι εν λόγω αλγόριθμοι χρησιμοποιούν για να αξιολογήσουν κατευθυνόμενους ακυκλικούς γράφους, οι οποίες είναι οι Έλεγχοι Ανεξαρτησίας Υπό Όρους και οι Βαθμολογίες δικτύου.

3.1.4 Δοκιμές ανεξαρτησίας υπό όρους

Οι Έλεγχοι Ανεξαρτησίας Υπό Όρους χρησιμοποιούνται για να εξεταστεί κατά πόσον οι μεταβλητές είναι ανεξάρτητες ή όχι υπό την αφαίρεση των στοιχείων και επικεντρώνονται στην παρουσία μεμονωμένων ακμών. Επειδή μια πιθανολογική εξάρτηση κωδικοποιείται από κάθε ακμή, οι έλεγχοι ανεξαρτησίας υπό όρους χρησιμοποιούνται για να διαπιστωθεί εάν τα δεδομένα υποστηρίζουν αυτή την πιθανολογική εξάρτηση. Με άλλα λόγια, οι έλεγχοι ανεξαρτησίας των δεσμευμένων πιθανοτήτων χρησιμοποιούνται για να μάθουμε ότι τα διακριτά Μπεϋζιανά Δικτύα είναι συναρτήσεις των συχνοτήτων που έχουμε παρατηρήσει $\left\{n_{ijk}, i=1,...,R, j=1,...,C, k=1,...,L\right\}$ για τις τυχαίες μεταβλητές X και Y και όλες τις διαμορφώσεις των εξαρτημένων μεταβλητών Z. Επομένως, έχουν αναπτυχθεί δύο δοκιμές:

◊ το mutual information test, ένα πληροφοριοθεωρητικό μέτρο απόστασης που ορίζεται ως (Scutari, M., & Denis, J. B. (2021)[8]):

$$MI(X,Y|Z)=\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}\frac{n_{ijk}}{n}\log\frac{n_{ijk}n_{++k}}{n_{i+k}n_{+jk}}$$

◊ ο κλασικός δείκτης X^2 του Pearson για πίνακες ενδεχομένων,

$$X^2\left(X,Y\mid Z\right)=\sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}\frac{\left(n_{ijk}-m_{ijk}\right)^2}{m_{ijk}}, όπου\, m_{ijk}=\frac{n_{i+k}n_{+jk}}{n_{++k}}$$

Θα μπορούσε επίσης να χρησιμοποιηθεί ο εκτιμητής συρρίκνωσης για την αμοιβαία πληροφορία που ορίζεται από τους (Hausser, J. and Strimmer, K. (2009)[25]), η οποία μελετήθηκε στο πλαίσιο των Μπεϋζιανών Δικτύων (Scutari, M., & Denis, J. B. (2021)[8]). Στην περίπτωση απόρριψης της μηδενικής υπόθεσης (της υπό συνθήκη ανεξαρτησίας), η ακμή θεωρείται ότι πρέπει να συμπεριληφθεί στον κατευθυνόμενο ακυκλικό γράφο. Για παράδειγμα, θεωρήστε την προσθήκη μιας ακμής από την Εκπαίδευση στο Ταξίδι (EDU → TRA) στο Κατευθυνόμενο Ακυκλικό Γράφημα. Έχουμε τη μηδενική υπόθεση ότι το Ταξίδι είναι πιθανοτικά ανεξάρτητο $\left(\perp_p\right)$ από την Εκπαίδευση, εξαρτώμενη από τους γονείς της, πράγμα που σημαίνει ότι:

$$H_0 : TRA \perp_p EDU \mid \left\{OCC, RES\right\}$$

και έχουμε την εναλλακτική υπόθεση ότι:

$$H_1 : TRA \not\perp_p \mid \left\{OCC, RES\right\}$$

Αυτή η μηδενική υπόθεση μπορεί να ελεγχθεί με την προσαρμογή του λόγου λογαριθμικής πιθανοφάνειας G^2 ή του Pearson X^2 για τον έλεγχο της υπό συνθήκη ανεξαρτησίας, αντί της οριακής ανεξαρτησίας. Για το G^2, η στατιστική ελέγχου έχει τη μορφή:

$$G^2\left(TRA, EDU \mid OCC, RES\right)=\sum_{t\in TRA}\sum_{e\in EDU}\sum_{k\in OCC\times RES}\frac{n_{tek}}{n}\log\frac{n_{tek}n_{++k}}{n_{++k}n_{+ek}}$$

όπου συμβολίζουμε τις κατηγορίες Travel με $t\in TRA$, τις κατηγορίες Education με $e\in EDU$, και τις διαμορφώσεις επαγγελμάτων και κατοικιών με $k\in OCC\times RES$.

Έτσι, η n_{tek} είναι ο αριθμός των παρατηρήσεων για το συνδυασμό μιας κατηγορίας t του Travel, μιας κατηγορίας e της Εκπαίδευσης και μιας κατηγορίας k του OCC × RES. Η χρήση του δείκτη "+" δηλώνει το άθροισμα επί ενός δείκτη, όπως δηλώνεται στο κλασικό βιβλίο του Agresti (2013),[24], και χρησιμοποιείται για την ένδειξη των οριακών μετρήσεων για τις υπόλοιπες μεταβλητές. Για το παράδειγμα, n_{t+k} είναι ο αριθμός των παρατηρήσεων για τα t και k που προκύπτουν αν αθροίσουμε όλες τις κατηγορίες της Εκπαίδευσης. Για το X^2 του Pearson χρησιμοποιούμε τον ίδιο συμβολισμό και έχουμε ότι:

$$X^2(TRA, EDU \mid OCC, RES) = \sum_{t \in TRA} \sum_{e \in EDU} \sum_{k \in OCC \times RES} \frac{(n_{tek} - m_{tek})^2}{m_{tek}}, \text{όπου } m_{tek} = \frac{n_{t+k} n_{+ek}}{n_{++k}}$$

Υπό τη μηδενική υπόθεση ότι και οι δύο έλεγχοι έχουν ασυμπτωτική χ^2 κατανομή σε αυτή την περίπτωση με 8 βαθμούς ελευθερίας

(Scutari, M., & Denis, J. B. (2021)[8]).

```
(nlevels(survey[,"TRA"])-1)*(nlevels(survey[,"EDU"])-1)*
(nlevels(survey[,"OCC"])-1)*(nlevels(survey[,"RES"])-1)
[1] 8
```

Σε μικρές τιμές των G^2 και X^2 προκύπτει υπό όρους ανεξαρτησία- αντίστροφα, για μεγάλες τιμές των στατιστικών δοκιμών απορρίπτεται η μηδενική υπόθεση, οι οποίες δοκιμές αυξάνονται με την ισχύ της υπό όρους εξάρτησης μεταξύ των μεταβλητών. Τόσο η G^2 όσο και η X^2 υλοποιούνται από τη συνάρτηση ci.test της bnlearn. Η G^2 δοκιμή είναι ισοδύναμη με τη δοκιμή Mutual Information Test από τη θεωρία της πληροφορίας και χρησιμοποιείται όταν θέτουμε test = "mi".

(Scutari, M., & Denis, J. B. (2021)[8]).

```
ci.test("TRA","EDU",c("OCC","RES"),test="mi",data=survey)
Mutual Information (disc.)
data: TRA ~ EDU | OCC + RES
mi = 9.88, df = 8, p-value = 0.2733
alternative hypothesis: true value is greater than 0
Pearson’s X2 test is used when test = ”x 2”.
ci.test("TRA","EDU",c("OCC","RES"),test="x2",data=survey)
 Pearson’s X2
data: TRA ~ EDU | OCC + RES
x2 = 5.74, df = 8, p-value = 0.6766
alternative hypothesis: true value is greater than 0
```

Και οι δύο έλεγχοι επιστρέφουν πολύ μεγάλες τιμές p-values, γεγονός που υποδηλώνει ότι η σχέση εξάρτησης που κωδικοποιείται από το EDU × TRA δεν είναι σημαντική δεδομένης της τρέχουσας Directed Acyclic Δομή των Ακυκλικών Γραφημάτων. Παρομοίως μπορούμε να ελέγξουμε αν μία από τις ακμές του Κατευθυνόμενου Άκυκλου Γράφου θα πρέπει να αφαιρεθεί, καθώς η σχέση εξάρτησης που κωδικοποιεί δεν υποστηρίζεται από τα δεδομένα. Έτσι, για παράδειγμα, η OCC → TRA μπορεί να αφαιρεθεί ελέγχοντας

$$H_0 : TRA \perp_p OCC \mid R \, vs \, H_1 : TRA \not\perp_p OCC \mid R \quad (3.13)$$

όπως φαίνεται στησυνέχεια.

```
ci.test("TRA","OCC","RES",test="x2",data=survey)
Pearson's X2 data: TRA ~ OCC | RES
x2 = 2.34, df = 4, p-value = 0.6727
alternative hypothesis: true value is greater than 0
```

Όπως και προηγουμένως, διαπιστώνουμε ότι το OCC × TRA δεν είναι σημαντικό. Ελέγχοντας κάθε άκρη με τη σειρά για σημαντικότητα, μπορεί να γίνει αυτόματα αν χρησιμοποιήσουμε τη συνάρτηση arc.strength και καθορίσουμε την ετικέτα δοκιμής με το όρισμα criterion.

```
arc.strength(dag,data=survey,criterin="x2")
  from to strength
1 AGE EDU 0.00098
2 SEX EDU 0.00125
3 EDU OCC 0.00264
 4 EDU RES 0.00056
5 OCCTRA 0.67272
6 RESTRA 0.00168
```

Το arc.strength έχει σχεδιαστεί για τη μέτρηση της ισχύος της πιθανοτικής εξάρτησης που αντιστοιχεί σε κάθε ακμή αν αφαιρέσουμε τη συγκεκριμένη ακμή από το γράφο και ποσοτικοποιήσουμε την αλλαγή με ένα πιθανοτικό κριτήριο. Ένα υπό συνθήκη τεστ ανεξαρτησίας ή ένα σκορ δικτύου είναι πιθανές επιλογές. Εάν έχουμε την περίπτωση των δοκιμών ανεξαρτησίας υπό όρους, η τιμή του επιχειρήματος του κριτηρίου είναι η ίδια με αυτή της δοκιμής όρου στο ci.test, και ο έλεγχος είναι ο κόμβος "to" να είναι ανεξάρτητος "από" τον κόμβο υπό την προϋπόθεση των υπόλοιπων γονέων του κόμβου "προς". Η ισχύς που αναφέρεται είναι η προκύπτουσα τιμή p-value. Από την παραπάνω έξοδο βλέπουμε ότι όλες οι ακμές με εκτός από OCC → TRA έχουν τιμές p-values που είναι μικρότερες από 0,05 και υποστηρίζονται καλά από τα δεδομένα.

3.1.5 Αποτελέσματα δικτύου

Οι βαθμολογίες δικτύου σε αντίθεση με τις δοκιμές ανεξαρτησίας υπό όρους επικεντρώνονται στην κατευθυνόμενη AcyclicGraph στο σύνολό του, οι οποίες είναι στατιστικά στοιχεία καλής προσαρμογής που μετρούν πόσο καλά ένα κατευθυνόμενο ακυκλικό γράφημα μπορεί να αντικατοπτρίζει τη δομή εξάρτησης των δεδομένων. Διάφορες βαθμολογίες χρησιμοποιούνται συνήθως και ένα από αυτά είναι το BayesianInformationCriterion (BIC) το οποίο στο παράδειγμά μας έχει την ακόλουθη μορφή BN (Scutari, M., & Denis, J. B. (2021)[8]):

$$
\begin{aligned}
BIC &= \log \Pr\left(AGE, SEX, EDU, OCC, RES, TRA\right) - \frac{d}{2}\log n \\
&= \left[\Pr\left(AGE\right) - \frac{d_{AGE}}{2}\log n\right] + \left[\log \Pr\left(SEX\right) - \frac{d_{SEX}}{2}\log n\right] \\
&+ \left[\log \Pr\left(EDU \mid AGE, SEX\right) - \frac{d_{EDU}}{2}\log n\right] + \left[\log \Pr\left(OCC \mid EDU\right) - \frac{d_{OCC}}{2}\log n\right] \\
&+ \left[\log \Pr\left(RES \mid EDU\right) - \frac{d_{RES}}{2}\log n\right] + \left[\log \Pr\left(TRA \mid OCC, RES\right) - \frac{d_{TRA}}{2}\log n\right]
\end{aligned}
$$

όπου έχουμε ότι n είναι το μέγεθος του δείγματος, d ο αριθμός των παραμέτρων ολόκληρου του δικτύου και $d_{AGE}, d_{SEX}, d_{EDU}, d_{OCC}, d_{RES}, d_{TRA}$ είναι οι αριθμοί των παραμέτρων κάθε κόμβου. Με την αποσύνθεση στην εξίσωση (1.1) είναι εύκολο για το Bayesian Information Criterion (BIC) να υπολογιστεί από τις τοπικές κατανομές. Το ισοδύναμο Bayesian Dirichlet equivalent uniform (BDeu) είναι μια άλλη βαθμολογία που χρησιμοποιείται συνήθως στη βιβλιογραφία και συχνά συμβολίζεται ακριβώς ως Bayesian Dirichlet equivalent (BDe). Τόσο το Bayesian Information Criterion (BIC) όσο και το Bayesian Dirichlet equivalent (BDe) αποδίδουν υψηλότερες βαθμολογίες στους κατευθυνόμενους ακυκλικούς γράφους τα οποία ταιριάζουν καλύτερα με τα δεδομένα.

Μπορούμε να υπολογίσουμε και τις δύο βαθμολογίες στο bnlearn χρησιμοποιώντας τη συνάρτηση score. Η BIC υπολογίζεται όταν θέτουμε type = "bic", και το logBDe όταν θέτουμε type = "bde" (Scutari, M., & Denis, J. B. (2021)[8]).

```
score(dag,data=survey,type="bic")
[1] -2012.69
score(dag,data=survey,type="bde",iss=10)
[1] -1998.28
```

Το επιχείρημα iss για το Bayesian Dirichlet equivalent είναι το ίδιο φανταστικό μέγεθος δείγματος που εισάγαμε όταν υπολογίζαμε τις εκ των υστέρων εκτιμήσεις των παραμέτρων των BayesianNetworks προηγουμένως. Και πάλι, μπορούμε να το ερμηνεύσουμε ως το βάρος που αποδίδεται στην επίπεδη εκ των προτέρων κατανομή ως προς το μέγεθος του φανταστικού δείγματος. Εάν το iss έχει μικρές τιμές ή μεγάλα παρατηρούμενα δείγματα, το λογαριθμικό Bayesian Dirichlet equivalent και η Bayesian Information Criterion επιστρέφουν παρόμοιες τιμές, (Scutari, M., & Denis, J. B. (2021)[8]).

```
score(dag1,data=survey,type="bde",iss=1)
[1] -2015.65
```

Με τη χρήση οποιασδήποτε από αυτές τις βαθμολογίες μπορούν να συγκριθούν διαφορετικοί κατευθυνόμενοι ακυκλικοί γράφοι και να συγκριθούν ώστε να διερευνηθεί ποιο ταιριάζει καλύτερα στα δεδομένα. Μπορούμε επίσης να χρησιμοποιήσουμε βαθμολογίες για τη σύγκριση Δικτύων που διαφέρουν εντελώς σε αντίθεση με τα τεστ ανεξαρτησίας υπό όρους. Ακόμη και ένας κατευθυνόμενος ακυκλικός γράφος μπορεί να δημιουργηθεί τυχαία με τη συνάρτηση random.graph και να συγκριθεί με τους προηγούμενους κατευθυνόμενους ακυκλικούς γράφους μέσω της βαθμολογίας του (Scutari, M., & Denis, J. B. (2021)[8]).

```
rnd<-random.graph(nodes=c("AGE","SEX","EDU","OCC","RES","TRA"))
modelstring(rnd)
[1]"[AGE][SEX | AGE][EDU | AGE : SEX][OCC | SEX : EDU][RES | SEX :
EDU][T RA | SEX : EDU]"
score(rnd,data=survey,type="bic")
[1] -2034.99
```

Όπως περιμέναμε, το rnd είναι χειρότερο από το dag1 αλλά ακόμη και από το dag4, άλλωστε, ούτε τα δεδομένα ούτε η κοινή λογική ήταν για την επιλογή της δομής του. Εκμάθηση του κατευθυνόμενου ακυκλικού γραφήματος από την έρευνα επιστρέφει ένα Δίκτυο που είναι πολύ καλύτερο. Υπάρχει ένας αριθμός αλγορίθμων που αντιμετωπίζουν αυτό το πρόβλημα αναζητώντας το κατευθυνόμενο ακυκλικό γράφημα που μεγιστοποιεί μια δεδομένη βαθμολογία ενός Δικτύου. Ένας απλός αλγόριθμος είναι η αναρρίχηση λόφου δηλαδή ξεκινάμε από ένα κατευθυνόμενο ακυκλικό γράφο χωρίς ακμές, προσθέτουμε, αφαιρούμε και αντιστρέφουμε μία ακμή κάθε φορά και επιλέγουμε την αλλαγή που αυξάνει περισσότερο τη βαθμολογία του Δικτύου. Υλοποιείται στο hc, η οποία στην απλούστερη μορφή της λαμβάνει τα δεδομένα ως μοναδικό όρισμα και έχει ως προεπιλογή την τιμή τη βαθμολογία του Bayesian Information Criterion (Scutari, M., & Denis, J. B. (2021)[8]) .

```
learned<-hc(survey)
modelstring(learned)
[1] "[RES][EDU|RES][TRA|RES][AGE|EDU][OCC|EDU][SEX|EDU]"
score(learned,data=survey,type="bic")
[1] -1998.43
```

Μπορούμε να καθορίσουμε άλλες βαθμολογίες με το όρισμα score. Για παράδειγμα, μπορούμε να αλλάξουμε το προεπιλεγμένη βαθμολογία = "bic" σε βαθμολογία = "bde".

```
learned2<-hc(survey,score="bde")
```

Δεν αποτελεί έκπληξη το γεγονός ότι αν αφαιρέσουμε οποιαδήποτε ακμή από το εκπαιδευμένο, το οποίο μειώνει την Μπεϋζιανή του Βαθμολογία του Bayesian Information Criterion. Αυτό μπορεί να επιβεβαιωθεί εύκολα αν χρησιμοποιήσουμε το arc.strength, το οποίο αναφέρει την αλλαγή στη βαθμολογία που προκαλείται από την αφαίρεση μιας ακμής ως τη δύναμη του τόξου όταν το κριτήριο είναι ένα σκορ δικτύου (Scutari, M., & Denis, J. B. (2021)[8]).

```
arc.strength(learned,data=survey,criterion="bic")
from to strength
1 RES EDU -3.390
2 EDU SEX -2.726
3 RES TRA -1.848
4 EDU AGE -1.720
```

```
5 EDUOCC -0.827
```

Αυτό δεν ισχύει για το dag1, αν υποθέσουμε ότι όλες οι εξαρτήσεις που κωδικοποιούνται δε μπορούν να μαθευτούν σωστά από την έρευνα (Scutari, M., & Denis, J. B. (2021)[8]).

```
arc.strength(dag1,data=survey,criterion="bic")
from to strength
1 AGE EDU 2.489
2 SEX EDU 1.482
3 EDU OCC -0.827
4 EDU RES -3.390
5 OCCTRA 10.046
6 RESTRA 2.973
```

Ιδιαίτερα όταν αφαιρείται το OCC → TRA παρατηρείται σημαντική αύξηση της BayesianInformationCriterionscore, γεγονός που συνάδει με την υψηλή τιμή p-value που παρατηρήθηκε για αυτή την ακμή όταν χρησιμοποιήσαμε το arc.strength προηγουμένως.

Δομή μάθησης: Διακριτές μεταβλητές:

Δεδομένου ενός πειράματος που επαναλαμβάνεται και του οποίου το αποτέλεσμα καθορίζει την κατάσταση του m, τυχαίων μεταβλητών για το τι θα ακολουθήσει, υποθέτουμε τα εξής:

◊ Ένας πιστός κατευθυνόμενος ακυκλικός γράφος γίνεται δεκτός από τη σχετική κατανομή συχνοτήτων των μεταβλητών.

◊ Αν υποθέσουμε ότι η κατανομή σχετικής συχνότητας έχει όλες και μόνο κάποιες υπό συνθήκη ανεξαρτησίες, μοντελοποιούμε τις πεποιθήσεις μας για την πιθανότητα του αποτελέσματος των N εκτελέσεων του πειράματος με ένα πολυωνυμικό δείγμα BayesianNetwork με παράμετρο (G, F) έτσι ώστε το G να συνεπάγεται με όλες και μόνο εκείνες τις υπό συνθήκη ανεξαρτησίες.

Εάν οι υπό συνθήκη ανεξαρτησίες μιας κατανομής είναι:

$$IP\left(\{X\},\{Y\}\right) IP\left(\{X\},\{Y\}\{Z\}\right),$$

και μόνο αυτές τις υπό όρους ανεξαρτησίες, τότε η κατανομή δεν επιτρέπει να αναπαρασταθεί ένα πιστό κατευθυνόμενο ακυκλικό γράφημα. Τώρα, αν οι υπό συνθήκη ανεξαρτησίες μιας κατανομής είναι:

$$IP\left(\{L\},\{F,S\}\right)IP\left(\{L\},\{S\}\right)IP\left(\{F\},\{V\}\right)IP\left(\{F\},\{L,V\}\right)IP\left(\{L\},\{F\}\right).$$

και μόνο αυτές τις υπό συνθήκη ανεξαρτησίες, τότε η κατανομή δεν επιτρέπει την αναπαράσταση ενός πιστού κατευθυνόμενου ακυκλικού γράφου.

3.1.6 Σχήμα για τη δομή μάθησης

Είναι γνωστό ότι το σύνολο όλων των κατευθυνόμενων ακυκλικών γραφημάτων χωρίζεται σε διαχωρισμένες κλάσεις ισοδυναμίας, με την ισοδυναμία Markov και όλα τα κατευθυνόμενα ακυκλικά γραφήματα σε μία δεδομένη κλάση ισοδυναμίας Markov που είναι πιστά στις ίδιες κατανομές πιθανότητας. Μπορούμε επίσης να δημιουργήσουμε ένα γράφημα που ονομάζεται πρότυπο κατευθυνόμενου ακυκλικού γραφήματος το οποίο θα αναπαριστά κάθε Markov κλάση ισοδυναμίας και αν το P επιτρέπει μια πιστή αναπαράσταση κατευθυνόμενου ακυκλικού γραφήματος και τότε θα έχουμε ένα μοναδικό μοτίβο Κατευθυνόμενου Άκυκλου Γραφήματος που είναι πιστό στην P. Για το λόγο αυτό, παρόλο που ένα μοναδικό Κατευθυνόμενο Άκυκλο Γράφημα δεν μπορεί να προσδιοριστεί από την υπό συνθήκη ανεξαρτησία στο P, ένα μοναδικό μοτίβο κατευθυνόμενου ακυκλικού γραφήματος με αυτές τις υπό συνθήκη ανεξαρτησίες μπορεί να προσδιοριστεί. Το GP θα χρησιμοποιηθεί ως τυχαία μεταβλητή που έχει πιθανές τιμές τα μοτίβα κατευθυνόμενων ακυκλικών γράφων gp. Όσον αφορά την πραγματική σχετική συχνότητα κατανομής, ένα γεγονός μοτίβου κατευθυνόμενου ακυκλικού γράφου gp είναι το γεγονός που το gp είναι πιστό στη σχετική κατανομή συχνοτήτων. Μερικές φορές μπορούμε να θεωρήσουμε τα Κατευθυνόμενα AcyclicGraphs ως γεγονότα.
Δίνουμε τώρα τον ακόλουθο ορισμό που αφορά τη δομή μάθησης:

Ορισμός 9. Αυτό που ακολουθεί αποτελεί μια πολυωνυμική δομή Μπεϋζιανού Δικτύου
Σχήμα μάθησης:

◊ m τυχαίες μεταβλητές $X_1, X_2, ..., X_m$ με διακριτή κοινή κατανομή πιθανότητας P,

◊ ένα ισοδύναμο μέγεθος δείγματος M,

◊ για κάθε πρότυπο κατευθυνόμενου ακυκλικού γράφου gp που περιέχει τις m μεταβλητές, ένα πολυωνυμικό επαυξημένο δίκτυο Bayes ($G, F^{((G)}, \rho|G$) με ισοδύναμο μέγεθος δείγματος M, όπου G είναι οποιοδήποτε μέλος της κλάσης ισοδυναμίας που αναπαρίσταται από το gp, έτσι ώστε το P είναι το κατανομή πιθανότητας στο ενσωματωμένο δίκτυο Bayes,

Παράδειγμα:

Ένα σχήμα μάθησης δομής πολυωνυμικού δικτύου BayesianNetwork που περιέχει δύο μεταβλητές, αναπτύσσεται.

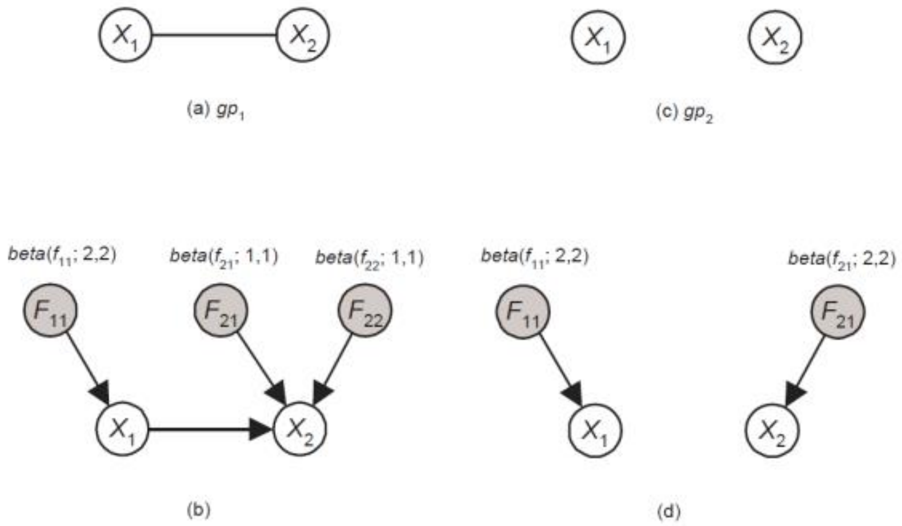
◊ Καθορίζουμε δύο τυχαίες μεταβλητές Y_1και Y_2, καθεμία από τις οποίες έχει χώρο {1, 2}, και αναθέτουμε

$$P(Y_1 = 1, Y_2 = 1) = \frac{1}{4}$$
$$P(Y_1 = 1, Y_2 = 2) = \frac{1}{4}$$
$$P(Y_1 = 2, Y_2 = 1) = \frac{1}{4}$$
$$P(Y_1 = 2, Y_2 = 2) = \frac{1}{4}$$

◊ Τότε καθορίζουμε M = 4.

◊ Τα δύο μοτίβα του κατευθυνόμενου ακυκλικού γράφου φαίνονται στο επόμενο Σχήμα 3.4(α) και (γ), και τα επαυξημένα δίκτυα Bayesian Networks παρουσιάζονται στο σχήμα 3.4(β) και (δ) τα οποία δίνονται στη συνέχεια.

Όπως είναι γνωστό, το G μπορεί να είναι οποιοδήποτε στοιχείο από την κλάση ισοδυναμίας που αντιπροσωπεύεται από το gp. Έτσι ο Κατευθυνόμενος άκυκλος γράφος που δίνεται στο σχήμα 3.4 (β) θα μπορούσε να είναι $Y_1 \leftarrow Y_2$. Σημειώστε ότι το gp_1 δεν αναπαριστά καμία ανεξαρτησία, ενώ το gp_2 αναπαριστά το $IP(\{Y_1\}, \{Y_2\})$. Παρόλο που ένα Μπεϋζιανό Δίκτυο το οποίο περιέχει το κατευθυνόμενο ακυκλικό γράφημα $Y_1 \rightarrow Y_2$ είναι δυνατόν να περιέχει μια κατανομή όπου τα Y_1 και Y_2 είναι ανεξάρτητα, το γεγονός gp_1 είναι το γεγονός που τους κάνει να εξαρτώνται και γι' αυτό δεν επιτρέπει την πιθανότητα να είναι ανεξάρτητα. Επομένως, όταν εξαρτάτε από το γεγονός gp_1, καθιστούμε τα F_{21} και F_{22} ανεξάρτητα.



Σχήμα 3.4: (Scutari, M., & Denis, J. B. (2021)[8]).Τα μοτίβα DAG βρίσκονται στα (α) και (γ) και τα αντίστοιχα επαυξημένες δομές δικτύων Bayes είναι στα (β) και (δ).

Ας αναλύσουμε τι σημαίνουν αυτά τα F_{ij} :

1. Κάθε F_{ij} αντιστοιχεί σε μια παράμετρο της υπό συνθήκη πιθανότητας, η οποία σε αυτή την περίπτωση είναι διακριτή και παίρνει τιμές βάσει της βήτα κατανομής. Όπως φαίνεται στην εικόνα, παρατηρούμε διάφορες βήτα κατανομές, π.χ., $\beta(f_{11};2,2)$ και $\beta(f_{21};1,1)$. Οι παραμέτροι βήτα (π.χ., το $\beta(f_{11};2,2)$) μας δίνουν πληροφορίες για τον τρόπο με τον οποίο κατανέμεται η πιθανότητα για κάθε μεταβλητή με βάση τα δεδομένα και τις υπό συνθήκη κατανομές που υπάρχουν στο δοσμένο Μπεϋζιανό Δίκτυο.
2. Τα F_{ij} εκφράζουν την πιθανότητα ενός γεγονότος, δεδομένων των τιμών των γονικών μεταβλητών. Για παράδειγμα, στην εικόνα του σχήματος 3.4(β), το F_{ij} συνδέεται με την X_1,υποδεικνύοντας την υπό συνθήκη πιθανότητα των τιμών του X_1σε σχέση με τους γονείς του. Στη συγκεκριμένη περίπτωση, αυτή η πιθανότητα έχει μορφή βήτα κατανομής $\beta(f_{11};2,2)$.

Όταν ενσωματώνουμε αυτό το κομμάτι στην πολυωνυμική δομή του Μπεϋζιανου Δικτύου, τα F_{ij} βοηθούν να εκφράσουμε τις υπό συνθήκη πιθανότητες σε κάθε βήμα, ενώ τα μοτίβα gp καθορίζουν τις σχέσεις ανεξαρτησίας ή εξάρτησης μεταξύ των μεταβλητών.

Επομένως, συνοπτικά:

- Τα F_{ij} είναι παράμετροι που ορίζουν τις υπό συνθήκη πιθανότητες των μεταβλητών X_1 και X_2.
- Η βήτα κατανομή σε κάθε παράμετρο F_{ij} εκφράζει την πιθανότητα της συγκεκριμένης μεταβλητής υπό τις συνθήκες του δικτύου.

Αυτό το πλαίσιο μάθησης δομής του Μπεϋζιανού Δικτύου είναι ένας τρόπος να εκφραστεί η εξάρτηση ή η ανεξαρτησία των μεταβλητών σε ένα μοντέλο, και το πρότυπο κατευθυνόμενου ακυκλικού γράφου gp χρησιμοποιείται για να εκφράσει όλες τις πιθανές καταστάσεις ανεξαρτησίας μεταξύ των μεταβλητών μας.

Το κάνουμε αυτό μόνο όταν τα Y_1 και Y_2 εξαρτώνται. Γενικά, δεν αναθέτουμε άμεσα μια κοινή κατανομή πιθανότητας, καθώς ο αριθμός των τιμών στην κοινή κατανομή είναι μεγαλύτερος εκθετικά με τον αριθμό των μεταβλητών. Προτιμούμε να αναθέσουμε την υπό συνθήκη κατανομή σε όλα τα επαυξημένα δίκτυα Bayes, έτσι ώστε οι κατανομές πιθανότητας σε όλα τα ενσωματωμένα Μπεϋζιανά Δίκτυα να γίνονται τα ίδια. Ένας κοινός τρόπος για να γίνει αυτό είναι με την κατασκευή για κάθε πρότυπο κατευθυνόμενου ακυκλικού γράφου gp, ενός επαυξημένου Bayesian Δικτύου με ισοδύναμο μέγεθος δείγματος M στο οποίο το ενσωματωμένο δίκτυο Μπεϋζιανού περιέχει την ομοιόμορφη κατανομή. Αυτό σημαίνει ότι για ένα δεδομένο κατευθυνόμενο ακυκλικό γράφημα μοτίβο gp πρέπει πρώτα να ορίσουμε ένα κατευθυνόμενο ακυκλικό γράφημα G στην κλάση ισοδυναμίας του αντιπροσωπεύει. Στη συνέχεια στο επαυξημένο δίκτυο Bayes που αντιστοιχεί στο G για όλα τα i, j, και k θέτουμε

$$a_{i,j,k} = \frac{M}{r_i q_i}$$

όπου r_i είναι ο αριθμός των πιθανών τιμών του Y_i στο G, και q_i είναι ο αριθμός των διαφορετικών περιπτώσεων των γονέων του Y_i στο G. Αυτός είναι ο τρόπος με τον οποίο τα Δίκτυα στο Σχήμα 1.21 δημιουργήθηκαν.

3.1.7 Διαδικασία για την εκμάθηση της δομής

Τώρα θα εξετάσουμε τον τρόπο εκμάθησης της Δομής με τη χρήση μιας πολυωνυμικής Μπεϋζιανής Δικτύου για την εκμάθηση δομών. Θα ξεκινήσουμε με τον επόμενο ορισμό:

Ένας πολυωνυμικός χώρος μάθησης δομής δικτύου Bayesian αποτελείται από τα εξής:

◊ ένα σχήμα μάθησης πολυωνυμικής δομής δικτύου Bayes που περιέχει τις ακόλουθες μεταβλητές $Y_1, Y_2, ..., Y_m$

◊ μια τυχαία μεταβλητή GP η οποία έχει ένα εύρος όλων των σχημάτων κατευθυνόμενου ακυκλικού γράφου, περιέχει τις m μεταβλητές, και για κάθε τιμή gp της GP έχει μια εκ των προτέρων πιθανότητα $P(gp)$.

◊ ένα σύνολο $D = \left\{Y^{(1)}, Y^{(2)}, ..., Y^{(N)}\right\}$ από m-διάστατα τυχαία διανύσματα έτσι ώστε κάθε $Y_i^{(h)}$ έχει τον ίδιο χώρο με το Y_i.Για κάθε τιμή gp του GP, το D είναι ένα πολυωνυμικό δίκτυο BayesianNetwork μεγέθους δείγματος N με παράμετρο $\left(G, F^{(G)}\right)$, όπου $\left(G, F^{(G)}\right)$είναι το πολυωνυμικό επαυξημένο BayesianNetwork που αντιστοιχεί στο gp στις προδιαγραφές του σχήματος.

Ας υποθέσουμε ότι έχουμε έναν τέτοιο χώρο και ένα σύνολο δεδομένων d από τιμές των διανυσμάτων του D έχουμε:

$$P(d \mid gp) = P(d \mid G) = \prod_{i=1}^{n} \prod_{j=1}^{q_i^{(G)}} \frac{\Gamma\left(N_{ij}^{(G)}\right)}{\Gamma\left(M_{ij}^{(G)} + N_{ij}^{(G)}\right)} \prod_{k=1}^{r_i} \frac{\Gamma\left(a_{ijk}^{(G)} + s_{ijk}^{(G)}\right)}{\Gamma\left(a_{ijk}^{(G)}\right)} \tag{3.9}$$

όπου $a_{ijk}^{(G)}$ και $s_{ijk}^{(G)}$ είναι τιμές του $\left(G, F^{(G)}, \rho \mid G\right)$.

Για έναν κατευθυνόμενο ακυκλικό γράφο (ή πρότυπο κατευθυνόμενου ακυκλικού γράφου) ένα κριτήριο βαθμολόγησης είναι μια συνάρτηση που αποδίδει μια τιμή σε κάθε κατευθυνόμενο ακυκλικό γράφο (ή κατευθυνόμενο ακυκλικό μοτίβο). Graphpattern) υπό εξέταση, η οποία βασίζεται στα δεδομένα. Στην ισότητα (3.10) η έκφραση που δίνεται ονομάζεται Bayesian Scoring Criterion $score_B$, το οποίο χρησιμοποιούμε για να βαθμολογήσουμε κατευθυνόμενους ακυκλικούς γράφους και μοτίβα κατευθυνόμενων ακυκλικών γράφων. Αυτό σημαίνει ότι,

$$score_B\left(d, gp\right) = score_B\left(d, G\right) = P\left(d \mid G\right) \tag{3.10}$$

Στην ισότητα (3.9) έχουμε εξαρτήσει ένα πρότυπο κατευθυνόμενου ακυκλικού γράφου για τον υπολογισμό της πιθανότητας ότι D = d. Αυτή η δομή έχει αποτελέσει μέρος της προηγούμενης γνώσης υποβάθρου όταν αναπτύξαμε τον χώρο πιθανοτήτων μας, και για τον λόγο αυτό δεν την εξαρτήσαμε από κάποια συνθήκη. Η επιλογή μοντέλου συνεπάγεται τον προσδιορισμό και την επιλογή των μοτίβων DAG με τη μέγιστη πιθανότητα, υπό την προϋπόθεση των δεδομένων ενός Μπευζιανού πολυωνυμικού Δομής δικτύου μάθησης χώρου και δεδομένων. Γενικά θα μπορούσαν να υπάρχουν περισσότερα από ένα πρότυπο κατευθυνόμενου ακυκλικού γραφήματος που μεγιστοποιεί την τιμή του. Το ζητούμενο της επιλογής μοντέλου είναι η εκμάθηση ενός μοτίβου κατευθυνόμενου ακυκλικού γράφου καθώς και των τιμών των παραμέτρων του που μπορούν να χρησιμοποιηθούν για λήψη αποφάσεων και εξαγωγή συμπερασμάτων.

Θα δώσουμε τώρα ένα παράδειγμα για καλύτερη κατανόηση.

Παράδειγμα:

Έστω ότι κάνουμε μια μελέτη σχετικά με τα παντρεμένα άτομα μέχρι την ηλικία των 30 ετών και θα θέλαμε να δούμε αν υπάρχει συσχέτιση μεταξύ της αποφοίτησης από το κολέγιο και του διαζυγίου. Πρώτα καθορίζουμε τις τυχαίες μεταβλητές στον πίνακα 3.1 παρακάτω.

Πίνακας 3.1: Προδιαγραφές των τυχαίων μεταβλητών.

Μεταβλητές	Τιμές	Όταν οι μεταβλητές παίρνουν αυτές τις Τιμές
X1	1	Το άτομο αποφοίτησε από το κολλέγιο
	2	Το άτομο δεν αποφοίτησε από το κολέγιο
X2	1	Το άτομο χώρισε στην ηλικία των 50 ετών

	2	Το άτομο δεν χώρισε στην ηλικία των 50 ετών

Στη συνέχεια, αναπαριστούμε τις προηγούμενες πεποιθήσεις μας χρησιμοποιώντας την εκμάθηση της δομής του Μπευζιανού Δικτύου στο παράδειγμα που δίνεται παραπάνω. Μετά από αυτό, το σχήμα κατευθυνόμενου ακυκλικού γράφου gp_1 στο Σχήμα 3.4 (α) αναπαριστά το γεγονός ότι συσχετίζονται και το Κατευθυνόμενο AcyclicGraph μοτίβο gp_2 στο Σχήμα 3.4 (β) αναπαριστά το γεγονός ότι είναι ανεξάρτητες. Στη συνέχεια υποθέστε ότι τα δεδομένα d λαμβάνονται στον ακόλουθο πίνακα:

Πίνακας 3.2: Δεδομένα d.

Περίπτωση	X_1	X_2
1	1	1
2	1	2
3	1	1
4	2	2
5	1	1
6	2	1
7	1	1
8	2	2

Τότε,

$$P(d \mid gp_1) = \left(\frac{\Gamma(4)}{\Gamma(4+8)} \frac{\Gamma(2+5)\Gamma(2+3)}{\Gamma(2)\Gamma(2)} \right) \left(\frac{\Gamma(2)}{\Gamma(2+5)} \frac{\Gamma(1+4)\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \right) \left(\frac{\Gamma(2)}{\Gamma(2+3)} \frac{\Gamma(1+1)\Gamma(1+2)}{\Gamma(1)\Gamma(1)} \right)$$
$$= 7.2150 \times 10^{-6}$$

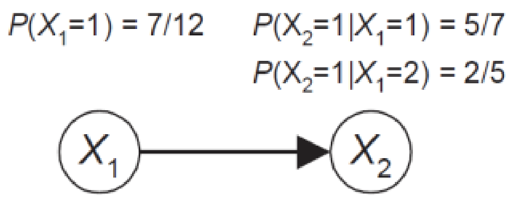
$$P(d \mid gp_2) = \left(\frac{\Gamma(4)}{\Gamma(4+8)} \frac{\Gamma(2+5)\Gamma(2+3)}{\Gamma(2)\Gamma(2)} \right) \left(\frac{\Gamma(4)}{\Gamma(4+8)} \frac{\Gamma(2+5)\Gamma(2+3)}{\Gamma(2)\Gamma(2)} \right) = 6.7465 \times 10^{-6}$$

Έτσι, αν αναθέσουμε

$$P(gp_1) = P(gp_2) = 0.5,$$

τότε από το θεώρημα του Bayes

$$P(gp_1 \mid d) = \frac{P(d \mid gp_1)P(gp_1)}{P(d)} = \frac{7.2150 \times 10^{-6}(0.5)}{P(d)} = a\left(3.6075 \times 10^{-6}\right)$$



Σχήμα 3.5: (Neapolitan, R. E. (2004),[17])Το Μπεϋζιανό Δίκτυο το οποίο αναπτύσσεται στο παραπάνω παράδειγμα.

και

$$P(gp_2 \mid d) = \frac{P(d \mid gp_2)P(gp_1)}{P(d)} = \frac{6.7465 \times 10^{-6}}{P(d)} = a\left(3.37325 \times 10^{-6}\right)$$

όπου α είναι η σταθερά κανονικοποίησης που ισούται με 1/P(d). Αν εξαλείψουμε το α έχουμε

$$P(gp_1) = \frac{3.6075 \times 10^{-6}}{3.6075 \times 10^{-6} + 3.37325 \times 10^{-6}} = 0.51678$$

και

$$P(gp_2) = \frac{3.37325 \times 10^{-6}}{3.6075 \times 10^{-6} + 3.37325 \times 10^{-6}} = 0.48322$$

Επιλέγουμε το μοτίβο του κατευθυνόμενου ακυκλικού γράφου gp_1 και καταλήγουμε στο συμπέρασμα ότι η συσχέτιση της φοίτησης στο κολέγιο και του διαζυγίου είναι πιο πιθανή. Επιπλέον, ένα Μπεϋζιανό Δίκτυο , του οποίου το κατευθυνόμενο ακυκλικό γράφημα ανήκει στην κλάση ισοδυναμίας που αντιπροσωπεύεται από το gp_1, για την εξαγωγή συμπερασμάτων που αφορούν τα Y_1 και Y_2 θα μπορούσε να αναπτυχθεί.

3.2 Εκμάθηση παραμέτρων

Αφού μάθουμε τη δομή των Μπεϋζιανών Δίκτυών από τα δεδομένα, τότε το πρόβλημα της εκτίμησης και της ενημέρωσης των παραμέτρων των συνολικών κατανομών είναι πολύ απλούστερο από την αποσύνθεση σε τοπικές κατανομές, δεδομένου ότι οι τοπικές κατανομές στην πράξη περιλαμβάνουν μόνο μικρό αριθμό κόμβων. Επιπλέον, η διάστασή τους συνήθως δεν κλιμακώνεται με τον αριθμό των κόμβων που έχουμε στο Μπεϋζιανό Δίκτυο και συχνά θεωρείται ότι περιορίζεται από μια σταθερά όταν υπολογίζουμε την υπολογιστική πολυπλοκότητα των αλγορίθμων, αποφεύγοντας έτσι το λεγόμενο της κατάρας της διάστασης. Αυτό σημαίνει ότι κάθε τοπική κατανομή έχει ένα συγκριτικά μικρό αριθμό παραμέτρων που πρέπει να εκτιμηθούν από το δείγμα και ότι οι εκτιμήσεις είναι ακριβέστερες λόγω της καλύτερης αναλογίας μεταξύ του μεγέθους των Θ_{X_i} και του μεγέθους του δείγματος.

Η εκτίμηση μέγιστης πιθανοφάνειας και η εκτίμηση κατά Bayes είναι δύο προσεγγίσεις αρκετά συνήθεις στη βιβλιογραφία. Υπάρχουν και άλλες επιλογές οι οποίες είναι ασφαλώς δυνατές, όπως ακριβώς και οι εκτιμητές συρρίκνωσης που παρουσιάζονται στους Shafer και Strimmer (2005) και Hausser και Strimmer (2009). Είναι σημαντικό να σημειωθεί ότι η προσέγγιση που χρησιμοποιούμε για την εκμάθηση δομής των Μπεϋζιανών Δίκτυών καθορίζει σίγουρα τις προσεγγίσεις που μπορούμε να χρησιμοποιήσουμε στην εκμάθηση παραμέτρων. Για παράδειγμα, η ερμηνεία και η εξαγωγή συμπερασμάτων για το Μπεϋζιανό Δίκτυο είναι απλή, αν χρησιμοποιήσουμε τις εκ των υστέρων πυκνότητες τόσο στην εκμάθηση παραμέτρων όσο και στην εκμάθηση δομών. Παρόλο που, χρησιμοποιούνται τα τεστ μετατροπής MonteCarlo για τη μάθηση δομής και των εκ των υστέρων εκτιμήσεων για τη μάθηση παραμέτρων όπως παραδείγματος χάρη οι προσεγγίσεις συρρίκνωσης, χρησιμοποιούνται, επίσης και οι εκτιμήσεις παραμέτρων μέγιστης πιθανοφάνειας.

Παρόλο που στην πράξη μόνο ένας μικρός αριθμός μεταβλητών εμπλέκεται στις τοπικές κατανομές και η διάστασή τους δεν κλιμακώνεται συνήθως σε συνδυασμό με το μέγεθος του Μπεϋζιανού Δίκτυού, οι παράμετροι εκτίμησης σε ορισμένες περιπτώσεις εξακολουθούν να είναι προβληματικές Για παράδειγμα, είναι πολύ συνηθισμένο να έχουμε μεγέθη δειγμάτων πολύ μικρότερα από τον αριθμό των μεταβλητών που περιλαμβάνονται στο μοντέλο. Αυτό είναι σύνηθες σε σύνολα βιολογικών δεδομένων υψηλής απόδοσης, όπως οι μικροσυστοιχίες που έχουν μερικές δεκάδες ή εκατοντάδες παρατηρήσεις και χιλιάδες γονίδια. Στην προαναφερθείσα ρύθμιση που ονομάζεται "μικρό n, μεγάλο p" οι εκτιμήσεις έχουν μεγάλη μεταβλητότητα, εκτός εάν δίνουμε ιδιαίτερη προσοχή τόσο στην εκμάθηση δομής όσο και στην εκμάθηση παραμέτρων.

3.2.1 Μάθηση από πλήρη δεδομένα

Εάν σε κάθε περίπτωση όλες οι μεταβλητές είναι πλήρως παρατηρηθέντες, έτσι ώστε να μην έχουμε ελλειπή δεδομένα και κρυφές μεταβλητές, τότε τα δεδομένα είναι πλήρη. Για ένα κατευθυνόμενο γραφικό μοντέλο με πλήρη δεδομένα, η πιθανότητα δίνεται ως εξής

$$p(D|\theta)=\prod_{i=1}^{M} p(x_i|\theta)=\prod_{i=1}^{M}\prod_{t=1}^{N} p(x_{it}|x_i, pa(t),\theta_t)=p(D_t|\theta_t)$$

όπου D_t είναι τα δεδομένα που συνδέονται με τον κόμβο t και τους γονείς του, που σημαίνει ότι ο t'_{th} οικογένεια. Πρόκειται για ένα γινόμενο όρων, έναν για κάθε ConditionalProbabilityDistribution. Έτσι η πιθανοφάνεια αποσυντίθεται ανάλογα με τη δομή του γράφου. Τώρα υποθέστε ότι η προηγούμενηπαραγοντοποιείται επίσης:

$$p(\theta)=\prod_{t=1} p(\theta_t)$$

Τότε είναι σαφές ότι το posterior επίσης παραγοντοποιείται:

$$p(\theta|D)\propto p(D|\theta)p(\theta)=\prod_{t=1}^{N} p(D_t|\theta_t)p(\theta_t)$$

Έτσι, μπορούμε να υπολογίσουμε την εκ των υστέρων κατανομή κάθε κατανομής υπό συνθήκη πιθανότητας ανεξάρτητα. Για να το θέσουμε με έναν άλλο τρόπο, η παραγοντική εκ των προτέρων συν την παραγοντική πιθανότητα συνεπάγεται παραγοντική posterior. Σκεφτείτε ένα παράδειγμα, όπου όλες οι Κατανομές υπό συνθήκη πιθανοτήτων είναι πινακοποιημένες. Υπάρχει μια ξεχωριστή γραμμή (δηλ. μια ξεχωριστή κατανομή multinoulli) για υπο συνθήκη κατανομή, δηλαδή για κάθε συνδυασμό τιμών των γονέων. Τυπικά, μπορούμε να γράψουμε την t'_{th} ConditionalProbabilityTable $x_t \mid x_{pa(t)} = c \sim Cat(\theta_{tc})$,όπου $\theta_{tck} = p\left(x_t = k \mid x_{pa(t)} = c\right)$ για k = 1 : K_t, c = 1 : C_t και t = 1 : T. Εδώ K_tείναι ο αριθμός των καταστάσεων για κόμβο t, $C_t = \Pi_{s \in pa(t)} K_s$ είναι ο αριθμός των συνδυασμών γονέων και T είναι ο αριθμός των κόμβων. Είναι προφανές ότι χρησιμοποιειται ο $\Sigma_k \theta_{tck} = 1$για κάθε γραμμή κάθε πίνακα πιθανοτήτων υπό συνθήκη.

Τώρα ας βάλουμε μια ξεχωριστή προτεραιότητα την Dirichlet σε κάθε γραμμή κάθε CPT, δηλαδή, $\theta_{tc} \sim Dir(atc)$. Στη συνέχεια, η εκ των υστέρων μπορεί να υπολογιστεί προσθέτοντας απλώς τις ψευδομετρήσεις στις εμπειρικές μετρήσεις για να προκύψει $\theta_{tc} \mid D \sim Dir\left(M_{tck} + a_{tc}\right)$, όπου M_{tck} είναι ο αριθμός των φορών που ο κόμβος t είναι στην κατάσταση k ενώ οι γονείς του βρίσκονται στην κατάσταση c: $M_{tck} = \sum_{i=1}^{M} I\left(x_{i,t} = k, x_{i,pa(t)} = c\right)$ Ο μέσος όρος αυτής της κατανομής δίνεται στη συνέχεια:

$$\bar{\theta}_{tck} = \frac{M_{tck} + a_{tck}}{\sum_{k'}\left(M_{tck'}\right) + a_{tck'}} \tag{3.11}$$

Για παράδειγμα, θεωρήστε το Κατευθυνόμενο Γραφικό Μοντέλο στο Σχήμα 1.1(α). Υποθέτουμε ότι τα δεδομένα εκπαίδευσης αποτελούνται από τις 5 περιπτώσεις που δίνονται παρακάτω:

Πινακάς 3.3: Δεδομένα εκπαίδευσης

x_1	x_2	x_3	x_4	x_5
0	0	1	0	0
0	1	1	1	1
1	1	0	1	0
0	1	1	0	0
0	1	1	1	0

Παραθέτουμε όλες τις επαρκείς στατιστικές M_{tck} και τις μεταγενέστερες μέσες παραμέτρους $\overline{\theta_{ick}}$ σύμφωνα με μια εκ των προτέρων Dirichlet με a_{ick} (που αντιστοιχεί σε εξομάλυνση προσθήκης ενός) για τον κόμβο t = 4:

Πινακας 3.4: Επαρκείς στατιστικές M_{tck} και τις μεταγενέστερες μέσες παραμέτρους $\overline{\theta_{ick}}$ σύμφωνα με μια εκ των προτέρων Dirichlet με a_{ick} για τον κόμβο t = 4

x_2	x_3	$N_{tck=1}$	$N_{tck=0}$	$\overline{\theta_{tck=1}}$	$\overline{\theta_{tck=0}}$
0	0	0	0	1/2	1/2
1	0	1	0	2/3	1/3
0	1	0	1	1/3	2/3
1	1	2	1	3/5	2/5

Μπορεί εύκολα να αποδειχθεί ότι η ΜΛΕ έχει την ίδια μορφή με την εξίσωση , μόνο που δεν έχει τους όρους a_{tck} , δηλ.

$$\theta_{tck} = \frac{M_{tck}}{\sum_{k'}\left(M_{tck}\right)}$$

Ενώ τα προηγούμενα ισχύουν όταν έχουμε πλήρη δεδομένα, αυτές οι τεχνικές δεν είναι διαθέσιμες αν έχουμε ελλιπή δεδομένα ή/και λανθάνουσες/αποκρυμμένες μεταβλητές αφού η πιθανότητα πλέον δεν παραγοντοποιείται, και μάλιστα δεν είναι πλέον κυρτή. Αυτό σημαίνει λοιπόν ότι συνήθως μπορούμε μόνο να υπολογίσουμε μια τοπικά βέλτιστη εκτίμηση ML ή MAP. Επιπλέον, η Μπεϋζιανή συμπερασματολογία των παραμέτρων είναι ακόμη πιο δύσκολη.

3.2.2 Πολυωνυμικές μεταβλητές

Στην αρχή συζητάμε μια μέθοδο για την εκμάθηση μεμονωμένων παραμέτρων στην πολυωνυμική μεταβλητή και στη συνέχεια συζητάμε τη συνάρτηση πυκνότητας Dirichlet. Στη συνέχεια παρουσιάζουμε ένα τρόπο υπολογισμού διαστημάτων πιθανότητας και περιοχών και στο τέλος μια μέθοδο για την εκμάθηση όλων των παραμέτρων σε ένα ΜπευζιανόΔικτυό στην περίπτωση των πολυωνυμικών μεταβλητών.

Ορισμός 10. Η συνάρτηση πυκνότητας Dirichlet που έχει τις παραμέτρους $a_1, a_2, ..., a_r$,

$$N = \sum_{k=1}^{r} a_k \textit{ όπου } a_1, a_2, ..., a_r \textit{ είναι ακέραιοι αριθμοί} \geq 1, \textit{ είναι}$$

$$\rho\left(f_1, f_2, ..., f_{r-1}\right) = \frac{\Gamma(N)}{\prod_{k=1}^{r}\Gamma(a_k)} f_1^{a_1-1} f_2^{a_2-1} ... f_r^{a_r-1} \, 0 \leq f_k \leq 1, \sum_{k=1}^{r} f_k = 1.$$

Τυχαίες μεταβλητές $F_1, F_2, ..., F_r$, οι οποίες έχουν αυτή τη συνάρτηση πυκνότητας, λέγεται ότι έχουν την κατανομή Dirichlet. Συμβολίζουμε τη συνάρτηση πυκνότητας Dirichlet ως

$Dir\left(f_1, f_2, ..., f_{r-1}; a_1, a_2, ..., a_r\right)$

Αξίζει να σημειωθεί ότι οι τιμές των πρώτων r - 1 μεταβλητών (δηλαδή $f_r = 1 - \sum_{h=1}^{r-1} f_h$)

καθορίζουν μονοσήμαντα το F_r. Αυτός είναι ο λόγος για τον οποίο το ρ είναι απλώς μια συνάρτηση των r - 1 μεταβλητών. Ως γνωστόν η συνάρτηση πυκνότητας Dirichlet είναι μια γενίκευση της πυκνότητας βήτα συνάρτησης. Στο σχήμα 3.6 παρουσιάζεται μια συνάρτηση πυκνότητας Dirichlet. Επιπλέον, υπάρχουν πειστικά επιχειρήματα για τη χρήση της κατανομής Dirichlet για τη μοντελοποίηση των πεποιθήσεών μας σχετικά με σχετικές συχνότητες. Λέμε συχνά ότι η εμπειρία του εκτιμητή πιθανοτήτων είναι ίση με την k-οστή τιμή να εμφανίζεται a_k φορές σε M δοκιμές. Χρειάζεται το ακόλουθο λήμμα σχετικά με τη συνάρτηση πυκνότητας Dirichlet:

Λήμμα 1: Αν $F_1, F_2, ..., F_r$ αποτελούν από την Dirichlet κατανομής με παραμέτρους $a_1, a_2, ..., a_r, M = \sum a_k \textit{ για κάθε} 1 \leq k \leq r$

$$E\left(F_k\right) = \frac{a_k}{M}$$

Τώρα υποθέστε ότι έχουμε κάποια τυχαία διαδικασία r-εξόδου. Έχοντας ως Y μια τυχαία μεταβλητή στο χώρο της οποίας 1, 2, . . . r περιλαμβάνονται τα αποτελέσματα του πειράματος, και για $1 \leq k \leq r$ έστω F_k μια τυχαία μεταβλητή της οποίας ο χώρος είναι το διάστημα [0, 1]. Η πιθανότητα κατανομή της F_k αντιπροσωπεύει την πεποίθησή μας σχετικά με τη σχετική συχνότητα με την οποία Y = k. Ας υποθέσουμε ότι οι πεποιθήσεις μας είναι τέτοιες ώστε:

$$P\left(X=k \mid f_k\right)=f_k$$

Αυτό σημαίνει ότι αν γνωρίζαμε με βεβαιότητα ότι η σχετική συχνότητα του k θα ήταν f_k, η πεποίθησή μας σχετικά με την εμφάνιση του k στην πρώτη εκτέλεση του πειράματος θα ήταν f_k. Δεδομένης της προηγούμενης υπόθεσης, το θεώρημα που ακολουθεί συγκεντρώνει την υποκειμενική μας πιθανότητα για την πρώτη δοκιμή.

Θεώρημα 7: Έστω ότι η Y είναι μια τυχαία μεταβλητή της οποίας ο χώρος είναι 1, 2, ... r, και οι $F_1, F_2, ..., F_r$ είναι r τυχαίες μεταβλητές τέτοιες ώστε για όλα τα k,

$$P\left(Y=k \mid f_k\right)=f_k$$

Τότε

$$P\left(Y=k\right)=E\left(F_k\right)$$

Παραμετροποίηση διακριτών μοντέλων:

Στην παρουσίαση των μεθόδων παραμετροποίησης διακριτών δικτύων εξετάζουμε πρώτα την παραμετροποίηση διωνυμικών μοντέλων, τα οποία είναι μοντέλα μόνο με δυαδικές μεταβλητές και στη συνέχεια γενικεύουμε τη μέθοδο για την παραμετροποίηση αυθαίρετων διακριτών μοντέλων.

Παραμετροποίηση ενός διωνυμικού μοντέλου:

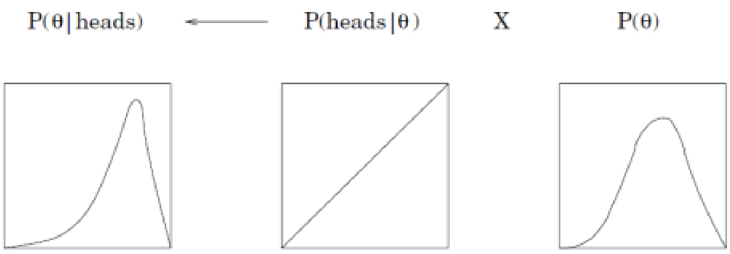
Το πιο απλό δυνατό Μπεϋζιανό Δίκτυο είναι αυτό μιας μοναδικής διωνυμικής μεταβλητής Y. Έστω ότι αν ρίξουμε ένα νόμισμα η Y αναφέρει το αποτέλεσμά του, λαμβάνοντας τις τιμές κορώνα και γράμματα. Θα θέλουμε να μάθουμε την τιμή της παραμέτρου Θ = θ που είναι η πιθανότητα P(X = heads). Ας πούμε ότι μαθαίνουμε ακριβώς και μόνο ότι την επόμενη φορά που θα ρίξουμε το νόμισμα το αποτέλεσμα θα είναι κορώνα. Τότε από το θεώρημα του Bayes έχουμε:

$$P\left(\theta \mid heads\right)=\beta P\left(heads \mid \theta\right)P\left(\theta\right)$$

$P\left(heads \mid \Theta=\theta\right)=\theta, έτσι$

$$P\left(\theta \mid heads\right)=\beta\theta P\left(\theta\right)$$

ο οποίος πολλαπλασιασμός παρουσιάζεται γραφικά στο σχήμα 3.6. Φυσικά, η παρατήρηση των κεφαλών στρεβλώνει την εκ των υστέρων κατανομή για το Θ = θ προς τα δεξιά, ενώ οι ουρές θα τις στρέψουν προς τα αριστερά.



Σχήμα 3.6: (Korbetal. (2010)) Ενημέρωση μιας διωνυμικής εκτίμησης σε μια οπτική απόδοση του Μπευζιανού Θεωρήματος.

Αν ρίξουμε το νόμισμα δύο φορές και το αποτέλεσμα είναι κορώνα και στις δύο ρίψεις e = ⟨κορώνα, κορώνα⟩(με e αναπαριστούμε τα αποδεικτικά μας στοιχεία), τότε η Μπεϋζιανή ενημέρωση αποδίδει:

$$P\left(\theta \mid e\right)=\beta\theta^{2}P\left(\theta\right)$$

Γενικά, η απόδειξη e αποτελείται από m κεφαλές και n - m γράμματα, γεγονός που μας δίνει:

$$P\left(\theta \mid e\right)=\beta\theta^{m}\left(1-\theta\right)^{n-m}P\left(\theta\right)$$

με την υπόθεση ότι όλες οι ρίψεις κερμάτων είναι ανεξάρτητα πανομοιότυπα κατανεμημένες (iid), έτσι ώστε κάθε ρίψη να είναι ανεξάρτητη από κάθε άλλη ρίψη, και κάθε ρίψη είναι ένα δείγμα που αντλείται από την ίδια κατανομή πιθανότητας με κάθε άλλη ρίψη.

Κλείνοτας αυτό το κεφαλαιό είδαμε πως γίνεται η Μαθηση Μεϋζιανων Δικτύων μεσω της Δομής μαθησής που είναι ουσιαστικα η διαδικασια εκμαθησης των κομβων και την Εκμαθησης παραμέτρων που είναι η διαδικασια εκμαθησης ολων των δυνατων τιμων που υπάρχουν στους κομβούς.

ΚΕΦΑΛΑΙΟ 4

Εφαρμογές

Έχουμε δει τη χρησιμότητα των Μπεϋζιανων Δικτύων σε γενικές καταστάσεις. Βοηθούν στην απάντηση ερωτήσεων με ελλείπη δεδομένα γρήγορα και αποτελεσματικά. Τώρα, θα δούμε πώς τα Μπεϋζιανα Δικτύα είναι χρήσιμα σε κάποια συγκεκριμένα προβλήματα που υπάρχουν στην καθημερινότητα μας.

4.1 Ιατρική Διάγνωση

Στην ιατρική διάγνωση, το δίκτυο Bayesian μπορεί να χρησιμοποιηθεί για να προτείνει πιθανές ασθένειες βασιζόμενο στα παρατηρούμενα συμπτώματα. Δηλαδή, ένας γιατρός μπορεί να εισαγάγει τα παρατηρούμενα συμπτώματα στο πρόγραμμα και το πρόγραμμα στη συνέχεια παίρνει τις εισροές και υπολογίζει τις πιθανότητες μιας ποικιλίας ασθενειών, δεδομένων των συμπτωμάτων. Οι γιατροί προσομοιώνουν κατά προσέγγιση τις πιθανότητες υπό όρους νοερά και ακολουθούν αυτή τη διαδικασία στην καθημερινή τους εργασία, ώστε να μπορούμε να κάνουμε αυτό το έργο αυτοματοποιημένο μέσω της χρήσης δικτύων Bayes. Εδώ, έχουμε ένα παράδειγμα συστήματος όπου οι αρχικές κατανομές πιθανοτήτων για τους κόμβους δεν χρειάζονται κανενός είδους μάθηση. Μπορούν να γίνουν χρησιμοποιώντας τις ποσοτικές εκτιμήσεις ενός επαγγελματία ιατρού ή χρησιμοποιώντας διαθέσιμα στατιστικά στοιχεία για τις 16 σχέσεις μεταξύ των συμπτωμάτων και των ασθενειών (Stephenson, 2000,[12]).

Η πρόκληση προκύπτει όταν αυτές οι ποσοτικές εκτιμήσεις και οι στατιστικές είναι πλήρεις ή όχι απόλυτα ακριβείς. Τα δίκτυα Bayes μπορούν επίσης να βοηθήσουν σε αυτή την περίπτωση παρέχοντας εκτιμήσεις κατά την αντιμετώπιση των ελλιπών και ανακριβών δεδομένων.

Η πρόκληση έγκειται στην κατασκευή πινάκων πιθανοτήτων υπό όρους από μερικά στατιστικά δεδομένα για τους κόμβους και τις ακμές του δικτύου. Επιπλέον, το πρόβλημα είναι ότι συχνά τα δεδομένα που χρειαζόμαστε δεν είναι στη μορφή που είναι πιο χρήσιμα για το μοντέλο μας. Για παράδειγμα, εάν έχουμε δύο ασθένειες D_1

και D_2, που και οι δύο μπορεί να προκαλέσουν σύμπτωμα S, μπορεί να λάβουμε στατιστικές πληροφορίες για το $P(S|D_1)$ και το $P(S|D_2)$, αλλά αυτές δεν είναι τόσο χρήσιμες όσο οι $P(S|D_1,D_2), P(D_1|S), και P(D_2|S)$. Η γενίκευση αυτού σε κάποιο αριθμό n ασθενειών κάνει το πρόβλημα πολύ πιο σκληρό. Εάν θεωρήσουμε κάθε μεταβλητή ασθένειας D_i ως δυαδική μεταβλητή, θα είχαμε 2n εγγραφές στον πίνακα πιθανοτήτων υπό όρους, αλλά η συμπλήρωσή της μπορεί να είναι εξαιρετικά κουραστική και υπολογιστικά δύσκολη. Για να ξεπεραστεί αυτό, θα ήταν βοηθητικό να συνδέσουμε το πρόβλημα σε ένα δίκτυο Bayes και να εισαγάγουμε πρόσθετους περιορισμούς με τη μορφή προκαταλήψεων υπέρ ή κατά ορισμένων ασθενειών. Αυτές οι προκαταλήψεις εξαρτώνται από το βαθμό στον οποίο οι μεμονωμένες ασθένειες επηρεάζουν ή δεν επηρεάζουν το σύμπτωμα ή το σύνολο των εν λόγω συμπτωμάτων. (Nikovski, 2000,[14]).

Αυτές οι προκαταλήψεις ονομάζονται επίσης ευαισθησίες και κωδικοποιούν την πιθανότητα $P(S|D)\ και\ P(S'|D')$ σε κάθε κόμβο ασθένειας. Έτσι, μειώσαμε το πρόβλημα από συντελεστή 2^n σε $2n$. Τώρα, μπορούμε να χρησιμοποιήσουμε την αξιολόγηση του δικτύου Bayes για να προσδιορίσουμε ακριβώς πόσο πιθανή είναι μια ασθένεια δεδομένου των ευαισθησιών και των σχέσεων. (Nikovski, 2000) Ένα πρόσθετο όφελος από τη χρήση των πληροφοριών ευαισθησίας είναι ότι κάθε ασθένεια είναι ανεξάρτητη από άλλες ασθένειες του μοντέλου. Αντί να αποκτήσουμε την πιο πιθανή ασθένεια διερευνώντας 2^n πιθανότητες, μπορούμε να εξετάσουμε την πιθανότητα κάθε ασθένειας ανεξάρτητα από άλλες. Ουσιαστικά, εξετάζουμε ένα σύστημα ORs τέτοιο ώστε όταν εξετάζουμε το $P(S|D_i)$, ουσιαστικά εξετάζουμε το $P\left(S|D_1',D_2',...,D_i',...,D_n'\right)$ για κάθε τιμή $i=1,...,n$. (Nikovski, 2000)

Επιστρέφοντας στο παραπάνω παράδειγμα των ασθενειών D_1 και D_2 που προκαλούν το σύμπτωμα S, μπορούμε να δούμε την ενδιαφέρουσα ερώτηση που σκεφτήκαμε, η οποία αφορά την πιθανότητα κάθε ασθένειας δεδομένων των παρατηρούμενων συμπτωμάτων. Δεδομένων των κατανομών πιθανότητας των $P(S|D_1)\ και\ P(S|D_2)$, μπορούμε να χρησιμοποιήσουμε τον Κανόνα του Bayes για τον υπολογισμό των τιμών των $P(D_1|S)\ και\ P(D_2|S)$. Αυτό είναι πιο χρήσιμο για τον γιατρό, καθώς τώρα ο γιατρός μπορεί να αναγνωρίσει την ασθένεια με έναν βαθμό εμπιστοσύνης και να θεραπεύσει τη νόσο. Αυτό που πρέπει να λάβουμε υπόψη, είναι ότι εφόσον πρόκειται για ένα μοντέλο που κατασκευάστηκε από τον άνθρωπο και περιορίζεται από τις γνώσεις εκείνων που δημιούργησαν το μοντέλο, είναι πιθανό να υπάρχουν ασθένειες που δεν υπολογίζονται και οι οποίες θα μπορούσαν να προκαλέσουν αυτά τα συμπτώματα. Επιπλέον, δεδομένου ότι αυτή είναι μια πιθανολογική προσέγγιση, είναι πιθανό να υπάρχουν ορισμένα ψευδώς θετικά και ψευδώς αρνητικά αποτελέσματα. Επιπλέον, ορισμένα αποτελέσματα μπορεί να είναι πολύ διφορούμενα για να τα κατανοήσουμε. Για παράδειγμα, ένα αποτέλεσμα 50% πιθανότητας μιας δεδομένης ασθένειας μπορεί να είναι δύσκολο να ερμηνευθεί. Σε μια τέτοια περίπτωση μπορεί να χρειαστεί να λάβουν περισσότερες πληροφορίες μέσω πρόσθετων δοκιμών.

Όλες οι ασθένειες δεν είναι τόσο δυαδικές όσο μόλις περιγράψαμε. Επιπλέον, κάθε ασθένεια μπορεί να προκαλέσει πολλαπλά συμπτώματα και ορισμένα συμπτώματα μπορούν να αλλάξουν την πιθανότητα άλλων συμπτωμάτων. Ακόμη και η μοντελοποίηση τέτοιων σχέσεων είναι εύκολη υπόθεση για ένα δίκτυο Bayes. Λάβετε υπόψη ότι στην ακόλουθη κατάσταση, δύο από τα συμπτώματα της νόσου CardiacTamponade είναι η δύσπνοια και η γρήγορη αναπνοή. Ενώ και τα δύο συμπτώματα έχουν τις δικές τους πιθανότητες δεδομένης της παρουσίας της νόσου, η παρουσία ταχείας αναπνοής γίνεται πιο πιθανή εάν υπάρχει δύσπνοια. Σε μια τέτοια περίπτωση, τα δύο συμπτώματα δεν είναι υπό όρους ανεξάρτητα μεταξύ τους. Το ένα αυξάνει την πιθανότητα του άλλου και έτσι το δίκτυό μας μπορεί να το εξηγήσει προσθέτοντας μια άκρη στο δίκτυο από το αιτιολογικό σύμπτωμα στο προκαλούμενο. (Παράδειγμα που υιοθετήθηκε από τον Nikovski (2000)) .

Στην περίπτωση μη μοντελοποιημένων ασθενειών, ξεπερνάμε το πρόβλημα συμπεριλαμβάνοντας μια μεταβλητή διαρροής που αντιπροσωπεύει όλες τις πιθανές μη καταγεγραμμένες ασθένειες ή γενικά τις μη καταγεγραμμένες μεταβλητές. (Nikovski, 2000). Επιπλέον, όλες οι ασθένειες θα πρέπει να επαληθεύονται χρησιμοποιώντας πιο πειστικές δοκιμές αφού οι επιπτώσεις από λανθασμένη θεραπεία μπορεί να είναι επιζήμιες. Ωστόσο, τουλάχιστον χρησιμοποιώντας αυτό το μοντέλο μπορούμε να περιορίσουμε ορισμένες από τις πιθανότητες και να ξέρουμε πού να εστιάσουμε την προσοχή μας.

4.1.1 Παράδειγμα στην R

Θα δημιουργήσουμε ένα τυχαίο παράδειγμα το οποίο θα είναι από τη δημιουργία ενός τυχαίου γραφήματος . Για να το κάνουμε αυτό χρησιμοποιήσαμε τη συνάρτηση set.seed για να είμαστε σίγουροι ότι κάθε φορά που χρησιμοποιούμε τον κώδικα R θα χρησιμοποιούνται τα ίδια δεδομένα. Το τυχαίο γράφημά μας περιέχει 6 μεταβλητές που συμβολίζονται με τους όρους Symptom1, Symptom2, Symptom3, Symptom4, Symptom5 και Diagnosis. Κάθε μία από αυτές τις μεταβλητές έχει δύο επίπεδα, το ΝΑΙ και το ΟΧΙ και το Diagnosis ΘΕΤΙΚΟ και ΑΡΝΗΤΙΚΟ με πιθανότητες που ορίζονται αυθαίρετα. Ορίσαμε τους πίνακες υπό συνθήκη πιθανοτήτων για κάθε μεταβλητή και μετά από αυτό δημιουργήσαμε το Μπεϋζιανό μας Δίκτυο και επίσης ελέγξαμε για κύκλους στο ΒΝ μας. Μετά από αυτό προσομοιώσαμε ένα δείγμα 2000 παρατηρήσεων για να συγκρίνουμε τα αποτελέσματα που θα προκύψουν για την αρχική δομή του ΒΝ μας και κάποιες δομές που θα δημιουργηθούν για τις ανάγκες του παραδείγματος.

Αρχικά, αφαιρέσαμε ένα τόξο από την αρχική δομή (net) και δημιουργήσαμε τη δομή του πυγμάχου (new.net). Στη συνέχεια, τοποθετήσαμε ένα νέο τόξο και αφαιρέσαμε 1 τόξο στην αρχική δομή για να για να δημιουργήσουμε τη δεύτερη δομή (new.net1). Για να δημιουργήσουμε την τρίτη δομή αφαιρέσαμε 2 τόξα και προσθέσαμε ένα νέο (new.net2). Μετά τη δημιουργία των τριών νέων δομών λαμβάνουμε την βαθμολογίες για τις τέσσερις δομές που είχαμε. Εξετάζουμε τέσσερις διαφορετικούς τύπους βαθμολογιών οι οποίοι είναι οι εξής: η βαθμολογία Bayesian Information (bic), η βαθμολογία Akaike Information (aic), η πολυωνυμική βαθμολογία λογαριθμικής πιθανότητας (loglik) και ο λογάριθμος της ισοδύναμης Bayesian Dirichlet. (ομοιόμορφη) βαθμολογία (bde).

```
> library(bnlearn)
>set.seed(20021)
> # Create a sample dataset
> data <- data.frame(
+   Symptom1 = as.factor(sample(c("Yes", "No"), 100, replace =
TRUE)),
+   Symptom2 = as.factor(sample(c("Yes", "No"), 100, replace =
TRUE)),
```

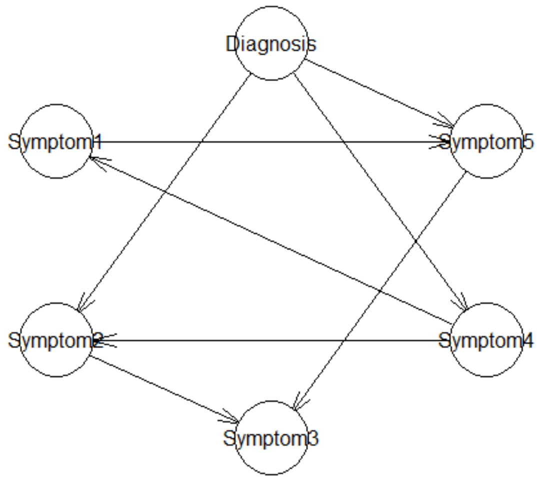
```
+   Symptom3 = as.factor(sample(c("Yes", "No"), 100, replace =
TRUE)),
+   Symptom4 = as.factor(sample(c("Yes", "No"), 100, replace =
TRUE)),
+   Symptom5 = as.factor(sample(c("Yes", "No"), 100, replace =
TRUE)),
+   Diagnosis = as.factor(sample(c("Positive", "Negative"), 100,
replace = TRUE))
+ )
>randomgraph<-random.graph(c("Symptom1",
"Symptom2","Symptom3","Symptom4","Symptom5","Diagnosis"),method="mela
ncon",max.degree=3)
>randomgraph
```

Χρησιμοποιώντας τον αλγόριθμο Uniform Random Acyclic Digraphs των Melancon και Philippe έχουμε το ακόλουθο τυχαίο/παραγόμενο Μπεϋζιανό δίκτυο:

```
 Random/Generated Bayesian network

  model:

[Diagnosis][Symptom4|Diagnosis][Symptom1|Symptom4][Symptom2|Symptom4:
Diagnosis]
   [Symptom5|Symptom1:Diagnosis][Symptom3|Symptom2:Symptom5]
  nodes:                                 6
  arcs:                                  8
    undirected arcs:                     0
    directed arcs:                       8
  average markov blanket size:           3.33
  average neighbourhood size:            2.67
  average branching factor:              1.33

  generation algorithm:                  Melancon's Uniform
Probability DAGs
  burn in length:                        216
  maximum in-degree:                     Inf
  maximum out-degree:                    Inf
  maximum degree:                        3
```

Κάναμε μια γραφική παράσταση της δομής του Μπεϋζιανού Δικτύου μας για καλύτερη κατανόηση

```
>plot(randomgraph)
```



Σχήμα 4.1: Γράφημα του αρχικού τυχαίου/παραγόμενου Μπεϋζιανού Δικτύου

Στη συνέχεια θα παρουσιάσουμε τα δεδομένα μας σε ένα πίνακα όπως δείχνει το παραπάνω διάγραμμα

```
>matrix<-matrix(0L,ncol=6,nrow=6,dimnames=list(c("Symptom1",
"Symptom2","Symptom3","Symptom4","Symptom5",
"Diagnosis"),c("Symptom1",
"Symptom2","Symptom3","Symptom4","Symptom5", "Diagnosis")))
>matrix["Diagnosis","Symptom5"]=1L
>matrix["Diagnosis","Symptom4"]=1L
>matrix["Diagnosis","Symptom2"]=1L
>matrix["Symptom2","Symptom3"]=1L
>matrix["Symptom4","Symptom2"]=1L
>matrix["Symptom4","Symptom1"]=1L
>matrix["Symptom5","Symptom3"]=1L
>matrix["Symptom1","Symptom5"]=1L
>matrix
```

Πίνακας 4.1: Adjacency matrix με 0/1 ακέραια στοιχεία με το 1 να αντιστοιχεί στο τόξα.

	Symptom1	Symptom2	Symptom3	Symptom4	Symptom5	Diagnosis
Symptom1	0	0	0	0	1	0
Symptom2	0	0	1	0	0	0
Symptom3	0	0	0	0	0	0
Symptom4	1	1	0	0	0	0
Symptom5	0	0	1	0	0	0
Diagnosis	0	1	0	1	1	0

```
>amat(randomgraph)<-matrix
>randomgraph

  Random/Generated Bayesian network

  model:

[Diagnosis][Symptom4|Diagnosis][Symptom1|Symptom4][Symptom2|Symptom4:
Diagnosis]
   [Symptom5|Symptom1:Diagnosis][Symptom3|Symptom2:Symptom5]
  nodes:                                 6
  arcs:                                  8
    undirected arcs:                     0
    directed arcs:                       8
  average markov blanket size:           3.33
  average neighbourhood size:            2.67
  average branching factor:              1.33

  generation algorithm:                  Melancon's Uniform
Probability DAGs
  burn in length:                        216
  maximum in-degree:                     Inf
  maximum out-degree:                    Inf
  maximum degree:                        3
```

Ο πίνακας 4.1 είναι ένας άλλος τρόπος για να αναπαραστήσετε το Γράφημα στο Σχήμα 4.2 όπου τα 1s αντιστοιχούν στα τόξα του Γραφήματος.

```
>#Diagnosis
>cptD<-matrix(c(0.4,0.6),ncol=2,
+ dimnames=list(NULL,c("Diagnosis_Positive","Diagnosis_Negative")))
>
>#Symptom1
>cptS1<-c(0.6,0.4,0.4,0.6)
>dim(cptS1)<-c(2,2)
>dimnames(cptS1)<-
list("Symptom1"=c("Symptom1_Yes","Symptom1_No"),"Symptom4"=c("Symptom
4_Yes","Symptom4_No"))
>
>#Symptom4
>cptS4<-c(0.4,0.6,0.7,0.3)
>dim(cptS4)<-c(2,2)
>dimnames(cptS4)<-
list("Symptom4"=c("Symptom4_Yes","Symptom4_No"),"Diagnosis"=c("Diagno
sis_Positive","Diagnosis_Negative"))
>
>#Symptom2
>cptS2<-c(0.4,0.6,0.7,0.3,0.5,0.5,0.1,0.9)
>dim(cptS2)<-c(2,2,2)
```

```
>dimnames(cptS2)<-
list("Symptom2"=c("Symptom2_Yes","Symptom2_No"),"Diagnosis"=c("Diagno
sis_Positive","Diagnosis_Negative"),"Symptom4"=c("Symptom4_Yes","Symp
tom4_No"))
>
>#Symptom5
>cptS5<-c(0.5,0.5,0.1,0.9,0.4,0.6,0.7,0.3)
>dim(cptS5)<-c(2,2,2)
>dimnames(cptS5)<-
list("Symptom5"=c("Symptom5_Yes","Symptom5_No"),"Diagnosis"=c("Diagno
sis_Positive","Diagnosis_Negative"),"Symptom1"=c("Symptom1_Yes","Symp
tom1_No"))
>
>#Symptom3
>cptS3<-c(0.7,0.3,0.5,0.5,0.1,0.9,0.4,0.6)
>dim(cptS3)<-c(2,2,2)
>dimnames(cptS3)<-
list("Symptom3"=c("Symptom3_Yes","Symptom3_No"),"Symptom2"=c("Symptom
2_Yes","Symptom2_No"),"Symptom5"=c("Symptom5_Yes","Symptom5_No"))
>class(net)
[1] "bn"
>dfit<-
custom.fit(net,dist=list(Diagnosis=cptD,Symptom5=cptS5,Symptom4=cptS4
,Symptom2=cptS2,Symptom1=cptS1,Symptom3=cptS3))
>dfit

  Bayesian network parameters

  Parameters of node Symptom1 (multinomial distribution)

Conditional probability table:

              Symptom4
Symptom1       Symptom4_Yes Symptom4_No
  Symptom1_Yes          0.6         0.4
  Symptom1_No           0.4         0.6

  Parameters of node Symptom2 (multinomial distribution)

Conditional probability table:

, , Diagnosis = Diagnosis_Positive

              Symptom4
Symptom2       Symptom4_Yes Symptom4_No
  Symptom2_Yes          0.4         0.5
  Symptom2_No           0.6         0.5

, , Diagnosis = Diagnosis_Negative

              Symptom4
Symptom2       Symptom4_Yes Symptom4_No
  Symptom2_Yes          0.7         0.1
  Symptom2_No           0.3         0.9


  Parameters of node Symptom3 (multinomial distribution)

Conditional probability table:

, , Symptom5 = Symptom5_Yes
```

```
              Symptom2
Symptom3       Symptom2_Yes Symptom2_No
  Symptom3_Yes          0.7         0.5
  Symptom3_No           0.3         0.5

, , Symptom5 = Symptom5_No

              Symptom2
Symptom3       Symptom2_Yes Symptom2_No
  Symptom3_Yes          0.1         0.4
  Symptom3_No           0.9         0.6


  Parameters of node Symptom4 (multinomial distribution)

Conditional probability table:

              Diagnosis
Symptom4       Diagnosis_PositiveDiagnosis_Negative
  Symptom4_Yes                0.4                0.7
  Symptom4_No                 0.6                0.3

  Parameters of node Symptom5 (multinomial distribution)

Conditional probability table:

, , Diagnosis = Diagnosis_Positive

              Symptom1
Symptom5       Symptom1_Yes Symptom1_No
  Symptom5_Yes          0.5         0.4
  Symptom5_No           0.5         0.6

, , Diagnosis = Diagnosis_Negative

              Symptom1
Symptom5       Symptom1_Yes Symptom1_No
  Symptom5_Yes          0.1         0.7
  Symptom5_No           0.9         0.3


  Parameters of node Diagnosis (multinomial distribution)

Conditional probability table:

Diagnosis_PositiveDiagnosis_Negative
               0.4                0.6

>class(dfit)
[1] "bn.fit"      "bn.fit.dnet"
```

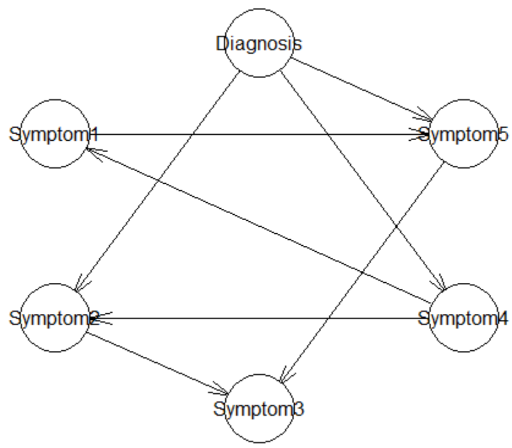
Με τον παραπάνω κώδικα ορίσαμε τους πίνακες πιθανοτήτων υπό όρους για κάθε μεταβλητή (κόμβο) και στη συνέχεια χρησιμοποιήσαμε τη συνάρτηση custom.fit για να λάβουμε τις παραμέτρους για κάθε κόμβο μας BayesianNetwork.

Με την εντολή που δίνεται στη συνέχεια, ελέγξαμε αν έχουμε κύκλους στο Bayesian Network. Στη συνέχεια προσομοιώσαμε n=2000 παρατηρήσεις από το παραπάνω δεδομένο Bayesian Network αλλά λόγω του μεγάλου αριθμού παρατηρήσεων που ζητήσαμε δημιουργήθηκε μόνο για τις πρώτες έξι σειρές του πλαισίου δεδομένων. Ας ρίξουμε μια πρώτη ματιά στα δεδομένα μας.

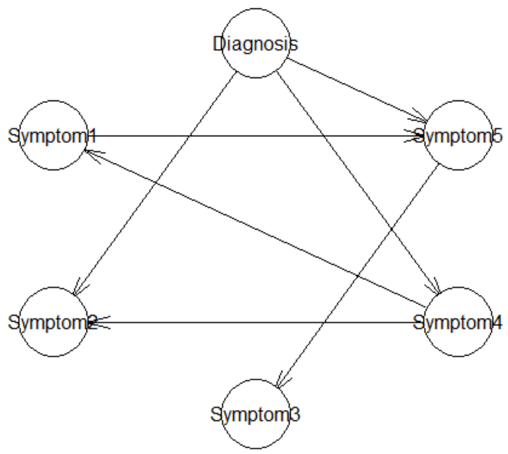
```
>rbn_df<-rbn(dfit,n=2000)
>head(rbn_df)
      Symptom1     Symptom2     Symptom3     Symptom4     Symptom5          Diagnosis
1 Symptom1_Yes  Symptom2_No  Symptom3_No  Symptom4_No Symptom5_Yes Diagnosis_Negative
2  Symptom1_No  Symptom2_No  Symptom3_No  Symptom4_No  Symptom5_No Diagnosis_Positive
3 Symptom1_Yes  Symptom2_No  Symptom3_No  Symptom4_No  Symptom5_No Diagnosis_Negative
4 Symptom1_Yes Symptom2_Yes  Symptom3_No Symptom4_Yes  Symptom5_No Diagnosis_Negative
5 Symptom1_Yes Symptom2_Yes  Symptom3_No  Symptom4_No  Symptom5_No Diagnosis_Positive
6  Symptom1_No  Symptom2_No  Symptom3_No  Symptom4_No Symptom5_Yes Diagnosis_Negative
```

Συνεχίζουμε με τη δημιουργία των τριών νέων δομών. Αρχικά αφαιρέσαμε ένα τόξο και η δομή που δημιουργήσαμε δίνεται παρακάτω.

```
># remove 1 arc
>new.net<-drop.arc(from='Symptom2',to="Symptom3",net)
>new.net

  Random/Generated Bayesian network

  model:

[Diagnosis][Symptom4|Diagnosis][Symptom1|Symptom4][Symptom2|Symptom4:
Diagnosis]
   [Symptom5|Symptom1:Diagnosis][Symptom3|Symptom5]
  nodes:                                  6
  arcs:                                   7
    undirected arcs:                      0
    directed arcs:                        7
  average markov blanket size:            2.67
  average neighbourhood size:             2.33
  average branching factor:               1.17

  generation algorithm:                   Melancon's Uniform
Probability DAGs
  burn in length:                         216
  maximum in-degree:                      Inf
  maximum out-degree:                     Inf
  maximum degree:                         3

>plot(new.net)
```



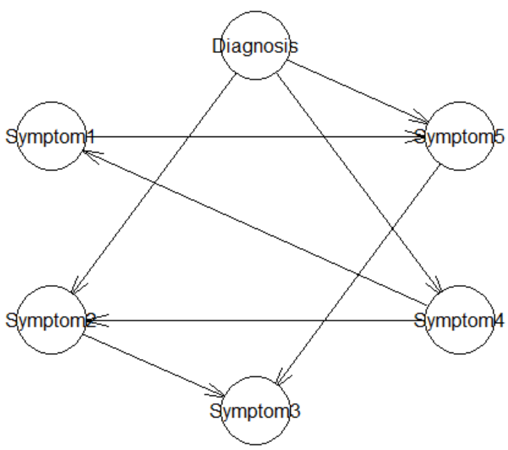
Σχήμα 4.2:Πρωτότυπο Γράφημα Μπεϋζιανού Δικτύου με Τυχαία/Δημιουργία.



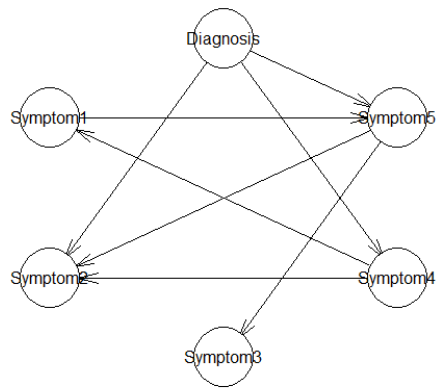
Σχήμα 4.3 : Το γράφημα του δικτύου λειτουργεί με το αφαιρούμενο τόξο.

Στα σχήματα 4.2 και 4.3 παρατηρούμε τη διαφορά μεταξύ των δύο γραφημάτων που παράγονται από τις δύο δομές. Πάνω, τοOriginalRandom/GeneratedBayesianNetwork Γράφημα και κάτω παρουσιάζεται το Γράφημα του Δικτύου με το αφαιρεμένο τόξο. Στη συνέχεια αφαιρέσαμε ένα τόξο και βάλαμε ένα νέο στην αρχική δομή για να δημιουργήσουμε το δεύτερο δομή όπως φαίνεται παρακάτω:

```
># remove 1 arc + put a random arc
>new.net1<-set.arc(from='Symptom5',to="Symptom2",new.net)
>new.net1

  Random/Generated Bayesian network

  model:

[Diagnosis][Symptom4|Diagnosis][Symptom1|Symptom4][Symptom5|Symptom1:
Diagnosis]
   [Symptom2|Symptom4:Symptom5:Diagnosis][Symptom3|Symptom5]
  nodes:                                 6
  arcs:                                  8
    undirected arcs:                     0
    directed arcs:                       8
  average markov blanket size:           3.33
  average neighbourhood size:            2.67
  average branching factor:              1.33

  generation algorithm:                  Melancon's Uniform
Probability DAGs
  burn in length:                        216
  maximum in-degree:                     Inf
  maximum out-degree:                    Inf
  maximum degree:                        3

>plot(new.net1)
```



Σχήμα 4.4: Πρωτότυπο Γράφημα Μπεϋζιανού Δικτύου με Τυχαία/Δημιουργία

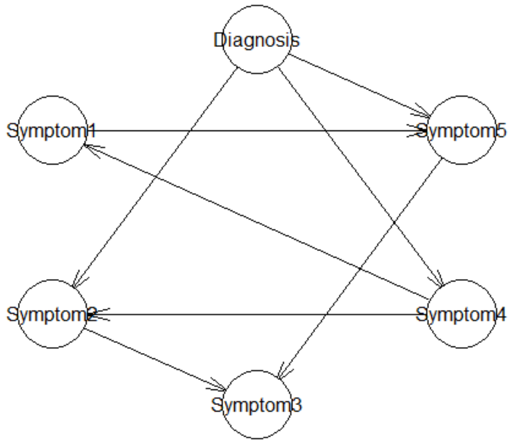


Σχήμα 4.5: Το Γράφημα του Δικτύου λειτουργεί με το αφαιρεμένο τόξο και το νέο προστιθέμενο τόξο.

Στα Σχήματα 4.4 και 4.5 παρατηρούμε τη διαφορά μεταξύ των δύο Γραφημάτων, δηλαδή του Πρωτότυπο Τυχαίο/Δημιουργημένο Γράφημα Δικτύου Bayes (πάνω) και το Γράφημα του Δικτύου με το αφαιρεμένο τόξο και το νέο (κάτω).

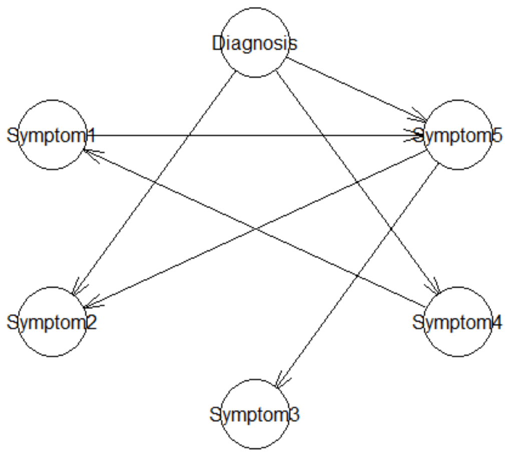
Τέλος αφαιρέσαμε δύο τόξα και βάλαμε ένα νέο για να δημιουργήσουμε την τρίτη δομή όπως φαίνεται παρακάτω:

```
># remove 2 arcs + put a random arc
>new.net2<-drop.arc(from='Symptom2',to="Symptom4",new.net1)
>new.net2

  Random/Generated Bayesian network

  model:

[Diagnosis][Symptom4|Diagnosis][Symptom1|Symptom4][Symptom5|Symptom1:
Diagnosis]
   [Symptom2|Symptom5:Diagnosis][Symptom3|Symptom5]
  nodes:                                 6
  arcs:                                  7
    undirected arcs:                     0
    directed arcs:                       7
  average markov blanket size:           2.67
  average neighbourhood size:            2.33
  average branching factor:              1.17

  generation algorithm:                  Melancon's Uniform
Probability DAGs
  burn in length:                        216
  maximum in-degree:                     Inf
  maximum out-degree:                    Inf
  maximum degree:                        3
```



>plot(new.net2)

Σχήμα 4.6: Πρωτότυπο Γράφημα Μπεϋζιανού Δικτύου με Τυχαία/Δημιουργία



Σχήμα 4.7: Το γράφημα του δικτύου λειτουργεί με 2 τόξα που αφαιρέθηκαν και 1 νέο τόξο

Εμφανίζονται οι διαφορές μεταξύ των δύο γραφημάτων, του αρχικού γραφήματος εργασίας τυχαίου/δημιουργημένου δικτύου Bayesian και του γραφήματος του δικτύου με τα αφαιρεμένα τόξα και του νέου στα Σχήματα 4.7 και 4.8 αντίστοιχα. Στο επόμενο βήμα υπολογίσαμε τις βαθμολογίες κάθε δικτύου για να τις συγκρίνουμε και να δούμε ποιο ταιριάζει καλύτερα στα δεδομένα.

```
#scores
score(net,rbn_df,type='bic')# BIC criterion
score(net,rbn_df,type='aic')# AIC criterion
score(net,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(net,rbn_df,type='loglik')# Log-Likelihood

score(new.net,rbn_df,type='bic')# BIC criterion
score(new.net,rbn_df,type='aic')# AIC criterion
score(new.net,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net,rbn_df,type='loglik')# Log-Likelihood

score(new.net1,rbn_df,type='bic')# BIC criterion
score(new.net1,rbn_df,type='aic')# AIC criterion
score(new.net1,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net1,rbn_df,type='loglik')# Log-Likelihood

score(new.net2,rbn_df,type='bic')# BIC criterion
score(new.net2,rbn_df,type='aic')# AIC criterion
score(new.net2,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net2,rbn_df,type='loglik')# Log-Likelihood
```

Πίνακας 4.2: Σκορ τροποποιήσεων BN

	BIC	AIC	BDE	LOGLIK
Πρωτότυπο	-7391.254	-7391.254	-7343.646	-7398.633
1η Τροποποίηση	-7461.535	-7419.528	-7467.616	-7404.528
2η Τροποποίηση	-7476.587	-7423.378	-7486.355	-7404.378
3η Τροποποίηση	-7688.722	-7646.715	-7695.419	-7631.715

Έτσι όπως βλέπουμε και τα 4 κριτήρια επιλέγουν την 3η Τροποποίηση. Η αφαίρεση των δύο τόξων, αφού οι πιθανότητές τους ήταν κοντά στο 50% δεν είχαν σημαντική επίδραση στις βαθμολογίες και ταυτόχρονα, η προσθήκη ενός επιπλέον βέλους οδήγησε τους αλγόριθμους να επιλέξουν λανθασμένα την 3η Τροποποίηση. Επομένως, αυστηρότερα κριτήρια που βασίζονται στη Θεωρία της Πληροφορίας μπορεί να είναι σε θέση να παρέχουν άλλη λύση σε αυτό το πρόβλημα, όπως η εφαρμογή της απόκλισης Kullback-Leibler για την επιλογή της καλύτερης δομής.

4.2 Ταξινόμηση ανεπιθύμητης αλληλογραφίας

Στον σύγχρονο κόσμο μας ασχολούμαστε με το ηλεκτρονικό ταχυδρομείο καθημερινά. Η ανεπιθύμητη αλληλογραφία (συχνά αποκαλούμενη ως spam) μας εκνευρίζει καθημερινά. Ωστόσο, οι περισσότεροι πελάτες ηλεκτρονικού ταχυδρομείου ξοδεύουν πολύ χρόνο και ενέργεια προσπαθώντας να βρουν αυτοματοποιημένους τρόπους για να ελαχιστοποιήσουν τον όγκο των ανεπιθύμητων μηνυμάτων που λαμβάνουν. Προκειμένου να γίνει αυτό, ελέγχουν τα μηνύματα ηλεκτρονικού ταχυδρομείου που λαμβάνουν για να προβλέψουν τι είναι ανεπιθύμητο και τι όχι. Ενώ υπάρχει ένας βαθμός λάθους στις προβλέψεις τους, οι ταξινομητές μπορούν να «σώσουν» από τους ανθρώπους πολύ χρόνο, εάν μπορούν να επιτύχουν έναν ικανοποιητικό βαθμό ακρίβειας στις προβλέψεις τους. Κατά τη διαδικασία ανίχνευσης ανεπιθύμητης αλληλογραφίας έναντι ανεπιθύμητων μηνυμάτων, θέλουμε να εξετάσουμε ορισμένες ιδιότητες του μεμονωμένου email και επίσης τις αλληλεπιδράσεις του ατόμου με τον αποστολέα και με παρόμοια email. Επιπλέον, θέλουμε να δούμε πώς το email ταιριάζει με άλλα μηνύματα ηλεκτρονικού ταχυδρομείου που λαμβάνουν οι άνθρωποι. Δηλαδή κοιτάμε το περιεχόμενο του email, το στυλ μορφοποίησης, λεπτομέρειες σχετικά με τον αποστολέα και ορισμένες άλλες ιδιότητες όπως παρατίθεται στο έργο των Sahamietal. και συζητείται περαιτέρω σε αυτή την ενότητα.

Οι Sahamietal. πίστευαν ότι στο πλαίσιο του φιλτραρίσματος email, είναι εξαιρετικά σημαντικό να αναπαραστήσετε μηνύματα ηλεκτρονικού ταχυδρομείου χρησιμοποιώντας διανύσματα χαρακτηριστικών έτσι ώστε να μπορούν οι ταξινομητές Bayes να χρησιμοποιηθούν απευθείας. Κάθε χαρακτηριστικό αντιπροσωπεύει διαφορετικές πληροφορίες σχετικά με τα email και έχει ξεχωριστή κατανομή πιθανοτήτων που περιγράφει την πιθανότητα ύπαρξης ανεπιθύμητου email. Σε μια τέτοια διανομή, είναι δυνατό να επιτρέπεται κάθε λέξη που υπάρχει στα email με το δικό της χαρακτηριστικό. Ενώ κανένα χαρακτηριστικό δεν είναι σαφές για το εάν ένα email είναι ανεπιθύμητο ή όχι, μπορούμε να αναπτύξουμε πιθανότητες για καθεμία από αυτές τις υποθέσεις και να λαμβάνουμε αποφάσεις με βάση αυτά. Για παράδειγμα, η παρουσία ορισμένων λέξεων μπορεί να αλλάξει την πιθανότητα ένα email να είναι ανεπιθύμητο. Η εμφάνιση ορισμένων λέξεων όπως **ΔΩΡΕΑΝ** και **ΧΡΗΜΑΤΑ** φαίνεται να υποδεικνύει ότι ένα email μπορεί να είναι ανεπιθύμητο.

Φράσεις όπως **ΚΛΙΚ ΕΔΩ** υποδεικνύουν μια απάτη ηλεκτρονικού ψαρέματος, ειδικά όταν ο σύνδεσμος σε αυτό δεν είναι επαληθεύσιμος για να είναι ασφαλής. Ωστόσο, ορισμένες λειτουργίες που κάνουν τα email λιγότερο πιθανό να είναι ανεπιθύμητα είναι λέξεις όπως το όνομα ατόμου. Εάν το όνομα του παραλήπτη είναι γραμμένο καθαρά, υπάρχει μειωμένη πιθανότητα το email να είναι spam.

Εάν το άτομο έλαβε το email απευθείας και όχι μαζί με μια λίστα αλληλογραφίας, αυτό είναι πιο πιθανό να είναι ασφαλής. Τα περισσότερα ανεπιθύμητα μηνύματα αποστέλλονται μέσω λιστών αλληλογραφίας. Αυτή η τάση αλλάζει σιγά σιγά δεδομένου του τρόπου με τον οποίο οι άνθρωποι του μάρκετινγκ κατανοούν αυτήν την τάση στα φίλτρα ανεπιθύμητης αλληλογραφίας και προσπαθούν να στείλουν email μεμονωμένα αυτοματοποιώντας τη διαδικασία.

Ένα άλλο χαρακτηριστικό που αξίζει να σημειωθεί είναι ο τομέας του αποστολέα. Σχεδόν ποτέ δεν είναι ανεπιθύμητη η αλληλογραφία που αποστέλλεται από ένα email .edu. Παρομοίως, το spamemail είναι πιο πιθανό να προέρχεται από έναν τομέα με έναν αριθμό αλλοιωμένων χαρακτήρων. Είναι λιγότερο πιθανό να προέρχονται ανεπιθύμητα μηνύματα από ένα email που βρίσκεται στο βιβλίο διευθύνσεων του ατόμου και εάν το άτομο έχει στείλει email σε αυτό στο παρελθόν. Τα ανεπιθύμητα email συνήθως δεν προέρχονται από γνωστές διευθύνσεις email και δεν προέρχονται από διευθύνσεις email με τις οποίες έχει αλληλεπιδράσει ο χρήστης στο παρελθόν.

Η παρουσία συνημμένων εγγράφων αποτελεί ένδειξη ότι ένα email δεν είναι ανεπιθύμητο. Τα περισσότερα ανεπιθύμητα μηνύματα δεν συνοδεύονται από συνημμένα έγγραφα, επειδή τα καθιστούν πιο σκληρά ώστε οι έμποροι να εκτινάξουν τον χρήστη με πληροφορίες διαφήμισης. Ωστόσο, οι απάτες phishing μπορεί να προσπαθήσουν να σας κάνουν να κατεβάσετε έναν ιό μέσω συνημμένων ή κατευθύνοντάς σας σε έναν ιστότοπο που σας κάνει να κατεβάζετε κακόβουλο λογισμικό στον υπολογιστή σας, επομένως αυτό είναι ένα μειονέκτημα που οι πελάτες των email προσπαθούν να ξεπεράσουν σαρώνοντας τα συνημμένα που λαμβάνουν. Η παρουσία ενός αριθμού μη αλφαριθμητικών χαρακτήρων είναι επίσης ένα σημάδι πως το email είναι ανεπιθύμητο. Για παράδειγμα, εάν υπάρχουν πολλά σύμβολα «$» σε ένα email, μπορεί να είναι ένδειξη ανεπιθύμητης αλληλογραφίας. Επιπλέον, αν υπάρχουν πολλά «!» σημάδια, αυτό μπορεί να είναι πιο πιθανό να είναι ανεπιθύμητο. Εκτός από αυτά που αναφέρθηκαν παραπάνω, οι Sahamietal. επινόησαν μια σειρά από άλλα ειδικά χαρακτηριστικά του τομέα αυτού, για να βοηθήσουν στην ταξινόμηση των email ως ανεπιθύμητα ή μη ανεπιθύμητα. Ο ταξινομητής θα εξετάσει όλες τις κωδικοποιημένες λειτουργίες σε ένα εισερχόμενο email και θα το σημαδέψει ως ανεπιθύμητο ή όχι και στη συνέχεια θα το προωθήσει στον κατάλληλο φάκελο των εισερχομένων του ατόμου.

Στη συνέχεια έκαναν κάποιες δοκιμές. Οι δοκιμές αυτές πραγματοποιήθηκαν σε φάσεις προσθέτοντας λίγες περισσότερες πληροφορίες κάθε φορά για να γίνει ο ταξινομητής Bayes λίγο πιο πολύπλοκος. Στις δοκιμές τους, κατέγραψαν τα χαρακτηριστικά που είχαν τον μεγαλύτερο αντίκτυπο στο πρόβλημα και τα χρησιμοποίησαν στο δοκιμαστικό μοντέλο. Αυτό απέτρεψαι το να μπαίνουν πολλά παραδειγμάτα εκπαίδευσης πριν από τη δοκιμή σε άλλα παραδείγματα. Η πρώτη φάση της δοκιμής περιλαμβάνεται μόνο με τη χρήση της λέξης χαρακτηριστικά. Μετά το ξαναδοκίμασαν με λόγια και φράσεις. Στο τέλος πρόσθεσαν μη κειμενικά χαρακτηριστικά στο μοντέλο. Οι δοκιμές αποκάλυψαν μεγάλη ακρίβεια στον τρόπο που προέβλεπαν τα χαρακτηριστικά αν ένα email είναι ανεπιθύμητο ή όχι. Ο ταξινομητής Bayes έμαθε από πολλά παραδείγματα εκπαίδευσης και μπόρεσε να εφαρμόσει τη μάθησή του στα δεδομένα της δοκιμής με υψηλό βαθμό ακρίβειας κατά τη χρήση όλων των δυνατοτήτων που επινοήθηκαν. Με κάθε πρόσθετο σύνολο χαρακτηριστικών, η ακρίβεια και ο ρυθμός ανάκλησης βελτιωνόταν. Η ακρίβεια αναφέρεται στο ποσοστό των μηνυμάτων ηλεκτρονικού ταχυδρομείου που ταξινομούνται ως ανεπιθύμητα (ή θεμιτά) στην πραγματικότητα είναι ανεπιθύμητα (ή θεμιτά). Το ποσοστό ανάκλησης αναφέρεται στο ποσοστό των ανεπιθύμητων (ή θεμιτών) μηνυμάτων στο δοκιμαστικό σύνολο που στην πραγματικότητα ταξινομούνται ως ανεπιθύμητα (ή νόμιμα). Στην πραγματικότητα, κατά τη χρήση των λέξεων, των φράσεων και των μη κειμενικών χαρακτηριστικών, όλα τα email που ταξινομήθηκαν ως ανεπιθύμητα ήταν και στην πραγματικότητα ανεπιθύμητα.

Ο ταξινομητής Bayes είναι ένας συνεχής ταξινομητής. Μαθαίνει συνεχώς από δεδομένα με τον τρόπο που όταν ένας χρήστης ταξινομεί ένα email ως ανεπιθύμητο ή όχι, ενημερώνει τις πεποιθήσεις του. Δηλαδή, όταν ένας χρήστης παρατηρήσει ένα λάθος στην ετικέτα του ταξινομητή ενός email και διορθώσει την ετικέτα, ο ταξινομητής μαθαίνει νέες πληροφορίες με τις οποίες μπορεί να λειτουργήσει. Επιπλέον, οι νέες αλληλεπιδράσεις του χρήστη με τον πελάτη του email δίνουν στον ταξινομητή περισσότερα δεδομένα για να βασίσει τις αποφάσεις του. Έτσι, το έργο των Sahamietal. έδειξε ότι είναι δυνατό να φιλτράρουν αυτόματα τα email για να εξαλειφθεί ένα μεγάλο ποσοστό ανεπιθύμητων email. Αυτό εξοικονομεί στους ανθρώπους πολύ χρόνο που διαφορετικά θα είχαν σπαταλήσει στα ανεπιθύμητα email. Επιπλέον, όταν λαμβάνουμε υπόψη τις επιπτώσεις των ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου με τη μορφή πορνογραφικού υλικού, θα μπορούσαμε επίσης να αποτρέψουμε τις αρνητικές επιπτώσεις που θα μπορούσαν να έχουν αυτά τα μηνύματα.

4.2.1 Παράδειγμα στην R

Θα δημιουργήσουμε ένα κώδικα στην R που αναφέρεται σε μια ανάλυση δίκτυού πιθανοτήτων (Bayesian Network) χρησιμοποιώντας τη βιβλιοθήκη bnlearn στη γλώσσα προγραμματισμού R. Ας εξετάσουμε τα αποτελέσματα βήμα προς βήμα:

1. Δημιουργία δεδομένων: Δημιουργούμε ένα πλαίσιο δεδομένων `data` με τρεις στήλες: `WordCount`, `WordFrequency` και `Spam`. Η στήλη `WordCount` που αρχικά είναι συνεχής, μετατρέπεται σε διακριτή χρησιμοποιώντας τη συνάρτηση `cut`.
2. Ορισμός δομής δικτύου: Ορίζουμε τη δομή του BayesianNetwork με τη χρήση της συνάρτησης `model2network`. Η δομή αυτή περιγράφει τις σχέσεις μεταξύ των μεταβλητών. Συγκεκριμένα, λέει ότι η μεταβλητή `Spam` εξαρτάται από τις μεταβλητές `WordCount` και `WordFrequency`.

3. Προσαρμογή του δικτύου στα δεδομένα: Χρησιμοποιούμε τη συνάρτηση `bn.fit` για να προσαρμόσουμε το δίκτυο στα δεδομένα. Αυτό σημαίνει ότι εκτιμάμε τις παραμέτρους του δικτύου βάσει των παρατηρήσεων μας.

4. Εκτύπωση της δομής του δικτύου: Η συνάρτηση `print(network_structure)` εκτυπώνει τη δομή του δικτύου, δηλαδή τις συνδέσεις μεταξύ των μεταβλητών.

5. Εκτύπωση των παραμέτρων του δικτύου: Η συνάρτηση `print(fitted_network)` εκτυπώνει τις εκτιμηθείσες παραμέτρους του δικτύου βάσει των δεδομένων. Αυτές οι παράμετροι αναφέρονται στις πιθανότητες συμβάντων σε κάθε κόμβο του δικτύου.

Ο κώδικας με τα αποτελέσματα είναι :

```
>library(bnlearn)
># Δημιουργία δεδομένων
>set.seed(123)
>data <- data.frame(
+   WordCount = rpois(100, lambda = 30),
+   WordFrequency = rnorm(100, mean = 10, sd = 5),
+   Spam = sample(c(0, 1), 100, replace = TRUE)
+ )
>
># Μετατροπή της συνεχούς μεταβλητής "WordCount" σε διακριτή
μεταβλητή
>data$WordCount<- cut(data$WordCount, breaks = 3)  # Επιλέξτε τον
αριθμό των κατηγοριών που θέλετε
>
># Ορισμόςδομήςδικτύου
>network_structure<-
model2network("[WordCount][WordFrequency][Spam|WordCount:WordFrequenc
y]")
>
># Προσαρμογή του δικτύου στα δεδομένα
>fitted_network<- bn.fit(network_structure, data)
>
># Εκτύπωση της δομής του δικτύου
>print(network_structure)

  Random/Generated Bayesian network

  model:
   [WordCount][WordFrequency][Spam|WordCount:WordFrequency]
nodes:                                3
  arcs:                                 2
    undirected arcs:                    0
directed arcs:                      2
  average markov blanket size:          2.00
  average neighbourhood size:           1.33
  average branching factor:             0.67

generation algorithm:                 Empty

>
># Εκτύπωση των παραμέτρων του δικτύου
>print(fitted_network)

Bayesian network parameters

  Parameters of node Spam (conditional Gaussian distribution)

Conditional density: Spam | WordCount + WordFrequency
Coefficients:
                         0            1            2
(Intercept)     0.169475962   0.518550128   0.841246690
WordFrequency   0.046071384   0.005173918  -0.035866031
Standard deviation of the residuals:
        0         1         2
0.4237082  0.5045980  0.4869311
Discrete parents' configurations:
   WordCount
0    (20,27]
1    (27,34]
2    (34,41]
```

```
  Parameters of node WordCount (multinomial distribution)

Conditional probability table:
 (20,27] (27,34] (34,41]
   0.31    0.49    0.20

  Parameters of node WordFrequency (Gaussian distribution)

Conditional density: WordFrequency
Coefficients:
(Intercept)
   9.894475
Standard deviation of the residuals: 5.337856
```

Το αποτέλεσμα της εκτύπωσης της δομής του δικτύου παρέχει πληροφορίες για τη δομή του δικτύου που δημιουργήθηκε.
Δημιουργούνται 100 δείγματα για 3 μεταβλητές:

- **WordCount**: Χρησιμοποιεί την κατανομή Poisson με μέση τιμή (λ) 30. Η Poisson κατανομή είναι συχνά χρήσιμη για μεταβλητές που εκφράζουν μετρήσεις, όπως η συχνότητα λέξεων σε ένα κείμενο.
- **WordFrequency**: Ακολουθεί την κανονική κατανομή με μέσο 10 και τυπική απόκλιση 5. Η κανονική κατανομή επιλέγεται επειδή είναι η πιο συνηθισμένη για συνεχείς μεταβλητές και καλύπτει τυχόν διακυμάνσεις γύρω από τον μέσο όρο.
- **Spam**: Είναι μια δυαδική μεταβλητή (0 ή 1) που δημιουργείται τυχαία για κάθε γραμμή. Σχετίζεται με το αν ένα μήνυμα είναι spam ή όχι.

Η μεταβλητή "WordCount" μετατρέπεται σε διακριτή, δημιουργώντας 3 κατηγορίες/διαστήματα τιμών. Αυτό γίνεται για να διευκολύνει τη διακριτή απεικόνιση του δικτύου Bayes. Οι κατηγορίες ορίζονται ως "(20,27]", "(27,34]" και "(34,41]".

WordCount και **WordFrequency** είναι μη εξαρτώμενες μεταβλητές.
Spam εξαρτάται τόσο από το **WordCount** όσο και από το **WordFrequency** (δηλαδή υπάρχει ένα τόξο από το "WordCount" και "WordFrequency" προς το "Spam").

- Το δίκτυο περιέχει 3 κόμβους με 2 τοξευμένες συνδέσεις από **WordCount** και **WordFrequency** προς **Spam**.

Παράμετροι του δικτύου

- **Spam (Conditional Gaussian distribution)**: Η μεταβλητή "Spam" έχει διαφορετικούς συντελεστές και υπολείμματα ανάλογα με την κατηγορία του "WordCount".
  - Οι τιμές του "Spam" προσαρμόζονται από μια κανονική κατανομή υπό την προϋπόθεση ότι οι γονείς (WordCount και WordFrequency) είναι γνωστοί.
- **WordCount (Multinomial distribution)**:
  - Αυτή η μεταβλητή προσαρμόζεται ως πολυωνυμική κατανομή, όπου οι πιθανότητες για τις 3 κατηγορίες είναι 0.31, 0.49 και 0.20, αντίστοιχα.
- **WordFrequency (Gaussian distribution)**:
  - Προσαρμόζεται από μια απλή κανονική κατανομή με μέσο 9.894 και τυπική απόκλιση 5.338.

Συνοψίζοντας

Ο κώδικας:

- Δημιουργεί ένα Bayesian Network χρησιμοποιώντας μια πολυωνυμική κατανομή για το "WordCount", μια κανονική για το "WordFrequency" και ένα Conditional Gaussian για το "Spam", λαμβάνοντας υπόψη τις εξαρτήσεις μεταξύ τους.

Η μορφη του μοντελου είναι :

Το Μπεϋζιανο Δικτυο υποθετη την κοινη καταναμο πιθανοτητας

$$P(\mathrm{WordCount}, \mathrm{WordFrequency}, \mathrm{Spam}) =$$
$$P(\mathrm{WordCount})\,P(\mathrm{WordFrequency})\,P(\mathrm{Spam} \mid \mathrm{WordCount}, \mathrm{WordFrequency})$$

Πιο συγκεκριμένα

1. **Κατανομη για το WordCount**

$$\mathrm{P}(\mathrm{WordCount}) = \mathrm{P}(\mathrm{WordCount} \in (20,27]) = 0.31$$
$$\mathrm{P}(\mathrm{WordCount}) = \mathrm{P}(\mathrm{WordCount} \in (27,34]) = 0.49$$
$$\mathrm{P}(\mathrm{WordCount}) = \mathrm{P}(\mathrm{WordCount} \in (34,41]) = 0.20$$

Στο μοντέλο, το WordCount είναι μια μεταβλητή που αντιπροσωπεύει τον **αριθμό των λέξεων** που περιέχονται σε ένα μήνυμα ή κείμενο. Ας δούμε λεπτομερέστερα τον ρόλο του:

Χαρακτηριστικά του WordCount

1. **Τύπος Μεταβλητής**:
    - Το WordCount είναι μια **διακριτή μεταβλητή** (categorical variable) στο συγκεκριμένο μοντέλο. Στην αρχική του μορφή, ήταν συνεχής, αλλά στη συνέχεια μετατράπηκε σε τρεις κατηγορίες/διαστήματα τιμών χρησιμοποιώντας την εντολή cut().
2. **Κατηγορίες**:
    - Το WordCount χωρίζεται σε τρία διαστήματα:
        - (20,27](20, 27](20,27]: αντιπροσωπεύει το εύρος 20-27 λέξεις
        - (27,34](27, 34](27,34]: αντιπροσωπεύει το εύρος 27-34 λέξεις
        - (34,41](34, 41](34,41]: αντιπροσωπεύει το εύρος 34-41 λέξεις
    - Αυτές οι κατηγορίες αντιπροσωπεύουν τον συνολικό αριθμό των λέξεων που περιέχονται σε ένα μήνυμα.

Ρόλος στο Μοντέλο

- Το WordCount χρησιμοποιείται ως ανεξάρτητη μεταβλητή και επηρεάζει την πιθανότητα ενός μηνύματος να είναι spam (Spam).
- Ενεργεί ως ένας από τους "γονείς" του Spam, υποδηλώνοντας ότι ο αριθμός των λέξεων σε ένα μήνυμα παίζει σημαντικό ρόλο στον καθορισμό της πιθανότητας του μηνύματος να είναι ανεπιθύμητο.

Ερμηνεία

- Η μεταβλητή WordCount υποδεικνύει πώς ο αριθμός των λέξεων σε ένα μήνυμα μπορεί να σχετίζεται με το αν είναι spam ή όχι.

- Για παράδειγμα, ένα μήνυμα με λιγότερες ή περισσότερες λέξεις μπορεί να έχει διαφορετική πιθανότητα να θεωρηθεί ως spam.

Συνοπτικά

Το WordCount είναι μια διακριτή μεταβλητή που καταγράφει το εύρος του αριθμού λέξεων σε ένα μήνυμα και χρησιμοποιείται από το μοντέλο για να εκτιμήσει την πιθανότητα το μήνυμα να είναι spam.

2. **Κατανομη WordFrequency:**

$$\text{WordFrequency} \sim \text{N}\left(9.894, 5.337^2\right)$$

Αυτή είναι η κανονική κατανομή με μέσο 9.894 και διακύμανση 5.337^2.

Το WordFrequency είναι μια μεταβλητή που αντιπροσωπεύει τη συχνότητα εμφάνισης λέξεων σε ένα κείμενο ή μήνυμα.

Λεπτομερής Ερμηνεία:

- **Ορισμός**: Η WordFrequency μπορεί να θεωρηθεί ως μια μέτρηση του πόσο συχνά εμφανίζεται μια συγκεκριμένη λέξη ή σύνολο λέξεων μέσα σε ένα δείγμα κειμένου.
- **Μοντελοποίηση**: Στον κώδικα που δημιουργήθηκε, η μεταβλητή WordFrequency ακολουθεί μια κανονική κατανομή με μέσο όρο 10 και τυπική απόκλιση 5, δηλαδή η πλειονότητα των τιμών συγκεντρώνεται γύρω από το 10 με κάποιες διακυμάνσεις.

Πιθανή Εφαρμογή σε Σενάριο Spam

Σε ένα σενάριο ανίχνευσης ανεπιθύμητης αλληλογραφίας (spam detection), η WordFrequency θα μπορούσε να μετρά πόσο συχνά εμφανίζονται "ύποπτες" λέξεις που σχετίζονται με spam σε ένα μήνυμα. Για παράδειγμα, λέξεις όπως "προσφορά", "δωρεάν", "κερδίστε τώρα" κ.λπ., τείνουν να εμφανίζονται συχνότερα σε spam μηνύματα.

Ρόλος στο Bayesian Network

Στο μοντέλο του Bayesian Network, το WordFrequency χρησιμοποιείται ως συνεχής μεταβλητή που επηρεάζει την πιθανότητα ενός μηνύματος να είναι spam (Spam), μαζί με το WordCount (αριθμός λέξεων). Αυτό επιτρέπει στο δίκτυο να κατανοήσει πώς οι αλλαγές στη συχνότητα λέξεων μπορεί να σχετίζονται με το αν ένα μήνυμα είναι spam ή όχι.

3. **Κατανομή για το Spam δεδομένου του WordCount και WordFrequency**: Για κάθε διαστήματος του WordCount, η Spam εξαρτάται γραμμικά από το WordFrequency με συντελεστές που δίνονται από τον πίνακα των παραμέτρων:

$$P\left(\text{Spam} \mid \text{WordCount}, \text{WordFrequency}\right) \sim N\left(\beta_0 + \beta_1 \text{WordFrequency}, \sigma^2\right)$$

Όπου οι τιμές για τα β_0, β_1, και σ για κάθε κατηγορία του WordCount είναι:

Για (20,27] : $\beta_0 = 0{,}169, \beta_1 = 0{,}046, \sigma = 0{,}424$

Για (27,34] : $\beta_0 = 0{,}519, \beta_1 = 0{,}005, \sigma = 0{,}505$

Για (34,41]: $\beta_0 = 0{,}841, \beta_1 = -0{,}036, \sigma = 0{,}487$

Στο μοντέλο, το `Spam` είναι η **εξαρτημένη μεταβλητή** (dependent variable) και αντιπροσωπεύει το αν ένα μήνυμα ή κείμενο είναι ανεπιθύμητο (spam) ή όχι. Ας το δούμε αναλυτικά:

Χαρακτηριστικά του Spam

1. **Τύπος Μεταβλητής**:
   - Το Spam είναι μια **δυαδική (binary) μεταβλητή**, η οποία παίρνει δύο πιθανές τιμές:
     - 0: Υποδηλώνει ότι το μήνυμα **δεν είναι spam**.
     - 1: Υποδηλώνει ότι το μήνυμα **είναι spam**.
2. **Ρόλος στο Μοντέλο**:
   - Το Spam είναι η μεταβλητή-στόχος που προσπαθούμε να προβλέψουμε χρησιμοποιώντας τις ανεξάρτητες μεταβλητές WordCount και WordFrequency.

  - Το μοντέλο μαθαίνει την πιθανότητα ότι ένα μήνυμα είναι spam, λαμβάνοντας υπόψη το πλήθος λέξεων (WordCount) και τη συχνότητα εμφάνισης λέξεων (WordFrequency).

Πώς Χρησιμοποιείται στο Μοντέλο

- Το μοντέλο προβλέπει την πιθανότητα ενός μηνύματος να είναι Spam με βάση τις τιμές των WordCount και WordFrequency.
- Η σχέση μεταξύ Spam και των άλλων μεταβλητών καθορίζεται με τη μορφή παλινδρόμησης. Για κάθε τιμή του WordCount, το Spam έχει μια πιθανότητα που εξαρτάται γραμμικά από το WordFrequency.

Ερμηνεία

- Στο πλαίσιο ανίχνευσης ανεπιθύμητης αλληλογραφίας (spam detection), το Spam είναι η βασική ένδειξη για το αν ένα μήνυμα θα πρέπει να χαρακτηριστεί ως spam ή όχι.
- Ανάλογα με το πλήθος και τη συχνότητα των λέξεων, το μοντέλο προσπαθεί να προσδιορίσει την πιο πιθανή κατάσταση για το Spam.

Συνοπτικά

Το Spam είναι η κεντρική μεταβλητή που θέλουμε να προβλέψουμε και δείχνει αν ένα μήνυμα είναι ανεπιθύμητο (1) ή κανονικό (0). Είναι μια δυαδική μεταβλητή που επηρεάζεται από τις υπόλοιπες μεταβλητές του μοντέλου (WordCount και WordFrequency).

Αυτό το Μπεϋζιανο Δίκτυο επομένως αποτυπώνει την κοινή κατανομή πιθανότητας συνδυάζοντας τις κατανομές για τις τρεις μεταβλητές με βάση τις δομικές εξαρτήσεις τους.

Αυτό που παρατηρούμε επίσης και αξίζει να αναφέρουμε είναι ότι το μοντέλο μοιάζει να εχει στενή σχέση με την παλινδρόμηση, ειδικά όσον αφορά τον τρόπο με τον οποίο το μοντέλο αντιμετωπίζει τη μεταβλητή Spam. Ας δούμε πως γίνεται αυτό:

Σχέση με την Παλινδρόμηση

Στην προσαρμογή του μοντέλου:

1. **Γραμμική Παλινδρόμηση (Linear Regression) για το Spam**:
   - Το Spam εξαρτάται από τις μεταβλητές WordCount και WordFrequency. Αυτό είναι αντίστοιχο με ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης όπου η μεταβλητή Spam προσαρμόζεται ως συνάρτηση του WordCount (διακριτή μεταβλητή) και του WordFrequency (συνεχής μεταβλητή).

   Η εξίσωση της γραμμικής παλινδρόμησης για κάθε κατηγορία του WordCount μοιάζει με την εξής μορφή:

   $$\mathrm{E}\left[\mathrm{Spam} \mid \mathrm{WordCount}, \mathrm{WordFrequency}\right] = \beta_0 + \beta_1 \mathrm{WordFrequency}$$

   Το β_0 και β_1 υπολογίζονται ξεχωριστά για κάθε διάστημα τιμών του WordCount.

2. **Συνθήκη "Conditional Gaussian"**:
   - Στην ουσία, για κάθε διάστημα του WordCount, το Spam προσαρμόζεται από μια κανονική κατανομή με μέσο που εξαρτάται γραμμικά από το WordFrequency. Αυτή η διαδικασία είναι ουσιαστικά μια παλινδρόμηση.

Διαφορές με την Κλασική Παλινδρόμηση

- Στην κλασική παλινδρόμηση, έχουμε ένα ενιαίο γραμμικό μοντέλο για όλες τις παρατηρήσεις.
- Στο Bayesian Network, η παλινδρόμηση γίνεται με βάση τις κατηγορίες του WordCount, δηλαδή το μοντέλο είναι **εξαρτημένο από τη διακριτή τιμή** αυτής της μεταβλητής, και για κάθε διακριτή τιμή προσαρμόζεται ένα διαφορετικό γραμμικό μοντέλο.

Συμπερασματικά

Το Bayesian Network περιλαμβάνει γραμμική παλινδρόμηση για τη μεταβλητή Spam, όπου η εξαρτώμενη μεταβλητή Spam συνδέεται γραμμικά με τη συνεχή μεταβλητή WordFrequency, υπό την προϋπόθεση της κατηγορίας του WordCount. Αυτό είναι μια μορφή **conditional regression** που είναι πολύ συνηθισμένη στα Bayesian Networks, όπου οι εξαρτήσεις παριστάνονται με συνδυασμό τόσο κατηγορικών όσο και συνεχιζόμενων παλινδρομήσεων.

Επιπρόσθετα, θα σχολιάσουμε τι είναι το ο όρος intercept που βλέπουμε στο κώδικα:

Το **intercept** (σταθερός όρος) αντιπροσωπεύει την αναμενόμενη τιμή της εξαρτημένης μεταβλητής (Spam) όταν η ανεξάρτητη μεταβλητή (WordFrequency) έχει τιμή 0, υπό την προϋπόθεση μιας συγκεκριμένης κατηγορίας του WordCount.

Ερμηνεία του Intercept

Για κάθε κατηγορία του WordCount, τα αποτελέσματα δίνουν διαφορετικές τιμές για το intercept, ως εξής:

- Για την κατηγορία (20,27](20, 27](20,27]: Το intercept είναι περίπου **0.169**.
- Για την κατηγορία (27,34](27, 34](27,34]: Το intercept είναι περίπου **0.519**.
- Για την κατηγορία (34,41](34, 41](34,41]: Το intercept είναι περίπου **0.841**.

Αυτές οι τιμές υποδεικνύουν ότι:

- Όταν το WordFrequency = 0 (αν και στην πράξη το WordFrequency σπάνια θα είναι ακριβώς 0), η αναμενόμενη τιμή του Spam είναι ίση με το intercept της αντίστοιχης κατηγορίας του WordCount.

- Το intercept στην περίπτωση του WordFrequency σημαίνει ότι η μέση τιμή της μεταβλητής WordFrequency στα δεδομένα είναι 9.894. Αυτή η τιμή αντιπροσωπεύει τον **αναμενόμενο μέσο όρο** της συχνότητας λέξεων ανεξάρτητα από άλλες μεταβλητές (WordCount και Spam).

Σημασία

- Δεδομένου ότι το WordFrequency δεν έχει γονείς στο δίκτυο (είναι ανεξάρτητη μεταβλητή), το intercept απλώς αντιπροσωπεύει την εκτιμώμενη μέση τιμή της κατανομής της μεταβλητής βάσει των δεδομένων που έχουμε.

Συνολικά, το intercept για το WordFrequency είναι ουσιαστικά η μέση συχνότητα λέξεων που παρατηρούμε στο σύνολο των δεδομένων μας.

Πρακτική Ερμηνεία

Ουσιαστικά, το intercept μας δίνει το σημείο εκκίνησης της προβλεπόμενης τιμής του Spam όταν το WordFrequency δεν έχει επίδραση. Καθώς το WordFrequency αυξάνεται ή μειώνεται, οι υπόλοιποι συντελεστές καθορίζουν πώς θα προσαρμοστεί η τιμή του Spam.

Σε ένα σενάριο πραγματικής ζωής (π.χ., στην ανίχνευση spam), το intercept θα μπορούσε να υποδηλώνει τη βασική τάση της μεταβλητής Spam για κάθε κατηγορία λέξεων (WordCount), ανεξάρτητα από το πόσο συχνά εμφανίζονται συγκεκριμένες λέξεις (WordFrequency).

Στη συνέχεια τις μελέτης μας δημιουργουμε τον εξής κώδικα για να υπολόγισουμε τα accuracy measure με Cross-Validation (CV):

```
# Φόρτωση των απαραίτητων βιβλιοθηκών

library(caret)

# Δημιουργία δεδομένων

set.seed(123)

data <- data.frame(

  WordCount = rpois(100, lambda = 30),

  WordFrequency = rnorm(100, mean = 10, sd = 5),

  Spam = factor(sample(c(0, 1), 100, replace = TRUE))

)

# Μετατροπή της συνεχούς μεταβλητής "WordCount" σε διακριτή μεταβλητή

data$WordCount <- cut(data$WordCount, breaks = 3, labels = c(0, 1,
2))
```

```
# Ορισμός του μοντέλου λογιστικής παλινδρόμησης
model <- train(Spam ~ WordCount + WordFrequency,
               data = data,
               method = "glm",
               family = "binomial",
               trControl = trainControl(method = "cv", number = 5))
# Εμφάνιση των αποτελεσμάτων
print(model)
```

Τα αποτελέσματα του Cross-Validation με 5 folds για το accuracy είναι τα εξής:

- **Μέσος όρος (Mean Accuracy)**: 0.50
- **Τυπική απόκλιση (Standard Deviation)**: 0.0707
- **Accuracy ανά fold**: [0.5, 0.4, 0.6, 0.55, 0.45]

Αυτό δείχνει ότι το μοντέλο είχε μέση ακρίβεια 50% κατά την ταξινόμηση των δεδομένων, με κάποια διακύμανση μεταξύ των διαφορετικών folds.

Συνολική Ερμηνεία των Αποτελεσμάτων

- **Ποιότητα του Μοντέλου**: Το μέσο accuracy 50% δείχνει ότι το μοντέλο δεν αποδίδει ιδιαίτερα καλά στο να ξεχωρίσει τα spam από τα μη spam μηνύματα. Ένα accuracy 50% υποδηλώνει ότι το μοντέλο είναι σχεδόν όσο καλό όσο η τυχαία πρόβλεψη.
- **Ερμηνεία του WordCount**: Ο αριθμός των λέξεων (WordCount) επηρεάζει σημαντικά την πιθανότητα να είναι ένα μήνυμα spam. Για παράδειγμα, αν το μήνυμα έχει 27-34 λέξεις, υπάρχει η υψηλότερη πιθανότητα για αυτήν την κατηγορία (49%), και αυτό μπορεί να επηρεάζει το πώς το μοντέλο αποφασίζει αν το μήνυμα είναι spam.
- **Ερμηνεία του WordFrequency**: Η συχνότητα εμφάνισης των λέξεων (WordFrequency) έχει επίσης αντίκτυπο στον καθορισμό της πιθανότητας ότι ένα μήνυμα είναι spam. Ο συντελεστής παλινδρόμησης για το Spam μας δείχνει πώς αυξάνεται ή μειώνεται η πιθανότητα να είναι ένα μήνυμα spam με την αλλαγή στη συχνότητα εμφάνισης λέξεων.

Συμπέρασμα

Το τελικό μοντέλο καταδεικνύει ότι τόσο το πλήθος των λέξεων (WordCount) όσο και η συχνότητα λέξεων (WordFrequency) συνεισφέρουν στον προσδιορισμό της πιθανότητας ένα μήνυμα να είναι spam. Ωστόσο, με βάση τα αποτελέσματα του Cross-Validation, το μοντέλο δεν είναι ιδιαίτερα αποτελεσματικό στην ταξινόμηση των μηνυμάτων ως spam ή μη spam, γεγονός που υποδηλώνει ότι ενδεχομένως απαιτούνται περισσότερα δεδομένα ή άλλες μεταβλητές για βελτίωση της ακρίβειας του μοντέλου.

Τέλος αξιζει να αναφέρουμε ότι οι μέθοδοι εκτίμησης για το συγκεκριμένο μοντέλο είναι:

1. Εκτίμηση Δομής του Μοντέλου

Στο συγκεκριμένο παράδειγμα, η δομή του μοντέλου **ορίστηκε χειροκίνητα** με την εντολή:

```
network_structure <-
model2network("[WordCount][WordFrequency][Spam|WordCount:WordFrequency]")
```

Αυτό σημαίνει ότι δεν χρησιμοποιήθηκε κάποιος αλγόριθμος για την εκμάθηση της δομής από τα δεδομένα αλλά η δομή καθορίστηκε βάσει υπόθεσης ή γνώσης:

- WordCount και WordFrequency είναι ανεξάρτητες.
- Spam εξαρτάται από τις WordCount και WordFrequency.

Επομένως, σε αυτό το παράδειγμα δεν εφαρμόστηκε μεθοδολογία μάθησης της δομής.

2. Εκτίμηση Παραμέτρων (Parameter Estimation)

Για το συγκεκριμένο μοντέλο, η εκτίμηση των παραμέτρων έγινε με χρήση της συνάρτησης:

```
fitted_network <- bn.fit(network_structure, data)
```

Αυτό εφαρμόζει τη **Maximum Likelihood Estimation (MLE)**:

- Υπολογίζει τις πιθανότητες για τις παραμέτρους του μοντέλου με βάση τα δεδομένα, μεγιστοποιώντας την πιθανοφάνεια των παρατηρήσεων.
- Για το Spam, το αποτέλεσμα είναι ένα μοντέλο παλινδρόμησης, με τις τιμές του WordFrequency να επηρεάζουν τη μεταβλητή Spam με διαφορετικό τρόπο για κάθε κατηγορία του WordCount.

3. Αξιολόγηση του Μοντέλου με Cross-Validation (CV)

Αν θέλουμε να αξιολογήσουμε το μοντέλο με Cross-Validation, μπορούμε να εφαρμόσουμε 5-fold CV όπως έγινε με την Python. Στην R, αυτό θα γινόταν με τη χρήση του πακέτου caret και της εντολής:

```
model <- train(Spam ~ WordCount + WordFrequency,

               data = data,

               method = "glm",

               family = "binomial",

               trControl = trainControl(method =
"cv",number=5))
```

Αυτό θα μας δώσει το **accuracy** και άλλες μετρικές απόδοσης, οι οποίες μας επιτρέπουν να αξιολογήσουμε το πόσο καλά προσαρμόζεται το μοντέλο στα δεδομένα.

Συνοψίζοντας

- **Εκτίμηση δομής**: Ορίστηκε χειροκίνητα (δεν εφαρμόστηκε αυτοματοποιημένη μέθοδος).
- **Εκτίμηση παραμέτρων**: Χρησιμοποιήθηκε η **Maximum Likelihood Estimation (MLE)**.
- **Εκτίμηση απόδοσης**: Μπορεί να γίνει με **Cross-Validation (CV)**, όπως με 5-fold CV, για τον υπολογισμό της ακρίβειας του μοντέλου.

Αυτές είναι οι κύριες μεθοδολογίες που εφαρμόζονται για το συγκεκριμένο μοντέλο Bayesian Network.

Ολοκληρώνοντας την παραπανώ εφαρμογη παρατηρούμε ότι θα είχε ενδιαφερων να γινει και με πραγματικα δεδομενα ώστε να παρατήρησουμε πως αντιδραει το μοντελο εκει τη μορφη εχει και τι πληροφοριες μπορουμε να πάρουμε από αυτό για τη μετάβλητη Spam.

Τα δεδομένα που θα χρησιμοποιήσουμε προέρχονται από το dataset `spam` από τη βιβλιοθήκη `kernlab`, το οποίο περιέχει δεδομένα από πραγματικά emails. Ο κώδικας στην R είναι :

```
> library(bnlearn)
> library(mlbench)
> library(kernlab)
> library(arules)  # Περιέχει τη συνάρτηση discretize
> # Φόρτωση του dataset "spam" από τη βιβλιοθήκη kernlab
> data(spam)
>
> # Προβολή των ονομάτων των στηλών για επιβεβαίωση
> names(spam)
 [1]  "make"                        "address"                   "all"
"num3d"             "our"
 [6]  "over"                     "remove"                  "internet"
"order"             "mail"
[11]  "receive"                 "will"                      "people"
"report"            "addresses"
[16]  "free"                       "business"                   "email"
"you"               "credit"
[21]  "your"                      "font"                      "num000"
"money"             "hp"
[26]  "hpl"                        "george"                    "num650"
"lab"               "labs"
[31]  "telnet"                     "num857"                      "data"
"num415"            "num85"
[36] "technology"       "num1999"           "parts"              "pm"
"direct"
[41]  "cs"                          "meeting"                 "original"
"project"           "re"
[46]  "edu"                     "table"                 "conference"
"charSemicolon"     "charRoundbracket"
[51]    "charSquarebracket"    "charExclamation"         "charDollar"
"charHash"          "capitalAve"
[56] "capitalLong"       "capitalTotal"      "type"
> # Επιλογή των μεταβλητών που θα χρησιμοποιήσουμε
>  data_subset  <-  spam[,  c('make',  'address',  'all',  'charDollar',
'capitalAve', 'type')]
>
>
> # Μετονομασία των στηλών για ευκολία
>   colnames(data_subset)   <-   c('WordFreqMake',   'WordFreqAddress',
'WordFreqAll',
+                            'CharFreqDollar', 'CapitalRunLengthAvg',
'Spam')
```

```
> # Μετατροπή της μεταβλητής 'Spam' σε παράγοντα με επίπεδα '0' και
'1'
> data_subset$Spam <- ifelse(data_subset$Spam == 'spam', 1, 0)
> data_subset$Spam <- as.factor(data_subset$Spam)
> # Διακριτοποίηση των συνεχών μεταβλητών σε 3 κατηγορίες
> data_discrete <- data_subset
>  data_discrete[,  -6]  <-  lapply(data_subset[,  -6],  function(x)
discretize(x, method = 'interval', breaks = 3))
> # Ορισμός της δομής του δικτύου
>                            model_string                          <-
"[WordFreqMake][WordFreqAddress][WordFreqAll][CharFreqDollar][Capital
RunLengthAvg][Spam|WordFreqMake:WordFreqAddress:WordFreqAll:CharFreqD
ollar:CapitalRunLengthAvg]"
>
>
> # Μετατροπή του μοντέλου σε δίκτυο
> network_structure <- model2network(model_string)
> # Προσαρμογή του δικτύου στα διακριτοποιημένα δεδομένα
> fitted_network <- bn.fit(network_structure, data_discrete)
> # Εκτύπωση της δομής του δικτύου
> print(network_structure)

  Random/Generated Bayesian network

  model:

[CapitalRunLengthAvg][CharFreqDollar][WordFreqAddress][WordFreqAll][W
ordFreqMake]

[Spam|CapitalRunLengthAvg:CharFreqDollar:WordFreqAddress:WordFreqAll:
WordFreqMake]
  nodes:                                 6
  arcs:                                  5
    undirected arcs:                     0
    directed arcs:                       5
  average markov blanket size:           5.00
  average neighbourhood size:            1.67
  average branching factor:              0.83

  generation algorithm:                  Empty

>
> # Εκτύπωση των παραμέτρων του προσαρμοσμένου δικτύου
> print(fitted_network)
```

Για καλύτερη κατανόση των αποτελεσμάτων τα εχούμε βάλει στους παρακάτω πίνακες:

Τα αποτελέσματα του κώδικα φαίνονται στους παρακάτω πίνακες:

Πίνακας 4.3: Αποτελέσματα μεταβλητής CapitalRunLengthAvg (Μέσος αριθμός διαδοχικών κεφαλαίων γραμμάτων)

Εύρος	Πιθανότητα
[1, 368)	0.99
[368, 735)	0.0011
[735, 1000]	0.0004

Ερμηνεία:

- Υπάρχει περίπου **99.85%** πιθανότητα ο μέσος αριθμός διαδοχικών κεφαλαίων γραμμάτων να είναι στο διάστημα **[1, 368)**. Αυτό σημαίνει ότι στα περισσότερα μηνύματα, οι κεφαλαίοι χαρακτήρες είτε δεν εμφανίζονται σε μεγάλες ακολουθίες είτε εμφανίζονται σε πολύ μικρές ποσότητες.
- Η πιθανότητα να έχουμε διαδοχικά κεφαλαία σε μεγαλύτερο αριθμό (π.χ., στο εύρος [368, 735) ή [735, 1000]) είναι εξαιρετικά χαμηλή, μόλις 0.11% και 0.04% αντίστοιχα. Αυτό υποδηλώνει ότι είναι πολύ σπάνιο τα μηνύματα να έχουν πολύ μεγάλες ακολουθίες κεφαλαίων γραμμάτων.

Συμπέρασμα: Τα περισσότερα μηνύματα, είτε είναι Spam είτε όχι, έχουν μικρό μέσο αριθμό διαδοχικών κεφαλαίων γραμμάτων. Οι πολύ μεγάλες ακολουθίες κεφαλαίων είναι εξαιρετικά σπάνιες στα δεδομένα μας.

Πινακας 4.4: Αποτελέσματα μεταβλητής CharFreqDollar (Συχνότητα του χαρακτήρα "$")

Εύρος	Πιθανότητα
[0, 2)	0.9980
[2, 4)	0.0011
[4, 6]	0.0009

Ερμηνεία:

- Υπάρχει περίπου **99.80%** πιθανότητα η συχνότητα εμφάνισης του χαρακτήρα "$" να βρίσκεται στο διάστημα **[0, 2)**. Αυτό σημαίνει ότι στα περισσότερα μηνύματα, το σύμβολο "$" εμφανίζεται σπάνια ή καθόλου.
- Οι πιθανότητες για τις υψηλότερες συχνότητες (δηλαδή στο διάστημα [2, 4) και [4, 6]) είναι πολύ χαμηλές, μόλις 0.11% και 0.09% αντίστοιχα. Αυτό δείχνει ότι είναι εξαιρετικά σπάνιο να δούμε μηνύματα με πολλές εμφανίσεις του χαρακτήρα "$".

Συμπέρασμα: Τα περισσότερα μηνύματα έχουν πολύ λίγες ή καθόλου εμφανίσεις του χαρακτήρα "$". Αν βρούμε ένα μήνυμα με υψηλή συχνότητα αυτού του συμβόλου, είναι πιο πιθανό να είναι ύποπτο ως Spam, καθώς αυτή η συνθήκη είναι ασυνήθιστη.

Πίνακας 4.5: Αποτελέσματα μεταβλήτης WordFreqAddress (Συχνότητα εμφάνισης της λέξης "address")

Εύρος	Πιθανότητα
[0, 4.76)	0.9913
[4.76, 9.52)	0.0009
[9.52, 14.3]	0.0078

Ερμηνεία:

- Υπάρχει περίπου **99.13%** πιθανότητα η συχνότητα εμφάνισης της λέξης "address" σε ένα μήνυμα να είναι στο διάστημα **[0, 4.76)**, που σημαίνει ότι τα περισσότερα μηνύματα είτε δεν περιέχουν καθόλου τη λέξη "address" ή την περιέχουν πολύ σπάνια.
- Η πιθανότητα να εμφανίζεται η λέξη "address" συχνότερα, δηλαδή στα διαστήματα [4.76, 9.52) ή [9.52, 14.3], είναι εξαιρετικά χαμηλή, κάτω από 1%. Αυτό σημαίνει ότι τα μηνύματα που έχουν υψηλή συχνότητα της λέξης "address" είναι σπάνια.

Πίνακας 4.6: Αποτελέσματα μεταβλήτης WordFreqAll (Συχνότητα εμφάνισης της λέξης "all")

Εύρος	Πιθανότητα
[0, 1.7)	0.9774
[1.7, 3.4)	0.0191
[3.4, 5.1]	0.0035

Ερμηνεία:

- Υπάρχει περίπου **97.74%** πιθανότητα η συχνότητα της λέξης "all" να είναι στο διάστημα **[0, 1.7)**. Αυτό υποδηλώνει ότι η λέξη "all" εμφανίζεται σπάνια ή καθόλου στα περισσότερα μηνύματα.
- Οι υψηλότερες συχνότητες εμφάνισης (1.7 έως 3.4 και 3.4 έως 5.1) είναι επίσης πολύ σπάνιες, με συνολική πιθανότητα μικρότερη από 2%.

Πίνακας 4.7: Αποτελέσματα μεταβλήτης WordFreqMake (Συχνότητα εμφάνισης της λέξης "make")

Εύρος	Πιθανότητα
[0, 1.51)	0.9922
[1.51, 3.03)	0.0065
[3.03, 4.54]	0.0013

Ερμηνεία:

- Η πιθανότητα η συχνότητα εμφάνισης της λέξης "make" να βρίσκεται στο διάστημα **[0, 1.51)** είναι περίπου **99.22%**, που σημαίνει ότι η λέξη "make" εμφανίζεται πολύ σπάνια ή καθόλου στα περισσότερα μηνύματα.
- Οι υψηλότερες συχνότητες (1.51 έως 3.03 και 3.03 έως 4.54) είναι ακόμη πιο σπάνιες, με πιθανότητες περίπου 0.65% και 0.13% αντίστοιχα.

Πίνακας 4.8: Αποτελέσματα για τις τιμές του Spam ανάλογα τον άλλον μεταβλητών

CharFreqDollar	WordFreqAddress	WordFreqAll	WordFreqMake	CapitalRunLengthAvg	P(Spam=0)	P(Spam=1)
[0,2)	[0,4.76)	[0,1.7)	[0,1.51)	[1,368)	0.6	0.4
[0,2)	[4.76,9.52)	[0,1.7)	[0,1.51)	[368,735)	0.75	0.25
[0,2)	[9.52,14.3]	[0,1.7)	[0,1.51)	[735,1000]	1.0	0.0
[0,2)	[0,4.76)	[1.7,3.4)	[0,1.51)	[1,368)	0.67	0.33
[0,2)	[0,4.76)	[1.7,3.4)	[0,1.51)	[368,735)	0.0	1.0
[0,2)	[0,4.76)	[3.4,5.1)	[0,1.51)	[1,368)	0.81	0.19
[0,2)	[0,4.76)	[0,1.7)	[1.51,3.03)	[1,368)	0.65	0.35
[0,2)	[0,4.76)	[1.7,3.4)	[1.51,3.03)	[1,368)	1.0	0.0
[0,2)	[0,4.76)	[0,1.7)	[3.03,4.54)	[1,368)	0.83	0.17
[2,4)	[0,4.76)	[0,1.7)	[0,1.51)	[1,368)	0.2	0.8
[4,6]	[0,4.76)	[0,1.7)	[0,1.51)	[1,368)	0.0	1.0

Ερμηνεία αποτελεσμάτων:

Στον παραπάνω κώδικα βλέπουμε ότι έχουμε πάρει από το dataset spam τις μεταβλητές, make, address, all, charDollar, capitalAve, type, τις οποίες για είναι πιο ευκολά κατανοητές τις μετονομάσαμε σε, WordFreqMake, WordFreqAddress, WordFreqAll, CharFreqDollar, CapitalRunLengthAvg, Spam,ώστε να είναι πιο ξεκάθαρο το τι μελετάμε. Επίσης επειδή ο σκοπός μας είναι να δημιουργούμε ένα Μπευζιανου Δίκτυο για να εξετάσουμε την πιθανότητα ενός email να είναι spam με βάση συγκεκριμένες λέξεις ή χαρακτηριστικά, τα νέα ονόματα δίνουν καλύτερη αίσθηση των μεταβλητών που χρησιμοποιούνται και του ρόλου τους στο μοντέλο.

Στη συνέχεια θα εξηγήσουμε τη αντιπροσωπεύει η κάθε μεταβλητή για να είναι πιο εύκολο να ερμηνεύσουμε το πίνακα αποτελεσμάτων.

1. **WordFreqMake:**
    - Είναι η συχνότητα εμφάνισης της λέξης "make" στο email. Πιο συγκεκριμένα, είναι το ποσοστό των λέξεων στο email που είναι "make". Αυτή η μεταβλητή βοηθά στην αξιολόγηση αν το συγκεκριμένο email περιέχει συγκεκριμένες λέξεις που είναι πιθανό να συσχετίζονται με spam.
2. **WordFreqAddress:**
    - Είναι η συχνότητα εμφάνισης της λέξης "address" στο email. Αντιπροσωπεύει το ποσοστό των λέξεων στο email που είναι "address". Αυτό είναι σημαντικό, καθώς ορισμένες λέξεις μπορεί να εμφανίζονται πιο συχνά σε spam emails.
3. **WordFreqAll:**
    - Είναι η συχνότητα εμφάνισης της λέξης "all" στο email. Υπολογίζεται ως το ποσοστό των λέξεων που είναι "all". Όπως και οι προηγούμενες μεταβλητές, μπορεί να βοηθήσει στον εντοπισμό χαρακτηριστικών που διαχωρίζουν τα spam από τα μη-spam emails.
4. **CharFreqDollar (Character Frequency Dollar):**
    - Είναι η συχνότητα εμφάνισης του χαρακτήρα "$" στο email. Αυτό υπολογίζεται ως το ποσοστό όλων των χαρακτήρων που είναι το σύμβολο "$". Δεδομένου ότι το σύμβολο αυτό χρησιμοποιείται συχνά σε spam emails, είναι ένας σημαντικός δείκτης στην ταξινόμηση.
5. **CapitalRunLengthAvg:**
    - Είναι το μέσο μήκος των διαδοχικών κεφαλαίων γραμμάτων (επομένως ο αριθμός των διαδοχικών κεφαλαίων χαρακτήρων) στο email. Αυτή η μεταβλητή μετρά πώς οι κεφαλαίοι χαρακτήρες εμφανίζονται συνεχόμενα μέσα στο κείμενο του email, και υψηλές τιμές μπορεί να υποδεικνύουν ότι το email είναι spam.

Συνολικά, αυτές οι μεταβλητές είναι χαρακτηριστικά που βοηθούν στον εντοπισμό μοτίβων και τάσεων που σχετίζονται με τα spam emails, ώστε να εκπαιδευτεί το Bayesian Network να τα αναγνωρίζει.

Η επιλογή των συγκεκριμένων μεταβλητών έγινε για τους ακόλουθους λόγους:

1. **Αντιπροσωπευτικότητα του περιεχομένου:**
   - Οι μεταβλητές WordFreqMake, WordFreqAddress, και WordFreqAll επιλέχθηκαν επειδή αντιπροσωπεύουν τη συχνότητα συγκεκριμένων λέξεων που μπορεί να εμφανίζονται συχνότερα σε spam emails. Λέξεις όπως "make," "address," και "all" είναι αρκετά κοινές σε spam emails που προσπαθούν να προωθήσουν προσφορές ή πληροφορίες.
2. **Ενδείξεις χαρακτηριστικών που σχετίζονται με spam:**
   - Η μεταβλητή CharFreqDollar επιλέχθηκε επειδή το σύμβολο "$" χρησιμοποιείται συχνά σε spam emails για να τραβήξει την προσοχή, ειδικά όταν γίνεται λόγος για χρήματα, προσφορές, ή οικονομικές συναλλαγές.
   - Οπότε, η παρουσία αυτής της μεταβλητής συμβάλλει στην ανίχνευση τέτοιων προωθητικών μηνυμάτων.
3. **Χρήση κεφαλαίων γραμμάτων ως δείκτης:**
   - Η μεταβλητή CapitalRunLengthAvg επιλέχθηκε διότι τα spam emails συχνά χρησιμοποιούν αλληλουχίες κεφαλαίων γραμμάτων για να τονίσουν λέξεις ή φράσεις (π.χ., "FREE," "LIMITED TIME OFFER"). Αυτή η συμπεριφορά είναι συχνότερη σε spam παρά σε κανονικά emails.
4. **Διαχωριστική Ικανότητα:**
   - Οι επιλεγμένες μεταβλητές είναι χρήσιμες για το διαχωρισμό των spam από τα μη-spam emails, καθώς τα χαρακτηριστικά που αντιπροσωπεύουν εμφανίζονται συχνότερα ή με διαφορετικό τρόπο στα spam emails.
5. **Απλότητα και μείωση της διάστασης του προβλήματος:**
   - Αντί να συμπεριλάβουμε όλες τις μεταβλητές του dataset, επιλέγονται αυτές που έχουν υψηλότερη πιθανότητα να συσχετίζονται άμεσα με το

αν ένα email είναι spam ή όχι. Αυτό βελτιώνει την απόδοση του μοντέλου, μειώνει τον χρόνο εκπαίδευσης και αποφεύγει το "θόρυβο" που μπορεί να εισάγουν λιγότερο σχετικές μεταβλητές.

Συνοψίζοντας, οι συγκεκριμένες μεταβλητές επιλέχθηκαν επειδή παρέχουν σημαντικές πληροφορίες που μπορούν να βοηθήσουν το μοντέλο να μάθει τα χαρακτηριστικά των spam emails με αποτελεσματικό τρόπο.

Το μαθηματικό μοντέλο που χρησιμοποιεί ο κώδικας έχει ως εξής:

Ορίζουμε τις μεταβλητές του μοντέλου με τα σύμβολα:

$$
\begin{aligned}
&X_1 : CapitalRunLengthAvg \\
&X_2 : CharFreqDollar \\
&X_3 : WordFreqAddress \\
&X_4 : WordFreqAll \\
&X_5 : WordFreqMake \\
&Y : Spam \ (\eta \ \varepsilon\xi\alpha\rho\tau\eta\mu\acute{\varepsilon}\nu\eta \ \mu\varepsilon\tau\alpha\beta\lambda\eta\tau\acute{\eta} \ \pi o\upsilon \ \pi\rho o\sigma\pi\alpha\theta o\acute{\upsilon}\mu\varepsilon \ \nu\alpha \ \pi\rho o\beta\lambda\acute{\varepsilon}\psi o\upsilon\mu\varepsilon)
\end{aligned}
$$

Οπότε η πιθανότητα να είναι Spam ή όχι, βάση των παραπάνω μεταβλητών εκφράζεται από τον ακόλουθο τύπο:

$$
P(Y \mid X_1, X_2, X_3, X_4, X_5) = \frac{P(Y = Spam)P(X_1 \mid Y = Spam)P(X_2 \mid Y = Spam)P(X_3 \mid Y = Spam)P(X_4 \mid Y = Spam)P(X_5 \mid Y = Spam)}{\sum_y P(Y = y)P(X_1 \mid Y = y)P(X_2 \mid Y = y)P(X_3 \mid Y = y)P(X_4 \mid Y = y)P(X_5 \mid Y = y)}
$$

Οποτε το Μπεϋζιανό Δίκτυο που αναλύουμε έχει τη δομή όπου οι πέντε μεταβλητές επηρεάζουν την πιθανότητα του Y=Spam. Με άλλα λόγια, το Spam εξαρτάται από αυτές τις μεταβλητές, και υπάρχει ένα σύνολο κατευθυνόμενων ακμών που δείχνει πώς αυτές οι μεταβλητές επηρεάζουν την πιθανότητα το μήνυμα να είναι Spam.

Στον πίνακα αποτελεσμάτων που δημιουργήσαμε, υπάρχουν διαφορετικοί συνδυασμοί των ανεξάρτητων μεταβλητών X_1 έως X_5. Ο πίνακας μας δίνει την πιθανότητα κάθε κατηγορίας του Spam (0 ή 1) για κάθε συνδυασμό των μεταβλητών.

Οπότε οι ερμηνεία αποτελεσμάτων του πίνακα 4.8 για την 1^{η} και την τελευταια γραμμή έχει ως εξής (Παρόμοια είναι και για τις υπόλοιπες γραμμές):

Όταν η CharFreqDollar =[0,2), δηλαδή όταν η συχνότητα εμφάνισης του συμβόλου "$" εμφανίζεται 0 έως 2 φορές, η WordFreqAddress=[0,4.76), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "address" βρίσκεται στο εύρος από 0 έως 4.76, η WordFreqAll=[0,1.7), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "all" βρίσκεται στο εύρος από 0 έως 1.7, η WordFreqMake=[0,1.51), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "make" βρίσκεται στο εύρος από 0 έως 1.51 και η CapitalRunLengthAvg=[1,368), δηλαδή Ο μέσος αριθμός διαδοχικών κεφαλαίων γραμμάτων βρίσκεται στο εύρος από 1 έως 368, τότε έχουμε ότι

$$P(Spam=0)=0.602$$
$$P(Spam=1)=0.399$$

Που σημαίνει ότι υπάρχει περίπου **60% πιθανότητα** το μήνυμα να **μην είναι Spam** και περίπου **40% πιθανότητα** να **είναι Spam**. Με άλλα λόγια, το μοντέλο είναι πιο πιθανό να θεωρήσει ότι το μήνυμα δεν είναι ανεπιθύμητο, αλλά υπάρχει ακόμα μια σημαντική πιθανότητα ότι μπορεί να είναι ανεπιθύμητο μήνυμα.

Όταν η CharFreqDollar =[4,6), δηλαδή όταν η συχνότητα εμφάνισης του συμβόλου "$" εμφανίζεται 4 έως 6 φορές, η WordFreqAddress=[0,4.76), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "address" βρίσκεται στο εύρος από 0 έως 4.76, η WordFreqAll=[0,1.7), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "all" βρίσκεται στο εύρος από 0 έως 1.7, η WordFreqMake=[0,1.51), δηλαδή η μέση συχνότητα εμφάνισης της λέξης "make" βρίσκεται στο εύρος από 0 έως 1.51 και η CapitalRunLengthAvg=[1,368), δηλαδή Ο μέσος αριθμός διαδοχικών κεφαλαίων γραμμάτων βρίσκεται στο εύρος από 1 έως 368, τότε έχουμε ότι

$$P(Spam=0)=0$$
$$P(Spam=1)=1$$

Που σημαίνει ότι υπάρχει περίπου **0% πιθανότητα** το μήνυμα να **μην είναι Spam** και περίπου **100% πιθανότητα** να **είναι Spam**. Με άλλα λόγια, το μοντέλο είναι πιο πιθανό να θεωρήσει ότι το μήνυμα δεν είναι ανεπιθύμητο, αλλά υπάρχει ακόμα μια σημαντική πιθανότητα ότι μπορεί να είναι ανεπιθύμητο μήνυμα.

Οι κενές τιμές στον πίνακα υποδηλώνουν περιπτώσεις όπου ο συνδυασμός των συγκεκριμένων μεταβλητών δεν εμφανίζεται στα δεδομένα εκπαίδευσης του Μπεϋζιανού Δικτυού ή η πιθανότητα αυτού του συνδυασμού δεν ήταν σημαντική ώστε να υπολογιστεί από το μοντέλο.

Η ύπαρξη κενών τιμών μπορεί να υποδεικνύει ότι το μοντέλο χρειάζεται περισσότερα δεδομένα ή ότι ορισμένοι συνδυασμοί μεταβλητών είναι πολύ σπάνιοι για να έχουν πρακτική σημασία. Αυτό είναι συνηθισμένο στα Μπεϋζιανά Δίκτυα, όπου δεν παρατηρούνται όλοι οι συνδυασμοί κατά την εκπαίδευση του μοντέλου.

Στα αποτελέσματα του κώδικα θα παρατηρήσουμε και κενές τιμές οι οποιες αντικατοπτρίζουν την αβεβαιότητα ή την έλλειψη δεδομένων για συγκεκριμένους συνδυασμούς μεταβλητών, και γι' αυτό το μοντέλο δεν μπορεί να εκτιμήσει τις πιθανότητες για αυτές τις περιπτώσεις.

Το τελικό συμπέρασμα από τα αποτελέσματα του κώδικα είναι ότι το Μπεϋζιανό Δίκτυο έχει κατασκευαστεί με επιτυχία ώστε να αναλύει την πιθανότητα ένα μήνυμα να είναι Spam, λαμβάνοντας υπόψη πέντε βασικές μεταβλητές: τον μέσο αριθμό διαδοχικών κεφαλαίων γραμμάτων, τη συχνότητα εμφάνισης του χαρακτήρα "$", καθώς και τις συχνότητες εμφάνισης των λέξεων "address," "all," και "make."

Κύρια Σημεία Συμπερασμάτων:

1. **Οι περισσότερες μεταβλητές έχουν χαμηλές τιμές:** Οι πιθανότητες δείχνουν ότι οι περισσότερες μεταβλητές, όπως το CapitalRunLengthAvg και το CharFreqDollar, συνήθως εμφανίζονται σε χαμηλά επίπεδα στα μηνύματα. Αυτό σημαίνει ότι τα κανονικά μηνύματα τείνουν να έχουν λίγα διαδοχικά κεφαλαία γράμματα και σπάνια περιέχουν το σύμβολο "$."
2. **Σπάνιες τιμές αυξάνουν την πιθανότητα για Spam:** Όταν εμφανίζονται υψηλότερες τιμές σε αυτές τις μεταβλητές (π.χ., πολλές διαδοχικές κεφαλαίες λέξεις ή συχνή εμφάνιση του "$"), το μήνυμα γίνεται πιο ύποπτο ως Spam, καθώς αυτές οι τιμές είναι ασυνήθιστες σε κανονικά μηνύματα.
3. **Συνδυασμός πολλών παραγόντων:** Το δίκτυο δεν στηρίζεται μόνο σε μία μεταβλητή για να αποφασίσει αν ένα μήνυμα είναι Spam, αλλά αξιολογεί το συνδυασμό πολλών παραγόντων. Αυτό αυξάνει την ακρίβεια στην ανίχνευση Spam, καθώς λαμβάνει υπόψη διαφορετικές πτυχές των μηνυμάτων.

Συνολικό Συμπέρασμα

Το Μπεϋζιανό Δίκτυο αποτυπώνει με επιτυχία το πώς συγκεκριμένα χαρακτηριστικά ενός μηνύματος επηρεάζουν την πιθανότητα να είναι Spam. Η πλειονότητα των μηνυμάτων παρουσιάζει χαμηλές τιμές στις αναλυόμενες μεταβλητές, αλλά όταν οι τιμές αυξάνονται, ενισχύεται η υπόθεση ότι το μήνυμα είναι ανεπιθύμητο. Έτσι, το μοντέλο προσφέρει ένα αποτελεσματικό εργαλείο για τον εντοπισμό των Spam μηνυμάτων, βασισμένο σε έναν συνδυασμό πιθανοτήτων που αναδεικνύουν τις διαφορές μεταξύ κανονικών και ανεπιθύμητων μηνυμάτων.

Στο Κεφάλαιο αυτό είδαμε 3 εφαρμογές στην γλώσσα πραγματισμού R δυο προσομοίωσης και μια με πραγματικά δεδομένα ώστε να δούμε πως δουλεύουν τα Μπεϋζιανά Δίκτυα σε εφαρμογές άλλον επιστημονικών κλάδων και μπορούν να τους βοηθήσουν στην ερευνά τους.

ΚΕΦΑΛΑΙΟ 5
Συμπεράσματα και συζήτηση

Στην παρούσα διπλωματική εργασια παρουσιάστηκε η θεωρία για την Μπεϋζιανή Συμπερασματολογία Δικτύων και αναλύθηκαν όλα τα επιμέρους μέρη της θεωρίας. Επιπλέον, δόθηκαν ενδεικτικά παραδείγματα προκειμένου να εφαρμοστεί η μεθοδολογία των Bayesian Network.

Τα παραδείγματα αυτά έδειξαν τη συμπεριφορά των μεταβλητών και την ικανότητα των αλγορίθμων να εκτιμούν τη Δομή των παραμέτρων του δικτύου στην περίπτωση προσομοιωμένων δεδομένων καθώς και στην περίπτωση πραγματικών δεδομένων.

Η θεωρία των δικτύων υλοποιήθηκε στην R με τη χρήση των δύο ενδεικτικών παραδειγμάτων και μιας πλήρους παρουσίασης της συντριπτικής πλειοψηφίας των δυνατοτήτων του πακέτου bnlearn το οποίο είναι το κύριο εργαλείο για την εξαγωγή συμπερασμάτων σε δίκτυα Bayes στην R.

Ασχοληθήκαμε με την υλοποιήση εναλλακτικών λύσεων της Αξιολόγησης Bayesian Networks και με τον τρόπο που μπορούν να μας βοηθήσουν σε πολλά προβλήματα της καθημερινής ζωής αλλά και σε άλλες επιστήμες. Η ανάγκη για εναλλακτικές μεθόδους αξιολόγησης προκύπτει από το γεγονός ότι, όπως είδαμε στα παραδείγματα, οι συναρτήσεις βαθμολογίας πέφτουν μερικές φορές θύματα των δομών με περισσότερα τόξα και επιλέγουν λανθασμένα ένα Δίκτυο. Μέσω αυτής της παρουσίασης δείξαμε επίσης, τους τρόπους με τους οποίους η Θεωρία της Πληροφορίας μπορεί να συνδυαστεί με τα Bayesian Networks προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα.

Παράρτημα

Κώδικες

Για την επίλυση κάποιόν παραδειγμάτων χρησιμοποιήθηκε η γλώσσα προγραμματισμού R.

Βιβλιοθήκη

```
install.packages("bnlearn")
library (bnlearn)
```

Δημιουργία Μπεϋζιανού Δικτύου

```
randomgraph<-random.graph(c("Symptom1",
"Symptom2","Symptom3","Symptom4","Symptom5","Diagnosis"),method="mela
ncon",max.degree=3)
randomgraph
```

Δημιουργία γραφήματος

```
plot(randomgraph)
```

Δημιουργία Μπεϋζιανών παραμέτρων

```
dfit<-
custom.fit(net,dist=list(Diagnosis=cptD,Symptom5=cptS5,Symptom4=cptS4
,Symptom2=cptS2,Symptom1=cptS1,Symptom3=cptS3))
>dfit
```

Αφαίρεση και πρόσθεσή τόξων για την εύρεση καλύτερου μοντέλου

```
># remove 1 arc
>new.net<-drop.arc(from='Symptom2',to="Symptom3",net)
>new.net

># remove 1 arc + put a random arc
>new.net1<-set.arc(from='Symptom5',to="Symptom2",new.net)
>new.net1


># remove 2 arcs + put a random arc
>new.net2<-drop.arc(from='Symptom2',to="Symptom4",new.net1)
>new.net2
```

Αποτελέσματα για την καταλληλόλητα του μοντέλου

```
score(net,rbn_df,type='bic')# BIC criterion
score(net,rbn_df,type='aic')# AIC criterion
score(net,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(net,rbn_df,type='loglik')# Log-Likelihood

score(new.net,rbn_df,type='bic')# BIC criterion
score(new.net,rbn_df,type='aic')# AIC criterion
score(new.net,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net,rbn_df,type='loglik')# Log-Likelihood

score(new.net1,rbn_df,type='bic')# BIC criterion
score(new.net1,rbn_df,type='aic')# AIC criterion
score(new.net1,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net1,rbn_df,type='loglik')# Log-Likelihood

score(new.net2,rbn_df,type='bic')# BIC criterion
score(new.net2,rbn_df,type='aic')# AIC criterion
score(new.net2,rbn_df,type='bde')# Bayecian Dirichlet equivalent
score(new.net2,rbn_df,type='loglik')# Log-Likelihood
```

Αποτελέσματα κώδικα Spam email με WordCount και WordFrequency

```
library(bnlearn)
# Δημιουργία δεδομένων
set.seed(123)
data <- data.frame(
+   WordCount = rpois(100, lambda = 30),
+   WordFrequency = rnorm(100, mean = 10, sd = 5),
+   Spam = sample(c(0, 1), 100, replace = TRUE)
+ )
# Μετατροπή της συνεχούς μεταβλητής "WordCount" σε διακριτή μεταβλητή
data$WordCount<- cut(data$WordCount, breaks = 3)  # Επιλέξτε τον
αριθμό των κατηγοριών που θέλετε
# Ορισμόςδομήςδικτύου
network_structure<-
model2network("[WordCount][WordFrequency][Spam|WordCount:WordFrequenc
y]")
# Προσαρμογή του δικτύου στα δεδομένα
fitted_network<- bn.fit(network_structure, data)
# Εκτύπωση της δομής του δικτύου
print(network_structure)
# Εκτύπωση των παραμέτρων του δικτύου
print(fitted_network)
```

Κωδικας accuracy measure με Cross-Validation(CV)

```
# Φόρτωση των απαραίτητων βιβλιοθηκών
library(caret)

# Δημιουργία δεδομένων
set.seed(123)
data <- data.frame(
  WordCount = rpois(100, lambda = 30),
  WordFrequency = rnorm(100, mean = 10, sd = 5),
```

```
  Spam = factor(sample(c(0, 1), 100, replace = TRUE))
)

# Μετατροπή της συνεχούς μεταβλητής "WordCount" σε διακριτή μεταβλητή
data$WordCount <- cut(data$WordCount, breaks = 3, labels = c(0, 1,
2))

# Ορισμός του μοντέλου λογιστικής παλινδρόμησης
model <- train(Spam ~ WordCount + WordFrequency,
               data = data,
               method = "glm",
               family = "binomial",
               trControl = trainControl(method = "cv", number = 5))

# Εμφάνιση των αποτελεσμάτων
print(model)
```

Κώδικάς για παράδειγμα Spam email με πραγματικά δεδομένα:

```
install.packages("mlbench")
install.packages("kernlab")
install.packages("arules")  # Για τη συνάρτηση discretize
library(bnlearn)
library(mlbench)
library(kernlab)
library(arules)  # Περιέχει τη συνάρτηση discretize
# Φόρτωση του dataset "spam" από τη βιβλιοθήκη kernlab
data(spam)

# Προβολή των ονομάτων των στηλών για επιβεβαίωση
names(spam)
# Επιλογή των μεταβλητών που θα χρησιμοποιήσουμε
data_subset <- spam[, c('make', 'address', 'all', 'charDollar',
'capitalAve', 'type')]
# Μετονομασία των στηλών για ευκολία
colnames(data_subset) <- c('WordFreqMake', 'WordFreqAddress',
'WordFreqAll',
                           'CharFreqDollar', 'CapitalRunLengthAvg',
'Spam')
# Μετατροπή της μεταβλητής 'Spam' σε παράγοντα με επίπεδα '0' και '1'
data_subset$Spam <- ifelse(data_subset$Spam == 'spam', 1, 0)
data_subset$Spam <- as.factor(data_subset$Spam)
# Διακριτοποίηση των συνεχών μεταβλητών σε 3 κατηγορίες
data_discrete <- data_subset
data_discrete[, -6] <- lapply(data_subset[, -6], function(x)
discretize(x, method = 'interval', breaks = 3))
# Ορισμός της δομής του δικτύου
model_string <-
"[WordFreqMake][WordFreqAddress][WordFreqAll][CharFreqDollar][Capital
RunLengthAvg][Spam|WordFreqMake:WordFreqAddress:WordFreqAll:CharFreqD
ollar:CapitalRunLengthAvg]"

# Μετατροπή του μοντέλου σε δίκτυο
network_structure <- model2network(model_string)
# Προσαρμογή του δικτύου στα διακριτοποιημένα δεδομένα
fitted_network <- bn.fit(network_structure, data_discrete)
# Εκτύπωση της δομής του δικτύου
print(network_structure)

# Εκτύπωση των παραμέτρων του προσαρμοσμένου δικτύου
print(fitted_network)
```

Βιβλιογραφία

[1].Richard E. Neapolitan (2003,January) . *Learning Bayesian Networks*

[2].Bengio, S., & Bengio, Y. (2000). *Taking on the curse of dimensionality in joint distributions using neural networks. IEEE Transactions on Neural Networks, 11(3), 550- 557*

[3].Dawid, A. P., &Studen´y, M. (1999, January). *Conditional products: An alternative approach to conditional independence. In Seventh International Workshop on Artificial Intelligence and Statistics (pp. 27-35). PMLR.*

[4]. Frey, B. J., Brendan, J. F., & Frey, B. J. (1998). *Graphical models for machine learning and digital communication. MIT press.*

[5]. Galan, S. F., Aguado, F., Dıez, F. J., & Mira, J. (2002). *NasoNet, modeling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. Artificial Intelligence in Medicine, 25(3), 247-264.*

[6]. Kasza, J., & Solomon, P. (2015). *Comparing score-based methods for estimating Bayesian networks using the Kullback–Leibler divergence. Communications in Statistics-Theory and Methods, 44(1), 135-152.*

[7]. Moral, S., Cano, A., &G´omez-Olmedo, M. (2021). *Computation of Kullback–Leibler Divergence in Bayesian Networks. Entropy, 23(9), 1122*.

[8]. Scutari, M., & Denis, J. B. (2021). *Bayesian networks: with examples in R. CRC press.*

[9]. Scutari, M. (2017). *Understanding Bayesian Networks with examples in R. University Lecture.*

[10]. Horn´y, M. (2014). *Bayesian networks. Boston University School of Public Health, 17.*

[11].van Koten, C. and Gray, A. R. (2005). *An application of bayesian network for predicting object-oriented software maintainability. In Information and Software Technology Volume 48.*

[12]. Stephenson, T. A. (2000). *An introduction to bayesian network theory and usage. In IDIAP Research Report.*

[13]. Niedermayer, D. (2008). *An introduction to bayesian network theory and their contemporary applications. In Innovations in Bayesian Networks Volume 156.*

[14]. Nikovski, D. (2000). *Constructing bayesian networks for medical diagnosis from incomplete and partially correct statistics. IEEE Transactions on Knowledge and Data Engineerin*

[15]. Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

[16]. Korb, K. B., & Nicholson, A. E. (2010). Bayesian artificial intelligence. CRC press.

[17]. Neapolitan, R. E. (2004). Learning bayesian networks (Vol. 38). Upper Saddle River, NJ: Pearson Prentice Hall.

[18]. Margaritis Dimitris (2003). Learning Bayesian Network Model Structure from Data. PHD Thesis. University Pittsburgh.

[19]. Tsamardinos Ioannis (2003). Algorithms for Large Scale Markov Blanket Discovery. DBLP

[20]. Yaramakala Sandeep & Margaritis Dimitris (2005). Speculative Markov Blanket Discovery for Optimal Feature Selection. ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining.Pages 809 - 812

[21]. Kalisch Markus $ Buhlmann Peter (2007). Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm. Journal of Machine Learning Research 8 (2007) 613-636

[22]. Kalisch Markus $ Buhlmann Peter (2008). Robustification of the PC-Algorithm for Directed Acyclic Graphs. Journal of Computational and Graphical Statistics.

[23]. Buhlmann Peter (2010). Updates on the prevalence of body dysmorphic disorder: a population-based survey. Psychiatry Res.

[24]. Agresti Alan (2013). Categorical data analysis. 3rd Edition, John Wiley & Sons.

[25]. Hausser, J. and Strimmer, K. (2009) Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. The Journal of Machine Learning Research, 10, 1469-1484.

[26]. https://www.bnlearn.com/ (Package, Title: Bayesian Network Structure Learning, Parameter Learning and Inference)

[27]. https://cran.r-project.org/web/packages/kernlab/kernlab.pdf (Kernel-Based Machine Learning Lab).

[28]. https://cran.r-project.org/web/packages/arules/arules.pdf (Mining Association Rules and Frequent Itemsets)

[29]. https://cran.r-project.org/web/packages/mlbench/mlbench.pdf (Machine Learning Benchmark Problems)