

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΜΠΕΪΖΙΑΝΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΣΕ
ΜΟΝΤΕΛΑ ΠΕΠΕΡΑΣΜΕΝΩΝ ΜΕΙΞΕΩΝ
ΚΑΤΑΝΟΜΩΝ ΜΕ ΧΡΗΣΗ
ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΤΕΧΝΙΚΩΝ**

Παναγιώτης Β. Παπασταμούλης

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

Πειραιάς
Ιανουάριος 2006

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΜΠΕΥΪΖΙΑΝΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΣΕ
ΜΟΝΤΕΛΑ ΠΕΠΕΡΑΣΜΕΝΩΝ ΜΕΙΞΕΩΝ
ΚΑΤΑΝΟΜΩΝ ΜΕ ΧΡΗΣΗ
ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΤΕΧΝΙΚΩΝ**

Παναγιώτης Β. Παπασταμούλης

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

Πειραιάς
Ιανουάριος 2006

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

Τα μέλη της Επιτροπής ήταν:

- Ηλιόπουλος Γιώργος (Επιβλέπων)
- Κούτρας Μάρκος
- Μπούτσικας Μιχάλης

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

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

**BAYESIAN INFERENCE ON
FINITE MIXTURES OF
DISTRIBUTIONS USING
COMPUTATIONAL METHODS**

By

Panagiotis V. Papastamoulis

MSc Dissertation

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

Piraeus, Greece
January 2006

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

*Στους γονείς μου Χρυσούλα και Βασίλη
και στον αδελφό μου Γρηγόρη*

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Ευχαριστίες

Έχοντας ολοκληρώσει την παρούσα διπλωματική εργασία αισθάνομαι την ανάγκη να ευχαριστήσω τους ανθρώπους που συνετέλεσαν στην ολοκλήρωσή της. Κατ' αρχάς θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Γιώργο Ηλιόπουλο, αφ' ενός μεν για την εμπιστοσύνη που μου έδειξε με την ανάθεση της εργασίας αυτής, αφ' ετέρου δε για την καθοδήγηση και τις εύστοχες υποδείξεις του, αλλά και για την ευκαιρία που μου έδωσε να γνωρίσω την Μπεϋζιανή πλευρά της Στατιστικής και για αυτό θεωρώ την συμβολή του ανεκτίμητη. Ακόμη θα ήθελα να ευχαριστήσω τον φίλο και συμφοιτητή μου Σωτήρη Αδαμάκη για εκείνες τις 25 ώρες κατάληψης του υπολογιστή του για το τρέξιμο ενός αλγορίθμου, όταν ο δικός μου δεν άντεχε κάτι τέτοιο. Τέλος, ο κυριότερος λόγος ύπαρξης αυτής της μικρής παραγράφου οφείλεται στους γονείς μου Χρυσούλα και Βασίλη, για την αμέριστη υποστήριξη και συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Παναγιώτης Β. Παπασταμούλης

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Περίληψη

Αντικείμενο αυτής της εργασίας είναι η εκτιμητική πεπερασμένων μείξεων κατανομών, μία περιοχή της Στατιστικής Συμπερασματολογίας, όπου μέχρι πρόσφατα λίγα πράγματα ήταν γνωστά. Τα τελευταία χρόνια η εικόνα αυτή έχει ανατραπεί, κυρίως λόγω του συγκερασμού προϋπαρχόντων μαθηματικών εργαλείων (όπως η θεωρία μέτρου και οι στοχαστικές ανελίξεις) με τις σημερινές διαθέσιμες υπολογιστικές τεχνικές και δυνατότητες· ο συνδυασμός αυτών των δύο γέννησε αυτό που ονομάζουμε μεθόδους MCMC. Η αφετηρία αυτών των μεθόδων είναι η θεώρηση των παραμέτρων των μοντέλων ως τυχαίες μεταβλητές, κάτι που φανερώνει τον Μπεϋζιανό χαρακτήρα της εργασίας. Αξιοσημείωτο είναι δε το γεγονός ότι πριν την ανάπτυξη των συγκεκριμένων μεθόδων η Μπεϋζιανή προσέγγιση (ως Στατιστική μεθοδολογία γενικότερα) είχε υποστεί έντονη κριτική από τους «κλασσικούς» Στατιστικούς, και οδήγησε σε ιστορικές διαμάχες οι οποίες - εκ των υστέρων - άρχισαν να κοπάζουν.

Η σειρά παρουσίασης που ακολουθήθηκε στην εργασία αυτή είναι η εξής:

Στο πρώτο κεφάλαιο περιγράφουμε κάποια βασικά χαρακτηριστικά των μείξεων κατανομών και εξηγούμε γιατί μία προσέγγιση που οδηγεί σε κλειστούς τύπους εκτιμητών είναι πρακτικά ασύμφορη. Επίσης σημειώνουμε κάποια παράδοξα που εμφανίζονται όταν έχουμε να κάνουμε με μοντέλα μείξεων με αποτέλεσμα το «ξεγέλασμα» κλασσικών μεθοδολογιών όπως ο αλγόριθμος EM. Τέλος, για λόγους πληρότητας, αναφερόμαστε σε κάποια εργαλεία που χρησιμοποιούνται από την παραμετρική Μπεϋζιανή στατιστική.

Στο δεύτερο κεφάλαιο πραγματευόμαστε τις συνήθεις μεθόδους συμπερασματολογίας των μοντέλων μείξεων όταν ο αριθμός των συνιστωσών θεωρείται γνωστός: διεύρυνση δεδομένων και δειγματολήπτης Gibbs, αλγόριθμος Metropolis - Hastings, με αντίστοιχα παραδείγματα εφαρμογής αυτών. Επί πλέον κάνουμε μια σύντομη αναφορά στην μέθοδο Population Monte Carlo.

Στο τρίτο κεφάλαιο παρουσιάζουμε δύο βασικές μεθόδους συμπερασματολογίας στην περίπτωση που ο αριθμός των συνιστωσών της μείξης είναι μέρος των προς εκτίμηση ποσοτήτων. Πρόκειται για την μέθοδο Reversible Jump MCMC (RJMCMC) των Richardson & Green και την εναλλακτική πρόταση του Stephens, την μέθοδο Birth - Death MCMC (BDMCMC).

Στο Παράρτημα (κεφάλαιο 4) βρίσκονται κάποιες αποδείξεις προτάσεων που δεν υπάρχουν στην πρωτότυπη εργασία των Richardson & Green. Ακόμη παρατίθεται ο κώδικας υλοποίησης των αλγορίθμων RJMCMC και BDMCMC στην γλώσσα FORTRAN και Mathematica αντίστοιχα, για το πρόβλημα μείξης κανονικών κατανομών.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ABSTRACT

The subject of this dissertation is the estimation of finite mixtures of distributions, an area of Statistical Inference where few things were known until recently. Nonetheless, the last years things have changed, due to the combination of already existing mathematical tools (measure theory and stochastic processes) with the contemporary computational techniques and possibilities; this led in the so called MCMC methods. In order to apply these methods it is required to consider the model parameters as random variables, so we have followed the Bayesian approach. The Bayesian paradigm was a cause for conflict between the Statisticians of the 20th century, a conflict that began to cease after the consolidation of MCMC methods.

This dissertation is organised as follows:

In the first chapter we detail some basic properties of mixtures and we explain why an approach that leads to closed forms of estimators is computationally too expensive to be used for more than few observations. Also we illustrate the paradoxes of mixture models and their detrimental results to classical methodologies as the EM algorithm. We close this chapter with a review of the tools of non - parametric Bayesian Statistics.

In the second chapter we discuss the two basic simulation methodologies when the number of components of the mixture is known: data augmentation and the Gibbs sampler and the Metropolis - Hastings algorithm. Also we make a short reference to the Population Monte Carlo methodology.

In the third chapter we present the additional methodologies when the number of components of the mixture model is unknown: Reversible Jump MCMC (RJMCMC) methodology of Richardson and Green and the alternative approach of Stephens which is called Birth - Death MCMC (BDMCMC).

In the Appendix (chapter 4) we give the proof of some propositions that cannot be found in the original paper of Richardson and Green. Also we give the RJMCMC algorithm in FORTRAN language and the BDMCMC algorithm in Mathematica package, in a normal mixture framework.

Εισαγωγή

Σήμερα οι αναλυτές δεδομένων είναι πιο ικανοί στην ακριβέστερη περιγραφή, εκτίμηση, πρόβλεψη και συμπερασματολογία περίπλοκων συστημάτων, χάρη στις ισχυρότερες από ποτέ υπολογιστικές τεχνικές αλλά και στις ευρύτερες κατανομές μοντελοποίησης. Τα μοντέλα μείξης κατανομών αποτελούν ένα χαρακτηριστικό παράδειγμα της παραπάνω κατάστασης: από την μία, στα πλαίσια μιας παραμετρικής οικογένειας, προσφέρουν ικανοποιητικές προσεγγίσεις σε μη παραμετρικά προβλήματα, αν και βασίζονται σε συνήθεις κατανομές, από την άλλη θέτουν υπολογιστικές προκλήσεις υψηλής πολυπλοκότητας και είναι εύκολο να αντιμετωπίσουν προβλήματα ταυτοποίησης και να πέσουν στην κλάση των 'ill posed' προβλημάτων. Επίσης παρέχουν την ευκαιρία εφαρμογής νέων τεχνικών, από τον αλγόριθμο EM μέχρι και την μέθοδο *reversible jump MCMC*.

Ειδικότερα, τα μοντέλα μείξεων αναδεικνύουν παραδειγματικά την τρομερή ευχέρεια που παρέχεται από νέες υπολογιστικές τεχνικές όπως οι αλγόριθμοι *Markov Chain Monte Carlo (MCMC)*. Πριν γίνουν δημοφιλείς οι μέθοδοι MCMC, απλά δεν υπήρχε κάποια ικανοποιητική προσέγγιση στον υπολογισμό των εκτιμητών Bayes για μείξεις κατανομών, αν και αργότερα ανακαλύφθηκαν αλγόριθμοι που βασίζονται στην προγενέστερη μεθοδολογία *importance sampling* για την προσομοίωση των εκ των υστέρων κατανομών των παραμέτρων της μείξης (Casella et al., 2002).

Οι μείξεις κατανομών συμβιβάζουν έναν πεπερασμένο ή άπειρο αριθμό συνιστωσών, πιθανώς διαφορετικών τύπων κατανομών, οι οποίες μπορούν να περιγράψουν διαφορετικά χαρακτηριστικά των δεδομένων. Έτσι διευκολύνουν την προσεκτική περιγραφή περίπλοκων συστημάτων, κάτι που στοιχειοθετείται από τον ενθουσιασμό με τον οποίο έχουν υιοθετηθεί σε τόσο ευρείες περιοχές όπως η αστρονομία, οικολογία, βιοπληροφορική, υπολογιστική επιστήμη, μηχανική, ρομποτική και βιοστατιστική. Για παράδειγμα στην γενετική, η θέση των ποσοτικών χαρακτηριστικών σε ένα χρωμόσωμα και η μετάφραση των μικρο-συστοιχιών σχετίζονται με μείξεις, ενώ στην επιστήμη των υπολογιστών, τα 'spam filters' και η ανάλυση κειμένων web ξεκινούν από μία υπόθεση μείξης για να ξεχωρίσουν τα 'spams' από τα κανονικά e-mails και τις ομαδικές σελίδες από τα θέματα, αντίστοιχα (Jordan, 2004).

Οι Μπεϋζιανές προσεγγίσεις σε μοντέλα μείξεων κατανομών έχουν προσελκύσει το ενδιαφέρον πολλών ερευνητών. Η Μπεϋζιανή μεθοδολογία (Berger, 1985, Besag et al., 1995, Robert, 2001), συμπεριλαμβάνει άμεση συμπερασματολογία για τις άγνωστες παραμέτρους, εκ των προτέρων πληροφορία και ιεραρχικές δομές στα τοπικά και συνολικά χαρακτηριστικά του μοντέλου. Επίσης αυτό το πλαίσιο επιτρέπει την αποσύνθεση της πολύπλοκης δομής του μοντέλου μείξης σε απλούστερες δομές μέσω της χρήσης κρυφών ή εικονικών μεταβλητών. Τέλος μπορεί να ειπωθεί ότι όταν ο αριθμός των συνιστωσών είναι άγνωστος, η Μπεϋζιανή μεθοδολογία είναι η μόνη λογική προσέγγιση στην εκτίμηση αυτού (Richardson & Green, 1997).

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1. ΣΤΟΙΧΕΙΑ ΠΕΠΕΡΑΣΜΕΝΩΝ ΜΕΙΞΕΩΝ ΚΑΤΑΝΟΜΩΝ ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.

1.1 ΟΡΙΣΜΟΣ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.2 ΠΡΟΣΕΓΓΙΣΗ ΜΕΣΩ ΕΛΛΙΠΤΩΝ ΔΕΔΟΜΕΝΩΝ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.3 ΠΡΑΚΤΙΚΕΣ ΔΥΣΚΟΛΙΕΣ ΚΑΙ ΠΑΡΑΔΟΞΟ ΠΛΗΡΟΦΟΡΙΑΣ ΣΤΗΝ ΕΚΤΙΜΗΤΙΚΗ ΤΩΝ ΜΕΙΞΕΩΝ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.4 Ο ΑΛΓΟΡΙΘΜΟΣ EM **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.5 ΤΟ ΠΡΟΒΛΗΜΑ LABEL SWITCHING **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.5.1 Ορισμός **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

1.5.2 Περιορισμοί διάκρισης **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

1.5.3 Αλγόριθμοι αναδιάταξης και προσεγγίσεις μέσω της Θεωρίας Αποφάσεων
Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.

1.6 ΗΜΙΠΑΡΑΜΕΤΡΙΚΗ ΠΡΟΣΕΓΓΙΣΗ..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

1.6.1 Η κατανομή Dirichlet **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

1.6.2 Η διαδικασία Dirichlet..... **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

2. ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΜΕΙΞΕΩΝ ΜΕ ΓΝΩΣΤΟ ΠΛΗΘΟΣ ΣΥΝΙΣΤΩΣΩΝ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

2.1 ΔΙΕΥΡΥΝΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ Ο ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

2.2 Ο ΑΛΓΟΡΙΘΜΟΣ METROPOLIS - HASTINGS **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

2.3 Η ΜΕΘΟΔΟΣ POPULATION MONTE CARLO..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

3. ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΜΕΙΞΕΩΝ ΜΕ ΑΓΝΩΣΤΟ ΠΛΗΘΟΣ ΣΥΝΙΣΤΩΣΩΝ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

3.1 REVERSIBLE JUMP MCMC **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

3.1.1 Η γενική περίπτωση..... **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

3.1.2 Εφαρμογή RJMCMC σε κανονικές μειξεις **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

3.2 BIRTH - DEATH MCMC..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

3.2.1 Οι παράμετροι ως σημειακή διαδικασία **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

3.2.2 Διαδικασίες γεννήσεων - θανάτων για τις συνιστώσες του μοντέλου μειξης
Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.

3.2.3 Η κατασκευή της Μαρκοβιανής αλυσίδας..... **Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.**

3.3 ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ RJMCMC ΚΑΙ BDMCMC **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

ΣΥΝΟΨΗ ΚΑΙ ΣΥΣΤΑΣΕΙΣ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

4. ΠΑΡΑΡΤΗΜΑ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

4.1 ΑΠΟΔΕΙΞΗ ΠΡΟΤΑΣΗΣ 3.1.1..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

- 4.2 ΑΠΟΔΕΙΞΗ ΠΡΟΤΑΣΗΣ 3.1.3..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**
- 4.3 ΚΩΔΙΚΑΣ ΑΛΓΟΡΙΘΜΟΥ RJMCMC ΣΤΗΝ FORTRAN..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**
- 4.4 ΚΩΔΙΚΑΣ ΑΛΓΟΡΙΘΜΟΥ BDMCMC ΣΤΟ MATHEMATICA... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

ΒΙΒΛΙΟΓΡΑΦΙΑ **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

ΕΥΡΕΤΗΡΙΟ ΑΛΓΟΡΙΘΜΩΝ..... **ΣΦΑΛΜΑ! ΔΕΝ ΕΧΕΙ ΟΡΙΣΤΕΙ ΣΕΛΙΔΟΔΕΙΚΤΗΣ.**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑΣ

ΚΕΦΑΛΑΙΟ 1

ΣΤΟΙΧΕΙΑ ΠΕΠΕΡΑΣΜΕΝΩΝ ΜΕΙΞΕΩΝ ΚΑΤΑΝΟΜΩΝ

*Βλέπουμε την αλήθεια,
αλλά βασιλεύει η πλάνη...*
K. Gödel

1. Στοιχεία πεπερασμένων μίξεων κατανομών

1.1 Ορισμός

Κάθε κυρτός συνδυασμός της μορφής

$$\sum_{i=1}^k p_i f_i(x) \text{ με } \sum_{i=1}^k p_i = 1 \text{ και } k > 1, \quad (1.1)$$

όπου f_1, \dots, f_k είναι πυκνότητες πιθανότητας καλείται *πεπερασμένη μείξη κατανομών*. Αν και έχουν οριστεί στην βιβλιογραφία και *συνεχείς μίξεις*

$$g(x) = \int_{\Theta} f(x|\theta)h(\theta)d\theta \text{ με } \int_{\Theta} h(\theta)d\theta = 1$$

δεν θα τις αντιμετωπίσουμε εδώ. Στις περισσότερες περιπτώσεις οι κατανομές f_i θεωρούμε ότι ανήκουν σε μία δεδομένη παραμετρική οικογένεια με άγνωστη παράμετρο θ_i , κάτι που οδηγεί στον ορισμό του *παραμετρικού μοντέλου μείξης*

$$\sum_{i=1}^k p_i f_i(x|\theta_i). \quad (1.2)$$

Στην ειδική περίπτωση όπου τα $f_i(\cdot|\theta_i)$ είναι όλα κανονικές κατανομές με το θ_i να δηλώνει την άγνωστη μέση τιμή και διασπορά, το εύρος του σχήματος και των χαρακτηριστικών της μείξης (1.2) μπορεί να ποικίλλει αρκετά όπως φαίνεται στο Σχήμα 1.

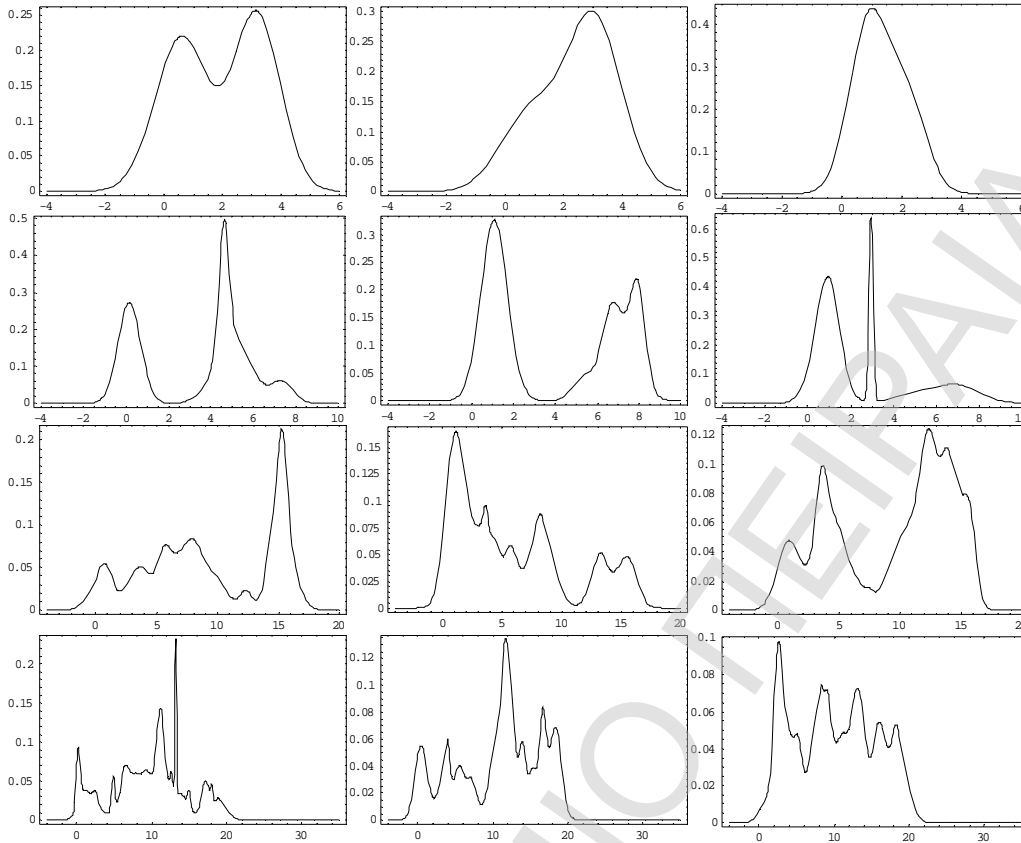
Εφ' όσον θα χρησιμοποιήσουμε μίξεις για να προσεγγίσουμε άγνωστες κατανομές, σημειώνουμε σε αυτό το σημείο ότι η συμπεριφορά των ουρών μιας μείξης περιγράφεται πάντα από μία ή δύο συνιστώσες αυτής, κάτι που δικαιολογεί την επιλογή της παραμετρικής οικογένειας $f_i(\cdot|\theta_i)$. Επίσης η αναπαράσταση μιας μείξης ως κυρτό συνδυασμό κατανομών συνεπάγεται ότι οι ροπές της (1.1) είναι κυρτοί συνδυασμοί των ροπών των $f_i(\cdot|\theta_i)$:

$$E(X^m) = \sum_{i=1}^k p_i E^{f_i}(X^m).$$

Ο Karl Pearson εκμεταλλεύτηκε αυτό το γεγονός για να παραγάγει έναν εκτιμητή ροπών των παραμέτρων μιας κανονικής μείξης με δύο συνιστώσες:

$$p\varphi(x; \mu_1, \sigma_1) + (1-p)\varphi(x; \mu_2, \sigma_2), \quad (1.3)$$

όπου με $\varphi(\cdot; \mu, \sigma)$ συμβολίζουμε την πυκνότητα της κατανομής $N(\mu, \sigma^2)$.



ΣΧΗΜΑ 1. Μερικές πυκνότητες¹ κανονικών μίξεων για $k=2$ (πρώτη γραμμή), $k=10$ (δεύτερη γραμμή), $k=25$ (τρίτη γραμμή) και $k=50$ (τελευταία γραμμή).

Δυστυχώς, η αναπαράσταση του μοντέλου μείξης (1.2) είναι επιζήμια στην εξαγωγή του εκτιμητή μεγίστης πιθανοφάνειας (όταν αυτός υπάρχει) και του εκτιμητή Bayes. Ας θεωρήσουμε την περίπτωση n ανεξαρτήτων και ισόνομων παρατηρήσεων $\underline{x} = (x_1, x_2, \dots, x_n)$ από αυτό το μοντέλο. Ορίζοντας $\underline{p} = (p_1, \dots, p_k)$ και $\underline{\theta} = (\theta_1, \dots, \theta_k)$, βλέπουμε ότι ακόμη και στην περίπτωση που χρησιμοποιήσουμε συζυγείς εκ των προτέρων κατανομές για κάθε παράμετρο των συνιστωσών (p_i, θ_i) , η αναπαράσταση της αντίστοιχης εκ των υστέρων κατανομής συμπεριλαμβάνει το ανάπτυγμα της πιθανοφάνειας

$$L(\underline{\theta}, \underline{p} | \underline{x}) = \prod_{i=1}^n \sum_{j=1}^k p_j f_j(x_i | \theta_j) \quad (1.4)$$

σε k^n όρους, το οποίο είναι πρακτικά ανέφικτο να υπολογιστεί για παραπάνω από λίγες παρατηρήσεις.

¹ Για να σχηματίσουμε αυτές τις πυκνότητες, παραγάγαμε τα βάρη από μία Dirichlet κατανομή $D(1, 1, \dots, 1)$, τους μέσους από μια ομοιόμορφη $U(0, 5 \log k)$ και τις διασπορές από μία Βήτα $B(1/(0.5 + 0.1 \log k), 1)$ (που σημαίνει ότι οι διασπορές είναι μικρότερες του 1). Οι κατανομές που πήραμε αντικατοπτρίζουν αυτή την επιλογή.

1.2 Προσέγγιση μέσω ελλειπών δεδομένων

Υπάρχουν αρκετά κίνητρα για την θεώρηση των μοντέλων μείξεων ως μια χρήσιμη επέκταση συνήθων κατανομών. Η πλέον φυσική προσέγγιση είναι η ενσωμάτωση ενός παράγοντα ο οποίος να χωρίζει τον πληθυσμό σε διάφορα στρώματα ή υποπληθυσμούς. Ένα από τα πρώτα παραδείγματα μοντέλου μείξης συναντάται στον Bertillon (1887) όπου η διωνυμική δομή στο ύψος των (στρατιωτικών) υποκειμένων της κεντρικής Γαλλίας μπορεί να εξηγηθεί μέσω της μείξης δύο πληθυσμών νέων αντρών, ο πρώτος από τα πεδινά και ο δεύτερος από τα ορεινά. Η συμπεριφορά μείξης εμφανίζεται διότι έχει χαθεί η προέλευση κάθε παρατήρησης, δηλαδή η κατάταξη σε ένα συγκεκριμένο στρώμα (πεδινά ή ορεινά). Έτσι κάθε παρατήρηση x_i είναι εκ των προτέρων κατανομημένη σε κάθε μία από τις $f_i(\cdot|\theta_i)$ με πιθανότητα p_i . Ανάλογα με την εκάστοτε περίπτωση, ο στόχος της συμπερασματολογίας είναι η ανασύνθεση της κατάταξης των παρατηρήσεων στα στρώματα (clustering), η κατασκευή εκτιμητών για τις παραμέτρους των συνιστωσών ή ακόμα και η εκτίμηση του ίδιου του πλήθους των συνιστωσών.

Αν και όπως φαίνεται παρακάτω δεν είναι πάντα αυτός ο λόγος για την μοντελοποίηση μέσω μείξεων κατανομών, η έμφυτη δομή των ελλειπών δεδομένων στην κατανομή μπορεί να εκμεταλλευτεί για να διευκολύνει την διαδικασία των εκτιμήσεων. Μέσω μιας διαδικασίας αποπεριθωριοποίησης, είναι πάντα εφικτό να αντιστοιχίσουμε σε μία τυχαία μεταβλητή X_i από μία μείξη k κατανομών της μορφής (1.2) μία άλλη τυχαία μεταβλητή Z_i με τιμές στο σύνολο $\{1,2,\dots,k\}$, τέτοια ώστε:

$$X_i|Z_i = z \sim f(x|\theta_z) \text{ με } Z_i \sim M_k(1; p_1, \dots, p_k) \quad (1.5)$$

όπου με $M_k(1; p_1, \dots, p_k)$ συμβολίζουμε την πολυωνυμική κατανομή με k κατηγορίες και 1 παρατήρηση, ώστε $P(Z_i = j) = p_j$. Έτσι αυτή η φανταστική μεταβλητή καθορίζει σε ποια συνιστώσα ανήκει η παρατήρηση x_i . Ανάλογα με την περίπτωση οι μεταβλητές Z_i θα είναι ή δεν θα είναι μέρος των προς εκτίμηση ποσοτήτων. Στην συνέχεια θεωρούμε ότι έχουμε παρατηρήσει τα διευρυμένα δεδομένα $\{(x_1, z_1), \dots, (x_n, z_n)\}$ τα οποία αποτελούνται από ανεξάρτητα ζεύγη με κατανομή

$$P(X_i \in dx, Z_i = j) = p_j f(x|\theta_j) dx. \quad (1.6)$$

Από τα παραπάνω έπεται ότι η δεσμευμένη κατανομή των φανταστικών μεταβλητών Z_i δοθέντος $X_i = x$ είναι

$$P(Z_i = j|X_i = x) = \frac{p_j f_j(x)}{\sum_{i=1}^k p_i f_i(x)} \propto p_j f_j(x). \quad (1.7)$$

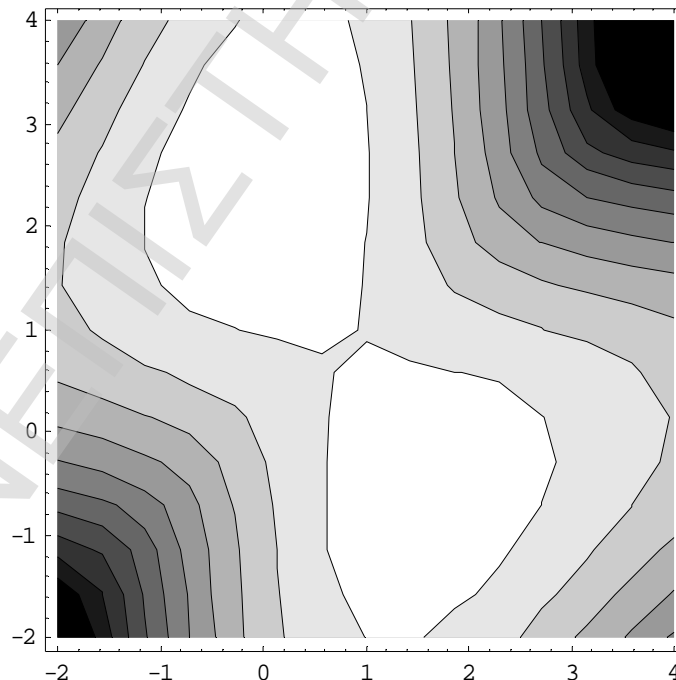
1.3 Πρακτικές δυσκολίες και παράδοξο πληροφορίας στην εκτιμητική των μείξεων

Σε αυτή την παράγραφο δείχνουμε ότι είναι δυνατή μία Μπεϋζιανή προσέγγιση, η οποία οδηγεί σε κλειστούς τύπους των εκτιμητών Bayes, αλλά ταυτόχρονα παρουσιάζεται και ένα «παράδοξο πληροφορίας» (information paradox): όταν ο αριθμός των παρατηρήσεων γίνεται μεγάλος, οι εκτιμητές δεν μπορούν να υπολογιστούν στην πράξη.

Παράδειγμα 1.3.1 Για μία πρώτη επίδειξη της πολυπλοκότητας στην διαδικασία των εκτιμήσεων, ας θεωρήσουμε την απλή περίπτωση μείξης δύο κανονικών κατανομών:

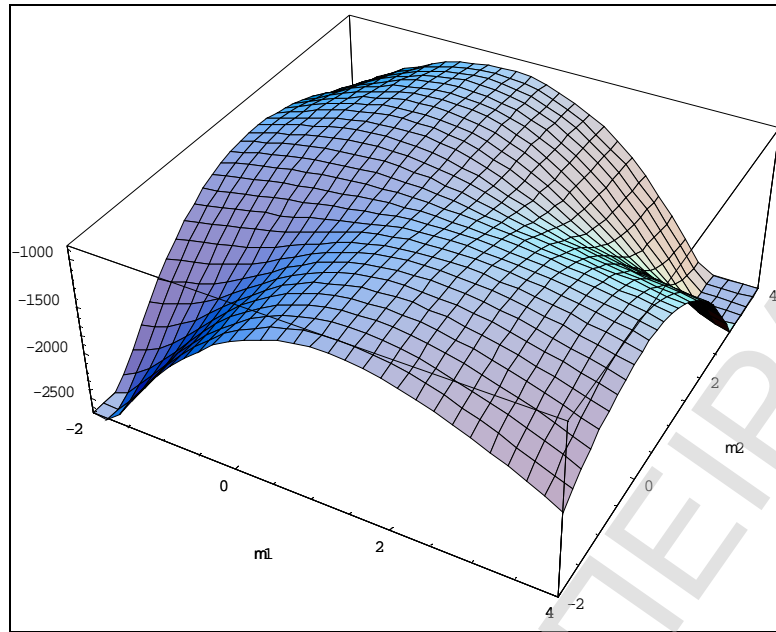
$$pN(\mu_1, 1) + (1-p)N(\mu_2, 1) \quad (1.8)$$

όπου τα βάρη $p \neq 0.5$ είναι γνωστά. Για τις ανάγκες του παραδείγματος χρησιμοποιήθηκαν 500 προσομοιωμένες παρατηρήσεις από το μοντέλο (1.8) με πραγματικές τιμές παραμέτρων $(\mu_1, \mu_2, p) = (0, 2.5, 0.7)$. Ο παραμετρικός χώρος τότε είναι το \mathcal{R}^2 και οι παράμετροι μ_1, μ_2 δεν μπορούν να μπερδευτούν μεταξύ τους (στο γεγονός αυτό θα επανέλθουμε στην Παράγραφο 1.5). Παρ' όλα αυτά, η επιφάνεια του λογαρίθμου της πιθανοφάνειας εμφανίζει δύο κορυφές: μία κοντά στην πραγματική τιμή των παραμέτρων που χρησιμοποιήθηκαν για την προσομοίωση των αντίστοιχων δεδομένων, και μία «ψευδοκορυφή», η οποία δεν σημαίνει και πολλά σχετικά με την πραγματική τιμή των παραμέτρων, αλλά είναι πάντα παρούσα λόγω της λογαριθμοποίησης της πιθανοφάνειας (Σχήμα 2).



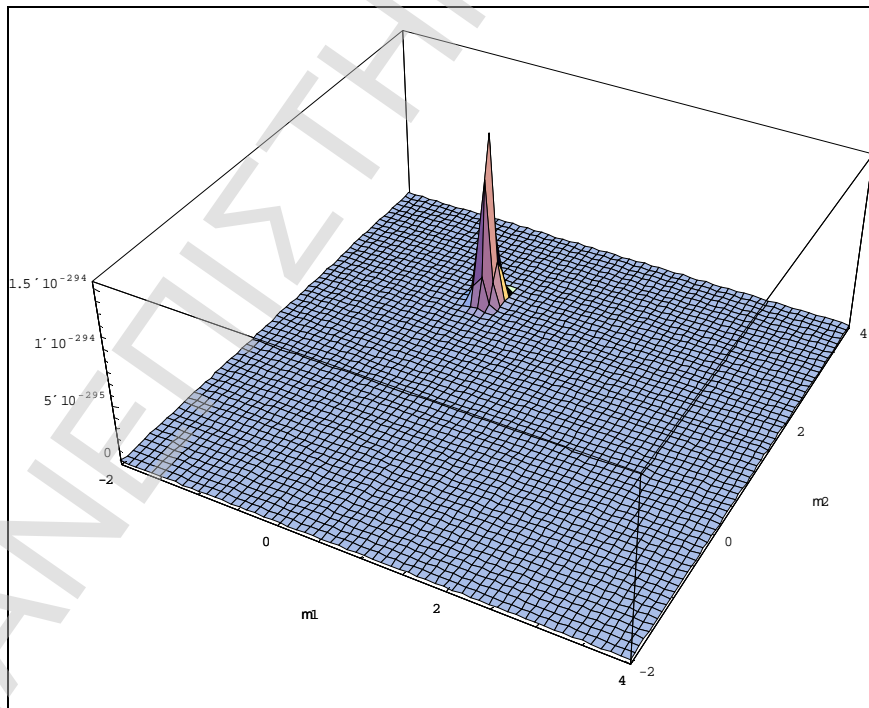
ΣΧΗΜΑ 2. Αναπαράσταση του χάρτη επιφάνειας του λογαρίθμου της πιθανοφάνειας της μείξης (1.8) για 500 προσομοιωμένες παρατηρήσεις και πραγματικές τιμές παραμέτρων $(\mu_1, \mu_2, p) = (0, 2.5, 0.7)$.

Η αντίστοιχη επιφάνεια έχει ως εξής:



ΣΧΗΜΑ 3. Η αντίστοιχη τρισδιάστατη αναπαράσταση του Σχήματος 2.

Όταν όμως απεικονίσουμε την επιφάνεια της πιθανοφάνειας (και όχι του λογαρίθμου της) παρατηρούμε ότι μόνο η πραγματική κορυφή είναι ορατή και αυτό συμβαίνει λόγω της διαφοράς της επίδρασης στην πιθανοφάνεια των τιμών των κορυφών. Λογαριθμίζοντας την πιθανοφάνεια απαλείφεται αυτή η διαφορά και έτσι σχηματίζονται οι δύο κορυφές (Σχήμα 4).



ΣΧΗΜΑ 4. Τρισδιάστατη αναπαράσταση της επιφάνειας της πιθανοφάνειας (πολλαπλασιασμένη με 10^{100}) της μείξης (1.8) για 500 προσομοιωμένες παρατηρήσεις και πραγματικές τιμές παραμέτρων $(\mu_1, \mu_2, p) = (0, 2.5, 0.7)$.

Παράδειγμα 1.3.2 Ας θεωρήσουμε n ανεξάρτητες και ισόνομες παρατηρήσεις $\underline{x} = (x_1, \dots, x_n)$ από το μοντέλο (1.2) και έστω $\pi(\underline{\theta}, \underline{p})$ η εκ των προτέρων κατανομή στο $(\underline{\theta}, \underline{p})$. Τότε για την εκ των υστέρων θα έχουμε:

$$\pi(\underline{\theta}, \underline{p} | \underline{x}) \propto \prod_{i=1}^n \sum_{j=1}^k p_j f(x_i | \theta_j) \pi(\underline{\theta}, \underline{p}) \quad (1.9)$$

Σαν μία απλή εφαρμογή του παραπάνω, ας υποθέσουμε ότι οι παρατηρήσεις προέρχονται από την (1.8) με εκ των προτέρων κατανομές ($i = 1, 2$):

$$\mu_i | \sigma_i \sim N(\zeta_i, \sigma_i^2 / \lambda_i), \quad \sigma_i^{-2} \sim \Gamma(\nu_i / 2, s_i^2 / 2), \quad p \sim B(\alpha, \beta)$$

όπου τα $\zeta_i, \nu_i, \lambda_i, \alpha, \beta$ είτε είναι υπερπαραμέτροι (οπότε θα πρέπει να οριστούν οι αντίστοιχες εκ των προτέρων κατανομές), είτε είναι σταθερές. Σε αυτή την περίπτωση $\underline{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, $\underline{p} = p$ και η εκ των υστέρων κατανομή είναι:

$$\pi(\underline{\theta}, \underline{p} | \underline{x}) \propto \prod_{i=1}^n (p \varphi(x_i; \mu_1, \sigma_1) + (1-p) \varphi(x_i; \mu_2, \sigma_2)) \pi(\underline{\theta}, \underline{p})$$

Για τον υπολογισμό της τελευταίας έκφρασης απαιτείται το ανάπτυγμα σε 2^n όρους, οπότε είναι πρακτικά ανέφικτο να καταλήξουμε σε αναλυτικές εκφράσεις του μεγίστου της πιθανοφάνειας και των εκτιμητών Bayes.

Στην συνέχεια περιγράφουμε μία άλλη αναπαράσταση της έκφρασης (1.9), κύριο χαρακτηριστικό της οποίας είναι ότι δείχνει ότι πολύ λίγες τιμές των k^n όρων έχουν μη αμελητέο ρόλο. Ας θεωρήσουμε τις φανταστικές μεταβλητές $\underline{z} = (z_1, \dots, z_n)$ όπως αυτές ορίστηκαν στην Παράγραφο 1.2. Επί πλέον, ας συμβολίσουμε με \mathfrak{S} το σύνολο όλων των δυνατών k^n διανυσμάτων κατάταξης \underline{z} . Το σύνολο \mathfrak{S} έχει ενδιαφέρουσα και πλούσια δομή. Συγκεκριμένα για k διακριτές συνιστώσες μπορούμε να διασπάσουμε το \mathfrak{S} ως εξής: Δοθέντος ενός συγκεκριμένου διανύσματος κατάταξης $\mathbf{n}_j = (n_1, \dots, n_k)$, με $n_1 + \dots + n_k = n$, ορίζουμε το σύνολο:

$$\mathfrak{S}_j = \left\{ \underline{z} : \sum_{i=1}^n I_{\{z_i=1\}} = n_1, \dots, \sum_{i=1}^n I_{\{z_i=k\}} = n_k \right\}$$

όπου αποτελείται από όλες τις δυνατές θέσεις για την δεδομένη κατάταξη (n_1, \dots, n_k) αναδιατεταγμένες για $j \in N$. Το πλήθος των μη αρνητικών ακέραιων λύσεων της διάσπασης του n σε k μέρη, τέτοια ώστε $n_1 + \dots + n_k = n$ ισούται με

$$r = \binom{n+k-1}{n}$$

Έτσι, προκύπτει η διαμέριση $\mathfrak{S} = \bigcup_{j=1}^r \mathfrak{S}_j$. Αν και ο συνολικός αριθμός των στοιχείων του \mathfrak{S} είναι k^n , ο αριθμός των συνόλων που το διαμερίζουν είναι $n^{k-1}/(k-1)!$. Ωστε η εκ των υστέρων κατανομή μπορεί να γραφτεί ως εξής:

$$\pi(\underline{\theta}, \underline{p} | \underline{x}) \propto \sum_{i=1}^r \sum_{\underline{z} \in \mathfrak{S}_i} \omega(\underline{z}) \pi(\underline{\theta}, \underline{p} | \underline{x}, \underline{z}) \quad (1.10)$$

όπου το $\omega(\underline{z})$ παριστάνει την εκ των υστέρων πιθανότητα της δοθείσης κατάταξης \underline{z} . Σημειώνουμε ότι ένας εκτιμητής Bayes σε αυτή την περίπτωση μπορεί να γραφεί ως:

$$\sum_{i=1}^r \sum_{\underline{z} \in \mathfrak{S}_i} \omega(\underline{z}) E^\pi(\underline{\theta}, \underline{p} | \underline{x}, \underline{z}).$$

Συνοψίζοντας, αυτή η αναπαράσταση της εκ των υστέρων κατανομής, θεωρεί κάθε δυνατή ανακατάταξη \underline{z} των δεδομένων στις συνιστώσες της μείξης, στην οποία αντιστοιχεί μία εκ των υστέρων πιθανότητα $\omega(\underline{z})$ και κατασκευάζει μία εκ των υστέρων κατανομή $\pi(\underline{\theta}, \underline{p} | \underline{x}, \underline{z})$ για τις παραμέτρους, δεσμεύοντας σε αυτή την συγκεκριμένη ανακατάταξη. Δυστυχώς όμως και εδώ δεν γλιτώνουμε από τις υπολογιστικές δυσκολίες εφόσον και πάλι υπάρχουν k^n όροι σε αυτό το άθροισμα. Οι Casella et al. (2000), σε μία μελέτη Monte Carlo, έδειξαν ότι πολύ λίγες διαμερίσεις έχουν μη αμελητέο βάρος. Στην συνέχεια δίνουμε ένα παράδειγμα υπολογισμού των εκ των υστέρων βαρών $\omega(\underline{z})$ σε μία απλή περίπτωση, που αναδεικνύει την πολυπλοκότητα των παραπάνω τύπων.

Παράδειγμα 1.3.1 Στην ειδική περίπτωση του μοντέλου (1.8), όπου θεωρούμε (συνέχεια) την ίδια κανονική εκ των προτέρων κατανομή στους μέσους, $\mu_1, \mu_2 \sim N(0,10)$, θα υπολογίσουμε το εκ των υστέρων βάρος $\omega(\underline{z})$ για μία κατάταξη \underline{z} , όπου στην πρώτη συνιστώσα βρίσκονται l παρατηρήσεις, δηλαδή $\sum_{i=1}^n I_{\{z_i=1\}} = l$, $(n_1, n_2) = (l, n-l)$. Έχουμε:

$$\begin{aligned} \pi(\underline{z}, \mu_1, \mu_2 | \underline{x}, n_1, n_2) &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n [I_{\{z_i=1\}}(x_i - \mu_1)^2 + (1 - I_{\{z_i=1\}})(x_i - \mu_2)^2] - \frac{\mu_1^2}{20} - \frac{\mu_2^2}{20} \right\} \\ &\times \frac{1}{(2\pi)^{n/2}} p^{\sum_{i=1}^n I_{\{z_i=1\}}} (1-p)^{n - \sum_{i=1}^n I_{\{z_i=1\}}}. \end{aligned} \quad (1.11)$$

Το εκ των υστέρων βάρος $\omega(\underline{z})$ προκύπτει ολοκληρώνοντας την (1.11) στο $\mathfrak{R} \times \mathfrak{R}$ ως προς μ_1 και μ_2 , το οποίο πρόκειται από ένα διπλό ολοκλήρωμα, που υπολογίζεται εύκολα. Για την ολοκλήρωση ως προς μ_1 , εξαιρώντας τα μέρη που δεν το περιέχουν, αρκεί να υπολογίσουμε το:

$$I_1 = \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n I_{z_i=1} (x_i - \mu_1)^2 - \frac{\mu_1^2}{20} \right\} d\mu_1. \quad (1.12)$$

Όμως:

$$\begin{aligned}
 & \exp\left\{-\frac{1}{2}\sum_{i=1}^n I_{\{z_i=1\}}(x_i - \mu_1)^2 - \frac{\mu_1^2}{20}\right\} = \\
 & = \exp\left\{-\frac{1}{2}\mu_1^2\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right) - \frac{1}{2}\sum_{i=1}^n x_i^2 I_{\{z_i=1\}} + \mu_1 \sum_{i=1}^n x_i I_{\{z_i=1\}}\right\} \\
 & = \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\mu_1^2 + \frac{\sum_{i=1}^n x_i^2 I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}} - 2\mu_1 \frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)\right\} \\
 & = \exp\left\{\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)^2\right\} \\
 & \times \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\mu_1^2 + \frac{\sum_{i=1}^n x_i^2 I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}} - 2\mu_1 \frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}} + \left(\frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)^2\right)\right\} \\
 & = c_1 \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\mu_1^2 - 2\mu_1 \frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}} + \left(\frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)^2\right)\right\} \\
 & = c_1 \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\mu_1 - \frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)^2\right\} \\
 & \text{όπου } c_1 = \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2 I_{\{z_i=1\}} + \frac{\left(\sum_{i=1}^n x_i I_{\{z_i=1\}}\right)^2}{2\left(l + \frac{1}{10}\right)}\right\}.
 \end{aligned}$$

Οπότε για τον υπολογισμό του ολοκληρώματος (1.12) έχουμε:

$$\begin{aligned}
 I_1 & = c_1 \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}\right)\left(\mu_1 - \frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}\right)^2\right\} d\mu_1 \Rightarrow \\
 I_1 & = c_1 \frac{\sqrt{2\pi}}{\sqrt{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}} = c_1 \frac{\sqrt{2\pi}}{\sqrt{l + \frac{1}{10}}} \tag{1.13}
 \end{aligned}$$

διότι το τελευταίο ολοκλήρωμα, πρόκειται για το ολοκλήρωμα σε όλο το στήριγμα,

$$\text{του εκθετικού της κατανομής } N \left(\frac{\sum_{i=1}^n x_i I_{\{z_i=1\}}}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}}, \frac{1}{\sum_{i=1}^n I_{\{z_i=1\}} + \frac{1}{10}} \right).$$

Για την ολοκλήρωση της (1.11) ως προς μ_2 , εξαιρώντας τα μέρη που δεν το περιέχουν, αρκεί να υπολογίσουμε το:

$$I_2 = \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (1 - I_{\{z_i=1\}}) (x_i - \mu_2)^2 - \frac{\mu_2^2}{20} \right\} d\mu_2$$

Ακολουθώντας την ίδια μεθοδολογία με προηγουμένως καταλήγουμε ότι:

$$I_1 = c_2 \frac{\sqrt{2\pi}}{\sqrt{n-l + \frac{1}{10}}} \quad (1.14)$$

$$\text{όπου } c_2 = \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 (1 - I_{\{z_i=1\}}) + \frac{\left(\sum_{i=1}^n x_i (1 - I_{\{z_i=1\}}) \right)^2}{2 \left(n-l + \frac{1}{10} \right)} \right\}.$$

Οπότε η εκ των υστέρων πιθανότητα $\omega(\underline{z})$ βρίσκεται αντικαθιστώντας τις σχέσεις (1.13) και (1.14) στην $\iint_{\mathfrak{R} \times \mathfrak{R}} \pi(\underline{z}, \mu_1, \mu_2 | \underline{x}, n_1, n_2) d\mu_1 d\mu_2$. Ωστε:

$$\begin{aligned} \omega(\underline{z}) &= c_1 \frac{\sqrt{2\pi}}{\sqrt{l + \frac{1}{10}}} c_2 \frac{\sqrt{2\pi}}{\sqrt{n-l + \frac{1}{10}}} p^{\sum_{i=1}^n I_{\{z_i=1\}}} (1-p)^{n - \sum_{i=1}^n I_{\{z_i=1\}}} \\ &= c_1 c_2 \frac{2\pi}{\sqrt{\left(l + \frac{1}{10} \right) \left(n-l + \frac{1}{10} \right)}} p^l (1-p)^{n-l} \end{aligned}$$

και αν αντικαταστήσουμε τα c_1, c_2 προκύπτει η - όχι και τόσο κομψή - σχέση:

$$\begin{aligned} \omega(\underline{z}) &= \frac{2\pi}{\sqrt{\left(l + \frac{1}{10} \right) \left(n-l + \frac{1}{10} \right)}} \\ &\times \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i I_{\{z_i=1\}} \right)^2}{l + \frac{1}{10}} - \frac{\left(\sum_{i=1}^n x_i (1 - I_{\{z_i=1\}}) \right)^2}{n-l + \frac{1}{10}} \right] \right\} p^l (1-p)^{n-l} \end{aligned}$$

Από αυτό το παράδειγμα (το οποίο αποτελεί και απλή περίπτωση αφού οι εκ των προτέρων κατανομές των παραμέτρων είναι ίδιες) διαπιστώνουμε ότι παρά αυτή την ενδιαφέρουσα ως προς την κατασκευή της προσέγγιση, οι τύποι που προκύπτουν είναι τόσο δύσχρηστοι που την καθιστούν πρακτικά ασύμφορη ακόμη και στις απλές εφαρμογές.

1.4 Ο αλγόριθμος EM

Μετά από μία πρώτη εντύπωση της δυσκολίας στην εκτίμηση των παραμέτρων των μείξεων, μέσω αναλυτικών τύπων, λογικό είναι να στρέψουμε την προσοχή μας (και τις ελπίδες μας για καλές εκτιμήσεις) σε υπολογιστικές τεχνικές. Μία από τις πιο κλασικές αριθμητικές μεθόδους βελτιστοποίησης, σχετικά με διαδικασίες που περιέχουν σε κάποιο βήμα το πρόβλημα της μεγιστοποίησης της πιθανοφάνειας, είναι ο αλγόριθμος EM (*Expectation - Maximization*) (Dempster et al., 1977). Παρ' όλα αυτά θα απογοητευτούμε για άλλη μία φορά καθώς, όπως θα δούμε παρακάτω, πολλές φορές αποτυγχάνει να συγκλίνει στην κύρια κορυφή της πιθανοφάνειας.

Αρχικά περιγράφουμε κάποια βασικά χαρακτηριστικά. Έστω $g(\cdot, \cdot | \theta)$ η από κοινού συνάρτηση πυκνότητας του $(\underline{X}, \underline{Z})$ δοθέντος του παραμετρικού διανύσματος θ , $f(\cdot | \theta)$ η συνάρτηση πυκνότητας του \underline{X} δοθέντος θ και $k(\cdot | \underline{x}, \theta)$ η συνάρτηση πυκνότητας της δεσμευμένης κατανομής του \underline{Z} δοθεισών των παρατηρήσεων \underline{x} και του θ . Ο αλγόριθμος βασίζεται στην χρησιμοποίηση των ελλιπών δεδομένων που περιγράψαμε στην Παράγραφο 1.2, δηλαδή μπορούμε να γράψουμε την κατανομή του δείγματος \underline{x} ως εξής:

$$\begin{aligned} f(\underline{x} | \theta) &= \int g(\underline{x}, \underline{z} | \theta) d\underline{z} \\ &= \int f(\underline{x} | \theta) k(\underline{z} | \underline{x}, \theta) d\underline{z} \end{aligned} \quad (1.15)$$

οπότε λογαριθμίζοντας την $g(\underline{x}, \underline{z} | \theta) = f(\underline{x} | \theta) k(\underline{z} | \underline{x}, \theta)$ καταλήγουμε σε έναν πλήρη (μη παρατηρούμενο) λογάριθμο της πιθανοφάνειας

$$L^c(\theta | \underline{x}, \underline{z}) = L(\theta | \underline{x}) + \log k(\underline{z} | \underline{x}, \theta),$$

όπου το L είναι ο παρατηρηθείς λογάριθμος της πιθανοφάνειας. Αυτό που κάνει ο αλγόριθμος EM είναι να συμπληρώνει τις ελλιπείς μεταβλητές \underline{z} βασιζόμενος στην $k(\underline{z} | \underline{x}, \theta)$ και στην συνέχεια να μεγιστοποιεί ως προς θ την αναμενόμενη πλήρη λογαριθμο-πιθανοφάνεια:

Γενικός αλγόριθμος EM

0. Δώσε κάποιες αρχικές τιμές στο $\theta^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots$:

1.1. Το E - βήμα, υπολόγισε:

$$Q(\theta|\theta^{(t-1)}, \underline{x}) = E_{\theta^{(t-1)}}(L^c(\theta|\underline{x}, \underline{Z})),$$

όπου $\underline{Z} \sim k(\underline{z}|\underline{x}, \theta)$.

1.2. Το M - βήμα, μεγιστοποίησε ως προς θ το $Q(\theta|\theta^{(t-1)}, \underline{x})$ και θέσε:

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

Το αποτέλεσμα κατά την εκτέλεση του παραπάνω αλγορίθμου είναι ότι σε κάθε επανάληψη, η (παρατηρηθείσα) $L(\theta|\underline{x})$ αυξάνει.

Παράδειγμα 1.4.1 Ως εφαρμογή του παραπάνω, ας θεωρήσουμε την ειδική περίπτωση του μοντέλου μείξης δύο κανονικών κατανομών (1.8), όπου όλες οι παράμετροι είναι γνωστές εκτός του $\underline{\theta} = (\mu_1, \mu_2)$. Για ένα προσομοιωμένο δείγμα 500 παρατηρήσεων και πραγματικές τιμές $p = 0.7$ και $(\mu_1, \mu_2) = (0, 2.5)$, ο λογάριθμος της πιθανοφάνειας παρουσιάζει όπως είδαμε δύο κορυφές. Εφαρμόζοντας τον αλγόριθμο EM για αυτό το μοντέλο, έχουμε ότι η πλήρης πιθανοφάνεια είναι:

$$p^{\sum_{i=1}^n I_{\{z_i=1\}}} (1-p)^{n-\sum_{i=1}^n I_{\{z_i=1\}}} (2\pi)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n [I_{\{z_i=1\}}(x_i - \mu_1)^2 + (1 - I_{\{z_i=1\}})(x_i - \mu_2)^2]\right\}$$

όποτε ο λογάριθμός της είναι

$$L^c(\theta|\underline{x}, \underline{z}) = \sum_{i=1}^n I_{\{z_i=1\}} \log p + \left(n - \sum_{i=1}^n I_{\{z_i=1\}}\right) \log(1-p) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n [I_{\{z_i=1\}}(x_i - \mu_1)^2 + (1 - I_{\{z_i=1\}})(x_i - \mu_2)^2]$$

Για το E - βήμα πρέπει να υπολογίσουμε το $Q(\theta|\theta^{(t-1)}, \underline{x}) = E_{\theta^{(t-1)}}(\log L^c(\theta|\underline{x}, \underline{Z}))$, όπου η μέση τιμή λαμβάνεται ως προς $\underline{Z} \sim k(\underline{z}|\underline{x}, \theta)$. Από την σχέση (1.7) έχουμε ότι τα Z_i είναι ανεξάρτητα με:

$$P(Z_i = 1|\underline{\theta}, \underline{x}) = \frac{p \exp\left\{-\frac{(x_i - \mu_1)^2}{2}\right\}}{p \exp\left\{-\frac{(x_i - \mu_1)^2}{2}\right\} + (1-p) \exp\left\{-\frac{(x_i - \mu_2)^2}{2}\right\}} = 1 - P(Z_i = 2|\underline{\theta}, \underline{x})$$

Στο βήμα t οι αναμενόμενες κατατάξεις ισούνται με:

$$z_i^{(t-1)} = E\left(\sum_{i=1}^n I_{\{z_i=1\}} \mid \underline{\theta}^{(t-1)}, \underline{x}\right) = P(Z_i = 1 \mid \underline{\theta}^{(t-1)}, \underline{x}).$$

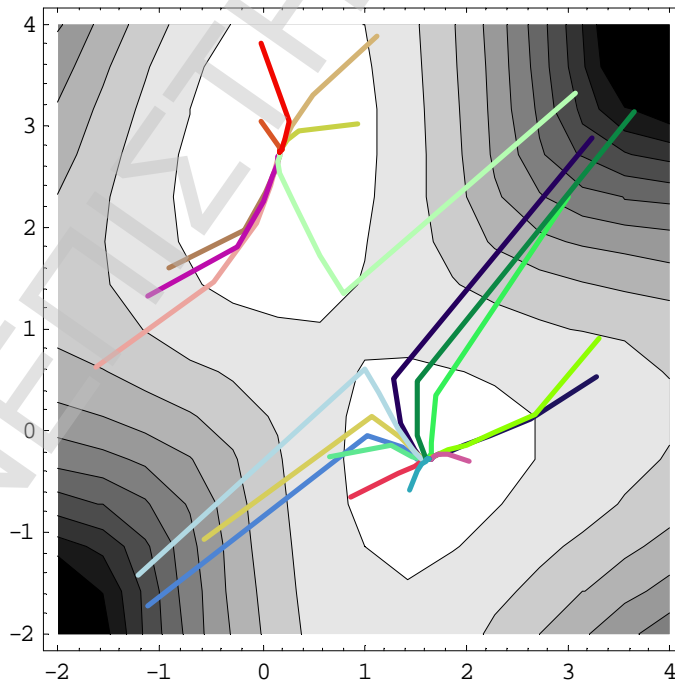
Συνεπώς,

$$Q(\theta \mid \theta^{(t-1)}, \underline{x}) = \sum_{i=1}^n z_i^{(t-1)} \log p + \left(n - \sum_{i=1}^n z_i^{(t-1)}\right) \log(1-p) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \left[z_i^{(t-1)} (x_i - \mu_1)^2 + (1 - z_i^{(t-1)}) (x_i - \mu_2)^2 \right]$$

το οποίο μεγιστοποιούμε στο M - βήμα ως προς (μ_1, μ_2) , και παίρνουμε

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

Τρέξαμε 20 φορές τον αλγόριθμο EM (με 100 επαναλήψεις την φορά) για το συγκεκριμένο παράδειγμα με τυχαία επιλογή των αρχικών τιμών. Όπως φαίνεται στο Σχήμα 5, ο αλγόριθμος πάντα συγκλίνει σε μία κορυφή της λογαριθμοπιθανοφάνειας, αλλά μόνο στις 8 από τις 20 φορές προσελκύστηκε από την υψηλότερη και κύρια κορυφή, ενώ στις υπόλοιπες 12 συνέκλινε στην ψευδοκορυφή (αν και η πιθανοφάνεια είναι πολύ μικρότερη). Αυτό συνέβη διότι οι αρχικές τιμές έτυχε να βρίσκονται πιο κοντά στην μικρότερη κορυφή.



ΣΧΗΜΑ 5. Τα ίχνη των 20 εκτελέσεων του αλγορίθμου EM στην επιφάνεια του λογαρίθμου πιθανοφάνειας της μείξης (1.8) για 500 προσομοιωμένες παρατηρήσεις και πραγματικές τιμές παραμέτρων $(\mu_1, \mu_2, p) = (0, 2.5, 0.7)$.

Έτσι στην περίπτωση που ο αλγόριθμος συγκλίνει στην κυρίαρχη κορυφή της πιθανοφάνειας παίρνουμε τις εκτιμήσεις: $\hat{\mu}_1 = -0.0160536$ και $\hat{\mu}_2 = 2.56785$ οι οποίες είναι αρκετά κοντά στις πραγματικές τιμές των παραμέτρων (0 και 2.5 αντίστοιχα). Στην άλλη περίπτωση οι εκτιμήσεις που παίρνουμε είναι: $\hat{\mu}_1 = 1.51108$ και $\hat{\mu}_2 = -0.523689$, οι οποίες είναι τελείως λάθος.

1.5 Το πρόβλημα *Label Switching*

1.5.1 Ορισμός

Εκτός από τις ιδιαιτερότητες των μοντέλων που παρουσιάσαμε μέχρι τώρα, μελετάμε ξεχωριστά την πιο σημαντική ίσως από αυτές. Ένα βασικό χαρακτηριστικό των μοντέλων μίξεων είναι ότι είναι αναλλοίωτα ως προς τις μεταθέσεις των δεικτών που ονοματίζουν τις συνιστώσες. Για κάθε μετάθεση $\nu = (\nu_1, \dots, \nu_k)$ του συνόλου των δεικτών $\{1, 2, \dots, k\}$ ως συμβολίσουμε την αντίστοιχη μετάθεση του παραμετρικού διανύσματος $(\underline{\theta}, \underline{p}) = (\theta_1, \dots, \theta_k, p_1, \dots, p_k)$ με:

$$\nu(\underline{\theta}, \underline{p}) = (\theta_{\nu_1}, \dots, \theta_{\nu_k}, p_{\nu_1}, \dots, p_{\nu_k}). \quad (1.16)$$

Παρατηρούμε ότι για κάθε μετάθεση $\nu(\underline{q}, \underline{p})$, η πιθανοφάνεια:

$$L(\underline{\theta}, \underline{p} | x) = \prod_{i=1}^n (p_1 f_1(x_i | \theta_1) + \dots + p_k f_k(x_i | \theta_k))$$

παραμένει αναλλοίωτη, λόγω της συμμετρίας που παρουσιάζει ως προς τις παραμέτρους. Το φαινόμενο αυτό καλείται *label switching* (σύμφωνα με τον όρο που χρησιμοποίησαν οι Redner & Walker 1984).

Σε μία μπεύζιανή ανάλυση, εάν δεν έχουμε εκ των προτέρων πληροφορία η οποία να διακρίνει τις συνιστώσες μεταξύ τους (δηλαδή στην περίπτωση που οι εκ των προτέρων κατανομές είναι ίδιες μεταξύ τους), τότε και η εκ των υστέρων κατανομή των παραμέτρων θα είναι συμμετρική. Έτσι, στην περίπτωση που έχουμε προσομοιώσει ένα δείγμα από την εκ των υστέρων κατανομή (με την χρήση μεθόδων MCMC όπως θα περιγράψουμε σε επόμενο κεφάλαιο) συνεπάγεται ότι οι παράμετροι (θ_i, p_i) δεν είναι διακριτές (ταυτοποιήσιμες) περιθωριακά. Αυτό σημαίνει απλά ότι δεν μπορούμε να ξεχωρίσουμε την συνιστώσα 1 και το (θ_1, p_1) από την συνιστώσα 2 και το (θ_2, p_2) , διότι η εκ των υστέρων μέση τιμή του (θ_1, p_1) θα ισούται με την εκ των υστέρων μέση τιμή του (θ_2, p_2) .

Για την αντιμετώπιση αυτού του προβλήματος έχουν προταθεί αρκετές μέθοδοι, που βασίζονται είτε σε αλγόριθμους αναδιάταξης (*relabelling algorithms*) του προσομοιωμένου δείγματος είτε στην εφαρμογή πλασματικών περιορισμών για την διάκριση των παραμέτρων. Υπάρχουν βέβαια και περιπτώσεις που χρησιμοποιούνται διαφορετικές εκ των προτέρων κατανομές και το παραπάνω πρόβλημα

παρακάμπτεται. Στην συνέχεια περιγράφουμε κάποιες από αυτές τις μεθόδους για την εφαρμογή τους στα επόμενα κεφάλαια².

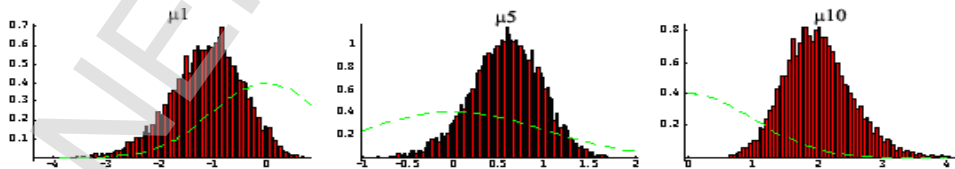
1.5.2 Περιορισμοί διάκρισης

Μία από τις πρώτες τακτικές (και η πιο απλή θα λέγαμε) που απαντώνται στην βιβλιογραφία για την αντιμετώπιση του προβλήματος αυτού, είναι να θέσουμε περιορισμούς διάκρισης (ή ταυτοποίησης) στον παραμετρικό χώρο. Για παράδειγμα, στην περίπτωση μείξης δύο κανονικών κατανομών (1.3) μπορούμε να ξεχωρίσουμε τις συνιστώσες σύμφωνα με την διάταξη των μέσων (ή των διασπορών, ή των βαρών). Από την μπεύζιανή σκοπιά αυτό σημαίνει τον μετασχηματισμό της αρχικής κανονικής εκ των προτέρων κατανομής $\pi(\underline{\theta}, \underline{p})$ στον περιορισμό της στο σύνολο $\{\underline{\mu} \in \mathcal{R}^k : \mu_1 \leq \dots \leq \mu_k\}$, δηλαδή:

$$\pi(\underline{\theta}, \underline{p}) I_{\{\mu_1 \leq \dots \leq \mu_k\}}. \quad (1.17)$$

Κάτι τέτοιο σημαίνει ότι το δείγμα που προσομοιώνουμε θα προκύπτει βάσει της (1.17). Ένα λογικό ερώτημα είναι: Δεν θα μπορούσαμε να χρησιμοποιήσουμε την $\pi(\underline{\theta}, \underline{p})$ και στην συνέχεια απλά να διατάσσουμε τις τιμές που προκύπτουν? Η απάντηση είναι ότι κάτι τέτοιο αποδίδει όταν τα μ_i (και συνεπώς οι αντίστοιχοι υποπληθυσμοί) απέχουν «αρκετά» μεταξύ τους. Σε αυτή την περίπτωση η διάκριση των συνιστωσών βάσει της διάταξης των μέσων της εκ των υστέρων κατανομής, θα αντικατοπτρίζει την πραγματική διάταξη των συνιστωσών του πληθυσμού. Στις περισσότερες περιπτώσεις κάτι τέτοιο δεν ισχύει αφού οι κατανομές των συνιστωσών «μπλέκονται» μεταξύ τους, με αποτέλεσμα την εμφάνιση του φαινομένου label switching.

Αν και η υιοθέτηση της (1.17) στο μοντέλο μας φαίνεται ασήμαντη (διότι όντως η κατανομή του δείγματος είναι ίδια με ή χωρίς την δείκτρια συνάρτηση), η εισαγωγή του παραπάνω περιορισμού έχει σοβαρές επιπτώσεις τόσο στην υπολογιστική διαδικασία όσο και στην συμπεραματολογία. Όταν περιορίζουμε τον παραμετρικό χώρο η πιθανοφάνεια επηρεάζεται άμεσα. Ακόμη, η εκ των υστέρων κατανομή εμφανίζει παραπάνω από μία κορυφές και οι εκ των υστέρων μέσοι μπορεί να βρίσκονται σε περιοχές χαμηλής εκ των υστέρων πιθανότητας.



ΣΧΗΜΑ 6. Κατανομές των $\mu_{(1)}, \mu_{(5)}$ και $\mu_{(10)}$, συγκρινόμενες με την εκ των προτέρων $N(0,1)$.

Συμβολίζουμε με $\mu_{(i)}$, την i διατεταγμένη τυχαία μεταβλητή. Στο Σχήμα 6 φαίνονται οι περιθωριακές εκ των προτέρων κατανομές των $\mu_{(1)}, \mu_{(5)}$ και $\mu_{(10)}$, όταν εκ των προτέρων $\mu_1, \dots, \mu_{10} \sim N(0,1)$. Η σύγκριση των αντίστοιχων κατανομών με την $N(0,1)$

² Παραπέμπουμε στα παραδείγματα των Παραγράφων 2.1 και 2.2.

δείχνει ξεκάθαρα τις επιπτώσεις της διάταξης. Έτσι οι εκτιμήσεις μας με βάση την εκ των υστέρων κατανομή μπορεί να είναι φτωχές, ειδικά όταν ο αριθμός των συνιστωσών είναι μεγάλος.

1.5.3 Αλγόριθμοι αναδιάταξης και προσεγγίσεις μέσω της Θεωρίας Αποφάσεων

Έστω ότι έχουμε n ανεξάρτητες και ισόνομες παρατηρήσεις $\underline{x} = (x_1, \dots, x_n)$ από το μοντέλο $\sum_{i=1}^k p_i f_i(x|\theta_i)$ όπου υπάρχουν k συνιστώσες, και ας υποθέσουμε ότι οι πυκνότητες $f(\cdot|\theta_i)$ είναι γνωστές εκτός από την παράμετρο θ_i . Στην Παράγραφο 1.5.2 είδαμε κάποια μειονεκτήματα της μεθόδου των περιορισμών διάταξης. Για αυτό θεωρούμε τώρα τον παραμετρικό χώρο ως έχει, πράγμα που σημαίνει ότι οι κορυφές της εκ των υστέρων κατανομής θα είναι ίσες με $k!$. Για να καταλήξουμε σε κατάλληλες εκτιμήσεις μπορούμε να χρησιμοποιήσουμε δύο μεθόδους: είτε μετά την ολοκλήρωση της διαδικασίας της προσομοίωσης να θέσουμε μία «λογική» συνθήκη αναδιάταξης του δείγματος είτε (για μία πιο αυστηρή ανάλυση) να χρησιμοποιήσουμε μία συνάρτηση ζημίας.

- **Αναδιάταξη μέσω του εσωτερικού γινομένου**

Για μία αναδιάταξη του $\{1, 2, \dots, k\}$ χρησιμοποιούμε τον συμβολισμό (1.16) για την αντίστοιχη διάταξη του παραμετρικού διανύσματος. Για να αντιμετωπίσουμε το πρόβλημα label switching εφαρμόζουμε την ακόλουθη διαδικασία στο προσομοιωμένο δείγμα μεγέθους M , $(\underline{\theta}, \underline{p})^{(1)}, \dots, (\underline{\theta}, \underline{p})^{(M)}$:

Αλγόριθμος αναδιάταξης μέσω του εσωτερικού γινομένου

1. Βρίσκουμε την Monte Carlo προσέγγιση του Maximum a Posteriori (MAP) εκτιμητή: $(\underline{\theta}, \underline{p})^{(i^*)}$ όπου

$$i^* = \arg \max_{i=1, \dots, M} \pi((\underline{\theta}, \underline{p})^{(i)} | \underline{x}).$$

2. Για $i = 1, 2, \dots, M$, βρίσκουμε όλες τις δυνατές ($k!$) μεταθέσεις $v(\underline{\theta}, \underline{p})^{(i)}$ του $(\underline{\theta}, \underline{p})^{(i)}$.

- a. Υπολογίζουμε το

$$v_m = \arg \max_v \left\langle v(\underline{\theta}, \underline{p})^{(i)}, v((\underline{\theta}, \underline{p})^{(i^*)}) \right\rangle_{2k}$$

όπου το $\langle \cdot, \cdot \rangle_l$ δηλώνει το εσωτερικό γινόμενο του διανυσματικού χώρου \mathcal{R}^l .

- b. Θέτουμε $(\underline{\theta}, \underline{p})^{(i)} = v_m((\underline{\theta}, \underline{p})^{(i)})$.

Το βήμα (2) επιλέγει εκείνη την αναδιάταξη $v_m((\underline{\theta}, \underline{p})^{(i)})$ του $(\underline{\theta}, \underline{p})^{(i)}$, η οποία είναι πιο κοντά στον MAP εκτιμητή, δηλαδή αυτή που μεγιστοποιεί το εσωτερικό της γινόμενο με το $(\underline{\theta}, \underline{p})^{(i)}$. Αυτή η μέθοδος είναι μία απλή «λογική» γεωμετρική προσέγγιση και λύνει το πρόβλημα του label switching χωρίς να επιβάλλει εκ των προτέρων κάποια διάταξη στις παραμέτρους του μοντέλου. Μετά την αναδιάταξη των προσομοιωμένων τιμών δεν έχουμε παρά να υπολογίσουμε την Monte Carlo εκτίμηση της εκ των υστέρων μέσης τιμής του θ_j , $E_x(\theta_j|x)$, η οποία ισούται με $\sum_{i=1}^M (v_m(\underline{\theta})^{(i)})_j / M$.

- **Αναδιάταξη μέσω μιας συνάρτησης ζημίας**

Διαδικασίες όπως η εκτίμηση παραμέτρων, η κατάταξη παρατηρήσεων σε σύνολα και ο καθορισμός των εκ των υστέρων κατανομών μπορούν να θεωρηθούν ως πρόβλημα επιλογής μιας απόφασης α από ένα σύνολο δυνατών αποφάσεων A . Στην θεωρία αποφάσεων ορίζεται μία συνάρτηση ζημίας $L: A \times \Theta \rightarrow \mathfrak{R}$, όπου το $L(\alpha; \theta)$ είναι η ζημία που προκύπτει όταν εκτιμήσουμε την ποσότητα θ με την τιμή α , ενώ στην συνέχεια επιλέγεται εκείνη η απόφαση $\hat{\alpha}$ η οποία ελαχιστοποιεί την εκ των υστέρων αναμενόμενη ζημία (ή κίνδυνο):

$$R(\alpha; \theta) = E(L(\alpha; \theta)|x)$$

Εφ' όσον στα προβλήματα που αφορούν σε μείξεις, όλες οι μεταθέσεις του $(\underline{\theta}, \underline{p})$ δίνουν την ίδια πιθανοφάνεια, είναι λογικό να ορίσουμε συναρτήσεις ζημίας οι οποίες είναι αναλλοίωτες ως προς τις μεταθέσεις του $(\underline{\theta}, \underline{p})$. Δίνουμε προσοχή λοιπόν σε συναρτήσεις ζημίας του τύπου

$$L(\alpha; (\underline{\theta}, \underline{p})) = \min_v L_0(\alpha; v(\underline{\theta}, \underline{p}))$$

για κάποιο $L_0: A \times \Theta \rightarrow \mathfrak{R}$.

Εάν τα $(\underline{\theta}, \underline{p})^{(1)}, \dots, (\underline{\theta}, \underline{p})^{(M)}$ είναι το προσομοιωμένο δείγμα από μία Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την εκ των υστέρων κατανομή του μοντέλου μας, τότε είναι λογικό να προσεγγίσουμε τον κίνδυνο Bayes $r(\alpha)$ από τον κίνδυνο Monte - Carlo:

$$\tilde{r}(\alpha) = \frac{1}{M} \sum_{t=1}^M \min_{v_t} L_0(\alpha; v_t((\underline{\theta}, \underline{p})^{(t)})) = \min_{v_1, \dots, v_M} \frac{1}{M} \sum_{t=1}^M L_0(\alpha; v_t((\underline{\theta}, \underline{p})^{(t)}))$$

και να επιλέξουμε το α που ελαχιστοποιεί το $\tilde{r}(\alpha)$. Κάθε επανάληψη του παρακάτω αλγορίθμου ελαττώνει το $\tilde{r}(\alpha)$:

Γενικός αλγόριθμος αναδιάταξης μέσω θεωρίας αποφάσεων

0. Θέσε v_i την ταυτοτική μετάθεση $i = 1, \dots, M$.

1. Προσδιόρισε εκείνο το $\alpha = \hat{\alpha}$ για το οποίο ελαχιστοποιείται το:

$$\sum_{t=1}^M L_0(\alpha; v_t, ((\underline{\theta}, \underline{p})^{(t)})).$$

2. Επίλεξε το v_t που ελαχιστοποιεί το

$$L_0(\alpha; v_t, ((\underline{\theta}, \underline{p})^{(t)}))$$

$\forall t = 1, 2, \dots, M$.

Ας υποθέσουμε τώρα ότι θέλουμε να χρησιμοποιήσουμε το μοντέλο μείξης για να κατατάξουμε τις παρατηρήσεις σε συνιστώσες και ταυτόχρονα να δώσουμε κάποια ένδειξη της αβεβαιότητας στην κατάταξη αυτή. Για αυτό τον σκοπό θεωρούμε έναν $n \times k$ πίνακα $\mathbf{Q} = (q_{ij})$, όπου το q_{ij} είναι η πιθανότητα η παρατήρηση i να ταξινομηθεί στην συνιστώσα j του πληθυσμού με $\sum_{j=1}^k q_{ij} = 1$. Ας συμβολίσουμε επίσης με $\mathbf{P}(\theta) = (\pi_{ij}(\underline{\theta}, \underline{p}))$ τον πίνακα των (πληθυσμιακών) πιθανοτήτων ταξινόμησης, όπου:

$$\pi_{ij}(\underline{\theta}, \underline{p}) = P(Z_i = j | \underline{x}, \underline{\theta}, \underline{p}) = \frac{p_j f_j(x_i | \theta_j)}{\sum_{i=1}^k p_i f_i(x | \theta_i)}$$

Μία κλασική μέθοδος μέτρησης της ζημίας για την εκτίμηση του \mathbf{P} με τον \mathbf{Q} , είναι η απόκλιση *Kullback - Leibler* του \mathbf{Q} από την πραγματική κατανομή της ταξινόμησης \mathbf{P} :

$$L_0(\mathbf{Q}; \mathbf{P}) = L_0(\mathbf{Q}; \theta) = \sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(\theta) \log \frac{\pi_{ij}(\theta)}{q_{ij}}$$

Για αυτή την επιλογή της συνάρτησης ζημίας, θα προσδιορίσουμε τα επιμέρους βήματα του προηγούμενου αλγορίθμου:

Κατ' αρχάς πρέπει να ελαχιστοποιήσουμε το $\sum_{t=1}^M L_0(\alpha; v_t, ((\underline{\theta}, \underline{p})^{(t)}))$, υπό τον περιορισμό

$\sum_{i=1}^n \sum_{j=1}^k \pi_{ij} = 1$. Χρησιμοποιώντας την μέθοδο Lagrange, έχουμε:

$$L = \sum_{t=1}^M L_0(\alpha; v_t((\underline{\theta}, \underline{p})^{(t)}))$$

$$= \sum_{t=1}^M \sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})) \log \frac{\pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)}))}{q_{ij}} + \lambda \left(\sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})) - 1 \right) \Rightarrow$$

$$\frac{\partial L}{\partial \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)}))} = -\frac{\pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)}))}{q_{ij}} + \lambda, \quad i = 1, \dots, n, j = 1, \dots, k$$

και θέτοντας $\frac{\partial L}{\partial \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)}))} = 0$ λαμβάνουμε ότι $\lambda = -\frac{\sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)}))}{q_{ij}}$. Όμως:

$$\lambda q_{ij} = -\sum_{t=1}^M \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})) \Rightarrow \lambda \sum_{i=1}^n \sum_{j=1}^k q_{ij} = -\sum_{t=1}^M \sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})) \Rightarrow \lambda = -M.$$

Ωστε:

$$q_{ij} = \frac{1}{M} \sum_{t=1}^M \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})).$$

Άρα ο αλγόριθμος παίρνει την εξής μορφή:

Αλγόριθμος αναδιάταξης με συνάρτηση ζημίας την απόκλιση Kullback - Leibler

0. Θέσε v_i την ταυτοτική μετάθεση $i = 1, \dots, M$.

1. Θέσε

$$q_{ij} = \frac{1}{M} \sum_{t=1}^M \pi_{ij}(v_t((\underline{\theta}, \underline{p})^{(t)})).$$

2. Επίλεξε το v_t που ελαχιστοποιεί το

$$L_0(\mathbf{Q}; \theta) = \sum_{i=1}^n \sum_{j=1}^k \pi_{ij}(\theta) \log \frac{\pi_{ij}(\theta)}{q_{ij}}$$

$$\forall t = 1, 2, \dots, M.$$

Σημειώνουμε ότι παρά την απλότητα σε επίπεδο προγραμματισμού αυτού του αλγορίθμου, ο χρόνος εκτέλεσής του για μία τυπική περίπτωση, π.χ. 20000 προσομοιωμένες τιμές, είναι αρκετά μεγάλος ακόμα και για γλώσσες προγραμματισμού όπως η FORTRAN. Πάντως για τέτοιου είδους προβλήματα υπάρχουν ισοδύναμοι αλγόριθμοι από την Θεωρία Βελτιστοποίησης οι οποίοι είναι αποδοτικοί καθώς συντομεύουν αρκετά την διαδικασία ελαχιστοποίησης (το

λεγόμενο 'Transportation Problem'). Προς αποφυγή αυτής της δυσκολίας, για την αντιμετώπιση του προβλήματος label switching, επιλέγουμε τον αλγόριθμο αναδιάταξης μέσω εσωτερικού γινομένου λόγω της απλότητας αλλά και της αποδοτικότητας αυτού.

1.6 Ημιπαραμετρική προσέγγιση

Μια εναλλακτική προσέγγιση στην ερμηνεία και εκτίμηση των μείξεων είναι ημιπαραμετρική. Σημειώνοντας ότι πολύ λίγα φαινόμενα υπακούουν στις συνήθεις κατανομές, η επιλογή του μοντέλου (1.2) μπορεί να αποτελεί και έναν συμβιβασμό μεταξύ της αντικειμενικής αναπαράστασης του προς μελέτη φαινομένου και της αποδοτικής εκτίμησης της άγνωστης κατανομής του. Εάν το k είναι μεγάλο, τότε η προσέγγιση (1.2) είναι συνήθως δικαιολογημένη.

Η πλέον συνηθισμένη προσέγγιση στην Μπεϋζιανή μη παραμετρική Στατιστική είναι η χρήση της διαδικασίας Dirichlet $\mathcal{D}(F_0, a)$ όπου F_0 είναι μία συνάρτηση κατανομής και a μία παράμετρος ακρίβειας. Αρχικά αναφέρουμε κάποιες γνωστές ιδιότητες της κατανομής Dirichlet.

1.6.1 Η κατανομή Dirichlet

Η κατανομή Dirichlet εμφανίζεται σε προβλήματα διατεταγμένων στατιστικών συναρτήσεων. Στην Μπεϋζιανή Στατιστική χρησιμοποιείται ως συζυγής εκ των προτέρων κατανομή για τις παραμέτρους της πολυωνυμικής κατανομής. Ο ακριβής ορισμός της είναι:

Ορισμός 1.6.1 Έστω Z_1, \dots, Z_k ανεξάρτητες τυχαίες μεταβλητές με $Z_i \sim \Gamma(\alpha_j, 1)$, όπου $\alpha_j \geq 0$ για κάθε j , $\alpha_j > 0$ για κάποια j , $j = 1, 2, \dots, k$ και $\Gamma(a, b)$ η κατανομή Γάμμα με μέση τιμή a/b . Η κατανομή Dirichlet με παράμετρο $(\alpha_1, \dots, \alpha_k)$, $D(\alpha_1, \dots, \alpha_k)$, ορίζεται ως η από κοινού κατανομή του (Y_1, \dots, Y_k) , όπου:

$$Y_j = \frac{Z_j}{\sum_{i=1}^k Z_i}, j = 1, \dots, k$$

Η κατανομή εκφυλίζεται ως προς το μέτρο Lebesgue στον k -διάστατο χώρο αφού $Y_1 + \dots + Y_k = 1$. Εάν κάποιο από τα $\alpha_j = 0$, το αντίστοιχο Y_j εκφυλίζεται στο μηδέν. Αν $\alpha_j > 0, \forall j = 1, \dots, k$, τότε η $(k-1)$ -διάστατη κατανομή του (Y_1, \dots, Y_{k-1}) είναι απολύτως συνεχής με πυκνότητα:

$$f(y_1, \dots, y_{k-1} | \alpha_1, \dots, \alpha_k) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \prod_{j=1}^{k-1} y_j^{\alpha_j - 1} \left(1 - \sum_{j=1}^{k-1} y_j \right)^{\alpha_k - 1} I_S(y_1, \dots, y_{k-1}) \quad (1.18)$$

$$\text{όπου } S = \left\{ (y_1, \dots, y_{k-1}) : y_j \geq 0, \sum_{j=1}^{k-1} y_j \leq 1 \right\}.$$

Για $k = 2$ η σχέση (1.18) δίνει την πυκνότητα της κατανομής βήτα $B(\alpha_1, \alpha_2)$. Οι κύριες ιδιότητες της κατανομής Dirichlet είναι:

I. Εάν $(Y_1, \dots, Y_k) \sim D(\alpha_1, \dots, \alpha_k)$ και r_1, \dots, r_l είναι ακέραιοι τέτοιοι ώστε $0 < r_1 < \dots < r_l = k$, τότε:

$$\left(\sum_{i=1}^{r_1} Y_i, \sum_{i=r_1+1}^{r_2} Y_i, \dots, \sum_{i=r_{l-1}+1}^{r_l} Y_i \right) \sim D \left(\sum_{i=1}^{r_1} \alpha_i, \sum_{i=r_1+1}^{r_2} \alpha_i, \dots, \sum_{i=r_{l-1}+1}^{r_l} \alpha_i \right)$$

Αυτό έπεται από τον ορισμό της κατανομής Dirichlet και την αναπαραγωγική ιδιότητα της κατανομής Γάμμα: αν $Z_1 \sim \Gamma(\alpha_1, 1)$ και $Z_2 \sim \Gamma(\alpha_2, 1)$ και αν Z_1, Z_2 ανεξάρτητες τότε $Z_1 + Z_2 \sim \Gamma(\alpha_1 + \alpha_2, 1)$. Ειδικότερα η περιθωριακή κατανομή κάθε

$$Y_j \text{ είναι η } B \left(\alpha_j, \sum_{i=1}^k \alpha_i - \alpha_j \right).$$

II. Οι δύο πρώτες ροπές των Y_i είναι:

$$\begin{aligned} EY_i &= \frac{\alpha_i}{\alpha} \\ EY_i^2 &= \frac{\alpha_i(\alpha_i + 1)}{\alpha(\alpha + 1)} \\ EY_i Y_j &= \frac{\alpha_i \alpha_j}{\alpha(\alpha + 1)}, \quad i \neq j \end{aligned}$$

$$\text{όπου } \alpha = \sum_{i=1}^k \alpha_i.$$

III. Εάν η εκ των προτέρων κατανομή του (Y_1, \dots, Y_k) είναι η $D(\alpha_1, \dots, \alpha_k)$ και αν:

$$P(X = j | Y_1, \dots, Y_k) = Y_j \text{ σχεδόν βεβαίως για } j = 1, \dots, k$$

τότε η εκ των υστέρων κατανομή του (Y_1, \dots, Y_k) δοθέντος ότι $X = j$ είναι $D(\alpha_1^{(j)}, \dots, \alpha_k^{(j)})$, όπου

$$\alpha_i^{(j)} = \begin{cases} \alpha_i & , i \neq j \\ \alpha_j + 1 & , i = j \end{cases}$$

Έστω $D(y_1, \dots, y_k | \alpha_1, \dots, \alpha_k)$ η συνάρτηση κατανομής της $D(\alpha_1, \dots, \alpha_k)$. Τότε η ισότητα

$$P(X = j, Y_1 \leq z_1, \dots, Y_k \leq z_k) = P(X = j)P(Y_1 \leq z_1, \dots, Y_k \leq z_k | X = j)$$

μπορεί να εκφραστεί συναρτήσει του $D(y_1, \dots, y_k | \alpha_1, \dots, \alpha_k)$ μέσω των ιδιοτήτων II και III ως εξής:

$$\int_0^{z_1} \dots \int_0^{z_k} y_j dD(y_1, \dots, y_k | \alpha_1, \dots, \alpha_k) = \frac{\alpha_j}{\alpha} D(z_1, \dots, z_k | \alpha_1^{(j)}, \dots, \alpha_k^{(j)}). \quad (1.19)$$

1.6.2 Η διαδικασία Dirichlet

Έστω Ξ ένα σύνολο και \mathcal{A} μία σ -άλγεβρα υποσυνόλων του Ξ . Περιγράφουμε παρακάτω τον ορισμό ενός τυχαίου μέτρου πιθανότητας P στον μετρήσιμο χώρο (Ξ, \mathcal{A}) , μέσω της από κοινού κατανομής των τυχαίων μεταβλητών $(P(A_1), \dots, P(A_m))$ για κάθε m και κάθε ακολουθία A_1, \dots, A_m μετρήσιμων υποσυνόλων $A_i \in \mathcal{A}$, $\forall i = 1, \dots, m$. Για τις λεπτομέρειες του ορισμού παραπέμπουμε στην εργασία του Ferguson (1973).

Όπως μπορεί να δειχθεί αρκεί να ορίσουμε το τυχαίο μέτρο P μέσω της από κοινού κατανομής του $(P(B_1), \dots, P(B_k))$ για κάθε k και για κάθε μετρήσιμη διαμέριση (B_1, \dots, B_k) του Ξ . Λέμε ότι το (B_1, \dots, B_k) αποτελεί μία μετρήσιμη διαμέριση του Ξ ανν:

$$B_i \in \mathcal{A} \quad \forall i, \quad B_i \cap B_j = \emptyset \quad \text{για } i \neq j, \quad \text{και} \quad \bigcup_{i=1}^k B_i = \Xi.$$

Εάν έχουμε κάποιο σύστημα κατανομών του $(P(B_1), \dots, P(B_k))$ για κάθε k και για κάθε μετρήσιμη διαμέριση (B_1, \dots, B_k) του Ξ , υπάρχει ένα «κριτήριο συνέπειας» που θα θέλαμε σίγουρα να ικανοποιείται:

Κριτήριο Ferguson Αν $(B'_1, \dots, B'_{k'})$ και (B_1, \dots, B_k) είναι μετρήσιμες διαμερίσεις του Ξ και αν το (B_1, \dots, B_k) προκύπτει από συνδυασμούς ενώσεων του $(B'_1, \dots, B'_{k'})$ με:

$$B_1 = \bigcup_{i=1}^{r_1} B'_i, \quad B_2 = \bigcup_{i=r_1+1}^{r_2} B'_i, \quad \dots, \quad B_k = \bigcup_{i=r_{k-1}+1}^{r_k} B'_i, \quad \text{τότε η κατανομή του:}$$

$$\left(\sum_{i=1}^{r_1} B'_i, \sum_{i=r_1+1}^{r_2} B'_i, \dots, \sum_{i=r_{k-1}+1}^{r_k} B'_i \right)$$

όπως αυτή ορίζεται από την κοινού κατανομή του $(P(B'_1), \dots, P(B'_{k'}))$ είναι ίδια με την κατανομή του $(P(B_1), \dots, P(B_k))$.

Σημειώνουμε ότι το παραπάνω κριτήριο αποτελεί ικανή συνθήκη για να ισχύουν οι σύνθηκες συνέπειας του Kolmogorov.

Ορισμός 1.6.2 Θα λέμε ότι το P είναι ένα τυχαίο μέτρο πιθανότητας στον (Ξ, \mathcal{A})

(α) εάν ικανοποιείται το κριτήριο Ferguson, (β) εάν το $P(A)$, όπου $A \in \mathcal{A}$, παίρνει τιμές μόνο στο $[0,1]$ και (γ) εάν $P(\Xi)=1$.

Ας συμβολίσουμε με $[0,1]_{\mathcal{A}}$ τον χώρο όλων των συναρτήσεων από το \mathcal{A} στο $[0,1]$ και με $F_{\mathcal{A}}$ την σ -άλγεβρα που γεννιέται από τον χώρο των κυλινδρικών συνόλων. Όπως αποδεικνύεται υπάρχει ένα μέτρο πιθανότητας P στον $([0,1]_{\mathcal{A}}, F_{\mathcal{A}})$ το οποίο ορίζει την από κοινού κατανομή του $(P(A_1), \dots, P(A_m))$.

Ορισμός 1.6.3 Έστω ότι α είναι ένα πεπερασμένο μέτρο στον χώρο $(\mathcal{X}, \mathcal{A})$. Λέμε ότι η P είναι μία διαδικασία Dirichlet στον $(\mathcal{X}, \mathcal{A})$ με παράμετρο α , εάν για κάθε $k=1,2,\dots$ και κάθε μετρήσιμη διαμέριση (B_1, \dots, B_k) του \mathcal{X} , η κατανομή του $(P(B_1), \dots, P(B_k))$ είναι Dirichlet, $D(\alpha(B_1), \dots, \alpha(B_k))$.

Παρατηρούμε ότι το κριτήριο Ferguson είναι ακριβώς η ιδιότητα I της κατανομής Dirichlet. Σημειώνουμε ότι ένας εναλλακτικός ορισμός της διαδικασίας Dirichlet επιλέγει αντί για το μέτρο α , μία κατανομή F_0 και στην συνέχεια χρησιμοποιεί έναν συντελεστή «ακρίβειας» α και την συμβολίζει με $\mathcal{D}(F_0, \alpha)$.

Υπάρχουν αρκετές ιδιότητες που δείχνουν την κοντινή σχέση μεταξύ του τυχαίου μέτρου P και της πιθανότητας α . Εμείς θα περιοριστούμε στην περιγραφή μιας από αυτές, την οποία θα χρειαστούμε και στην συνέχεια. Έστω ότι P είναι διαδικασία Dirichlet στον $(\mathcal{X}, \mathcal{A})$ με παράμετρο α και έστω $A \in \mathcal{A}$. Τότε εάν $\alpha(A) = 0$ έπεται ότι και $P(A) = 0$ με πιθανότητα 1. Επιπλέον εάν $\alpha(A) > 0$ έπεται ότι και $P(A) > 0$ με πιθανότητα 1. Τέλος έχουμε ότι:

$$E(P(A)) = \alpha(A) / \alpha(\mathcal{X}). \quad (1.20)$$

Αυτό ισχύει διότι αν θεωρήσουμε την διαμέριση (A, A^c) , έχουμε ότι το $P(A)$ ακολουθεί την κατανομή $B(\alpha(A), \alpha(A^c))$ (αυτό έπεται από τον Ορισμό 1.6.3 και την παρατήρηση μετά τον Ορισμό 1.6.1). Ωστε:

$$E(P(A)) = \frac{\alpha(A)}{\alpha(A) + \alpha(A^c)} = \frac{\alpha(A)}{\alpha(A \cup A^c)} \Rightarrow E(P(A)) = \alpha(A) / \alpha(\mathcal{X})$$

Τέλος σημειώνουμε ότι το στήριγμα της διαδικασίας Dirichlet περιέχει το σύνολο όλων των μέτρων πιθανότητας τα οποία είναι απολύτως συνεχή ως προς το α .

Ορισμός 1.6.4 Έστω ότι το P είναι ένα τυχαίο μέτρο πιθανότητας στον $(\mathcal{X}, \mathcal{A})$. Λέμε ότι το X_1, \dots, X_n είναι ένα δείγμα μεγέθους n από το P , εάν για κάθε $m \in \mathbb{N}$ και για οποιαδήποτε μετρήσιμα σύνολα A_1, \dots, A_m και C_1, \dots, C_n :

$$P\{X_1 \in C_1, \dots, X_n \in C_n | P(A_1), \dots, P(A_m), P(C_1), \dots, P(C_n)\} = \prod_{j=1}^n P(C_j) \quad (1.21)$$

σχεδόν βεβαίως.

Γενικά αυτό σημαίνει ότι το X_1, \dots, X_n είναι δείγμα μεγέθους n από το P , εάν δοθέντων των $P(C_1), \dots, P(C_n)$, τα γεγονότα $\{X_1 \in C_1\}, \dots, \{X_n \in C_n\}$ είναι ανεξάρτητα μεταξύ τους, αλλά και από την υπόλοιπη διαδικασία, με

$$P\{X_j \in C_j | P(C_1), \dots, P(C_n)\} = P(C_j) \text{ σχεδόν βεβαίως, } \forall j = 1, \dots, m.$$

Πρόταση 1.6.1 Έστω ότι η P είναι μία διαδικασία Dirichlet στον $(\mathcal{E}, \mathcal{A})$ με παράμετρο α και έστω ότι το X είναι δείγμα μεγέθους 1 από την P . Τότε για $A \in \mathcal{A}$ ισχύει

$$P(X \in A | P(A)) = \alpha(A) / \alpha(\mathcal{E}). \quad (1.22)$$

Απόδειξη

Από την σχέση (1.21) έχουμε ότι $\Pi(X \in A | P(A)) = P(A)$ σχεδόν βεβαίως. Οπότε:

$$\begin{aligned} P(X \in A | P(A)) &= E P(X \in A | P(A)) \\ &= E(P(A)) \\ &= \alpha(A) / \alpha(\mathcal{E}). \end{aligned}$$

Έστω ότι $x \in \mathcal{E}$ και με d_x συμβολίζουμε το μέτρο Dirac στο x ορισμένο στον $(\mathcal{E}, \mathcal{A})$, δηλαδή αυτό το οποίο δίνει πυκνότητα 1 στο σημείο x :

$$\delta_x(A) = \begin{cases} 1, & \text{αν } x \in A \\ 0, & \text{αν } x \notin A \end{cases}$$

Τότε ισχύει το κάτωθι θεώρημα:

Θεώρημα 1.6.1 Έστω ότι η P είναι μία διαδικασία Dirichlet στον $(\mathcal{E}, \mathcal{A})$ με παράμετρο α και έστω ότι το X_1, \dots, X_n είναι δείγμα μεγέθους n από αυτή. Τότε η δεσμευμένη κατανομή της P δοθέντος (X_1, \dots, X_n) είναι η διαδικασία Dirichlet με παράμετρο $\alpha + \sum_{i=1}^n \delta_{X_i}$.

Για την απόδειξη του παραπάνω θεωρήματος παραπέμπουμε στην εργασία του Ferguson (1973).

Ολοκληρώνουμε το κεφάλαιο αυτό με κάποιες εφαρμογές της διαδικασίας Dirichlet.

Εφαρμογή 1.6.1 Εκτίμηση της συνάρτησης κατανομής.

Έστω $(\mathcal{E}, \mathcal{A}) = (\mathcal{R}, \mathcal{B})$ και έστω ότι ο χώρος αποφάσεων είναι ο χώρος όλων των συναρτήσεων κατανομών στον \mathcal{R} . Θεωρούμε ότι η συνάρτηση ζημίας είναι η

$$L(P, \hat{F}) = \int (F(t) - \hat{F}(t))^2 dW(t)$$

όπου W είναι ένα δοσμένο πεπερασμένο μέτρο στον $(\mathcal{R}, \mathcal{B})$ (μία συνάρτηση βάρους), και ότι:

$$F(t) = P((-\infty, t]).$$

Εάν $P \sim \mathcal{D}(\alpha)$, τότε $F(t) \sim B(\alpha(-\infty, t], \alpha(t, +\infty))$ για κάθε t . Ο κίνδυνος Bayes για την περίπτωση που δεν έχουμε δείγμα, είναι:

$$E(L(P, \hat{F})) = \int E(F(t) - \hat{F}(t))^2 dW(t)$$

ο οποίος ελαχιστοποιείται επιλέγοντας το $\hat{F}(t)$ να ελαχιστοποιεί το $E(F(t) - \hat{F}(t))^2 \forall t$. Αυτό επιτυγχάνεται επιλέγοντας $\hat{F}(t) = E(F(t))$. Έτσι, λόγω της (1.6.20), η απόφαση Bayes είναι

$$\hat{F}(t) = EF(t) = E(P((-\infty, t])) = \frac{\alpha((-\infty, t])}{\alpha(\mathcal{R})} = F_0(t)$$

η οποία εκφράζει την εκ των προτέρων άποψή μας για την μορφή της άγνωστης $F(t)$.

Στην περίπτωση που διαθέτουμε δείγμα μεγέθους n , σύμφωνα με το Θεώρημα 1.6.1, ο κανόνας Bayes είναι ο

$$\begin{aligned} \hat{F}_n(t|X_1, \dots, X_n) &= \frac{\alpha((-\infty, t]) + \sum_{i=1}^n \delta_{X_i}(-\infty, t]}{\alpha(\mathcal{R}) + n} \\ &= p_n F_0(t) + (1 - p_n) F_n(t|X_1, \dots, X_n) \end{aligned}$$

όπου $p_n = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n}$ και $F_n(t|X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(-\infty, t]$ η εμπειρική συνάρτηση κατανομής. Παρατηρούμε ότι η εκτίμηση μας είναι μία μείξη της εκ των προτέρων άποψης μας και της εμπειρικής συνάρτησης κατανομής.

Πόρισμα Από την εφαρμογή 1.6.1 έχουμε ότι:

$$\begin{aligned} \hat{F}_n(t|X_1, \dots, X_n) &= \frac{\alpha((-\infty, t]) + \sum_{i=1}^n \delta_{X_i}(-\infty, t]}{\alpha(\mathcal{R}) + n} \\ &= \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n} \frac{\alpha((-\infty, t])}{\alpha(\mathcal{R})} + \frac{1}{\alpha(\mathcal{R}) + n} \sum_{i=1}^n I_{X_i \leq t} \end{aligned}$$

Οπότε η περιθωριακή δεσμευμένη κατανομή του X_1 δοθέντος του (X_2, \dots, X_n) είναι:

$$\hat{F}_{n-1}(X_1|X_2, \dots, X_n) = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n - 1} F_0(X_1) + \frac{1}{\alpha(\mathcal{R}) + n - 1} \sum_{i=2}^n I_{X_i \leq X_1} \quad (1.23)$$

Αναφέρουμε ότι εκφράσεις της μορφής (1.23) κατέχουν κεντρικό ρόλο στις μεθόδους προσομοίωσης της απαραμετρικής Μπεύζιανής Στατιστικής για συμπερασματολογία σε μοντέλα μείξεων, αλλά εμείς δεν θα επεκταθούμε σε αυτές. Για μία περιγραφή των μεθόδων αυτών παραπέμπουμε στην εργασία των Escobar & West (1995).

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΚΕΦΑΛΑΙΟ 2

ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΜΕΙΞΕΩΝ ΜΕ ΓΝΩΣΤΟ ΠΛΗΘΟΣ ΣΥΝΙΣΤΩΣΩΝ

Τα πάντα πρέπει να απλοποιούνται
όσο γίνεται πιο πολύ.
Όχι όμως περισσότερο.
A. Einstein

2. Συμπερασματολογία για μοντέλα μείξεων με γνωστό πλήθος συνιστωσών

Τα μοντέλα μείξεων κατανομών έχουν υπάρξει η αφορμή για την ανάπτυξη πολλών μεθοδολογιών στην Υπολογιστική Στατιστική. Εκτός από τον αλγόριθμο EM, μπορούμε να ξεχωρίσουμε την μέθοδο *διεύρυνσης των δεδομένων* (data augmentation) που προτάθηκε από τους Tanner & Wong (1987), και η οποία αποτελεί πρόδρομο του *δειγματολήπτη Gibbs* των Gelfand & Smith (1990). Αυτό το κεφάλαιο μελετά μοντέλα όπου ο αριθμός των συνιστωσών της μείξης είναι εκ των προτέρων γνωστός, ακολουθώντας μεθόδους Monte Carlo και - κυρίως - MCMC (Markov Chain Monte Carlo).

2.1 Διεύρυνση δεδομένων και ο δειγματολήπτης Gibbs

Ο δειγματολήπτης Gibbs είναι μία από τις πιο κλασσικές μεθόδους MCMC και η χρήση του είναι ιδιαίτερα διαδεδομένη και στην Μπεϋζιανή εκτίμηση μείξεων [Diebolt & Robert (1990a), (1994), Lavine & West (1992), Verdinelli & Wasserman (1992), Chib (1995), Escobar & West (1995)]. Η μέθοδος αυτή εκμεταλλεύεται τα ελλιπή δεδομένα (Παράγραφος 1.2), δηλαδή αντιστοιχίζει σε κάθε παρατήρηση x_j μία μη παρατηρηθείσα μεταβλητή από την πολυωνυμική κατανομή, $z_j \sim M_k(1; p_1, \dots, p_k)$, έτσι ώστε

$$x_j | z_j = i \sim f(x | q_i),$$

και στην συνέχεια προσομοιώνει κάθε παράμετρο από τις πλήρεις δεσμευμένες κατανομές, δοθεισών όλων των υπολοίπων.

Υπενθυμίζουμε ότι σε ετερογενείς πληθυσμούς αποτελούμενους από ομοιογενή στρώματα, είναι διαισθητικά λογικό να ερμηνεύουμε το z_j σαν τον δείκτη του πληθυσμού απ' όπου προέρχεται το x_j και ο οποίος έχει χαθεί στην διαδικασία παρατήρησης. Παρ' όλα αυτά, η διάκριση μεταξύ πραγματικής και φανταστικής συμπλήρωσης των δεδομένων, δεν υφίσταται στον δειγματολήπτη, ο οποίος απλά στοχεύει στο να παραγάγει μία Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την εκ των υστέρων κατανομή των παραμέτρων του μοντέλου μας.

Το πλαίσιο που θα κινηθούμε είναι το εξής: Έστω το μοντέλο (1.2) $\sum_{i=1}^k p_i f_i(x | q_i)$, όπου εδώ το $k \in N$ θεωρείται μία γνωστή σταθερά. Έστω επίσης ότι οι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_i, i=1, \dots, n$ ακολουθούν το μοντέλο (1.2) και $\underline{x} = (x_1, \dots, x_n)$ το τυχαίο δείγμα που παρατηρήσαμε. Η διεύρυνση των δεδομένων με την εισαγωγή του $\underline{Z} = (Z_1, \dots, Z_n)$ στο μοντέλο, συνεπάγεται ότι $X_j | \underline{\theta}, Z_j = i \sim f(x | \theta_i)$, αλλά και ότι:

$$(\underline{X}, \underline{Z}) | \underline{p}, \underline{\theta} \sim \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}).$$

Επειδή στις περισσότερες εφαρμογές τα βάρη είναι και αυτά άγνωστα, θεωρούμε την συζυγή εκ των προτέρων κατανομή $\underline{p} \sim D(\underline{g}_1, \dots, \underline{g}_{k-1}; \underline{g}_k)$. Αν υποθέσουμε ότι η πυκνότητα f ανήκει σε μία εκθετική οικογένεια έχουμε ότι

$$f(x|\theta) = h(x) \exp\{\langle r(\theta), t(x) \rangle_k - \varphi(\theta)\} \quad (2.1)$$

όπου $h: \mathfrak{X} \rightarrow \mathfrak{X}^+$, $r: \Theta \rightarrow \mathfrak{R}$, $t: \Theta \rightarrow \mathfrak{X}$ και $\langle \cdot, \cdot \rangle_k$ συμβολίζει το εσωτερικό γινόμενο του χώρου \mathfrak{R}^k . Σε αυτή την περίπτωση μία συζυγής κατανομή για το θ δίνεται από την σχέση:

$$\pi(\theta) \propto \exp\{\langle r(\theta), \alpha \rangle_k - \beta \varphi(\theta)\} \quad (2.2)$$

όπου $\alpha \in \mathfrak{R}^k$ και $\beta > 0$ είναι σταθερές είτε υπερπαραμέτροι με εκ των προτέρων κατανομή $\pi(\alpha, \beta)$ (Robert, 2001).

Σε αυτή την περίπτωση έχουμε ότι η εκ των υστέρων κατανομή είναι

$$\begin{aligned} \underline{p}, \underline{\theta}, \underline{z}, \underline{\beta} | \underline{x} &\sim \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}) \prod_{i=1}^k \pi(\theta_j | \alpha_j, \beta_j) \prod_{i=1}^k \pi(\alpha_j, \beta_j) \prod_{i=1}^k p_j^{\gamma_j - 1} \Rightarrow \\ \pi(\underline{p}, \underline{\theta}, \underline{z}, \underline{\beta} | \underline{x}) &\propto \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}) \prod_{i=1}^k \exp\{\langle r(\theta_j), \alpha_j \rangle_k - \beta_j \varphi(\theta_j)\} \prod_{i=1}^k \pi(\alpha_j, \beta_j) \prod_{i=1}^k p_j^{\gamma_j - 1} \end{aligned} \quad (2.3)$$

Ο δειγματολήπτης Gibbs προσομοιώνει σε κάθε επανάληψη τις παραμέτρους και υπερπαραμέτρους από τις αντίστοιχες πλήρεις δεσμευμένες κατανομές της (2.3), ώστε να παραγάγει μία Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την (2.3). Για τα βάρη έχουμε ότι

$$\begin{aligned} \pi(\underline{p} | \mathbf{K}) &\propto \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}) \prod_{i=1}^k \exp\{\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i)\} \prod_{i=1}^k \pi(\alpha_i, \beta_i) \prod_{i=1}^k p_i^{\gamma_i - 1} \\ &\propto \left(p_1^{\gamma_1 - 1} \prod_{z_i=1} p_{z_i} \right) \left(p_2^{\gamma_2 - 1} \prod_{z_i=2} p_{z_i} \right) \mathbf{K} \left(p_{k-1}^{\gamma_{k-1} - 1} \prod_{z_i=k-1} p_{z_i} \right) \left(p_k^{\gamma_k - 1} \prod_{z_i=k} p_{z_i} \right) \Rightarrow \\ \pi(\underline{p} | \mathbf{K}) &\propto p_1^{\gamma_1 - 1 + n_1} p_2^{\gamma_2 - 1 + n_2} \mathbf{K} p_{k-1}^{\gamma_{k-1} - 1 + n_{k-1}} (1 - p_1 - p_{k-1})^{\gamma_k - 1 + n_k} \Rightarrow \\ \underline{p} | \mathbf{K} &\sim D(\gamma_1 + n_1, \dots, \gamma_k + n_k) \end{aligned} \quad (2.4)$$

όπου με $\pi(p_j | \mathbf{K})$ συμβολίζουμε την πλήρη δεσμευμένη εκ των υστέρων κατανομή του p_j δοθεισών όλων των υπολοίπων παραμέτρων, ενώ $n_j = \sum_{i=1}^n I_{\{z_i=j\}}$. Για τις παραμέτρους $\underline{\theta}$ έχουμε ότι:

$$\begin{aligned} \pi(\underline{\theta} | \mathbf{K}) &\propto \prod_{i=1}^n p_{z_i} f(x_i | \theta_{z_i}) \prod_{i=1}^k \exp\{\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i)\} \prod_{i=1}^k \pi(\alpha_i, \beta_i) \prod_{i=1}^k p_i^{\gamma_i - 1} \\ &\propto \prod_{i=1}^n p_{z_i} h(x_i) \exp\{\langle r(\theta_{z_i}), t(x_{z_i}) \rangle_k - \varphi(\theta_{z_i})\} \prod_{i=1}^k \exp\{\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i)\} \\ &\propto \prod_{i=1}^k \exp\{\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i)\} \prod_{i=1}^k \prod_{z_j=i} \exp\{\langle r(\theta_{z_j}), t(x_{z_j}) \rangle_k - \varphi(\theta_{z_j})\} \\ &\propto \prod_{i=1}^k \exp\{\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i)\} \prod_{i=1}^k \exp\left\{\left\langle r(\theta_i), \sum_{j=1}^n I_{\{z_j=i\}} t(x_j) \right\rangle_k - \sum_{j=1}^n I_{\{z_j=i\}} \varphi(\theta_i) \right\} \end{aligned}$$

$$\begin{aligned} &\propto \prod_{i=1}^k \exp \left(\langle r(\theta_i), \alpha_i \rangle_k - \beta_i \varphi(\theta_i) + \left\langle r(\theta_i), \sum_{j=1}^n I_{\{z_j=i\}} t(x_j) \right\rangle_k - \sum_{j=1}^n I_{z_j=i} \varphi(\theta_i) \right) \Rightarrow \\ \underline{\theta} | \mathbf{K} &\sim \prod_{i=1}^k \exp \left(\left\langle r(\theta_i), \alpha_i + \sum_{j=1}^n I_{\{z_j=i\}} t(x_j) \right\rangle_k - (\beta_i + n_i) \varphi(\theta_i) \right) \end{aligned} \quad (2.5)$$

Τέλος από την σχέση (1.7) έχουμε ότι:

$$P(Z_i | \mathbf{K}) \propto p_j f(x_i | \theta_j). \quad (2.6)$$

Λόγω των (2.4), (2.5) και (2.6), τα βήματα του γενικού δειγματολήπτη Gibbs είναι:

Δειγματολήπτης Gibbs για μείξεις εκθετικών οικογενειών

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}^{(0)}, \underline{p}^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots$:

1.1. Προσομοίωσε τα $z_i^{(t)}$ ($i = 1, \dots, n$) από την

$$P(z_i^{(t)} = j | p_j^{(t-1)}, \theta_j^{(t-1)}, x_i) \propto p_j^{(t-1)} f(x_i | \theta_j^{(t-1)}).$$

1.2. Υπολόγισε τα $n_j^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}}$ και $s_j^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}} t(x_i)$.

1.3. Προσομοίωσε τα $\underline{p}^{(t)}$ από την $D(\gamma_1 + n_1, \dots, \gamma_k + n_k)$.

1.4. Προσομοίωσε τα $\theta_j^{(t)}$ ($j = 1, \dots, k$) από την

$$\pi(\theta_j | \underline{z}^{(t)}, \underline{x}) \propto \exp \left(\langle r(\theta_j), \alpha_j + s_j^{(t)} \rangle_k - (\beta_j + n_j) \varphi(\theta_j) \right).$$

Όπως με όλες τις μεθόδους Monte Carlo, πρέπει να εξεταστεί προσεκτικά η αποδοτικότητα των αλγορίθμων MCMC. Εδώ η αποδοτικότητα έχει να κάνει με διάφορα θέματα όπως η αυτοσυσχέτιση των προσομοιωμένων αλυσίδων και το σφάλμα Monte Carlo. Προφανώς η σύγκλιση του αλγορίθμου εξαρτάται και από την επιλογή των εκ των προτέρων κατανομών. Για την αντιμετώπιση τέτοιων ζητημάτων παραπέμπουμε στους Mengersen et al. (1999) και Robert & Casella (2004, κεφάλαιο 12). Στην συνέχεια παραθέτουμε την εφαρμογή του παραπάνω αλγορίθμου σε κάποια παραδείγματα.

Παράδειγμα 2.1.1 Ως μία απλή εφαρμογή, ας θεωρήσουμε την απλή περίπτωση μείξης δύο κανονικών κατανομών (μοντέλο (1.8)):

$$pN(\mu_1, 1) + (1 - p)N(\mu_2, 1)$$

όπου τα βάρη είναι γνωστά. Σε αυτή την περίπτωση ο παραμετρικός χώρος είναι διδιάστατος και έτσι μπορούμε να σχεδιάσουμε το γράφημα της εκ των υστέρων

κατανομής. Θεωρούμε ότι εκ των προτέρων οι μέσοι είναι ανεξάρτητοι και ακολουθούν την κατανομή $N(\delta, 1/\lambda)$, όπου τα δ και λ είναι γνωστές υπερπαραμέτροι. Τότε:

$$\begin{aligned} \mu_1, \mu_2 | \underline{z}, \underline{x} &\propto \prod_{i=1}^n \exp\left(-\frac{(x_i - \mu_{z_i})^2}{2}\right) \exp\left(-\frac{\lambda(\mu_1 - \delta)^2}{2} - \frac{\lambda(\mu_2 - \delta)^2}{2}\right) \\ &\propto \exp\left(-\frac{1}{2} \sum_{i=1}^n I_{\{z_i=1\}} (x_i - \mu_1)^2 - \frac{\lambda(\mu_1 - \delta)^2}{2}\right) \\ &\quad \times \exp\left(-\frac{1}{2} \sum_{i=1}^n I_{\{z_i=2\}} (x_i - \mu_2)^2 - \frac{\lambda(\mu_2 - \delta)^2}{2}\right). \end{aligned} \quad (2.7)$$

Από την (2.7) παρατηρούμε ότι οι μέσοι είναι εκ των υστέρων επίσης ανεξάρτητοι. Για τον πρώτο έχουμε:

$$\begin{aligned} \mu_1 | \underline{z}, \underline{x} &\propto \exp\left(-\frac{1}{2} \sum_{i=1}^n I_{\{z_i=1\}} x_i^2 - \frac{1}{2} \mu_1^2 \sum_{i=1}^n I_{\{z_i=1\}} + \mu_1 \sum_{i=1}^n I_{\{z_i=1\}} x_i - \frac{\mu_1^2 \lambda}{2} + \mu_1 \lambda \delta\right) \\ &\propto \exp\left(-\frac{1}{2} \left\{ \mu_1^2 \left(\sum_{i=1}^n I_{\{z_i=1\}} + \lambda \right) - 2\mu_1 \left(\sum_{i=1}^n I_{\{z_i=1\}} x_i + \lambda \delta \right) \right\}\right) \\ &= \exp\left(-\frac{\left(\sum_{i=1}^n I_{\{z_i=1\}} + \lambda \right) \left\{ \mu_1^2 - 2\mu_1 \frac{\sum_{i=1}^n I_{\{z_i=1\}} x_i + \lambda \delta}{\sum_{i=1}^n I_{\{z_i=1\}} + \lambda} \right\}}{2}\right) \\ &\propto \exp\left(-\frac{\lambda + n_1}{2} \left(\mu_1 - \frac{s_1^x + \lambda \delta}{n_1 + \lambda} \right)^2\right) \end{aligned}$$

όπου $s_1^x = \sum_{i=1}^n I_{\{z_i=1\}} x_i$. Με ανάλογο τρόπο δουλεύουμε και για το μ_2 . Οπότε οι πλήρεις δεσμευμένες κατανομές των μ_1 και μ_2 είναι:

$$\mu_1 | \underline{z}, \underline{x} \sim N\left(\frac{\lambda \delta + s_1^x}{\lambda + n_1}, \frac{1}{\lambda + n_1}\right) \text{ και } \mu_2 | \underline{z}, \underline{x} \sim N\left(\frac{\lambda \delta + s_2^x}{\lambda + n_2}, \frac{1}{\lambda + n_2}\right).$$

Επίσης η εκ των υστέρων δεσμευμένη κατανομή του z_i δοθέντος του (μ_1, μ_2) , προκύπτει από τον τύπο (2.6):

$$\begin{aligned} P(z_i = 1 | \mu_1, \mu_2, x_i) &= \frac{p \exp\left\{-\frac{1}{2}(x_i - \mu_1)^2\right\}}{p \exp\left\{-\frac{1}{2}(x_i - \mu_1)^2\right\} + (1-p) \exp\left\{-\frac{1}{2}(x_i - \mu_2)^2\right\}} \\ &= 1 - P(z_i = 2 | \mu_1, \mu_2, x_i) \end{aligned}$$

δηλαδή τα z_i είναι ανεξάρτητες τυχαίες μεταβλητές και ακολουθούν την κατανομή Bernoulli στο $\{1,2\}$ με πιθανότητα επιτυχίας που δίνεται από την τελευταία έκφραση. Οπότε ο αλγόριθμος για την εφαρμογή του δειγματολήπτη Gibbs σε αυτή την περίπτωση έχει ως εξής:

Δειγματολήπτης Gibbs για το μοντέλο (1.8)

0. Δώσε κάποιες αρχικές τιμές στο $\mu_1^{(0)}, \mu_2^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots$:

1.1. Προσομοίωσε τα $z_i^{(t)}$ ($i = 1, \dots, n$) από την

$$P(z_i^{(t)} = 1) = 1 - P(z_i^{(t)} = 2) \propto p \exp\left(-0.5(x_i - \mu_1^{(t-1)})^2\right).$$

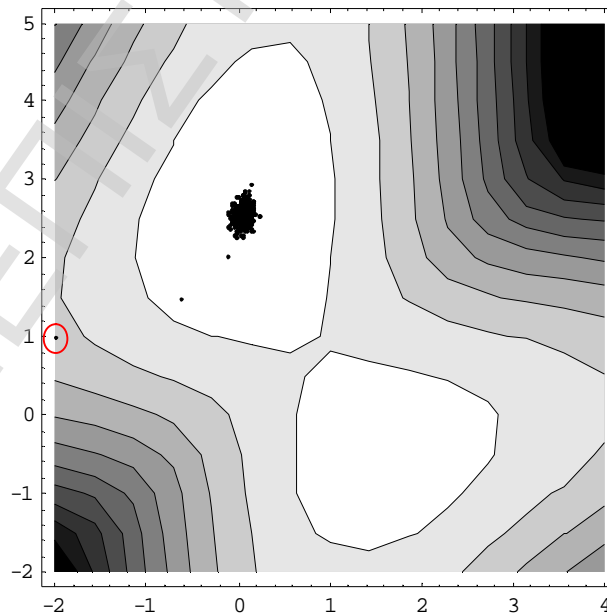
1.2. Υπολόγισε τα $n_j^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}}$ και $(s_j^x)^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}} x_i$.

1.3. Προσομοίωσε τα $\underline{p}^{(t)}$ από την $D(\gamma_1 + n_1, \dots, \gamma_k + n_k)$.

1.4. Προσομοίωσε τα $\mu_j^{(t)}$ ($j = 1, 2$) από την

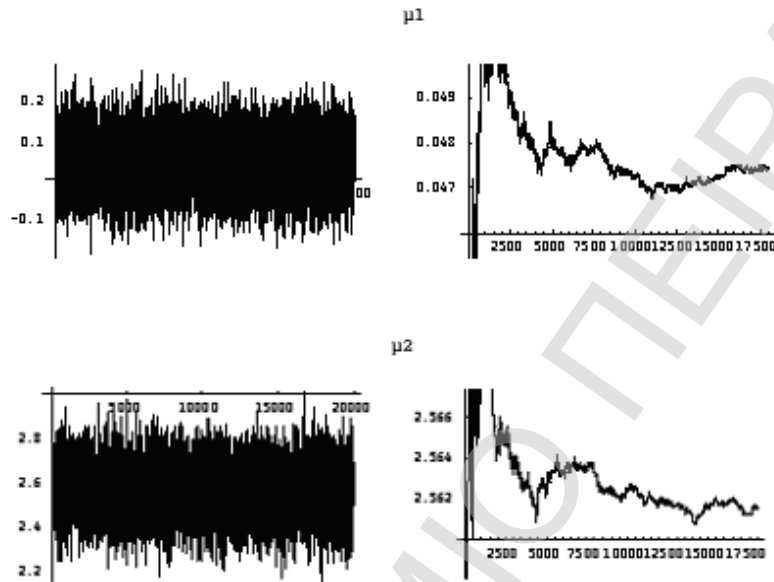
$$\mu_j | \underline{z}^{(t)}, \underline{x} \sim N\left(\frac{\lambda \delta + s_j^x}{\lambda + n_j}, \frac{1}{\lambda + n_j}\right).$$

Στο Σχήμα 7 φαίνεται η απόδοση του παραπάνω αλγορίθμου για ένα προσομοιωμένο δείγμα παρατηρήσεων από την $0.7N(0,1) + 0.3N(2.5,1)$.



ΣΧΗΜΑ 7. Χάρτης επιφάνειας του λογαρίθμου πιθανοφάνειας και το αντίστοιχο δείγμα Gibbs για το μοντέλο (1.8) βασισμένο σε 20000 επαναλήψεις. Οι αρχικές τιμές είναι $\mu_1^{(0)} = -2, \mu_2^{(0)} = 1$.

Όπως παρατηρούμε στο Σχήμα 7, η συμπεριφορά του αλγορίθμου είναι σε πλήρη συμφωνία με την αντίστοιχη επιφάνεια της πιθανοφάνειας. Βλέπουμε ότι η κύρια κορυφή έχει προσελκύσει όλες τις τιμές που παράγονται από τον δειγματολήπτη Gibbs. Ειδικότερα, οι τιμές των μέσων και τα αντίστοιχα γραφήματα των εργοδικών μέσων έχουν ως εξής:



ΣΧΗΜΑ 8. Προσομοιωμένες τιμές και τα αντίστοιχα γραφήματα των εργοδικών μέσων για το μοντέλο (1.8), με περίοδο *burn in* τις πρώτες 2000 επαναλήψεις.

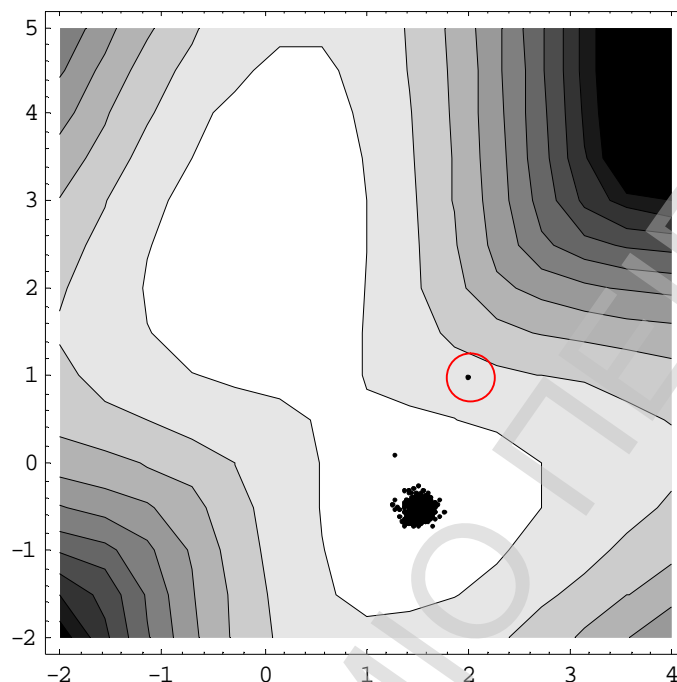
Μετά το «κάψιμο» (*burn-in*¹) των πρώτων 2000 επαναλήψεων παρατηρούμε από το γράφημα των εργοδικών μέσων, ότι ο αλγόριθμος έχει συγκλίνει, ενώ οι εκτιμήσεις που παίρνουμε είναι: για το $E(\mu_1|x)$: $\hat{\mu}_1 = 0.04743$ και για το $E(\mu_2|x)$: $\hat{\mu}_2 = 2.56148$.

Το παράδειγμα αυτό δίνει μία εσφαλμένη αίσθηση ασφάλειας για την αποδοτικότητα του αλγορίθμου, γιατί δεν καταδεικνύει την εξάρτηση του από τις αρχικές συνθήκες. Λόγω της χρησιμοποίησης των δεσμευμένων κατανομών, ο δειγματολήπτης Gibbs πολλές φορές είναι περιορισμένος σχετικά με το εύρος των κινήσεων του. Εδώ δοθέντος του z έχουμε ότι οι προτεινόμενες τιμές για το (μ_1, μ_2) είναι αρκετά συγκεντρωμένες και δεν επιτρέπουν δραστικές αλλαγές στις τιμές του z για το επόμενο βήμα. Για να παρατηρήσουμε μία σημαντική αλλαγή στην κατάταξη απαιτείται μεγάλος αριθμός επαναλήψεων, άπαξ και έχει προσεγγιστεί μία «σταθερή» θέση. Στο Σχήμα 9 φαίνεται ότι όταν αρχικοποιήσουμε τον αλγόριθμο σε μία κοντινή περιοχή της ψεύτικης κορυφής, ο δειγματολήπτης Gibbs είναι ανίκανος να ξεφύγει από αυτή, ακόμα και μετά από έναν μεγάλο αριθμό επαναλήψεων.

Είναι αρκετά ενδιαφέρον ότι ο δειγματολήπτης Gibbs παρουσιάζει το ίδιο πρόβλημα με τον αλγόριθμο EM, αν και αυτό δεν είναι περίεργο αφού και οι δύο μέθοδοι βασίζονται στην ίδια αρχή συμπλήρωσης των δεδομένων. Έτσι με αυτό το

¹ *burn-in period*: Όταν προσομοιώνουμε μια Μαρκοβιανή αλυσίδα με σκοπό να προσεγγίσουμε την στάσιμη κατανομή π αυτής συνηθίζεται να «πετάμε» τις πρώτες τιμές που παράγονται, γιατί απέχουν πολύ από την π . Οι τιμές αυτές εξαιρούνται στον υπολογισμό των εργοδικών μέσων και το πλήθος αυτών ονομάζεται *burn-in period*.

παράδειγμα καταλαβαίνουμε ότι αν και η μέθοδος αυτή είναι αρκετά λογική (διότι είναι κατά κάποιο τρόπο μέρος του ορισμού του μοντέλου), η ωφελιμότητα αυτής δεν μεταφέρεται απαραίτητα και στον αλγόριθμο προσομοίωσης.



ΣΧΗΜΑ 9. Χάρτης επιφάνειας του λογαρίθμου πιθανοφάνειας και το αντίστοιχο δείγμα Gibbs για το μοντέλο (1.8) βασισμένο σε 20000 επαναλήψεις. Οι αρχικές τιμές είναι $\mu_1^{(0)} = -2, \mu_2^{(0)} = 1$.

Παράδειγμα 2.1.2 Ας θεωρήσουμε την μείξη 3 κατανομών Poisson, και έστω ότι διαθέτουμε ένα τυχαίο δείγμα \underline{x} από το μοντέλο $\sum_{j=1}^3 p_j P(\lambda_j)$ (2.8), όπου με $P(\lambda_j)$ συμβολίζουμε την κατανομή Poisson με παράμετρο λ_j . Εδώ έχουμε ότι $\underline{\theta} = (\lambda_1, \lambda_2, \lambda_3)$, $\underline{p} = (p_1, p_2, p_3)$ και θεωρούμε επιπλέον ότι και τα βάρη είναι άγνωστα. Εκ των προτέρων έχουμε ότι: $\lambda_j \sim \Gamma(\alpha_j, \beta_j)$ και $\underline{p} \sim D(\gamma_1, \gamma_2, \gamma_3)$, όπου τα $(\alpha_j, \beta_j, \gamma_j), j = 1, 2, 3$ είναι γνωστές υπερπαράμετροι. Τότε:

$$\begin{aligned} \lambda_1, \lambda_2, \lambda_3 | \underline{z}, \underline{x} &\propto \prod_{i=1}^n e^{-\lambda_{z_i}} \lambda_{z_i}^{x_i} \prod_{j=1}^3 \lambda_j^{\alpha_j - 1} e^{-\beta_j \lambda_j} \\ &\propto \prod_{j=1}^3 \exp\left(-\lambda_j \left(\sum_{i=1}^n I_{\{z_i=j\}} + \beta_j\right)\right) \lambda_j^{\alpha_j - 1 + \sum_{i=1}^n I_{\{z_i=j\}} x_i} \\ &= \prod_{j=1}^3 \exp(-\lambda_j (n_j + \beta_j)) \lambda_j^{\alpha_j + s_j^x - 1} \end{aligned} \quad (2.9)$$

Από την σχέση (2.9) έχουμε ότι εκ των υστέρων τα λ_j είναι ανεξάρτητα και οι δεσμευμένες κατανομές αυτών είναι: $\lambda_j | \underline{z}, \underline{x} \sim \Gamma(\alpha_j + s_j^x, n_j + \beta_j), j = 1, 2, 3$. Οπότε ο

αλγόριθμος για την εφαρμογή του δειγματολήπτη Gibbs σε αυτή την περίπτωση έχει ως εξής:

Δειγματολήπτης Gibbs για μείξη Poisson

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}^{(0)}, \underline{p}^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots$:

1.1. Προσομοίωσε τα $z_i^{(t)}, (i = 1, \dots, n)$ από την

$$P(z_i^{(t)} = j) \propto p_j^{(t-1)} (\lambda_j^{(t-1)})^{x_i} \exp(-\lambda_j^{(t-1)}).$$

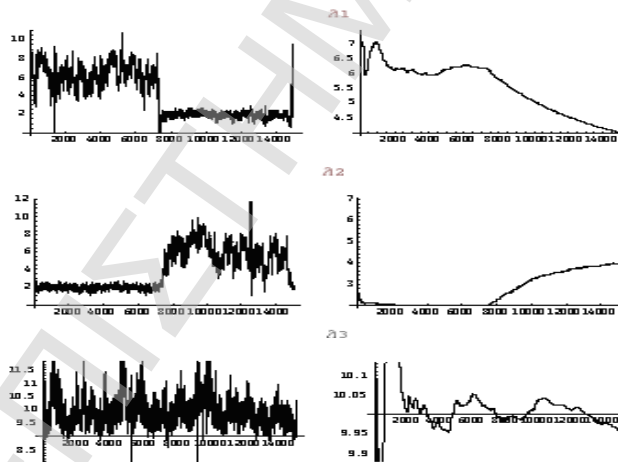
1.2. Υπολόγισε τα $n_j^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}}$ και $(s_j^x)^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}} x_i, j = 1, 2, 3$.

1.3. Προσομοίωσε τα $\underline{p}^{(t)}$ από την $D(\gamma_1 + n_1, \dots, \gamma_k + n_k)$.

1.4. Προσομοίωσε τα $\lambda_j^{(t)} (j = 1, 2, 3)$ από την

$$\lambda_j | \underline{z}^{(t)}, \underline{x} \sim \Gamma(\alpha_j + s_j^x, n_j + \beta_j).$$

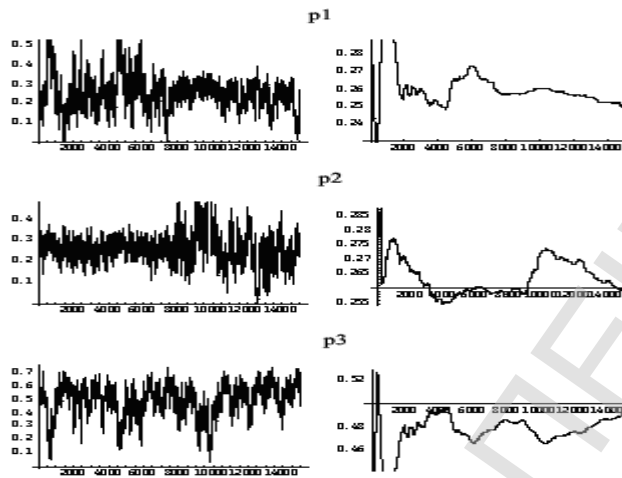
Στο Σχήμα 10 φαίνονται οι τιμές και οι εργοδικοί μέσοι του παραπάνω αλγορίθμου για ένα προσομοιωμένο δείγμα 800 παρατηρήσεων από την $0.25P(2) + 0.25P(6) + 0.5P(10)$.



ΣΧΗΜΑ 10. Προσομοιωμένες τιμές και τα αντίστοιχα γραφήματα των εργοδικών μέσων των $\lambda_j, j = 1, 2, 3$ για το μοντέλο (2.8), για 15000 επαναλήψεις του δειγματολήπτη.

Για την εφαρμογή αυτή επιλέξαμε σκόπιμα την ίδια εκ των προτέρων κατανομή στις παραμέτρους των κατανομών Poisson της μείξης (συγκεκριμένα θέσαμε $\lambda_j \sim \Gamma(0.7, 0.1), j = 1, 2, 3$). Από την εκτέλεση 15000 επαναλήψεων του αλγορίθμου παρατηρούμε ότι ο δειγματολήπτης Gibbs εναλλάσσει συνεχώς κορυφές τις πιθανοφάνειας διότι έπεσε θύμα του φαινομένου label switching: στα Σχήματα 10 και 11 φαίνονται οι διακριτές αναπηδήσεις που συμβαίνουν στα ίχνη των μέσων των λ_j και p_j αντίστοιχα, καθώς η αλυσίδα προχωράει. Διαισθητικά, οι περιοχές όπου

«επισκέπτεται» ο αλγόριθμος αντιστοιχούν σε κάποιους από τους 3! δυνατούς τρόπους για να ονοματίσουμε τις συνιστώσες. Οι εκτιμήσεις των περιθωριακών εκ των υστέρων κατανομών είναι παρόμοιες μεταξύ τους ώστε η εκτίμηση των εργοδικών μέσων δεν είναι άμεση.



ΣΧΗΜΑ 11. Τα αντίστοιχα γραφήματα των p_j , $j = 1, 2, 3$ για το μοντέλο (2.8).

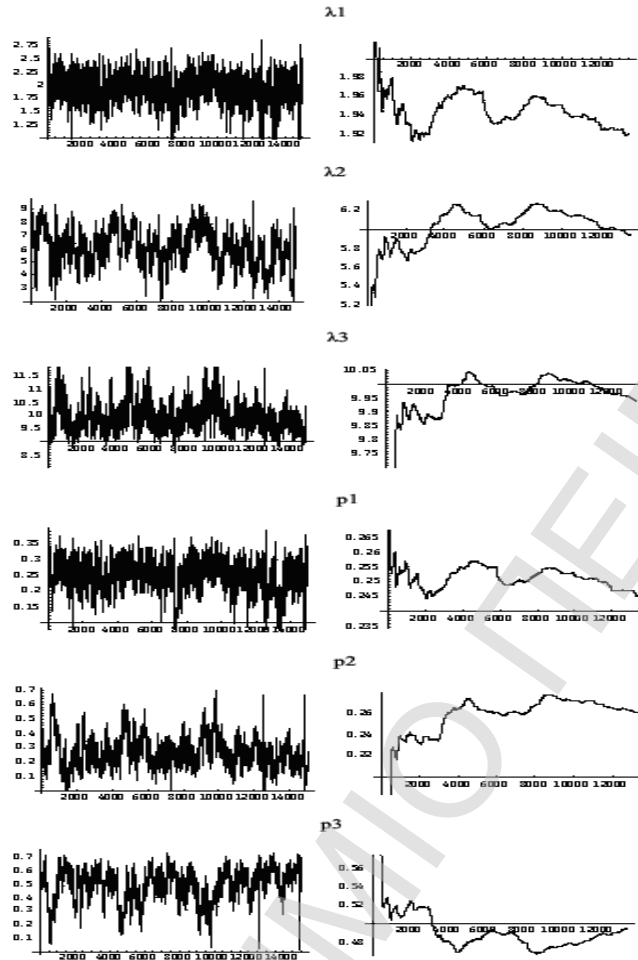
Σε αυτή την περίπτωση οι εκτιμήσεις που παίρνουμε είναι:

$$\begin{aligned} \hat{E}(\lambda_1 | \underline{x}) = \hat{\lambda}_1 &= 4.02416 & \hat{E}(p_1 | \underline{x}) = \hat{p}_1 &= 0.248063 \\ \hat{E}(\lambda_2 | \underline{x}) = \hat{\lambda}_2 &= 3.91039 & \hat{E}(p_2 | \underline{x}) = \hat{p}_2 &= 0.259743 \\ \hat{E}(\lambda_3 | \underline{x}) = \hat{\lambda}_3 &= 9.95358 & \hat{E}(p_3 | \underline{x}) = \hat{p}_3 &= 0.492195 \end{aligned}$$

Βέβαια αυτές οι εκτιμήσεις δεν έχουν καν νόημα και τις παραθέτουμε μόνο και μόνο για μία εντύπωση των προβλημάτων που προκαλεί το φαινόμενο label switching.

Για την αντιμετώπιση του προβλήματος label switching επιλέγουμε να χρησιμοποιήσουμε τον αλγόριθμο αναδιάταξης των παραγόμενων τιμών μέσω του εσωτερικού γινομένου που περιγράψαμε στην Παράγραφο 1.5.3. Το διάνυσμα των προσομοιωμένων τιμών που μεγιστοποιεί την εκ των υστέρων πυκνότητα είναι το: $\{I_1, I_2, I_3, p_1, p_2, p_3\} = \{1.9422, 5.35077, 9.60246, 0.247285, 0.163929, 0.588785\}^2$, το οποίο ορίζεται ως οδηγός για την αναδιάταξη των υπολοίπων τιμών. Μετά λοιπόν την αναδιάταξη των προηγούμενων τιμών λαμβάνουμε την εξής εικόνα (δες Σχήμα 12):

² Προσέξτε ότι ο ίδιος ο οδηγός αποτελεί μία σχετικά καλή εκτίμηση, συγκρίνοντας τον με τις πραγματικές τιμές των παραμέτρων που χρησιμοποιήθηκαν για την προσομοίωση των δεδομένων του παραδείγματος.



ΣΧΗΜΑ 12. Αναδιατεταγμένες τιμές και τα αντίστοιχα γραφήματα των εργοδικών μέσων των $\lambda_j, j=1,2,3$ και $p_j, j=1,2,3$ με περίοδο burn in τις πρώτες 1500 επαναλήψεις, για το μοντέλο (2.8), μέσω του εσωτερικού γινομένου.

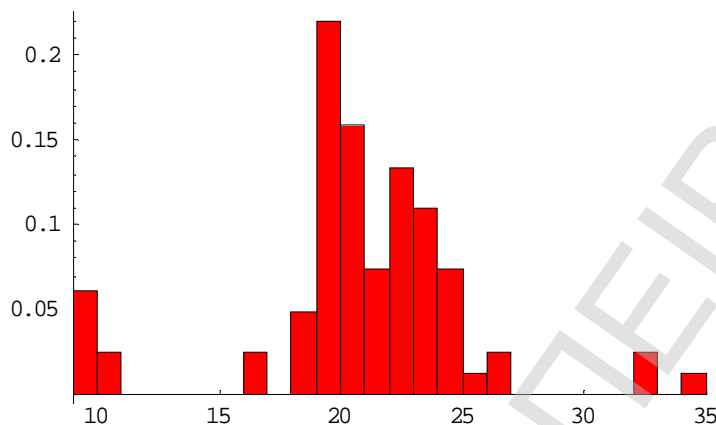
Στο Σχήμα 12 παρατηρούμε την συγκέντρωση των τιμών σε ξεχωριστές περιοχές μεταξύ τους, το οποίο συνεπάγεται την σύγκλιση των αντίστοιχων εργοδικών μέσων. Συγκεκριμένα οι εκτιμήσεις που παίρνουμε, είναι:

$$\begin{aligned} \hat{E}(I_1|x) &= \hat{I}_1 = 1.92151 & \hat{E}(p_1|x) &= \hat{p}_1 = 0.245559 \\ \hat{E}(I_2|x) &= \hat{I}_2 = 5.94996 & \hat{E}(p_2|x) &= \hat{p}_2 = 0.258544 \\ \hat{E}(I_3|x) &= \hat{I}_3 = 9.94219 & \hat{E}(p_3|x) &= \hat{p}_3 = 0.495896 \end{aligned}$$

Συμπεραίνουμε ότι ο δειγματολήπτης Gibbs σε συνδυασμό με την συγκεκριμένη μέθοδο αναδιάταξης των τιμών που παράγονται από αυτόν, είναι αρκετά αποδοτικός, αφενός μεν διότι εξαλείφεται το πρόβλημα label switching, αφ' ετέρου δε διότι επιτυγχάνεται η σύγκλιση των εργοδικών μέσων στις πραγματικές εκ των υστέρων μέσες τιμές των παραμέτρων.

Παράδειγμα 2.1.3 Αυτό το παράδειγμα πραγματεύεται μία εφαρμογή - ορόσημο στην εκτίμηση μείξεων, το γαλαξιακό σύνολο δεδομένων του Roeder (1990). Το δείγμα αποτελείται από τις ταχύτητες (σε 10^3 km/sec) 82 γαλαξιών που

απομακρύνονται από τον δικό μας, και οι οποίοι προέρχονται από τις κωνικές περιοχές του *Corona Borealis*. Έχει αναλυθεί υπό μία πλειάδα μοντέλων μείξεων από ερευνητές όπως οι Crawford (1994), Chib (1995), Philips & Smith (1996), Richardson & Green (1997) και Stephens (2000). Στο Σχήμα 13 παραθέτουμε το ιστόγραμμα των δεδομένων αυτών.



ΣΧΗΜΑ 13. Το ιστόγραμμα των 82 γαλαξιακών ταχυτήτων του *Corona Borealis*.

Όλοι οι ερευνητές θεωρούν ότι οι ταχύτητες αυτές είναι πραγματοποιήσεις ανεξάρτητων και ισόνομων τυχαίων μεταβλητών, κατανεμημένες σύμφωνα με μία μείξη k κανονικών κατανομών. Οι εκτιμήσεις για τον αριθμό των συνιστωσών k ποικίλλουν από 3 (για τους Roeder & Wasserman), σε 5 ή 6 (Richardson & Green), ενώ από μία καθαρά αστροφυσική σκοπιά έχει υποστηριχθεί ότι πρόκειται τουλάχιστον για 7. Προς το παρόν θα περιοριστούμε στην παραδοχή ότι ο αριθμός των συνιστωσών είναι εκ των προτέρων γνωστός και ίσος με 3, κυρίως για λόγους κατανόησης της μεθόδου, αλλά και ιστορικούς αφού αυτή ήταν η πρώτη προσέγγιση για το συγκεκριμένο πρόβλημα (για μία ολοκληρωμένη αντιμετώπιση όπου ο αριθμός k αποτελεί μέρος των προς εκτίμηση παραμέτρων παραπέμπουμε στο Κεφάλαιο 3). Έτσι θεωρούμε το μοντέλο:

$$\pi(\underline{x}|\underline{p}, \underline{\mu}, \underline{\sigma}^2) = \sum_{i=1}^k p_i \mathfrak{N}(\underline{x}|\mu_i, \sigma_i^2). \quad (2.10)$$

Εδώ έχουμε ότι $\underline{\theta} = (\underline{\mu}, \underline{\sigma}^2)$ και $k = 3$ (βέβαια ότι ακολουθεί ισχύει για όλα τα k). Για τις εκ των προτέρων κατανομές ακολουθούμε τους Richardson & Green, όπου:

$$\mu_j \sim N(\xi, \kappa^{-1}) \text{ και } \sigma_j^{-2} \sim \Gamma(\alpha, \beta)$$

με το ξ να λαμβάνεται ίσο με τον δειγματική διάμεσο (20.8315) ενώ το $\kappa = 1/R^2 \cong 0.00158$, $R = \max\{X_i; i = 1, \dots, n\} - \min\{X_i; i = 1, \dots, n\} = 25.107$. Επίσης θέτουμε $\alpha = 2$, ενώ το $\beta \sim \Gamma(g, h)$ όπου $g = 0.2$ και $h = 10/R^2 = 0.0158$. Ως εκ των προτέρων κατανομή των βαρών της μείξης λαμβάνεται η μη πληροφοριακή Dirichlet(1,1,1). Με αυτές τις υποθέσεις προκύπτουν συζυγείς εκ των υστέρων κατανομές.

Για το διάνυσμα των μέσων έχουμε:

$$\begin{aligned} \underline{\mu} | \underline{z}, \underline{x}, \underline{\sigma}^2 &\propto \prod_{i=1}^n \exp\left(-\frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right) \prod_{j=1}^k \exp\left(-\frac{1}{2} \sum_{i=1}^n \kappa(\mu_j - \zeta)^2\right) \\ &\propto \prod_{j=1}^k \exp\left\{-\frac{1}{2} \left(\kappa(\mu_j - \zeta)^2 + \sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} (x_i - \mu_j)^2 \right)\right\} \end{aligned} \quad (2.11)$$

Από την (2.11) παρατηρούμε ότι οι μέσοι είναι εκ των υστέρων ανεξάρτητοι. Οπότε για $j = 1, \dots, k$, έχουμε:

$$\begin{aligned} \mu_j | \underline{z}, \underline{x}, \underline{\sigma}^2 &\propto \exp\left(-\frac{1}{2} \sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} x_i^2 - \frac{1}{2} \sigma_j^{-2} \mu_j^2 \sum_{i=1}^n I_{\{z_i=j\}} + \sigma_j^{-2} \mu_j \sum_{i=1}^n I_{\{z_i=j\}} x_i - \frac{\mu_j^2 \kappa}{2} + \mu_j \kappa \zeta\right) \\ &\propto \exp\left(-\frac{1}{2} \left\{ \mu_j^2 \left(\sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} + \kappa \right) - 2\mu_j \left(\sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} x_i + \kappa \zeta \right) \right\}\right) \\ &= \exp\left(-\frac{\left(\sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} + \kappa \right)}{2} \left\{ \mu_j^2 - 2\mu_j \frac{\sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} x_i + \kappa \zeta}{\sigma_j^{-2} \sum_{i=1}^n I_{\{z_i=j\}} + \kappa} \right\}\right) \\ &\propto \exp\left(-\frac{\kappa + \sigma_j^{-2} n_j}{2} \left(\mu_j - \frac{\sigma_j^{-2} s_j^x + \kappa \zeta}{\kappa + \sigma_j^{-2} n_j} \right)^2\right) \end{aligned}$$

οπότε οι πλήρεις δεσμευμένες κατανομές των μ_j είναι:

$$\mu_j | \underline{z}, \underline{x}, \underline{\sigma}^2 \sim N\left(\frac{\sigma_j^{-2} s_j^x + \kappa \zeta}{\kappa + \sigma_j^{-2} n_j}, \frac{1}{\kappa + \sigma_j^{-2} n_j}\right) \quad j = 1, \dots, k. \quad (2.12)$$

Στην συνέχεια υπολογίζουμε την εκ των υστέρων δεσμευμένη κατανομή του διανύσματος των διασπορών:

$$\begin{aligned} \underline{\sigma}^2 | \underline{z}, \underline{x}, \underline{\mu}, \beta &\propto \prod_{i=1}^n \frac{1}{\sigma_{z_i}} \exp\left(-\frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right) \prod_{j=1}^k \frac{1}{(\sigma_j^2)^{\alpha+1}} \exp(-\beta/\sigma_j^2) \\ &\propto \prod_{j=1}^k \frac{1}{(\sigma_j^2)^{\alpha + \frac{1}{2} \sum_{i=1}^n I_{\{z_i=j\}} + 1}} \exp\left\{-\left(\beta + \frac{1}{2} \sum_{i=1}^n I_{\{z_i=j\}} (x_i - \mu_j)^2\right) / \sigma_j^2\right\} \\ &= \prod_{j=1}^k \frac{1}{(\sigma_j^2)^{\alpha + \frac{1}{2} n_j + 1}} \exp\left\{-\left(\beta + \frac{1}{2} ss_j^x\right) / \sigma_j^2\right\} \end{aligned} \quad (2.13)$$

όπου: $ss_j^x = \sum_{i=1}^n I_{\{z_i=j\}} (x_i - \mu_j)^2$.

Η σχέση (2.13) δείχνει ότι τα σ_j^2 είναι ανεξάρτητα και ακολουθούν κατανομές αντίστροφες Γάμμα. Οπότε:

$$\sigma_j^{-2} | \dots \sim \Gamma\left(\alpha + \frac{1}{2}n_j, \beta + \frac{1}{2}ss_j^x\right), \quad j = 1, 2, \dots, k. \quad (2.14)$$

Επίσης η εκ των υστέρων δεσμευμένη κατανομή του \underline{z} δοθέντος του (μ_1, \dots, μ_k) , προκύπτει από τον τύπο (1.6):

$$P(z_i = j | \dots) \propto \frac{p_j}{\sigma_j} \exp\left\{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right\}, \quad j = 1, 2, \dots, k.$$

Τέλος, για την εκ των υστέρων κατανομή της υπερπαραμέτρου β , λαμβάνουμε:

$$\begin{aligned} \beta | \underline{\sigma}^{-2} &\propto \beta^{g-1} \exp(-h\beta) \prod_{j=1}^k \beta^\alpha \exp(-\beta\sigma_j^{-2}) = \beta^{g+\kappa\alpha-1} \exp\left\{-\left(h + \sum_{j=1}^k \sigma_j^{-2}\right)\beta\right\} \Rightarrow \\ \beta | \underline{\sigma}^{-2} &\sim \Gamma\left(g + \kappa\alpha, h + \sum_{j=1}^k \sigma_j^{-2}\right). \end{aligned} \quad (2.15)$$

Ύστερα από αυτήν την ανάλυση ο αλγόριθμος για την εκτέλεση του δειγματολήπτη Gibbs για αυτή την εφαρμογή, έχει ως εξής:

Δειγματολήπτης Gibbs για κανονικές μείξεις (μοντέλο 2.10)

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}^{(0)}$, $\underline{p}^{(0)}$ και $\beta^{(0)}$.

1. Στην επανάληψη t , $t = 1, 2, \dots$:

1.1. Προσομοίωσε τα $z_i^{(t)}$, ($i = 1, \dots, n$) από την

$$P(z_i^{(t)} = j) \propto \frac{p_j^{(t-1)}}{\sigma_j^{(t-1)}} \exp \left\{ -\frac{(x_i - \mu_j^{(t-1)})^2}{2(\sigma_j^2)^{(t-1)}} \right\}.$$

1.2. Υπολόγισε τα $n_j^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}}$ και $(s_j^x)^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}} x_i$.

1.3. Προσομοίωσε τα $\underline{p}^{(t)}$ από την $D(\gamma_1 + n_1, \dots, \gamma_k + n_k)$.

1.4. Προσομοίωσε τα $\mu_j^{(t)}$ ($j = 1, \dots, k$) από την

$$\mu_j | \underline{z}^{(t)}, \underline{x}, (\underline{\sigma}^2)^{(t-1)} \sim N \left(\frac{\sigma_j^{-2} s_j^x + \kappa \zeta}{\kappa + \sigma_j^{-2} n_j}, \frac{1}{\kappa + \sigma_j^{-2} n_j} \right).$$

1.5. Υπολόγισε τα $(ss_j^x)^{(t)} = \sum_{i=1}^n I_{\{z_i^{(t)}=j\}} (x_i - \mu_j^{(t)})^2$, $j = 1, \dots, k$.

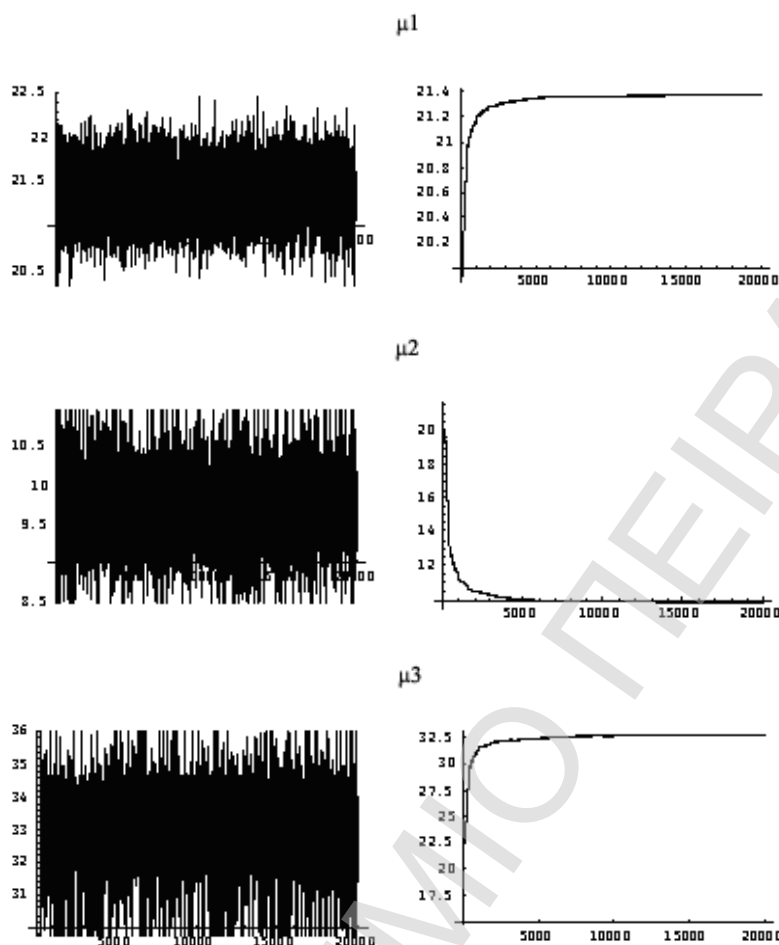
1.6. Προσομοίωσε τα $(\sigma_j^2)^{(t)}$ ($j = 1, \dots, k$) από την

$$\sigma_j^2 | \underline{z}^{(t)}, \underline{x}, \mu_j^{(t)}, \beta^{(t-1)} \sim \Gamma \left(\alpha + \frac{1}{2} n_j, \beta + \frac{1}{2} ss_j^x \right).$$

1.7. Προσομοίωσε την υπερπαράμετρο $\beta^{(t)}$ από την

$$\beta | (\underline{\sigma}^2)^{(t)} \sim \Gamma \left(g + k\alpha, h + \sum_{j=1}^k \sigma_j^{-2} \right).$$

Από την εκτέλεση 20000 επαναλήψεων του αλγορίθμου παρατηρούμε την απουσία του φαινομένου label switching. Αυτό εν μέρει, οφείλεται στο ότι ο αλγόριθμος δεν ολοκλήρωσε το δυνατό εύρος κινήσεων του, αλλά και στο μικρό πλήθος του αριθμού των συνιστωσών (δες και την παρατήρηση μετά το παράδειγμα 2.2.2). Στο Σχήμα 14 φαίνονται οι τιμές των μ_j καθώς η αλυσίδα εξελίσσεται.



ΣΧΗΜΑ 14. Προσομοιωμένες τιμές των $\mu_j, j=1,2,3$ για το μοντέλο (2.10), για 30000 επαναλήψεις του δειγματολήπτη Gibbs.

Η εικόνα για τους αντίστοιχους εργοδικούς μέσους είναι ξεκάθαρη: όπως παρατηρούμε στο Σχήμα 14 η σύγκλιση είναι αρκετά γρήγορη. Συγκεκριμένα οι εκτιμήσεις των εκ των υστέρων μέσων τιμών που παίρνουμε είναι οι ακόλουθες:

$$\begin{array}{lll}
 \hat{E}(\mu_1|\underline{x}) \approx 21.3926 & \hat{E}(\sigma_1^2|\underline{x}) \approx 4.81412 & \hat{E}(p_1|\underline{x}) \approx 0.855719 \\
 \hat{E}(\mu_2|\underline{x}) \approx 9.71465 & \hat{E}(\sigma_2^2|\underline{x}) \approx 0.899851 & \hat{E}(p_2|\underline{x}) \approx 0.0944184 \\
 \hat{E}(\mu_3|\underline{x}) \approx 32.7642 & \hat{E}(\sigma_3^2|\underline{x}) \approx 2.90852 & \hat{E}(p_3|\underline{x}) \approx 0.0498626
 \end{array}$$

Σε αυτό το παράδειγμα παρουσιάσαμε μία από τις πιο γνωστές εφαρμογές των μοντέλων μείξεων στην οποία θα επανέλθουμε και στην συνέχεια. Προς το παρόν τονίζουμε ότι τα δεδομένα αυτά έχουν αναλυθεί από αρκετές διαφορετικές σκοπιές (όπως η επιλογή των εκ των προτέρων κατανομών, ακόμα και του αριθμού των συνιστωσών). Η επιλογή του μοντέλου των Richardson & Green είναι μία παραδοχή η οποία συμφωνεί και με άλλες προσεγγίσεις, για αυτό και θα επιμείνουμε σε αυτήν. Η θεώρηση του πλήθους των συνιστωσών ίσου με 3 είναι επίσης μία παραδοχή και η εξέταση της βασιμότητάς της θα μας απασχολήσει ιδιαίτερα στα επόμενα κεφάλαια.

2.2 Ο αλγόριθμος Metropolis - Hastings

Όπως φαίνεται στο Σχήμα 9, ο δειγματολήπτης Gibbs μπορεί να αποτύχει να αποδράσει από την έλξη μιας τοπικής κορυφής ακόμα και σε απλές σχετικά περιπτώσεις, όπως αυτή του Παραδείγματος 2.1.1 όπου τα βάρη και οι διασπορές είναι γνωστές αλλά και οι παράμετροι είναι διακριτές μεταξύ τους. Αυτό οφείλεται εν μέρει στην συμπλήρωση των δεδομένων με την εισαγωγή των ελλιπών μεταβλητών \underline{z} . Από την μία αυτή η προσέγγιση λύνει το πρόβλημα των υπολογιστικών δυσκολιών, αλλά από την άλλη αυξάνει την διάσταση του χώρου προσομοίωσης και ταυτόχρονα περιορίζει αρκετά τις κινήσεις της Μαρκοβιανής αλυσίδας.

Μία εναλλακτική μέθοδος είναι ο αλγόριθμος *Metropolis - Hastings*, ο οποίος δεν απαιτεί την διεύρυνση των δεδομένων και κατ' επέκταση την διόγκωση του χώρου προσομοίωσης (βέβαια απαιτείται ο υπολογισμός της πιθανοφάνειας, κάτι το οποίο συνεπάγεται την αφιέρωση $O(kn)$ μονάδων χρόνου). Αυτό αποτελεί και το μειονέκτημα του αλγορίθμου αυτού, τόσο από άποψη χρόνου, όσο και από το γεγονός ότι είναι πολύ πιθανή η υπερχειλίση (overflow) των υπολογιστικών πακέτων ή της γλώσσας που χρησιμοποιείται για τον προγραμματισμό του. Ενδεικτικά αναφέρουμε ότι σε μία τυπική περίπτωση ενός μοντέλου μείξης όπως αυτή του μοντέλου (1.8) η τιμή της πιθανοφάνειας στην κορυφή της είναι της τάξης του 10^{-390} (!).

Η βασική ιδέα: Έστω π η στάσιμη κατανομή της αλυσίδας. Η μέθοδος αυτή βασίζεται σε μία *κατανομή πρότασης* q από την οποία παράγονται οι *υποψήφιος* τιμές. Έτσι αν στο βήμα t είμαστε στην κατάσταση $X^{(t)} = x$, προσομοιώνουμε μία τιμή $y \sim q(\cdot|x)$ την οποία την αποδεχόμαστε με πιθανότητα:

$$a(y|x) = \min \left\{ 1, \frac{p(y)q(x|y)}{p(x)q(y|x)} \right\}.$$

Σε αντίθετη περίπτωση παραμένουμε στην προηγούμενη κατάσταση. Η διαφορά με τον δειγματολήπτη Gibbs είναι ότι η Μαρκοβιανή αλυσίδα που παράγεται έχει και μία επιπλέον ιδιότητα: αυτή της *αντιστρεψιμότητας*.

Όπως ίσως γίνεται εύκολα αντιληπτό, η επιλογή της κατανομής πρότασης q , η οποία μπορεί να είναι οποιαδήποτε³, έχει καθοριστική σημασία για τον αλγόριθμο. Αυτό από την μία είναι καλό διότι όποια κατανομή και να επιλέξουμε - θεωρητικά τουλάχιστον - ο αλγόριθμος συγκλίνει, από την άλλη όμως λίγες είναι εκείνες που δεν οδηγούν σε υψηλό ρυθμό απόρριψης των προτεινόμενων κινήσεων, με αποτέλεσμα την σύγκλιση του αλγορίθμου σε πρακτικά πλαίσια. Μία συνηθισμένη επιλογή για την $q(\cdot|x)$, είναι ο τυχαίος περίπατος, όπου κάθε υποψήφια τιμή των παραμέτρων προκύπτει από την μέση τιμή της κατανομής πρότασης, δηλαδή: $\tilde{q}_j = q_j^{(t-1)} + u_j$ (2.16), όπου $E(u_j) = 0$. Συνηθίζεται δε η επιλογή $u_j \sim N(0, \zeta^2)$.

Για να επανατοποθετηθούμε στο πλαίσιο του προβλήματος της Μπεϋζιανής εκτίμησης μείξεων, υπενθυμίζουμε ότι η στάσιμη κατανομή της αλυσίδας που προσπαθούμε να προσεγγίσουμε μέσω των αλγορίθμων προσομοίωσης είναι η εκ των υστέρων κατανομή του μοντέλου. Ο αλγόριθμος Metropolis - Hastings σε αυτή την περίπτωση έχει ως εξής:

³ Ελάχιστη απαίτηση: η κατανομή πρότασης πρέπει να έχει το ίδιο στήριγμα με την π .

Γενικός αλγόριθμος Metropolis - Hastings για μοντέλα μείξεων

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}^{(0)}, \underline{p}^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots$:

1.1. Προσομοίωσε τις προτεινόμενες τιμές των παραμέτρων $(\tilde{\theta}, \tilde{p})$
από την $(\tilde{\theta}, \tilde{p}) \sim q(\underline{\theta}, \underline{p} | \underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})$.

1.2. Υπολόγισε το

$$r = \frac{f(x|\tilde{\theta}, \tilde{p})\pi(\tilde{\theta}, \tilde{p})q(\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)}|\tilde{\theta}, \tilde{p})}{f(x|\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})\pi(\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})q(\tilde{\theta}, \tilde{p}|\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})}$$

1.3. Προσομοίωσε έναν τυχαίο αριθμό $u \sim U(0,1)$.

Εάν $u < r$ τότε θέσε $(\underline{\theta}^{(t)}, \underline{p}^{(t)}) = (\tilde{\theta}, \tilde{p})$

αλλιώς θέσε $(\underline{\theta}^{(t)}, \underline{p}^{(t)}) = (\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})$.

Πριν προχωρήσουμε σε κάποια παραδείγματα σημειώνουμε ότι όταν έχουμε να προτείνουμε τιμές για παραμέτρους όπως οι διασπορές ή τα βάρη, το στήριγμα των οποίων είναι περιορισμένο, η επιλογή (2.16) για την κατανομή πρότασης είναι αναποτελεσματική. Αυτή η δυσκολία μπορεί να ξεπεραστεί αν αναπαραμετροποιήσουμε το μοντέλο (1.2) θέτοντας

$$p_j = \frac{\omega_j}{\sum_{i=1}^k \omega_i}, \quad \omega_j > 0, \quad \forall j \in \{1, \dots, k\}. \quad (2.17)$$

Έτσι η προτεινόμενη κίνηση για τα $\omega_j, j = 1, \dots, k$, είναι $\log(\tilde{\omega}_j) = \log(\omega_j^{(t-1)}) + u_j$, όπου $E(u_j) = 0$.

Παράδειγμα 2.2.1 Ως πρώτη εφαρμογή θεωρούμε το μοντέλο (1.8) και για κατανομή πρότασης έναν κανονικό τυχαίο περίπατο, δηλαδή:

$$\tilde{\mu}_1 \sim N(\mu_1^{(t-1)}, \zeta^2) \text{ και } \tilde{\mu}_2 \sim N(\mu_2^{(t-1)}, \zeta^2).$$

Παρατηρούμε ότι στην σχέση που δίνει το r , οι πυκνότητες των κατανομών πρότασης απλοποιούνται λόγω συμμετρίας. Σε αυτήν την περίπτωση λοιπόν, εκτελούμε τον ακόλουθο αλγόριθμο:

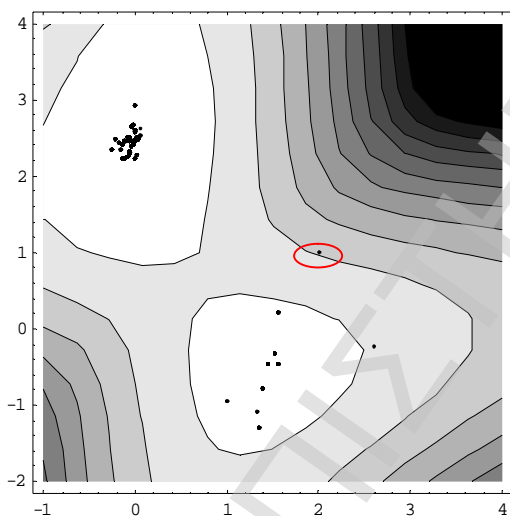
Αλγόριθμος Metropolis - Hastings για το μοντέλο 1.8

0. Δώσε κάποιες αρχικές τιμές στο $\mu_1^{(0)}, \mu_2^{(0)}$.
1. Στην επανάληψη $t, t = 1, 2, \dots$:
 - 1.1. Προσομοίωσε τα $\tilde{\mu}_j \sim N(\mu_j^{(t-1)}, \zeta^2), j = 1, 2$.
 - 1.2. Υπολόγισε το

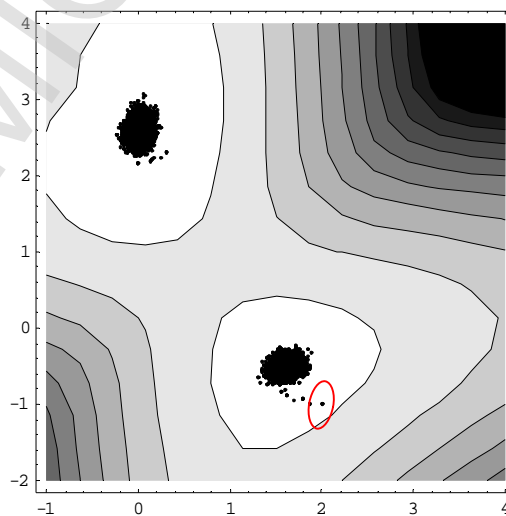
$$r = \frac{f(x|\tilde{\mu}_1, \tilde{\mu}_2)\pi(\tilde{\mu}_1, \tilde{\mu}_2)}{f(x|\tilde{\mu}_1^{(t-1)}, \tilde{\mu}_2^{(t-1)})\pi(\tilde{\mu}_1^{(t-1)}, \tilde{\mu}_2^{(t-1)})}$$

- 1.3. Προσομοίωσε έναν τυχαίο αριθμό $u \sim U(0,1)$.
Εάν $u < r$ τότε θέσε $(\mu_1^{(t)}, \mu_2^{(t)}) = (\tilde{\mu}_1, \tilde{\mu}_2)$
αλλιώς θέσε $(\mu_1^{(t)}, \mu_2^{(t)}) = (\mu_1^{(t-1)}, \mu_2^{(t-1)})$.

Στο Σχήμα 17 παρατηρούμε την εξαιρετική απόδοση του παραπάνω αλγορίθμου. Για λόγους σύγκρισης δώσαμε τις ίδιες αρχικές τιμές με αυτές του Σχήματος 9, ενώ στο Σχήμα 18 δίνουμε διαφορετικές αρχικές τιμές:

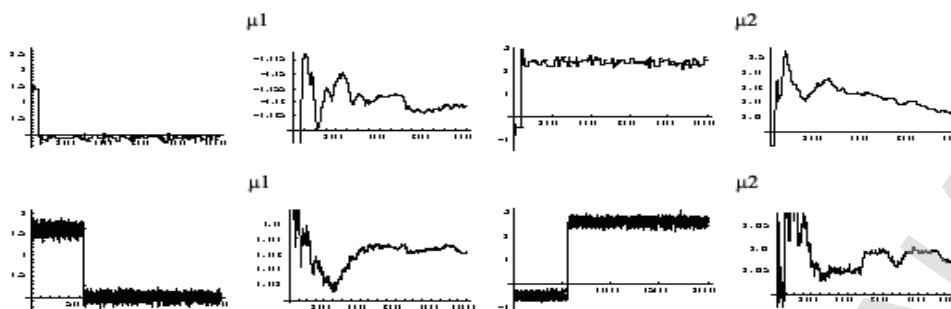


ΣΧΗΜΑ 17. Χάρτης επιφάνειας του λογαρίθμου πιθανοφάνειας και το αντίστοιχο δείγμα τυχαίου περιπάτου Metropolis - Hastings για το μοντέλο (1.8) βασισμένο σε 10000 επαναλήψεις. Οι αρχικές τιμές είναι $\mu_1^{(0)} = 2, \mu_2^{(0)} = 1$, ενώ η διασπορά του τυχαίου περιπάτου ισούται με 1.



ΣΧΗΜΑ 18. Το ίδιο γράφημα βασισμένο σε 20000 επαναλήψεις. Οι αρχικές τιμές είναι $\mu_1^{(0)} = 2, \mu_2^{(0)} = -1$, ενώ η διασπορά του τυχαίου περιπάτου ισούται με 0.01.

Το εντυπωσιακό είναι ότι μετά από λίγες επαναλήψεις ο αλγόριθμος Metropolis - Hastings καταφέρνει να ξεφύγει από την έλξη της τοπικής κορυφής και να προσεγγίσει την πραγματική κορυφή της πιθανοφάνειας, σε αντίθεση με τον δειγματολήπτη Gibbs, ο οποίος άπαξ και πλησιάσει την πρώτη είναι αδύνατο να αποδράσει από αυτή σε «λογικό» πλήθος επαναλήψεων. Τα αντίστοιχα γραφήματα των τιμών και των εργοδικών μέσων έχουν ως εξής:



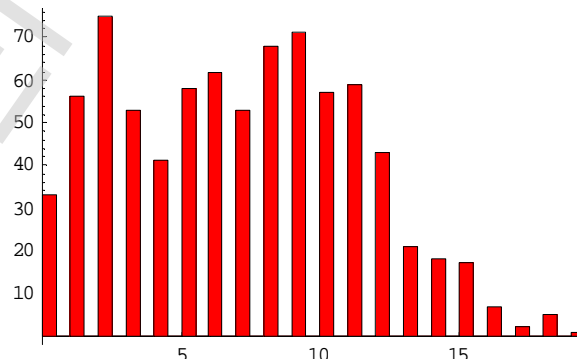
ΣΧΗΜΑ 19. Προσομοιωμένες τιμές και τα αντίστοιχα γραφήματα των εργοδικών μέσων για το μοντέλο (1.8), του Σχήματος 17 με περίοδο burn in τις πρώτες 1000 επαναλήψεις (πρώτη σειρά), και του Σχήματος 18 με περίοδο burn in τις πρώτες 10000 επαναλήψεις.

Μετά το «κάψιμο» των πρώτων 1000 επαναλήψεων παρατηρούμε από το γράφημα των εργοδικών μέσων, ότι ο αλγόριθμος έχει συγκλίνει, ενώ οι εκτιμήσεις που παίρνουμε είναι: για το $E(\mu_1|x)$: $\hat{\mu}_1 = -0.05002$ και για το $E(\mu_2|x)$: $\hat{\mu}_2 = 2.44102$. Εδώ η διασπορά του τυχαίου περιπάτου ήταν ίση με 1, το οποίο εκ των υστέρων κρίνεται καλή επιλογή μια και ο αλγόριθμος πλησιάζει την κατάλληλη περιοχή της πιθανοφάνειας σε λιγότερες από 500 επαναλήψεις, κάτι που φαίνεται στο Σχήμα 18. Εκτελέσαμε τον ίδιο αλγόριθμο με $\zeta^2 = 0.01$, αρχικές τιμές $\mu_1^{(0)} = 2, \mu_2^{(0)} = -1$ και οι εκτιμήσεις που πήραμε είναι και πάλι αρκετά καλές, όμως χρειάστηκαν κάτι παραπάνω από 5500 επαναλήψεις για την απόδραση από την έλξη της ψευτικής κορυφής (Σχήματα 18 και 19). Αυτό είναι αναμενόμενο διότι περιορίσαμε κατά πολύ το εύρος των κινήσεων με αυτή την τιμή για την διασπορά.

Παράδειγμα 2.2.2 Θεωρούμε το μοντέλο του παραδείγματος 2.1.2 και εξετάζουμε την απόδοση του αντίστοιχου αλγόριθμου Metropolis - Hastings, με τις εξής κατανομές πρότασης:

$$\tilde{\lambda}_j \sim LN(\log(\lambda_j^{(t-1)}), \zeta^2) \text{ και } \tilde{\omega}_j \sim LN(\log(\omega_j^{(t-1)}), \zeta^2)$$

όπου με $LN(\mu, \sigma^2)$ συμβολίζουμε την λογαριθμοκανονική κατανομή με παραμέτρους μ και σ^2 .



ΣΧΗΜΑ 20. Το ιστόγραμμα 800 προσομοιωμένων παρατηρήσεων από την $0.25P(2)+0.25P(6)+0.5P(10)$.

Για τον λόγο πυκνοτήτων των κατανομών προτάσεων στην περίπτωση αυτή έχουμε:

$$\begin{aligned}
 & \frac{q(\underline{\theta}^{(t-1)}, \underline{p}^{(t-1)} | \tilde{\underline{\theta}}, \tilde{\underline{p}})}{q(\tilde{\underline{\theta}}, \tilde{\underline{p}} | \underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})} = \\
 & = \frac{\prod_{j=1}^k \frac{1}{\lambda_j^{(t-1)} \sqrt{2\pi\zeta^2}} \exp\left\{-\frac{(\log(\lambda_j^{(t-1)}) - \log(\tilde{\lambda}_j))^2}{2\zeta^2}\right\}}{\prod_{j=1}^k \frac{1}{\tilde{\lambda}_j \sqrt{2\pi\zeta^2}} \exp\left\{-\frac{(\log(\tilde{\lambda}_j) - \log(\lambda_j^{(t-1)}))^2}{2\zeta^2}\right\}} \cdot \frac{\prod_{j=1}^k \frac{1}{\omega_j^{(t-1)} \sqrt{2\pi\zeta^2}} \exp\left\{-\frac{(\log(\omega_j^{(t-1)}) - \log(\tilde{\omega}_j))^2}{2\zeta^2}\right\}}{\prod_{j=1}^k \frac{1}{\tilde{\omega}_j \sqrt{2\pi\zeta^2}} \exp\left\{-\frac{(\log(\tilde{\omega}_j) - \log(\omega_j^{(t-1)}))^2}{2\zeta^2}\right\}} \\
 & = \frac{\prod_{j=1}^k \tilde{\lambda}_j \tilde{\omega}_j}{\prod_{j=1}^k \lambda_j^{(t-1)} \omega_j^{(t-1)}}.
 \end{aligned}$$

Οπότε ο αλγόριθμος έχει ως εξής:

Αλγόριθμος Metropolis - Hastings για μείξη Poisson

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\omega}^{(0)}, \underline{\lambda}^{(0)}$.
1. Στην επανάληψη $t, t = 1, 2, \dots$:
 - 1.1. Προσομοίωσε τα $\tilde{\lambda}_j \sim LN(\log \lambda_j^{(t-1)}, \zeta^2), j = 1, \dots, k$.
 - 1.2. Προσομοίωσε τα $\tilde{\omega}_j \sim LN(\log \omega_j^{(t-1)}, \zeta^2), j = 1, \dots, k$.
 - 1.3. Υπολόγισε το

$$r = \frac{f(\underline{x} | \tilde{\underline{\omega}}, \tilde{\underline{\lambda}}) \pi(\tilde{\underline{\omega}}, \tilde{\underline{\lambda}}) \prod_{j=1}^k \tilde{\lambda}_j \tilde{\omega}_j}{f(\underline{x} | \underline{\omega}^{(t-1)}, \underline{\lambda}^{(t-1)}) \pi(\underline{\omega}^{(t-1)}, \underline{\lambda}^{(t-1)}) \prod_{j=1}^k \lambda_j^{(t-1)} \omega_j^{(t-1)}}.$$

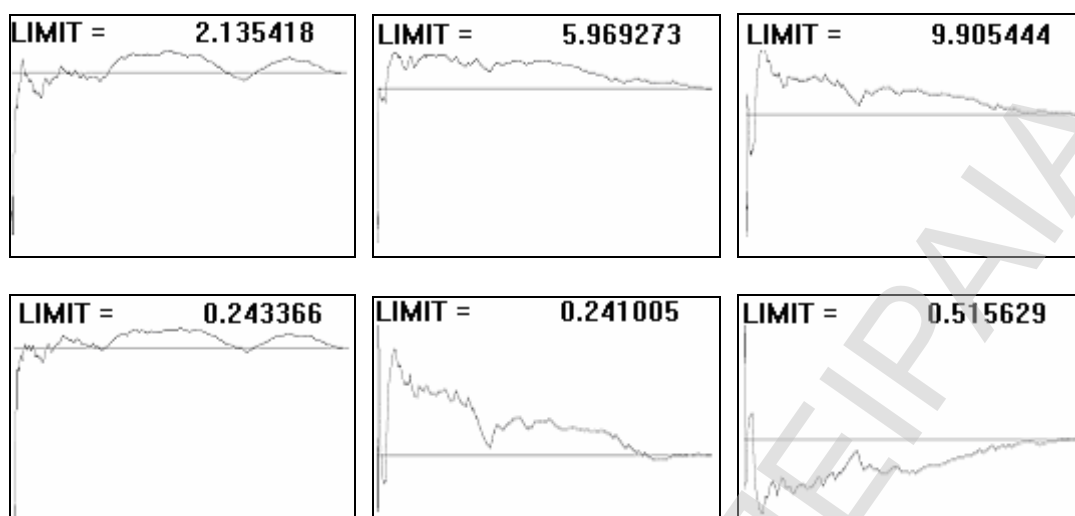
- 1.4. Προσομοίωσε έναν τυχαίο αριθμό $u \sim U(0,1)$.
 Εάν $u < r$ τότε θέσε $(\underline{\lambda}^{(t)}, \underline{\omega}^{(t)}) = (\tilde{\underline{\lambda}}, \tilde{\underline{\omega}})$
 αλλιώς θέσε $(\underline{\lambda}^{(t)}, \underline{\omega}^{(t)}) = (\underline{\lambda}^{(t-1)}, \underline{\omega}^{(t-1)})$.

Εφαρμόσαμε τον παραπάνω αλγόριθμο για ένα προσομοιωμένο σύνολο δεδομένων 800 παρατηρήσεων από την $0.25P(2) + 0.25P(6) + 0.5P(10)$, με την κλίμακα του τυχαίου περιπάτου ίση με $\zeta^2 = 0.05$, και μετά την αναδιάταξη των τιμών, παρατηρήσαμε τα εξής αποτελέσματα:

$$\begin{aligned}
 \hat{E}(I_1 | \underline{x}) &= \hat{I}_1 = 2.135418 \\
 \hat{E}(I_2 | \underline{x}) &= \hat{I}_2 = 5.969273 \\
 \hat{E}(I_3 | \underline{x}) &= \hat{I}_3 = 9.905444
 \end{aligned}$$

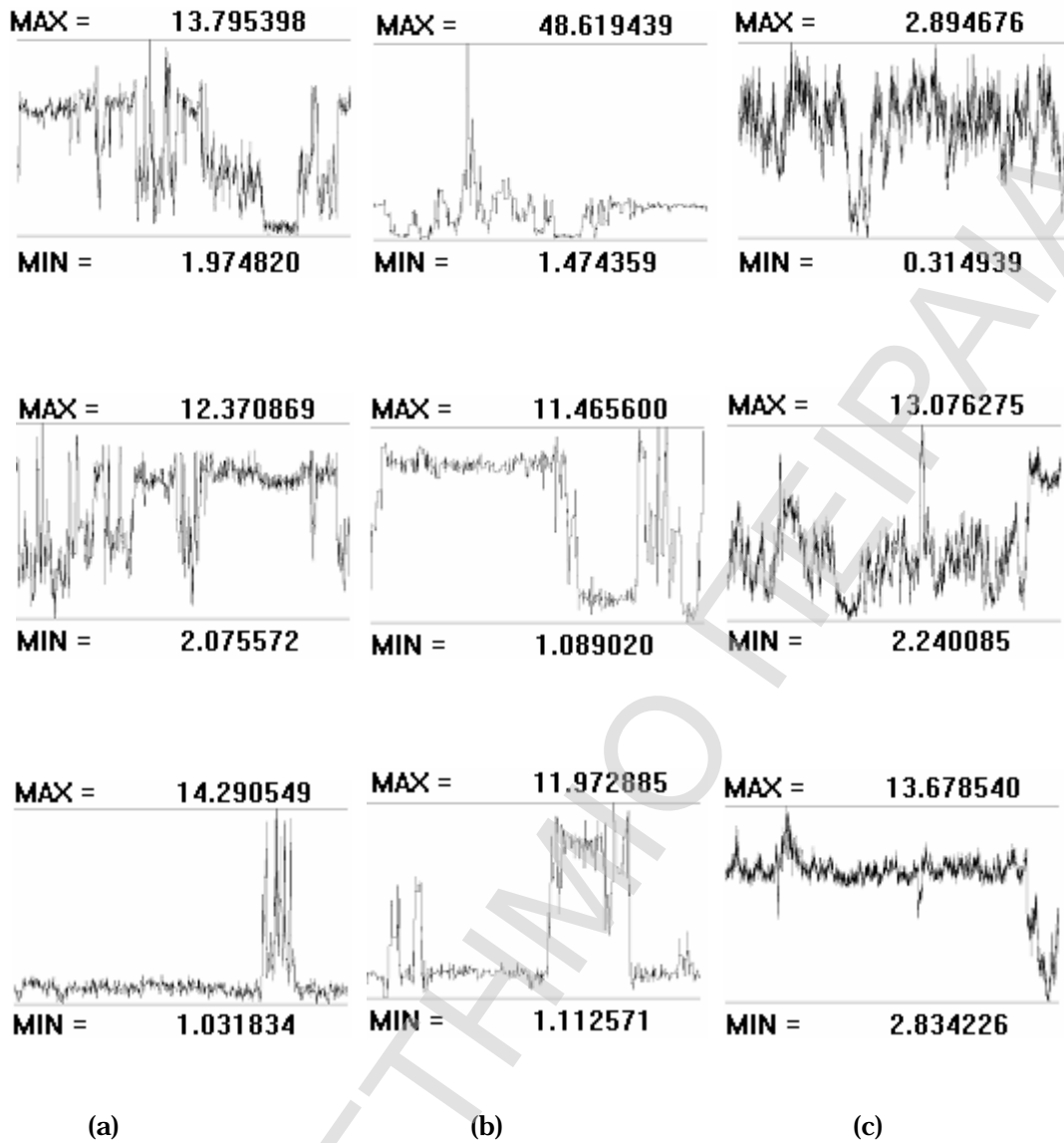
$$\begin{aligned}
 \hat{E}(p_1 | \underline{x}) &= \hat{p}_1 = 0.243366 \\
 \hat{E}(p_2 | \underline{x}) &= \hat{p}_2 = 0.241005 \\
 \hat{E}(p_3 | \underline{x}) &= \hat{p}_3 = 0.515629
 \end{aligned}$$

Οι εργοδικοί μέσοι φαίνονται στο επόμενο σχήμα.



ΣΧΗΜΑ 21. Τα γραφήματα των εργοδικών μέσων των λ_j (πρώτη γραμμή) και $p_j, j=1,2,3$ για 100000 επαναλήψεις του αλγόριθμου Metropolis – Hastings, με περίοδο burn in τις πρώτες 20000 επαναλήψεις και με την κλίμακα του τυχαίου περιπάτου ίση με $\zeta^2 = 0.05$.

ΠΑΡΑΤΗΡΗΣΗ: Σε αυτό το σημείο τονίζουμε ότι για να καταλήξουμε σε ικανοποιητικές εκτιμήσεις απαιτήθηκε πολύ μεγαλύτερος αριθμός επαναλήψεων (100000) σε σχέση με το αντίστοιχο πρόβλημα του δειγματολήπτη Gibbs (15000). Επί πλέον το φαινόμενο label switching δεν είχε έντονη παρουσία. Ίδια εικόνα εμφανίζεται και όταν επιλέξαμε $\zeta^2 = 0.01$ (Σχήμα 22.c). Εκτελέσαμε τον ίδιο αλγόριθμο αλλάζοντας την κλίμακα του τυχαίου περιπάτου σε $\zeta^2 = 0.1$ και οι εκτιμήσεις των παραμέτρων των κατανομών Poisson ήταν ικανοποιητικές ως έναν βαθμό, δυστυχώς όμως δεν μπορούμε να ισχυριστούμε το ίδιο και για τα βάρη των συνιστωσών. Αυτό συνέβη λόγω του υψηλού ρυθμού απορρίψεων που είχε σαν συνέπεια αυτή η επιλογή, το οποίο από την μία περιόρισε την επαρκή εξερεύνηση των περιοχών που επισκέφθηκε ο αλγόριθμος, από την άλλη όμως κινήθηκε σε όλες τις δυνατές περιοχές. Από το Σχήμα 22 παρατηρούμε την επίδραση της παραμέτρου κλίμακας του τυχαίου περιπάτου στην εξέλιξη των τιμών κατά την εκτέλεση του αλγορίθμου. Στην περίπτωση όπου η κλίμακα είναι μεγαλύτερη (Σχήμα 22.b) ο αλγόριθμος περνάει από όλες τις κορυφές, λόγω της έντονης παρουσίας του προβλήματος label switching, κάτι που δεν γίνεται στην περίπτωση που η παράμετρος κλίμακος είναι μικρή. Αυτό διαπιστώνεται στο Σχήμα 22.c όπου το φαινόμενο label switching δεν εμφανίζεται. Αυτό το γεγονός υπογραμμίζει το παράδοξο του label switching: όταν συμβαίνει, η συμπερασματολογία καθίσταται αρκετά δύσκολη, ενώ όταν δεν συμβαίνει (ή συμβαίνει σε μικρό βαθμό), η διαδικασία των εκτιμήσεων είναι θέμα ρουτίνας αλλά είναι βασισμένη σε έναν αλγόριθμο που - στην καλύτερη περίπτωση - συγκλίνει πολύ αργά (υπό την έννοια ότι δεν έχει εξαντλήσει όλες τις δυνατές περιοχές που μπορεί να επισκεφτεί)! Από αυτό λοιπόν το παράδειγμα διαπιστώνουμε την ευαισθησία του δειγματολήπτη Metropolis - Hastings στην επιλογή της κλίμακας του τυχαίου περιπάτου και την ανάγκη χρησιμοποίησης διαφορετικών τιμών για την ανίχνευση τοπικών και συνολικών χαρακτηριστικών.



ΣΧΗΜΑ 22. Τα γραφήματα των λ_j για 100000 επαναλήψεις του αλγόριθμου Metropolis – Hastings, μετά το burn in των πρώτων 20000 επαναλήψεων, με την παράμετρο κλίμακος του τυχαίου περιπάτου ίση με $\zeta^2 = 0.05$ (a), $\zeta^2 = 0.1$ (b) και $\zeta^2 = 0.01$ (c).

Παράδειγμα 2.2.3 Επανερχόμαστε στα δεδομένα των γαλαξιακών ταχυτήτων, όπου $\underline{\theta} = (\underline{\mu}, \underline{\sigma})$ και εξετάζουμε την απόδοση του αντίστοιχου αλγόριθμου Metropolis - Hastings, αλλά αυτή την φορά παρουσιάζουμε ένα εναλλακτικό μοντέλο (Casella et al., 2000), σύμφωνα με το οποίο οι εκ των προτέρων κατανομές των παραμέτρων είναι:

$$\sigma_j^{-2} \sim \Gamma(a_j, \beta_j), \mu_j | \sigma_j^2 \sim N(\lambda_j, \sigma_j^2 / \tau_j), \quad j = 1, \dots, k \quad (2.18)$$

όπου τα a_j, β_j, λ_j και τ_j είναι γνωστές σταθερές, ενώ η εκ των προτέρων κατανομή των ποσοστών της μείξης λαμβάνεται η μη πληροφοριακή Dirichlet(1,1,1).

Εφαρμόσαμε τον αλγόριθμο Metropolis - Hastings με τις εξής κατανομές πρότασης:

$$\tilde{\mu}_j \sim N(\mu_j^{(t-1)}, \zeta^2), \tilde{\sigma}_j^2 \sim LN(\log((\sigma_j^2)^{(t-1)}), \zeta^2), \tilde{\omega}_j \sim LN(\log(\omega_j^{(t-1)}), \zeta^2).$$

Αλγόριθμος Metropolis - Hastings για κανονικές μειξεις (μοντέλο 2.18)

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}^{(0)}, \underline{\omega}^{(0)}$.
1. Στην επανάληψη $t, t = 1, 2, \dots$:
 - 1.1. Προσομοίωσε τα $\tilde{\mu}_j \sim N(\mu_j^{(t-1)}, \zeta^2), j = 1, \dots, k$.
 - 1.2. Προσομοίωσε τα $\tilde{\sigma}_j^2 \sim LN(\log((\sigma_j^2)^{(t-1)}), \zeta^2), j = 1, \dots, k$.
 - 1.3. Προσομοίωσε τα $\tilde{\omega}_j \sim LN(\log(\omega_j^{(t-1)}), \zeta^2), j = 1, \dots, k$.
 - 1.4. Υπολόγισε το

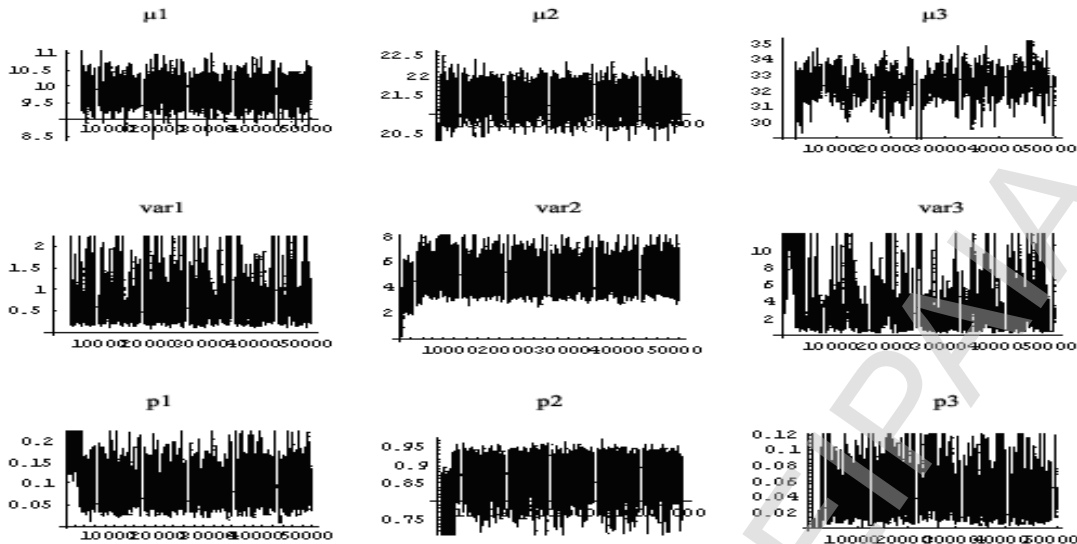
$$r = \frac{f(x|\tilde{\omega}, \tilde{\theta}) \pi(\tilde{\omega}, \tilde{\theta}) \prod_{j=1}^k \tilde{\sigma}_j^2 \tilde{\omega}_j}{f(x|\underline{\omega}^{(t-1)}, \underline{\theta}^{(t-1)}) \pi(\underline{\omega}^{(t-1)}, \underline{\theta}^{(t-1)}) \prod_{j=1}^k (\sigma_j^2)^{(t-1)} \omega_j^{(t-1)}}.$$

- 1.5. Προσομοίωσε έναν τυχαίο αριθμό $u \sim U(0,1)$.
 Εάν $u < r$ τότε θέσε $(\underline{\theta}^{(t)}, \underline{\omega}^{(t)}) = (\tilde{\theta}, \tilde{\omega})$
 αλλιώς θέσε $(\underline{\theta}^{(t)}, \underline{\omega}^{(t)}) = (\underline{\theta}^{(t-1)}, \underline{\omega}^{(t-1)})$.

Εφαρμόσαμε τον παραπάνω αλγόριθμο επιλέγοντας να θέσουμε διαφορετική εκ των προτέρων πληροφορία στους μέσους των τριών κανονικών κατανομών. Συγκεκριμένα θέσαμε $\lambda_1 = 12, \lambda_2 = 20$ και $\lambda_3 = 30$, ενώ όλες οι υπόλοιπες υπερπαραμέτροι ήταν ίσες μεταξύ τους. Όπως ήταν αναμενόμενο το φαινόμενο label switching δεν παρουσιάστηκε λόγω της ταυτοποίησης των παραμέτρων. Οι τελικές εκτιμήσεις που πήραμε είναι:

$\hat{E}(\mu_1 x) \approx 9.848$	$\hat{E}(\sigma_1^2 x) \approx 4.84384$	$\hat{E}(p_1 x) \approx 0.875538$
$\hat{E}(\mu_2 x) \approx 21.391$	$\hat{E}(\sigma_2^2 x) \approx 0.673615$	$\hat{E}(p_2 x) \approx 0.0869834$
$\hat{E}(\mu_3 x) \approx 32.3641$	$\hat{E}(\sigma_3^2 x) \approx 3.61005$	$\hat{E}(p_3 x) \approx 0.0374788$

Συγκρίνοντας τα παραπάνω αποτελέσματα με τα αντίστοιχα του Παραδείγματος 2.1.3 συμπεραίνουμε την συμφωνία των σε μεγάλο βαθμό. Η εναλλακτική προσέγγιση που επιλέξαμε σε αυτό το παράδειγμα αναδεικνύει την ανθεκτικότητα των εκτιμήσεων τόσο από πλευράς μοντελοποίησης (εκ των προτέρων πληροφορία και κατανομές, επιλογή υπερπαραμέτρων), όσο και από την μέθοδο που χρησιμοποιούμε (Gibbs sampler ή Metropolis - Hastings).



ΣΧΗΜΑ 23. Τα γραφήματα των $\underline{\mu}, \underline{\sigma}^2$ και \underline{p} για 50000 επαναλήψεις του αλγόριθμου Metropolis – Hastings, με την παράμετρο κλίμακος του τυχαίου περιπάτου ίση με $\zeta^2 = 0.05$.

2.3 Η μέθοδος Population Monte Carlo

Μία εναλλακτική προσέγγιση για την εκτίμηση των μοντέλων μείξεων είναι η μέθοδος *Population Monte Carlo* ή PMC (Carré et al., 2003). Η μεγάλη διαφορά με τις MCMC προσεγγίσεις είναι ότι αποφεύγεται η μελέτη των θεωρητικών ιδιοτήτων μιας και εγκαταλείπουν το σκηνικό των Μαρκοβιανών αλυσίδων. Βασίζονται σε προγενέστερες και απλούστερες τακτικές, όπως αυτή της *δειγματοληψίας σπουδαιότητας* (importance sampling). Η κύρια αρχή του PMC είναι η διαδοχική παραγωγή δειγμάτων σπουδαιότητας από συναρτήσεις σπουδαιότητας εξαρτώμενες από τα προηγούμενα δείγματα. Κάποια πλεονεκτήματα ως προς τις MCMC μεθόδους είναι η ασθενέστερη εξάρτηση από τις αρχικές τιμές, ενώ οι εκτιμητές που προκύπτουν είναι αμερόληπτοι σε κάθε επανάληψη, κάτι που σημαίνει ότι η διαδικασία μπορεί να τερματιστεί μετά από λίγες επαναλήψεις (χωρίς να λαμβάνουμε υπ' όψη κριτήρια σύγκλισης κτλ), και βέβαια όσο τις αυξάνουμε τόσο βελτιώνονται και τα δείγματα σπουδαιότητας.

Η βασική ιδέα: Το σημείο αναφοράς της μεθόδου PMC είναι η ακόλουθη γνωστή ιδιότητα από την δειγματοληψία σπουδαιότητας: Έστω ότι π είναι η εκ των υστέρων κατανομή (κατανομή στόχος) και ότι θέλουμε να εκτιμήσουμε την ποσότητα:

$$\mathfrak{I} = \int h(\underline{x})\pi(\underline{x})d\underline{x}$$

Έστω ότι στην επανάληψη t παράγουμε το δείγμα $\underline{y}^{(t)}$ μεγέθους M , με $y_i^{(t)} \sim q_{it}$ ανεξάρτητα μεταξύ τους δοθέντων των προηγούμενων δειγμάτων. Σε κάθε $y_i^{(t)}$ αντιστοιχίζουμε το λεγόμενο *λόγο σπουδαιότητας* (importance ratio):

$$\rho_i^{(t)} = \frac{\pi(y_i^{(t)})}{q_{it}(y_i^{(t)})}, i = 1, \dots, M \quad (2.19)$$

όπου π είναι η εκ των υστέρων κατανομή (κατανομή στόχος). Τότε οι εκτιμητές του \mathfrak{S} της μορφής:

$$\mathfrak{S}_t = \frac{1}{M} \sum_{i=1}^M \rho_i^{(t)} h(y_i^{(t)}) \quad (2.20)$$

είναι αμερόληπτοι για κάθε ολοκληρώσιμη συνάρτηση h και για κάθε επανάληψη t . Κατόπιν, παίρνουμε δείγμα M τιμών (με επανάθεση) από το $\underline{y}_{(t)}$, χρησιμοποιώντας⁴ τα βάρη $\omega_i^{(t)}$ που υπολογίζονται σταθμίζοντας τους λόγους σπουδαιότητας ώστε να αθροίζουν στη μονάδα, δηλαδή

$$\omega_i^{(t)} = \frac{\rho_i^{(t)}}{\sum_{i=1}^M \rho_i^{(t)}}, i = 1, \dots, M.$$

Το κομβικό σημείο για την καλή απόδοση του αλγορίθμου αυτού είναι βεβαίως η κατάλληλη επιλογή της κατανομής πρότασης. Αρκούμαστε να αναφέρουμε ότι η αμεροληψία των παραπάνω εκτιμητών ισχύει για όλες τις κατανομές q_{it} των οποίων το στήριγμα περιέχει αυτό της π .

Για να επανέλθουμε στα θέματα των μείξεων για το μοντέλο (1.2) η αντίστοιχη μορφή της (2.20) είναι η:

$$\mathfrak{S}_t = \frac{1}{M} \sum_{i=1}^M \frac{\rho_i^{(i)}}{\sum_{i=1}^M \rho_i^{(i)}} h(\underline{\theta}^{(i)}, \underline{p}^{(i)})$$

όπου

$$\rho_i^{(i)} = \frac{f(x|\underline{\theta}^{(i)}, \underline{p}_i^{(i)}) \pi(\underline{\theta}^{(i)}, \underline{p}^{(i)})}{q_{it}(\underline{\theta}^{(i)}, \underline{p}^{(i)})}, i = 1, \dots, M$$

(εδώ με π εννοούμε την εκ των προτέρων κατανομή των παραμέτρων). Ο παρακάτω ψευδο - κώδικας περιγράφει την γενική διαδικασία που ακολουθούμε για αυτήν την προσέγγιση.

⁴ Αυτό σημαίνει ότι κάθε $y_i^{(t)}$ επιλέγεται από το σύνολο $Y = \{y_1^{(t)}, \dots, y_M^{(t)}\}$ με πιθανότητα $\omega_i^{(t)}$, $i=1, \dots, M$. Έτσι παράγουμε έναν τυχαίο αριθμό u και αν $\sum_{j=1}^{i-1} \omega_j^{(t)} < u \leq \sum_{j=1}^i \omega_j^{(t)}$ τότε λαμβάνουμε το $y_i^{(t)}$ χωρίς να το εξαιρούμε από το σύνολο Y . Αυτό το επαναλαμβάνουμε μέχρις ότου επιλέξουμε M στοιχεία.

Γενικός αλγόριθμος Population Monte Carlo

0. Δώσε κάποιες αρχικές τιμές στο $\underline{\theta}_1^{(0)}, \dots, \underline{\theta}_M^{(0)}, \underline{p}_1^{(0)}, \dots, \underline{p}_M^{(0)}$.

1. Στην επανάληψη $t, t = 1, 2, \dots, T$:

1.1. Για $i = 1, \dots, M$:

1.1.1. Προσομοίωσε τα $(\underline{\theta}_i^{(t)}, \underline{p}_i^{(t)})$ από την q_{ii} .

1.1.2. Υπολόγισε το

$$\rho_i^{(t)} = \frac{f(\underline{x} | \underline{\theta}_i^{(t)}, \underline{p}_i^{(t)}) \pi(\underline{\theta}_i^{(t)}, \underline{p}_i^{(t)})}{q_{ii}(\underline{\theta}_i^{(t)}, \underline{p}_i^{(t)})}, i = 1, \dots, M.$$

1.2. Υπολόγισε το $\omega_i^{(t)} = \frac{\rho_i^{(t)}}{\sum_{i=1}^M \rho_i^{(t)}}$, $i = 1, \dots, M$.

1.3. Εκτέλεσε δειγματοληψία M τιμών με επανάθεση από τα $(\underline{\theta}_i^{(t)}, \underline{p}_i^{(t)})$, $i = 1, \dots, M$ χρησιμοποιώντας τα βάρη $\omega_i^{(t)}$.

Τέλος, σημειώνουμε ότι οι κατανομές πρότασης q_{ii} , μπορούν να επιλεγούν βάσει της απόδοσης των προηγούμενων $q_{i(t-1)}$, ή ακόμα και δοθέντων όλων των προηγούμενων δειγμάτων, εάν κάτι τέτοιο είναι εφικτό από άποψη υπολογιστικών δυνατοτήτων. Αυτά τα θέματα όμως γίνονται σαφέστερα στο παράδειγμα που ακολουθεί.

Παράδειγμα 2.3.1 Στην περίπτωση του μοντέλου $pN(\mu_1, 1) + (1-p)N(\mu_2, 1)$ (1.8) κατασκευάζουμε έναν δειγματολήπτη PMC, χωρίς το βήμα αύξησης των δεδομένων που απαιτεί ο δειγματολήπτης Gibbs, χρησιμοποιώντας έναν κανονικό τυχαίο περίπατο, ο οποίος βασίζεται στο προηγούμενο δείγμα των (μ_1, μ_2) , δηλαδή:

$$\mu_{1i}^{(t)} \sim N(\mu_{1i}^{(t-1)}, \zeta_l^2) \text{ και } \mu_{2i}^{(t)} \sim N(\mu_{2i}^{(t-1)}, \zeta_l^2),$$

όπου $l = 1, \dots, L$. Η δυσκολία με τους τυχαίους περιπάτους είναι πάντα ο καθορισμός μιας κατάλληλης κλίμακας (διασποράς στην συγκεκριμένη περίπτωση). Εδώ το πρόβλημα αυτό μπορεί να ξεπεραστεί θεωρώντας L διαφορετικές τιμές, και αυτό λόγω της ευχέρειας⁵ που δίνεται από την μέθοδο PMC. Εδώ τα πράγματα ευνοούν κάτι τέτοιο σκεφτόμενοι απλά ως εξής: σε κάθε επανάληψη του αλγορίθμου διαλέγουμε την νέα διασπορά ανάλογα με την απόδοση των ήδη διαθέσιμων κλιμάκων ζ_l^2 στις προηγούμενες επαναλήψεις. Ο πιο απλός τρόπος να κάνουμε κάτι τέτοιο είναι να επιλέγουμε κάθε κλίμακα με πιθανότητα ανάλογη του ρυθμού μη - εκφυλισμού (non - degeneracy rate) αυτής κατά την προηγούμενη επανάληψη. Η πιθανότητα αυτή είναι το ποσοστό των δειγματοσημείων που παρήχθησαν με

⁵ Κάτι τέτοιο θα μπορούσε να γίνει και στον αλγόριθμο Metropolis - Hastings, θεωρώντας την διασπορά του τυχαίου περιπάτου τυχαία μεταβλητή, το οποίο σημαίνει ότι θα έπρεπε να γίνει περαιτέρω μοντελοποίηση.

κλίμακα ζ_l^2 και επέζησαν μετά την επαναδειγματοληψία. Με λίγα λόγια θεωρούμε ότι οι δείκτες των κλιμάκων ακολουθούν μία πολυωνυμική κατανομή, τέτοια ώστε: $l \sim M(1; r_1, \dots, r_L)$, όπου $r_l = (\# \text{ των } x_i^{(t)} \text{ με διασπορά } \zeta_l^2 \text{ που επέζησαν})/L$. Όταν ο ρυθμός επιβίωσης είναι μηδέν τότε θέτουμε το αντίστοιχο r_l ίση με μία μικρή τιμή (π.χ. 1%), απλά και μόνο για να αποφύγουμε την κατάργηση μιας δοθείσης κλίμακας ζ_l^2 . Ο αλγόριθμος για την εκτέλεση της μεθόδου αυτής έχει ως εξής:

Αλγόριθμος Population Monte Carlo για το μοντέλο (1.8)

0. Δώσε κάποιες αρχικές τιμές στο $\mu_{11}^{(0)}, \dots, \mu_{1M}^{(0)}, \mu_{21}^{(0)}, \dots, \mu_{2M}^{(0)}$.
1. Στην επανάληψη $t, t = 1, 2, \dots, T$:
 - 1.1. Για $i = 1, \dots, M$:
 - 1.1.1. Προσομοίωσε το l από την $M(1; r_1, \dots, r_L)$.
 - 1.1.2. Προσομοίωσε τα $\mu_{j_i}^{(t)}, j = 1, 2$, από την $N(\mu_{j_i}^{(t-1)}, \zeta_l^2)$.
 - 1.1.3. Υπολόγισε το

$$\rho_i^{(t)} = \frac{f(x | \mu_{1_i}^{(t)}, \mu_{2_i}^{(t)}) \pi(\mu_{1_i}^{(t)}, \mu_{2_i}^{(t)})}{\sum_{l=1}^L \sum_{j=1}^2 \varphi(\mu_{j_i}^{(t)}; \mu_{j_i}^{(t-1)}, \zeta_l^2)}, i = 1, \dots, M$$

- 1.2. Υπολόγισε το $\omega_i^{(t)} = \frac{\rho_i^{(t)}}{\sum_{i=1}^M \rho_i^{(t)}}, i = 1, \dots, M$.

- 1.3. Εκτέλεσε δειγματοληψία M τιμών με επανάθεση από τα $\mu_{1_i}^{(t)}, \mu_{2_i}^{(t)}, i = 1, \dots, M$ χρησιμοποιώντας τα βάρη $\omega_i^{(t)}$.

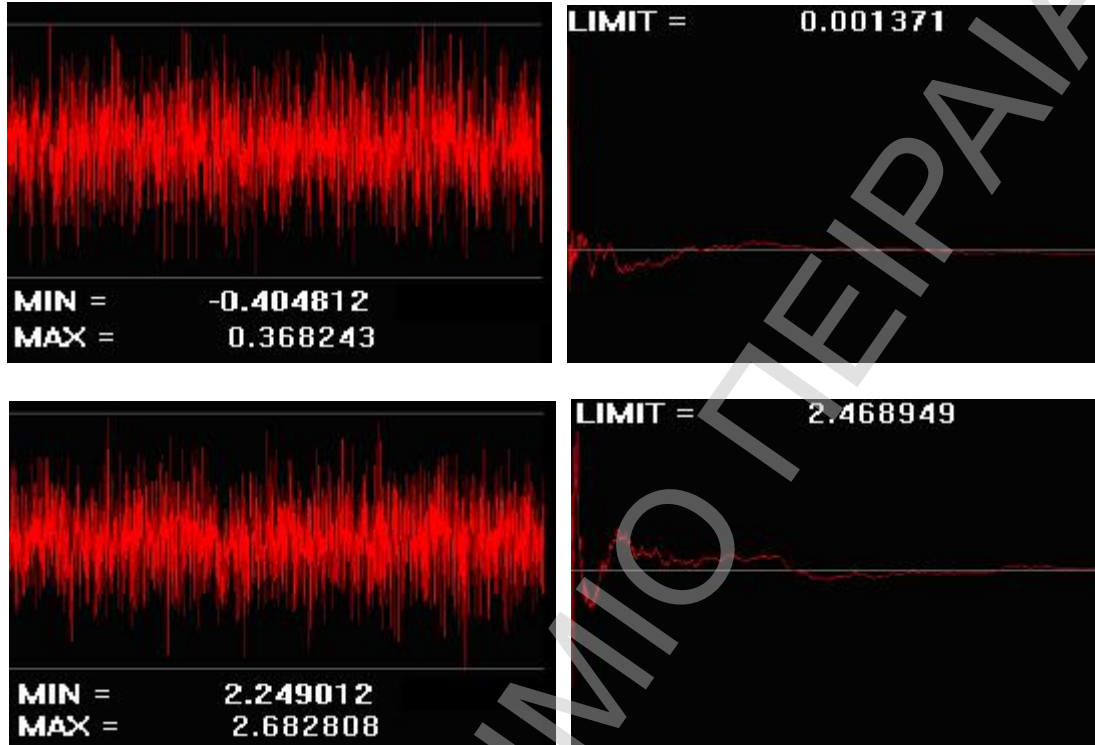
- 1.4. Ανανέωσε τις παραμέτρους της πολυωνυμικής κατανομής: το r_l είναι ανάλογο του ποσοστού των $x_i^{(t)}$ με διασπορά ζ_l^2 που επέζησαν.

Στο παραπάνω συμβολίζουμε με $\varphi(\mu_{j_i}^{(t)}; \mu_{j_i}^{(t-1)}, \zeta_l^2)$ την τιμή της συνάρτησης πυκνότητας της κατανομής $N(\mu_{j_i}^{(t-1)}, \zeta_l^2)$ στο σημείο $\mu_{j_i}^{(t)}$. Για την εκτέλεση του Βήματος 1.3 παραπέμπουμε στην υποσημείωση της προηγούμενης σελίδας.

Στο Σχήμα 23 αναπαρίσταται η συμπεριφορά των τιμών και των αντίστοιχων μέσων του δείγματος σπουδαιότητας ($M = 3000$) μετά από 60 δειγματοληψίες ($T = 60$) για το μοντέλο (1.8). Εκτελέσαμε τον παραπάνω αλγόριθμο για 6 διαφορετικές κλίμακες: $\zeta_1^2 = 1, \zeta_2^2 = 0.5, \zeta_3^2 = 0.1, \zeta_4^2 = 0.01, \zeta_5^2 = 5$ και $\zeta_6^2 = 25$. Μετά τις πρώτες 4 επαναλήψεις οι ρυθμοί επιβίωσης των $\zeta_1^2, \zeta_2^2, \zeta_3^2, \zeta_5^2, \zeta_6^2$ είναι όλοι μικρότεροι του

4%. Στο τέλος των 60 επαναλήψεων η ζ_4^2 έχει μέση τιμή επιβίωσης περίπου ίση με 0.899. Τέλος οι εκτιμήσεις που παίρνουμε είναι

για το $E(\mu_1|x)$: $\hat{m}_1 = 0.001371$ και για το $E(\mu_2|x)$: $\hat{m}_2 = 2.468949$.



ΣΧΗΜΑ 23. Εκτέλεση του αλγορίθμου PMC για το μοντέλο (1.8), μετά από 60 επαναλήψεις και δείγμα σπουδαιότητας μεγέθους 3000 τιμών. Στη πρώτη γραμμή φαίνονται οι τιμές και οι μέσοι του δείγματος σπουδαιότητας $\underline{\mu}_1^{(60)}$ και στην δεύτερη τα αντίστοιχα για το $\underline{\mu}_2^{(60)}$.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΚΕΦΑΛΑΙΟ 3

ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΜΕΙΞΕΩΝ ΜΕ ΑΓΝΩΣΤΟ ΠΛΗΘΟΣ ΣΥΝΙΣΤΩΣΩΝ

*Η κινητήρια δύναμη της
Μαθηματικής ανακάλυψης
δεν είναι η λογική αλλά
η φαντασία.*
A. De Morgan

3. Συμπερασματολογία για μοντέλα μείξεων με άγνωστο πλήθος συνιστωσών

Σε αυτό το κεφάλαιο θεωρούμε ότι ο αριθμός k των συνιστωσών του μοντέλου $\sum_{i=1}^k p_i f_i(x|\theta_i)$ είναι άγνωστος. Μία άμεση συνέπεια αυτού είναι ότι «ο αριθμός των πραγμάτων που δεν γνωρίζουμε, είναι κάτι που δεν γνωρίζουμε». Η εκτίμηση του k αποτελεί ειδική περίπτωση του προβλήματος επιλογής μοντέλου, για το οποίο προτείνονται οι εξής δυνατές λύσεις:

- (i) Παράγοντας Bayes (Kass & Raftery, 1995, Richardson & Green, 1997).
- (ii) Απόσταση εντροπίας ή απόκλιση Kullback - Leibler (Mengersen & Robert, 1996, Sahu & Cheng, 2003).
- (iii) MCMC αντιστρέψιμης μετάβασης (reversible jump) (Richardson & Green, 1997, Gruet et al., 1999).
- (iv) Διαδικασίες γεννήσεων - θανάτων (Stephens 2000, Cappé et al. 2002).

Οι προσεγγίσεις αυτές εξαρτώνται από το αν ο σκοπός της ανάλυσης είναι έλεγχος ή εκτίμηση. Εμείς θα εστιάσουμε στον δεύτερο (μέθοδοι (iii) και (iv)), διότι αναδεικνύει την Μπεϋζιανή προσέγγιση και προσφέρει ένα ευρύτερο πλαίσιο συμπερασματολογίας.

Όταν ο αριθμός των συνιστωσών είναι άγνωστος, πρέπει να θεωρήσουμε επιμέρους μοντέλα M_k , με αντίστοιχους παραμετρικούς χώρους Ω_k . Έτσι αντιμετωπίζουμε μία συλλογή μοντέλων με πιθανώς μη φραγμένους παραμετρικούς χώρους (και φυσικά τις αντίστοιχες εκ των προτέρων κατανομές σε αυτούς τους χώρους), με συνέπεια η υπολογιστική πρόκληση να είναι πολύ μεγαλύτερη από αυτή του προηγούμενου κεφαλαίου.

Η μέθοδος MCMC που προτάθηκε από τον Green (1995) καλείται Reversible Jump MCMC (RJMCMC), επειδή βασίζεται σε έναν περιορισμό αντιστρεψιμότητας κατά την κίνηση αλλαγής διάστασης της Μαρκοβιανής αλυσίδας που προσομοιώνεται, η οποία γεφυρώνει τα σύνολα Ω_k . Συγκεκριμένα, η κύρια δυσκολία σε σχέση με τις προηγούμενες μεθόδους είναι ο ορισμός κινήσεων (ή μεταβάσεων) μεταξύ των Ω_k , μια και οι προτεινόμενες κινήσεις για δοθέν Ω_k υπακούουν στην συνήθη θεωρία του προηγούμενου κεφαλαίου.

Η μέθοδος Birth - Death MCMC (BDMCMC) του Stephens (2000) θεωρεί τις παραμέτρους του μοντέλου ως μία μαρκαρισμένη σημειακή διαδικασία (Ripley, 1987) και επιτρέπει τη γέννηση νέων και τον θάνατο υπαρχόντων συνιστωσών σε συνεχή χρόνο σύμφωνα με κατάλληλα ορισμένους ρυθμούς γεννήσεων και θανάτων αντίστοιχα. Έτσι η στάσιμη κατανομή της διαδικασίας είναι η εκ των υστέρων κατανομή του μοντέλου. Εδώ, οι μεταβάσεις ανάμεσα σε διαφορετικές διαστάσεις της διαδικασίας γίνονται πάντα αποδεκτές εφόσον η συνηθισμένη πιθανότητα αποδοχής που περιλαμβάνουν οι μέθοδοι MCMC αντικαθίσταται από το χρονικό διάστημα που η διαδικασία βρίσκεται σε κάθε κατάσταση.

3.1 Reversible Jump MCMC

Έστω $\pi(dy)$ η κατανομή στόχος (π.χ. η εκ των υστέρων κατανομή), των παραμέτρων δοθέντων των δεδομένων. Στις παραμέτρους συμπεριλαμβάνουμε και τον αριθμό των συνιστωσών k (μαζί με το $\theta^{(k)}$). Στις μεθόδους MCMC κατασκευάζουμε μία Μαρκοβιανή αλυσίδα, ορίζοντας έναν πυρήνα μετάβασης $P(y, dy')$ που είναι απεριοδικός και ανάγωγος (ως προς ένα μέτρο φ), που επιτυγχάνει την στάσιμη κατανομή της αλυσίδας π , για την οποία: $\pi(A) = \int \pi(dy)P(y, A)$. Αν επί πλέον η αλυσίδα έχει την ιδιότητα της αντιστρεψιμότητας ως προς την π τότε ισχύει η *συνθήκη λεπτομερούς ισορροπίας* (detailed balance condition)

$$\iint_{AB} \pi(dy)P(y, dy') = \iint_{BA} \pi(dy')P(y', dy) \quad (3.1)$$

για όλα τα A, B και προσομοιώνουμε αυτή την αλυσίδα ώστε να παρατηρήσουμε ένα εξαρτημένο δείγμα (προσεγγιστικά) από την $\pi(dy)$.

Στις απλές περιπτώσεις, η $\pi(dy)$, είτε είναι μία διακριτή συνάρτηση κατανομής, είτε έχει μία πυκνότητα αναφορικά με κάποιο μέτρο, συνήθως το Lebesgue, και οι πιο γνωστοί τρόποι κατασκευής του $P(y, dy')$, είναι ο δειγματολήπτης Gibbs και η μέθοδος Metropolis - Hastings, με τις οποίες ασχοληθήκαμε στο προηγούμενο Κεφάλαιο.

Ο δειγματολήπτης Gibbs δεν έχει νόημα όταν το y έχει μη σταθερή διάσταση. Για αυτό επικεντρωνόμαστε στην κατηγορία των αλγορίθμων Metropolis - Hastings για την κατασκευή του πλαισίου αντιμετώπισης της περίπτωσης όπου δεν υπάρχει κάποιο απλό μέτρο. Η μέθοδος Reversible Jump MCMC αποτελεί μία γενίκευση του αλγορίθμου Metropolis - Hastings για την προσομοίωση μίας Μαρκοβιανής αλυσίδας όπου ο χώρος καταστάσεων αυτής δεν έχει σταθερή διάσταση.

3.1.1 Η γενική περίπτωση

Σε μία τυπική εφαρμογή με διαφορετικούς χώρους $\{\Omega_k\}$ διαφορετικών διαστάσεων είναι απαραίτητο να ορίσουμε εναλλακτικούς τύπους κινήσεων μεταξύ αυτών (move types). Αυτό επιτυγχάνεται μέσω τυχαίας επιλογής των τύπων κινήσεων σε κάθε μετάβαση, ώστε να διασχίζουμε ελεύθερα τον συνδυασμένο παραμετρικό χώρο $\Omega = \bigcup_{k \in \mathbb{K}} \Omega_k$. Δίνουμε βάση σε Μαρκοβιανές αλυσίδες των οποίων η στάσιμη κατανομή επιτυγχάνεται για κάθε τύπο κινήσεων.

Όταν είμαστε στην κατάσταση y , προτείνουμε μια κίνηση τύπου m που θα μας πάει στην κατάσταση dy' με πιθανότητα $q_m(y, dy')$. Προς το παρόν αυτό είναι ένα αυθαίρετο μέτρο στο m και y (όχι απαραίτητα πιθανότητας αλλά υπό-πιθανότητας). Αυτό σημαίνει ότι $\sum_m q_m(y, \Omega) \leq 1$. Επίσης με πιθανότητα $1 - \sum_m q_m(y, \Omega)$ παραμένουμε απλώς στην ίδια κατάσταση. Τέλος δεν είναι δυνατοί όλοι οι τύποι κινήσεων για την αρχική κατάσταση y , (για κάθε y), έτσι για κάποια m θα ισχύει $q_m(y, \Omega) = 0$.

Όπως στους αλγορίθμους Metropolis - Hastings η προτεινόμενη κίνηση δεν γίνεται αμέσως αποδεκτή, οπότε συμβολίζουμε με $\alpha_m(y, y')$ την πιθανότητα

αποδοχής. Στην ανάλυση που ακολουθεί σκοπός μας είναι να ορίσουμε τον ακριβή τύπο αυτής της πιθανότητας, ώστε για κάθε τύπο κίνησης να επιτυγχάνεται η συνθήκη της ισορροπίας.

Ο πυρήνας μετάβασης γράφεται ως εξής:

$$P(y, B) = \sum_m \int_B q_m(y, dy') \alpha_m(y, y') + s(y) I(y \in B) \quad (3.2)$$

για κάθε σύνολο Borel B στον Ω και όπου με I συμβολίζουμε την δείκτρια συνάρτηση. Επίσης η πιθανότητα παραμονής στην ίδια κατάσταση $s(y)$ θα ισούται με το άθροισμα της πιθανότητας μη αποδοχής της προτεινόμενης κίνησης και της πιθανότητας καμμίας κίνησης, όποτε:

$$s(y) = \left\{ \sum_m \int_{\Omega} q_m(y, dy') (1 - \alpha_m(y, y')) \right\} + \left\{ 1 - \sum_m q_m(y, \Omega) \right\}.$$

Από (3.1) και (3.2) λαμβάνουμε:

$$\begin{aligned} \sum_m \int_A \pi(dy) \int_B q_m(y, dy') \alpha_m(y, y') + \int_{A \cap B} \pi(dy) s(y) &= \\ &= \sum_m \int_B \pi(dy') \int_A q_m(y', dy) \alpha_m(y', y) + \int_{B \cap A} \pi(dy') s(y'). \end{aligned} \quad (3.3)$$

Όποτε μια ικανή και όχι αναγκαία συνθήκη για να ισχύει το παραπάνω είναι:

$$\int_A \pi(dy) \int_B q_m(y, dy') \alpha_m(y, y') = \int_B \pi(dy') \int_A q_m(y', dy) \alpha_m(y', y) \quad \forall m, A, B$$

Πριν ορίσουμε την πιθανότητα αποδοχής, διατυπώνουμε την υπόθεση του Green που οδηγεί στον ορισμό του κατάλληλου μέτρου:

DIMENSION MATCHING ASSUMPTION:

Έστω ότι το $\pi(dy) q_m(y, dy')$ έχει πεπερασμένη πυκνότητα $f_m(y, y')$ αναφορικά με ένα συμμετρικό μέτρο ξ_m στον $\Omega \times \Omega$.

Τότε μπορούμε να γράψουμε ότι:

$$\begin{aligned} \int_A \pi(dy) \int_B q_m(y, dy') \alpha_m(y, y') &= \iint_{AB} \xi_m(y, dy') f_m(y, y') \alpha_m(y, y') = \\ &= \iint_{BA} \xi_m(y', dy) f_m(y', y) \alpha_m(y', y) = \int_B \pi(dy') \int_A q_m(y', dy) \alpha_m(y', y) \end{aligned} \quad (3.4)$$

η οποία - ορίζοντας κατάλληλα το $\alpha_m(y, y')$ - σύμφωνα με τα παραπάνω επιτυγχάνει την στάσιμη κατανομή. Σημειώνουμε ότι η μετάβαση από την δεύτερη ισότητα στην τρίτη ισχύει λόγω της υποτιθέμενης συμμετρίας του ξ_m . Όποτε για να ισχύει η (3.4) αρκεί να ισχύει η:

$$f_m(y, y') \alpha_m(y, y') = f_m(y', y) \alpha_m(y', y). \quad (3.5)$$

Αλλά αν θέσουμε ότι:

$$\alpha_m(y, y') = \min \left\{ 1, \frac{f_m(y', y)}{f_m(y, y')} \right\}$$

παρατηρούμε ότι ισχύει η (3.5). Ισοδύναμα γράφουμε

$$\alpha_m(y, y') = \min \left\{ 1, \frac{\pi(dy')q_m(y', dy')}{\pi(dy)q_m(y, dy')} \right\}. \quad (3.6)$$

Μιλώντας αυστηρά το παραπάνω πρόκειται για λόγο παραγώγων Radon - Nikodym ως προς ένα κατάλληλα επιλεγμένο κοινό μέτρο. Η ύπαρξη αυτού του μέτρου εξασφαλίζεται από την dimension matching assumption, κάτι που θα καταστεί σαφές στην συνέχεια. Πριν όμως από αυτό θα δώσουμε κάποια παραδείγματα για μια διαισθητική κατανόηση της υπόθεσης του Green:

Έστω ότι υπάρχουν μόνο δύο υπόχωροι $\Omega_1 = \{1\} \times \mathfrak{X}$ και $\Omega_2 = \{2\} \times \mathfrak{X}^2$, όπου δοθέντος ότι $k=1$ και 2 η π έχει κατάλληλες πυκνότητες στον \mathfrak{X} και \mathfrak{X}^2 αντίστοιχα. Ας θεωρήσουμε την εξής κατάσταση: από το σημείο $(2, \theta_1, \theta_2) \in \Omega_2$ μια καλή κίνηση θα μπορούσε να είναι η μετάβαση στο $\left(1, \frac{\theta_1 + \theta_2}{2}\right) \in \Omega_1$. Για την κίνηση

αυτού του τύπου η από κοινού πιθανότητα στην κατάσταση ισορροπίας $\int_B \pi(dy) \int_A q_m(y, dy')$ όπου $A \subset \Omega_1$ και $B \subset \Omega_2$, πρέπει να έχει πυκνότητα αναφορικά

με ένα μέτρο στον $\mathfrak{X} \times \mathfrak{X}^2$ το οποίο συγκεντρώνει όλη του τη μάζα στο $\left\{(\theta, \theta_1, \theta_2) : \theta = \frac{\theta_1 + \theta_2}{2}\right\}$ αντί για το σύνθετες μέτρο Lebesgue στον \mathfrak{X}^3 . Για να

επιτευχθεί η συνθήκη λεπτομερούς ισορροπίας απαιτείται η αντίστροφη κίνηση (από το A στο B) να οριστεί μέσω μιας κατανομής πρότασης $q_m(y, dy')$, τέτοια ώστε $\forall y = (1, \theta) \in \Omega_1$ να είναι μοναδική, δηλαδή όλη η πιθανότητα να βρίσκεται στο

$\left\{(2, \theta_1, \theta_2) : \theta = \frac{\theta_1 + \theta_2}{2}\right\}$. Για παράδειγμα, αυτό μπορεί να γίνει, παράγοντας μια επι

πλέον τυχαία μεταβλητή u από κάποια αυθαίρετη κατανομή και να θέσουμε $\theta_1 = \theta + u$ και $\theta_2 = \theta - u$. Έτσι, αυτό που κάνει η υπόθεση, είναι να εξασφαλίζει ότι οι απαιτήσεις μοναδικότητας του παραπάνω τύπου είναι συνεπείς με τον ορισμό τους.

Έστω τώρα ότι υπάρχουν δύο υπόχωροι για $k=1,2$, και ότι οι $p(\theta^{(1)}|k=1)$ και $p(\theta^{(2)}|k=2)$ είναι κατάλληλες πυκνότητες στον \mathfrak{X}^{n_1} και \mathfrak{X}^{n_2} αντίστοιχα. Θεωρούμε μόνο ένα τύπο κινήσεων, όπου πάντα εναλλάσσει υποχώρους, οπότε $q(y, \Omega_1) = 0$ όταν $y \in \Omega_1$ και $q(y, \Omega_2) = 0$ όταν $y \in \Omega_2$. Συμβολίζουμε με $r(y)$ την πιθανότητα επιλογής αυτής της κίνησης. Ένας τυπικός τρόπος πραγματοποίησης μιας μετάβασης από τον Ω_1 στον Ω_2 , είναι να παράγουμε ένα διάνυσμα τυχαίων μεταβλητών $u^{(1)}$ διάστασης m_1 , ανεξάρτητα από το $\theta^{(1)}$, και στην συνέχεια να θέσουμε το $\theta^{(2)}$, ως μια ντετερμινιστική συνάρτηση των $\theta^{(1)}$, $u^{(1)}$: $\theta^{(2)} = \theta^{(2)}(\theta^{(1)}, u^{(1)})$. Ομοίως για την αντίστροφη κίνηση παράγουμε ένα $u^{(2)}$ διάστασης m_2 , ανεξάρτητα από το $\theta^{(2)}$, και

στην συνέχεια θέτουμε το $\theta^{(1)}$, ως μια ντετερμινιστική συνάρτηση των $\theta^{(2)}, u^{(2)}$. Η υπόθεση απαιτεί να υπάρχει ισότητα των διαστάσεων του $(\theta^{(1)}, u^{(1)})$ και του $(\theta^{(2)}, u^{(2)})$. Ειδικότερα θα πρέπει:

$$n_1 + m_1 = n_2 + m_2.$$

Με απλά λόγια η υπόθεση του Green εξασφαλίζει ότι οι από κοινού κατανομές $\pi(dy)q_m(y, dy')$ και $\pi(dy')q_m(y', dy)$ έχουν πυκνότητες σε χώρους ίσης διάστασης!

Η κατανομή πρότασης $q_m(y, dy')$ μπορεί να οριστεί από τις κατανομές των $u^{(1)}$ και $u^{(2)}$, όπου υποθέτουμε ότι δίνονται από κατάλληλες πυκνότητες q_1 και q_2 στον \mathfrak{R}^{m_1} και \mathfrak{R}^{m_2} αντίστοιχα. Για $A \subset \Omega_1$ και $B \subset \Omega_2$, θέτουμε:

$$\xi(A \times B) = \xi(B \times A) = \lambda\left(\left(\theta^{(1)}, u^{(1)}\right) : \theta^{(1)} \in A, \theta^{(2)}\left(\theta^{(1)}, u^{(1)}\right) \in B\right)$$

όπου το λ είναι το $(n_1 + m_1)$ - διάστατο μέτρο Lebesgue. Γενικώς αν $A, B \subset \Omega$ θέτουμε:

$$\xi(A \times B) = \xi((A \cap \Omega_1) \times (B \cap \Omega_2)) + \xi((A \cap \Omega_2) \times (B \cap \Omega_1))$$

το οποίο είναι συμμετρικό όπως απαιτείται.

Οπότε αν $y = (1, \theta^{(1)}) \in \Omega_1$ και $y' = (2, \theta^{(2)}) \in \Omega_2$, έστω:

$$f(y, y') = p\left(\left(1, \theta^{(1)}\right) \Big|_x\right) r\left(1, \theta^{(1)}\right) q_1\left(u^{(1)}\right)$$

$$f(y', y) = p\left(\left(2, \theta^{(2)}\right) \Big|_x\right) r\left(2, \theta^{(2)}\right) q_2\left(u^{(2)}\right) \left| \frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right|$$

και $f(y, y') = 0$ διαφορετικά. Τότε $\forall y, y' \in \Omega$ η $f(y, y')$ είναι η από κοινού πυκνότητα (ως προς το μέτρο ξ) της από κοινού κατανομής πρότασης στην ισορροπία: $\pi(dy)q_m(y, dy')$.

Σύμφωνα με την (3.6), η κατάλληλη πιθανότητα αποδοχής για την προτεινόμενη μετάβαση από το $y = (1, \theta^{(1)}) \in \Omega_1$ στο $y' = (2, \theta^{(2)}) \in \Omega_2$ είναι:

$$\alpha_m(y, y') = \min \left\{ 1, \frac{p\left(\left(2, \theta^{(2)}\right) \Big|_x\right) r\left(2, \theta^{(2)}\right) q_2\left(u^{(2)}\right) \left| \frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right|}{p\left(\left(1, \theta^{(1)}\right) \Big|_x\right) r\left(1, \theta^{(1)}\right) q_1\left(u^{(1)}\right)} \right\}$$

Στην πράξη θεωρούμε κινήσεις στις οποίες το m_1 ή το m_2 ισούται με το 0. Έτσι στην μία κατεύθυνση δεν χρειάζεται να παράγουμε τα αντίστοιχα $u^{(i)}$ και τότε η πιθανότητα αποδοχής απλοποιείται. Για παράδειγμα όταν $m_2 = 0$:

$$\alpha_m(y, y') = \min \left\{ 1, \frac{p\left(\left(2, \theta^{(2)}\right) \Big|_x\right) r\left(2, \theta^{(2)}\right)}{p\left(\left(1, \theta^{(1)}\right) \Big|_x\right) r\left(1, \theta^{(1)}\right) q_1\left(u^{(1)}\right) \left| \frac{\partial \theta^{(2)}}{\partial(\theta^{(1)}, u^{(1)})} \right|} \right\}. \quad (3.7)$$

Τέλος σημειώνουμε ότι η ορίζουσα προκύπτει από την αλλαγή μεταβλητής $\theta^{(1)} \rightarrow \theta^{(2)}(\theta^{(1)}, u^{(1)})$.

3.1.2 Εφαρμογή RJMCMC σε κανονικές μείξεις

Στην παράγραφο αυτή παρουσιάζεται αναλυτικά ο τρόπος εφαρμογής του αλγόριθμου reversible jump στα δεδομένα των ταχυτήτων 82 γαλαξιών. Για τον σκοπό αυτό οι Richardson & Green θεώρησαν το μοντέλο $X|\underline{p}, \underline{\theta}, k \sim \sum_{i=1}^k p_i f_i(x|\theta_i)$ όπου:

Η παράμετρος θ είναι το διάνυσμα $(\mu_j, \sigma_j^2), j = 1, 2, \dots, k$ έτσι ώστε

$$f(x|\theta_j) = f(x|\mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right\}.$$

Εκ των προτέρων έχουμε ότι τα μ_j, σ_j^2 είναι ανεξάρτητα με εκ των προτέρων κατανομές:

$$\mu_j \sim N(\xi, \kappa^{-1}) \text{ και } \sigma_j^2 \sim \Gamma(\alpha, \beta).$$

Οι επιλογές αυτές παρέχουν κάποια πλεονεκτήματα μιας και προκύπτουν εκ των υστέρων κατανομές που ανήκουν στην οικογένεια των συζυγών κατανομών των εκ των προτέρων, αν και δεν πρόκειται για την «φυσική συζυγή» κατά την οποία οι παράμετροι σε κάθε ζευγάρι είναι εξαρτημένες εκ των προτέρων. Οι Richardson & Green ακολουθούν μια εμπειρική ανάλυση καθώς το ξ λαμβάνεται ίσο με την διάμεσο των παρατηρήσεων, ενώ $\kappa^{-1} = R^2$ όπου R είναι το δειγματικό εύρος. Επίσης θέτουν ένα επιπλέον ιεραρχικό επίπεδο στην ανάλυση επιτρέποντας στο β να ακολουθεί την κατανομή $\Gamma(g, h)$, όπου θεωρούν $\alpha > 1 > g$ για να εκφράσουν την πεποίθηση ότι τα σ_j^2 είναι παρόμοια χωρίς να δίνεται πληροφορία για το μέγεθός τους. Η παράμετρος κλίμακος h θα είναι ένα μικρό πολλαπλάσιο του $1/R^2$.

Όπως συνήθως η εκ των προτέρων κατανομή του διανύσματος $(p_1, p_2, \dots, p_{k-1})$ των βαρών των συνιστωσών είναι η Dirichlet $D(\delta, \delta, \dots, \delta)$, όπου το πλήθος των παραμέτρων $(\delta, \delta, \dots, \delta)$ είναι ίσο με k . Επίσης είναι απαραίτητο να ορίσουμε μια κατάλληλη εκ των προτέρων κατανομή για τον αριθμό των συνιστωσών της μείξης k . Μια συνήθης επιλογή είναι η $P(I)$ όμως, για λόγους ευκολίας στην παρουσίαση και ερμηνεία, για το σκοπό αυτό θα χρησιμοποιηθεί επίσης η διακριτή ομοιόμορφη κατανομή μεταξύ του 1 και ενός καθορισμένου ακεραίου k_{\max} .

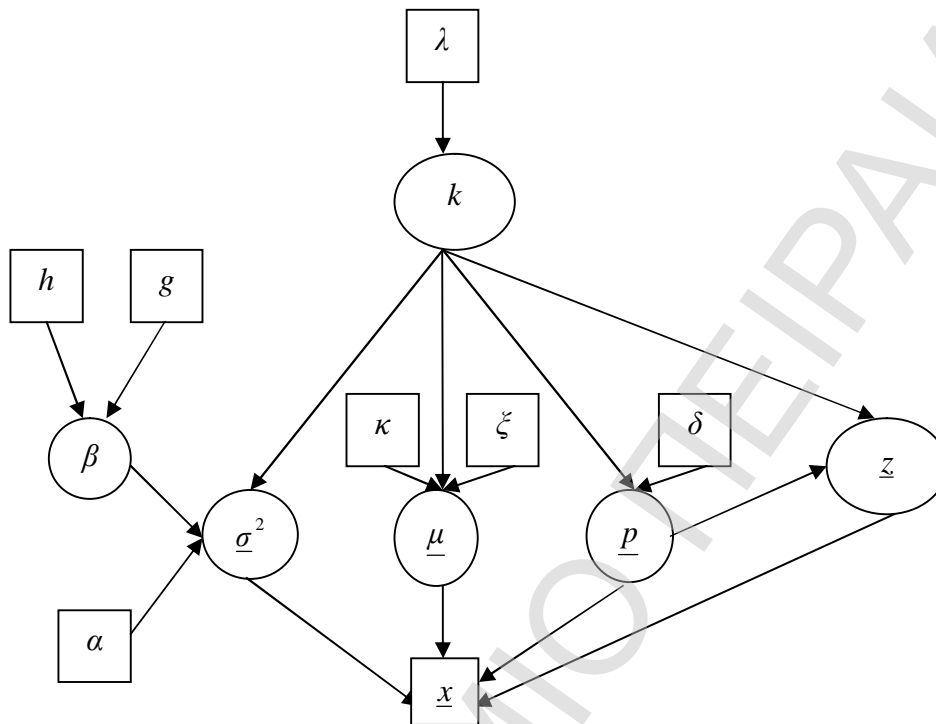
Τέλος για το διάνυσμα z όπου z_j είναι η θέση της j παρατήρησης έχουμε ότι:

$$p(z_i = j|\underline{p}, k) = p_j, \quad j = 1, \dots, k \quad i = 1, \dots, n.$$

Συμβολίζοντας με η το διάνυσμα $(\xi, \kappa, \alpha, \beta)$, έχουμε ότι η από κοινού κατανομή όλων των μεταβλητών μπορεί να γραφεί ως:

$$p(\eta, k, \underline{p}, \underline{z}, \underline{x}|\lambda, \delta) = p(\eta)p(k|\lambda)p(\underline{p}|k, \delta)p(\underline{z}|\underline{p}, k)p(\underline{\theta}|k, \eta)p(\underline{x}|\underline{\theta}, \underline{z}). \quad (3.8)$$

Το μοντέλο αυτό μπορεί να παρουσιαστεί και με ένα κατευθυνόμενο ακυκλικό γράφημα, στο οποίο τα τετράγωνα αναπαριστούν σταθερές ή παρατηρούμενες ποσότητες και οι κύκλοι άγνωστες, ως εξής:



ΣΧΗΜΑ 24. Αναπαράσταση του ιεραρχικού μοντέλου (3.8) ως ένα κατευθυνόμενο ακυκλικό γράφημα.

Πριν προχωρήσουμε, αναφερόμαστε στο σημαντικό ζήτημα της ταυτοποίησης των συνιστωσών της μείξης. Επειδή το μοντέλο είναι αναλλοίωτο ως προς τις μεταθέσεις των δεικτών $j=1,2,\dots,k$, για λόγους ταυτοποίησης πρέπει να χρησιμοποιήσουμε μια μοναδική ετικέτα για τον καθορισμό των συνιστωσών: θεωρούμε αυτή όπου τα μ_j είναι σε αύξουσα τάξη, έτσι ώστε η από κοινού εκ των υστέρων κατανομή των παραμέτρων είναι $k!$ φορές το γινόμενο κανονικών και Γάμμα πυκνοτήτων, περιορισμένες στο σύνολο $\mu_1 < \mu_2 < \dots < \mu_k$.

Αρχίζοντας την περιγραφή του RJMCMC, υπενθυμίζουμε ότι προτείνεται μια κίνηση τύπου m από ένα σημείο y σε ένα y' μεγαλύτερης διάστασης. Αυτό επιτυγχάνεται παράγοντας ένα διάνυσμα τυχαίων μεταβλητών u , ανεξάρτητα από το y και λαμβάνοντας το y' μέσω ενός αντιστρέψιμου μετασχηματισμού: $y' = y'(y, u)$. Η αντίστροφη κίνηση (από το y' στο y) γίνεται χρησιμοποιώντας τον αντίστροφο μετασχηματισμό, έτσι ώστε η προτεινόμενη κίνηση να είναι ντετερμινιστική. Η πιθανότητα αποδοχής γενικά είναι:

$$\alpha_m(y, y') = \min \left\{ 1, \frac{p(y'|x)r_m(y')}{p(y|x)r_m(y)q(u)} \left| \frac{\partial(y')}{\partial(y, u)} \right| \right\} \quad (3.9)$$

όπου $r_m(y)$ είναι η πιθανότητα επιλογής κίνησης τύπου m όταν είμαστε στην κατάσταση y και $q(u)$ είναι η πυκνότητα του u . Ο τελευταίος όρος είναι η ιακωβιανή του μετασχηματισμού από το (y, u) στο $y' = y'(y, u)$.

• **Reversible jump τύποι κινήσεων για κανονικές μείξεις**

Για το ιεραρχικό μοντέλο (3.8) οι Richardson & Green προτείνουν διαδοχικά τους παρακάτω τύπους κινήσεων:

- Τύπος 1 *Ανανέωση των βαρών \underline{p} .*
- Τύπος 2 *Ανανέωση των παραμέτρων $(\underline{\mu}, \underline{\sigma}^2)$.*
- Τύπος 3 *Ανανέωση του διανύσματος θέσης \underline{z} .*
- Τύπος 4 *Ανανέωση της υπερπαραμέτρου β .*
- Τύπος 5 *Διαχωρισμός (split) μιας συνιστώσας σε δύο, ή συνδυασμός (combine) δύο συνιστωσών σε μία.*
- Τύπος 6 *Γέννηση (birth) ή θάνατος (death) μιας κενής συνιστώσας.*

Οι κινήσεις 5 και 6 αλλάζουν το k κατά 1 και κάνουν κατάλληλες μετατροπές στα $\underline{\mu}, \underline{\sigma}^2, \underline{z}, \underline{p}$ οι οποίες θα περιγραφούν διεξοδικά στην συνέχεια.

Η τυχαιότητα στην διαδοχή των παραπάνω κινήσεων βρίσκεται στην επιλογή μεταξύ split και combine στο βήμα 5 και birth ή death στο 6. Μία ολοκλήρωση των βημάτων 1 - 6 καλείται sweep και είναι το βασικό βήμα του αλγορίθμου.

Από την άλλη οι κινήσεις 1,2,3 και 4 δεν αλλάζουν την διάσταση του διανύσματος των παραμέτρων και δεν διαφέρουν σε τίποτα από την ανάλυση που προηγήθηκε (παράδειγμα 2.1.3 σελ. 39). Για τα βήματα αυτά λοιπόν χρησιμοποιούμε έναν πυρήνα Gibbs στον οποίο οι πλήρεις δεσμευμένες κατανομές είναι αντίστοιχα:

$$1. \quad \underline{p} | \dots \sim D(\delta + n_1, \dots, \delta + n_k) \text{ όπου } n_j = \sum_{i=1}^n I_{\{z_i^{(r)}=j\}}, \quad j=1,2,\dots,k.$$

$$2. \quad \mu_j | \dots \sim N\left(\frac{\sigma_j^{-2} \sum_{i:z_i=j} x_i + \kappa \zeta}{\sigma_j^{-2} n_j + \kappa}, \frac{1}{\sigma_j^{-2} n_j + \kappa}\right), \quad j=1,2,\dots,k. \text{ Μια βασική παρατήρηση}$$

είναι ότι για την διατήρηση της διάταξης των μ_j , η συγκεκριμένη κατανομή χρησιμοποιείται μόνο για να προτείνει ένα καινούριο μ_j και αυτό γίνεται αποδεκτό αν και μόνο αν δεν παραβιάζεται η διάταξη. Διαφορετικά κρατάμε την προηγούμενη τιμή.

$$\sigma_j^{-2} | \dots \sim \Gamma\left(\alpha + \frac{1}{2} n_j, \beta + \frac{1}{2} \sum_{i:z_i=j} (x_i - \mu_j)^2\right), \quad j=1,2,\dots,k.$$

$$3. \quad p(z_i = j | \dots) \propto \frac{p_j}{\sigma_j} \exp\left\{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right\}, \quad j=1,2,\dots,k.$$

$$4. \quad \beta \sim \Gamma\left(g + \kappa \alpha, h + \sum_{j=1}^k \sigma_j^{-2}\right).$$

Στην συνέχεια περιγράφουμε αναλυτικά τα βήματα 5 και 6 του αλγορίθμου.

✚ Κίνηση (5) τύπου Split - Combine

Για την κίνηση αυτή απαιτείται ο μηχανισμός reversible jump. Πρέπει να σχεδιάσουμε αυτές τις κινήσεις στην σειρά μια και αποτελούν ένα αντιστρέψιμο ζευγάρι. Οι κατανομές πρότασης που εμπλέκονται σε αυτά τα βήματα μπορεί να είναι οποιεσδήποτε, αρκεί να είμαστε συνεπείς με την υπόθεση dimension matching.

Αρχικά κάνουμε μια τυχαία επιλογή μεταξύ του διαχωρισμού και συνδυασμού, με πιθανότητες b_k και $d_k = 1 - b_k$ αντίστοιχα, ανάλογα με το εκάστοτε k . Φυσικά έχουμε ότι $b_1 = 1$ και $b_{k_{\max}} = 0$, διαφορετικά θεωρούμε ότι $b_k = 0.5$, $\forall k = 2, \dots, k_{\max} - 1$.

Στην περίπτωση που προτείνουμε συνδυασμό δυο συνιστωσών, πρώτα απ' όλα επιλέγουμε τυχαία ένα ζευγάρι δεικτών (j_1, j_2) οι οποίοι αντιστοιχούν σε γειτονικές συνιστώσες, δηλαδή:

$$\mu_{j_1} < \mu_{j_2} \text{ με κανένα άλλο } \mu_j \text{ στο } [\mu_{j_1}, \mu_{j_2}].$$

Στην συνέχεια αυτές οι δύο συνιστώσες ενώνονται, μειώνοντας το k κατά 1, και δημιουργούν μια καινούρια συνιστώσα j^* . Στην συνέχεια πρέπει να ανακατατάξουμε εκείνες τις παρατηρήσεις x_i με $z_i = j_1$ ή $z_i = j_2$ και να δημιουργήσουμε τιμές για το $(p_{j^*}, \mu_{j^*}, \sigma_{j^*})$. Η ανακατανομή γίνεται απλά θέτοντας $z_i = j^*$, ενώ οι υπόλοιπες παράμετροι λαμβάνονται ταιριάζοντας την μηδενική, πρώτη και δεύτερη ροπή της καινούριας συνιστώσας με το άθροισμα των άλλων δύο που αντικαθιστά:

$$\begin{aligned} p_{j^*} &= p_{j_1} + p_{j_2} \\ p_{j^*} \mu_{j^*} &= p_{j_1} \mu_{j_1} + p_{j_2} \mu_{j_2} \\ p_{j^*} (\mu_{j^*}^2 + \sigma_{j^*}^2) &= p_{j_1} (\mu_{j_1}^2 + \sigma_{j_1}^2) + p_{j_2} (\mu_{j_2}^2 + \sigma_{j_2}^2) \end{aligned} \quad (3.10)$$

Για την αντίστροφη κίνηση (split) έχουμε:

Αρχικά επιλέγουμε τυχαία μια συνιστώσα j^* και την χωρίζουμε στα δύο, συμβολίζοντας τις καινούριες συνιστώσες j_1, j_2 . Το διάνυσμα των παραμέτρων για αυτές προκύπτει από τις εξισώσεις (3.10). Στο σύστημα αυτό όμως έχουμε 3 εξισώσεις και 6 άγνωστους, οπότε υπάρχουν 3 βαθμοί ελευθερίας και έτσι πρέπει να παραγάγουμε ένα τυχαίο διάνυσμα διάστασης 3, το οποίο θα καθορίσει τις καινούριες παραμέτρους μέσω ενός αντιστρέψιμου μετασχηματισμού.

Για το λόγο αυτό παράγουμε το τυχαίο διάνυσμα $u = (u_1, u_2, u_3)$ όπου:

$$u_1 \sim B(2,2) \quad u_2 \sim B(2,2) \quad u_3 \sim B(1,1)$$

και θέτουμε:

$$\begin{aligned} p_{j_1} &= p_{j^*} u_1 & p_{j_2} &= p_{j^*} (1 - u_1) \\ \mu_{j_1} &= \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{\frac{p_{j_2}}{p_{j_1}}} & \mu_{j_2} &= \mu_{j^*} + u_2 \sigma_{j^*} \sqrt{\frac{p_{j_1}}{p_{j_2}}} \\ \sigma_{j_1}^2 &= u_3 (1 - u_2^2) \sigma_{j^*}^2 \frac{p_{j^*}}{p_{j_1}} & \sigma_{j_2}^2 &= (1 - u_3) (1 - u_2^2) \sigma_{j^*}^2 \frac{p_{j^*}}{p_{j_2}} \end{aligned} \quad (3.11)$$

Είναι εύκολο να διαπιστωθεί ότι οι σχέσεις αυτές ικανοποιούν τις εξισώσεις του συστήματος (3.10).

Σε αυτό το σημείο ελέγχουμε αν η συνθήκη της διάταξης μέσω παραμένει αναλλοίωτη. Εάν στο διάστημα $[\mu_{j_1}, \mu_{j_2}]$ υπάρχει κάποιο από τα μ_{j^*-1} ή μ_{j^*+1} , τότε η κίνηση απορρίπτεται αμέσως, διότι τότε το ζευγάρι (split,combine) δεν θα ήταν αντιστρέψιμο. Σε αντίθετη περίπτωση μένει να προτείνουμε την ανακατανομή στις δυο καινούριες συνιστώσες εκείνων των x_i με $z_i = j^*$. Αυτό γίνεται πολύ απλά σύμφωνα με το αντίστοιχο βήμα στον Gibbs sampler:

$$p(z_i = j_1 | \dots) = \frac{\frac{p_{j_1}}{\sigma_{j_1}} \exp\left\{-\frac{(x_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right\}}{\frac{p_{j_1}}{\sigma_{j_1}} \exp\left\{-\frac{(x_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right\} + \frac{p_{j_2}}{\sigma_{j_2}} \exp\left\{-\frac{(x_i - \mu_{j_2})^2}{2\sigma_{j_2}^2}\right\}}$$

και $p(z_i = j_2 | \dots) = 1 - p(z_i = j_1 | \dots)$.

Στην συνέχεια θεωρούμε το πρόβλημα καθορισμού της πλήρους έκφρασης της πιθανότητας αποδοχής αυτών των κινήσεων. Συγκεκριμένα θα αποδείξουμε ότι :

Πρόταση 3.1.1: Για την κίνηση τύπου διαχωρισμού η πιθανότητα αποδοχής ισούται με:

$$\alpha_5(y, y') = \min\{1, A\}$$

όπου :

$$\begin{aligned} A = (\text{likelihood ratio}) & \frac{p(k+1)}{p(k)} (k+1) \frac{P_{j_1}^{\delta-1+l_1} P_{j_2}^{\delta-1+l_2}}{P_{j^*}^{\delta-1+l_1+l_2} B(\delta, k\delta)} \\ & \times \sqrt{\frac{\kappa}{2\pi}} \exp\left[-\frac{1}{2} \kappa \left\{ (\mu_{j_1} - \xi)^2 + (\mu_{j_2} - \xi)^2 - (\mu_{j^*} - \xi)^2 \right\}\right] \\ & \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j_1}^2 \sigma_{j_2}^2}{\sigma_{j^*}^2} \right)^{-\alpha-1} \exp\left[-\beta \left\{ \sigma_{j_1}^{-2} + \sigma_{j_2}^{-2} - \sigma_{j^*}^{-2} \right\}\right] \\ & \times \frac{d_{k+1}}{b_k P_{alloc}} \left\{ g_{2,2}(u_1) g_{2,2}(u_2) g_{1,1}(u_3) \right\}^{-1} \\ & \times \frac{P_{j^*} |\mu_{j_1} - \mu_{j_2}| \sigma_{j_1}^2 \sigma_{j_2}^2}{u_2 (1-u_2) u_3 (1-u_3) \sigma_{j^*}^2} \end{aligned} \quad (3.12)$$

Υπενθυμίζουμε ότι k είναι ο αριθμός των συνιστωσών πριν τον διαχωρισμό, l_1 και l_2 είναι ο αριθμός των παρατηρήσεων που προτείνεται να ανακαταταχθούν στις j_1 και j_2 αντίστοιχα, $B(\cdot, \cdot)$ είναι η συνάρτηση βήτα, P_{alloc} είναι η πιθανότητα της συγκεκριμένης ανακατανομής, $g_{p,q}$ είναι η συνάρτηση πυκνότητας της κατανομής

$B(p, q)$ και likelihood ratio είναι ο λόγος της πιθανοφάνειας του μοντέλου διάστασης $(k + 1)$ προς αυτής του μοντέλου διάστασης k .

Απόδειξη: Η απόδειξη παρατίθεται στην Παράγραφο 4.1 του Παραρτήματος.

Τώρα για την αντίστοιχη κίνηση συνδυασμού η πιθανότητα αποδοχής είναι $\alpha_m(y, y') = \min\{1, A^{-1}\}$ χρησιμοποιώντας την ίδια έκφραση για το A με κάποιες αλλαγές στις αντικαταστάσεις. Εδώ τα πράγματα έχουν ως εξής: είμαστε στην κατάσταση y διάστασης k και επιλέγουμε να κάνουμε συνδυασμό δύο γειτονικών συνιστωσών και να πάμε στην κατάσταση y' διάστασης $k - 1$. Οπότε:

Πρόταση 3.1.2: Για την κίνηση τύπου συνδυασμού η πιθανότητα αποδοχής ισούται με:

$$\alpha_s(y, y') = \min\{1, A^{-1}\}$$

όπου :

$$\begin{aligned} A^{-1} = (\text{likelihood ratio}) & \frac{p(k-1)}{p(k)} \frac{1}{k} \frac{P_{j^*}^{\delta-1+l_1+l_2} B(\delta, (k-1)\delta)}{P_{j_1}^{\delta-1+l_1} P_{j_2}^{\delta-1+l_2}} \\ & \times \sqrt{\frac{2\pi}{\kappa}} \exp\left[-\frac{1}{2} \kappa \left\{ (\mu_{j^*} - \xi)^2 - (\mu_{j_1} - \xi)^2 - (\mu_{j_2} - \xi)^2 \right\}\right] \\ & \times \frac{\Gamma(\alpha)}{\beta^\alpha} \left(\frac{\sigma_{j^*}^2}{\sigma_{j_1}^2 \sigma_{j_2}^2} \right)^{-\alpha-1} \exp\left[-\beta \left\{ \sigma_{j^*}^{-2} - \sigma_{j_1}^{-2} - \sigma_{j_2}^{-2} \right\}\right] \\ & \times \frac{b_{k-1} P_{alloc}}{d_k} \left\{ g_{2,2}(u_1) g_{2,2}(u_2) g_{1,1}(u_3) \right\} \\ & \times \frac{u_2 (1-u_2^2) u_3 (1-u_3) \sigma_{j^*}^2}{P_{j^*} |\mu_{j_1} - \mu_{j_2}| \sigma_{j_1}^2 \sigma_{j_2}^2} \end{aligned} \quad (3.13)$$

ΣΧΟΛΙΑ: Ολοκληρώνουμε με κάποιες παρατηρήσεις που είναι σημαντικές για την κίνηση συνδυασμού, οι οποίες θεωρούνται αυτονόητες στην εργασία των Richardson & Green:

Πριν τον συνδυασμό, έχουμε l_1 και l_2 παρατηρήσεις καταταγμένες στις συνιστώσες j_1 και j_2 αντίστοιχα, οπότε P_{alloc} είναι η πιθανότητα να παρατηρηθεί η συγκεκριμένη κατανομή κατά την αντίστροφη κίνηση, δηλαδή κατά τον διαχωρισμό της j^* συνιστώσας στην κατάσταση y' διάστασης $k - 1$.

Προσοχή θα πρέπει να δοθεί στις τιμή του τυχαίου διανύσματος u . Στην κίνηση αυτή δεν το παράγουμε από τις αντίστοιχες κατανομές βήτα, αλλά το θέτουμε ίσο με την λύση των εξισώσεων (3.11), δηλαδή:

$$u_1 = \frac{P_{j_1}}{P_{j^*}}, \quad u_2 = \frac{|\mu_{j_1} - \mu_{j^*}|}{\sigma_{j^*}} \sqrt{\frac{P_{j_1}}{P_{j_2}}}, \quad u_3 = \frac{\sigma_{j_1}^2 P_{j_1}}{(1-u_2^2) \sigma_{j^*}^2 P_{j^*}}.$$

Στην πιθανότητα αποδοχής όμως θεωρούμε ότι κατά την αντίστροφη κίνηση συνδυασμού αυτές οι τιμές έχουν παραχθεί από τις κατανομές βήτα, ώστε οι τιμές τους να εμφανίζονται στις αντίστοιχες πυκνότητες. Σε πολλές περιπτώσεις όμως η τιμή του u_3 τυχαίνει να είναι μεγαλύτερη της μονάδας, κάτι που έχει συνέπεια τον μηδενισμό της πιθανότητας αποδοχής και κατ' επέκταση την άμεση απόρριψη της κίνησης συνδυασμού.

Κίνηση (6) τύπου Birth - Death

Για την κίνηση αυτή απαιτείται επίσης ο μηχανισμός reversible jump, αλλά ευτυχώς τα πράγματα εδώ είναι λιγάκι πιο απλά. Αρχικά κάνουμε μια τυχαία επιλογή μεταξύ γέννησης και θανάτου, χρησιμοποιώντας τις ίδιες πιθανότητες b_k και d_k με πριν.

Για την κίνηση γέννησης μιας κενής συνιστώσας, παράγουμε το βάρος και τις παραμέτρους για αυτήν χρησιμοποιώντας τις κατανομές:

$$p_{j^*} \sim B(1, k), \quad \mu_{j^*} \sim N(\xi, \kappa^{-1}), \quad \sigma_{j^*}^2 \sim \Gamma(\alpha, \beta).$$

Η διαφορά με πριν είναι ότι για να «κάνουμε χώρο» για την καινούρια συνιστώσα, διορθώνουμε όλα τα υπάρχοντα βάρη, ώστε να αθροίζονται στην μονάδα, σύμφωνα με την σχέση: $p'_j = p_j(1 - p_{j^*})$, $\forall j = 1, 2, \dots, k$.

Εάν επιλέξουμε θάνατο τότε διαλέγουμε τυχαία μία από τις κενές συνιστώσες (εφ' όσον υπάρχουν), έστω την j^* , και την διαγράφουμε. Τα βάρη που απομένουν διορθώνονται ώστε να αθροίζονται στην μονάδα, σύμφωνα με την σχέση: $p'_j = p_j / (1 - p_{j^*})$, $\forall j \in \{1, 2, \dots, k\} \setminus \{j^*\}$. Στην περίπτωση που δεν υπάρχουν κενές συνιστώσες η κίνηση απορρίπτεται και οι τιμές των παραμέτρων παραμένουν ίδιες.

Τα πράγματα σε αυτή την περίπτωση είναι πιο απλά σε σχέση με την κίνηση split - combine, διότι καμμία άλλη αλλαγή δεν προτείνεται στις μεταβλητές. Ειδικότερα, η κατανομή των παρατηρήσεων στις συνιστώσες παραμένει αναλλοίωτη.

Η στάσιμη κατανομή της αλυσίδας επιτυγχάνεται, καθώς αποδεχόμαστε γεννήσεις και θανάτους σύμφωνα με την έκφραση (3.9), στην οποία τον ρόλο του u παίζει το $(p_{j^*}, \mu_{j^*}, \sigma_{j^*}^2)$. Η χρησιμοποίηση των εκ των προτέρων κατανομών για τις προτεινόμενες τιμές των $\mu_{j^*}, \sigma_{j^*}^2$ απλοποιεί τον λόγο αυτό. Συγκεκριμένα:

Πρόταση 3.1.3: Για την κίνηση τύπου γέννησης η πιθανότητα αποδοχής ισούται με:

$$\alpha_6(y, y') = \min\{1, A\}$$

όπου:

$$A = \frac{p(k+1)}{p(k)} \frac{1}{B(k\delta, \delta)} p_{j^*}^{\delta-1} (1 - p_{j^*})^{n+k(\delta-1)} (k+1) \times \frac{d_{k+1}}{(k_0+1)b_k} \frac{1}{g_{1,k}(p_{j^*})} (1 - p_{j^*})^{k-1} \quad (3.14)$$

Απόδειξη: Η απόδειξη παρατίθεται στην Παράγραφο 4.2 του Παραρτήματος.

Η πιθανότητα αποδοχής της αντίστροφης κίνησης τύπου θανάτου παρατίθεται στην επόμενη πρόταση.

Πρόταση 3.1.4: Για την κίνηση τύπου θανάτου η πιθανότητα αποδοχής ισούται με:

$$\alpha_6(y, y') = \min \{1, A^{-1}\}$$

όπου:

$$A^{-1} = \frac{p(k-1)}{p(k)} B((k-1)\delta, \delta) \frac{1}{p_{j^*}^{\delta-1} (1-p_{j^*})^{n+(k-1)(\delta-1)}} \frac{1}{k} \times \frac{b_{k-1}k_0}{d_k} g_{1,k-1}(p_{j^*}) \frac{1}{(1-p_{j^*})^{(k-1)-1}} \quad (3.15)$$

Εδώ ολοκληρώνεται ο ορισμός των τύπων κινήσεων. Η εγκυρότητα του αλγορίθμου δεν είναι ασυμβίβαστη με την επιλογή των προτεινόμενων κινήσεων. Εφόσον έχει επιτευχθεί η στάσιμη κατανομή της αλυσίδας, μένει μόνο να ελέγξουμε εάν αυτή είναι ανάγωγη και απεριοδική. Το δεύτερο είναι προφανές, διότι για κάθε «μικρή» γειτονιά της παρούσας κατάστασης $(\beta, \underline{\mu}, \underline{\sigma}^2, k, \underline{p}, \underline{z})$ υπάρχει θετική πιθανότητα η αλυσίδα να βρίσκεται σε αυτή την γειτονιά, μετά από μία επανάληψη (sweep) του αλγορίθμου.

Η αλυσίδα επίσης είναι ξεκάθαρο ότι είναι και ανάγωγη, διότι κινείται από οποιοδήποτε k σε κάποιο άλλο με βήματα μήκους 1 κάθε φορά, ενώ στην κίνηση (3) όλες οι δυνατές κατανομές των παρατηρήσεων στις συνιστώσες έχουν θετική πιθανότητα, ενώ όλες οι παράμετροι και υπερπαράμετροι ανανεώνονται χρησιμοποιώντας συνεχείς κατανομές με στηρίγματα τους φυσικούς παραμετρικούς χώρους.

Στην συνέχεια δίνουμε τον αλγόριθμο για την υλοποίηση του αλγορίθμου που περιγράψαμε:

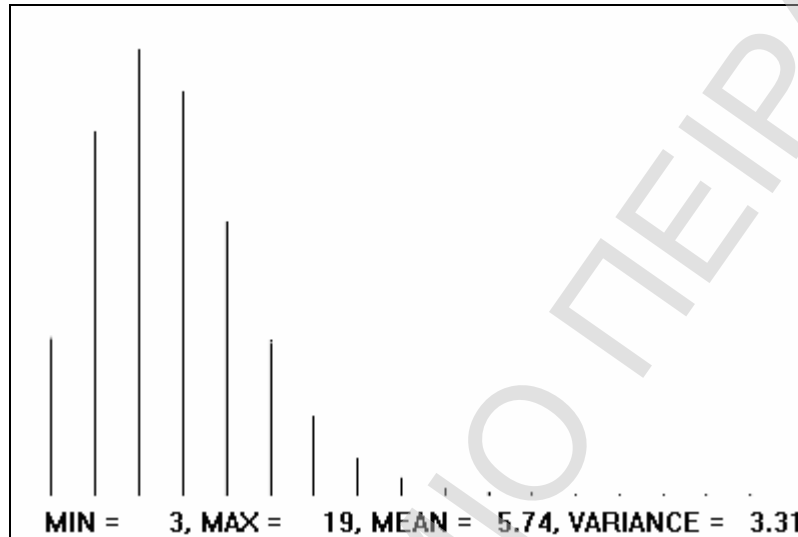
Αλγόριθμος Reversible jump MCMC για κανονικές μείξεις

0. Δώσε κάποιες αρχικές τιμές στο $(\underline{\mu}, \underline{\sigma}, k, \underline{p})$.
1. Στην επανάληψη t : $y^{(t)} = (\underline{\mu}^{(t)}, \underline{\sigma}^{(t)}, k^{(t)}, \underline{p}^{(t)})$
- 1.1. Προσομοίωσε το διάνυσμα θέσης \underline{z} σύμφωνα με την κίνηση Τύπου 3.
- 1.2. Προσομοίωσε το διάνυσμα των μέσων και των διασπορών $(\underline{\mu}, \underline{\sigma}^2)$ σύμφωνα με την κίνηση Τύπου 2.
- 1.3. Προσομοίωσε το διάνυσμα βαρών \underline{p} σύμφωνα με την κίνηση Τύπου 1.
- 1.4. Προσομοίωσε το β σύμφωνα με την κίνηση Τύπου 4.
2. Κίνηση Τύπου 5 (Split - Combine): Με πιθανότητα b_k πήγαινε στο βήμα 2.1 αλλιώς πήγαινε στο βήμα 2.2
- 2.1. Επίλεξε μία συνιστώσα j^* για διαχωρισμό.
- 2.1.1. Προσομοίωσε το διάνυσμα u .

$$u_1 \sim B(2,2), \quad u_2 \sim B(2,2), \quad u_3 \sim B(1,1).$$
 Πρότεινε τις καινούριες τιμές των μεταβλητών $y' = (\underline{\mu}', \underline{\sigma}', k+1, \underline{p}', \underline{z}')$ σύμφωνα με τις σχέσεις (3.11) και έλεγξε για τυχόν παραβίαση της διάταξης των μέσων. Αν ναι, θέσε $y^{(t+1)} = y^{(t)}$ και πήγαινε στο βήμα 3 αλλιώς συνέχισε.
- 2.1.2. Με πιθανότητα $\alpha_5(y, y') = \min\{1, A\}$ θέσε $y^{(t+1)} = y'$, αλλιώς $y^{(t+1)} = y^{(t)}$.
- 2.2. Επίλεξε δύο γειτονικές συνιστώσες j_1 και j_2 για συνδυασμό.
- 2.2.1. Θέσε το διάνυσμα u ίσο με την λύση των εξισώσεων (3.11).
 Πρότεινε τις καινούριες τιμές των μεταβλητών $y' = (\underline{\mu}', \underline{\sigma}', k-1, \underline{p}', \underline{z}')$ σύμφωνα με τις σχέσεις (3.10).
- 2.2.2. Με πιθανότητα $\alpha_5(y, y') = \min\{1, A^{-1}\}$ θέσε $y^{(t+1)} = y'$, αλλιώς $y^{(t+1)} = y^{(t)}$.
3. Κίνηση Τύπου 6 (Birth - Death): Με πιθανότητα b_k πήγαινε στο βήμα 3.1 αλλιώς πήγαινε στο βήμα 3.2.
- 3.1. Κίνηση γέννησης κενής συνιστώσας:
- 3.1.1. Προσομοίωσε το διάνυσμα $(p_{j^*}, \mu_{j^*}, \sigma_{j^*}^2)$:

$$p_{j^*} \sim B(1, k), \quad \mu_{j^*} \sim N(\xi, \kappa^{-1}), \quad \sigma_{j^*}^{-2} \sim \Gamma(\alpha, \beta).$$
 Πρότεινε τις καινούριες τιμές των μεταβλητών $y' = (\underline{\mu}', \underline{\sigma}', k+1, \underline{p}', \underline{z}')$.
- 3.1.2. Με πιθανότητα $\alpha_6(y, y') = \min\{1, A'\}$ θέσε $y^{(t+1)} = y'$, αλλιώς $y^{(t+1)} = y^{(t)}$.
- 3.2. Επίλεξε μία κενή συνιστώσα για θάνατο με πιθανότητα $1/(k_0 + 1)$. Αν δεν υπάρχουν κενές συνιστώσες θέσε $y^{(t+1)} = y^{(t)}$ και σταμάτα.
- 3.2.1. Πρότεινε τις καινούριες τιμές των μεταβλητών $y' = (\underline{\mu}', \underline{\sigma}', k-1, \underline{p}', \underline{z}')$.
- 3.2.2. Με πιθανότητα $\alpha_6(y, y') = \min\{1, (A')^{-1}\}$ θέσε $y^{(t+1)} = y'$, αλλιώς $y^{(t+1)} = y^{(t)}$.

Παράδειγμα 3.1.1 Εφαρμόσαμε τον παραπάνω αλγόριθμο στα δεδομένα των ταχυτήτων των 82 γαλαξιών. Οι τιμές των σταθερών που δεν είχαν οριστεί πριν είναι: $\kappa = 1/R^2$, $\alpha = 2$, $g = 0.2$, $h = 10/R^2$ και $\delta = 1$. Η εκ των προτέρων στο k είναι η διακριτή ομοιόμορφη στους ακέραιους $1, 2, \dots, k_{\max} = 30$. Ο αλγόριθμος reversible jump εκτελέστηκε για 200000 επαναλήψεις, μετά από το burn-in των πρώτων 100000 επαναλήψεων, κάτι το οποίο κρίνεται αρκετό για να λάβουμε αξιόπιστα αποτελέσματα.

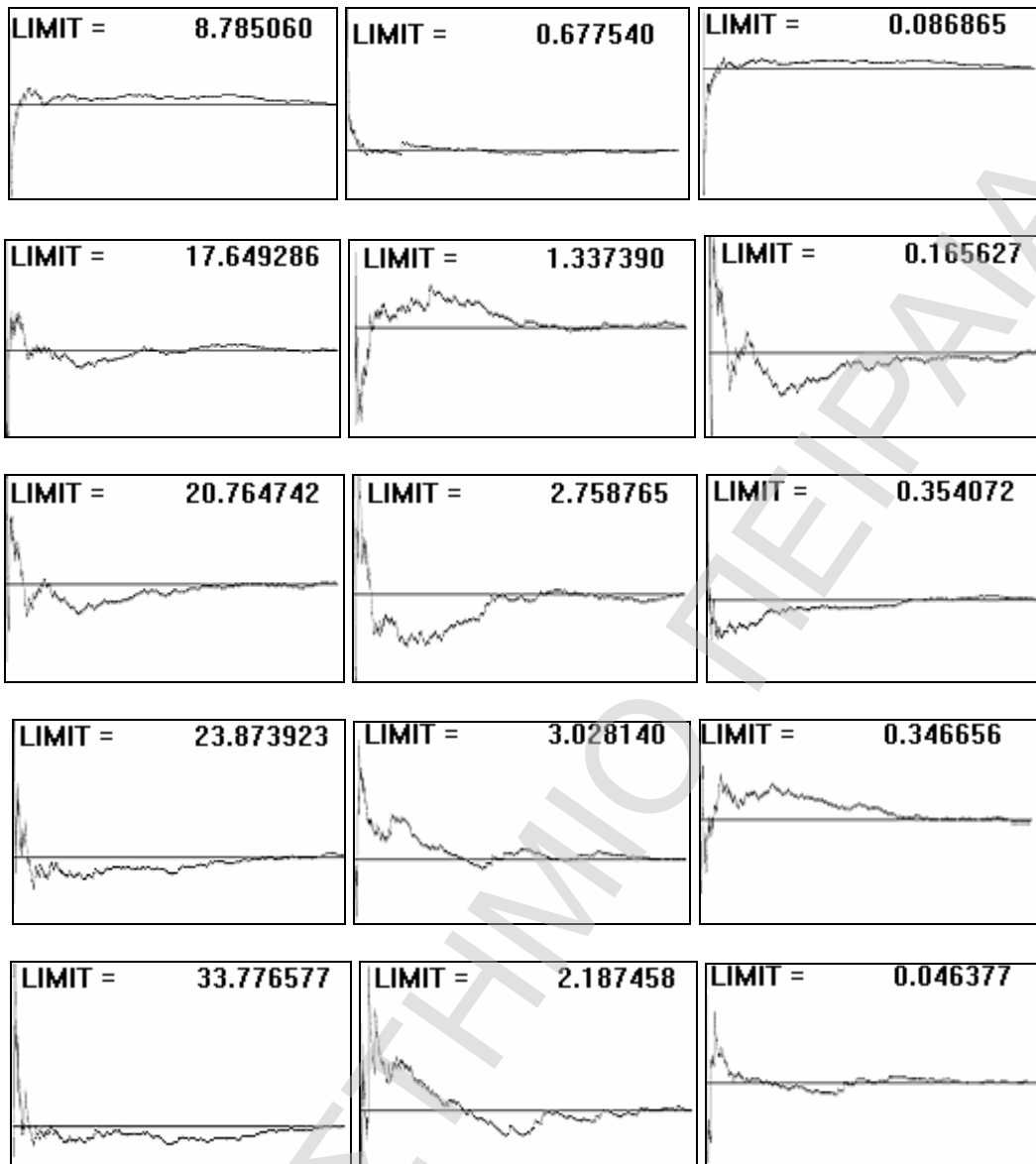


ΣΧΗΜΑ 25. Το ιστόγραμμα των εκ των υστέρων πιθανοτήτων του αριθμού των συνιστωσών.

Ο αλγόριθμος αρχικοποιήθηκε στην διάσταση 1, ενώ η μεγαλύτερη τιμή που παρατηρήθηκε ήταν ίση με 19. Στο Σχήμα 25 φαίνεται το ιστόγραμμα των εκ των υστέρων πιθανοτήτων του αριθμού των συνιστωσών και συμπεραίνουμε ότι η κατανομή του έχει αρκετά μεγάλο εύρος. Τα κυριότερα μοντέλα που «ανταγωνίζονται» για την εξήγηση των δεδομένων είναι αυτά που αποτελούνται από 4 έως 7 συνιστώσες. Αναλυτικά η εκ των υστέρων κατανομή του φαίνεται στον παρακάτω πίνακα:

$p(k x)$			Ποσοστό (%) των αποδεκτών κινήσεων	
			Split – Combine	Birth - Death
$p(1) = 0.0000$	$p(5) = 0.2213$	$p(9) = 0.0390$	4.9	15.98
$p(2) = 0.0000$	$p(6) = 0.2094$	$p(10) = 0.0174$		
$p(3) = 0.0901$	$p(7) = 0.1472$	$p(11) = 0.0073$		
$p(4) = 0.1847$	$p(8) = 0.0788$	$\sum_{i \geq 12} p(i) = 0.0049$		

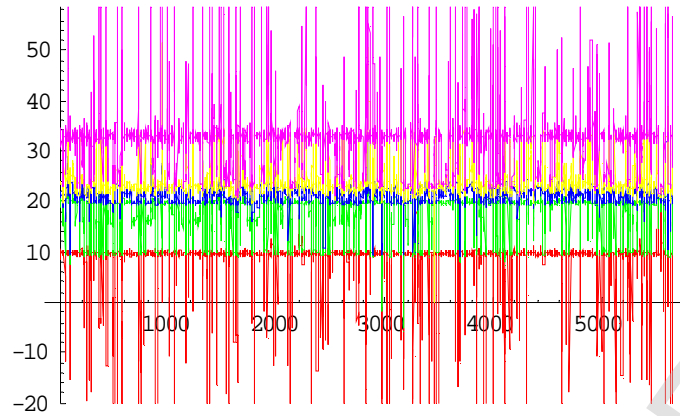
Παρατηρούμε ότι η πιο πιθανή (εκ των υστέρων) τιμή του πλήθους των συνιστωσών είναι ίση με 5. Σε αυτή την περίπτωση, δεσμεύοντας ως προς $k = 5$, έχουμε την παρακάτω εικόνα:



ΣΧΗΜΑ 26. Τα ίχνη των εργοδικών μέσων των παραμέτρων, δεσμεύοντας ως προς $k=5$.

Αναλυτικά, δεσμεύοντας ως προς $k = 5$, οι εκτιμήσεις των παραμέτρων είναι:

$$\begin{array}{lll}
 \hat{E}(\mu_1|x) \cong 8.7851 & \hat{E}(\sigma_1^2|x) \cong 0.6775 & \hat{E}(p_1|x) \cong 0.08687 \\
 \hat{E}(\mu_2|x) \cong 17.6493 & \hat{E}(\sigma_2^2|x) \cong 1.3374 & \hat{E}(p_2|x) \cong 0.16563 \\
 \hat{E}(\mu_3|x) \cong 20.7647 & \hat{E}(\sigma_3^2|x) \cong 2.7588 & \hat{E}(p_3|x) \cong 0.35407 \\
 \hat{E}(\mu_4|x) \cong 23.8739 & \hat{E}(\sigma_4^2|x) \cong 3.0281 & \hat{E}(p_4|x) \cong 0.34666 \\
 \hat{E}(\mu_5|x) \cong 33.7766 & \hat{E}(\sigma_5^2|x) \cong 2.1875 & \hat{E}(p_5|x) \cong 0.04638
 \end{array}$$

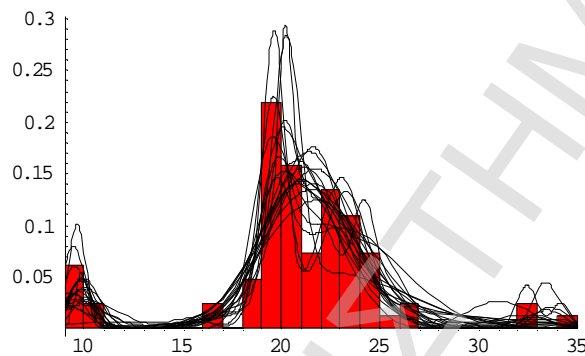


ΣΧΗΜΑ 27. Τα ίχνη των τιμών των μέσων των συνιστωσών, δεσμεύοντας ως προς $k=5$.

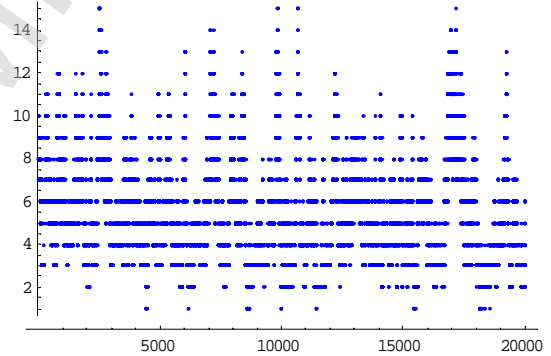
Σε κάθε επανάληψη του αλγορίθμου παράγονται τιμές των $(\underline{p}, \underline{\theta})$, από τις οποίες μπορούμε να υπολογίσουμε τις αντίστοιχες πυκνότητες

$$f(\cdot | k, \underline{p}, \underline{\theta}) = \sum_{j=1}^k p_j f(\cdot | \theta_j)$$

Η μορφή αυτών για 50 επαναλήψεις αναπαρίσται στο σχήμα 28.

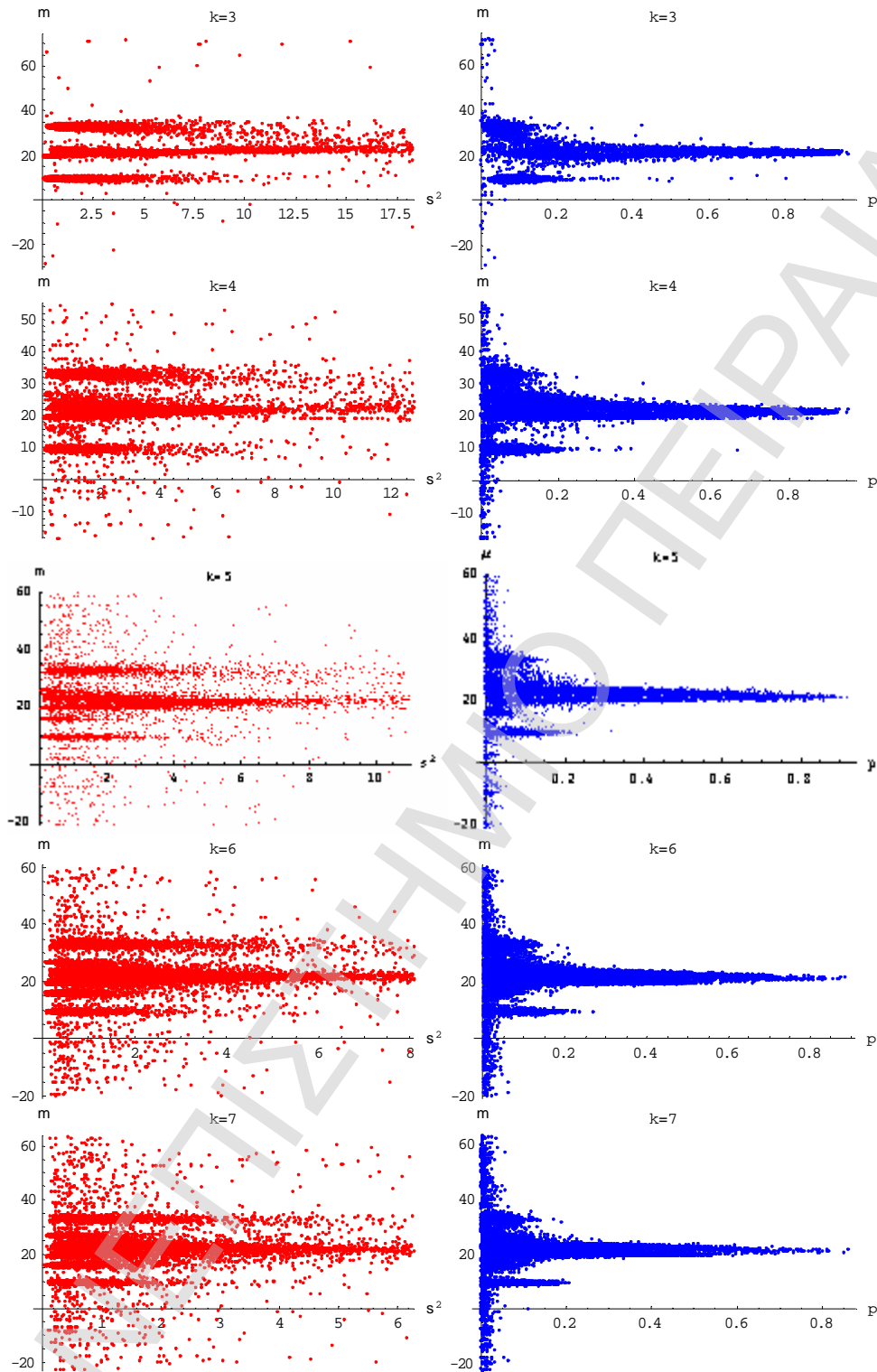


ΣΧΗΜΑ 28. Δείγμα από την εκ των υστέρων κατανομή $f(\cdot | k, \underline{p}, \underline{\theta})$.



ΣΧΗΜΑ 29. Οι αναπηδήσεις της αλυσίδας σε διαφορετικές διαστάσεις.

Για τις κυριότερες τιμές του αριθμού των συνιστωσών παραθέτουμε στο Σχήμα 30 το διάγραμμα συνδιακύμανσης (scatterplot) των μέσων με τις αντίστοιχες διασπορές και των βαρών τους.

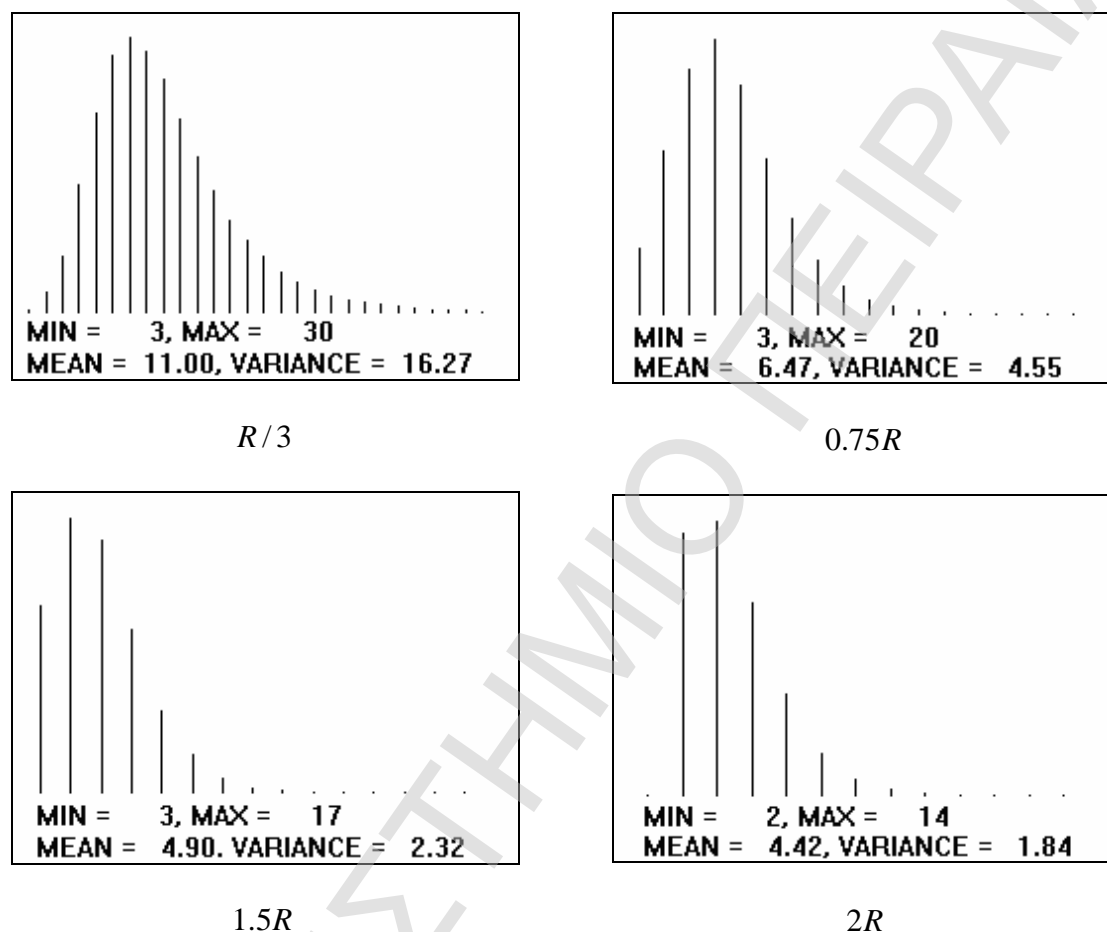


ΣΧΗΜΑ 30. Τα scatterplots των μέσων με τις αντίστοιχες διασπορές (1^η στήλη) και των μέσων με τα αντίστοιχα βάρη (2^η στήλη) για $k=3,4,5,6,7$.

Εναισθησία των αποτελεσμάτων στις εκ των προτέρων υποθέσεις.

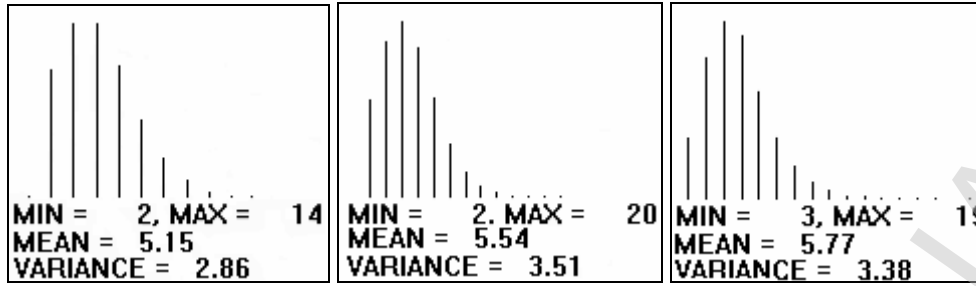
Εδώ θα παρουσιάσουμε κάποιες συγκρίσεις των αποτελεσμάτων που προκύπτουν για εναλλακτικές τιμές των υπερπαραμέτρων που εμπλέκονται στο μοντέλο. Κατ' αρχάς

εξετάζουμε πώς ο αριθμός των συνιστωσών σχετίζεται με την εκ των προτέρων πληροφορία στις διασπορές αυτών. Από μια διαισθητική σκοπιά είναι φυσικό να συσχετίζουμε τα σ_j με το εύρος R των δεδομένων και αναμένουμε ότι αυξάνοντας τις τιμές των μέσων των αντίστοιχων κατανομών γάμμα, θα καταλήξουμε σε μοντέλα με λιγότερες συνιστώσες. Στο Σχήμα 31 φαίνεται η αλλαγή της εκ των υστέρων κατανομής του k , καθώς η τιμή του R μεταβάλλεται σε $R/3$, $0.75R$, $1.5R$ και $2R$.



ΣΧΗΜΑ 31. Τα ιστογράμματα των εκ των υστέρων ποσοστών του k , καθώς μεταβάλλεται η υπερπαραμέτρος R .

Παρατηρούμε ότι ακόμα και μικρές μεταβολές της υπερπαραμέτρου R , επιδρούν σημαντικά στην εκ των υστέρων κατανομή του πλήθους των συνιστωσών. Έτσι, στο μοντέλο (3.8), η επιλογή του R , παίζει μεγάλο ρόλο στην εκ των υστέρων κατανομή του k , οπότε είναι δύσκολο να είμαστε ασθενώς πληροφοριακοί. Παρ' όλα αυτά, η εκ των υστέρων κατανομή του k είναι αρκετά ανθεκτική στις μεταβολές του λόγου g/h (το οποίο σχετίζεται με το εύρος R), οι οποίες όμως οδηγούν στον ίδιο μέσο και τυπική απόκλιση για το β . Αυτό το γεγονός απεικονίζεται στο Σχήμα 32 στο οποίο παρατηρούμε περίπου την ίδια συμπεριφορά της εκ των υστέρων κατανομής του k , για 3 τιμές των α , g και h επιλεγμένες έτσι ώστε η εκ των προτέρων τάξη των σ_j του υψηλότερου επιπέδου, $\sqrt{g/ha}$, να κυμαίνεται μεταξύ των $R/5$, $R/10$ και $R/20$ (η επιλογή $R/10$ είναι αυτή που χρησιμοποιήσαμε τελικά).

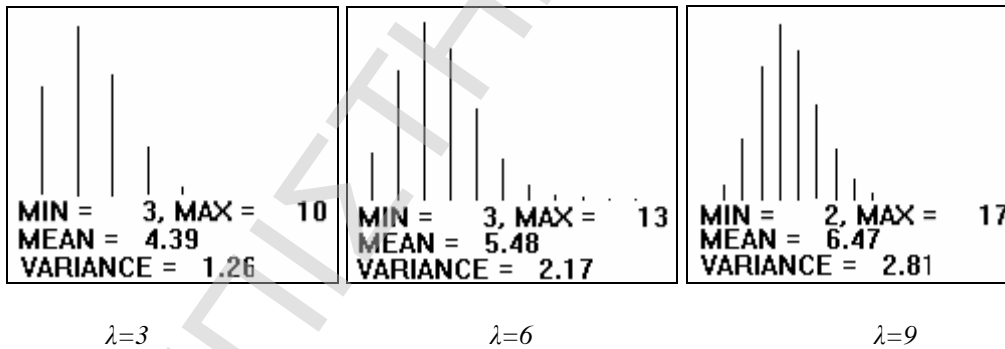


ΣΧΗΜΑ 32. Η εκ των υστέρων κατανομή του k , καθώς ο λόγος $\sqrt{g/ha}$ μεταβάλλεται στις τιμές $R/5$, $R/10$ και $R/20$.

Άρα, το ιεραρχικό μοντέλο (3.8) μας επιτρέπει να είμαστε αρκετά μη πληροφοριακοί σχετικά με την κατανομή των διασπορών της μείξης.

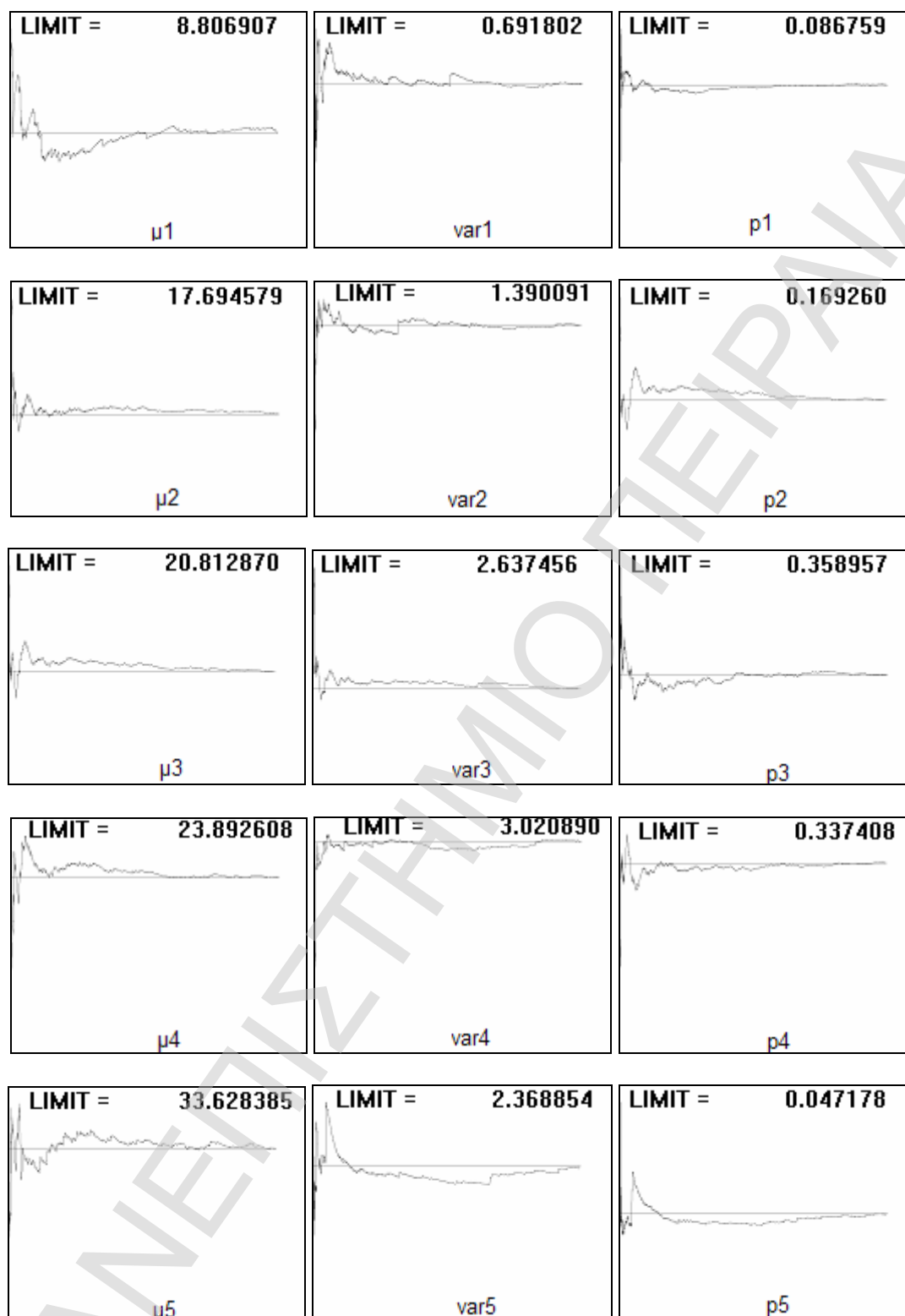
Επίσης, συμπληρώνουμε ότι οι εκτιμήσεις όλων των παραμέτρων είναι αρκετά σταθερές μεταβάλλοντας τον μέσο ξ των εκ των προτέρων κατανομών των μ_j . Εκτελέσαμε τον αλγόριθμο με το ξ να παίρνει τιμές 5, 10 και 30 (η αρχική μας επιλογή ήταν ίση με 20.8335, δηλαδή η διάμεσος των παρατηρήσεων). Η εκ των υστέρων κατανομή του k παραμένει αναλλοίωτη, ενώ απλώς απαιτούνται περισσότερες επαναλήψεις του αλγορίθμου για την σύγκλιση των εργοδικών μέσων.

Τέλος, μελετήσαμε την επίδραση στην εκ των υστέρων κατανομή του αριθμού των συνιστωσών, εάν εκ των προτέρων θεωρήσουμε ότι ακολουθεί την κατανομή Poisson. Στο Σχήμα 33 παρατηρούμε το ιστόγραμμα της εκ των υστέρων κατανομής όταν η παράμετρος θέσης της εκ των προτέρων κατανομής Poisson είναι ίση με 3, 6 και 9 αντίστοιχα.



ΣΧΗΜΑ 33. Η εκ των υστέρων κατανομή του αριθμού των συνιστωσών, όταν εκ των προτέρων $k \sim P(\lambda)$, με $\lambda=3, 6, 9$ αντίστοιχα.

Από την πλευρά της εκτέλεσης του αλγορίθμου το μόνο που αλλάζει είναι να θέσουμε $p(k+1)/p(k) = \lambda/(k+1)$ στις πιθανότητες αποδοχής (3.12) και (3.14), και επίσης $p(k-1)/p(k) = k/\lambda$ στις (3.13) και (3.15). Από τα αποτελέσματα που πήραμε, συμπεραίνουμε ότι η εκ των υστέρων κατανομή του k , είναι αρκετά ευαίσθητη σε μεταβολές τις παραμέτρου λ . Πάντως, τα αποτελέσματα της προηγούμενης ανάλυσης συμφωνούν σε μεγάλο βαθμό όταν το λ κυμαίνεται από 5 έως 7. Μικρότερες τιμές αυτού έχουν ως αποτέλεσμα μικρότερο αριθμό συνιστωσών, ενώ αυξάνοντας το λ η εκ των υστέρων κατανομή συγκεντρώνεται σε μεγαλύτερες τιμές. Στην περίπτωση όπου $\lambda=6$ παραθέτουμε τις εκτιμήσεις των εργοδικών μέσων για $k=5$ και τις αντίστοιχες γραφικές παραστάσεις στο επόμενο σχήμα.



ΣΧΗΜΑ 34. Τα γραφήματα των εργοδικών μέσων των παραμέτρων μαζί με τις αντίστοιχες εκτιμήσεις, δεσμεύοντας ως προς $k=5$, όταν εκ των προτέρων $k \sim P(6)$.

Συγκρίνοντας τις εκτιμήσεις των παραμέτρων που φαίνονται στο Σχήμα 34 συμπεραίνουμε την πλήρη συμφωνία με την περίπτωση όπου η εκ των προτέρων κατανομή του k είναι η ομοιόμορφη.

3.2 Birth - Death MCMC

Σε αυτή την παράγραφο περιγράφουμε μία εναλλακτική μέθοδο MCMC, για την Μπεϋζιανή ανάλυση δεδομένων από μείξεις κατανομών με άγνωστο πλήθος παραμέτρων. Η προσέγγιση αυτή λέγεται *Birth - Death Markov Chain Monte Carlo* (BDMCMC Stephens, 2000) και θεωρεί τις παραμέτρους του μοντέλου ως μία μαρκαρισμένη σημειακή διαδικασία και σκοπός της είναι να δημιουργήσει μια Μαρκοβιανή διαδικασία γεννήσεων - θανάτων συνεχούς χρόνου, με κατάλληλη στάσιμη κατανομή.

3.2.1 Οι παράμετροι ως σημειακή διαδικασία

Η στρατηγική είναι να θεωρήσουμε κάθε συνιστώσα της μείξης ως ένα σημείο στον παραμετρικό χώρο και εκμεταλλευόμενοι την θεωρία προσομοίωσης σημειακών διαδικασιών (Ripley, 1977), να κατασκευάσουμε μία Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την εκ των υστέρων κατανομή του μοντέλου. Αγνοώντας τις ετικέτες των συνιστωσών, θεωρούμε κάθε σύνολο k τιμών των παραμέτρων $\{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\}$, ως ένα σύνολο k σημείων στο $\Theta \times [0,1]$, υπό τον περιορισμό ότι $\sum_{i=1}^k p_i = 1$. Έτσι η εκ των υστέρων κατανομή μπορεί να θεωρηθεί ως η (κατάλληλα περιορισμένη) κατανομή των σημείων στο $\Theta \times [0,1]$, ή με άλλα λόγια, σαν μία σημειακή διαδικασία στο $\Theta \times [0,1]$. Ισοδύναμα, αν αντιστοιχίσουμε σε κάθε σημείο $(\underline{\theta}_i)$ το πρόσημο $p_i \in [0,1]$, λέμε ότι η εκ των υστέρων κατανομή αποτελεί μια μαρκαρισμένη σημειακή διαδικασία (*marked point process*) στον χώρο Θ , υπό τον περιορισμό ότι τα πρόσημα αθροίζουν στην μονάδα.

3.2.2 Διαδικασίες γεννήσεων - θανάτων για τις συνιστώσες του μοντέλου μείξης

Έστω ότι το Ω_k είναι ο παραμετρικός χώρος του μοντέλου (1.2) με k συνιστώσες, αγνοώντας τα ονόματα των συνιστωσών και έστω $\Omega = \bigcup_k \Omega_k$. Γενικά, αναφερόμαστε σε Μαρκοβιανές διαδικασίες γεννήσεων - θανάτων συνεχούς χρόνου στον χώρο Ω , οι οποίες περιλαμβάνουν πεπερασμένο αριθμό μεταβάσεων στους υποχώρους Ω_i σε πεπερασμένα χρονικά διαστήματα. Οι μεταβάσεις αυτές είναι δύο τύπων: «γεννήσεις» από το Ω_k στο Ω_{k+1} , και «θάνατοι» από το Ω_k στο Ω_{k-1} . Όταν η διαδικασία βρίσκεται στο $y \in \Omega_k$, η συμπεριφορά της ορίζεται από τον ρυθμό γεννήσεων $\beta(y)$, τον ρυθμό θανάτων $\delta(y)$, και από τους πυρήνες μετάβασης γεννήσεων και θανάτων $K_\beta^k(y, \cdot)$ και $K_\delta^k(y, \cdot)$, οι οποίοι είναι μέτρα πιθανότητας στους χώρους Ω_{k+1} και Ω_{k-1} αντίστοιχα. Οι γεννήσεις και οι θάνατοι συμβαίνουν σύμφωνα με ανεξάρτητες διαδικασίες Poisson, με ρυθμούς $\beta(y)$ και $\delta(y)$ αντίστοιχα. Εάν συμβεί μία γέννηση, τότε η διαδικασία μεταβαίνει σε ένα μετρήσιμο σύνολο $F \subseteq \Omega_{k+1}$ με πιθανότητα $K_\beta^k(y, F)$, ενώ αν συμβεί θάνατος τότε η αντίστοιχη πιθανότητα για κάθε μετρήσιμο υποσύνολο G του Ω_{k-1} ισούται με $K_\delta^k(y, G)$. Για

τέτοιου είδους διαδικασίες έχει αποδειχθεί (Preston, 1976) ότι η επόμενη συνθήκη είναι ικανή για την ύπαρξη της στάσιμης κατανομής π αυτών:

Συνθήκη λεπτομερούς ισορροπίας: Λέμε ότι η π ικανοποιεί τις συνθήκες λεπτομερούς ισορροπίας αν:

$$\int_F \beta(y) \pi_k(dy) = \int_{\Omega_{k+1}} \delta(z) K_\delta^{k+1}(z, F) \pi_{k+1}(dz) \text{ για } k \geq 0, F \subseteq \Omega_k \quad (3.16)$$

και

$$\int_G \delta(z) \pi_{k+1}(dz) = \int_{\Theta_k} \beta(y) K_\beta^k(y, G) \pi_k(dy) \text{ για } k \geq 0, G \subseteq \Omega_{k+1}. \quad (3.17)$$

Διαισθητικά, αυτό σημαίνει ότι ο ρυθμός με τον οποίο η διαδικασία εγκαταλείπει οποιαδήποτε περιοχή μέσω μίας γέννησης, ισούται με τον ρυθμό με τον οποίο επισκέπτεται αυτή την περιοχή μέσω ενός θανάτου και αντιστρόφως.

Συμβολίζουμε με:

$$y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\} \in \Omega_k$$

τις παραμέτρους του μοντέλου (θεωρώντας σταθερές τις υπερπαραμέτρους αν υπάρχουν), έτσι μπορούμε να γράψουμε $(\underline{\theta}_i, p_i) \in y$. Αναλυτικά, ορίζουμε ως γεννήσεις και θανάτους στον Ω τα παρακάτω γεγονότα:

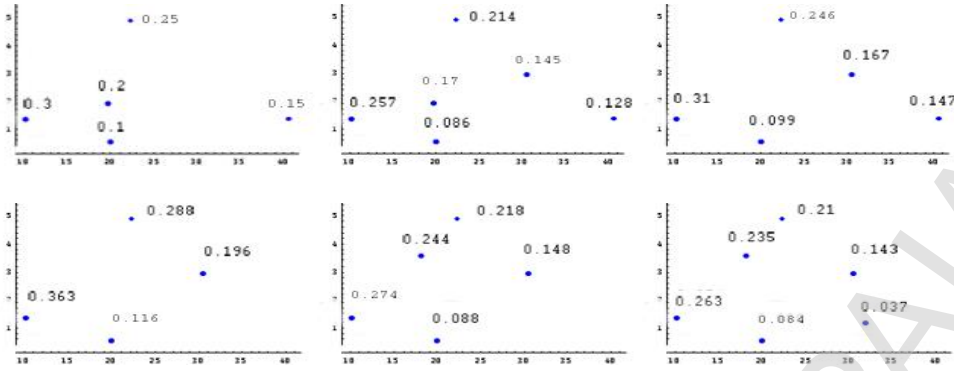
Γεννήσεις: Οι γεννήσεις συμβαίνουν με συνολικό ρυθμό $\beta(y)$, ενώ μία γέννηση συμβαίνει στο σημείο $(\underline{\theta}, p) \in \Theta \times [0,1]$, επιλεγμένο σύμφωνα με μία πυκνότητα $b(y; (\underline{\theta}, p))$. Εάν στον χρόνο t η διαδικασία βρίσκεται στο $y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\} \in \Omega_k$ και πρόκειται να γίνει μία γέννηση στο $(\underline{\theta}, p) \in \Theta \times [0,1]$ τότε η διαδικασία μεταβαίνει στο

$$y \cup (\underline{\theta}, p) := \{(\underline{\theta}_1, p_1(1-p)), \dots, (\underline{\theta}_k, p_k(1-p)), (\underline{\theta}, p)\} \in \Omega_{k+1}. \quad (3.18)$$

Θάνατοι: Όταν η διαδικασία βρίσκεται στο $y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\} \in \Omega_k$ κάθε σημείο $(\underline{\theta}_i, p_i) \in y$ πεθαίνει ανεξάρτητα από τα άλλα, σύμφωνα με μία διαδικασία Poisson με ρυθμό $\delta_i(y) = d(y \setminus (\underline{\theta}_i, p_i); (\underline{\theta}_i, p_i))$ για κάποιο $d: \Omega \times (\Theta \times [0,1]) \rightarrow \mathbb{R}^+$. Έτσι, ο συνολικός αριθμός θανάτων ισούται με $\sum_i \delta_i(y)$. Εάν στον χρόνο t η διαδικασία βρίσκεται στο $y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\} \in \Omega_k$ και πρόκειται να συμβεί ένας θάνατος στο $(\underline{\theta}_i, p_i) \in y$ τότε η διαδικασία μεταβαίνει στο

$$y \setminus (\underline{\theta}_i, p_i) := \left\{ \left(\underline{\theta}_1, \frac{p_1}{1-p_i} \right), \dots, \left(\underline{\theta}_{i-1}, \frac{p_{i-1}}{1-p_i} \right), \left(\underline{\theta}_{i+1}, \frac{p_{i+1}}{1-p_i} \right), \dots, \left(\underline{\theta}_k, \frac{p_k}{1-p_i} \right) \right\} \in \Omega_{k-1}. \quad (3.19)$$

Προφανώς οι γεννήσεις και οι θάνατοι έχουν επιλεγεί κατά τέτοιο τρόπο ώστε να είναι αντίστροφες διαδικασίες μεταξύ τους, ενώ ο περιορισμός $\sum p_i = 1$ εξακολουθεί να ισχύει και μετά από αυτές.



ΣΧΗΜΑ 35. Αναπαράσταση γεννήσεων και θανάτων: στην πρώτη εικόνα έχουμε $k=4$ και στην επόμενη γεννιέται μία συνιστώσα με βάρος 0.145, ενώ τα υπόλοιπα πρόσημα διορθώνονται ώστε να αθροίζονται στην μονάδα.

Στην συνέχεια ο χρόνος για το επόμενο γεγονός ακολουθεί την εκθετική κατανομή με μέσο $1/(\beta(y)+\delta(y))$. Θα είναι γέννηση με πιθανότητα $\beta(y)/(\beta(y)+\delta(y))$ και θάνατος της συνιστώσας j με πιθανότητα $\delta_j(y)/(\beta(y)+\delta(y))$.

Επιστρέφοντας στις μείξεις κατανομών, θεωρούμε το εξής ιεραρχικό μοντέλο: Έστω ότι η εκ των προτέρων κατανομή του $(k, \underline{p}, \underline{\theta})$, δοθισών των υπερπαραμέτρων ω και η , είναι η $r(k, \underline{p}, \underline{\theta} | \omega, \eta)$. Επί πλέον, υποθέτουμε ότι το k έχει εκ των προτέρων κατανομή $r(k | \omega, \eta)$ και ότι τα $\underline{p}, \underline{\theta}_1, \dots, \underline{\theta}_k$ είναι εκ των προτέρων ανεξάρτητα, με τα $\underline{\theta}_i$ να είναι ταυτοτικά κατανομημένα από μία κατανομή με πυκνότητα $\pi(\cdot | \omega, \eta)$, ενώ το (p_1, \dots, p_{k-1}) κατανέμεται ομοιόμορφα στον τόπο l^{k-1} και $p_k = 1 - \sum_{i=1}^{k-1} p_i$, όπου $l^{k-1} = \{(p_1, \dots, p_{k-1}) : p_1, \dots, p_{k-1} \geq 0 \text{ και } p_1 + \dots + p_{k-1} \leq 1\}$. Οπότε:

$$r(k, \underline{p}, \underline{\theta} | \omega, \eta) = r(y) \propto r(k | \omega, \eta) \prod_{i=1}^k \pi(\theta_i | \omega, \eta).$$

Το επόμενο θεώρημα δίνει ικανές συνθήκες για να έχει η παραπάνω Μαρκοβιανή διαδικασία στάσιμη κατανομή την εκ των υστέρων κατανομή $\pi(k, \underline{p}, \underline{\theta} | x, \omega, \eta)$.

Θεώρημα 3.2.1 Θεωρώντας το τελευταίο ιεραρχικό μοντέλο και για σταθερά ω και η σταθερά, η παραπάνω διαδικασία γεννήσεων - θανάτων έχει στάσιμη κατανομή $\pi(k, \underline{p}, \underline{\theta} | x, \omega, \eta)$ εάν τα b και d ικανοποιούν την

$$(k+1)d(y; (\tilde{\theta}, p))r(y \cup (\tilde{\theta}, p))L(y \cup (\tilde{\theta}, p))k(1-p)^{k-1} = \beta(y)b(y; (\tilde{\theta}, p))r(y)L(y) \quad (3.20)$$

για κάθε $y \in \Omega_k$ και $(\tilde{\theta}, p) \in \Theta \times [0, 1]$.

Απόδειξη: Για την απόδειξη παραπέμπουμε στην εργασία του Stephens (2000).

Ο Stephens (2000) θεωρεί την διαδικασία με τον ρυθμό γεννήσεων σταθερό: $\beta(y) = \lambda_b$, ενώ $b(y; (\tilde{\theta}, p)) = k(1-p)^{k-1} \pi(\tilde{\theta} | \omega, \eta)$. Αυτό σημαίνει ότι το βάρος p

«γεννιέται» ανεξάρτητα από τις παραμέτρους $\tilde{\theta}$. Εφαρμόζοντας το Θεώρημα 3.2.1 βρίσκουμε ότι η διαδικασία έχει την κατάλληλη στάσιμη κατανομή όταν, ενώ είμαστε στην κατάσταση $y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\}$, κάθε σημείο πεθαίνει ανεξάρτητα από τα υπόλοιπα σύμφωνα με μία διαδικασία Poisson με ρυθμό:

$$\begin{aligned} kd(y \setminus (\underline{\theta}_j, p_j); (\underline{\theta}_j, p_j)) r(k|\omega, \eta) \prod_{i=1}^k \pi(\underline{\theta}_i|\omega, \eta) L(y) k(1-p)^{k-1} &= \\ &= \lambda_b k(1-p)^{k-1} \pi(\underline{\theta}_j|\omega, \eta) r(k-1|\omega, \eta) \prod_{i \neq j} \pi(\underline{\theta}_i|\omega, \eta) L(y \setminus (\underline{\theta}_j, p_j)) \Rightarrow \\ d(y \setminus (\underline{\theta}_j, p_j); (\underline{\theta}_j, p_j)) &= \lambda_b \frac{L(y \setminus (\underline{\theta}_j, p_j))}{L(y)} \frac{r(k-1|\omega, \eta)}{kr(k|\omega, \eta)}, \quad j = 1, \dots, k. \end{aligned} \quad (3.21)$$

Ο παρακάτω αλγόριθμος προσομοιώνει την παραπάνω διαδικασία στην περίπτωση όπου οι υπερπαραμέτροι θεωρούνται σταθερές.

Αλγόριθμος προσομοίωσης διαδικασίας γεννήσεων - θανάτων με σταθερές υπερπαραμέτρους

0. Δώσε κάποιες αρχικές τιμές στο k και στο $y = \{(\underline{\theta}_1, p_1), \dots, (\underline{\theta}_k, p_k)\} \in \Omega_k$.
1. Θέσε $\beta(y) = \lambda_b$.
2. Στο βήμα t :
 - 2.1. Υπολόγισε το:

$$d(y \setminus (\underline{q}_j, p_j); (\underline{q}_j, p_j)) = \lambda_b \frac{L(y \setminus (\underline{q}_j, p_j))}{L(y)} \frac{r(k-1|w, h)}{kr(k|w, h)}, \quad j = 1, \dots, k.$$

- 2.2. Υπολόγισε τον συνολικό ρυθμό θανάτων $\delta(y) = \sum_i \delta_i(y)$.
- 2.3. Προσομοίωσε τον χρόνο παραμονής σε αυτή την κατάσταση από την εκθετική κατανομή με μέσο $1/(\beta(y) + \delta(y))$.
- 2.4. Κίνηση birth - death. Με πιθανότητα $\beta(y)/(\beta(y) + \delta(y))$ πήγαινε στο βήμα
 - 2.4.1 αλλιώς πήγαινε στο βήμα 2.4.2.
 - 2.4.1. *Γέννηση*: Προσομοίωσε το βάρος και τις παραμέτρους ανεξάρτητα από τις κατανομές: $p \sim k(1-p)^{k-1}$ και $\tilde{\theta} \sim \pi(\tilde{\theta}|\omega, \eta)$. Διόρθωσε το y σύμφωνα με την σχέση (3.18).
 - 2.4.2. *Θάνατος*: Η συνιστώσα $(\underline{\theta}_j, p_j) \in y$ ($j = 1, \dots, k$) πεθαίνει με πιθανότητα $\delta_j(y)/(\beta(y) + \delta(y))$. Διόρθωσε το y σύμφωνα με την σχέση (3.19).

Σημειώνουμε ότι ο παραπάνω αλγόριθμος προσομοιώνει την διαδικασία γέννησης - θανάτων για το μοντέλο μείξεων στην περίπτωση όπου οι υπερπαραμέτροι είναι σταθερές. Για την περίπτωση μη σταθερών υπερπαραμέτρων θα ολοκληρωθεί στην συνέχεια. Παρατηρούμε την αμεσότητα εκτέλεσής του εφόσον τα μόνα κρίσιμα σημεία είναι ο υπολογισμός της πιθανοφάνειας $L(y)$ και η δυνατότητα

προσομοίωσης από την $\pi(\tilde{\theta}|\omega, \eta)$. Τέλος, προσθέτουμε ότι στο βήμα 2.4.1. η κατανομή που χρησιμοποιείται για τα βάρη είναι η βήτα, $B(1, k)$.

Ένα σημείο που χρήζει περαιτέρω διευκρίνησης είναι ότι οι γεννήσεις προκύπτουν από την εκ των προτέρων κατανομή και δεν εξαρτώνται από την εκάστοτε κατάσταση y , κάτι που μπορεί να οδηγήσει σε πολλές γεννήσεις συνιστωσών που δεν θα βοηθούν στην ερμηνεία των δεδομένων. Το γεγονός αυτό δικαιολογείται από το ότι αυτές οι συνιστώσες θα έχουν υψηλό ρυθμό θανάτου και έτσι θα πεθαίνουν αρκετά γρήγορα, το οποίο είναι το ίδιο (από πλευράς απόδοσης) με έναν αλγόριθμο αποδοχής - απόρριψης με υψηλό ρυθμό απορρίψεων. Παρ' όλα αυτά στις περισσότερες εφαρμογές αυτή η επιλογή έχει ικανοποιητική απόδοση, οπότε δεν κρίνεται απαραίτητη η εύρεση μιας πολυπλοκότερης προσέγγισης.

3.2.3 Η κατασκευή της Μαρκοβιανής αλυσίδας

Η παραπάνω διαδικασία μπορεί να συνδυαστεί με ενδιάμεσα βήματα MCMC για την ανανέωση των ω, η και έτσι να προκύψει μία Μαρκοβιανή αλυσίδα με στάσιμη κατανομή $\pi(k, \underline{p}, \underline{\theta}, \omega, \eta | \underline{x})$. Διερύνοντας τα δεδομένα με την εισαγωγή των ελλειπών μεταβλητών \underline{z} (Παράγραφος 1.2) και υποθέτοντας την ύπαρξη των κατάλληλων συζυγών εκ των προτέρων κατανομών, μπορούμε να χρησιμοποιήσουμε έναν δειγματολήπτη Gibbs, ή έναν αλγόριθμο Metropolis - Hastings, αν θέλουμε να αποφύγουμε την διεύρυνση των δεδομένων.

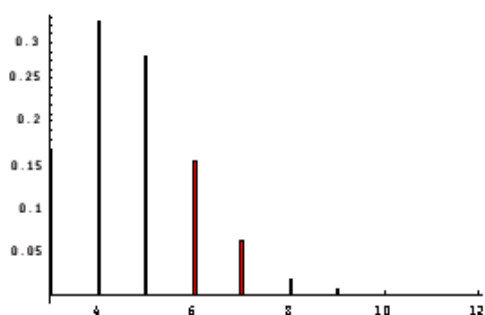
Αλγόριθμος Birth Death MCMC

0. Δώσε κάποιες αρχικές τιμές $(k^{(0)}, \underline{\theta}^{(0)}, \underline{p}^{(0)}, \omega^{(0)}, \eta^{(0)})$.
1. Στο βήμα t , δοθείσης της κατάστασης $(k^{(t-1)}, \underline{\theta}^{(t-1)}, \underline{p}^{(t-1)}, \omega^{(t-1)}, \eta^{(t-1)})$, προσομοίωσε το $(k^{(t)}, \underline{\theta}^{(t)}, \underline{p}^{(t)}, \omega^{(t)}, \eta^{(t)})$ ως εξής:
 - 1.1. Θεωρώντας $(\omega, \eta) = (\omega^{(t-1)}, \eta^{(t-1)})$, προσομοίωσε νέες τιμές $(\tilde{k}^{(t)}, \tilde{\theta}^{(t)}, \tilde{p}^{(t)})$ τρέχοντας την διαδικασία γεννήσεων - θανάτων για ένα προκαθορισμένο χρονικό διάστημα t_0 , ξεκινώντας από το $(k^{(t-1)}, \underline{\theta}^{(t-1)}, \underline{p}^{(t-1)})$. Θέσε $k^{(t)} = \tilde{k}^{(t)}$.
 - 1.2. Προσομοίωσε το $\underline{z}^{(t)}$ από την $p(\underline{z} | k^{(t)}, \tilde{\theta}^{(t)}, \tilde{p}^{(t)}, \omega^{(t-1)}, \eta^{(t-1)}, \underline{x})$.
 - 1.3. Προσομοίωσε το $(\omega^{(t)}, \eta^{(t)})$ από την $p(\omega, \eta | k^{(t)}, \tilde{\theta}^{(t)}, \tilde{p}^{(t)}, \underline{x}, \underline{z}^{(t)})$.
 - 1.4. Προσομοίωσε το $(\theta^{(t)}, \underline{p}^{(t)})$ από την $p(\theta, \underline{p} | k^{(t)}, \omega^{(t)}, \eta^{(t)}, \underline{x}, \underline{z}^{(t)})$.

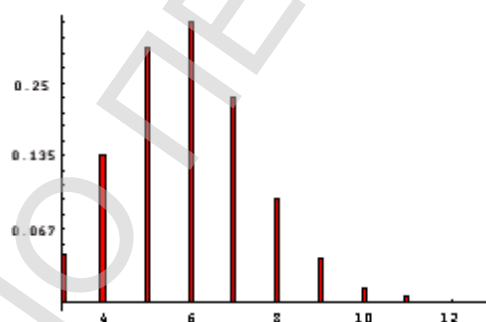
Υπό την προϋπόθεση ότι οι πλήρεις δεσμευμένες κατανομές έχουν στήριγμα σε κατάλληλους περιορισμούς του παραμετρικού χώρου, ο παραπάνω αλγόριθμος ορίζει μία ανάγωγη, απεριοδική και επαναληπτική Μαρκοβιανή αλυσίδα με στάσιμη κατανομή $\pi(k, \underline{p}, \underline{\theta}, \omega, \eta, \underline{z} | \underline{x})$. Το βήμα 1.1 περιλαμβάνει κινήσεις μεταξύ διαφορετικών τιμών του k , μέσω της γέννησης καινούργιων και του θανάτου

υπαρχόντων συνιστωσών. Τα βήματα 1.2, 1.3 και 1.4 επιτρέπουν την αλλαγή των τιμών των παραμέτρων κρατώντας το k σταθερό. Το βήμα 1.4 δεν είναι απαραίτητο για την σύγκλιση στην σωστή κατανομή, αλλά βελτιώνει την απόδοση του αλγορίθμου.

Παράδειγμα 3.2.1 Εφαρμόσαμε τον παραπάνω αλγόριθμο στα δεδομένα των ταχυτήτων των 82 γαλαξιών, υποθέτοντας εκ των προτέρων ότι ο αριθμός των συνιστωσών ακολουθεί την κατανομή Poisson $P(I)$, με $\lambda=3$ και $\lambda=6$. Οι υπόλοιπες εκ των προτέρων υποθέσεις, δεσμεύοντας ως προς k , είναι ίδιες με το μοντέλο των Richardson & Green (Παράγραφος 3.1.2, σελ. 64). Ο αλγόριθμος BDMCMC εκτελέστηκε για 30000 επαναλήψεις, μετά από το burn in των πρώτων 20000 επαναλήψεων, κάτι το οποίο κρίνεται αρκετό για να λάβουμε αξιόπιστα αποτελέσματα.



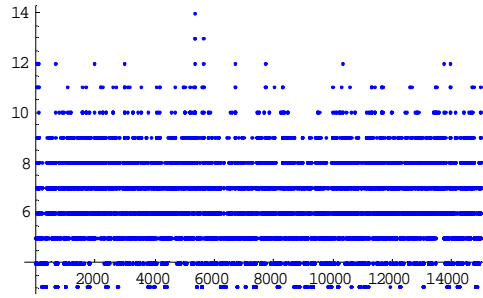
ΣΧΗΜΑ 36. Η εκ των υστέρων κατανομή του αριθμού των συνιστωσών, με $\lambda=3$.



ΣΧΗΜΑ 37. Το αντίστοιχο γράφημα για $\lambda=6$.

Στα Σχήματα 36 και 37 φαίνεται το ιστόγραμμα των εκ των υστέρων ποσοστών του αριθμού των συνιστωσών για $I = 3$ και $I = 6$, και αρχικά συμπεραίνουμε ότι η κατανομή και στις δύο περιπτώσεις έχει μικρότερο εύρος από την αντίστοιχη με την μέθοδο reversible jump. Έπειτα παρατηρούμε ότι η κατανομή δεν είναι ανεπηρέαστη από την εκ των προτέρων πληροφορία. Τα αποτελέσματά μας συμφωνούν με την μέθοδο reversible jump στην περίπτωση όπου $I = 6$. Συνδυάζοντας αυτό το γεγονός με το ότι σε αυτή την περίπτωση είμαστε λιγότερο πληροφοριακοί από ότι όταν $I = 3$ (υπό την έννοια της εκ των προτέρων διασποράς), επιλέγουμε να συνεχίσουμε την ανάλυση με $I = 6$. Ο αλγόριθμος αρχικοποιήθηκε στην διάσταση 1, ενώ η μεγαλύτερη τιμή που παρατηρήθηκε ήταν ίση με 14. Τα κυριότερα μοντέλα που «ανταγωνίζονται» για την εξήγηση των δεδομένων είναι αυτά που αποτελούνται από 4 έως 7 συνιστώσες. Αναλυτικά η εκ των υστέρων κατανομή του αριθμού των συνιστωσών φαίνεται στον παρακάτω πίνακα:

$p(k x)$		
$p(1) = 0$	$p(5) \cong 0.231$	$p(9) \cong 0.0392$
$p(2) = 0$	$p(6) \cong 0.2548$	$p(10) \cong 0.0122$
$p(3) \cong 0.0431$	$p(7) \cong 0.1865$	$p(11) \cong 0.046$
$p(4) \cong 0.1339$	$p(8) \cong 0.0936$	$\sum_{i \geq 12} p(i) \cong 0.0011$

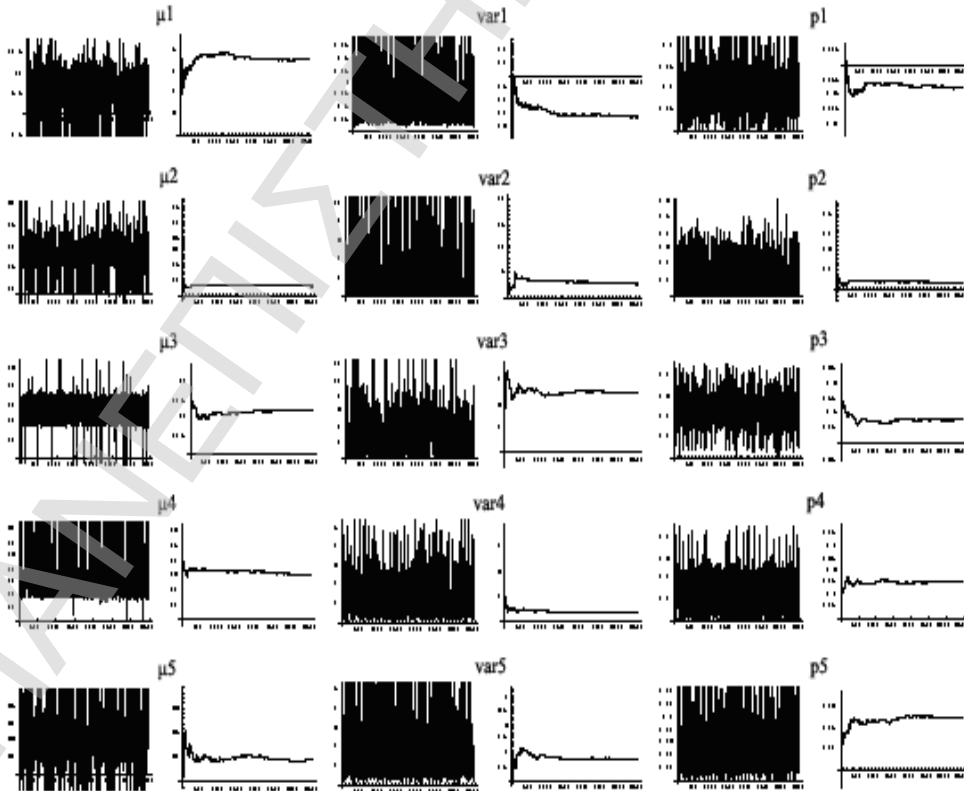


ΣΧΗΜΑ 38. Οι αναπηδήσεις σε διαφορετικές διαστάσεις με την μέθοδο BDMCMC.

Εφαρμόσαμε τον αλγόριθμο αναδιάταξης μέσω του εσωτερικού γινομένου για την αντιμετώπιση του φαινομένου label switching. Δεσμεύοντας ως προς $k = 5$ λάβαμε τις παρακάτω εκτιμήσεις:

$$\begin{array}{lll}
 \hat{E}(\mu_1|x) \cong 8.5939 & \hat{E}(\sigma_1^2|x) \cong 0.6710 & \hat{E}(p_1|x) \cong 0.0884 \\
 \hat{E}(\mu_2|x) \cong 18.002 & \hat{E}(\sigma_2^2|x) \cong 1.4535 & \hat{E}(p_2|x) \cong 0.1695 \\
 \hat{E}(\mu_3|x) \cong 21.0097 & \hat{E}(\sigma_3^2|x) \cong 2.3455 & \hat{E}(p_3|x) \cong 0.3371 \\
 \hat{E}(\mu_4|x) \cong 23.9116 & \hat{E}(\sigma_4^2|x) \cong 3.32643 & \hat{E}(p_4|x) \cong 0.3561 \\
 \hat{E}(\mu_5|x) \cong 33.8662 & \hat{E}(\sigma_5^2|x) \cong 2.3629 & \hat{E}(p_5|x) \cong 0.0533
 \end{array}$$

Στο Σχήμα 39 φαίνονται οι αναδιατεταγμένες τιμές των παραμέτρων καθώς και οι αντίστοιχοι εργοδικοί μέσοι.



ΣΧΗΜΑ 39. Οι αναδιατεταγμένες τιμές των παραμέτρων και οι αντίστοιχοι εργοδικοί μέσοι, δεσμεύοντας ως προς $k=5$, με την μέθοδο BDMCMC.

Συγκρίνοντας τα παραπάνω αποτελέσματα με τα αντίστοιχα της μεθόδου RJMCMC, συμπεραίνουμε αρχικά ότι όσον αφορά στην εκ των υστέρων κατανομή του αριθμού των συνιστωσών k , την ικανοποιητική σύγκλιση των δύο διαφορετικών προσεγγίσεων, στην περίπτωση όπου έχει υιοθετηθεί παρόμοια εκ των προτέρων πληροφορία. Κάποιες αναμενόμενες αποκλίσεις παρουσιάζονται στις εκτιμήσεις, κυρίως των διασπορών και των βαρών της 3^{ης} και 4^{ης} συνιστώσας (οι οποίες είναι οι πιο κοντινές ως προς την απόσταση των μέσων) και αυτό ίσως να οφείλεται εν μέρει στον αλγόριθμο αναδιάταξης που χρησιμοποιήσαμε στην μέθοδο BDMCMC. Επίσης ο Stephens στην εργασία του υποστηρίζει ότι η παραπάνω ανάλυση δίνει πιο αξιόπιστα αποτελέσματα, εάν θεωρήσουμε ότι το μοντέλο μας απαρτίζεται από συνιστώσες που ακολουθούν την κατανομή t_4 .

3.3 Σύγκριση μεταξύ RJMCMC και BDMCMC

Το παράδειγμα που διαπραγματευτήκαμε και με τις δύο μεθόδους, δείχνει ότι είναι δυνατή μία Μπεϋζιανή προσέγγιση για την αντιμετώπιση του προβλήματος μείξεων κανονικών κατανομών με άγνωστο αριθμό συνιστωσών. Σημειώνουμε ότι τα αποτελέσματα που λάβαμε δεσμεύοντας ως προς k συμφωνούν με τα αντίστοιχα της περίπτωσης όπου ο αριθμός των συνιστωσών είναι σταθερός. Παρ' όλα αυτά έχει υποστηριχθεί ότι όταν το k θεωρείται γνωστό ο δειγματολήπτης καθυστερεί να συγκλίνει διότι επισκέπτεται πιο συχνά περιοχές χαμηλής πιθανότητας, έτσι η προσέγγιση με άγνωστο k πρέπει να προτιμάται (Richardson & Green, 1997, p. 751). Τέλος προσθέτουμε ότι και οι δύο μεθοδολογίες μπορούν να εφαρμοστούν και στην περίπτωση πολυδιάστατων κανονικών κατανομών (Stephens, 2000).

Ένα κοινό χαρακτηριστικό (ή μειονέκτημα) και των δύο μεθόδων είναι η ισχυρή εξάρτηση της εκ των υστέρων κατανομής του αριθμού των συνιστωσών, από την εκ των προτέρων επιλογή μας τόσο για αυτόν, όσο και για τις υπόλοιπες παραμέτρους του μοντέλου. Καθώς φαίνεται (δες και την ανάλυση ευαισθησίας των αποτελεσμάτων με την μέθοδο RJMCMC, σελ.77) γενικά δεν υπάρχει τρόπος να εκφράσουμε «εκ των προτέρων άγνοια» στην ιεραρχική δομή του μοντέλου μας. Για αυτό το γεγονός ο Aitkin (2001), μετά από μία σειρά εναλλακτικών προσεγγίσεων στο γαλαξιακό σύνολο δεδομένων, σημειώνει ότι:

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

Όπως φαίνεται από τα αποτελέσματα, η μέθοδος BDMCMC αποτελεί μία ελκυστική εναλλακτική λύση ως προς την μέθοδο του Green, για προβλήματα πεπερασμένων μείξεων. Οι κύριες διαφορές μεταξύ των δύο αυτών προσεγγίσεων είναι οι εξής:

- A. Η μέθοδος BDMCMC λειτουργεί σε συνεχή χρόνο, αντικαθιστώντας το βήμα αποδοχής - απόρριψης με τον χρόνο παραμονής σε κάθε κατάσταση.
- B. Οι μεταπηδήσεις ανάμεσα σε διαφορετικές διαστάσεις της αλυσίδας, στην μέθοδο RJMCMC γίνεται με την χρήση των ελλিপών δεδομένων, ενώ στην μέθοδο BDMCMC εξαλείφονται, εφόσον ολοκληρώνουμε ως προς αυτά, για τον υπολογισμό της πιθανοφάνειας.
- C. Η μέθοδος BDMCMC δεν περιέχει μετασηματισμούς παραμέτρων, έτσι αποφεύγεται ο υπολογισμός πολύπλοκων οριζουσών.
- D. Η μέθοδος BDMCMC αντιμετωπίζει τις παραμέτρους ως σημειακή διαδικασία και δεν επιβάλλει εικονικούς περιορισμούς (διάταξη των μέσων) όπως η μέθοδος RJMCMC.
- E. Το φαινόμενο label switching δεν εμφανίζεται στην μέθοδο RJMCMC, συνεπώς η συμπερασματολογία είναι πιο άμεση.

Η πιο σημαντική διαφορά θεωρείται η B, ενώ στην άλλη άκρη βρίσκεται η A, μια και οι δύο μεθοδολογίες θα μπορούσαν να διαφοροποιηθούν έτσι ώστε να λειτουργούν σε συνεχή ή σε διακριτό χρόνο.

Αναφορικά με την εκτέλεση του αλγορίθμου από προγραμματιστικής πλευράς θα λέγαμε ότι η μέθοδος BDMCMC είναι αρκετά πιο απλή, από την άλλη όμως παρουσιάζει το μειονέκτημα του υπολογισμού της πιθανοφάνειας, κάτι που έχει αρκετές επιπτώσεις. Για παράδειγμα, στην γλώσσα FORTRAN ο υπολογισμός αυτός έχει ως αποτέλεσμα την «υπερχείλισή» της, με συνέπεια τα αποτελέσματα που πήραμε να είναι αρκετά εσφαλμένα. Έτσι αναγκαστήκαμε να εκτελέσουμε τον αλγόριθμο BDMCMC στο υπολογιστικό πακέτο Mathematica, εκμεταλλευόμενοι την ευχέρεια που παρέχει για ακριβέστερους υπολογισμούς σε περιπτώσεις πολύ μικρών αριθμών, κάτι που απαίτησε βέβαια πολύ μεγαλύτερο χρόνο.

Παρ' όλα αυτά, σημειώνουμε ότι ο αλγόριθμος BDMCMC μπορεί να γενικευθεί συμπεριλαμβάνοντας και άλλου είδους μεταβάσεις συνεχούς χρόνου, όπως αυτές του τύπου split - combine των Richardson & Green. Τέλος, οι Carré, Robert και Rydén (2003) απέδειξαν ότι με κατάλληλους χρονικούς μετασχηματισμούς, η αλυσίδα reversible jump συγκλίνει σε μία διαδικασία γεννήσεων - θανάτων συνεχούς χρόνου, επιβεβαιώνοντας την ισχυρή ομοιότητα των δύο μεθόδων.

4. Παράρτημα

4.1 Απόδειξη Πρότασης 3.1.1

Σύμφωνα με την σχέση (3.9), έχουμε ότι γενικά η πιθανότητα αποδοχής δίνεται από την έκφραση:

$$\alpha_m(y, y') = \min \left\{ 1, \frac{p(y'|x)r_m(y')}{p(y|x)r_m(y)q(u)} \left| \frac{\partial(y')}{\partial(y, u)} \right| \right\}$$

Αρχικά υπολογίζουμε τον λόγο $\frac{p(y'|x)}{p(y|x)}$, ο οποίος σύμφωνα με το ιεραρχικό μοντέλο (3.8) γράφεται αναλυτικότερα ως εξής:

$$\begin{aligned} \frac{p(y'|x)}{p(y|x)} &= \frac{p(\eta, k', p', z', \theta'|x)}{p(\eta, k, p, z, \theta|x)} \\ &= \frac{p(x|\theta', z') p(\theta'|k', \eta) p(z'|p', k') p(p'|k', \delta) p(k') p(\eta)}{p(x|\theta, z) p(\theta|k, \eta) p(z|p, k) p(p|k, \delta) p(k) p(\eta)} \\ &= \frac{p(x|\theta', z') p(\theta'|k+1, \eta) p(z'|p', k+1) p(p'|k+1, \delta) p(k+1)}{p(x|\theta, z) p(\theta|k, \eta) p(z|p, k) p(p|k, \delta) p(k)} \end{aligned} \quad (4.1)$$

Στην τελευταία έκφραση ο πρώτος όρος είναι ο λόγος πιθανοφανειών των δυο μοντέλων και ακολουθούν οι εκ των προτέρων κατανομές.

Για τον λόγο των πιθανοφανειών έχουμε ότι:

$$\text{likelihood ratio} = \frac{p(x|\theta', z')}{p(x|\theta, z)} = \frac{\prod_{i=1}^n \varphi(x_i | \mu'_i, \sigma'^2_i)}{\prod_{i=1}^n \varphi(x_i | \mu_i, \sigma^2_i)} \quad (4.2)$$

όπου με $\varphi(\cdot)$ συμβολίζουμε την πυκνότητα της αντίστοιχης κανονικής κατανομής. Στην συνέχεια λαμβάνουμε (συμβολίζουμε με $\text{INVT}(\cdot)$ την πυκνότητα της αντίστροφης Γάμμα κατανομής):

$$\frac{p(\theta'|k+1, \eta)}{p(\theta|k, \eta)} = \frac{(k+1)! \prod_{i=1}^{k+1} (\mu'_i | k+1, \xi) \text{INVT}(\sigma'_i | k+1, \xi)}{k! \prod_{i=1}^k (\mu_i | k, \xi) \text{INVT}(\sigma_i | k, \xi)} =$$

$$\begin{aligned}
 &= (k+1) \frac{\frac{1}{\sqrt{2\pi\kappa^{-1}}} \exp\left\{-\frac{(\mu_{j_1} - \zeta)^2 \kappa}{2}\right\}}{\frac{1}{\sqrt{2\pi\kappa^{-1}}} \exp\left\{-\frac{(\mu_{j_2} - \zeta)^2 \kappa}{2}\right\}} \\
 &\quad \times \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_{j_1}^{-2})^{\alpha+1} \exp(-\beta\sigma_{j_1}^{-2})}{\frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_{j_2}^{-2})^{\alpha+1} \exp(-\beta\sigma_{j_2}^{-2})} \\
 &\quad \times \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_{j^*}^{-2})^{\alpha+1} \exp(-\beta\sigma_{j^*}^{-2})}{\frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_{j^*}^{-2})^{\alpha+1} \exp(-\beta\sigma_{j^*}^{-2})} \\
 &= (k+1) \sqrt{\frac{\kappa}{2\pi}} \exp\left[-\frac{1}{2} \kappa \{(\mu_{j_1} - \zeta)^2 + (\mu_{j_2} - \zeta)^2 - (\mu_{j^*} - \zeta)^2\}\right] \\
 &\quad \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j_1}^2 \sigma_{j_2}^2}{\sigma_{j^*}^2}\right)^{-\alpha-1} \exp\left[-\beta\{\sigma_{j_1}^{-2} + \sigma_{j_2}^{-2} - \sigma_{j^*}^{-2}\}\right] \tag{4.3}
 \end{aligned}$$

Για τον επόμενο όρο στην (4.1) έχουμε:

$$\frac{p(\underline{z}' | \underline{p}', k+1)}{p(\underline{z} | \underline{p}, k)} = \frac{\prod_{i=1}^n p'_{z_i}}{\prod_{i=1}^n p_{z_i}} = \frac{p_{j_1}^{l_1} p_{j_2}^{l_2}}{p_{j^*}^{l_1+l_2}} \tag{4.4}$$

όπου l_1 και l_2 είναι ο αριθμός των παρατηρήσεων που προτείνεται να ανακαταταχθούν στις j_1 και j_2 αντίστοιχα.

Τέλος:

$$\begin{aligned}
 \frac{p(\underline{p}' | k+1, \delta)}{p(\underline{p} | k, \delta)} &= \frac{\frac{\Gamma((k+1)\delta)}{(\Gamma(\delta))^{k+1}} p_1^{\delta-1} \dots p_{j^*-1}^{\delta-1} p_{j_1}^{\delta-1} p_{j_2}^{\delta-1} p_{j^*+1}^{\delta-1} \dots p_k^{\delta-1}}{\frac{\Gamma(k\delta)}{(\Gamma(\delta))^k} p_1^{\delta-1} \dots p_{j^*}^{\delta-1} \dots p_k^{\delta-1}} = \\
 &= \frac{\Gamma((k+1)\delta)}{\Gamma(k\delta)\Gamma(\delta)} \frac{p_{j_1}^{\delta-1} p_{j_2}^{\delta-1}}{p_{j^*}^{\delta-1}} = \frac{1}{B(\delta, k\delta)} \frac{p_{j_1}^{\delta-1} p_{j_2}^{\delta-1}}{p_{j^*}^{\delta-1}} \tag{4.5}
 \end{aligned}$$

Αντικαθιστώντας τις (4.2), (4.3), (4.4) και (4.5) στην (4.1) λαμβάνουμε ότι:

$$\begin{aligned}
 \frac{p(y' | \underline{x})}{p(y | \underline{x})} &= (\text{likelihood ratio}) \frac{p(k+1)}{p(k)} (k+1) \frac{p_{j_1}^{\delta-1+l_1} p_{j_2}^{\delta-1+l_2}}{p_{j^*}^{\delta-1+l_1+l_2} B(\delta, k\delta)} \\
 &\quad \times \sqrt{\frac{\kappa}{2\pi}} \exp\left[-\frac{1}{2} \kappa \{(\mu_{j_1} - \zeta)^2 + (\mu_{j_2} - \zeta)^2 - (\mu_{j^*} - \zeta)^2\}\right] \\
 &\quad \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j_1}^2 \sigma_{j_2}^2}{\sigma_{j^*}^2}\right)^{-\alpha-1} \exp\left[-\beta\{\sigma_{j_1}^{-2} + \sigma_{j_2}^{-2} - \sigma_{j^*}^{-2}\}\right] \tag{4.6}
 \end{aligned}$$

Στην συνέχεια υπολογίζουμε τον λόγο $\frac{r_m(y')}{r_m(y)q(u)}$. Υπενθυμίζουμε ότι $r_m(y)$ είναι

η πιθανότητα να επιλέξουμε κίνηση τύπου m όταν είμαστε στην κατάσταση y . Εφόσον επιλέγουμε να κάνουμε διαχωρισμό μιας συνιστώσας και στην συνέχεια να ανακατατάξουμε τις παρατηρήσεις αυτής στις δύο καινούργιες, έχουμε ότι:

$r_m(y) = b_k P_{alloc}$, όπου k είναι η διάσταση όταν είμαστε στο y και P_{alloc} είναι η πιθανότητα να παρατηρηθεί η συγκεκριμένη ανακατάταξη. Οπότε:

$$P_{alloc} = \prod_{i \in \mathcal{I}} \frac{\frac{p_{z_i}}{\sigma_{z_i}} \exp\left\{-\frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2}\right\}}{\frac{p_{j_1}}{\sigma_{j_1}} \exp\left\{-\frac{(x_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right\} + \frac{p_{j_2}}{\sigma_{j_2}} \exp\left\{-\frac{(x_i - \mu_{j_2})^2}{2\sigma_{j_2}^2}\right\}} \quad (4.7)$$

όπου $\mathcal{I} = \{i : z_i = j_1\} \cup \{i : z_i = j_2\}$.

Από την άλλη, το $r_m(x')$ είναι η πιθανότητα να επιλέξουμε την αντίστροφη της προηγούμενης κίνησης, το οποίο ισοδυναμεί με το να είμαστε στην κατάσταση y' διάστασης $k+1$ και να επιλέξουμε συνδυασμό (των συνιστωσών που γεννηθήκανε στην κίνηση διαχωρισμού). Ωστε $r_m(y') = d_{k+1}$.

Επίσης, το $q(u)$ είναι η πυκνότητα του τυχαίου διανύσματος (u_1, u_2, u_3) , δηλαδή: $q(u) = g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)$ όπου $g_{p,q}$ είναι η συνάρτηση πυκνότητας της κατανομής $B(p, q)$. Επομένως

$$\frac{r_m(y')}{r_m(y)q(u)} = \frac{d_{k+1}}{b_k P_{alloc}} \{g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)\}^{-1} \quad (4.8)$$

Τέλος μένει να υπολογίσουμε την ιακωβιανή ορίζουσα του μετασχηματισμού από το $(p_{j^*}, \mu_{j^*}, \sigma_{j^*}, u_1, u_2, u_3)$ στο $(p_{j_1}, \mu_{j_1}, \sigma_{j_1}, \omega_{j_2}, \mu_{j_2}, \sigma_{j_2})$, το οποίο είναι και το πιο χρονοβόρο σημείο από θέμα πράξεων μια και ο πίνακας των μερικών παραγώγων από τον οποίο προκύπτει η παραπάνω ορίζουσα είναι διαστάσεων 6×6 . Σημειώνουμε ότι το (i, j) στοιχείο του παρακάτω πίνακα ισούται με την μερική παράγωγο του j στοιχείου του $(p_{j_1}, \mu_{j_1}, \sigma_{j_1}, \omega_{j_2}, \mu_{j_2}, \sigma_{j_2})$ ως προς το i στοιχείο του $(p_{j^*}, \mu_{j^*}, \sigma_{j^*}, u_1, u_2, u_3)$, η οποία δίνεται από την κατάλληλη εξίσωση των σχέσεων (3.11). Είναι:

$$J = \begin{vmatrix} u_1 & 0 & u_3(1-u_2^2)\frac{\sigma_j^2}{p_{j_1}} & 1-u_1 & 0 & (1-u_3)(1-u_2^2)\frac{\sigma_j^2}{p_{j_2}} \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & -\frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_2}}{p_{j_1}}} & u_3(1-u_2^2)\frac{p_{j_1}^*}{p_{j_1}} & 0 & \frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_1}}{p_{j_2}}} & (1-u_3)(1-u_2^2)\frac{p_{j_1}^*}{p_{j_2}} \\ p_{j_1}^* & 0 & 0 & -p_{j_1}^* & 0 & 0 \\ 0 & -\sigma_j^*\sqrt{\frac{p_{j_2}}{p_{j_1}}} & -2u_2u_3\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & \sigma_j^*\sqrt{\frac{p_{j_1}}{p_{j_2}}} & -2u_2(1-u_3)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \\ 0 & 0 & (1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & 0 & -(1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \end{vmatrix}$$

Ξεκινάμε την διαδικασία υπολογισμού αναπτύσσοντας κατά την πρώτη στήλη, οπότε:

$$J = u_1 A - p_{j_1}^* B \quad (4.9)$$

όπου:

$$A = \begin{vmatrix} 1 & 0 & 0 & 1 & 0 \\ -\frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_2}}{p_{j_1}}} & u_3(1-u_2^2)\frac{p_{j_1}^*}{p_{j_1}} & 0 & \frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_1}}{p_{j_2}}} & (1-u_3)(1-u_2^2)\frac{p_{j_1}^*}{p_{j_2}} \\ 0 & 0 & -p_{j_1}^* & 0 & 0 \\ -\sigma_j^*\sqrt{\frac{p_{j_2}}{p_{j_1}}} & -2u_2u_3\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & \sigma_j^*\sqrt{\frac{p_{j_1}}{p_{j_2}}} & -2u_2(1-u_3)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \\ 0 & (1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & 0 & -(1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \end{vmatrix}$$

και

$$B = \begin{vmatrix} 0 & u_3(1-u_2^2)\frac{\sigma_j^2}{p_{j_1}} & 1-u_1 & 0 & (1-u_3)(1-u_2^2)\frac{\sigma_j^2}{p_{j_2}} \\ 1 & 0 & 0 & 1 & 0 \\ -\frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_2}}{p_{j_1}}} & u_3(1-u_2^2)\frac{p_{j_1}^*}{p_{j_1}} & 0 & \frac{u_2}{2\sigma_j^*}\sqrt{\frac{p_{j_1}}{p_{j_2}}} & (1-u_3)(1-u_2^2)\frac{p_{j_1}^*}{p_{j_2}} \\ -\sigma_j^*\sqrt{\frac{p_{j_2}}{p_{j_1}}} & -2u_2u_3\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & \sigma_j^*\sqrt{\frac{p_{j_1}}{p_{j_2}}} & -2u_2(1-u_3)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \\ 0 & (1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_1}} & 0 & 0 & -(1-u_2^2)\sigma_j^2\frac{p_{j_1}^*}{p_{j_2}} \end{vmatrix}$$

Η υποορίζουσα A ισούται με $A = \Gamma - \Delta$, όπου οι Γ και Δ είναι ορίζουσες διαστάσεων 4×4 και οι οποίες αναπτύσσοντας τις κατάλληλα ισούνται με:

$$\Gamma = -p_{j^*} \begin{vmatrix} u_3(1-u_2^2) \frac{P_{j^*}}{p_{j_1}} & \frac{u_2}{2\sigma_{j^*}} \sqrt{\frac{P_{j_1}}{p_{j_2}}} & (1-u_3)(1-u_2^2) \frac{P_{j^*}}{p_{j_2}} \\ -2u_2u_3\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} & \sigma_{j^*} \sqrt{\frac{P_{j_1}}{p_{j_2}}} & -2u_2(1-u_3)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} \\ (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} & 0 & -(1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} \end{vmatrix}$$

και

$$\Delta = p_{j^*} \begin{vmatrix} -\frac{u_2}{2\sigma_{j^*}} \sqrt{\frac{P_{j_2}}{p_{j_1}}} & u_3(1-u_2^2) \frac{P_{j^*}}{p_{j_1}} & (1-u_3)(1-u_2^2) \frac{P_{j^*}}{p_{j_2}} \\ -\sigma_{j^*} \sqrt{\frac{P_{j_2}}{p_{j_1}}} & -2u_2u_3\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} & -2u_2(1-u_3)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} \\ 0 & (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} & -(1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} \end{vmatrix}$$

Για την υποορίζουσα Γ έχουμε:

$$\begin{aligned} -\frac{\Gamma}{p_{j^*}} &= (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} \left[-\frac{u_2}{2\sigma_{j^*}} \sqrt{\frac{P_{j_1}}{p_{j_2}}} 2u_2(1-u_3)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} - \sigma_{j^*} \sqrt{\frac{P_{j_1}}{p_{j_2}}} (1-u_3)(1-u_2^2) \frac{P_{j^*}}{p_{j_2}} \right] \\ &\quad - (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} \left[u_3(1-u_2^2) \frac{P_{j^*}}{p_{j_1}} \sigma_{j^*} \sqrt{\frac{P_{j_1}}{p_{j_2}}} + \frac{u_2}{2\sigma_{j^*}} \sqrt{\frac{P_{j_1}}{p_{j_2}}} 2u_2u_3\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} \right] \Rightarrow \\ \Gamma &= (1-u_2^2)\sigma_{j^*}^3 \frac{P_{j^*}}{p_{j_1}p_{j_2}} \sqrt{\frac{P_{j_1}}{p_{j_2}}}. \end{aligned} \quad (4.10)$$

Για την υποορίζουσα Δ έχουμε:

$$\begin{aligned} \frac{\Delta}{p_{j^*}} &= -\frac{u_2}{2\sigma_{j^*}} \sqrt{\frac{P_{j_2}}{p_{j_1}}} \left[2u_2u_3\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} + 2u_2(1-u_3)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} \right] \\ &\quad + \sigma_{j^*} \sqrt{\frac{P_{j_2}}{p_{j_1}}} \left[-u_3(1-u_2^2) \frac{P_{j^*}}{p_{j_1}} (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_2}} - (1-u_3)(1-u_2^2) \frac{P_{j^*}}{p_{j_2}} (1-u_2^2)\sigma_{j^*}^2 \frac{P_{j^*}}{p_{j_1}} \right] \Rightarrow \\ \Delta &= -(1-u_2^2)\sigma_{j^*}^3 \frac{P_{j^*}}{p_{j_1}p_{j_2}} \sqrt{\frac{P_{j_2}}{p_{j_1}}}. \end{aligned} \quad (4.11)$$

$$\text{Από τις (4.10) και (4.11) έχουμε ότι: } A = \Gamma - \Delta = \dots = (1-u_2^2)\sigma_{j^*}^3 \frac{P_{j^*}^4}{\sqrt{p_{j_1}p_{j_2}}}. \quad (4.15)$$

Για τον υπολογισμό της υποορίζουσας B , αναπτύσσοντας κατά την τρίτη στήλη, έχουμε ότι:

$$B = (1 - u_1)(E + Z) \quad (4.16)$$

όπου με E και Z συμβολίζουμε τις 3×3 υποορίζουσες:

$$E = \begin{vmatrix} u_3(1-u_2^2)\frac{P_{j^*}}{P_{j_1}} & \frac{u_2}{2\sigma_{j^*}}\sqrt{\frac{P_{j_1}}{P_{j_2}}} & (1-u_3)(1-u_2^2)\frac{P_{j^*}}{P_{j_2}} \\ -2u_2u_3\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} & \sigma_{j^*}\sqrt{\frac{P_{j_1}}{P_{j_2}}} & -2u_2(1-u_3)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} \\ (1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} & 0 & -(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} \end{vmatrix}$$

και

$$Z = \begin{vmatrix} -\frac{u_2}{2\sigma_{j^*}}\sqrt{\frac{P_{j_2}}{P_{j_1}}} & u_3(1-u_2^2)\frac{P_{j^*}}{P_{j_1}} & (1-u_3)(1-u_2^2)\frac{P_{j^*}}{P_{j_2}} \\ -\sigma_{j^*}\sqrt{\frac{P_{j_2}}{P_{j_1}}} & -2u_2u_3\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} & -2u_2(1-u_3)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} \\ 0 & (1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} & -(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} \end{vmatrix}$$

Για την υποορίζουσα E έχουμε:

$$\begin{aligned} E &= (1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} \left[-\frac{u_2}{2\sigma_{j^*}}\sqrt{\frac{P_{j_1}}{P_{j_2}}} 2u_2(1-u_3)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} - (1-u_3)(1-u_2^2)\frac{P_{j^*}}{P_{j_2}}\sigma_{j^*}\sqrt{\frac{P_{j_1}}{P_{j_2}}} \right] \\ &\quad - (1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} \left[u_3(1-u_2^2)\frac{P_{j^*}}{P_{j_1}}\sigma_{j^*}\sqrt{\frac{P_{j_1}}{P_{j_2}}} + \frac{u_2}{2\sigma_{j^*}}\sqrt{\frac{P_{j_1}}{P_{j_2}}} 2u_2u_3\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} \right] \\ &= -(1-u_2^2)\sigma_{j^*}^3\frac{P_{j^*}^2}{P_{j_1}P_{j_2}}\sqrt{\frac{P_{j_1}}{P_{j_2}}}(1-u_3) - (1-u_2^2)\sigma_{j^*}^3\frac{P_{j^*}^2}{P_{j_1}P_{j_2}}\sqrt{\frac{P_{j_1}}{P_{j_2}}}u_3 \Rightarrow \\ E &= -(1-u_2^2)\sigma_{j^*}^3\frac{P_{j^*}^2}{P_{j_1}P_{j_2}}\sqrt{\frac{P_{j_1}}{P_{j_2}}} \end{aligned} \quad (4.17)$$

Για την υποορίζουσα Z έχουμε:

$$\begin{aligned} Z &= -\frac{u_2}{2\sigma_{j^*}}\sqrt{\frac{P_{j_2}}{P_{j_1}}} \left[2u_2u_3\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}}(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} + 2u_2(1-u_3)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}}(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} \right] \\ &\quad + \sigma_{j^*}\sqrt{\frac{P_{j_2}}{P_{j_1}}} \left[-u_3(1-u_2^2)\frac{P_{j^*}}{P_{j_1}}(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_2}} - (1-u_3)(1-u_2^2)\frac{P_{j^*}}{P_{j_2}}(1-u_2^2)\sigma_{j^*}^2\frac{P_{j^*}}{P_{j_1}} \right] \end{aligned}$$

$$\begin{aligned}
&= -\frac{u_2}{\sigma_{j^*}} \sqrt{\frac{p_{j_2}}{p_{j_1}}} u_2 \sigma_{j^*}^4 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} (1-u_2^2) - \sigma_{j^*} \sqrt{\frac{p_{j_2}}{p_{j_1}}} (1-u_2^2)^2 \sigma_{j^*}^2 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \\
&= -u_2^2 \sqrt{\frac{p_{j_2}}{p_{j_1}}} (1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} - (1-u_2^2)^2 \sqrt{\frac{p_{j_2}}{p_{j_1}}} \sigma_{j^*}^3 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \Rightarrow \\
Z &= -\sqrt{\frac{p_{j_2}}{p_{j_1}}} (1-u_2^2) \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \sigma_{j^*}^3 \tag{4.18}
\end{aligned}$$

Αντικαθιστώντας τις (4.17) και (4.18) στην (4.16) παίρνουμε ότι:

$$\begin{aligned}
B &= (1-u_1) \left(- (1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \sqrt{\frac{p_{j_1}}{p_{j_2}}} - \sqrt{\frac{p_{j_2}}{p_{j_1}}} (1-u_2^2) \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \sigma_{j^*}^3 \right) \\
&= -(1-u_1)(1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \left(\sqrt{\frac{p_{j_1}}{p_{j_2}}} + \sqrt{\frac{p_{j_2}}{p_{j_1}}} \right) \\
&= -(1-u_1)(1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^2}{p_{j_1} p_{j_2}} \frac{p_{j^*}}{\sqrt{p_{j_1} p_{j_2}}} \Rightarrow \\
B &= -(1-u_1)(1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^3}{p_{j_1} p_{j_2} \sqrt{p_{j_1} p_{j_2}}} \tag{4.19}
\end{aligned}$$

Οπότε αντικαθιστώντας τις (4.15) και (4.19) στην (4.9) λαμβάνουμε τελικά ότι:

$$\begin{aligned}
J &= u_1 (1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^4}{\sqrt{p_{j_1} p_{j_2}}} + p_{j^*} (1-u_1)(1-u_2^2) \sigma_{j^*}^3 \frac{p_{j^*}^3}{p_{j_1} p_{j_2} \sqrt{p_{j_1} p_{j_2}}} \Rightarrow \\
J &= \frac{(1-u_2^2) \sigma_{j^*}^3 p_{j^*}^4}{p_{j_1} p_{j_2} \sqrt{p_{j_1} p_{j_2}}} \tag{4.20}
\end{aligned}$$

η οποία ολοκληρώνει την διαδικασία υπολογισμού της ιακωβιανής του μετασχηματισμού. Σημειώνουμε τέλος ότι στην σχέση (4.20) μπορούμε να κάνουμε αντικατάσταση των σχέσεων:

$$\begin{aligned}
\mu_{j_1} &= \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{\frac{p_{j_2}}{p_{j_1}}} & \mu_{j_2} &= \mu_{j^*} + u_2 \sigma_{j^*} \sqrt{\frac{p_{j_1}}{p_{j_2}}} \\
\sigma_{j_1}^2 &= u_3 (1-u_2^2) \sigma_{j^*}^2 \frac{p_{j^*}}{p_{j_1}} & \sigma_{j_2}^2 &= (1-u_3)(1-u_2^2) \sigma_{j^*}^2 \frac{p_{j^*}}{p_{j_2}}
\end{aligned}$$

και

ώστε να προκύψει η ισοδύναμη έκφραση:

$$J = \frac{p_{j^*} |\mu_{j_1} - \mu_{j_2}| \sigma_{j_1}^2 \sigma_{j_2}^2}{u_2 (1 - u_2^2) u_3 (1 - u_3) \sigma_{j^*}^2} \quad (4.21)$$

η οποία είναι αυτή που χρησιμοποιείται από τους Richardson & Green.

Ολοκληρώνουμε τώρα την απόδειξη της πρότασης, αντικαθιστώντας τις σχέσεις (4.6), (4.7), (4.8) και (4.21) στον τύπο της πιθανότητας αποδοχής:

$$\begin{aligned} A = (\text{likelihood ratio}) & \frac{p(k+1)}{p(k)} (k+1) \frac{p_{j_1}^{\delta-1+l_1} p_{j_2}^{\delta-1+l_2}}{p_{j^*}^{\delta-1+l_1+l_2} B(\delta, k\delta)} \\ & \times \sqrt{\frac{\kappa}{2\pi}} \exp\left[-\frac{1}{2} \kappa \left\{ (\mu_{j_1} - \xi)^2 + (\mu_{j_2} - \xi)^2 - (\mu_{j^*} - \xi)^2 \right\}\right] \\ & \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j_1}^2 \sigma_{j_2}^2}{\sigma_{j^*}^2} \right)^{-\alpha-1} \exp\left[-\beta \left\{ \sigma_{j_1}^{-2} + \sigma_{j_2}^{-2} - \sigma_{j^*}^{-2} \right\}\right] \\ & \times \frac{d_{k+1}}{b_k P_{alloc}} \left\{ g_{2,2}(u_1) g_{2,2}(u_2) g_{1,1}(u_3) \right\}^{-1} \\ & \times \frac{p_{j^*} |\mu_{j_1} - \mu_{j_2}| \sigma_{j_1}^2 \sigma_{j_2}^2}{u_2 (1 - u_2^2) u_3 (1 - u_3) \sigma_{j^*}^2} \end{aligned}$$

η οποία είναι και η ζητούμενη σχέση.

Τέλος για την αντίστροφη κίνηση συνδυασμού εφαρμόζουμε την ανάλογη πορεία και καταλήγουμε στην σχέση (3.13).

4.2 Απόδειξη Πρότασης 3.1.3

Για την περίπτωση γέννησης μιας κενής συνιστώσας, έχουμε ότι:
 $\alpha_m(y, y') = \min\{1, A\}$ όπου

$$A = \frac{p(y'|x)r_m(y')}{p(y|x)r_m(y)q(u)} \left| \frac{\partial(y')}{\partial(y, u)} \right|.$$

Αρχικά υπολογίζουμε τον λόγο $\frac{p(y'|x)}{p(y|x)q(u)}$. Κατ' αρχάς από την (4.1) έχουμε ότι:

$$\frac{p(y'|x)}{p(y|x)} = \frac{p(x|\theta', \underline{z}')}{p(x|\underline{\theta}, \underline{z})} \frac{p(\theta'|k+1, \eta)}{p(\theta|k, \eta)} \frac{p(\underline{z}'|\underline{p}', k+1)}{p(\underline{z}|\underline{p}, k)} \frac{p(\underline{p}'|k+1, \delta)}{p(\underline{p}|k, \delta)} \frac{p(k+1)}{p(k)}$$

Εφόσον η κατανομή των παρατηρήσεων στις συνιστώσες δεν αλλάζει, έπεται ότι ο λόγος πιθανοφανειών θα είναι ίσος με 1, μια και $\underline{z}' = \underline{z}$.

Στην συνέχεια λαμβάνουμε:

$$\begin{aligned} \frac{p(\theta'|k+1, \eta)}{p(\theta|k, \eta)q(u)} &= \frac{(k+1)! \prod_{i=1}^{k+1} (\mu_i|k+1, \xi) \text{INVT}(\sigma_i|k+1, \xi)}{k! \prod_{i=1}^k (\mu_i|k, \xi) \text{INVT}(\sigma_i|k, \xi)} \frac{1}{q(u)} \\ &= \frac{k+1}{q(u)} \sqrt{\frac{\kappa}{2\pi}} \exp\left\{-\frac{1}{2}\kappa(\mu_{j^*} - \xi)^2\right\} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\sigma_{j^*}^2)^{\alpha+1}} \exp\{-\beta\sigma_{j^*}^{-2}\} \end{aligned}$$

Επειδή όμως

$$q(u) = g_{1,k}(p_{j^*}) \sqrt{\frac{\kappa}{2\pi}} \exp\left\{-\frac{1}{2}\kappa(\mu_{j^*} - \xi)^2\right\} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\sigma_{j^*}^2)^{\alpha+1}} \exp\{-\beta\sigma_{j^*}^{-2}\}$$

προκύπτει ότι:

$$\frac{p(\theta'|k+1, \eta)}{p(\theta|k, \eta)q(u)} = \frac{k+1}{g_{1,k}(p_{j^*})} \quad (4.22)$$

Έπειτα έχουμε:

$$\frac{p(\underline{z}'|\underline{p}', k+1)}{p(\underline{z}|\underline{p}, k)} = \frac{\prod_{i=1}^n p'_{z_i}}{\prod_{i=1}^n p_{z_i}} = \frac{\prod_{i=1}^n p_{z_i} (1-p_{j^*})}{\prod_{i=1}^n p_{z_i}} = (1-p_{j^*})^n \quad (4.23)$$

Για τον επόμενο όρο στο γινόμενο έχουμε:

$$\frac{p(\underline{p}'|k+1, \delta)}{p(\underline{p}|k, \delta)} = \frac{\Gamma((k+1)\delta) \prod_{i=1}^{k+1} (p'_i)^{\delta-1}}{(\Gamma(\delta))^{k+1}} = \frac{1}{B(\delta, k\delta)} \frac{p_{j^*}^{\delta-1} \prod_{i \in \{1, 2, \dots, k+1\} \setminus \{j^*\}} p_i^{\delta-1} (1-p_{j^*})^{\delta-1}}{\prod_{i=1}^k p_i^{\delta-1}} \Rightarrow$$

$$\frac{p(\underline{p}'|k+1, \delta)}{p(\underline{p}|k, \delta)} = \frac{p_{j^*}^{\delta-1} (1-p_{j^*})^{k(\delta-1)}}{B(\delta, k\delta)} \quad (4.24)$$

Από τις (4.22), (4.23) και (4.24) έχουμε ότι:

$$\frac{p(y'|x)}{p(y|x)q(u)} = \frac{p(k+1)}{p(k)} \frac{1}{B(k\delta, \delta)} p_{j^*}^{\delta-1} (1-p_{j^*})^{n+k(\delta-1)} \frac{k+1}{g_{1,k}(p_{j^*})} \quad (4.25)$$

Στην συνέχεια υπολογίζουμε τον λόγο: $\frac{r_m(y')}{r_m(y)}$. Υπενθυμίζουμε ότι $r_m(y)$ είναι η πιθανότητα να επιλέξουμε κίνηση τύπου m όταν είμαστε στην κατάσταση y . Εφόσον επιλέγουμε να κάνουμε γέννηση μιας (κενής) συνιστώσας έχουμε ότι: $r_m(y) = b_k$, όπου k είναι η διάσταση όταν είμαστε στο y . Για τον υπολογισμό του $r_m(y')$ έστω ότι πριν την γέννηση οι κενές συνιστώσες είναι πλήθους $k_0 < k$ το πλήθος, και από την κατάσταση y' διάστασης $k+1$ θα πάμε στην y μέσω του θανάτου μιας από τις k_0+1 κενές συνιστώσες, έστω της j^* . Η κίνηση θανάτου συμβαίνει με πιθανότητα d_{k+1} ενώ η συνιστώσα j^* επιλέγεται για να διαγραφεί με πιθανότητα $1/(k_0+1)$.

Οπότε $r_m(y') = d_{k+1} \frac{1}{k_0+1}$. Ωστε:

$$\frac{r_m(y')}{r_m(y)} = \frac{d_{k+1}}{(k_0+1)b_k} \quad (4.26)$$

Τέλος μένει ο υπολογισμός της ιακωβιανής του μετασχηματισμού από το $\underline{p} = (p_1, p_2, \dots, p_k)$ στο $\underline{p}' = (p_1(1-p_{j^*}), p_2(1-p_{j^*}), \dots, p_{j^*}, \dots, p_k(1-p_{j^*}))$. Εφόσον

$p'_j = p_j(1-p_{j^*})$ έχουμε ότι: $\frac{\partial p'_j}{\partial p_j} = 1-p_{j^*}, \forall j=1, 2, \dots, k$. Προσοχή θα πρέπει να

δοθεί στον περιορισμό $\sum_j p_j = 1$ που στην ουσία έχει σαν συνέπεια η ιακωβιανή ορίζουσα να είναι διάστασης $(k-1) \times (k-1)$ (ένα γεγονός που διέφυγε της προσοχής των Richardson & Green στην εργασία τους το 1997 ώσπου να διορθωθεί αργότερα):

$$J = \begin{vmatrix} 1-p_{j^*} & \mathbf{K} & \mathbf{K} & \mathbf{K} \\ \mathbf{K} & 1-p_{j^*} & \mathbf{K} & \mathbf{K} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{K} & \mathbf{K} & \mathbf{K} & 1-p_{j^*} \end{vmatrix} \Rightarrow J = (1-p_{j^*})^{k-1}.$$

Συνεπώς για την κίνηση γέννησης η πιθανότητα αποδοχής ισούται με $\alpha_m(y, y') = \min\{1, A\}$ όπου:

$$A = \frac{p(k+1)}{p(k)} \frac{1}{B(k\delta, \delta)} p_{j^*}^{\delta-1} (1-p_{j^*})^{n+k(\delta-1)} (k+1) \\ \times \frac{d_{k+1}}{(k_0+1)b_k} \frac{1}{g_{1,k}(p_{j^*})} (1-p_{j^*})^{k-1}.$$

Τέλος για την αντίστροφη κίνηση θανάτου, εφαρμόζουμε την ανάλογη πορεία και καταλήγουμε στην σχέση $\alpha_m(y, y') = \min\{1, A^{-1}\}$, όπου

$$A^{-1} = \frac{p(k-1)}{p(k)} B((k-1)\delta, \delta) \frac{1}{p_{j^*}^{\delta-1} (1-p_{j^*})^{n+(k-1)(\delta-1)}} \frac{1}{k} \\ \times \frac{b_{k-1}k_0}{d_k} g_{1,k-1}(p_{j^*}) \frac{1}{(1-p_{j^*})^{(k-1)-1}}.$$

4.3 Κώδικας αλγορίθμου RJMCMC στην Fortran

USE NUMERICAL_LIBRARIES

IMPLICIT NONE

```

INTEGER N, F, I, ITER, J, M, Q, IND, WN, K, INDE, SPLIT, COND1, COND2,      &
      COMBINE1, COMBINE2, AST, SKCOMB, EMPTY, BIRTH, THESI, DEATH,      &
      BMLA, SMLA, CMLA, DMLA, RAM, BURN, ODIM, PL
INTEGER ,ALLOCATABLE :: SK(:), KSIM(:,:), KMAX(:), KMAXPROP(:), KSIMNEW(:,:), EMPTIES(:)
INTEGER SKNEW(2), KSIMPROP(82), KSIMNEWWORD(82)
REAL(8) PSIMPROP(2), MUPROP(2), S2PROP(2), DIST(1)
REAL(8) ,ALLOCATABLE :: X(:), BHTA(:), SKX(:), KDIST(:), SKU(:), B(:),      &
      D(:), KLIST(:), W(:), PINAK(:), MIU(:), VAR(:),      &
      PITH(:), VARPINAK(:), PROPPINAK(:), DIMWEIGHTS(:)
REAL(8) ,ALLOCATABLE :: PSIM(:,:), S2(:,:), MU(:,:), MUDOK(:), S2DOK(:),      &
      PSIMDOK(:), MUPROPOSED(:,:), S2PROPOSED(:,:),      &
      PSIMPROPOSED(:,:), MEAN5(:,:), VAR5(:,:), WEIGHT5(:,:)
REAL(8) MY_GAMMA_JOHNK, MY_BETA_JOHNK, MY_NORMAL, MY_NORMAL, DEN, DENKDIST, DRNUNF,      &
      U, PSIM, R, KSI, KAPPA, G, H, DELTA, PALLOC1, SB, ALPHA, U1, U2,      &
      U3, PALLOC, PI, LOGLIKELIHOODRATIO1, LOGLIKELIHOODRATIO2, LAP,      &
      LAP1, PDFBETA, PSIMCOMB, MUCOMB, S2COMB, PBIRTH, MUBIRTH, S2BIRTH,      &
      S, PDEATH, MUDEATH, S2DEATH, SPROP, CPROP, BPROP, DPROP, SCPROP, BDPROP, sw

```

PI=3.14151926535897932d0

N=82

ALLOCATE (X(N))

!***** *GALAXY DATA SET*****

```

X(1)=9.172d0; X(43)=20.875d0;
X(2)=9.35d0; X(44)=20.986d0;
X(3)=9.483d0; X(45)=21.137d0;
X(4)=9.558d0; X(46)=21.492d0;
X(5)=9.775d0; X(47)=21.701d0;
X(6)=10.227d0; X(48)=21.814d0;
X(7)=10.406d0; X(49)=21.921d0;
X(8)=16.084d0; X(50)=21.96d0;
X(9)=16.170d0; X(51)=22.185d0;
X(10)=18.419d0; X(52)=22.209d0;
X(11)=18.552d0; X(53)=22.242d0;
X(12)=18.6d0; X(54)=22.249d0;
X(13)=18.927d0; X(55)=22.314d0;
X(14)=19.052d0; X(56)=22.374d0;
X(15)=19.07d0; X(57)=22.495d0;
X(16)=19.33d0; X(58)=22.746d0;
X(17)=19.343d0; X(59)=22.747d0;
X(18)=19.349d0; X(60)=22.888d0;
X(19)=19.440d0; X(61)=22.914d0;
X(20)=19.473d0; X(62)=23.206d0;
X(21)=19.529d0; X(63)=23.241d0;
X(22)=19.541d0; X(64)=23.263d0;
X(23)=19.547d0; X(65)=23.484d0;
X(24)=19.663d0; X(66)=23.538d0;
X(25)=19.846d0; X(67)=23.542d0;
X(26)=19.856d0; X(68)=23.666d0;
X(27)=19.863d0; X(69)=23.706d0;
X(28)=19.914d0; X(70)=23.711d0;
X(29)=19.918d0; X(71)=24.129d0;
X(30)=19.973d0; X(72)=24.285d0;
X(31)=19.989d0; X(73)=24.289d0;
X(32)=20.166d0; X(74)=24.366d0;
X(33)=20.175d0; X(75)=24.717d0;
X(34)=20.179d0; X(76)=24.990d0;
X(35)=20.196d0; X(77)=25.633d0;
X(36)=20.215d0; X(78)=26.96d0;
X(37)=20.221d0; X(79)=26.995d0;
X(38)=20.415d0; X(80)=32.065d0;
X(39)=20.629d0; X(81)=32.789d0;
X(40)=20.795d0; X(82)=34.279d0;
X(41)=20.821d0;
X(42)=20.846d0;

```

```
CALL REAL_HISTOGRAM(10,'GALAXY VELOCITIES',X,N,0,N,45)
```

```
!*****GREEN'S MODEL ASSUMPTIONS*****
```

```
K=30
ALLOCATE ( B(K), D(K), DIMWEIGHTS(K) )
B(1)=1.0d0
D(K)=1.0d0
B(K)=0.0d0
D(1)=0.0d0
DO I=2, K-1
    B(I)=0.5d0
    D(I)=0.5d0
END DO
R=(X(N)-X(1))
KSI=0.0d0
DO I=1, N
    KSI=KSI+X(I)
END DO
KSI=KSI/N
KAPPA=1/(R**2)
G=0.2d0
H=10.0d0/(R**2)
ALPHA=2.0d0
DELTA=1.0d0
PL=20
```

```
!*****NUMBER OF ITERATIONS*****
```

```
M=300000
```

```
!*****Burn in period*****
```

```
BURN=M/10
```

```
!*****INITIAL VALUES*****
```

```
ALLOCATE ( KMAX(M), KMAXPROP(M), BHTA(M) )
```

```
BHTA(1)=MY_GAMMA_JOHNK(G,H)
KMAX(1)=1
ODIM=5
```

```
ALLOCATE ( PSIM(K,M), S2(K,M), MU(K,M), KSIM(N,M), KSIMNEW(N,M), &
    MUPROPOSED(K,M), S2PROPOSED(K,M), PSIMPROPOSED(K,M), MEAN5(ODIM,M), &
    VAR5(ODIM,M), WEIGHT5(ODIM,M) )
```

```
SMLA=0
CMLA=0
BMLA=0
DMLA=0
RAM=0
SPROP=0.0d0
CPROP=0.0d0
BPROP=0.0d0
DPROP=0.0d0
SCPROP=0.0d0
BDPROP=0.0d0
DO I=1, KMAX(1)
    PSIM(I,1)=1.0d0/KMAX(1)
    S2(I,1)=1.0d0/MY_GAMMA_JOHNK(ALPHA,B(1))
    MU(I,1)=MY_NORMAL(KSI,1/KAPPA)
END DO
```

```
CALL DSVRGN (KMAX(1), MU(:,1), MU(:,1))
```

```
DO I=1, K
```

Κεφάλαιο 4: Παράρτημα

```

DIMWEIGHTS(I)=0.0d0
END DO

!*****
!*****REVERSIBLE JUMP MCMC*****
!*****

DO ITER=2, M
  ALLOCATE ( SK(KMAX(ITER-1)), SKX(KMAX(ITER-1)), KDIST(KMAX(ITER-1)), &
    SKU(KMAX(ITER-1)), MUDOK(KMAX(ITER-1)), S2DOK(KMAX(ITER-1)), &
    PSIMDOK(KMAX(ITER-1)))

  DO I=1, KMAX(ITER-1)
    SK(I)=0
    SKX(I)=0.0d0
  END DO

!*****SIMULATION OF ALLOCATION VARIABLES K(1),...,K(N)*****

  DO J=1, N
    DENKDIST=0.0d0
    DO Q=1, KMAX(ITER-1)
      DENKDIST=DENKDIST+PSIM(Q,ITER-1)* &
        DEXP(-(X(J)-MU(Q,ITER-1))*2.0d0)/(2.0d0*S2(Q,ITER-1)))/(DSQRT(S2(Q,ITER-1)))
    END DO

    KDIST(1)=(PSIM(1,ITER-1)* &
      DEXP(-(X(J)-MU(1,ITER-1))*2.0d0)/(2.0d0*S2(1,ITER-1)))/ &
      (DSQRT(S2(1,ITER-1)))/DENKDIST

    DO I=2, KMAX(ITER-1)
      KDIST(I)=KDIST(I-1)+(PSIM(I,ITER-1)* &
        DEXP(-(X(J)-MU(I,ITER-1))*2.0d0)/(2.0d0*S2(I,ITER-1)))/ &
        (DSQRT(S2(I,ITER-1)))/DENKDIST
    END DO

    F=1
    U=DRNUNF()
    DO WHILE (U>KDIST(F))
      F=F+1
    END DO
    KSIM(J,ITER)=F
    SK(F)=SK(F)+1
    SKX(F)=SKX(F)+X(J)

  END DO

  DO J=1, KMAX(ITER-1)
    SKU(J)=0.0d0
  END DO

!*****SIMULATION OF MU(1),...,MU(KMAX)*****

  MUDOK(1)=MY_NORMAL(((SKX(1)/S2(1,ITER-1))+KAPPA*KSI)/(KAPPA+(SK(1)/S2(1,ITER-1))), &
    1.0d0/(KAPPA+((SK(1)+0.0d0)/S2(1,ITER-1))))

  DO J=1, (KMAX(ITER-1)-1)
    MUDOK(J+1)= &
      MY_NORMAL(((SKX(J+1)/S2(J+1,ITER-1))+KAPPA*KSI)/(KAPPA+(SK(J+1)/S2(J+1,ITER-1))),
      &
      1.0d0/(KAPPA+((SK(J+1)+0.0d0)/S2(J+1,ITER-1))))

    IF ( MUDOK(J+1)<MUDOK(J) ) THEN
      MUDOK(J+1)=MU(J+1,ITER-1)
    IF ( MUDOK(J)>MUDOK(J+1) ) THEN
      DO I=1, J+1
        MUDOK(I)=MU(I,ITER-1)
      END DO
    END IF
  END DO

```

```

                END IF
            END IF

        END DO

    DO J= 1, KMAX(ITER-1)

        DO I=1, N

            IF ( KSIM(I,ITER)==J ) THEN
                INDE=1
            ELSE
                INDE=0
            END IF

            SKU(J)=SKU(J)+(INDE+0.0d0)*(X(I)-MUDOK(J))*2.0d0

        END DO

    END DO

!*****SIMULATION OF P(1),...,P(KMAX)*****

    DO I=1, KMAX(ITER-1)
        PSIMDOK(I)=MY_GAMMA_JOHNK(DELTA+SK(I)+0.0d0,1.0d0)
    END DO

    SPSIM=0.0d0

    DO I=1, KMAX(ITER-1)
        SPSIM=SPSIM+PSIMDOK(I)
    END DO

    DO I=1, KMAX(ITER-1)
        PSIMDOK(I)=PSIMDOK(I)/SPSIM
    END DO

!*****SIMULATION OF S2(1),...,S2(KMAX)*****

    DO J=1, KMAX(ITER-1)
        S2DOK(J)=1/MY_GAMMA_JOHNK(ALPHA+(SK(J)+0.0d0)/2.0d0,BHTA(ITER-1)+0.5d0*SKU(J))
    END DO

!*****Simulation of β*****

    S=0.0d0
    DO J=1, KMAX(ITER-1)
        S=S+1.0d0/S2DOK(J)
    END DO
    BHTA(ITER)=MY_GAMMA_JOHNK(G+(KMAX(ITER-1)+0.0d0)*ALPHA,H+S)

!*****MOVE TYPE SPLIT - COMBINE*****
!*****

    IF ( DRNUNF(<B(KMAX(ITER-1))) ) THEN

!*****SPLIT MOVE*****

        SMLA=SMLA+1

        U1=MY_BETA_JOHNK(2.0d0,2.0d0)
        U2=MY_BETA_JOHNK(2.0d0,2.0d0)
        U3=MY_BETA_JOHNK(1.0d0,1.0d0)

```



```
KMAXPROP(ITER)=KMAX(ITER-1)+1
SPLIT=FLOOR(KMAX(ITER-1)*DRNUNF()+1
```

!*****Deterministic proposals for the split move*****

```
PSIMPROP(1)=PSIMDOK(SPLIT)*U1
PSIMPROP(2)=PSIMDOK(SPLIT)*(1.0d0-U1)
MUPROP(1)=MUDOK(SPLIT)-U2*DSQRT(S2DOK(SPLIT)*(PSIMPROP(2)/PSIMPROP(1)))
MUPROP(2)=MUDOK(SPLIT)+U2*DSQRT(S2DOK(SPLIT)*(PSIMPROP(1)/PSIMPROP(2)))
S2PROP(1)=U3*(1.0d0-U2**2.0d0)*S2DOK(SPLIT)*PSIMDOK(SPLIT)/PSIMPROP(1)
S2PROP(2)=(1.0d0-U3)*(1.0d0-U2**2.0d0)*S2DOK(SPLIT)*PSIMDOK(SPLIT)/PSIMPROP(2)
```

!*****Checking ordering of the proposals for the means*****

```
IF ( KMAX(ITER-1)==1 ) THEN
  COND1=2
ELSE
  IF ( SPLIT==1 ) THEN

    IF ( MAX(MUPROP(1),MUPROP(2))>MUDOK(SPLIT+1) ) THEN
      COND1=1
    ELSE
      COND1=2
    END IF

  ELSE

    IF ( SPLIT==KMAX(ITER-1) ) THEN
      IF ( MIN(MUPROP(1),MUPROP(2))<MUDOK(SPLIT-1) ) THEN
        COND1=1
      ELSE
        COND1=2
      END IF
    ELSE
      IF ( ( MIN(MUPROP(1),MUPROP(2))<MUDOK(SPLIT-1) ) .OR. &
        ( MAX(MUPROP(1),MUPROP(2))>MUDOK(SPLIT+1) ) ) THEN
        COND1=1
      ELSE
        COND1=2
      END IF
    END IF
  END IF
END IF
```

!*****If cond1 equals 1 then at least one ordering constraint is violated, so we remain at the previous state.*****

```
IF ( COND1==1 ) THEN
  DO I=1, KMAX(ITER-1)
    MU(I,ITER)=MUDOK(I)
    S2(I,ITER)=S2DOK(I)
    PSIM(I,ITER)=PSIMDOK(I)
    KMAX(ITER)=KMAX(ITER-1)
  END DO
  KMAX(ITER)=KMAX(ITER-1)
  DO J=1, N
    KSIMNEW(J,ITER)=KSIM(J,ITER)
  END DO
ELSE
```

!*****Reallocation proposals for the split move.*****

```
DO I=1, 2
  SKNEW(I)=0
```

```

END DO

PALLOC1=1.0d0

DO J=1, N

  IF ( KSIM(J,ITER)<SPLIT ) THEN
    KSIMNEW(J,ITER)=KSIM(J,ITER)
  ELSE
    IF ( KSIM(J,ITER)>SPLIT ) THEN
      KSIMNEW(J,ITER)=KSIM(J,ITER)+1
    ELSE
      IF ( KSIM(J,ITER)==SPLIT ) THEN

        DEN=0.0d0
        DO Q=1, 2
          DEN=DEN+PSIMPROP(Q)* &
            DEXP(-(X(J)- &
              MUPROP(Q))*2.0d0)/ &
              (2.0d0*S2PROP(Q))/ &
              (DSQRT(ABS(S2PROP(Q))))
        END DO

        DIST(1)=(PSIMPROP(1)* &
          DEXP(-(X(J)- &
            MUPROP(1))*2.0d0)/ &
            (2.0d0*S2PROP(1))/ &
            (DSQRT(ABS(S2PROP(1)))))/DEN

        U=DRNUNF()
        IF ( U<DIST(1) ) THEN
          KSIMPROP(J)=1
          KSIMNEW(J,ITER)=KSIM(J,ITER)
          SKNEW(1)=SKNEW(1)+1
        ELSE
          KSIMPROP(J)=2
          KSIMNEW(J,ITER)=KSIM(J,ITER)+1
          SKNEW(2)=SKNEW(2)+1
        END IF

        palloc1=palloc1*PSIMPROP(KSIMPROP(J))/&
          DSQRT( ABS(S2PROP(KSIMPROP(J))) ) * &
          dexp(-(X(J)-
            MUPROP(KSIMPROP(J))*2.0d0)/ &
            (2.0d0*S2PROP(KSIMPROP(J))))/DEN
      END IF
    END IF
  END IF

  END DO

  DO I=1, SPLIT-1
    MUPROPOSED(I,ITER)=MUDOK(I)
    S2PROPOSED(I,ITER)=S2DOK(I)
    PSIMPROPOSED(I,ITER)=PSIMDOK(I)
  END DO

  DO I=1, 2
    MUPROPOSED(SPLIT-1+I,ITER)=MUPROP(I)
    S2PROPOSED(SPLIT-1+I,ITER)=S2PROP(I)
    PSIMPROPOSED(SPLIT-1+I,ITER)=PSIMPROP(I)
  END DO

  DO I=SPLIT+1, KMAX(ITER-1), 1
    MUPROPOSED(I+1,ITER)=MUDOK(I)
    S2PROPOSED(I+1,ITER)=S2DOK(I)
    PSIMPROPOSED(I+1,ITER)=PSIMDOK(I)
  END DO

  !*****Logarithm of Acceptance Probability (LAP) for the split move*****

  LOGLIKELIHOODRATIO2=1.0d0
  DO J=1, N
    LOGLIKELIHOODRATIO2=LOGLIKELIHOODRATIO2*&
      DSQRT(S2DOK(KSIM(J,ITER)))/&
      S2PROPOSED(KSIMNEW(J,ITER),ITER))* &

```

666

```

DEXP( ((X(J)-MUDOK( KSIM(J,ITER)
))*2.0d0)/(2.0d0*S2DOK(KSIM(J,ITER))))- (((X(J)- &
MUPROPOSED( KSIMNEW(J,ITER),ITER ))**2.0d0)/&
(2.0d0*S2PROPOSED(KSIMNEW(J,ITER),ITER))) )

END DO

LAP1=LOGLIKELIHOODRATIO2*(KMAX(ITER-1)+1.0d0)* &
( (PSIMPROP(1)**(DELTA-1.0d0+SKNEW(1))) * &
(PSIMPROP(2)**(DELTA-1.0d0+SKNEW(2)))/ &
(PSIMDOK(SPLIT)**(DELTA-1.0d0+SKNEW(1)+SKNEW(2))) ) * &
DSQRT( KAPPA/(2.0d0*PI) ) *DEXP( -0.5d0*KAPPA*( ((MUPROP(1)-KSI)**2)+ &
((MUPROP(2)-KSI)**2)-((MUDOK(SPLIT)-KSI)**2)) ) &
*((BHTA(ITER)**ALPHA)/DGAMMA(ALPHA)) * &
(( S2PROP(1)*S2PROP(2)/S2DOK(SPLIT) ) **(-ALPHA-1.0d0)) * &
DEXP( -BHTA(ITER)*(1.0d0/S2PROP(1))+1.0d0/S2PROP(2) ) &
-(1.0d0/S2DOK(SPLIT))) ) * D(KMAX(ITER-1)+1)/(B(KMAX(ITER-1))*PALLOC1) ) &
*(1.0d0/( PDFBETA(2.0d0, 2.0d0, U1)*PDFBETA(2.0d0, 2.0d0, U2)* &
PDFBETA(1.0d0, 1.0d0, U3) ))*( PSIMDOK(SPLIT)*ABS(MUPROP(1)-MUPROP(2)) ) &
S2PROP(1)*S2PROP(2)/(U2*(1.0d0-U2**2.0d0)*U3*(1.0d0-U3)*S2DOK(SPLIT)) )

IF ( DRNUNF()<LAP1 ) THEN
DO I=1, KMAXPROP(ITER)
MU(I,ITER)=MUPROPOSED(I,ITER)
S2(I,ITER)=S2PROPOSED(I,ITER)
PSIM(I,ITER)=PSIMPROPOSED(I,ITER)
END DO
KMAX(ITER)=KMAXPROP(ITER)
SPROP=SPROP+1.0d0
SCPROP=SCPROP+1.0d0

ELSE
DO I=1, KMAX(ITER-1)
MU(I,ITER)=MUDOK(I)
S2(I,ITER)=S2DOK(I)
PSIM(I,ITER)=PSIMDOK(I)
END DO
DO J=1, N
KSIMNEW(J,ITER)=KSIM(J,ITER)
END DO
KMAX(ITER)=KMAX(ITER-1)
END IF
END IF
ELSE
|*****COMBINE MOVE*****
CMLA=CMLA+1
KMAXPROP(ITER)=KMAX(ITER-1)-1
COMBINE1=FLOOR(KMAX(ITER-1)*DRNUNF()+1)
IF ( COMBINE1==1 ) THEN
COMBINE2=2
ELSE
IF ( COMBINE1==KMAX(ITER-1) ) THEN
COMBINE2=KMAX(ITER-1)-1
ELSE
IF ( DRNUNF()<0.50d0 ) THEN
COMBINE2=COMBINE1-1
ELSE
COMBINE2=COMBINE1+1
END IF
END IF
END IF
AST=MIN(COMBINE1,COMBINE2)
|*****Deterministic proposals for the combine move*****
PALLOC1=1.0d0
SKCOMB=0
DO J=1, N
IF ( ( KSIM(J,ITER)==COMBINE1).OR.(KSIM(J,ITER)==COMBINE2) ) THEN

```

```

DEN=PSIMDOK(COMBINE1)*DEXP(-(X(J)-
MUDOK(COMBINE1)**2)/(2*S2DOK(COMBINE1)))/ &
(DSQRT(ABS(S2DOK(COMBINE1)))+PSIMDOK(COMBINE2)* &
DEXP(-(X(J)-MUDOK(COMBINE2)**2)/(2*S2DOK(COMBINE2)))/ &
(DSQRT(ABS(S2DOK(COMBINE2)))))

KSIMPROP(J)=AST
SKCOMB=SKCOMB+1
palloc1=palloc1*(PSIMDOK(KSIM(J,ITER))/DSQRT(S2DOK(KSIM(J,ITER)))*&
dexp(-(X(J)-MUDOK(KSIM(J,ITER))**2.0d0)/ &
(2.0d0*S2DOK(KSIM(J,ITER)))/DEN

END IF
END DO

PSIMCOMB=PSIMDOK(COMBINE1)+PSIMDOK(COMBINE2)
MUCOMB=(PSIMDOK(COMBINE1)*MUDOK(COMBINE1)+ &
PSIMDOK(COMBINE2)*MUDOK(COMBINE2))/PSIMCOMB &
S2COMB=(PSIMDOK(COMBINE1)*(MUDOK(COMBINE1)**2.0d0+S2DOK(COMBINE1)) &
+PSIMDOK(COMBINE2)*(MUDOK(COMBINE2)**2.0d0+S2DOK(COMBINE2))- &
PSIMCOMB*MUCOMB**2.0d0)/PSIMCOMB
U1=PSIMDOK(COMBINE1)/PSIMCOMB
U2=(MUCOMB-MUDOK(COMBINE1))/ &
(DSQRT(S2COMB*PSIMDOK(COMBINE2)/PSIMDOK(COMBINE1)))
U3=S2DOK(COMBINE1)*PSIMDOK(COMBINE1)/(PSIMCOMB*S2COMB*(1.0d0-U2**2.0d0))

IF ((U1<0.0d0).OR.(U2<0.0d0).OR.(U3<0.0d0).OR.(U1>1.0d0).OR.(U2>1.0d0).OR.(U3>1.0d0)) THEN
DO J=1, N
KSIMNEW(J,ITER)=KSIM(J,ITER)
END DO
KMAX(ITER)=KMAX(ITER-1)
GOTO 666
END IF

DO J=1, N
IF (KSIM(J,ITER)<AST) THEN
KSIMNEW(J,ITER)=KSIM(J,ITER)
ELSE
IF (KSIM(J,ITER)>AST+1) THEN
KSIMNEW(J,ITER)=KSIM(J,ITER)-1
ELSE
IF ((KSIM(J,ITER)==COMBINE1).OR.(KSIM(J,ITER)==COMBINE2)) THEN
KSIMNEW(J,ITER)=AST
END IF
END IF
END IF
END DO

DO I=1, AST-1
MUPROPOSED(I,ITER)=MUDOK(I)
S2PROPOSED(I,ITER)=S2DOK(I)
PSIMPROPOSED(I,ITER)=PSIMDOK(I)
END DO
MUPROPOSED(AST,ITER)=MUCOMB
S2PROPOSED(AST,ITER)=S2COMB
PSIMPROPOSED(AST,ITER)=PSIMCOMB
DO I=AST+2, KMAX(ITER-1)
MUPROPOSED(I-1,ITER)=MUDOK(I)
S2PROPOSED(I-1,ITER)=S2DOK(I)
PSIMPROPOSED(I-1,ITER)=PSIMDOK(I)
END DO

!*****Logarithm of Acceptance Probability (LAP) for the combine move*****

LOGLIKELIHOODRATIO1=1.0d0
DO J=1, N
LOGLIKELIHOODRATIO1=LOGLIKELIHOODRATIO1 * &
DSQRT(S2DOK(KSIM(J,ITER))/ &
S2PROPOSED(KSIMNEW(J,ITER),ITER)) * &
DEXP(((X(J)-MUDOK(KSIM(J,ITER))**2.0d0)/(2.0d0*S2DOK(KSIM(J,ITER)))- &
((X(J)-MUPROPOSED(KSIMNEW(J,ITER),ITER))**2.0d0)/ &
(2.0d0*S2PROPOSED(KSIMNEW(J,ITER),ITER))))
END DO

```

```
LAP1=LOGLIKELIHOODRATIO*(1.0d0/KMAX(ITER-1))*( PSIMCOMB**((DELTA-1.0d0+SKCOMB)* &
DBETA(DELTA,(KMAX(ITER-1)-1.0d0)*DELTA)/ &
((PSIMDOK(COMBINE1)**(DELTA-1.0d0+SK(COMBINE1)))** &
(PSIMDOK(COMBINE2)**(DELTA-1.0d0+SK(COMBINE2)))))*DSQRT(2.0d0*PI/KAPPA)* &
DEXP(-0.5d0*KAPPA*(-(MUDOK(COMBINE1)-KSI)**2.0d0- &
(MUDOK(COMBINE2)-KSI)**2.0d0+(MUCOMB-KSI)**2.0d0))* &
(DGAMMA(ALPHA)/BH(A(ITER))* &
((S2COMB/(S2DOK(COMBINE1)*S2DOK(COMBINE2)))**(-ALPHA-1.0d0))* &
DEXP(-BH(A(ITER))*(-(1.0d0/S2DOK(COMBINE1))- &
(1.0d0/S2DOK(COMBINE2))+1.0d0/S2COMB ))* &
(B(KMAX(ITER-1)-1)*PALLOC1/D(KMAX(ITER-1)))*PDFBETA(2.0d0, 2.0d0, U1)* &
PDFBETA(2.0d0, 2.0d0, U2)*PDFBETA(1.0d0, 1.0d0, U3)* &
(U2*(1-U2**2)*U3*(1-U3)*S2COMB/ &
( PSIMCOMB*ABS(MUDOK(COMBINE1)- &
MUDOK(COMBINE2))*S2DOK(COMBINE1)*S2DOK(COMBINE2) ))
```

```
IF ( DRNUNF()<LAP1 ) THEN
DO I=1, KMAXPROP(ITER)
MU(I,ITER)=MUPROPOSED(I,ITER)
S2(I,ITER)=S2PROPOSED(I,ITER)
PSIM(I,ITER)=PSIMPROPOSED(I,ITER)
END DO
KMAX(ITER)=KMAXPROP(ITER)
CPROP=CPROP+1.0d0
SCPROP=SCPROP+1.0d0
ELSE
DO I=1, KMAX(ITER-1)
MU(I,ITER)=MUDOK(I)
S2(I,ITER)=S2DOK(I)
PSIM(I,ITER)=PSIMDOK(I)
END DO
DO J=1, N
KSIMNEW(J,ITER)=KSIM(J,ITER)
END DO
KMAX(ITER)=KMAX(ITER-1)
END IF
```

!*****telos split combine*****

END IF

!*****
!*****MOVE TYPE BIRTH - DEATH*****
!*****

CALL SVIGN (N, KSIMNEW(:,ITER), KSIMNEWORD)

ALLOCATE (EMPTYES(KMAX(ITER)), W(KMAX(ITER)), PINAK(KMAX(ITER)+1), &
MIU(KMAX(ITER)+1), VAR(KMAX(ITER)+1), PITH(KMAX(ITER)+1))

```
EMPTY=0
DO I=1, KMAX(ITER)
CALL ISRCH (N, I, KSIMNEWORD, 1, IND)
IF ( IND<0 ) THEN
EMPTY=EMPTY+1
EMPTYES(I)=I
ELSE
EMPTYES(I)=K+1
END IF
PINAK(I)=MU(I,ITER)
END DO
```

PINAK(KMAX(ITER)+1)=0.0d0

IF (DRNUNF()<B(KMAX(ITER))) THEN

!*****BIRTH MOVE*****

```
BMLA=BMLA+1
PBIRTH=MY_BETA_JOHNK( 1.0d0,KMAX(ITER)+0.0d0)
MUBIRTH=MY_NORMAL(KSI,1/KAPPA)
S2BIRTH=1.0d0/MY_GAMMA_JOHNK(ALPHA,BH(A(ITER)))
PINAK(KMAX(ITER)+1)=MUBIRTH
DO I=1, KMAX(ITER)
```

```

      W(I)=PSIM(I,ITER)*(1.0d0-PBIRTH)
END DO

CALL DSVRGN (KMAX(ITER)+1, PINAK, PINAK)
CALL DSRCH (KMAX(ITER)+1, MUBIRTH, PINAK, 1, INDE)

DO I=1, INDE-1
  MIU(I)=MU(I,ITER)
  VAR(I)=S2(I,ITER)
  PITH(I)=W(I)
END DO
MIU(INDE)=MUBIRTH
VAR(INDE)=S2BIRTH
PITH(INDE)=PBIRTH
DO I=INDE, KMAX(ITER)
  MIU(I+1)=MU(I,ITER)
  VAR(I+1)=S2(I,ITER)
  PITH(I+1)=W(I)
END DO

```

!*****Logarithm of Acceptance Probability (LAP) for the BIRTH move*****

```

LAP1=((PBIRTH**(DELTA-1.0d0))*((1.0d0-PBIRTH)**(N+KMAX(ITER)*DELTA-1.0d0))/ &
DBETA(KMAX(ITER)*DELTA,DELTA)*(KMAX(ITER)+1.0d0)*( D(KMAX(ITER)+1)/ &
(B(KMAX(ITER))*(EMPTY+1.0d0)))*1.0d0/(PDFBETA(1.0d0, KMAX(ITER)+0.0d0, PBIRTH))

```

```

IF ( DRNUNF(<LAP1 ) THEN
  DO I=1, KMAX(ITER)+1
    MU(I,ITER)=MIU(I)
    S2(I,ITER)=VAR(I)
    PSIM(I,ITER)=PITH(I)
  END DO
  KMAX(ITER)=KMAX(ITER)+1
  BPROP=BPROP+1.0d0
  BDPROP=BDPROP+1.0d0
END IF

```

ELSE

!*****DEATH MOVE*****

```

DMLA=DMLA+1

IF ( EMPTY==0 ) THEN
  GOTO 687
ELSE
  DEATH=K+1
  DO WHILE ( DEATH>K )
    CALL DRNUN(1,U)
    DEATH=EMPTYIES(FLOOR(KMAX(ITER)*U)+1)
  END DO

  DO I=1, KMAX(ITER)
    W(I)=PSIM(I,ITER)/(1.0d0-PDEATH)
  END DO

  PDEATH=PSIM(DEATH,ITER)
  MUDEATH=MU(DEATH,ITER)
  S2DEATH=S2(DEATH,ITER)

  DO I=DEATH, KMAX(ITER)-1
    PINAK(I)=MU(I+1,ITER)
  END DO
  PINAK(KMAX(ITER))=0.0d0

```

!*****Logarithm of Acceptance Probability (LAP) for the DEATH move*****

```

LAP1=DBETA((KMAX(ITER)-1.0d0)*DELTA,DELTA)* &
PDFBETA(1.0d0, KMAX(ITER)-1.0d0, PDEATH)*B(KMAX(ITER)-1)* &
(EMPTY+0.0d0)/((1.0d0-PDEATH)**(N+0.0d0+(KMAX(ITER)-1.0d0)*DELTA-1.0d0))*&
(PDEATH**(DELTA-1.0d0))*(KMAX(ITER)+0.0d0)*D(KMAX(ITER))

```

```

IF ( DRNUNF(<LAP1 ) THEN

```

```

  ALLOCATE ( VARPINAK(KMAX(ITER)-1), PROPPINAK(KMAX(ITER)-1) )

```

```

DO I=1, DEATH-1
    VARPINAK(I)=S2(I,ITER)
    PROPPINAK(I)=W(I)
END DO

DO I=DEATH, KMAX(ITER)-1
    VARPINAK(I)=S2(I+1,ITER)
    PROPPINAK(I)=W(I+1)
END DO

DO I=1, KMAX(ITER)-1
    MU(I,ITER)=PINAK(I)
    S2(I,ITER)=VARPINAK(I)
    PSIM(I,ITER)=PROPPINAK(I)
END DO
KMAX(ITER)=KMAX(ITER)-1
DPROP=DPROP+1.0d0
BDPROP=BDPROP+1.0d0

DEALLOCATE ( VARPINAK, PROPPINAK )
END IF
END IF

!*****telos birth death*****

687     END IF

444     FORMAT ( 'MU(', I2, ',', I5,')=', F9.5, ' P(', I2, ',', I5,')=', F6.4, ' S2(', I2, ',', I5,')=', F10.4 )
DEALLOCATE ( SK, SKX, KDIST, SKU, MUDOK, S2DOK, PSIMDOK, EMPTIES, W, PINAK, MIU, VAR, PITH )

IF ( KMAX(ITER)==ODIM ) THEN
    RAM=RAM+1
    DO I=1, ODIM
        MEAN5(I,RAM)=MU(I,ITER)
        VAR5(I,RAM)=S2(I,ITER)
        WEIGHT5(I,RAM)=PSIM(I,ITER)
    END DO
END IF

IF ( MOD(ITER,M/10)==0.0d0 ) THEN
    U=(ITER+0.0d0)/(M+0.0d0)
    WRITE (*,41), 100.0d0*U
    FORMAT (F4.0, '% completed')
41     END IF

IF ( ITER>BURN ) THEN
    DO I=1, K
        IF ( KMAX(ITER)==I ) THEN
            DIMWEIGHTS(I)=DIMWEIGHTS(I)+1.0d0
        END IF
    END DO
END IF

!*****TELOS RJMCMC*****

END DO

DO I=1, K
    DIMWEIGHTS(I)=DIMWEIGHTS(I)/(M-BURN+0.0d0)
END DO

!*****Output*****

SPROP=(SPROP+0.0d0)/(SMLA+0.0d0)
CPROP=(CPROP+0.0d0)/(CMLA+0.0d0)
BPROP=(BPROP+0.0d0)/(BMLA+0.0d0)
DPROP=(DPROP+0.0d0)/(DMLA+0.0d0)
SCPROP=(SCPROP+0.0d0)/(M+0.0d0)
BDPROP=(BDPROP+0.0d0)/(M+0.0d0)

```

```

PRINT *, 'ACCEPTED:'
WRITE (*,905), SCPROP, BDPROP
905  FORMAT ( 'SPLIT & COMBINE=', F6.4, ' BIRTH & DEATH=', F6.4 )

WRITE (*,906), SPROP, CPROP, BPROP, DPROP
906  FORMAT ( 'SPLITS=', F6.4, ' COMBINE=', F6.4, ' BIRTH=', F6.4, ' DEATH=', F6.4 )

DO I=1, 11
    WRITE (*,654), I, DIMWEIGHTS(I)
END DO
sw=0.0d0
DO I= 12, K
    SW=SW+DIMWEIGHTS(I)
END DO
write (*,655), SW
655  FORMAT ( 'SW=', F8.6 )

654  FORMAT ( 'p(', I2, ')=', F6.4 )

CALL DISCRETE_HISTOGRAM(20,'DIMENSION HISTOGRAM',KMAX,M,BURN,M)

CALL REAL_SEQUENCE(30,'M1 ERGODIC MEANS ',MEAN5(1,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(40,'M2 ERGODIC MEANS ',MEAN5(2,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(50,'M3 ERGODIC MEANS ',MEAN5(3,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(60,'M4 ERGODIC MEANS ',MEAN5(4,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(70,'M5 ERGODIC MEANS ',MEAN5(5,:),M,BURN,RAM,2)

CALL REAL_SEQUENCE(130,'RAW MEAN M1',MEAN4(1,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(140,'RAW MEAN M2',MEAN4(2,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(150,'RAW MEAN M3',MEAN4(3,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(160,'RAW MEAN M4',MEAN4(4,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(170,'RAW MEAN M5',MEAN5(5,:),M,BURN,RAM,1)

CALL REAL_SEQUENCE(230,'VAR1 ERGODIC MEANS ',VAR5(1,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(240,'VAR2 ERGODIC MEANS ',VAR5(2,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(250,'VAR3 ERGODIC MEANS ',VAR5(3,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(260,'VAR4 ERGODIC MEANS ',VAR5(4,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(270,'VAR5 ERGODIC MEANS ',VAR5(5,:),M,BURN,RAM,2)

CALL REAL_SEQUENCE(430,'RAW MEAN VAR 1', VAR5 (1,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(440,'RAW MEAN VAR 2', VAR5 (2,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(450,'RAW MEAN VAR 3', VAR5 (3,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(460,'RAW MEAN VAR 4', VAR5 (4,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(470,'RAW MEAN VAR 5', VAR5 (5,:),M,BURN,RAM,1)

CALL REAL_SEQUENCE(330,'W1 ERGODIC MEANS ',WEIGHT5(1,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(340,'W2 ERGODIC MEANS ',WEIGHT5(2,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(350,'W3 ERGODIC MEANS ',WEIGHT5(3,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(360,'W4 ERGODIC MEANS ',WEIGHT5(4,:),M,BURN,RAM,2)
CALL REAL_SEQUENCE(370,'W5 ERGODIC MEANS ',WEIGHT5(5,:),M,BURN,RAM,2)

CALL REAL_SEQUENCE(530,'RAW MEAN W1 ',WEIGHT5(1,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(540,'RAW MEAN W2 ',WEIGHT5(2,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(550,'RAW MEAN W3 ',WEIGHT5(3,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(560,'RAW MEAN W4 ',WEIGHT5(4,:),M,BURN,RAM,1)
CALL REAL_SEQUENCE(570,'RAW MEAN W5 ',WEIGHT5(5,:),M,BURN,RAM,1)

45  END

```


4.4 Κώδικας αλγορίθμου BDMCMC στο Mathematica

```

birthdeathmcmc[
  Clear msim, s2sim, matrix, ktable
  Do
  "with death process";
  8
  Clear deathrate
  t = 0;
  dim = 1;
  While[
    If[
      d[1] == deathrate = 0
      loglikelihooddeath q[1] m Log Sum
      psi[i, iter-1]
      - p[im, q[iter-1]] * s2[i, iter-1]
      Exp[
        - 2 * s2[i, iter-1]
        loglikelihood
        S
        m
        d q
        deathrate = Sum
        t = t +
        If
        Random
        birthrate
        deathrate
        psi[m, dim+1, iter-1]
        RandomBetaDistribution
        m dim
        iter-1
        RandomNormalDistribution
        s2 dim +
        iter-1
        RandomGammaDistribution
        Do psi[m, dim+1, iter-1]
        p[im, dim+1, iter-1]
        s2, dim
        dim = dim + 1
        deathlist =
        deathrate
        Do deathlist =
        deathrate
    ]
  
```


Simulation of χ^2 Distribution
 b iter, kmax iter

GammaDistribution, kmax iter
 + Sum, s2, j iter, kmax iter

single, j iter, kmax iter

Do print, kmax iter

Spsting Samples, kmax iter

If print, kmax iter

Simulation of m 1, kmax iter

Do

m iter, Distribution, k + $\frac{s2}{j iter}$, $\frac{s2}{j iter}$, E

Do

Do

if ksim i iter, 1, 0

, n, kmax iter

Simulation of s^2 1, 2, kmax iter

Do, 2, j iter, GammaDistribution, $\frac{j}{2}$, $\frac{b}{2}$, b iter, sku j

Print Number, M, 1, M

"Or izw touV (fa) ..."

msim = Table m, j

pmatrix = Table, k

s2sim = Table s, kmax

ktable = Table ksim i, M, n

;

Σε αυτή την εργασία παρουσιάσαμε αρχικά τον ορισμό και τις δυσκολίες που αντιμετωπίζουν κλασσικές μέθοδοι στην συμπερασματολογία (όπως η εύρεση αναλυτικών εκφράσεων των εκτιμητών Bayes και ο αλγόριθμος EM), καθώς και την αντιμετώπιση του προβλήματος label switching. Στην συνέχεια διαπραγματευτήκαμε μεθόδους MCMC για τις περιπτώσεις γνωστού και άγνωστου αριθμού συνιστωσών και συμπεράναμε την ευαισθησία της εκ των υστέρων κατανομής αυτού, ως προς τις εκ των προτέρων υποθέσεις.

Για την κατασκευή ενός δειγματολήπτη MCMC όταν η διάσταση του μοντέλου μείξης θεωρείται σταθερή, η εμπειρία μας προτείνει την ακόλουθη στρατηγική. Ξεκινάμε με την εφαρμογή του αλγορίθμου Metropolis - Hastings χωρίς την διεύρυνση των δεδομένων όπου είναι και η αποδοτικότερη (όταν μπορούμε να υπολογίσουμε την πιθανοφάνεια). Μετά την εκτέλεση του αλγορίθμου με διαφορετικές τιμές των παραμέτρων των κατανομών πρότασης (π.χ. την τιμή της κλίμακας του τυχαίου περίπατου), βρίσκουμε εκείνες τις τιμές που βελτιστοποιούν την απόδοση του δειγματολήπτη (όπως μεγάλη πιθανότητα αποδοχής των προτεινόμενων κινήσεων). Εάν ο δειγματολήπτης βρίσκεται σε μία περιοχή της πιθανοφάνειας τότε συμπεραίνουμε ότι δεν έχει επισκεφτεί τις συμμετρικές κορυφές αυτής και είτε θα πρέπει να συνεχίσουμε τις επαναλήψεις του αλγορίθμου, είτε να «πειράζουμε» τις παραμέτρους των κατανομών πρότασης. Όπως είδαμε, ο δειγματολήπτης Gibbs θα πρέπει να αποτελεί την δεύτερη σκέψη μας, καθώς πολλές φορές αποτυγχάνει να επισκεφτεί την κύρια κορυφή της πιθανοφάνειας μια και οι κινήσεις του είναι πολύ περιορισμένες και μια σημαντική αλλαγή στο διάνυσμα κατάταξης των παρατηρήσεων απαιτεί πολλές επαναλήψεις.

Εάν δεν είμαστε σίγουροι για τον αριθμό των συνιστωσών που αποτελούν την μείξη τότε καταφεύγουμε στις μεθόδους reversible jump ή birth - death MCMC. Όταν μπορούμε να υπολογίσουμε την πιθανοφάνεια θα λέγαμε ότι η μέθοδος BDMCMC είναι πιο απλή στον προγραμματισμό. Από την άλλη, η προσέγγιση reversible jump δίνει άμεσα αποτελέσματα αφού δεν χρειάζεται να αναδιατάξουμε τις παρατηρήσεις. Πάντως και στις δύο μεθόδους απαιτείται η προσεκτική μελέτη των αποτελεσμάτων και η ανάλυση θα πρέπει να γίνει βάσει της εκ των προτέρων γνώσης για το υπό μελέτη φαινόμενο.

Τέλος για την αντιμετώπιση του φαινομένου label switching πρέπει να έχουμε λάβει υπόψη τα ακόλουθα κριτήρια. Στις περιπτώσεις όπου επιθυμούμε να κάνουμε συμπερασματολογία μόνο για τις παραμέτρους του μοντέλου, αρκεί η υιοθέτηση εικονικών περιορισμών, ή ακόμα καλύτερα ο αλγόριθμος αναδιάταξης μέσω του εσωτερικού γινομένου. Όταν ο σκοπός της ανάλυσης συμπεριλαμβάνει και την κατάταξη των παρατηρήσεων στις συνιστώσες (clustering), τότε θα πρέπει να χρησιμοποιήσουμε μία συνάρτηση ζημίας. Κάτι τέτοιο θα μπορούσε να γίνει και με την μέθοδο του εσωτερικού γινομένου, αλλά από την Μπεϋζιανή σκοπιά αυτό δεν είναι αρκετό διότι δεν έχουμε κάποια μέτρηση της αβεβαιότητας των αποτελεσμάτων που παίρνουμε.

Μετά την ιστορική παρουσίαση της εξέλιξης των Μπεϋζιανών μεθόδων για τα μοντέλα μείξεων, παραμένουν ανοιχτές κάποιες προκλήσεις που το μέλλον θα δείξει αν αυτές θα απαντηθούν. Μία περιοχή σύγχρονης έρευνας στην βιοπληροφορική είναι η κατάταξη γονιδίων, η οποία βασίζεται στην χρήση μοντέλων μείξεων. Ένα εμπόδιο στην χρήση μιας Μπεϋζιανής προσέγγισης είναι η δυσκολία προσομοίωσης από μία πολυδιάστατη (μη σταθερή) μέτρηση - στόχο, η οποία είναι χαρακτηριστική τέτοιων προβλημάτων (ένα παράδειγμα ενός μικρού συνόλου δεδομένων είναι 2000

παρατηρήσεις αποτελούμενων από 6 διαστάσεις). Οι υπάρχοντες δειγματολήπτες (reversible jump ή συνεχούς χρόνου) είναι ανίκανοι να κινηθούν αποδοτικά στον παραμετρικό χώρο και για την εφαρμογή μιας Μπεϋζιανής μεθοδολογίας σε τέτοιου είδους προβλήματα απαιτούνται νέες μέθοδοι προσομοίωσης (Jasra, Stephens & Holmes, 2005).

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Βιβλιογραφία

- Aitkin, M. (2001). *Likelihood and Bayesian analysis of mixtures*.
Statistical Modelling, Vol. 1, No. 4: 287-304.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*.
Springer - Verlag, 2nd edition.
- Cappé, O., Guillin, A., Marin, J., and Robert, C. (2003). *Population Monte Carlo*.
Journal of Computational and Graphical Statistics, Vol. 13, No.4: 907-929.
- Cappé, O., Guillin, A., Marin, J., and Robert, C. (2002). *Reversible jump MCMC converging to birth-and-death MCMC and more general continuous time samplers*.
Journal of the Royal Statistical Society B, Vol. 65, No.3: 679-700.
- Casella, G., Mengersen, K., Robert, C., and Titterton, D. (2002). *Perfect slice samplers for mixtures of distributions*.
Journal of the Royal Statistical Society B, Vol. 64, No.4: 777-790.
- 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 B, Vol. 39: 1-38.
- Diebolt, J., and Robert, C. (1994). *Estimation of finite mixture distributions through Bayesian sampling*.
Journal of the Royal Statistical Society B, Vol. 56, No.2: 363-375.
- Escobar, M., and West, M. (1995). *Bayesian density estimation and inference using mixtures*.
Journal of the American Statistical Association, Vol. 90, No.430: 577-588.
- Ferguson, T. (1973). *A Bayesian analysis of some nonparametric problems*.
The Annals Of Statistics 1973, Vol.1, No.2: 209-230.
- Gelfand, A., and Smith, A. (1990). *Sampling based approaches to calculating marginal densities*.
Journal of the American Statistical Association, Vol. 85: 398-409.
- Green, P. (1995). *Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination*.
Biometrika, Vol. 82, No.4: 711-712.
- Jasra, A., Holmes, C.C., and Stephens, D.A. (2005). *Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling*.
Statistical Science, Vol. 20, No.1: 50-67.
- Kass, R., and Raftery, A. (1995). *Bayes factors*.
Journal of the American Statistical Association, Vol. 90: 773-795.

Marin, J.M., Mengersen, K., and Robert, C (2004). *Bayesian modelling and inference on mixtures of distributions*.

Handbook of Statistics, Vol. 25, Elsevier Sciences.

Richardson, S., and Green, P. (1997). *On Bayesian analysis with an unknown number of components*.

Journal of the Royal Statistical Society B, Vol. 59, No.4: 731-792.

Ripley, B. (1987). *Stochastic Simulation*.

John Wiley, New York.

Robert, C., and Casella, G., (2004). *Monte Carlo Statistical Methods*.

Springer - Verlag, New York, NY, second edition.

Robert, C., and Mengersen, K. (1999). *Reparameterization issues in mixture estimation and their bearings on the Gibbs sampler*.

Computational Statistics and Data Analysis, Vol. 29: 325-343.

Roeder, K. (1992). *Density estimation with confidence sets exemplified by super-clusters and voids in galaxies*.

Journal of the American Statistical Association, Vol. 85: 617-624.

Roeder, K., and Wasserman, L. (1997). *Practical Bayesian density estimation using mixtures of normal distributions*.

Journal of the American Statistical Association, Vol. 92: 894-902.

Stephens, M. (2000a). *On Bayesian analysis with an unknown number of component - an alternative to reversible jump methods*.

The Annals of Statistics, Vol. 28: 40-74.

Stephens, M. (2000b). *Dealing with label switching in mixture models*.

Journal of the Royal Statistical Society B, Vol. 62, No.4: 795-809.

Tanner, M., and Wong, W. (1987). *The calculation of posterior distributions by data augmentation*.

Journal of the American Statistical Association, Vol. 82: 528-550.

Waagepetersen, R., and Sorensen, D. (2001). *A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping*.

International Statistical Review, Vol.69: 49-62.

Ευρετήριο αλγορίθμων

Αλγόριθμος Birth Death MCMC.....	85
Αλγόριθμος Metropolis - Hastings για κανονικές μείξεις.....	52
Αλγόριθμος Metropolis - Hastings για μείξη Poisson.....	49
Αλγόριθμος Metropolis - Hastings για το μοντέλο (1.8).....	47
Αλγόριθμος Population Monte Carlo για το μοντέλο (1.8).....	56
Αλγόριθμος Reversible jump MCMC για κανονικές μείξεις.....	73
Αλγόριθμος αναδιάταξης με συνάρτηση ζημίας την απόκλιση Kullback - Leibler.....	21
Αλγόριθμος αναδιάταξης μέσω του εσωτερικού γινομένου.....	18
Αλγόριθμος προσομοίωσης διαδικασίας γεννήσεων - θανάτων με σταθερές υπερπαραμέτρους.....	84
Γενικός αλγόριθμος EM.....	14
Γενικός αλγόριθμος Metropolis - Hastings για μοντέλα μείξεων.....	46
Γενικός αλγόριθμος Population Monte Carlo.....	55
Γενικός αλγόριθμος αναδιάταξης μέσω θεωρίας αποφάσεων.....	20
Δειγματολήπτης Gibbs για κανονικές μείξεις.....	43
Δειγματολήπτης Gibbs για μείξεις εκθετικών οικογενειών.....	32
Δειγματολήπτης Gibbs για μείξη Poisson.....	37
Δειγματολήπτης Gibbs για το μοντέλο (1.8).....	34