

Επίλυση του Προβλήματος Εναλλαγής Ετικετών στην Μπεϋζιανή Ανάλυση Μείξεων Κατανομών

Παναγιώτης Παπασταμούλης

Διδακτορική Διατριβή

Πανεπιστήμιο Πειραιώς
Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

Επιβλέπων: Αναπληρωτής Καθηγητής Γιώργος Ηλιόπουλος

Πειραιάς 2010

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον δάσκαλό μου, Αναπληρωτή Καθηγητή κ. Γιώργο Ηλιόπουλο, για την εμπιστοσύνη που μου έδειξε. Η υποδειγματική καθοδήγησή του μου έδινε την ελευθερία να εξερευνώ νέα μονοπάτια αλλά και να μαθαίνω από τα λάθη μου, ενώ ταυτόχρονα με κρατούσε στην σωστή πορεία. Τον ευχαριστώ επίσης για την βοήθειά του στο να ξεπεραστούν οι ποικίλες δυσκολίες που προέκυψαν σε κρίσιμα σημεία της διατριβής μου.

Ευχαριστώ επίσης τον Καθηγητή κ. Μάρκο Κούτρα και τον Αναπληρωτή Καθηγητή κ. Ευστάθιο Χατζηκωνσταντινίδη για την συμμετοχή τους στην τριμελή επιτροπή μου, καθώς επίσης και τα υπόλοιπα μέλη της επταμελούς επιτροπής, Καθηγητή κ. Π. Δελλαπόρτα, Καθηγητή κ. Γ. Κοκολάκη, Επίκουρο Καθηγητή κ. Δ. Καρλή, Επίκουρο Καθηγητή κ. Ι. Ντζούφρα.

Επί πλέον, ευχαριστώ το Ίδρυμα Κρατικών Υποτροφιών (Ι.Κ.Υ) για την οικονομική ενίσχυση που μου παρείχε κατά τα τρία πρώτα έτη εκπόνησης της διατριβής μου.

Όλα αυτά τα χρόνια είχα ως στήριγμα την ανεκτίμητη βοήθεια και ενθάρρυνση των γονιών μου, Βασίλη και Χρυσούλας Παπασταμούλη. Το παρόν σύγγραμμα είναι αφιερωμένο σε αυτούς, σαν ελάχιστη ανταπόδοση για την αμέριστη συμπαράστασή τους.

Παναγιώτης Παπασταμούλης

Πειραιάς, 2010

РАНЕКЪМЪО РЕПАА

Περιεχόμενα

1	Εισαγωγή	1
2	Συμπερασματολογία σε Μοντέλα Μείξεων Κατανομών	7
2.1	Εισαγωγή	7
2.2	Στοιχεία πεπερασμένων μίξεων κατανομών	8
2.2.1	Η διατύπωση του μοντέλου μέσω ελλειπών δεδομένων	12
2.2.2	Μοντελοποιώντας άγνωστες κατανομές	14
2.3	Ο αλγόριθμος EM	17
2.4	Η Μπεϋζιανή προσέγγιση	20
2.5	Μέθοδοι MCMC και μίξεις κατανομών	23
2.5.1	Δειγματολήπτης Gibbs	24
2.5.2	Αλγόριθμος Metropolis-Hastings	27
2.5.3	Αλγόριθμος RJMCMC για μίξεις κανονικών κατανομών	28
2.6	Το πρόβλημα label switching	34
2.6.1	Εικονικοί περιορισμοί διάκρισης	38
2.6.2	Αλγόριθμος Πιλοτικής Αναδιάταξης	39
2.6.3	Αλγόριθμος αναδιάταξης μέσω της απόκλισης Kullback-Leibler	40
2.6.4	Αναδιάταξη μέσω συναρτήσεων ζημίας	42
2.7	Ένα μοντέλο εκτίμησης μίξεων πολυμεταβλητών παλινδρομήσεων	43
2.7.1	Παραδείγματα	47
3	RJMCMC σε Μίξεις Κανονικών Κατανομών με Κοινές Μέσες Τιμές	57
3.1	Εισαγωγή	57
3.2	Βασικοί ορισμοί	58
3.3	Το μπεϋζιανό ιεραρχικό μοντέλο	60

3.4	Ο προτεινόμενος αλγόριθμος RJMCMC	62
3.4.1	Οι τροποποιημένες κινήσεις αλλαγής αριθμού συνιστωσών	63
3.4.2	Ο δειγματολήπτης Gibbs	64
3.4.3	Η κίνηση διαχωρισμού-συνδυασμού των μέσων	65
3.5	Εφαρμογές	68
3.5.1	Προσομοιωμένο σύνολο από την κατανομή (2.31)	70
3.5.2	Έκτροπες παρατηρήσεις	74
3.5.3	Γαλαξιακό σύνολο δεδομένων	75
3.5.4	Εφαρμογή σε μη κανονικά δεδομένα	79
3.6	Σύνοψη και συμπεράσματα	81
4	Επίλυση του προβλήματος Label Switching	85
4.1	Εισαγωγή	85
4.2	Ο αλγόριθμος ECR	86
4.2.1	Η εκ των υστέρων κατανομή ως ισοβαρής μείξη μη συμμετρικών κατανομών	88
4.2.2	Επιλογή του συνόλου αντιπροσώπων	95
4.2.3	Μία αναλυτική εφαρμογή της προτεινόμενης μεθόδου	97
4.2.4	Αναδιατάσσοντας ένα προσομοιωμένο δείγμα	100
4.3	Ταχύτητα σύγκλισης του αλγορίθμου	101
4.4	Ο αλγόριθμος ECR και το πρόβλημα εκχώρησης	109
5	Εφαρμογές του Αλγορίθμου ECR	111
5.1	Εισαγωγή	111
5.2	Εφαρμογές σε κλασικά μοντέλα μείξεων κατανομών	112
5.2.1	Αλγόριθμος ECR έναντι Αλγορίθμου Πιλοτικής Αναδιάταξης	113
5.2.2	Αλγόριθμος ECR έναντι αλγορίθμων KL: Γαλαξιακό σύνολο δεδομένων	116
5.2.3	Μείξεις Πολυμεταβλητών Κανονικών Κατανομών	120
5.2.4	Μείξεις πολυμεταβλητών παλινδρομήσεων	123
5.2.5	Εφαρμογή σε έναν μη διευρυμένο αλγόριθμο	125
5.3	Εφαρμογές σε Hidden Markov Models	127

5.3.1	Εισαγωγή	127
5.3.2	Συμβολισμοί και το μπεϋζιανό μοντέλο	129
5.3.3	Το φαινόμενο label switching στα HMM	132
5.3.4	Αυτοπαλίνδρομο Μοντέλο Μαρκοβιανών Εναλλαγών	133
5.3.5	Μη ομογενής διαδικασία Poisson διακριτού χρόνου	137
5.3.6	Μη συζυγή HMM	140
5.4	Εφαρμογές σε Τυχαία Μαρκοβιανά Πεδία	151
5.4.1	Μπεϋζιανή κατάτμηση εικόνων	156
A	Παράρτημα Κεφαλαίου 2	161
A.1	Η γενική μορφή του αλγορίθμου RJMCMC	161
A.2	Απόδειξη Λήμματος 2.7.1	163
B	Παράρτημα Κεφαλαίου 3	165
B.1	Απόδειξη σχέσης (3.4)	165
B.2	Απόδειξη Λήμματος 3.4.1	166
B.3	Απόδειξη Πρότασης 3.4.1	166
B.4	Κανονικές μείξεις με κοινές μέσες τιμές έναντι μείξεων κατανομών t_4 στο γαλαξιακό σύνολο δεδομένων	169
Γ	Παράρτημα Κεφαλαίου 4	173
Γ.1	Απόδειξη Πρότασης 4.2.2	173
Γ.2	Εξασφάλιση της (4.21)	175
	Βιβλιογραφία	179

ΓΑΛΕΡΙΣΤΗΜΟ ΠΕΡΑΙΑ

Κεφάλαιο 1

Εισαγωγή

Οι πεπερασμένες μείξεις κατανομών αποτελούν μια μαθηματική προσέγγιση για τη στατιστική μοντελοποίηση ενός μεγάλου πλήθους τυχαιών φαινομένων. Κατά τις δύο τελευταίες δεκαετίες, τα συγκεκριμένα μοντέλα επικεντρώνουν το ενδιαφέρον αρκετών ερευνητών τόσο σε πρακτικό όσο και θεωρητικό επίπεδο. Οι μείξεις κατανομών παρέχουν το κατάλληλο στατιστικό εργαλείο για την περιγραφή διαφορετικών χαρακτηριστικών που εμφανίζονται σε ετερογενή σύνολα δεδομένων. Εφαρμογές των μοντέλων μείξεων κατανομών απαντούν σε αρκετά επιστημονικά πεδία όπως η Αστρονομία, Βιολογία, Γενετική, Ιατρική, Ψυχολογία, Οικονομία, Μηχανική, Κοινωνιολογία, Τεχνητή Νοημοσύνη, Πληροφορική, Ρομποτική κ.α. Σε αυτές τις εφαρμογές, οι μείξεις κατανομών υποστηρίζουν μια πληθώρα στατιστικών τεχνικών, όπως η ανάλυση συστάδων (cluster analysis), η διαχωριστική ανάλυση (discriminant analysis), η ανάλυση ψηφιακών εικόνων και η ανάλυση επιβίωσης. Πιο άμεσος είναι ο ρόλος τους στην ανάλυση δεδομένων και στην συμπερασματολογία παρέχοντας περιγραφικά μοντέλα για διάφορες «μη τυπικές» κατανομές.

Η εφαρμογή μπεϋζιανών μεθόδων σε προβλήματα μείξεων κατανομών αποτελεί αντικείμενο πολλών ερευνητικών εργασιών. Η Μπεϋζιανή Συμπερασματολογία επιτρέπει την εξαγωγή πιθανοθεωρητικών συμπερασμάτων σχετικά με τις άγνωστες παραμέτρους του μοντέλου, την ενσωμάτωση σε αυτό εκ των προτέρων γνώσης, και την ιεραρχική περιγραφή τόσο των τοπικών όσο και των συνολικών χαρακτηριστικών του. Το πλαίσιο αυτό καθιστά δυνατή την αποσύνθεση της πολύπλοκης δομής ενός μοντέλου μείξης σε απλούστερες δομές μέσω της χρήσης εικονικών (μη παρατηρήσιμων) μεταβλητών κατάταξης. Όταν ο αριθμός των συνιστωσών που αποτελούν την μείξη είναι άγνωστος, η Μπεϋζιανή Συμπερασματολογία αποτελεί και την μοναδική δαισθητικά λογική προσέγγιση για την εκτίμηση αυτού (Richardson and Green, 1997).

Η μπεϋζιανή συμπερασματολογία σε προβλήματα μείξεων κατανομών άρχισε να γίνεται δημοφιλής μετά την σημαντική εργασία των Gelfand and Smith (1990) η οποία εισήγαγε στη στατιστική κοινότητα τον δειγματολήπτη Gibbs ως μια γενική μέθοδο προσομοίωσης από μία κατανομή-στόχο. Πριν την διάδοση των μεθόδων Markov Chain Monte Carlo (MCMC) δεν υπήρχε κάποια ικανοποιητική τεχνική για τον υπολογισμό των εκτιμητών Bayes. Τα μοντέλα μείξεων συνέβαλαν αρκετά στην εξάπλωση των μεθόδων αυτών, καθώς δεν αποτελεί σύμπτωση το γεγονός ότι ο δειγματολήπτης Gibbs για την εκτίμηση των παραμέτρων μείξεων κατανομών είχε προταθεί λίγο πριν (Tanner and Wong, 1987) και αμέσως μετά (Diebolt and Robert, 1990) από την εργασία των Gelfand and Smith (1990). Για την περίπτωση όπου ο αριθμός συνιστωσών θεωρείται άγνωστος, η πιο σημαντική εργασία ήταν αυτή των Richardson and Green (1997) όπου εφάρμοσαν την μέθοδο Reversible Jump MCMC (RJMCMC) του Green (1995) για την εκτίμηση του αριθμού συνιστωσών μιας μείξης κανονικών κατανομών.

Παρά την πληθώρα των τεχνικών MCMC που έχουν αναπτυχθεί για την προσομοίωση ενός δείγματος με στόχο την εκ των υστέρων κατανομή ενός μοντέλου μείξης, η μπεϋζιανή εκτίμηση των παραμέτρων του δεν αποτελεί τετριμμένη διαδικασία. Πιο συγκεκριμένα, σε τυπικές περιπτώσεις, οι παράμετροι ενός μοντέλου μείξης δεν είναι περιθωριακά ταυτοποιήσιμες, ανεξαρτήτως αν ο αριθμός των συνιστωσών θεωρείται γνωστός ή όχι. Το φαινόμενο που ευθύνεται για αυτό μεταφράζεται ως «εναλλαγή ετικετών». Σε ό,τι ακολουθεί χρησιμοποιείται ο όρος “label switching” που προτάθηκε από τους Redner and Walker (1984). Συνήθως το πρόβλημα αυτό παρακάμπτεται μέσω της εισαγωγής διάφορων εικονικών περιορισμών στον παραμετρικό χώρο. Παρ’ όλ’ αυτά, υπάρχει και μια πληθώρα εναλλακτικών αλλά πιο πολύπλοκων και δύσκολων μεθόδων που αντιμετωπίζουν το πρόβλημα αυτό πιο αποδοτικά (Jasra et al., 2005).

Στόχος και περίγραμμα της διατριβής

Ο πρωταρχικός στόχος της διατριβής είναι η θεμελίωση μιας γενικής μεθόδου αντιμετώπισης του προβλήματος label switching η οποία να είναι ανεξάρτητη της διάστασης των παραμέτρων του μοντέλου μείξης. Παρά την πληθώρα των μεθόδων που υπάρχουν στην βιβλιογραφία για την αντιμετώπιση του προβλήματος, καμία δεν είναι ταυτόχρονα απλή αλλά και αποδοτική ώστε να εφαρμοστεί σε ένα γενικό πλαίσιο. Μέσω της μελέτης του συνόλου των εικονικών μεταβλητών κατάταξης, κατασκευάζονται υποσύνολα αυτού που οδηγούν σε κλάσεις μη συμμετρικών εκ των

υστέρων κατανομών οι οποίες έχουν την ιδιότητα να αναπαράγουν την συμμετρική εκ των υστέρων κατανομή μέσω μεταθέσεων. Η μέθοδος αναπτύσσεται σε ένα ευρύτερο πλαίσιο μοντέλων ελλিপών δεδομένων με την προϋπόθεση ότι πληρούνται κάποιες αρκετά γενικές ιδιότητες. Τα μοντέλα αυτά περιλαμβάνουν τόσο τις κλασσικές μείξεις κατανομών όσο και τα πιο γενικά Hidden Markov Models (HMM) και Markov Random Fields (MRF). Η προτεινόμενη μέθοδος δύναται να θεωρηθεί ως μία γενική τεχνική αντιμετώπισης του προβλήματος label switching, η οποία είναι ανεξάρτητη από την διάσταση του παραμετρικού χώρου και είναι αρκετά απλή στην εφαρμογή της. Επί πλέον, η θεμελίωση της μεθόδου παρέχει ένα ισχυρό θεωρητικό υπόβαθρο που κάνει δυνατή την μελέτη περαιτέρω ιδιοτήτων όπως η ταχύτητα σύγκλισης του αλγορίθμου, κάτι που δεν είναι άμεσο για τις υπάρχουσες μεθόδους αντιμετώπισης του φαινομένου label switching.

Αρχικά, η μελέτη των επιπτώσεων που συνεπάγεται ένας περιορισμός διάκρισης για την ταυτοποίηση των παραμέτρων οδήγησε στην αναθεώρηση του αλγορίθμου RJMCMC για την εκτίμηση του αριθμού συνιστωσών μιας μείξης κανονικών κατανομών. Οι κινήσεις που είναι υπεύθυνες για την ανανέωση του αριθμού συνιστωσών εκμεταλλεύονται την εκ των προτέρων διάταξη των μέσων τιμών των συνιστωσών του μοντέλου διότι έτσι είναι πιο εύκολος ο σχεδιασμός αυτών των μεταβάσεων. Παρ' όλ' αυτά, όταν η περιοχή υψηλής εκ των υστέρων πυκνότητας βρίσκεται κοντά στο όριο του περιορισμού, η συγκεκριμένη υπόθεση έχει αρκετές επιπτώσεις. Από την μία πλευρά, ο αλγόριθμος δεν εκτιμά σωστά τον αριθμό των συνιστωσών και κατ' επέκταση το ίδιο ισχύει και για τις εκτιμήσεις των παραμέτρων. Κρίνεται λοιπόν απαραίτητη η προσαρμογή του αλγορίθμου αυτού ώστε να βελτιώνονται οι εκτιμήσεις σε τέτοιου είδους περιπτώσεις. Αυτό επιτυγχάνεται θεωρώντας ότι μοντέλα με κοινές μέσες τιμές έχουν θετική εκ των προτέρων πιθανότητα. Αφού οριστούν κατάλληλες κινήσεις που γεφυρώνουν αυτά τα μοντέλα, κατασκευάζεται ένας αλγόριθμος RJMCMC ο οποίος σε συνδυασμό με τις κινήσεις των Richardson and Green δρα στον χώρο των μοντέλων αυτών.

Τα κεφάλαια της διατριβής έχουν οργανωθεί ως εξής:

Στο Κεφάλαιο 2 δίνονται οι απαραίτητοι ορισμοί για τα κλασσικά μοντέλα μίξεων κατανομών καθώς και η θεώρησή τους ως μοντέλα ελλিপών δεδομένων. Στη συνέχεια γίνεται μια σύντομη περιγραφή του αλγορίθμου Expectation - Maximization (EM) της Κλασσικής Στατιστικής. Ακολούθως γίνεται μια ανασκόπηση των γνωστών τεχνικών MCMC για την εκτίμηση ενός μοντέλου μίξης θεωρώντας γνωστό ή άγνωστο τον αριθμό συνιστωσών (δειγματολήπτης Gibbs, αλγόριθμος Metropolis-Hastings

και RJMCMC), όπου διαπιστώνονται τα προβλήματα που αντιμετωπίζει η μέθοδος RJMCMC σε συγκεκριμένες περιπτώσεις. Κατόπιν περιγράφεται το φαινόμενο label switching και γίνεται μια ανασκόπηση των διαθέσιμων μεθοδολογιών αντιμετώπισής του. Το κεφάλαιο ολοκληρώνεται με τη μελέτη των μείξεων κατανομών σε μοντέλα γραμμικής παλινδρόμησης. Αφού εισαχθεί ένα μοντέλο εκτίμησης μείξεων πολυμεταβλητών παλινδρομήσεων, δίνονται μερικά παραδείγματα που υπογραμμίζουν τα προβλήματα που δημιουργεί το φαινόμενο label switching.

Στο Κεφάλαιο 3 περιγράφεται μία νέα θεώρηση στην μοντελοποίηση μείξεων κανονικών κατανομών κατά την οποία συμπεριλαμβάνονται μοντέλα με κοινές μέσες τιμές. Έτσι, ο αλγόριθμος RJMCMC δεν επηρεάζεται από τον τεχνητό περιορισμό διάταξης των μέσων για την ταυτοποίηση των παραμέτρων. Για τον σκοπό αυτό, γίνεται μια αναθεώρηση του ιεραρχικού μοντέλου που ορίζεται από τους Richardson and Green (1997) ώστε να έχουν θετική εκ των προτέρων πιθανότητα μείξεις με κοινές μέσες τιμές. Αφού οριστούν μοντέλα με διαφορετικό πλήθος παραμέτρων για κάθε αριθμό συνιστωσών, στην συνέχεια εισάγεται μία νέα κίνηση που γεφυρώνει τα μοντέλα αυτά. Τελικά, κατασκευάζεται ένας αλγόριθμος RJMCMC για την προσομοίωση μιας μαρκοβιανής αλυσίδας με στάσιμη κατανομή την αντίστοιχη από κοινού εκ των υστέρων κατανομή. Μέσω των παραδειγμάτων που παρουσιάζονται γίνεται σαφές ότι βελτιώνεται η εκτίμηση αρκετών μείξεων κατανομών ή μοντέλων που μπορούν να προσεγγιστούν από μείξεις κανονικών κατανομών όπου ο αρχικός αλγόριθμος δεν δίνει ικανοποιητικά αποτελέσματα.

Στο Κεφάλαιο 4 μελετάται το πρόβλημα label switching σε ένα γενικό πλαίσιο μοντέλων ελλিপών δεδομένων. Σύμφωνα με το βασικό θεώρημα που αποδεικνύεται εκεί, η συμμετρική εκ των υστέρων κατανομή γράφεται ως μείξη μη συμμετρικών κατανομών οι οποίες έχουν το ίδιο στήριγμα με αυτή. Η αναπαράσταση της εκ των υστέρων κατανομής κατ' αυτόν τον τρόπο οδηγεί στην επίλυση του προβλήματος μη ταυτοποίησης επιλέγοντας μία από τις συνιστώσες της μείξης. Ακολούθως προτείνεται ένας αλγόριθμος αναδιάταξης προσομοιωμένων δειγμάτων από την εκ των υστέρων κατανομή. Αποδεικνύεται ότι ο προτεινόμενος αλγόριθμος προσομοιώνει μια μαρκοβιανή αλυσίδα η οποία συγκλίνει στη νέα κατανομή-στόχο. Επί πλέον, μελετώνται οι ιδιότητες της ταχύτητας σύγκλισης του προτεινόμενου αλγορίθμου ως προς τη νόρμα ολικής μεταβολής. Μεταξύ άλλων, αποδεικνύεται ότι η ταχύτητα σύγκλισης της αναδιατεταγμένης ακολουθίας είναι τουλάχιστον ίση με την αντίστοιχη της αρχικής ακολουθίας.

Στο Κεφάλαιο 5 εφαρμόζεται ο προτεινόμενος αλγόριθμος για την επίλυση του

label switching στις εξής κατηγορίες μοντέλων ελλειπών δεδομένων: κλασσικά μοντέλα μείξεων κατανομών και παλινδρομήσεων, hidden Markov models και Markov random fields. Χρησιμοποιείται μια πληθώρα παραδειγμάτων πραγματικών ή προσομοιωμένων δεδομένων (μονοδιάστατων ή πολυμεταβλητών), τα οποία καλύπτουν ένα ευρύ φάσμα εφαρμογών. Ταυτόχρονα η προτεινόμενη μέθοδος συγκρίνεται με τις ήδη υπάρχουσες και αναδεικνύονται τα πλεονεκτήματά της έναντι αυτών, είτε ως προς την ποιότητα των εκτιμήσεων είτε ως προς την πολυπλοκότητα του αλγορίθμου.

Τα αποτελέσματα αυτής της διατριβής οδήγησαν στην συγγραφή τριών εργασιών. Οι δύο από αυτές έχουν ήδη δημοσιευθεί στα διεθνή περιοδικά Στατιστικής *Computational Statistics and Data Analysis* (Papastamoulis and Iliopoulos, 2009) και *Journal of Computational and Graphical Statistics* (Papastamoulis and Iliopoulos, 2010a), ενώ η τρίτη (Papastamoulis and Iliopoulos, 2010b) έχει υποβληθεί προς δημοσίευση.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Κεφάλαιο 2

Συμπερασματολογία σε Μοντέλα Μείξεων Κατανομών

2.1 Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζονται τα βασικά χαρακτηριστικά των μοντέλων μίξεων κατανομών. Αρχικά, δίνεται μια σύντομη περιγραφή των μεθόδων συμπερασματολογίας της κλασσικής Στατιστικής (όπως ο αλγόριθμος EM) και διαπιστώνονται τα μειονεκτήματα αυτής της προσέγγισης. Κατόπιν, περιγράφεται η μπεϋζιανή θεμελίωση ενός μοντέλου μίξης και οι υπολογιστικές δυσκολίες των κλειστών τύπων των εκτιμητών Bayes. Ακολούθως, παρουσιάζονται οι υπολογιστικές μέθοδοι Markov Chain Monte Carlo (MCMC) προσομοίωσης ενός δείγματος από την εκ των υστέρων κατανομή των παραμέτρων ενός μοντέλου μίξης θεωρώντας γνωστό αριθμό συνιστωσών. Στην συνέχεια περιγράφονται μέθοδοι συμπερασματολογίας όταν ο αριθμός συνιστωσών θεωρείται άγνωστος, όπως η μέθοδος reversible jump mcmc για μίξεις μονοδιάστατων κανονικών κατανομών. Η μέθοδος εφαρμόζεται σε παραδείγματα με προσομοιωμένα σύνολα δεδομένων όπου γίνεται σαφές ότι σε συγκεκριμένες περιπτώσεις η μέθοδος δεν εκτιμά σωστά το πλήθος των συνιστωσών.

Κατόπιν, παρουσιάζεται ένα βασικό χαρακτηριστικό της εκ των υστέρων κατανομής των παραμέτρων το οποίο ονομάζεται φαινόμενο label switching. Το εν λόγω φαινόμενο καθιστά μη τετριμμένη την διαδικασία εκτίμησης των παραμέτρων. Αφού δοθεί ένα απλό παράδειγμα για την κατανόηση του προβλήματος, γίνεται μια ανασκόπηση των διαθέσιμων μεθόδων αντιμετώπισής του. Το κεφάλαιο ολοκληρώνεται παρουσιάζοντας μία γενίκευση των μίξεων κατανομών σε μοντέλα πολυμεταβλητών παλινδρομήσεων. Εισάγεται ένα μοντέλο εκτίμησης των παραμέτρων για το συγκεκριμένο πρόβλημα, θεωρώντας γνωστό αριθμό συνιστωσών. Μέσω των παραδειγμάτων που παρατίθενται αναδεικνύεται η ανάγκη εύρεσης μιας γενικής μεθόδου αντιμετώπισης του προβλήματος label switching.

2.2 Στοιχεία πεπερασμένων μίξεων κατανομών

Ορισμός 2.2.1. Έστω $k \in \mathbb{N}$ με $k > 1$ και $f_1(\cdot), \dots, f_k(\cdot)$ συναρτήσεις πυκνότητας πιθανότητας ως προς κάποιο κατάλληλο μέτρο. Ένας κυρτός συνδυασμός της μορφής

$$\sum_{j=1}^k p_j f_j(\cdot), \quad (2.1)$$

με $p_j > 0$, $j = 1, \dots, k$, και $\sum_{j=1}^k p_j = 1$, καλείται (πεπερασμένη) μίξη των k κατανομών.

Οι ποσότητες p_j , $j = 1, \dots, k$, καλούνται βάρη της μίξης (2.1), ενώ ο θετικός ακέραιος k δηλώνει τον αριθμό των συνιστωσών αυτής. Από τον Ορισμό 2.2.1 είναι προφανές ότι η (2.1) ορίζει μία συνάρτηση πυκνότητας. Η $f_j(\cdot)$ θα αποκαλείται η πυκνότητα της j -συνιστώσας της μίξης. Επειδή θα ασχοληθούμε μόνο με πεπερασμένες μίξεις κατανομών, θα αναφερόμαστε σε ένα πεπερασμένο μοντέλο μίξης κατανομών απλώς με τον όρο μοντέλο μίξης.

Στις περισσότερες περιπτώσεις θεωρούμε ότι η f_j ανήκει σε κάποια δεδομένη παραμετρική οικογένεια με άγνωστη παράμετρο $\theta_j \in \Theta_j$, κάτι που οδηγεί στον ορισμό του *παραμετρικού μοντέλου μίξης*

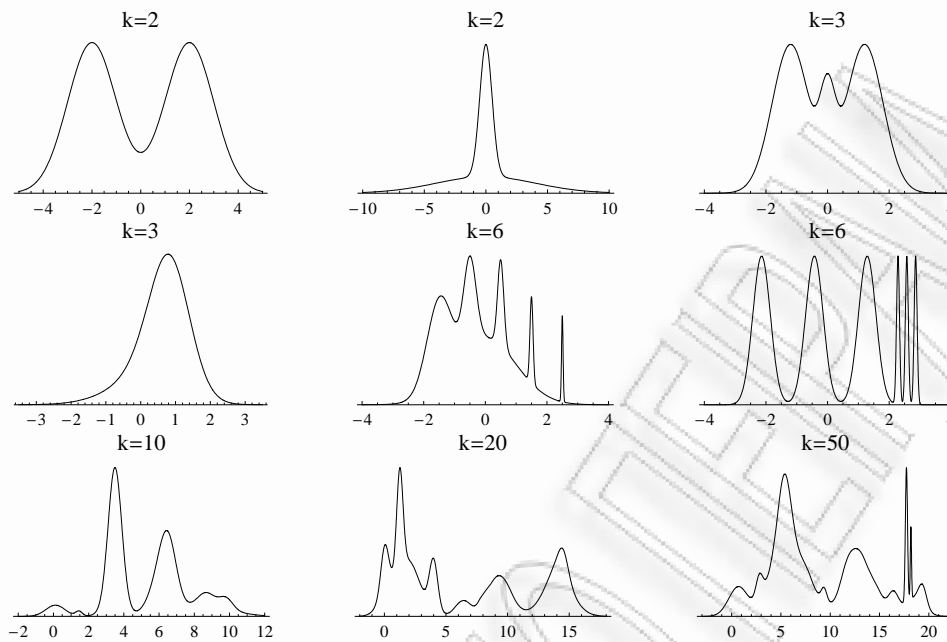
$$\sum_{j=1}^k p_j f_j(\cdot | \theta_j). \quad (2.2)$$

Συνήθως, οι συνιστώσες της μίξης (2.2) ανήκουν στην ίδια παραμετρική οικογένεια, δηλαδή ισχύει $f_j(\cdot | \theta_j) = f(\cdot | \theta_j)$ καθώς και $\Theta_j = \Theta$, $\forall j = 1, \dots, k$. Σε αυτήν την περίπτωση θα γράφουμε

$$\sum_{j=1}^k p_j f(\cdot | \theta_j). \quad (2.3)$$

Στο Σχήμα 2.1 αναπαρίστανται οι συναρτήσεις πυκνότητας διαφόρων μίξεων κανονικών κατανομών και παρατηρούμε την μεγάλη ποικιλία της μορφής αυτών.

Έστω $\mathbf{x} = (x_1, \dots, x_n)$ ένα τυχαίο δείγμα μεγέθους n από το παραμετρικό μοντέλο μίξης (2.3), με $x_i \in \mathcal{X} \subseteq \mathbb{R}^p$, $i = 1, \dots, n$. Το πρόβλημα της εκτίμησης ενός μοντέλου μίξεων κατανομών, με γνωστό πλήθος συνιστωσών, ανάγεται στην εκτίμηση των παραμέτρων του, $(\mathbf{p}, \boldsymbol{\theta}) \in \mathcal{P}_k \times \Theta^k$, όπου $\mathcal{P}_k := \{(p_1, \dots, p_k) \in (0, 1)^k : \sum_{j=1}^k p_j = 1\}$, δοθέντος του παρατηρηθέντος δείγματος \mathbf{x} . Στην παρούσα διατριβή θα ασχοληθούμε με την μπεϋζιανή εκτίμηση των εκ των υστέρων κατανομών των παραμέτρων των μοντέλων μίξης. Η μπεϋζιανή θεμελίωση ενός μοντέλου μίξης θα



Σχήμα 2.1: Μερικές πυκνότητες μείξεων κανονικών κατανομών για διάφορους αριθμούς συνιστώσων. Παρατηρούμε ότι ο αριθμός των κορυφών της πυκνότητας δεν είναι ενδεικτικός του αριθμού των συνιστώσων, εκτός του ότι μπορεί να θεωρηθεί ως ένα κάτω φράγμα αυτού.

συζητηθεί στην Ενότητα 2.4. Πρώτα όμως κρίνεται απαραίτητη μία εισαγωγή στις μεθόδους εκτίμησης υπό το πρίσμα της κλασσικής Στατιστικής.

Η πρώτη δημοσιευμένη έρευνα σχετικά με την εκτίμηση μιας μείξης κατανομών αποδίδεται στον Karl Pearson (1894). Ο Pearson θεώρησε το πρόβλημα εύρεσης του εκτιμητή ροπών των πέντε ανεξάρτητων παραμέτρων μιας μείξης κανονικών κατανομών με δύο συνιστώσες. Από τον Ορισμό 2.2.1 είναι σαφές ότι οι ροπές της (2.1) είναι κυρτοί συνδυασμοί των ροπών των συνιστώσων της, καθώς είναι προφανές ότι

$$\mathbb{E}(X^m) = \sum_{j=1}^k p_j \mathbb{E}_{f_j}(X^m), \quad \forall m \in \mathbb{Z}^+.$$

Έτσι λοιπόν, ο Pearson θεώρησε τις εξισώσεις που ορίζονται από τις πρώτες πέντε ροπές και ύστερα από επίπονες αλγεβρικές πράξεις κατέληξε στις εκτιμήσεις που εξαρτώνται από μία εκ των ριζών ενός πολυωνύμου ενάτου βαθμού. Χαρακτηριστικά, ο Charlier (1906) περιγράφει την προσπάθεια του Pearson ως μία «ηρωική πράξη». Για τα επόμενα πενήντα χρόνια περίπου, η μέθοδος των ροπών υιοθετήθηκε και από άλλους ερευνητές, όπως οι Charlier (1906), Burrau (1934) και Preston (1953), σε απλά προβλήματα μείξεων κατανομών με σκοπό κυρίως την απλοποίηση της μεθόδου

του Pearson.

Εν τω μεταξύ, εμφανίζονται οι πρώτες εργασίες σχετικά με την εκτίμηση ενός μοντέλου μείξης μέσω της μεθόδου μέγιστης πιθανοφάνειας. Η πιθανοφάνεια του μοντέλου (2.3) ορίζεται ως

$$L(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) = \prod_{i=1}^n \sum_{j=1}^k p_j f(x_i | \theta_j), \quad (\mathbf{p}, \boldsymbol{\theta}) \in \mathcal{P}_k \times \Theta^k. \quad (2.4)$$

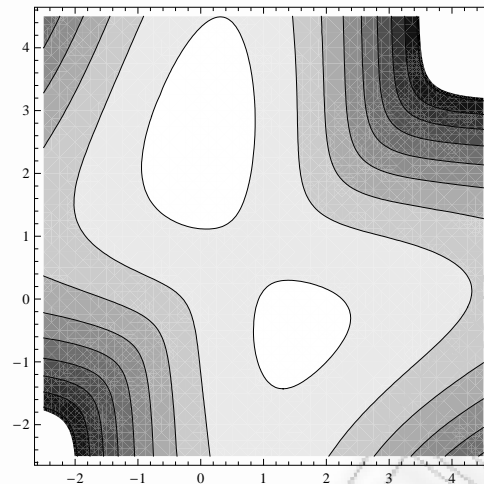
Από την (2.4) είναι σαφές ότι το ανάπτυγμα της πιθανοφάνειας περιλαμβάνει k^n όρους και είναι γενικά μη γραμμικό. Έτσι, ακόμα και για μικρό μέγεθος δείγματος, η εύρεση του εκτιμητή μέγιστης πιθανοφάνειας μέσω αναλυτικών τύπων είναι αδύνατη. Οι Rao (1948) και Mendenhall and Hader (1958) χρησιμοποίησαν κατάλληλες επαναληπτικές διαδικασίες ώστε να προσεγγίσουν τη λύση των εξισώσεων πιθανοφάνειας αλλά παρά το γεγονός ότι διαπραγματεύονταν πολύ απλές περιπτώσεις, αντιμετώπιζαν σοβαρά υπολογιστικά εμπόδια.

Με την εξάπλωση των ηλεκτρονικών υπολογιστών κατά την δεκαετία του 1960 οι υπολογιστικές απαιτήσεις έπαψαν σιγά σιγά να αποτελούν εμπόδιο. Έτσι, η έρευνα σχετικά με την προσέγγιση των λύσεων των εξισώσεων πιθανοφάνειας ενός μοντέλου μείξης επανέκαμψε. Οι Hasselblad (1966) και Wolfe (1970) χρησιμοποίησαν επαναληπτικές αριθμητικές μεθόδους για την εκτίμηση μίξεων μονοδιάστατων και πολυδιάστατων κανονικών κατανομών, αντίστοιχα. Τελικά, οι Dempster et al. (1977) εισάγουν τον αλγόριθμο EM (δες Ενότητα 2.3) ο οποίος έπαιξε καθοριστικό ρόλο στην υιοθέτηση των μίξεων κατανομών ως ένα χρήσιμο στατιστικό εργαλείο.

Παρά τις προαναφερθείσες σημαντικές εξελίξεις και την παράκαμψη των υπολογιστικών δυσκολιών, άρχισαν να προκύπτουν κάποιες άλλες δυσκολίες σχετικά με την μέθοδο της μέγιστης πιθανοφάνειας. Ο εκτιμητής μέγιστης πιθανοφάνειας ενός μοντέλου μείξης δεν είναι πάντα διαθέσιμος, καθώς στις περιπτώσεις όπου τα θ_j , $j = 1, \dots, k$, παριστάνουν ζεύγη παραμέτρων θέσης-κλίμακος, η πιθανοφάνεια δεν είναι φραγμένη. Ένα τέτοιο παράδειγμα δίνεται στους Cox and Hinkley (1974), όπου θεωρούν ένα τυχαίο δείγμα $\mathbf{x} = (x_1, \dots, x_n)$ από την μείξη

$$10^{-10} \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} + (1 - 10^{-10}) \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2} \right\}.$$

Είναι σαφές ότι η αντίστοιχη πιθανοφάνεια τείνει στο άπειρο στην περίπτωση όπου το μ συμπίπτει με κάποιο από τα x_i , και $\sigma^2 \downarrow 0$ (δες και Kiefer and Wolfowitz, 1956). Προφανώς, ανάλογες ιδιάζουσες περιπτώσεις προκύπτουν όταν είναι άγνωστα τα βάρη και οι διασπορές και των δύο συνιστωσών. Το γεγονός αυτό μπορεί να αποφευχθεί



Σχήμα 2.2: Ο λογάριθμος της πιθανοφάνειας στο επίπεδο των (μ_1, μ_2) για τις 500 προσομοιωμένες παρατηρήσεις από την κατανομή (2.5).

Θεωρώντας ότι οι διασπορές των κανονικών συνιστωσών είναι ίσες μεταξύ τους ή αν είναι γνωστό εκ των προτέρων ότι κάθε συνιστώσα της μείξης έχει ένα ελάχιστο πλήθος παρατηρήσεων που προέρχονται από αυτή (Hosmer, 1973a). Βέβαια, σε γενικά πλαίσια, τέτοιου είδους υποθέσεις είτε είναι αρκετά περιοριστικές είτε είναι αδύνατο να εξασφαλιστούν.

Σε περιπτώσεις όπου η πιθανοφάνεια δεν είναι φραγμένη, είναι σαφές ότι η εύρεση του ΕΜΠ ενός μοντέλου μείξης έχει να κάνει με την ανακάλυψη τοπικών μεγίστων της πιθανοφάνειας. Παρακάμπτοντας τις υπολογιστικές δυσκολίες του εγχειρήματος αυτού, συχνά η πιθανοφάνεια ενός μοντέλου παρουσιάζει αρκετά τοπικά μέγιστα κάθε ένα από τα οποία παράγει διαφορετικές σημειακές εκτιμήσεις των άγνωστων παραμέτρων. Μία τέτοια περίπτωση απεικονίζεται στο Σχήμα 2.2 όπου (θεωρώντας γνωστά τα βάρη και τις διασπορές) ο λογάριθμος της πιθανοφάνειας $L(\boldsymbol{\mu}|\mathbf{x})$ παρουσιάζει δύο κορυφές, για $n = 500$ προσομοιωμένες παρατηρήσεις από την μείξη δύο κανονικών κατανομών

$$0.7\mathcal{N}(0, 1) + 0.3\mathcal{N}(2.5, 1). \quad (2.5)$$

Γενικά, η ύπαρξη πολλών κορυφών εγείρει και το ερώτημα για το αν έχουν εντοπιστεί όλες ή όχι. Σε επόμενη ενότητα θα εξηγηθεί ότι κάθε κορυφή της πιθανοφάνειας ενός μοντέλου μείξης δεν αντιστοιχεί σε μία μοναδική τιμή των παραμέτρων και κάτι τέτοιο έχει επιπτώσεις στην ασυμπτωτική θεωρία του εκτιμητή μέγιστης πιθανοφάνειας.

Οι Redner and Walker (1984) έχουν διατυπώσει κάποιες ικανές συνθήκες υπό την ισχύ των οποίων εξασφαλίζεται η ύπαρξη μιας ισχυρά συνεπούς λύσης των εξισώσεων πιθανοφάνειας, η οποία μεγιστοποιεί τοπικά την πιθανοφάνεια και είναι

ασυμπτωτικά κανονικά κατανεμημένη. Δυστυχώς όμως, στις πρακτικές εφαρμογές το μέγεθος του δείγματος δεν είναι πάντοτε αρκετά μεγάλο. Έτσι, η συμπεριφορά του ΕΜΠ είναι εξαιρετικά ασταθής σε περιπτώσεις όπου οι συνιστώσες των μείξεων δεν είναι επαρκώς διαχωρισμένες μεταξύ τους. Γενικά, ένα πρόβλημα καλείται «ασθενώς ορισμένο» (ill-conditioned) όταν η λύση του είναι αρκετά ευαίσθητη σε μικρές διαφοροποιήσεις των δεδομένων. Το πρόβλημα εύρεσης τοπικών μεγίστων της πιθανοφάνειας ενός μοντέλου μείξης ανήκει ακριβώς σε αυτήν την κατηγορία προβλημάτων. Για παράδειγμα, αν έχουμε δεδομένα από μία μείξη δύο κανονικών πληθυσμών με παραμέτρους $p_1 = 0.3, \sigma_1^2 = \sigma_2^2 = 1$ και $\mu_1 - \mu_2 = 1$, τότε απαιτείται μέγεθος δείγματος της τάξης του 10^6 προκειμένου να εξασφαλιστεί ότι η τυπική απόκλιση κάθε στοιχείου του ΕΜΠ είναι μικρότερη από 0.1 (Redner and Walker, 1984). Πιο γενικά, ο Hosmer (1973b), ύστερα από μελέτη της συμπεριφοράς του ΕΜΠ σε περιπτώσεις μικρού μεγέθους δείγματος και μικρής διαφοροποίησης των συνιστωσών, καταλήγει στο ότι

«οι εκτιμήσεις μέγιστης πιθανοφάνειας θα πρέπει να χρησιμοποιούνται με μεγάλη επιφύλαξη ή να μην χρησιμοποιούνται καθόλου».

2.2.1 Η διατύπωση του μοντέλου μέσω ελλিপών δεδομένων

Σε πολλές περιπτώσεις, η πολυπλοκότητα ενός μοντέλου που υιοθετείται για την περιγραφή των διαθέσιμων δεδομένων προέρχεται από την έλλειψη πληροφορίας για την υιοθέτηση κάποιου απλούστερου μηχανισμού. Τυπικά, όλα τα μοντέλα που ορίζονται μέσω μιας διαδικασίας περιθωριοποίησης, δηλαδή όταν η πυκνότητα των παρατηρήσεων \mathbf{x} , $f(\mathbf{x}|\boldsymbol{\eta})$, μπορεί να δοθεί από ένα ολοκλήρωμα της μορφής

$$f(\mathbf{x}|\boldsymbol{\eta}) = \int_{\mathbf{z}} f(\mathbf{x}, \mathbf{z}|\boldsymbol{\eta})\nu(d\mathbf{z}), \quad (2.6)$$

δύναται να θεωρηθούν ως μοντέλα ελλিপών δεδομένων. Αν μία τέτοια αναπαράσταση έχει νόημα για την κατασκευή του μοντέλου, τότε τα (\mathbf{x}, \mathbf{z}) αναφέρονται ως *πλήρη δεδομένα* αν και το διάνυσμα \mathbf{z} είναι μη παρατηρήσιμο. Παρά το γεγονός ότι οποιαδήποτε πυκνότητα μπορεί να γραφτεί κατ' αυτόν τον τρόπο, στα μοντέλα μείξεων μία τέτοιου είδους αναπαραστάση παρέχει το κατάλληλο κίνητρο για την ερμηνεία των μείξεων κατανομών και την χρήση αυτών ως μοντέλα που περιγράφουν ετερογενή σύνολα δεδομένων.

Υπάρχουν αρκετά κίνητρα για την θεώρηση των μοντέλων μείξεων ως μία χρήσιμη επέκταση «συνήθων» κατανομών. Η πλέον φυσική προσέγγιση είναι η ενσωμάτωση

ενός παράγοντα ο οποίος χωρίζει τον πληθυσμό σε (πεπερασμένου πλήθους) στρώματα ή υποπληθυσμούς. Μία από τις πρώτες καταγραφές τέτοιας συμπεριφοράς συναντάται στον Bertillon (1887) όπου η διωνυμική δομή στο ύψος των στρατιωτών της κεντρικής Γαλλίας εξηγήθηκε μέσω της μείξης δύο πληθυσμών νέων αντρών, ο πρώτος από τα πεδινά και ο δεύτερος από τα ορεινά. Η συμπεριφορά της μείξης εμφανίστηκε διότι δεν είχε καταγραφεί η προέλευση κάθε παρατήρησης, που στην προκειμένη περίπτωση μπορεί να θεωρηθεί ως μία δίτιμη μεταβλητή που καταγράφει την κατάταξη κάθε στρατιώτη σε έναν από τους δύο πληθυσμούς (πεδινά ή ορεινά). Έτσι, κάθε παρατήρηση x_i είναι *εκ των προτέρων* κατανεμημένη σε κάθε μία από τις συνιστώσες $f(\cdot|\theta_j)$ που περιγράφουν το ύψος κάθε περιοχής με πιθανότητα p_j , $j = 1, 2$. Πολύ συχνά, στις πρακτικές εφαρμογές δεν είναι πάντα δυνατόν να καταγράφεται η μεταβλητή κατάταξης κάθε παρατήρησης στους πληθυσμούς που συνιστούν τη μείξη. Για παράδειγμα, σύμφωνα με το International Halibut Commission of Seattle, η κατανομή του μήκους του ιππόγλωσσου¹ μιας συγκεκριμένης ηλικίας προσεγγίζεται από μία μείξη δύο κανονικών κατανομών. Εδώ η συμπεριφορά της μείξης εμφανίζεται διότι δεν καταγράφεται το φύλο του ψαριού, μία διαδικασία που είναι ασύμφορη λόγω του μεγάλου κόστους της.

Μέσω μιας διαδικασίας αποπεριθωριοποίησης λοιπόν, είναι πάντα εφικτό να αντιστοιχίσουμε σε μια τυχαία μεταβλητή X_i μία άλλη τυχαία μεταβλητή Z_i με τιμές στο σύνολο $\{1, \dots, k\}$, τέτοια ώστε για κάθε $j = 1, \dots, k$

$$X_i|Z_i = j, \boldsymbol{\theta} \sim f(x|\theta_j), \quad (2.7)$$

$$Z_i|\mathbf{p} \sim M_k(1; p_1, \dots, p_k) \quad (2.8)$$

για $i = 1, \dots, n$, όπου $M_k(1; p_1, \dots, p_k)$ συμβολίζει την πολυωνυμική κατανομή με k κατηγορίες, ώστε $P(Z_i = j) = p_j$. Οι τυχαίες μεταβλητές Z_i , $i = 1, \dots, n$, θα καλούνται *μεταβλητές κατάταξης* καθώς καθορίζουν σε ποια συνιστώσα της μείξης ανήκει κάθε μία παρατήρηση. Εφ' όσον δε θεωρούνται ανεξάρτητες μεταξύ τους, είναι προφανές ότι το ίδιο ισχύει και για τις X_i . Λόγω των (2.7) και (2.8) προκύπτει ότι η σχέση (2.6), στην περίπτωση των μείξεων, μεταφράζεται στο ότι η περιθωριακή κατανομή του X_i είναι η (2.3). Υπό τη συγκεκριμένη οπτική, το κλασσικό μοντέλο μείξης κατανομών (2.3) μπορεί να θεωρηθεί ως ένα μοντέλο ελλειπών δεδομένων, υπό την έννοια ότι τα Z_i είναι μη παρατηρήσιμα. Η υπόθεση ανεξαρτησίας των μεταβλητών κατάταξης θα υιοθετηθεί στο μεγαλύτερο μέρος της παρούσας διατριβής. Ωστόσο, αν οι μεταβλητές κατάταξης θεωρηθούν εξαρτημένες μεταξύ τους τότε προκύπτουν αρ-

¹Είδος ψαριών.

κετά ενδιαφέρουσες κλάσεις μοντέλων όπως τα hidden Markov models ή τα Markov random fields που θα συζητηθούν στο Κεφάλαιο 5.

Ανάλογα με την περίπτωση, οι μη παρατηρηθείσες μεταβλητές κατάταξης θα είναι ή δεν θα είναι μέρος των προς εκτίμηση ποσοτήτων. Η από κοινού κατανομή των πλήρων δεδομένων (x_i, z_i) , $i = 1, \dots, n$, τα οποία αποτελούνται από ανεξάρτητα ζεύγη, είναι η

$$P(X_i \in dx, Z_i = j) = p_j f(x|\theta_j) dx, \quad (2.9)$$

για $j = 1, \dots, k$. Σύμφωνα με τα παραπάνω, η πλήρης δεσμευμένη εκ των υστέρων κατανομή των μεταβλητών κατάταξης είναι

$$P(Z_i = j | X_i = x, \mathbf{p}, \boldsymbol{\theta}) = \frac{p_j f(x|\theta_j)}{p_1 f(x|\theta_1) + \dots + p_k f(x|\theta_k)} \propto p_j f(x|\theta_j), \quad (2.10)$$

ανεξάρτητα για κάθε $i = 1, \dots, n$.

Η αναπαράσταση (2.9) μέσω των μη παρατηρήσιμων πλήρων δεδομένων ορίζει την *πλήρη πιθανοφάνεια* του μοντέλου, η οποία γράφεται σε λογαριθμική μορφή

$$\log L^c(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z}) = \log L(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) + \log f(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}, \mathbf{p}), \quad (2.11)$$

όπου λόγω της (2.10) και της εκ των υστέρων ανεξαρτησίας των μεταβλητών κατάταξης, έχουμε ότι

$$f(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}, \mathbf{p}) \propto \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}),$$

για $\mathbf{z} = (z_1, \dots, z_n) \in \mathcal{Z} := \{1, \dots, k\}^n$.

2.2.2 Μοντελοποιώντας άγνωστες κατανομές

Όπως συζητήθηκε στην Ενότητα 2.2.1, οι μείξεις κατανομών μπορούν να χρησιμοποιηθούν για την ερμηνεία της ετερογένειας στον υπό μελέτη πληθυσμό. Εν τούτοις, η συγκεκριμένη οπτική δεν αποτελεί το μόνο λόγο για την εφαρμογή ενός μοντέλου μείξης στα διαθέσιμα δεδομένα. Πολλές φορές το δείγμα μπορεί να παρουσιάζει χαρακτηριστικά όπως ασυμμετρία, κύρτωση, λοξότητα κλπ. και γενικά η μορφή αυτού να μη δικαιολογεί την προσαρμογή κάποιας από τις «συνηθισμένες» οικογένειες κατανομών. Αρκετά συνηθισμένη δε, είναι η προσπάθεια εξάλειψης αυτών των χαρακτηριστικών μέσω κατάλληλων μετασχηματισμών (όπως π.χ. η λογαρίθμηση των δεδομένων). Είναι κατανοητό ότι τέτοιου είδους απόπειρες δεν είναι οι πλέον ενδεδειγμένες διότι στηρίζονται σε εμπειρικούς κανόνες και δεν είναι σαφής η εφαρμογή τους σε γενικά πλαίσια. Σε αυτές τις περιπτώσεις μπορεί να επιστρατευτεί ένα μοντέλο μείξης κατανομών με σκοπό την προσέγγιση της άγνωστης κατανομής, αφού

αποφασισθεί με κάποιον τρόπο ο κατάλληλος αριθμός συνιστωσών και εκτιμηθούν οι παράμετροί του.

Επανερχόμενοι στο Σχήμα 2.1, είναι εμφανές ότι η πυκνότητα μιας μείξης κανονικών κατανομών δύναται να έχει σχεδόν οποιαδήποτε μορφή επιθυμούμε με την κατάλληλη επιλογή του αριθμού συνιστωσών και των παραμέτρων. Επομένως, τουλάχιστον γραφικά, δικαιολογείται ο ισχυρισμός ότι οποιαδήποτε πυκνότητα μπορεί να προσεγγιστεί μέσω μιας κανονικής μείξης (δες Marron and Wand, 1992). Το γεγονός αυτό είχε παρατηρηθεί αρκετά νωρίς. Συγκεκριμένα, ο Pearson (1895) σε μια προσπάθεια εύρεσης κάποιου κριτηρίου που να διαφοροποιεί μία λοξή κατανομή από μία μείξη κανονικών κατανομών (που παρουσιάζει λοξότητα), καταλήγει στο ότι

«...δεν είναι εύκολη η διάκριση μιας (αληθινά) λοξής καμπύλης και μιας μείξης, υποθέτοντας ότι δεν έχουμε κάποια ιδιαίτερη εκ των προτέρων υποψία για την ύπαρξη μείξης. Προς το παρόν, δεν έχω καταφέρει να βρω κάποια γενική συνθήκη μεταξύ των ροπών, όπου θα ήταν αδύνατη για μία λοξή καμπύλη και δυνατή για μία σύνθετη, ώστε να διαπιστώνεται η ύπαρξη μείξης. Παρ' όλ' αυτά, δεν απελπίζομαι για την ύπαρξη μιας τέτοιας συνθήκης.»

Για να γίνει αντιληπτή η θέση του Pearson για την δυσκολία διάκρισης μεταξύ (λοξών ή μη) κατανομών και μείξεων κατανομών, στο Σχήμα 2.3 απεικονίζονται οι συναρτήσεις πυκνότητας των εξής μείξεων δύο κανονικών κατανομών

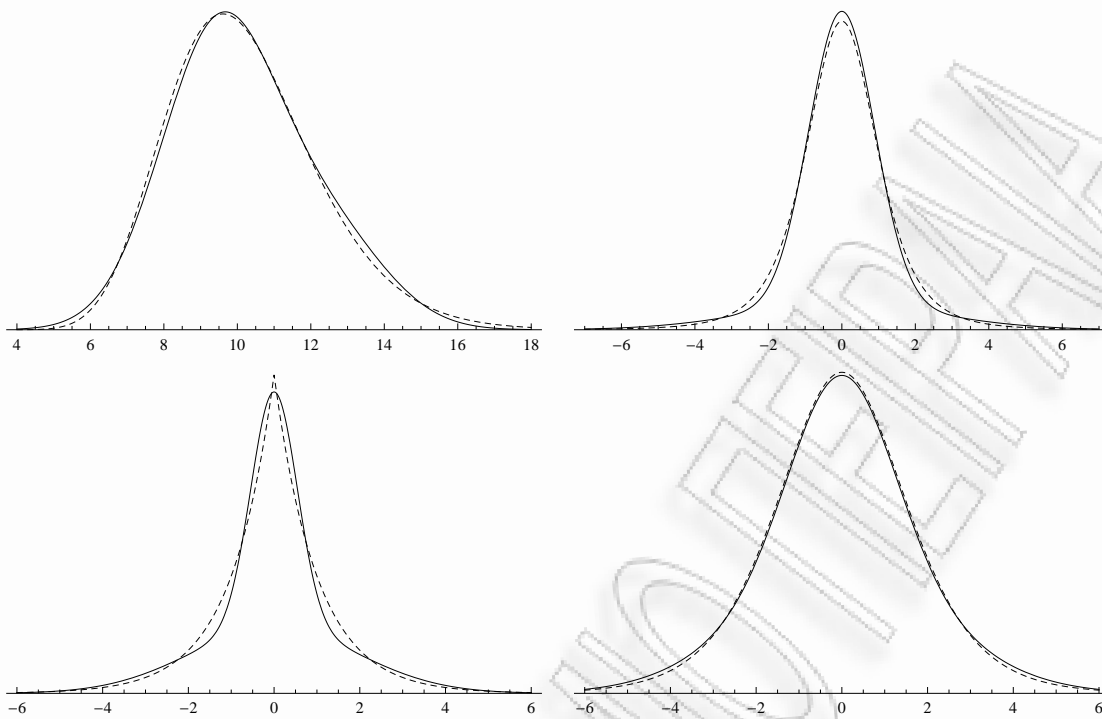
$$0.9\mathcal{N}(9.5, 2.5) + 0.1\mathcal{N}(12.5, 2.5) \quad (2.12)$$

$$0.8\mathcal{N}(0, 0.8) + 0.2\mathcal{N}(0, 7) \quad (2.13)$$

$$0.5\mathcal{N}(0, 0.29) + 0.5\mathcal{N}(0, 3.8) \quad (2.14)$$

$$0.47\mathcal{N}(0, 1.45) + 0.53\mathcal{N}(0, 5.3). \quad (2.15)$$

Στο ίδιο Σχήμα, με διακεκομμένες γραμμές απεικονίζονται οι πυκνότητες των κατανομών $\mathcal{LN}(\log 10, 0.04)$, t_4 , $\text{Laplace}(0, 1)$ και $\text{Logistic}(0, 1)$, αντίστοιχα (με \mathcal{LN} συμβολίζεται η λογαριθμο-κανονική κατανομή). Το γεγονός ότι οι συναρτήσεις πυκνότητας των μείξεων σχεδόν συμπίπτουν με τις πυκνότητες αυτών των κατανομών σημαίνει ότι στην πράξη είναι πολύ δύσκολο να τις ξεχωρίσουμε μεταξύ τους. Υπό αυτήν την οπτική, η προσέγγιση διάφορων μοντέλων πιθανοτήτων με την χρήση ενός μοντέλου μείξεων κατανομών είναι απόλυτα δικαιολογημένη. Η ευχέρεια που παρέχουν τα μοντέλα μείξεων γίνεται ακόμα πιο προφανής όταν σκεφτούμε ότι σε πολλές περιπτώσεις δεν είναι βολικό να μοντελοποιούμε τα δεδομένα μας με μη τυπικές



Σχήμα 2.3: Πάνω: Πυκνότητες των μίξεων (2.12) και (2.13) (συνεχείς γραμμές), μαζί με τις πυκνότητες των κατανομών $\mathcal{LN}(\log 10, 0.04)$ και t_4 (διακεκομμένες γραμμές), αντίστοιχα. Κάτω: τα αντίστοιχα γραφήματα για τις μίξεις (2.14) και (2.15) με τις κατανομές $\text{Laplace}(0, 1)$ και $\text{Logistic}(0, 1)$.

κατανομές όπως οι προηγούμενες. Για παράδειγμα, η εκτίμηση των παραμέτρων μιας κατανομής $t_\nu(\mu, \sigma^2)$ δεν αποτελεί μια απλή υπόθεση όταν οι βαθμοί ελευθερίας αυτής (ν) είναι άγνωστοι. Χαρακτηριστικά αναφέρουμε ότι σε αυτήν την περίπτωση η σύγκλιση του αλγορίθμου EM είναι πολύ αργή (Liu and Rubin, 1994), ενώ το βήμα για την ανανέωση του ν είναι αρκετά χρονοβόρο (Lange et al., 1989).

Έτσι λοιπόν, αντί να σκέφτεται κάποιος ερευνητής μια πιθανή παραμετρική οικογένεια κατανομών που να προσαρμόζεται καλά στα δεδομένα του, οι μίξεις κατανομών παρέχουν μια ελκυστική εναλλακτική δυνατότητα, αρκεί να είναι διαθέσιμες κατάλληλες μέθοδοι για την εκτίμηση αυτών. Βέβαια, από υπολογιστικής άποψης υπάρχει ταυτόχρονα το μειονέκτημα ότι χρειάζεται λίγη παραπάνω προσπάθεια, καθώς αντί να εκτιμηθούν π.χ. δύο παράμετροι (όπως στην περίπτωση των κατανομών Log-Normal, Logistic και Laplace), πρέπει να εκτιμηθούν πέντε (στην περίπτωση όπου επιλέγεται μία μίξη δύο κανονικών κατανομών). Σε αυτό το σημείο ας υπογραμμίσουμε το γεγονός ότι στις μίξεις (2.13), (2.14) και (2.15) οι μέσες τιμές των δύο κανονικών συνιστωσών είναι ίδιες μεταξύ τους. Αν μια τέτοιου είδους πληρο-

φορία ληφθεί υπ' όψιν κατά την εκτίμηση ενός μοντέλου μείζης, το υπολογιστικό κόστος μπορεί να μειωθεί, καθώς τότε εκτιμάται ένα μοντέλο τεσσάρων παραμέτρων, αντί για πέντε. Αυτή η παρατήρηση αποτελεί ένα από τα κίνητρα για την ανάπτυξη της μεθόδου που προτείνεται στο Κεφάλαιο 3 της διατριβής.

2.3 Ο αλγόριθμος EM

Ο αλγόριθμος Expectation - Maximization (EM) των Dempster et al. (1977) είναι μία από τις πιο γνωστές αριθμητικές μεθόδους μεγιστοποίησης της πιθανοφάνειας. Εφαρμόζεται σε μοντέλα όπου υπάρχουν (ή μπορεί να θεωρηθεί ότι υπάρχουν) ελλιπή δεδομένα. Μέχρι την ραγδαία εξάπλωση των μεθόδων προσομοίωσης MCMC, ο αλγόριθμος EM αποτελούσε ίσως το μοναδικό εργαλείο για εκτίμηση παραμέτρων στα μοντέλα μείξεων κατανομών. Για μία ολοκληρωμένη παρουσίαση της μεθόδου στα προβλήματα μείξεων κατανομών ο αναγνώστης παραπέμπεται στην μονογραφία των McLachlan and Peel (2000). Εδώ θα περιγραφεί εν συντομία η μέθοδος και θα εφαρμοστεί σε ένα απλό παράδειγμα ώστε να αναδειχθούν τα μειονεκτήματα της κλασσικής συμπερασματολογίας σε προβλήματα μείξεων κατανομών.

Αλγόριθμος 2.3.1. Αλγόριθμος EM για μείξεις κατανομών

Δώσε κάποιες αρχικές τιμές $(\mathbf{p}, \boldsymbol{\theta})^{(0)}$ και επανάλαβε τα ακόλουθα βήματα μέχρι να ικανοποιηθεί κάποιο (προκαθορισμένο) κριτήριο σύγκλισης, για $t = 1, 2, \dots$:

1. *Βήμα E*: Υπολόγισε

$$Q(\mathbf{p}, \boldsymbol{\theta} | (\mathbf{p}, \boldsymbol{\theta})^{(t-1)}, \mathbf{x}) := \mathbb{E}_{(\mathbf{p}, \boldsymbol{\theta})^{(t-1)}} \log \{L^c(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{Z})\},$$

όπου $\mathbf{Z} \sim f(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}, \mathbf{p})$.

2. *Βήμα M*: Μεγιστοποίησε ως προς $\mathbf{p}, \boldsymbol{\theta}$ το $Q(\mathbf{p}, \boldsymbol{\theta} | (\mathbf{p}, \boldsymbol{\theta})^{(t-1)}, \mathbf{x})$ και θέσε

$$(\mathbf{p}, \boldsymbol{\theta})^{(t)} = \arg \max_{(\mathbf{p}, \boldsymbol{\theta})} Q(\mathbf{p}, \boldsymbol{\theta} | (\mathbf{p}, \boldsymbol{\theta})^{(t-1)}, \mathbf{x}).$$

Μία γενική μορφή του αλγορίθμου EM για προβλήματα μείξεων κατανομών περιγράφεται στον Αλγόριθμο 2.3.1. Ο αλγόριθμος ξεκινάει από κάποιες αυθαίρετες τιμές των παραμέτρων και παράγει μία ακολουθία τιμών $(\mathbf{p}, \boldsymbol{\theta})^{(t)}$, $t = 1, 2, \dots$, η οποία σε κάθε επανάληψη αυξάνει την τιμή της (παρατηρηθείσας) πιθανοφάνειας (Dempster et al., 1977). Σε κάθε βήμα, οι μη παρατηρηθείσες μεταβλητές \mathbf{z} αντι-

καθίστανται από τις μέσες τιμές τους σύμφωνα με την κατανομή (2.10) και κατόπιν μεγιστοποιείται η αναμενόμενη πλήρης πιθανοφάνεια (2.11).

Η μέθοδος είναι αρκετά αποδοτική όταν η πιθανοφάνεια έχει μοναδική κορυφή. Στην Ενότητα 2.2 είδαμε όμως ότι η πιθανοφάνεια μιας μείζης κανονικών κατανομών ενδέχεται να μην είναι φραγμένη, οπότε σε αυτές τις περιπτώσεις ο αλγόριθμος EM αποτυγχάνει να συγκλίνει όταν αρχικοποιηθεί κοντά σε αυτή την περιοχή. Εκτός από αυτές τις ιδιόζουσες περιπτώσεις, η πιθανοφάνεια ενός μοντέλου μείζης μπορεί να παρουσιάζει αρκετές κορυφές και επομένως, ακόμα και σε απλές περιπτώσεις, ο αλγόριθμος συγκλίνει σε ένα τοπικό μέγιστο της πιθανοφάνειας χωρίς να εξερευνήσει τις υπόλοιπες περιοχές. Μία τέτοια περίπτωση θεωρείται στο Παράδειγμα 2.3.1.

Παράδειγμα 2.3.1. Εκτίμηση μείζης (2.5).

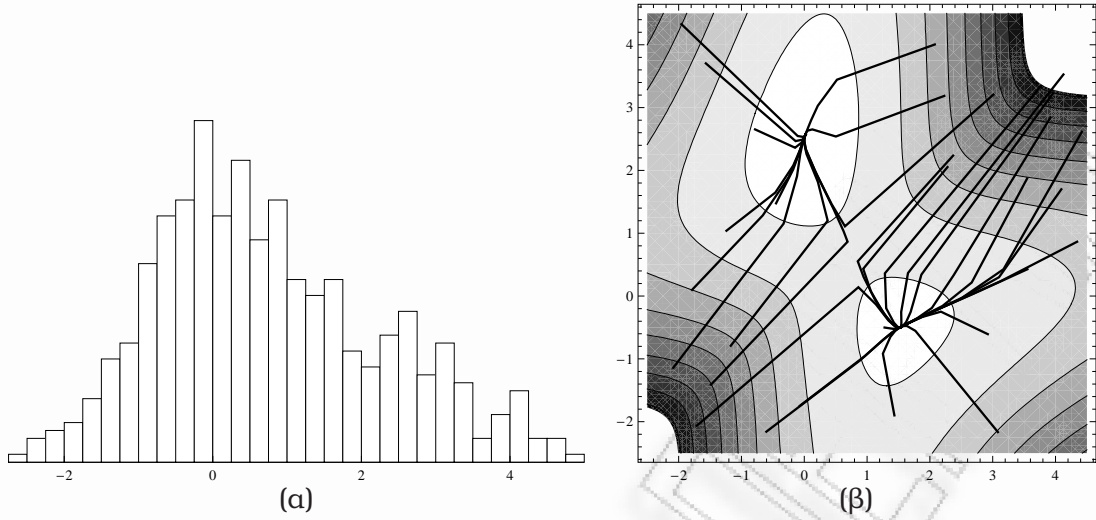
Ως εφαρμογή του Αλγορίθμου 2.3.1 σε μια σχετικά απλή περίπτωση, επιστρέφουμε στα προσομοιωμένα δεδομένα από την μείξη (2.5), όπου τα βάρη και οι διασπορές των δύο συνιστωσών θεωρούνται γνωστά. Το ιστόγραμμα των δεδομένων παρατίθεται στο Σχήμα 2.4(α). Σε αυτήν την περίπτωση, το παραμετρικό διάνυσμα αποτελείται μόνο από τους μέσους των κανονικών κατανομών, δηλαδή $\theta = (\mu_1, \mu_2) \in \mathbb{R}^2$. Ήδη στο Σχήμα 2.2 είδαμε ότι ο λογάριθμος της παρατηρηθείσας πιθανοφάνειας παρουσιάζει δύο κόρυφες. Η κύρια κορυφή βρίσκεται κοντά στην περιοχή $(0, 2.5)$ όπου αντιστοιχεί στις πραγματικές τιμές των παραμέτρων που χρησιμοποιήθηκαν για την προσομοίωση των δεδομένων. Επιπροσθέτως, υπάρχει και μία μικρότερη κορυφή κοντά στην περιοχή του $(1.5, -0.5)$. Αν και η διαφορά των τιμών της πιθανοφάνειας ανάμεσα σε αυτές τις δύο κορυφές είναι πολύ μεγάλη (της τάξης του 75.64 σε λογαριθμική κλίμακα), η ύπαρξη της μικρότερης κορυφής έχει σημαντική επίπτωση στη συμπεριφορά του αλγορίθμου EM.

Για να προσεγγιστεί το μέγιστο της πιθανοφάνειας, εκτελείται ο Αλγόριθμος 2.3.1 30 φορές ανεξάρτητα (με 100 επαναλήψεις την φορά), επιλέγοντας τυχαία τις αρχικές τιμές. Είναι εύκολο να δειχτεί ότι κατά την t -οστή επανάληψη του Αλγορίθμου 2.3.1, στο βήμα E θέτουμε

$$Q(\boldsymbol{\mu} | \boldsymbol{\mu}^{(t-1)}, \mathbf{x}) = \sum_{i=1}^n \hat{z}_i^{(t-1)} \log 0.7 + (n - \sum_{i=1}^n \hat{z}_i^{(t-1)}) \log 0.3 - \frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^n \{ \hat{z}_i^{(t-1)} (x_i - \mu_1)^2 + (1 - \hat{z}_i^{(t-1)}) (x_i - \mu_2)^2 \}$$

όπου

$$\hat{z}_i^{(t-1)} = \frac{0.7 \exp\{-(x_i - \mu_1^{(t-1)})^2/2\}}{0.7 \exp\{-(x_i - \mu_1^{(t-1)})^2/2\} + 0.3 \exp\{-(x_i - \mu_2^{(t-1)})^2/2\}}.$$



Σχήμα 2.4: (α) Ιστογράμμο των 500 προσομοιωμένων παρατηρήσεων από την κατανομή (2.5). (β) Τα ίχνη των 30 ανεξάρτητων εκτελέσεων του αλγορίθμου EM.

Ακολουθώντας, στο βήμα M θέτουμε

$$\mu_1^{(t)} = \frac{\sum_{i=1}^n \hat{z}_i^{(t-1)} x_i}{\sum_{i=1}^n \hat{z}_i^{(t-1)}} \quad \text{και} \quad \mu_2^{(t)} = \frac{\sum_{i=1}^n (1 - \hat{z}_i^{(t-1)}) x_i}{n - \sum_{i=1}^n \hat{z}_i^{(t-1)}}.$$

Όπως φαίνεται στο Σχήμα 2.4(β), ο αλγόριθμος πάντα συγκλίνει σε μία κορυφή της πιθανοφάνειας, αλλά μόνο στις 12 από τις 30 επαναλήψεις η ακολουθία τιμών συνέκλινε στην κύρια κορυφή. Πιο συγκεκριμένα, η εκτίμηση μέγιστης πιθανοφάνειας σε αυτήν την περίπτωση είναι $\hat{\theta} = (0.0209747, 2.57561)$ και όπως αναμένεται είναι αρκετά κοντά στις πραγματικές τιμές των παραμέτρων. Αντίθετα, οι υπόλοιπες 18 φορές συνέκλιναν στην μικρότερη κορυφή που αντιστοιχεί στο σημείο $\hat{\theta} = (1.52786, -0.497546)$. Αυτό προφανώς συμβαίνει διότι έτυχε οι αρχικές τιμές να βρίσκονται στην περιοχή έλξης της μικρότερης κορυφής.

Η εξάρτηση του αλγορίθμου EM από τις αρχικές τιμές είναι ένα γενικό μειονέκτημα αυτής της μεθόδου και συμβαίνει πάντα σε περιπτώσεις όπου η πιθανοφάνεια έχει αρκετές κορυφές. Γενικά, συνιστάται η εκτέλεση του αλγορίθμου με διαφορετικές αρχικές τιμές και η σύγκριση των αποτελεσμάτων μεταξύ τους ώστε να διαπιστωθεί ποιά είναι το μέγιστο της πιθανοφάνειας. Σε πιο πολύπλοκες περιστάσεις όμως, όπως όταν είναι άγνωστες όλες οι παράμετροι του μοντέλου ή σε πολυμεταβλητές περιπτώσεις, κάτι τέτοιο είναι αρκετά χρονοβόρα διαδικασία. Επιπροσθέτως, κανένας δεν εγγυάται αν οι διάφορες εκτελέσεις του αλγορίθμου έχουν πράγματι βρει την υψηλότερη κορυφή και όχι μόνο κάποιες από τις μικρότερες. Από την άλλη πλευρά, σε πολλές εφαρμογές οι μικρότερες κορυφές μπορεί να παρέχουν εναλλακτικές

εξηγήσεις των δεδομένων και η διαδικασία επιλογής μιας κορυφής έναντι κάποιας άλλης δεν μπορεί να αιτιολογηθεί υπό το πρίσμα της κλασσικής συμπερασματολογίας. Γίνεται λοιπόν σαφές ότι υπάρχουν περιπτώσεις όπου η κλασσική Στατιστική δεν δίνει ικανοποιητικές απαντήσεις και έτσι κρίνεται απαραίτητη η χρήση μπεϋζιανών μεθόδων.

2.4 Η Μπεϋζιανή προσέγγιση

Στη μπεϋζιανή συμπερασματολογία, κάθε άγνωστη παράμετρος θεωρείται τυχαία μεταβλητή η οποία εκ των προτέρων (δηλαδή πριν την παρατήρηση των δεδομένων) κατανέμεται σύμφωνα με μία εκ των προτέρων κατανομή. Η συμπερασματολογία δοθέντος ενός δείγματος τιμών από τον υπό μελέτη πληθυσμό γίνεται βάσει της αντίστοιχης εκ των υστέρων κατανομής. Συχνά, η εκ των προτέρων κατανομή καθορίζεται βάσει πρακτικών ευκολιών που παρέχουν μερικές επιλογές για αυτήν παρά βάσει υποκειμενικών πληροφοριών που μπορεί να έχει κάποιος ερευνητής. Μία τέτοια βολική επιλογή είναι η χρήση *συζυγών εκ των προτέρων κατανομών* (conjugate priors), οι οποίες οδηγούν σε γνωστές εκ των υστέρων κατανομές.

Πολλές φορές, δεν υπάρχει καμμία κοινά αποδεκτή εκ των προτέρων κατανομή και σε μερικές από αυτές τις περιπτώσεις μπορούν να χρησιμοποιηθούν *μη πληροφοριακές εκ των προτέρων κατανομές*, όπως για παράδειγμα η ομοιόμορφη κατανομή ή οι κατανομές του Jeffreys (1946). Παρ' όλ' αυτά, όταν ο παραμετρικός χώρος δεν είναι ένα σύνολο πεπερασμένου μέτρου, τότε οι μη πληροφοριακές εκ των προτέρων κατανομές είναι *καταχρηστικές*. Γενικά, οι *καταχρηστικές κατανομές* (improper priors) είναι θετικά μέτρα παρά μέτρα πιθανότητας. Για περισσότερες πληροφορίες, ο αναγνώστης παραπέμπεται στους Bernardo and Smith (1994).

Στα μοντέλα μείξεων κατανομών, η χρήση καταχρηστικών εκ των προτέρων κατανομών είναι απαγορευτική. Όπως θα γίνει σαφές στην συνέχεια, αν η εκ των προτέρων κατανομή των παραμέτρων είναι καταχρηστική τότε το ίδιο ισχύει και για την εκ των υστέρων κατανομή τους, κάτι που κάνει αδύνατη την συμπερασματολογία. Το γεγονός αυτό δεν παίζει ουσιάδη ρόλο, καθώς η χρησιμοποίηση εκ των προτέρων κατανομών με αρκετά μεγάλες διασπορές αποτελεί μια συνηθισμένη και αποδοτική υπόθεση. Σε ό,τι ακολουθεί θα χρησιμοποιηθεί ο γενικός συμβολισμός $f(\cdot|y)$ για να δηλώσει την δεσμευμένη συνάρτηση πυκνότητας της μεταβλητής που θα μπαίνει ως όρισμα στην f και δεν θα πρέπει να συγχέεται με την κατανομή των δεδομένων.

Έτσι λοιπόν, θα υποτεθεί ότι οι παράμετροι του μοντέλου (2.3) ακολουθούν εκ

των προτέρων μία κατανομή

$$(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\delta}, \boldsymbol{\beta}) \sim f(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\delta}, \boldsymbol{\beta}). \quad (2.16)$$

Τα $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)$ και $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$ είναι υπερπαραμέτροι, οι οποίες ανάλογα με την περίπτωση θα θεωρούνται ως σταθερές ποσότητες ή τυχαίες μεταβλητές. Συνήθως υποτίθεται εκ των προτέρων ανεξαρτησία μεταξύ των βαρών και των παραμέτρων των συνιστωσών και αυτό σημαίνει ότι η (2.16) μπορεί να γραφτεί ως

$$(\mathbf{p}, \boldsymbol{\theta} | \boldsymbol{\delta}, \boldsymbol{\beta}) \sim f(\mathbf{p} | \boldsymbol{\delta}) f(\boldsymbol{\theta} | \boldsymbol{\beta}). \quad (2.17)$$

Μία τυπική επιλογή για την εκ των προτέρων κατανομή των βαρών είναι η κατανομή Dirichlet, ενώ η συνηθισμένη πρακτική υποθέτει εκ των προτέρων ανεξαρτησία μεταξύ των παραμέτρων των συνιστωσών. Έτσι λοιπόν σε ό,τι ακολουθεί θα υποθεθεί ότι

$$\mathbf{p} | \boldsymbol{\delta} \sim \mathcal{D}(\delta_1, \dots, \delta_k), \quad (2.18)$$

$$\boldsymbol{\theta} | \boldsymbol{\beta} \sim \prod_{j=1}^k f(\theta_j | \beta_j). \quad (2.19)$$

Σε αυτό το σημείο τονίζεται ότι στην περίπτωση του μοντέλου (2.3) όπου όλες οι συνιστώσες ανήκουν στην ίδια παραμετρική οικογένεια κατανομών, δεν είναι δυνατόν να υπάρξει κάποιου είδους εκ των προτέρων πληροφορία που να ξεχωρίζει τις συνιστώσες μεταξύ τους. Έτσι, για λόγους συνέπειας σχετικά με την όσο το δυνατόν περισσότερο μη πληροφοριακή θεμελίωση των εκ των προτέρων υποθέσεων, η συνηθισμένη υπόθεση που γίνεται εδώ είναι ότι $\delta_1 = \dots = \delta_k = \delta$, όπως και $\beta_1 = \dots = \beta_k = \beta$. Αυτό σημαίνει ότι οι παράμετροι των συνιστωσών είναι εκ των προτέρων ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Μια περαιτέρω υπόθεση που χρησιμοποιείται επίσης εδώ είναι ότι $\delta = 1$ κάτι που αντιστοιχεί στην ομοιόμορφη κατανομή στον χώρο \mathcal{P}_k .

Η γενική παρουσίαση των εκ των προτέρων υποθέσεων μπορεί να διευρυνθεί συμπεριλαμβάνοντας και τα διανύσματα κατάταξης των παρατηρήσεων στις συνιστώσες της μείξης. Κάτι τέτοιο δεν είναι απαραίτητο βέβαια για τη μπεϋζιανή εξαγωγή συμπερασμάτων για τις παραμέτρους του μοντέλου, αλλά διευκολύνει αρκετά την διαδικασία. Από την άλλη πλευρά, σε πολλές περιπτώσεις ενδιαφέρει η εξαγωγή συμπερασμάτων όχι μόνο για τις παραμέτρους των συνιστωσών αλλά και για τις κατατάξεις των παρατηρήσεων στις συνιστώσες (clustering). Όπως ήδη συζητήθηκε στην Ενότητα 2.2.1, η εκ των προτέρων κατανομή των μεταβλητών κατάταξης είναι η πολυωνυμική κατανομή (2.8).

Κατ' αυτόν τον τρόπο, η προσέγγιση που ακολουθείται μπορεί να θεωρηθεί ως ένα ιεραρχικό μοντέλο. Στο πιο υψηλό επίπεδο βρίσκεται φυσικά ο αριθμός των συνιστωσών όπου προς το παρόν θεωρείται γνωστός. Ακολουθώντας, οι υπερπαραμέτροι δ και β καθορίζουν το αμέσως επόμενο επίπεδο στην ιεραρχία, όπου βρίσκονται οι παράμετροι του μοντέλου $(\mathbf{p}, \boldsymbol{\theta})$. Κατόπιν, μία πραγματοποίηση των βαρών \mathbf{p} καθορίζει την συμπεριφορά του διανύσματος κατάταξης \mathbf{z} του οποίου η κατανομή εξαρτάται από αυτά. Δοθέντος \mathbf{z} και μιας πραγματοποίησης του $\boldsymbol{\theta}$, παράγονται τα δεδομένα \mathbf{x} τα οποία βρίσκονται στο χαμηλότερο ιεραρχικό επίπεδο. Είναι σαφές ότι η εκ των υστέρων κατανομή των παραμέτρων του μοντέλου γράφεται ως

$$f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) = C^{-1} L(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) f(\mathbf{p}, \boldsymbol{\theta}) \quad (2.20)$$

$$= C^{-1} \sum_{\mathbf{z} \in \mathcal{Z}} L^c(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z}) w(\mathbf{z} | \mathbf{x}) f(\mathbf{p}, \boldsymbol{\theta}), \quad (2.21)$$

όπου

$$C = \int_{\mathcal{P}_k \times \Theta^k} \sum_{\mathbf{z} \in \mathcal{Z}} L^c(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z}) w(\mathbf{z} | \mathbf{x}) f(\mathbf{p}, \boldsymbol{\theta}) d\mathbf{p} d\boldsymbol{\theta}.$$

Το $w(\mathbf{z} | \mathbf{x})$ είναι το εκ των υστέρων βάρος του \mathbf{z} , το οποίο δίνεται από την

$$w(\mathbf{z} | \mathbf{x}) = \int_{\mathcal{P}_k \times \Theta^k} f(\mathbf{z}, \mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) d\mathbf{p} d\boldsymbol{\theta} \quad (2.22)$$

$$\propto \int_{\mathcal{P}_k \times \Theta^k} f(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}) f(\mathbf{z} | \mathbf{p}) f(\mathbf{p}, \boldsymbol{\theta}) d\mathbf{p} d\boldsymbol{\theta}. \quad (2.23)$$

Λόγω των (2.7) και (2.8) προκύπτει ότι

$$f(\mathbf{z} | \mathbf{p}) = \prod_{i=1}^n M_k(1; p_1, \dots, p_k)$$

και

$$f(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i | \theta_{z_i}).$$

Από την (2.21) προκύπτει άμεσα ότι αν η εκ των προτέρων κατανομή είναι καταχρηστική, τότε το ίδιο θα ισχύει και για την εκ των υστέρων κατανομή. Πράγματι, αν

$$\int_{\Theta^k} f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty,$$

τότε για κάθε n ,

$$\int_{\mathcal{P}_k \times \Theta^k} f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) d\mathbf{p} d\boldsymbol{\theta} = \infty$$

διότι μεταξύ των k^n όρων του αθροίσματος στη (2.21) υπάρχουν συνολικά

$$\sum_{j=1}^{k-1} \binom{k}{j} (k-j)^n$$

όροι όπου αντιστοιχούν στα διανύσματα \mathbf{z} που περιέχουν j κενές συνιστώσες, για $j = 1, \dots, k - 1$. Γενικά, η συνιστώσα j θα λέγεται «κενή» όταν το διάνυσμα \mathbf{z} είναι τέτοιο ώστε $z_i \neq j, i = 1, \dots, n$. Επομένως, λόγω της (2.19), οι αντίστοιχες δεσμευμένες εκ των υστέρων κατανομές $f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z})$ περιέχουν j όρους που είναι ίσοι με την εκ των προτέρων κατανομή. Γι' αυτό, σε ό,τι ακολουθεί υποθέτουμε ότι η εκ των προτέρων κατανομή είναι μη καταχρηστική.

Από την έκφραση (2.21) μπορούμε να δούμε ότι ένας εκτιμητής Bayes των παραμέτρων $\mathbf{p}, \boldsymbol{\theta}$ είναι ο:

$$(\hat{\mathbf{p}}, \hat{\boldsymbol{\theta}}) = \sum_{\mathbf{z} \in \mathcal{Z}} w(\mathbf{z} | \mathbf{x}) \mathbb{E}(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z}). \quad (2.24)$$

Στην περίπτωση που χρησιμοποιούνται συζυγείς εκ των προτέρων κατανομές, είναι διαθέσιμες οι δεσμευμένες εκ των υστέρων κατανομές $f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}, \mathbf{z})$. Η (2.24) ερμηνεύεται ως εξής: Σε κάθε δυνατή κατάταξη \mathbf{z} των παρατηρήσεων αντιστοιχίζεται το εκ των υστέρων βάρος και στην συνέχεια η εκτίμηση Bayes των παραμέτρων προκύπτει ως σταθμισμένος μέσος των εκτιμήσεων για τις δεσμευμένες μέσες τιμές αυτών. Επί πλέον, μία εκτίμηση Bayes για την κατάταξη της i -παρατήρησης, $i = 1, \dots, n$, δίνεται από τον τύπο

$$\hat{z}_i = \max_j \{P(Z_i = j | \mathbf{x}), j = 1, \dots, k\}, \quad (2.25)$$

όπου

$$P(Z_i = j | \mathbf{x}) = \int_{\mathcal{P}_k \times \Theta^k} P(Z_i = j | \mathbf{x}, \mathbf{p}, \boldsymbol{\theta}) f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) d\mathbf{p} d\boldsymbol{\theta}, \quad (2.26)$$

ενώ η $P(Z_i = j | \mathbf{x}, \mathbf{p}, \boldsymbol{\theta})$ δίνεται από την (2.10). Βέβαια, οι παραπάνω τύποι δεν έχουν μεγάλη πρακτική αξία. Για παράδειγμα, στην (2.24) υπάρχουν k^n όροι και οι υπολογιστικές δυσκολίες είναι μεγάλες ακόμα και για μικρά μεγέθη δείγματος. Τέλος, σημειώνεται ότι οι Casella et al. (2000) σε μια μελέτη Monte Carlo, έδειξαν ότι πολύ λίγα διανύσματα κατάταξης έχουν μη αμελητέο βάρος, συγκρίνοντας με τον συνολικό αριθμό τους. Συνεπώς, κρίνεται απαραίτητη η επικέντρωση σε υπολογιστικές μεθόδους που προσομοιώνουν ένα δείγμα από την εκ των υστέρων κατανομή των παραμέτρων (2.21). Έτσι είναι δυνατή η εκτίμηση εκ των υστέρων μέσω των τιμών της μορφής (2.24) και (2.26) μέσω κατάλληλων μέσων όρων.

2.5 Μέθοδοι MCMC και μείξεις κατανομών

Τα μοντέλα μείξεων κατανομών έχουν υπάρξει η αφορμή για την ανάπτυξη πολλών γενικών μεθοδολογιών στην Υπολογιστική Στατιστική. Εκτός από τον αλγόριθμο EM,

ξεχωρίζει η μέθοδος διεύρυνσης των δεδομένων (data augmentation) που προτάθηκε από τους Tanner and Wong (1987) και η οποία αποτελεί πρόδρομο του δειγματολήπτη Gibbs των Gelfand and Smith (1990). Μία εναλλακτική μέθοδος είναι ο αλγόριθμος Metropolis-Hastings (Metropolis et al., 1953, Hastings, 1970) που δεν απαιτεί την διεύρυνση των δεδομένων. Οι δύο αυτές μέθοδοι ανήκουν στην κατηγορία των αλγορίθμων Markov Chain Monte Carlo (MCMC). Η γενική ιδέα είναι ότι σε αυτούς τους αλγορίθμους προσομοιώνεται μία μαρκοβιανή αλυσίδα με οριακή κατανομή την από κοινού εκ των υστέρων κατανομή των παραμέτρων του μοντέλου. Μία εναλλακτική προσέγγιση είναι η τεχνική Population Monte Carlo (Carpe et al., 2003) η οποία βασίζεται στην τεχνική της δειγματοληψίας σπουδαιότητας (importance sampling). Για μία γενική εισαγωγή στις μεθόδους προσομοίωσης Monte Carlo και MCMC ο αναγνώστης παραπέμπεται στους Robert and Casella (2004).

Στην περίπτωση που ο αριθμός συνιστωσών της μείξης είναι άγνωστος, η μπεϋζιανή προσέγγιση αποτελεί και την μόνη δυνατότητα. Μετά την εξάπλωση των μεθόδων MCMC αναπτύχθηκαν αρκετές μέθοδοι εκτίμησης του αριθμού των συνιστωσών. Οι πρώτες προσπάθειες προς αυτήν την κατεύθυνση βασίζονταν στην σύγκριση αποτελεσμάτων από ξεχωριστές εκτελέσεις αλγορίθμων θεωρώντας διάφορες τιμές του k . Πιο συγκεκριμένα, οι Kass and Raftery (1995) και Chib (1996) πρότειναν τη χρήση του παράγοντα Bayes, ενώ οι Mengersen and Robert (1996) σύγκριναν τα επικρατέστερα μοντέλα μέσω της απόκλισης Kullback - Leibler. Η πραγματική όμως επανάσταση στη μπεϋζιανή μοντελοποίηση με μείξεις κατανομών ήρθε το 1997 με την εργασία των Richardson and Green όπου εφάρμοσαν την μέθοδο Reversible Jump MCMC (RJMCMC) του Green (1995). Η καινοτόμος ιδέα αυτής της μεθόδου ήταν η προσομοίωση μιας αλυσίδας Markov που κινείται στον χώρο καταστάσεων όλων των πιθανών μοντέλων ταυτόχρονα και όχι σε κάθε μοντέλο ξεχωριστά. Μία άλλη σημαντική εναλλακτική μέθοδος όπου έχει παρόμοια βασική ιδέα είναι ο αλγόριθμος Birth-Death MCMC του Stephens (2000b). Σημειώνεται ότι οι Carré et al. (2003) έδειξαν ότι ο αλγόριθμος Birth-Death MCMC είναι οριακή περίπτωση του αλγορίθμου RJMCMC.

2.5.1 Δειγματολήπτης Gibbs

Ο δειγματολήπτης Gibbs είναι μία από τις πιο κλασσικές μεθόδους MCMC και η χρήση του είναι ιδιαίτερα διαδεδομένη και στην μπεϋζιανή εκτίμηση μίξεων κατανομών. Η πρώτη εφαρμογή του δειγματολήπτη Gibbs σε προβλήματα μίξεων κατανομών με γνωστό αριθμό συνιστωσών έγινε από τους Diebolt and Robert (1990).

Από εκεί και μετά ακολούθησαν αρκετοί ερευνητές μεταξύ των οποίων οι Lavine and West (1992), Verdinelli and Wasserman (1992), Chib (1995) και Escobar and West (1995).

Η μέθοδος αυτή εκμεταλλεύεται την αναπαράσταση του μοντέλου μέσω των ελλিপών δεδομένων (δες Ενότητα 2.2.1). Σκοπός της είναι η προσομοίωση μιας ακολουθίας τιμών $(\mathbf{z}, \mathbf{p}, \boldsymbol{\theta})^{(t)}$, $t = 1, \dots, M$, που αποτελεί πραγματοποίηση μιας μαρκοβιανής διαδικασίας, τέτοιας ώστε η περιθωριακή οριακή κατανομή της ακολουθίας $(\mathbf{p}, \boldsymbol{\theta})^{(t)}$ να είναι η (2.21). Ο δειγματολήπτης προσομοιώνει σε κάθε βήμα τις ελλιπούσες μεταβλητές χρησιμοποιώντας την πλήρη δεσμευμένη εκ των υστέρων κατανομή τους στη (2.10). Ακολούθως, τα βάρη και οι παράμετροι των συνιστωσών προσομοιώνονται από τις αντίστοιχες πλήρεις δεσμευμένες κατανομές. Να σημειωθεί ότι όταν τα βάρη και οι παράμετροι είναι εκ των προτέρων ανεξάρτητες τυχαίες μεταβλητές τότε ισχύει $f(\mathbf{p}|\mathbf{x}, \boldsymbol{\theta}, \mathbf{z}) = f(\mathbf{p}|\mathbf{z})$. Πιο συγκεκριμένα, αν η εκ των προτέρων κατανομή των βαρών είναι η (2.18), τότε είναι εύκολο να δειχτεί ότι

$$\mathbf{p}|\mathbf{z}, \boldsymbol{\delta} \sim \mathcal{D}(\delta_1 + n_1, \dots, \delta_k + n_k) \quad (2.27)$$

όπου

$$n_j := \sum_{i=1}^n I(z_i = j), \quad j = 1, \dots, k. \quad (2.28)$$

Εδώ με $I(\cdot)$ συμβολίζεται η δείκτρια συνάρτηση. Οι πλήρεις δεσμευμένες εκ των υστέρων κατανομές των παραμέτρων $\boldsymbol{\theta}$ εξαρτώνται από την φύση του εκάστοτε μοντέλου. Στις περισσότερες περιπτώσεις χρησιμοποιούνται συζυγείς εκ των προτέρων κατανομές, οπότε η προσομοίωση από την αντίστοιχη πλήρη δεσμευμένη εκ των υστέρων κατανομή είναι πολύ εύκολη. Έτσι λοιπόν, ένας γενικός δειγματολήπτης Gibbs για μείξεις κατανομών έχει την ακόλουθη μορφή:

Αλγόριθμος 2.5.1. Δειγματολήπτης Gibbs για μείξεις κατανομών

Δώσε κάποιες αρχικές τιμές $(\mathbf{z}, \mathbf{p}, \boldsymbol{\theta})^{(0)}$ και για $t \in \{1, \dots, M\}$ επανάλαβε τα ακόλουθα βήματα:

1. Προσομοίωσε τα $z_i^{(t)}$ από την κατανομή $z_i|\mathbf{p}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}$ στη (2.10), ανεξάρτητα για $i = 1, \dots, n$.
 2. Προσομοίωσε τα βάρη $\mathbf{p}^{(t)}$ από την κατανομή $\mathbf{p}|\mathbf{z}^{(t)}$ στη (2.27).
 3. Προσομοίωσε τις παραμέτρους $\boldsymbol{\theta}^{(t)}$ από τις πλήρεις δεσμευμένες εκ των υστέρων κατανομές τους.
-

Όπως τονίστηκε, το βήμα 3 του Αλγορίθμου 2.5.1 εξαρτάται από την φύση του εκάστοτε μοντέλου. Σημειώνεται ότι αν οι συνιστώσες της μείξης ανήκουν στην εκθετική οικογένεια κατανομών, τότε υπάρχει μια οικογένεια εκ των προτέρων κατανομών που είναι συζυγής της πιθανοφάνειας του μοντέλου. Γενικά, όταν είναι διαθέσιμες συζυγείς κατανομές, οι αντίστοιχες πλήρεις δεσμευμένες εκ των υστέρων κατανομές των παραμέτρων δίνονται από κλειστούς τύπους και έτσι είναι εύκολη η προσομοίωση από αυτές. Για περισσότερες λεπτομέρειες, ο αναγνώστης παραπέμπεται στον Robert (2001). Στην συνέχεια περιγράφονται οι εκ των προτέρων υποθέσεις και οι αντίστοιχες πλήρεις δεσμευμένες κατανομές των παραμέτρων στην περίπτωση μείξεων κανονικών κατανομών.

Παράδειγμα 2.5.1. *Random beta model.*

Έστω $\mathbf{x} = (x_1, \dots, x_n)$ ένα τυχαίο δείγμα από μία μείξη k κανονικών κατανομών

$$x_i \sim \sum_{j=1}^k p_j \mathcal{N}(\mu_j, \sigma_j^2), \quad i = 1, \dots, n.$$

Το ιεραρχικό μοντέλο των Richardson and Green (1997) για την εκτίμηση μείξεων κανονικών κατανομών θα χρησιμοποιηθεί αρκετές φορές σε αυτή την διατριβή. Οι παράμετροι του μοντέλου, εκτός από τα βάρη, είναι η μέση τιμή και η διασπορά κάθε συνιστώσας, συνεπώς $\theta_j = (\mu_j, \sigma_j^2)$, $j = 1, \dots, k$, όπου το k προς το παρόν θεωρείται γνωστή σταθερά. Οι εκ των προτέρων κατανομές των διανυσμάτων κατάταξης και των βαρών είναι οι (2.8) και (2.18) που ήδη συζητήθηκαν. Στην συνέχεια παρατίθενται οι εκ των προτέρων υποθέσεις για τις παραμέτρους θ_j , $j = 1, \dots, k$:

$$\mu_j \sim \mathcal{N}(\xi, \kappa^{-1}), \text{ ανεξάρτητα για } j = 1, \dots, k$$

$$\sigma_j^{-2} \sim \mathcal{G}(\alpha, \beta), \text{ ανεξάρτητα για } j = 1, \dots, k$$

$$\beta \sim \mathcal{G}(g, h)$$

$$\alpha = 2$$

$$\kappa = 1/R^2$$

$$g = 0.2$$

$$h = 10/R^2$$

$$R = x_{(n)} - x_{(1)}.$$

Σημειώνεται ότι με $\mathcal{G}(\alpha, \beta)$ συμβολίζεται η κατανομή γάμμα με μέση τιμή α/β . Επειδή η υπερπάρμετρος β είναι τυχαία μεταβλητή, το παραπάνω ιεραρχικό μοντέλο

καλείται random beta model. Με τις παραπάνω υποθέσεις προκύπτουν συζυγείς εκ των υστέρων κατανομές, οι οποίες έχουν ως εξής:

$$\begin{aligned}\mu_j | \mathbf{z}, \boldsymbol{\sigma}^2, \mathbf{x} &\sim \mathcal{N} \left(\frac{\sigma_j^{-2} s_j + \kappa \xi}{\kappa + \sigma_j^{-2} n_j}, \frac{1}{\kappa + \sigma_j^{-2} n_j} \right), \text{ ανεξάρτητα για } j = 1, \dots, k \\ \sigma_j^{-2} | \mathbf{z}, \boldsymbol{\mu}, \beta, \mathbf{x} &\sim \mathcal{G} \left(\alpha + \frac{n_j}{2}, \beta + r_j \right), \text{ ανεξάρτητα για } j = 1, \dots, k \\ \beta | \boldsymbol{\sigma}^2 &\sim \mathcal{G} \left(g + \kappa \alpha, h + \sum_{j=1}^k \sigma_j^{-2} \right),\end{aligned}$$

όπου

$$n_j := \sum_{i=1}^n I(z_i = j), \quad s_j := \sum_{i=1}^n I(z_i = j) x_i, \quad r_j := \sum_{i=1}^n I(z_i = j) (x_i - \mu_j)^2, \quad j = 1, \dots, k. \quad (2.29)$$

Σε αυτήν την περίπτωση λοιπόν, στο βήμα 3 του Αλγορίθμου 2.5.1 προσομοιώνονται οι μέσες τιμές και οι διασπορές των κανονικών συνιστωσών χρησιμοποιώντας τις παραπάνω πλήρεις δεσμευμένες εκ των υστέρων κατανομές. Επίσης προστίθεται ένα επί πλέον βήμα για την ανανέωση της υπερπαραμέτρου β . Μία εφαρμογή του random beta model δίνεται στην Ενότητα 2.5.3.

2.5.2 Αλγόριθμος Metropolis-Hastings

Όπως περιγράφηκε στην Ενότητα 2.5.1, ο δειγματολήπτης Gibbs προσομοιώνει εκτός από τις παραμέτρους και τις ελλειπούσες μεταβλητές κατάταξης. Η διαδικασία αυτή μπορεί να αποφευχθεί με την εναλλακτική μέθοδο του αλγορίθμου Metropolis-Hastings, ο οποίος δεν απαιτεί την διεύρυνση των δεδομένων.

Έστω π η οριακή κατανομή της αλυσίδας, η οποία στο πλαίσιο των μείξεων είναι η (2.20). Η μέθοδος βασίζεται σε μία κατανομή πρότασης q από την οποία παράγονται οι υποψήφιος τιμές. Στην συνέχεια οι τιμές αυτές γίνονται αποδεκτές με κάποια πιθανότητα. Σε αντίθετη περίπτωση η αλυσίδα παραμένει στην προηγούμενη κατάσταση. Στην περίπτωση των μείξεων κατανομών, μία γενική μορφή του αλγορίθμου Metropolis-Hastings δίνεται στον Αλγόριθμο 2.5.2.

Η επιλογή της κατανομής πρότασης q έχει καθοριστική σημασία για τον αλγόριθμο. Η ελάχιστη απαίτηση για αυτήν είναι ότι πρέπει να έχει τουλάχιστον το ίδιο στήριγμα με την κατανομή-στόχο του αλγορίθμου, δηλαδή την π . Μία συνηθισμένη επιλογή για την κατανομή πρότασης είναι μια συμμετρική κατανομή (π.χ. κανονική με μέση τιμή την προηγούμενη κατάσταση). Βέβαια, τέτοιου είδους επιλογές δεν είναι πάντοτε αποτελεσματικές: Για παράδειγμα, αυτό ισχύει για τα βάρη των συνι-

στωσών ή τις διασπορές μιας μείξης κανονικών κατανομών, λόγω ενδεχόμενων περιορισμών του στηρίγματος κάποιων παραμέτρων. Αυτή η δυσκολία ωστόσο μπορεί να ξεπεραστεί εύκολα με αναπαραμετροποίηση του μοντέλου μείξης (για περισσότερες λεπτομέρειες δεξ Marin et al., 2005).

Αλγόριθμος 2.5.2. Αλγόριθμος Metropolis-Hastings για μίξεις κατανομών

Δώσε κάποιες αρχικές τιμές $(\mathbf{p}, \boldsymbol{\theta})^{(0)}$ και για $t \in \{1, \dots, M\}$ επανάλαβε τα ακόλουθα βήματα :

1. Πρότεινε τις τιμές των παραμέτρων

$$(\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}) \sim q(\cdot | (\mathbf{p}, \boldsymbol{\theta})^{(t-1)}).$$

2. Υπολόγισε το

$$r = \frac{L(\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}} | \mathbf{x}) f(\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}) q((\mathbf{p}, \boldsymbol{\theta})^{(t-1)} | \tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}})}{L((\mathbf{p}, \boldsymbol{\theta})^{(t-1)} | \mathbf{x}) f((\mathbf{p}, \boldsymbol{\theta})^{(t-1)}) q(\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}} | (\mathbf{p}, \boldsymbol{\theta})^{(t-1)})}.$$

3. Θέσε

$$(\mathbf{p}, \boldsymbol{\theta})^{(t)} = (\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}})$$

με πιθανότητα $\min\{1, r\}$, διαφορετικά θέσε

$$(\mathbf{p}, \boldsymbol{\theta})^{(t)} = (\mathbf{p}, \boldsymbol{\theta})^{(t-1)}.$$

2.5.3 Αλγόριθμος RJMCMC για μίξεις κανονικών κατανομών

Είδαμε ότι ο δειγματολήπτης Gibbs και ο αλγόριθμος Metropolis-Hastings μπορούν να εφαρμοστούν για την εκτίμηση των παραμέτρων μίξεων κατανομών, με την προϋπόθεση ότι ο αριθμός συνιστωσών k είναι γνωστός. Σε αυτήν την ενότητα θεωρείται το γενικότερο πρόβλημα όπου το k είναι μέρος των προς εκτίμηση παραμέτρων. Ο αλγόριθμος RJMCMC επινοήθηκε από τον Green (1995) και εφαρμόστηκε σε μίξεις κανονικών κατανομών από τους Richardson and Green (1997). Σε γενικές γραμμές, αποτελεί μία προσαρμογή του αλγορίθμου Metropolis-Hastings για την προσομοίωση μιας μαρκοβιανής αλυσίδας που έχει χώρο καταστάσεων με μη σταθερή διάσταση. Βέβαια, η επέκταση αυτή παρουσιάζει αρκετές απαιτήσεις τόσο θεωρητικές όσο και υπολογιστικές. Για μία γενική περιγραφή της μεθόδου RJMCMC ο αναγνώστης

παραπέμπεται στην Ενότητα A.1 του Παραρτήματος.

Σε ό,τι ακολουθεί υποτίθεται ότι η πυκνότητα (2.2) των δεδομένων αποτελείται από κανονικές συνιστώσες όπου το k πλέον είναι άγνωστο. Δοθέντος k , η εκ των προτέρων δομή στις παραμέτρους του μοντέλου ακολουθεί τις ίδιες υποθέσεις με το random beta model που ορίστηκε στην Ενότητα 2.5.1. Μία μικρή διαφοροποίηση ωστόσο, είναι η εκ των προτέρων διάταξη των μέσων κατά αύξουσα τάξη: Για κάθε k οι παράμετροι του μοντέλου, $(\mathbf{p}, \boldsymbol{\theta})$, θα ανήκουν στον περιορισμένο χώρο $\mathcal{P}_k \times \Theta_{\mu_1 < \dots < \mu_k}^k$. Αυτό γίνεται για δύο λόγους. Πρώτον, η διάταξη των μέσων κάνει πιο εύκολο τον σχεδιασμό προτεινόμενων κινήσεων εναλλαγής διάστασης. Δεύτερον, μέσω αυτού του περιορισμού οι παράμετροι του μοντέλου γίνονται ταυτοποιήσιμες (για περισσότερες λεπτομέρειες δες Ενότητα 2.6).

Οι Richardson and Green (1997) θεωρούν ως εκ των προτέρων κατανομή του αριθμού συνιστωσών την ομοιόμορφη κατανομή στο σύνολο των ακεραίων από το 1 έως κάποιο προκαθορισμένο άνω φράγμα. Στις εφαρμογές που θα παρουσιαστούν ως άνω φράγμα θεωρείται η τιμή $k_{max} = 30$. Οι κινήσεις αλλαγής διάστασης γεφυρώνουν γειτονικά μοντέλα, δηλαδή μοντέλα όπου το πλήθος των συνιστωσών τους διαφέρει κατά μία μονάδα. Το σημείο αυτό είναι πολύ σημαντικό για την καλή απόδοση του αλγορίθμου RJMCMC. Μία κίνηση αλλαγής διάστασης θα πρέπει να σχεδιάζεται ως ένα καλώς ορισμένο ζεύγος αντιστρέψιμων μεταβάσεων συγκεκριμένου τύπου, με την έννοια ότι θα πρέπει να ικανοποιούν την υπόθεση συμφωνίας διαστάσεων (A.1) του Green. Αυτό αποτελεί μια ικανή συνθήκη για την σύγκλιση του αλγορίθμου. Επί πλέον, οι κινήσεις αυτές θα πρέπει να είναι προσεκτικά σχεδιασμένες, έτσι ώστε ο αλγόριθμος να εξερευνά επαρκώς το σύνολο των δυνατών μοντέλων. Οι Richardson and Green (1997) χρησιμοποίησαν δύο τύπους κινήσεων αλλαγής μοντέλου: τον διαχωρισμό μιας συνιστώσας σε δύο ή τον συνδυασμό δύο συνιστωσών σε μία (split-combine moves) και την γέννηση ή τον θάνατο μιας κενής συνιστώσας (birth-death moves). Οι κινήσεις διαχωρισμού και γέννησης αυξάνουν τον αριθμό των συνιστωσών ενώ οι κινήσεις συνδυασμού και θανάτου τον μειώνουν.

Τονίζεται το γεγονός ότι οι κινήσεις αυτές δρουν συμπληρωματικά μεταξύ τους. Το πρώτο ζευγάρι κινήσεων (split-combine) δρα «τοπικά», με την έννοια ότι προτείνει αλλαγές μεταξύ γειτονικών συνιστωσών (δηλαδή οι μέσες τιμές αυτών είναι τέτοιες ώστε να μην παρεμβάλλεται ο μέσος κάποιας άλλης συνιστώσας αναμεσά τους). Επί πλέον η κίνηση αυτού του τύπου αντλεί πληροφορία από την εκάστοτε κατάσταση της αλυσίδας καθώς οι προτεινόμενες τιμές είναι τέτοιες ώστε να μένουν αναλλοίωτες οι δύο πρώτες ροπές της μείξης. Σημειώνουμε ότι η εκ των προτέρων υπόθεση διά-

ταξης των μέσων είναι απαραίτητη για την κίνηση αυτή καθώς έτσι ταυτοποιούνται οι γειτονικές συνιστώσες που συμμετέχουν στην προτεινόμενη μετάβαση. Αντίθετα, το δεύτερο ζευγάρι κινήσεων (birth-death) προτείνει πιο «τυχαίες» αλλαγές στην κατάσταση της αλυσίδας.

Αλγόριθμος 2.5.3. *Αλγόριθμος RJMCMC για μείξεις κανονικών κατανομών*

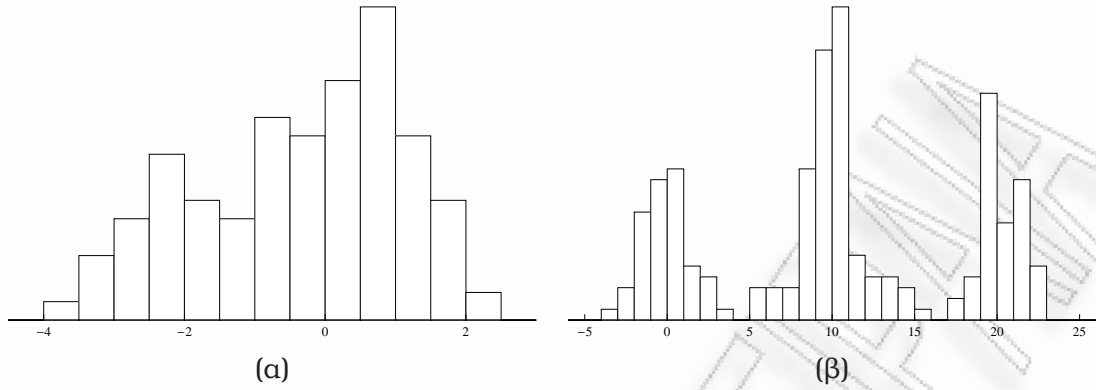
Δώσε κάποιες αρχικές τιμές $k^{(0)}, z^{(0)}, (\mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^{(0)}$, και για $t \in \{1, \dots, M\}$, επανάλαβε τα παρακάτω βήματα:

1. Διαχωρισμός μιας συνιστώσας σε δύο ή συνδυασμός δύο συνιστωσών σε μία.
 2. Γέννηση ή θάνατος μιας κενής συνιστώσας.
 3. Ανανέωση των βαρών $\mathbf{p} = (p_1, \dots, p_k)$.
 4. Ανανέωση των μέσων $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$.
 5. Ανανέωση των διασπορών $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2)$.
 6. Ανανέωση του διανύσματος κατάταξης $\mathbf{z} = (z_1, \dots, z_n)$.
 7. Ανανέωση της υπερπαραμέτρου β .
-

Στα βήματα 3 – 7 του Αλγορίθμου 2.5.3 χρησιμοποιούνται οι πλήρεις δεσμευμένες εκ των υστέρων κατανομές των παραμέτρων που παρουσιάστηκαν στην Ενότητα 2.5.1. Μία βασική διαφορά εδώ είναι ότι η πλήρης δεσμευμένη κατανομή των μέσων χρησιμοποιείται μόνο για την πρόταση μιας νέας τιμής. Αυτή γίνεται αποδεκτή μόνο αν δεν παραβιάζεται η διάταξη των μέσων, διαφορετικά επαναλαμβάνεται η προηγούμενη τιμή. Για μία αναλυτική περιγραφή των βημάτων του αλγορίθμου, ο αναγνώστης παραπέμπεται στην εργασία των Richardson and Green (1997) (δες επίσης και Παπασταμούλης, 2006). Γενικά, ο παραπάνω αλγόριθμος είναι σχεδιασμένος για να εκτιμά το πλήθος συνιστωσών και τις παραμέτρους αυτών σε δεδομένα που μπορεί να θεωρηθεί ότι προσεγγίζονται από μία μείξη κανονικών κατανομών και αποδίδει καλά σε τυπικές περιπτώσεις. Παρ' όλ' αυτά, στα παραδείγματα που ακολουθούν γίνεται σαφές ότι υπάρχουν κάποιες περιπτώσεις όπου ο αλγόριθμος των Richardson and Green (1997) δεν εκτιμά σωστά τις συνιστώσες της μείξης.

Παράδειγμα 2.5.2. *Εφαρμογή του Αλγορίθμου 2.5.3 σε προσομοιωμένα σύνολα δεδομένων.*

Ας υποθέσουμε ότι διαθέτουμε τα δείγματα μεγέθους 200 παρατηρήσεων τα ιστογράμματα των οποίων απεικονίζονται στο Σχήμα 2.5. Τα δεδομένα αυτά προσομοιώ-



Σχήμα 2.5: Προσομοιωμένα δεδομένα από τις μείξεις (2.30) (α) και (2.31) (β).

θηκαν από τις μείξεις κανονικών κατανομών

$$\frac{1}{3}\mathcal{N}(-2, 1) + \frac{1}{3}\mathcal{N}(0, 0.5) + \frac{1}{3}\mathcal{N}(1, 0.25) \quad (2.30)$$

$$\frac{1}{4}\mathcal{N}(0, 2) + \frac{1}{4}\mathcal{N}(10, 7) + \frac{1}{4}\mathcal{N}(10, 1) + \frac{1}{4}\mathcal{N}(20, 1), \quad (2.31)$$

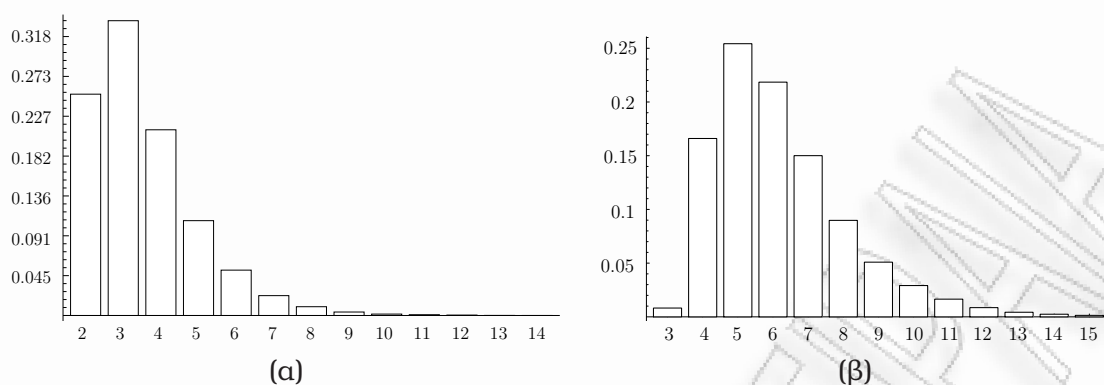
αντίστοιχα. Προκειμένου να εκτιμήσουμε τον αριθμό συνιστωσών που αποτελούν τις μείξεις αυτές, εκτελέσαμε τον Αλγόριθμο 2.5.3 για 300000 επαναλήψεις (αφού προηγήθηκαν 30000 επαναλήψεις ως περίοδος burn-in) και προέκυψε η εκ των υστέρων κατανομή του αριθμού των συνιστωσών του Σχήματος 2.6. Για τα δεδομένα από την μείξη (2.30) παρατηρούμε ότι η εκ των υστέρων κατανομή υποδεικνύει ότι ο πιο πιθανός αριθμός συνιστωσών είναι τρεις, με $\hat{P}(k = 3|\mathbf{x}) = 0.333$. (Αυτός είναι και ο πραγματικός αριθμός συνιστωσών.) Στην συνέχεια, μπορούμε να εκτιμήσουμε τις παραμέτρους του μοντέλου τριών συνιστωσών δεσμεύοντας στο ενδεχόμενο $k = 3$. Έτσι καταλήγουμε στις ακόλουθες εκτιμήσεις των εκ των υστέρων μέσων τιμών:

$$\hat{\mathbb{E}}(\boldsymbol{\mu}|\mathbf{x}) = (-2.33, -0.41, 0.87)$$

$$\hat{\mathbb{E}}(\boldsymbol{\sigma}^2|\mathbf{x}) = (0.52, 0.67, 0.46)$$

$$\hat{\mathbb{E}}(\mathbf{p}|\mathbf{x}) = (0.258, 0.326, 0.416).$$

Για τα δεδομένα της μείξης (2.31), στο Σχήμα 2.6 παρατηρούμε ότι, παρά το σχετικά μεγάλο μέγεθος δείγματος, η κορυφή της εκτιμηθείσας εκ των υστέρων κατανομής αντιστοιχεί στο $k = 5$ (ενώ στην πραγματικότητα υπάρχουν τέσσερεις συνιστώσες). Αυτό είναι μάλλον μη αναμενόμενο, καθώς μόνο τρεις κορυφές είναι εμφανείς στο ιστόγραμμα. Βέβαια, αφού η πραγματική μείξη έχει ίσους τους δύο «μεσαίους» μέσους, ένα μοντέλο τριών συνιστωσών θα είχε πιθανότητα φτωχή προσαρμογή στα



Σχήμα 2.6: Οι εκτιμηθείσες εκ των υστέρων κατανομές των αριθμού συνιστώσων μέσω του αλγορίθμου RJMCMC για τα δεδομένα του Παραδείγματος 2.5.2. (α) Μείξη (2.30). (β) Μείξη (2.31).

δεδομένα, μια που το κύριο μέρος αυτών προκύπτει από τις δύο «μεσαίες» συνιστώσες που έχουν αρκετά μεγάλη διαφορά στις διασπορές. Από την άλλη πλευρά, ένα μοντέλο τεσσάρων συνιστώσων (δηλαδή το πραγματικό) αποτυγχάνει να κυριαρχήσει στην εκ των υστέρων κατανομή διότι η περιοχή υψηλής εκ των υστέρων πυκνότητας βρίσκεται στο σύνορο του περιορισμένου παραμετρικού χώρου, καθώς ο δεύτερος και τρίτος μέσος ταυτίζονται. Πράγματι, αν δεσμεύσουμε στο μοντέλο με $k = 4$, τότε οι εκτιμήσεις των εκ των υστέρων μέσων τιμών των μέσων της δεύτερης και τρίτης συνιστώσας είναι αρκετά μακριά από την πραγματική κοινή τιμή τους. Φυσικά κάτι τέτοιο ήταν αναμενόμενο, λόγω της διάταξης των μέσων. Εδώ αναφέρουμε ότι αυτό το πρόβλημα έχει σχολιαστεί από τον Stephens (1997b) στην συζήτηση της εργασίας των Richardson and Green, ο οποίος (σε ένα παρόμοιο παράδειγμα) αναφέρει χαρακτηριστικά ότι

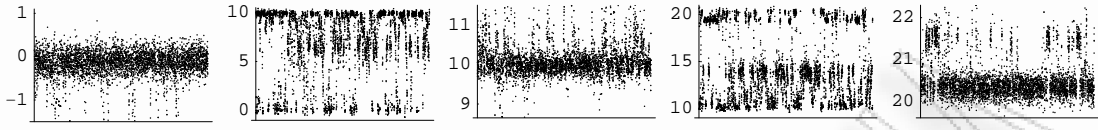
«...η αναδιάταξη βάσει του περιορισμού $\mu_1 < \mu_2 < \mu_3$ αποτυγχάνει να χωρίσει τις δύο συνιστώσες με κοινό μέσο...»

Εδώ βλέπουμε ότι κάτι τέτοιο δεν επηρεάζει μόνο την διαδικασία των εκτιμήσεων, αλλά κυρίως έχει επιπτώσεις στην εκ των υστέρων κατανομή, η κορυφή της οποίας υπερεκτιμά τον πραγματικό αριθμό συνιστώσων στο συγκεκριμένο παράδειγμα.

Δεσμεύοντας στο μοντέλο με $k = 5$ συνιστώσες, οι εργοδικοί μέσοι των τιμών που προσομοιώθηκαν από τον τυπικό αλγόριθμο RJMCMC ήταν

$$\hat{\mathbb{E}}(\boldsymbol{\mu}|\mathbf{x}) = (-0.59, 7.11, 10.09, 14.27, 20.93)$$

$$\hat{\mathbb{E}}(\boldsymbol{\sigma}^2|\mathbf{x}) = (1.98, 3.88, 3.41, 3.69, 1.57)$$



Σχήμα 2.7: Μείξη (2.31): Οι προσομοιωμένες τιμές των μέσων για τις τελευταίες 5000 επαναλήψεις του αλγορίθμου 2.5.3 δεσμεύοντας στο μοντέλο των πέντε συνιστωσών.

$$\widehat{\mathbb{E}}(\mathbf{p}|\mathbf{x}) = (0.23, 0.14, 0.28, 0.12, 0.23)$$

Συγκρίνοντας τις παραπάνω εκτιμήσεις με τις πραγματικές τιμές των παραμέτρων του αληθινού μοντέλου είναι λογικό να διερωτηθούμε αν οι εκτιμήσεις των μέσων για την δεύτερη και τέταρτη συνιστώσα είναι κατά κάποιον τρόπο δικαιολογημένες. Για να απαντήσουμε σε αυτό θα πρέπει να κοιτάξουμε πιο προσεκτικά την εκτιμηθείσα εκ των υστέρων κατανομή.

Στο Σχήμα 2.7 αναπαρίστανται οι προσομοιωμένες τιμές των μέσων για τις τελευταίες 5000 επαναλήψεις του τυπικού αλγορίθμου RJMCMC δεσμεύοντας στο μοντέλο των πέντε συνιστωσών. Παρατηρούμε ότι οι τιμές των μ_2 και μ_4 αντιστοιχούν σε αρκετές περιοχές υψηλής εκ των υστέρων πυκνότητας. Με άλλα λόγια η εκ των υστέρων κατανομή παρουσιάζει πολλές κορυφές. Το φαινόμενο αυτό είναι τυπικό των εκ των υστέρων κατανομών των παραμέτρων ενός μοντέλου μείξης. Κάτι τέτοιο συμβαίνει όταν το μοντέλο που προσαρμόζεται στα δεδομένα μας είναι υπερπαραμετροποιημένο. Ας θυμηθούμε ότι στην περίπτωσή μας το σωστό μοντέλο έχει μόνο τρεις διακεκριμένους μέσους. Επομένως, ένα μοντέλο με $k = 5$ συνιστώσες είναι υπερπαραμετροποιημένο διότι περιλαμβάνει τουλάχιστον δύο επί πλέον παραμέτρους. Εφ' όσον οι προσομοιωμένες τιμές αυτών των παραμέτρων δεν έχουν στην ουσία πραγματικό στόχο, «περιπλανώνται» από έναν «υπάρχοντα» μέσο σε έναν άλλον «υπάρχοντα» μέσο, ενώ ταυτόχρονα επισκέπτονται κάποιες ενδιάμεσες περιοχές. Κάτι τέτοιο απεικονίζεται στο Σχήμα 2.7. Για παράδειγμα, οι προσομοιωμένες τιμές του μ_2 εναλλάσσονται μεταξύ των περιοχών κοντά στο 0, 7 και 10. Επίσης, παρόμοια είναι και η συμπεριφορά του μ_4 . Τέτοιου είδους καταστάσεις παρατηρήσαμε σε αρκετά υπερπαραμετροποιημένα μοντέλα που προσαρμόσαμε σε διάφορα σύνολα δεδομένων. Συμπεραίνουμε ότι η υπερπαραμετροποίηση συνήθως οδηγεί σε μία εκ των υστέρων κατανομή η οποία έχει αρκετές «γνήσιες» κορυφές ανεξαρτήτως αν χρησιμοποιούνται εικονικοί περιορισμοί διάταξης ή όχι.

2.6 Το πρόβλημα label switching

Στην Ενότητα 2.5.3, είδαμε ότι οι Richardson and Green θεώρησαν εκ των προτέρων τον περιορισμό $\mu_1 < \dots < \mu_k$. Αυτό έγινε πρωτίστως για τον εύκολο σχεδιασμό των προτεινόμενων κινήσεων τύπου split-combine. Ανεξάρτητα από αυτό, αναφέραμε προηγουμένως ότι ένας τέτοιος περιορισμός κάνει τις παραμέτρους του μοντέλου μείζης ταυτοποιήσιμες και έτσι στα παραδείγματα που παρουσιάσαμε υπάρχει μια λανθασμένη εντύπωση σχετικά με το πόσο τυπική υπόθεση μπορεί να είναι η διαδικασία των εκτιμήσεων. Στην συνέχεια εξηγείται γιατί συμβαίνει αυτό, επιστρέφοντας στην γενική περίπτωση όπου για κάθε δοθείσα τιμή k του αριθμού συνιστωσών οι παράμετροι κάθε συνιστώσας παίρνουν τιμές στο ίδιο σύνολο, δηλαδή $(\mathbf{p}, \boldsymbol{\theta}) \in \mathcal{P}_k \times \Theta^k$.

Εστω \mathcal{T}_k το σύνολο των μεταθέσεων των δεικτών των συνιστωσών $\{1, \dots, k\}$. Για κάποιο $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$ ας θεωρήσουμε την αντίστοιχη μετάθεση του παραμετρικού διανύσματος

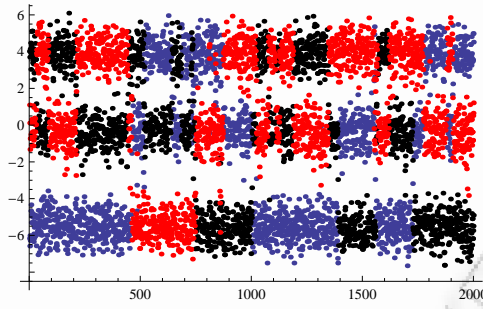
$$\tau(\mathbf{p}, \boldsymbol{\theta}) \equiv (\tau\mathbf{p}, \tau\boldsymbol{\theta}) := (p_{t_1}, \dots, p_{t_k}, \theta_{t_1}, \dots, \theta_{t_k}).$$

Η ρίζα του φαινομένου *label switching* (σύμφωνα με τον όρο που χρησιμοποιήθηκε από τους Redner and Walker, 1984) εντοπίζεται στο γεγονός ότι η πιθανοφάνεια (2.4) είναι αναλλοίωτη ως προς τις μεταθέσεις των δεικτών των συνιστωσών, καθώς είναι προφανές ότι

$$L(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) = L(\tau(\mathbf{p}, \boldsymbol{\theta}) | \mathbf{x}), \quad \forall \tau \in \mathcal{T}_k \quad (2.32)$$

(δες και Stephens, 1997a).

Αν και το παραπάνω φαινόμενο παίζει μεγάλο ρόλο στην μπεϋζιανή συμπερασματολογία όπως θα εξηγηθεί στην συνέχεια, σε αυτό το σημείο αναφέρεται ότι έχει επιπτώσεις και στην κλασσική συμπερασματολογία. Για παράδειγμα, οι συνηθισμένες συνθήκες ομαλότητας (Cramer, 1946) σχετικά με τις ασυμπτωτικές ιδιότητες του εκτιμητή μέγιστης πιθανοφάνειας του $(\mathbf{p}, \boldsymbol{\theta})$ δεν ισχύουν. Αυτό συμβαίνει διότι οι παράμετροι $(p_1, \theta_1), \dots, (p_k, \theta_k)$ δεν είναι ταυτοποιήσιμες, λόγω του ότι μία μετάθεση των δεικτών των συνιστωσών δεν αλλάζει την τιμή της πιθανοφάνειας. Το γεγονός αυτό έχει επιπτώσεις και σε ελέγχους υποθέσεων σχετικούς με μείξεις κατανομών, καθώς ο κλασσικός έλεγχος λόγου πιθανοφανειών δεν έχει γνωστή ασυμπτωτική κατανομή (Hartigan, 1985a, 1985b). Για περισσότερες λεπτομέρειες για τις επιπτώσεις του προβλήματος μη ταυτοποίησης των παραμέτρων στην συμπεριφορά του εκτιμητή μέγιστης πιθανοφάνειας ο αναγνώστης παραπέμπεται στους McLachlan and Peel (2000) και Redner (1981).



Σχήμα 2.9: Η ακολουθία των προσομοιωμένων τιμών $(\mu_1, \mu_2, \mu_3)^{(t)}$, $t = 1, \dots, 2000$.

Εδώ, ο αναλυτικός υπολογισμός της εκ των υστέρων κατανομής είναι δυνατός λόγω του πολύ μικρού μεγέθους δείγματος και της μικρής διάστασης του παραμετρικού χώρου. Τονίζεται το γεγονός ότι σε γενικές περιπτώσεις κάτι τέτοιο είναι απαγορευτικό· αυτός εξ άλλου είναι και ο λόγος που καταφεύγουμε στην προσομοίωση ενός δείγματος από την εκ των υστέρων κατανομή. Ας υποθέσουμε λοιπόν ότι επιθυμούμε να προσομοιώσουμε ένα δείγμα από την κατανομή (2.34). Ας επιλέξουμε τον Αλγόριθμο 2.5.2 χρησιμοποιώντας ως κατανομή πρότασης στο βήμα t τον τυχαίο περίπατο

$$\tilde{\mu}_j \sim \mathcal{N}(\mu_j^{(t-1)}, 2^2), \quad j = 1, 2, 3.$$

Υστερα από $M = 2000$ επαναλήψεις (μετά την διαγραφή των πρώτων 200 τιμών ως περίοδο burn-in), το προσομοιωμένο δείγμα $(\mu_1, \mu_2, \mu_3)^{(t)}$, $t = 1, \dots, M$ έχει ανακαλύψει τις 6 συμμετρικές περιοχές υψηλής εκ των υστέρων πιθανότητας, όπως απεικονίζεται στο Σχήμα 2.8(β). Αυτή είναι και η επιθυμητή συμπεριφορά για την καλή απόδοση ενός αλγορίθμου MCMC που χρησιμοποιείται για την προσέγγιση της εκ των υστέρων κατανομής. Η συμμετρία θα πρέπει να υφίσταται στο προσομοιωμένο δείγμα, διότι πολύ απλά είναι ένα χαρακτηριστικό αυτού που θέλουμε να προσεγγίσουμε.

Παρ' όλ' αυτά, τα προβλήματα που επιφέρει η συμμετρία εμφανίζονται όταν εξετάζουμε περιθωριακά την ακολουθία των προσομοιωμένων τιμών $(\mu_1, \mu_2, \mu_3)^{(t)}$, $t = 1, \dots, M$, η οποία απεικονίζεται στο Σχήμα 2.9. Όπως είναι προφανές οι παράμετροι δεν είναι ταυτοποιήσιμες διότι οι τιμές αυτών εναλλάσσονται συνεχώς μεταξύ των περιοχών υψηλής εκ των υστέρων πυκνότητας. Για παράδειγμα, είναι εύκολο να φανταστεί κάποιος ότι η μέση τιμή των «κόκκινων» τιμών είναι περίπου ίση με τη μέση τιμή είτε των «μπλέ» είτε των «μαύρων» τιμών. Επί πλέον, αυτές οι εκτιμήσεις δεν αποκαλύπτουν κάτι για το τί είναι «διαφορετικό» ανάμεσα στις συνιστώσες της μείξης. Γενικά, είναι απαραίτητη η εξάλειψη του φαινομένου label switching για την

μπεϋζιανή συμπερασματολογία βάσει της εκ των υστέρων κατανομής οποιασδήποτε παραμετρικής συνάρτησης που εξαρτάται² από την ταυτοποίηση των παραμέτρων. Αυτό επιτυγχάνεται είτε καταφεύγοντας σε μεθόδους που αποτρέπουν την εμφάνιση του label switching είτε μέσω κατάλληλων αλγορίθμων αναδιάταξης του προσομοιωμένου δείγματος.

Σύμφωνα με τα παραπάνω, το πρόβλημα μη ταυτοποίησης των παραμέτρων παρουσιάζεται και σε προσομοιωμένα δείγματα από την εκ των υστέρων κατανομή. Κάτι τέτοιο έχει τόσο πλεονεκτήματα όσο και μειονεκτήματα. Από την μία πλευρά αποτελεί αναγκαία συνθήκη για την σύγκλιση των αλγορίθμων MCMC στην κατανομή-στόχο ενώ από την άλλη πλευρά η παρουσία του φαινομένου αυτού καθιστά μη τετριμμένη τη διαδικασία των εκτιμήσεων. Στην βιβλιογραφία έχουν προταθεί αρκετές μέθοδοι αντιμετώπισης του προβλήματος label switching, οι οποίες ποικίλλουν από την υιοθέτηση απλών περιορισμών διάκρισης (Diebolt and Robert, 1994, Richardson and Green, 1997, Frühwirth-Schnatter, 2001) έως πιο σύνθετους αλγορίθμους που βασίζονται στην απόκλιση Kullback-Leibler (Stephens, 1997a, 2000a) ή σε μεθόδους αναδιάταξης μέσω κατάλληλων συναρτήσεων ζημίας (Celeux et al., 2000). Στην συνέχεια περιγράφουμε εν συντομία τις προαναφερθείσες μεθόδους.

2.6.1 Εικονικοί περιορισμοί διάκρισης

Ενας περιορισμός διάκρισης (identifiability constraint) είναι μία συνθήκη στον παραμετρικό χώρο των (p, θ) η οποία ικανοποιείται από ακριβώς³ μία μετάθεση του παραμετρικού διανύσματος. Οι περιορισμοί διάκρισης έχουν χρησιμοποιηθεί κατά κόρον στην βιβλιογραφία: ενδεικτικά αναφέρουμε τους Diebolt and Robert (1994), Dellaportas et al. (1996) και Richardson and Green (1997). Αν και πρόκειται για την πιο κοινή μέθοδο, η μέθοδος έχει υποστεί έντονη κριτική (δες για παράδειγμα τους Celeux, 1997, Celeux et al., 2000, και Stephens, 1997a, 1997b, 2000a). Ένα πρόβλημα της προσέγγισης αυτής είναι η επιλογή του περιορισμού. Επίσης δεν είναι σαφής η εφαρμογή των περιορισμών διάκρισης σε ένα γενικό πλαίσιο, όπως για παράδειγμα σε πολυμεταβλητά προβλήματα. Επί πλέον, υπάρχουν αρκετές περιπτώσεις όπου η εκ των υστέρων κατανομή των παραμέτρων παρουσιάζει παραπάνω

²Υπάρχουν και παραμετρικές συναρτήσεις που δεν εξαρτώνται από την ταυτοποίηση των παραμέτρων, συνεπώς το φαινόμενο label switching δεν αποτελεί πρόβλημα σε αυτές τις περιπτώσεις. Στο Παράδειγμα 2.6.1, η $g(\mu) = \mu_1 + \mu_2 + \mu_3$ είναι μία τέτοια.

³Εξαιρώντας κάποια υποσύνολα του παραμετρικού χώρου που έχουν μηδενικό μέτρο, όπως για παράδειγμα το όριο του περιορισμού.

από μία (μη συμμετρικές) κορυφές⁴. Τότε προκύπτει ότι δεν υπάρχει περιορισμός διάκρισης που να μπορεί να απομονώσει ταυτόχρονα την κύρια και τις μικρότερες κορυφές (δες για παράδειγμα τους Grün and Leisch, 2008).

Μία εναλλακτική προσέγγιση προτάθηκε από την Frühwirth-Schnatter (2001) η οποία χρησιμοποίησε μια μεθοδολογία δύο βημάτων. Πρώτον, προσομοιώνεται ένα δείγμα με στόχο την συμμετρική εκ των υστέρων κατανομή χρησιμοποιώντας έναν δειγματολήπτη τυχαίας μετάθεσης (ΔΤΜ). Σε κάθε βήμα ενός ΔΤΜ, προσομοιώνονται οι παράμετροι του μοντέλου και ακολούθως μετατίθενται τυχαία. Αν ο αλγόριθμος προσομοιώνει και τις μεταβλητές κατάταξης, τότε η τυχαία μετάθεση εφαρμόζεται και σε αυτές. Η Frühwirth-Schnatter (2001) έδειξε ότι το βήμα της τυχαίας μετάθεσης δεν μεταβάλλει την κατανομή-στόχο. Μέσω αυτού όμως εξασφαλίζεται ότι έχουν εξερευνηθεί επαρκώς και οι $k!$ συμμετρικές περιοχές, αφού «εξαναγκάζεται» η παρουσία του φαινομένου label switching. Περισσότερες λεπτομέρειες για τον αλγόριθμο τυχαίας μετάθεσης θα παρουσιαστούν στην Ενότητα 4.3, καθώς μεταξύ άλλων δίνεται ένα νέο αποτέλεσμα για αυτόν. Εξετάζοντας γραφικά (μέσω διαγραμμάτων διασποράς) το δείγμα που προκύπτει από τον ΔΤΜ, η Frühwirth-Schnatter προσπαθεί να ανακαλύψει κατάλληλους περιορισμούς που (ενδεχομένως να) απομονώνουν επαρκώς τις συμμετρικές περιοχές. Ακολούθως, ένας δειγματολήπτης περιορισμένης μετάθεσης (ΔΠΜ) παράγει ένα δείγμα από την περιορισμένη εκ των υστέρων κατανομή, δηλαδή την εκ των υστέρων κατανομή περιορισμένη σε εκείνο το υποσύνολο του παραμετρικού χώρου στο οποίο ικανοποιείται ο περιορισμός διάκρισης που έχει επιλεγεί από τον ΔΤΜ στο προηγούμενο βήμα. Παρ' όλ' αυτά, τα προηγούμενα μειονεκτήματα εξακολουθούν να ισχύουν και σε αυτήν την περίπτωση.

2.6.2 Αλγόριθμος Πιλοτικής Αναδιάταξης

Ο αλγόριθμος πιλοτικής αναδιάταξης (Pivotal Reordering Algorithm, PRA) εισήχθη από τους Marin et al. (2005) (δες επίσης Marin and Robert, 2007) και πρόκειται για μια απλή μέθοδο αντιμετώπισης του φαινομένου label switching. Η γενική ιδέα είναι να αναδιατάξουμε με τέτοιο τρόπο τα προσομοιωμένα παραμετρικά διανύσματα έτσι ώστε να μοιάζουν μεταξύ τους όσο γίνεται περισσότερο. Το πρώτο βήμα για την εφαρμογή της μεθόδου είναι η επιλογή ενός διανύσματος που δρα ως οδηγός (pivot) για την αναδιάταξη των υπολοίπων. Αυτό το διάνυσμα επιλέγεται έτσι ώστε να αντιστοιχεί σε περιοχή μεγάλης εκ των υστέρων πυκνότητας. Η επιλογή που προτεί-

⁴Θα αναφερόμαστε σε τέτοιου είδους εκ των υστέρων κατανομές με τον όρο γνήσιως πολυκόρυφη, σύμφωνα με τον αγγλικό όρο (genuinely multimodal).

νεται από τους Marin et al. (2005), είναι ο Maximum A Posteriori (MAP) εκτιμητής, ο οποίος γενικά ισοδυναμεί με έναν ποινικοποιημένο εκτιμητή μέγιστης πιθανοφάνειας (Ciuperca et al., 2003). Δοθέντος ενός προσομοιωμένου δείγματος μεγέθους M ο συγκεκριμένος οδηγός μπορεί να εκτιμηθεί μέσω της προσέγγισης Monte Carlo του MAP εκτιμητή, δηλαδή, με εκείνη την προσομοιωμένη τιμή $(\theta^{(t^*)}, p^{(t^*)})$ η οποία μεγιστοποιεί την εκ των υστέρων κατανομή:

$$t^* = \arg \max_{t=1, \dots, M} f(\theta^{(t)}, p^{(t)} | \mathbf{x}).$$

Στην συνέχεια, όλα τα προσομοιωμένα παραμετρικά διανύσματα αναδιατάσσονται έτσι ώστε να ελαχιστοποιείται κάποια προεπιλεγμένη απόσταση d από τον οδηγό. Στις μονομεταβλητές περιπτώσεις η απόσταση αυτή είναι η Ευκλείδεια. Έτσι, ο αλγόριθμος αναδιάταξης έχει ως εξής:

Αλγόριθμος 2.6.1. Αλγόριθμος Πιλοτικής Αναδιάταξης

Για $t \in \{1, \dots, M\}$:

1. Υπολόγισε

$$\tau_t = \arg \min_{\tau \in \mathcal{T}_k} d(\tau(\theta^{(t)}, p^{(t)}), (\theta^{(t^*)}, p^{(t^*)})),$$

όπου d είναι μία απόσταση στον παραμετρικό χώρο.

2. Θέσε

$$(\theta^{(t)}, p^{(t)}) = \tau_t(\theta^{(t)}, p^{(t)}).$$

Η μέθοδος αυτή αποδίδει σχετικά καλά σε απλές περιπτώσεις. Όταν όμως η εκ των υστέρων κατανομή είναι γνησίως πολυκόρυφη έχει αρκετά μειονεκτήματα λόγω της αδυναμίας του οδηγού να παράσχει και λιγότερο πιθανές ερμηνείες για τα δεδομένα (δες Jasra et al., 2005).

2.6.3 Αλγόριθμος αναδιάταξης μέσω της απόκλισης Kullback–Leibler

Ο Stephens (1997a, 2000a) ήταν από τους πρώτους που αντιμετώπισε το πρόβλημα label switching θεωρητικά χωρίς να καταφύγει σε απλοϊκές μεθόδους όπως αυτές των περιορισμών διάκρισης. Συγκεκριμένα, ανέπτυξε έναν αλγόριθμο αναδιάταξης του προσομοιωμένου δείγματος με τέτοιο τρόπο ώστε να μεγιστοποιείται η συμφωνία των $n \times k$ πινάκων που περιέχουν τις πιθανότητες κατάταξης $\pi_{ij} = p_j f(x_i | \theta_j) / \sum_{l=1}^k p_l f(x_i | \theta_l)$. Ο Stephens χρησιμοποίησε ως «απόσταση» μεταξύ δύο πινάκων $\Pi = (\pi_{ij})$ και $Q = (q_{ij})$ πιθανοτήτων κατάταξης την απόκλιση

Kullback–Leibler (KL) η οποία ορίζεται ως

$$D(\Pi||Q) = \sum_{i=1}^n \sum_{j=1}^k \pi_{ij} \log \frac{\pi_{ij}}{q_{ij}}.$$

Ο αλγόριθμος αναδιάταξης βασίζεται σε ένα προσομοιωμένο δείγμα μεγέθους M από την εκ των υστέρων κατανομή και βρίσκει κατάλληλες μεταθέσεις τ_t , $t = 1, \dots, M$, και έναν πίνακα πιθανοτήτων κατάταξης $\hat{\Pi}$, έτσι ώστε να ελαχιστοποιείται η $\mathcal{D} = \sum_{t=1}^M D(\tau_t \Pi^{(t)} || \hat{\Pi})$. Έτσι, ο αλγόριθμος που προτείνει ο Stephens έχει ως εξής:

Αλγόριθμος 2.6.2. *Αλγόριθμος αναδιάταξης μέσω της απόκλισης Kullback–Leibler*

Δώσε κάποιες αρχικές τιμές τ_1, \dots, τ_M και επανάλαβε τα παρακάτω βήματα μέχρι να ικανοποιηθεί κάποιο (προκαθορισμένο) κριτήριο σύγκλισης

1. Ελαχιστοποίησε ως προς $\hat{\Pi}$ το

$$\sum_{t=1}^M D(\tau_t \Pi^{(t)} || \hat{\Pi}).$$

2. Για $t = 1, \dots, M$ ελαχιστοποίησε ως προς τ_t το

$$D(\tau_t \Pi^{(t)} || \hat{\Pi}).$$

Ο αναγνώστης παραπέμπεται στον Stephens (1997a) για περισσότερες λεπτομέρειες στα Βήματα 1 και 2 του αλγορίθμου. Η συγκεκριμένη μέθοδος είναι αρκετά αποδοτική καθώς δεν παρουσιάζει τα προβλήματα των προαναφερθεισών μεθόδων. Έχει όμως αρκετά υπολογιστικά μειονεκτήματα, καθώς ο ίδιος ο Stephens αναφέρει ότι ο αλγορίθμος έχει υψηλές απαιτήσεις μνήμης ακόμα και με την σύγχρονη διαθέσιμη υπολογιστική τεχνολογία. Επί πλέον, πολλές φορές κρίνεται απαραίτητη η εκκίνηση του αλγορίθμου από διαφορετικές αρχικές τιμές. Πρόσφατα, οι Grün and Leisch (2008) πρότειναν μία γενίκευση του αλγορίθμου του Stephens στις περιπτώσεις όπου η εκ των υστέρων κατανομή είναι γνησίως πολυκόρυφη. Πιο συγκεκριμένα, εισήγαγαν μία μέθοδο όπου, εκτός από την αναδιάταξη του προσομοιωμένου δείγματος, απομονώνονται επίσης οι μικρότερες κορυφές της εκ των υστέρων κατανομής. Αυτό γίνεται χρησιμοποιώντας ένα επί πλέον βήμα που ομαδοποιεί το δείγμα σε ετερογενείς περιοχές (clusters).

2.6.4 Αναδιάταξη μέσω συναρτήσεων ζημίας

Πρόκειται για μέθοδο που εφαρμόζει ιδέες από την Θεωρία Αποφάσεων. Αναπτύχθηκε από τους Celeux et al. (2000) και εφαρμόστηκε επίσης από τους Hurn et al. (2003). Η μέθοδος βασίζεται στον ορισμό μιας συνάρτησης ζημίας

$$\mathcal{L} : \mathcal{A} \times \mathcal{P}_k \times \Theta^k \rightarrow [0, +\infty),$$

με την ιδιότητα

$$\mathcal{L}(a, \mathbf{p}, \boldsymbol{\theta}) = \mathcal{L}(a, \tau(\mathbf{p}, \boldsymbol{\theta})), \quad \forall \tau \in \mathcal{T}_k,$$

όπου $a \in \mathcal{A}$ συμβολίζει μία απόφαση που ανήκει στον χώρο αποφάσεων \mathcal{A} . Φυσικά, το πρώτο κρίσιμο σημείο για την καλή απόδοση της μεθόδου είναι η επιλογή της συνάρτησης ζημίας \mathcal{L} . Δεδομένης της \mathcal{L} , η βέλτιστη απόφαση είναι αυτή που ελαχιστοποιεί το εκ των υστέρων αναμενόμενο κόστος

$$\begin{aligned} \mathbb{E}\{\mathcal{L}(a, \mathbf{p}, \boldsymbol{\theta}) | \mathbf{x}\} &= \int_{\mathcal{P} \times \Theta^k} \mathcal{L}(a, \mathbf{p}, \boldsymbol{\theta}) f(\mathbf{p}, \boldsymbol{\theta} | \mathbf{x}) d\mathbf{p} d\boldsymbol{\theta} \\ &\approx \frac{1}{M} \sum_{t=1}^M \mathcal{L}(a, (\mathbf{p}, \boldsymbol{\theta})^{(t)}). \end{aligned}$$

Στην πραγματικότητα, η τελευταία σχέση προσεγγίζει το εκ των υστέρων αναμενόμενο κόστος, δοθέντος ενός προσομοιωμένου δείγματος M τιμών από την εκ των υστέρων κατανομή των παραμέτρων της μείξης. Συνήθως, το βήμα της ελαχιστοποίησης δεν μπορεί να γίνει με αναλυτικούς τρόπους και γι' αυτό εφαρμόζονται μέθοδοι стоχαστικής βελτιστοποίησης όπως ο αλγόριθμος simulated annealing (Kirkpatrick et al., 1983).

Από την μπεϋζιανή οπτική, η συγκεκριμένη μέθοδος είναι η πλέον αποδεκτή καθώς η συμπερασματολογία εξαρτάται μόνο από τα παρατηρηθέντα δεδομένα. Από την άλλη πλευρά, το μειονέκτημα των αλγορίθμων αυτών είναι το τεράστιο υπολογιστικό κόστος. Επιπροσθέτως, μπορεί να μην είναι δυνατή η ελαχιστοποίηση του εκ των υστέρων αναμενόμενου κόστους για κάποιες συναρτήσεις ζημίας, κάτι που περιορίζει το πεδίο εφαρμογών της μεθόδου σε κλάσεις συναρτήσεων ζημίας που μπορεί να μην είναι «λογικές» για το πρόβλημα απόφασης που μας ενδιαφέρει (δες Jasra, et al., 2005).

Σύμφωνα με την προηγηθείσα ανασκόπηση, ο διαθέσιμες μέθοδοι αντιμετώπισης του προβλήματος label switching κρίνονται ανεπαρκείς. Οι μεν οι απλοϊκές τεχνικές (περιορισμοί διάκρισης, αλγόριθμος πιλοτικής αναδιάταξης) δεν μπορούν

να εφαρμοστούν σε ένα γενικό πλαίσιο, οι δε πιο αποδοτικές τεχνικές (αλγόριθμος K-L, αλγόριθμοι συναρτήσεων ζημίας) δεν αποτελούν ελκυστικές εναλλακτικές λόγω του υψηλού υπολογιστικού κόστους αυτών. Χαρακτηριστική είναι η θέση των Celeux et al. (2000) σχετικά με την συμπερασματολογία στα μοντέλα μείξεων κατανομών, σύμφωνα με την οποία

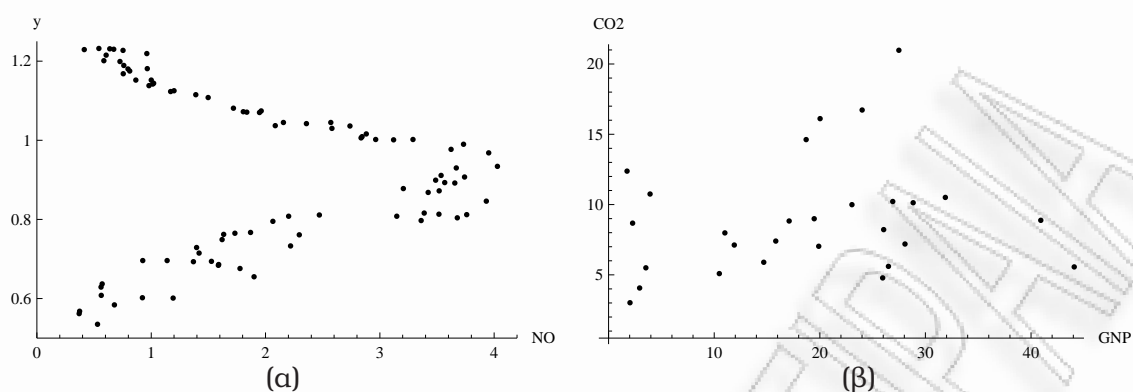
«Το γενικό συμπέρασμα είναι ότι πρόκειται για ένα πρόβλημα υψηλής πολυπλοκότητας, αλλά κάτι τέτοιο δεν πρέπει να θεωρείται αποτρεπτικό διότι οι μείξεις κατανομών (και άλλες κλάσεις μοντέλων ελλειπών δεδομένων) είναι μία σύνθετη περιοχή όπου ενδείκνυται η χρήση προηγμένων μεθοδολογιών.»

Η παραπάνω θέση γίνεται σαφής αρχικά στην Ενότητα 2.7 όπου αναδεικνύονται οι αδυναμίες των περιορισμών διάκρισης αλλά και πιο γενικά στο Κεφάλαιο 5 όπου εφαρμόζεται η πλειοψηφία των διαθέσιμων μεθοδολογιών. Έτσι, κρίνεται απαραίτητη η θεμελίωση μιας μεθόδου που να συμβιβάζει τις δύο παραπάνω οπτικές: Να είναι αρκετά απλή στην εφαρμογή της και να μπορεί να θεωρηθεί ως μία γενική τεχνική αποδοτικής αντιμετώπισης του προβλήματος. Αυτός είναι και ο στόχος του Κεφαλαίου 4.

2.7 Ένα μοντέλο εκτίμησης μείξεων πολυμεταβλητών παλινδρομήσεων

Υπάρχουν αρκετά ενδιαφέρουσες γενικεύσεις των κλασικών μοντέλων μείξεων κατανομών που συζητήσαμε ως τώρα. Σε αυτήν την ενότητα θα μελετήσουμε πώς μπορούν να χρησιμοποιηθούν οι μείξεις κατανομών σε μοντέλα γραμμικής παλινδρόμησης. Για τον σκοπό αυτό εισάγεται ένα μοντέλο εκτίμησης μείξεων πολυμεταβλητών παλινδρομήσεων, θεωρώντας γνωστό αριθμό συνιστωσών. Παρόμοια μοντέλα έχουν μελετηθεί για την μονοδιάστατη περίπτωση από τους Hurn et al. (2003).

Η ανάγκη θεώρησης μείξεων (πολυμεταβλητών ή μη) παλινδρομήσεων είναι επιτακτική όταν αντιμετωπίζουμε ετερογενή σύνολα δεδομένων που δεν δικαιολογούν την προσαρμογή ενός απλού γραμμικού μοντέλου στο σύνολό τους, αλλά σε συγκεκριμένα υποσύνολα αυτών. Για παράδειγμα, στο Σχήμα 2.10(a) απεικονίζονται δεδομένα που περιγράφουν τη σχέση μεταξύ του λόγου ισοδυναμίας, δηλαδή του λόγου της περιεκτικότητας αέρα - αιθανόλης της μηχανής, έναντι της συγκέντρωσης νιτρικού μονοξειδίου στην εξάτμιση για $n = 88$ αυτοκίνητα (Hurvich et al. 1998).



Σχήμα 2.10: (α): Λόγος ισοδυναμίας μηχανής (y) έναντι της συγκέντρωσης νιτρικού μονοξειδίου στην εξάτμιση για $n = 88$ αυτοκίνητα (β): ακαθάριστο εθνικό προϊόν (GNP) και επίπεδα εκπομπής CO_2 για $n = 28$ χώρες το 1996.

Παρατηρούμε ότι στο διάγραμμα διασποράς των δεδομένων καταγράφονται δύο υποομάδες παρατηρήσεων στις οποίες φαίνεται να προσαρμόζονται διαφορετικές ευθείες παλινδρόμησης. Βέβαια, η συμπεριφορά της μείξης παρατηρείται διότι δεν γνωρίζουμε την κατάταξη κάθε σημείου σε μία από τις (υποτιθέμενες δύο) ομάδες. Αν και σε αυτήν την περίπτωση η κατάταξη κάθε παρατήρησης μπορεί να είναι προφανής με το «μάτι» καθώς οι δύο υποομάδες είναι επαρκώς διαχωρισμένες μεταξύ τους, κάτι τέτοιο δεν ισχύει πάντα όπως όπως διαπιστώνεται στο Σχήμα 2.10(β). Εκεί απεικονίζεται το διάγραμμα διασποράς του ακαθάριστου εθνικού προϊόντος και των εκτιμώμενων εκπομπών CO_2 για $n = 28$ χώρες κατά το 1996. Η διασπορά των σημείων δεν δείχνει ξεκάθαρα κάποια γραμμική τάση μεταξύ των δύο μεταβλητών. Παρ' όλ' αυτά διακρίνονται κάποιες ομάδες σημείων που θα δικαιολογούσαν την προσαρμογή ενός μοντέλου γραμμικής παλινδρόμησης σε αυτά.

Έστω $\{(y_i, x_i), i = 1, \dots, n\}$ ένα δείγμα n παρατηρήσεων m μεταβλητών απόκρισης $y_i = (y_{1i}, \dots, y_{mi})'$ σε q ερμηνευτικές μεταβλητές $x_i = (x_{1i}, \dots, x_{qi})'$. Θεωρούμε ότι αυτές προκύπτουν από ένα μοντέλο μείξης k πολυμεταβλητών παλινδρομήσεων, δηλαδή για κάθε $i = 1, \dots, n$ η παρατήρηση y_i μπορεί να εκφραστεί ως

$$y_i = \begin{cases} \beta'_1 x_i + e_{i1}, & \text{με πιθανότητα } p_1 \\ \vdots \\ \beta'_k x_i + e_{ik}, & \text{με πιθανότητα } p_k. \end{cases}$$

Για κάθε $j = 1, \dots, k$, το β_j είναι ένας πίνακας διάστασης $q \times m$, ενώ τα σφάλματα e_{ij} , $i = 1, \dots, n$, υποθέτουμε ότι είναι ανεξάρτητα και ισόνομα τυχαία διανύσματα τα οποία ακολουθούν μια m -διάστατη κανονική κατανομή με μέση τιμή $(0, \dots, 0)'$

και πίνακα διασπορών-συνδιασπορών Σ_j . Το παραπάνω μοντέλο είναι ισοδύναμο με το εξής:

$$y_i \sim \sum_{j=1}^k p_j \mathcal{N}_m(\beta'_j x_i, \Sigma_j), \quad i = 1, \dots, n, \text{ ανεξάρτητα.} \quad (2.35)$$

Σύμφωνα με τα παραπάνω, οι παράμετροι του μοντέλου είναι το διάνυσμα

$$(\mathbf{p}, \boldsymbol{\theta}) = ((p_1, \dots, p_k), (\beta'_1, \dots, \beta'_k), (\Sigma_1, \dots, \Sigma_k)) \in \mathcal{P}_k \times \mathbf{M}_{q \times m}^k \times \mathcal{M}_m^k,$$

όπου με $\mathbf{M}_{q \times m}$ συμβολίζουμε τον χώρο των πινάκων διάστασης $q \times m$, ενώ με \mathcal{M}_m συμβολίζουμε τον χώρο των θετικά ορισμένων πινάκων διάστασης $m \times m$. Για την εκτίμηση των παραμέτρων του μοντέλου μπορούμε να ακολουθήσουμε μία μπεϋζιανή προσέγγιση θεωρώντας το ακόλουθο ιεραρχικό μοντέλο:

$$\begin{aligned} \beta'_j | \Sigma_j &\sim \mathcal{MN}_{m \times q}(M', \Omega, \Sigma_j), \quad j = 1, \dots, k, \text{ ανεξάρτητα} \\ \Sigma_j &\stackrel{\text{iid}}{\sim} \mathcal{IW}(\Psi, \zeta), \quad j = 1, \dots, k, \\ P(Z_i = j | \mathbf{p}) &= p_j, \quad i = 1, 2, \dots, n, \quad j = 1, \dots, k, \text{ ανεξάρτητα} \\ \mathbf{p} &\sim \mathcal{D}(1, \dots, 1) \\ y_i | Z_i = j &\sim \mathcal{N}_m(\beta'_j x_i, \Sigma_j), \quad i = 1, 2, \dots, n, \text{ ανεξάρτητα.} \end{aligned}$$

Εδώ με \mathcal{MN} και \mathcal{IW} συμβολίζουμε την κανονική κατανομή στον χώρο των πινάκων και την αντίστροφη κατανομή Wishart αντίστοιχα. Έστω \mathbf{x} ο $n \times q$ πίνακας σχεδιασμού του μοντέλου (δηλαδή η i -οστή γραμμή του \mathbf{x} είναι το x'_i) και \mathbf{y} ο $n \times m$ πίνακας με i -οστή γραμμή το y'_i . Για τις παραμέτρους των εκ των προτέρων κατανομών θεωρούμε ότι

$$\begin{aligned} M' &= (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y} \\ \Omega &= \text{diag}(\omega_1, \dots, \omega_q), \end{aligned}$$

όπου ο τελεστής diag μετασχηματίζει ένα διάνυσμα στον αντίστοιχο διαγώνιο πίνακα. Για $\lambda = 2, \dots, q$ υποθέτουμε ότι $\omega_\lambda = x_{\lambda;(n)} - x_{\lambda;(1)}$. Στις περισσότερες περιπτώσεις, το μοντέλο (2.35) θα περιλαμβάνει σταθερό όρο (δηλαδή η πρώτη στήλη του πίνακα σχεδιασμού θα αποτελείται από μονάδες) και τότε θεωρούμε $\omega_1 = c \sum_{\lambda=2}^q \omega_\lambda / (q-1)$, με c κάποια θετική σταθερά (για τις εφαρμογές που θα παρουσιαστούν θέσαμε $c = 3$). Σε διαφορετικές περιπτώσεις θεωρούμε $\omega_1 = x_{1;(n)} - x_{1;(1)}$. Επί πλέον, για τις υπερπαραμέτρους των διασπορών ακολουθούμε τους Dellaportas and Papageorgiou (2005), σύμφωνα με τους οποίους $\Psi = \text{diag}(\psi_1, \dots, \psi_m)$, με $\psi_j \sim \mathcal{G}(g, \rho)$, δηλαδή μία κατανομή γάμμα με μέση τιμή g/ρ , για την οποία $g = 1$ και $\rho = 36^{-1}$. Οι βαθμοί ελευθερίας της αντίστροφης κατανομής Wishart λαμβάνονται ίσοι με $\zeta = m + 1$.

Βάσει των παραπάνω υποθέσεων, η απο κοινού εκ των υστέρων κατανομή των παραμέτρων του μοντέλου μπορεί να γραφτεί ως εξής

$$\begin{aligned} f(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{z}, \boldsymbol{\psi} | \mathbf{x}, \mathbf{y}, g, \rho) &\propto f(\mathbf{y} | \mathbf{z}, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) f(\mathbf{z} | \mathbf{p}) f(\mathbf{p}) \\ &\times f(\boldsymbol{\psi} | g, \rho) \prod_{j=1}^k f(\beta_j | \Sigma_j) f(\Sigma_j | \boldsymbol{\psi}) \end{aligned} \quad (2.36)$$

Από την τελευταία σχέση είναι προφανές ότι η εκ των υστέρων κατανομή των παραμέτρων του μοντέλου, όπως όλων των άλλων μοντέλων μείζης, είναι συμμετρική στις μεταθέσεις οπότε το φαινόμενο label switching υφίσταται και εδώ.

Για να προσομοιώσουμε από την εκ των υστέρων κατανομή (2.36) θα χρησιμοποιήσουμε τον δειγματολήπτη Gibbs. Προσομοιώνουμε δηλαδή κάθε παράμετρο από την πλήρη δεσμευμένη κατανομή της. Έστω $I(A)$ η δείκτρια συνάρτηση του A . Αν για $j = 1, \dots, k$,

$$\begin{aligned} U_j &= M' \Omega^{-1} + \sum_{i=1}^n I(z_i = j) y_i x_i', \\ V_j &= \Omega^{-1} + \sum_{i=1}^n I(z_i = j) x_i x_i', \\ T_j &= \sum_{i=1}^n I(z_i = j) (y_i - \beta_j' x_i) (y_i - \beta_j' x_i)' \end{aligned}$$

και τα n_j είναι όπως στην (2.28), τότε ισχύει το ακόλουθο.

Λήμμα 2.7.1. *Οι πλήρεις δεσμευμένες εκ των υστέρων κατανομές του ιεραρχικού μοντέλου (2.36) έχουν ως εξής:*

$$\beta_j' | \dots \sim \mathcal{MN}_{m \times q}(U_j V_j^{-1}, V_j^{-1}, \Sigma_j), \quad (2.37)$$

$$\Sigma_j | \dots \sim \mathcal{W}^{-1}(\Psi + T_j + (\beta_j' - M') \Omega^{-1} (\beta_j - M), \zeta + q + n_j), \quad (2.38)$$

$$Z_i | \dots \propto \{p_j f(y_i | \beta_j' x_i, \Sigma_j); j = 1, \dots, k\}, \quad (2.39)$$

$$\mathbf{p} | \dots \sim \mathcal{D}(1 + n_1, \dots, 1 + n_k), \quad (2.40)$$

$$\boldsymbol{\psi} | \dots \sim \mathcal{G}\left(g + k\zeta/2, \rho + \sum_{j=1}^k \Sigma_j^{-1}(l, l)/2\right), \quad (2.41)$$

ανεξάρτητα για $j = 1, \dots, k$, $i = 1, \dots, n$ και $l = 1, \dots, m$.

Απόδειξη. Για την απόδειξη ο αναγνώστης παραπέμπεται στην Ενότητα A.2 στο Παράρτημα A. □

Η προσομοίωση από την εκ των υστέρων κατανομή (2.36), γίνεται χρησιμοποιώντας τις κατανομές του Λήμματος 2.7.1 μέσω του δειγματολήπτη Gibbs. Έτσι, ο αλγόριθμος που ακολουθεί παράγει μία ακολουθία τιμών $(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{z}, \boldsymbol{\psi})^{(t)}$, $t = 1, \dots, M$ η οποία αποτελεί μία μαρκοβιανή αλυσίδα με οριακή κατανομή την (2.36).

Αλγόριθμος 2.7.1. Δειγματολήπτης Gibbs για την κατανομή στόχο (2.36)

Δώσε κάποιες αρχικές τιμές $(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{z}, \boldsymbol{\psi})^{(0)}$ και για $t = 1, \dots, M$ επανάλαβε τα παρακάτω βήματα :

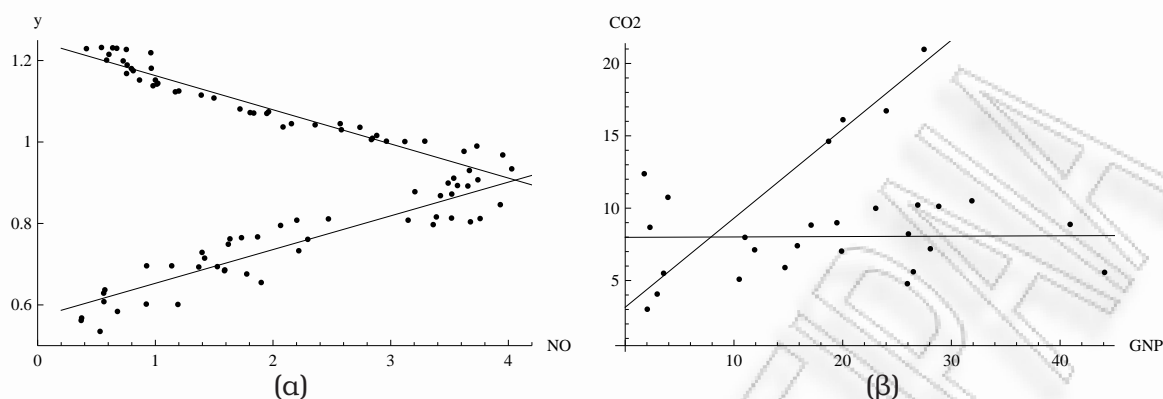
1. Για $j = 1, \dots, k$ προσομοίωσε ανεξάρτητα τα $\beta_j^{(t)} | \Sigma_j^{(t-1)}, \mathbf{z}^{(t-1)}$ από την (2.37).
 2. Για $j = 1, \dots, k$ προσομοίωσε ανεξάρτητα τα $\Sigma_j^{(t)} | \boldsymbol{\psi}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{z}^{(t-1)}$ από την (2.38).
 3. Για $i = 1, \dots, n$ προσομοίωσε ανεξάρτητα τα $z_i^{(t)} = j | \mathbf{p}^{(t-1)}, \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\beta}^{(t)}$ από την (2.39).
 4. Προσομοίωσε τα $\mathbf{p}^{(t)} | \mathbf{z}^{(t)}$ από την (2.40).
 5. Για $l = 1 \dots, m$ προσομοίωσε ανεξάρτητα τα $\psi_l^{(t)} | \Sigma_j^{(t)}$ από την (2.41).
-

2.7.1 Παραδείγματα

Σε αυτήν την ενότητα, εφαρμόζεται ο αλγόριθμος εκτίμησης μείξεων παλινδρομήσεων 2.7.1, χρησιμοποιώντας προσομοιωμένα αλλά και πραγματικά δεδομένα. Το Παράδειγμα 2.7.1 διαπραγματεύεται μια περίπτωση όπου είναι δυνατή η άμεση εκτίμηση των εκ των υστέρων περιθωριακών κατανομών των παραμέτρων του μοντέλου (2.35) διότι δεν εμφανίζεται το φαινόμενο label switching. Χρησιμοποιείται για να αναδείξει την καλή απόδοση του προτεινόμενου αλγορίθμου στην εξερεύνηση τουλάχιστον μίας εκ των συμμετρικών περιοχών, ανεξάρτητα από το φαινόμενο label switching. Στο Παράδειγμα 2.7.2, εμφανίζεται το φαινόμενο label switching και εφαρμόζεται η μέθοδος των εικονικών περιορισμών διάκρισης για την αντιμετώπισή του, ενώ στο Παράδειγμα 2.7.3 αναδεικνύονται οι «περιορισμοί των περιορισμών» και η ανάγκη εύρεσης μιας γενικής μεθόδου αντιμετώπισης του προβλήματος.

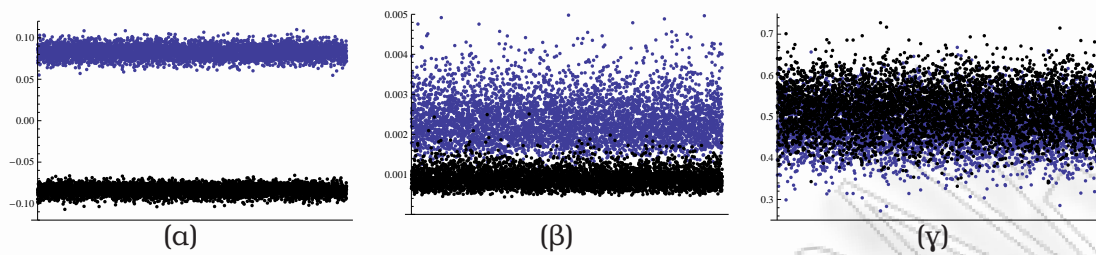
Παράδειγμα 2.7.1. Σύνοψη δεδομένων NO.

Οι Hurn, Justel and Robert (2003) χρησιμοποίησαν τα συγκεκριμένα δεδομένα θε-



Σχήμα 2.11: Εκτιμηθείσες ευθείες παλινδρόμησης με $k = 2$ συνιστώσες για τα δύο μονοδιάστατα παραδείγματα. (α): Σύνολο δεδομένων NO. (β): Σύνολο δεδομένων CO₂ (ύστερα από την αναδιάταξη βάσει του περιορισμού $p_1 < p_2$).

ωρώντας $k = 2$ συνιστώσες, κάτι που θα υποθέσουμε και εμείς. Στο πλαίσιο του μοντέλου (2.35) για $k = 2$ συνιστώσες έχουμε μια μεταβλητή απόκρισης ($m = 1$, η οποία είναι ο λόγος ισοδυναμίας) και δύο ερμηνευτικές μεταβλητές ($q = 2$, η πρώτη αντιστοιχεί στον σταθερό όρο των ευθειών παλινδρόμησης, ενώ η δεύτερη στην συγκέντρωση νιτρικού οξέος). Εκτελέσαμε τον δειγματολήπτη Gibbs του Αλγορίθμου 2.7.1 για 55000 επαναλήψεις, αφού προηγήθηκαν 5000 επαναλήψεις ως περίοδος burn-in. Στα Σχήματα 2.12(α), (β) και (γ) φαίνονται οι ακολουθίες των προσομοιωμένων τιμών των $(\beta_{12}, \beta_{22})^{(t)}$, $(\sigma_1^2, \sigma_2^2)^{(t)}$ και $(p_1, p_2)^{(t)}$, αντίστοιχα, για $t = 5000(10)55000$ (με διαφορετικό χρώμα για κάθε μία από τις δύο συνιστώσες της μείξης). Επειδή στο συγκεκριμένο παράδειγμα έχουμε $m = 1$, οι «πίνακες» διασπορών-συνδιασπορών Σ_j , $j = 1, 2$, είναι θετικοί πραγματικοί αριθμοί και συμβολίζονται με μικρά γράμματα. Παρατηρούμε ότι οι προσομοιωθείσες τιμές των παραμέτρων κάθε μιας συνιστώσας καταλαμβάνουν διαφορετικές περιοχές. Άρα το φαινόμενο label switching δεν έχει παρουσιαστεί μέχρι το πέρας των προσομοιώσεων. Κανονικά, θα περιμέναμε οι τιμές των παραμέτρων να εναλλάσσονται μεταξύ των δύο συμμετρικών περιοχών της εκ των υστέρων κατανομής όπως προβλέπει η θεωρία. Επομένως, ο δειγματολήπτης Gibbs δεν έχει συγκλίνει, με την έννοια ότι έχει καταφέρει να εντοπίσει μόνο την μία συμμετρική περιοχή των παραμέτρων (από τις δύο). Η συμπεριφορά αυτή εξηγείται από το γεγονός ότι στο συγκεκριμένο παράδειγμα οι συμμετρικές περιοχές της εκ των υστέρων κατανομής είναι αρκετά μακριά μεταξύ τους. Αυτό σημαίνει ότι μεταξύ των δύο συμμετρικών περιοχών η εκ των υστέρων πυκνότητα των παραμέτρων είναι σχεδόν μηδενική. Έτσι, είναι πολύ δύσκολο για την προσομοιωθείσα μαρκοβιανή

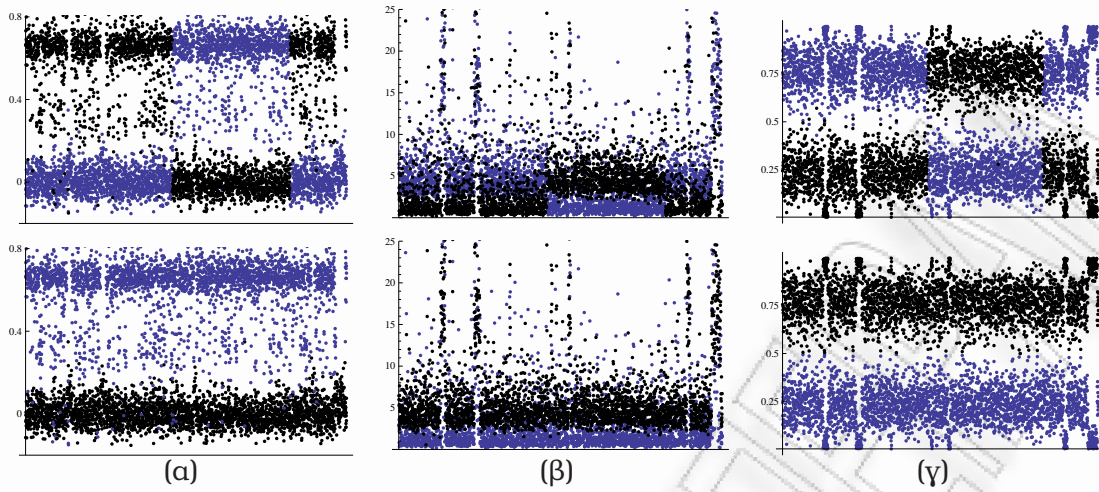


Σχήμα 2.12: Οι προσομοιωμένες τιμές των παραμέτρων των συνιστωσών για το Παράδειγμα 2.7.1: συντελεστές παλινδρόμησης (α), διασπορές σφαλμάτων (β) και βάρη (γ).

αλύσιδα να μετακινηθεί από την μία περιοχή στην άλλη, καθώς για να γίνει αυτό θα πρέπει να περάσει από την περιοχή πολύ χαμηλής εκ των υστέρων πυκνότητας. Η περίπτωση αυτή είναι πολύ τυπική για τον δειγματολήπτη Gibbs όταν οι συμμετρικές περιοχές είναι αρκετά μακριά μεταξύ τους. Έτσι, ενώ εξερευνείται επαρκώς μία από τις συμμετρικές περιοχές, ο αλγόριθμος δεν έχει συγκλίνει διότι θα έπρεπε να τις είχε εξαντλήσει όλες.

Το γεγονός αυτό υπογραμμίζει το παράδοξο του φαινομένου label switching. Σε αυτήν την περίπτωση είναι δυνατή η συμπερασματολογία, καθώς οι τιμές των παραμέτρων δεν εναλλάσσονται μεταξύ των δύο συμμετρικών περιοχών. Αν υποθέσουμε λοιπόν ότι η περιοχή που έχει «κολλήσει» ο δειγματολήπτης έχει εξερευνηθεί επαρκώς, τότε αν εκτιμήσουμε τις παραμέτρους με την συνηθισμένη πρακτική των εργοδικών μέσων οι εκτιμήσεις που θα πάρουμε θα είναι πολύ κοντά στις πραγματικές τιμές των παραμέτρων. Το παράδοξο έγκειται στο γεγονός όμως ότι αυτές οι εκτιμήσεις θα έχουν προκύψει από έναν αλγόριθμο που δεν έχει συγκλίνει στην εκ των υστέρων κατανομή (αλλά μόνο σε μία από τις συμμετρικές περιοχές αυτής). Από την άλλη πλευρά, αν το φαινόμενο label switching είχε εμφανιστεί, δεν θα ήταν άμεση η συμπερασματολογία, παρά το γεγονός ότι σε αυτήν την περίπτωση θα είχε επιτευχθεί η σύγκλιση του αλγορίθμου. Αυτό οφείλεται στο γεγονός ότι οι εργοδικοί μέσοι των παραμέτρων θα ήταν σχεδόν ίσοι μεταξύ τους, και έτσι δεν θα είχε νόημα η εκτίμηση των πραγματικών εκ των υστέρων μέσων τιμών με αυτούς.

Επιστρέφοντας στο παράδειγμα, μπορούμε να εκμεταλλευτούμε την απουσία του φαινομένου label switching και να εκτιμήσουμε τις παραμέτρους του μοντέλου. Υπολογίζοντας τους εργοδικούς μέσους των παραμέτρων καταλήγουμε στις εξής εκτιμήσεις:



Σχήμα 2.13: Πάνω: προσομοιωμένες τιμές των παραμέτρων για το Παράδειγμα 2.7.2: συντελεστές παλινδρόμησης (α), διασπορές σφαλμάτων (β) και βάρη (γ). Κάτω: οι αντίστοιχες αναδιατεταγμένες τιμές σύμφωνα με τον περιορισμό $p_1 < p_2$.

$$\begin{aligned} \hat{\beta}_1 &= (0.570, 0.0829)' & \hat{\sigma}_1^2 &= 0.00239 & \hat{p}_1 &= 0.479 \\ \hat{\beta}_2 &= (1.247, -0.0838)' & \hat{\sigma}_2^2 &= 0.000906 & \hat{p}_2 &= 0.521. \end{aligned}$$

Στο Σχήμα 2.11(α) παρατίθεται το διάγραμμα διασποράς των παρατηρήσεων μαζί με τις εκτιμηθείσες ευθείες παλινδρόμησης που αποτελούν την μείξη. Διαπιστώνεται η αρκετά καλή προσαρμογή αυτών στα δεδομένα.

Παράδειγμα 2.7.2. Σύνοψη δεδομένων CO_2 .

Οι Hurn et al. (2003) εκτίμησαν την εκ των υστέρων κατανομή του k με την μέθοδο birth-death MCMC και συμπέραναν ότι τα δεδομένα υποστηρίζουν την ύπαρξη δύο παλινδρομήσεων. Συγκεκριμένα εκτίμησαν $\hat{P}(k = 2 | \mathbf{y}, \mathbf{x}) = 0.873$. Στην ανάλυση που ακολουθεί θέτουμε επίσης $k = 2$ και εφαρμόζουμε τον Αλγόριθμο 2.7.1.

Οι ακολουθίες των προσομοιωμένων τιμών των κλίσεων των δύο ευθειών παλινδρόμησης $\beta_1^{(t)}, \beta_2^{(t)}$, των διασπορών των σφαλμάτων και των βαρών των δύο συνιστωσών παρατίθενται στο Σχήμα 2.13, για 30000 επαναλήψεις του αλγορίθμου (αφού προηγήθηκαν 5000 επαναλήψεις ως περίοδος burn-in). Εδώ παρατηρούμε την παρουσία του φαινομένου label switching μεταξύ των περιοχών υψηλής εκ των υστέρων πυκνότητας. Πιο συγκεκριμένα, οι κλίσεις των δύο ευθειών παλινδρόμησης εναλλάσσονται μεταξύ των περιοχών κοντά στο 0 και 0.7, οι διασπορές των σφαλμάτων μεταξύ των περιοχών κοντά στο 5 και 2 και τα βάρη μεταξύ των 0.8 και 0.2, αντίστοιχα. Από τα γραφήματα αυτά είναι ξεκάθαρο ότι η ακολουθία των προσομοιωμένων τιμών

$((p_1, p_2)^{(t)}, (\beta_1, \beta_2)^{(t)}, (\sigma_1^2, \sigma_2^2)^{(t)})$, $t = 1, \dots, M$, δεν μπορεί να χρησιμοποιηθεί για την εκτίμηση των περιθωριακών εκ των υστέρων κατανομών των παραμέτρων.

Παρ' όλ' αυτά, ο περιορισμός $p_1 < p_2$ είναι ικανός να απομονώσει επιτυχώς τις συμμετρικές περιοχές της εκ των υστέρων κατανομής των παραμέτρων. Συνεπώς, αν κατά την επανάληψη t η μετάθεση $\tau_t = (r_1, r_2) \in \{(1, 2), (2, 1)\}$ είναι τέτοια ώστε $p_{r_1}^{(t)} < p_{r_2}^{(t)}$, τότε το παραμετρικό διάνυσμα μετατίθεται στο

$$(\tau_t(p_1, p_2)^{(t)}, \tau_t(\beta_1, \beta_2)^{(t)}, \tau_t(\sigma_1^2, \sigma_2^2)^{(t)}).$$

Αναδιατάσσοντας την ακολουθία των προσομοιωμένων τιμών κατ' αυτόν τον τρόπο καταλήγουμε στα γραφήματα της δεύτερης γραμμής του Σχήματος 2.13. Οι αντίστοιχες εκτιμήσεις των παραμέτρων της μείξης παλινδρομήσεων έχουν ως εξής:

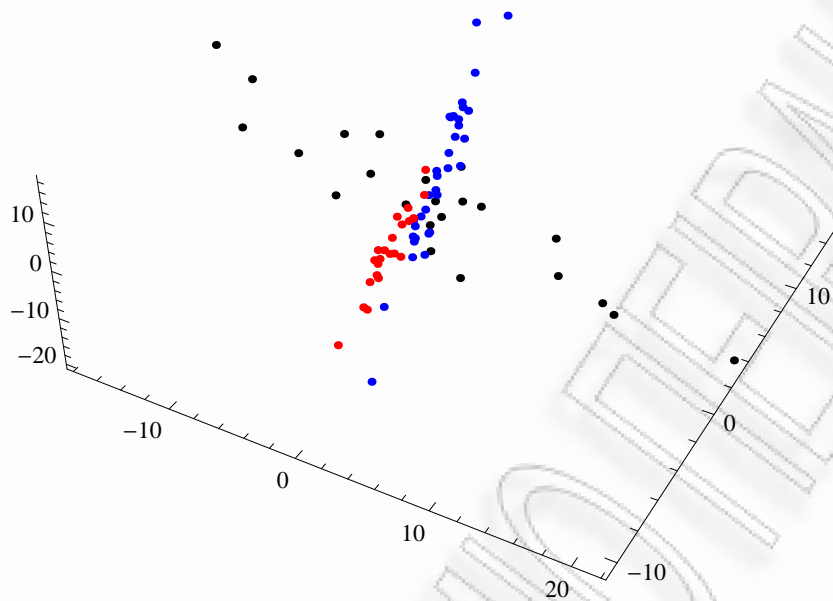
$$\begin{aligned} \hat{\beta}_1 &= (3.15, 0.62)' & \hat{\sigma}_1^2 &= 2.81 & \hat{p}_1 &= 0.220 \\ \hat{\beta}_2 &= (7.99, 0.00)' & \hat{\sigma}_2^2 &= 5.44 & \hat{p}_2 &= 0.880. \end{aligned}$$

Στο Σχήμα 2.11(β) παρατίθεται η εκτιμηθείσα μείξη ευθειών παλινδρόμησης. Επειδή οι συμμετρικές περιοχές της εκ υστέρων κατανομής είναι αρκετά μακριά μεταξύ τους, η μέθοδος των εικονικών περιορισμών διάταξης αναδιατάσσει επιτυχώς το προσομοιωμένο δείγμα. Έτσι, οι εκτιμήσεις των παραμέτρων που προκύπτουν, βάσει της συγκεκριμένης μεθόδου αντιμετώπισης του label switching, δίνουν μία αρκετά καλή προσαρμογή στα δεδομένα. Αυτό όμως δεν ισχύει πάντα, όπως γίνεται σαφές στο παράδειγμα που ακολουθεί.

Παράδειγμα 2.7.3. Προσομοιωμένα πολυμεταβλητά δεδομένα.

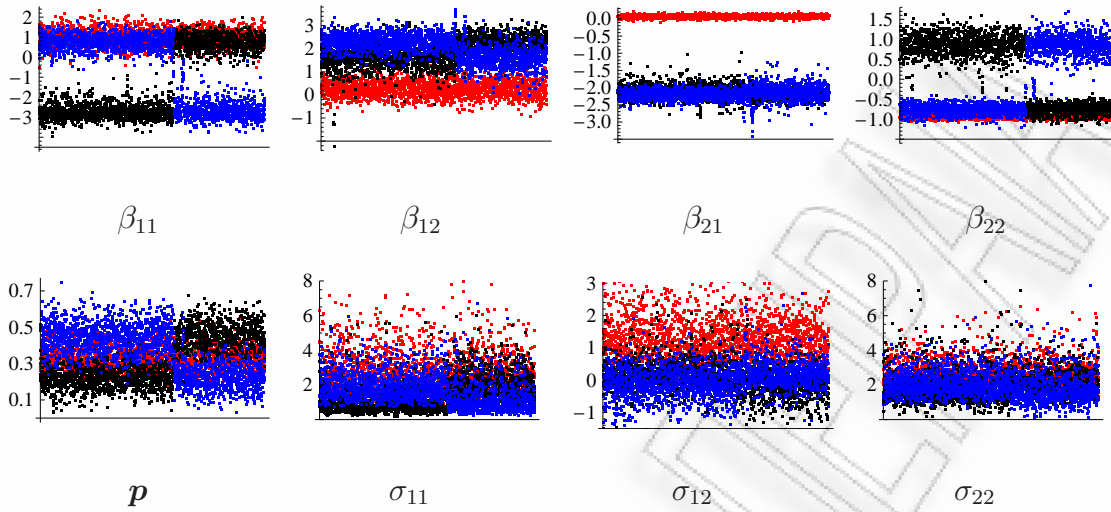
Το τελευταίο παράδειγμα δείχνει ότι υπάρχουν περιπτώσεις ετερογενών συνόλων δεδομένων όπου δεν είναι δυνατόν να καταλήξουμε σε «καλές» εκτιμήσεις των παραμέτρων μέσω εικονικών περιορισμών διάταξης. Για το σκοπό αυτό θεωρούμε μία μείξη πολυμεταβλητών παλινδρομήσεων. Προσομοιώσαμε $n = 80$ παρατηρήσεις από το μοντέλο (2.35) θεωρώντας $k = 3$ συνιστώσες, μία ερμηνευτική μεταβλητή και τον σταθερό όρο ($q = 2$) και δύο μεταβλητές απόκρισης ($m = 2$). Οι πραγματικές τιμές των παραμέτρων που χρησιμοποιήθηκαν για την προσομοίωση, έχουν ως εξής:

$$\begin{aligned} \beta_1 &= \begin{pmatrix} -2 & -2 \\ 2 & 1 \end{pmatrix} & \Sigma_1 &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} & p_1 &= 1/3 \\ \beta_2 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & \Sigma_2 &= \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} & p_2 &= 1/3 \\ \beta_3 &= \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix} & \Sigma_3 &= \begin{pmatrix} 2 & 0.5 \\ 0.5 & 3 \end{pmatrix} & p_3 &= 1/3. \end{aligned}$$



Σχήμα 2.14: Τα προσομοιωμένα δεδομένα του Παραδείγματος 2.7.3. Κάθε χρώμα αντιστοιχεί στην (υποτιθέμενη άγνωστη) πραγματική κατάταξη κάθε παρατήρησης: κόκκινο για την πρώτη, μαύρο για την δεύτερη και μπλε για την τρίτη συνιστώσα.

Τα προσομοιωμένα δεδομένα φαίνονται στο Σχήμα 2.14, όπου οι παρατηρήσεις έχουν διαφορετικό χρώμα αναλόγως από το ποια συνιστώσα προέρχονται. Αυτό είναι δυνατόν λόγω του ότι τα δεδομένα είναι προσομοιωμένα και έτσι γνωρίζουμε από ποια συνιστώσα της μείξης προήλθε κάθε παρατήρηση. Φυσικά αυτή η πληροφορία δεν θα θεωρηθεί διαθέσιμη στην διαδικασία της εκτίμησης. Στο Σχήμα 2.15 απεικονίζονται οι προσομοιωμένες τιμές των παραμέτρων του μοντέλου μείξης (με διαφορετικό χρώμα για κάθε μια συνιστώσα) ύστερα από 50000 επαναλήψεις του Αλγορίθμου 2.7.1 (αφού προηγήθηκαν 5000 επαναλήψεις ως περίοδος burn-in). Όπως παρατηρούμε εκεί, το φαινόμενο label switching έχει εμφανιστεί μεταξύ της μαύρης και μπλε συνιστώσας, ενώ η κόκκινη φαίνεται να παραμένει σε μία σταθερή περιοχή της εκ των υστέρων κατανομής. Αυτό σημαίνει ότι ο δειγματολήπτης έχει εξερευνήσει δύο (από τις 3!) συμμετρικές περιοχές της εκ των υστέρων κατανομής. Επί πλέον, στην προσέγγιση της εκ των υστέρων κατανομής δεν υπάρχει ένδειξη μικρότερων (γνήσιων) κορυφών, καθώς οι προσομοιωμένες τιμές συγκεντρώνονται σε σταθερές περιοχές (χωρίς να λάβουμε υπ' όψιν της εναλλαγής μεταξύ αυτών λόγω του label switching). Παρ' όλ' αυτά, για όλες τις παραμέτρους παρατηρούμε ότι δεν υπάρχει



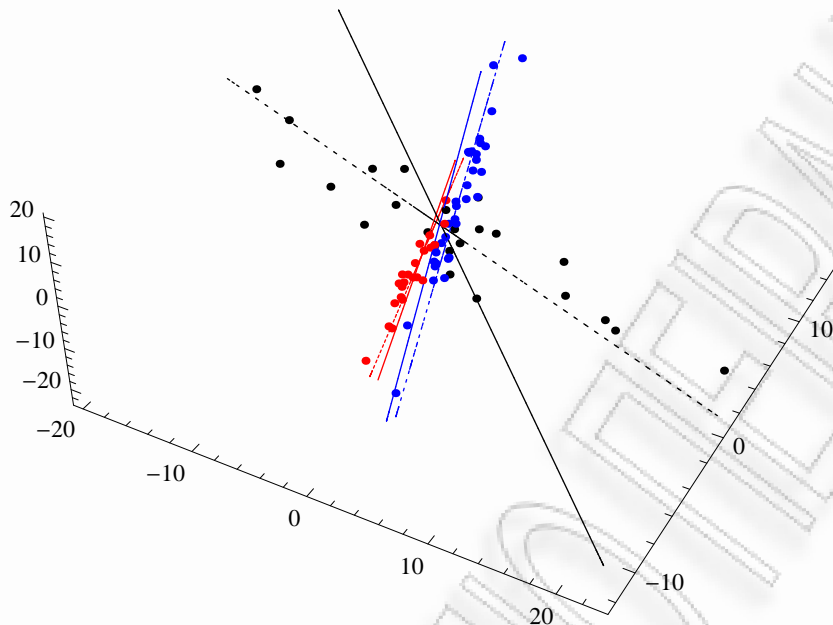
Σχήμα 2.15: Η ακολουθία των προσομοιωμένων τιμών των παραμέτρων για το Παράδειγμα 2.7.3.

κάποιος περιορισμός που να διαχωρίζει επαρκώς τις συμμετρικές περιοχές.

Ας υποθέσουμε τώρα ότι κάποιος επιθυμούσε να απαλαχτεί από το label switching χρησιμοποιώντας έναν περιορισμό διάκρισης. Με μια προσεκτική εξέταση στα γραφήματα του Σχήματος 2.15 θα λέγαμε ότι η μεγαλύτερη διαφοροποίηση μεταξύ των παραμέτρων των τριών συνιστωσών εμφανίζεται στα βάρη τους. Για να δούμε τις επιπτώσεις που έχει στην συμπερασματολογία ένας τέτοιος περιορισμός αναδιατάσσουμε το προσομοιωμένο δείγμα χρησιμοποιώντας τον περιορισμό $p_1 < p_2 < p_3$. Οι εκτιμήσεις που προκύπτουν είναι οι

$$\begin{aligned} \hat{\beta}_1 &= \begin{pmatrix} -2.09 & -1.84 \\ 1.47 & 0.58 \end{pmatrix} & \hat{\Sigma}_1 &= \begin{pmatrix} 1.58 & 0.25 \\ 0.25 & 2.18 \end{pmatrix} & \hat{p}_1 &= 0.207 \\ \hat{\beta}_2 &= \begin{pmatrix} 0.55 & -0.56 \\ 0.76 & -0.71 \end{pmatrix} & \hat{\Sigma}_2 &= \begin{pmatrix} 2.34 & 0.94 \\ 0.94 & 2.35 \end{pmatrix} & \hat{p}_2 &= 0.328 \\ \hat{\beta}_3 &= \begin{pmatrix} 0.68 & -1.85 \\ 1.98 & -0.76 \end{pmatrix} & \hat{\Sigma}_3 &= \begin{pmatrix} 2.58 & 1.30 \\ 1.30 & 6.77 \end{pmatrix} & \hat{p}_3 &= 0.465, \end{aligned}$$

οι οποίες έχουν ως αποτέλεσμα την εκτιμηθείσα μείξη παλινδρομήσεων (στον χώρο των (x_1, y_1, y_2)) που απεικονίζεται στο Σχήμα 2.16. Η εκτίμηση που προκύπτει δεν είναι αρκετά καλή, συγκρίνοντάς την τόσο με το διάγραμμα διασποράς των δεδομένων όσο και με τις πραγματικές ευθείες παλινδρόμησης που αποτελούν την μείξη. Όπως παρατηρούμε συγκρίνοντας τις εκτιμήσεις με τις πραγματικές τιμές, το γεγονός αυτό οφείλεται στην εκτίμηση των παραμέτρων της δεύτερης συνιστώσας,



Σχήμα 2.16: Η εκτιμηθείσα μείξη παλινδρομήσεων (μαζί με την πραγματική κατάσταση κάθε παρατήρησης) αναδιατάσσοντας το προσομοιωμένο δείγμα βάσει του περιορισμού $p_1 < p_2 < p_3$ για το Παράδειγμα 2.7.3. Οι διακεκομμένες γραμμές αποτελούν τις πραγματικές ευθείες παλινδρομήσεων που αποτελούν την μείξη.

εκτίμηση που είναι αρκετά μακριά από τις αντίστοιχες πραγματικές τιμές. Σε αυτή την περίπτωση λοιπόν, η οποία είναι αρκετά τυπική, παρ' όλο που είναι πολύ εύκολο για τον δειγματολήπτη να εντοπίσει τις περιοχές υψηλής εκ των υστέρων πιθανότητας (ή τουλάχιστον κάποιες από αυτές), δεν είναι δυνατή η αντιμετώπιση του φαινομένου label switching με την υιοθέτηση της απλοϊκής τακτικής των εικονικών περιορισμών.

Είναι σαφές ότι το φαινόμενο label switching παίζει πρωταγωνιστικό ρόλο στην εκτίμηση μείξεων κατανομών (αλλά και πιο γενικών μοντέλων ελλειπών δεδομένων). Συνοψίζοντας την ανάλυση που προηγήθηκε, προκύπτουν τα εξής ζητήματα που χρήζουν περαιτέρω διερεύνησης.

Αρχικά, κρίνεται απαραίτητη η ανάπτυξη μιας γενικής μεθόδου αντιμετώπισης του προβλήματος. Οι περιορισμοί διάταξης αποδίδουν καλά μόνο σε πολύ απλές περιπτώσεις, ενώ οι υπόλοιπες διαθέσιμες μέθοδοι δεν αποτελούν μία ελκυστική εναλλακτική λύση λόγω του μεγάλου υπολογιστικού κόστους αυτών. Στο Κεφάλαιο 4 αναπτύσσεται μία νέα μέθοδος αντιμετώπισης του label switching η οποία συμβιβάζει τις παραπάνω δύο οπτικές. Είναι αρκετά απλή στην εφαρμογή της και δύναται να

θεωρηθεί ως μία γενική και αποδοτική τεχνική επίλυσης του προβλήματος.

Από την άλλη πλευρά, ο Αλγόριθμος RJMCMC για την εκτίμηση του αριθμού συνιστωσών μείξεων κανονικών κατανομών, εκμεταλλεύεται την εκ των προτέρων ταυτοποίηση των συνιστωσών μέσω ενός περιορισμού. Παρ' όλ' αυτά, είδαμε ότι η συγκεκριμένη υπόθεση οδηγεί σε μη αναμενόμενες εκτιμήσεις όταν η περιοχή υψηλής εκ των υστέρων πυκνότητας βρίσκεται κοντά στο σύνορο του περιορισμού $\mu_1 < \dots < \mu_k$. Είναι εύλογο να αναρωτηθεί κανείς πώς θα μπορούσε να απαλλαχτεί ο Αλγόριθμος RJMCMC από τα μειονεκτήματα της υπόθεσης διάταξης, η οποία όμως είναι απαραίτητη για τον σχεδιασμό των μεταβάσεων μεταξύ διαφορετικών τιμών του k . Σε αυτό απαντάει το Κεφάλαιο 3.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ

Κεφάλαιο 3

RJMCMC σε Μείξεις Κανονικών Κατανομών με Κοινές Μέσες Τιμές

3.1 Εισαγωγή

Σκοπός αυτού του κεφαλαίου είναι η μελέτη μίξεων κανονικών κατανομών στην περίπτωση που κάποιες συνιστώσες μπορεί να έχουν κοινές μέσες τιμές. Όπως είδαμε στην Ενότητα 2.5.3, ο Αλγόριθμος 2.5.3 των Richardson and Green ταυτοποιεί τις συνιστώσες της μείξης χρησιμοποιώντας τον περιορισμό $\mu_1 < \dots < \mu_k$ για κάθε δυνατή τιμή του k . Σε τυπικές περιπτώσεις, όπως στα δεδομένα από την μείξη (2.30), ο αλγόριθμος εκτιμά σωστά τον αριθμό συνιστωσών. Αυτό όμως δεν ισχύει όταν η περιοχή υψηλής εκ των υστέρων πυκνότητας βρίσκεται κοντά στο όριο του περιορισμού, όπως στα δεδομένα από την μείξη (2.31). Γενικά, σε αυτές τις περιπτώσεις, ο αλγόριθμος είτε τείνει να συνδυάζει σε μία τις συνιστώσες με τον κοινό μέσο είτε αντικαθιστά τις συνιστώσες αυτές με περισσότερες από όσες χρειάζονται. Έτσι η εκ των υστέρων κατανομή του αριθμού των συνιστωσών της μείξης έχει κορυφή σε μοντέλα όπου είτε υποεκτιμούν, είτε υπερεκτιμούν τον πραγματικό αριθμό συνιστωσών.

Αυτό το πρόβλημα μπορεί να ξεπεραστεί αν στα ανταγωνιζόμενα μοντέλα συμπεριληφθούν και αυτά που περιέχουν συνιστώσες με κοινούς μέσους. Για κάθε αριθμό συνιστωσών, ορίζουμε τέτοιου είδους μοντέλα με διαφορετικό αριθμό παραμέτρων. Ταυτόχρονα, εισάγεται ένα νέο είδος κίνησης που δρα στον χώρο των μοντέλων αυτών. Στην συνέχεια η προτεινόμενη μέθοδος συνδυάζεται κατάλληλα με τις κινήσεις split-combine και birth-death των Richardson and Green. Έτσι προκύπτει μία γενίκευση του αλγορίθμου RJMCMC όπου γίνονται μεταβάσεις μεταξύ μοντέλων που διαφέρουν τόσο σε αριθμό συνιστωσών, όσο και στις συνιστώσες που (πιθανόν) έχουν κοινούς μέσους. Σημειώνουμε ότι τέτοιου είδους μοντέλα έχουν μελετηθεί πρόσφατα υπό το πρίσμα της Κλασικής Συμπερασματολογίας από τον Garel (2007) σε πολύ απλές περιπτώσεις.

3.2 Βασικοί ορισμοί

Για να συμπεριλάβουμε πιο σύνθετα μοντέλα στην εκ των υστέρων κατανομή, θα πρέπει να υιοθετηθεί μία νέα προσέγγιση στα μοντέλα μείξεων κανονικών κατανομών. Για παράδειγμα, ας θεωρήσουμε τις μείξεις κανονικών κατανομών

$$0.3\mathcal{N}(1, 2) + 0.2\mathcal{N}(2, 4) + 0.5\mathcal{N}(5, 2.5) \quad (3.1)$$

και

$$0.3\mathcal{N}(1, 2) + 0.2\mathcal{N}(1, 4) + 0.5\mathcal{N}(5, 2.5). \quad (3.2)$$

Η συνήθης διαδικασία μοντελοποίησης θεωρεί τις κατανομές (3.1) και (3.2) ως μοντέλα με τρεις συνιστώσες. Παρ' όλ' αυτά, αν υποθέσουμε ότι διαθέτουμε δεδομένα από αυτές τις κατανομές, τότε η διαδικασία εκτίμησης δεν θα είναι ίδια αν γνωρίζουμε ότι στην μείξη (3.1) υπάρχουν τρεις διαφορετικές συνιστώσες και ότι στην μείξη (3.2) υπάρχουν δύο συνιστώσες με ίδιο μέσο. Προκειμένου όμως να συμπεριληφθούν τέτοιου είδους πληροφορίες σε μία μπεϋζιανή ανάλυση προκύπτει ένα ζήτημα μοντελοποίησης. Η «τυπική» εκ των προτέρων κατανομή του $\mu = (\mu_1, \dots, \mu_k)$ (δοθέντος του αριθμού των συνιστωσών k) είναι απολύτως συνεχής ως προς το μέτρο Lebesgue. Επομένως μοντέλα με συνιστώσες με κοινούς μέσους έχουν μηδενική εκ των προτέρων πιθανότητα αφού σε αυτήν την περίπτωση το μ ανήκει σε έναν χώρο μικρότερης διάστασης απ' ό,τι ο αριθμός των συνιστωσών. Επιπροσθέτως, ακόμα και αν τέτοιου είδους μοντέλα είχαν θετική εκ των προτέρων πιθανότητα, ο τυπικός αλγόριθμος RJMCMC δεν μπορεί να τα επισκεφτεί.

Εφ' όσον λοιπόν είναι δυνατόν να έχουμε διαφορετικά μοντέλα, δοθέντος του αριθμού των συνιστωσών της μείξης, πρέπει να οριστεί ένας τρόπος ταυτοποίησης αυτών. Για παράδειγμα, είναι βολικό να αναφερόμαστε στην κατανομή (3.1) ως ένα μοντέλο με τρεις διαφορετικές συνιστώσες, ενώ στην κατανομή (3.2) ως ένα μοντέλο όπου οι δύο πρώτες συνιστώσες μοιράζονται τον ίδιο μέσο. Τυπικά, μπορούμε να γράψουμε ότι στις κατανομές (3.1) και (3.2) αντιστοιχούν οι καταστάσεις $(1, 1, 1)$ και $(2, 1)$. Πιο αυστηρά, δίνουμε τον επόμενο ορισμό.

Ορισμός 3.2.1. *Εστώ το μοντέλο μείξης κανονικών κατανομών με k συνιστώσες*

$$x_i \sim \sum_{j=1}^k p_j \mathcal{N}(\mu_j, \sigma_j^2), \quad i = 1, \dots, n, \quad (3.3)$$

όπου $\mu_1 \leq \dots \leq \mu_k$. Αν $\kappa_0 \leq k$ είναι ο αριθμός των διακεκριμένων μέσων, τότε το διάνυσμα $\mathbf{c} = (c_1, \dots, c_{\kappa_0})$ με $c_i \in \{1, \dots, k\}$, $i = 1, \dots, \kappa_0$, και

- $\sum_{i=1}^{\kappa_0} c_i = k$
- $c_i = t$ αν $\mu_{h+1} = \dots = \mu_{h+t}$, όπου $h = \sum_{j=1}^{i-1} c_j$ (με την σύμβαση $\sum_{j=1}^0 \equiv 0$)

θα καλείται το διάνυσμα κατάστασης του μοντέλου.

Όπως είδαμε στην Ενότητα 2.2.2, χρησιμοποιώντας μόλις δύο κανονικές συνιστώσες με κοινό μέσο είναι δυνατή η ικανοποιητική προσέγγιση αρκετών κατανομών, όπως η κατανομή t_ν η οποία ουσιαστικά πρόκειται για μία συνεχή μείξη κανονικών κατανομών με κοινή μέση τιμή. Έτσι, θα θεωρήσουμε μοντέλα για τα οποία $c \in \{1, 2\}^{\kappa_0}$, δηλαδή κάθε μέσος δεν μπορεί να επαναλαμβάνεται σε παραπάνω από δύο συνιστώσες. Σύμφωνα με την προηγούμενη παρατήρηση η υπόθεση αυτή δεν πρέπει να θεωρείται πολύ περιοριστική. Οι συνιστώσες με ίδιο μέσο είναι στην ουσία μείξεις κανονικών κατανομών με διαφορετικές διασπορές. Θα αναφερόμαστε σε τέτοιου είδους συνιστώσες με τον όρο *σύνθετες συνιστώσες*. Στην αντίθετη περίπτωση θα χρησιμοποιούμε τον όρο *απλές συνιστώσες*. Το σύνολο όλων των δυνατών μοντέλων προκύπτει επιτρέποντας στο k (και επομένως στο κ_0) να μεταβάλλεται εντός κάποιου προκαθορισμένου εύρους. Έτσι, ένα συγκεκριμένο μοντέλο, καθορίζεται από τρία χαρακτηριστικά: τον συνολικό αριθμό συνιστωσών k , τον αριθμό των διακεκριμένων μέσων κ_0 και την κατανομή των κ_0 μέσων στις k συνιστώσες. Παρατηρούμε ότι και τα τρία αυτά χαρακτηριστικά περιέχονται στο διάνυσμα κατάστασης c : το άθροισμα των στοιχείων του ισούται με k , η διάστασή του είναι ίση με το κ_0 και τα στοιχεία του αυτά καθ' αυτά μας πληροφορούν σχετικά με την κατανομή των κ_0 διαφορετικών μέσων στις k συνιστώσες. Εφ' όσον λοιπόν ένα μοντέλο καθορίζεται πλήρως από το μοντέλο κατάστασης, θα αναφερόμαστε στο εξής στο μοντέλο με διάνυσμα κατάστασης c απλά με τον όρο μοντέλο c .

Πριν προχωρήσουμε στον προσδιορισμό των εκ των προτέρων κατανομών, ορίζουμε τις εξής ποσότητες.

- $N_{\kappa_0, k}$: Το πλήθος των μοντέλων δοθέντος του αριθμού συνιστωσών k και του αριθμού των διαφορετικών μέσων κ_0 , όπου $k \in \{1, \dots, K \equiv k_{\max}\}$ και $\kappa_0 \in \mathcal{K}_0(k) := \{(k+j)/2 : j = k, k-2, \dots, j_0\}$ με $j_0 = 0$ αν k άρτιος και $j_0 = 1$ αν k περιττός.
- N_k : Το πλήθος των μοντέλων δοθέντος του αριθμού συνιστωσών k . Είναι εύκολο να δούμε ότι

$$N_k = \sum_{\kappa_0 \in \mathcal{K}_0(k)} N_{\kappa_0, k} = \begin{cases} \sum_{\ell=0}^{k/2} \binom{k/2+\ell}{2\ell}, & \text{αν } k \text{ άρτιος,} \\ \sum_{\ell=0}^{(k-1)/2} \binom{(k+2\ell+1)/2}{2\ell+1}, & \text{αν } k \text{ περιττός.} \end{cases} \quad (3.4)$$

- N : Ο συνολικός αριθμός των μοντέλων, δηλαδή: $N = \sum_{k=1}^K N_k$.

Για την απόδειξη της σχέσης (3.4) παραπέμπουμε στην Ενότητα B.1 του Παραρτήματος B.

3.3 Το μπεϋζιανό ιεραρχικό μοντέλο

Στην συνέχεια, θα διατηρήσουμε τον συμβολισμό μ_j για τον μέσο της j συνιστώσας. Παρ' όλ' αυτά, θα χρησιμοποιήσουμε τον συμβολισμό $\mu_1^*, \dots, \mu_{\kappa_0}^*$ για να αναφερόμαστε στους διακεκριμένους μέσους. Παρατηρούμε ότι τα μ και μ^* συνδέονται μοναδικά μεταξύ τους μέσω του διανύσματος κατάστασης του μοντέλου αφού $\mu_j = \mu_i^*$ όπου

$$i = \min\{t \in \{1, \dots, \kappa_0\} : \sum_{s=1}^t c_s \geq j\}$$

για $j = 1, \dots, k$.

Σύμφωνα με τα παραπάνω, το μοντέλο που εισάγουμε έχει τις εξής παραμέτρους:

- τον αριθμό συνιστωσών k ,
- τον αριθμό διαφορετικών μέσων κ_0 ,
- το διάνυσμα κατάστασης \mathbf{c} ,
- τα βάρη των συνιστωσών $\mathbf{p} = (p_1, \dots, p_k)$,
- τους (διαφορετικούς) μέσους $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_{\kappa_0}^*)$ και
- τις διασπορές των συνιστωσών $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2)$.

Για τον αριθμό των συνιστωσών k θεωρούμε την μη πληροφοριακή εκ των προτέρων κατανομή $f(k) = 1/K$, $k = 1, \dots, K$, η οποία είναι η πιο συνηθισμένη επιλογή στην βιβλιογραφία (δες και Richardson and Green, 1997). Η δεσμευμένη εκ των προτέρων κατανομή του κ_0 , δοθέντος k , θα είναι η $f(\kappa_0|k) = N_{\kappa_0,k}/N_k$, $\kappa_0 \in \mathcal{K}_0(k)$. Τέλος, η εκ των προτέρων κατανομή του \mathbf{c} δοθέντος του κ_0 και του k θα είναι επίσης ομοιόμορφη, $f(\mathbf{c}|\kappa_0, k) = 1/N_{\kappa_0,k}$, $\mathbf{c} \in \{1, 2\}^{\kappa_0}$. Οδηγός για τις συγκεκριμένες εκ των προτέρων υποθέσεις είναι το γεγονός ότι επιθυμούμε να είμαστε μη πληροφοριακοί τόσο για τον αριθμό των συνιστωσών όσο και για την κατανομή των μοντέλων δοθέντος του αριθμού των συνιστωσών. Πράγματι, παρατηρούμε ότι σε αυτήν την περίπτωση προκύπτει ότι η εκ των προτέρων από κοινού κατανομή των (\mathbf{c}, κ_0) είναι

ομοιόμορφη (δοθέντος του k). Από την άλλη πλευρά, η από κοινού εκ των προτέρων κατανομή του $(\mathbf{c}, \kappa_0, k)$ είναι

$$f(\mathbf{c}, \kappa_0, k) = f(k)f(\kappa_0|k)f(\mathbf{c}|\kappa_0, k) = \frac{1}{K} \frac{N_{\kappa_0, k}}{N_k} \frac{1}{N_{\kappa_0, k}} = \frac{1}{KN_k}.$$

Εφ' όσον το N_k είναι αύξουσα συνάρτηση του k , η από κοινού εκ των προτέρων κατανομή είναι πληροφοριακή στο σύνολο όλων των πιθανών μοντέλων, υπό την έννοια ότι δίνει μεγαλύτερο βάρος σε μοντέλα με μικρότερο αριθμό συνιστωσών.

Οι εκ των προτέρων υποθέσεις για τις υπόλοιπες παραμέτρους είναι ίδιες με αυτές των Richardson and Green (1997). Πιο συγκεκριμένα,

- $\mathbf{p}|k \sim \mathcal{D}(\delta, \dots, \delta)$, δηλαδή, μία συμμετρική $(k-1)$ -διάστατη κατανομή Dirichlet.
- $\mu_j^*|\kappa_0 \sim \mathcal{N}(\xi, \kappa^{-1})$, $j = 1, \dots, \kappa_0$, ανεξάρτητα, δοθέντος όμως ότι $\mu_1^* < \dots < \mu_{\kappa_0}^*$ για λόγους ταυτοποίησης των παραμέτρων,
- $\sigma_j^{-2}|k \stackrel{\text{iid}}{\sim} \mathcal{G}(\alpha, \beta)$, $j = 1, \dots, k$, ανεξάρτητα.

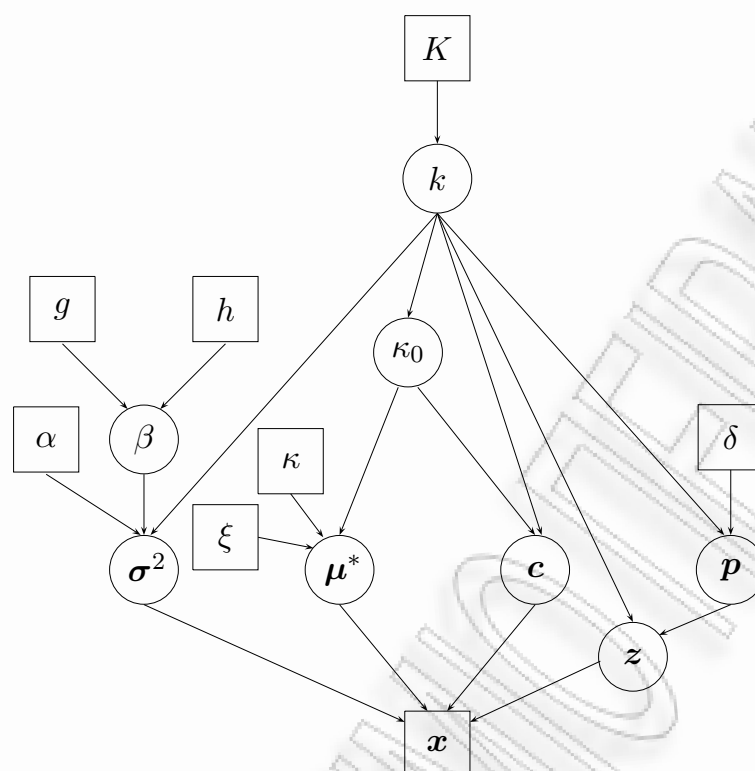
Σύμφωνα με τους Richardson and Green (1997), η παράμετρος ξ είναι η δειγματική διάμεσος και $\kappa = 1/R^2$, όπου R είναι το δειγματικό εύρος. Επί πλέον θέτουμε $\delta = 1$, $\alpha = 2$, ενώ το β είναι υπερπαραμέτρος για την οποία υποθέτουμε ότι $\beta \sim \mathcal{G}(g, h)$ με $g = 0.2$ και $h = 10/R^2$.

Το ιεραρχικό μοντέλο ολοκληρώνεται με τις εκ των προτέρων υποθέσεις για τις εικονικές μεταβλητές κατάταξης $\mathbf{z} = (z_1, \dots, z_n)$, για τις οποίες ισχύει ότι $x_i|z_i \sim f_{z_i}$, $i = 1, \dots, n$. Ως συνήθως, εκ των προτέρων έχουμε ότι $P(z_i = j|\mathbf{p}, k) = p_j$, $j = 1, \dots, k$, $i = 1, \dots, n$.

Σύμφωνα με τα παραπάνω, η από κοινού εκ των υστέρων κατανομή των παραμέτρων είναι

$$f(k, \kappa_0, \mathbf{c}, \mathbf{p}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2, \mathbf{z}, \beta|K, \delta, \xi, \kappa, \alpha, g, h, \mathbf{x}) = f(k|K)f(\kappa_0|k)f(\mathbf{c}|\kappa_0, k)f(\mathbf{p}|k, \delta)f(\boldsymbol{\mu}^*|\kappa_0, \xi, \kappa) \times f(\boldsymbol{\sigma}^2|k, \alpha, \beta)f(\mathbf{z}|k, \mathbf{p})f(\beta|g, h)f(\mathbf{x}|\mathbf{c}, \mathbf{z}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2). \quad (3.5)$$

Το μοντέλο (3.5) απεικονίζεται στο κατευθυνόμενο ακυκλικό γράφημα του Σχήματος 3.1, όπου οι κύκλοι παριστάνουν τυχαίες ποσότητες ενώ τα τετράγωνα σταθερές ή παρατηρούμενες ποσότητες.



Σχήμα 3.1: Αναπαράσταση του ιεραρχικού μοντέλου (3.5) ως κατευθυνόμενο ακυκλικό γράφημα.

3.4 Ο προτεινόμενος αλγόριθμος RJMCMC

Ο Αλγόριθμος 3.4.1 που ακολουθεί είναι ένας αλγόριθμος Reversible Jump MCMC (RJMCMC) που δρα στον χώρο των δυνατών μοντέλων. Οι κινήσεις αλλαγής μοντέλου είναι τέτοιες ώστε να εξασφαλίζεται ότι κάθε μοντέλο έχει θετική πιθανότητα επίσκεψης από οποιοδήποτε άλλο σε πεπερασμένο αριθμό βημάτων. Επί πλέον, είναι σχεδιασμένες έτσι ώστε να ικανοποιείται η υπόθεση συμφωνίας των διαστάσεων (A.1) του Green (1995). Οι παράμετροι κάθε μοντέλου προσομοιώνονται από τις πλήρεις δεσμευμένες εκ των υστέρων κατανομές αυτών. Έτσι, παράγεται μία ακολουθία τιμών

$$(k, \kappa_0, \mathbf{c}, \mathbf{p}, \boldsymbol{\mu}^*, \sigma^2, \mathbf{z}, \beta)^{(t)}, \quad t = 1, 2, \dots,$$

η οποία αποτελεί μία μαρκοβιανή αλυσίδα με οριακή κατανομή την (3.5) και επιτυγχάνει τη συνθήκη λεπτομερούς ισορροπίας (A.2). Μετά το πέρας των προσομοιώσεων μπορεί να γίνει συμπερασματολογία τόσο για την από κοινού κατανομή των ανταγω-

νιζόμενων μοντέλων όσο και για τις περιθωριακές κατανομές, όπως αυτή του αριθμού συνιστωσών, ή του αριθμού διακεκριμένων μέσων των κανονικών συνιστωσών. Επί πλέον, δεσμεύοντας σε ένα μοντέλο, είναι δυνατή η συμπερασματολογία για τις παραμέτρους αυτού.

Αλγόριθμος 3.4.1. *RJMCMC για μείξεις κανονικών κατανομών με κοινές μέσες τιμές*

Δώσε κάποιες αρχικές τιμές $k^{(0)}, \kappa_0^{(0)}, \mathbf{c}^{(0)}, (\mathbf{p}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2)^{(0)}$, και για $t \in \{1, \dots, M\}$, επανάλαβε τα παρακάτω βήματα:

1. Ανανέωση του αριθμού των συνιστωσών k .
 2. Ανανέωση του αριθμού των διακεκριμένων μέσων κ_0 και του διανύσματος κατάστασης \mathbf{c} .
 3. Ανανέωση των βαρών $\mathbf{p} = (p_1, \dots, p_k)$.
 4. Ανανέωση των διακεκριμένων μέσων $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_{\kappa_0}^*)$.
 5. Ανανέωση των διασπορών $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2)$.
 6. Ανανέωση του διανύσματος κατάταξης $\mathbf{z} = (z_1, \dots, z_n)$.
 7. Ανανέωση της υπερπαραμέτρου β .
-

Το βήμα 1 είναι υπεύθυνο για την κίνηση μεταξύ γειτονικών τιμών του k . Αυτό γίνεται χρησιμοποιώντας τους δύο τύπους κινήσεων που εισήχθησαν από τους Richardson and Green (1997), δηλαδή, τις κινήσεις split-combine και birth-death, τροποποιημένες κατάλληλα για τις ανάγκες του μοντέλου μας, όπως περιγράφεται στην Ενότητα 3.4.1. Στο βήμα 2 προτείνεται η μεταβολή του αριθμού των διακριτών μέσων κατά έναν, ενώ ταυτόχρονα μετατρέπονται κατάλληλα τα $\mathbf{p}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2$ όπως θα περιγραφεί στην Ενότητα 3.4.3 που ακολουθεί. Τα υπόλοιπα βήματα (3-7) δεν αλλάζουν την διάσταση του μοντέλου και πραγματοποιούνται μέσω του τυπικού δειγματολήπτη Gibbs που περιγράψαμε στην Ενότητα 3.4.2.

3.4.1 Οι τροποποιημένες κινήσεις αλλαγής αριθμού συνιστωσών

Οι κινήσεις που εισήχθησαν από τους Richardson and Green (1997) ενσωματώνονται στον Αλγόριθμο 3.4.1 για την αλλαγή του αριθμού των συνιστωσών του μοντέλου. Για μία πλήρη περιγραφή αυτών των κινήσεων ο αναγνώστης παραπέμπεται στην εργασία των Richardson and Green (1997). Στην συνέχεια περιγράφουμε τις τροποποιήσεις που γίνονται σε αυτές ώστε να μπορούν να χρησιμοποιηθούν υπό το πλαίσιο μας.

Η πρώτη κίνηση των Richardson and Green αποτελείται από τον διαχωρισμό μιας συνιστώσας σε δύο καινούργιες, ή τον συνδυασμό δύο (γειτονικών) συνιστωσών σε μία. Ωστόσο, στον Αλγόριθμο 3.4.1 η συγκεκριμένη κίνηση προτείνεται μόνο όταν το μοντέλο αποτελείται αποκλειστικά από απλές συνιστώσες, δηλαδή όταν ισχύει $\kappa_0 = k$. Αποφεύγουμε την περαιτέρω εφαρμογή αυτής της κίνησης για δύο λόγους: Πρώτα απ' όλα, είναι πολύ πιο πρακτικό στο υπολογιστικό μέρος του αλγορίθμου. Δεύτερον, δεν είναι απαραίτητο να υιοθετήσουμε πιο πολύπλοκες στρατηγικές, μια και η απόδοση του αλγορίθμου μας κρίνεται ικανοποιητική ως έχει.

Κατά τον δεύτερο τύπο κίνησης των Richardson and Green προτείνεται η γέννηση μιας κενής συνιστώσας, οι παράμετροι της οποίας προσομοιώνονται από την εκ των προτέρων κατανομή. Κατά την αντίστροφη κίνηση επιλέγεται στην τύχη μία κενή συνιστώσα και προτείνεται να διαγραφεί. Υπό το δικό μας πλαίσιο μία τέτοια κίνηση χρησιμοποιείται μόνο για να προτείνει απλές (και κενές) συνιστώσες. Συνεπώς κατά την αντίστροφη κίνηση δικαιούμαστε να προτείνουμε τον θάνατο μόνο κάποιας από τις υπάρχουσες απλές και κενές συνιστώσες. Αν δεν υπάρχουν τέτοιου είδους συνιστώσες, η κίνηση απορρίπτεται αμέσως. Επομένως, σε κάθε βήμα πρέπει να υπολογίζουμε τον αριθμό των απλών και κενών συνιστωσών αντί των κενών συνιστωσών όπως στην τυπική μέθοδο.

Σε αυτό το σημείο ολοκληρώνεται η περιγραφή των κινήσεων που είναι υπεύθυνες για τις μεταβάσεις μεταξύ των μοντέλων με διαφορετικό πλήθος συνιστωσών. Σημειώνεται ότι προκειμένου τελικά να καταλήξουμε στην επιλογή των κινήσεων αυτών επιστρατεύτηκαν και πιο σύνθετες μέθοδοι, χωρίς όμως να καταγραφεί κάποια ουσιαστική διαφορά στην απόδοση του Αλγορίθμου 3.4.1. Συγκεκριμένα, εφαρμόστηκαν οι τύποι κινήσεων που περιγράφονται από τους Viallefont et al. (2000), αλλά μια τέτοια προσέγγιση απλώς αύξησε την πολυπλοκότητα του αλγορίθμου χωρίς τα αποτελέσματα να αλλάζουν σημαντικά.

3.4.2 Ο δειγματολήπτης Gibbs

Οι πλήρεις δεσμευμένες εκ των υστέρων κατανομές των \mathbf{p} , σ^2 , \mathbf{z} και β είναι ακριβώς ίδιες με αυτές των Richardson and Green (1997) και περιγράφηκαν στην Ενότητα 2.5.1. Στο επόμενο λήμμα δίνουμε την πλήρη δεσμευμένη κατανομή του μ^* .

Λήμμα 3.4.1. *Η πλήρης δεσμευμένη κατανομή του μ^* είναι τέτοια ώστε*

$$\mu_j^* | \dots \sim \mathcal{N} \left(\frac{\sum_{i=l_j}^{u_j} s_i / \sigma_i^2 + \kappa \xi}{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa}, \frac{1}{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa} \right), \quad j = 1, \dots, \kappa_0,$$

όπου $l_j = \sum_{i=1}^{j-1} c_i + 1$, $u_j = \sum_{i=1}^j c_i$, με τα $\mu_1^*, \dots, \mu_{k_0}^*$ ανεξάρτητα και περιορισμένα στο σύνολο $\mu_1^* < \dots < \mu_{k_0}^*$.

Απόδειξη. Η απόδειξη παρατίθεται στην Ενότητα Β.2 του Παραρτήματος Β. □

Ο δειγματολήπτης Gibbs εφαρμόζεται, ως συνήθως, προσομοιώνοντας στην σειρά από τις παραπάνω πλήρεις δεσμευμένες εκ των υστέρων κατανομές. Παρ' όλ' αυτά, αν η νέα τιμή του μ^* παραβιάζει την διάταξη των (διακεκριμένων) μέσων τότε η τιμή αυτή απορρίπτεται και επαναλαμβάνεται η προηγούμενη.

3.4.3 Η κίνηση διαχωρισμού-συνδυασμού των μέσων

Εδώ περιγράφεται το βήμα 2 του Αλγορίθμου 3.4.1, δηλαδή η ταυτόχρονη ανανέωση του κ_0 και του c . Κατ' αρχάς, για πρακτικούς λόγους, οι προτεινόμενες μεταβάσεις γίνονται μόνο μεταξύ γειτονικών μοντέλων.

Ορισμός 3.4.1. Δύο μοντέλα με διανύσματα κατάστασης c , c' θα λέγονται γειτονικά, αν το c' προκύπτει από το c αντικαθιστώντας είτε ένα ζευγάρι γειτονικών 1 με το 2 (δηλαδή συνδιάζοντας δύο απλές γειτονικές συνιστώσες σε μία σύνθετη) είτε αντικαθιστώντας ένα 2 με ένα ζευγάρι από 1 (δηλαδή διαχωρίζοντας μία σύνθετη συνιστώσα σε δύο απλές).

Το παράδειγμα που ακολουθεί επεξηγεί τον Ορισμό 3.4.1. Ας θεωρήσουμε τα μοντέλα $c = (1, 2, 2, 1)$ και $c' = (1, 2, 1, 1, 1)$ τα οποία είναι γειτονικά και αποτελούνται από έξι συνιστώσες. Η διαφορά τους έγκειται στο ότι η τέταρτη και πέμπτη συνιστώσα του πρώτου μοντέλου μοιράζονται τον ίδιο μέσο, ενώ στο δεύτερο μοντέλο οι μέσοι αυτών των συνιστωσών διαφέρουν. Η μετάβαση από το μοντέλο c στο c' είναι μία κίνηση διαχωρισμού των μέσων ενώ η αντίστροφη μετάβαση είναι μία κίνηση συνδυασμού των μέσων. Άρα, δύο γειτονικά μοντέλα (α) έχουν τον ίδιο αριθμό συνιστωσών και (β) διαφέρουν μόνο ως προς τους μέσους δύο γειτονικών συνιστωσών. Από την άλλη πλευρά, είναι σαφές ότι τα μοντέλα $(2, 1, 2, 1)$ και $(1, 2, 2)$ δεν είναι γειτονικά στο μοντέλο c . Προφανώς το μοντέλο μίας συνιστώσας δεν έχει γειτονικά μοντέλα.

Έστω $c \in \{1, 2\}^{\kappa_0}$. Συμβολίζουμε με m_1 τον αριθμό των ζευγών γειτονικών 1 και με m_2 τον συνολικό αριθμό των 2 που υπάρχουν στο c . Προφανώς, ο αριθμός των συνολικών γειτονικών μοντέλων είναι $m_1 + m_2$. Συνεπώς, κατά το βήμα 2 του Αλγορίθμου 3.4.1 προχωρούμε ως εξής: Δοθέντος ότι το τρέχον μοντέλο είναι το c ,

προτείνουμε μία μετάβαση στο c' ομοιόμορφα στον χώρο των γειτονικών μοντέλων, δηλαδή με πιθανότητα ίση με $1/(m_1 + m_2)$.

Χωρίς βλάβη της γενικότητας, ας θεωρήσουμε την μετάβαση μεταξύ των μοντέλων $c = (c_1, \dots, c_{k_0})$ και $c' = (c'_1, c'_2, \dots, c'_{k_0+1})$ όπου $c_1 = 2$ και $c'_1 = c'_2 = 1$ και $c'_j = c_{j-1}$ για $j = 3, \dots, k_0 + 1$. Το c' προκύπτει από το c μέσω μιας κίνησης διαχωρισμού μέσων και αντιστρόφως. Ας συμβολίσουμε τα βάρη, μέσους και διασπορές των δύο πρώτων συνιστωσών του μοντέλου μείξης υπό το μοντέλο c με (q_1, ν_1^*, τ_1^2) , (q_2, ν_1^*, τ_2^2) , αντίστοιχα, ενώ υπό το μοντέλο c' με $(p_1, \mu_1^*, \sigma_1^2)$, $(p_2, \mu_2^*, \sigma_2^2)$. Οι παράμετροι των υπόλοιπων $k - 2$ συνιστωσών παραμένουν ίδιες. Εφ' όσον το c' έχει μόνο μία παράμετρο επί πλέον από το μοντέλο c , η υπόθεση συμφωνίας των διαστάσεων (Green, 1995) απαιτεί την παραγωγή μιας παραπάνω τυχαίας μεταβλητής κατά την μετάβαση από το c στο c' σε σχέση με το πλήθος των τυχαίων μεταβλητών που παράγουμε κατά την αντίστροφη μετάβαση. Εδώ, προτείνουμε να μεταβούμε από το $(q_1, q_2, \nu_1^*, \tau_1^2, \tau_2^2)$ στο $(p_1, p_2, \mu_1^*, \mu_2^*, \sigma_1^2, \sigma_2^2)$ παράγοντας δύο τυχαίες μεταβλητές, ενώ για την αντίστροφη μετάβαση θα παράγουμε μόλις μία.

Στην συνέχεια, περιγράφουμε τους σχετικούς μετασχηματισμούς και περιορισμούς που εξασφαλίζουν ότι οι κινήσεις διαχωρισμού-συνδυασμού των μέσων είναι αντιστρέψιμες μεταξύ τους. Φυσικά, ένα ζεύγος αντιστρέψιμων μετασχηματισμών θα πρέπει να επιλεγεί προσεκτικά, καθώς σε αντίθετη περίπτωση οι προτεινόμενες κινήσεις θα έχουν πολύ μικρή πιθανότητα αποδοχής με αποτέλεσμα την αργή σύγκλιση του αλγορίθμου στην εκ των υστέρων κατανομή. Οδηγός για τους μετασχηματισμούς αυτούς είναι το κριτήριο συμφωνίας της μηδενικής, πρώτης και δεύτερης ροπής (δες επίσης Richardson and Green, 1997). Αυτό σημαίνει ότι οι συγκεκριμένες ροπές θα παραμείνουν ίδιες πριν και μετά την μετάβαση, δηλαδή θα ισχύει ότι

$$\begin{aligned} q_1 + q_2 &= p_1 + p_2 \\ q_1 \nu_1^* + q_2 \nu_1^* &= p_1 \mu_1^* + p_2 \mu_2^* \\ q_1 (\nu_1^{*2} + \tau_1^2) + q_2 (\nu_1^{*2} + \tau_2^2) &= p_1 (\mu_1^{*2} + \sigma_1^2) + p_2 (\mu_2^{*2} + \sigma_2^2). \end{aligned}$$

Οι μετασχηματισμοί που παρουσιάζονται στην συνέχεια εξασφαλίζουν την ισχύ των παραπάνω εξισώσεων, κάτι που μπορεί να δειχτεί πολύ εύκολα μέσω στοιχειώδους άλγεβρας.

Μετάβαση από το c στο c' : Παράγουμε $u_1, u_2 \stackrel{\text{iid}}{\sim} \text{Beta}(2, 2)$ και θέτουμε

$$\begin{aligned} (p_1, p_2) &= ((q_1 + q_2)u_1, (q_1 + q_2)(1 - u_1)), \\ (\sigma_1^2, \sigma_2^2) &= \left(\frac{p_2 q_1 \tau_1^2 u_2^2 - p_1 (p_2 + q_2 \tau_2^2 u_2^2) + \sqrt{D}}{2p_1 p_2 u_2^2}, \frac{p_1 q_2 \tau_2^2 u_2^2 - p_2 (p_1 + q_1 \tau_1^2 u_2^2) + \sqrt{D}}{2p_1 p_2 u_2^2} \right), \end{aligned}$$

$$(\mu_1^*, \mu_2^*) = (\nu_1^* - u_2 \sigma_1 \sigma_2 \sqrt{p_2/p_1}, \nu_1^* + u_2 \sigma_1 \sigma_2 \sqrt{p_1/p_2}),$$

όπου $D = 4p_1^2 p_2 q_2 \tau_2^2 u_2^2 + \{p_1 p_2 + (p_2 q_1 \tau_1^2 - p_1 q_2 \tau_2^2) u_2^2\}^2$. Σε περίπτωση που παραβιάζεται η διάταξη των $k_0 + 1$ μέσων, η προτεινόμενη κίνηση απορρίπτεται. Διαφορετικά, πρώτα επανακατατάσσουμε μεταξύ τους τις παρατηρήσεις που ήδη ανήκαν σε αυτές τις συγκεκριμένες συνιστώσες όπως περιγράφεται στην Ενότητα Β.3 του Παραρτήματος Β.

Πρόταση 3.4.1. Η πιθανότητα αποδοχής της κίνησης διαχωρισμού-συνδυασμού των μέσων είναι ίση με $\min\{1, A_1\}$, όπου

$$\begin{aligned} A_1 &= \frac{f(\mathbf{x}|\mathbf{c}', \mathbf{z}', \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2)}{f(\mathbf{x}|\mathbf{c}, \mathbf{z}, \boldsymbol{\nu}^*, \boldsymbol{\tau}^2)} \times \frac{(\kappa_0 + 1)f(\kappa_0 + 1|k)f(\mathbf{c}'|\kappa_0 + 1, k)}{f(\kappa_0|k)f(\mathbf{c}|\kappa_0, k)} \\ &\times \sqrt{\frac{\kappa}{2\pi}} \exp\left\{-\frac{\kappa}{2} [(\mu_1^* - \xi)^2 + (\mu_2^* - \xi)^2 - (\nu_1^* - \xi)^2]\right\} \\ &\times \left(\frac{\tau_1^2 \tau_2^2}{\sigma_1^2 \sigma_2^2}\right)^{\alpha+1} \exp\left\{\beta(\tau_1^{-2} + \tau_2^{-2} - \sigma_1^{-2} - \sigma_2^{-2})\right\} \\ &\times \left(\frac{p_1^{\delta-1+n'_1} p_2^{\delta-1+n'_2}}{q_1^{\delta-1+n_1} q_2^{\delta-1+n_2}}\right) \left(\frac{m_1 + m_2}{m'_1 + m'_2}\right) \left(\frac{P'_{\text{realloc}}}{P_{\text{realloc}}}\right) \frac{g_{2,2}(\tilde{u}_1)}{g_{2,2}(u_1)g_{2,2}(u_2)} |J|. \end{aligned} \quad (3.6)$$

Για την απόδειξη παραπέμπουμε στην Ενότητα Β.3 του Παραρτήματος Β. Στην (3.6)

$$|J| = \frac{q_1 q_2 (p_1 + p_2)^3}{(p_1^2 + p_2^2) \sigma_1^2 \sigma_2^2 + p_1 p_2 \{2\sigma_1^2 \sigma_2^2 + (\sigma_1^2 + \sigma_2^2)(\mu_1^* - \mu_2^*)^2\}} \left(\frac{\sigma_1^2 \sigma_2^2}{p_1 p_2}\right)^{3/2}$$

είναι η Ιακωβιανή του μετασχηματισμού, $g_{2,2}(\cdot)$ είναι η πυκνότητα πιθανότητας της κατανομής $Beta(2, 2)$ και

- $P_{\text{realloc}}, P'_{\text{realloc}}$ είναι οι πιθανότητες της συγκεκριμένης ανακατάταξης όπως προσδιορίζονται στο Παράρτημα Β,
- \mathbf{z}' είναι το διάνυσμα κατάταξης των παρατηρήσεων στην προτεινόμενη κατάσταση,
- (n_1, n_2) είναι όπως ορίζονται στην (2.29) και (n'_1, n'_2) είναι οι αντίστοιχες συχνότητες στο προτεινόμενο μοντέλο,
- $m'_1 + m'_2$ είναι το πλήθος των γειτονικών μοντέλων του \mathbf{c}' ,
- $\tilde{u}_1 = q_1/(q_1 + q_2)$, έτσι ώστε ο μετασχηματισμός να είναι αντιστρέψιμος (δες παρακάτω για την αντίστροφη μετάβαση).

Μετάβαση από το c' στο c : Παράγουμε $\tilde{u}_1 \sim \text{Beta}(2, 2)$ και θέτουμε

$$\begin{aligned} (q_1, q_2) &= ((p_1 + p_2)\tilde{u}_1, (p_1 + p_2)(1 - \tilde{u}_1)) \\ (\tau_1^2, \tau_2^2) &= \left(\frac{p_1}{q_1} \sigma_1^2 (1 + u_2^2 \sigma_2^2), \frac{p_2}{q_2} \sigma_2^2 (1 + u_2^2 \sigma_1^2) \right), \\ \nu_1^* &= \frac{p_1 \mu_1^* + p_2 \mu_2^*}{p_1 + p_2}, \end{aligned}$$

όπου

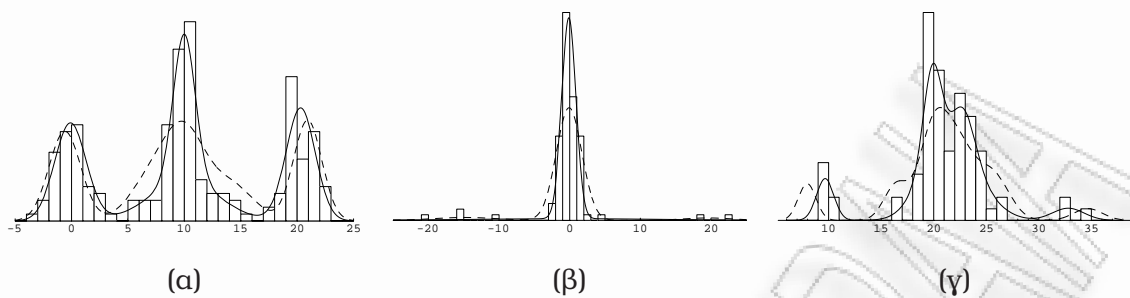
$$u_2 = \frac{\sqrt{p_1 p_2}}{p_1 + p_2} \frac{\mu_2^* - \mu_1^*}{\sigma_1 \sigma_2}.$$

Σαφώς, σε αυτήν την περίπτωση η διάταξη των μέσων δεν μπορεί να παραβιαστεί. Η πιθανότητα αποδοχής αυτής της κίνησης είναι ίση με $\min\{1, A_1^{-1}\}$, με το A_1 να ορίζεται όπως και στην (3.6) αλλά με το $\kappa_0 - 1$ στην θέση του κ_0 .

3.5 Εφαρμογές

Σε αυτήν την ενότητα εφαρμόζουμε την προτεινόμενη μέθοδο χρησιμοποιώντας παραδείγματα με προσομοιωμένα σύνολα δεδομένων καθώς και το γνωστό γαλαξιακό σύνολο δεδομένων (Roeder, 1992). Σε όλες τις περιπτώσεις τα αποτελέσματα συγκρίνονται με τα αντίστοιχα του Αλγορίθμου 2.5.3 των Richardson and Green (1997). Οι Αλγόριθμοι 2.5.3 και 3.4.1 εκτελέστηκαν στην γλώσσα προγραμματισμού Fortran 90 σε έναν τυπικό υπολογιστή Pentium IV.

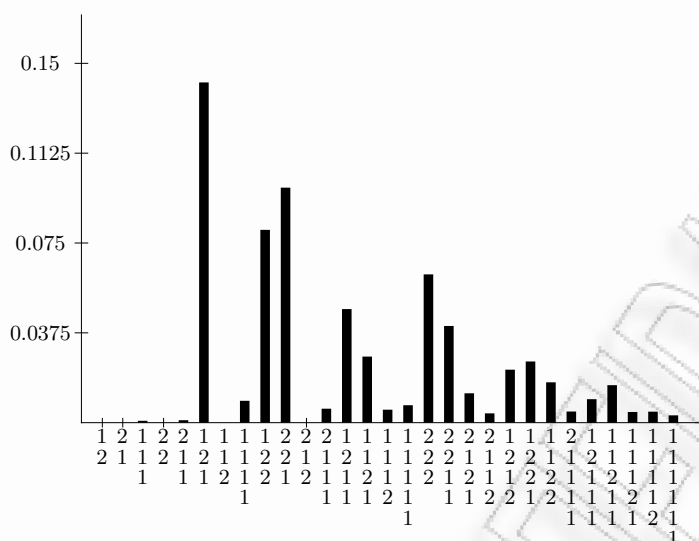
Ο πρώτος στόχος της ανάλυσης είναι η εξερεύνηση της εκ των υστέρων κατανομής των μοντέλων. Ακολουθώς, μας απασχολεί η εκτίμηση των παραμέτρων των μοντέλων, κάτι που είναι μια πολύ απλή διαδικασία στον αλγόριθμο των Richardson and Green λόγω της ταυτοποίησης όλων των παραμέτρων εξαιτίας του περιορισμού διάταξης των μέσων. Κάτι τέτοιο δεν ισχύει στην μέθοδό μας. Ας θεωρήσουμε για παράδειγμα ότι ενδιαφερόμαστε να εκτιμήσουμε τις παραμέτρους δεσμεύοντας σε ένα συγκεκριμένο μοντέλο. Στην περίπτωση των απλών συνιστωσών, τόσο οι διασπορές όσο και τα βάρη είναι πάντα ταυτοποιήσιμα διότι οι προσομοιωμένες τιμές των παραμέτρων ταυτοποιούνται αναδιατάσσοντας εκ των προτέρων τους διαφορετικούς μέσους. Αντιθέτως, αυτές οι παράμετροι παύουν να είναι ταυτοποιήσιμες εντός μιας σύνθετης συνιστώσας λόγω του φαινομένου label switching. Εν τούτοις, δεν υιοθετούμε επί πλέον περιορισμούς διάκρισης στις παράμετρους των σύνθετων συνιστωσών (όπως για παράδειγμα μία εκ των προτέρων διάταξη των διασπορών). Κάτι τέτοιο θα περιοριζε ακόμα περισσότερο τον παραμετρικό χώρο και θα είχαμε επιπτώσεις όπως η απόρριψη προτεινόμενων κινήσεων που θα γίνονταν δεκτές σε διαφορετική περίπτωση. Συνεπώς θεωρούμε μη περιορισμένο τον παραμετρικό χώρο των διασπορών



Σχήμα 3.2: Ιστογράμματα των συνόλων δεδομένων που χρησιμοποιούνται στα παραδείγματα μαζί με τις εκτιμηθείσες πυκνότητες, δεσμεύοντας στο μοντέλο με την μεγαλύτερη εκ των υστερών πιθανότητα, σύμφωνα με τον Αλγόριθμο 3.4.1 (συνεχόμενη γραμμή) και τον Αλγόριθμο 2.5.3 (διακεκομμένη γραμμή): Προσομοιωμένα σύνολα δεδομένων από τις κατανομές: (α) (2.31), (β) (3.7), (γ) Γαλαξιακό σύνολο δεδομένων.

και βαρών, κάτι που συνεπάγεται ότι η εκ των υστερών κατανομή θα έχει ένα ακέραιο πολλαπλάσιο 2^{m_2} συμμετρικών κορυφών (δοθέντος ότι δεσμεύουμε σε ένα μοντέλο με διάνυσμα κατάστασης που περιέχει m_2 φορές το στοιχείο 2).

Ως εκ τούτου, για να λάβουμε κατάλληλες εκτιμήσεις των παραμέτρων μιας σύνθετης συνιστώσας θα πρέπει να λύσουμε το πρόβλημα label switching εντός των σύνθετων συνιστωσών. Κάτι τέτοιο μπορεί να γίνει απλά μέσω αναδιάταξης του προσομοιωμένου δείγματος, δηλαδή, μετά το πέρας των προσομοιώσεων. Συγκεκριμένα, χρησιμοποιούμε τον αλγόριθμο 2.6.1 Πιλοτικής Αναδιάταξης βασισμένοι στον Maximum A Posteriori (MAP) εκτιμητή (δες Marin et al., 2005, ή Marin and Robert, 2007). Όπως περιγράψαμε στην Ενότητα 2.6.2, αυτή η τεχνική επιλέγει εκείνη την αναδιάταξη του εκάστοτε προσομοιωμένου διανύσματος παραμέτρων που ελαχιστοποιεί την απόστασή του από τον προσεγγιστικό MAP εκτιμητή. Έτσι, μπορούμε να λάβουμε ταυτοποιήσιμες παραμέτρους που θα μας οδηγήσουν σε κατάλληλες εκτιμήσεις χωρίς να απαιτήσουμε κάποια επί πλέον διάταξη στον παραμετρικό χώρο. Παρά τα μειονεκτήματα της μεθόδου αυτής σε γενικά πλαίσια, εδώ πρόκειται για μια αποδοτική λύση καθώς εφαρμόζεται μόνο σε δύο συνιστώσες κάθε φορά (αυτές δηλαδή που αντιστοιχούν στα ζευγάρια των σύνθετων συνιστωσών του μοντέλου) και έτσι δεν δημιουργούνται προβλήματα όπως η ύπαρξη πολλών κορυφών. Εξ άλλου, εφ' όσον οι σύνθετες συνιστώσες διαφέρουν μόνο ως προς την διασπορά τους, είναι σαφές ότι η μέθοδος πιλοτικής αναδιάταξης είναι ικανή να βρει την σωστή αναδιάταξη. Τέλος, αναφέρουμε ότι σε όλα τα παραδείγματα που εξετάσαμε, η μέθοδος της Πιλοτικής Αναδιάταξης εντός των σύνθετων συνιστωσών καταλήγει σε σχεδόν ταυτό-



Σχήμα 3.3: Η εκτιμηθείσα εκ των υστέρων κατανομή του (c, κ_0, k) για το προσομοιωμένο σύνολο δεδομένων από την κατανομή (2.31).

σημα αποτελέσματα με την αναδιάταξη των προσομοιωμένες τιμών αυτών βάσει ενός περιορισμού διάταξης στις διασπορές τους.

3.5.1 Προσομοιωμένο σύνολο από την κατανομή (2.31)

Στην Ενότητα 2.5.3 είδαμε ότι για το προσομοιωμένο σύνολο δεδομένων από την κατανομή (2.31) ο Αλγόριθμος 2.5.3 δεν επιλέγει τον σωστό αριθμό συνιστωσών. Πιο συγκεκριμένα, ενώ τα δεδομένα προσομοιώθηκαν από μία μείξη τεσσάρων κατανομών, η μεγαλύτερη εκ των υστέρων πιθανότητα αντιστοιχίζεται στις πέντε συνιστώσες. Επί πλέον, στο Σχήμα 3.2(α) φαίνεται η εκτίμηση της πυκνότητας των δεδομένων (διακεκομμένη γραμμή) δεσμεύοντας στο πιο πιθανό μοντέλο κατά τον τυπικό αλγόριθμο. Από αυτήν συμπεραίνουμε την όχι και τόσο καλή προσαρμογή του στα δεδομένα.

Στην συνέχεια παρουσιάζουμε τα αντίστοιχα αποτελέσματα σύμφωνα με την προτεινόμενη μέθοδο. Εκτελέσαμε τον Αλγόριθμο 3.4.1 για τον ίδιο αριθμό επαναλήψεων (300000), ενώ θέσαμε $1 \leq k \leq K = 15$. Διαγράφοντας τις πρώτες 30000 τιμές ως περίοδο burn-in, καταλήξαμε στην εκτιμηθείσα εκ των υστέρων κατανομή του (c, k, κ_0) που απεικονίζεται στο Σχήμα 3.3 (παρουσιάζονται μόνο τα μοντέλα εντός του εύρους $3 \leq k \leq 6$, τα οποία συγκεντρώνουν τη μεγαλύτερη μάζα της εκ των υστέρων κατανομής). Παρατηρούμε ότι η μέθοδός μας υποδεικνύει ότι το πραγματικό μοντέλο έχει την μεγαλύτερη εκ των υστέρων πιθανότητα. Πιο συγκεκριμένα, η κορυφή αντιστοιχεί στο μοντέλο (1, 2, 1), το οποίο είναι αυτό που παράγαγε τα δε-

Δεδομένα	Mean S-C	S-C	B-D
Μείξη (2.31)	6.85%	6.86%	7.05%
Μείξη (3.7)	2.46%	4.52%	4.74%
Γαλαξιακό	8.69%	10.07%	13.95%

Πίνακας 3.1: Ποσοστό των αποδεκτών κινήσεων για τα τρία σύνολα δεδομένων, για κάθε τύπο μεταβάσεων: διαχωρισμός-συνδυασμός μέσω (Mean S-C), διαχωρισμός-συνδυασμός (S-C), γέννηση-θάνατος B-D.

δομένα. Επί πλέον, η εκτιμηθείσα εκ των υστέρων πιθανότητα αυτού του μοντέλου είναι μιάμιση φορά μεγαλύτερη από την αμέσως επόμενη, που αντιστοιχεί στο μοντέλο (2,2,1). Από την άλλη πλευρά, είναι αξιοσημείωτο ότι τώρα το μοντέλο (1,1,1,1,1) που αντιστοιχεί στο μοντέλο πέντε συνιστώσων το οποίο ήταν το κυρίαρχο σύμφωνα με τον τυπικό αλγόριθμο RJMCMC έχει πολύ μικρότερη εκ των υστέρων πιθανότητα υπό την μέθοδό μας. Βεβαίως, βλέπουμε ότι τα μοντέλα με τις αμέσως μεγαλύτερες εκ των υστέρων πιθανότητες είναι τα (2,2,1) και (1,2,2) τα οποία αντιστοιχούν σε $k = 5$ συνιστώσες (και $\kappa_0 = 3$). Η πιθανότητα αποδοχής των προτεινόμενων κινήσεων, κυμαίνεται γύρω στο 7% για όλους τους τύπους κινήσεων (βλ. Πίνακα 3.1).

Ακολουθως θεωρούμε το πρόβλημα εκτίμησης των παραμέτρων δεσμεύοντας στο κυρίαρχο μοντέλο, δηλαδή το (1,2,1). Όπως περιγράφηκε πριν, οι παράμετροι που αντιστοιχούν στις συνιστώσες με τον ίδιο μέσο δεν είναι ταυτοποιήσιμες. Έτσι, σε αυτό το παράδειγμα το φαινόμενο label switching συμβαίνει μεταξύ των διασπορών και βαρών της τρίτης και τέταρτης συνιστώσας. Οι αντίστοιχες παράμετροι των υπόλοιπων συνιστώσων είναι ταυτοποιήσιμες. Στο Σχήμα 3.4 απεικονίζονται οι συνεχείς εναλλαγές μεταξύ των τιμών της δεύτερης και τρίτης διασποράς. Η MAP εκτίμηση που παρατηρήσαμε έχει ως εξής:

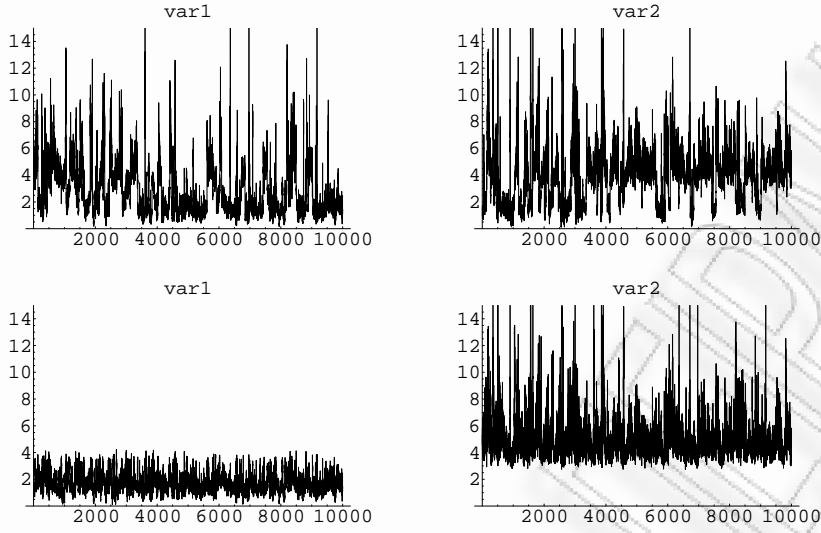
$$\boldsymbol{\mu}(\text{MAP}) \approx (0.01, 10.09, 10.09, 20.28)$$

$$\boldsymbol{\sigma}^2(\text{MAP}) \approx (1.86, 0.704, 6.36, 1.51)$$

$$\boldsymbol{p}(\text{MAP}) \approx (.248, .249, .226, 276).$$

Σημειώνουμε ότι η εκτίμηση αυτή δεν είναι καθόλου κακή, συγκρίνοντάς την με τις πραγματικές τιμές των παραμέτρων. Ύστερα από την αναδιάταξη των προσομοιωμένων τιμών των βαρών και διασπορών της δεύτερης και τρίτης συνιστώσας, σύμφωνα με τον αλγόριθμο Πιλοτικής Αναδιάταξης παρατηρούμε ότι εξαλείφεται το πρόβλημα label switching (βλ. Σχήμα 3.4) και καταλήγουμε στις εξής εκτιμήσεις:

$$\widehat{\mathbb{E}}(\boldsymbol{\mu}|\boldsymbol{x}) = (-0.10, 9.99, 9.99, 20.31)$$



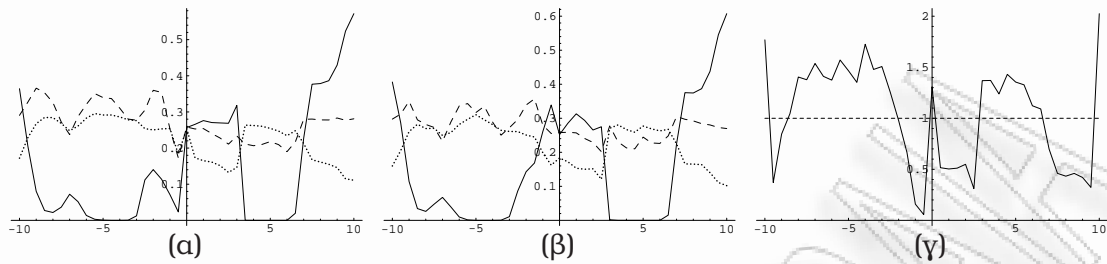
Σχήμα 3.4: Πάνω: Το φαινόμενο label switching για τις διασπορές της δεύτερης και τρίτης συνιστώσας του προσομοιωμένου συνόλου δεδομένων από την μείξη (2.31). Κάτω: Οι αντίστοιχες αναδιατεταγμένες τιμές ύστερα από την εφαρμογή του αλγορίθμου Πιλοτικής Αναδιάταξης.

$$\hat{\mathbb{E}}(\sigma^2 | \mathbf{x}) = (1.96, 0.95, 9.45, 1.68)$$

$$\hat{\mathbb{E}}(\mathbf{p} | \mathbf{x}) = (.250, .262, .221, .267).$$

Τονίζουμε ότι στην παραπάνω αναδιάταξη δεν συμμετέχουν οι τιμές της πρώτης και τέταρτης συνιστώσας. Παρατηρούμε ότι οι εκ των υστέρων μέσες τιμές είναι αρκετά κοντά στις πραγματικές τιμές των παραμέτρων, με μία εξαίρεση ίσως για τις διασπορές της τρίτης και τέταρτης συνιστώσας. Στο Σχήμα 3.2(a) φαίνεται η εκτιμηθείσα συνάρτηση πυκνότητας δεδομένων δεσμεύοντας στο μοντέλο (1,2,1). Στο ίδιο Σχήμα φαίνεται και η αντίστοιχη εκτίμηση με τον τυπικό αλγόριθμο RJMCMC για λόγους σύγκρισης. Βλέπουμε ότι η μέθοδός μας καταλήγει σε μία εκτίμηση που προσαρμόζεται πολύ καλά στα δεδομένα σε αντίθεση με την προηγούμενη μέθοδο όπου καταγράφεται κακή προσαρμογή ιδιαίτερα στο κεντρικό τμήμα τους.

Ακολουθώντας, ασχολούμαστε με την ανάλυση ευαισθησίας της λύσης μας. Επικεντρωνόμαστε στην εξέταση της απόδοσης του Αλγορίθμου 3.4.1 στην περίπτωση όπου όλες οι συνιστώσες έχουν διαφορετικούς μέσους ως συνάρτηση της διαφοράς $\mu_2 - \mu_3$. Ειδικότερα, τρέξαμε τον αλγόριθμο προσθέτοντας μία σταθερά ϵ στις παρατηρήσεις που ήταν αρχικώς καταταγμένες στην δεύτερη ή τρίτη συνιστώσα. Έτσι, για κάθε $\epsilon \notin \{-10, 0, 10\}$ το πραγματικό μοντέλο είναι το (1, 1, 1, 1) (με $k = \kappa_0 = 4$), δηλαδή



Σχήμα 3.5: (α) Εκ των υστέρων πιθανότητες των $k = 3$ (συνεχής γραμμή), 4 (διακεκομμένη γραμμή) και 5 (εσιγμένη γραμμή) σύμφωνα με τον τυπικό αλγόριθμο RJMCMC. (β) Οι αντίστοιχες εκτιμήσεις της εκ των υστέρων πιθανότητας του κ_0 σύμφωνα με την προτεινόμενη μέθοδο. (γ) Εκ των υστέρων λόγοι της πιθανότητας του πραγματικού μοντέλου είτε ως προς το πιο πιθανό μοντέλο (όταν η μέθοδος μας δεν βρίσκει το πραγματικό μοντέλο) είτε ως προς το δεύτερο πιο πιθανό μοντέλο (όταν η μέθοδος μας βρίσκει το πραγματικό μοντέλο).

η μείξη κανονικών κατανομών

$$0.25\mathcal{N}(0, 2) + 0.25\mathcal{N}(10 + \epsilon I_{\{\epsilon < 0\}}, 7) + 0.25\mathcal{N}(10 + \epsilon I_{\{\epsilon > 0\}}, 1) + 0.25\mathcal{N}(20, 1).$$

Θεωρήσαμε $\epsilon = -10(.5)10$ και για κάθε τιμή εκτελέσαμε για 300000 επαναλήψεις (ύστερα από την διαγραφή των πρώτων 30000 επαναλήψεων ως περίοδος burn-in) τον Αλγόριθμο 3.4.1 αλλά και τον τυπικό Αλγόριθμο 2.5.3.

Στο Σχήμα 3.5(β) φαίνονται οι εκ των υστέρων πιθανότητες των $\kappa_0 = 3, 4$ και 5 συναρτήσει του ϵ . Για λόγους σύγκρισης παραθέτουμε και τις αντίστοιχες εκ των υστέρων πιθανότητες του αριθμού των συνιστωσών k σύμφωνα με την μέθοδο των Richardson and Green. Παρατηρούμε ότι οι συγκεκριμένες πιθανότητες είναι αρκετά κοντά όταν το $|\epsilon|$ είναι μακριά από το μηδέν, ενώ για μικρές τιμές του $|\epsilon|$ κάθε αλγόριθμος συμπεριφέρεται όπως ακριβώς και στην περίπτωση $\epsilon = 0$. Επί πλέον, όταν $\epsilon = -10$ ο αλγόριθμος υποδεικνύει τον σωστό αριθμό διακεκριμένων μέσων ενώ η τυπική μέθοδος υποεκτιμά το k . Μία μάλλον περίεργη συμπεριφορά παρατηρείται όταν $\epsilon \in [3.5, 6.5]$ όπου και οι δύο μέθοδοι υποδεικνύουν ότι η επικρατέστερη τιμή του k ή του κ_0 είναι ίση με 5. Εν τούτοις, η εκ των υστέρων κατανομή υπό την προτεινόμενη προσέγγιση έχει την κορυφή της στο σωστό μοντέλο. Γενικά, παρατηρούμε ότι το σύνολο όπου η μέθοδος μας υποδεικνύει το σωστό μοντέλο είναι το $\{-10\} \cup [-8.5, -2.5] \cup \{0\} \cup [2.5, 6.5] \cup \{10\}$.

Για να πάρουμε μια ιδέα για την πιθανότητα του σωστού μοντέλου έχουμε απεικονίσει τους εκ των υστέρων λόγους της, είτε ως προς την πιθανότητα του πιο πιθανού μοντέλου (όταν η μέθοδος μας αποτυγχάνει) είτε ως προς το δεύτερο πιο

ϵ	Πραγματικό μοντέλο	k (RG)	k_0	Επικρατέστερο μοντέλο
-10	(2,1,1)	3	3	(2,1,1)
-9.5	(1,1,1,1)	4	4	(2,1,1)
-9	(1,1,1,1)	4	4	(2,1,1,1)
[-8.5, -7.5]	(1,1,1,1)	4	4	(1,1,1,1)
-7	(1,1,1,1)	5	5	(1,1,1,1)
[-6.5, -4]	(1,1,1,1)	4	4	(1,1,1,1)
-3.5	(1,1,1,1)	4	5	(1,1,1,1)
[-3, -2.5]	(1,1,1,1)	4	4	(1,1,1,1)
[-2, -1.5]	(1,1,1,1)	4	4	(1,1,2,1)
-1	(1,1,1,1)	5	4	(1,2,1)
-0.5	(1,1,1,1)	5	3	(1,2,1)
0	(1,2,1)	5	3	(1,2,1)
0.5	(1,1,1,1)	5	3	(1,2,1)
1	(1,1,1,1)	5	5	(1,2,1,1)
1.5	(1,1,1,1)	4	4	(1,2,1)
2	(1,1,1,1)	4	4	(1,2,1,1)
[2.5, 3]	(1,1,1,1)	5	4	(1,1,1,1)
[3.5, 6.5]	(1,1,1,1)	5	5	(1,1,1,1)
7	(1,1,1,1)	4	4	(1,1,1)
[7.5, 9.5]	(1,1,1,1)	3	3	(1,1,1)
10	(1,1,1)	3	3	(1,1,1)

Πίνακας 3.2: Αποτελέσματα ανάλυσης ευαισθησίας. ϵ : διαφορά μέσω δεύτερης και τρίτης συνιστώσας, k (RG): η κορυφή της εκ των υστέρων κατανομής του αριθμού των συνιστωσών σύμφωνα με τον αλγόριθμο των Richardson and Green (1997), k_0 : η κορυφή της εκ των υστέρων κατανομής των διαφορετικών μέσων σύμφωνα με την προτεινόμενη μέθοδο.

πιθανό μοντέλο (όταν η μέθοδος μας επιτυγχάνει), βλ. Σχήμα 3.5(γ). Τέλος, στον Πίνακα 3.2 παρουσιάζονται τα πιο πιθανά μοντέλα μαζί με την εκ των υστέρων κορυφή του αριθμού των συνιστωσών και διαφορετικών μέσων υπό την τυπική και την προτεινόμενη μέθοδο αντίστοιχα.

3.5.2 Έκτροπες παρατηρήσεις

Τα μοντέλα μείξεων μπορούν να εφαρμοστούν στην περίπτωση όπου το υπό μελέτη σύνολο δεδομένων περιέχει έκτροπες παρατηρήσεις. Μία τέτοια περίπτωση εμφανίζεται όταν σε μία από τις συνιστώσες της μείξης αντιστοιχεί πολύ μικρό βάρος. Τότε, η εκ των υστέρων κατανομή των παραμέτρων μπορεί να έχει αρκετές κορυφές, καθώς δημιουργούνται προβλήματα ταυτοποίησης (Hoogerheide et al., 2007). Υπό το πλαίσιο που εισάγουμε, μπορούμε να ελέγξουμε την απόδοση του τυπικού και του προτεινόμενου αλγορίθμου σε τέτοιου είδους καταστάσεις. Θεωρούμε ένα

προσομοιωμένο σύνολο δεδομένων $n = 100$ παρατηρήσεων από την μείζη κατανομών

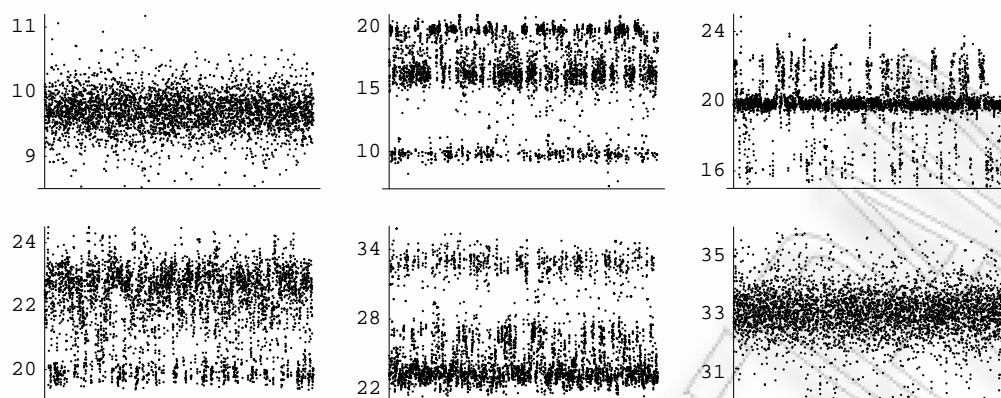
$$0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(3, 100), \quad (3.7)$$

το ιστόγραμμα των οποίων απεικονίζεται στο Σχήμα 3.2(β). Παρατηρούμε ότι πρόκειται για αρκετά προκλητική περίπτωση για τους συνήθεις δειγματολήπτες, καθώς οι έξι ακραίες παρατηρήσεις μπορούν να σχηματίσουν ξεχωριστές συνιστώσες αντί να καταταχθούν σε μία όπως ισχύει στην πραγματικότητα.

Η προτεινόμενη μέθοδος υποδεικνύει το μοντέλο (2) ως αυτό με την μεγαλύτερη εκ των υστέρων πιθανότητα. Εξαιτίας του μικρού αριθμού παρατηρήσεων, είναι αρκετά δύσκολο να κυριαρχήσει το μοντέλο (1,1), το οποίο είναι το σωστό. Παρ' όλα αυτά, η κυριαρχία του μοντέλου (2) έχει ως αποτέλεσμα οι έξι έκτροπες παρατηρήσεις να είναι όντως καταταγμένες σε μία συνιστώσα. Αυτό δεν ισχύει για τον τυπικό αλγόριθμο RJMCMC, όπου το μεγάλο εύρος των ακραίων παρατηρήσεων και ο περιορισμός διάταξης των μέσων, έχει ως συνέπεια την ομαδοποίηση των πολύ μικρών και πολύ μεγάλων παρατηρήσεων σε διαφορετικές συνιστώσες. Έτσι, ο τυπικός αλγόριθμος καταλήγει σε μία εκ των υστέρων κατανομή όπου κυριαρχούν τα μοντέλα τριών και τεσσάρων συνιστώσων, ενώ το μοντέλο δύο συνιστωσών έρχεται τρίτο. Όπως βλέπουμε από την εκτιμηθείσα συνάρτηση πυκνότητας (διακεκομμένη γραμμή) στο Σχήμα 3.2(β), η προσαρμογή του κυρίαρχου μοντέλου ($k = 3$) είναι κακή, σε αντίθεση με το μοντέλο με μία σύνθετη συνιστώσα όπου προσαρμόζεται πολύ καλά ακόμα και στις ουρές της κατανομής.

3.5.3 Γαλαξιακό σύνολο δεδομένων

Το γαλαξιακό σύνολο δεδομένων παρουσιάστηκε για πρώτη φορά από τους Postman et al. (1986). Οι συγκεκριμένες παρατηρήσεις καταγράφουν την ταχύτητα (σε km^3/sec) 82 γαλαξιών από τις έξι κωνικές περιοχές του αστερισμού Corona Borealis. Το 1992 η Roeder ήταν η πρώτη που εφάρμοσε ένα μοντέλο μείζης σε αυτά τα δεδομένα και έκτοτε αποτελεί ένα παράδειγμα-ορόσημο στην μοντελοποίηση με μείζεις κατανομών. Έχουν χρησιμοποιηθεί από αρκετούς ερευνητές για επίδειξη μεθοδολογιών, μεταξύ των οποίων είναι οι Crawford (1994), Carlin and Chib (1995), Escobar and West (1995), Philips and Smith (1996), Richardson and Green (1997), Roeder and Wasserman (1997), Stephens (1997a, 2000a, 2000b) και Nobile (2007). Για μία συγκριτική παρουσίαση των διαφορετικών προσεγγίσεων στο συγκεκριμένο σύνολο δεδομένων παραπέμπουμε στον Aitkin (2001). Το ιστόγραμμα των δεδομένων απεικονίζεται στο Σχήμα 3.2(γ).



Σχήμα 3.6: Οι προσομοιωμένες τιμές από τον τυπικό αλγόριθμο RJMCMC για τους μέσους των συνιστωσών των γαλαξιακών δεδομένων, δεσμεύοντας στο μοντέλο με $k = 6$. Παρατηρούνται αρκετές κορυφές για τους μέσους της δεύτερης και πέμπτης συνιστώσας.

Οι περισσότεροι από τους παραπάνω ερευνητές θεωρούν ότι οι συγκεκριμένες παρατηρήσεις προέρχονται από μία μείξη k κανονικών κατανομών. Μία διαφορετική προσέγγιση ακολούθησε ο Stephens (1997a), ο οποίος χρησιμοποίησε μία μείξη κατανομών t_4 . Με μια πρώτη ματιά στο ιστόγραμμα είναι προφανή τα κενά μεταξύ του κυρίου τμήματος των παρατηρήσεων (από το 15 έως το 26 περίπου) και των ακραίων παρατηρήσεων που είναι συγκεντρωμένες κοντά στο 10 και 32 αντίστοιχα. Επομένως, είναι φανερό ότι πρόκειται για μία μείξη τουλάχιστον τριών κατανομών. Δυστυχώς, αυτή η παρατήρηση είναι το μοναδικό σημείο στο οποίο φαίνεται να συμφωνούν οι προαναφερθείσες εργασίες. Χαρακτηριστικά αναφέρουμε ότι οι εκτιμήσεις του αριθμού των συνιστωσών ποικίλλουν από 3 (Roeder and Wasserman) ή 4 (Carlin and Chib), έως 5-7 (Richardson and Green) ή 5-9 (Escobar and West). Για αυτό το γεγονός, ο Aitkin (2001), συγκρίνοντας τα αποτελέσματα των προαναφερθεισών εργασιών καταλήγει στο ότι

«Η πολυπλοκότητα των απαιτούμενων δομών για μία μπεϋζιανή ανάλυση, η συγκεχυμένη επίδραση αυτών στην πιθανοφάνεια, καθώς και τα διαφορετικά συμπεράσματα που προκύπτουν από διαφορετικές προσεγγίσεις, αφήνουν τον αναλυτή τελείως μπερδεμένο για το «τί λένε τα δεδομένα» για τον αριθμό των συνιστωσών από μία μπεϋζιανή σκοπιά.»

Ας επικεντρωθούμε αρχικά στην μέθοδο των Richardson and Green (1997), δεσμεύοντας στο μοντέλο με $k = 6$ (κανονικές) συνιστώσες που είναι η κορυφή της

Μοντέλο	$\hat{P}(c, k, \kappa_0 x)$	Μοντέλο	$\hat{P}(c, k, \kappa_0 x)$	Μοντέλο	$\hat{P}(c, k, \kappa_0 x)$
(1,2,2,1)	6.02%	(1,1,1,2,1)	2.92%	(1,1,1,1,1)	2.42%
(1,1,2,1)	3.76%	(1,2,2,1,1)	2.78%	(1,2,1,1,1)	2.31%
(1,1,2,2,1)	3.63%	(1,1,1,1)	2.72%	(1,1,2,2,1,1)	2.27%
(1,2,1,1)	3.60%	(1,1,2,1,1)	2.66%	(1,1,1)	2.15%

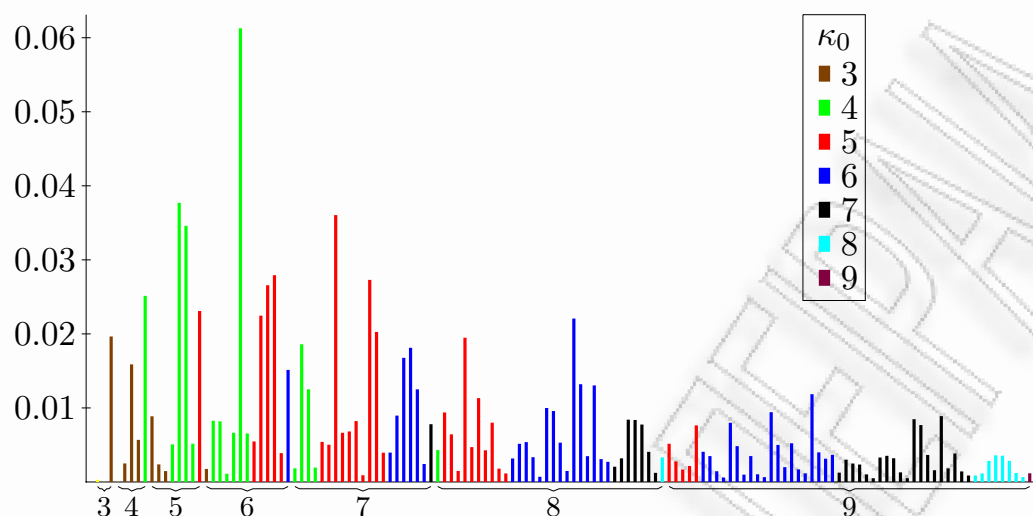
Πίνακας 3.3: Τα 12 επικρατέστερα (από τα 142) μοντέλα για το γαλαξιακό σύνολο δεδομένων, σύμφωνα με την προτεινόμενη μέθοδο.

εκτιμηθείσας εκ των υστέρων κατανομής. Στο Σχήμα 3.6 απεικονίζεται η ακολουθία των προσομοιωμένων μέσων για τις τελευταίες 5000 επαναλήψεις του Αλγορίθμου 2.5.3. Παρατηρούμε ότι, παρά την εκ των προτέρων διάταξη των μέσων, η εκ των υστέρων κατανομή παρουσιάζει αρκετές κορυφές ειδικά για την δεύτερη και πέμπτη συνιστώσα. Ενώ τα μ_1, μ_3 και μ_6 κινούνται συνεχώς γύρω από τις τιμές 10, 20 και 33, αντίστοιχα, τα μ_2, μ_4 και μ_5 εναλλάσσονται μεταξύ κάποιων τιμών. Πιο συγκεκριμένα, μεγάλο μέρος των προσομοιωμένων τιμών του μ_2 εμφανίζονται γύρω από το 16 αλλά αρκετές φορές πλησιάζουν το 10 (που είναι η περιοχή του μ_1) ή το 20 (που είναι η περιοχή του μ_3). Ένδειξη παρόμοιας συμπεριφοράς καταγράφεται και για τις προσομοιωμένες τιμές του μ_4 που εναλλάσσονται μεταξύ των περιοχών 23 και 20 όπως και για του μ_5 στις περιοχές 23, 26 και 33. Φυσικά, η εκ των υστέρων κατανομή δίνει μεγάλο βάρος στο μοντέλο έξι (ή και περισσότερων) συνιστωσών, αλλά μήπως οι πολλές παρατηρηθείσες κορυφές αποτελούν ένδειξη για την ύπαρξη σύνθετων συνιστωσών στο γαλαξιακό σύνολο δεδομένων;

Εφαρμόσαμε την προτεινόμενη μέθοδο, αρχικά για μοντέλα όπου $1 \leq k \leq 9$ στο συγκεκριμένο εύρος του k αντιστοιχούν 142 διαφορετικά ανταγωνιζόμενα μοντέλα. Εκτελέσαμε τον Αλγόριθμο 3.4.1 για 2000000 επαναλήψεις μετά από την διαγραφή των πρώτων 200000 ως περίοδο burn-in. Τα 12 επικρατέστερα μοντέλα παρουσιάζονται στον Πίνακα 3.3. Η εκτιμηθείσα εκ των υστέρων κατανομή των 142 μοντέλων απεικονίζεται στο Σχήμα 3.7. Παρατηρούμε ότι η κορυφή της εκ των υστέρων κατανομής αντιστοιχεί στο μοντέλο (1, 2, 2, 1), δηλαδή, ένα μοντέλο με έξι συνιστώσες και τέσσερις διαφορετικούς μέσους, όπου η δεύτερη και τρίτη όπως και η τέταρτη και πέμπτη συνιστώσα έχουν κοινούς μέσους:

$$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 < \mu_6.$$

Σημειώνουμε ότι το συγκεκριμένο μοντέλο παραμένει η κορυφή της εκ των υστέρων κατανομής ακόμα και αν θεωρήσουμε ένα ευρύτερο διάστημα τιμών για το k . Συγκεκριμένα εκτελέσαμε τον Αλγόριθμο 3.4.1 με $k_{max} = 25$ και η διαφορά στα αποτελέσματα ήταν αμελητέα.



Σχήμα 3.7: Η εκ των υστέρων κατανομή του (c, k, κ_0) για $1 \leq k \leq 9$. Τα μοντέλα παρουσιάζονται κατά αύξουσα διάταξη του k και του $\kappa_0|k$. Η κορυφή αντιστοιχεί στο μοντέλο $(1, 2, 2, 1)$.

Προκειμένου να εκτιμήσουμε τις εκ των υστέρων μέσες τιμές των παραμέτρων δεσμεύοντας στο μοντέλο $(1, 2, 2, 1)$, πρέπει να απαλλαγούμε από το πρόβλημα label switching εντός των σύνθετων συνιστωσών (όπως περιγράψαμε στην Ενότητα 3.5). Μετά την εφαρμογή του Αλγορίθμου Πιλοτικής αναδιάταξης στις παραμέτρους των συνιστωσών 2,3 και 4,5 καταλήγουμε στις εξής εκτιμήσεις:

$$\hat{\mathbb{E}}(\boldsymbol{\mu}|\mathbf{x}) = (9.67, 19.87, 19.87, 22.81, 22.81, 32.78)$$

$$\hat{\mathbb{E}}(\boldsymbol{\sigma}^2|\mathbf{x}) = (0.697, 0.695, 3.60, 1.37, 6.14, 2.51)$$

$$\hat{\mathbb{E}}(\mathbf{p}|\mathbf{x}) = (.091, .237, .166, .177, .280, .049)$$

Η εκτίμηση της πυκνότητας πιθανότητας δεσμεύοντας στο μοντέλο $(1, 2, 2, 1)$ απεικονίζεται στο Σχήμα 3.2(γ). Στο ίδιο σχήμα φαίνεται η αντίστοιχη εκτίμηση με την μέθοδο των Richardson and Green, δεσμεύοντας στο επικρατέστερο μοντέλο (έξι συνιστωσών) εκείνης της μεθόδου. Παρατηρούμε ότι η προσαρμογή στο ιστόγραμμα των δεδομένων είναι σαφώς καλύτερη κάτω από την προτεινόμενη μέθοδο.

Υπάρχουν κάποια ενδιαφέροντα συμπεράσματα σχετικά με το πιο πιθανό μοντέλο σύμφωνα με την μέθοδό μας. Κατ' αρχάς, ο αριθμός των συνιστωσών συμφωνεί με τον αριθμό των συνιστωσών του επικρατέστερου μοντέλου υπό την μέθοδο των Richardson and Green. Παρ' όλ' αυτά, η κυριαρχία του μοντέλου $(1, 2, 2, 1)$ συνε-

πάγεται ότι αν θεωρήσουμε ότι τα γαλαξιακά δεδομένα κατανέμονται ως μία μείξη έξι κανονικών συνιστωσών, τότε μόνο οι τέσσερις από τους έξι μέσους είναι διαφορετικοί μεταξύ τους. Από αυτήν την άποψη, το μοντέλο $(1, 2, 2, 1)$ δικαιολογεί την ύπαρξη των πολλών κορυφών που καταγράφονται στο μοντέλο έξι συνιστωσών του τυπικού αλγορίθμου. Επί πλέον, το γεγονός ότι καταλήγουμε σε ένα μοντέλο που έχει τέσσερις διαφορετικούς μέσους φαίνεται να συμφωνεί με την λύση του Stephens (1997a), η μέθοδος του οποίου υποδεικνύει ένα μοντέλο τεσσάρων κατανομών t_4 . Έτσι, συγκρίναμε το μοντέλο $(1, 2, 2, 1)$ με αυτό του Stephens. Για αυτόν τον σκοπό, εκτελέσαμε έναν αλγόριθμο RJMCMC στον χώρο αυτών των δύο μοντέλων θεωρώντας ότι εκ των προτέρων έχουν ίσες πιθανότητες. Οι λεπτομέρειες του αλγορίθμου παρατίθενται στην ενότητα B.4 του Παραρτήματος B.

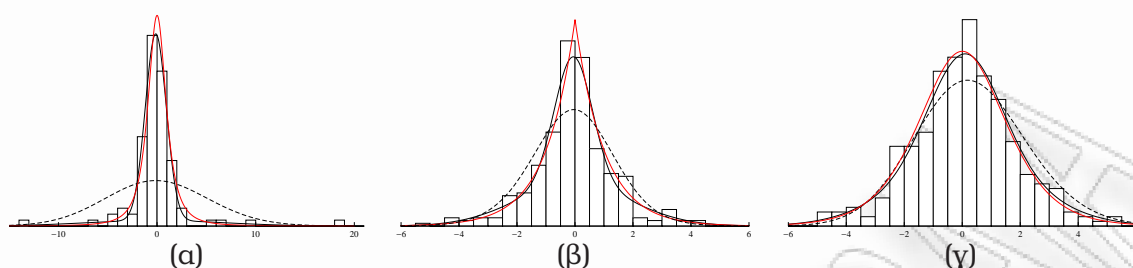
Οι εκ των υστέρων πιθανότητες των δύο μοντέλων εκτιμήθηκαν .452 και .548, αντίστοιχα. Επομένως, ο παράγοντας Bayes υπέρ του μοντέλου του Stephens εκτιμάται

$$\frac{0.548/0.452}{0.5/0.5} = 1.21,$$

τιμή που «δεν είναι άξια αναφοράς» σύμφωνα με τις κλίμακες των Jeffreys και Kass and Raftery (1995). Είναι όμως άξιο αναφοράς το γεγονός ότι ο Stephens εκτιμά τον παράγοντα Bayes υπέρ ενός μοντέλου μείξης κατανομών t_4 έναντι ενός μοντέλου κανονικών μείξεων ως 10.91 (δες Stephens, 1997a, σελ. 95). Έτσι λοιπόν, παρ' όλο που το μοντέλο $c = (1, 2, 2, 1)$ δεν προσαρμόζεται καλύτερα έναντι ενός μοντέλου μείξης τεσσάρων κατανομών t_4 , το γεγονός ότι ο προκύπτων παράγοντας Bayes δεν δίνει μεγάλη διαφορά σε αυτά τα δύο μοντέλα είναι αρκετά ενθαρρυντικό. Συνεπώς, το συγκεκριμένο αποτέλεσμα είναι ενδεικτικό της βελτιωμένης προσαρμογής που παρέχεται από την μέθοδό μας έναντι των κλασικών μοντέλων μείξεων κανονικών κατανομών. Επί πλέον, η μοντελοποίηση μέσω κατανομών t_ν παρουσιάζει γενικά τη δυσκολία της εκτίμησης των βαθμών ελευθερίας. Η μη ύπαρξη συζυγούς εκ των προτέρων κατανομής για το ν (δες Stephens, 1997a σελ. 94) σημαίνει ότι δεν είναι άμεση η εκτίμηση αυτού, ακόμα και υπό ένα μπεϋζιανό πλαίσιο.

3.5.4 Εφαρμογή σε μη κανονικά δεδομένα

Αρκετά συχνά, ένας ερευνητής καλείται να προσαρμόσει μία κατανομή στα δεδομένα τα οποία όμως αποκλίνουν αρκετά από την κανονικότητα. Όπως ήδη αναφέρθηκε στην Ενότητα 2.2.2, μια από τις πιο συνηθισμένες εφαρμογές των μείξεων κατανομών είναι η επιστράτευσή τους για την μοντελοποίηση δεδομένων που ακολουθούν κάποια άγνωστη κατανομή. Για τον σκοπό αυτό, αν θεωρηθεί άγνωστος αριθμός συ-



Σχήμα 3.8: Προσομοιωμένα δείγματα από τις κατανομές (α) t_2 , (β) $\text{Laplace}(0, 1)$, (γ) $\text{Logistic}(0, 1)$ και οι αντίστοιχες εκτιμήσεις των πυκνοτήτων σύμφωνα με τον προτεινόμενο αλγόριθμο (συνεχής μαύρη γραμμή) και τον τυπικό αλγόριθμο RJMCMC (διακεκομμένη γραμμή) μαζί με την πραγματική συνάρτηση πυκνότητας (κόκκινη γραμμή).

Δείγμα	Μοντέλο	$\hat{P}(c, k, k_0 x)$	\hat{p}	$\hat{\mu}$	$\hat{\sigma}^2$
t_2	(2)	37.53%	(0.839, 0.161)	(-0.13, -0.13)	(1.14, 48.68)
$\text{Laplace}(0, 1)$	(2)	49.95%	(0.500, 0.500)	(-0.06, -0.06)	(0.43, 3.79)
$\text{Logistic}(0, 1)$	(2)	50.20%	(0.472, 0.528)	(0.10, 0.10)	(1.48, 5.30)

Πίνακας 3.4: Εκτιμηθείσα εκ των υστέρων πιθανότητα του επικρατέστερου μοντέλου υπό τον προτεινόμενο αλγόριθμο και οι αντίστοιχες εκτιμήσεις των εκ των υστέρων μέσων τιμών των παραμέτρων αυτού για τα τρία σύνολα μη κανονικών δεδομένων.

νιστώσων, ο τυπικός αλγόριθμος RJMCMC των Richardson and Green ενδεχομένως να αποτελεί την πρώτη σκέψη του αναλυτή. Ας θεωρήσουμε για παράδειγμα τα δεδομένα των Σχημάτων 3.8(α), (β) και (γ) όπου απεικονίζονται τα ιστόγραμμα 100, 300 και 300 προσομοιωμένων τιμών από τις κατανομές t_2 , $\text{Laplace}(0, 1)$ και $\text{Logistic}(0, 1)$, αντίστοιχα. Όπως έγινε σαφές στην Ενότητα 2.2.2 (δες Σχήμα 2.3), οι κατανομές αυτές θα μπορούσαν να προσεγγιστούν από μείξεις κανονικών κατανομών.

Το γεγονός βέβαια ότι οι παραπάνω κατανομές διαφέρουν σημαντικά από την κανονική κατανομή, έχει αρκετές επιπτώσεις στον τυπικό αλγόριθμο RJMCMC των Richardson and Green. Σε όλες τις περιπτώσεις ο συγκεκριμένος αλγόριθμος επιλέγει το μοντέλο δύο συνιστώσων ως το πλέον πιθανό βάσει της εκτιμηθείσας εκ των υστέρων κατανομής. Παρ' όλ' αυτά, η προσαρμογή των μοντέλων αυτών είναι πολύ κακή συγκρίνοντας τις εκτιμηθείσες πυκνότητες (διακεκομμένες γραμμές) τόσο με το ιστόγραμμα των δεδομένων, όσο και με την πραγματική συνάρτηση πυκνότητας (κόκκινες γραμμές). Αντίθετα, ο προτεινόμενος αλγόριθμος σε όλες τις περιπτώσεις επιλέγει το μοντέλο $c = (2)$, δηλαδή το μοντέλο με δύο συνιστώσες που έχουν κοινό

μέσο. Η προσαρμογή που έχουν τα μοντέλα αυτά είναι αισθητά βελτιωμένη, καθώς οι εκτιμηθείσες πυκνότητες (μαύρες γραμμές) σχεδόν ταυτίζονται με τις αντίστοιχες πραγματικές. Σημειώνεται ότι και οι δύο αλγόριθμοι εκτελέστηκαν για 200000 επαναλήψεις (αφού προηγήθηκαν 20000 επαναλήψεις ως περίοδος burn-in). Ο Πίνακας 3.4 συγκεντρώνει τις εκτιμηθείσες εκ των υστέρων πιθανότητες του επικρατέστερου μοντέλου καθώς και τις εκτιμήσεις των εκ των υστέρων μέσων τιμών αυτού για κάθε ένα από τα τρία σύνολα δεδομένων.

3.6 Σύνοψη και συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάσαμε μία μέθοδο που απαντά στο ερώτημα σχετικά με το πότε μία μείξη κανονικών κατανομών με άγνωστο αριθμό συνιστωσών μπορεί να έχει συνιστώσες με ίδιους μέσους. Η προσέγγιση που ακολουθήθηκε είναι μία επέκταση του αλγορίθμου των Richardson and Green (1997). Έτσι έχουμε την δυνατότητα να εφαρμόσουμε πλήρη μπεϋζιανή συμπερασματολογία τόσο για την εκ των υστέρων κατανομή των μοντέλων όσο και για τις παραμέτρους της μείξης δεσμεύοντας σε συγκεκριμένα μοντέλα. Φυσικά, η μέθοδος μπορεί να εφαρμοστεί και σε περιπτώσεις όπου ο αριθμός των συνιστωσών θεωρείται γνωστός.

Ένα κρίσιμο σημείο για τον προσδιορισμό του ιεραρχικού μοντέλου είναι η εκ των προτέρων κατανομή του (c, k, κ_0) . Υιοθετήσαμε την προσέγγιση που δίνει την ίδια εκ των προτέρων πιθανότητα σε κάθε τιμή του k και σε κάθε μοντέλο $(c, \kappa_0) | k$, δηλαδή μία ομοιόμορφη κατανομή για τον αριθμό των συνιστωσών και για τα μοντέλα εντός κάθε αριθμού συνιστωσών. Μία άλλη επιλογή που δοκιμάστηκε επίσης ήταν η ομοιόμορφη κατανομή για όλα τα μοντέλα (c, k, κ_0) . Αυτό γίνεται πολύ απλά θέτοντας $f(k) = N_k/N, \forall k = 1, \dots, K$, ενώ η εκ των προτέρων κατανομή στο $(c, \kappa_0) | k$ παραμένει ίδια. Αυτή η επιλογή υποστηρίζει την πλήρως μη πληροφοριακή προσέγγιση σε προβλήματα επιλογής μοντέλου, καθώς κάθε μοντέλο έχει την ίδια εκ των προτέρων πιθανότητα. Ένα μειονέκτημα αυτής της υπόθεσης είναι το γεγονός ότι η εκ των προτέρων πιθανότητα του αριθμού των συνιστωσών k είναι αύξουσα συνάρτηση του k . Έτσι, παρ' όλο που σε αυτήν την περίπτωση δεν είμαστε πληροφοριακοί για την τριάδα (c, k, κ_0) , είμαστε πληροφοριακοί για τον αριθμό των συνιστωσών. Κάτι τέτοιο όμως είναι πρόβλημα στην περίπτωση όπου ο σκοπός της ανάλυσης είναι η συμπερασματολογία για την περιθωριακή κατανομή του αριθμού συνιστωσών. Πράγματι, με αυτήν την εκ των προτέρων κατανομή, ο αλγόριθμος επισκέπτεται πιο συχνά μοντέλα με μεγαλύτερες τιμές του k από ό,τι πριν. Παρ' όλ' αυτά, παρατηρήθηκε ότι η κορυφή της εκ των υστέρων κατανομής των μοντέλων (c, k, κ_0) παραμένει

ίδια στην συντριπτική πλειοψηφία των παραδειγμάτων που αναλύσαμε. Πιο συγκεκριμένα, η κορυφή έμεινε ίδια τόσο στα δύο προσομοιωμένα σύνολα δεδομένων που αναλύθηκαν στις Ενότητες 3.5.1 και 3.5.2, όσο και στα γαλαξιακά δεδομένα.

Για το γαλαξιακό σύνολο δεδομένων, συμπεράναμε ότι υπάρχει ένας σχετικά μικρός αριθμός ομάδων για τους μέσους των συνιστωσών (αυτό συμφωνεί με τις εργασίες των Roeder and Wasserman και των Carlin and Chib) ενώ ταυτόχρονα υπάρχει ένας μεγαλύτερος αριθμός ομάδων για τις διασπορές των συνιστωσών (και αυτό συμφωνεί με τις εργασίες των Richardson and Green και των Escobar and West). Υπό τη συγκεκριμένη οπτική, η προσέγγισή μας μπορεί να θεωρηθεί ως ένας συμβιβασμός των διαφορετικών αποτελεσμάτων που υπάρχουν στη βιβλιογραφία, καθώς μπορεί να μοντελοποιήσει τα δεδομένα μέσω ενός σχετικά μικρού πλήθους ετερογενών ομάδων με ίδιους μέσους. Επί πλέον, η μέθοδός μας μπορεί να συσχετιστεί με την προσέγγιση των τεσσάρων κατανομών t_4 του Stephens (1997a). Η βασική διαφορά είναι ότι οι κατανομές t είναι συνεχείς παρά διακριτές μείξεις κανονικών κατανομών (όπως δηλαδή είναι οι σύνθετες συνιστώσες της προσέγγισής μας). Εν τούτοις, ένα πλεονέκτημα της μεθόδου μας έναντι αυτής του Stephens είναι το γεγονός ότι δεν χρειάζεται να εκτιμήσουμε επί πλέον παράμετρους όπως οι βαθμοί ελευθερίας των κατανομών t .

Συνοψίζοντας κάποια εμπειρικά συμπεράσματα που παρατηρήθηκαν συστηματικά τόσο στα προσομοιωμένα σύνολα δεδομένων που παρουσιάσαμε στις Ενότητες 3.5.1 και 3.5.2, όπως και σε αρκετά ακόμα που θεωρήσαμε, έχουμε τα εξής:

- Όταν οι συνιστώσες της μείξης είναι επαρκώς διαχωρισμένες, τότε η εκ των υστέρων κατανομή του αριθμού των διαφορετικών μέσων k_0 υπό την μέθοδό μας και του αριθμού των συνιστωσών k υπό την μέθοδο των Richardson and Green, δίνουν σχεδόν τα ίδια αποτελέσματα. Συγκεκριμένα, όταν η τυπική μέθοδος επιλέγει τον πραγματικό αριθμό των συνιστωσών, τότε το ίδιο ισχύει για τον αριθμό των διαφορετικών μέσων υπό την προτεινόμενη μέθοδο. Το αντίστροφο δεν ισχύει απαραίτητα.
- Το προηγούμενο δεν ισχύει όταν υπάρχει τουλάχιστον μία σύνθετη συνιστώσα ή όταν οι μέσοι δύο γειτονικών συνιστωσών είναι κοντά μεταξύ τους. Τότε η κορυφή της εκ των υστέρων κατανομής υπό την μέθοδό μας αντιστοιχεί είτε στο σωστό μοντέλο (όπως στο πρώτο προσομοιωμένο σύνολο δεδομένων) είτε σε ένα μοντέλο με τον σωστό αριθμό συνιστωσών (όπως στο δεύτερο προσομοιωμένο σύνολο δεδομένων). Από την άλλη πλευρά, η τυπική μέθοδος υπερεκτιμά το k όταν οι διασπορές των δύο συνιστωσών διαφέρουν αρκετά μεταξύ τους, ενώ το

υποεκτιμά στην αντίθετη περίπτωση.

- Όταν η εκ των υστέρων κατανομή παρουσιάζει κύριες ή μικρότερες κορυφές κοντά στο όριο του παραμετρικού χώρου που δημιουργείται από τον περιορισμό διάταξης, η μέθοδος μας έχει δύο πλεονεκτήματα έναντι της τυπικής. Κατ' αρχάς, αυτές οι κορυφές εντοπίζονται πιο γρήγορα και επιπροσθέτως εξερευνώνται επαρκώς λόγω της νέας κίνησης διαχωρισμού-συνδυασμού των μέσων. Δεύτερον, οι αντίστοιχες εκτιμήσεις είναι αρκετά κοντά στις πραγματικές τιμές των παραμέτρων σε αντίθεση με την τυπική μέθοδο όπου οι εκτιμήσεις επηρεάζονται από τον περιορισμό του παραμετρικού χώρου και αυτό έχει αποτέλεσμα υπερεκτιμήσεις και υποεκτιμήσεις.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Κεφάλαιο 4

Επίλυση του προβλήματος Label Switching

4.1 Εισαγωγή

Στην Ενότητα 2.6 είδαμε ότι το πρόβλημα label switching παίζει σημαντικό ρόλο στην Μπεϋζιανή Συμπερασματολογία για μοντέλα μείξεων κατανομών. Οι υπάρχουσες μέθοδοι επίλυσης του προβλήματος κρίνονται ως ανεπαρκείς, υπό την έννοια ότι καμμία εξ αυτών δεν είναι απλή στην εφαρμογή και αποδοτική ταυτόχρονα. Έτσι κρίνεται απαραίτητη η ανάπτυξη μιας μεθόδου που θα λύνει επιτυχώς το πρόβλημα σε γενικά πλαίσια ενώ θα απαιτεί συγχρόνως μικρό υπολογιστικό κόστος.

Σε αυτό το κεφάλαιο εισάγουμε την θεωρητική θεμελίωση μιας νέας μεθόδου για την αντιμετώπιση του προβλήματος label switching, θεωρώντας τον χώρο \mathcal{Z} των εικονικών μεταβλητών κατάταξης. Η μέθοδος αναπτύσσεται σε ένα ευρύ πλαίσιο μοντέλων ελλিপών δεδομένων, γενικότερο από τα μοντέλα μείξεων κατανομών. Αρχικά, αποδεικνύεται ότι αν ένα μοντέλο ελλিপών δεδομένων πληροί κάποιες συγκεκριμένες ιδιότητες, τότε τόσο η πιθανοφάνεια των παρατηρηθέντων δεδομένων όσο και η εκ των υστέρων κατανομή των παραμέτρων αυτού είναι συμμετρική. Ως εκ τούτου, το φαινόμενο label switching θα εμφανίζεται στην εκ των υστέρων κατανομή. Έτσι, τα αποτελέσματα που παρουσιάζονται μπορούν να εφαρμοστούν σε οποιοδήποτε μοντέλο ανήκει στην συγκεκριμένη κατηγορία.

Αρχικά, μελετάται ο χώρος των διανυσμάτων κατάταξης \mathcal{Z} και ορίζεται μία σχέση ισοδυναμίας μεταξύ των στοιχείων του. Ακολούθως, το \mathcal{Z} διαμερίζεται στις προκύπτουσες κλάσεις ισοδυναμίας και αποδεικνύεται ότι υπάρχουν συγκεκριμένα υποσύνολα αυτού που οδηγούν σε κλάσεις μη συμμετρικών κατανομών οι οποίες έχουν το ίδιο στήριγμα με την συμμετρική εκ των υστέρων κατανομή των παραμέτρων. Σύμφωνα με το βασικό θεώρημα που αποδεικνύεται, αυτές οι μη συμμετρικές κατανομές έχουν την ιδιότητα ότι αναπαράγουν την συμμετρική εκ των υστέρων κατανομή απλώς μεταθέτοντας τους δείκτες των συνιστωσών.

Κατόπιν προτείνουμε έναν αλγόριθμο αναδιάταξης ενός δείγματος από την συμμετρική εκ των υστέρων κατανομή, που έχει ως στόχο τη νέα κατανομή που είναι απαλλαγμένη από το φαινόμενο label switching. Αφού αποδείξουμε ότι ο συγκεκριμένος αλγόριθμος όντως συγκλίνει στην επιθυμητή κατανομή-στόχο, στην συνέχεια προτείνεται ένας συγκεκριμένος τρόπος κατασκευής του υποσυνόλου που αναπαράγει την εκ των υστέρων κατανομή ο οποίος είναι αρκετά αποδοτικός και απλός στην εφαρμογή του. Επί πλέον, μελετώνται κάποιες θεωρητικές ιδιότητες σχετικά με την ταχύτητα σύγκλισης του προτεινόμενου αλγορίθμου στην νέα κατανομή-στόχο. Σύμφωνα με αυτές, η ταχύτητα σύγκλισης (ως προς τη νόρμα ολικής μεταβολής) της αναδιατεταγμένης ακολουθίας είναι τουλάχιστον ίση με την ταχύτητα σύγκλισης της αρχικής. Το κεφάλαιο ολοκληρώνεται με την προσαρμογή ενός βήματος του προτεινόμενου αλγορίθμου στην κατηγορία προβλημάτων εκχώρησης (assignment problem) του ακέραιου προγραμματισμού.

4.2 Ο αλγόριθμος ECR

Σε αυτήν την ενότητα περιγράφεται το θεωρητικό υπόβαθρο που αποτελεί το κίνητρο για την ανάπτυξη του αλγορίθμου «Equivalence Classes Representatives» (ECR) ως μία μέθοδος αντιμετώπισης του προβλήματος label switching. Το πλαίσιο που θα ακολουθηθεί θα είναι λίγο πιο γενικό από αυτό που είδαμε μέχρι τώρα ώστε να είναι άμεση η εφαρμογή των αποτελεσμάτων και σε πιο σύνθετα μοντέλα πέραν των μείξεων κατανομών, όπως για παράδειγμα τα HMM. Για αυτόν τον λόγο, θα χρησιμοποιήσουμε τον συμβολισμό $\eta = (\eta_1, \dots, \eta_k)$ για τις παραμέτρους των k συνιστώσων του μοντέλου. Για παράδειγμα, στα κλασσικά μοντέλα μείξεων κατανομών έχουμε ότι $\eta_j = (p_j, \theta_j)$, $j = 1, \dots, k$. Το z εξακολουθεί να συμβολίζει το διάνυσμα των εικονικών μεταβλητών κατάταξης των n παρατηρήσεων x .

Έστω ότι η πυκνότητα των παρατηρηθέντων δεδομένων, x , προκύπτει από τα μη παρατηρηθέντα πλήρη δεδομένα $z \in \mathcal{Z} = \{1, \dots, k\}^n$ μέσω περιθωριοποίησης, δηλαδή

$$f(x|\eta) = \sum_{z \in \mathcal{Z}} f(x, z|\eta). \quad (4.1)$$

Σε ό,τι ακολουθεί βασιζόμαστε σε μεταθέσεις στον χώρο των διανυσμάτων κατάταξης.

Ορισμός 4.2.1. Έστω $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$ μία μετάθεση του συνόλου των δεικτών $\{1, \dots, k\}$. Η αντίστοιχη αναδιάταξη τz του διανύσματος κατάταξης $z = (z_1, \dots, z_n) \in \mathcal{Z} := \{1, \dots, k\}^n$ είναι $\eta \tau z = (t_{z_1}, \dots, t_{z_n}) \in \mathcal{Z}$.

Παράδειγμα 4.2.1. Αναδιατάξεις διανυσμάτων κατάταξης.

Έστω $n = 10$, $k = 3$ και τα διανύσματα κατάταξης

$$z_1 = (2, 2, 1, 1, 2, 2, 3, 3, 3, 1) \quad \text{και} \quad z_2 = (2, 2, 2, 2, 2, 2, 2, 2, 2, 2).$$

Σύμφωνα με τον Ορισμό 4.2.1, η μετάθεση $\tau = (2, 1, 3)$ μετατρέπει τα παραπάνω διανύσματα κατάταξης στα $\tau z_1 = (1, 1, 2, 2, 1, 1, 3, 3, 3, 2)$ και $\tau z_2 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$. Σε αυτό το σημείο ας παρατηρήσουμε ότι αν $\tau' = (3, 1, 2)$ τότε $\tau z_2 = \tau' z_2$.

Επίσης σημειώνουμε ότι

$$\begin{aligned} \tau'(\tau z_1) &= (3, 1, 2)(1, 1, 2, 2, 1, 1, 3, 3, 3, 2) \\ &= (3, 3, 1, 1, 3, 3, 2, 2, 2, 1). \end{aligned}$$

Όμως $\tau\tau' = (2, 1, 3)(3, 1, 2) = (1, 3, 2)$, οπότε

$$\begin{aligned} (\tau\tau')z_1 &= (1, 3, 2)(2, 2, 1, 1, 2, 2, 3, 3, 3, 1) \\ &= (3, 3, 1, 1, 3, 3, 2, 2, 2, 1). \end{aligned}$$

Παρατηρούμε ότι $\tau'(\tau z_1) = (\tau\tau')z_1$.

Στη συνέχεια θα θεωρηθεί ένα ευρύ πλαίσιο μοντέλων ελλিপών δεδομένων, όπου θα υποθέσουμε την ισχύ των παρακάτω αρκετά γενικών ιδιοτήτων:

- Ο παραμετρικός χώρος \mathcal{H} είναι αναλλοίωτος ως προς τις μεταθέσεις, δηλαδή

$$\tau\mathcal{H} = \mathcal{H}, \quad \forall \tau \in \mathcal{T}_k, \quad (4.2)$$

με την έννοια ότι $\eta \in \mathcal{H} \Leftrightarrow \tau\eta \in \mathcal{H}, \forall \tau \in \mathcal{T}_k$.

- Η εκ των προτέρων κατανομή των παραμέτρων των συνιστωσών είναι συμμετρική, δηλαδή

$$f(\eta) = f(\tau\eta), \quad \forall \tau \in \mathcal{T}_k. \quad (4.3)$$

- Η πυκνότητα των πλήρων δεδομένων (\mathbf{x}, \mathbf{z}) έχει την ιδιότητα

$$f(\mathbf{x}, \tau\mathbf{z}|\eta) = f(\mathbf{x}, \mathbf{z}|\tau\eta), \quad \forall \tau \in \mathcal{T}_k. \quad (4.4)$$

Η ισχύς της (4.4) θα εξακριβωθεί τόσο για τα κλασσικά μοντέλα μείξεων όσο και για τα γενικότερα μοντέλα HMM.

Πρόταση 4.2.1. Κάθε μοντέλο ελιπιτών δεδομένων (4.1) που πληροί τις ιδιότητες (4.2), (4.3) και (4.4), έχει συμμετρική πιθανοφάνεια και συμμετρική εκ των υστέρων κατανομή.

Απόδειξη. Από την (4.2) εξασφαλίζεται ότι το σύνολο των δυνατών τιμών των παραμέτρων είναι ίδιο για όλες τις συνιστώσες. Εξ άλλου, λόγω της (4.4), προκύπτει ότι η παρατηρηθείσα πιθανοφάνεια είναι συμμετρική, καθώς $\forall \tau \in \mathcal{T}_k$ έχουμε

$$\begin{aligned} f(\mathbf{x}|\tau\eta) &= \sum_{z \in \mathcal{Z}} f(\mathbf{x}, z|\tau\eta) \\ &= \sum_{z \in \mathcal{Z}} f(\mathbf{x}, \tau z|\eta) \\ &= \sum_{\xi \in \mathcal{Z}} f(\mathbf{x}, \xi|\eta) \\ &= f(\mathbf{x}|\eta), \end{aligned} \quad (4.5)$$

για κάθε $\eta \in \mathcal{H}$. Τέλος, η εκ των υστέρων κατανομή του η είναι συμμετρική διότι, $\forall \tau \in \mathcal{T}_k$

$$\begin{aligned} f(\tau\eta|\mathbf{x}) &= \frac{f(\mathbf{x}|\tau\eta)f(\tau\eta)}{f(\mathbf{x})} \\ &= \frac{f(\mathbf{x}|\eta)f(\eta)}{f(\mathbf{x})} \\ &= f(\eta|\mathbf{x}), \end{aligned}$$

για κάθε $\eta \in \mathcal{H}$, λόγω των (4.5) και (4.3). \square

Από τα παραπάνω έπεται ότι το φαινόμενο label switching εμφανίζεται στην εκ των υστέρων κατανομή οποιουδήποτε μοντέλου πληροί τις προϋποθέσεις της Πρότασης 4.2.1. Σημειώνουμε ότι η πυκνότητα των δεδομένων, $f(\mathbf{x}|\dots)$, ενδέχεται να εξαρτάται και από κάποιες (παρατηρήσιμες) ερμηνευτικές μεταβλητές (π.χ. συμμεταβλητές). Ο λόγος που δεν τις συμπεριλαμβάνουμε στον συμβολισμό μας είναι ότι οι συγκεκριμένες μεταβλητές δεν εξαρτώνται από τους δείκτες των συνιστωσών $\{1, \dots, k\}$, συνεπώς δεν επηρεάζουν σε τίποτα την ανάλυση που ακολουθεί.

4.2.1 Η εκ των υστέρων κατανομή ως ισοβαρής μείξη μη συμμετρικών κατανομών

Από το Θέωρημα Ολικής Πιθανότητας, η εκ των υστέρων κατανομή των παραμέτρων η μπορεί να γραφεί ως

$$f(\eta|\mathbf{x}) = \sum_{z \in \mathcal{Z}} w(z|\mathbf{x})f(\eta|\mathbf{x}, z), \quad (4.6)$$

όπου με $w(z|\mathbf{x})$ συμβολίζουμε το εκ των υστέρων βάρος του διανύσματος κατάταξης \mathbf{z} , ενώ με $f(\boldsymbol{\eta}|\mathbf{x}, z)$ συμβολίζουμε την εκ των υστέρων κατανομή του $\boldsymbol{\eta}$ δοθέντος του (\mathbf{x}, z) (δες επίσης Marin et al., 2005).

Για τα αποτελέσματα που ακολουθούν, θα φανεί χρήσιμο το επόμενο λήμμα.

Λήμμα 4.2.1. Η εκ των υστέρων κατανομή του $(z, \boldsymbol{\eta})$ ικανοποιεί την σχέση

$$f(\tau z, \boldsymbol{\eta}|\mathbf{x}) = f(z, \tau \boldsymbol{\eta}|\mathbf{x}), \quad \forall \tau \in \mathcal{T}_k. \quad (4.7)$$

Επί πλέον, για κάθε $z \in \mathcal{Z}$

$$w(\tau z|\mathbf{x}) = w(z|\mathbf{x}), \quad \forall \tau \in \mathcal{T}_k. \quad (4.8)$$

Για την δεσμευμένη κατανομή του $\boldsymbol{\eta}$ δοθέντος (\mathbf{x}, z) ισχύει ότι

$$f(\boldsymbol{\eta}|\mathbf{x}, \tau z) = f(\tau \boldsymbol{\eta}|\mathbf{x}, z), \quad \forall \tau \in \mathcal{T}_k. \quad (4.9)$$

Απόδειξη. Σε κάθε περίπτωση μπορούμε να γράψουμε

$$f(z, \boldsymbol{\eta}|\mathbf{x}) = \frac{f(\mathbf{x}, z|\boldsymbol{\eta})f(\boldsymbol{\eta})}{f(\mathbf{x})}.$$

Συνεπώς, για κάθε $\tau \in \mathcal{T}_k$ παίρνουμε ότι

$$f(\tau z, \boldsymbol{\eta}|\mathbf{x}) = \frac{f(\mathbf{x}, \tau z|\boldsymbol{\eta})f(\boldsymbol{\eta})}{f(\mathbf{x})}. \quad (4.10)$$

Αντικαθιστώντας τις (4.4) και (4.3) στην (4.10) λαμβάνουμε ότι

$$f(\tau z, \boldsymbol{\eta}|\mathbf{x}) = f(z, \tau \boldsymbol{\eta}|\mathbf{x})$$

και ολοκληρώνεται η απόδειξη της (4.7). Για το δεύτερο σκέλος του λήμματος, λόγω της (4.7) έχουμε για κάθε $\tau \in \mathcal{T}_k$ ότι

$$w(\tau z|\mathbf{x}) = \int_{\mathcal{H}} f(\tau z, \boldsymbol{\eta})d\boldsymbol{\eta} = \int_{\mathcal{H}} f(z, \tau \boldsymbol{\eta})d\boldsymbol{\eta}.$$

Λόγω της (4.2), με την αλλαγή μεταβλητής $\tau \boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}$ η τελευταία σχέση γράφεται

$$\int_{\tau \mathcal{H}} f(z, \tilde{\boldsymbol{\eta}})d\tilde{\boldsymbol{\eta}} = \int_{\mathcal{H}} f(z, \tilde{\boldsymbol{\eta}})d\tilde{\boldsymbol{\eta}} = w(z|\mathbf{x})$$

επειδή η απόλυτη τιμή Ιακωβιανής του τελευταίου μετασχηματισμού ισούται με την μονάδα και τούτο ολοκληρώνει την απόδειξη της (4.8). Τέλος, η (4.9) προκύπτει άμεσα από την (4.7) και το γεγονός ότι τα $f(z, \boldsymbol{\eta}|\mathbf{x}) = w(z|\mathbf{x})f(\boldsymbol{\eta}|\mathbf{x}, z)$ και $w(z|\mathbf{x})$ είναι αναλλοίωτα ως προς τις μεταθέσεις των δεικτών των συνιστωσών. \square

Είναι εύκολο να δούμε ότι ο Ορισμός 4.2.1 οδηγεί σε μία σχέση ισοδυναμίας στον χώρο \mathcal{Z} των διανυσμάτων κατάταξης.

Ορισμός 4.2.2. Δύο διανύσματα κατάταξης $z_1, z_2 \in \mathcal{Z}$ θα καλούνται *ισοδύναμα* αν υπάρχει $\tau \in \mathcal{T}_k$ τέτοιο ώστε $z_1 = \tau z_2$. Σε αυτήν την περίπτωση θα γράφουμε $z_1 \sim z_2$.

Προφανώς, η σχέση \sim αποτελεί μία σχέση ισοδυναμίας, καθώς είναι

- αυτοπαθητική: $z \sim z$, για κάθε $z \in \mathcal{Z}$
- συμμετρική: αν $z_1 \sim z_2$ τότε $z_2 \sim z_1$, για κάθε $z_1, z_2 \in \mathcal{Z}$
- μεταβατική: αν $z_1 \sim z_2$ και $z_2 \sim z_3$ τότε $z_1 \sim z_3$, για κάθε $z_1, z_2, z_3 \in \mathcal{Z}$.

Ας συμβολίσουμε με $\Xi_z = \{\xi \in \mathcal{Z} : \xi \sim z\}$ την κλάση ισοδυναμίας του z . Χρησιμοποιώντας ένα επιχείρημα εγκλεισμού–αποκλεισμού, μπορεί ναδειχτεί η επόμενη πρόταση σχετικά με το πλήθος των στοιχείων του συνόλου ηλικίου \mathcal{Z}/\sim .

Πρόταση 4.2.2. Για δεδομένα k και n ο συνολικός αριθμός των κλάσεων ισοδυναμίας ισούται με

$$\sum_{i=1}^k \frac{i^n}{i!} \sum_{j=0}^{k-i} \frac{(-1)^j}{j!}. \quad (4.11)$$

Απόδειξη. Για την απόδειξη της (4.11) παραπέμπουμε στην ενότητα Γ.1 του Παραρτήματος Γ. \square

Ορισμός 4.2.3. Οποιοδήποτε υποσύνολο $\mathcal{Z}_0 \subset \mathcal{Z}$, με τις ιδιότητες

- $\bigcup_{z \in \mathcal{Z}_0} \Xi_z = \mathcal{Z}$ και
- $z_1 \not\sim z_2, \forall z_1, z_2 \in \mathcal{Z}_0$,

θα καλείται *σύνολο αντιπροσώπων των κλάσεων ισοδυναμίας*.

Από τον Ορισμό 4.2.3 είναι φανερό ότι ένα σύνολο αντιπροσώπων των κλάσεων ισοδυναμίας περιέχει ένα και μόνο ένα στοιχείο από κάθε κλάση. Για τη συνέχεια θα φανεί χρήσιμο το επόμενο λήμμα.

Λήμμα 4.2.2. Για κάθε δυνατή συλλογή αντιπροσώπων \mathcal{Z}_0 και για κάθε συνάρτηση $g(z)$ ισχύει

$$\sum_{z \in \mathcal{Z}} g(z) = \sum_{z \in \mathcal{Z}_0} \sum_{\tau \in \mathcal{T}_k} \frac{g(\tau z)}{(k - k_0(z))!}, \quad (4.12)$$

όπου το $k_0(z)$ ισούται με το πλήθος των διαφορετικών συμβόλων που εμφανίζονται στο $z \in \mathcal{Z}$.

Απόδειξη. Για κάθε $z \in \mathcal{Z}$, έστω η απεικόνιση $\tau z : \mathcal{T}_k \mapsto \Xi_z$, η οποία είναι επί του Ξ_z . Κάθε στοιχείο του Ξ_z είναι εικόνα του z για τόσα $\tau \in \mathcal{T}_k$ όσο και το πλήθος των μεταθέσεων $\tau \in \mathcal{T}_k$ που δίνουν το ίδιο τz . Το πλήθος αυτών των μεταθέσεων ισούται με $(k - k_0(z))!$. Έτσι, ενώ συνολικά γίνονται $k!$ απεικονίσεις, κάθε στοιχείο του πεδίου τιμών αντιστοιχεί σε $(k - k_0(z))!$ στοιχεία του πεδίου ορισμού. Αυτό σημαίνει ότι κάθε κλάση Ξ_z περιέχει $k!/(k - k_0(z))!$ το πλήθος στοιχεία. Συνεπώς, μετά την αναδιάταξη των k^n όρων, για κάθε Z_0 έχουμε διαδοχικά ότι

$$\begin{aligned} \sum_{z \in \mathcal{Z}} g(z) &= \sum_{z \in Z_0} \sum_{\xi \in \Xi_z} g(\xi) \\ &= \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \frac{g(\tau z)}{(k - k_0(z))!} \end{aligned}$$

και η απόδειξη ολοκληρώθηκε. \square

Έστω τώρα

$$f_{Z_0}(\eta|\mathbf{x}) := \sum_{z \in Z_0} \frac{k! w(z|\mathbf{x})}{(k - k_0(z))!} f(\eta|\mathbf{x}, z). \quad (4.13)$$

Επειδή, για όλα τα δυνατά σύνολα αντιπροσώπων Z_0 , έχουμε ότι

$$\begin{aligned} \sum_{z \in Z_0} \frac{k! w(z|\mathbf{x})}{(k - k_0(z))!} &= \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \frac{w(\tau z|\mathbf{x})}{(k - k_0(z))!} \\ &= \sum_{z \in Z_0} \sum_{z^* \in \Xi_z} w(z^*|\mathbf{x}) \\ &= \sum_{z \in \mathcal{Z}} w(z|\mathbf{x}) \\ &= 1, \end{aligned}$$

η $f_{Z_0}(\eta|\mathbf{x})$ είναι μία συνάρτηση πυκνότητας πιθανότητας. Προφανώς, η $f_{Z_0}(\eta|\mathbf{x})$ έχει το ίδιο στήριγμα με την εκ των υστέρων κατανομή $f(\eta|\mathbf{x})$, καθώς είναι μία μείξη των δεσμευμένων κατανομών $f(\eta|\mathbf{x}, z)$, για $z \in Z_0$. Παρατηρούμε επίσης ότι η f_{Z_0} είναι μη συμμετρική για οποιαδήποτε επιλογή του συνόλου των αντιπροσώπων Z_0 . Επί πλέον, ισχύει το ακόλουθο:

Θεώρημα 4.2.1. Η εκ των υστέρων κατανομή του η μπορεί να αναπαρασταθεί ως

$$f(\eta|\mathbf{x}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{Z_0}(\tau\eta|\mathbf{x}) \quad (4.14)$$

για οποιαδήποτε επιλογή του συνόλου Z_0 των αντιπροσώπων των κλάσεων ισοδυναμίας.

Απόδειξη. Για κάθε \mathcal{Z}_0 , η εκ των υστέρων κατανομή του η στην (4.6) γράφεται

$$f(\eta|\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}_0} \sum_{\tau \in \mathcal{T}_k} \frac{w(\tau\mathbf{z}|\mathbf{x})}{(k - k_0(\tau\mathbf{z}))!} f(\eta|\mathbf{x}, \tau\mathbf{z}),$$

λόγω του Λήμματος 4.2.2. Αλλά από τις (4.8), (4.9) και χρησιμοποιώντας την προφανή ιδιότητα $k_0(\mathbf{z}) = k_0(\tau\mathbf{z})$, για κάθε $\tau \in \mathcal{T}_k$, $\mathbf{z} \in \mathcal{Z}$, η τελευταία σχέση γράφεται

$$\begin{aligned} f(\eta|\mathbf{x}) &= \sum_{\mathbf{z} \in \mathcal{Z}_0} \frac{w(\mathbf{z}|\mathbf{x})}{(k - k_0(\mathbf{z}))!} \sum_{\tau \in \mathcal{T}_k} f(\tau\eta|\mathbf{x}, \mathbf{z}) \\ &= \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} k! \sum_{\mathbf{z} \in \mathcal{Z}_0} \frac{w(\mathbf{z}|\mathbf{x})}{(k - k_0(\mathbf{z}))!} f(\tau\eta|\mathbf{x}, \mathbf{z}) \\ &= \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0}(\tau\eta|\mathbf{x}) \end{aligned}$$

και τούτο ολοκληρώνει την απόδειξη. \square

Σύμφωνα με το Θεώρημα 4.2.1, η εκ των υστέρων κατανομή του η γράφεται ως ισοβαρής μείξη μη συμμετρικών κατανομών. Αυτή η αναπαράσταση έχει μία ενδιαφέρουσα εξήγηση ως προς το φαινόμενο label switching. Ένα δείγμα από την κατανομή $f_{\mathcal{Z}_0}(\eta|\mathbf{x})$ μπορεί να μετασχηματιστεί σε ένα δείγμα από την εκ των υστέρων κατανομή $f(\eta|\mathbf{x})$ απλώς μεταθέτοντας τυχαία τα παραμετρικά διανύσματα, όπως δείχνει η σχέση (4.14). Η αντίστροφη κατεύθυνση είναι πιο ενδιαφέρουσα και υποδεικνύει το πώς μπορούμε να επιλύσουμε το φαινόμενο label switching. Δεδομένου ότι διαθέτουμε ένα δείγμα από την συμμετρική εκ των υστέρων κατανομή, είναι δυνατόν να παραγάγουμε ένα δείγμα από την $f_{\mathcal{Z}_0}(\eta|\mathbf{x})$ σύμφωνα με το πλαίσιο που περιγράφουμε στην συνέχεια, αρκεί το δείγμα να είναι διευρυμένο με τις μεταβλητές κατάταξης: Ας υποθέσουμε ότι στην i -οστή επανάληψη του αρχικού δειγματολήπτη προσομοιώνεται $(\mathbf{z}^{(i)}, \eta^{(i)})$ με $\mathbf{z}^{(i)} \in \mathcal{Z}$ και η μετάθεση $\tau_i \in \mathcal{T}_k$ είναι τέτοια ώστε $\tau_i \mathbf{z}^{(i)} \in \mathcal{Z}_0$. Προφανώς, ισχύει ότι $\mathbf{z}^{(i)} = \tau_i^{-1} \tau_i \mathbf{z}^{(i)}$. Τότε, σύμφωνα με το Λήμμα 4.2.1,

$$f(\mathbf{z}^{(i)}, \eta^{(i)}|\mathbf{x}) = f(\tau_i \mathbf{z}^{(i)}, \tau_i^{-1} \eta^{(i)}|\mathbf{x}). \quad (4.15)$$

Η τελευταία σχέση αποτελεί το κλειδί ώστε να μεταβούμε σε ένα δείγμα από την $f_{\mathcal{Z}_0}(\eta|\mathbf{x})$, ενώ διαθέτουμε ένα (διευρυμένο) δείγμα από την από κοινού εκ των υστέρων κατανομή των (\mathbf{z}, η) , όπως αποδεικνύεται στο επόμενο λήμμα.

Λήμμα 4.2.3. Έστω $(\mathbf{z}, \eta) \sim f(\mathbf{z}, \eta|\mathbf{x})$ και, δοθέντος (\mathbf{z}, η) , η μετάθεση τ είναι ομοιόμορφα κατανεμημένη στο σύνολο

$$\mathcal{T}_{\mathbf{z}} := \{\tau \in \mathcal{T}_k : \tau\mathbf{z} \in \mathcal{Z}_0\}, \quad \mathbf{z} \in \mathcal{Z}. \quad (4.16)$$

Τότε,

$$\tau^{-1}\boldsymbol{\eta} \sim f_{\mathcal{Z}_0}(\boldsymbol{\eta}|\mathbf{x}).$$

Απόδειξη. Ας παρατηρήσουμε αρχικά ότι η μετάθεση τ εξαρτάται μόνο από το \mathbf{z} και ότι υπάρχουν ακριβώς $(k - k_0(\mathbf{z}))!$ δυνατές μεταθέσεις που αν εφαρμοστούν στο \mathbf{z} το μεταφέρουν στο \mathcal{Z}_0 . Έτσι, η από κοινού συνάρτηση πυκνότητας πιθανότητας των $\tau, \mathbf{z}, \boldsymbol{\eta}$ είναι

$$\begin{aligned} f(\tau, \mathbf{z}, \boldsymbol{\eta}|\mathbf{x}) &= f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{z})f(\mathbf{z}|\mathbf{x})f(\tau|\mathbf{z}) \\ &= \frac{w(\mathbf{z}|\mathbf{x})}{(k - k_0(\mathbf{z}))!} f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{z}) I(\tau \in \mathcal{T}_{\mathbf{z}}). \end{aligned} \quad (4.17)$$

Έστω $(\tau^*, \mathbf{z}^*, \boldsymbol{\eta}^*) = (\tau, \tau\mathbf{z}, \tau^{-1}\boldsymbol{\eta})$. Τότε, για κάθε $\tau \in \mathcal{T}_k$, $\mathbf{u} \in \mathcal{Z}_0$ και για οποιοδήποτε μετρήσιμο υποσύνολο C του χώρου \mathcal{H} του $\boldsymbol{\eta}$ έχουμε ότι

$$\begin{aligned} \mathbf{P}(\tau^* = t, \mathbf{z}^* = \mathbf{u}, \boldsymbol{\eta}^* \in C|\mathbf{x}) &= \mathbf{P}(\tau = t, \tau\mathbf{z} = \mathbf{u}, \tau^{-1}\boldsymbol{\eta} \in C|\mathbf{x}) \\ &= \mathbf{P}(\tau = t, \mathbf{z} = \tau^{-1}\mathbf{u}, \boldsymbol{\eta} \in \tau C|\mathbf{x}) \\ &= \frac{w(t^{-1}\mathbf{u}|\mathbf{x})}{(k - k_0(t^{-1}\mathbf{u}))!} \int_{\tau C} f(\boldsymbol{\eta}|\mathbf{x}, t^{-1}\mathbf{u}) d\boldsymbol{\eta}, \end{aligned} \quad (4.18)$$

αφού $t \in T_{t^{-1}\mathbf{u}}$ για κάθε $t \in \mathcal{T}_k$ και $\mathbf{u} \in \mathcal{Z}_0$. Αλλά λόγω της (4.8) ισχύει ότι $w(t^{-1}\mathbf{u}|\mathbf{x}) = w(\mathbf{u}|\mathbf{x})$ και $k_0(t^{-1}\mathbf{u}) = k_0(\mathbf{u})$. Επομένως, χρησιμοποιώντας την (4.9), η (4.18) γράφεται

$$\frac{w(\mathbf{u}|\mathbf{x})}{(k - k_0(\mathbf{u}))!} \int_{\tau C} f(\boldsymbol{\eta}|\mathbf{x}, t^{-1}\mathbf{u}) d\boldsymbol{\eta} = \frac{w(\mathbf{u}|\mathbf{x})}{(k - k_0(\mathbf{u}))!} \int_C f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{u}) d\boldsymbol{\eta}.$$

Συνεπώς, η συνάρτηση πυκνότητας του $(\tau^*, \mathbf{z}^*, \boldsymbol{\eta}^*)$ (ως προς το κατάλληλο μέτρο) είναι

$$f^*(t, \mathbf{u}, \boldsymbol{\eta}|\mathbf{x}) = \frac{w(\mathbf{u}|\mathbf{x})}{(k - k_0(\mathbf{u}))!} f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{u}), \quad t \in \mathcal{T}_k, \mathbf{u} \in \mathcal{Z}_0, \boldsymbol{\eta} \in \mathcal{H}, \quad (4.19)$$

ενώ η περιθωριακή κατανομή του $\boldsymbol{\eta}^* = \tau^{-1}\boldsymbol{\eta}$ είναι

$$\begin{aligned} f^*(\boldsymbol{\eta}|\mathbf{x}) &= \sum_{t \in \mathcal{T}_k} \sum_{\mathbf{u} \in \mathcal{Z}_0} \frac{w(\mathbf{u}|\mathbf{x})}{(k - k_0(\mathbf{u}))!} f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{u}) \\ &= k! \sum_{\mathbf{u} \in \mathcal{Z}_0} \frac{w(\mathbf{u}|\mathbf{x})}{(k - k_0(\mathbf{u}))!} f(\boldsymbol{\eta}|\mathbf{x}, \mathbf{u}) \end{aligned}$$

δηλαδή η $f_{\mathcal{Z}_0}(\boldsymbol{\eta}|\mathbf{x})$. □

Το παρακάτω αποτέλεσμα εξασφαλίζει την ισχύ της μεθόδου μας κατά την εφαρμογή της σε ένα προσομοιωμένο δείγμα από έναν έναν αλγόριθμο MCMC.

Πρόταση 4.2.3. Έστω η μαρκοβιανή αλυσίδα $(z^{(t)}, \eta^{(t)})$, $t = 1, 2, \dots$ με οριακή κατανομή $f(z, \eta|x)$. Αν για κάθε t , η μετάθεση τ_t κατανέμεται ομοιόμορφα στο σύνολο $\mathcal{T}_{z^{(t)}}$, τότε η οριακή κατανομή της ακολουθίας $\tau_t^{-1}\eta^{(t)}$ είναι η $f_{Z_0}(\eta|x)$.

Απόδειξη. Είναι σαφές ότι η επαυξημένη ακολουθία $(\tau_t, z^{(t)}, \eta^{(t)})$ είναι μία μαρκοβιανή αλυσίδα με οριακή κατανομή την $f(\tau, z, \eta|x)$ που ορίστηκε στην (4.17). Τώρα, αφού το $(\tau^*, z^*, \eta^*)^{(t)}$ είναι ένας αντιστρέψιμος μετασχηματισμός του $(\tau_t, z^{(t)}, \eta^{(t)})$, έχουμε ότι και η προκύπτουσα ακολουθία θα είναι επίσης μία μαρκοβιανή αλυσίδα, με οριακή κατανομή την $f^*(\tau^*, z^*, \eta^*|x)$ που ορίσαμε στην (4.19). Το αποτέλεσμα προκύπτει παίρνοντας την περιθωριακή κατανομή του η^* (δοθέντος x). \square

Σύμφωνα με την Πρόταση 4.2.3, η οριακή κατανομή της ακολουθίας $\eta^{(t)}$, $t = 1, \dots, M$, που κατασκευάζεται στο βήμα 2 του αλγόριθμου που ακολουθεί, είναι η f_{Z_0} .

Αλγόριθμος 4.2.1. Αλγόριθμος Equivalence Classes Representatives (ECR)

1. Καθόρισε το σύνολο αντιπροσώπων Z_0 και προσομοίωσε ένα δείγμα $(z, \eta)^{(t)}$, $t = 1, \dots, M$, με κατανομή-στόχο $f(z, \eta|x)$.
 2. Για $t = 1, \dots, M$:
 - (α) Επίλεξε τυχαία μία μετάθεση τ_t από το σύνολο $\mathcal{T}_{z^{(t)}}$ που ορίστηκε στην (4.16).
 - (β) Θέσε $(z', \eta')^{(t)} = (\tau_t z^{(t)}, \tau_t^{-1} \eta^{(t)})$.
 3. Προσέγγισε τις περιθωριακές εκ των υστέρων κατανομές των (η_1, \dots, η_k) και των διανυσμάτων κατάταξης z χρησιμοποιώντας το αναδιατεταγμένο δείγμα $\eta'^{(t)}$ και $z'^{(t)}$, $t = 1, \dots, M$, αντίστοιχα.
-

Από τον ορισμό του συνόλου Z_0 , η f_{Z_0} είναι μία μη συμμετρική κατανομή και μπορεί να αναπαραγάγει ένα δείγμα από την συμμετρική εκ των υστέρων κατανομή μεταθέτοντας απλώς τους δείκτες των συνιστωσών. Έτσι, η f_{Z_0} μπορεί να χρησιμοποιηθεί για να λύσει το πρόβλημα label switching. Με βάση την ανάλυση που προηγήθηκε, αυτό επιτυγχάνεται μέσω του Αλγορίθμου ECR. Επί πλέον, το στήριγμα της κατανομής-στόχου του Αλγορίθμου ECR είναι ίδιο με αυτό της συμμετρικής εκ των υστέρων κατανομής. Αυτό δεν ισχύει για τις περιορισμένες εκ των υστέρων κατανομές που προκύπτουν ως λύσεις μέσω του ΔΠΜ της Frühwirth-Schnatter (2001), ή τον Αλγόριθμο Πιλοτικής αναδιάταξης των Marin et al., (2005).

Μέγεθος δείγματος n	Αριθμός συνιστωσών k					
	2	3	4	5	6	7
20	5.72	8.76	10.66	11.90	12.71	13.21
40	11.74	18.31	22.70	25.88	28.27	30.10
60	17.76	27.85	34.74	39.86	43.83	47.00
80	23.78	37.39	46.78	53.84	59.39	63.91
100	29.80	46.93	58.83	67.82	74.96	80.81
120	35.82	56.48	70.87	82.80	90.52	97.71
140	41.84	66.02	82.91	95.78	106.08	114.61
160	47.86	75.56	94.94	109.76	121.65	131.51
180	53.88	85.10	106.99	123.74	137.21	148.42
200	59.91	94.65	119.03	137.72	152.77	165.32

Πίνακας 4.1: Δεκαδικός λογάριθμος του πλήθους κλάσεων ισοδυναμίας για διάφορες τιμές του n και k .

4.2.2 Επιλογή του συνόλου αντιπροσώπων

Για την εφαρμογή του Αλγορίθμου ECR πρέπει να καθοριστεί ένα σύνολο αντιπροσώπων Z_0 . Στον Πίνακα 4.1 φαίνεται ο δεκαδικός λογάριθμος του πλήθους των κλάσεων ισοδυναμίας για ενδεικτικές τιμές των n και k , σύμφωνα με την (4.11). Βλέπουμε ότι ο αριθμός των κλάσεων είναι πολύ μεγάλος και επομένως καταλαβαίνουμε ότι οι δυνατοί τρόποι επιλογής ενός αντιπροσώπων από τις κλάσεις είναι ακόμα περισσότεροι. Γίνεται σαφές λοιπόν ότι η διαδικασία επιλογής ενός συνόλου αντιπροσώπων πρέπει να γίνει προσεκτικά. Σε αυτή την παράγραφο δίνονται απαντήσεις σχετικά με τις διαφορές δύο διαφορετικών συνόλων αντιπροσώπων, πότε ένα τέτοιο σύνολο θα θεωρείται «καλό» και φυσικά πώς μπορεί να εφαρμοστεί πρακτικά μία διαδικασία επιλογής αντιπροσώπων.

Όπως είδαμε στην (4.13), η κατανομή f_{Z_0} του αναδιατεταγμένου δείγματος η είναι μία μείξη των δεσμευμένων κατανομών $f(\eta|\mathbf{x}, z)$, $z \in Z_0$. Σε αυτό το σημείο παρατηρήσουμε ότι τα βάρη της μείξης αυτής δεν επηρεάζονται από την επιλογή των αντιπροσώπων, καθώς η ποσότητα $w(\tau z|\mathbf{x})/(k - k_0(\tau z))!$ είναι σταθερή ως προς $\tau \in \mathcal{T}_k$. Συνεπώς, οι κατανομές f_{Z_0} και $f_{Z'_0}$ που αντιστοιχούν σε δύο διαφορετικά σύνολα αντιπροσώπων Z_0 και Z'_0 , διαφέρουν μόνον ως προς τις συνιστώσες που τις απαρτίζουν, δηλαδή τις $f(\eta|\mathbf{x}, z)$. Επί πλέον, αν $Z_0 \neq \tau Z'_0$ για κάθε $\tau \in \mathcal{T}_k$, τότε η f_{Z_0} και $f_{Z'_0}$ θα έχουν τελειώς διαφορετικές μορφές και η πρώτη δεν μπορεί να προκύψει από την δεύτερη μεταθέτοντας απλώς τους δείκτες του συνόλου $\{1, \dots, k\}$.

Μία αποδοτική αντιμετώπιση του προβλήματος label switching δεν σημαίνει απλώς την εξάλειψη της συμμετρίας της εκ των υστέρων κατανομής, αλλά επί πλέον,

την εξουδετέρωση όσο το δυνατόν περισσότερο των συμμετρικών περιοχών της (εκτός μιας φυσικά). Είναι σαφές λοιπόν ότι δεν μπορούμε να αρκεστούμε μόνο στο ότι η κατανομή f_{Z_0} είναι μη συμμετρική για όλα τα δυνατά Z_0 . η αυθαίρετη επιλογή του αντιπροσώπου κάθε κλάσης ισοδυναμίας είναι πολύ πιθανόν να οδηγήσει σε μία κατανομή όπου θα διατηρείται σε κάποιον βαθμό η επίδραση των συμμετρικών περιοχών. Για να αποφύγουμε αυτό το σενάριο, ο αντιπρόσωπος κάθε κλάσης θα πρέπει να επιλεγεί με τέτοιο τρόπο έτσι ώστε όλες οι συνιστώσες που απαρτίζουν την μείξη f_{Z_0} να έχουν το μεγαλύτερο μέρος της μάζας τους συγκεντρωμένο σε μία από τις $k!$ συμμετρικές περιοχές. Είναι σαφές ότι οι περιοχές στις οποίες βρίσκεται το μεγαλύτερο μέρος της μάζας των συνιστωσών $f(\eta|x, z)$ εξαρτάται μόνον από το διάνυσμα z . Σαφώς λοιπόν, αν οι αντιπρόσωποι z είναι «όμοιοι» μεταξύ τους τότε οι αντίστοιχες κατανομές $f(\eta|x, z)$ θα επικεντρώνουν τις μάζες τους στην ίδια περιοχή και όχι στις υπόλοιπες συμμετρικές. Ένας απλός τρόπος να μετρήσουμε την ομοιότητα μεταξύ δύο διανυσμάτων κατάταξης είναι ο αριθμός των παρατηρήσεων που έχουν την ίδια κατάταξη:

Ορισμός 4.2.4. Το μέτρο ομοιότητας S μεταξύ δύο διανυσμάτων κατάταξης $z_1 = (z_{11}, \dots, z_{1n})$ και $z_2 = (z_{21}, \dots, z_{2n})$ ορίζεται ως

$$S(z_1, z_2) := \sum_{i=1}^n I(z_{1i} = z_{2i}),$$

όπου $I(A)$ είναι η δείκτη συνάρτηση του A .

Προφανώς ισχύει $0 \leq S(z_1, z_2) \leq n$. Επί πλέον, παρατηρούμε ότι η ποσότητα $n - S(z_1, z_2)$ είναι η απόσταση απλής συμφωνίας (simple matching distance) μεταξύ των z_1 και z_2 . Παρ' ότι θα μπορούσαμε να υιοθετήσουμε πιο σύνθετα μέτρα ομοιότητας, όπως θα δούμε η ομοιότητα S αποδίδει εξαιρετικά για το πρόβλημά μας.

Ίσως να είναι διαισθητικά λογικό ότι για την κατασκευή του Z_0 όλοι οι αντιπρόσωποι (και συνεπώς οι αντίστοιχες συνιστώσες f_{Z_0}) θα πρέπει να επιλεγούν έτσι ώστε να είναι όσο το δυνατόν περισσότερο όμοιοι μεταξύ τους. Παρ' όλ' αυτά, εμείς ενδιαφερόμαστε για την επιλογή μιας εκ των συμμετρικών περιοχών παρά για την συνολική ομοιότητα των συνιστωσών. Αυτό μπορεί να γίνει επιλέγοντας αρχικά ένα κατάλληλο $z^* \in Z$, και αφού περιληφθεί πρώτο στο Z_0 , στην συνέχεια να επιλέξουμε όλους τους υπόλοιπους αντιπροσώπους έτσι ώστε να είναι όσο το δυνατόν περισσότερο όμοιοι σε αυτό.

Υπό την παραπάνω οπτική, η διαδικασία προσδιορισμού του συνόλου αντιπροσώπων Z_0 καταλήγει στην εύρεση ενός «καλού» διανύσματος κατάταξης z^* το οποίο θα

χρησιμοποιηθεί ως οδηγός για την επιλογή των υπολοίπων αντιπροσώπων. Ένας τέτοιος καθορισμός του Z_0 μπορεί να θεωρηθεί ότι δανείζεται ιδέες από τον αλγόριθμο Πιλοτικής Αναδιάταξης των Marin et al. (2005) (δες Ενότητα 2.6.2), με την διαφορά ότι αναφέρεται στο σύνολο Z και όχι στον παραμετρικό χώρο. Ας υποθέσουμε λοιπόν ότι θέλουμε να επιλέξουμε τον αντιπρόσωπο τz της κλάσης Ξ_z . Συμβολίζουμε με

$$\tau(\Xi_z) = \arg \max_{\tau \in \mathcal{T}_k} S(\tau z, z^*)$$

την μετάθεση εκείνη που μεγιστοποιεί την ομοιότητα S μεταξύ των z και z^* . Έτσι, στην συλλογή αντιπροσώπων Z_0 εισέρχεται το στοιχείο $\tau(\Xi_z)z$. Στην περίπτωση όπου υπάρχουν παραπάνω από μία τέτοιες μεταθέσεις, τ_1, τ_2, \dots , τότε διατάσσουμε λεξικογραφικά τα $\tau_1 z, \tau_2 z, \dots$ και θέτουμε ως $\tau(\Xi_z)$ την μετάθεση που αντιστοιχεί στον πρώτο όρο. Αυτό γίνεται έτσι ώστε να γνωρίζουμε κάθε φορά ποιος ακριβώς αντιπρόσωπος έχει συμπεριληφθεί στο Z_0 . Επομένως, το σύνολο αντιπροσώπων θα είναι το

$$Z_0 = \cup \{ \tau(\Xi_z)z; z \in Z \}.$$

Παρατηρούμε την διαφορά μεταξύ των δυνατών $\tau(\Xi_z)$ και των \mathcal{T}_z που οριστήκαν στην προηγούμενη ενότητα: Τα $\tau(\Xi_z)$ ορίζουν το Z_0 ενώ τα σύνολα \mathcal{T}_z έχουν νόημα μετά τον ορισμό του Z_0 . Από την άλλη πλευρά, το \mathcal{T}_z είναι προφανώς υποσύνολο του συνόλου των $\tau(\Xi_z)$.

4.2.3 Μία αναλυτική εφαρμογή της προτεινόμενης μεθόδου

Προσομοιώσαμε $n = 5$ παρατηρήσεις $\mathbf{x} = (6, 12, 9, 4, 6)$ από μία μείξη δύο ($k = 2$) κατανομών Poisson με γνωστά και ίσα βάρη,

$$0.5\mathcal{P}(\theta_1) + 0.5\mathcal{P}(\theta_2),$$

όπου οι πραγματικές τιμές των παραμέτρων ήταν $\boldsymbol{\theta} = (\theta_1, \theta_2) = (5, 7)$. Σύμφωνα με τους συμβολισμούς μας, σε αυτήν την περίπτωση έχουμε $\boldsymbol{\eta} = \boldsymbol{\theta}$ και $\mathcal{H} = (0, \infty)^2$. Αρχικά, ας δούμε ότι η (4.2) ισχύει διότι εδώ έχουμε $\mathcal{T}_k = \{(1, 2), (2, 1)\}$ και προφανώς $\tau((0, \infty)^2) = (0, \infty)^2$, για κάθε $\tau \in \mathcal{T}_k$. Υποθέτουμε ότι τα θ_1, θ_2 είναι εκ των προτέρων ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κατανομή $\mathcal{G}(\alpha, \beta)$, δηλαδή, μία κατανομή γάμμα με μέση τιμή α/β . Θέσαμε $\alpha = 1.2$ και $\beta = 0.2$. Λόγω του μικρού μεγέθους δείγματος και των προηγούμενων υποθέσεων, είναι εύκολο να προσδιορίσουμε αναλυτικά την (συμμετρική) εκ των υστέρων κατανομή του $(\theta_1, \theta_2) | \mathbf{x}$, η οποία αναπαρίσταται στο Σχήμα 4.1(α). Αυτό μπορεί να γίνει χρησιμοποιώντας την αναπαράσταση της εκ των υστέρων κατανομής που δίνεται στην (4.6),

αφού προηγουμένως έχουμε υπολογίσει τα εκ των υστέρων βάρη,

$$\begin{aligned}
w(\mathbf{z}|\mathbf{x}) &= \int_{(0,\infty)^2} f(\mathbf{z}, \boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\
&\propto \int_{(0,\infty)^2} f(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}) f(\boldsymbol{\theta}) f(\mathbf{z}) d\boldsymbol{\theta} \\
&= \int_{(0,\infty)^2} \prod_{j=1}^2 \frac{e^{-n_j(\mathbf{z})\theta_j} \theta_j^{s_j(\mathbf{z})}}{\prod_{i:z_i=j} x_i!} \frac{\beta^\alpha \theta_j^{\alpha-1} e^{-\beta\theta_j}}{\Gamma(\alpha)} 0.5^{n_j(\mathbf{z})} d\boldsymbol{\theta} \\
&= \frac{\beta^{2\alpha} 0.5^n}{\Gamma(\alpha)^2 \prod_{i=1}^n x_i!} \prod_{j=1}^2 \int_0^{+\infty} \theta_j^{\alpha+s_j(\mathbf{z})-1} e^{-(\beta+n_j(\mathbf{z}))\theta_j} d\theta_j \\
&\propto \prod_{j=1}^2 \frac{\Gamma(\alpha + s_j(\mathbf{z}))}{(\beta + n_j(\mathbf{z}))^{\alpha+s_j(\mathbf{z})}},
\end{aligned}$$

όπου τα n_j και s_j , για $j = 1, 2$ ορίζονται όπως στην (2.29). Σημειώνουμε ότι η ακριβής έκφραση των εκ των υστέρων βαρών προκύπτει κανονικοποιώντας τα βάρη που μόλις υπολογίσαμε, συνεπώς

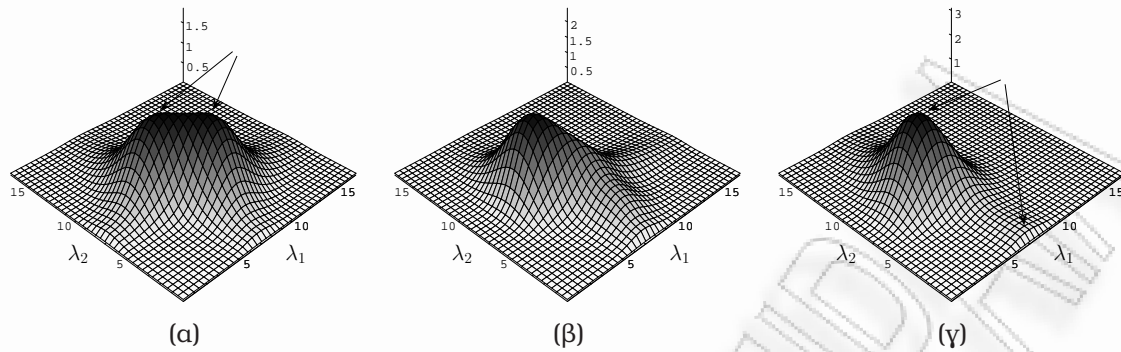
$$w(\mathbf{z}|\mathbf{x}) = \frac{\prod_{j=1}^2 \frac{\Gamma(\alpha+s_j(\mathbf{z}))}{(\beta+n_j(\mathbf{z}))^{\alpha+s_j(\mathbf{z})}}}{\sum_{\boldsymbol{\xi} \in \mathcal{Z}} \prod_{j=1}^2 \frac{\Gamma(\alpha+s_j(\boldsymbol{\xi}))}{(\beta+n_j(\boldsymbol{\xi}))^{\alpha+s_j(\boldsymbol{\xi})}}}, \quad \mathbf{z} \in \mathcal{Z}.$$

Από την παραπάνω ανάλυση είναι σαφές επίσης ότι

$$\theta_j | \mathbf{z}, \mathbf{x} \sim \mathcal{G}(\alpha + s_j(\mathbf{z}), \beta + n_j(\mathbf{z})), \quad j = 1, 2.$$

Έτσι η εκ των υστέρων κατανομή του $(\theta_1, \theta_2) | \mathbf{x}$ είναι μία μείξη $2^5 = 32$ κατανομών, όπου κάθε συνιστώσα αποτελείται από δύο ανεξάρτητες κατανομές γάμμα, με τα βάρη που υπολογίσαμε. Χρησιμοποιώντας το Mathematica βρήκαμε αριθμητικά ότι οι δύο συμμετρικές κορυφές της κατανομής αυτής είναι οι (7.75, 6.01) και (6.01, 7.75) (σημειώνονται στο Σχήμα 4.1(a) με βέλη).

Στρεφόμενοι τώρα στην προτεινόμενη μέθοδο, έχουμε ότι κάθε κλάση ισοδυναμίας αποτελείται από 2 στοιχεία, συνεπώς οποιοδήποτε σύνολο αντιπροσώπων θα περιέχει 16 διανύσματα κατάταξης· ένα από κάθε κλάση ισοδυναμίας. Χρησιμοποιώντας λοιπόν 16 από τα παραπάνω εκ των υστέρων βάρη και δεσμευμένες κατανομές, είναι εύκολο να βρούμε την αναλυτική έκφραση της κατανομής $f_{\mathcal{Z}_0}$, για κάποιο προκαθορισμένο σύνολο αντιπροσώπων \mathcal{Z}_0 , μέσω της (4.13). Στα Σχήματα 4.1(β) και 4.1(γ) φαίνονται οι (μη-συμμετρικές) κατανομές $f_{\mathcal{Z}_0}(\boldsymbol{\theta}|\mathbf{x})$ που αντιστοιχούν σε δύο διαφορετικές επιλογές του συνόλου αντιπροσώπων \mathcal{Z}_0 . Στην πρώτη περίπτωση, οι αντιπρόσωποι των κλάσεων επιλέχτηκαν τυχαία ενώ στην δεύτερη κάθε αντιπρόσωπος είναι το μέλος κάθε κλάσης το οποίο μεγιστοποιεί το μέτρο ομοιότητας S με το



Σχήμα 4.1: (α) Η συμμετρική εκ των υστέρων κατανομή του $(\theta_1, \theta_2)|\mathbf{x}$. (β,γ) Οι κατανομές $f_{z_0}(\boldsymbol{\theta}|\mathbf{x})$ που αντιστοιχούν σε τυχαία επιλεγμένους αντιπροσώπους και μεγιστοποιώντας το μέτρο ομοιότητας S με το $\mathbf{z}^* = (1, 2, 2, 1, 1)$, αντίστοιχα. (Όλες οι πυκνότητες αναφέρονται στην ίδια πολλαπλασιαστική σταθερά.)

διάνυσμα κατάταξης $\mathbf{z}^* = (1, 2, 2, 1, 1)$ που είναι εκείνο το διάνυσμα που αντιστοιχεί στο μέγιστο της $f(\mathbf{z}, \boldsymbol{\theta}|\mathbf{x})$. Είναι σαφές ότι ενώ και στις δύο περιπτώσεις προκύπτουν μη συμμετρικές κατανομές, θα πρέπει να προτιμηθεί η δεύτερη καθώς η επίδραση της συμμετρικής περιοχής έχει εξουδετερωθεί πλήρως, σε αντίθεση με την πρώτη περίπτωση όπου βλέπουμε να διατηρείται ένα αρκετά μεγάλο μέρος αυτής.

Γενικά, οι κορυφές που φαίνονται σε μια μείξη κατανομών μπορεί να είναι αρκετά μακριά από τις κορυφές των επιμέρους συνιστωσών που την απαρτίζουν. Όντως, παρατηρούμε στο Σχήμα 4.1(γ) ότι η κορυφή της f_{z_0} βρίσκεται στο σημείο $(5.35, 8.36)$. Επί πλέον, αν είμαστε λίγο πιο προσεκτικοί θα παρατηρήσουμε ότι η f_{z_0} έχει και μία μικρότερη κορυφή στο σημείο $(7.15, 1.00)$ η οποία όμως δεν είναι εμφανής στην συμμετρική εκ των υστέρων κατανομή. Η ύπαρξή της οφείλεται στο γεγονός ότι η εκ των υστέρων πιθανότητα μιας κενής συνιστώσας είναι αρκετά μεγάλη: Συγκεκριμένα, με αναλυτικό υπολογισμό λαμβάνουμε $w(1, 1, 1, 1, 1|\mathbf{x}) = w(2, 2, 2, 2, 2|\mathbf{x}) \approx .0394$. Η θέση της δεύτερης κορυφής επίσης δικαιολογείται πλήρως από το γεγονός ότι η δεύτερη συντεταγμένη της αντιστοιχεί στην ουσία στην κορυφή της εκ των προτέρων κατανομής. Αυτή είναι που καθορίζει πλήρως την συμπεριφορά της δεύτερης συντεταγμένης καθώς είναι κενή και δεν εξαρτάται από τα δεδομένα.

Το γεγονός ότι η μέθοδός μας καταφέρνει και απομονώνει τόσο την κύρια κορυφή όσο και το ότι αποκαλύπτει και μικρότερες κορυφές της εκ των υστέρων κατανομής, είναι αρκετά ενθαρρυντικό. Γενικά, οι μικρότερες κορυφές στην ουσία αποκαλύπτουν λιγότερο πιθανές (αλλά παρ' όλ' αυτά πιθανές) ερμηνείες των δεδομένων. Στο συγκεκριμένο παράδειγμα λοιπόν, κάτι τέτοιο μας βοηθά να συμπεράνουμε εκ των υστέρων ότι τα δεδομένα μας προέρχονται από μία μείξη δύο κατανομών Poisson

με πιθανότητα 92.12%, ενώ με πιθανότητα 7.88% αποτελούν ένα τυχαίο δείγμα από μία κατανομή Poisson (εξαιτίας της κενής συνιστώσας). Αυτό το συμπέρασμα δεν θα ήταν τόσο άμεσο βλέποντας την συμμετρική εκ των υστέρων κατανομή όπου οι (δύο) μικρότερες (και συμμετρικές) κορυφές είναι κρυμμένες κάτω από την επίδραση των δύο κυρίαρχων κορυφών. Τέλος, τονίζουμε ότι σε αυτό το παράδειγμα οι συμμετρικές περιοχές της εκ των υστέρων κατανομής είναι αρκετά κοντά μεταξύ τους καθώς δεν απέχουν αρκετά από την διαγώνιο του τεταρτημορίου που είναι συγκεντρωμένη η μάζα τους. Προφανώς λοιπόν, σε τέτοιου είδους περιπτώσεις διαπιστώνεται η αδυναμία απομόνωσης των συμμετρικών περιοχών μέσω ενός περιορισμού διάκρισης. Για παράδειγμα, αν υιοθετήσουμε τον (μοναδικό δυνατό) περιορισμό που υποδεικνύει το πρόβλημα, δηλαδή τον $\theta_1 < \theta_2$, τότε είναι εύκολο να φανταστούμε ότι η κατανομή που θα προκύψει θα οδηγήσει σε μεροληπτικά συμπεράσματα, καθώς θα αγνοηθεί η ουρά της μίας συμμετρικής περιοχής η οποία βρίσκεται πάνω στο όριο του περιορισμού.

4.2.4 Αναδιατάσσοντας ένα προσομοιωμένο δείγμα

Είναι σαφές ότι η εκ των υστέρων κατανομή $f(\mathbf{z}, \boldsymbol{\eta} | \mathbf{x})$ έχει πολλαπλές κορυφές. Πράγματι, η σχέση (4.15) υποδεικνύει ότι αν το $(\mathbf{z}^{(\text{MAP})}, \boldsymbol{\eta}^{(\text{MAP})}) \equiv (\mathbf{z}, \boldsymbol{\eta})^{(\text{MAP})}$ είναι μία κορυφή, τότε το ίδιο θα ισχύει και για κάθε μετάθεση αυτής, $(\tau \mathbf{z}^{(\text{MAP})}, \tau^{-1} \boldsymbol{\eta}^{(\text{MAP})})$. Για τον ορισμό του οδηγού \mathbf{z}^* , είναι λογικό ότι μία πολύ καλή επιλογή είναι το $\mathbf{z}^{(\text{MAP})}$. Αυτό δικαιολογείται από την σχέση (4.6) και το γεγονός ότι, γενικά, πολύ λίγα διανύσματα κατάταξης έχουν μη αμελητέα εκ των υστέρων πιθανότητα συγκρινόμενα με το συνολικό πλήθος τους (δες Casella et al., 2004). Παρ' όλ' αυτά, το υπολογιστικό κόστος για τον προσδιορισμό των κορυφών μέσω αναλυτικών μεθόδων αυξάνει πολύ γρήγορα σε σχέση με το μέγεθος του δείγματος, οπότε, ακόμα και στην περίπτωση που είναι δυνατή μία τέτοια προσέγγιση, αυτόματα καθίσταται πρακτικά απαγορευτική η εφαρμογή της.

Έστω το προσομοιωμένο δείγμα $(\mathbf{z}, \boldsymbol{\eta})^{(i)}$, $i = 1, \dots, M$, με κατανομή-στόχο $f(\mathbf{z}, \boldsymbol{\eta} | \mathbf{x})$. Ας συμβολίσουμε επίσης με $\hat{\mathbf{z}}^{(\text{MAP})}$ το διάνυσμα κατάταξης που αντιστοιχεί στην προσέγγιση Monte Carlo του MAP εκτιμητή

$$(\hat{\mathbf{z}}, \hat{\boldsymbol{\eta}})^{(\text{MAP})} = \arg \max_{1 \leq i \leq M} f((\mathbf{z}, \boldsymbol{\eta})^{(i)} | \mathbf{x}).$$

Ο $(\hat{\mathbf{z}}, \hat{\boldsymbol{\eta}})^{(\text{MAP})}$ είναι ένας συνεπής εκτιμητής μιας κορυφής της $f(\mathbf{z}, \boldsymbol{\eta} | \mathbf{x})$. Επομένως, καθώς αυξάνει το μέγεθος M του προσομοιωμένου δείγματος, έχουμε ότι το $S(\hat{\mathbf{z}}^{(\text{MAP})}, \tau \mathbf{z}^{(\text{MAP})})$ θα συγκλίνει στο n για κάποιο $\tau \in \mathcal{T}_k$. Βεβαίως, η ομοιότητα S

μεταξύ των $\hat{z}^{(\text{MAP})}$ και $\tau z^{(\text{MAP})}$ (για κάποιο $\tau \in \mathcal{T}_k$) αργά ή γρήγορα θα πάρει την τιμή n , λόγω του ότι ο χώρος των κατατάξεων είναι διακριτός. Σε αυτό το σημείο όμως υπενθυμίζουμε ότι το μόνο που χρειαζόμαστε για την εφαρμογή της μεθόδου μας είναι ένας καλός οδηγός z^* και όχι απαραίτητα «ο καλύτερος». Άρα, το $\hat{z}^{(\text{MAP})}$ αντιστοιχεί σε μία καλή επιλογή επίσης.

Πρέπει να τονίσουμε ότι επιλέγοντας ως οδηγό το $\hat{z}^{(\text{MAP})}$, το σενάριο ύπαρξης δύο (ή και περισσότερων) διαφορετικών στοιχείων μιας κλάσης που μεγιστοποιούν την ομοιότητα S (κάτι που συνεπάγεται την επιλογή του αντιπροσώπου μέσω λεξικογραφικής διάταξης αυτών) είναι εξαιρετικά σπάνιο. Συγκεκριμένα, μία τέτοια συμπεριφορά δεν παρατηρήθηκε σε κανένα από τα παραδείγματα που εφαρμόσαμε τον αλγόριθμο. Σημειώνουμε επίσης ότι με οδηγό το $\hat{z}^{(\text{MAP})}$ κάθε $z \in \mathcal{Z}$ ισχύει

$$\mathcal{T}_z = \{\tau = \arg \max_{\tau \in \mathcal{T}_k} S(\tau z, \hat{z}^{(\text{MAP})})\}.$$

Τέλος, πρέπει να αναφέρουμε ότι κάποιες διαφορετικές καλές επιλογές για τον οδηγό, εκτός της εκτίμησης MAP, είναι η πιο πιθανή κατάταξη ή η κατάταξη που αντιστοιχεί στο μέγιστο της πλήρους πιθανοφάνειας. Επί πλέον, εφ' όσον το \mathcal{Z}_0 μπορεί να είναι οποιοδήποτε σύνολο αντιπροσώπων, θα μπορούσε να χρησιμοποιηθεί ως οδηγός και οποιοδήποτε διάνυσμα κατάταξης έχει εμφανιστεί συχνά στο προσομοιωμένο δείγμα, αρκεί το πλήθος n των παρατηρήσεων να μην είναι πολύ μικρό. Πράγματι, παρατηρήσαμε ότι τέτοιου είδους διαφορετικές επιλογές του οδηγού οδήγησαν σε σχεδόν ταυτόσημα αποτελέσματα. Τονίζουμε το γεγονός ότι αποφύγαμε πιο πολύπλοκες μεθόδους για το βήμα 1 του αλγορίθμου μας, διότι η επιλογή ενός οδηγού z^* δεν παρουσιάζει τα μειονεκτήματα του Αλγορίθμου Πιλοτικής Αναδιάταξης. Εφ' όσον ασχολούμαστε με τον χώρο των εικονικών μεταβλητών κατάταξης παρά τον ίδιο τον παραμετρικό χώρο, εκμεταλλευόμαστε την διακριτή φύση του και τον μικρό αριθμό των κατατάξεων με μη αμελητέο εκ των υστέρων βάρος. Αυτό σημαίνει ότι η πλειοψηφία των κλάσεων ισοδυναμίας έχουν σχεδόν μηδενικό βάρος, άρα δεν συνεισφέρουν ουσιαστικά στην εκ των υστέρων κατανομή.

4.3 Ταχύτητα σύγκλισης του αλγορίθμου

Σε αυτήν την ενότητα μελετάται η ταχύτητα σύγκλισης του Αλγορίθμου ECR. Αρχικά, συγκρίνεται η ταχύτητα σύγκλισης στην από κοινού κατανομή σε σχέση με την ταχύτητα σύγκλισης του αντίστοιχου αρχικού αλγορίθμου (που έχει σαν στόχο την συμμετρική εκ των υστέρων κατανομή). Ακολούθως, συγκρίνεται η ταχύτητα σύ-

γκλισης στην οριακή κατανομή των διανυσμάτων κατάταξης. Τέλος, αποδεικνύεται η υπό συνθήκη ομοιόμορφη εργοδικότητα του αλγορίθμου μας.

Σε ό,τι ακολουθεί, για ευκολία στον συμβολισμό αποφεύγεται η ρητή έκφραση της εξάρτησης των κατανομών από τα παρατηρηθέντα δεδομένα (\mathbf{x}). Έτσι, έστω ότι $f_{\mathcal{Z}_0}^{(m)}(\mathbf{z}, \boldsymbol{\eta})$ και $f^{(m)}(\mathbf{z}, \boldsymbol{\eta})$ είναι οι πυκνότητες των παραγόμενων $(\mathbf{z}^{(m)}, \boldsymbol{\eta}^{(m)})$, σύμφωνα με τον αλγόριθμο ECR και τον αρχικό αλγόριθμο, αντίστοιχα. Εξ ορισμού του Αλγορίθμου ECR, κάθε παραγόμενο ζευγάρι $(\mathbf{z}, \boldsymbol{\eta})$ από τον αρχικό αλγόριθμο μετασχηματίζεται στο $(\tau\mathbf{z}, \tau^{-1}\boldsymbol{\eta})$, όπου η μετάθεση τ επιλέγεται μεταξύ των $(k - k_0(\mathbf{z}))!$ μεταθέσεων που μετασχηματίζουν το \mathbf{z} στο καθορισμένο \mathcal{Z}_0 με πιθανότητα $1/(k - k_0(\mathbf{z}))!$. Αυτό σημαίνει ότι

$$f_{\mathcal{Z}_0}^{(m)}(\mathbf{z}, \boldsymbol{\eta}) = \sum_{\tau \in \mathcal{T}_k} \frac{f^{(m)}(\tau\mathbf{z}, \tau^{-1}\boldsymbol{\eta})}{(k - k_0(\mathbf{z}))!}, \quad \forall \mathbf{z} \in \mathcal{Z}_0, \boldsymbol{\eta} \in \mathcal{H}. \quad (4.20)$$

Γενικά, παρουσιάζει ενδιαφέρον η σύγκριση του ρυθμού σύγκλισης των ακολουθιών $(\mathbf{z}, \boldsymbol{\eta})$ βάσει του Αλγορίθμου ECR και της αντίστοιχης ακολουθίας του αρχικού αλγορίθμου στις κατανομές στόχους αυτών. Η απάντηση δίνεται στην Πρόταση 4.3.1. Προκειμένου να αποδειχθεί το αποτέλεσμα αυτό, θα θεωρηθεί επίσης ο δειγματολήπτης τυχαίας μετάθεσης (που αναφέρθηκε στην Ενότητα 2.6.1) της Frühwirth-Schnatter (2001) ο οποίος, υπό το πλαίσιο που αναπτύσσουμε, ορίζεται ως ακολούθως. Μετά την προσομοίωση του $(\mathbf{z}, \boldsymbol{\eta})$ παράγεται μια τυχαία μετάθεση ομοιόμορφα στο σύνολο \mathcal{T}_k η οποία μετασχηματίζει την δοθείσα κατάσταση της αλυσίδας στο $(\tau\mathbf{z}, \tau^{-1}\boldsymbol{\eta})$. Σύμφωνα με την Frühwirth-Schnatter (2001), η συγκεκριμένη κίνηση δεν αλλάζει την κατανομή στόχο της αλυσίδας. Παρομοίως με τον παραπάνω συμβολισμό, έστω $f_*^{(m)}(\mathbf{z}, \boldsymbol{\eta})$ η πυκνότητα του m -οστού βήματος σύμφωνα με τον συγκεκριμένο διευρυμένο δειγματολήπτη.

Για την συνέχεια θα υποθέσουμε ότι ο πυρήνας μετάβασης της αρχικής αλυσίδας είναι αναλλοίωτος ως προς τις μεταθέσεις των δεικτών των συνιστωσών, με την έννοια ότι ισχύει

$$f(\tau\mathbf{z}', \tau^{-1}\boldsymbol{\eta}' | \tau\mathbf{z}, \tau^{-1}\boldsymbol{\eta}) = f(\mathbf{z}', \boldsymbol{\eta}' | \mathbf{z}, \boldsymbol{\eta}), \quad \forall \tau \in \mathcal{T}_k. \quad (4.21)$$

Η παραπάνω υπόθεση δεν περιορίζει την ισχύ των αποτελεσμάτων μας καθώς είναι μια αρκετά γενική ιδιότητα που έχουν όλοι οι αλγόριθμοι MCMC που παρουσιάζονται στις εφαρμογές. Είναι εύκολο να διαπιστωθεί ότι η υπόθεση (4.21) ισχύει όταν οι προσομοιώσεις των παραμέτρων γίνονται βάσει των πλήρων δεσμευμένων κατανομών (δες Λήμμα Γ.2.1 του Παραρτήματος Γ). Γενικά, η (4.21) ισχύει για την πλειοψηφία των αλγορίθμων MCMC για τα κλασσικά μοντέλα μείξεων αλλά και για πιο γενικά μη

παρατηρήσιμα μαρκοβιανά μοντέλα. Ειδικότερα, η συγκεκριμένη σχέση ισχύει για όλους τους αλγορίθμους που θα παρουσιαστούν στις εφαρμογές που παρουσιάζονται στο επόμενο κεφάλαιο. Για παράδειγμα, μπορεί να δειχτεί ότι η συγκεκριμένη σχέση ισχύει και για τον δειγματολήπτη των Robert et al. (2000), όπου δεν είναι άμεση η προσομοίωση κάποιων παραμέτρων από τις πλήρεις δεσμευμένες εκ των υστέρων κατανομές τους.

Δεδομένης της υπόθεσης (4.21), αποδεικνύουμε στην συνέχεια δύο βοηθητικά λήμματα που οδηγούν στο βασικό μας αποτέλεσμα που διατυπώνεται στην Πρόταση 4.3.1, σύμφωνα με το οποίο η ταχύτητα σύγκλισης του αλγορίθμου ECR είναι τουλάχιστον ίση με την ταχύτητα σύγκλισης του αρχικού αλγορίθμου. Η πορεία που ακολουθείται για την απόδειξη είναι η εξής: Αρχικά αποδεικνύεται ότι αν ο αρχικός αλγόριθμος είναι συμπληρωμένος με την κίνηση label switching της Frühwirth-Schnatter, τότε η ταχύτητα σύγκλισης αυτού βελτιώνεται (Λήμμα 4.3.2). Κατόπιν, αποδεικνύεται ότι η ταχύτητα σύγκλισης του Αλγορίθμου ECR είναι πάντοτε ίση με την ταχύτητα του αρχικού αλγορίθμου όταν αυτός συμπληρώνεται με την κίνηση label switching (Λήμμα 4.3.3). Το Λήμμα 4.3.1 είναι μια συνέπεια της (4.21) και θα φανεί χρήσιμο στην πορεία της απόδειξης του αποτελέσματός μας.

Λήμμα 4.3.1. Έστω ότι ισχύει η (4.21). Τότε, για κάθε $z \in \mathcal{Z}$, $\eta \in \mathcal{H}$,

$$f_*^{(m)}(z, \eta) = \frac{1}{k!} \sum_{\tau \in \mathcal{I}_k} f^{(m)}(\tau z, \tau^{-1} \eta). \quad (4.22)$$

Επί πλέον, ισχύει ότι $f_*^{(m)}(\tau z, \tau^{-1} \eta) = f_*^{(m)}(z, \eta)$ για κάθε $\tau \in \mathcal{I}_k$.

Απόδειξη. Το αποτέλεσμα θα δειχθεί μέσω επαγωγής. Από τον ορισμό της κίνησης label switching το αποτέλεσμα ισχύει για $m = 1$. Τότε, υποθέτοντας ότι ισχύει για κάποιο $m \geq 1$ έχουμε ότι

$$\begin{aligned} f_*^{(m+1)}(z', \eta') &= \sum_{z \in \mathcal{Z}} \int_{\mathcal{H}} f(z', \eta' | z, \eta) f_*^{(m)}(z, \eta) d\eta \\ &\stackrel{(4.22)}{=} \sum_{z \in \mathcal{Z}} \int_{\mathcal{H}} f(z', \eta' | z, \eta) \left\{ \frac{1}{k!} \sum_{\tau \in \mathcal{I}_k} f^{(m)}(\tau z, \tau^{-1} \eta) \right\} d\eta \\ &\stackrel{(4.21)}{=} \frac{1}{k!} \sum_{\tau \in \mathcal{I}_k} \sum_{z \in \mathcal{Z}} \int_{\mathcal{H}} f(\tau z', \tau^{-1} \eta' | \tau z, \tau^{-1} \eta) f^{(m)}(\tau z, \tau^{-1} \eta) d\eta \\ &= \frac{1}{k!} \sum_{\tau \in \mathcal{I}_k} \sum_{\tilde{z} \in \mathcal{Z}} \int_{\mathcal{H}} f(\tau z', \tau^{-1} \eta' | \tilde{z}, \tilde{\eta}) f^{(m)}(\tilde{z}, \tilde{\eta}) d\tilde{\eta} \end{aligned}$$

(μέσω της αλλαγής μεταβλητής $(\tilde{z}, \tilde{\eta}) = (\tau z, \tau^{-1} \eta)$ και λόγω του ότι $\tau \mathcal{Z} = \mathcal{Z}$, $\tau^{-1} \mathcal{H} = \mathcal{H}$, $\forall \tau \in \mathcal{I}_k$)

$$= \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f^{(m+1)}(\tau \mathbf{z}', \tau^{-1} \boldsymbol{\eta}')$$

και τούτο ολοκληρώνει την απόδειξη της (4.22). Για το δεύτερο επιχείρημα, αρχικά ως παρατηρήσουμε ότι από τον Ορισμό 4.2.1 έπεται ότι για οποιεσδήποτε μεταθέσεις $\tau = (t_1, \dots, t_k)$ και $\rho = (r_1, \dots, r_k)$ έχουμε ότι

$$\rho(\tau \mathbf{z}) = \rho(t_{z_1}, \dots, t_{z_n}) = (r_{t_{z_1}}, \dots, r_{t_{z_n}}) = (\tau \rho) \mathbf{z}, \quad (4.23)$$

διότι $\tau \rho = (r_{t_1}, \dots, r_{t_k})$. Από την άλλη πλευρά, για κάθε $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)$ έχουμε ότι

$$\rho(\tau \boldsymbol{\eta}) = \rho(\eta_{t_1}, \dots, \eta_{t_k}) = (\eta_{t_{r_1}}, \dots, \eta_{t_{r_k}}) = (\rho \tau) \boldsymbol{\eta}. \quad (4.24)$$

Επομένως, λόγω της (4.22), έχουμε ότι για κάθε $\tau \in \mathcal{T}_k$

$$\begin{aligned} f_*^{(m)}(\tau \mathbf{z}, \tau^{-1} \boldsymbol{\eta}) &= \frac{1}{k!} \sum_{\rho \in \mathcal{T}_k} f^{(m)}(\rho(\tau \mathbf{z}), \rho^{-1}(\tau^{-1} \boldsymbol{\eta})) \\ &\stackrel{(4.23), (4.24)}{=} \frac{1}{k!} \sum_{\rho \in \mathcal{T}_k} f^{(m)}((\tau \rho) \mathbf{z}, (\rho^{-1} \tau^{-1}) \boldsymbol{\eta}) \\ &= \frac{1}{k!} \sum_{\rho \in \mathcal{T}_k} f^{(m)}((\tau \rho) \mathbf{z}, (\tau \rho)^{-1} \boldsymbol{\eta}) \\ &\quad \text{(από την γνωστή ιδιότητα } (\tau \rho)^{-1} = \rho^{-1} \tau^{-1}) \\ &= \frac{1}{k!} \sum_{\gamma \in \mathcal{T}_k} f^{(m)}(\gamma \mathbf{z}, \gamma^{-1} \boldsymbol{\eta}) \\ &\quad \text{(θέτοντας } \tau \rho = \gamma) \\ &= f_*^{(m)}(\mathbf{z}, \boldsymbol{\eta}) \end{aligned}$$

και η απόδειξη ολοκληρώθηκε. \square

Στην βιβλιογραφία συχνά αναφέρεται ότι αν ένας δειγματολήπτης είναι διευρυμένος με την κίνηση label switching της Frühwirth-Schnatter τότε επιταχύνεται η σύγκλιση στην κατανομή-στόχο (δες Jasra et al., 2005). Τούτο είναι διαισθητικά λογικό γιατί αυτή η κίνηση εξαναγκάζει την ακολουθία να επισκέπτεται όλες τις συμμετρικές περιοχές της εκ των υστέρων κατανομής. Παρ' όλ' αυτά, δεν έχει δοθεί ακόμη μια αυστηρή απόδειξη αυτής της εικασίας. Στο λήμμα που ακολουθεί αποδεικνύεται ότι είναι όντως αληθής.

Πριν διατυπώσουμε το λήμμα θα πρέπει να αναφέρουμε ότι για θέματα σύγκλισης των αλυσίδων που προσομοιώνονται από αλγορίθμους MCMC χρησιμοποιείται η απόσταση ολικής μεταβολής. Η συγκεκριμένη απόσταση μεταξύ δύο κατανομών G_1 και G_2 ισούται με

$$\|G_1 - G_2\| = \sup_A |G_1(A) - G_2(A)|.$$

Σε περίπτωση που οι G_1, G_2 έχουν αντίστοιχες πυκνότητες g_1, g_2 ως προς κάποιο μέτρο ν , είναι εύκολο να δειχτεί ότι

$$\|G_1 - G_2\| \equiv \|g_1 - g_2\| = \frac{1}{2} \int |g_1(x) - g_2(x)| \nu(\mathbf{d}x).$$

Λήμμα 4.3.2. Έστω ότι η (4.21) ισχύει. Τότε,

$$\|f_*^{(m)} - f\| \leq \|f^{(m)} - f\|, \quad \forall m \geq 1.$$

Απόδειξη. Έστω ότι Z_0 είναι ένα οποιοδήποτε σύνολο αντιπροσώπων των κλάσεων ισοδυναμίας. Τότε,

$$\begin{aligned} 2\|f_*^{(m)} - f\| &= \sum_{z \in Z} \int_{\mathcal{H}} |f_*^{(m)}(z, \eta) - f(z, \eta)| d\eta \\ &\stackrel{(4.12)}{=} \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f_*^{(m)}(\tau z, \eta)}{(k - k_0(z))!} - \frac{f(\tau z, \eta)}{(k - k_0(z))!} \right| d\eta \\ &= \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f_*^{(m)}(\tau z, \tau^{-1}\tilde{\eta})}{(k - k_0(z))!} - \frac{f(\tau z, \tau^{-1}\tilde{\eta})}{(k - k_0(z))!} \right| d\tilde{\eta} \\ &\quad (\text{μέσω της αλλαγής μεταβλητής } \tilde{\eta} = \tau\eta) \\ &= \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f_*^{(m)}(z, \tilde{\eta})}{(k - k_0(z))!} - \frac{f(z, \tilde{\eta})}{(k - k_0(z))!} \right| d\tilde{\eta} \\ &\quad (\text{διότι } f_*^{(m)}(\tau z, \tau^{-1}\tilde{\eta}) = f_*^{(m)}(z, \tilde{\eta}) \text{ σύμφωνα με το Λήμμα 4.3.1}) \\ &= \sum_{z \in Z_0} \int_{\mathcal{H}} \left| \frac{k! f_*^{(m)}(z, \tilde{\eta})}{(k - k_0(z))!} - \frac{k! f(z, \tilde{\eta})}{(k - k_0(z))!} \right| d\tilde{\eta} \tag{4.25} \\ &\stackrel{(4.22)}{=} \sum_{z \in Z_0} \int_{\mathcal{H}} \left| \frac{\sum_{\tau \in \mathcal{T}_k} f^{(m)}(\tau z, \tau^{-1}\tilde{\eta})}{(k - k_0(z))!} - \frac{k! f(z, \tilde{\eta})}{(k - k_0(z))!} \right| d\tilde{\eta} \\ &\leq \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f^{(m)}(\tau z, \tau^{-1}\tilde{\eta})}{(k - k_0(z))!} - \frac{f(z, \tilde{\eta})}{(k - k_0(z))!} \right| d\tilde{\eta} \\ &= \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f^{(m)}(\tau z, \eta)}{(k - k_0(z))!} - \frac{f(z, \tau\eta)}{(k - k_0(z))!} \right| d\eta \\ &\quad (\text{μέσω της αλλαγής μεταβλητής } \tau\eta = \tilde{\eta}) \\ &\stackrel{(4.15)}{=} \sum_{z \in Z_0} \sum_{\tau \in \mathcal{T}_k} \int_{\mathcal{H}} \left| \frac{f^{(m)}(\tau z, \eta)}{(k - k_0(z))!} - \frac{f(\tau z, \eta)}{(k - k_0(z))!} \right| d\eta \\ &\stackrel{(4.12)}{=} \sum_{z \in Z} \int_{\mathcal{H}} |f^{(m)}(z, \eta) - f(z, \eta)| d\eta \\ &= 2\|f^{(m)} - f\|. \end{aligned}$$

και τούτο ολοκληρώνει την απόδειξη. □

Λήμμα 4.3.3. Για κάθε σύνολο αντιπροσώπων των κλάσεων ισοδυναμίας \mathcal{Z}_0 ισχύει ότι

$$\|f_{\mathcal{Z}_0}^{(m)} - f_{\mathcal{Z}_0}\| = \|f_*^{(m)} - f\|, \quad \forall m \geq 1.$$

Απόδειξη. Έχουμε ότι

$$\begin{aligned} 2\|f_{\mathcal{Z}_0}^{(m)} - f_{\mathcal{Z}_0}\| &= \sum_{z \in \mathcal{Z}_0} \int_{\mathcal{H}} |f_{\mathcal{Z}_0}^{(m)}(z, \eta) - f_{\mathcal{Z}_0}(z, \eta)| d\eta \\ &\stackrel{(4.20)}{=} \sum_{z \in \mathcal{Z}_0} \int_{\mathcal{H}} \left| \frac{\sum_{\tau \in \mathcal{I}_k} f^{(m)}(\tau^{-1}z, \tau\eta)}{(k - k_0(z))!} - \frac{k! f(z, \eta)}{(k - k_0(z))!} \right| d\eta \\ &= \sum_{z \in \mathcal{Z}_0} \int_{\mathcal{H}} \left| \frac{k! f_*^{(m)}(z, \eta)}{(k - k_0(z))!} - \frac{k! f(z, \eta)}{(k - k_0(z))!} \right| d\eta \\ &= 2\|f_*^{(m)} - f\| \end{aligned}$$

από την (4.25). □

Από την απόδειξη του Λήμματος 4.3.3 είναι σαφές ότι η ταχύτητα σύγκλισης της μαρκοβιανής αλυσίδας που παράγεται από τον Αλγόριθμο ECR παραμένει ίδια, ανεξαρτήτως από το αν ο αρχικός αλγόριθμος είναι διευρυμένος με την κίνηση label switching ή όχι. Τώρα είμαστε σε θέση να διατυπώσουμε το βασικό αποτέλεσμα που αναφέρθηκε στην αρχή της ενότητας.

Πρόταση 4.3.1. Έστω $f^{(m)}$ και $f_{\mathcal{Z}_0}^{(m)}$ οι κατανομές των ακολουθιών $(z, \eta)^{(m)}$ βάσει του αρχικού αλγόριθμου MCMC και του αλγόριθμου ECR, αντίστοιχα, και έστω $f, f_{\mathcal{Z}_0}$ οι αντίστοιχες κατανομές στόχοι αυτών. Τότε, υπό την ισχύ της υπόθεσης (4.21), ισχύει

$$\|f_{\mathcal{Z}_0}^{(m)} - f_{\mathcal{Z}_0}\| \leq \|f^{(m)} - f\|, \quad \forall m \geq 1.$$

Απόδειξη. Το αποτέλεσμα είναι άμεση συνέπεια των Λημμάτων 4.3.2 και 4.3.3. □

Σε αρκετές εφαρμογές έχει ενδιαφέρον η ομαδοποίηση των παρατηρήσεων σε ομογενή σύνολα. Στο δικό μας πλαίσιο αυτό μεταφράζεται στην εκτίμηση της εκ των υστέρων κατανομής των διανυσμάτων κατάταξης. Ακολουθώντας αποδεικνύουμε ότι σε κάθε περίπτωση (και ανεξάρτητα από την ισχύ της (4.21)), η σύγκλιση στην εκ των υστέρων περιθωριακή κατανομή των διανυσμάτων κατάταξης είναι μικρότερη ή ίση από την αντίστοιχη του αρχικού αλγορίθμου.

Πρόταση 4.3.2. Για κάθε $m \geq 1$ ισχύουν τα εξής:

$$1. \|w_{\mathcal{Z}_0}^{(m)}(\cdot|\mathbf{x}) - w_{\mathcal{Z}_0}(\cdot|\mathbf{x})\| \leq \|w^{(m)}(\cdot|\mathbf{x}) - w(\cdot|\mathbf{x})\|.$$

2. Μία κανή συνθήκη για να ισχύει η ισότητα είναι $w^{(m)}(\tau z|\mathbf{x}) = w^{(m)}(z|\mathbf{x})$ για κάθε $\tau \in \mathcal{T}_k$, $z \in \mathcal{Z}$.

Απόδειξη. Επειδή $w(\tau z|\mathbf{x}) = w(z|\mathbf{x})$, $\forall \tau \in \mathcal{T}_k$, έχουμε ότι

$$\begin{aligned}
2\|w_{\mathcal{Z}_0}^{(m)}(\cdot|\mathbf{x}) - w_{\mathcal{Z}_0}(\cdot|\mathbf{x})\| &= \sum_{z \in \mathcal{Z}_0} |w_{\mathcal{Z}_0}^{(m)}(z|\mathbf{x}) - w_{\mathcal{Z}_0}(z|\mathbf{x})| \\
&= \sum_{z \in \mathcal{Z}_0} \left| \sum_{\tau \in \mathcal{T}_k} \frac{w^{(m)}(\tau z|\mathbf{x})}{(k - k_0(z))!} - \frac{k!w(z|\mathbf{x})}{(k - k_0(z))!} \right| \\
&= \sum_{z \in \mathcal{Z}_0} \left| \sum_{\tau \in \mathcal{T}_k} \frac{w^{(m)}(\tau z|\mathbf{x})}{(k - k_0(z))!} - \sum_{\tau \in \mathcal{T}_k} \frac{w(z|\mathbf{x})}{(k - k_0(z))!} \right| \\
&= \sum_{z \in \mathcal{Z}_0} \left| \sum_{\tau \in \mathcal{T}_k} \frac{1}{(k - k_0(z))!} (w^{(m)}(\tau z|\mathbf{x}) - w(z|\mathbf{x})) \right| \quad (4.26) \\
&\leq \sum_{z \in \mathcal{Z}_0} \sum_{\tau \in \mathcal{T}_k} \frac{1}{(k - k_0(z))!} |w^{(m)}(\tau z|\mathbf{x}) - w(z|\mathbf{x})| \\
&= \sum_{z \in \mathcal{Z}} |w^{(m)}(z|\mathbf{x}) - w(z|\mathbf{x})| \\
&= 2\|w^{(m)}(\cdot) - w(\cdot|\mathbf{x})\|.
\end{aligned}$$

Για το δεύτερο μέρος της πρότασης παρατηρούμε ότι αν αντικαταστήσουμε την $w^{(m)}(\tau z|\mathbf{x}) = w^{(m)}(z|\mathbf{x})$ στην (4.26) λαμβάνουμε

$$\begin{aligned}
2\|w_{\mathcal{Z}_0}^{(m)}(\cdot|\mathbf{x}) - w_{\mathcal{Z}_0}(\cdot|\mathbf{x})\| &= \sum_{z \in \mathcal{Z}_0} \left| \frac{k!}{(k - k_0(z))!} (w^{(m)}(z|\mathbf{x}) - w(z|\mathbf{x})) \right| \\
&= \sum_{z \in \mathcal{Z}_0} \frac{k!}{(k - k_0(z))!} |(w^{(m)}(z|\mathbf{x}) - w(z|\mathbf{x}))| \\
&= \sum_{z \in \mathcal{Z}} |(w^{(m)}(z|\mathbf{x}) - w(z|\mathbf{x}))| \\
&= 2\|w^{(m)}(\cdot|\mathbf{x}) - w(\cdot|\mathbf{x})\|.
\end{aligned}$$

□

Το τελευταίο αποτέλεσμα της παρούσας ενότητας αφορά στην ομοιόμορφη εργοδικότητα της ακολουθίας των η που παράγεται από τον Αλγόριθμο ECR. Αποδεικνύεται ότι, σε περίπτωση που ο αρχικός αλγόριθμος MCMC ανανεώνει τις παραμέτρους με έναν συγκεκριμένο τρόπο, τότε η ακολουθία των προσομοιωμένων τιμών των παραμέτρων είναι ομοιόμορφα εργοδική.

Πρόταση 4.3.3. Έστω ότι κατά την m -οστή επανάληψη ο αρχικός αλγόριθμος ανανεώνει τα $(z^{(m)}, \eta^{(m)})$ ως εξής. Πρώτα παράγεται το $\eta^{(m)}$ από την κατανομή $f(\eta|\mathbf{x}, z^{(m-1)})$ και ακολούθως παράγονται τα $z_i^{(m)} \sim f(z_i|x_i, \eta, z_{i-1})$, $i = 1, \dots, n$. Τότε υπάρχουν σταθερές $C > 0$ και $0 \leq \rho < 1$ τέτοιες ώστε, για οποιαδήποτε αρχική τιμή $\eta^{(0)}$, η ακολουθία $(\eta_m)_{m \geq 1}$ είναι ομοιόμορφα εργοδική, δηλαδή για την περιθωριακή κατανομή αυτών ισχύει ότι

$$\|f_{z_0}^{(m)} - f_{z_0}\| \leq C\rho^m. \quad (4.27)$$

Απόδειξη. Γνωρίζουμε ότι αν η ανανέωση των παραμέτρων γίνεται με τον συγκεκριμένο τρόπο, τότε, τα $(z^{(m)})$ αποτελούν μία ομογενή μαρκοβιανή διαδικασία η οποία είναι ομοιόμορφα εργοδική (δες Robert et al., 1993). Έτσι, υπάρχουν σταθερές C και $0 \leq \rho < 1$ τέτοιες ώστε

$$\|w^{(m)}(\cdot) - w(\cdot|\mathbf{x})\| \leq C\rho^m \quad (4.28)$$

για κάθε αρχική κατανομή $w^{(0)}(z)$. Είναι

$$\begin{aligned} \|f_{z_0}^{(m+1)}(\cdot) - f_{z_0}(\cdot|\mathbf{x})\| &= \int_{\mathcal{H}} |f_{z_0}^{(m+1)}(\eta) - f_{z_0}(\eta|\mathbf{x})| d\eta \\ &= \int_{\mathcal{H}} \left| \sum_{z \in \mathcal{Z}_0} (f(\eta|z, \mathbf{x}) w_{z_0}^{(m)}(z) - f(\eta|z, \mathbf{x}) w_{z_0}(z|\mathbf{x})) \right| d\eta \\ &= \int_{\mathcal{H}} \left| \sum_{z \in \mathcal{Z}_0} f(\eta|z, \mathbf{x}) \left(\sum_{\tau \in \mathcal{T}_k} \frac{w^{(m)}(\tau z)}{(k - k_0(z))!} - \frac{k! w(z|\mathbf{x})}{(k - k_0(z))!} \right) \right| d\eta \end{aligned}$$

Λόγω της (4.8), η τελευταία έκφραση γράφεται ως εξής

$$\begin{aligned} \|f_{z_0}^{(m+1)}(\cdot) - f_{z_0}(\cdot|\mathbf{x})\| &= \int_{\mathcal{H}} \left| \sum_{z \in \mathcal{Z}_0} f(\eta|z, \mathbf{x}) \frac{1}{(k - k_0(z))!} \sum_{\tau \in \mathcal{T}_k} (w^{(m)}(\tau z) - w(z|\mathbf{x})) \right| d\eta \\ &\leq \int_{\mathcal{H}} \sum_{z \in \mathcal{Z}_0} f(\eta|z, \mathbf{x}) \frac{1}{(k - k_0(z))!} \sum_{\tau \in \mathcal{T}_k} |(w^{(m)}(\tau z) - w(z|\mathbf{x}))| d\eta \\ &= \sum_{z \in \mathcal{Z}_0} \frac{1}{(k - k_0(z))!} \sum_{\tau \in \mathcal{T}_k} |(w^{(m)}(\tau z) - w(z|\mathbf{x}))| \int_{\mathcal{H}} f(\eta|z, \mathbf{x}) d\eta \\ &= \sum_{z \in \mathcal{Z}_0} \frac{1}{(k - k_0(z))!} \sum_{\tau \in \mathcal{T}_k} |(w^{(m)}(\tau z) - w(z|\mathbf{x}))| \\ &= \sum_{z \in \mathcal{Z}} |(w^{(m)}(z) - w(z|\mathbf{x}))| \\ &= \|w^{(m)}(\cdot) - w(\cdot|\mathbf{x})\|, \end{aligned}$$

και η ανισότητα (4.27) προκύπτει από την (4.28). \square

4.4 Ο αλγόριθμος ECR και το πρόβλημα εκχώρησης

Στο Βήμα 2(a) του Αλγορίθμου ECR πρέπει κάθε φορά να επιλέξουμε εκείνο το μέλος της κλάσης ισοδυναμίας το οποίο μεγιστοποιεί το μέτρο ομοιότητας S με τον $z^{\text{MAP}} = z^*$. Πιο συγκεκριμένα έχουμε να επιλύσουμε το πρόβλημα

$$\text{μεγιστοποίησε το } S(\tau z, z^*), \quad \text{για } \tau \in \mathcal{T}_k, \quad (4.29)$$

όπου $\tau = (t_1, \dots, t_k)$ είναι μια μετάθεση του $\{1, \dots, k\}$. Η επίλυση του συγκεκριμένου προβλήματος απαιτεί υψηλό υπολογιστικό κόστος εάν ψάχνουμε μία-μία όλες τις δυνατές μεταθέσεις όταν το πλήθος των συνιστωσών την μείξης k λαμβάνει μεγάλες τιμές. Για παράδειγμα, όταν $k = 6$ πρέπει να υπολογίσουμε το μέτρο ομοιότητας S για $6! = 720$ μεταθέσεις.

Κρίνεται λοιπόν σκόπιμο να αποφύγουμε την παραπάνω χρονοβόρα διαδικασία. Αυτό μπορεί να γίνει μετατρέποντας το πρόβλημα (4.29) σε ένα ισοδύναμο πρόβλημα ακέραιου προγραμματισμού. Γενικά, όταν το ζητούμενο είναι η αντιστοίχιση των k στοιχείων ενός (πεπερασμένου) συνόλου στα στοιχεία ένα άλλου (ισοπληθούς) συνόλου, η λύση ανάγεται στην εύρεση μιας κατάλληλης μετάθεσης του συνόλου των δεικτών $\{1, \dots, k\}$. Αν επί πλέον ο κανόνας βάσει του οποίου θα πρέπει να γίνει η αντιστοίχιση περιγράφεται μέσω γραμμικών περιορισμών, τότε το πρόβλημα εμπίπτει στην κατηγορία των προβλημάτων εκχώρησης (assignment problems) του ακέραιου προγραμματισμού. Για μία εισαγωγή στη συγκεκριμένη περιοχή ο αναγνώστης παραπέμπεται στην μονογραφία των Burkard et al. (2009). Σημειώνουμε ότι η λύση του προβλήματος εκχώρησης έχει ενσωματωθεί και από τον Stephens στο Βήμα 2 του Αλγορίθμου 2.6.2.

Η αναγωγή του αρχικού προβλήματος (4.29) στο αντίστοιχο πρόβλημα εκχώρησης γίνεται ορίζοντας έναν κατάλληλο πίνακα κόστους $C = (c_{ij})$ διάστασης $k \times k$. Έτσι, το ισοδύναμο πρόβλημα μπορεί να επιλυθεί αποδοτικά εφαρμόζοντας κάποιον από τους διαθέσιμους αλγορίθμους επίλυσης του προβλήματος εκχώρησης. Ενδεικτικά, αναφέρουμε τη ρουτίνα του Carpaneto (1980) στη Fortran και τη ρουτίνα `lp.assign` των Berkelaar et al. (2008) που περιέχεται στο πακέτο `lp.Solve` της R.

Για την κατασκευή του πίνακα κόστους C ορίζουμε αρχικά τα σύνολα

$$\mathcal{I}_i := \{\ell \in \{1, \dots, n\} : z_\ell = i\}, \quad i = 1, \dots, k.$$

Το (i, j) στοιχείο του C ορίζεται ως

$$c_{ij} := |\mathcal{I}_i| - \sum_{\ell \in \mathcal{I}_i} I(z_\ell^* = j), \quad i, j = 1, \dots, k,$$

όπου με $|\mathcal{A}|$ συμβολίζουμε τον πληθάρημο του (διακριτού) συνόλου \mathcal{A} .

Για $i, j = 1, \dots, k$, ορίζουμε τις δίτιμες μεταβλητές απόφασης

$$\delta_{ij} := \begin{cases} 1, & \text{αν ο δείκτης } i \text{ αντιστοιχιστεί στον δείκτη } j \\ 0, & \text{διαφορετικά.} \end{cases}$$

Έτσι, το πρόβλημα (4.29) είναι ισοδύναμο με το εξής πρόβλημα εκχώρησης:

$$\begin{aligned} & \min_{\delta_{ij} \in \{0,1\}, i,j=1,\dots,k} \sum_{i=1}^k \sum_{j=1}^k \delta_{ij} c_{ij} \\ & \text{υπό τους περιορισμούς} \quad \sum_{i=1}^k \delta_{ij} = 1, \quad \forall j = 1, \dots, k \\ & \quad \quad \quad \sum_{j=1}^k \delta_{ij} = 1, \quad \forall i = 1, \dots, k. \end{aligned}$$

Η λύση της (4.29) προκύπτει θέτοντας

$$t_i = j \Leftrightarrow \delta_{ij} = 1.$$

Κεφάλαιο 5

Εφαρμογές του Αλγορίθμου ECR

5.1 Εισαγωγή

Σε αυτό το Κεφάλαιο εφαρμόζεται ο Αλγόριθμος ECR σε μία πληθώρα παραδειγμάτων που καλύπτουν τρεις μεγάλες κατηγορίες μοντέλων ελλειπών δεδομένων της μορφής (4.1). Η Ενότητα 5.2 αφορά σε κλασσικά μοντέλα μείζων κατανομών αλλά και (πολυμεταβλητών) παλινδρομήσεων. Στην Ενότητα 5.3 η προτεινόμενη μέθοδος εφαρμόζεται σε hidden Markov models, αφού πρώτα γίνει μία εισαγωγή σε αυτήν την κατηγορία μοντέλων. Τέλος, στην Ενότητα 5.4 παρουσιάζονται κάποιες εφαρμογές σε μοντέλα που ανήκουν στην κατηγορία των Markov random fields. Σε κάθε μία από αυτές τις περιπτώσεις εξακριβώνονται οι ιδιότητες (4.2), (4.3) και (4.4), οι οποίες εξασφαλίζουν την ισχύ των αποτελεσμάτων του Κεφαλαίου 4.

Στα παραδείγματα χρησιμοποιούνται τόσο προσομοιωμένα όσο και πραγματικά σύνολα δεδομένων η διάσταση των οποίων ποικίλλει. Μελετώνται προβλήματα από την Αστρονομία, Γεωλογία, Χημεία, Οικονομία, Βιολογία, Μετεωρολογία και Ψηφιακή Κατάτμηση εικόνων, και αναδεικνύεται το ευρύ φάσμα εφαρμογών της προτεινόμενης μεθόδου. Επί πλέον, γίνεται σύγκριση των αποτελεσμάτων του Αλγορίθμου ECR με τα αντίστοιχα των διαθέσιμων εναλλακτικών μεθόδων που αναφέρθηκαν στην Ενότητα 2.6. Σε όλες τις περιπτώσεις, η υπεροχή του Αλγορίθμου ECR φαίνεται ξεκάθαρα κρίνοντας είτε από την ποιότητα των εκτιμήσεων που προκύπτουν είτε από την πολυπλοκότητα των αλγορίθμων.

Σχεδόν όλοι οι αλγόριθμοι προσομοίωσης και αναδιάταξης έχουν προγραμματιστεί στη γλώσσα Fortran 90. Η βέλτιστη μετάθεση στο βήμα 2 του Αλγορίθμου ECR έχει βρεθεί χρησιμοποιώντας την ρουτίνα του Carpaneto (1980) η οποία επιλύει αποδοτικά το πρόβλημα εκχώρησης. Σημειώνεται ότι η συγκεκριμένη ρουτίνα ενσωματώθηκε και στον Αλγόριθμο 2.6.2 του Stephens. Αυτό έγινε διότι στις Ενότητες 5.2.2 και 5.2.3 συγκρίνεται ο Αλγόριθμος ECR με τον Αλγόριθμο 2.6.2 ως προς

τον μέσο χρόνο αναδιάταξης του προσομοιωμένου δείγματος. Έτσι, η σύγκριση των δύο αλγορίθμων γίνεται επί ίσοις όροις και καθίσταται σαφής η μείωση του υπολογιστικού κόστους που επιφέρει η προτεινόμενη μέθοδος. Τέλος, οι προσομοιώσεις της Ενότητας 5.4 έγιναν στη γλώσσα R, χρησιμοποιώντας το πακέτο `lp.Solve` για τη λύση του προβλήματος εκχώρησης.

5.2 Εφαρμογές σε κλασσικά μοντέλα μείξεων κατανομών

Σε αυτήν την ενότητα εφαρμόζεται η προτεινόμενη μέθοδος για την επίλυση του προβλήματος label switching μέσω μονοδιάστατων και πολυδιάστατων συνόλων δεδομένων. Ο Αλγόριθμος ECR χρησιμοποιείται για την αναδιάταξη προσομοιωμένων δειγμάτων από την εκ των υστέρων κατανομή τόσο σε προσομοιωμένα όσο και πραγματικά σύνολα δεδομένων. Παράλληλα, γίνονται συγκρίσεις των αποτελεσμάτων με υπάρχουσες μεθόδους όπως ο αλγόριθμος Πιλότικης Αναδιάταξης και Αναδιάταξης μέσω των αλγορίθμων Kullback-Leibler. Τέλος, η μέθοδός μας εφαρμόζεται και σε ένα παράδειγμα όπου ο αρχικός δειγματολήπτης δεν προσομοιώνει τις μεταβλητές κατάταξης, ώστε να δείξουμε ότι είναι δυνατόν να εφαρμόσουμε τον Αλγόριθμο ECR και σε τέτοιου είδους περιπτώσεις. Σε όλες τις περιπτώσεις θεωρείται γνωστός ο αριθμός συνιστωσών της μείξης. Σημειώνεται επίσης ότι σε όλους τους αλγορίθμους έχει προστεθεί μία κίνηση label switching ώστε να εξασφαλιστεί η παρουσία του φαινομένου. Τέλος, κάθε εκτίμηση συνοδεύεται από το αντίστοιχο τυπικό σφάλμα αυτής, το οποίο έχει εκτιμηθεί από 100 ανεξάρτητες αλυσίδες που έχουν ξεκινήσει από διαφορετικές αρχικές τιμές.

Πριν προχωρήσουμε, θα πρέπει να επαληθεύσουμε τις ιδιότητες της Ενότητας 4.2 ώστε να εξασφαλίσουμε ότι ισχύουν τα θεωρητικά αποτελέσματα του προηγούμενου κεφαλαίου. Αρχικά, ας παρατηρήσουμε ότι για τα κλασσικά παραμετρικά μοντέλα μείξεων (2.3), έχουμε $\eta = ((p_1, \theta_1), \dots, (p_k, \theta_k))$, όπου $\mathbf{p} \in \mathcal{P}_k$ και $\boldsymbol{\theta} \in \Theta^k$. Επομένως, αν $\eta \in \mathcal{H}$, τότε έχουμε ότι και $\tau\eta \in \mathcal{H}$ για κάθε $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$ διότι προφανώς $\tau\mathbf{p} = (p_{t_1}, \dots, p_{t_k}) \in \mathcal{P}_k$ και $\tau\boldsymbol{\theta} = (\theta_{t_1}, \dots, \theta_{t_k}) \in \Theta^k$, και έτσι εξασφαλίζεται η (4.2). Επίσης, οι εκ των προτέρων υποθέσεις (2.18), (2.19) με $\beta_j = \beta, \forall j = 1, \dots, k$, εξασφαλίζουν την (4.3). Η σχέση (2.32) είναι η (4.5) στην περίπτωση των κλασσικών μείξεων κατανομών. Τέλος, είναι εύκολο να δούμε ότι η από κοινού συνάρτηση πυκνότητας των πλήρων δεδομένων έχει πράγματι την ιδιότητα (4.4). Από την (2.9), λαμβάνουμε διαδοχικά ότι

$$f(\mathbf{x}, \tau\mathbf{z}|\eta) \equiv f(\mathbf{x}, \tau\mathbf{z}|\mathbf{p}, \boldsymbol{\theta})$$

$$\begin{aligned}
&= \prod_{i=1}^n p_{t_{z_i}} f(x_i | \theta_{t_{z_i}}) \\
&= \prod_{i=1}^n (\tau \mathbf{p})_{z_i} f(x_i | (\tau \boldsymbol{\theta})_{z_i}) \\
&= f(\mathbf{x}, \mathbf{z} | \tau(\mathbf{p}, \boldsymbol{\theta})) \\
&\equiv f(\mathbf{x}, \mathbf{z} | \tau \boldsymbol{\eta}), \quad \forall \tau = (t_1, \dots, t_k) \in \mathcal{T}_k.
\end{aligned}$$

5.2.1 Αλγόριθμος ECR έναντι Αλγορίθμου Πιλοτικής Αναδιάταξης

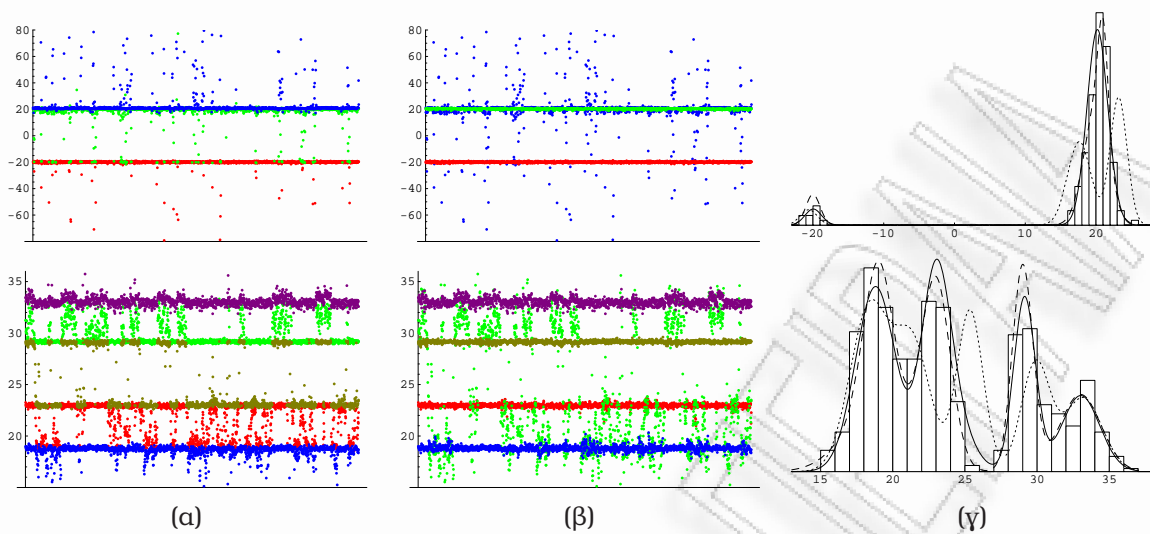
Είδαμε στην Ενότητα 4.2.2 ότι ο τρόπος καθορισμού του συνόλου αντιπροσώπων στο Βήμα 1 του Αλγορίθμου 4.2.1 δανείζεται παρόμοιες ιδέες από τον Αλγόριθμο Πιλοτικής Αναδιάταξης. Παρ' όλ' αυτά τονίσαμε ότι ο αλγόριθμός μας είναι απαλλαγμένος από τα μειονεκτήματα αυτής της μεθόδου. Στις εφαρμογές που ακολουθούν δικαιολογούμε τον ισχυρισμό μας χρησιμοποιώντας δύο προσομοιωμένα σύνολα δεδομένων από τις εξής μείξεις κανονικών κατανομών:

$$0.10\mathcal{N}(-20, 1) + 0.65\mathcal{N}(20, 3) + 0.25\mathcal{N}(21, 0.5), \quad (5.1)$$

$$0.20\mathcal{N}(19, 5) + 0.20\mathcal{N}(19, 1) + 0.25\mathcal{N}(23, 1) + 0.20\mathcal{N}(29, 0.5) + 0.15\mathcal{N}(33, 2). \quad (5.2)$$

Συγκεκριμένα, προσομοιώσαμε $n = 160$ και $n = 600$ παρατηρήσεις από τις κατανομές (5.1) και (5.2) αντίστοιχα. Και στις δύο περιπτώσεις μοντελοποιήσαμε τα δεδομένα χρησιμοποιώντας το random beta model των Richardson and Green (1997) (δες Παράδειγμα 2.5.1) θεωρώντας τον αριθμό συνιστωσών ίσο με την πραγματική του τιμή (3 και 5 αντίστοιχα). Αφού προσομοιώσαμε μέσω του δειγματολήπτη Gibbs μία αλυσίδα Markov με οριακή κατανομή την αντίστοιχη εκ των υστέρων κατανομή, αναδιατάξαμε τα παραμετρικά διανύσματα σύμφωνα με τον Αλγόριθμο Πιλοτικής Αναδιάταξης και την προτεινόμενη μέθοδο. Οι αναδιατεταγμένες προσομοιωμένες τιμές των μέσων των κανονικών συνιστωσών φαίνονται στο Σχήμα 5.1. Στην τρίτη στήλη παρουσιάζονται τα ιστογράμματα των προσομοιωμένων δεδομένων μαζί με την προκύπτουσα εκτίμηση της συναρτήσης πυκνότητας αυτών. Επί πλέον, οι εκτιμήσεις των εκ των υστέρων μέσων τιμών των παραμέτρων από τους αντίστοιχους εργοδικούς μέσους παρατίθενται στις δύο πρώτες γραμμές του Πίνακα 5.1. Είναι εμφανές ότι υπάρχουν πολύ μεγάλες διαφορές μεταξύ των αποτελεσμάτων που δίνουν οι δύο μέθοδοι, κάτι το οποίο οφείλεται όπως θα δούμε στην ύπαρξη μικρότερων κορυφών στην εκ των υστέρων κατανομή των παραμέτρων των δύο παραδειγμάτων.

Ας παρατηρήσουμε αρχικά ότι στην κατανομή (5.1), οι δύο τελευταίες συνιστώσες



Σχήμα 5.1: Πάνω: Αποτελέσματα για την μείξη (5.1). Κάτω: Αποτελέσματα για την μείξη (5.2). Αναδιατεταγμένες τιμές των προσομοιωμένων μέσων τιμών σύμφωνα με τους αλγορίθμους (α) Πιλοτικής Αναδιάταξης και (β) ECR. (γ) Εκτιμήσεις των συναρτήσεων πυκνότητας σύμφωνα με τον PRA (εστιαγμένη γραμμή) και τον ECR (συνεχής γραμμή) και η πραγματική πυκνότητα (διακεκομμένη γραμμή).

είναι αρκετά κοντά μεταξύ τους. Για αυτόν τον λόγο, αναμένεται ότι ο δειγματοληπτής θα συνδυάζει συχνά αυτές τις δύο συνιστώσες σε μία, αφήνοντας μία κενή συνιστώσα όπου οι παράμετροί της θα παράγονται από την εκ των προτέρων κατανομή. Πράγματι, σε 20000 επαναλήψεις (μετά από την περίοδο burn-in), η σχετική συχνότητα του ενδεχομένου ύπαρξης μιας κενής συνιστώσας ήταν περίπου 17%. Αφού δε η διασπορά της εκ των προτέρων κατανομής των μέσων είναι αρκετά μεγάλη (υπενθυμίζουμε ότι έχει ληφθεί ίση με το τετράγωνο του παρατηρηθέντος εύρους σύμφωνα με το μοντέλο των Richardson and Green) έπεται ότι ο μέσος μιας κενής συνιστώσας έχει πιθανότητα 95% να ανήκει στο διάστημα $(-110, 151)$. Σημειώνουμε ότι στο συγκεκριμένο παράδειγμα ισχύει εκ των προτέρων ότι $\mu_j \sim \mathcal{N}(\xi, R^2)$, $j = 1, 2, 3$, με $\xi = 20.52$ και $R = 66.59$. Συνεπώς, η προσομοιωμένη τιμή του μέσου μιας κενής συνιστώσας μπορεί να είναι αρκετά μακριά από την αντίστοιχη περιοχή υψηλής εκ των υστέρων πυκνότητας. Όπως φαίνεται στην πρώτη γραμμή του Σχήματος 5.1(a), η συμπεριφορά αυτή οδηγεί τον Αλγόριθμο Πιλοτικής Αναδιάταξης να λειτουργεί ως εξής: Αν η προσομοιωμένη τιμή από την εκ των προτέρων κατανομή είναι αρκετά μικρή (αντίστοιχα, μεγάλη) τότε η κενή συνιστώσα θα αντιστοιχιστεί σε εκείνη με τον μικρότερο μέσο (αντίστοιχα, μεγαλύτερο) μέσο. Αυτό έχει ως τελικό αποτέλεσμα την υποεκτίμηση του μικρότερου μέσου και υπερεκτίμηση του μεγαλύτερου μέσου. Επιπροσθέτως, αν η συγκεκριμένη προσομοιωμένη τιμή ανήκει στο

	PRA					ECR				
$\hat{\mathbb{E}}(\boldsymbol{\mu} \boldsymbol{x})$	23.08	17.62	-20.59			20.46	19.61	-19.96		
	(.419)	(.391)	(.122)			(.016)	(.194)	(.003)		
$\hat{\mathbb{E}}(\boldsymbol{\sigma}^2 \boldsymbol{x})$	1.83	2.27	1.42			2.01	2.15	1.36		
	(.278)	(.270)	(.021)			(.025)	(.071)	(.009)		
$\hat{\mathbb{E}}(\boldsymbol{p} \boldsymbol{x})$.522	.418	.060			.723	.216	.061		
	(.016)	(.016)	(.000)			(.011)	(.011)	(.000)		
$\hat{\mathbb{E}}(\boldsymbol{\mu} \boldsymbol{x})$	21.61	30.06	18.40	33.25	25.82	22.96	25.28	18.80	32.95	29.13
	(.260)	(.179)	(.101)	(.059)	(.536)	(.005)	(1.007)	(.009)	(.025)	(.007)
$\hat{\mathbb{E}}(\boldsymbol{\sigma}^2 \boldsymbol{x})$	1.47	1.05	1.86	1.69	0.80	1.01	1.57	1.92	1.80	0.56
	(.115)	(.109)	(.063)	(.064)	(.062)	(.006)	(.056)	(.015)	(.038)	(.008)
$\hat{\mathbb{E}}(\boldsymbol{p} \boldsymbol{x})$.222	.141	.317	.128	.192	.255	.070	.360	.138	.178
	(.001)	(.010)	(.015)	(.005)	(.007)	(.003)	(.005)	(.007)	(.003)	(.003)

Πίνακας 5.1: Εργοδικοί μέσοι και οι εκτιμήσεις των τυπικών σφαλμάτων για τις μείξεις της Ενότητας 5.2.1. Πάνω: μείξη (5.1). Κάτω: μείξη (5.2).

διάστημα $(-20, 20)$, τότε η κενή συνιστώσα θα καταταχθεί σε αυτήν που αντιστοιχεί στον μεσαίο μέσο και κάτι τέτοιο οδηγεί σε υποεκτίμηση της μέσης τιμής της δεύτερης συνιστώσας επίσης. Αντιθέτως, ο Αλγόριθμος ECR αντιμετωπίζει με πολύ διαφορετικό τρόπο την παραπάνω κατάσταση. Όπως φαίνεται στο Σχήμα 5.1(β), η συνιστώσα που έχει τον μέσο με την μέγιστη εκ των υστέρων διασπορά αντιστοιχίζεται πάντα στην ίδια ετικέτα (συγκεκριμένα στη «μπλέ»). Παρατηρούμε ότι οι τιμές της «μπλε» συνιστώσας παρουσιάζουν μία διττή συμπεριφορά: Είτε συγκεντρώνονται σε μία σταθερή περιοχή κοντά στο 20 είτε κυμαίνονται στο διάστημα $(-80, 80)$. Είναι ξεκάθαρο λοιπόν ότι υπάρχει μία μικρότερη κορυφή που αντιστοιχεί στην ύπαρξη μιας κενής συνιστώσας. Επί πλέον, η συγκεκριμένη αναδιάταξη παράγει μία καλύτερη προσαρμογή στα δεδομένα καθώς οι δύο πρώτες συνιστώσες έχουν τις μέσες τιμές τους συγκεντρωμένες σε μία πολύ σταθερή περιοχή υψηλής εκ των υστέρων πυκνότητας. Οι προσομοιωμένες τιμές από την εκ των προτέρων κατανομή πάντα αναδιατάσσονται με τέτοιο τρόπο ώστε να αντιστοιχίζονται στην τρίτη συνιστώσα και οι ακραίες τιμές αλληλοεξουδετερώνονται στον υπολογισμό του εργοδικού μέσου ώστε να μην έχουμε προβλήματα υποεκτίμησης ή υπερεκτίμησης. Τα συγκεκριμένα σημεία καταγράφονται συνοπτικά συγκρίνοντας τις αντίστοιχες εκτιμηθείσες συναρτήσεις πυκνότητας με την πραγματική αλλά και με το ισόγραμμα των δεδομένων στην πρώτη γραμμή του Σχήματος 5.1(γ). Είναι φανερό ότι, η προσαρμογή υπό την προτεινόμενη μέθοδο είναι αρκετά καλή, κάτι που δεν ισχύει για την αντίστοιχη υπό τον αλγόριθμο Πιλοτικής Αναδιάταξης.

Στην περίπτωση της μείξης (5.2), το φαινόμενο ύπαρξης κενών συνιστωσών είναι αρκετά σπάνιο λόγω του μεγάλου μεγέθους του δείγματος. Πράγματι, η σχετική

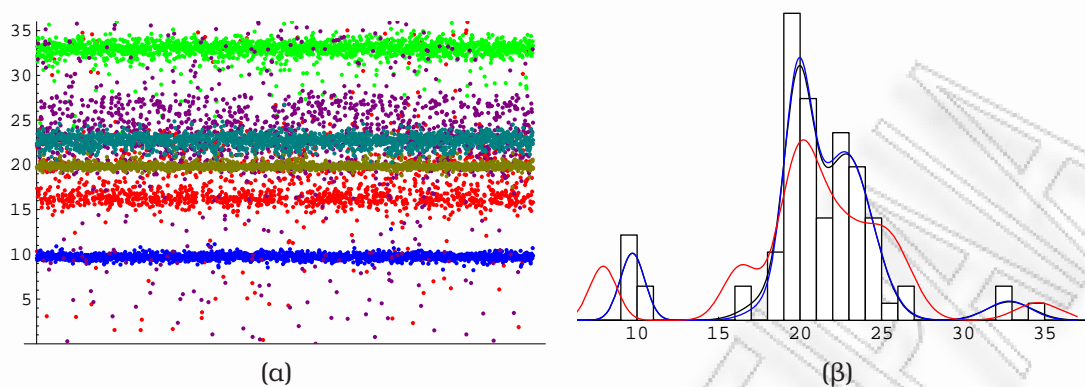
συχνότητα κενών συνιστωσών εκτιμήθηκε γύρω στο 1.5% και συνεπώς οι προσομοιωμένες τιμές από την εκ των προτέρων κατανομή δεν επιδρούν όσο προηγουμένως στις εκτιμήσεις των παραμέτρων. Έτσι λοιπόν θα ήταν αναμενόμενο οι δύο αλγόριθμοι αναδιάταξης να συμπεριφέρονται παρόμοια. Παρ' όλ' αυτά, κάτι τέτοιο δεν συμβαίνει διότι σε αυτό το παράδειγμα αντιμετωπίζουμε μία διαφορετική περίπτωση γνησίως πολυκόρυφης εκ των υστέρων κατανομής. Τα αναδιατεταγμένα δείγματα των μέσων τιμών των συνιστωσών σύμφωνα με τους δύο αλγορίθμους απεικονίζονται στην δεύτερη γραμμή του Σχήματος 5.1.

Ας επικεντρώθουμε αρχικά στην αναδιάταξη μέσω του Αλγορίθμου ECR. Σύμφωνα με τη δεύτερη γραμμή του Σχήματος 5.1(β), υπάρχουν τέσσερις «σταθερές» συνιστώσες και μία συνιστώσα όπου οι προσομοιωμένες τιμές της μέσης τιμής αυτής συγκεντρώνονται σε δύο διαφορετικές περιοχές. Η πρώτη είναι κοντά στο 19 (περιοχή που αντιστοιχεί στην πραγματική τιμή του μέσου της δεύτερης συνιστώσας) και σε άλλη μία —όχι και τόσο δικαιολογημένη— περιοχή εντός του διαστήματος (30, 33). Αυτό οφείλεται στο γεγονός ότι ο δειγματολήπτης σε αρκετές επαναλήψεις συνδυάζει τις δύο πρώτες συνιστώσες σε μία και, αντί να δημιουργεί μία κενή συνιστώσα, διαχωρίζει την πέμπτη σε δύο που έχουν κοντινούς μέσους και διαφορετικές διασπορές. Έτσι, μπορούμε να συμπεράνουμε εδώ ότι υπάρχει ένδειξη για την καλή προσαρμογή των δεδομένων σε δύο διαφορετικά μοντέλα πέντε συνιστωσών: Ένα μοντέλο που έχει τις δυο πρώτες συνιστώσες με κοντινούς μέσους και ένα μοντέλο όπου οι μέσοι των δύο τελευταίων συνιστωσών είναι ακόμα πιο κοντά μεταξύ τους (ενδεχομένως και ίσοι). Από την άλλη πλευρά, ο αλγόριθμος Πιλοτικής Αναδιάταξης αντιμετωπίζει το προσομοιωμένο δείγμα όπως και πριν, καταλήγοντας σε ένα αναδιατεταγμένο δείγμα όπου δεν φαίνονται καν οι δύο κορυφές της εκ των υστέρων κατανομής.

5.2.2 Αλγόριθμος ECR έναντι αλγορίθμων KL: Γαλαξιακό σύνολο δεδομένων

Αφού στην προηγούμενη ενότητα διαπιστώθηκε η υπεροχή της προτεινόμενης μεθόδου έναντι απλών μεθόδων αναδιάταξης όπως ο αλγόριθμος Πιλοτικής Αναδιάταξης, συνεχίζουμε συγκρίνοντας τώρα τον αλγόριθμό μας με πιο εκλεπτυσμένες μεθόδους όπως αυτή του Stephens που βασίζεται στην απόκλιση Kullback–Leibler. Η σύγκριση θα γίνει χρησιμοποιώντας το γαλαξιακό σύνολο δεδομένων. Επίσης θα γίνει σύγκριση και με τον αλγόριθμο Πιλοτικής Αναδιάταξης.

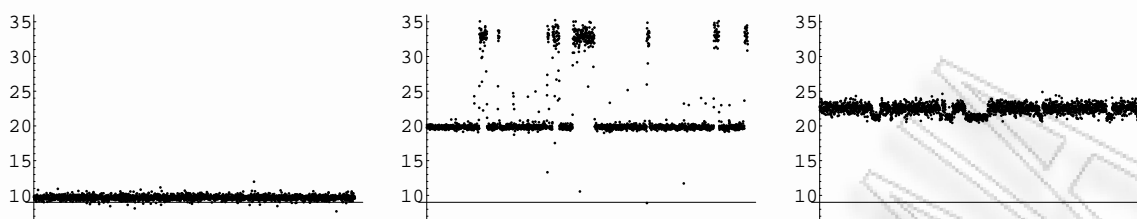
Θεωρούμε αρχικά ότι οι $n = 82$ παρατηρήσεις προέρχονται από μία μείξη $k = 6$ κανονικών κατανομών και τις μοντελοποιούμε σύμφωνα με το random beta model των Richardson and Green (1997). Όπως είδαμε στην Ενότητα 3.5.3 ο συγκεκριμέ-



Σχήμα 5.2: Γαλαξιακό σύνολο δεδομένων με $k = 6$ κανονικές συνιστώσες. (α) Αναδιατεταγμένες τιμές των προσομοιωμένων μέσων τιμών μέσω του Αλγορίθμου ECR. (β) Εκτιμήσεις της συνάρτησης πυκνότητας βάσει των Αλγορίθμων ECR (μαύρη γραμμή), Kullback-Leibler (μπλε γραμμή) και Πιλοτικής Αναδιάταξης (κόκκινη γραμμή).

νος αριθμός συνιστωσών αντιστοιχεί στο πιο πιθανό μοντέλο σύμφωνα με τον τυπικό Αλγόριθμο RJMCMC. Εκτελέσαμε τον δειγματολήπτη Gibbs για 60000 επαναλήψεις (αφού διαγράφηκαν οι πρώτες 10000 επαναλήψεις ως περίοδος burn-in). Στην συνέχεια το προσομοιωμένο δείγμα των (μ, σ^2, p, z) αναδιατάχθηκε σύμφωνα με τον Αλγόριθμο ECR, τον Αλγόριθμο Kullback-Leibler και τον Αλγόριθμο Πιλοτικής Αναδιάταξης. Στο Σχήμα 5.2(α) φαίνονται οι αναδιατεταγμένες τιμές των μέσων των συνιστωσών σύμφωνα με τον αλγόριθμο ECR. Στο Σχήμα 5.2(β) παρατίθεται το ιστογράμμο των δεδομένων και οι αντίστοιχες εκτιμηθείσες συναρτήσεις πυκνότητας σύμφωνα με τους τρεις αλγορίθμους αναδιάταξης. Επί πλέον, οι εκτιμήσεις των εκ των υστέρων μέσων τιμών των παραμέτρων του μοντέλου μέσω των αντίστοιχων εργοδικών μέσων της αλυσίδας παρατίθενται στον Πίνακα 5.2. Αρχικά παρατηρούμε ότι ο αλγόριθμος Πιλοτικής Αναδιάταξης οδηγεί σε σοβαρότατες υποεκτιμήσεις και υπερεκτιμήσεις των μέσων της πρώτης και τελευταίας συνιστώσας, κάτι που υποστηρίζεται και από την πολύ κακή προσαρμογή της εκτιμηθείσας συνάρτησης πυκνότητας στο ιστογράμμο των δεδομένων. Αυτό οφείλεται στο γεγονός ότι έχουμε μεγάλο ποσοστό κενών συνιστωσών (πάνω από 30%), με συνέπεια η εκ των υστέρων κατανομή να είναι γνησίως πολυκόρυφη.

Η μέθοδος του Stephens οδηγεί σε σχεδόν ταυτόσημα αποτελέσματα με τον Αλγόριθμο ECR. Όπως βλέπουμε στο Σχήμα 5.2(β) η εκτίμηση της πυκνότητας των δεδομένων βάσει του Αλγορίθμου ECR και του Αλγορίθμου Kullback-Leibler σχεδόν συμπίπτουν. Οι εκτιμήσεις για τους μέσους όπως αναφέρθηκαν από τους Jasra et



Σχήμα 5.3: Γαλαξιακά δεδομένα: Αναδιατεταγμένο δείγμα για τους μέσους των συνιστωσών σύμφωνα με τον Αλγόριθμο ECR, θεωρώντας ένα μοντέλο τριών συνιστωσών t_4 .

al. (2005) είναι

$$\hat{\mathbb{E}}(\boldsymbol{\mu}|\mathbf{x}) = (9.71, 19.01, 19.88, 22.71, 22.86, 32.92).$$

Οι Jasra et al. (2005) εφάρμοσαν τον αλγόριθμο του Stephens στο συγκεκριμένο σύνολο δεδομένων με τις ίδιες υποθέσεις. Σημειώνουμε ότι για λόγους σύγκρισης έχουμε εκτελέσει και εμείς τον αλγόριθμο του Stephens και βρήκαμε ότι όντως όλες οι εκτιμήσεις των παραμέτρων συμφωνούν τόσο με τα αποτελέσματα των Jasra et al. (2005) όσο και με τα αποτελέσματα του Αλγορίθμου ECR που παρατίθενται στον Πίνακα 5.2. Παρ' ότι στο συγκεκριμένο παράδειγμα η αλήθεια για την κατανομή των δεδομένων είναι άγνωστη, το γεγονός ότι οι δύο μέθοδοι δίνουν ουσιαστικά τις ίδιες απαντήσεις αποτελεί μεγάλο πλεονέκτημα για την προτεινόμενη μέθοδο, καθώς το υπολογιστικό κόστος της είναι πολύ μικρότερο. Πιο συγκεκριμένα, ο μέσος απαιτούμενος χρόνος CPU για την ολοκλήρωση των αναδιατάξεων (στο ίδιο δείγμα), σύμφωνα με τον Αλγόριθμο ECR και τον αλγόριθμο του Stephens, εκτιμήθηκε ίσος με 1.56 και 256.68 δευτερόλεπτα, αντίστοιχα.

Στην συνέχεια ακολουθούμε την εναλλακτική προσέγγιση του Stephens (1997a) ο οποίος θέωρησε ότι τα συγκεκριμένα δεδομένα προέρχονται από μία μείξη κατανομών t_4 . Τα αποτελέσματά μας συγκρίνονται με αυτά που παρουσιάστηκαν πρόσφατα από τους Grün and Leisch (2008). Ακολουθώντας αυτούς, θέσαμε τον αριθμό των συνιστωσών ίσο με $k = 3$ και εκτελέσαμε τον δειγματολήπτη Gibbs του Stephens. Οι αναδιατεταγμένες τιμές των μέσων των συνιστωσών απεικονίζονται στο Σχήμα 5.3.

Ας παρατηρήσουμε αρχικά ότι το φαινόμενο label switching έχει εξαλειφθεί επιτυχώς, καθώς οι αναδιατεταγμένες τιμές των προσομοιωμένων μέσων των τριών κατανομών καταλαμβάνουν διαφορετικές περιοχές. Παρ' όλ' αυτά, η εκ των υστέρων κατανομή παρουσιάζει έντονα πολλές γνήσιες (μη συμμετρικές) κορυφές. Αυτό είχε εντοπιστεί τόσο από τον Stephens (1997a), όσο και από τους Grün and Leisch (2008). Όπως παρατηρούμε στο αναδιατεταγμένο δείγμα, οι γνήσιες κορυφές έχουν

	PRA						ECR					
$\widehat{\mathbb{E}}(\boldsymbol{\mu} \boldsymbol{x})$	7.92 (.100)	16.35 (.083)	19.86 (.046)	22.21 (.038)	25.53 (.071)	34.60 (.094)	9.71 (.002)	18.29 (.129)	19.88 (.009)	22.75 (.015)	23.00 (.133)	32.84 (.039)
$\widehat{\mathbb{E}}(\sigma^2 \boldsymbol{x})$	0.70 (.017)	1.19 (.034)	1.40 (.133)	3.17 (.219)	1.94 (.068)	1.87 (.087)	0.57 (.009)	2.15 (.101)	0.79 (.028)	2.63 (.055)	2.10 (.115)	2.05 (.125)
$\widehat{\mathbb{E}}(\boldsymbol{p} \boldsymbol{x})$.081 (.000)	.104 (.004)	.286 (.004)	.307 (.004)	.179 (.005)	.043 (.000)	.090 (.000)	.064 (.003)	.335 (.003)	.387 (.005)	.077 (.003)	.047 (.000)

Πίνακας 5.2: Εργοδικοί μέσοι και οι εκτιμήσεις των τυπικών σφαλμάτων τους για το γαλαξιακό σύνολο δεδομένων με $k = 6$ κανονικές συνιστώσες.

Συστάδα	Βάρος	μ_1	μ_2	μ_3	σ_1^2	σ_2^2	σ_3^2	p_1	p_2	p_3
1	.831 (.053)	9.69 (.002)	19.77 (.035)	22.56 (.009)	0.47 (.010)	0.60 (.029)	4.05 (.016)	.093 (.001)	.319 (.002)	.588 (.002)
2	.169 (.053)	9.70 (.003)	33.06 (.107)	21.28 (.004)	0.78 (.027)	2.14 (.148)	3.53 (.019)	.093 (.000)	.045 (.000)	.862 (.000)

Πίνακας 5.3: Γαλαξιακά δεδομένα: Βάρη και κεντροειδή των δύο συστάδων του αναδιατεταγμένου δείγματος MCMC σύμφωνα με τον αλγόριθμο ECR, για ένα μοντέλο τριών κατανομών t_4 .

καταγραφεί επιτυχώς από την προτεινόμενη μέθοδο αναδιάταξης. Πιο συγκεκριμένα, ο μέσος της πρώτης συνιστώσας κυμαίνεται σε μία σταθερή περιοχή υψηλής εκ των υστέρων πυκνότητας, περίπου στο 9.7, ενώ οι άλλοι δύο μέσοι έχουν τιμές που εναλλάσσονται μεταξύ του 19.8 και 32.8 (δεύτερη συνιστώσα) και μεταξύ του 21.3 και 22.6 (τρίτη συνιστώσα). Αυτά τα αποτελέσματα συμφωνούν πλήρως με εκείνα που παρουσιάστηκαν πρόσφατα από τους Grün and Leisch (2008).

Εφ' όσον η εκ των υστέρων κατανομή φαίνεται να παρουσιάζει δύο διαφορετικές κορυφές, οι Grün and Leisch (2008) συμπεριέλαβαν ένα επί πλέον βήμα στον αλγόριθμό τους όπου κάθε (αναδιατεταγμένη) προσομοιωμένη τιμή αντιστοιχίζεται σε μία από τις κορυφές της κατανομής. Μπορούμε κι εδώ να ακολουθήσουμε ένα παρόμοιο σκεπτικό για να συγκρίνουμε τα αποτελέσματά μας. Έχοντας αναδιατάξει τις τιμές των $(\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ μέσω του Αλγορίθμου ECR, εφαρμόσαμε έναν αλγόριθμο K -means (θεωρώντας δύο συστάδες). Τα αποτελέσματα που πήραμε παρουσιάζονται στον Πίνακα 5.3. Στην κύρια συστάδα (με το μεγαλύτερο βάρος) έχουμε ότι $\hat{\mu}_2 = 19.77$ και $\hat{\mu}_3 = 22.56$, ενώ στην άλλη συστάδα έχουμε ότι $\hat{\mu}_2 = 33.06$ και $\hat{\mu}_3 = 21.28$. Αντίστοιχες αλλαγές καταγράφονται και στις διασπορές και βάρη αυτών των συνιστωσών. Η πρώτη συνιστώσα δεν συμπεριφέρεται διαφορετικά ανάμεσα στις δύο συστάδες. Ως εκ τούτου, οι δύο (γνήσιες) κορυφές της εκ των υστέρων κατανομής των παραμέτρων διαφέρουν ως προς τις παραμέτρους της δεύτερης και τρίτης συνιστώσας. Τέλος, αναφέρουμε ότι και τα αποτελέσματα της ανάλυσης συστάδων βρίσκονται σε πλήρη συμφωνία με εκείνα που παρουσιάστηκαν από τους Grün and Leisch. Λαμβάνοντας

παράμετρος	πραγματική τιμή	KL	ECR	τυπ. σφάλμα
p	(.25, .25, .25, .25)	(.30, .21, .24, .25)	(.30, .21, .24, .25)	$(4, 3, 4, 4) \times 10^{-4}$
μ	(4.5, -2.5)	(4.43, -2.36)	(4.43, -2.36)	(.0014, .0015)
	(-3.0, 4.0)	(-2.91, 4.04)	(-2.91, 4.04)	(.0022, .0012)
	(6.5, 7.0)	(6.73, 7.34)	(6.73, 7.34)	(.0028, .0036)
	(7.0, -3.0)	(6.99, -2.77)	(6.99, -2.77)	(.0036, .0065)
Σ	$\begin{pmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.54 & -0.20 \\ -0.20 & 0.81 \end{pmatrix}$	$\begin{pmatrix} 0.54 & -0.20 \\ -0.20 & 0.81 \end{pmatrix}$	$\begin{pmatrix} .0022 & .0013 \\ .0013 & .0027 \end{pmatrix}$
	$\begin{pmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 1.74 & -0.77 \\ -0.77 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 1.74 & -0.77 \\ -0.77 & 0.69 \end{pmatrix}$	$\begin{pmatrix} .0045 & .0026 \\ .0026 & .0018 \end{pmatrix}$
	$\begin{pmatrix} 4 & 2.5 \\ 2.5 & 4 \end{pmatrix}$	$\begin{pmatrix} 3.30 & 2.14 \\ 2.14 & 4.09 \end{pmatrix}$	$\begin{pmatrix} 3.30 & 2.14 \\ 2.14 & 4.09 \end{pmatrix}$	$\begin{pmatrix} .0071 & .0072 \\ .0072 & .0106 \end{pmatrix}$
	$\begin{pmatrix} 4 & 2.5 \\ 2.5 & 9 \end{pmatrix}$	$\begin{pmatrix} 3.55 & 2.27 \\ 2.27 & 10.13 \end{pmatrix}$	$\begin{pmatrix} 3.55 & 2.27 \\ 2.27 & 10.13 \end{pmatrix}$	$\begin{pmatrix} .0097 & .0119 \\ .0119 & .0384 \end{pmatrix}$

Πίνακας 5.4: Προσομοιωμένο σύνολο δεδομένων της Ενότητας 5.2.3: Εργοδικοί μέσοι των προσομοιωμένων παραμετρικών τιμών, σύμφωνα με την μέθοδο KL του Stephens και τον Αλγόριθμο ECR (10000 επαναλήψεις μετά την διαγραφή των πρώτων 1000 ως burn-in).

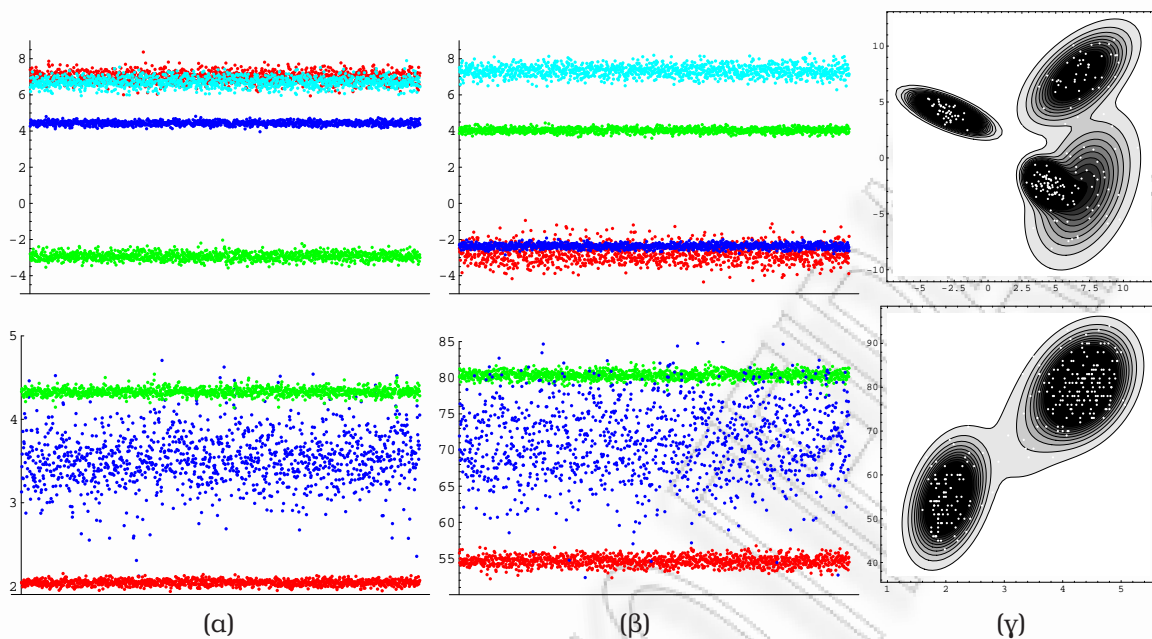
υπ' όψιν το γεγονός ότι η μέθοδος των Grün and Leisch βασίζεται στον χρονοβόρο αλγόριθμο KL του Stephens και επομένως απαιτείται πολύ μεγαλύτερη υπολογιστική προσπάθεια προκειμένου να εφαρμοστεί, γίνεται σαφές ότι η προτεινόμενη μέθοδος αποτελεί μία προτιμότερη λύση.

5.2.3 Μείξεις Πολυμεταβλητών Κανονικών Κατανομών

Σε αυτήν την ενότητα ελέγχουμε την απόδοση της προτεινόμενης μεθόδου σε πολυμεταβλητές περιπτώσεις. Για τον σκοπό αυτό θεωρούμε δεδομένα από μείξεις πολυμεταβλητών κανονικών κατανομών και αρχικά προσομοιώνεται δείγμα από την εκ των υστέρων κατανομή των παραμέτρων μέσω της γενίκευσης του random beta model που προτάθηκε από τους Dellaportas and Papageorgiou (2005), θεωρώντας γνωστό τον αριθμό συνιστωσών. Το πρώτο παράδειγμα βασίζεται σε ένα προσομοιωμένο σύνολο $n = 200$ παρατηρήσεων από τη μείξη διδιάστατων κανονικών κατανομών

$$\sum_{j=1}^4 p_j \mathcal{N}_2(\mu_j, \Sigma_j)$$

με τις πραγματικές τιμές των παραμέτρων που παρατίθενται στον Πίνακα 5.4. Πρόκειται για μία ενδιαφέρουσα περίπτωση αντιμετώπισης του φαινομένου label switching, καθώς οι πραγματικές τιμές των παραμέτρων έχουν επιλεγεί έτσι ώστε να μην υπάρχει κάποιος περιορισμός διάταξης που να διαχωρίζει τις συνιστώσες. Το δεύτερο παράδειγμα είναι το σύνολο δεδομένων Old Faithful το οποίο έχει αναλυθεί από τον Stephens (1997a) αλλά και από τους Dellaportas and Papageorgiou



Σχήμα 5.4: Αναδιατεταγμένες τιμές των (α) μ_{1j} και (β) μ_{2j} , $j = 1, \dots, k$, σύμφωνα με τον αλγόριθμο ECR και (γ) διάγραμμα διασποράς των διδιάστατων παρατηρήσεων μαζί με τις αντίστοιχες εκτιμήσεις των συναρτήσεων πυκνότητας. Πάνω: Προσομοιωμένο σύνολο δεδομένων από την μείξη του Πίνακα 5.4 ($k = 4$). Κάτω: Σύνολο δεδομένων του ηφαιστείου Old Faithful ($k = 3$).

(2005). Πρόκειται για 272 διδιάστατες παρατηρήσεις (x_t, y_t) που σχετίζονται με τη δράση του ηφαιστείου Old Faithful στο Yellowstone National Park του Wyoming στις ΗΠΑ. Τα δεδομένα καταγράφουν τα εξής χαρακτηριστικά: το x_t είναι η διάρκεια της t -οστής έκρηξης του ηφαιστείου, ενώ το y_t είναι ο χρόνος αναμονής μέχρι την έκρηξη $t + 1$. Σύμφωνα με τους Dellaportas and Papageorgίου (2005) ο πιο πιθανός αριθμός συνιστωσών ισούται με $k = 3$.

Στο Σχήμα 5.4(γ) φαίνονται τα διαγράμματα διασποράς των δύο συνόλων δεδομένων. Παρατηρούμε ότι στα προσομοιωμένα δεδομένα δεν φαίνονται ξεκάθαρα τέσσερις διαφορετικές συνιστώσες, συνεπώς η διαδικασία εκτίμησης των παραμέτρων αποτελεί μία προκλητική περίπτωση. Στο ίδιο σχήμα παρατίθενται οι εκτιμήσεις των πυκνοτήτων που προκύπτουν αναδιατάσσοντας τα προσομοιωμένα δείγματα μεγέθους 10000 και 30000 (μετά την περίοδο burn-in), αντίστοιχα, με τον Αλγόριθμο ECR. Επί πλέον, οι αναδιατεταγμένες τιμές των μέσων των συνιστωσών φαίνονται στα Σχήματα 5.4(α) και 5.4(β). Παρατηρούμε ότι τα δείγματα έχουν επιτυχώς αναδιαταχθεί. Οι αντίστοιχες εκτιμήσεις των εκ των υστέρων μέσων τιμών των παραμέτρων παρουσιάζονται στους Πίνακες 5.4 και 5.5.

Για τα προσομοιωμένα δεδομένα οι εκτιμήσεις των εκ των υστέρων μέσων τιμών

παράμετρος	D&P	ECR	τυπ. σφάλμα
p	(.572, .340, .087)	(.590, .351, .059)	(.0052, .0004, .0052)
μ	(4.34, 80.34)	(4.33, 80.33)	(.0019, .0129)
	(2.02, 54.48)	(2.04, 54.63)	(.0006, .0080)
	(3.44, 70.19)	(3.53, 71.33)	(.0198, .3182)
Σ	$\begin{pmatrix} 0.14 & 0.47 \\ 0.47 & 32.86 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0.63 \\ 0.63 & 32.46 \end{pmatrix}$	$\begin{pmatrix} .0010 & .0073 \\ .0073 & .1368 \end{pmatrix}$
	$\begin{pmatrix} 0.06 & 0.32 \\ 0.32 & 34.58 \end{pmatrix}$	$\begin{pmatrix} 0.09 & 0.66 \\ 0.66 & 38.55 \end{pmatrix}$	$\begin{pmatrix} .0003 & .0023 \\ .0023 & .0488 \end{pmatrix}$
	$\begin{pmatrix} 0.29 & 3.32 \\ 3.32 & 85.98 \end{pmatrix}$	$\begin{pmatrix} 0.30 & 1.75 \\ 1.75 & 82.07 \end{pmatrix}$	$\begin{pmatrix} .0820 & .7422 \\ .7422 & 16.66 \end{pmatrix}$

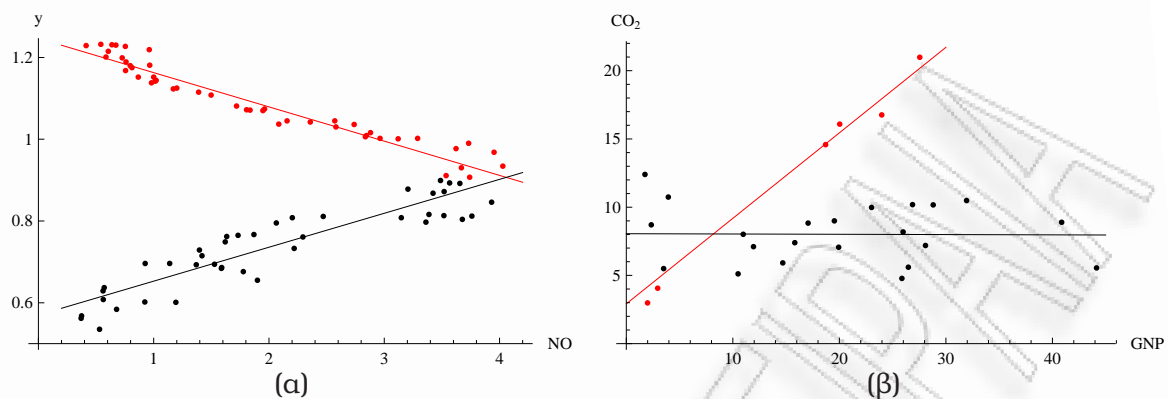
Πίνακας 5.5: Σύνολο δεδομένων Old Faithful: Εκτιμήσεις των εκ των υστέρων μέσω των τιμών των παραμέτρων σύμφωνα με τους Dellaportas and Parageorgiou (2005) και σύμφωνα με τον αλγόριθμο ECR, μαζί με τα εκτιμηθέντα τυπικά σφάλματα (30000 επαναλήψεις μετά το burn-in των πρώτων 20000).

είναι αρκετά κοντά στις πραγματικές τιμές που χρησιμοποιήθηκαν· δες Πίνακα 5.4. Είναι σημαντικό να τονιστεί ότι ένας περιορισμός στους μέσους των συνιστωσών θα αποτύγχανε να απομονώσει τις συμμετρικές περιοχές της εκ των υστέρων κατανομής, κάτι το οποίο φαίνεται από την πρώτη γραμμή του Σχήματος 5.4(α,β). Το ίδιο ισχύει τόσο για τις διασπορές όσο και για τις συνδιασπορές. Επί πλέον, η περιοχή υψηλής εκ των υστέρων πυκνότητας για τα βάρη των συνιστωσών βρίσκεται εκεί όπου όλα τα βάρη είναι σχεδόν ίσα μεταξύ τους, συνεπώς ούτε ένας περιορισμός στα βάρη θα ήταν κατάλληλος για την αντιμετώπιση του φαινομένου label switching. Όπως βλέπουμε στον Πίνακα 5.4, τα αποτελέσματά μας συμφωνούν απολύτως με τα αντίστοιχα του αλγορίθμου KL του Stephens, ο οποίος αποτελεί και την μόνη εναλλακτική λύση για την αντιμετώπιση του φαινομένου label switching στην προκειμένη περίπτωση. Παρ' όλ' αυτά, για άλλη μια φορά βλέπουμε την ανωτερότητα της προτεινόμενης μεθόδου έναντι αυτής του Stephens, καθώς ο εκτιμηθείς μέσος απαιτούμενος χρόνος CPU για την ολοκλήρωση των αναδιατάξεων είναι σαφώς μικρότερος υπό την μέθοδό μας (0.45 έναντι 16.31 δευτερολέπτων, αντίστοιχα).

Για το σύνολο δεδομένων Old Faithful οι εκτιμήσεις μας βρίσκονται σε συμφωνία με αυτές που αναφέρθηκαν από τους Dellaportas and Parageorgiou (2005), οι οποίοι αναδιέταξαν το δείγμα από την εκ των υστέρων κατανομή σύμφωνα με τον περιορισμό

$$\mu_{11}^{(i)} < \mu_{12}^{(i)} < \mu_{13}^{(i)}, \quad i = 1, \dots, M$$

(δες Πίνακα 5.5). Αυτό συμβαίνει εδώ διότι οι προσομοιωμένες τιμές αυτής της παραμέτρου βρίσκονται σε συμμετρικές περιοχές οι οποίες είναι αρκετά μακριά μεταξύ τους, όπως φαίνεται και από τη δεύτερη σειρά του Σχήματος 5.4. Βεβαίως, όπως είδαμε στο προηγούμενο παράδειγμα, τέτοιοι περιορισμοί διάκρισης δύνανται



Σχήμα 5.5: Εκτιμηθείσες ευθείες παλινδρόμησης με $k = 2$ συνιστώσες μαζί με την εκτιμηθείσα κατάταξη κάθε παρατήρησης, αναδιατάσσοντας το προσομοιωμένο δείγμα βάσει του αλγορίθμου ECR. (α) Σύνολο δεδομένων NO. (β) Σύνολο δεδομένων CO₂.

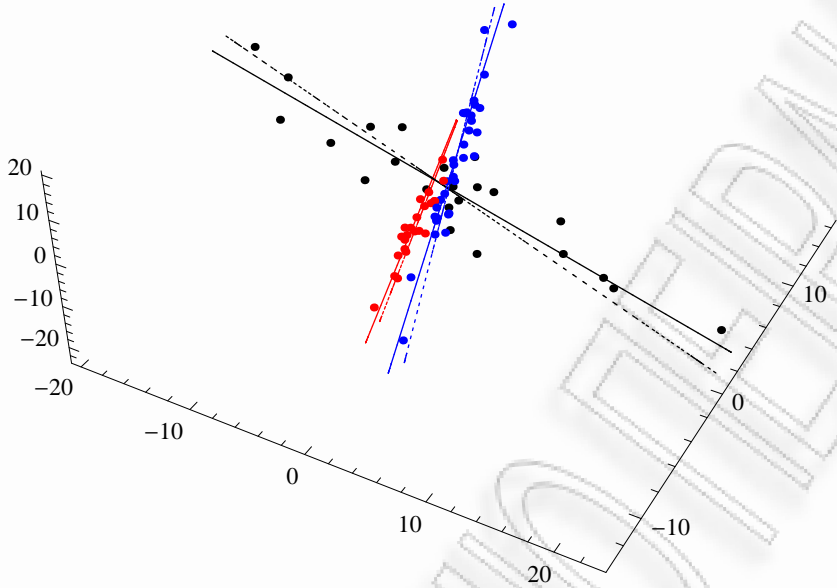
γενικά να αποδειχθούν αρκετά αναποτελεσματικοί.

5.2.4 Μείξεις πολυμεταβλητών παλινδρομήσεων

Σε αυτήν την ενότητα εφαρμόζουμε τον Αλγόριθμο ECR στα παραδείγματα που παρουσιάσαμε στην Ενότητα 2.7.1. Σημειώνεται ότι ο αλγόριθμος εφαρμόστηκε στα ίδια προσομοιωμένα δείγματα των παραμέτρων για κάθε παράδειγμα με αυτά που παρουσιάστηκαν στην Ενότητα 2.7.1.

Για τα σύνολα δεδομένων NO και CO₂ η αναδιάταξη του προσομοιωμένου δείγματος σύμφωνα με τον Αλγόριθμο ECR καταλήγει στα ίδια αποτελέσματα με αυτά που παρουσιάστηκαν στην Ενότητα 2.7.1. Το αποτέλεσμα αυτό είναι αναμενόμενο, καθώς οι συμμετρικές περιοχές είναι αρκετά μακριά μεταξύ τους και κατά συνέπεια ο περιορισμός διάταξης είναι επιτυχής. Βέβαια το προτέρημα που έχει η μέθοδος μας έναντι των περιορισμών είναι ότι δεν χρειάζεται να ψάξουμε την ακολουθία των προσομοιωμένων τιμών από τον δειγματολήπτη ώστε να ανακαλύψουμε κάποια κατάλληλη σχέση διάταξης μεταξύ των παραμέτρων. Οι εκτιμηθείσες ευθείες παλινδρόμησης που αποτελούν την μείξη απεικονίζονται στο Σχήμα 5.5 μαζί με την εκτιμηθείσα κατάταξη κάθε παρατήρησης σε μία από τις δύο συνιστώσες.

Για το παράδειγμα των πολυμεταβλητών παλινδρομήσεων είδαμε ότι δεν υπάρχει κάποιος περιορισμός που να ταυτοποιεί επιτυχώς τις συμμετρικές περιοχές. Πιο συγκεκριμένα, η αναδιάταξη βάσει του περιορισμού στα βάρη δίνει μια πολύ κακή εκτίμηση των μείξεων παλινδρομήσεων. Η αναδιάταξη όμως του προσομοιωμένου



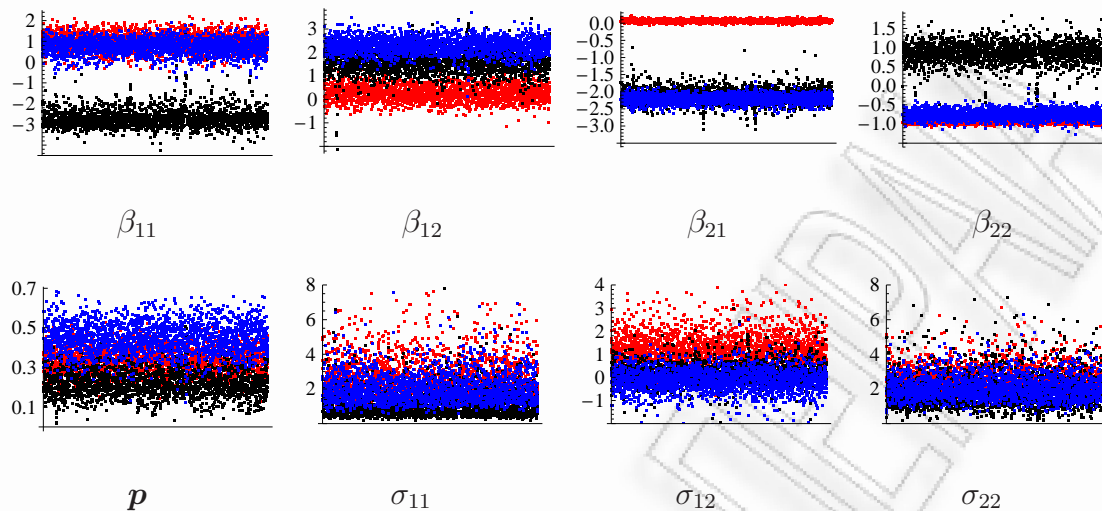
Σχήμα 5.6: Η εκτιμηθείσα μείξη παλινδρομήσεων μαζί με την εκτιμηθείσα κατάταξη κάθε παρατήρησης, αναδιατάσσοντας το προσομοιωμένο δείγμα βάσει του αλγορίθμου ECR. Οι διακεκομμένες γραμμές αποτελούν τις πραγματικές ευθείες παλινδρομήσεων που αποτελούν την μείξη.

δείγματος βάσει του Αλγορίθμου ECR καταλήγει στις εκτιμήσεις

$$\begin{aligned} \hat{\beta}_1 &= \begin{pmatrix} -2.70 & -2.11 \\ 1.58 & 0.84 \end{pmatrix} & \hat{\Sigma}_1 &= \begin{pmatrix} 1.13 & 0.18 \\ 0.18 & 1.85 \end{pmatrix} & \hat{p}_1 &= 0.220 \\ \hat{\beta}_2 &= \begin{pmatrix} 1.01 & 0.07 \\ 0.26 & -0.95 \end{pmatrix} & \hat{\Sigma}_2 &= \begin{pmatrix} 2.66 & 1.30 \\ 1.30 & 2.50 \end{pmatrix} & \hat{p}_2 &= 0.333 \\ \hat{\beta}_3 &= \begin{pmatrix} 0.74 & -2.21 \\ 2.30 & -0.77 \end{pmatrix} & \hat{\Sigma}_3 &= \begin{pmatrix} 2.02 & -0.12 \\ -0.12 & 2.08 \end{pmatrix} & \hat{p}_3 &= 0.448. \end{aligned}$$

Η αντίστοιχη εκτιμηθείσα μείξη παλινδρομήσεων απεικονίζεται στο Σχήμα 5.6. Βλέπουμε ότι αυτή προσαρμόζεται πολύ καλά στα δεδομένα, κάτι που ισχυροποιείται περαιτέρω και από την σύγκριση με τις πραγματικές ευθείες παλινδρομής (διακεκομμένες γραμμές). Στο ίδιο σχήμα καταγράφονται και οι εκτιμηθείσες εκ των υστέρων κατατάξεις των παρατηρήσεων στις συνιστώσες, βάσει του αλγορίθμου ECR. Οι εκτιμήσεις αυτές αντιστοιχούν στις μέγιστες εκ των υστέρων πιθανότητες των (αναδιατεταγμένων) κατατάξεων για κάθε μία παρατήρηση, δηλαδή στις:

$$\hat{z}_i = \operatorname{argmax}_{j \in \{1, \dots, k\}} \hat{P}(z_i = j | \mathbf{y}, \mathbf{x}), \quad i = 1, \dots, n.$$



Σχήμα 5.7: Η ακολουθία των αναδιατεταγμένων τιμών των παραμέτρων μέσω του αλγορίθμου ECR για το παράδειγμα μείζης τριών πολυμεταβλητών παλινδρομήσεων.

Συγκρίνοντας τις εκτιμηθείσες κατατάξεις με τις αντίστοιχες πραγματικές (Σχήμα 2.14) βλέπουμε ότι συμφωνούν αρκετά. Πιο συγκεκριμένα, ο αριθμός των παρατηρήσεων που έχουν την σωστή κατάταξη ισούται με 68 (από τις 80). Οι αντίστοιχες αναδιατεταγμένες τιμές των παραμέτρων παρατίθενται στο Σχήμα 5.7.

5.2.5 Εφαρμογή σε έναν μη διευρυμένο αλγόριθμο

Όπως είδαμε στην Ενότητα 2.5.2, είναι δυνατόν να εκτιμήσουμε τις παραμέτρους μιας μείζης κατανομών χρησιμοποιώντας έναν μη διευρυμένο αλγόριθμο όπως είναι ο αλγόριθμος Metropolis–Hastings. Η μέθοδος μας βασίζεται σε αναδιατάξεις στον χώρο των διανυσμάτων κατάταξης οπότε είναι λογικό να αναρωτηθεί κανείς για τη δυνατότητα εφαρμογής της σε τέτοιου είδους δείγματα. Σε αυτό το παράδειγμα λοιπόν, ασχολούμαστε με την εφαρμογή της προτεινόμενης μεθόδου για την αναδιάταξη ενός προσομοιωμένου δείγματος από την εκ των υστέρων κατανομή ενός μοντέλου μείζης στην περίπτωση όπου αυτό έχει προκύψει χωρίς την προσομοίωση των εικονικών μεταβλητών κατάταξης.

Γενικά, η διαδικασία προσομοίωσης των μεταβλητών κατάταξης μπορεί να γίνει μετά το πέρας των προσομοιώσεων των $(\mathbf{p}, \boldsymbol{\theta})^{(t)}$, $t = 1, \dots, M$. Πράγματι, αν το $\mathbf{z}^{(t)}$ προσομοιώνεται δοθέντων των $(\mathbf{p}, \boldsymbol{\theta})^{(t)}$, τότε το διευρυμένο δείγμα $(\mathbf{z}, \mathbf{p}, \boldsymbol{\theta})^{(t)}$, $t = 1, \dots, M$, έχει ως στόχο πάλι την εκ των υστέρων κατανομή $f(\mathbf{z}, \mathbf{p}, \boldsymbol{\theta} | \mathbf{x})$. Ακολούθως, ο αλγόριθμος ECR μπορεί να εφαρμοστεί στο επαυξημένο δείγμα όπως και πριν.

Μέθοδος MCMC	θ_1	θ_2	θ_3	θ_4	θ_5	p_1	p_2	p_3	p_4	p_5
Metropolis - Hastings	1.17	6.31	8.43	13.66	22.65	.216	.222	.077	.303	.182
Gibbs Sampler	1.17	6.26	8.95	13.58	22.59	.216	.223	.076	.298	.186

Πίνακας 5.6: Εργοδικοί μέσοι των δύο αναδιατεταγμένων δειγμάτων από την εκ των υστέρων κατανομή μέσω του δειγματολήπτη Gibbs και του επαυξημένου αλγορίθμου MH σύμφωνα με τον Αλγόριθμο ECR για τα προσομοιωμένα δεδομένα από τη μείξη Poisson της Ενότητας 5.2.5.

Για τον σκοπό αυτό λοιπόν, θεωρήσαμε ένα προσομοιωμένο σύνολο δεδομένων $n = 250$ παρατηρήσεων μία μείξη πέντε κατανομών Poisson,

$$0.2\mathcal{P}(1) + 0.2\mathcal{P}(6) + 0.2\mathcal{P}(10) + 0.2\mathcal{P}(14) + 0.2\mathcal{P}(22).$$

Εδώ οι παράμετροι του μοντέλου είναι τα βάρη των συνιστωσών $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5)$ και οι μέσες τιμές των κατανομών $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$. Εκ των προτέρων υποθέτουμε ότι $\mathbf{p} \sim \mathcal{D}(1, 1, 1, 1; 1)$ και ότι $\theta_j \sim \mathcal{G}(0.9, 0.1)$, ανεξάρτητα για $j = 1, \dots, 5$. Ακολουθώντας, αναπαραμετροποιούμε το μοντέλο σύμφωνα με την διαδικασία που περιγράφεται στους Marin et al. (2005) και εκτελούμε σε αυτό τον αλγόριθμο Metropolis-Hastings. Σύμφωνα με αυτούς, θέτουμε

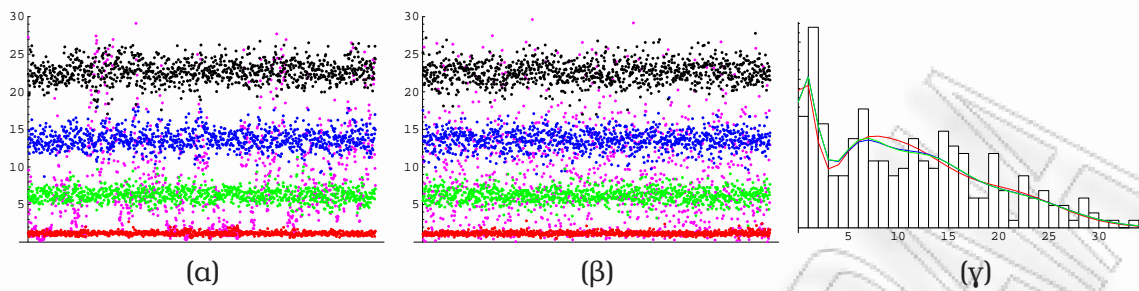
$$p_j = \frac{w_j}{\sum_{\ell=1}^k w_\ell}, \quad w_j > 0,$$

και έτσι ο αλγόριθμος Metropolis-Hastings ανανεώνει τα w_j , $j = 1, \dots, k$ επιλύοντας, το πρόβλημα του περιορισμού των βαρών. Στο βήμα t , $t = 1, 2, \dots$, χρησιμοποιούνται οι κατανομές πρότασης

$$\tilde{\lambda}_j \sim \mathcal{LN}(\log \lambda_j^{(t-1)}, 1) \quad \tilde{w}_j \sim \mathcal{LN}(\log w_j^{(t-1)}, 1).$$

Ο αλγόριθμος εκτελέστηκε για 100000 επαναλήψεις μετά την διαγραφή των πρώτων 20000 τιμών ως περίοδο burn-in. Μετά το πέρας των προσομοιώσεων, παράγουμε τις τιμές των εικονικών μεταβλητών κατάταξης z χρησιμοποιώντας την πλήρη δεσμευμένη κατανομή τους και το διευρυμένο δείγμα αναδιατάσσεται βάσει του Αλγορίθμου ECR. Οι εκτιμήσεις των εκ των υστέρων μέσων τιμών προκύπτουν υπολογίζοντας τους εργοδικούς μέσους του αναδιατεταγμένου δείγματος (δες Πίνακα 5.6). Για λόγους σύγκρισης παραθέτουμε και τις αντίστοιχες εκτιμήσεις που προκύπτουν αναδιατάσσοντας ένα (εξ ορισμού διευρυμένο) δείγμα που προσομοιώθηκε μέσω του δειγματολήπτη Gibbs.

Είναι σαφές ότι οι προκύπτουσες εκτιμήσεις συμφωνούν μεταξύ τους. Αυτό υποστηρίζεται περαιτέρω από το Σχήμα 5.8(γ) όπου φαίνεται ότι οι δύο εκτιμήσεις



Σχήμα 5.8: Αναδιατεταγμένα δείγματα των θ_j , $j = 1, \dots, 5$, τα οποία προσομοιώθηκαν βάσει του διευρυμένου αλγορίθμου MH (α) και του δειγματολήπτη Gibbs (β) σύμφωνα με τον Αλγόριθμο ECR. (γ) Το ιστόγραμμα των δεδομένων μαζί με την πραγματική συνάρτηση μάζας πιθανότητας (κόκκινη γραμμή) και οι εκτιμήσεις αυτής βάσει του Αλγορίθμου ECR, στο αναδιατεταγμένο δείγμα του διευρυμένου αλγορίθμου MH (πράσινη γραμμή) και του δειγματολήπτη Gibbs (μπλέ γραμμή).

της συνάρτησης μάζας πιθανότητας σχεδόν συμπίπτουν, ενώ ταυτόχρονα βρίσκονται αρκετά κοντά στην πραγματική συνάρτηση μάζας πιθανότητας. Στον Πίνακα 5.6 παρατίθενται και οι αναδιατεταγμένες τιμές των μέσων των συνιστωσών, σύμφωνα με τους δύο διαφορετικούς αλγόριθμους προσομοίωσης. Παρατηρούμε στα Σχήματα 5.8(α) και 5.8(β) ότι οι αναδιατεταγμένες τιμές του θ_3 (αντιστοιχεί στο μωβ χρώμα) έχουν σαφώς μεγαλύτερη διασπορά από τις υπόλοιπες. Αυτό συμβαίνει διότι ο αριθμός των παρατηρήσεων που είναι κατεταγμένες σε αυτήν την συνιστώσα είναι συχνά πολύ μικρός σε σχέση με τις υπόλοιπες οπότε η εκ των υστέρων διασπορά αυτής είναι αισθητά μεγαλύτερη. Πράγματι, η εκ των υστέρων διασπορά της συγκεκριμένης παραμέτρου εκτιμήθηκε περίπου είκοσι φορές μεγαλύτερη από την αντίστοιχη των άλλων συνιστωσών. Ίσως εδώ να έχουμε μία ένδειξη ότι το μοντέλο πέντε συνιστωσών να είναι υπερπαραμετροποιημένο για τα συγκεκριμένα δεδομένα, και ένα μοντέλο με λιγότερες συνιστώσες ενδέχεται να προσαρμόζεται καλύτερα (παρ' ό,τι η πραγματική κατανομή έχει πέντε συνιστώσες).

5.3 Εφαρμογές σε Hidden Markov Models

5.3.1 Εισαγωγή

Τα Hidden Markov Models (HMM) αποτελούν γενίκευση των κλασικών μοντέλων μείξεων κατανομών. Είδαμε ότι σε ένα κλασικό μοντέλο μείξης υποτίθεται ανεξαρτησία των παρατηρήσεων. Σε πάρα πολλές εφαρμογές η συγκεκριμένη υπόθεση δεν ισχύει λόγω της φύσης του προβλήματος. Η διαφορά των HMM με τα κλασικά μο-

ντέλα μείξεων έγκειται στο ότι τώρα οι παρατηρήσεις θεωρούνται εξαρτημένες μεταξύ τους. Ειδικότερα, υποθέτουμε ότι οι εικονικές μεταβλητές κατάταξης αποτελούν μία μαρκοβιανή διαδικασία. Η συγκεκριμένη υπόθεση, μαζί με το γεγονός ότι οι εικονικές μεταβλητές δεν είναι παρατηρήσιμες δικαιολογεί και την ονομασία των μοντέλων αυτών.

Ένα τυπικό HMM αποτελείται από μια διδιάστατη διαδικασία $(z_t, y_t)_{t \in \mathbb{N}}$, όπου η $(z_t)_{t \in \mathbb{N}}$ είναι μία μη παρατηρήσιμη ομογενής μαρκοβιανή αλυσίδα σε χώρο $k \geq 2$ καταστάσεων. Δοθείσης της $(z_t)_{t \in \mathbb{N}}$, θεωρούμε ότι τα $(y_t)_{t \in \mathbb{N}}$ αποτελούν μία πραγματοποίηση μιας ακολουθίας τυχαίων μεταβλητών οι οποίες μπορεί να είναι είτε ανεξάρτητες είτε όχι. Τα HMM παρέχουν την κατάλληλη μεθοδολογία για την μοντελοποίηση ετερογενών δεδομένων από ένα ευρύ φάσμα εφαρμογών, που δεν ισχύει η υπόθεση ανεξαρτησίας των παρατηρήσεων. Έχουν αποδειχθεί πολύ χρήσιμα σε αρκετούς τομείς όπως είναι η Οικονομία, Τηλεπικοινωνίες, Γενετική, Μετεωρολογία, Ανάλυση Ψηφιακών εικόνων, Ψυχολογία κ.α. Ο αναγνώστης παραπέμπεται στους MacDonald and Zucchini (1997) και στους Carré et al. (2004).

Οι πρώτες εφαρμογές Μπεϋζιανής Συμπερασματολογίας σε HMM με γνωστό αριθμό καταστάσεων αναπτύχθηκαν από τους Albert and Chib (1993), Robert et al. (1993) και Chib (1996). Αργότερα, το πρόβλημα αντιμετωπίστηκε στο γενικότερο πλαίσιο αγνώστου αριθμού καταστάσεων με πιο σημαντικές εργασίες αυτές των Robert et al. (2001) και Green and Richardson (2002). Ανεξαρτήτως αν το k θεωρείται γνωστό ή όχι, υπάρχει ένα κοινό πρόβλημα στην συμπερασματολογία στην περίπτωση όπου η εκ των πληροφορία δεν διακρίνει τις καταστάσεις μεταξύ τους: Όπως και στα κλασικά μοντέλα μείξεων, η εκ των υστέρων κατανομή των παραμέτρων παρουσιάζει ένα πολλαπλάσιο $k!$ συμμετρικών κορυφών, όπου το k είναι ο αριθμός των καταστάσεων της μη παρατηρήσιμης μαρκοβιανής διαδικασίας. Επομένως, η διαδικασία εκτίμησης των παραμέτρων ενός HMM μέσω της τυπικής διαδικασίας των εργοδικών μέσων ενός προσομοιωμένου δείγματος MCMC δεν οδηγεί σε σωστές εκτιμήσεις, ως συνέπεια του φαινομένου label switching.

Οι υπάρχουσες μέθοδοι για την αντιμετώπιση του φαινομένου label switching στα HMM έχουν να κάνουν στην συντριπτική τους πλειοψηφία με την χρήση εικονικών περιορισμών. Οι Robert et al. (2001) εφάρμοσαν στο προσομοιωμένο δείγμα τη μέθοδο απόρριψης βάσει της οποίας απορρίπτονται όλες οι προτεινόμενες παραμετρικές τιμές αν δεν ικανοποιούν έναν προεπιλεγμένο περιορισμό. Οι Green and Richardson (2002) πρότειναν την χρήση της μεθόδου partial reordering έτσι ώστε να εξασφαλίζουν ότι θα προτείνονται πάντα τιμές που ικανοποιούν τον προεπιλεγ-

μένο περιορισμό, αποφεύγοντας έτσι τις απορρίψεις τιμών. Η Frühwirth-Schnatter (2001) αντιμετώπισε το φαινόμενο label switching μέσω του δειγματολήπτη περιορισμένης μετάθεσης, αφού πρώτα είχε προσομοιώσει ένα δείγμα με τον δειγματολήπτη τυχαίας μετάθεσης (δες Ενότητα 2.6.1). Πρόσφατα, ο Sprezia (2009) σύγκρινε όλες τις προαναφερθείσες μεθόδους στην περίπτωση αγνώστου αριθμού καταστάσεων της αλυσίδας. Το συμπέρασμα ήταν ότι όλες οι μέθοδοι δουλεύουν καλά στην περίπτωση όπου οι συμμετρικές περιοχές της εκ των υστέρων κατανομής είναι «επαρκώς μακριά» μεταξύ τους. Παρ' όλ' αυτά θα δούμε ότι κάτι τέτοιο δεν ισχύει πάντα.

Αρχικά θα διαπιστώσουμε ότι ο Αλγόριθμος ECR για την αντιμετώπιση του φαινομένου label switching μπορεί να εφαρμοστεί και στο πλαίσιο των HMM. Τόσο οι θεωρητικές ιδιότητες του αλγορίθμου όσο και η πρακτική υλοποίησή του παραμένουν ίδιες. Χρησιμοποιώντας αρκετά παραδείγματα δείχνουμε ότι το φαινόμενο label switching μπορεί να λυθεί επιτυχώς στα HMM με μικρό υπολογιστικό κόστος. Η μέθοδός μας αποδίδει εξ ίσου καλά ακόμα και στην περίπτωση όπου η εκ των υστέρων κατανομή παρουσιάζει αρκετές γνήσιες κορυφές ή όταν δεν υπάρχει ένας περιορισμός που να απομονώνει επαρκώς τις συμμετρικές περιοχές. Τα αποτελέσματα συγκρίνονται με τα αντίστοιχα που προκύπτουν αναδιατάσσοντας το προσομοιωμένο δείγμα σύμφωνα με κάποιον εικονικό περιορισμό διάταξης που όπως αναφέρθηκε είναι η πιο διαδεδομένη διαδικασία αντιμετώπισης του φαινομένου label switching ως τώρα. Οι εφαρμογές που θα παρουσιαστούν καλύπτουν διάφορους τομείς όπως η Οικονομία, Μετεωρολογία και Βιολογία.

5.3.2 Συμβολισμοί και το μπεϋζιανό μοντέλο

Έστω $\mathbf{y} = (y_1, \dots, y_n)$ μία ακολουθία μονοδιάστατων ή πολυδιάστατων παρατηρήσεων που παίρνουν τιμές στο διακριτό ή συνεχή χώρο \mathcal{Y} . Εκτός των \mathbf{y} , ενδέχεται να παρατηρούμε και άλλη μία ακολουθία $\mathbf{x} = (x_1, \dots, x_n)$ μονοδιάστατων ή πολυδιάστατων μεταβλητών. Υποθέτουμε ότι η κατανομή του \mathbf{y} εξαρτάται από μία ακολουθία $\mathbf{z} = (z_1, \dots, z_n)$ εικονικών (μη παρατηρήσιμων) μεταβλητών με τιμές στο πεπερασμένο σύνολο $\{1, \dots, k\}$, όπου το $k \geq 2$ είναι μια γνωστή σταθερά. Οι μεταβλητές z_t , $t = 1, \dots, n$, αποτελούν μία ομογενή μαρκοβιανή αλυσίδα με k καταστάσεις. Συμβολίζουμε με $\mathbf{P} = (p_{ij})$, $i, j = 1, \dots, k$, τον πίνακα μετάβασης της αλυσίδας, δηλαδή

$$p_{ij} := P(z_t = j | z_{t-1} = i).$$

Η ονομασία των hidden Markov models δικαιολογείται από το γεγονός ότι η μαρκοβιανή διαδικασία $(z_t)_{t=1, \dots, n}$ είναι μη παρατηρήσιμη. Ας σημειωθεί εδώ ότι αν τα

ζευγάρια $(z_1, y_1), \dots, (z_n, y_n)$ ήταν ανεξάρτητα και δεν υπήρχαν συμμεταβλητές \mathbf{x} τότε το μοντέλο καταλήγει σε μία κλασσική μείζη κατανομών. Υπό αυτήν την έννοια, τα HMM θεωρούνται γενίκευση των μοντέλων μείζων.

Δοθεισών των εικονικών μεταβλητών z_t , $t = 1, \dots, n$, και παραμέτρων $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ οι οποίες αντιστοιχούν σε όλες τις πιθανές καταστάσεις, η δεσμευμένη πυκνότητα του \mathbf{y} γράφεται ως

$$f(y_1, \dots, y_n | y_{1-s:0}, \mathbf{z}, \mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^n f(y_t | y_{t-s:t-1}, x_t, z_t, \theta_{z_t}).$$

Εδώ χρησιμοποιούμε τον συμβολισμό $y_{t-s:t-1} = (y_{t-s}, \dots, y_{t-1})$. Η περίπτωση $s = 0$ καταλήγει σε δεσμευμένη ανεξαρτησία των παρατηρήσεων. Όταν όμως $s > 0$, το μοντέλο ορίζεται δεσμεύοντας στις πρώτες s παρατηρήσεις, όπου για χάρην ευκολίας συμβολίζονται με y_{1-s}, \dots, y_0 . Επιπροσθέτως υποθέτουμε ότι η $f(y_t | \cdot)$ ανήκει σε μια παραμετρική οικογένεια κατανομών στο χώρο \mathcal{Y} ως προς κάποιο κατάλληλο μέτρο.

Θεωρώντας ότι η μη παρατηρήσιμη μαρκοβιανή αλυσίδα (z_t) βρίσκεται σε στασιμότητα, η από κοινού κατανομή των πλήρων δεδομένων είναι η

$$f(\mathbf{z}, \mathbf{y} | \mathbf{P}, \mathbf{x}, \boldsymbol{\theta}) = \pi_{z_1} f(y_1 | y_{t-s:t-1}, x_1, z_1, \theta_{z_1}) \prod_{t=2}^n p_{z_{t-1}, z_t} f(y_t | y_{t-s:t-1}, x_t, z_t, \theta_{z_t}), \quad (5.3)$$

όπου με π_{z_1} συμβολίζεται η στάσιμη πιθανότητα της κατάστασης z_1 , δηλαδή, το z_1 -οστό στοιχείο του αριστερού ιδιοδιανύσματος του πίνακα μετάβασης \mathbf{P} το οποίο αντιστοιχεί στην ιδιοτιμή 1. Η συνηθισμένη επιλογή της εκ των προτέρων κατανομής του πίνακα μετάβασης είναι ένα γινόμενο κατανομών Dirichlet. Πιο συγκεκριμένα, υποτίθεται ότι οι γραμμές του \mathbf{P} είναι ανεξάρτητες με

$$p_i \equiv (p_{i1}, \dots, p_{ik}) \sim \mathcal{D}(a_{i1}, \dots, a_{ik}), \quad i = 1, \dots, k,$$

όπου $a_{ii} \equiv a_p$, $i = 1, \dots, k$, και $a_{ij} \equiv a_q$, $\forall i, j \in \{1, \dots, k; i \neq j\}$ και a_p και a_q είναι προκαθορισμένοι θετικοί αριθμοί. Τότε η εκ των προτέρων κατανομή του πίνακα μετάβασης είναι αναλλοίωτη ως προς τις μεταθέσεις των δεικτών των καταστάσεων. Οι Robert et al. (1993, 2000) και Spezia (2009) θεωρούν $a_p = a_q = 1$ ώστε η κατανομή να είναι μη πληροφοριακή, ενώ η Frühwirth-Schnatter (2001) θέτει $a_p = 2$ και $a_q = 1/(k-1)$ ώστε να εκφράσει την εκ των προτέρων πεποίθηση ότι η πιθανότητα παραμονής σε κάθε κατάσταση είναι μεγαλύτερη από την πιθανότητα μετάβασης σε κάποια άλλη. Υπό αυτές τις υποθέσεις, η εκ των προτέρων κατανομή του πίνακα μετάβασης είναι συζυγής της δεσμευμένης εκ των υστέρων κατανομής

και έτσι προκύπτει η πλήρης δεσμευμένη εκ των υστέρων κατανομή

$$p_i | \mathbf{z} \sim \mathcal{D}(a_{ii} + n_{i1}, \dots, a_{ik} + n_{ik}), \quad i = 1, \dots, k, \quad (5.4)$$

με $n_{ij} = \sum_{t=2}^n I(z_{t-1} = i, z_t = j)$ να δηλώνει τον αριθμό μεταβάσεων από την κατάσταση i στην j .

Η μοναδική εκ των προτέρων κατανομή που εξαρτάται από τον ορισμό του εκάστοτε μοντέλου είναι αυτή των παραμέτρων των καταστάσεων θ . Σύμφωνα με τους Diebolt and Robert (1994), δεν είναι δυνατόν να είμαστε τελειώς μη πληροφοριακοί για το θ , διότι καταχρηστικές εκ των προτέρων κατανομές στο θ καταλήγουν σε καταχρηστικές εκ των υστέρων κατανομές. Συνήθως λοιπόν τα $\theta_1, \dots, \theta_k$ υποτίθενται εκ των προτέρων ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από κάποια κατανομή

$$\pi(\theta | \phi) = \pi(\theta_1 | \phi) \dots \pi(\theta_k | \phi), \quad (5.5)$$

όπου ϕ είναι μία υπερπαραμέτρος κοινή για όλες τις καταστάσεις και αυτή με τη σειρά της μπορεί να είναι είτε γνωστή είτε τυχαία μεταβλητή. Σε πολλές περιπτώσεις η εκ των προτέρων κατανομή $\pi(\theta | \phi)$ επιλέγεται συζυγής της πλήρους πιθανοφάνειας. Τότε είναι διαθέσιμες οι πλήρεις δεσμευμένες κατανομές των θ_i , οπότε η προσομοίωση από την κατανομή-στόχο γίνεται μέσω του δειγματολήπτη Gibbs. Παρ' όλ' αυτά, κάτι τέτοιο μπορεί να μην ισχύει πάντα, οπότε απαιτούνται πιο εξειδικευμένες μέθοδοι, όπως ο slice sampler των Damien et al. (1999). Τέλος, τονίζουμε το γεγονός ότι δεν υποθέτουμε κάποια εκ των προτέρων διάταξη των παραμέτρων του μοντέλου και έτσι η εκ των προτέρων κατανομή είναι επίσης αναλλοίωτη ως προς τις μεταθέσεις των δεικτών των καταστάσεων.

Ο αλγόριθμος MCMC ολοκληρώνεται με την ανανέωση των μεταβλητών (z_1, \dots, z_n) . Μία βασική διαφορά με το κλασικό μοντέλο μείζωνων κατανομών είναι ότι στα HMM δεν είναι άμεση η χρήση της τεχνικής διεύρυνσης των δεδομένων (Tanner and Wong, 1987), δηλαδή η προσομοίωση από τις κατανομές

$$z_t | \mathbf{y}, z_{t-1}, \dots, z_1, \theta, \mathbf{P},$$

διότι η πολυπλοκότητα αυξάνεται με το t . Οι Robert et al. (1993) χρησιμοποιούν στη θέση αυτής την πλήρη δεσμευμένη κατανομή κάθε μιας μεταβλητής και χρησιμοποιούν τον δειγματολήπτη Gibbs. Αργότερα, ο Chib (1996) πρότεινε την ταυτόχρονη ανανέωση της ακολουθίας των εικονικών μεταβλητών μέσω της τεχνικής "forward filtering-backward sampling". Εμείς θα ακολουθήσουμε την μέθοδο των Robert et al. (1993) και κάθε εικονική μεταβλητή θα προσομοιώνεται από την πλήρη

δεσμευμένη κατανομή

$$P(z_t = i | \dots) \propto p_{z_{t-1}, i} f(y_t | \theta_i) p_{i, z_{t+1}}.$$

Στην παραπάνω έκφραση, για $t = 1$ ο πρώτος όρος αντικαθίσταται με την εκ των υστέρων στάσιμη πιθανότητα π_i (δηλαδή την i -συντεταγμένη του αριστερού ιδιοδιάνυσματος του πίνακα μετάβασης \mathbf{P}), ενώ για $t = n$ ο τελευταίος όρος αντικαθίσταται με τη μονάδα.

5.3.3 Το φαινόμενο label switching στα HMM

Έστω \mathcal{T}_k το σύνολο μεταθέσεων των καταστάσεων $\{1, \dots, k\}$. Για κάποια μετάθεση $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$ θεωρούμε την αντίστοιχη μετάθεση του παραμετρικού διανύσματος $\tau(\mathbf{P}, \boldsymbol{\theta})$, όπου $\tau\boldsymbol{\theta} = (\theta_{t_1}, \dots, \theta_{t_k})$ και $\tau\mathbf{P} = (p_{t_i, t_j})$. Όπως στα κλασσικά μοντέλα μείξης έτσι και στα HMM, η πιθανοφάνεια

$$\begin{aligned} L(\mathbf{P}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}) &= \sum_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{z}, \mathbf{y} | \mathbf{P}, \boldsymbol{\theta}, \mathbf{x}) \\ &= \sum_{\mathbf{z} \in \mathcal{Z}} \pi_{z_1} f(y_1 | y_{t-1:s:t-1}, x_1, z_1, \theta_{z_1}) \prod_{t=2}^n p_{z_{t-1}, z_t} f(y_t | y_{t-s:t-1}, x_t, z_t, \theta_{z_t}) \end{aligned} \quad (5.6)$$

είναι αναλλοίωτη ως προς τους δείκτες των καταστάσεων καθώς έχουμε $L(\mathbf{P}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}) = L(\tau(\mathbf{P}, \boldsymbol{\theta}) | \mathbf{y}, \mathbf{x}), \forall \tau \in \mathcal{T}_k$. Συνεπώς, αν η εκ των προτέρων κατανομή είναι επίσης συμμετρική ως προς τις μεταθέσεις των δεικτών των συνιστωσών, έχουμε ότι και η εκ των υστέρων κατανομή των παραμέτρων θα είναι συμμετρική, καθώς θα είναι ανάλογη με το γινόμενο μιας συμμετρικής πιθανοφάνειας με την συμμετρική εκ των προτέρων κατανομή. Αυτό οδηγεί στο ότι οι παράμετροι του μοντέλου δεν είναι περιθωριακά διακριτές μεταξύ τους, καθώς οι αντίστοιχες περιθωριακές εκ των υστέρων κατανομές είναι ίδιες. Αυτή η ιδιότητα λοιπόν θα διατηρείται και σε αλγορίθμους που προσομοιώνουν ένα δείγμα από την εκ των υστέρων κατανομή με αποτέλεσμα να μην μπορούμε να εκτιμήσουμε σωστά τις περιθωριακές κατανομές. Για παράδειγμα, οι εργοδικοί μέσοι θα προκύπτουν ίσοι μεταξύ τους και έτσι δεν θα καταλήγουν σε λογικές εκτιμήσεις των αντίστοιχων εκ των υστέρων μέσων τιμών.

Όπως αναφέραμε, η πλειοψηφία της βιβλιογραφίας στα HMM βασίζεται σε μεθόδους όπου αποτρέπουν το φαινόμενο label switching και αυτό συνήθως κάνει πολύ αργή την σύγκλιση στην εκ των υστέρων κατανομή. Ταυτόχρονα αυτές οι μέθοδοι βασίζονται στην χρήση εικονικών περιορισμών διάταξης των παραμέτρων που συχνά

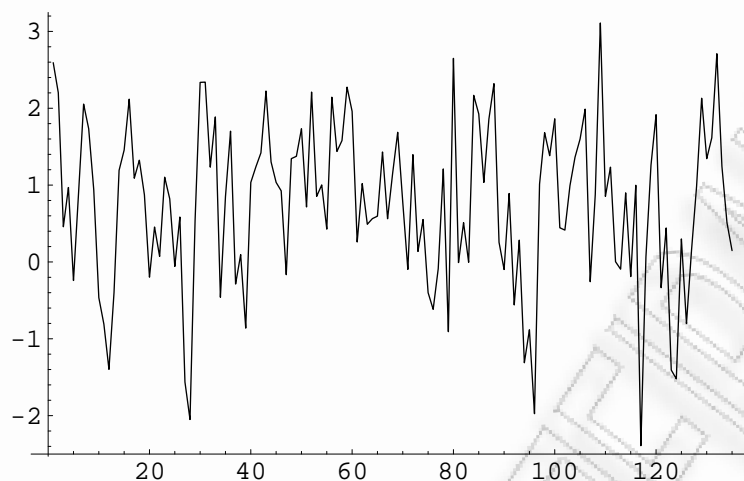
δεν είναι ικανοί να απομονώσουν τις συμμετρικές περιοχές της εκ των υστέρων κατανομής, παρά μόνον σε πολύ απλές περιπτώσεις. Επί πλέον, όπως ήδη είδαμε, όταν υπάρχουν περισσότερες από μία γνήσιες κορυφές στην εκ των υστέρων κατανομή, τότε ένας περιορισμός διάκρισης αγνοεί τις μικρότερες κορυφές. Παρ' ό,τι είναι διαθέσιμες εκλεπτυσμένες μέθοδοι για την αντιμετώπιση του φαινομένου label switching, όπως ο αλγόριθμος του Stephens (2000a) ή η αναδιάταξη μέσω συναρτήσεων ζημίας των Celeux et al. (2000), το μεγάλο υπολογιστικό κόστος αυτών των μεθόδων κάνει σχεδόν απαγορευτική την χρήση τους. Αυτό υποστηρίζεται περαιτέρω και από το γεγονός ότι δεν έχουν εφαρμοστεί ποτέ σε HMM.

Πριν προχωρήσουμε στις εφαρμογές, θα πρέπει να επαληθεύσουμε τις ιδιότητες της Ενότητας 4.2 ώστε να εξασφαλιστούν τα θεωρητικά αποτελέσματα του προηγούμενου κεφαλαίου. Αρχικά, ας παρατηρήσουμε ότι από την συζήτηση που προηγήθηκε, έχουμε $\boldsymbol{\eta} = ((p_1, \theta_1), \dots, (p_k, \theta_k))$ (υπενθυμίζουμε ότι εδώ το p_i συμβολίζει την i -γραμμή του πίνακα μετάβασης), συνεπώς $\mathcal{H} = (\mathcal{P}_k \times \Theta)^k$, και επομένως για κάθε $\tau \in \mathcal{T}_k$ εξασφαλίζεται η ισχύς της (4.2). Ακολουθώντας, οι εκ των προτέρων υποθέσεις (5.4), (5.5) εξασφαλίζουν την ισχύ της ιδιότητας (4.3). Η σχέση (5.6) είναι η (4.5) στην περίπτωση των HMM. Τέλος, είναι εύκολο να δούμε ότι η απο κοινού συνάρτηση πυκνότητας των πλήρων δεδομένων έχει πράγματι την ιδιότητα (4.4). Από την (5.3), λαμβάνουμε διαδοχικά

$$\begin{aligned}
 f(\mathbf{y}, \tau \mathbf{z} | \boldsymbol{\eta}) &= f(\mathbf{y}, \tau \mathbf{z} | \mathbf{P}, \boldsymbol{\theta}) \\
 &= \pi_{t_{z_1}} \prod_{i=2}^n p_{t_{z_{i-1}} t_{z_i}} f(y_i | \theta_{t_{z_i}}) \\
 &= (\tau \boldsymbol{\pi})_{z_1} \prod_{i=2}^n (\tau \mathbf{P})_{z_{i-1} z_i} f(y_i | \tau \boldsymbol{\theta}_{z_i}) \\
 &= f(\mathbf{y}, \mathbf{z} | \tau(\mathbf{P}, \boldsymbol{\theta})) \\
 &= f(\mathbf{y}, \mathbf{z} | \tau \boldsymbol{\eta}), \quad \forall \tau = (t_1, \dots, t_k) \in \mathcal{T}_k.
 \end{aligned}$$

5.3.4 Αυτοπαλίνδρομο Μοντέλο Μαρκοβιανών Εναλλαγών

Τα αυτοπαλίνδρομα μοντέλα μαρκοβιανών εναλλαγών (Markovian Switching Autoregressive (MSAR) models) εισήχθησαν από τους Shumway and Stoffer (1991). Θα δούμε ότι σε ένα μπεϋζιανό πλαίσιο αποτελούν ένα τυπικό παράδειγμα εμφάνισης του φαινομένου label switching. Σε ένα τέτοιο μοντέλο υποθέτουμε ότι, δοθισών των μη παρατηρήσιμων μεταβλητών κατατάξης $z_t = j, j \in \{1, \dots, k\}$, οι παρατηρήσεις



Σχήμα 5.9: Ποσοστιαία μεταβολή του τετραμηνιαίου ακαθάριστου εθνικού προϊόντος των ΗΠΑ από το δεύτερο τετράμηνο του 1951 έως το τελευταίο τετράμηνο του 1984.

$y_t, t = 1, \dots, n$, προκύπτουν από το αυτοπαλίνδρομο μοντέλο

$$y_t = \beta_{j0} + \sum_{i=1}^s \beta_{ji} y_{t-i} + \epsilon_t, \quad t = 1, \dots, n,$$

όπου $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$. Έτσι, δεσμεύοντας στη μη παρατηρηθείσα διαδικασία, έχουμε ότι

$$y_t | z_t = j, y_{t-1}, \dots, y_{t-s} \sim \mathcal{N}(\mu_{jt}, \sigma_j^2),$$

με $\mu_{jt} = \beta_{j0} + \sum_{i=1}^s \beta_{ji} y_{t-i}$. Παρατηρούμε επίσης ότι το μοντέλο ορίζεται δοθεισών των πρώτων s παρατηρήσεων y_{1-s}, \dots, y_0 . Θεωρούμε ότι η υποδιαδικασία Z_1, \dots, Z_n αποτελεί μία μαρκοβιανή αλυσίδα k καταστάσεων και πίνακα μεταβασης $\mathbf{P} = (p_{ij})$, $i, j = 1, \dots, k$. Θα αναφερόμαστε σε τέτοιου είδους μοντέλα με τον όρο MSAR(k, s), όπου το k δηλώνει τον αριθμό των καταστάσεων και το s τη χρονική υστέρηση του μοντέλου. Η παράμετρος του μοντέλου MSAR(k, s) είναι το διάνυσμα $\theta = (\mathbf{P}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$, όπου

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{10} & \beta_{11} & \dots & \beta_{1s} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k0} & \beta_{k1} & \dots & \beta_{ks} \end{pmatrix} \quad \text{και} \quad \boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2).$$

Ένα χαρακτηριστικό παράδειγμα εφαρμογής τέτοιου είδους μοντέλων αποτελεί η χρονοσειρά GNP (Growth National Product) (δες Hamilton, 1989), η οποία αποτελείται από δεδομένα σχετικά με την ποσοστιαία μεταβολή του τετραμηνιαίου

ακαθάριστου εθνικού προϊόντος των ΗΠΑ. Πρόκειται για $n = 135$ παρατηρήσεις της μορφής

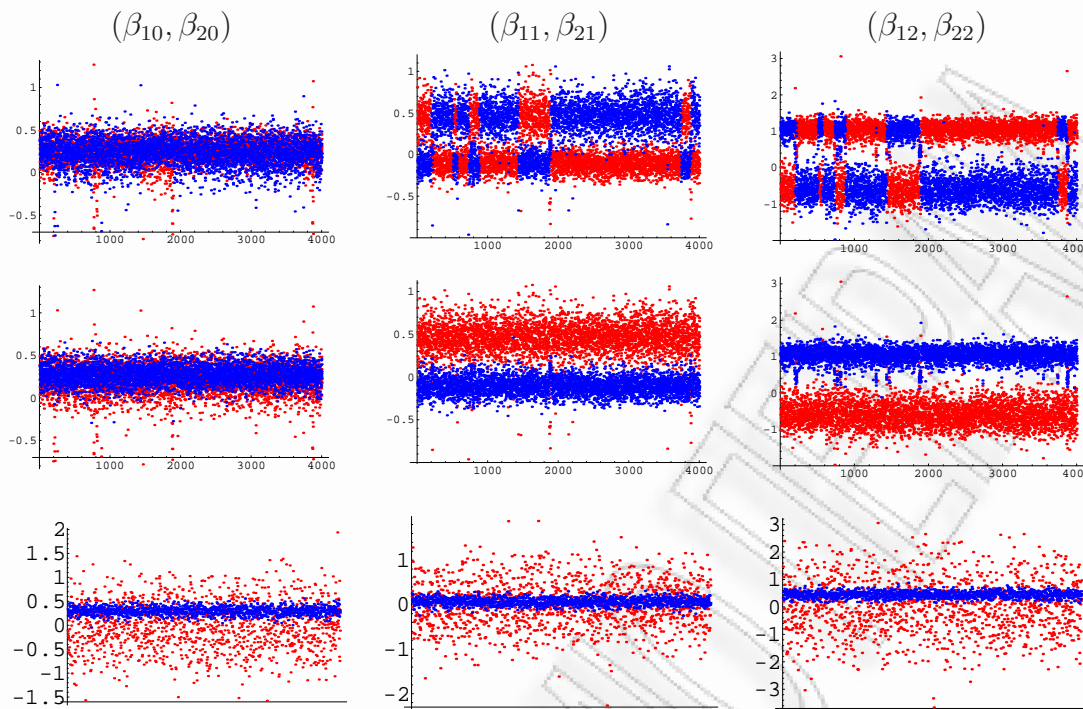
$$y_t = 100\{\log(GNP_t) - \log(GNP_{t-1})\}, \quad t = 1, \dots, 135,$$

όπου το ακαθάριστο εθνικό προϊόν κατά το t -οστό τετράμηνο συμβολίζεται με GNP_t . Η χρονοσειρά απεικονίζεται στο Σχήμα 5.9. Ο Chib (1996) μοντελοποίησε τα συγκεκριμένα δεδομένα ως ένα μοντέλο MSAR(k, s), θεωρώντας $k = 4, s = 4$ και που η μόνη διαφοροποίηση ως προς $j = 1, 2, 3, 4$ εντοπίζεται στον σταθερό όρο β_{j0} . Αργότερα η Frühwirth-Schnatter (2001) χρησιμοποίησε $k = 2, s = 2$ σε ένα μοντέλο όπου όλοι οι αυτοπαλίνδρομοι συντελεστές όπως και οι διασπορές των σφαλμάτων στις διαφορετικές καταστάσεις επιτρέπεται να πάρουν διαφορετικές τιμές. Σημειώνουμε τέλος ότι ο Chib χρησιμοποίησε διαφορετικές εκ των προτέρων κατανομές για τις καταστάσεις της μη παρατηρήσιμης διαδικασίας με σκοπό να αποτρέψει την παρουσία του φαινομένου label switching. Παρ' ολ' αυτά, δεν αρκεί κάτι τέτοιο για την αντιμετώπιση του προβλήματος, καθώς η πιθανοφάνεια παραμένει συμμετρική και αυτό επηρεάζει την αντίστοιχη εκ των υστέρων κατανομή. Έτσι, οι εκτιμήσεις του Chib κρίνονται αναξιόπιστες (δες και Frühwirth-Schnatter, 2001). Εμείς χρησιμοποιούμε το μοντέλο της Frühwirth-Schnatter, δηλαδή ένα μοντέλο MSAR(2, 2) με εκ των προτέρων κατανομές

$$\begin{aligned} (\beta_{j0}, \dots, \beta_{js})^T &\sim \mathcal{N}_{s+1}(b_0, B_0) \\ \sigma_j^2 &\sim \mathcal{IG}(\nu_0, D_0), \end{aligned}$$

ανεξάρτητα για $j = 1, \dots, k$. Με τις συγκεκριμένες υποθέσεις προκύπτουν συζυγείς πλήρεις δεσμευμένες εκ των υστέρων κατανομές. Για περισσότερες λεπτομέρειες ο αναγνώστης παραπέμπεται στην Frühwirth-Schnatter (2001).

Εκτελέσαμε τον δειγματολήπτη Gibbs που περιγράφεται στην Frühwirth-Schnatter (2001) αλλά χωρίς εκ των προτέρων περιορισμούς στις παραμέτρους του μοντέλου. Ύστερα από 80000 επαναλήψεις (μετά το burn-in των πρώτων 20000), ο αλγόριθμος παράγαγε τις προσομοιωμένες τιμές των αυτοπαλίνδρομων συντελεστών που φαίνονται στην πρώτη γραμμή του Σχήματος 5.10 (στο σχήμα απεικονίζεται κάθε 20η επανάληψη). Βλέπουμε ότι καταγράφονται πολλές εναλλαγές των προσομοιωμένων τιμών μεταξύ των δύο καταστάσεων της αλυσίδας. Επί πλέον, παρατηρούμε ότι οι συντελεστές (β_{10}, β_{20}) έχουν σχεδόν την ίδια περιοχή υψηλής εκ των υστέρων πυκνότητας. Επίσης φαίνεται ότι οι τιμές του δεύτερου και τρίτου αυτοπαλίνδρομου συντελεστή ανάμεσα στις δύο καταστάσεις της κρυμμένης διαδικασίας ικανοποιούν κάποιον περιορισμό διάταξης. Η Frühwirth-Schnatter (2001) εκμεταλλεύτηκε το



Σχήμα 5.10: Πάνω: Οι προσομοιωμένες τιμές των αυτοπαλίνδρομων συντελεστών. Μέση: Αναδιατεταγμένες τιμές σύμφωνα με τον αλγόριθμο ECR. Κάτω: Οι αναδιατεταγμένες τιμές που αντιστοιχούν στην μικρότερη κορυφή της εκ των υστέρων κατανομής.

συγκεκριμένο γεγονός και αντιμετώπισε το φαινόμενο label switching χρησιμοποιώντας ένα δειγματολήπτη περιορισμένης μετάθεσης έτσι ώστε όλες οι τιμές να ικανοποιούν τον περιορισμό $\beta_{12} < \beta_{22}$. Παρ' όλ' αυτά, μία πιο προσεκτική ματιά στο προσομοιωμένο δείγμα αποκαλύπτει ότι υπάρχει και μία μικρότερη κορυφή στην εκ των υστέρων κατανομή των παραμέτρων. Κάτι τέτοιο μπορεί να γίνει αντιληπτό παρατηρώντας ότι υπάρχουν κάποιες προσομοιωμένες τιμές που συγκεντρώνονται στην περιοχή που βρίσκεται ανάμεσα στις περιοχές υψηλής εκ των υστέρων πυκνότητας των παραμέτρων β_{12}, β_{22} . Μία τέτοια συμπεριφορά έχουμε δει ότι είναι αρκετά συχνή σε δείγματα MCMC όταν υπάρχουν μικρότερες κορυφές στην εκ των υστέρων κατανομή. Στο συγκεκριμένο παράδειγμα η μικρότερη κορυφή οφείλεται στην μη αμελητέα πιθανότητα του ενδεχομένου κενής συνιστώσας. Πράγματι, η σχετική συχνότητα μιας κενής συνιστώσας εκτιμήθηκε βάσει των προσομοιωμένων τιμών των διανυσμάτων κατάταξης ίση με 1.6%. Επομένως ένας περιορισμός διάταξης θα αναδιατάξει σωστά μόνο τις τιμές που αντιστοιχούν στην κύρια κορυφή της κατανομής.

Κάτι τέτοιο δεν θα ισχύει για τις μικρότερες κορυφές καθώς οι τιμές της κενής συνιστώσας παράγονται από την εκ των προτέρων κατανομή και το μεγάλο εύρος αυτών έχει ως αποτέλεσμα να κοίτωνται εκατέρωθεν των τιμών της άλλης.

Τα παραπάνω σημεία γίνονται ακόμα πιο ξεκάθαρα αν το δείγμα αναδιαταχθεί μέσω του Αλγορίθμου ECR. Όπως φαίνεται στην δεύτερη γραμμή του Σχήματος 5.10, το φαινόμενο label switching έχει αντιμετωπιστεί επιτυχώς. Επί πλέον, στο αναδιαταγεμένο δείγμα διατηρείται και η μικρότερη κορυφή της εκ των υστέρων κατανομής. Αυτό υποστηρίζεται από τις αναδιατεταγμένες τιμές των β_{12}, β_{22} όπου βλέπουμε ότι όταν εμφανίζονται οι τιμές με την μεγάλη διασπορά (δηλαδή αυτές που παράγονται από την εκ των προτέρων κατανομή) κατατάσσονται πάντα στην «κόκκινη» συνιστώσα. Προκειμένου να γίνει ακόμα πιο σαφής η τελευταία παρατήρηση, η τρίτη γραμμή του Σχήματος (5.10) απεικονίζει μόνο εκείνες τις τιμές που αντιστοιχούν στην μικρότερη κορυφή, δηλαδή τις τιμές που παράγονται στις επαναλήψεις που καταγράφεται κενή συνιστώσα. Σύμφωνα με την ανάλυση που προηγήθηκε, προκύπτει ότι δεν υπάρχει περιορισμός που να είναι ικανός να απομονώσει τόσο την κύρια όσο και την μικρότερη κορυφή της εκ των υστέρων κατανομής. Ως εκ τούτου, η χρήση του Αλγορίθμου ECR σε αυτή την περίπτωση κρίνεται απαραίτητη ώστε να περιγράψουμε σωστά την συμπεριφορά των παραμέτρων.

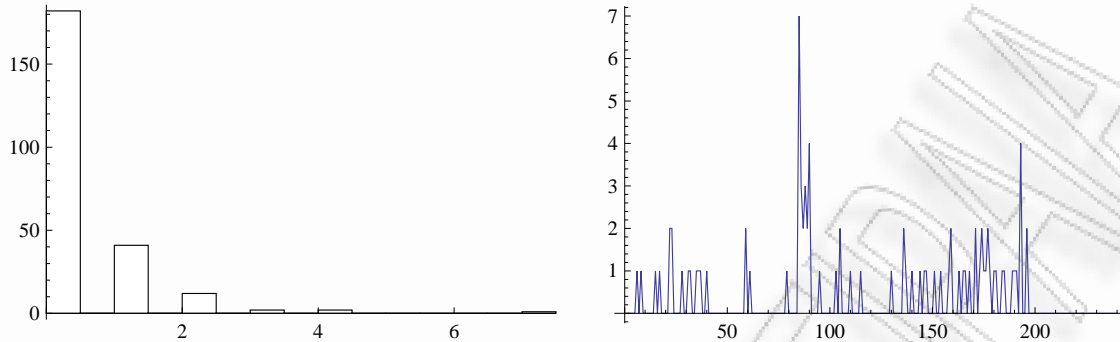
Ολοκληρώνουμε την ενότητα παραθέτοντας τις εκτιμήσεις των εκ των υστέρων μέσων τιμών των παραμέτρων του μοντέλου, έτσι όπως προκύπτουν αν υπολογίσουμε τους εργοδικούς μέσους των αναδιατεταγμένων τιμών βάσει του Αλγορίθμου ECR:

$$\hat{\beta} = \begin{pmatrix} 0.21 & 0.46 & -0.57 \\ 0.27 & -0.11 & 1.06 \end{pmatrix}, \quad \hat{P} = \begin{pmatrix} 0.538 & 0.462 \\ 0.277 & 0.723 \end{pmatrix}, \quad \hat{\sigma}^2 = (0.59, 0.53).$$

5.3.5 Μη ομογενής διαδικασία Poisson διακριτού χρόνου

Τα δεδομένα Poisson fetal lamb data χρησιμοποιήθηκαν από τους Chib (1996) και Frühwirth-Schnatter (2001). Πρόκειται για 240 παρατηρήσεις που καταγράφουν το πλήθος κινήσεων ενός εμβρύου αμνού σε διαδοχικά χρονικά διαστήματα πέντε δευτερολέπτων. Η χρονοσειρά απεικονίζεται στο Σχήμα 5.11.

Έστω ότι το y_t , $t = 1, \dots, n$, παριστάνει το πλήθος των κινήσεων στο t -οστό χρονικό διάστημα. Εφ' όσον οι μετρήσεις αναφέρονται σε διαδοχικά διαστήματα είναι λογικό να υποθέσουμε ότι θα είναι εξαρτημένες. Τα συγκεκριμένα δεδομένα μοντελοποιήθηκαν από την Frühwirth-Schnatter (2001) ως ένα HMM Poisson με $k = 3$ δυνατές καταστάσεις. Πιο συγκεκριμένα, έστω z_1, \dots, z_n η μη παρατηρήσιμη μαρκοβιανή αλυσίδα με πίνακα μετάβασης $P = (p_{ij})$, $i, j = 1, 2, 3$. Τότε, δοθείσης της



Σχήμα 5.11: Πλήθος κινήσεων του εμβρύου αμνού.

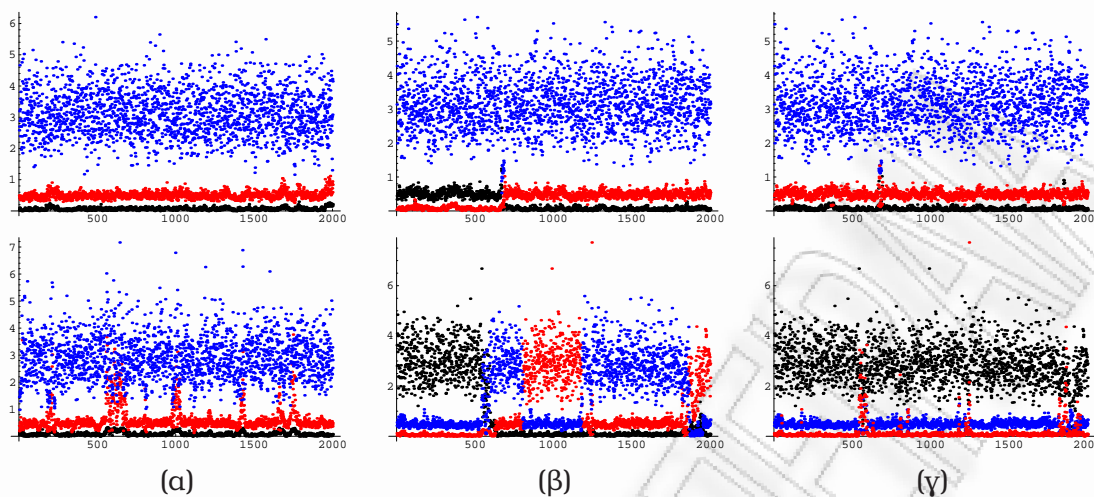
διαδικασίας αυτής, η ακολουθία του πλήθους των κινήσεων θεωρείται πραγματοποίηση ανεξάρτητων τυχαίων μεταβλητών με $y_t | z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \dots, n$.

Ακολουθώντας την Frühwirth-Schnatter (2001), θεωρούμε εκ των προτέρων ότι $\lambda_j \sim \mathcal{G}(1, 0.5)$, ανεξάρτητα για $j = 1, 2, 3$. Για την εκ των προτέρων κατανομή του πίνακα μετάβασης θεωρήσαμε και τις δύο επιλογές που συζητήθηκαν στην Ενότητα 5.3.2. Σύμφωνα με τα αποτελέσματα της Frühwirth-Schnatter, που προκύπτουν βάσει της πληροφοριακής κατανομής στον πίνακα πιθανοτήτων μετάβασης, ο περιορισμένος αλγόριθμος είναι αρκετά αποδοτικός. Πράγματι, στην πρώτη γραμμή του Σχήματος 5.11(a) βλέπουμε ότι ο περιορισμός $\lambda_1 < \lambda_2 < \lambda_3$ ταυτοποιεί επιτυχώς τις συμμετρικές περιοχές. Δεν υπάρχει ένδειξη μικρότερων κορυφών στην εκ των υστέρων κατανομή οπότε οι εκτιμήσεις που αναφέρονται στον Πίνακα 3 της Frühwirth-Schnatter (2001) είναι αρκετά αξιόπιστες.

Ωστόσο, μία διαφορετική συμπεριφορά καταγράφεται αν επιλέξουμε την μη πληροφοριακή εκ των προτέρων κατανομή για τον πίνακα πιθανοτήτων μετάβασης. Πιο συγκεκριμένα, ύστερα από 20000 επαναλήψεις (αφού προηγήθηκαν 2000 επαναλήψεις ως περίοδος burn-in) του δειγματολήπτη περιορισμένης μετάθεσης προκύπτει η ακολουθία της δεύτερης γραμμής του Σχήματος 5.11(a). Οι αντίστοιχες εκτιμήσεις των παραμέτρων είναι

$$\hat{\lambda} = (0.09, 0.60, 2.92), \quad \hat{P} = \begin{pmatrix} 0.890 & 0.084 & 0.026 \\ 0.138 & 0.803 & 0.059 \\ 0.242 & 0.213 & 0.545 \end{pmatrix}.$$

Παρατηρούμε στη δεύτερη γραμμή του Σχήματος 5.11(a) ότι οι τιμές των λ_2, λ_3 παρουσιάζουν υψηλή συγκέντρωση στο όριο του περιορισμού $\lambda_2 < \lambda_3$. Συνεπώς, ο



Σχήμα 5.12: (α) Προσομοιωμένες τιμές του δειγματολήπτη περιορισμένης μετάθεσης (Frühwirth–Schnatter, 2001). Προσομοιωμένες τιμές (β) και αναδιατεταγμένες τιμές (γ) των $(\lambda_1, \lambda_2, \lambda_3)$ σύμφωνα με τον αλγόριθμο ECR για δύο εκτελέσεις του δειγματολήπτη Gibbs με διαφορετικές εκ των προτέρων κατανομές στον πίνακα μεταβάσεων της μη παρατηρήσιμης διαδικασίας. Πάνω: η εκ των προτέρων κατανομή της Frühwirth–Schnatter. Κάτω: μη πληροφοριακή εκ των προτέρων κατανομή.

συγκεκριμένος περιορισμός δεν είναι ικανός να απομονώσει επιτυχώς τις συμμετρικές περιοχές της εκ των υστέρων κατανομής και έτσι αυτές οι εκτιμήσεις δεν κρίνονται αξιόπιστες. Συμπερασματικά, η μέθοδος των περιορισμών είναι αρκετά ευαίσθητη σε αλλαγές των εκ των προτέρων υποθέσεων: Στην περίπτωση της πληροφοριακής κατανομής στον πίνακα μετάβασης ένας περιορισμός ταυτοποιεί επαρκώς τις συνιστώσες, ενώ στην περίπτωση της μη πληροφοριακής κατανομής κάτι τέτοιο δεν είναι τόσο ξεκάθαρο.

Ακολούθως παρουσιάζονται τα αποτελέσματα βάσει του Αλγορίθμου ECR σε δείγματα μεγέθους 20000 (αφού προηγήθηκαν 2000 επαναλήψεις ως περίοδος burn-in) του δειγματολήπτη Gibbs. Στο Σχήμα 5.11(β) φαίνονται οι προσομοιωμένες τιμές των παραμέτρων των συνιστωσών για τις δύο διαφορετικές επιλογές της εκ των προτέρων κατανομής του πίνακα μετάβασης. Βλέπουμε ότι το φαινόμενο label switching εμφανίζεται και στις δύο περιπτώσεις, σε μεγαλύτερο βαθμό δε όταν χρησιμοποιούμε την μη πληροφοριακή εκ των προτέρων κατανομή. Το προσομοιωμένο δείγμα αναδιατάχθηκε σύμφωνα με τον Αλγόριθμο ECR και οι προκύπτουσες τιμές απεικονίζονται στο Σχήμα 5.11(γ). Οι εργοδικοί μέσοι των αναδιατεταγμένων τιμών του αλγορίθμου με την πληροφοριακή εκ των προτέρων κατανομή του πίνακα μετάβασης

βρέθηκαν ίσοι με

$$\hat{\lambda} = (0.07, 0.49, 3.11), \quad \hat{P} = \begin{pmatrix} 0.939 & 0.043 & 0.018 \\ 0.046 & 0.942 & 0.012 \\ 0.162 & 0.123 & 0.715 \end{pmatrix}.$$

Οι παραπάνω εκτιμήσεις είναι σχεδόν ταυτόσημες με αυτές που αναφέρει η Frühwirth-Schnatter (2001). Αυτό ήταν αναμενόμενο, καθώς ο περιορισμός $\lambda_1 < \lambda_2 < \lambda_3$ είναι λογικός σε αυτή την περίπτωση, όπως ήδη συζητήθηκε.

Για την μη πληροφοριακή κατανομή στον πίνακα πιθανοτήτων μετάβασης, η μέθοδός μας καταλήγει στις εκτιμήσεις

$$\hat{\lambda} = (0.14, 0.47, 2.85), \quad \hat{P} = \begin{pmatrix} 0.846 & 0.113 & 0.041 \\ 0.090 & 0.883 & 0.027 \\ 0.237 & 0.199 & 0.563 \end{pmatrix}.$$

Στο αναδιατεταγμένο δείγμα βάσει του αλγορίθμου ECR (δες Σχήμα 5.11(γ), δεύτερη γραμμή) υπάρχει ένδειξη μικρότερων κορυφών στην εκ των υστέρων κατανομή. Βλέπουμε ξεκάθαρα ότι υπάρχουν περιπτώσεις όπου το λ_1 μπορεί να πάρει πολύ μεγάλες τιμές και αυτό επηρεάζει την συμπεριφορά του λ_2 . Όπως ήδη συζητήθηκε, σε αυτήν την περίπτωση ο δειγματολήπτης περιορισμένης μετάθεσης της Frühwirth-Schnatter περιορίζει σαφώς τις τιμές αυτής της συνιστώσας, καθώς δεν φαίνεται να ικανοποιείται κάποιος περιορισμός. Για άλλη μια φορά βλέπουμε ξεκάθαρα την υπεροχή της προτεινόμενης μεθόδου έναντι των περιορισμών διάταξης: Ανεξαρτήτως αν οι συμμετρικές περιοχές της εκ των υστέρων κατανομής είναι μακριά μεταξύ τους ή όχι, ο Αλγόριθμος ECR είναι ικανός να βρει την σωστή αναδιάταξη του προσομοιωμένου δείγματος.

5.3.6 Μη συζυγή HMM

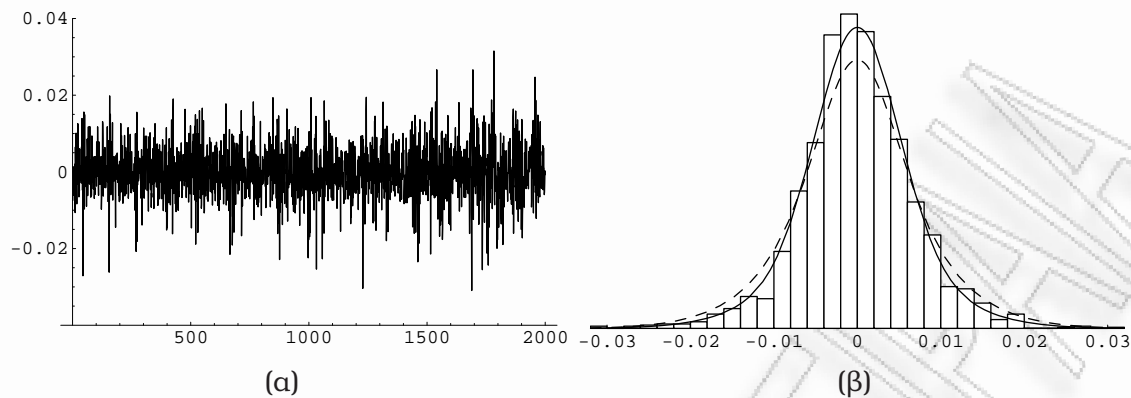
Οι Robert et al. (2000) επέκτειναν τον αλγόριθμο RJMCMC των Richardson and Green (1997) στην κλάση των HMM. Η κύρια συνεισφορά της έρευνάς τους βεβαίως βρίσκεται στο γεγονός ότι είναι δυνατή η συμπερασματολογία για το άγνωστο πλήθος των καταστάσεων k της μη παρατηρήσιμης διακριτής μαρκοβιανής διαδικασίας. Πιο συγκεκριμένα, το μοντέλο που θεώρησαν έχει ως εξής. Για κάθε σταθερό $k > 1$, υποθέτουν ότι δοθέντος του διανύσματος κατάταξης $z = (z_1, \dots, z_n) \in \{1, \dots, k\}^n$, οι παρατηρήσεις y_1, \dots, y_n είναι πραγματοποιήσεις ανεξάρτητων τυχαίων μεταβλητών

$$y_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2), \quad t = 1, \dots, n, \quad j = 1, \dots, k. \quad (5.7)$$

Όπως προηγουμένως, έστω $\mathbf{P} = P(z_{t+1} = j | z_t = i) = (p_{ij})$ ο πίνακας μετάβασης της μη παρατηρήσιμης μαρκοβιανής αλυσίδας. Επομένως, δοθέντος του k , οι παράμετροι του μοντέλου είναι $(\mathbf{P}, \boldsymbol{\sigma})$. Ο πίνακας \mathbf{P} αποτελείται εκ των προτέρων από ανεξάρτητες γραμμές οι οποίες θεωρούνται τυχαία διανύσματα που ακολουθούν την κατανομή Dirichlet, $\mathcal{D}(1, \dots, 1)$. Επίσης, οι τυπικές αποκλίσεις $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_k)$ είναι ανεξάρτητες τυχαίες μεταβλητές $\mathcal{U}(0, \alpha)$ που στο μοντέλο των Robert et al. (2000) θεωρούνται εκ των προτέρων διατεταγμένες κατά αύξουσα τάξη έτσι ώστε να είναι ταυτοποιήσιμες. Τέλος, για την υπερπαράμετρο α υποτίθεται ότι $\alpha \sim \mathcal{E}(1/\xi)$, με $1/\xi = c \max |y_t|$ όπου το $c > 0$ είναι μία προκαθορισμένη σταθερά. Παρατηρούμε ότι η ομοιόμορφη κατανομή στις διασπορές δεν είναι συζυγής στο μοντέλο, κάτι που ισχύει επίσης και για την εκ των προτέρων κατανομή του α . Ως εκ τούτου, η προσομοίωση από τις πλήρεις εκ των υστέρων δεσμευμένες κατανομές δεν αποτελεί μία τυπική διαδικασία. Οι Robert et al. (2000) προτείνουν την χρήση του slice sampler των Damien et al. (1999). Η ανανέωση των σ_j , $j = 1, \dots, k$, γίνεται προτείνοντας μία τιμή από τις πλήρεις δεσμευμένες εκ των υστέρων κατανομές η οποία όμως απορρίπτεται ευθύς αμέσως όταν δεν ικανοποιεί τους περιορισμούς διάταξης και τότε επαναλαμβάνεται η προηγούμενη τιμή. Κάτι τέτοιο όμως επιφέρει —στην καλύτερη περίπτωση— πολύ αργή σύγκλιση του αλγορίθμου. Επί πλέον, όπως είδαμε στο Κεφάλαιο 3, ένας περιορισμός στον αλγόριθμο RJMCMC έχει και επιπτώσεις στην εκ των υστέρων κατανομή του k όταν οι περιοχές υψηλής εκ των υστέρων πιθανότητας βρίσκονται κοντά στο όριο του περιορισμού.

Εδώ μας ενδιαφέρει η συμπερασματολογία για τις περιθωριακές κατανομές των παραμέτρων δεσμεύοντας σε συγκεκριμένες τιμές του k . Προκειμένου να δείξουμε τα μειονεκτήματα του περιορισμού διάταξης τρέχουμε δύο δειγματολήπτες. Ο πρώτος ενσωματώνει τον περιορισμό διάταξης και θα αναφερόμαστε σε αυτόν με τον όρο περιορισμένος αλγόριθμος ενώ ο δεύτερος λειτουργεί ελεύθερα σε ολόκληρο τον παραμετρικό χώρο. Μόλις ολοκληρωθεί η διαδικασία των προσομοιώσεων, εφαρμόζουμε τον Αλγόριθμο ECR στο δείγμα που παράγεται από τον ελεύθερο αλγόριθμο, στο οποίο βέβαια θα εμφανίζεται το φαινόμενο label switching. Τονίζουμε ότι στον ελεύθερο αλγόριθμο έχουμε το πλεονέκτημα ότι δεν γίνονται απορρίψεις τιμών, συνεπώς η σύγκλισή του είναι πιο γρήγορη.

Στα παραδείγματα που ακολουθούν χρησιμοποιούμε προσομοιωμένα και πραγματικά σύνολα δεδομένων που θεωρούνται πραγματοποιήσεις του μοντέλου (5.7). Αρχικά, μέσω της μελέτης στα προσομοιωμένα δεδομένα αναδεικνύονται οι επιπτώσεις του περιορισμένου δειγματολήπτη και η βελτίωση που επιφέρει στην συμπε-



Σχήμα 5.13: (α): Προσομοιωμένα δεδομένα 2000 παρατηρήσεων (β): εκτιμήσεις της περιθωριακής συνάρτησης πυκνότητας σύμφωνα με τον αλγόριθμο ECR (—) και τον αλγόριθμο περιορισμού (- - -).

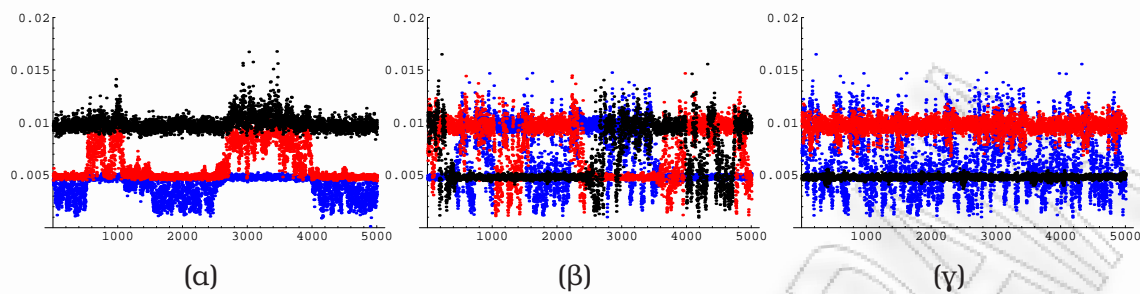
ρασατολογία η χρήση του Αλγορίθμου ECR για την αντιμετώπιση του φαινομένου label switching. Ακολουθώς, η προτεινόμενη μέθοδος εφαρμόζεται σε πραγματικά σύνολα οικονομικών και μετεωρολογικών δεδομένων. Στο σύνολο των οικονομικών δεδομένων η εκ των υστέρων κατανομή των παραμέτρων παρουσιάζει πολλές γνήσιες κορυφές και είναι σαφές ότι η μέθοδος των περιορισμών δεν είναι ικανή να απομονώσει τις συμμετρικές περιοχές της. Το σύνολο των μετεωρολογικών δεδομένων αποτελεί τυπικότερη περίπτωση και η μέθοδος των περιορισμών οδηγεί στα ίδια συμπεράσματα με τον Αλγόριθμο ECR.

Παράδειγμα 5.3.1. Προσομοιωμένα δεδομένα.

Χρησιμοποιούμε αρχικά ένα προσομοιωμένο σύνολο δεδομένων ώστε να γίνει σαφές ότι η χρήση εικονικών περιορισμών έχει σοβαρότατες επιπτώσεις στην συμπερασατολογία και ταυτόχρονα να αναδειχθεί η ανωτερότητα της προτεινόμενης μεθόδου. Έτσι λοιπόν προσομοιώσαμε 2000 παρατηρήσεις από το μοντέλο (5.7) με $k = 3$ καταστάσεις χρησιμοποιώντας τις εξής τιμές των παραμέτρων:

$$\sigma = (0.0045, 0.0060, 0.0090), \quad P = \begin{pmatrix} 0.95 & 0.03 & 0.02 \\ 0.10 & 0.85 & 0.05 \\ 0.025 & 0.025 & 0.95 \end{pmatrix}.$$

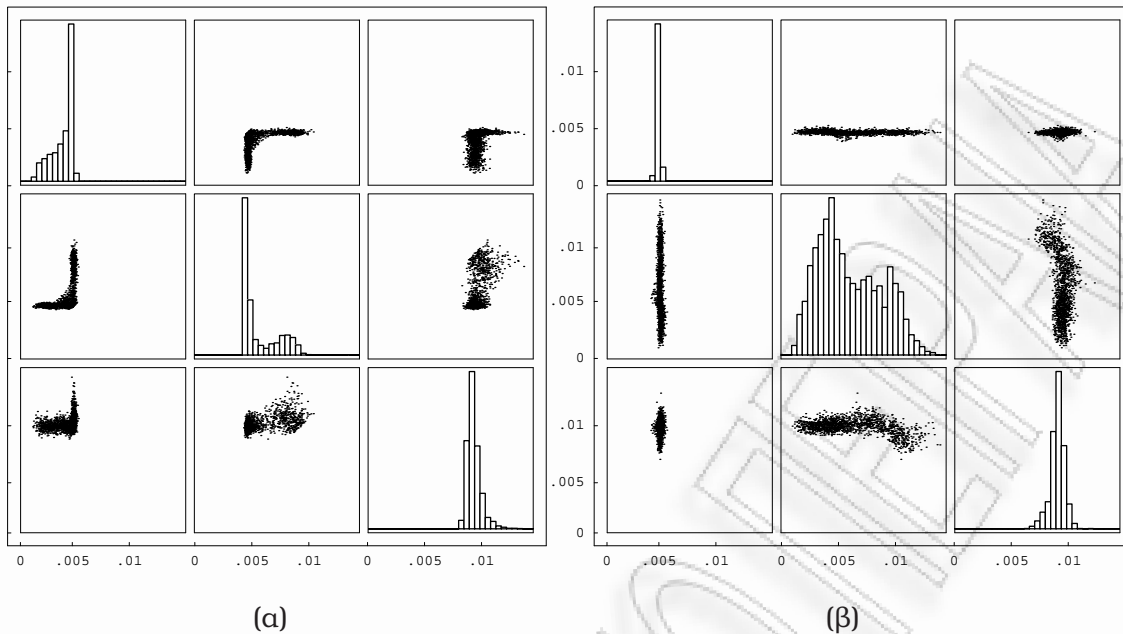
Η χρονοσειρά παρουσιάζεται στο Σχήμα 5.13(α). Στο Σχήμα 5.14 απεικονίζεται το προσομοιωμένο δείγμα σύμφωνα με τους δύο αλγόριθμους για 60000 επαναλήψεις, αφού προηγήθηκαν 40000 επαναλήψεις ως περίοδος burn-in (φαίνεται κάθε 12η επανάληψη). Είναι προφανές ότι υπάρχουν μεγάλες διαφορές μεταξύ των δύο μεθόδων. Παρατηρούμε ότι ο αλγόριθμος περιορισμού παράγει πολλές τιμές που



Σχήμα 5.14: Οι προσομοιωθείσες τιμές των $(\sigma_1, \sigma_2, \sigma_3)$ για το Παράδειγμα 5.3.1. (α) αλγόριθμος περιορισμού, (β) ελεύθερος αλγόριθμος, (γ) αναδιατεταγμένο δείγμα του ελεύθερου αλγορίθμου, σύμφωνα με τον Αλγόριθμο ECR.

βρίσκονται στο όριο του περιορισμού διάταξης $\sigma_1 < \sigma_2 < \sigma_3$. Είναι σαφές ότι κάτι τέτοιο περιορίζει το εύρος των τιμών που βρίσκονται στην περιοχή υψηλής πυκνότητας με συνέπεια την αμφίβολη εκτίμηση αυτών των περιοχών. Αντιθέτως, στα Σχήματα 5.14(β) και 5.14(γ) φαίνεται μία πολύ διαφορετική εικόνα καθώς συμπεραίνουμε ότι οι τυπικές αποκλίσεις συγκεντρώνονται σε δύο διακριτές περιοχές υψηλής εκ των υστέρων πυκνότητας (γύρω στο 0.0045 και 0.009, αντίστοιχα) στις οποίες παίρνουν τιμές τα σ_1 και σ_3 . Οι τιμές του σ_2 καλύπτουν και τις δύο αυτές περιοχές και εναλλάσσονται συνεχώς μεταξύ αυτών. Αυτό πιθανώς αποτελεί ένδειξη ότι για την εκτίμηση του συγκεκριμένου HMM ίσως να ήταν αρκετές δύο (αντί για τρεις) καταστάσεις. Επί πλέον, στο Σχήμα 5.14(β) καταγράφονται αρκετές εναλλαγές μεταξύ των καταστάσεων και προκειμένου να απαλλαγούμε από αυτές εφαρμόζουμε τον Αλγόριθμο ECR για την αναδιάταξη των προσομοιωμένων τιμών.

Οι παραπάνω παρατηρήσεις καταγράφονται και στο Σχήμα 5.15 το οποίο περιέχει τα διαγράμματα διασποράς των τυπικών αποκλίσεων μαζί με τα αντίστοιχα ιστογράμματα σύμφωνα με τις δύο μεθόδους. Ειδικότερα, στο Σχήμα 5.15(α) αποκαλύπτεται ότι υπάρχει μεγάλη συγκέντρωση τιμών των $(\sigma_2, \sigma_1)^{(i)}$ και $(\sigma_3, \sigma_2)^{(i)}$ στην διαγώνιο του πρώτου τεταρτημορίου που αποτελεί το σύνορο του περιορισμού του πρώτου αλγορίθμου. Έτσι φαίνεται ότι η χρήση περιορισμών διάταξης σε αυτήν την περίπτωση είναι μάλλον ανίκανη να απομονώσει τις συμμετρικές περιοχές της εκ των υστέρων κατανομής. Επιπροσθέτως, το γεγονός ότι αυτά τα διαγράμματα παρουσιάζουν σχήμα ορθής γωνίας οδηγεί στο συμπέρασμα ότι το σ_1 όχι μόνο μπορεί να είναι ίσο με το σ_2 αλλά και ακόμα μεγαλύτερο από αυτό. Αντιθέτως, στο Σχήμα 5.15(β) όπου βρίσκονται τα αντίστοιχα γραφήματα σύμφωνα με τον Αλγόριθμο ECR, καταγράφεται μια τελείως διαφορετική εικόνα. Από τα διαγράμματα διασποράς των



Σχήμα 5.15: Παράδειγμα 5.3.1: Η περιθωριακή εκ των υστέρων κατανομή των τυπικών αποκλίσεων για κάθε κατάσταση με τα αντίστοιχα διαγράμματα διασποράς. (α) Περιορισμένος αλγόριθμος, (β): αναδιατεταγμένο δείγμα σύμφωνα με τον Αλγόριθμο ECR.

$(\sigma_2, \sigma_1)^{(i)}$ και $(\sigma_3, \sigma_2)^{(i)}$ συμπεραίνουμε ότι δεν υπάρχει κανένας περιορισμός μεταξύ των τυπικών αποκλίσεων καθώς το σ_1 βρίσκεται σταθερά σε μία περιοχή ενώ την ίδια στιγμή το σ_2 παίρνει τιμές σε ένα πολύ μεγαλύτερο διάστημα. Παρόμοια είναι και η εικόνα μεταξύ των σ_3 και σ_2 . Βλέπουμε ότι η ύπαρξη πολλών κορυφών στην περιθωριακή κατανομή του σ_2 επηρεάζει και τις τιμές του σ_3 : Όταν το σ_2 παίρνει μεγάλες τιμές το σ_3 αυξάνεται.

Ο περιορισμένος αλγόριθμος οδηγεί στις ακόλουθες εκτιμήσεις των παραμέτρων (μέσω των εργοδικών μέσων):

$$\hat{\sigma} = (0.0039, 0.0060, 0.0099), \quad \hat{P} = \begin{pmatrix} 0.515 & 0.224 & 0.260 \\ 0.151 & 0.652 & 0.197 \\ 0.086 & 0.183 & 0.732 \end{pmatrix}.$$

Συγκρίνοντας τις παραπάνω εκτιμήσεις με τις πραγματικές τιμές των αντίστοιχων παραμέτρων, συμπεραίνουμε την υποεκτίμηση του σ_1 και την υπερεκτίμηση του σ_3 . Επί πλέον, παρατηρούμε ότι η εκτίμηση του πίνακα πιθανοτήτων μετάβασης της μη παρατηρήσιμης διαδικασίας είναι αρκετά κακή. Από την άλλη πλευρά, οι εργοδικοί μέσοι του αναδιατεταγμένου δείγματος βάσει του αλγορίθμου ECR δίνουν

τις ακόλουθες εκτιμήσεις:

$$\hat{\sigma} = (0.0047, 0.0060, 0.0096), \quad \hat{P} = \begin{pmatrix} 0.868 & 0.106 & 0.026 \\ 0.305 & 0.256 & 0.439 \\ 0.057 & 0.160 & 0.783 \end{pmatrix}.$$

Όπως ήταν αναμενόμενο, οι εκτιμήσεις της πρώτης και τρίτης απόκλισης είναι πιο κοντά στις πραγματικές τους τιμές, ενώ το ίδιο ισχύει για τις αντίστοιχες γραμμές του πίνακα μετάβασης. Η κακή εκτίμηση της δεύτερης γραμμής παρ' όλα αυτά ίσως να είναι συνέπεια του γεγονότος ότι η δεύτερη συνιστώσα δεν συνεισφέρει και πολύ στην ερμηνεία των δεδομένων. Όπως περιγράψαμε και προηγουμένως, δύο συνιστώσες φαίνεται να προσαρμόζονται εξ ίσου καλά σε αυτά.

Μία άλλη ποσότητα που παρουσιάζει ενδιαφέρον σε ένα HMM, είναι η εκτίμηση της στάσιμης κατανομής των κρυμμένων καταστάσεων $\pi = (\pi_1, \pi_2, \pi_3)$. Έτσι λοιπόν, υπολογίζοντας το κατάλληλο ιδιοδιάνυσμα του P , έχουμε ότι η πραγματική στάσιμη κατανομή των καταστάσεων της αλυσίδας είναι η

$$\pi \approx (.490, .157, .353).$$

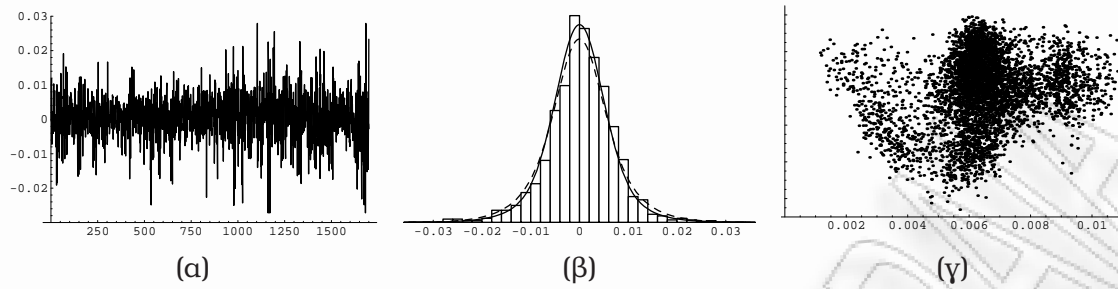
Ας παρατηρήσουμε εδώ ότι η πιθανότητα επίσκεψης της δεύτερης κατάστασης είναι αρκετά μικρότερη σε σχέση με των άλλων δύο. Τα αντίστοιχα ιδιοδιανύσματα των δύο πινάκων που χρησιμοποιήθηκαν για την εκτίμηση του P είναι

$$(.191, .359, .449),$$

για τον περιορισμένο αλγόριθμο και

$$(.494, .148, .359)$$

για τον ελεύθερο αλγόριθμο, αφού εφαρμοστεί στο προσομοιωμένο δείγμα ο αλγόριθμος ECR. Είναι εμφανές ότι η εκτίμηση υπό την μέθοδό μας προσεγγίζει την π πολύ καλύτερα σε σχέση με την αντίστοιχη εκτίμηση των περιορισμών διάταξης. Το γεγονός αυτό είναι αρκετά ενθαρρυντικό για την μέθοδό μας καθώς φαίνεται ότι οδηγεί σε αρκετά καλές εκτιμήσεις για τις εξομαλυμένες πιθανότητες των κρυμμένων καταστάσεων για κάθε μια παρατήρηση (δες Hamilton, 1989) κάτι που είναι αρκετά σημαντικό για τις εφαρμογές στα HMM. Οι εκτιμήσεις των στάσιμων πιθανοτήτων των καταστάσεων μαζί με τις εκτιμήσεις των τυπικών αποκλίσεων δίνουν την εκτίμηση της περιθωριακής συνάρτησης πυκνότητας των δεδομένων που παρατίθεται στο Σχήμα 5.13(β). Βλέπουμε ξεκάθαρα ότι τα αποτελέσματα της μεθόδου μας προσαρμόζονται καλύτερα στα δεδομένα.

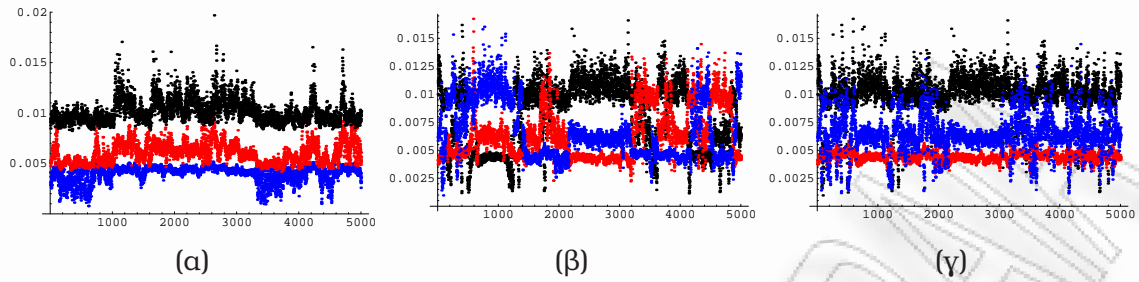


Σχήμα 5.16: (α) Δεδομένα Standard and Poors 500 (series E) και (β) Εκτιμήσεις της περιθωριακής συνάρτησης πυκνότητας σύμφωνα με τον αλγόριθμο ECR (—) και τον αλγόριθμο περιορισμού (- - -). (γ) Διάγραμμα διασποράς του $(\sigma_2, \log f(\mathbf{z}, \boldsymbol{\sigma}, \mathbf{P}, \alpha|\mathbf{y}))^{(i)}$, $i = 400000(20)500000$ (η εκ των υστέρων κατανομή υπολογίζεται ως προς μία πολλαπλασιαστική σταθερά) για τις αναδιατεταγμένες τιμές του σ_2 βάσει του Αλγορίθμου ECR.

Παράδειγμα 5.3.2. Σύνολο δεδομένων Standard and Poors 500.

Το πρώτο πραγματικό σύνολο δεδομένων που χρησιμοποιούμε είναι τα δεδομένα Standard and Poors 500, τα οποία έχουν χρησιμοποιηθεί και από τους Robert et al. (2000). Το σύνολο δεδομένων αποτελείται από 1700 παρατηρήσεις ημερήσιων αποδόσεων κατά την δεκαετία του 1950. Ο αναγνώστης παραπέμπεται στους Robert et al. (2000) και τις αναφορές εκεί για περισσότερες πληροφορίες για τα δεδομένα. Η αντίστοιχη χρονοσειρά απεικονίζεται στο Σχήμα 5.16(a). Στην εργασία των Robert et al. υποστηρίζεται ότι η εκ των υστέρων κατανομή του k υποδεικνύει ότι ο πιο πιθανός αριθμός των καταστάσεων της κρυμμένης μαρκοβιανής διαδικασίας είναι ίσος με 2 ή 3. Οι δύο αυτές τιμές φαίνονται σχεδόν ισοπίθανες καθώς εκτιμάται ότι $\hat{P}(k = 2|\mathbf{y}) = 0.4877$ και $\hat{P}(k = 3|\mathbf{y}) = 0.4521$. Όταν ο αριθμός των συνιστωσών ισούται με $k = 2$ οι συμμετρικές περιοχές υψηλής εκ των υστέρων πυκνότητας είναι αρκετά μακριά μεταξύ τους και έτσι τα αποτελέσματα αναδιάταξης του δείγματος συμπίπτουν με αυτά του περιορισμένου αλγορίθμου, βάσει της διάταξης των τυπικών αποκλίσεων. Έτσι, εδώ παρουσιάζουμε τα αποτελέσματα για $k = 3$.

Στο Σχήμα 5.17 απεικονίζονται τα προσομοιωμένα δείγματα από την εκ των υστέρων κατανομή που πήραμε από τους δύο αλγορίθμους για 500000 επαναλήψεις μετά το burn-in των πρώτων 200000 (απεικονίζεται κάθε 20η επανάληψη από τις τελευταίες 100000). Ο μεγάλος αριθμός επαναλήψεων σχολιάζεται στο τέλος της ενότητας. Προφανώς υπάρχουν μεγάλες διαφορές μεταξύ των δύο μεθόδων. Παρατηρούμε ότι οι προσομοιωμένες τιμές σύμφωνα με τον περιορισμένο αλγόριθμο συγκεντρώνονται



Σχήμα 5.17: Προσομοιωμένες τιμές από τον δειγματολήπτη Gibbs για το σύνολο δεδομένων Standard and Poors. (α) Περιορισμένος αλγόριθμος, (β) ελεύθερος αλγόριθμος, (γ) αναδιατεταγμένες τιμές του ελεύθερου αλγορίθμου βάσει του Αλγορίθμου ECR.

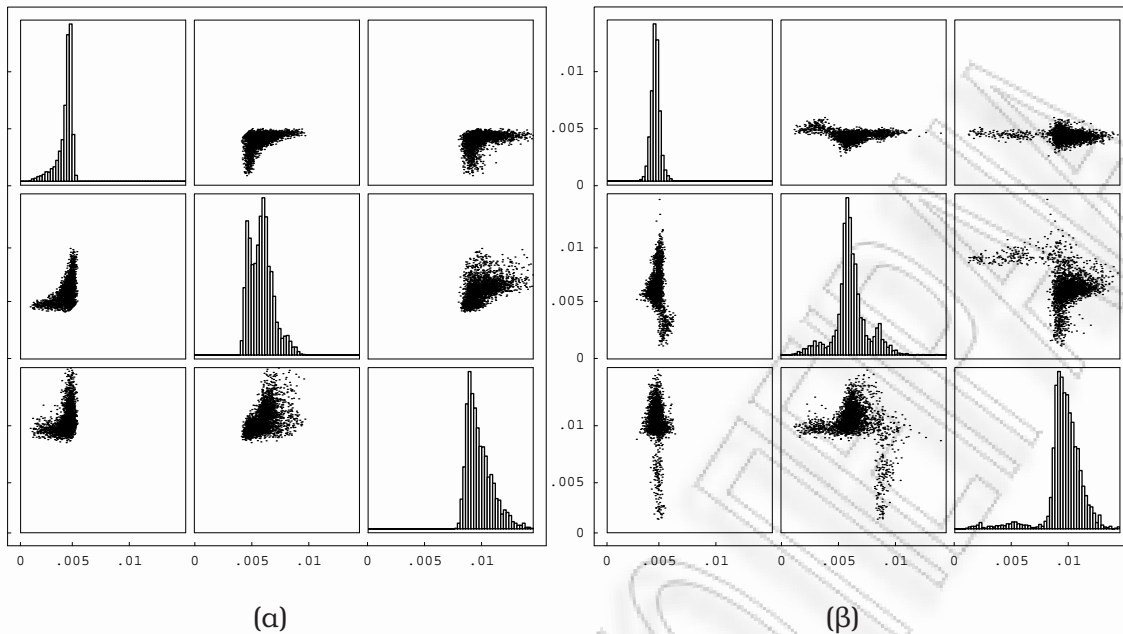
στο όριο του περιορισμού διάταξης των τυπικών αποκλίσεων $\sigma_1 < \sigma_2$, ενώ η τρίτη τυπική απόκλιση παίρνει τιμές σε μία πιο απομακρυσμένη περιοχή. Αυτό αποτελεί ένδειξη ότι η υπόθεση διάταξης των τυπικών αποκλίσεων περιορίζει τις τιμές των δύο πρώτων. Από την άλλη πλευρά, στα Σχήματα 5.17(β) και 5.17(γ) φαίνεται μια διαφορετική εικόνα. Παρατηρούμε ότι ενώ υπάρχουν δύο συνιστώσες οι τυπικές αποκλίσεις των οποίων βρίσκονται συγκεντρωμένες σε δύο περιοχές (η πρώτη βρίσκεται κοντά στο 0.0045 και η δεύτερη γύρω στο 0.01), οι τιμές της τυπικής απόκλισης της άλλης συνιστώσας εναλλάσσονται συνεχώς μεταξύ αυτών των δύο περιοχών. Στο Σχήμα 5.17(β) βλέπουμε και το φαινόμενο label switching μεταξύ αυτών των τιμών.

Προκειμένου να απαλλαχτούμε από το label switching το δείγμα αναδιατάσσεται βάσει του Αλγορίθμου ECR και προκύπτει το δείγμα που απεικονίζεται στο Σχήμα 5.17(γ). Τα ιστογράμματα των σ_i , $i = 1, \dots, k$, και τα αντίστοιχα διαγράμματα διασποράς παρατίθενται στο Σχήμα 5.18, σύμφωνα με τις δύο μεθόδους. Παρατηρούμε ότι η αναδιάταξη βάσει του Αλγορίθμου ECR οδηγεί σε περιθωριακές κατανομές που διατηρούν τόσο την κύρια κορυφή όσο και τις μικρότερες. Αυτό δεν ισχύει για το δείγμα που παράγεται από τον περιορισμένο αλγόριθμο, καθώς η πλειοψηφία των τιμών συγκεντρώνεται κοντά στην διχοτόμο του πρώτου τεταρτημορίου. Αυτό δείχνει ξεκάθαρα ότι ο περιορισμός που χρησιμοποιείται δεν είναι ικανός να ξεχωρίσει τις συμμετρικές περιοχές.

Ο περιορισμένος αλγόριθμος κατέληξε στις εξής εκτιμήσεις:

$$\hat{\sigma} = (0.0040, 0.0060, 0.010), \quad \hat{P} = \begin{pmatrix} 0.701 & 0.206 & 0.093 \\ 0.181 & 0.722 & 0.096 \\ 0.048 & 0.141 & 0.812 \end{pmatrix}.$$

Η αναδιάταξη του δείγματος που προκύπτει από τον ελεύθερο αλγόριθμο, υπό την



Σχήμα 5.18: Παράδειγμα 5.3.2: Περιθωριακές κατανομές και διαγράμματα διασποράς των διασπορών. (α): αλγόριθμος περιορισμού, (β): αναδιάταξη βάσει του Αλγορίθμου ECR σε προσομοιωμένο δείγμα από την συμμετρική εκ των υστέρων κατανομή.

προτεινόμενη μέθοδο δίνει τις εκτιμήσεις:

$$\hat{\sigma} = (0.0045, 0.0064, 0.010), \quad \hat{P} = \begin{pmatrix} 0.894 & 0.083 & 0.023 \\ 0.173 & 0.669 & 0.158 \\ 0.061 & 0.214 & 0.725 \end{pmatrix}.$$

Όπως αναμένεται σύμφωνα με τα παραπάνω και το γράφημα της περιθωριακής κατανομής, ο περιορισμένος αλγόριθμος υποεκτιμά το σ_1 . Αντίθετα, η αναδιάταξη βάσει του Αλγορίθμου ECR αποκαλύπτει ότι υπάρχει μία συνιστώσα της οποίας η διασπορά μπορεί να είναι είτε περίπου ίση με 0.0060 είτε θα είναι ίση με την διασπορά των άλλων δύο συνιστωσών. Αυτό αποτελεί ένδειξη ότι το μοντέλο τριών συνιστωσών ίσως είναι υπερπαραμετροποιημένο, μία συμπεριφορά που δεν μπορεί να αποκαλυφθεί χρησιμοποιώντας τον περιορισμό διάκρισης των Robert et al. (2000), καθώς τονίζουμε για άλλη μία φορά ότι ο αλγόριθμός τους δεν μπορεί να ξεχωρίσει την προσαρμογή μεταξύ ενός μοντέλου με $k = 2$ και ενός μοντέλου με $k = 3$ συνιστώσες. Πιστεύουμε ότι κάτι τέτοιο συμβαίνει λόγω της ανικανότητας του περιορισμού να διακρίνει τις συνιστώσες, αλλά επί πλέον και λόγω των μικρών ποσοστών αποδοχής των κινήσεων αλλαγής διάστασης. Εξ άλλου κάτι τέτοιο υποστηρίζεται και

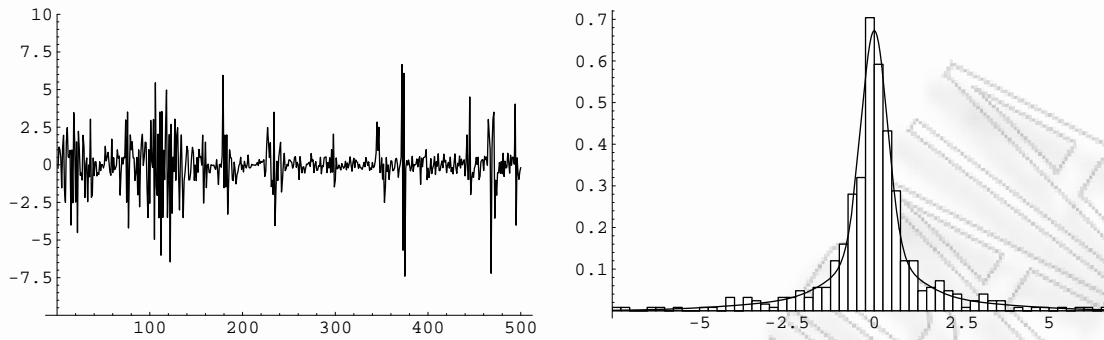
από τα συμπεράσματα του Κεφαλαίου 3, όπου είδαμε ότι όταν ο παραμετρικός χώρος είναι περιορισμένος και η εκ των υστέρων κατανομή παρουσιάζει μέγιστο κοντά στο σύνορο του περιορισμού, τότε ο δειγματολήπτης τείνει να επισκέπτεται πιο συχνά μοντέλα μεγαλύτερων διαστάσεων, κάτι που οδηγεί σε υπερεκτίμηση του άγνωστου αριθμού καταστάσεων k .

Κλείνουμε αυτήν την ενότητα με τον σχολιασμό του μεγάλου πλήθους επαναλήψεων του δειγματολήπτη Gibbs. Σε αυτό το παράδειγμα, η παρουσία πολλών κορυφών στην εκ των υστέρων κατανομή καθιστά αρκετά δύσκολη τη διαδικασία εκτίμησης του μεγίστου της καθώς απαιτήθηκε πολύ μεγάλο πλήθος επαναλήψεων για να παρατηρήσουμε μία σταθερή εκτίμηση του $(z, \sigma, P)^{(MAP)}$. Στο Σχήμα 5.16(γ) απεικονίζεται το διάγραμμα διασποράς του $(\sigma_2, \log f(z, \sigma, P, \alpha|y))^{(i)}$, $i = 400000(20)500000$, για τις αναδιατεταγμένες τιμές του σ_2 σύμφωνα με τον Αλγόριθμο ECR. Ο λόγος που επιλέξαμε το σ_2 για να δείξουμε την ύπαρξη αρκετών κορυφών στην εκ των υστέρων κατανομή προκύπτει από το γεγονός ότι αυτή η τυπική απόκλιση παρουσιάζει έντονη ετερογένεια στις προσομοιωθείσες τιμές της. Αυτό φαίνεται στο Σχήμα 5.18(β). Είναι προφανές ότι η εκ των υστέρων κατανομή έχει περίπου την ίδια τιμή στις τρεις κορυφές της περιθωριακής κατανομής του σ_2 . Σύμφωνα με τις προσομοιώσεις που κάναμε, όταν ο αριθμός των επαναλήψεων είναι μικρότερος του 100000 τότε η εκτίμηση του μεγίστου της εκ των υστέρων κατανομής φαίνεται να μην έχει συγκλίνει και δείχνει να αντιστοιχεί σε οποιαδήποτε από τις τρεις κορυφές. Όταν αυξήσαμε τον αριθμό των επαναλήψεων στις 500000 η εκτίμηση αυτή συνέκλινε τελικά στην κορυφή όπου οι τυπικές αποκλίσεις έχουν τιμή $\sigma^{(MAP)} = (0.0045, 0.0063, 0.0114)$. Έτσι, η αντίστοιχη εκτίμηση $z^{(MAP)}$ αναδιατάσσει επιτυχώς το δείγμα. Στην περίπτωση όμως που ο αριθμός των επαναλήψεων δεν αρκεί ώστε να επιτευχθεί σύγκλιση στην κορυφή της εκ των υστέρων κατανομής, η αναδιάταξη βάσει του Αλγορίθμου ECR δεν κατορθώνει να εξαλείψει πλήρως την συμμετρία. Αυτή η συμπεριφορά δεν πρέπει να μας προβληματίζει καθώς ήδη έχει εξηγηθεί στο «ακριβές» παράδειγμα με την μείξη δύο κατανομών Poisson της Ενότητας 4.2.3. Είδαμε εκεί ότι υπάρχουν επιλογές του οδηγού z^* που ενώ οδηγούν σε μη συμμετρικές κατανομές, δεν κατορθώνουν να εξαλείψουν πλήρως τη συμμετρία της εκ των υστέρων κατανομής.

Παράδειγμα 5.3.3. Ένταση ανέμων στην Αθήνα.

Το συγκεκριμένο σύνολο δεδομένων¹ καταγράφει 500 ωριαίες μετρήσεις της έντασης του ανέμου στην Αθήνα κατά τον Ιανουάριο του 1990. Οι παρατηρήσεις αυτές έχουν

¹ Ευχαριστώ τον Christian Francq για την ευγενική παραχώρηση των δεδομένων.



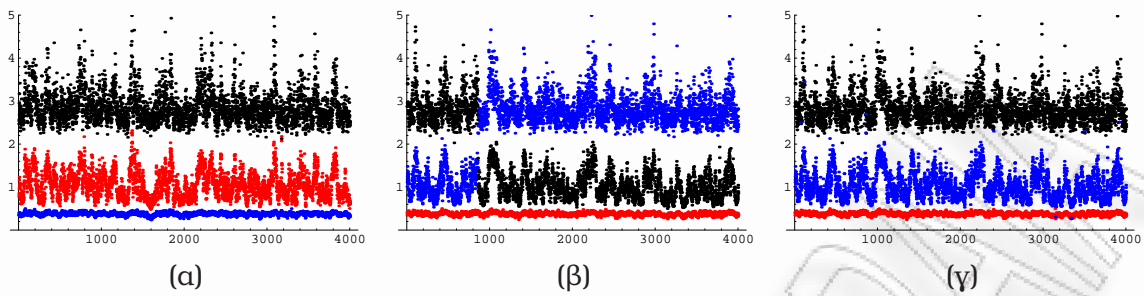
Σχήμα 5.19: Σύνολο δεδομένων ανέμου μαζί με την εκτίμηση της περιθωριακής συνάρτησης πυκνότητας σύμφωνα με τον Αλγόριθμο ECR.

αναλυθεί από τους Robert et al. (2000) και από τους Francq and Roussignol (1997). Η ανάλυση γίνεται χρησιμοποιώντας τις πρώτες διαφορές των μετρήσεων οι οποίες όμως είναι διακριτοποιημένες σε διαστήματα μήκους 0.1. Για το λόγο αυτό προσθήσαμε στις μετρήσεις έναν ομοιόμορφο θόρυβο στο $(-0.05, 0.05)$ · δεξ και Robert et al. (2000) για λεπτομέρειες. Η αντίστοιχη χρονοσειρά απεικονίζεται στο Σχήμα 5.19 και σύμφωνα με τους Robert et al. (2000), η εκ των υστέρων κατανομή του πλήθους των καταστάσεων της μη παρατηρήσιμης μαρκοβιανής αλυσίδας έχει κορυφή στην τιμή $k = 3$ με εκ των υστέρων πιθανότητα 0.8725.

Στο Σχήμα 5.20 φαίνεται το προσομοιωμένο δείγμα σύμφωνα με τους δύο αλγορίθμους MCMC για 60000 επαναλήψεις μετά το burn-in των πρώτων 20000 (αναπαρίσταται κάθε 10η παρατήρηση). Είναι προφανές ότι σε αυτήν την περίπτωση ο περιορισμένος αλγόριθμος και ο ελεύθερος αλγόριθμος δίνουν ακριβώς την ίδια εικόνα, εκτός βέβαια από το label switching που καταγράφηκε μεταξύ των τιμών των δύο συνιστωσών με τις μεγαλύτερες τυπικές αποκλίσεις. Ύστερα από την αναδιάταξη του δείγματος που προέκυψε από τον ελεύθερο αλγόριθμο βάσει του Αλγορίθμου ECR, λάβαμε το δείγμα που απεικονίζεται στο Σχήμα 5.20(γ). Οι προσομοιωθείσες τιμές για κάθε συνιστώσα βρίσκονται σε αρκετά απομακρυσμένες περιοχές μεταξύ τους, συνεπώς σε αυτό το παράδειγμα είναι αναμενόμενο ότι οι δύο μέθοδοι αντιμετώπισης του φαινομένου label switching θα οδηγούν στα ίδια ακριβώς αποτελέσματα.

Οι εργοδικοί μέσοι σύμφωνα με τον Αλγόριθμο ECR δίνουν τις εκτιμήσεις

$$\begin{aligned}
 (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3) &= (1.04, 0.37, 2.83) \\
 \begin{pmatrix} \hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\ \hat{p}_{21} & \hat{p}_{22} & \hat{p}_{23} \\ \hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33} \end{pmatrix} &= \begin{pmatrix} 0.813 & 0.107 & 0.079 \\ 0.037 & 0.932 & 0.031 \\ 0.111 & 0.034 & 0.86 \end{pmatrix}.
 \end{aligned}$$



Σχήμα 5.20: Προσομοιωμένο δείγμα βάσει του δειγματολήπτη Gibbs για τα δεδομένα ανέμου. (α) Περιορισμένος αλγόριθμος, (β) Ελεύθερος αλγόριθμος, (γ) Αναδιάταξη του δείγματος του ελεύθερου αλγορίθμου βάσει του Αλγορίθμου ECR.

Σημειώνουμε ότι οι παραπάνω εκτιμήσεις είναι ακριβώς ίδιες με αυτές του περιορισμένου αλγορίθμου κάτι που, όπως τονίστηκε πριν, σε αυτήν την περίπτωση είναι αναμενόμενο.

5.4 Εφαρμογές σε Τυχαία Μαρκοβιανά Πεδία

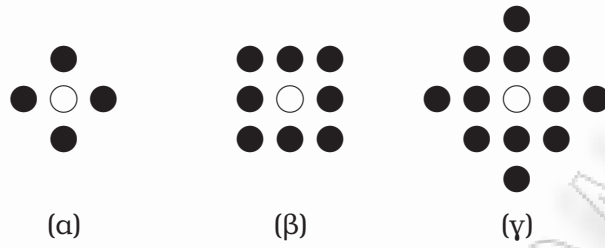
Σε αυτήν την ενότητα αντιμετωπίζουμε το φαινόμενο label switching σε μία γενίκευση των HMM και συγκεκριμένα σε μία κατηγορία τυχαίων μαρκοβιανών πεδίων (Markov random fields). Ειδικότερα, θα μας απασχολήσει το πρόβλημα της κατάτμησης διαταραγμένων ψηφιακών εικόνων, ένα σύγχρονο πεδίο εφαρμογής της μπεϋζιανής μεθοδολογίας αλλά και των μοντέλων μείζων κατανομών.

Γενικά, μία ψηφιακή εικόνα αποτελείται από ένα σύνολο $r \times c$ εικονοστοιχείων (pixels) τα οποία θα συμβολίζουμε με $z_i, i \in \mathcal{I}$, όπου το \mathcal{I} παριστάνει μία αυθαίρετη διάταξη του συνόλου των εικονοστοιχείων. Αυτό γίνεται για πρακτικούς λόγους στον συμβολισμό προκειμένου να δουλεύουμε σε ένα υποσύνολο των θετικών ακεραίων αντί ενός υποσυνόλου του $\mathbb{N} \times \mathbb{N}$. Κάθε εικονοστοιχείο παίρνει τιμές στο σύνολο $\{1, \dots, k\}$, όπου το k ισούται με το πλήθος των χρωμάτων που αποτελούν την εικόνα και συνήθως αναφέρεται στην κλίμακα αποχρώσεων του γκριζου (grayscale). Το πρόβλημα της ψηφιακής κατάτμησης εικόνας περιγράφεται ως ακολούθως: Σε πολλές πρακτικές εφαρμογές, αντί να παρατηρούμε την πραγματική εικόνα, δηλαδή το σύνολο των εικονοστοιχείων z , παρατηρούμε μία διαταραγμένη εικόνα y , στην οποία φαίνονται πολύ περισσότερα χρώματα από τα πραγματικά. Τέτοια προβλήματα παρατηρούνται σε δορυφορικές εικόνες (όπου ο θόρυβος μπορεί να οφείλεται στην ύπαρξη συννεφιάς στην ατμόσφαιρα), στα εγκεφαλογραφήματα (όπου ο θόρυ-

βος μπορεί να προέρχεται από σφάλματα του οργάνου), σε εικόνες παρακολούθησης της κυκλοφορίας αυτοκινητοδρόμων (όπου καταγράφεται μια διαταραγμένη εικόνα του αριθμού κυκλοφορίας κάποιου αυτοκινήτου λόγω του ότι αυτό βρίσκεται εν κινήσει), σε μηχανήματα αυτόματης αναγνώρισης προσώπων (όπου ο θόρυβος προέρχεται από το γεγονός ότι ένα πρόσωπο δεν έχει ποτέ την ίδια ακριβώς στάση με αυτή που είναι αποθηκευμένη στην μνήμη του μηχανήματος αναγνώρισης) κλπ. Ο σκοπός της ανάλυσης σε τέτοιου είδους προβλήματα είναι η κατάταξη κάθε παρατηρούμενου εικονοστοιχείου σε κάποιο από τα k δυνατά χρώματα ή, με άλλα λόγια, να γίνει συμπερασματολογία για την πραγματική (αλλά μη παρατηρηθείσα) εικόνα z , δοθείσης της παρατηρούμενης θορυβώδους εκδοχής αυτής y .

Είναι σαφές ότι το παραπάνω πρόβλημα έχει πολλές ομοιότητες με τα προβλήματα σε μοντέλα μείξεων κατανομών. Εδώ οι παρατηρήσεις που ανήκουν στην ίδια χρωματική περιοχή έχουν ενδεχομένως διαφορετικό θόρυβο από τις παρατηρήσεις που ανήκουν σε κάποια άλλη χρωματική περιοχή. Το γεγονός αυτό θυμίζει την ετερογένεια που καταγράφεται σε ένα σύνολο δεδομένων που προέρχεται από μία μείξη k κατανομών. Πιο συγκεκριμένα, η ανάκτηση της πραγματικής εικόνας z είναι το ίδιο πρόβλημα με την ανάκτηση του πιο πιθανού διανύσματος κατάταξης ενός συνόλου παρατηρήσεων από μία μείξη κατανομών. Η διαφορά βέβαια με τα απλά μοντέλα μείξεων κατανομών ή και με τα πιο σύνθετα HMM έγκειται στην υπόθεση εξάρτησης των παρατηρήσεων μεταξύ τους. Είναι προφανές ότι για να μοντελοποιηθεί στατιστικά μία εικόνα θα πρέπει να γενικευθεί η έννοια της μαρκοβιανής εξάρτησης στο διδιάστατο επίπεδο των εικονοστοιχείων. Η κατάλληλη θεωρία για το συγκεκριμένο πρόβλημα παρέχεται από τα τυχαία μαρκοβιανά πεδία.

Για μία πλήρη θεωρητική παρουσίαση των ιδιοτήτων των τυχαίων μαρκοβιανών πεδίων, ο αναγνώστης παραπέμπεται στο βιβλίο του Bremaud (1999). Στην συνέχεια περιγράφουμε εν συντομία κάποια βασικά χαρακτηριστικά τους. Καθώς πρόκειται για μια γενίκευση των μαρκοβιανών διαδικασιών σε περισσότερες διαστάσεις, μία πρώτη απαίτηση είναι ο ορισμός μιας κατάλληλης έννοιας γεινίασης. Το $i \in \mathcal{I}$ στοιχείο της εικόνας είναι γείτονας με το j (συμβολικά $i \sim j$) αν και μόνον αν το j είναι γείτονας του i . Συμβατικά, θεωρούμε ότι το i δεν είναι γείτονας του εαυτού του. Οι πιο συνηθισμένες επιλογές γειτονικών δομών είναι οι τέσσερις, οκτώ και δώδεκα πλησιέστεροι γείτονες (δες Σχήμα 5.21). Ένα *τυχαίο πεδίο* στο \mathcal{I} είναι μια συλλογή τυχαίων μεταβλητών $\{z_i; i \in \mathcal{I}\}$ όπου κάθε z_i παίρνει τιμές σε ένα πεπερασμένο σύνολο. Προφανώς, η ενδιαφέρουσα περίπτωση για τις πραγματικές εφαρμογές είναι η περίπτωση όπου τα z_i είναι εξαρτημένες τυχαίες μεταβλητές.



Σχήμα 5.21: Μερικά συστήματα γειννίασης στο επίπεδο των εικονοστοιχείων: (α) 4, (β) 8 και (γ) 12 πλησιέστεροι γείτονες.

Αν συμβολίσουμε με $n(i)$ το σύνολο γειτόνων του $i \in \mathcal{I}$ και αν το $z_A = \{z_i; i \in A\}$ είναι το υποσύνολο του z για τους δείκτες εκείνους που ανήκουν στο υποσύνολο $A \subset \mathcal{I}$, τότε το $z_{n(i)}$ δηλώνει το σύνολο τιμών που έχουν οι γείτονες του i . Η επέκταση της μαρκοβιανής εξάρτησης σε τέτοιου είδους δομές λοιπόν, υποθέτει μόνο εξάρτηση από τους γείτονες. Ας συμβολίσουμε με $z_{-A} = \{z_i; i \notin A\}$ τα στοιχεία εκείνα που δεν ανήκουν σε έναν δοθέν υποσύνολο $A \subset \mathcal{I}$. Τότε, ένα τυχαίο μαρκοβιανό πεδίο είναι ένα τυχαίο πεδίο όπου η δεσμευμένη κατανομή κάθε εικονοστοιχείου δοθέντων των υπολοίπων, εξαρτάται μόνο από τις τιμές των γειτονικών του εικονοστοιχείων, δηλαδή όταν ισχύει ότι

$$\pi(z_i | z_{-i}) = \pi(z_i | z_{n(i)}), \quad \forall i \in \mathcal{I}.$$

Ας υποθέσουμε ότι ο αριθμός των δυνατών χρωμάτων στην πραγματική εικόνα είναι ίσος με k . Η συνήθης εκ των προτέρων κατανομή στο z είναι ένα μοντέλο Potts με k κατηγορίες, δηλαδή ένα τυχαίο μαρκοβιανό πεδίο με από κοινού εκ των προτέρων κατανομή

$$f(z|\beta) = \frac{1}{C(\beta)} \exp\left\{\beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i)\right\},$$

όπου $C(\beta)$ είναι η σταθερά κανονικοποίησης. Η παράμετρος β παίζει τον ρόλο της «θερμοκρασίας» του μοντέλου: Όσο μεγαλώνει η τιμή του, τόσο πιο ομογενής γίνεται η εικόνα. Να σημειωθεί εδώ ότι υπάρχει μία κρίσιμη τιμή του β , $\beta_c = 2.269185$, τέτοια ώστε αν $\beta > \beta_c$ τότε η μαρκοβιανή αλυσίδα που προσομοιώνεται από τον δειγματολήπτη Gibbs που θα περιγραφεί στην συνέχεια, συγκλίνει σε μία από k δυνατές στάσιμες κατανομές οι οποίες αντιστοιχούν σε εκ των υστέρων εικόνες ενός χρώματος, αναλόγως την αρχική της τιμή. Το φαινόμενο αυτό καλείται phase transition· δες Marin and Robert (2007).

Για μία πλήρως μπεϋζιανή προσέγγιση του προβλήματος η παράμετρος β θεωρείται τυχαία μεταβλητή. Ακολουθώντας τους Marin and Robert (2007), υποθέτουμε εκ των προτέρων ότι $\beta \sim \mathcal{U}(0, 2)$. Παρ' όλ' αυτά, υπάρχει μία προφανής δυσκολία στην ανανέωση αυτής της παραμέτρου μέσω κάποιου δειγματολήπτη καθώς ο όρος $C(\beta)$, όπου εξ ορισμού ισχύει ότι

$$C(\beta) = \sum_{\mathbf{z}} \exp\left\{\beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i)\right\},$$

είναι πρακτικώς αδύνατο να υπολογιστεί αναλυτικά ακόμα και για πολύ μικρές εικόνες. Για να αντιμετωπίσουμε τη συγκεκριμένη δυσκολία ακολουθούμε την μέθοδο που προτείνεται στους Marin and Robert (2007) ώστε να καταλήξουμε σε μια προσέγγιση, $\hat{r}(\beta_1, \beta_0)$, του λόγου $C(\beta_1)/C(\beta_0)$ για οποιοδήποτε ζεύγος τιμών β_1, β_0 . Ακολούθως, αυτή η προσέγγιση ενσωματώνεται σε ένα βήμα Metropolis-Hastings για την ανανέωση του β . Για περισσότερες λεπτομέρειες ο αναγνώστης παραπέμπεται στους Marin and Robert (2007, σελ. 240).

Δοθέντος \mathbf{z} , υποθέτουμε ότι το σύνολο παρατηρήσεων (δηλαδή, το παρατηρηθέν χρώμα για κάθε ένα από τα εικονοστοιχεία), \mathbf{y} , αποτελείται από ανεξάρτητες κανονικές τυχαίες μεταβλητές,

$$f(\mathbf{y}|\mathbf{z}, \sigma^2, \boldsymbol{\mu}) = \prod_{i \in \mathcal{I}} \frac{1}{(2\pi\sigma_{z_i}^2)^{1/2}} \exp\left\{-\frac{(y_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right\}, \quad (5.8)$$

περιορισμένες στο διάστημα $[0, 255]$ (όπου αντιστοιχεί στην κλίμακα αποχρώσεων του γκριζου). Σύμφωνα με τους Marin and Robert (2007), η υπόθεση αυτή είναι ελαφρώς προσεγγιστική καθώς τα y_i είναι ακέραιοι μεταξύ του 0 και 255, αλλά κάτι τέτοιο είναι πιο πρακτικό από μία παραμετροποιημένη διακριτή κατανομή στο σύνολο $\{0, \dots, 255\}$. Η απο κοινού συνάρτηση πυκνότητας των \mathbf{y}, \mathbf{z} γράφεται ως

$$f(\mathbf{y}, \mathbf{z}|\sigma^2, \boldsymbol{\mu}, \beta) = \frac{1}{C(\beta)} \exp\left\{\beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i)\right\} \prod_{i \in \mathcal{I}} \frac{1}{(2\pi\sigma_{z_i}^2)^{1/2}} \exp\left(-\frac{(y_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right). \quad (5.9)$$

Σημειώνουμε ότι το παραπάνω μοντέλο είναι πιο γενικό απ' ό,τι αυτό των Marin and Robert (2007) διότι θεωρούμε διαφορετικές διασπορές σε κάθε κατάσταση, ενώ οι Marin and Robert (2007) θεωρούν $\sigma_j^2 \equiv \sigma^2, \forall j = 1, \dots, k$. Ωστόσο, στις εφαρμογές που ακολουθούν θα χρησιμοποιηθούν και οι δύο επιλογές για την διασπορά.

Οι παράμετροι του θορύβου στην εικόνα, $\boldsymbol{\mu}, \sigma^2$, θεωρούνται εκ των προτέρων ανεξάρτητες τυχαίες μεταβλητές με $\mu_j \sim \mathcal{U}(0, 255)$, $\sigma_j^2 \sim \mathcal{IG}(g_0, G_0)$, $j = 1, \dots, k$. Σημειώνουμε ότι η εκ των προτέρων κατανομή στα μ_j είναι ίδια με αυτή των Marin

and Robert (2007) εκτός του ότι δεν υποθέτουμε εκ των προτέρων διάταξη μεταξύ τους. Από την άλλη πλευρά, οι διασπορές ακολουθούν μια αντίστροφη κατανομή γάμμα αντί της καταχρηστικής κατανομής που χρησιμοποιείται από τους Marin and Robert (2007) λόγω του ότι εδώ οι διασπορές είναι διαφορετικές μεταξύ των καταστάσεων και έτσι μία καταχρηστική εκ των προτέρων κατανομή θα οδηγούσε σε μία καταχρηστική εκ των υστέρων κατανομή.

Σύμφωνα με τα παραπάνω έχουμε ότι η απο κοινού εκ των υστέρων κατανομή όλων των παραμέτρων του μοντέλου είναι

$$\begin{aligned} \pi(\mathbf{z}, \beta, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 | \mathbf{y}) &\propto \pi(\beta, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \times \frac{1}{C(\beta)} \exp\left\{\beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i)\right\} \\ &\times \prod_{i \in \mathcal{I}} \frac{1}{(2\pi\sigma_{z_i}^2)^{1/2}} \exp\left\{-\frac{(y_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right\} \end{aligned}$$

και είναι προφανές ότι πρόκειται για μια αναλλοίωτη κατανομή ως προς τις μεταθέσεις των δεικτών $\{1, \dots, k\}$, καθώς

$$\pi(\mathbf{z}, \beta, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 | \mathbf{y}) = \pi(\boldsymbol{\tau}^{-1}\mathbf{z}, \beta, \boldsymbol{\tau}\boldsymbol{\mu}, \boldsymbol{\tau}\boldsymbol{\sigma}^2 | \mathbf{y}), \quad \forall \boldsymbol{\tau} \in \mathcal{T}_k.$$

Ως εκ τούτου, το φαινόμενο label switching συμβαίνει και σε τέτοια τυχαία μαρκοβιανά πεδία. Οι Marin and Robert (2007) είναι οι μοναδικοί που έχουν αναφέρει κάτι σχετικά με το φαινόμενο label switching σε τέτοιου είδους εφαρμογές, αλλά απλώς χρησιμοποιούν την εκ των προτέρων διάταξη των μέσων, $\mu_1 < \dots < \mu_k$, για την ταυτοποίηση των παραμέτρων. Στην συνέχεια θα συγκρίνουμε σε παραδείγματα την συγκεκριμένη μεθοδολογία με τον Αλγόριθμο ECR.

Για την προσομοίωση από την παραπάνω κατανομή-στόχο θα χρησιμοποιήσουμε τον αλγόριθμο Metropolis-within-Gibbs που περιγράφεται από τους Marin and Robert (2007) προσαρμοσμένο στις αλλαγές που υπάρχουν στο μοντέλο μας. Έτσι, αφού δώσουμε κάποιες αρχικές τιμές στις παραμέτρους, σε κάθε βήμα του αλγορίθμου το παραμετρικό διάνυσμα ανανεώνεται προσομοιώνοντας τα $\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ από τις αντίστοιχες πλήρεις δεσμευμένες εκ των υστέρων κατανομές τους, οι οποίες είναι

$$\begin{aligned} P(z_i = j | \dots) &\propto \exp\left\{\beta \sum_{j \sim i} I(z_i = j) - \frac{1}{2\sigma_j^2}(y_i - \mu_j)^2\right\}, \quad i \in \mathcal{I}, \\ \mu_j | \dots &\sim \mathcal{N}\left(\frac{\sum_{i \in \mathcal{I}} I(z_i = j)y_i}{\sum_{i \in \mathcal{I}} I(z_i = j)}, \frac{\sigma_j^2}{\sum_{i \in \mathcal{I}} I(z_i = j)}\right), \quad j = 1, \dots, k, \\ \sigma_j^2 | \dots &\sim \mathcal{IG}\left(g_0 + \frac{|\mathcal{I}|}{2}, \left[\frac{1}{G_0} + \frac{\sum_{i \in \mathcal{I}} I(z_i = j)(y_i - \mu_j)^2}{2}\right]^{-1}\right), \quad j = 1, \dots, k. \end{aligned}$$

Τέλος, η ανανέωση της παραμέτρου β κατά την επανάληψη t βασίζεται στην κατανομή πρότασης $\tilde{\beta} \sim \mathcal{U}(\beta^{(t-1)} - 0.05, \beta^{(t-1)} + 0.05)$, η οποία έχει την (προσεγγιστική) πιθανότητα αποδοχής που δίνεται από τους Marin and Robert (2007)

$$\min \left\{ 1, \hat{r}(\beta^{(t-1)}, \tilde{\beta}) \exp \left[(\tilde{\beta} - \beta^{(t-1)}) \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i) \right] \right\}.$$

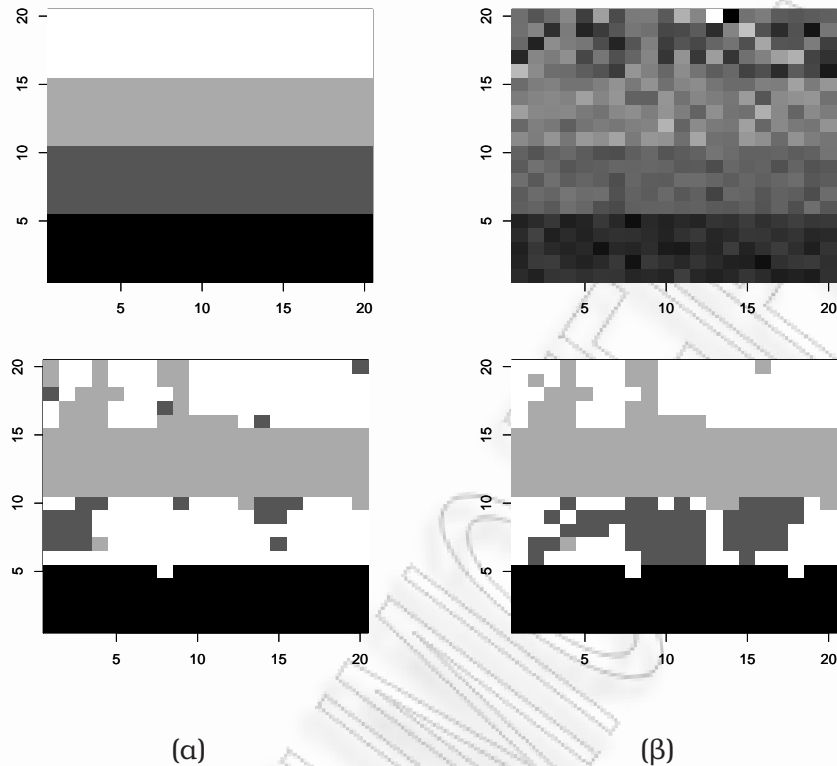
Πριν προχωρήσουμε στις εφαρμογές, θα πρέπει να επαληθεύσουμε τις ιδιότητες της Ενότητας 3.2 ώστε να εξασφαλίσουμε τα θεωρητικά αποτελέσματα του προηγούμενου κεφαλαίου. Αρχικά, ως παρατηρήσουμε ότι από την συζήτηση που προηγήθηκε, έχουμε ότι $\eta = ((\mu_1, \sigma_1^2), \dots, (\mu_k, \sigma_k^2))$ και συνεπώς $\mathcal{H} = (\mathbb{R} \times (0, \infty))^k$. Επομένως, για κάθε $\tau \in \mathcal{T}_k$ εξασφαλίζεται η ισχύς της (4.2). Προφανώς, οι εκ των προτέρων υποθέσεις στις παραμέτρους εξασφαλίζουν την ισχύ της ιδιότητας (4.3). Επίσης, η (5.8) συνεπάγεται ότι η (4.5) ισχύει. Τέλος, είναι εύκολο να δούμε ότι η αποκοινού συνάρτηση πυκνότητας των πλήρων δεδομένων έχει πράγματι την ιδιότητα (4.4). Από την (5.9), λαμβάνουμε διαδοχικά ότι για κάθε $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$

$$\begin{aligned} f(\mathbf{y}, \tau \mathbf{z} | \eta) &= f(\mathbf{y}, \tau \mathbf{z} | \beta, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \\ &= \frac{1}{C(\beta)} \exp \left\{ \beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(t_{z_j} = t_{z_i}) \right\} \\ &\quad \times \prod_{i \in \mathcal{I}} \frac{1}{(2\pi\sigma_{t_{z_i}}^2)^{1/2}} \exp \left(-\frac{(y_i - \mu_{t_{z_i}})^2}{2\sigma_{t_{z_i}}^2} \right) \\ &= \frac{1}{C(\beta)} \exp \left\{ \beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i) \right\} \\ &\quad \times \prod_{i \in \mathcal{I}} \frac{1}{(2\pi(\tau\boldsymbol{\sigma}^2)_{z_i})^{1/2}} \exp \left(-\frac{(y_i - (\tau\boldsymbol{\mu})_{z_i})^2}{2(\tau\boldsymbol{\sigma}^2)_{z_i}} \right) \\ &= f(\mathbf{y}, \mathbf{z} | \beta, \tau\boldsymbol{\mu}, \tau\boldsymbol{\sigma}^2) \\ &= f(\mathbf{y}, \mathbf{z} | \tau\eta). \end{aligned}$$

5.4.1 Μπεϋζιανή κατάτμηση εικόνων

Παράδειγμα 5.4.1. Προσομοιωμένα δεδομένα.

Σε αυτήν την ενότητα συγκρίνεται η προτεινόμενη μέθοδος αντιμετώπισης του φαινομένου label switching με τα αποτελέσματα των περιορισμών διάκρισης, την μοναδική τεχνική που έχει εφαρμοστεί μέχρι στιγμής σε προβλήματα μπεϋζιανής κατάτμησης ψηφιακών εικόνων. Για να δείξουμε τις επιπτώσεις που επιφέρει στην συμπερασματολογία η χρήση περιορισμών διάκρισης χρησιμοποιούμε αρχικά μία μικρή εικόνα

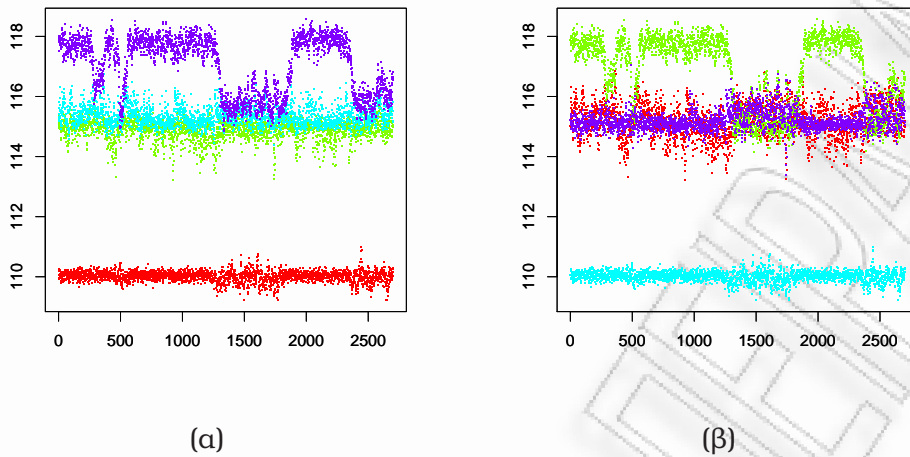


Σχήμα 5.22: Πρώτη γραμμή: Πραγματική (μη παρατηρήσιμη) εικόνα (α) και παρατηρήσιμη (θορυβώδης) εικόνα (β). Δεύτερη γραμμή: κατάτμηση εικόνας βάσει του (α) περιορισμού διάταξης των μέσων και (β) του Αλγορίθμου ECR.

μεγέθους 20×20 εικονοστοιχείων με $k = 4$ αποχρώσεις του γκριζου, η οποία απεικονίζεται στο Σχήμα 5.22. Αυτή παίζει τον ρόλο της μη παρατηρηθείσας πραγματικής εικόνας, την οποία όπως είπαμε στην προηγούμενη ενότητα συμβολίζουμε με z . Η παρατηρούμενη θορυβώδης εικόνα προκύπτει από την πραγματική προσθέτοντας ανεξάρτητο θόρυβο σε κάθε μια από τις διαφορετικές χρωματικές ομάδες αυτής. Οι θόρυβοι που προσθέσαμε στις τέσσερις αποχρώσεις ακολουθούν κανονικές κατανομές με παραμέτρους

$$\mu = (110, 115, 118, 115), \quad \sigma^2 = (1, 1, 2, 4).$$

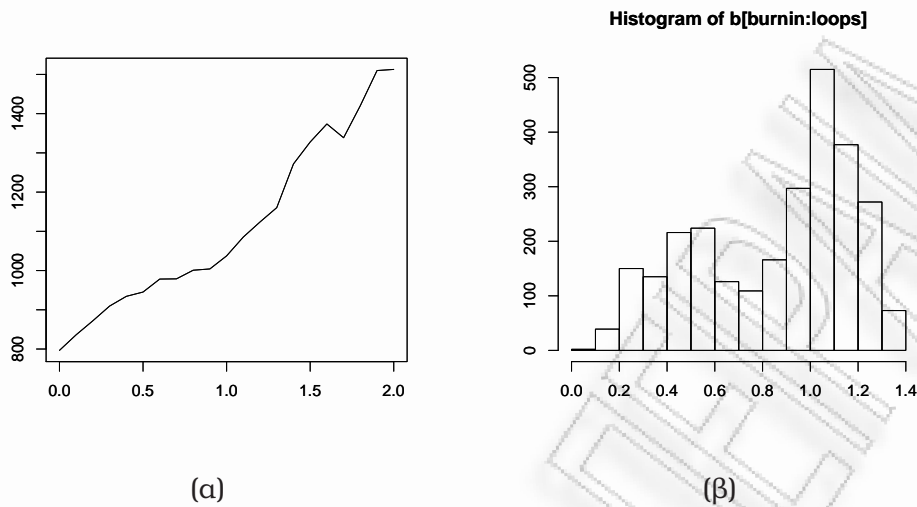
Στην πρώτη γραμμή του Σχήματος 5.22(β) εμφανίζεται η διαταραγμένη εικόνα. Σκοπός της ανάλυσής μας είναι να ανακατασκευάσουμε την πραγματική εικόνα και αυτό θα γίνει εκτελώντας τον αλγόριθμο MCMC που περιγράφηκε στην προηγούμενη ενότητα. Για τις παραμέτρους της εκ των προτέρων κατανομής των διασπορών θέτου-



Σχήμα 5.23: Αναδιατεταγμένες τιμές του $(\mu_1, \mu_2, \mu_3, \mu_4)^{(i)}$ σύμφωνα με τον (α) περιορισμό διάκρισης και (β) τον αλγόριθμο ECR.

με $g_0 = 1$ και $G_0 = 1/5$. Σημειώνουμε ότι οι συγκεκριμένες τιμές αποτελούν αρκετά μη πληροφοριακές επιλογές. Επίσης επιλέξαμε το σύστημα τεσσάρων πλησιέστερων γειτόνων. Στη συνέχεια το δείγμα αναδιατάσσεται βάσει του Αλγορίθμου ECR αλλά και βάσει του περιορισμού διάταξης των μέσων που χρησιμοποιούν οι Marin and Robert (2007). Η κατατημένη εικόνα προκύπτει από τις πιο πιθανές αναδιατεταγμένες κατατάξεις των εικονοστοιχείων στις $k = 4$ χρωματικές περιοχές. Για την εκτέλεση του αλγορίθμου θα πρέπει πρώτα να κατασκευάσουμε μία προσέγγιση του $f(\beta) = \mathbb{E}_\beta[\sum_{\mathbf{z}} \exp\{\beta \sum_{i \in \mathcal{I}} \sum_{j \sim i} I(z_j = z_i)\}]$. Ακολουθώντας την μεθοδολογία που περιγράφεται στους Marin and Robert (2007), καταλήξαμε στην προσέγγιση που φαίνεται στο Σχήμα 5.24(a).

Ο αλγόριθμος Metropolis-within-Gibbs εκτελέστηκε για 3000 επαναλήψεις (οι πρώτες 300 διεγράφησαν ως περίοδος burn-in) και ακολούθως αναδιατάξαμε το δείγμα που προέκυψε. Στη δεύτερη γραμμή του Σχήματος 5.22 φαίνονται οι προκύπτουσες κατατημήσεις της παρατηρηθείσας εικόνας βάσει του περιορισμού $\mu_1 < \mu_2 < \mu_3 < \mu_4$ και του Αλγορίθμου ECR, αντίστοιχα. Επί πλέον, στο Σχήμα 5.23 παρουσιάζονται οι αναδιατεταγμένες τιμές των μέσων τιμών κάθε συνιστώσας. Από τις αναδιατεταγμένες τιμές των μέσων, παρατηρούμε ότι η εκ των υστέρων κατανομή παρουσιάζει αρκετές κορυφές, καθώς υπάρχουν έντονες εναλλαγές μεταξύ των περιοχών υψηλής εκ των υστέρων πυκνότητας. Αυτό ισχύει και για την εκ των υστέρων κατανομή του β επίσης, όπου υπάρχει μία κύρια κορυφή κοντά στο 1.15 και μία



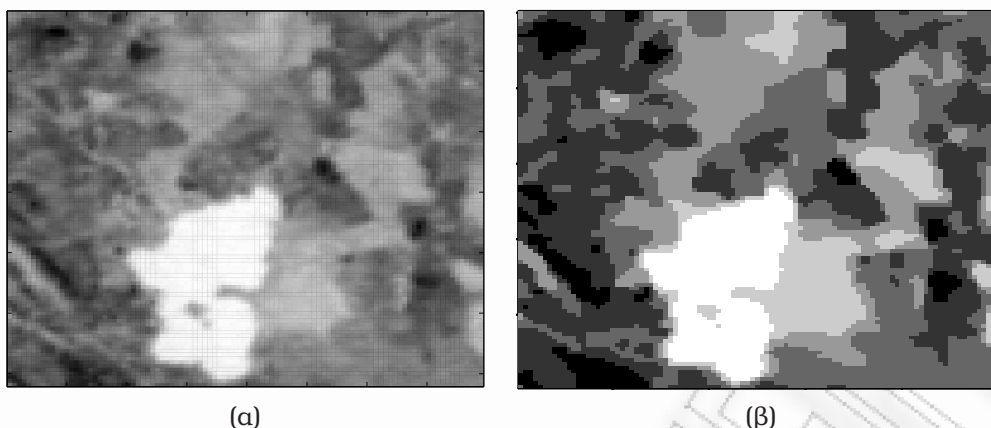
Σχήμα 5.24: (α) Προσέγγιση του $f(\beta)$ του μοντέλου Potts με σύστημα τεσσάρων πλησιέστερων γειτόνων για την εικόνα διαστάσεων 20×20 . Η προσέγγιση βασίζεται σε διαμέριση του διαστήματος $(0, 2)$ σε 20 υποδιαστήματα ίσου μήκους και 3000 επαναλήψεις για κάθε υποδιάστημα του αλγορίθμου MCMC με σταθερό β μετά το burn-in των πρώτων 300. (β) Το ιστόγραμμα των προσομοιωμένων τιμών του β .

μικρότερη κοντά στο 0.5. Αυτό φαίνεται στο Σχήμα 5.24(β). Όπως είναι φανερό από τις κατατμημένες εικόνες, συμπεραίνουμε ότι η κατάτμηση βάσει του Αλγορίθμου ECR πλησιάζει περισσότερο την πραγματική εικόνα. Για να υποστηρίξουμε αυτόν τον ισχυρισμό, εκτελέσαμε 50 ανεξάρτητες επαναλήψεις της διαδικασίας των προσομοιώσεων και αναδιατάξεων. Το 99% ασυμπτωτικό διάστημα εμπιστοσύνης για τον λόγο των μέσων λανθασμένων κατατάξεων που προκύπτουν από τον Αλγόριθμο ECR προς εκείνων που προκύπτουν από τους περιορισμούς διάκρισης υπολογίστηκε ως $(0.628, 0.742)$. Συνεπώς, είναι ξεκάθαρη η ανώτερότητα της προτεινόμενης μεθόδου σε αυτό το παράδειγμα.

Παράδειγμα 5.4.2. Λίμνη Menteith.

Στην συνέχεια παρουσιάζουμε ένα παράδειγμα πραγματικής εικόνας. Στο Σχήμα 5.25(a) φαίνεται μία δορυφορική φωτογραφία της λίμνης Menteith στη Σκωτία. Η ψηφιακή εικόνα έχει διάσταση 100×100 . Ακολουθώντας την προσέγγιση των Marin and Robert (2007) θεωρούμε $k = 6$ ομογενείς περιοχές με ίδιες διασπορές με (καταχρηστική) εκ των προτέρων κατανομή

$$\sigma^2 \sim \frac{1}{\sigma^2} I(\sigma^2 > 0).$$



Σχήμα 5.25: Η περιοχή της λίμνης Menteith: Παρατηρηθείσα δορυφορική εικόνα (α) και κατατημένη εικόνα (β) σε $k = 6$ ομογενείς περιοχές σύμφωνα με τον Αλγόριθμο ECR.

Σημειώνουμε ότι το παραπάνω αντιστοιχεί σε μία (καταχρηστική) ομοιόμορφη εκ των προτέρων κατανομή για το $\log \sigma$. Οι υπόλοιπες εκ των προτέρων υποθέσεις παραμένουν ίδιες με αυτές της Ενότητας 5.4. Η αντίστοιχη πλήρης δεσμευμένη εκ των υστέρων κατανομή της κοινής διασποράς είναι

$$\sigma^2 \sim \mathcal{IG}(|\mathcal{I}|^2/2, \sum_{i \in \mathcal{I}} (y_i - \mu_{z_i})^2/2).$$

Υστερα από 1000 επαναλήψεις (και μετά το burn-in των πρώτων 500 επαναλήψεων) του αλγορίθμου MCMC που περιγράφεται στους Marin and Robert (2007), χωρίς βέβαια την υιοθέτηση των εκ των προτέρων περιορισμών στις μέσες τιμές, οι παράμετροι συνέκλιναν στις περιοχές υψηλής εκ των υστέρων πυκνότητας και ακολούθως οι προσομοιωμένες τιμές των $\mu_j^{(t)}$, $j = 1, \dots, 6$, $t = 1, \dots, 1000$ αναδιατάχθηκαν μέσω του Αλγορίθμου ECR. Σημειώνεται ότι η παράμετρος β τέθηκε ίση με $\beta = 1.3$. Η υπόθεση αυτή δικαιολογείται από το γεγονός ότι η εκτιμηθείσα περιοχή υψηλής εκ των υστέρων πιθανότητας σύμφωνα με τους Marin and Robert (2007) εξαντλείται στο διάστημα $(1.295, 1.315)$ και από το ότι η παράμετρος αυτή δεν επηρεάζεται από το label switching.

Η κατατημένη εικόνα που προκύπτει βάσει της μεθόδου μας, φαίνεται στο Σχήμα 5.25.(β). Τα αποτελέσματά μας είναι σχεδόν ίδια με αυτά των Marin and Robert (2007) όπου υποτίθεται εκ των προτέρων διάταξη των μέσων. Αυτό φυσικά συμβαίνει εδώ διότι οι περιοχές υψηλής εκ των υστέρων πυκνότητας των έξι μέσων είναι αρκετά μακριά μεταξύ τους.

Παράρτημα Α

Παράρτημα Κεφαλαίου 2

A.1 Η γενική μορφή του αλγορίθμου RJMCMC

Τα μοντέλα μείξεων κατανομών με άγνωστο πλήθος συνιστωσών είναι μόνο μία ειδική περίπτωση μοντέλων μεταβλητής διάστασης. Η διαδικασία εκτίμησης του αριθμού συνιστωσών k είναι ένα πρόβλημα επιλογής μοντέλου. Στην Μπεϋζιανή Συμπεραματολογία αυτό ισοδύναμι με την εκτίμηση των εκ των υστέρων πιθανοτήτων των υποψήφιων μοντέλων και επιλογή εκείνου που έχει την μεγαλύτερη εκ των υστέρων πιθανότητα. Έστω \mathcal{K} ένα αριθμήσιμο σύνολο και $\{\mathcal{M}_k, k \in \mathcal{K}\}$ μία συλλογή μοντέλων η διάσταση των οποίων ενδέχεται να ποικίλλει. Οι παράμετροι του μοντέλου \mathcal{M}_k , $\omega = (k, \omega_k) \in \Omega_k$, ορίζονται δοθέντος του k .

Η αδυναμία των τυπικών μεθοδολογιών MCMC, όπως ο δειγματολήπτης Gibbs και ο αλγόριθμος Metropolis-Hastings, σε προβλήματα επιλογής μοντέλου έγκειται στο εξής: Ενώ μπορούν να χρησιμοποιηθούν για να εξερευνήσουν κάθε ένα μοντέλο ξεχωριστά, δεν μπορούν να γεφυρώσουν τα ανταγωνιστικά μοντέλα μεταξύ τους και, κατ' επέκταση, να εξερευνήσουν τον χώρο $\Omega = \cup_{k \in \mathcal{K}} \Omega_k$ όλων των δυνατών μοντέλων ταυτόχρονα. Αυτό έγινε δυνατό με τον αλγόριθμο reversible jump MCMC του Green (1995). Ο συγκεκριμένος αλγόριθμος προσομοιώνει μία εργοδική μαρκοβιανή αλυσίδα $\omega^{(t)}$, $t = 1, 2, \dots$ με οριακή κατανομή την εκ των υστέρων κατανομή των ανταγωνιστικών μοντέλων. Σε κάθε επανάληψη προτείνονται κατάλληλες κινήσεις μεταπήδησης από ένα μοντέλο σε ένα άλλο οι οποίες γίνονται δεκτές βάσει κάποιας συγκεκριμένης πιθανότητας αποδοχής.

Ας υποθέσουμε ότι βρισκόμαστε στο μοντέλο \mathcal{M}_m και θέλουμε να προτείνουμε μία μετάβαση στο μοντέλο \mathcal{M}_n , με παραμετρικούς χώρους Ω_m και Ω_n αντίστοιχα, οι οποίοι μπορεί να είναι διαφορετικών διαστάσεων¹. Η κίνηση αυτού του τύπου θα

¹Αν τα ανταγωνιστικά μοντέλα καθορίζονται πλήρως από την διάστασή τους (κάτι που δεν ισχύει γενικά), τότε καλό είναι να συσχετίζουμε το m με την διάσταση του μοντέλου. Η παρατήρηση αυτή

είναι μία από τις δυνατές μεταπηδήσεις, οπότε έστω ότι π_{mn} είναι η πιθανότητα της συγκεκριμένης πρότασης. Αντίστοιχα, έστω ότι π_{nm} είναι η πιθανότητα πρότασης της αντίστροφης κίνησης. Η βασική ιδέα του Green (1995) είναι η συμπλήρωση των χώρων Ω_m και Ω_n με εικονικούς χώρους, \mathcal{U}_n και \mathcal{U}_m αντίστοιχα, με σκοπό τον ορισμό ενός ένα-προς-ένα μετασχηματισμού μεταξύ αυτών. Αυτό συνήθως (αλλά όχι απαραίτητα) γίνεται συμπληρώνοντας μόνο τον χώρο μικρότερης διάστασης. Στην γενική περίπτωση όπου και οι δύο χώροι συμπληρώνονται, η προτεινόμενη ευθεία και αντίστροφη μετάβαση μπορεί να γίνει με την παραγωγή κάποιων επί πλέον τυχαίων διανυσμάτων $u_n \sim g_n(u_n)$ και $u_m \sim g_m(u_m)$. Οι g_n και g_m ενδέχεται να εξαρτώνται και από τα ω_m και ω_n , αντίστοιχα, αλλά αυτό δεν ενσωματώνεται στην έκφρασή τους για λόγους ευκολίας στον συμβολισμό. Η μετάβαση από τον χώρο Ω_m στον χώρο Ω_n γίνεται μέσω κάποιων μετασχηματισμών

$$\begin{aligned} (\omega_n, u_m) &= \mathcal{F}_{m \rightarrow n}(\omega_m, u_n), \quad u_n \in \mathcal{U}_n \\ (\omega_m, u_n) &= \mathcal{F}_{n \rightarrow m}(\omega_n, u_m), \quad u_m \in \mathcal{U}_m \end{aligned}$$

όπου $\mathcal{F}_{m \rightarrow n} : \Omega_m \times \mathcal{U}_n \mapsto \Omega_n \times \mathcal{U}_m$ και $\mathcal{F}_{n \rightarrow m} : \Omega_n \times \mathcal{U}_m \mapsto \Omega_m \times \mathcal{U}_n$, που γεφυρώνουν τις παραμέτρους των δύο διαφορετικών χώρων. Η συγκεκριμένη διαδικασία έχει οριακή κατανομή την εκ των υστέρων κατανομή των μοντέλων αν ικανοποιείται η υπόθεση συμφωνίας των διαστάσεων του Green. Σύμφωνα με αυτήν, εξασφαλίζεται ότι οι εμπλεκόμενοι μετασχηματισμοί δρουν σε χώρους ίσων διαστάσεων και ο ένας είναι ο αντίστροφος του άλλου. Μία πιο απλοποιημένη διατύπωση της υπόθεσης συμφωνίας διαστάσεων συνοψίζεται στο ότι θα πρέπει

$$\begin{cases} \dim \Omega_m + \dim \mathcal{U}_n = \dim \Omega_n + \dim \mathcal{U}_m \\ \mathcal{F}_{n \rightarrow m}(\cdot) = \mathcal{F}_{m \rightarrow n}^{-1}(\cdot). \end{cases} \quad (\text{A.1})$$

Έτσι, υπό την ισχύ της (A.1), αποδεικνύεται (Green, 1995) ότι η αλυσίδα έχει οριακή κατανομή $\pi(\cdot)$ στον χώρο των δυνατών μοντέλων $\Omega = \cup_k \Omega_k$, η οποία επιτυγχάνει την συνθήκη λεπτομερούς ισορροπίας

$$\int_A \int_B \pi(d\omega_m) P(\omega_m, d\omega_n) = \int_B \int_A \pi(d\omega_n) P(\omega_n, d\omega_m), \quad (\text{A.2})$$

για κάθε $A, B \in \mathcal{B}(\Omega)$, όπου $\mathcal{B}(\Omega)$ είναι η σ -άλγεβρα Borel του Ω και $P(\omega_m, d\omega_n)$ ο πυρήνας μετάβασης από το ω_m στο $d\omega_n$.

Κατ' αυτόν τον τρόπο, η μετάβαση από το (ω_m, u_n) στο (ω_n, u_m) μπορεί να θεωρηθεί ως μία συνηθισμένη κίνηση Metropolis-Hastings με την διαφορά βέβαια ότι δεν είναι απαραίτητη αλλά βοηθά στην κατανόηση της μεθόδου.

αλλάζει η κατανομή-στόχος: Αντί της στάσιμης κατανομής $\pi_m(\omega_m)g_n(u_n)$ η μετάβαση έχει στόχο την κατανομή $\pi_n(\omega_n)g_m(u_m)$, κάτι το οποίο όμως δεν αποτελεί πρόβλημα για τον αλγόριθμο Metropolis-Hastings (δες Marin and Robert, 2007). Έτσι λοιπόν, η γενική μορφή του αλγορίθμου RJMCMC έχει ως εξής.

Αλγόριθμος A.1.1. *Ο γενικός αλγόριθμος RJMCMC*

Έστω ότι κατά την επανάληψη $t = 1, 2, \dots$ η αλυσίδα βρίσκεται στην κατάσταση $\omega^{(t)} = (m, \omega_m)$:

1. Πρότεινε μετάβαση στο μοντέλο \mathcal{M}_n με πιθανότητα π_{mn} .
2. Παρήγαγε $u_n \sim g_n(u_n)$ και θέσε $(\omega_n, u_m) = \mathcal{F}_{m \rightarrow n}(\omega_m, u_n)$.
3. Θέσε $\omega^{(t+1)} = (n, \omega_n)$ με πιθανότητα

$$\min \left(1, \frac{\pi(n, \omega_n) \pi_{nm} g_m(u_m)}{\pi(m, \omega_m) \pi_{mn} g_n(u_n)} \left| \frac{\partial \mathcal{F}_{m \rightarrow n}(\omega_m, u_n)}{\partial (\omega_m, u_n)} \right| \right),$$

διαφορετικά, θέσε $\omega^{(t+1)} = \omega^{(t)}$.

Στον παραπάνω αλγόριθμο ο τελευταίος όρος της πιθανότητας αποδοχής συμβολίζει την Ιακωβιανή του (αντιστρέψιμου) μετασχηματισμού $(\omega_m, u_n) \mapsto (\omega_n, u_m)$, ο οποίος γίνεται μέσω της $\mathcal{F}_{m \rightarrow n}$.

A.2 Απόδειξη Λήμματος 2.7.1

Για την πλήρη δεσμευμένη εκ των υστέρων κατανομή των συντελεστών παλινδρόμησης έχουμε

$$\begin{aligned} \beta'_j | \dots &\propto \exp\left\{-\frac{1}{2} \text{tr}[\Omega^{-1}(\beta'_j - M')' \Sigma_j^{-1}(\beta_j - M)]\right\} \times \\ &\exp\left\{-\frac{1}{2} \left[\sum_{i=1}^n I(z_i = j) (y_i - \beta'_j x_i)' \Sigma_j^{-1} (y_i - \beta'_j x_i) \right]\right\} \\ &= \exp\left\{-\frac{1}{2} \text{tr}[\Sigma_j^{-1}(\beta'_j - M') \Omega^{-1}(\beta_j - M)]\right\} \times \\ &\exp\left\{-\frac{1}{2} \sum_{i=1}^n I(z_i = j) \text{tr}[\Sigma_j^{-1} (y_i y_i' - y_i x_i' \beta'_j - \beta'_j x_i y_i' + \beta'_j x_i x_i' \beta_j)]\right\} \\ &= \exp\left\{\text{tr}[\Sigma_j^{-1}(\beta'_j \Omega^{-1} \beta_j - \beta'_j \Omega^{-1} \beta_j - \beta'_j \Omega^{-1} M - M' \Omega^{-1} \beta_j + M' \Omega^{-1} M)]\right\} \times \\ &\exp\left\{-\frac{1}{2} \sum_{i=1}^n I(z_i = j) \text{tr}[\Sigma_j^{-1} (y_i y_i' - y_i x_i' \beta_j - \beta'_j x_i y_i' \beta_j x_i' \beta_j)]\right\} \end{aligned}$$

Μετά από κατάλληλες αναδιατάξεις όρων στην παραπάνω έκφραση λαμβάνουμε ότι

$$\begin{aligned}
\beta'_j | \dots &\propto \exp\left\{-\frac{1}{2}\text{tr}[\Sigma_j^{-1}(\beta'_j(\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i')\beta_j - \right. \\
&\quad \left. - \beta'_j(\Omega^{-1}M + \sum_{i=1}^n I(z_i = j)x_i y_i') - (M'\Omega^{-1} + \sum_{i=1}^n I(z_i = j)y_i x_i')\beta_j)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\text{tr}[\Sigma_j^{-1}(\beta'_j(\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i') - \right. \\
&\quad \beta'_j(\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i')(\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i')^{-1} \times \\
&\quad (\Omega^{-1}M + \sum_{i=1}^n I(z_i = j)x_i y_i') - (M'\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i y_i') \times \\
&\quad \left. (\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i')^{-1}(\Omega^{-1} + \sum_{i=1}^n I(z_i = j)x_i x_i')^{-1})\beta_j)\right\}.
\end{aligned}$$

Αυτό σημαίνει ότι

$$\beta'_j | \mathbf{x}, \mathbf{y}, \Sigma, \Omega, M' \sim \mathcal{MN}_{m \times q}(U_j V_j^{-1}, V_j^{-1}, \Sigma_j), \quad j = 1, \dots, k \text{ ανεξάρτητα.}$$

Για την πλήρη δεσμευμένη εκ των υστέρων κατανομή των πινάκων διασπορών-συνδιασπορών έχουμε

$$\begin{aligned}
\Sigma_j | \dots &\sim |\Sigma_j|^{-(\zeta+m+1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\Psi \Sigma_j^{-1})\right\} \times \\
&\exp\left\{-\frac{1}{2}\text{tr}[\Sigma_j \sum_{i=1}^n I(z_i = j)(y_i - \beta'_j x_i)(y_i - \beta'_j x_i)']\right\} |\Sigma_j|^{-n_j/2} \times \\
&\exp\left\{-\frac{1}{2}\text{tr}[(\beta'_j - M')\Omega^{-1}(\beta_j - M)\Sigma_j^{-1}]\right\} |\Sigma_j|^{-q/2}
\end{aligned}$$

με τον πρώτο όρο να αντιστοιχεί στην εκ των προτέρων κατανομή του Σ_j , τον δεύτερο στην πιθανοφάνεια και τον τελευταίο στην εκ των προτέρων κατανομή του β'_j . Ύστερα από αναδιατάξεις όρων παίρνουμε

$$\begin{aligned}
\Sigma_j | \dots &\sim |\Sigma_j|^{-(\zeta+m+q+n_j+1)/2} \exp\left\{-\frac{1}{2}\text{tr}[\Psi + \sum_{i=1}^n I(z_i = j)(y_i - \beta'_j x_i)(y_i \right. \\
&\quad \left. - \beta'_j x_i)' + (\beta'_j - M')\Omega^{-1}(\beta_j - M)]\Sigma_j^{-1}\right\}.
\end{aligned}$$

Αυτό σημαίνει ότι

$$\Sigma_j | \dots \sim \mathcal{W}^{-1}(\Psi + T_j + (\beta'_j - M')\Omega^{-1}(\beta_j - M), \zeta + q + n_j)$$

όπου $T_j = \sum_{i=1}^n I(z_i = j)(y_i - \beta'_j x_i)(y_i - \beta'_j x_i)', j = 1, \dots, k$.

Παράρτημα Β

Παράρτημα Κεφαλαίου 3

Β.1 Απόδειξη σχέσης (3.4)

Έστω $k \in \mathbb{N}$ ο αριθμός των συνιστωσών και $\kappa_0 \in \mathbb{N}$ ο αριθμός των διαφορετικών μέσων με $\kappa_0 \leq k$. Θεωρούμε το διάνυσμα κατάστασης του μοντέλου $c = \{c_1, \dots, c_{\kappa_0}\} \in \{1, 2\}^{\kappa_0}$ με

$$\sum_{i=1}^{\kappa_0} c_i = k. \quad (\text{B.1})$$

Το πρόβλημα είναι να βρεθεί το πλήθος των διανυσμάτων c με τις παραπάνω ιδιότητες για όλα τα $k \in \mathbb{N}$. Αυτό είναι ισοδύναμο με το εξής: Έστω k σφαιρίδια τα οποία ρίπνται σε κ_0 δοχεία (με $\kappa_0 \leq k$). Θεωρώντας ότι το κ_0 μπορεί να πάρει όλες τις δυνατές του τιμές, πόσες είναι οι δυνατές κατατάξεις των σφαιριδίων έτσι ώστε κάθε ένα από τα κ_0 δοχεία να περιέχει 1 ή 2 σφαιρίδια;

Αν στα κ_0 δοχεία υπάρχουν συνολικά $x \leq \kappa_0$ δοχεία με ένα σφαιρίδιο τότε θα έχουμε $\kappa_0 - x$ δοχεία με δύο σφαιρίδια. Λόγω της (B.1) έχουμε: $\sum_{i=1}^{\kappa_0} c_i = k \Rightarrow x + 2(\kappa_0 - x) = k$ και επομένως

$$\kappa_0 = \frac{k + x}{2}. \quad (\text{B.2})$$

Από την (B.2) έχουμε ότι το x είναι άρτιος αν και μόνον αν το k είναι άρτιος. Άρα οι δυνατές τιμές του x βρίσκονται διακρίνοντας τις εξής περιπτώσεις:

$$x \in \begin{cases} \{0, 2, \dots, k-2, k\}, & \text{αν } k \text{ άρτιος} \\ \{1, 3, \dots, k-2, k\}, & \text{αν } k \text{ περιττός.} \end{cases}$$

Για δοθέντα x και κ_0 , το πλήθος των δυνατών τρόπων κατάληψης x θέσεων από κ_0 δίνεται από τους συνδυασμούς $\binom{\kappa_0}{x}$. Άρα, οι συνολικοί τρόποι θα δίνονται αθροίζοντας για τις δυνατές τιμές του x . Οπότε το πλήθος των δυνατών μοντέλων δίνεται από

την σχέση

$$N_k = \begin{cases} \sum_{l=0}^{k/2} \binom{k/2+l}{2l}, & \text{αν } k \text{ άρτιος} \\ \sum_{l=0}^{(k-1)/2} \binom{(k+2l+1)/2}{2l+1}, & \text{αν } k \text{ περιττός.} \end{cases} \quad (\text{B.3})$$

B.2 Απόδειξη Λήμματος 3.4.1

Ας συμβολίσουμε με $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_{\kappa_0}^*)$ το κ_0 -διάστατο διάνυσμα των διαφορετικών μέσων, με $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$ το k -διάστατο διάνυσμα των μέσων των συνιστωσών και έστω $\boldsymbol{c} = (c_1, \dots, c_{\kappa_0})$ το διάνυσμα κατάστασης. Έχουμε

$$\begin{aligned} \boldsymbol{\mu}^* | \boldsymbol{z}, \boldsymbol{\sigma}^2, \boldsymbol{x} &\propto \prod_{t=1}^n \exp\{-(x_t - \mu_{z_t})^2 / (2\sigma_{z_t}^2)\} \prod_{j=1}^{\kappa_0} \exp\{-\kappa(\mu_j^* - \xi)^2 / 2\} \\ &\propto \prod_{j=1}^{\kappa_0} \exp\left\{-\frac{1}{2} \left[\kappa(\mu_j^* - \xi)^2 + \sum_{i=l_j}^{u_j} \left(\sigma_i^{-2} \sum_{t=1}^n I(z_t = i)(x_t - \mu_j^*)^2 \right) \right]\right\}, \end{aligned}$$

όπου $l_j = \sum_{t=1}^{j-1} c_t + 1$ και $u_j = \sum_{t=1}^j c_t$, $j = 1, \dots, \kappa_0$. Από την τελευταία έκφραση προκύπτει ότι οι κ_0 διαφορετικοί μέσοι είναι εκ των υστέρων ανεξάρτητοι. Συνεπώς, για $j = 1, \dots, \kappa_0$,

$$\begin{aligned} \mu_j^* | \boldsymbol{z}, \boldsymbol{\sigma}^2, \boldsymbol{x} &\propto \exp\left\{-\frac{1}{2} \left[\kappa(\mu_j^* - \xi)^2 + \sum_{i=l_j}^{u_j} \left(\sigma_i^{-2} \sum_{t=1}^n I(z_t = i)(x_t - \mu_j^*)^2 \right) \right]\right\} \\ &\propto \exp\left\{-\frac{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa}{2} \left(\mu_j^* - \frac{\sum_{i=l_j}^{u_j} s_i / \sigma_i^2 + \kappa \xi}{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa} \right)^2\right\} \\ &\sim \mathcal{N}\left(\frac{\sum_{i=l_j}^{u_j} s_i / \sigma_i^2 + \kappa \xi}{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa}, \frac{1}{\sum_{i=l_j}^{u_j} n_i / \sigma_i^2 + \kappa}\right), \end{aligned}$$

όπου $s_i = \sum_{t=1}^n I(z_t = i)x_t$ και $n_i = \sum_{t=1}^n I(z_t = i)$.

B.3 Απόδειξη Πρότασης 3.4.1

Ας υποθέσουμε ότι η αλυσίδα βρίσκεται στην κατάσταση $y = (\boldsymbol{c}, k, k_0, \boldsymbol{q}, \boldsymbol{\nu}^*, \boldsymbol{\tau}^2, \boldsymbol{z})$ και ότι προτείνουμε μία μετάβαση στην κατάσταση $y' = (\boldsymbol{c}', k', k'_0, \boldsymbol{p}, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2, \boldsymbol{z}')$. Τα στοιχεία των u και u' είναι οι επί πλέον τυχαίες μεταβλητές που πρέπει να παραχθούν ώστε να ικανοποιούν την υπόθεση συμφωνίας των διαστάσεων για τα διανύσματα (y, u) και (y', u') . Σύμφωνα με τον Αλγόριθμο Α.1.1 η γενική μορφή της πιθανότητας αποδοχής είναι

$$A = \frac{f(\boldsymbol{x} | \boldsymbol{c}', \boldsymbol{z}', \boldsymbol{\mu}^*, \boldsymbol{\sigma}^2)}{f(\boldsymbol{x} | \boldsymbol{c}, \boldsymbol{z}, \boldsymbol{\nu}^*, \boldsymbol{\tau}^2)} \times \frac{f(k') f(k'_0 | k') f(\boldsymbol{c}' | k'_0, k') f(\boldsymbol{z}' | k', \boldsymbol{p})}{f(k) f(k_0 | k) f(\boldsymbol{c} | k_0, k) f(\boldsymbol{z} | k, \boldsymbol{q})}$$

$$\times \frac{f(\mathbf{p}|\delta, k')}{f(\mathbf{q}|\delta, k)} \frac{f(\boldsymbol{\mu}^*|k'_0)}{f(\boldsymbol{\nu}^*|k_0)} \frac{f(\boldsymbol{\sigma}^2|k')}{f(\boldsymbol{\tau}^2|k)} \frac{p[(y', u') \mapsto (y, u)]f(u')}{p[(y, u) \mapsto (y', u')]f(u)} \times \left| \frac{\partial y'}{\partial y} \right|, \quad (\text{B.4})$$

όπου $p[a \mapsto b]$ συμβολίζει την πιθανότητα μετάβασης στο b ενώ βρισκόμαστε στην κατάσταση a .

Στην συνέχεια θα υπολογίσουμε τις αναλυτικές εκφράσεις των όρων που εμπλέκονται στην (B.4). Από εδώ και στο εξής συμβολίζουμε με P_{realloc} την πιθανότητα μιας συγκεκριμένης ανακατάταξης μεταξύ των συνιστώσων, ενώ τα m_1 και m_2 είναι όπως ορίστηκαν στην Ενότητα 3.2.

Έστω ότι προτείνουμε τον διαχωρισμό μιας σύνθετης συνιστώσας σε δύο απλές. Χωρίς βλάβη της γενικότητας υποθέτουμε ότι η σύνθετη συνιστώσα αποτελείται από τις δύο πρώτες συνιστώσες της μείξης, δηλαδή τις συνιστώσες 1 και 2. Προφανώς, $k'_0 = k_0 + 1$, $k' = k$, οπότε $f(k')/f(k) = 1$.

Λόγος εκ των προτέρων πιθανοτήτων των κατατάξεων: Έστω n_1, n_2 οι αριθμοί των παρατηρήσεων που είναι ήδη κατεταγμένες στις συνιστώσες 1,2, αντίστοιχα, και q_1, q_2 τα αντίστοιχα βάρη. Αν n'_1, n'_2 και p_1, p_2 είναι οι αντίστοιχες ποσότητες μετά την κίνηση διαχωρισμού μέσω (με $n_1 + n_2 = n'_1 + n'_2$), έχουμε

$$\frac{f(\mathbf{z}'|k, \mathbf{p})}{f(\mathbf{z}|k, \mathbf{q})} = \frac{p_1^{n'_1} p_2^{n'_2}}{q_1^{n_1} q_2^{n_2}}. \quad (\text{B.5})$$

Λόγος των εκ των προτέρων κατανομών των βαρών: Τα μοναδικά βάρη που αλλάζουν κατά την κίνηση αυτή είναι εκείνα που αντιστοιχούν στις δύο συνιστώσες που αποτελούν την σύνθετη, άρα

$$\frac{f(\mathbf{p}|\delta, k)}{f(\mathbf{q}|\delta, k)} = \frac{p_1^{\delta-1} p_2^{\delta-1}}{q_1^{\delta-1} q_2^{\delta-1}}. \quad (\text{B.6})$$

Λόγος των εκ των προτέρων κατανομών των μέσων: Πριν την κίνηση διαχωρισμού, οι συνιστώσες 1 και 2 έχουν έναν κοινό μέσο (τον ν_1) και προτείνουμε την μετάβαση σε μία κατάσταση με δύο διαφορετικούς μέσους (μ_1^* και μ_2^*). Οι υπόλοιποι μέσοι παραμένουν ίδιοι, άρα

$$\begin{aligned} \frac{f(\boldsymbol{\mu}^*|k_0 + 1)}{f(\boldsymbol{\nu}^*|k_0)} &= \frac{(k_0 + 1)! \prod_{i=1}^{k_0+1} \varphi(\mu_i^*|\xi, \kappa^{-1})}{k_0! \prod_{i=1}^{k_0} \varphi(\nu_i^*|\xi, \kappa^{-1})} \\ &= (k_0 + 1) \sqrt{\frac{\kappa}{2\pi}} \exp \left\{ -\frac{\kappa}{2} \left[(\mu_1^* - \xi)^2 + (\mu_2^* - \xi)^2 - (\nu_1^* - \xi)^2 \right] \right\}. \end{aligned} \quad (\text{B.7})$$

Εδώ, με $\varphi(\cdot|\xi, \kappa^{-1})$ συμβολίζουμε την πυκνότητα πιθανότητας της $\mathcal{N}(\xi, \kappa^{-1})$.

Λόγος των εκ των προτέρων κατανομών των διασπορών: Έστω τ_1^2, τ_2^2 οι διασπορες των δύο πρώτων συνιστωσών πριν την κίνηση διαχωρισμού, και σ_1^2, σ_2^2 οι αντίστοιχες διασπορές μετά την προτεινόμενη μετάβαση. Τότε,

$$\frac{f(\sigma^2|k)}{f(\tau^2|k)} = \left(\frac{\tau_1^2 \tau_2^2}{\sigma_1^2 \sigma_2^2} \right)^{\alpha+1} \exp\{-\beta(\sigma_1^{-2} + \sigma_2^{-2} - \tau_1^{-2} - \tau_2^{-2})\}. \quad (\text{B.8})$$

Τα παραγόμενα τυχαία διανύσματα για την υπόθεση συμφωνίας των διαστάσεων είναι $u = (u_1, u_2)$ για την κίνηση διαχωρισμού μέσων, και $u' = \tilde{u}_1$ για την κίνηση συνδυασμού μέσων. Έστω επίσης m_2 το πλήθος των 2 που υπάρχουν στο διάνυσμα κατάστασης c . Έτσι, αν προτείνουμε διαχωρισμό μέσων με πιθανότητα $m_2/(m_1 + m_2)$, έχουμε m_2 δυνατές επιλογές για την επιλογή μιας σύνθετης συνιστώσας, οπότε το μοντέλο c' προτείνεται με πιθανότητα $1/(m_1 + m_2)$. Μόλις παραχθεί το u , έχουμε να ανακατατάξουμε τις παρατηρήσεις των συνιστωσών 1 και 2 (μετά την διάσπαση των μέσων) στις δύο συνιστώσες.

Κατά την αντίστροφη κίνηση, το μοντέλο c' έχει $k_0 + 1$ διακεκριμένους μέσους και m'_1 ζευγάρια απλών και γειτονικών συνιστωσών. Μετά την επιλογή συνδυασμού μέσων με πιθανότητα $m'_1/(m'_1 + m'_2)$, έχουμε να επιλέξουμε μεταξύ των m'_1 ζευγαριών. Έτσι, το μοντέλο c προτείνεται με πιθανότητα $1/(m'_1 + m'_2)$. Τέλος, μένει να προτείνουμε την ανακατάταξη των παρατηρήσεων που ανήκουν στις συνιστώσες 1 και 2 (μετά την συγχώνευση μέσων) μεταξύ τους. Συμβολίζουμε με P'_{realloc} την πιθανότητα της συγκεκριμένης ανακατάταξης. Έτσι:

$$\frac{p[(y', u') \mapsto (y, u)]f(u')}{p[(y, u) \mapsto (y', u')]f(u)} = \frac{m_1 + m_2}{m'_1 + m'_2} \frac{P'_{\text{realloc}}}{P_{\text{realloc}}} \frac{g_{2,2}(\tilde{u}_1)}{g_{2,2}(u_1)g_{2,2}(u_2)}. \quad (\text{B.9})$$

Λόγος πιθανοτήτων ανακατάταξης: Έστω z_i η τρέχουσα κατάταξη της παρατήρησης x_i , $i = 1, \dots, n$ και z'_i η προτεινόμενη ανακατάταξή της. Έστω επίσης $\mathcal{Z}_{12} = \{i : z_i = 1 \text{ ή } 2\}$ το σύνολο εκείνων των δεικτών των παρατηρήσεων που κατατάσσονται στις δύο πρώτες συνιστώσες. Έτσι:

$$P_{\text{realloc}} = \prod_{i \in \mathcal{Z}_{12}} \frac{(p_{z'_i}/\sigma_{z'_i}) \exp\{- (x_i - \mu_{z'_i})^2 / (2\sigma_{z'_i}^2)\}}{(p_1/\sigma_1) \exp\{- (x_i - \mu_1)^2 / (2\sigma_1^2)\} + (p_2/\sigma_2) \exp\{- (x_i - \mu_2)^2 / (2\sigma_2^2)\}} \quad (\text{B.10})$$

και

$$P'_{\text{realloc}} = \prod_{i \in \mathcal{Z}_{12}} \frac{(q_{z_i}/\tau_{z_i}) \exp\{- (x_i - \mu)^2 / (2\tau_{z_i}^2)\}}{(q_1/\tau_1) \exp\{- (x_i - \mu)^2 / (2\tau_1^2)\} + (q_2/\tau_2) \exp\{- (x_i - \mu)^2 / (2\tau_2^2)\}} \quad (\text{B.11})$$

Αντικαθιστώντας τις (B.5)–(B.11) και την Ιακωβιανή ορίζουσα στην (B.4) προκύπτει η (3.6) και ολοκληρώνεται η απόδειξη.

B.4 Κανονικές μείξεις με κοινές μέσες τιμές έναντι μείξεων κατανομών t_4 στο γαλαξιακό σύνολο δεδομένων

Στην Ενότητα 3.5.3 συγκρίναμε την προσαρμογή του μοντέλου $c = (1, 2, 2, 1)$ έναντι ενός μοντέλου μείξης τεσσάρων κατανομών t_ν (με $\nu = 4$), που αποτελεί την λύση που προτάθηκε από τον Stephens (1997). Αυτό έγινε κατασκευάζοντας έναν αλγόριθμο RJMCMC που δρα στον χώρο αυτών των δύο μοντέλων και θα περιγραφεί ακολούθως. Ωστόσο, η μεθοδολογία μπορεί να γενικευθεί άμεσα για τη σύγκριση οποιασδήποτε κανονικής μείξης με κοινές μέσες τιμές με το αντίστοιχο¹ μοντέλο μείξης κατανομών t_ν , αρκεί $\nu > 2$ ώστε να υπάρχουν οι δύο πρώτες ροπές αυτού.

Συμβολίζουμε με $p_j, \sigma_j^2, j = 1, \dots, 6$, και με $\mu_j^*, j = 1, \dots, 4$, τις παραμέτρους του μοντέλου $c = (1, 2, 2, 1)$. Έστω επίσης q_j, s_j^2 και $\gamma_j, j = 1, \dots, 4$, οι αντίστοιχες παράμετροι για το μοντέλο του Stephens, όπου και για τις δύο περιπτώσεις υποθέτουμε εκ των προτέρων διάταξη στους μέσους των μοντέλων αυτών, δηλαδή $\mu_1^* < \mu_2^* < \mu_3^* < \mu_4^*$ και $\gamma_1 < \gamma_2 < \gamma_3 < \gamma_4$.

Όταν η αλυσίδα βρίσκεται σε ένα από τα δύο ανταγωνιστικά μοντέλα, οι ανανεώσεις των παραμέτρων γίνονται χρησιμοποιώντας τις πλήρεις δεσμευμένες κατανομές αυτών μέσω ενός δειγματολήπτη Gibbs. Για το μοντέλο $c = (1, 2, 2, 1)$ οι πλήρεις δεσμευμένες κατανομές περιγράφονται στην Ενότητα 3.3, ενώ για το μοντέλο των κατανομών t_ν παραπέμπουμε στον Stephens (1997a), σελ. 33. Στην συνέχεια ορίζεται μία κατάλληλη κίνηση (μαζί με την αντίστροφή της) ώστε η αλυσίδα να εναλλάσσεται μεταξύ αυτών των δύο μοντέλων, με στόχο την από κοινού εκ των υστέρων κατανομή τους. Σύμφωνα με την υπόθεση συμφωνίας των διαστάσεων (A.1), η μετάβαση από το μοντέλο τεσσάρων συνιστώσων t_ν στο μοντέλο $c = (1, 2, 2, 1)$ απαιτεί την παραγωγή τεσσάρων τυχαίων μεταβλητών.

Η κίνηση από το μοντέλο $c = (1, 2, 2, 1)$ στο μοντέλο τεσσάρων κατανομών t_ν προτείνει την αντικατάσταση των δύο σύνθετων συνιστώσων του πρώτου μοντέλου από δύο συνιστώσες t_ν , ενώ οι παράμετροι των υπόλοιπων συνιστώσων παραμένουν ως έχουν. Επί πλέον, αλλάζει η παραμετρική οικογένεια από κανονικές κατανομές σε t_ν . Αντίστοιχα, κατά την αντίστροφη κίνηση προτείνεται η αντίστροφη πορεία, δηλαδή η αντικατάσταση των δύο μεσαίων συνιστώσων t_ν με δύο σύνθετες συνιστώσες. Οδηγός για τους παρακάτω μετασχηματισμούς, είναι η διατήρηση της μηδενικής, πρώτης

¹Δηλαδή η μείξη των κατανομών t_ν να έχει τόσες συνιστώσες όσοι και οι διακριτοί μέσοι της κανονικής μείξης.

και δεύτερης ροπής του μοντέλου μείζης. Θεωρούμε δηλαδή ότι ισχύει

$$\begin{aligned} p_2 + p_3 &= q_2 \\ p_2\mu_2^* + p_3\mu_3^* &= q_2\gamma_2 \\ p_2(\mu_2^{*2} + \sigma_2^2) + p_3(\mu_3^{*2} + \sigma_3^2) &= q_2 \left(\gamma_2^2 + \frac{\nu}{\nu-2} s_2^2 \right) \end{aligned}$$

και

$$\begin{aligned} p_4 + p_5 &= q_3 \\ p_4\mu_4^* + p_5\mu_5^* &= q_3\gamma_3 \\ p_4(\mu_4^{*2} + \sigma_4^2) + p_5(\mu_5^{*2} + \sigma_5^2) &= q_3 \left(\gamma_3^2 + \frac{\nu}{\nu-2} s_3^2 \right). \end{aligned}$$

Η μετάβαση από το μοντέλο τεσσάρων κατανομών t_ν στο $\mathbf{c} = (1, 2, 2, 1)$ γίνεται παράγοντας αρχικά τις ανεξάρτητες τυχαίες μεταβλητές $u_j \sim \mathcal{U}(0, 1)$, $j = 1, \dots, 4$, και στην συνέχεια προτείνοντας τους μετασχηματισμούς των παραμέτρων. Για τα βάρη έχουμε

$$\begin{aligned} p_1 &= q_1 & p_4 &= q_3 u_2 \\ p_2 &= q_2 u_1 & p_5 &= q_3 (1 - u_2) \\ p_3 &= q_2 (1 - u_1) & p_6 &= q_6. \end{aligned}$$

Οι μέσοι παραμένουν ίδιοι:

$$\mu_j^* = \gamma_j, \quad \forall j = 1, \dots, 4.$$

Για τις διασπορές ισχύει:

$$\begin{aligned} \sigma_1^2 &= s_1^2 & \sigma_4^2 &= \frac{\nu u_4 q_3 s_3^2}{(\nu-2)p_4} \\ \sigma_2^2 &= \frac{\nu u_3 q_2 s_2^2}{(\nu-2)p_2} & \sigma_5^2 &= \frac{\nu(1-u_4)q_3 s_3^2}{(\nu-2)p_5} \\ \sigma_3^2 &= \frac{\nu(1-u_3)q_2 s_2^2}{(\nu-2)p_3} & \sigma_6^2 &= s_4^2. \end{aligned}$$

Στην συνέχεια προτείνεται η ανακατάταξη των παρατηρήσεων που ανήκουν στην δεύτερη και στην τρίτη συνιστώσα t_ν στα ζευγάρια σύνθετων κανονικών συνιστωσών που αντικαθιστούν κάθε μια από αυτές, ενώ οι υπόλοιπες παρατηρήσεις παραμένουν καταταγμένες στις δύο ακραίες συνιστώσες. Τέλος, η Ιακωβιανή του μετασχηματισμού

$$(q_2, q_3, \gamma_2, \gamma_3, s_2^2, s_3^2, u_1, u_2, u_3, u_4) \mapsto (p_2, p_3, p_4, p_5, \mu_2^*, \mu_3^*, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2)$$

είναι

$$|J| = \frac{q_2 q_3 s_2^2 s_3^2 \nu^4}{(\nu-2)^4 (1-u_1) u_1 (1-u_2) u_2}.$$

Κατά την αντίστροφη κίνηση, θέτουμε

$$q_1 = p_1, q_2 = p_2 + p_3, q_3 = p_4 + p_5, q_6 = p_6$$

$$\gamma_j = \mu_j^*, \quad \forall j = 1, \dots, 4$$

$$s_1^2 = \sigma_1^2, s_2^2 = \frac{\nu - 2}{\nu}(p_2\sigma_2^2 + p_3\sigma_3^2), s_3^2 = \frac{\nu - 2}{\nu}(p_4\sigma_4^2 + p_5\sigma_5^2), s_4^2 = \sigma_6^2,$$

ενώ τα $u_j, j = 1, \dots, 4$, τίθενται ίσα με

$$u_1 = \frac{p_2}{q_2}, u_2 = \frac{p_4}{q_3}, u_3 = \frac{(\nu - 2)p_2\sigma_2^2}{\nu q_2 s_2^2}, u_4 = \frac{(\nu - 2)p_4\sigma_4^2}{\nu q_3 s_3^2}.$$

Αν κάποιο $u_j, j = 1, \dots, 4$, δεν ανήκει στο διάστημα $(0, 1)$, τότε η προτεινόμενη κίνηση απορρίπτεται αμέσως. Σε διαφορετική περίπτωση, οι παρατηρήσεις όπου αρχικά ανήκαν στην δεύτερη και τρίτη κανονική συνιστώσα ανακατατάσσονται στην δεύτερη συνιστώσα t_ν . Αντίστοιχα, οι παρατηρήσεις που αρχικά ανήκαν στην τέταρτη και πέμπτη κανονική συνιστώσα ανακατατάσσονται στην τρίτη συνιστώσα t_ν . Σε αυτήν την περίπτωση, η ιακωβιανή του μετασχηματισμού είναι απλώς αντίστροφη εκείνης της προηγούμενης μετάβασης.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΙΑ

Παράρτημα Γ

Παράρτημα Κεφαλαίου 4

Γ.1 Απόδειξη Πρότασης 4.2.2

Για να αποδείξουμε την Πρόταση 4.2.2 θα χρειαστούμε το επόμενο Λήμμα.

Λήμμα Γ.1.1. Έστω $m_{j,k}$, $1 \leq j \leq k$, το πλήθος των n -άδων που περιέχουν ακριβώς j από k δείκτες. Τότε,

$$m_{j,k} = \binom{k}{j} \sum_{i=1}^j (-1)^{j-i} \binom{j}{i} i^n. \quad (\Gamma.1)$$

Απόδειξη. Είναι προφανές ότι η (Γ.1) ισχύει για $j = 1$. Θα υποθέσουμε ότι ισχύει για κάποιο $j > 1$ και θα δείξουμε την ισχύ της για $j + 1$. Ας παρατηρήσουμε ότι η υπόθεση της επαγωγής ισοδυναμεί με το ότι η ποσότητα $\sum_{\lambda=1}^{j-1} m_{\lambda,j}$ εκφράζει το πλήθος των n -άδων που περιέχουν ακριβώς λ από j συγκεκριμένους δείκτες, για $\lambda = 1, \dots, j - 1$. Αν ισχύει λοιπόν η υπόθεσή μας, αφαιρώντας αυτήν την ποσότητα από το j^n θα πάρουμε το πλήθος των n -άδων που περιέχουν ακριβώς j συγκεκριμένους δείκτες. Με άλλα λόγια, η υπόθεση της επαγωγής είναι ισοδύναμη με την

$$j^n - \sum_{\lambda=1}^{j-1} m_{\lambda,j} = \sum_{i=1}^j (-1)^{j-i} \binom{j}{i} i^n.$$

Επομένως αν δείξουμε ότι η τελευταία σχέση ισχύει και για $j + 1$, δηλαδή αν

$$(j+1)^n - \sum_{\lambda=1}^j m_{\lambda,j+1} = \sum_{i=1}^{j+1} (-1)^{j+1-i} \binom{j+1}{i} i^n \quad (\Gamma.2)$$

τότε το αποτέλεσμα προκύπτει επαγωγικά θεωρώντας όλους τους πιθανούς συνδυασμούς των $j + 1$ από τους k δείκτες. Έχουμε λοιπόν ότι

$$\sum_{\lambda=1}^j m_{\lambda,j+1} = \sum_{\lambda=1}^j \binom{j+1}{\lambda} \sum_{i=1}^{\lambda} (-1)^{\lambda-i} \binom{\lambda}{i} i^n$$

$$\begin{aligned}
&= \sum_{i=1}^j \sum_{\lambda=i}^j \frac{(j+1)!}{\lambda!(j+1-\lambda)!} \frac{\lambda!}{i!(\lambda-i)!} (-1)^{\lambda-i} i^n \\
&= \sum_{i=1}^j \frac{(j+1)!(-1)^i i^n}{i!} \sum_{\lambda=i}^j \frac{(-1)^\lambda}{(\lambda-i)!(j+1-\lambda)!} \\
&= \sum_{i=1}^j \frac{(j+1)!(-1)^i}{i!} i^n \sum_{\lambda=0}^{j-i} \frac{(-1)^{\lambda+i}(j+1-i)!}{\lambda!(j+1-i-\lambda)!(j+1-i)!} \\
&= \sum_{i=1}^j \frac{(j+1)!}{i!(j+1-i)!} i^n \sum_{\lambda=0}^{j-i} (-1)^\lambda \binom{j+1-i}{\lambda} \\
&= \sum_{i=1}^j \binom{j+1}{i} i^n [0 - (-1)^{j-i+1}] \\
&= \sum_{i=1}^j (-1)^{j-i} \binom{j+1}{i} i^n
\end{aligned}$$

Αφαιρώντας την τελευταία ποσότητα από την $(j+1)^n$ προκύπτει η (Γ.2) και τούτο ολοκληρώνει την απόδειξη. \square

Απόδειξη Πρότασης 4.2.2. Ας συμβολίσουμε με c_{jk} τον αριθμό των κλάσεων που περιέχουν ακριβώς j από k στοιχεία. Επειδή κάθε τέτοια κλάση περιέχει ακριβώς $k!/(k-j)!$ στοιχεία, έχουμε ότι

$$c_{jk} = \frac{m_{jk}}{k!/(k-j)!},$$

όπου το m_{jk} ορίστηκε στην (Γ.1). Συνεπώς, ο συνολικός αριθμός κλάσεων θα προκύπτει αθροίζοντας τις κλάσεις για όλες τις δυνατές τιμές του $j = 1, \dots, k$, οπότε

$$\begin{aligned}
\sum_{j=1}^k c_{jk} &= \sum_{j=1}^k \frac{m_{jk}}{k!/(k-j)!} \\
&= \sum_{j=1}^k \frac{1}{j!} \sum_{i=1}^j (-1)^{j-i} \binom{j}{i} i^n \\
&= \sum_{j=1}^k \sum_{i=1}^j (-1)^{j-i} \frac{i^n}{i!(j-i)!} \\
&= \sum_{i=1}^k \frac{i^n}{i!} \sum_{j=0}^{k-i} \frac{(-1)^j}{j!}
\end{aligned}$$

και τούτο ολοκληρώνει την απόδειξη. \square

Γ.2 Εξασφάλιση της (4.21)

Στην συνέχεια περιγράφονται κάποιες αρκετά τυπικές περιπτώσεις όπου εξασφαλίζεται η ισχύς της σχέσης (4.21). Ας υποθέσουμε ότι οι παράμετροι των συνιστωσών/καταστάσεων (σε γενικότερες περιπτώσεις όπως στα HMM) κατηγοριοποιούνται στις ομάδες (blocks) $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m)$ με $\boldsymbol{\eta}_l = (\eta_{l1}, \dots, \eta_{lk})$ για κάθε $l = 1, \dots, m$. Για παράδειγμα, στις μείξεις κανονικών κατανομών τα blocks αυτά αποτελούνται από τα βάρη, μέσους και διασπορές, δηλαδή $m = 3$ και $\boldsymbol{\eta}_1 = \boldsymbol{\mu}, \boldsymbol{\eta}_2 = \boldsymbol{\sigma}^2$. Συνεπώς, το παραμετρικό διάνυσμα μπορεί να εκφραστεί ως $(\boldsymbol{z}, \boldsymbol{\eta})$. Στην συνέχεια θα φανεί χρήσιμη η αναλλοίωτη ιδιότητα (4.15) της εκ των υστέρων κατανομής ενός μοντέλου μείξης ως προς τις μεταθέσεις.

Λήμμα Γ.2.1. Έστω ένας δειγματολήπτης Gibbs κατά τον οποίο το παραμετρικό διάνυσμα προσομοιώνεται μέσω των πλήρων δεσμευμένων εκ των υστέρων κατανομών των blocks $\boldsymbol{\eta}_l, l = 1, \dots, m$ και $z_i, i = 1, \dots, n$ κατά τον ακόλουθο τρόπο. Αν $\boldsymbol{\eta}^{(m)} = \boldsymbol{\eta}'$ και $\boldsymbol{z}^{(m)} = \boldsymbol{z}'$, τότε κατά το βήμα $m + 1$, προσομοιώνεται

$$\begin{aligned} z_i &\sim f(z_i | \boldsymbol{z}_{1:(i-1)}, \boldsymbol{z}'_{(i+1):n}, \boldsymbol{\eta}', \boldsymbol{x}), \quad i = 1, \dots, n \\ \boldsymbol{\eta}_l &\sim f(\boldsymbol{\eta}_l | \boldsymbol{\eta}_{1:(l-1)}, \boldsymbol{\eta}'_{(l+1):m}, \boldsymbol{z}, \boldsymbol{x}), \quad l = 1, \dots, m. \end{aligned}$$

Τότε η (4.21) ισχύει.

Απόδειξη. Η πυκνότητα μετάβασης της μαρκοβιανής αλυσίδας μπορεί να γραφτεί ως εξής:

$$f(\boldsymbol{z}, \boldsymbol{\eta} | \boldsymbol{z}', \boldsymbol{\eta}') = \prod_{i=1}^n f(z_i | \boldsymbol{z}_{1:(i-1)}, \boldsymbol{z}'_{(i+1):n}, \boldsymbol{\eta}', \boldsymbol{x}) \times \prod_{l=1}^m f(\boldsymbol{\eta}_l | \boldsymbol{\eta}_{1:(l-1)}, \boldsymbol{\eta}'_{(l+1):m}, \boldsymbol{z}, \boldsymbol{x}).$$

Σύμφωνα με τον παραπάνω ορισμό της μετάβασης, έχουμε

$$\begin{aligned} f(\boldsymbol{z}, \boldsymbol{\eta} | \boldsymbol{z}', \boldsymbol{\eta}') &= \prod_{i=1}^n f(z_i | \boldsymbol{z}_{1:(i-1)}, \boldsymbol{z}'_{(i+1):n}, \boldsymbol{\eta}', \boldsymbol{x}) \prod_{l=1}^m f(\boldsymbol{\eta}_l | \boldsymbol{\eta}_{1:(l-1)}, \boldsymbol{\eta}'_{(l+1):m}, \boldsymbol{z}, \boldsymbol{x}) \\ &\propto \prod_{i=1}^n f(\boldsymbol{z}_{1:(i-1)}, z_i, \boldsymbol{z}'_{(i+1):n}, \boldsymbol{\eta}' | \boldsymbol{x}) \prod_{l=1}^m f(\boldsymbol{\eta}_{1:(l-1)}, \boldsymbol{\eta}_l, \boldsymbol{\eta}'_{(l+1):m}, \boldsymbol{z} | \boldsymbol{x}) \\ &\propto \prod_{i=1}^n f(\boldsymbol{z}_{1:i}, \boldsymbol{z}'_{(i+1):n}, \boldsymbol{\eta}' | \boldsymbol{x}) \prod_{l=1}^m f(\boldsymbol{\eta}_{1:l}, \boldsymbol{\eta}'_{(l+1):m}, \boldsymbol{z} | \boldsymbol{x}) \\ &\propto \prod_{i=1}^n f(\tau^{-1} \boldsymbol{z}_{1:i}, \tau^{-1} \boldsymbol{z}'_{(i+1):n}, \tau \boldsymbol{\eta}' | \boldsymbol{x}) \prod_{l=1}^m f(\tau \boldsymbol{\eta}_{1:l}, \tau \boldsymbol{\eta}'_{(l+1):m}, \tau^{-1} \boldsymbol{z} | \boldsymbol{x}) \end{aligned} \quad (\Gamma.3)$$

(λόγω της (4.15))

$$\propto f(\tau^{-1}\mathbf{z}, \tau\boldsymbol{\eta} | \tau^{-1}\mathbf{z}', \tau\boldsymbol{\eta}'), \quad \forall \tau \in \mathcal{I}_k$$

και τούτο ολοκληρώνει την απόδειξη, καθώς η σταθερά κανονικοποίησης είναι ίδια σε όλες τις περιπτώσεις. \square

Στην συνέχεια παρατίθεται η απόδειξη της ισχύος της (4.21) για δειγματολήπτες όπου η προσομοίωση των παραμέτρων μέσω των πλήρων δεσμευμένων εκ των υστέρων κατανομών δεν είναι άμεση και χρησιμοποιείται για τον σκοπό αυτό ο slice sampler των Damien et al. (1999). Ένα παράδειγμα τέτοιου δειγματολήπτη είναι ο αλγόριθμος των Robert et al. (2001) που εξετάσαμε στην Ενότητα 5.3.6. Η γενική ιδέα πίσω από την τεχνική αυτή είναι η εξής: Έστω ότι θέλουμε να προσομοιώσουμε ένα δείγμα από την κατανομή $f(x)$, η οποία μπορεί να είναι μη κανονικοποιημένη. Τότε αυτό μπορεί να γίνει παράγοντας τυχαίους αριθμούς που ακολουθούν την ομοιόμορφη κατανομή στο χωρίο κάτω από το γράφημα της $f(x)$. Η ιδέα αυτή υλοποιείται με την εισαγωγή μιας εικονικής μεταβλητής u και την παραγωγή τυχαίων αριθμών (x, u) ομοιόμορφα στο χωρίο $S = \{(x, u) : 0 < u < f(x)\}$.

Πιο συγκεκριμένα, ο slice sampler για την ανανέωση των παραμέτρων (τυπικών αποκλίσεων) των συνιστωσών για το μοντέλο των Robert et al. (2001) χρησιμοποιεί κάποιες επί πλέον τυχαίες μεταβλητές $\mathbf{u} = (u_1, \dots, u_k)$ και έχει ως εξής¹: Έστω $(\boldsymbol{\sigma}', \mathbf{u}') = ((\sigma'_1, \dots, \sigma'_k), (u'_1, \dots, u'_k))$ η παρούσα κατάσταση της μαρκοβιανής αλυσίδας. Τότε η επόμενη κατάσταση, $(\boldsymbol{\sigma}, \mathbf{u}) = ((\sigma_1, \dots, \sigma_k), (u_1, \dots, u_k))$, προσομοιώνεται σύμφωνα με το εξής σχήμα για $j = 1, \dots, k$:

$$(\sigma'_j, u'_j) \mapsto (\sigma'_j, u_j) \mapsto (\sigma_j, u_j),$$

όπου

$$u_j | \sigma'_j, \mathbf{z} \sim \mathcal{U}(0, f(\sigma'_j | \mathbf{z}))$$

και

$$\sigma_j | u_j, \mathbf{z} \sim \mathcal{U}(\mathcal{A}_j(\mathbf{z})),$$

με $\mathcal{A}_j(\mathbf{z}) := \{s : u_j \leq f(s | \mathbf{z}, a)\}$. Εδώ, το $f(s | \cdot)$ δηλώνει την πλήρη δεσμευμένη εκ των υστέρων κατανομή του σ_j , για την οποία είναι διαθέσιμη η αναλυτική μορφή της (εκτός από μία πολλαπλασιαστική σταθερά, η οποία όμως δεν επηρεάζει τον αλγόριθμο) και το a είναι μία υπερπαραμέτρος που όπως είδαμε στην Ενότητα 5.3.6 είναι κοινή για όλες τις καταστάσεις της μη παρατηρήσιμης μαρκοβιανής διαδικασίας. Είναι προφανές ότι

$$\mathcal{A}_j(\mathbf{z}) = \mathcal{A}_{t_j}(\tau^{-1}\mathbf{z}), \quad (\Gamma.4)$$

¹Σε ό,τι ακολουθεί δεν σημειώνουμε την εξάρτηση των κατανομών από το διάνυσμα κατάταξης \mathbf{z} για λόγους ευκολίας στον συμβολισμό.

για κάθε $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$. Έστω ότι $\text{Leb}(\mathcal{A})$ είναι το μέτρο Lebesgue του συνόλου \mathcal{A} . Το επόμενο Λήμμα δείχνει ότι, υπό το παραπάνω πλαίσιο, ο πυρήνας μετάβασης της αλυσίδας ικανοποιεί την συνθήκη (4.21).

Λήμμα Γ.2.2. Έστω ότι στον δειγματολήπτη Gibbs που περιγράφεται στο Λήμμα Γ.2.1 η ανανέωση του η_λ για κάποιον δείκτη $\lambda \in \{1, \dots, m\}$ γίνεται μέσω ενός βήματος slice sampling. Τότε η (4.21) εξακολουθεί να ισχύει.

Απόδειξη. Ο πυρήνας μετάβασης σε αυτήν την περίπτωση είναι

$$f(\mathbf{z}, \boldsymbol{\eta}, \mathbf{u} | \mathbf{z}', \boldsymbol{\eta}', \mathbf{u}') = \prod_{i=1}^n f(z_i | \mathbf{z}_{1:(i-1)}, \mathbf{z}'_{(i+1):n}, \boldsymbol{\eta}', \mathbf{x}) \times \prod_{l \neq \lambda} f(\eta_l | \boldsymbol{\eta}_{1:(l-1)}, \boldsymbol{\eta}'_{(l+1):m}, \mathbf{z}, \mathbf{x}) \times f(\mathbf{u}, \eta_\lambda | \mathbf{u}', \eta'_\lambda, \mathbf{z}).$$

Λόγω της (Γ.3), αρκεί να δείξουμε ότι για κάθε $\tau \in \mathcal{T}_k$ ισχύει

$$f(\tau \mathbf{u}, \tau \boldsymbol{\eta}_\lambda | \tau \mathbf{u}', \tau \boldsymbol{\eta}'_\lambda, \tau^{-1} \mathbf{z}) = f(\mathbf{u}, \eta_\lambda | \mathbf{u}', \eta'_\lambda, \mathbf{z})$$

Χωρίς βλάβη της γενικότητας ας υποθέσουμε ότι $\eta_{\lambda_j} > 0$ για κάθε $j = 1, \dots, k$. Εξ ορισμού έχουμε

$$\begin{aligned} f(\mathbf{u}, \eta_\lambda | \mathbf{u}', \eta'_\lambda, \mathbf{z}) &= f(\mathbf{u} | \boldsymbol{\eta}'_\lambda, \mathbf{z}) f(\eta_\lambda | \mathbf{u}, \mathbf{z}) = \prod_{j=1}^k f(u_j | \eta'_{\lambda_j}, \mathbf{z}) f(\eta_{\lambda_j} | u_j, \mathbf{z}) \\ &= \prod_{j=1}^k \frac{I(u_j \in (0, \eta'_{\lambda_j})) I(\eta_{\lambda_j} \in \mathcal{A}_j(\mathbf{z}))}{\eta'_{\lambda_j} \text{Leb}(\mathcal{A}_j(\mathbf{z}))} \\ &= \prod_{j=1}^k \frac{I(u_{t_j} \in (0, \eta'_{\lambda t_j})) I(\eta_{\lambda t_j} \in \mathcal{A}_{t_j}(\tau^{-1} \mathbf{z}))}{\eta'_{\lambda t_j} \text{Leb}(\mathcal{A}_{t_j}(\tau^{-1} \mathbf{z}))} \\ &= f(\tau \mathbf{u}, \tau \boldsymbol{\eta}_\lambda | \tau \mathbf{u}', \tau \boldsymbol{\eta}'_\lambda, \tau^{-1} \mathbf{z}) \end{aligned}$$

για κάθε $\tau = (t_1, \dots, t_k) \in \mathcal{T}_k$ και τούτο ολοκληρώνει την απόδειξη. \square

Ακολουθώς δίνουμε μια ικανή συνθήκη, υπό την ισχύ της οποίας εξασφαλίζεται η (4.21) για έναν υβριδικό αλγόριθμο Metropolis within Gibbs. Ας υποθέσουμε ότι δεν είναι δυνατή η προσομοίωση από την πλήρη δεσμευμένη κατανομή του η_λ για κάποιον δείκτη $\lambda \in \{1, \dots, m\}$. Έτσι, ενσωματώνουμε στον δειγματολήπτη Gibbs του Λήμματος Γ.2.1 ένα βήμα Metropolis-Hastings, ως εξής: Αρχικά προτείνεται μια υποψήφια τιμή $\tilde{\boldsymbol{\eta}}_\lambda = (\tilde{\eta}_{\lambda 1}, \dots, \tilde{\eta}_{\lambda k}) \sim q(\cdot | \boldsymbol{\eta}'_\lambda)$ και στην συνέχεια γίνεται αποδεκτή (δηλαδή θέτουμε $\eta_\lambda^{(m+1)} = \tilde{\boldsymbol{\eta}}_\lambda$) με πιθανότητα

$$\alpha(\boldsymbol{\eta}'_\lambda, \tilde{\boldsymbol{\eta}}_\lambda) = \min \left\{ 1, \frac{f(\mathbf{z}, \boldsymbol{\eta}_{1:(\lambda-1)}, \tilde{\boldsymbol{\eta}}_\lambda, \boldsymbol{\eta}'_{(\lambda+1):m} | \mathbf{x}) q(\tilde{\boldsymbol{\eta}}_\lambda | \boldsymbol{\eta}'_\lambda)}{f(\mathbf{z}, \boldsymbol{\eta}_{1:(\lambda-1)}, \boldsymbol{\eta}'_{\lambda:m} | \mathbf{x}) q(\boldsymbol{\eta}'_\lambda | \tilde{\boldsymbol{\eta}}_\lambda)} \right\},$$

διαφορετικά επαναλαμβάνεται η προηγούμενη τιμή, δηλαδή $\eta_\lambda^{(m+1)} = \eta_\lambda^{(m)}$.

Λήμμα Γ.2.3. Έστω ότι στον δειγματοληπτή Gibbs που περιγράφεται στο Λήμμα Γ.2.1 η ανανέωση του η_λ για κάποιον δείκτη $\lambda \in \{1, \dots, m\}$ γίνεται μέσω ενός βήματος Metropolis-Hastings με την κατανομή πρότασης $q(\tilde{\eta}_\lambda | \eta'_\lambda)$. Αν η κατανομή πρότασης έχει την ιδιότητα

$$q(\tilde{\eta}_\lambda | \eta'_\lambda) = q(\tau \tilde{\eta}_\lambda | \tau \eta'_\lambda), \quad \forall \tau \in \mathcal{I}_k, \quad (\Gamma.5)$$

η (4.21) εξακολουθεί να ισχύει.

Απόδειξη. Έστω \mathcal{H}_λ ο παραμετρικός χώρος του η_λ . Σε αυτήν την περίπτωση, ο πυρήνας μετάβασης της αλυσίδας γράφεται όπως στην (Γ.3) με την διαφορά ότι ο λ-οστός όρος του δεύτερου γινομένου αντικαθίσταται από την έκφραση

$$\int_{\mathcal{A}} q(\tau \tilde{\eta}_\lambda | \tau \eta'_\lambda) \alpha(\tau \eta'_\lambda, \tau \tilde{\eta}_\lambda) d\tilde{\eta}_\lambda + I(\tau \eta'_\lambda = \tau \tilde{\eta}_\lambda) \left\{ 1 - \int_{\mathcal{H}_\lambda} q(\tau \tilde{\eta}_\lambda | \tau \eta'_\lambda) \alpha(\tau \eta'_\lambda, \tau \tilde{\eta}_\lambda) d\tilde{\eta}_\lambda \right\},$$

για κάθε μετρήσιμο υποσύνολο \mathcal{A} του \mathcal{H}_λ . Είναι σαφές ότι αν ισχύει η (Γ.5), τότε ισχύει και $\alpha(\tau \eta'_\lambda, \tau \tilde{\eta}_\lambda) = \alpha(\eta'_\lambda, \tilde{\eta}_\lambda)$, οπότε η τελευταία έκφραση ισούται με την

$$\int_{\mathcal{A}} q(\tilde{\eta}_\lambda | \eta'_\lambda) \alpha(\eta'_\lambda, \tilde{\eta}_\lambda) d\tilde{\eta}_\lambda + I(\eta'_\lambda = \tilde{\eta}_\lambda) \left\{ 1 - \int_{\mathcal{H}_\lambda} q(\tilde{\eta}_\lambda | \eta'_\lambda) \alpha(\eta'_\lambda, \tilde{\eta}_\lambda) d\tilde{\eta}_\lambda \right\}$$

και τούτο ολοκληρώνει την απόδειξη. \square

Σημειώνουμε ότι η ισχύς της συνθήκης (4.21) δεν είναι πάντα δεδομένη, αν και οι περιπτώσεις που δεν ισχύει φαίνεται να μην παρουσιάζουν ιδιαίτερο πρακτικό ενδιαφέρον. Για παράδειγμα ας υποθέσουμε ότι $k = 2$ και η ανανέωση των παραμέτρων η_1, η_2 γίνεται με ένα βήμα Metropolis-Hastings χρησιμοποιώντας μία κατανομή πρότασης με πυκνότητα $q(\tilde{\eta}_1, \tilde{\eta}_2 | \eta'_1, \eta'_2) = \varphi(\tilde{\eta}_1 | \eta'_1, 2) \varphi(\tilde{\eta}_2 | \eta'_2, 4)$, όπου $\varphi(x | \mu, \sigma^2)$ συμβολίζει την πυκνότητα πιθανότητας της $\mathcal{N}(\mu, \sigma^2)$ υπολογισμένη στο σημείο x . Σε αυτήν την περίπτωση είναι σαφές ότι η ιδιότητα (Γ.5) δεν ισχύει, διότι θεωρώντας την μετάθεση $\tau = (2, 1)$ στους δείκτες των παραμέτρων, τότε $q(\tilde{\eta}_2, \tilde{\eta}_1 | \eta'_2, \eta'_1) = \varphi(\tilde{\eta}_2 | \eta'_2, 2) \varphi(\tilde{\eta}_1 | \eta'_1, 4) \neq q(\tilde{\eta}_1, \tilde{\eta}_2 | \eta'_1, \eta'_2)$. Όπως είπαμε όμως, τέτοιες περιπτώσεις είναι μάλλον αδικαιολόγητες από την πλευρά της μοντελοποίησης, καθώς δεν είναι συνηθισμένο (και λογικό) να χρησιμοποιούμε διαφορετικές κατανομές πρότασης για μια k -άδα παραμέτρων που παίζουν τον ίδιο ρόλο στο μοντέλο. Το λογικό θα ήταν να χρησιμοποιηθεί η ίδια κλίμακα στον τυχαίο περίπατο για την ανανέωση των η_1, η_2 οπότε τότε θα ίσχυε η (Γ.5) και επομένως και το Λήμμα (Γ.2.3).

Βιβλιογραφία

- [1] Aitkin M. (2001). Likelihood and Bayesian analysis of mixtures. *Statistical Modelling*, **1**, 287-304.
- [2] Albert, J. and Chib, S. (1993). Bayesian Inference via Gibbs Sampling of Autoregressive Time-Series Subject to Markov Mean and Variance Shifts. *Journal of Business and Economic Statistics*, **11**, 1-15.
- [3] Berkelaar, M. and others (2008). lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs. R package version 5.6.4.
- [4] Bernardo, J. and Smith, A. (1994). *Bayesian Theory*, John Wiley, New York.
- [5] Bertillon, A. (1887). De l'identification par les signalements anthropométriques. *Revue Pénitentiaire (Bulletin de la Société des Prisons)*, **11**, 272-297.
- [6] Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, New York.
- [7] Burkard, R., Amico, M.D. and Martello S. (2009). *Assignment Problems*, SIAM, ISBN 978-0-898716-63-4.
- [8] Burrau, C. (1934). The half-invariants of the sum of two typical laws of errors, with an application to the problem of dissecting a frequency curve into components. *Skandinavian Aktuarietidskrift*, **17**, 1-6.
- [9] Cappé, O., Guillin, A., Marin, J. and Robert, C.P. (2003). Population Monte Carlo. *Journal of Computational and Graphical Statistics*, **13**, 907-929.
- [10] Cappé, O., Moulines, E. and Rydén, T. (2005). *Hidden Markov Models*. Springer-Verlag, New York.

- [11] Cappé, O., Robert, C.P. and Rydén, T. (2003). Reversible jump MCMC, birth-and-death and more general continuous time samplers. *Journal of the Royal Statistical Society, Series B*, **65**, 679–700.
- [12] Carlin B.P. and Chib S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, **57**, 473–484.
- [13] Carpaneto, P. (1980). Algorithm 548: Solution of the assignment problem. *ACM Transactions on Mathematical Software*, **6**, 104–111.
- [14] Casella, G., Robert, C.P. and Wells, M. (2004). Mixture models latent variables and partitioned importance sampling. *Statistical Methodology*, **1**, 1–18.
- [15] Celeux, G. (1997). Discussion of “On Bayesian analysis of mixtures with an unknown number of components” by S. Richardson and P.J. Green. *Journal of the Royal Statistical Society, Series B*, **59**, 775–776.
- [16] Celeux, G., Hurn, M. and Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, **95**, 957–970.
- [17] Celeux, G., Forbes, F., Robert, C.P. and Titterton D.M. (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis*, **1**, 651–674.
- [18] Charlier, C.V.L. (1906). Researches into the theory of probability. *Lunds Universitets Arsskrift, Ny foljd 2*, **1**, 33–38.
- [19] Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**, 1313–1321.
- [20] Chib, S. (1996). Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics*, **75**, 79–97.
- [21] Ciuperca, G., Ridolfi, A. and Idier, J. (2003). Penalized maximum likelihood for normal mixtures. *Scandinavian Journal of Statistics*, **30**, 45–59.
- [22] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

- [23] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey.
- [24] Crawford, S.L. (1994). An application of the Laplace method to finite mixture distributions. *Journal of the American Statistical Association*, **89**, 259-267.
- [25] Damien, P., Wakefield, J. and Walker, S. (1999). Gibbs Sampling for Bayesian Non-conjugate and Hierarchical Models by using Auxiliary Variables. *Journal of the Royal Statistical Society, Series B*, **61**, 331-344.
- [26] Dellaportas, P., Stephens, D.A., Smith, A.F.M. and Guttman, I. (1996). A comparative study of perinatal mortality using a two-component mixture model. *Bayesian Biostatistics*, D.A. Berry and D.K. Stangl (eds), 601-616. Dekker, New York.
- [27] Dellaportas, P. and Papageorgiou, I. (2005). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, **16**, 57-68.
- [28] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum Likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- [29] Diebolt, J. and Robert, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the American Statistical Association*, **96**, 194-209.
- [30] Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577-588.
- [31] Ferguson, T.S. 1973. A Bayesian analysis of some non parametric problems. *Annals of Statistics*, **1**, 209-230.
- [32] Francq, C. and Roussinol, M. (1997). On white noises driven by Hidden Markov Chains. *Journal of Time Series Analysis*, **18**, 553-578.
- [33] Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, **56**, 363-375.

- [34] Garel, B. (2007). Recent asymptotic results in testing for mixtures. *Computational Statistics and Data Analysis*, **51**, 5295–5304.
- [35] Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.
- [36] Green, P.J. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
- [37] Green, P.J. and Richardson, S. (2002). Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*, **97**, 1055–1076.
- [38] Grün, B. and Leisch, F. (2009). Dealing with label switching in mixture models under genuine multimodality. *Journal of Multivariate Analysis*, **100**, 851–861.
- [39] Hartigan, J.A. (1985a). Statistical theory in clustering. *Journal of Classification*, **2**, 63–76.
- [40] Hartigan, J.A. (1985b). A failure of likelihood asymptotics for normal mixtures. *Proceedings of the Berkeley Conference in Honor of Jersy Neyman and Jack Kiefer*, **2**, Monterey: Wadsworth, 807–810.
- [41] Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, **8**, 431–444.
- [42] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- [43] Hoogerheide, L.F., Kaashoek, J.F. and van Dijk, H.K. (2007). On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks. *Journal of Econometrics*, **139**, 154–180.
- [44] Hosmer, D.W. (1973a). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, **29**, 761–770.
- [45] Hosmer, D.W. (1973b). On MLE of the parameters of a mixture of two normal distributions when the sample size is small. *Computational Statistics*, **1**, 217–227.

- [46] Hurn, M., Justel, A. and Robert, C.P. (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, **12**, 55–79.
- [47] Hurvich, C.M., Simonoff, S. and Tsai, C.L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, Series B*, **60**, 271–293.
- [48] Jasra, A., Holmes, C.C. and Stephens D.A. (2005). Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modelling. *Statistical Science*, **20**, 50–67.
- [49] Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.
- [50] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood equation in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, **27**, 887–906.
- [51] Kirkpatrick, S., Gellat C.D. and Vecchi M.P. (1983). Optimization by simulated annealing. *Science, New Series*, **220**, 671–680.
- [52] Lange, K.L, Little, R.J.A. and Taylor, J.M.G. (1989). Robust statistical modelling using the t distribution. *Journal of the American Statistical Association*, **84**, 881–896.
- [53] Lavine, M. and West, M. (1992). A Bayesian method for classification and discrimination. *Canadian Journal of Statistics*, **20**, 451–461.
- [54] Liu, C.H. and Rubin, D.B. (1994). The ECME Algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**, 633–648.
- [55] MacLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, John Wiley, New York.
- [56] Marin, J.M., Mengersen, K. and Robert, C.P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, **25**, D. Dey and C.R. Rao (eds). Elsevier-Sciences.
- [57] Marin, J.M. and Robert, C.P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*, Springer-Verlag, New York.

- [58] Marron, J.S. and Wand, M.P. (1992). Exact mean integrated square error. *Annals of Statistics*, **20**, 712–736.
- [59] MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. London: Chapman and Hall.
- [60] Mendenhall, W. and Hader, R.J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, **45**, 504–520.
- [61] Mengersen, K. and Robert, C.P. (1996). Testing for Mixtures: a Bayesian entropic approach (with discussion). *Bayesian Statistics*, **5**, 255–276.
- [62] Metropolis, N., Rosenbluth, A.W., Rosenbluth, N.M., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.
- [63] Nobile, A. and Fearnside, A.T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing*, **17**, 147–162.
- [64] Papaspiliopoulos, O. and Roberts, G.O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, **95**, 169–186.
- [65] Papastamoulis, P. and Iliopoulos, G. (2009). Reversible jump MCMC in mixtures of normal distributions with same component means. *Computational Statistics and Data Analysis*, **53**, 900–911.
- [66] Papastamoulis, P. and Iliopoulos, G. (2010a). An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions. *Journal of Computational and Graphical Statistics* (to appear).
- [67] Papastamoulis, P. and Iliopoulos, G. (2010b). Handling the label switching phenomenon in hidden Markov models via the ECR Algorithm. (submitted).
- [68] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Statistical Society of London, A*, **185**, 71–110.

- [69] Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: skew variation. *Philosophical Transactions of the Royal Statistical Society of London, A*, **186**, 343–414.
- [70] Philips, D.B. and Smith, A.F. (1996). Bayesian model comparison via jump diffusions, in *Markov Chain Monte Carlo in Practice*, eds. WR Gilks, S. Richardson, and DJ Spiegelhalter, New York: Chapman and Hall, 215–239.
- [71] Postman, M., Huchra, J.P. and Geller, M.J. (1986). Probes of large-scale structure in the Corona Borealis region. *The Astronomical Journal*, **92**, 1238–1247.
- [72] Preston, E.J. (1953). A graphical method for the analysis of statistical distributions into two normal components. *Biometrika*, **40**, 460–464.
- [73] R Development Core Team (2009). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [74] Rao, C.R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B*, **10**, 159–193.
- [75] Redner, R.A. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Annals of Statistics*, **9**, 225–228.
- [76] Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**, 195–239.
- [77] Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B*, **59**, 731–792.
- [78] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 2nd edition.
- [79] Robert, C.P. (2001). *The Bayesian Choice*. Springer-Verlag, New York.
- [80] Robert, C.P., Celeux, G. and Diebolt, J. (1993). Bayesian Estimation of Hidden Markov Models: a stochastic implementation. *Statistics and Probability Letters*, **16**, 77–83.

- [81] Robert, C.P. and Mengersen, K. (1999). Reparameterization issues in mixture estimation and their bearings on the Gibbs sampler. *Computational Statistics and Data Analysis*, **29**, 325–343.
- [82] Robert, C.P., Rydén, T. and Titterton, D.M. (2000). Bayesian Inference in Hidden Markov Models through the Reversible Jump Markov Chain Monte Carlo Method. *Journal of the Royal Statistical Society, Series B*, **62**, 57–75.
- [83] Roeder, K. (1992). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *Journal of the American Statistical Association*, **85**, 617–624.
- [84] Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, **92**, 894–902.
- [85] Shumway, R.H. and Stoffer, D.S. (1991). Dynamic Linear Models With Switching. *Journal of the American Statistical Association*, **86**, 763–769.
- [86] Spezia, L. (2009). Reversible Jump and the Label Switching Problem in Hidden Markov Models. *Journal of Statistical Planning and Inference*, **139**, 2305–2315.
- [87] Stephens, M. (1997a). Bayesian methods for mixtures of normal distributions. D. Phil dissertation, Department of Statistics, University of Oxford.
- [88] Stephens, M. (1997b). Discussion of “On Bayesian analysis of mixtures with an unknown number of components” by S. Richardson and P.J. Green. *Journal of the Royal Statistical Society, Series B*, **59**, 768–769.
- [89] Stephens, M. (2000a). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society Series B*, **62**, 795–809.
- [90] Stephens, M. (2000b). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Annals of Statistics*, **28**, 40–74.
- [91] Tanner, M. and Wong, W. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, **82**, 528–540.

- [92] Verdinelli, I. and Wasserman, L. (1992). Bayesian analysis of outliers problems using the Gibbs sampler. *Statistics and Computing*, **1**, 105–117.
- [93] Viallefont, V., Richardson, S. and Green, P.J. (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics*, **14**, 181–202.
- [94] Wolfe, J.H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, **5**, 329–350.