



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΝΑΛΟΓΙΣΤΙΚΩΝ-
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ ΜΑΘΗΜΑΤΙΚΩΝ**

**«Μη παραμετρική Μπεϋζιανή συμπερασματολογία σε στοχαστικά δυναμικά
συστήματα»**

Διπλωματική Εργασία για το Μεταπτυχιακό Πρόγραμμα Σπουδών

Η παρούσα Εργασία εκπονήθηκε
ως μερική ικανοποίηση των απαιτήσεων για την απόκτηση
του αντιστοίχου τίτλου σπουδών στην
Στατιστική και Αναλογιστικά-Χρηματοοικονομικά Μαθηματικά

Μόραλη Δήμητρα

Ημερομηνία
20/2/2024

ΣΑΜΟΣ

Μόραλη Δήμητρα

«Μη παραμετρική Μπεϋζιανή συμπερασματολογία σε στοχαστικά δυναμικά συστήματα»

Ημερομηνία
20/2/2024

Διπλωματική Εργασία για το Μεταπτυχιακό Πρόγραμμα Σπουδών

**Τμήμα Στατιστικής και Αναλογιστικών-Χρηματοοικονομικών
Μαθηματικών**

Συγγραφέας: Μόραλη Δήμητρα

.....

Επιβλέπων: Ξανθόπουλος Στυλιανός

.....

Μέλος Επιτροπής: Καλούδης Κωνσταντίνος

.....

Μέλος Επιτροπής: Μέρκατας Χρίστος

.....

ΣΑΜΟΣ

Ευχαριστίες

Με την ολοκλήρωση της εργασίας θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ξανθόπουλο Στυλιανό για την ευκαιρία που μου προσέφερε να υλοποιήσω την παρούσα εργασία στα πλαίσια του Μεταπτυχιακού Προγράμματος σπουδών στη Στατιστική και Αναλογιστικά Χρηματοοικονομικά Μαθηματικά. Επίσης, θα ήθελα να ευχαριστήσω τους καθηγητές μου και μέλη της τριμελούς επιτροπής κ. Καλούδη Κωνσταντίνο και κ. Μέρκατα Χρίστο για όλη την καθοδήγηση και στήριξή τους καθ'όλη της διάρκεια εκπόνησης της διπλωματικής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω θερμά την οικογένεια μου και τους φίλους μου για την αμέριστη συμπαράσταση και στήριξή τους κατά τη διάρκεια των μεταπτυχιακών σπουδών μου.

Abstract

In this thesis, a computational method of nonparametric Bayesian statistics is presented that aims both at estimating the various parameters and the density of the noise that perturbs stochastic dynamical systems. This approach is based on the Dirichlet process mixture model.

Chapter 1 provides some introductory information regarding Bayesian statistics, focusing on the key differences between Classical and Bayesian statistics. Furthermore, the chapter defines the statistical model, which serves as the foundation for statistical inference, before moving on to the concept of uncertainty and how it is expressed in terms of probabilities in Bayesian statistics. Another notion that plays an important role in the application of Bayesian statistics is exchangeability. That is why the definition, de Finetti's well-known representation theorem, and an example are presented for a deeper understanding.

Chapter 2 briefly examines the two types of computational methods used in Bayesian inference, namely the Variational Inference (VI) and Markov Chain Monte Carlo (MCMC) approaches. Due to the fact that this thesis focuses on the second class of methods, the Metropolis-Hastings, Gibbs, and Slice sampling algorithms are discussed in depth. Furthermore, an introduction to Bayesian finite mixture models is provided, as well as an example of parameter estimation for a Gaussian finite mixture model using the Gibbs sampling algorithm. The Label Switching Problem, which is covered at the end of the chapter, is one of the issues that arise while using this technique.

Chapter 3 provides the definition of the Dirichlet process and also some constructive definitions proposed for its representation, such as the Generalized Pólya urn, the Chinese Restaurant Process (CRP) and the Stick-Breaking representation. Then, both marginal and conditional computational inference methods are discussed, followed by an introduction to infinite Bayesian mixture models. This thesis focuses on the infinite Dirichlet process mixture models which are used for parameter estimation and density estimation. Walker (2007) proposed an algorithm for sampling from these models, which is presented in this chapter and is based on a subset of conditional methods. The algorithm is then used and the results are presented at the end of the chapter.

Chapter 4 deals with the application of the algorithm introduced in Chapter 3 to stochastic dynamical systems. For the definition of stochastic dynamical systems, an introduction to deterministic dynamical systems is provided, as well as the categories of random noise. Furthermore, the chaotic behavior observed in dynamical systems is discussed, followed by the definition of a stochastic dynamical system. The Dirichlet process reconstruction model is then presented, which is used to estimate the parameters of a stochastic dynamical system along with the density of the random noise perturbing it. The chapter completes with a presentation of the findings acquired

by applying the algorithm.

Chapter 5 provides a brief overview of the thesis, presents the conclusions that obtained and suggests a possible future research topic based on this thesis.

Finally, the Bibliography and an Appendix with the codes used in the paper are mentioned.

Keywords: MCMC Algorithms, Stick-Breaking Representation, Bayesian Non Parametric Mixture Models, Dynamical Systems, Dirichlet Process Reconstruction Models.

Περίληψη

Στην παρούσα εργασία παρουσιάζεται μία υπολογιστική μέθοδος της μη παραμετρικής Μπεϋζιανής στατιστικής που στοχεύει τόσο στην εκτίμηση των διάφορων παραμέτρων, όσο και στην εκτίμηση πυκνότητας του θορύβου ενός στοχαστικού δυναμικού συστήματος. Η μέθοδος αυτή βασίζεται στα μοντέλα μίξης της διαδικασίας Dirichlet.

Στο Κεφάλαιο 1, δίνονται κάποιες εισαγωγικές πληροφορίες για τη Μπεϋζιανή στατιστική, όπου και αναφέρονται οι βασικές διαφορές μεταξύ Κλασικής και Μπεϋζιανής στατιστικής. Στη συνέχεια, αναφέρεται ο ορισμός του στατιστικού μοντέλου, που αποτελεί θεμέλιο της στατιστικής συμπερασματολογίας, ενώ παράλληλα αναφέρεται και η έννοια της αβεβαιότητας και πως αυτή εκφράζεται με όρους πιθανοτήτων στη Μπεϋζιανή στατιστική. Μία ακόμα έννοια που έχει πρωταρχικό ρόλο στη χρήση της Μπεϋζιανής στατιστικής είναι αυτή της ανταλλαξιμότητας. Για το λόγο αυτό παρουσιάζονται ο ορισμός της, το διάσημο θεώρημα αναπαράστασης του de Finetti, καθώς και ένα παράδειγμα για την καλύτερη κατανόηση της.

Στο Κεφάλαιο 2, γίνεται αναφορά στις δύο κατηγορίες υπολογιστικών μεθόδων που χρησιμοποιούνται στην Μπεϋζιανή συμπερασματολογία, δηλαδή στις μεθόδους που βασίζονται στη μεταβολική συμπερασματολογία (Variational Inference) και στις μεθόδους Markov Chain Monte Carlo (MCMC). Λόγω του ότι η εργασία βασίζεται στην δεύτερη κατηγορία μεθόδων, δίνονται αναλυτικά οι αλγόριθμοι Metropolis-Hastings, Gibbs και Slice. Επιπλέον, γίνεται μια εισαγωγή στα πεπερασμένα Μπεϋζιανά μοντέλα μίξης, όπου με τη βοήθεια του αλγορίθμου Gibbs, πραγματοποιείται η εκτίμηση των παραμέτρων ενός πεπερασμένου μοντέλου μίξης κανονικών κατανομών. Ένα πρόβλημα που δημιουργείται κατά την εφαρμογή του αλγορίθμου είναι το Πρόβλημα Εναλλαγής Ετικέτας (Label Switching Problem), στο οποίο γίνεται αναφορά στο τέλος του κεφαλαίου.

Στο Κεφάλαιο 3, δίνεται ο ορισμός της διαδικασίας Dirichlet, καθώς και κάποιοι κατασκευαστικοί ορισμοί που προτάθηκαν για την αναπαράστασή της, όπως το Γενικευμένο Pólya urn, η διαδικασία Κινέζικου Εστιατορίου (Chinese Restaurant Process) και η αναπαράσταση Stick-Breaking. Έπειτα, γίνεται αναφορά τόσο στις περιθώριες, όσο και στις υπό συνθήκη υπολογιστικές μεθόδους συμπερασματολογίας, ενώ παράλληλα ακολουθεί μία εισαγωγή στα μη πεπερασμένα Μπεϋζιανά μοντέλα μίξης. Η παρούσα εργασία εστιάζει στα μη πεπερασμένα μοντέλα μίξης της διαδικασίας Dirichlet τα οποία χρησιμοποιούνται για την εκτίμηση παραμέτρων και την εκτίμηση πυκνότητας. Ο Walker (2007) πρότεινε έναν αλγόριθμό για τη δειγματοληψία από αυτά τα μοντέλα, ο οποίος παρουσιάζεται στο κεφάλαιο αυτό και βασίζεται σε μία υποκατηγορία των υπό συνθήκη μεθόδων. Στη συνέχεια, γίνεται χρήση του αλγορίθμου και τα αποτελέσματα παρουσιάζονται στο τέλος του κεφαλαίου.

Το Κεφάλαιο 4 αφορά την εφαρμογή του αλγορίθμου, που παρουσιάστηκε στο Κεφάλαιο 3, σε στοχαστικά δυναμικά συστήματα. Συγκεκριμένα, για τον ορισμό των στοχαστικών δυναμικών συστημάτων γίνεται μία εισαγωγή στα ντετερμινιστικά δυναμικά συστήματα

και παρουσιάζονται οι κατηγορίες τυχαίων θορύβων. Επιπλέον, αναφέρεται η χαοτική συμπεριφορά που παρατηρείται στα δυναμικά συστήματα και ακολουθεί ο ορισμός ενός στοχαστικού δυναμικού συστήματος. Στη συνέχεια, δίνεται ο αλγόριθμος ανακατασκευής της διαδικασίας Dirichlet, όπου χρησιμοποιείται για την εκτίμηση των παραμέτρων ενός στοχαστικού δυναμικού συστήματος, καθώς και για την εκτίμηση πυκνότητας του τυχαίου θορύβου που το διαταράσσει. Το κεφάλαιο κλείνει με την παρουσίαση των αποτελεσμάτων που προέκυψαν κατά την εφαρμογή του αλγορίθμου.

Στο Κεφάλαιο 5, γίνεται μία σύντομη επισκόπηση της εργασίας, παρουσιάζονται τα συμπεράσματα που προέκυψαν καθ' όλη τη διάρκειά της και προτείνεται μία πιθανή μελλοντική έρευνα με αφορμή το θέμα της παρούσας εργασίας.

Τέλος, παρατίθενται η Βιβλιογραφία και ένα Παράρτημα με τους κώδικες που χρησιμοποιήθηκαν στην εργασία.

Λέξεις-Κλειδιά: Αλγόριθμοι MCMC, Αναπαράσταση Stick-Breaking, Μη παραμετρικά Μπεϋζιανά μοντέλα μίξης, Δυναμικά Συστήματα, Μοντέλο ανακατασκευής διαδικασίας Dirichlet.

Περιεχόμενα

Abstract	ii
Περίληψη	iv
Κατάλογος Σχημάτων	viii
1 Εισαγωγή	1
1.1 Μπεϋζιανή και Κλασική Στατιστική	2
1.2 Στατιστικό Μοντέλο και Αβεβαιότητα	3
1.3 Ανταλλαξιμότητα	4
2 Μπεϋζιανή Εκτίμηση	8
2.1 Κατηγορίες υπολογιστικών μεθόδων Μπεϋζιανής εκτίμησης	8
2.1.1 Μέθοδοι Μεταβολών (Variational Inference)	8
2.1.2 Markov Chain Monte Carlo	9
2.2 Κατηγορίες MCMC μεθόδων	10
2.2.1 Ο αλγόριθμος Metropolis-Hastings	10
2.2.2 Ο αλγόριθμος Gibbs	12
2.2.3 Ο αλγόριθμος Slice	13
2.3 Πεπερασμένα Μπεϋζιανά μοντέλα μίξης	15
2.4 Πρόβλημα Εναλλαγής Ετικέτας	22
3 Μη παραμετρική Μπεϋζιανή Στατιστική	24
3.1 Διαδικασία Dirichlet	24
3.2 Αναπαραστάσεις της Διαδικασίας Dirichlet	27
3.3 Μη Παραμετρικά Μπεϋζιανά μοντέλα μίξης	30
3.4 Δειγματοληψία του μοντέλου μίξης διαδικασίας Dirichlet με slices	32
3.5 Προσομοίωση	35
4 Εφαρμογή σε στοχαστικά δυναμικά συστήματα	37
4.1 Δυναμικά Συστήματα	37
4.1.1 Ντετερμινιστικά Δυναμικά Συστήματα	37
4.1.2 Κατηγορίες Τυχαίου Θορύβου	38

4.1.3	Η έννοια του χάους στα δυναμικά συστήματα	40
4.1.4	Στοχαστικά Δυναμικά Συστήματα	43
4.2	Μοντέλο Ανακατασκευής Διαδικασίας Dirichlet	47
4.3	Προσομοίωση	51
5	Συμπεράσματα	57
	Βιβλιογραφία	58
	Παράρτημα	64

Κατάλογος Σχημάτων

2.1	Ιστόγραμμα περικομμένης κανονικής κατανομής στο διάστημα (0,4) . . .	15
2.2	Trace plot, ACF plot, διάγραμμα κινούμενου μέσου και Ιστόγραμμα για το w_2	21
2.3	Διαγράμματα ίχνους για τις μέσες τιμές της τριπλής μίξης (2.27)	23
3.1	Οι εμπειρικές συναρτήσεις κατανομών 20 τυχαίων μέτρων DP, με παράμετρο συγκέντρωσης $c = 5, 50, 500$ και μέτρο βάσης $G_0 = N(0, 1)$	26
3.2	Ιστόγραμμα του δείγματος για την εκτίμηση της πυκνότητας της μίξης (3.25) μαζί με την καμπύλη της θεωρητικής πυκνότητας (αριστερά) και διάγραμμα κινούμενου μέσου για τον αριθμό των κλάσεων (δεξιά).	36
4.1	Τροχιές με αρχικές συνθήκες $x_0 = 0.5$ (μπλε) και $x'_0 = 0.5001$ (κόκκινο) (αριστερά). Διάγραμμα της απόλυτης τιμής της διαφοράς των δύο τροχιών (δεξιά).	41
4.2	Διάγραμμα διακλάδωσης της λογιστικής απεικόνισης με αρχική συνθήκη $x_0 = 0.5$. Με κόκκινο χρώμα δίνεται η τιμή της παραμέτρου ελέγχου $\vartheta = 1.71$ που χρησιμοποιήθηκε στα προηγούμενα παραδείγματα.	43
4.3	Ιστογράμματα των τροχιών της λογιστικής απεικόνισης (4.8) για $x_0 = 0.5$ (αριστερά) και $x_0 = 0.5001$ (δεξιά).	44
4.4	Ιστογράμματα των τροχιών ενός ντετερμινιστικού (αριστερά) και ενός στοχαστικού (δεξιά) δυναμικού συστήματος με παράμετρο ελέγχου $\vartheta = 1.71$ και $\sigma = 0.01$	47
4.5	Τροχιά του δυναμικού συστήματος (4.29) για $\vartheta = 1.71$, $x_0 = 0.5$, και θόρυβο από τη μίξη (4.30).	52
4.6	Ιστογράμματα για τα posterior δείγματα των ϑ (αριστερά) και x_0 (δεξιά). . .	54
4.7	Διάγραμμα κινούμενου μέσου για το ϑ (αριστερά), διάγραμμα ίχνους για το x_0 (κέντρο) και διάγραμμα κινούμενου μέσου για τον αριθμό των στοιχείων μίξης (δεξιά).	55
4.8	Ιστόγραμμα του δείγματος από την posterior predictive του θορύβου (αριστερά) και λογαριθμικό διάγραμμα πυκνοτήτων (δεξιά).	56

Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια, οι μη παραμετρικές Μπεϋζιανές μέθοδοι αποτελούν ισχυρά εργαλεία για τη μοντελοποίηση σύνθετων και πολυδιάστατων δεδομένων. Σε αντίθεση με τις παραδοσιακές παραμετρικές προσεγγίσεις που υποθέτουν σταθερές δομές μοντέλων και πεπερασμένους χώρους παραμέτρων, οι μη παραμετρικές Μπεϋζιανές μέθοδοι έχουν την ικανότητα να προσαρμόζουν αυτόματα την πολυπλοκότητα του μοντέλου στην πολυπλοκότητα των δεδομένων, καθιστώντας τις ιδιαίτερα κατάλληλες για την ανάλυση μεγάλων και ετερογενών συνόλων δεδομένων.

Μεταξύ αυτών των μεθόδων, τα μοντέλα μίξης διαδικασίας Dirichlet (DPM) είναι αρκετά διαδεδομένα λόγω της ικανότητάς τους να μοντελοποιούν άγνωστους και δυνητικά άπειρους αριθμούς παραμέτρων. Τα μοντέλα DPM παρέχουν ένα ευέλικτο πλαίσιο για την εκτίμηση της πυκνότητας και των παραμέτρων, επιτρέποντας την ανακάλυψη κρυφών δομών στα δεδομένα χωρίς την απαραίτητη γνώση του αριθμού των στοιχείων της μίξης ή την κατανομή των δεδομένων.

Σκοπός της παρούσας εργασίας είναι να προσφέρει το θεωρητικό πλαίσιο των υπολογιστικών μεθόδων των μη παραμετρικών Μπεϋζιανών μοντέλων μίξης, με ιδιαίτερη έμφαση στα μοντέλα μίξης διαδικασίας Dirichlet. Για την επίτευξη του παραπάνω στόχου και για την καλύτερη κατανόηση του αναγνώστη πραγματοποιείται η εφαρμογή των υπολογιστικών μεθόδων μοντελοποίησης στα στοχαστικά δυναμικά συστήματα με στόχο την εκτίμηση των παραμέτρων του μοντέλου, καθώς και της πυκνότητας του τυχαίου θορύβου.

Στο κεφάλαιο που ακολουθεί, δίνονται κάποιες εισαγωγικές πληροφορίες της Μπεϋζιανής στατιστικής που είναι απαραίτητες για την κατανόηση του θεωρητικού πλαισίου που έπεται στη συνέχεια.

1.1 Μπεϋζιανή και Κλασική Στατιστική

Η στατιστική συμπερασματολογία ορίζεται ως η διαδικασία εκμάθησης κάποιων γενικών χαρακτηριστικών ενός πληθυσμού που εξάγεται από ένα υποσύνολό του. Οι αριθμητικές τιμές των χαρακτηριστικών του πληθυσμού εκφράζονται με κάποια παράμετρο θ , ενώ οι αριθμητικές περιγραφές αυτού του υποσυνόλου παράγουν το σύνολο δεδομένων (Hoff, 2009). Η στατιστική συμπερασματολογία μπορεί να γίνει είτε χρησιμοποιώντας την Κλασική προσέγγιση είτε την Μπεϋζιανή. Οι δύο προσεγγίσεις διαφέρουν ως προς τις θεμελιώδεις αρχές και μεθοδολογίες τους (Bolstad, 2011).

Όσον αφορά την Κλασική στατιστική, οι παράμετροι ενός μοντέλου θεωρούνται άγνωστες σταθερές και όχι τυχαίες, ενώ η έννοια της πιθανότητας ορίζεται ως μια μακροχρόνια σχετική συχνότητα ενός «επαναλαμβανόμενου» γεγονότος. Η εξαγωγή συμπερασμάτων στην Κλασική στατιστική περιλαμβάνει έναν συνδυασμό τεχνικών, εννοιών και μεθόδων δύο διαφορετικών θεωριών. Η πρώτη θεωρία είναι του Fisher που βασίζεται στη συνάρτηση πιθανοφάνειας εισάγοντας έννοιες όπως οι εκτιμητές μέγιστης πιθανοφάνειας (MLE) οι οποίοι χρησιμοποιούνται στην ανάλυση διακύμανσης και στους ελέγχους σημαντικότητας. Ενώ, η δεύτερη θεωρία αναπτύχθηκε από τους Neyman και Pearson στην οποία εισήχθη η έννοια της θεωρίας αποφάσεων δημιουργώντας διάφορα κριτήρια για την εύρεση βέλτιστων λύσεων (Bolstad, 2011). Αξίζει να αναφερθεί ότι ένα βασικό εργαλείο της Κλασικής στατιστικής είναι τα 95% διαστήματα εμπιστοσύνης, τα οποία δεδομένου ότι η παράμετρος θ είναι μια σταθερά, ερμηνεύονται ως εξής: κατά τη διεξαγωγή επαναλαμβανόμενων δοκιμών εύρεσης διαστημάτων εμπιστοσύνης, η πραγματική τιμή της παραμέτρου θ βρίσκεται στο 95% των διαστημάτων αυτών (Casella and Berger, 2017). Σημειώνεται επίσης, ότι το ποσοστό 95% είναι ενδεικτικό και ανάλογο με το πρόβλημα ο ερευνητής επιλέγει το κατάλληλο ποσοστό για την εξαγωγή συμπερασμάτων.

Όσον αφορά τη Μπεϋζιανή στατιστική, οι παράμετροι θεωρούνται τυχαίες μεταβλητές, σε αντίθεση με την Κλασική στατιστική στην οποία όπως αναφέρθηκε οι παράμετροι είναι άγνωστες σταθερές. Στην Μπεϋζιανή στατιστική για τις τυχαίες μεταβλητές $\theta \in \Theta$, όπου Θ το σύνολο των τιμών των παραμέτρων, καθορίζεται μια εκ των προτέρων (prior) κατανομή $p(\theta)$ η οποία ποσοτικοποιεί την πληροφορία για το θ πριν την παρατήρηση των δεδομένων. Έπειτα, ακολουθεί η περιγραφή της συσχέτισης μεταξύ των δεδομένων x και των παραμέτρων θ μέσω μιας συνάρτησης $p(x|\theta)$ η οποία καλείται συνάρτηση πιθανοφάνειας. Τέλος, με τη βοήθεια του κανόνα του Bayes ορίζεται η εκ των υστέρων (posterior) κατανομή

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)} \quad (1.1)$$

η οποία ποσοτικοποιεί την πληροφορία για το θ , μετά την παρατήρηση των δεδομένων (Hoff, 2009). Σε αντίθεση με την Κλασική στατιστική, η Μπεϋζιανή έναντι των διαστημάτων εμπιστοσύνης χρησιμοποιεί τα 95% διαστήματα αξιοπιστίας. Δεδομένου ότι η παράμετρος θ είναι τυχαία μεταβλητή που προέρχεται από μια συνάρτηση πιθανότητας, τα διαστήματα αξιοπιστίας ερμηνεύονται ως εξής: υπάρχει 95% πιθανότητα η πραγμα-

τική τιμή της παραμέτρου ϑ να βρίσκεται μέσα σε αυτό το διάστημα (Casella and Berger, 2002). Σημειώνεται και σε αυτή την περίπτωση, ότι το ποσοστό 95% είναι ενδεικτικό και είναι στην ευχέρεια του κάθε ερευνητή να επιλέξει το κατάλληλο ποσοστό για την εξαγωγή συμπερασμάτων.

1.2 Στατιστικό Μοντέλο και Αβεβαιότητα

Η χρήση των στατιστικών μοντέλων είναι απαραίτητη, καθώς χρησιμοποιούνται για την αναπαράσταση και την ερμηνεία πολύπλοκων φαινομένων, καθώς και για την πρόβλεψη μελλοντικών γεγονότων (Bernardo and Smith 1994). Πως ορίζεται όμως ένα στατιστικό μοντέλο;

Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ ένας χώρος πιθανοτήτων όπου Ω είναι ο δειγματικός χώρος, δηλαδή το σύνολο όλων των πιθανών αποτελεσμάτων, \mathcal{F} είναι μία σ -άλγεβρα, δηλαδή ο χώρος των ενδεχομένων του Ω και \mathbb{P} ένα μέτρο πιθανότητας ορισμένο στη σ -άλγεβρα. Επιπλέον, έστω \mathbf{X} ο χώρος καταστάσεων με \mathcal{X} τη Borel σ -άλγεβρα των υποσυνόλων του \mathbf{X} , $\mathbf{x} = (x_1, x_2, \dots, x_n)$ το παρατηρούμενο δείγμα από πραγματοποιήσεις των τυχαίων μεταβλητών X_i στο χώρο καταστάσεων \mathbf{X} και $\mathcal{P}_{\mathbf{X}}$ ο χώρος όλων των μέτρων πιθανότητας στον \mathbf{X} .

Ορισμός 1.2.1 (Στατιστικό μοντέλο, Wasserman, 2000). Έστω Θ ένας παραμετρικός χώρος και $\mathcal{P}_{\mathbf{X}}$ ο χώρος όλων των μέτρων πιθανότητας ενός χώρου καταστάσεων \mathbf{X} . Ένα στατιστικό μοντέλο ορίζεται να είναι ένα υποσύνολο \mathcal{M} του $\mathcal{P}_{\mathbf{X}}$ όπου κάθε στοιχείο του \mathcal{M} είναι παραμετροποιημένο ως προς κάποια παράμετρο $\vartheta \in \Theta$.

$$\mathcal{M} = \{P_{\vartheta} | \vartheta \in \Theta\} \subseteq \mathcal{P}_{\mathbf{X}}$$

Η στατιστική συμπερασματολογία περιλαμβάνει συμπεράσματα σχετικά με την παράμετρο ϑ . Στην περίπτωση που παραμετρικός χώρος Θ είναι πεπερασμένης διάστασης, τότε το μοντέλο είναι παραμετρικό, ενώ στην περίπτωση που είναι άπειρης διάστασης, τότε το μοντέλο είναι μη παραμετρικό.

Κατά τη στατιστική μοντελοποίηση συχνά εμφανίζεται η έννοια της αβεβαιότητας, είτε λόγω της τυχαιότητας ενός συστήματος που εξελίσσεται, είτε εξαιτίας του μεγάλου αριθμού των ενεργών βαθμών ελευθερίας. Η στατιστική συμπερασματολογία παρέχει πληροφορίες σχετικά με αυτή την αβεβαιότητα, η οποία εκφράζεται με όρους πιθανοτήτων. Συγκεκριμένα, η Μπεϋζιανή προσέγγιση, η οποία χρησιμοποιείται ευρέως λόγω της ραγδαίας εξέλιξης των υπολογιστικών μεθόδων της, εκφράζει την αβεβαιότητα μέσω κάποιων παραμέτρων οι οποίες μοντελοποιούνται ως τυχαίες μεταβλητές που χαρακτηρίζονται από μια κατανομή πιθανότητας $p(\vartheta)$, τη γνωστή prior κατανομή. Η prior κατανομή, όπως έχει ήδη αναφερθεί, περιέχει όλη τη γνώση για την παράμετρο ϑ πριν την παρατήρηση των δεδομένων. Στη συνέχεια, υπό την παρατήρηση των δεδομένων, εξαγονται συμπεράσματα για την παράμετρο ϑ μέσω της posterior κατανομής $p(\vartheta|x)$ (Berger, 1985; Robert et al. 2010; Kaloudis, 2019).

1.3 Ανταλλαξιμότητα

Η βασική προϋπόθεση που χρησιμοποιείται στη Μπεϋζιανή συμπερασματολογία είναι η ανταλλαξιμότητα, σε αντίθεση με την Κλασική στην οποία χρησιμοποιείται η ανεξαρτησία. Η υπόθεση της ανταλλαξιμότητας είναι ασθενέστερη από αυτή της ανεξαρτησίας, αλλά ταυτόχρονα ισχυρότερη από την υπόθεση των ισόνομων μεταβλητών. Επιπλέον, ισχύει ότι αν κάποιες μεταβλητές είναι ανεξάρτητες και ισόνομες, τότε είναι και αντάλλαξιμες, ενώ σύμφωνα με το θεώρημα αναπράστασης του de Finetti αν είναι ανταλλάξιμες, τότε είναι και υπό συνθήκη ανεξάρτητες και ισόνομες δοθέντος ενός μέτρου πιθανότητας (Heath and Sudderth, 1976; Poirier, 2011).

Ορισμός 1.3.1 (Ανταλλαξιμότητα, Heath and Sudderth, 1976). Μία πεπερασμένη ακολουθία τυχαίων μεταβλητών $X^{(n)} = (X_1, X_2, \dots, X_n)$ είναι ανταλλάξιμη αν οι $n!$ μεταθέσεις των τυχαίων μεταβλητών $(X_{k_1}, X_{k_2}, \dots, X_{k_n})$ προέρχονται από την ίδια κατανομή πιθανότητας με τις τυχαίες μεταβλητές (X_1, X_2, \dots, X_n) . Οι μεταβλητές μιας άπειρης ακολουθίας $(X_m)_{m \geq 1}$ είναι ανταλλάξιμες, αν κάθε πεπερασμένη υπακολουθία της $(X_m)_{m \geq 1}$ είναι επίσης ανταλλάξιμη.

Με άλλα λόγια, η κατανομή των μεταβλητών (X_1, X_2, \dots, X_n) παραμένει αναλλοίωτη σε οποιαδήποτε μετάθεση των δεικτών. Παρακάτω δίνεται το παράδειγμα του μοντέλου Ρόλγια για την καλύτερη κατανόηση του ορισμού της ανταλλαξιμότητας.

Παράδειγμα 1.3.1. (Briggs, 2016). Έστω ότι έχουμε ένα δοχείο με n άσπρες και m μαύρες μπάλες. Κατά τη διάρκεια του πειράματος επιλέγουμε τυχαία μια μπάλα, σημειώνουμε το χρώμα της μπάλας που τραβήξαμε και την επανατοποθετούμε μέσα στο δοχείο προσθέτοντας άλλη μία μπάλα του ίδιου χρώματος. Στη συνέχεια, επιλέγουμε μια δεύτερη μπάλα από το δοχείο και επαναλαμβάνουμε την ίδια διαδικασία. Σύμφωνα με τα δεδομένα, η πιθανότητα η πρώτη μπάλα που θα τραβήξουμε να είναι άσπρη είναι $\frac{n}{n+m}$. Έστω τώρα ότι η πρώτη μπάλα που επιλέγουμε είναι άσπρη, ακολουθώντας τη διαδικασία του πειράματος, οι άσπρες μπάλες που υπάρχουν στο δοχείο αυτή τη στιγμή είναι $n+1$. Λαμβάνοντας υπόψη τα δεδομένα του πειράματος έως τώρα, η πιθανότητα η δεύτερη μπάλα να είναι και αυτή άσπρη είναι $\frac{n+1}{n+m+1}$. Στην περίπτωση που η πρώτη μπάλα ήταν μαύρη, τότε η πιθανότητα η δεύτερη μπάλα να είναι άσπρη είναι $\frac{n}{n+m+1}$. Η γνώση του ποια μπάλα επιλέχθηκε πρώτα είναι σχετική με τη γνώση της πιθανότητας η δεύτερη μπάλα να είναι άσπρη. Δηλαδή, έχουμε ότι

$$\mathbb{P}(\text{πρώτη μπάλα άσπρη} \mid n, m) = \frac{n}{n+m}$$

Τότε,

$$\mathbb{P}(\text{δεύτερη μπάλα άσπρη} \mid n+1, m) = \frac{n+1}{n+m+1}$$

ή

$$\mathbb{P}(\text{δεύτερη μπάλα άσπρη} \mid n, m + 1) = \frac{n}{n + m + 1}$$

Δοθέντων των δεδομένων E , μπορούμε να υπολογίσουμε την πιθανότητα η δεύτερη μπάλα να είναι άσπρη.

$$\begin{aligned} \mathbb{P}(\text{δεύτερη μπάλα άσπρη} \mid E) &= \mathbb{P}(\text{πρώτη άσπρη-δεύτερη άσπρη ή πρώτη μαύρη-δεύτερη άσπρη} \mid E) \\ &= \frac{n}{n + m} \frac{n + 1}{n + m + 1} + \frac{m}{n + m} \frac{n}{n + m + 1} \\ &= \frac{n(n + 1) + mn}{(n + m)(n + m + 1)} \\ &= \frac{n(n + m + 1)}{(n + m)(n + m + 1)} \\ &= \frac{n}{n + m} \end{aligned}$$

Η πιθανότητα η πρώτη μπάλα να είναι άσπρη είναι ίση με την πιθανότητα η δεύτερη μπάλα να είναι άσπρη. Αν συνεχίσουμε να υπολογίζουμε τις πιθανότητες η τρίτη, τέταρτη κ.ο.κ μπάλα να είναι άσπρη θα δούμε ότι και αυτές οι πιθανότητες είναι επίσης ίσες, δηλαδή $\frac{n}{n+m}$. Αυτό συμβαίνει διότι στην ουσία αυξάνουμε το δοχείο με άσπρες και μαύρες μπάλες της ίδιας περίπου ποσότητας. Ωστόσο, το γεγονός αυτό είναι συνέπεια της ανταλλαξιμότητας. Για να κατανοήσουμε καλύτερα τον ορισμό της ανταλλαξιμότητας θα πρέπει να υπολογίσουμε τις εξής πιθανότητες $\mathbb{P}(\text{πρώτη άσπρη-δεύτερη μαύρη} \mid E)$ και $\mathbb{P}(\text{πρώτη μαύρη-δεύτερη άσπρη} \mid E)$.

$$\begin{aligned} \mathbb{P}(\text{πρώτη άσπρη-δεύτερη μαύρη} \mid E) &= \frac{n}{n + m} \frac{m}{n + m + 1} \\ &= \frac{nm}{(n + m)(n + m + 1)} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(\text{πρώτη μαύρη-δεύτερη άσπρη} \mid E) &= \frac{m}{n + m} \frac{n}{n + m + 1} \\ &= \frac{nm}{(n + m)(n + m + 1)} \end{aligned}$$

Παρατηρούμε ότι η πιθανότητα παραμένει αναλλοίωτη στην αλλαγή της σειράς με την οποία επιλέγονται οι μπάλες και αυτό ισχύει όσο μεγάλος κι αν είναι ο αριθμός των μπαλών που εξάγονται από το δοχείο. Συμπεραίνουμε λοιπόν, ότι η πιθανότητα δεν σχετίζεται με τη σειρά με την οποία επιλέγονται οι μπάλες αλλά εξαρτάται μόνο από τον αριθμό των άσπρων και τον αριθμό των μαύρων μπαλών που θα επιλεχθούν.

□

Το θεώρημα αναπαράστασης του De Finetti αποδείχθηκε αρχικά για ακολουθίες με δυαδικές τιμές και αργότερα επεκτάθηκε και για τυχαίες παραγματικές μεταβλητές. Το θεώρημα αυτό στην ουσία αποδεικνύει ότι μια άπειρη ακολουθία τυχαίων μεταβλητών είναι ανταλλάξιμη, αν και μόνο αν οι τυχαίες μεταβλητές είναι υπό συνθήκη ανεξάρτητες και ισόνομες δοθέντος ένα μέτρο πιθανότητας, (Bernardo and Smith, 1994) . Για να δοθεί το θεώρημα αναπαράστασης του De Finetti πρέπει να δοθεί πρώτα ο ορισμός του μέτρου πιθανότητας.

Ορισμός 1.3.2 (Μέτρο πιθανότητας, Billingsley 1995). Έστω δειγματικός χώρος Ω . Η συνολοσυνάρτηση είναι μία πραγματική συνάρτηση που ορίζεται από μία συλλογή υποσυνόλων του δειγματικού χώρου Ω , δηλαδή από μία σ -άλγεβρα. Τότε, μία συνολοσυνάρτηση P από μια σ -άλγεβρα \mathcal{F} είναι μέτρο πιθανότητας αν ικανοποιεί τις εξής συνθήκες:

- i. $0 \leq P(A) \leq 1$ για $A \in \mathcal{F}$
- ii. $P(\emptyset) = 0$ και $P(\Omega) = 1$
- iii. Αν $A_1, A_2, \dots \in \mathcal{F}$ με $A_i \cap A_j = \emptyset$ για κάθε $i \neq j$ και αν $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ τότε

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

□

Θεώρημα 1.3.1 (Θεώρημα Αναπαράστασης De Finetti (Bernardo and Smith 1994)). Αν $(X_n)_{n \geq 1}$ είναι μία άπειρη ακολουθία ανταλλάξιμων μεταβλητών με $X_i \in \{0, 1\} \forall i$ και μέτρο πιθανότητας P , τότε υπάρχει μια συνάρτηση κατανομής Q τέτοια ώστε η από κοινού συνάρτηση πιθανότητας $p(x_1, \dots, x_n)$ να έχει τη μορφή

$$p(x_1, \dots, x_n) = \int_0^1 \prod_{r=1}^n \vartheta^{x_i} (1 - \vartheta)^{n-x_i} dQ(\vartheta).$$

όπου,

$$Q(\vartheta) = \lim_{r \rightarrow \infty} P\left[\frac{y_r}{r} \leq \vartheta\right].$$

με

$$y_n = \sum_{i=1}^n x_i \text{ και } \vartheta = \lim_{n \rightarrow \infty} \frac{y_n}{n}$$

□

Όπως αναφέρθηκε νωρίτερα, το θεώρημα αυτό αναφέρεται σε ακολουθίες μεταβλητών Bernoulli, ωστόσο έχει επεκταθεί και για άλλες περιπτώσεις.

Θεώρημα 1.3.2 (Bernardo and Smith 1994). *Έστω X_1, X_2, \dots είναι μία άπειρη ακολουθία ανταλλάξιμων τυχαίων μεταβλητών με μέτρο πιθανότητας p . Τότε, υπάρχει ένα μέτρο πιθανότητας Q τέτοιο ώστε η από κοινού κατανομή των X_1, X_2, \dots να μπορεί να γραφτεί ως*

$$p(x_1, \dots, x_n) = \int_G \prod_{r=1}^n G(x_r) dQ(G)$$

όπου G είναι ο χώρος των συναρτήσεων πιθανότητας, $Q(G) = \lim_{n \rightarrow \infty} p(G_n)$ και G_n είναι η εμπειρική συνάρτηση κατανομής που ορίζεται από τα X_1, \dots, X_n .

Στα παραπάνω θεωρήματα, η συνάρτηση Q ονομάζεται μέτρο De Finetti και έχει το ρόλο της prior κατανομής, ενώ η G έχει το ρόλο της πιθανοφάνειας. Στην ουσία, το θεώρημα αναπαράστασης του De Finetti αναφέρει ότι μία ακολουθία τυχαίων μεταβλητών είναι ανταλλάξιμη αν και μόνο αν είναι μια μίξη ακολουθιών ανεξάρτητων και ισόνομων τυχαίων μεταβλητών (Bernardo et al. 1988).

Στο επόμενο κεφάλαιο θα δοθούν βασικές πληροφορίες σχετικά με διάφορες μεθόδους Μπεϋζιανής εκτίμησης, θα γίνει εφαρμογή αυτών των μεθόδων σε πεπερασμένα μοντέλα μίξης και θα αναφερθούν τα προβλήματα που μπορεί να παρουσιαστούν κατά την εφαρμογή τους.

Κεφάλαιο 2

Μπεϋζιανή Εκτίμηση

2.1 Κατηγορίες υπολογιστικών μεθόδων Μπεϋζιανής εκτίμησης

Όπως ήδη αναφέρθηκε στο προηγούμενο κεφάλαιο, στη Μπεϋζιανή στατιστική η αβεβαιότητα εκφράζεται μέσω της έννοιας των πιθανοτήτων. Απαιτείται, λοιπόν, ο υπολογισμός της posterior κατανομής. Όμως, για να γίνει αυτό χρειάζεται ο υπολογισμός του αθροίσματος ή του ολοκληρώματος κάποιων μεταβλητών που δεν είναι πάντα εύκολος. Έτσι, αναπτύχθηκαν δύο βασικές κατηγορίες υπολογιστικών μεθόδων Μπεϋζιανής εκτίμησης με σκοπό τον υπολογισμό των κατανομών που είναι δύσκολες στον προγραμματισμό τους. Η πρώτη μεγάλη κατηγορία περιλαμβάνει τις μεθόδους μεταβολών (Variational Inference), ενώ η δεύτερη περιλαμβάνει τις μεθόδους Markov Chain Monte Carlo (MCMC).

2.1.1 Μέθοδοι Μεταβολών (Variational Inference)

Η βασική ιδέα των VI μεθόδων για τον υπολογισμό της posterior κατανομής, στηρίζεται ως επί το πλείστον στην βελτιστοποίηση, κατά την οποία η posterior προσεγγίζεται μέσω μια απλούστερης κατανομής. Έστω θ η παράμετρος ενδιαφέροντος, η οποία προέρχεται από μια κατανομή, και x οι παρατηρήσεις. Στόχος είναι η προσέγγιση της posterior κατανομής $p(\theta|x)$. Πρώτο βήμα είναι ο καθορισμός μιας οικογένειας κατανομών Q , η οποία ανατίθεται στην παράμετρο θ . Επόμενο βήμα είναι η εύρεση κατάλληλης κατανομής, η οποία ανήκει στην οικογένεια Q , τέτοια ώστε να ελαχιστοποιεί το μέτρο απόκλισης Kullback–Leibler (KL)¹ από την πραγματική posterior κατανομή. Δηλαδή η

¹Το μέτρο αυτό είναι γνωστό και ως σχετική εντροπία και είναι ένα μέτρο ομοιότητας μεταξύ δύο κατανομών πιθανότητας. Αν f, g δύο κατανομές πιθανότητας τότε

$$KL(f||g) \stackrel{\text{def}}{=} \int f(x) \log \frac{f(x)}{g(x)} dx = E_f \left[\log \frac{f(x)}{g(x)} \right]$$

εύρεση της κατανομής:

$$q^*(\vartheta) = \arg \min_{q(\vartheta) \in \mathcal{Q}} KL(q(\vartheta) || p(\vartheta|x)) \quad (2.1)$$

Η ζητούμενη posterior δίνεται από την q^* .

Στις μεθόδους αυτές δεν χρειάζεται ο υπολογισμός των σταθερών κανονικοποίησης, παρά μόνο ο τύπος της από κοινού πυκνότητας της παραμέτρου ϑ και των δεδομένων x . Ένα ακόμα πλεονέκτημα των μεθόδων, είναι ότι ο αλγόριθμος είναι γρήγορος υπολογιστικά και για το λόγο αυτό μπορεί να εφαρμοστεί σε προβλήματα με μεγάλο σύνολο δεδομένων. Ωστόσο, μειονεκτούν στο γεγονός ότι δεν υπάρχει ακριβής υπολογισμός αλλά προσεγγιστικός. Επίσης, η επιλογή της συνάρτησης που ελαχιστοποιείται οδηγεί σε υποεκτιμήσεις της διασποράς (Blei et al. 2017; Zhang et al. 2019; Hershey and Olsen, 2007). Ενδεικτικά, κάποιοι γνωστοί αλγόριθμοι VI είναι οι εξής: Coordinate Ascent VI (CAVI), Automatic Differentiation Variational Inference (ADVI), Black Box Variational Inference (BBVI) και Stochastic Variational Inference (SVI) (Ganguly and Earp, 2021; Kucukelbir et al. 2016; Hoffman et al. 2012).

2.1.2 Markov Chain Monte Carlo

Η βασική ιδέα των MCMC μεθόδων στηρίζεται στην παραγωγή μιας εργοδικής Μαρκοβιανής αλυσίδας $\vartheta^{(m)}$ η οποία έχει ως στάσιμη κατανομή, την κατανομή ενδιαφέροντος $f(\vartheta|x)$, η οποία ονομάζεται κατανομή στόχος.

Ορισμός 2.1.1 (Μαρκοβιανή Αλυσίδα). Μια στοχαστική διαδικασία $X = \{X_n : n \in \mathbb{N}\}$ ονομάζεται μαρκοβιανή αλυσίδα όταν ισχύει ότι

$$P \{X_{n+1} = j | X_0, X_1, \dots, X_n\} = P \{X_{n+1} = j | X_n\} \quad (2.2)$$

για κάθε $j \in E$, όπου E ο χώρος καταστάσεων, και $n \in \mathbb{N}$.

Με άλλα λόγια, κάθε επόμενη κατάσταση της αλυσίδας εξαρτάται μόνο από την τελευταία κατάσταση και είναι ανεξάρτητη όλων των προηγούμενων καταστάσεων (Cinlar, 1975).

Όπως αναφέρθηκε, οι MCMC μέθοδοι παράγουν μια μαρκοβιανή αλυσίδα, η οποία είναι εργοδική. Με αυτή την ιδιότητα εξασφαλίζεται ότι η κατανομή της αλυσίδας $\vartheta^{(m)}$ συγκλίνει στην κατανομή στόχο f για κάθε αρχική τιμή $\vartheta^{(0)}$ και άρα δεν εξαρτάται από την αρχική τιμή. Επιπλέον, όταν μια μαρκοβιανή αλυσίδα έχει στάσιμη κατανομή, αυτό σημαίνει ότι η κατανομή στην οποία συγκλίνει η αλυσίδα παραμένει αναλλοίωτη ως προς τις μεταβάσεις της αλυσίδας (Robert, 2007; Robert and Casella, 2010).

Ένα μειονέκτημα των μεθόδων αυτών είναι το υπολογιστικό κόστος που δημιουργείται κατά την εφαρμογή τους. Ωστόσο, σε αντίθεση με τις VI μεθόδους, τα δείγματα που

παράγονται από τους αλγορίθμους MCMC είναι ασυμπτωτικά πιο ακριβή. Επιπλέον, διαχειρίζονται καλύτερα προβλήματα στα οποία η κατανομή στόχος είναι μια μίξη κατανομών (Blei et al. 2017). Για το λόγο αυτό, στην παρούσα εργασία, οι εκτιμήσεις των παραμέτρων σε παραμετρικά και μη παραμετρικά μοντέλα μίξης θα γίνουν αποκλειστικά με τη χρήση MCMC μεθόδων.

Παρακάτω, στην ενότητα 2.2 δίνονται αναλυτικότερα οι αλγόριθμοι Metropolis-Hastings, Gibbs και Slice, που είναι χαρακτηριστικά παραδείγματα MCMC αλγορίθμων και αποτελούν το βασικό εργαλείο που θα χρησιμοποιηθεί στα επόμενα κεφάλαια.

2.2 Κατηγορίες MCMC μεθόδων

2.2.1 Ο αλγόριθμος Metropolis-Hastings

Ο αλγόριθμος Metropolis-Hastings (MH) είναι ένας από τους πιο δημοφιλείς MCMC αλγορίθμους που προτάθηκε από τον Nicholas Metropolis το 1953 και επεκτάθηκε αργότερα από τον Wilfred Keith Hastings το 1970 (Tierney, 1994).

Σκοπός του αλγορίθμου είναι να εξασφαλίσει τη δειγματοληψία από μια κατανομή $f(\cdot)$ η οποία ονομάζεται κατανομή στόχος, όταν η άμεση δειγματοληψία από αυτή την κατανομή είναι δύσκολη. Η ιδέα του αλγορίθμου περιλαμβάνει μια κατανομή $q(\cdot|x)$, από την οποία είναι εύκολη η δειγματοληψία και ονομάζεται κατανομή πρότασης, σε συνδυασμό με ένα μηχανισμό ο οποίος αποδέχεται τα δείγματα που προτείνονται από την κατανομή $q(\cdot|x)$ ως δείγματα της f . Έχει αποδειχθεί ότι ο αλγόριθμος MH έχει στάσιμη κατανομή και δεν είναι άλλη από την κατανομή στόχο (Tierney, 1994).

Αναλυτικότερα, ο αλγόριθμος ξεκινάει με την αρχικοποίηση του $\vartheta^{(0)}$, τέτοιο ώστε $f(\vartheta) > 0$. Στη συνέχεια, σε κάθε επανάληψη t του αλγορίθμου παράγεται ένας αριθμός $\xi \sim q(\xi|\vartheta^{(t-1)})$ από την κατανομή πρότασης. Ο αριθμός αυτός γίνεται δεκτός με πιθανότητα

$$\alpha(\xi, \vartheta^{(t-1)}) = \min \left\{ 1, \frac{f(\xi)q(\vartheta^{(t-1)}|\xi)}{f(\vartheta^{(t-1)})q(\xi|\vartheta^{(t-1)})} \right\}. \quad (2.3)$$

Η πιθανότητα $\alpha(\xi, \vartheta^{(t-1)})$ ονομάζεται πιθανότητα αποδοχής και δεν εξαρτάται από τη σταθερά κανονικοποίησης. Παρακάτω στον Αλγόριθμο 1 δίνεται αναλυτικά ο αλγόριθμος MH για ένα k -διάστατο διάνυσμα παραμέτρων $\vartheta = (\theta_1, \dots, \theta_k)$.

Algorithm 1 Metropolis-Hastings

1: **procedure** ΔΕΙΓΜΑ ΑΠΟ $(\vartheta = (\theta_1, \dots, \theta_k))$ 2: Αρχικοποίηση $\vartheta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ 3: **for** $t = 1$ to N **do**4: Δείγμα από την κατανομή πρότασης $\xi \sim q(\xi|\vartheta^{(t-1)})$ 5: Υπολογισμός πιθανότητας αποδοχής $\alpha(\xi, \vartheta^{(t-1)})$ 6: Δείγμα από $u \sim \mathcal{U}(0, 1)$

7: Υπολογισμός

$$\vartheta^{(t)} = \begin{cases} \xi & \text{αν } u < \alpha(\xi, \vartheta^{(t-1)}) \\ \vartheta^{(t-1)} & \text{,αλλιώς} \end{cases} \quad (2.4)$$

8: **end for**9: **end procedure**

Ο αλγόριθμος εξαρτάται αποκλειστικά από τους λόγους

$$\frac{f(\xi)}{f(\vartheta^{(t-1)})} \quad \text{και} \quad \frac{q(\vartheta^{(t-1)}|\xi)}{q(\xi|\vartheta^{(t-1)})} \quad (2.5)$$

συνεπώς, είναι ανεξάρτητος από οποιαδήποτε σταθερά κανονικοποίησης εφόσον και η συνάρτηση πρότασης είναι γνωστή έως μια σταθερά που είναι ανεξάρτητη του x . Η επιλογή της κατανομής πρότασης είναι ελεύθερη, ωστόσο η κατανομή αυτή πρέπει να ικανοποιεί τον εξής περιορισμό

$$\text{supp} f \subseteq \bigcup_{x \in \text{supp} f} \text{supp} q(\cdot|x) \quad (2.6)$$

δηλαδή, το στήριγμα της κατανομής στόχου να είναι υποσύνολο της ένωσης των στηρίγμάτων της κατανομής πρότασης για κάθε x που ανήκει στο στήριγμα της κατανομής στόχου. Αυτή η προϋπόθεση εξασφαλίζει την πλήρη εξερεύνηση της αλυσίδας στο χώρο (Robert and Casella, 2010).

Ο αρχικός αλγόριθμος που πρότεινε ο Metropolis περιλάμβανε στις προϋποθέσεις για την επιλογή της κατανομής πρότασης και τη συμμετρία της κατανομής, δηλαδή έπρεπε να ισχύει $q(x|y) = q(y|x)$ (Metropolis et al., 1953). Αργότερα, ο Hastings επέκτεινε την εφαρμογή του αλγορίθμου και για μη συμμετρικές κατανομές πρότασης (Hastings, 1970). Στην περίπτωση που ικανοποιείται η συνθήκη της συμμετρίας, τότε η πιθανότητα αποδοχής μετατρέπεται σε

$$\alpha(\xi, \vartheta^{(t-1)}) = \min \left\{ \frac{f(\xi)}{f(\vartheta^{(t-1)})}, 1 \right\}. \quad (2.7)$$

Αυτή ορίζεται να είναι και η πιθανότητα αποδοχής του αλγορίθμου Random Walk Metropolis .

Έστω (x, y) δύο στοιχεία του χώρου της μαρκοβιανής αλυσίδας. Η μαρκοβιανή αλυσίδα που παράγεται από τον αλγόριθμο MH είναι αντιστρέψιμη με στάσιμη κατανομή την f , διότι ικανοποιεί τη συνθήκη της αναλυτικής ισορροπίας.

$$\text{Εξίσωση αναλυτικής ισορροπίας: } f(y)K(x, y) = f(x)K(y, x) \quad (2.8)$$

Ο πυρήνας μετάβασης της μαρκοβιανής αλυσίδας είναι:

$$K(x, y) = \alpha(x, y)q(y|x) + \left(1 - \underbrace{\int \alpha(x, y)q(y|x) dy}_{r(x)} \right) I(x = y) \quad (2.9)$$

όπου ο δεύτερος όρος του πυρήνα εμφανίζεται όταν το υποψήφιο σημείο απορρίπτεται και η αλυσίδα παραμένει στην ίδια κατάσταση.

Επομένως, προκύπτουν οι εξισώσεις:

$$\alpha(x, y)q(y|x)f(x) = \alpha(y, x)q(x|y)f(y) \quad (2.10)$$

$$(1 - r(x))I(x = y)f(x) = (1 - r(y))I(y = x)f(y) \quad (2.11)$$

δηλαδή, ικανοποιείται η συνθήκη της αναλυτικής ισορροπίας, επομένως η $f(\cdot)$ είναι η ζητούμενη στάσιμη κατανομή (Rizzo, 2007).

2.2.2 Ο αλγόριθμος Gibbs

Ο δειγματολήπτης Gibbs προτάθηκε από τους Geman και Geman το 1984 και μπορεί να θεωρηθεί ειδική περίπτωση του αλγορίθμου Metropolis-Hastings , με πιθανότητα αποδοχής ίση με τη μονάδα (Robert and Casella, 2010). Εφαρμόζεται ιδιαίτερα σε περιπτώσεις που η δειγματοληψία γίνεται από πολυμεταβλητές κατανομες. Αυτό συμβαίνει διότι ο αλγόριθμος μετατρέπει το πρόβλημα σε πρόβλημα μικρότερης διάστασης. Αναλυτικότερα, έστω ότι η κατανομή στόχος είναι η πολυμεταβλητή κατανομή $f(\theta_1, \dots, \theta_k|x)$. Η δειγματοληψία από την κατανομή αυτή, γίνεται μέσω των πλήρων δεσμευμένων κατανομών της (Full Conditionals ή αλλιώς FCs) $f(\theta_j|\theta_{-j})$, όπου

$$\theta_{(-j)} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k), \quad j = 1, \dots, k$$

Παρακάτω, στον Αλγόριθμο 2, δίνεται αναλυτικά ο αλγόριθμος Gibbs για ένα k -διάστατο διάνυσμα παραμέτρων.

Algorithm 2 Gibbs

```
1: procedure ΔΕΙΓΜΑ ΑΠΌ ( $\vartheta = (\theta_1, \dots, \theta_k)$ )
2:   Αρχικοποίηση  $\vartheta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ 
3:   for  $i = 1$  to  $N$  do
4:     for  $j = 1$  to  $d$  do
5:       Δείγμα από  $\theta_j^{(i)} \sim f_{\theta_j|\theta_{-j}}(\theta_j|\theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i)}, \dots, \theta_k^{(i)})$ 
6:     end for
7:     Υπολογισμός  $\vartheta^{(i)} = (\theta_1^{(i)}, \dots, \theta_k^{(i)})$ 
8:   end for
9: end procedure
```

Σημειώνεται ότι κάθε θ_j μπορεί να είναι και μεγαλύτερης διάστασης, οπότε στην περίπτωση αυτή η δειγματοληψία γίνεται σε κομμάτια (Gelman and Rubin, 1996). Επιπλέον, για την επιτυχή εφαρμογή του αλγορίθμου, χρειάζεται οι πλήρως δεσμευμένες κατανομές να είναι γνωστές μορφές κατανομών. Στην περίπτωση που η δειγματοληψία από κάποια από τις πλήρως δεσμευμένες κατανομές δεν είναι εφικτή, τότε εισάγουμε ένα βήμα Metropolis, δηλαδή η δειγματοληψία θα γίνει με τη βοήθεια μιας κατανομής πρότασης όπως αναφέρθηκε στην ενότητα 2.2.1. Αυτός είναι ο γνωστός αλγόριθμος Metropolis within Gibbs (Gilks et al. 1995).

2.2.3 Ο αλγόριθμος Slice

Υπάρχουν περιπτώσεις στις οποίες είναι δύσκολο να εφαρμοστούν οι αλγόριθμοι Metropolis-Hastings και Gibbs. Για το λόγο αυτό, ο Neal (2003) και οι Damien et al. (1999) πρότειναν μία νέα MCMC μέθοδο. Η βασική ιδέα της μεθόδου στηρίζεται στο γεγονός ότι κάποιος μπορεί να δημιουργήσει ένα δείγμα, κάνοντας ομοιόμορφη δειγματοληψία από την περιοχή που βρίσκεται κάτω από την καμπύλη της κατανομής ενδιαφέροντος. Ένας τρόπος, για να πραγματοποιηθεί αυτό είναι να γίνει ομοιόμορφη δειγματοληψία από τις οριζόντιες «φέτες» που δημιουργούνται στην τρέχουσα κατακόρυφη θέση. Η μέθοδος αυτή μπορεί να εφαρμοστεί τόσο σε μονοδιάστατα προβλήματα, όσο και σε πολυδιάστατα (Neal, 2003).

Το παρακάτω θεώρημα αποτελεί το βασικό θεμέλιο πάνω στο οποίο δημιουργήθηκε η μέθοδος αυτή.

Θεώρημα 2.2.1 (Damien et al. 1999). Έστω f η κατανομή από την οποία θα παραχθούν τυχαίες μεταβλητές και δίνεται από τη σχέση

$$f(x) \propto \pi(x) \prod_{i=1}^N l_i(x), \quad (2.12)$$

όπου π μία γνωστή κατανομή και τα l_i είναι μη αρνητικές αντιστρέψιμες συναρτήσεις για τις οποίες, αν $l_i(x) > u$ τότε μπορεί να δημιουργηθεί το σύνολο $A_u^i = \{x : l_i(x) > u\}$. Τότε μπορεί να εφαρμοστεί ο αλγόριθμος Gibbs, στην περίπτωση που όλες οι πλήρως δεσμευμένες κατανομές είναι ομοιόμορφες κατανομές εκτός από μία που θεωρείται μία περικομμένη εκδοχή της κατανομής π .

Σύμφωνα με τους Damien et al. 1999, έστω f η κατανομή από την οποία θα παραχθεί δείγμα

$$f(\vartheta) \propto \pi(\vartheta)l(\vartheta) \quad (2.13)$$

όπου π είναι μία γνωστή πυκνότητα και l μία μη αρνητική αντιστρέψιμη συνάρτηση. Έστω u μία νέα μεταβλητή, τότε η από κοινού πυκνότητα των u και ϑ γράφεται ως εξής

$$f(\vartheta, u) \propto \pi(\vartheta)I(u < l(\vartheta)). \quad (2.14)$$

όπου αν κάποιος ολοκληρώσει ως προς u την παραπάνω σχέση, τότε επιστρέφει πάλι στη σχέση (2.13). Εφαρμόζοντας τον αλγόριθμο Gibbs οι πλήρεις δεσμευμένες κατανομές που προκύπτουν για τα u και ϑ είναι:

- $u \sim \mathcal{U}(0, l(\vartheta))$
- μια περικομμένη π στο σύνολο

$$A_u = \{\vartheta : u < l(\vartheta)\}.$$

Παράδειγμα 2.2.1. (Damien and Walker, 2001). Έστω X τυχαίες μεταβλητές από μία περικομμένη τυπική κανονική κατανομή στο διάστημα (a, b) και έστω ότι θέλουμε να παράξουμε δείγμα από αυτές. Δηλαδή, θέλουμε να παράξουμε δείγμα από την

$$f_X(x) \propto \exp\left(-\frac{x^2}{2}\right)I(x \in (a, b)). \quad (2.15)$$

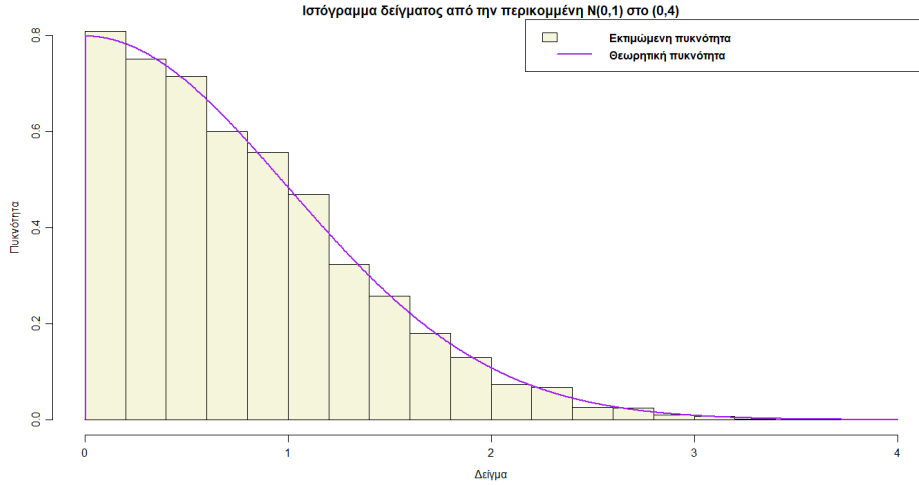
Εισάγουμε την τυχαία μεταβλητή Y . Η από κοινού πυκνότητα της X και της Y είναι

$$f_{X,Y}(x, y) \propto I_{(0, \exp(-\frac{x^2}{2}))}(y)I(x \in (a, b)). \quad (2.16)$$

Οπότε, οι πλήρεις δεσμευμένες κατανομές που προκύπτουν είναι

- $Y|(X = x) \sim \mathcal{U}\left(0, \exp\left(-\frac{x^2}{2}\right)\right)$
- $X|(Y = y) \sim \mathcal{U}\left(\max\{a, -\sqrt{-2\log y}\}, \min\{b, \sqrt{-2\log y}\}\right)$

Ακολουθώντας το παράδειγμα, εφαρμόστηκε ο αλγόριθμος Slice για την παραγωγή ενός δείγματος από την περικομμένη τυπική κανονική κατανομή στο διάστημα (0,4). Τα αποτελέσματα παρουσιάζονται στο Σχήμα 2.1 όπου φαίνεται το Ιστόγραμμα του δείγματος που παράχθηκε. Επιπλέον, με μωβ χρώμα φαίνεται η καμπύλη της θεωρητικής κατανομής.



Σχήμα 2.1: Ιστόγραμμα περικομμένης κανονικής κατανομής στο διάστημα (0,4)

2.3 Πεπερασμένα Μπεϋζιανά μοντέλα μίξης

Η ανάγκη για τη δημιουργία των μοντέλων μίξης γεννήθηκε καθώς τα δεδομένα σε διάφορα προβλήματα μπορεί να προέρχονται από δύο ή περισσότερους πληθυσμούς σε διαφορετικές αναλογίες (McLachlan and Basford, 1988).

Μία μίξη κατανομών είναι ο κυρτός συνδυασμός C κατανομών,

$$\sum_{c=1}^C w_c f_c(x) \quad (2.17)$$

όπου $0 < w_c < 1$ και $\sum_{c=1}^C w_c = 1$. Οι κατανομές f_c συνήθως ανήκουν σε μια οικογένεια κατανομών όπως για παράδειγμα η Κανονική, η Βήτα ή η Γάμμα με παραμέτρους ϑ_c . Επομένως προκύπτει το μοντέλο μίξης

$$\sum_{c=1}^C w_c f_c(x|\vartheta_c) \quad (2.18)$$

με παραμέτρους w_c και ϑ_c . Τα w_c ονομάζονται βάρη της μίξης. Ένα μοντέλο μίξης μπορεί να εκφραστεί με όρους ελλειπών δεδομένων. Δηλαδή, αν z_i , ($i = 1, \dots, n$) είναι

ένα διάνυσμα διάστασης C όπου κάθε στοιχείο του υποδεικνύει από ποιο στοιχείο της μίξης προήλθε η παρατήρηση x_i , τέτοιο ώστε $z_{ic} \in \{0, 1\}$ και $\sum_{c=1}^C z_{ic} = 1$, τότε η πυκνότητα των (x_i, z_i) μπορεί να γραφεί ως (Diebolt and Robert, 1994; Marin and Robert, 2014)

$$\prod_{c=1}^C w_c^{z_{ic}} f_c^{z_{ic}}(x_i). \quad (2.19)$$

Για την εκτίμηση των παραμέτρων σε Μπεϋζιανά μοντέλα μίξης εφαρμόζεται συνήθως ο αλγόριθμος Gibbs. Έστω $X = (X_1, \dots, X_n)$ δεδομένα που προέρχονται από το μοντέλο μίξης κανονικών κατανομών (Gaussian Mixture Model)

$$X|\vartheta \sim \sum_{c=1}^C w_c N(x|\mu_c, \tau_c), \quad (2.20)$$

όπου μ_c η μέση τιμή και $\tau_c = \frac{1}{\sigma_c^2}$ η ακρίβεια του στοιχείου c της μίξης. Το διάνυσμα των παραμέτρων του μοντέλου είναι

$$\vartheta = (\mu, \tau, w)$$

με $\mu = (\mu_1, \dots, \mu_C)$, $\tau = (\tau_1, \dots, \tau_C)$ και $w = (w_1, \dots, w_C)$.

Εισάγεται μία νέα βοηθητική μεταβλητή $Z_i = (Z_{i1}, \dots, Z_{iC})$ η οποία υποδεικνύει από ποιο στοιχείο της μίξης προήλθε η παρατήρηση X_i . Για τη Z_{ic} ισχύει ότι

$$z_{ic} = \begin{cases} 1 & \text{Αν ανήκει στο στοιχείο } c \text{ της μίξης} \\ 0 & \text{αλλιού} \end{cases} \quad (2.21)$$

με

$$Z_i|\vartheta \stackrel{\text{iid}}{\sim} \text{Multinom}(1, w = (w_1, \dots, w_C))$$

και

$$X_i|Z_i = z_i, \vartheta \stackrel{i}{\sim} N(x_i|\mu_{z_i}, \tau_{z_i}).$$

Η πιθανοφάνεια του μοντέλου δίνεται από:

$$\begin{aligned} L(\vartheta, z) &= \prod_{i=1}^n f(x_i|Z_i = z_i, \mu, \tau, w) P(Z_i = z_i|\mu, \tau, w) \\ &= \prod_{i=1}^n \left(\prod_{c=1}^C N(x_i|\mu_c, \tau_c)^{z_{ic}} \prod_{c=1}^C w_c^{z_{ic}} \right) \\ &= \prod_{i=1}^n \prod_{c=1}^C (w_c N(x_i|\mu_c, \tau_c))^{z_{ic}} \end{aligned} \quad (2.22)$$

Βασικό βήμα για την εκτίμηση των παραμέτρων του μοντέλου είναι η επιλογή των prior κατανομών για τις παραμέτρους αυτές.

$$\begin{aligned}\mu_c &\stackrel{\text{iid}}{\sim} N(\mu_0, \tau_0) \\ \tau_c &\stackrel{\text{iid}}{\sim} \mathcal{G}(\alpha_0, \beta_0) \\ w &\sim \text{Dirichlet}(\alpha_1, \dots, \alpha_C)\end{aligned}$$

Αφού έγινε ο καθορισμός των εκ των προτέρων κατανομών, μπορεί να ακολουθήσει ο υπολογισμός των πλήρων δεσμευμένων κατανομών για την εφαρμογή του αλγορίθμου Gibbs.

1. Για το μ_c :

$$\begin{aligned}\pi(\mu_c | \tau_c, z, x) &\propto \pi_{\mu_c}(\mu_c) L_c(\vartheta, z) \\ &\propto N(\mu_c | \mu_0, \tau_0) \prod_{i=1}^n (N(x_i | \mu_c, \tau_c))^{z_{ic}} \\ &\propto \exp \left\{ -\tau_0 \frac{(\mu_c - \mu_0)^2}{2} \right\} \prod_{i=1}^n \exp \left\{ -\tau_c z_{ic} \frac{(x_i - \mu_c)^2}{2} \right\} \\ &= \exp \left\{ -\tau_0 \frac{(\mu_c - \mu_0)^2}{2} \right\} \times \exp \left\{ -\tau_c \sum_{i=1}^n z_{ic} \frac{(x_i - \mu_c)^2}{2} \right\} \\ &= \exp \left\{ -\frac{\tau_0}{2} (\mu_c^2 - 2\mu_c\mu_0 + \mu_0^2) - \frac{\tau_c}{2} \sum_{i=1}^n z_{ic} (x_i^2 - 2x_i\mu_c + \mu_c^2) \right\} \\ &\propto \exp \left\{ -\frac{\tau_0}{2} \mu_c^2 + \tau_0 \mu_c \mu_0 + \tau_c \sum_{i=1}^n z_{ic} x_i \mu_c - \frac{\tau_c}{2} \sum_{i=1}^n z_{ic} \mu_c^2 \right\} \\ &= \exp \left\{ -\frac{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}}{2} \left(\mu_c^2 - 2 \frac{\frac{\tau_0}{2} \mu_0 + \frac{\tau_c}{2} \sum_{i=1}^n z_{ic} x_i}{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}} \mu_c \right) \right\} \\ &= \exp \left\{ -\frac{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}}{2} \left(\mu_c^2 - 2 \frac{\tau_0 \mu_0 + \tau_c \sum_{i=1}^n z_{ic} x_i}{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}} \mu_c \pm \left[\frac{\tau_0 \mu_0 + \tau_c \sum_{i=1}^n z_{ic} x_i}{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}} \right]^2 \right) \right\} \\ &\propto \exp \left\{ -\frac{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}}{2} \left(\mu_c - \frac{\tau_0 \mu_0 + \tau_c \sum_{i=1}^n z_{ic} x_i}{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}} \right)^2 \right\} \tag{2.23}\end{aligned}$$

προκύπτει ότι η σχέση (2.23) είναι ο πυρήνας μιας Κανονικής Κατανομής.

$$\text{Συνεπώς, } \mu_c | \tau_c, z, x \sim N \left(\frac{\tau_0 \mu_0 + \tau_c \sum_{i=1}^n z_{ic} x_i}{\tau_0 + \tau_c \sum_{i=1}^n z_{ic}}, (\tau_0 + \tau_c \sum_{i=1}^n z_{ic})^{-1} \right)$$

2. Για το τ_c :

$$\begin{aligned} \pi(\tau_c | \mu_c, z, x) &\propto \pi_{\tau_c}(\tau_c) L_c(\vartheta, z) \\ &\propto \tau_c^{\alpha_0 - 1} \exp\{-\beta_0 \tau_c\} \prod_{i=1}^n \left(\tau_c^{\frac{1}{2}} \exp\left\{-\tau_c \frac{(x_i - \mu_c)^2}{2}\right\} \right)^{z_{ic}} \\ &= \tau_c^{\alpha_0 - 1} \exp\{-\beta_0 \tau_c\} \prod_{i=1}^n \tau_c^{\frac{z_{ic}}{2}} \exp\left\{-z_{ic} \tau_c \frac{(x_i - \mu_c)^2}{2}\right\} \\ &= \tau_c^{\alpha_0 - 1} \tau_c^{\frac{\sum_{i=1}^n z_{ic}}{2}} \exp\{-\beta_0 \tau_c\} \exp\left\{-\tau_c \frac{\sum_{i=1}^n z_{ic} (x_i - \mu_c)^2}{2}\right\} \\ &= \tau_c^{\alpha_0 + \frac{\sum_{i=1}^n z_{ic}}{2} - 1} \exp\left\{-\left(\beta_0 + \frac{\sum_{i=1}^n z_{ic} (x_i - \mu_c)^2}{2}\right) \tau_c\right\} \end{aligned} \quad (2.24)$$

Είναι φανερό ότι η σχέση (2.24) είναι ο πυρήνας μιας Γάμμα Κατανομής.

$$\text{Συνεπώς, } \tau_c | \mu_c, z, x \sim \mathcal{G} \left(\alpha_0 + \frac{\sum_{i=1}^n z_{ic}}{2}, \beta_0 + \frac{\sum_{i=1}^n z_{ic} (x_i - \mu_c)^2}{2} \right)$$

3. Για τα βάρη w :

$$\begin{aligned} \pi(w | z) &\propto \pi_w(w) L_c(\vartheta, z) \\ &\propto \prod_{c=1}^C w_c^{\alpha_c - 1} \prod_{c=1}^C \prod_{i=1}^n w_c^{z_{ic}} \\ &= \prod_{c=1}^C w_c^{\alpha_c - 1} \prod_{c=1}^C w_c^{\sum_{i=1}^n z_{ic}} \end{aligned}$$

$$= \prod_{c=1}^C w_c^{\alpha_c + \sum_{i=1}^n z_{ic} - 1} \quad (2.25)$$

Η σχέση (2.25) είναι ο πυρήνας μιας *Dirichlet Κατανομής*, της οποίας ο ορισμός δίνεται αναλυτικά στο επόμενο κεφάλαιο στην ενότητα 3.1.

Συνεπώς, $w|z \sim \text{Dirichlet}(\alpha_c + \sum_{i=1}^n z_{ic})$.

4. Για το z_i :

$$\begin{aligned} \pi(z_i|\mu, \tau, w, x_i) &\propto L_c(\vartheta, z) \\ &\propto \prod_{c=1}^C (w_c N(x_i|\mu_c, \tau_c))^{z_{ic}} \\ &= \prod_{c=1}^C \left(\frac{w_c N(x_i|\mu_c, \tau_c)}{\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c)} \sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c) \right)^{z_{ic}} \\ &= \prod_{c=1}^C \left(\frac{w_c N(x_i|\mu_c, \tau_c)}{\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c)} \right)^{z_{ic}} \prod_{c=1}^C \left(\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c) \right)^{z_{ic}} \\ &= \prod_{c=1}^C \left(\frac{w_c N(x_i|\mu_c, \tau_c)}{\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c)} \right)^{z_{ic}} \left(\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c) \right)^{\sum_{c=1}^C z_{ic}} \\ &\propto \prod_{c=1}^C \left(\frac{w_c N(x_i|\mu_c, \tau_c)}{\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c)} \right)^{z_{ic}} \end{aligned} \quad (2.26)$$

Προκύπτει ότι η σχέση (2.26) είναι ο πυρήνας μιας *Multinomial Κατανομής*.

$$\text{Συνεπώς, } z_i|\mu, \tau, w, x_i \sim \text{Multinomial} \left(1, p_{ic} = \frac{w_c N(x_i|\mu_c, \tau_c)}{\sum_{c=1}^C w_c N(x_i|\mu_c, \tau_c)} \right)$$

Παράδειγμα. Έστω η μίξη των κανονικών κατανομών $N(0,1)$, $N(5,1)$ και $N(10,1)$ με βάρη 0.5, 0.3, 0.2 αντίστοιχα. Δηλαδή,

$$0.5N(0,1) + 0.3N(5,1) + 0.2N(10,1) \quad (2.27)$$

Οι παράμετροι του μοντέλου είναι

- $\boldsymbol{w} = (w_1, w_2, w_3) = (0.5, 0.3, 0.2)$
- $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3) = (0, 5, 10)$
- $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3) = (1, 1, 1)$

Με σκοπό την εκτίμηση των παραμέτρων του μοντέλου μίξης (2.27), εφαρμόστηκε ο αλγόριθμος Gibbs. Παράχθηκε δείγμα 4000 παρατηρήσεων ύστερα από περίοδο burn-in² 1000 παρατηρήσεων και λέπτυνση (thinning)³ ανά 5 παρατηρήσεις. Για τις εκτιμήσεις των παραμέτρων, υπολογίστηκε η μέση τιμή των τιμών που προέκυψαν από τις επαναλήψεις του αλγορίθμου και τα αποτελέσματα παρουσιάζονται παρακάτω.

Για τα βάρη \boldsymbol{w} οι εκτιμήσεις είναι οι εξής:

Για το w_1 η εκτίμηση είναι **0.4849**.

Για το w_2 η εκτίμηση είναι **0.3064**.

Για το w_3 η εκτίμηση είναι **0.2087**.

Για τις μέσες τιμές $\boldsymbol{\mu}$ οι εκτιμήσεις είναι:

Για το μ_1 η εκτίμηση είναι **0.0207**.

Για το μ_2 η εκτίμηση είναι **5.0214**.

Για το μ_3 η εκτίμηση είναι **9.9266**.

Για τις τυπικές αποκλίσεις $\boldsymbol{\sigma} = \frac{1}{\sqrt{\boldsymbol{\tau}}}$ οι εκτιμήσεις είναι οι εξής:

Για το σ_1 η εκτίμηση είναι **0.9267**.

Για το σ_2 η εκτίμηση είναι **0.9297**.

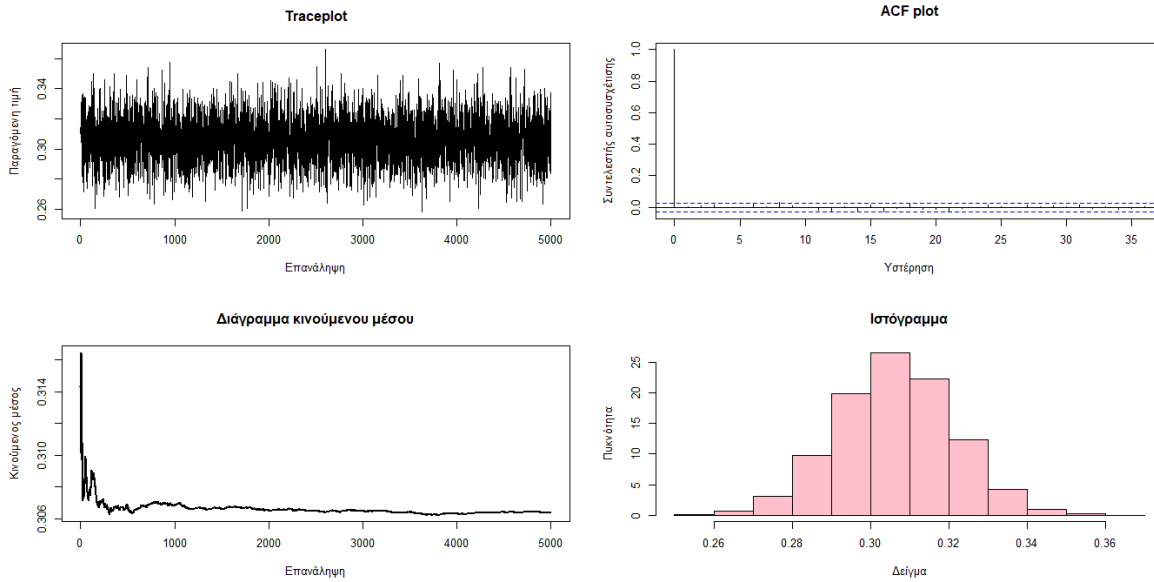
Για το σ_3 η εκτίμηση είναι **1.0257**.

Στο Σχήμα 2.2 φαίνεται το διάγραμμα ίχνους (Trace plot), το διάγραμμα αυτοσυσχέτισης (ACF plot), το διάγραμμα κινούμενου μέσου και το ιστογράμμο για τον έλεγχο σύγκλισης της αλυσίδας για την εκτίμηση της παραμέτρου w_2 .

²Απόρριψη του αρχικού μέρους του δείγματος μέχρι να επιτευχθεί σύγκλιση στην επιθυμητή κατανομή

³Επιλογή δειγματικού σημείου ανά k παρατηρήσεις

Διαγράμματα για το w_2



Σχήμα 2.2: Trace plot, ACF plot, διάγραμμα κινούμενου μέσου και Ιστόγραμμα για το w_2

Το διάγραμμα ίχνους είναι ένα διάγραμμα που στον οριζόντιο άξονα αποτυπώνει τους αριθμούς των επανηλήψεων και στον κατακόρυφο άξονα τις τιμές της παραμέτρου που εκτίμησε ο αλγόριθμος σε κάθε επανάληψη. Εάν στο διάγραμμα δεν υπάρχει ένδειξη κάποιας συστηματικής τάσης, τότε η αλυσίδα που παράχθηκε από τον αλγόριθμο συγκλίνει στην πραγματική κατανομή (Dong and Peng, 2013). Στο διάγραμμα ίχνους, στο Σχήμα 2.2, δεν φαίνεται κάποια τάση, επομένως το συμπέρασμα είναι ότι η αλυσίδα έχει καλή σύγκλιση.

Το διάγραμμα αυτοσυσχέτισης δείχνει τις αυτοσυσχετίσεις μεταξύ των τιμών της αλυσίδας σε διάφορες υστερήσεις. Εάν οι κατακόρυφες γραμμές δεν υπερβαίνουν κατά πολύ τις μπλε διακεκομμένες γραμμές, σημαίνει ότι δεν υπάρχει ένδειξη για αυτοσυσχέτιση μεταξύ των τιμών (Dong and Peng, 2013). Στο Σχήμα 2.2 φαίνεται ότι ικανοποιείται αυτή η συνθήκη, επομένως δεν υπάρχει αυτοσυσχέτιση μεταξύ των τιμών του w_2 .

Το διάγραμμα κινούμενου μέσου δείχνει σε κάθε επανάληψη, τη μέση τιμή των τιμών της αλυσίδας μέχρι την τρέχουσα επανάληψη. Στο Σχήμα 2.2 η τιμή φαίνεται να συγκλίνει στην τιμή 0.306, η οποία είναι πολύ κοντά στην πραγματική τιμή του w_2 που είναι 0.3.

Τέλος, στο Σχήμα 2.2 δίνεται επίσης και το Ιστόγραμμα της κατανομής της παραμέτρου w_2 που εκτίμησε ο αλγόριθμος Gibbs.

Τα αποτελέσματα για την παράμετρο w_2 είναι ενδεικτικά, καθώς προέκυψαν παρόμοια

αποτελέσματα και για τις υπόλοιπες παραμέτρους που δεν συμπεριλήφθηκαν στην εργασία για λόγους συντομίας.

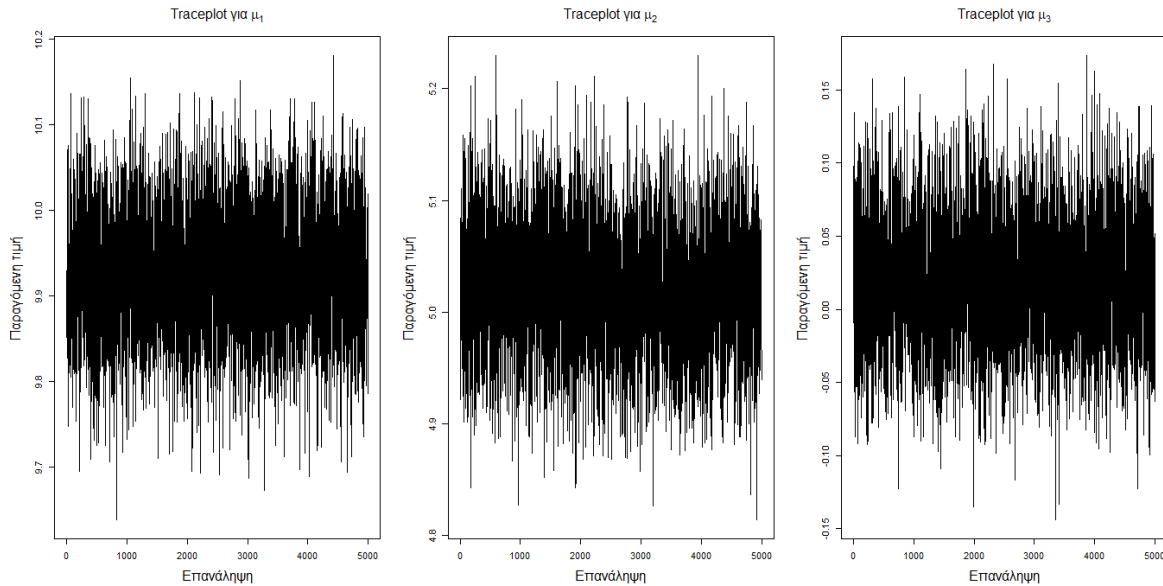
2.4 Πρόβλημα Εναλλαγής Ετικέτας

Ένα πρόβλημα που δημιουργείται κατά την εφαρμογή MCMC μεθόδων σε πεπερασμένα μοντέλα μίξης είναι αυτό της μη ταυτοποιησιμότητας. Αναλυτικότερα, πριν την εφαρμογή κάποιας μεθόδου, πρώτο βήμα είναι η ανάθεση prior κατανομών στις παραμέτρους του μοντέλου μίξης. Στην περίπτωση όμως που οι κατανομές αυτές είναι ανταλλάξιμες, η posterior, που είναι απόρροια της αλυσίδας που δημιουργήσε ο αλγόριθμος, παραμένει αναλλοίωτη στις μεταθέσεις των δεικτών των παραμέτρων της. Με άλλα λόγια, οι περιθώριες posterior κατανομές των παραμέτρων είναι ταυτόσημες για κάθε στοιχείο της μίξης (Jasra et al. 2005; Marin and Robert, 2014). Το πρόβλημα αυτό είναι γνωστό ως Πρόβλημα εναλλαγής ετικέτας (Label Switching problem) και παρατηρείται στις Μπεϋζιανές μεθόδους εκτίμησης αλλά όχι στις κλασικές μεθόδους, όπως για παράδειγμα η εκτίμηση μέγιστης πιθανοφάνειας (Frühwirth-Schnatter, 2006).

Το πρόβλημα αυτό δημιουργείται λόγω της συμμετρίας που παρουσιάζει η πιθανοφάνεια του μοντέλου, δηλαδή η πιθανοφάνεια παραμένει η ίδια στις διάφορες μεταθέσεις των δεικτών των παραμέτρων. Αν δεν υπάρχει a priori πληροφορία για τη διάκριση μεταξύ των στοιχείων της μίξης, τότε για όλες τις μεταθέσεις των παραμέτρων η prior είναι η ίδια και άρα η posterior είναι επίσης συμμετρική (Stephens, 2000).

Για παράδειγμα, στο Σχήμα 2.3 απεικονίζονται τα τρία διαγράμματα ίχνους για τις παραμέτρους μ_1 , μ_2 και μ_3 , του προηγούμενου παραδείγματος, αντίστοιχα. Σύμφωνα με τα διαγράμματα, φαίνεται ότι η συμπεριφορά της αλυσίδας που παράχθηκε για την παράμετρο μ_1 αντιστοιχεί στο τρίτο διάγραμμα για την παράμετρο μ_3 , ενώ η συμπεριφορά της αλυσίδας για το μ_3 αποτυπώνεται στο διάγραμμα για το μ_1 .

Διαγράμματα ίχνους των παραμέτρων μ_1, μ_2, μ_3



Σχήμα 2.3: Διαγράμματα ίχνους για τις μέσες τιμές της τριπλής μίξης (2.27)

Με την πάροδο των χρόνων προτάθηκαν διάφορες μέθοδοι για τον περιορισμό του προβλήματος εναλλαγής ετικέτας. Η πιο διαδεδομένη μέθοδος, που υποστήριξαν ιδιαίτερα οι Diebolt και Robert (1994), καθώς και οι Richardson και Green (1997), είναι αυτή των περιορισμών ταυτοποιησιμότητας (Identifiability Constraints) που τίθενται στον παραμετρικό χώρο, με στόχο να περιορίσουν τη συνθήκη της συμμετρίας της πιθανοφάνειας. Κατά τη μέθοδο αυτή, ορίζονται διατάξεις μεταξύ των παραμέτρων, για παράδειγμα $\theta_1 < \theta_2$, οι οποίες ικανοποιούν μόνο μια μετάθεση των δεικτών. Αργότερα, οι Celeux et al. (2000) πρότειναν μία θεωρητική μέθοδο που στηρίζεται στη χρήση μιας συνάρτησης απώλειας, η οποία μένει αναλλοίωτη στις μεταθέσεις των δεικτών, και στην ελαχιστοποίηση της εκ των υστέρων αναμενόμενης απώλειας (Papastamoulis and Iliopoulos, 2010; Jasra et al. 2005).

Στο επόμενο κεφάλαιο παρουσιάζεται τόσο η διαδικασία Dirichlet, όσο και κάποιες αναπαράστασεις της. Στη συνέχεια, γίνεται η εισαγωγή στα μη πεπερασμένα μοντέλα μίξης και δίνεται αναλυτικά ο αλγόριθμος για τη δειγματοληψία του μοντέλου μίξης Dirichlet με slices που πρότεινε ο Walker (2007).

Κεφάλαιο 3

Μη παραμετρική Μπεϋζιανή Στατιστική

Για τη Μπεϋζιανή συμπερασματολογία σε μη παραμετρικά προβλήματα χρειάζεται η ύπαρξη μιας εκ των προτέρων κατανομής για την παράμετρο, η οποία, όπως έχει ήδη αναφερθεί, είναι άπειρης διάστασης. Σε αυτό το κεφάλαιο, θα χρησιμοποιηθεί η διαδικασία Dirichlet ως η εκ των προτέρων κατανομή για την απειροδιάστατη παράμετρο σε μη παραμετρικά προβλήματα.

3.1 Διαδικασία Dirichlet

Η διαδικασία Dirichlet προτάθηκε από τον Ferguson το 1973 ως prior στο χώρο των μέτρων πιθανότητας και έκτοτε χρησιμοποιείται ευρέως σε μη παραμετρικά προβλήματα εκτίμησης (Jara, 2017). Για τον ορισμό της διαδικασίας Dirichlet θα πρέπει πρώτα να δοθεί ο ορισμός της κατανομής Dirichlet, η οποία χρησιμοποιείται ως συζυγής¹ prior της Πολυωνυμικής κατανομής.

Ορισμός 3.1.1 (Κατανομή Dirichlet, Kai Wang Ng et al. 2011). Έστω $Z_i \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha_i, 1)$, $1 \leq i \leq n$ και $Z = \sum_{i=1}^n Z_i$. Το τυχαίο διάνυσμα $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ με $Y_i = \frac{Z_i}{Z}$, $1 \leq i \leq n$, ακολουθεί την κατανομή Dirichlet με παραμέτρους $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ και συνάρτηση πυκνότητας πιθανότητας

$$\mathcal{D}(\mathbf{Y} | \alpha) = \frac{\Gamma\left(\sum_{j=1}^n \alpha_j\right)}{\prod_{j=1}^n \Gamma(\alpha_j)} \prod_{i=1}^n y_i^{\alpha_i - 1} \mathcal{I}(\mathbf{Y} \in \Delta^n)$$

με στήριγμα $\Delta^n = \{(y_1, y_2, \dots, y_n) \in \mathbb{R}_n^+ : \sum_{i=1}^n y_i = 1\}$.

¹Συζυγή prior ονομάζουμε την κατανομή η οποία σε συνδυασμό με την συνάρτηση πιθανοφάνειας, οδηγεί στην εκ των υστέρων κατανομή που ανήκει και αυτή στην ίδια οικογένεια κατανομών με την prior.

Ο Ferguson (1973) χρησιμοποιώντας το θεώρημα συνέπειας του Kolmogorov απέδειξε την ύπαρξη της διαδικασίας Dirichlet, η οποία θεωρείται η απειροδιάστατη γενίκευση της κατανομής Dirichlet (Müller et al. 2015). Για περισσότερες πληροφορίες σχετικά με το θεώρημα συνέπειας του Kolmogorov, ο αναγνώστης παραπέμπεται στη βιβλιογραφία (Billingsley, 1995). Επομένως, τώρα μπορεί να δοθεί και ο ορισμός της διαδικασίας Dirichlet.

Ορισμός 3.1.2 (Διαδικασία Dirichlet, Ferguson 1973). Το τυχαίο μέτρο \mathbb{P} ακολουθεί τη διαδικασία Dirichlet $\mathbb{P} \sim \mathcal{DP}(c, G_0)$ με παράμετρο συγκέντρωσης $c > 0$ και μέτρο βάσης $G_0 \in \mathcal{P}_{\mathbf{X}}$, αν και μόνο αν για κάθε πεπερασμένη διαμέριση A_1, A_2, \dots, A_n του \mathbf{X} το διάνυσμα των τυχαίων πιθανοτήτων $(\mathbb{P}(A_1), \mathbb{P}(A_2), \dots, \mathbb{P}(A_n))$ ακολουθεί την κατανομή Dirichlet

$$(\mathbb{P}(A_1), \mathbb{P}(A_2), \dots, \mathbb{P}(A_n)) \sim \mathcal{D}(cG_0(A_1), cG_0(A_2), \dots, cG_0(A_n)).$$

Δηλαδή, το τυχαίο μέτρο \mathbb{P} προέρχεται από μια διαδικασία Dirichlet αν κάθε μία από τις περιθώριες κατανομές είναι Dirichlet.

Έστω η διαμέριση του δειγματικού χώρου που αποτελείται από το γεγονός A και από το συμπληρωματικό του A^c , δηλαδή $\{A, A^c\}$. Τότε,

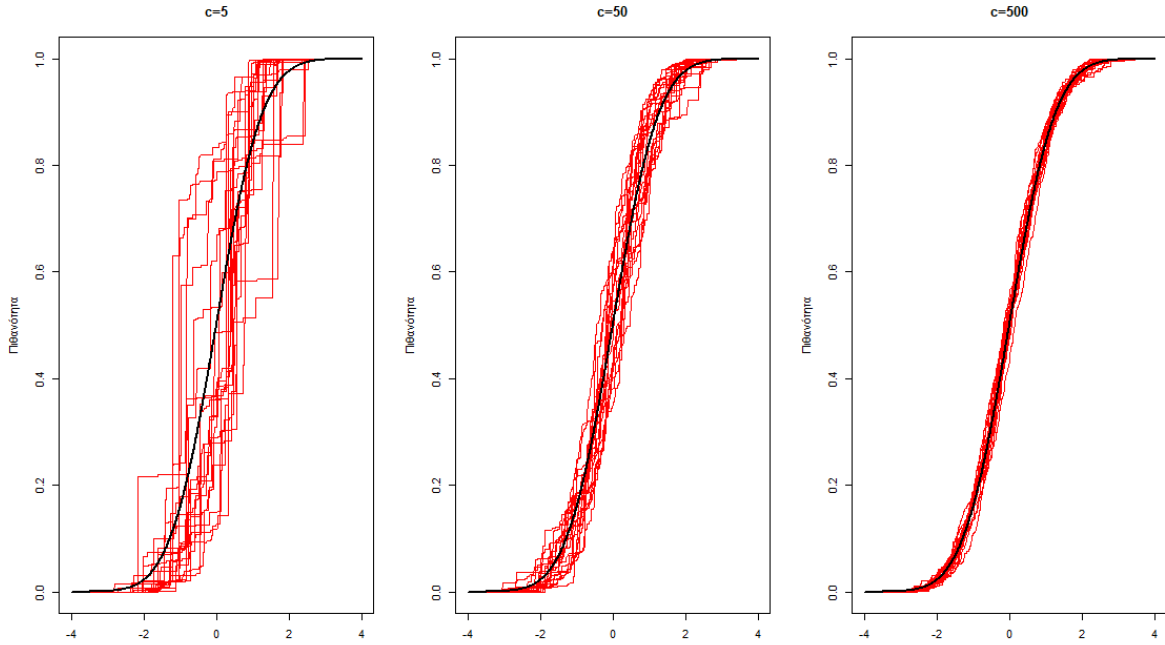
$$\mathbb{P}(A) \sim \mathcal{Be}(cG_0(A), cG_0(A^c)).$$

Από τις ιδιότητες της Βήτα κατανομής προκύπτει ότι

- $\mathbb{E}[\mathbb{P}(A)] = \frac{cG_0(A)}{(cG_0(A) + cG_0(A^c))} = G_0(A)$
- $Var[\mathbb{P}(A)] = \frac{G_0(A)(1 - G_0(A))}{c + 1}$

Αυτό σημαίνει ότι το μέτρο βάσης G_0 παίζει το ρόλο της μέσης τιμής στο DP, ενώ η παράμετρος συγκέντρωσης c έχει το ρόλο της ακρίβειας, δηλαδή της αντίστροφης διασποράς. Αυτό σημαίνει ότι όσο μεγαλύτερη είναι η παράμετρος συγκέντρωσης, τόσο τα μέτρα που παράγονται από το DP πλησιάζουν το μέτρο βάσης G_0 (Hjort et al. 2010).

Στο Σχήμα 3.1 παρουσιάζονται 3 διαγράμματα. Σε κάθε διάγραμμα αποτυπώνονται 20 τυχαία μέτρα που παράχθηκαν από ένα DP, αλλάζοντας κάθε φορά την παράμετρο συγκέντρωσης. Είναι φανερό ότι επιβεβαιώνεται ο παραπάνω ισχυρισμός, δηλαδή όσο μεγαλώνει η τιμή της παραμέτρου συγκέντρωσης, τόσο τα μέτρα που παράγονται από το DP τείνουν στο μέτρο βάσης. Σημειώνεται ότι στο συγκεκριμένο παράδειγμα ως μέτρο βάσης χρησιμοποιήθηκε η τυπική κανονική κατανομή.



Σχήμα 3.1: Οι εμπειρικές συναρτήσεις κατανομών 20 τυχαίων μέτρων DP, με παράμετρο συγκέντρωσης $c = 5, 50, 500$ και μέτρο βάσης $G_0 = N(0, 1)$.

Συζυγία

Έστω \mathbb{P} ένα τυχαίο μέτρο που προέρχεται από ένα DP, δηλαδή $\mathbb{P} \sim \mathcal{DP}(c, G_0)$ και έστω x_1, \dots, x_n ανεξάρτητες τιμές από το \mathbb{P} στο \mathbf{X} . Επιπλέον, έστω A_1, \dots, A_r μία πεπερασμένη μετρήσιμη διαμέριση του \mathbf{X} και $n_k = \#\{i : x_i \in A_k\}$. Λόγω του ότι η κατανομή Dirichlet είναι συζυγής με την Πολυωνυμική κατανομή, ισχύει ότι

$$(\mathbb{P}(A_1), \dots, \mathbb{P}(A_r)) | x_1, \dots, x_n \in \mathcal{D}(cG_0(A_1) + n_1, \dots, cG_0(A_r) + n_r). \quad (3.1)$$

το οποίο ισχύει για κάθε πεπερασμένη μετρήσιμη διαμέριση του \mathbf{X} . Αυτό σημαίνει ότι η posterior $\mathbb{P}|x_1, \dots, x_n$ είναι και αυτή ένα DP.

$$\mathbb{P}|x_1, \dots, x_n \sim \mathcal{DP}\left(c + n, \frac{c}{c + n}G_0 + \frac{n}{c + n} \frac{\sum_{i=1}^n \delta_{x_i}}{n}\right) \quad (3.2)$$

Επομένως, χρησιμοποιώντας ως prior ένα DP, η ζητούμενη posterior είναι και αυτή ένα DP. Δηλαδή, ικανοποιείται η συνθήκη της συζυγίας. Επιπλέον, από τη σχέση (3.2) φαίνεται ότι το μέτρο βάσης της posterior είναι ο σταθμισμένος μέσος του G_0 και της εμπειρικής συνάρτησης $\frac{\sum_{i=1}^n \delta_{x_i}}{n}$, όπου το βάρος του μέτρου βάσης G_0 είναι ανάλογο

της παραμέτρου c και το βάρος της εμπειρικής συνάρτησης είναι ανάλογο του αριθμού των παρατηρήσεων (Teh, 2017).

Predictive Κατανομή

Έστω x_1, \dots, x_n δείγμα από το τυχαίο μέτρο \mathbb{P} , όπου $\mathbb{P} \sim \mathcal{DP}(c, G_0)$. Η predictive κατανομή για μία νέα παρατήρηση $x_{n+1}|x_1, \dots, x_n$, που προκύπτει ολοκληρώνοντας ως προς το τυχαίο μέτρο \mathbb{P} , για ένα μετρήσιμο σύνολο $A \subset \mathbf{X}$ δίνεται από

$$\begin{aligned} P(x_{n+1} \in A|x_1, \dots, x_n) &= \mathbb{E}[\mathbb{P}(A)|x_1, \dots, x_n] \\ &= \frac{c}{c+n} G_0(A) + \frac{n}{c+n} \frac{\sum_{i=1}^n \delta_{x_i}(A)}{n} \\ &= \frac{1}{c+n} \left(c G_0(A) + \sum_{i=1}^n \delta_{x_i}(A) \right) \end{aligned} \quad (3.3)$$

Επομένως, ισχύει ότι

$$x_{n+1}|x_1, \dots, x_n \sim \frac{1}{c+n} \left(c G_0(A) + \sum_{i=1}^n \delta_{x_i} \right) \quad (3.4)$$

(Teh, 2017).

Στην επόμενη ενότητα, παρουσιάζονται κάποιες διαδικασίες, οι οποίες χρησιμοποιούνται ως κατασκευαστικοί ορισμοί για την αναπαράσταση της διαδικασίας Dirichlet.

3.2 Αναπαραστάσεις της Διαδικασίας Dirichlet

Η διαδικασία Dirichlet αποτελεί βασικό εργαλείο της Μπεϋζιανής μη παραμετρικής στατιστικής και έχει πρωταγωνιστικό ρόλο στην παρούσα εργασία. Για το λόγο αυτό, γεννήθηκε η ανάγκη αναπαράστασης του τυχαίου μέτρου DP, μέσω κάποιων κατασκευαστικών ορισμών με σκοπό την εφαρμογή του σε διάφορα προβλήματα. Στην ενότητα αυτή παρουσιάζονται τρεις διαφορετικές αναπαραστάσεις της διαδικασίας Dirichlet.

Γενικευμένο Pólya urn

Ένας τρόπος αναπαράστασης ενός τυχαίου μέτρου DP γίνεται με τη βοήθεια του γενικευμένου Pólya urn που προτάθηκε από τους Blackwell και MacQueen το 1973. Η ονομασία της διαδικασίας προέρχεται από τον παρακάτω συμβολισμό. Έστω ένα άδειο δοχείο, ένα σύνολο μοναδικών χρωμάτων \mathbf{X} και ένας άπειρος αριθμός μπαλών. Επιλέγεται τυχαία ένα χρώμα από το μέτρο βάσης G_0 , χρωματίζεται η πρώτη μπάλα $x_1 \sim G_0$ και τοποθετείται μέσα στο δοχείο. Στη συνέχεια, επιλέγεται μια δεύτερη μπάλα η οποία είτε θα χρωματιστεί με κάποιο νέο χρώμα με πιθανότητα $\frac{c}{c+1}$, είτε θα χρωματιστεί με

το χρώμα της μπάλας που υπάρχει ήδη μέσα στο δοχείο με πιθανότητα $\frac{1}{c+1}$ και στη συνέχεια θα τοποθετηθεί στο δοχείο. Επαναλαμβάνοντας τη διαδικασία n φορές, στο $n + 1$ βήμα επιλέγεται μία νέα μπάλα η οποία χρωματίζεται είτε με κάποιο νέο χρώμα από το μέτρο βάσης $x_{n+1} \sim G_0$ με πιθανότητα $\frac{c}{c+n}$ είτε με το χρώμα κάποιας μπάλας που υπάρχει ήδη μέσα στο δοχείο με πιθανότητα $\frac{n}{c+n}$ και τοποθετείται μέσα στο δοχείο. Επομένως προκύπτει η εξής αναπαράσταση:

Συμβολίζοντας με y_1, \dots, y_k τις μοναδικές τιμές των x_1, \dots, x_n και τη συχνότητα εμφάνισης με n_i , $1 \leq i \leq k$, τότε η κατανομή της νέας παρατήρησης γράφεται ως

$$x_{n+1} | x_1, \dots, x_n \sim \begin{cases} \delta_{y_i}, & \text{με πιθανότητα } \frac{n_i}{c+n}, i = 1, \dots, k \\ G_0, & \text{με πιθανότητα } \frac{c}{c+n}. \end{cases} \quad (3.5)$$

Σημειώνεται ότι όσο αυξάνεται ο αριθμός των μπαλών n , τόσο θα αυξάνεται και η πιθανότητα η νέα μπάλα να χρωματιστεί με το χρώμα μιας μπάλας που υπάρχει ήδη μέσα στο δοχείο. Αυτό αιτιολογεί και τον ισχυρισμό ότι το DP είναι διακριτό μέτρο.

Μία ακόμη σημαντική ιδιότητα της διαδικασίας είναι η συνθήκη της άπειρης ανταλλαξιμότητας. Σύμφωνα με το θεώρημα του de Finetti, για την ακολουθία x_1, \dots, x_n υπάρχει ένα τυχαίο μέτρο Q , τέτοιο ώστε

$$P(x_1, \dots, x_n) = \int_{\mathcal{P}_{\mathbf{x}}} \prod_{i=1}^n \mathbb{P}(x_i) Q(d\mathbb{P}).$$

Το τυχαίο μέτρο Q αποδεικνύεται ότι είναι η διαδικασία Dirichlet (Theodoridis, 2020).

Διαδικασία Chinese Restaurant (CRP)

Η διαδικασία Κινέζικου Εστιατορίου (Chinese Restaurant Process) έχει την ίδια λογική με τη διαδικασία του γενικευμένου Pólya urn. Η ονομασία της CRP προέρχεται από τον παρακάτω συμβολισμό. Έστω ένα εστιατόριο με άπειρο πλήθος τραπέζιων και μία ακολουθία πελατών που θέλει να εξυπηρετηθεί. Αρχικά, μπαίνει ο πρώτος πελάτης ο οποίος κάθεται στο πρώτο τραπέζι. Στη συνέχεια, μπαίνει ο δεύτερος πελάτης ο οποίος έχει την επιλογή να καθίσει και αυτός στο πρώτο τραπέζι που κάθισε και ο προηγούμενος πελάτης με πιθανότητα $\frac{1}{c+1}$ ή να καθίσει σε ένα κενό τραπέζι με πιθανότητα $\frac{c}{c+1}$. Ο n -οστός πελάτης που θα μπει στο εστιατόριο, μπορεί να καθίσει είτε σε κάποιο τραπέζι που κάθονται προηγούμενοι πελάτες με πιθανότητα ανάλογη του αριθμού των προηγούμενων πελατών, είτε σε ένα καινούριο τραπέζι με πιθανότητα ανάλογη του αριθμού c . Στην πραγματικότητα, η ανάθεση των πελατών στα τραπέζια καθορίζει μια τυχαία διαμέριση (Gershman and Blei, 2012).

Μια σημαντική ιδιότητα που ικανοποιείται κατά τη διαδικασία CRP είναι αυτή της ανταλλαξιμότητας, διότι η πιθανότητα για μια συγκεκριμένη ταξινόμηση δεν εξαρτάται από τη σειρά άφιξης των πελατών και παραμένει αμετάβλητη μετά την αναδιάταξή τους (Theodoridis, 2020).

Αναπαράσταση Stick Breaking

Ο Sethuraman το 1994 πρότεινε έναν κατασκευαστικό ορισμό για την αναπαράσταση ενός τυχαίου μέτρου \mathbb{P} από ένα $\mathcal{DP}(c, G_0)$. Έστω

$$v_1, v_2, \dots \stackrel{\text{iid}}{\sim} \mathcal{Be}(1, c)$$

και

$$x_1, x_2, \dots \stackrel{\text{iid}}{\sim} G_0.$$

Τότε το τυχαίο μέτρο \mathbb{P} ορίζεται ως

$$\mathbb{P} = \sum_{i=1}^{\infty} w_i \delta_{x_i} \quad (3.6)$$

όπου δ_{x_i} είναι το μέτρο Dirac και είναι το μέτρο με σημειακή μάζα 1 στο x_i , ενώ $w_1 = v_1$ και για κάθε $i > 1$,

$$w_i = v_i \prod_{k < i} (1 - v_k) \quad (3.7)$$

Τα w_i ονομάζονται stick-breaking βάρη και ισχύει ότι $\sum_{i=1}^{\infty} w_i = 1$.

Επίσης, αποδεικνύεται ότι τα τυχαία μέτρα που παράγονται από ένα DP σύμφωνα με την αναπαράσταση stick-breaking είναι σχεδόν βεβαίως διακριτά (Walker, 2007; Sethuraman, 1994).

Αργότερα, οι Ishwaran και James το 2001 πρότειναν μία επέκταση του παραπάνω κατασκευαστικού ορισμού, όπου τα $v_i \sim \mathcal{Be}(a_i, b_i)$ και απέδειξαν ότι το άθροισμα των βαρών $\sum_{i=1}^{\infty} w_i$ ισούται σχεδόν βεβαίως με τη μονάδα όταν ισχύει

$$\sum_{i=1}^{\infty} \log \left(1 + \frac{a_i}{b_i} \right) = +\infty$$

(Ishwaran and James, 2001).

Στην ενότητα που ακολουθεί, παρουσιάζονται τα μη παραμετρικά Μπεϋζιανά μοντέλα μίξης, καθώς και οι δύο βασικές κατηγορίες MCMC μεθόδων που χρησιμοποιούνται για τη δειγματοληψία από αυτά.

3.3 Μη Παραμετρικά Μπεϋζιανά μοντέλα μίξης

Όπως αναφέρθηκε στις προηγούμενες ενότητες, η διαδικασία Dirichlet παράγει σχεδόν βεβαίως διακριτά μέτρα πιθανοτήτων. Αυτό έχει ως αποτέλεσμα τη δυσκολία εφαρμογής της σε προβλήματα εκτίμησης συνεχών κατανομών. Το πρόβλημα αυτό μπορεί να λυθεί χρησιμοποιώντας ένα τυχαίο μέτρο DP ως μέτρο μίξης μιας μίξης απλών παραμετρικών κατανομών. Η Μπεϋζιανή μη παραμετρική στατιστική αντιμετωπίζει το τυχαίο μέτρο \mathbb{P} ως μία απειροδιάστατη παράμετρο η οποία χρειάζεται μία prior κατανομή. Ωστόσο, από τη στιγμή που η \mathbb{P} είναι κατανομή πιθανότητας, στην πραγματικότητα, η ανάθεση της prior κατανομής σε παραμέτρους, μετατρέπεται σε ανάθεση prior κατανομής σε κατανομές (McAuliffe, 2006; Müller, 2015; Theodoridis, 2020). Η έννοια του μοντέλου μίξης διαδικασίας Dirichlet (Dirichlet Process Mixture Model) εισήχθη από τον Antoniak (1974) και συνεχίστηκε αργότερα από τον Lo (1984) με σκοπό την κατασκευή συνεχών τυχαίων κατανομών (Walker, 2007).

Στα Μπεϋζιανά μοντέλα τα στοιχεία της μίξης θεωρούνται στην ουσία κάποιες τάξεις, οπότε το πρόβλημα μετατρέπεται σε πρόβλημα ταξινόμησης. Στα μη παραμετρικά προβλήματα ο αριθμός των στοιχείων μίξης, δηλαδή των τάξεων, είναι άγνωστος. Για το λόγο αυτό τα Μπεϋζιανά μη παραμετρικά μοντέλα μίξης θεωρούν ότι ο αριθμός των τάξεων είναι δυνητικά άπειρος. Ένας βασικός λόγος που τα μοντέλα αυτά είναι τόσο δημοφιλή, είναι η προσαρμοστικότητα και η ευελιξία τους, καθώς τα δεδομένα του προβλήματος είναι αυτά που καθορίζουν την πολυπλοκότητα του μοντέλου. Με άλλα λόγια, όσο αυξάνονται τα δεδομένα, τόσο αυξάνεται και η πολυπλοκότητα του μοντέλου (Joe and Gooyabadi, 2021; Li et al. 2019).

Για τη συμπερασματολογία των μοντέλων DPM χρησιμοποιούνται συνήθως MCMC μέθοδοι που παρουσιάστηκαν στο δεύτερο κεφάλαιο. Οι μέθοδοι αυτές χωρίζονται σε δύο βασικές κατηγορίες, στις περιθώριες (marginal) και στις υπό συνθήκη (conditional) μεθόδους.

Περιθώριες μέθοδοι

Όσον αφορά τις περιθώριες μεθόδους, η βασική ιδέα τους στηρίζεται στην ολοκλήρωση ως προς το τυχαίο μέτρο \mathbb{P} , με στόχο τη δειγματοληψία από τις πεπεραμένες, πλέον, περιθώριες κατανομές των υπόλοιπων μεταβλητών. Στη συνέχεια, με τη βοήθεια του αλγορίθμου Gibbs πραγματοποιείται η δειγματοληψία από την posterior κατανομή των υπόλοιπων μεταβλητών. Ένα σημαντικό μειονέκτημα που παρουσιάζουν οι περιθώριες μέθοδοι είναι η εξάρτηση που υπάρχει μεταξύ των παραμέτρων.

Υπό συνθήκη μέθοδοι

Σε αντίθεση με τις περιθώριες, οι υπό συνθήκη μέθοδοι διατηρούν το τυχαίο μέτρο \mathbb{P} κατά τη διάρκεια της δειγματοληψίας, η οποία πραγματοποιείται με τη βοήθεια του αλγορίθμου Gibbs. Σημειώνεται ότι σε κάθε επανάληψη του αλγορίθμου, η δειγματοληψία γίνεται από έναν πεπερασμένο αριθμό υποσυνόλων των μεταβλητών. Οι υπό συνθήκη μέθοδοι παρουσιάζουν αρκετά πλεονεκτήματα, έναντι των περιθώριων μεθόδων. Αρχικά, δεν υπάρχει η εξάρτηση μεταξύ των παραμέτρων, ενώ παράλληλα επιτρέπει τη συμπερασματολογία για το τυχαίο μέτρο \mathbb{P} (Papaspiliopoulos and Roberts, 2008).

Έστω μία παραμετρική οικογένεια πυρήνων

$$K_{\vartheta} = \{K(\cdot|\vartheta), \vartheta \in \Theta \subseteq \mathbb{R}^d\} \quad (3.8)$$

με παραμέτρους ϑ . Επιπλέον, έστω $x = (x_1, \dots, x_n)$ ένα δείγμα και $\mathbb{P} \sim \mathcal{DP}(c, G_0)$. Τότε, με χρήση του μοντέλου DPM, η άπειρη μίξη πυρήνων K_{ϑ} με μέτρο μίξης ένα τυχαίο μέτρο \mathbb{P} από ένα DP, δίνεται από τη σχέση

$$f(x|\mathbb{P}) = \int_{\vartheta \in \Theta} K(x|\vartheta)\mathbb{P}(d\vartheta). \quad (3.9)$$

Αν $(\vartheta_1, \dots, \vartheta_n) \in \Theta^n$, τότε η ιεραρχική αναπαράσταση του μοντέλου είναι η εξής

$$\begin{aligned} x_i|\vartheta_i &\stackrel{ind}{\sim} K(\cdot|\vartheta_i) \\ \vartheta_i|\mathbb{P} &\stackrel{iid}{\sim} \mathbb{P} \\ \mathbb{P} &\sim \mathcal{DP}(c, G_0) \end{aligned} \quad (3.10)$$

(Hjort et al. 2010; Fuentes-García et al. 2010).

Στην προηγούμενη ενότητα παρουσιάστηκαν διάφορες αναπαραστάσεις της διαδικασίας Dirichlet. Χρησιμοποιώντας, λοιπόν, την αναπαράσταση stick-breaking το μοντέλο της σχέσης (3.9) μπορεί να γραφεί

$$f(x|w, \vartheta) = \sum_{i=1}^{\infty} w_i K(x|\vartheta_i) \quad (3.11)$$

όπου w_i είναι τα stick-breaking βάρη που κατασκευάζονται σύμφωνα με τη σχέση (3.7) και $\sum_{i=1}^{\infty} w_i = 1$.

Εφαρμόζοντας μία κατηγορία MCMC υπό συνθήκη μεθόδων, προκειμένου η δειγματοληψία να γίνει από πεπερασμένα σύνολα, κάθε παρατήρηση x_i αντιστοιχίζεται με μία μεταβλητή d_i η οποία έχει το ρόλο της δείκτριας μεταβλητής, διότι προσδιορίζει από ποιο στοιχείο της μίξης προήλθε η εκάστοτε παρατήρηση. Η μεταβλητή αυτή δειγματοληπτείται από το σχεδόν βεβαίως πεπερασμένο τυχαίο σύνολο \mathbb{A}_i , το οποίο ονομάζεται slice σύνολο και για το οποίο ισχύει

$$d_i|\mathbb{A}_i \sim \mathcal{DU}(\mathbb{A}_i).$$

Επομένως, $f(d_i = j|\mathbb{A}_i) = \frac{1}{|\mathbb{A}_i|}I(j \in \mathbb{A}_i)$. Ενώ, η πυκνότητα της $x_i|\mathbb{A}_i$ δίνεται από

$$\begin{aligned} f(x_i|\mathbb{A}_i) &= \sum_{j=1}^{\infty} f(x_i, d_i = j|\mathbb{A}_i) \\ &= \sum_{j=1}^{\infty} f(d_i = j|\mathbb{A}_i)f(x_i|d_i = j) \\ &= \sum_{j \in \mathbb{A}_i} \frac{1}{|\mathbb{A}_i|} K(x_i|\vartheta_j). \end{aligned} \quad (3.12)$$

όπου πλέον το άπειρο άθροισμα μετατράπηκε σε πεπερασμένο και σύμφωνα με τη σχέση (3.12) πρόκειται για μίξη κατανομών με ίσα βάρη.

Στην επόμενη ενότητα παρουσιάζεται, αναλυτικά, η δειγματοληψία με slices που πρότεινε ο Walker το 2007, για την δειγματοληψία του μοντέλου DPM.

3.4 Δειγματοληψία του μοντέλου μίξης διαδικασίας Dirichlet με slices

Στην ενότητα αυτή, θα δοθεί αναλυτικά ο αλγόριθμος που πρότεινε ο Walker το 2007 και επεκτάθηκε αργότερα από τους Kalli et al. το 2009. Ο αλγόριθμος αυτός βασίζεται στη δειγματοληψία slice και ανήκει στις υπό συνθήκη μεθόδους που αναφέρθηκαν στην προηγούμενη ενότητα. Η εφαρμογή του αλγορίθμου βασίζεται στην επαύξηση του χώρου καταστάσεων με τις κατάλληλες μεταβλητές, με στόχο η δειγματοληψία σε κάθε επανάληψη του αλγορίθμου να γίνεται από έναν πεπερασμένο αριθμό παραμέτρων.

Έστω, λοιπόν, $x = (x_1, \dots, x_n)$ δείγμα από ένα μη παραμετρικό Μπεϋζιανό μοντέλο μίξης, στο οποίο χρησιμοποιείται ένα τυχαίο μέτρο DP ως prior στο μέτρο μίξης και δίνεται από τη σχέση (3.10). Εφαρμόζοντας την αναπράσταση stick-breaking, η δεσμευμένη κατανομή του x_i είναι η εξής:

$$f(x_i|w, \vartheta) = \sum_{j=1}^{\infty} w_j K(x_i|\vartheta_j) \quad (3.13)$$

με $w = (w_1, w_2, \dots)$ το απειροδιάστατο διάνυσμα των stick-breaking βαρών και $\vartheta = (\vartheta_1, \vartheta_2, \dots)$ το απειροδιάστατο διάνυσμα των παραμέτρων.

Εισάγοντας την τυχαία βοηθητική μεταβλητή u_i με $i = 1, \dots, n$, η από κοινού πυκνότητα των (x_i, u_i) γράφεται ως εξής:

$$f(x_i, u_i|w, \vartheta) = \sum_{j=1}^{\infty} I(w_j > u_i) K(x_i|\vartheta_j) = \sum_{j \in \mathbb{A}_j} K(x_i|\vartheta_j) \quad (3.14)$$

όπου

$$\mathbb{A}_i = \{j \in \mathbb{N} : 0 < u_i < w_j\} \quad (3.15)$$

είναι τα λεγόμενα slice σύνολα τα οποία είναι πεπεραμένα για κάθε $u_i > 0$.

Στη συνέχεια, εισάγοντας τη δείκτρια μεταβλητή d_i , $i = 1, \dots, n$, η σχέση (3.14) γράφεται ως εξής

$$f(x_i, u_i | w, \vartheta) = \sum_{j=1}^{\infty} f(d_i = j) f(x_i, u_i | d_i = j, w, \vartheta) = \sum_{j=1}^{\infty} w_j \mathcal{U}(u_i | 0, w_j) K(x_i | \vartheta_j). \quad (3.16)$$

Προφανώς, η ολοκλήρωση της σχέσης (3.16) ως προς u_i οδηγεί στη σχέση (3.13).

Η από κοινού πυκνότητα των (x_i, u_i, d_i) δίνεται από την παρακάτω σχέση:

$$\begin{aligned} f(x_i, u_i, d_i) &= f(d_i = j) f(u_i | d_i = j) f(x_i | d_i = j) \\ &= w_j \mathcal{U}(u_i | 0, w_j) K(x_i | \vartheta_j) \\ &= I(u_i < w_j) K(x_i | \vartheta_j). \end{aligned}$$

Επιπλέον, η πιθανοφάνεια του μοντέλου είναι

$$l(x, u, d | w, \vartheta) = \prod_{i=1}^n I(u_i < w_{d_i}) K(x_i | \vartheta_{d_i}), \quad (3.17)$$

ενώ η posterior κατανομή είναι

$$f(w, \vartheta | x, u, d) \propto f(w, \vartheta) \prod_{i=1}^n I(u_i < w_{d_i}) K(x_i | \vartheta_{d_i}). \quad (3.18)$$

Από τη σχέση (3.18) είναι φανερό ότι η δειγματοληψία των παραμέτρων του μοντέλου πραγματοποιείται από ένα πεπερασμένο πλέον σύνολο.

Παρακάτω δίνεται αναλυτικά ο αλγόριθμος του Walker (Walker, 2007; Kalli et al. 2009). Πριν τη δειγματοληψία από τις δεσμευμένες κατανομές των παραμέτρων για $j = 1, \dots, \max_i \{d_i\}$ χρειάζεται η αρχικοποίηση του n -διάστατου διανύσματος των μεταβλητών d_i . Στη συνέχεια ακολουθούν τα βήματα που πρέπει να πραγματοποιηθούν σε κάθε επανάληψη του αλγορίθμου.

1. Πρώτο βήμα είναι η δημιουργία των v_j η οποία θα γίνει ξεχωριστά από τη δημιουργία των u_i . Αυτή είναι και η βασική διαφοροποίηση των δύο αλγορίθμων. Η δεσμευσμένη κατανομή για τα v_j είναι

$$f(v_j | \dots \text{εκτός των } u_i) = \mathcal{B}e \left(1 + \sum_{i=1}^n I(d_i = j), c + \sum_{i=1}^n I(d_i > j) \right), \quad (3.19)$$

όπου ακολουθεί και η κατασκευή των stick-breaking βαρών w_j σύμφωνα με τη σχέση (3.7).

2. Επόμενο βήμα είναι η δημιουργία των u_i σύμφωνα με τη δεσμευμένη κατανομή

$$f(u_i|\cdot) = \mathcal{U}(u_i|0, w_{d_i}). \quad (3.20)$$

3. Στη συνέχεια ακολουθεί η δειγματοληψία των παραμέτρων των στοιχείων μίξης από τις εξής δεσμευμένες κατανομές

$$f(\vartheta_j|\cdot) \propto g_0(\vartheta_j) \prod_{d_i=j} K(x_i|\vartheta_j), \quad (3.21)$$

όπου g_0 η συνάρτηση πυκνότητας που αντιστοιχεί στο μέτρο βάσης G_0 . Στην περίπτωση που δεν υπάρχουν $d_i = j$, τότε η δειγματοληψία γίνεται από την prior G_0 .

4. Για τη δειγματοληψία των d_i χρειάζεται η κατασκευή των $\mathbb{A}_i = \{j \in \mathbb{N} : 0 < u_i < w_j\}$. Για να γίνει αυτό, πρέπει να υπάρχει ο κατάλληλος αριθμός των w_j , ώστε να υπάρχουν όλα τα $u_i < w_j$. Είναι βέβαιο ότι δεν υπάρχει κάποιο $k > k^i$ για το οποίο ισχύει $u_i < w_k$ όταν

$$\sum_{j=1}^{k^i} w_j > 1 - u_i.$$

Επομένως, για όλα τα i υπολογίζεται το μικρότερο k^* ώστε

$$\sum_{j=1}^{k^*} w_j > 1 - u^* \quad (3.22)$$

όπου $u^* = \min\{u_1, \dots, u_n\}$. Αυτό σημαίνει ότι τα βάρη που χρειάζονται για τη δειγματοληψία των d_i είναι w_1, \dots, w_{k^*} . Σε περίπτωση που χρειαστούν περισσότερα βάρη και παράμετροι τότε παράγονται από τις αντίστοιχες prior, δηλαδή τα βάρη από την $\mathcal{Be}(1, c)$ και οι υπόλοιπες παράμετροι από την G_0 . Έπειτα, η δειγματοληψία των d_i πραγματοποιείται από την διακριτή δεσμευμένη κατανομή

$$f(d_i|\cdot) \propto K(x_i|\vartheta_{d_i})I(u_i < w_{d_i}). \quad (3.23)$$

5. Επόμενο βήμα είναι η δειγματοληψία για την παράμετρο c της διαδικασίας Dirichlet σύμφωνα με τον West (1992). Τα βήματα είναι τα εξής:

i. Παραγωγή $s \sim \mathcal{Be}(c + 1, n)$

ii. Παραγωγή $c|s, \kappa \sim p_c \mathcal{G}(\alpha + \kappa, \beta - \log s) + (1 - p_c) \mathcal{G}(\alpha + \kappa - 1, \beta - \log s)$

όπου κ είναι το πλήθος των μοναδικών τιμών της μεταβλητής d_i . Δηλαδή, η παράμετρος c προέρχεται από μία μίξη Γάμμα κατανομών με $p_c = \frac{\alpha + \kappa - 1}{n(\beta - \log s) + \alpha + \kappa - 1}$.

6. Τέλος, για την εκτίμηση πυκνότητας χρειάζεται ο υπολογισμός της κατανομής predictive

$$x_{n+1}|x_1, \dots, x_n \sim \sum_{j=1}^{\infty} w_j^* K(\cdot|\vartheta_j^*). \quad (3.24)$$

Για να γίνει αυτό, χρησιμοποιείται κάποιο βάρος w_k^* με στόχο την επιλογή των αντίστοιχων παραμέτρων ϑ_k^* οι οποίες προσδιορίζουν από ποιο στοιχείο της μίξης $K(\cdot|\vartheta_k^*)$ θα γίνει η δειγματοληψία για την παρατήρηση x_{n+1} . Για τον προσδιορισμό του k τα βήματα είναι τα εξής:

i. Παραγωγή αριθμού $r \sim \mathcal{U}(0, 1)$

ii. Υπολογισμός k ώστε να ισχύει $\sum_{j=1}^{k-1} w_j < r < \sum_{j=1}^k w_j$ με $w_0 = 0$.

Στην περίπτωση που δεν ικανοποιείται η παραπάνω σχέση για κάποιο από τα υπάρχοντα βάρη, η επιλογή των παραμέτρων ϑ_{n+1}^* γίνεται από την prior G_0 .

3.5 Προσομείωση

Στην ενότητα αυτή, θα χρησιμοποιηθεί ο αλγόριθμος της προηγούμενης ενότητας με στόχο την εκτίμηση πυκνότητας. Συγκεκριμένα, ως πυρήνας μίξης χρησιμοποιείται η κανονική κατανομή $N(\cdot|\vartheta)$ όπου $\vartheta = (\mu, \tau^{-1})$.

Έστω λοιπόν, $x = (x_1, \dots, x_{50})$ παρατηρήσεις από τη μίξη κανονικών κατανομών

$$f(x) = \frac{1}{3}N(x|-4, 1) + \frac{1}{3}N(x|0, 1) + \frac{1}{3}N(x|8, 1). \quad (3.25)$$

Οι prior του μοντέλου είναι οι εξής

$$\begin{aligned} \mu_j &\sim N\left(\mu_0, \frac{1}{\tau_0}\right), j \geq 1 \\ \tau_j &\sim \mathcal{G}(\alpha, \beta), j \geq 1 \\ c &\sim \mathcal{G}(\alpha_c, \beta_c) \end{aligned}$$

όπου $\mu_0 = 0, \tau_0 = 0.1$ και $\alpha = \beta = \alpha_c = \beta_c = 0.1$.

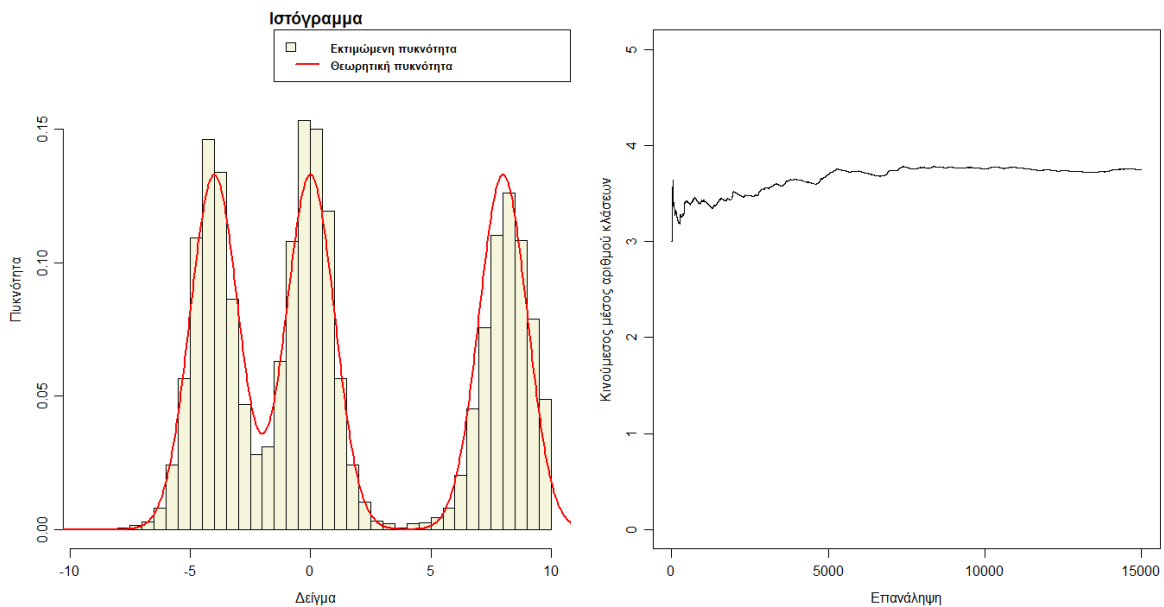
Οι δεσμευμένες κατανομές για τις παραμέτρους μ_j και τ_j είναι

$$f(\mu_j|\cdot) = N\left(\frac{\tau_0\mu_0 + \tau_j \sum_{d_i=j} x_i}{n^*\tau_j + \tau_0}, \frac{1}{n^*\tau_j + \tau_0}\right) \quad (3.26)$$

$$f(\tau_j|\cdot) = \mathcal{G}\left(\alpha + \frac{n^*}{2}, \beta + \frac{\sum_{d_i=j} (x_i - \mu_j)^2}{2}\right) \quad (3.27)$$

όπου $n^* = \sum_{i=1}^n I(d_i = j)$.

Στο Σχήμα 3.2 δίνονται δύο διαγράμματα. Το πρώτο διάγραμμα είναι το ιστόγραμμα του δείγματος της predictive κατανομής x_{n+1} που παράχθηκε για την εκτίμηση της πυκνότητας των 50 παρατηρήσεων, ενώ με κόκκινο χρώμα εφαρμόζεται η καμπύλη της θεωρητικής πυκνότητας των 50 παρατηρήσεων που προέρχονται από τη μίξη κατανομών της σχέσης (3.25). Το δεύτερο διάγραμμα είναι το διάγραμμα κινούμενου μέσου των αριθμών των στοιχείων μίξης που χρησιμοποιήθηκαν σε κάθε επανάληψη του αλγορίθμου. Σημειώνεται ότι ο αλγόριθμος έτρεξε για 20,000 επαναλήψεις και το δείγμα προέκυψε μετά από περίοδο burn-in 5000 επαναλήψεων. Είναι φανερό, ότι η μαρκοβιανή αλυσίδα συγκλίνει στη στάσιμη κατανομή και έτσι το δείγμα που παράχθηκε φαίνεται να προέρχεται από την predictive κατανομή.



Σχήμα 3.2: Ιστόγραμμα του δείγματος για την εκτίμηση της πυκνότητας της μίξης (3.25) μαζί με την καμπύλη της θεωρητικής πυκνότητας (αριστερά) και διάγραμμα κινούμενου μέσου για τον αριθμό των κλάσεων (δεξιά).

Κεφάλαιο 4

Εφαρμογή σε στοχαστικά δυναμικά συστήματα

4.1 Δυναμικά Συστήματα

Δυναμικά συστήματα ονομάζονται τα συστήματα τα οποία εξελίσσονται στο χρόνο. Η εφαρμογή τους βασίζεται κυρίως στην κατανόηση και την περιγραφή των συστημάτων που εξελίσσονται, καθώς και στην πρόβλεψη της μελλοντικής συμπεριφοράς τους. Κάποια παραδείγματα δυναμικών συστημάτων είναι η κίνηση των ουράνιων σωμάτων, ένα εκκρεμές που ταλαντώνεται, η ανάπτυξη ενός πληθυσμού κ.ά. (Hasselblatt and Katok, 2003). Τα δυναμικά συστήματα χωρίζονται σε δύο κατηγορίες, στα ντετερμινιστικά και στα στοχαστικά δυναμικά συστήματα. Στην πρώτη περίπτωση, τα ντετερμινιστικά συστήματα προσδιορίζουν την παρούσα κατάσταση του συστήματος μοναδικά, ως συνάρτηση των προηγούμενων καταστάσεων. Σε αντίθεση με τη δεύτερη περίπτωση, όπου στα στοχαστικά δυναμικά συστήματα εισάγεται η έννοια της τυχαιότητας για την περιγραφή της παρούσας κατάστασης (Alligood et al. 1996).

Στη συνέχεια, δίνονται, αναλυτικότερα, σημαντικοί ορισμοί και έννοιες τόσο για τα ντετερμινιστικά όσο και για τα στοχαστικά δυναμικά συστήματα.

4.1.1 Ντετερμινιστικά Δυναμικά Συστήματα

Ορισμός 4.1.1 (Ντετερμινιστικό Δυναμικό Σύστημα, Brin and Stuck 2002). Ένα ντετερμινιστικό δυναμικό σύστημα είναι μία τριάδα (X, T, g) , όπου X είναι ο χώρος καταστάσεων, T είναι ένα χρονικό σύνολο και $g : X \times T \rightarrow X$ είναι μια απεικόνιση για την οποία ισχύουν:

1. $g_0 = 1_X$
2. $g_{n+m} = g_n \circ g_m, \forall n, m \in T$

με $g_n := g(\cdot, n)$ και $\mathbf{1}_X$ την ταυτοτική συνάρτηση.

Στην περίπτωση που $T = \mathbb{Z}$, τότε το δυναμικό σύστημα είναι διακριτού χρόνου και η g ονομάζεται απεικόνιση. Τότε, το δυναμικό σύστημα μπορεί να αναπαρασταθεί ως

$$x_i = g(\vartheta, x_{i-1}, x_{i-2}, \dots, x_{i-d}), \quad i \geq 1 \quad (4.1)$$

όπου $(x_0, x_{-1}, \dots, x_{-d+1}) \in \mathbf{X}^d$ είναι οι αρχικές συνθήκες και η απεικόνιση g είναι παραμετροποιημένη ως προς κάποιες παραμέτρους ελέγχου $\vartheta \in \Theta$, με παραμετρικό χώρο Θ . Στο παρόν κεφάλαιο, για λόγους απλότητας, κάθε κατάσταση x_i εξαρτάται μόνο από την προηγούμενη κατάσταση x_{i-1} , επομένως το ντετερμινιστικό δυναμικό σύστημα διακριτού χρόνου δίνεται από τη σχέση

$$x_i = g(\vartheta, x_{i-1}), \quad i \geq 1 \quad (4.2)$$

Στην περίπτωση που $T = \mathbb{R}$ τότε το δυναμικό σύστημα είναι συνεχούς χρόνου και η απεικόνιση g ονομάζεται ροή. Τότε, το δυναμικό σύστημα μπορεί να αναπαρασταθεί με τη βοήθεια διαφορικών εξισώσεων

$$\dot{x}_i(t) = g_i(t, x_1, x_2, \dots, x_n), \quad i = 1, \dots, n \quad (4.3)$$

(Brin and Stuck, 2002; Merkatas, 2018; Afanas'ev et al. 1996). Αξίζει να σημειωθεί ότι ένα δυναμικό σύστημα συνεχούς χρόνου διάστασης N μπορεί να μετατραπεί σε σύστημα διακριτού χρόνου διάστασης $N-1$ με τη μέθοδο τομής της επιφάνειας Poincaré (Ott, 2002).

Ένας σημαντικός ορισμός που χρειάζεται να αναφερθεί είναι αυτός της τροχιάς.

Ορισμός 4.1.2 (Τροχιά, Alligood et al. 1996). Έστω x ένα σημείο και g μία απεικόνιση. Η τροχιά του x υπό την g ορίζεται να είναι το σύνολο των σημείων

$$\mathcal{O}_g(x) = \{x, g(x), g^2(x), \dots\} = \{g^n(x) : n \geq 0\}. \quad (4.4)$$

όπου $g^n := \underbrace{g \circ g \circ \dots \circ g}_{n\text{-φορές}}$. Το αρχικό σημείο x της τροχιάς ονομάζεται αρχική συνθήκη.

4.1.2 Κατηγορίες Τυχαίου Θορύβου

Σε πολλές περιπτώσεις που υπάρχει αβεβαιότητα τα δυναμικά συστήματα μπορούν να μοντελοποιηθούν σαν να εξελίσσονται με την παρουσία τυχαίων διαταραχών. Οι διαταραχές αυτές κατατάσσονται σε δύο τυχαίες κατηγορίες θορύβου, οι οποίες έχουν διαφορετικές επιδράσεις στα συστήματα.

Η πρώτη κατηγορία αφορά το θόρυβο μέτρησης ή αλλιώς το θόρυβο παρατήρησης (measurement/observational noise). Ο θόρυβος παρατήρησης προκύπτει από σφάλματα

που μπορεί να συμβούν κατά τη διαδικασία της μέτρησης και θεωρείται ανεξάρτητος από τη δυναμική του συστήματος. Σημειώνεται ότι εισάγεται μετά τη χρονική εξέλιξη των εξεταζόμενων τροχιών. Στην ουσία, ο θόρυβος προκύπτει σε εργαστηριακά ή πραγματικά δεδομένα χρονοσειρών λόγω ανακρίβειών στις μετρήσεις. Επίσης είναι σημαντικό να ανφερθεί ότι, ενώ ο θόρυβος παρατήρησης επηρεάζει τα παρατηρούμενα δεδομένα, δεν επηρεάζει εγγενώς τη μελλοντική εξέλιξη του συστήματος.

Η δεύτερη κατηγορία αφορά το δυναμικό θόρυβο (dynamical noise), ο οποίος μπορεί να είναι είτε προσθετικός είτε πολλαπλασιαστικός. Ο δυναμικός θόρυβος, που συχνά χρησιμοποιείται ως σφάλμα του μοντέλου, αποτελεί αναπόσπαστο κομμάτι σε ορισμένα μοντέλα διότι επηρεάζει σημαντικά τη μελλοντική εξέλιξη του δυναμικού συστήματος. Σε αντίθεση με τον θόρυβο παρατήρησης, ο δυναμικός θόρυβος δεν εισάγεται μετά τη χρονική εξέλιξη αλλά ενσωματώνεται στο ίδιο το μοντέλο, καθώς προστίθεται σε κάθε βήμα της χρονικής εξέλιξης των τροχιών. Με την ενσωμάτωση του δυναμικού θορύβου στα μοντέλα, οι ερευνητές στοχεύουν στην καλύτερη αποτύπωση της πολυπλοκότητας των συστημάτων του πραγματικού κόσμου και στη βελτίωση της ακρίβειας των προβλέψεων, ιδίως όταν πρόκειται για συστήματα που παρουσιάζουν εγγενή στοχαστική συμπεριφορά (Strumik, and Macek, 2008; Merkatas, 2018; Kaloudis, 2019).

Παρακάτω δίνονται ενδεικτικά κάποια παραδείγματα δυναμικών συστημάτων στα οποία ενσωματώνονται τα είδη θορύβων που μόλις αναφέρθηκαν.

Παράδειγμα 4.1.1. (Δυναμικό σύστημα με προσθετικό θόρυβο παρατήρησης). Δίνεται το μοντέλο

$$\begin{aligned} y_i &= g(\vartheta, y_{i-1}) \\ x_i &= y_i + e_i, \quad 1 \leq i \leq n \end{aligned} \quad (4.5)$$

όπου $e_i \stackrel{iid}{\sim} \pi(\cdot)$. Αυτό είναι ένα παράδειγμα δυναμικού συστήματος με προσθετικό θόρυβο παρατήρησης. Σε αυτή την περίπτωση, ο θόρυβος προστίθεται μετά τη δημιουργία της τροχιάς.

Παράδειγμα 4.1.2. (Δυναμικό σύστημα με πολλαπλασιαστικό δυναμικό θόρυβο). Δίνεται το μοντέλο

$$x_i = \epsilon_i g(\vartheta, x_{i-1}) \quad (4.6)$$

όπου $\epsilon_i \stackrel{iid}{\sim} \pi(\cdot)$. Συνήθως, ο πολλαπλασιαστικός δυναμικός θόρυβος εφαρμόζεται απευθείας στις παραμέτρους έλεγχου ϑ του συστήματος. Ενώ, ο προσθετικός δυναμικός θόρυβος προστίθεται σε κάθε επανάληψη της εξέλιξης της τροχιάς του συστήματος.

Παράδειγμα 4.1.3. (Δυναμικό σύστημα με προσθετικό θόρυβο παρατήρησης και δυναμικό θόρυβο). Τα μοντέλα που συνδυάζουν δυναμικό θόρυβο και θόρυβο παρατήρησης ονομάζονται state-space μοντέλα. Η χρήση των μοντέλων αυτών είναι ιδιαίτερα δημοφιλής στην περίπτωση προβλημάτων ανάλυσης δυναμικών συστημάτων. Ένα παράδειγμα

state-space μοντέλου είναι το ακόλουθο :

$$\begin{aligned} y_i &= g(\varphi, y_{i-1}) + e_i, \quad \text{όπου } e_i \stackrel{iid}{\sim} \pi(\cdot) \\ x_i &= f(\vartheta, y_i) + \epsilon_i, \quad \text{όπου } \epsilon_i \stackrel{iid}{\sim} q(\cdot) \end{aligned} \quad (4.7)$$

4.1.3 Η έννοια του χάους στα δυναμικά συστήματα

Τα αμιγώς ντετερμινιστικά δυναμικά συστήματα μπορεί να παρουσιάσουν ιδιαίτερα πολύπλοκη συμπεριφορά, η οποία καλείται χαοτική, και αυτό μπορεί να οφείλεται στην μη γραμμικότητα της απεικόνισης g (Gleick, 1987). Με άλλα λόγια, η μη γραμμικότητα αποτελεί αναγκαία αλλά όχι και ικανή συνθήκη για την εμφάνιση του χάους. Στην ενότητα αυτή θα δοθούν σημαντικοί ορισμοί και έννοιες σύμφωνα με τους Alligood et al. (1996), με στόχο την κατανόηση του ορισμού του χάους στα δυναμικά συστήματα.

Ορισμός 4.1.3. Έστω $g : X \rightarrow X$ μία απεικόνιση. Ένα σύνολο $S \subseteq X$ ονομάζεται αναλλοίωτο σύνολο υπό την g όταν $g(S) \subseteq S$.

Όταν η τροχιά εισέλθει σε ένα αναλλοίωτο σύνολο, τότε θα παραμείνει σε αυτό για πάντα.

Ορισμός 4.1.4. (Περιοδική τροχιά). Έστω $g : X \rightarrow X$ μία απεικόνιση. Ένα σημείο p ονομάζεται περιοδικό σημείο με περίοδο m όταν $g^m(p) = p$, όπου m είναι ο μικρότερος δυνατός θετικός ακέραιος. Η τροχιά με αρχικό σημείο το σημείο p που αποτελείται από m σημεία ονομάζεται περιοδική τροχιά με περίοδο m . Στην περίπτωση όπου $m = 1$, τότε το σημείο p καλείται σταθερό σημείο.

Ένα σταθερό σημείο p ενός δυναμικού συστήματος έχει την ικανότητα είτε να έλκει γειτονικά του σημεία είτε να τα απωθεί. Στην περίπτωση που το σταθερό σημείο p έλκει τα κοντινά του σημεία τότε ονομάζεται ελκτικό σταθερό σημείο, ενώ στην περίπτωση που τα απωθεί ονομάζεται απωστικό σταθερό σημείο.

Θεώρημα 4.1.1. (Ευστάθεια περιοδικών τροχιών). Έστω $g : X \rightarrow X$ μία απεικόνιση και $\{x_1, \dots, x_m\}$ μία m -περιοδική τροχιά. Τότε

1. Αν $|(g^m)'(x_1)| < 1$, η τροχιά είναι ελκτική.
2. Αν $|(g^m)'(x_1)| > 1$, η τροχιά είναι απωστική.

Ορισμός 4.1.5. (Τοπολογική μεταβατικότητα). Μία απεικόνιση $g : X \rightarrow X$ είναι τοπολογικά μεταβατική εάν για κάθε ζευγάρι ανοιχτών συνόλων $A, B \subseteq X$, υπάρχει $m > 0$ τέτοιο ώστε

$$g^m(A) \cap B \neq \emptyset.$$

Η τοπολογική μεταβατικότητα εξασφαλίζει την μετάβαση της τροχιάς που ξεκινάει από το $x_0 \in A$ σε κάποιο τυχαίο σύνολο γειτονικών σημείων B του x_0 . Επομένως, η τροχιά δεν θα απορροφηθεί από κάποιο $A \subseteq X$ (Devaney, 1989).

Ορισμός 4.1.6. (Ευαίσθητη εξάρτηση από τις αρχικές συνθήκες). Έστω $g : X \rightarrow X$ μία απεικόνιση. Ένα σημείο x_0 παρουσιάζει ευαίσθητη εξάρτηση από τις αρχικές συνθήκες όταν υπάρχει $d > 0$ τέτοιο ώστε για οποιοδήποτε σύνολο γειτονικών σημείων N του x_0 υπάρχει σημείο x ώστε

$$|g^m(x) - g^m(x_0)| \geq d$$

για κάποιο θετικό ακέραιο m . Το σημείο x_0 ονομάζεται ευαίσθητο σημείο.

Όταν υπάρχει ευαίσθητη εξάρτηση από τις αρχικές συνθήκες, δύο τροχιές του ίδιου συστήματος στις οποίες οι αρχικές συνθήκες διαφέρουν ελάχιστα, θα αποκλίνουν με εκθετικό ρυθμό. Αυτό μπορεί να έχει ως συνέπεια τη μεγέθυνση των σφαλμάτων στρογγυλοποίησης. Επίσης, ένα χαρακτηριστικό των χαοτικών συστημάτων είναι η μη προβλεψιμότητα, η οποία οφείλεται στην ευαίσθητη εξάρτηση από τις αρχικές συνθήκες. Ωστόσο, υπάρχει η δυνατότητα ποσοτικοποίησης της ευαίσθητης εξάρτησης από τις αρχικές συνθήκες με τη βοήθεια του εκθέτη Lyapunov, ο οποίος ορίζεται ως

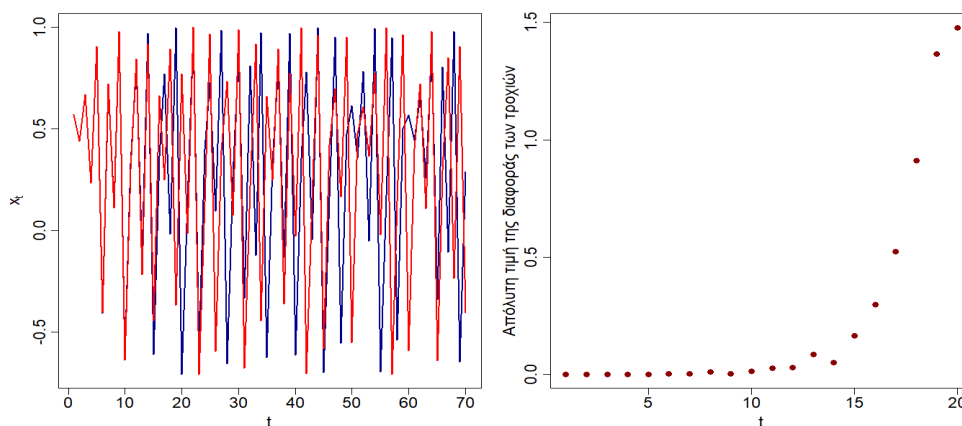
$$h(x_1) = \lim_{n \rightarrow \infty} \frac{1}{n} [\log |g'(x_1)| + \dots + \log |g'(x_n)|].$$

Για την καλύτερη κατανόηση της ευαίσθητης εξάρτησης από τις αρχικές συνθήκες παρατίθενται οι τροχιές του ίδιου δυναμικού συστήματος αλλάζοντας ελάχιστα την αρχική συνθήκη.

Έστω η λογιστική απεικόνιση¹

$$x_i = 1 - \vartheta x_{i-1}^2, \quad i = 1, \dots, 70 \quad (4.8)$$

με $\vartheta = 1.71$ και αρχικές συνθήκες $x_0 = 0.5$ και $x'_0 = 0.5001$.



Σχήμα 4.1: Τροχιές με αρχικές συνθήκες $x_0 = 0.5$ (μπλε) και $x'_0 = 0.5001$ (κόκκινο) (αριστερά). Διάγραμμα της απόλυτης τιμής της διαφοράς των δύο τροχιών (δεξιά).

¹Χρησιμοποιείται η αναπαράσταση $x = 1 - \vartheta x^2$, $\vartheta \in [0, 2]$ έναντι της $x = T(\mu, x) = \mu x(1 - x)$ οι οποίες είναι ισοδύναμες. Αυτό συμβαίνει λόγω του ότι υπάρχει απεικόνιση ϕ , τέτοια ώστε $g(\vartheta, \phi(x)) = \phi(T(\mu, x))$, όπου $\phi(x) = \frac{4x-2}{m-2}$ και $\vartheta = \frac{\mu(\mu-2)}{4}$.

Στο Σχήμα 4.1, απεικονίζονται δύο διαγράμματα. Στο πρώτο διάγραμμα με μπλε χρώμα αποτυπώνεται η τροχιά με αρχική συνθήκη $x_0 = 0.5$, ενώ με κόκκινο χρώμα παρουσιάζεται η τροχιά με αρχική συνθήκη $x'_0 = 0.5001$. Φαίνεται ότι οι δύο τροχιές διαφέρουν σημαντικά, παρόλο που η διαφορά στην αρχική συνθήκη είναι πολύ μικρή. Παράλληλα, στο δεύτερο διάγραμμα απεικονίζεται η απόλυτη τιμή της διαφοράς των δύο τροχιών. Φαίνεται ότι η καμπύλη αυξάνεται εκθετικά όσο μεγαλώνει ο αριθμός των επαναλήψεων.

Τώρα μπορεί να δοθεί ο ορισμός ενός χαοτικού δυναμικού συστήματος. Σημειώνεται ότι ο παρακάτω ορισμός δίνεται σύμφωνα με τον Devaney (1989).

Ορισμός 4.1.7. Έστω (X, T, g) ένα δυναμικό σύστημα. Τότε αυτό είναι χαοτικό εάν ισχύουν τα εξής:

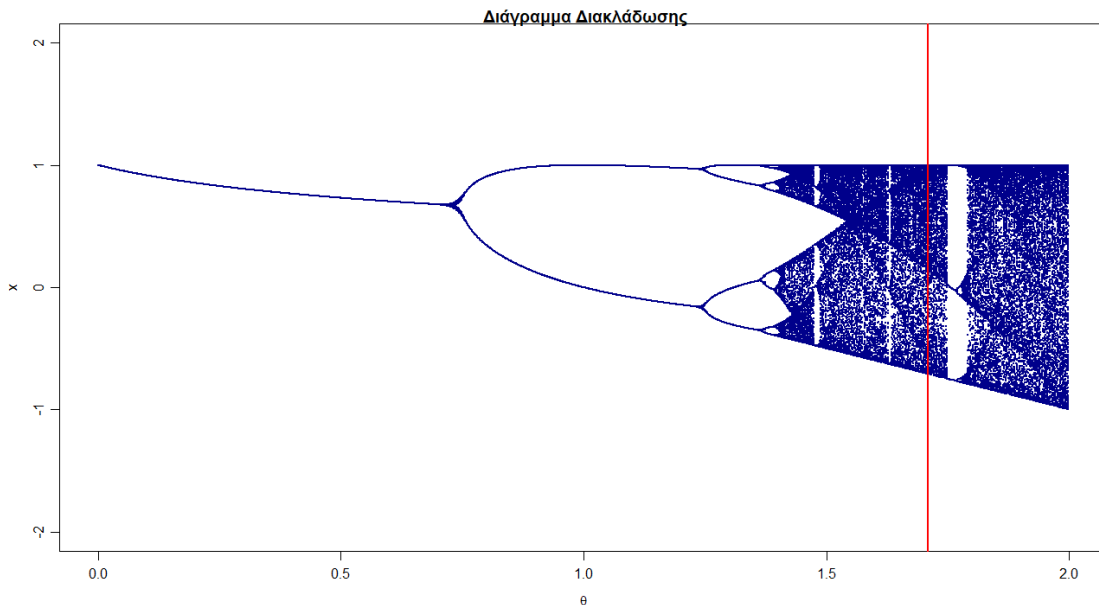
1. Η g είναι τοπολογικά μεταβατική,
2. Τα περιοδικά σημεία είναι πυκνά στον X ,
3. Η g παρουσιάζει ευαίσθητη εξάρτηση από τις αρχικές συνθήκες.

Σημειώνεται ότι οι δύο πρώτες συνθήκες του ορισμού συνδέονται με τοπολογικά χαρακτηριστικά, ενώ η τρίτη σύνθηκη εξαρτάται από κάποια μετρική. Σύμφωνα με τους Banks et al. η ευαίσθητη εξάρτηση από τις αρχικές συνθήκες είναι περιττή στον ορισμό του χάους, διότι αν ισχύει η τοπολογική μεταβατικότητα και η πυκνότητα των περιοδικών σημείων τότε συνεπάγεται και η ευαισθησία από τις αρχικές συνθήκες.

Ένας τρόπος για να αποτυπωθούν γραφικά οι διαφορετικές δυναμικές συμπεριφορές ενός χαοτικού δυναμικού συστήματος είναι με τη βοήθεια των διαγραμμάτων διακλάδωσης. Τα διαγράμματα αυτά περιγράφουν τη γέννηση, την εξέλιξη και την εξαφάνιση ελκτικών συνόλων. Με άλλα λόγια, απεικονίζουν τις ποιοτικές αλλαγές στη συμπεριφορά ενός δυναμικού συστήματος καθώς μεταβάλλεται η παράμετρος ελέγχου θ . Για την κατασκευή ενός διαγράμματος διακλάδωσης επιλέγεται η αρχική συνθήκη $x_0 \in X$ και υπολογίζεται η τροχιά του x_0 για έναν μεγάλο αριθμό επαναλήψεων, σύμφωνα με την απεικόνιση g . Η διαδικασία επαναλαμβάνεται αλλάζοντας κάθε φορά την τιμή της παραμέτρου ελέγχου θ . Σημειώνεται ότι δε λαμβάνονται υπόψιν όλες οι τιμές για τη δημιουργία του διαγράμματος, παρά μόνο ένα τελικό κομμάτι του δείγματος (Nusse et al. 1994). Στο Σχήμα 4.2 δίνεται ένα παράδειγμα διαγράμματος διακλάδωσης με αρχική συνθήκη $x_0 = 0.5$, σύμφωνα με τη λογιστική απεικόνιση που δόθηκε στη σχέση (4.8) και μεταβάλλοντας τις τιμές της παραμέτρου ελέγχου μεταξύ του διαστήματος $[0, 2]$. Επιπλέον, με την κάθετη κόκκινη γραμμή δίνεται η τιμή της παραμέτρου ελέγχου $\theta = 1.71$, που παρουσιάστηκε στα προηγούμενα παραδείγματα και για την οποία φαίνεται ότι το δυναμικό σύστημα παρουσιάζει χαοτική συμπεριφορά.

Όταν η τιμή του θ κυμαίνεται μεταξύ 0 και 0.75 οι τροχιές έλκονται από ένα σημείο, αμέσως μετά προκύπτουν τροχιές περιόδου 2 και στη συνέχεια τροχιές περιόδου 4. Αυτή η συμπεριφορά καλείται διαδρομή διπλασιασμού περιόδου του χάους. Σύμφωνα με το Σχήμα 4.2, για τιμές του θ μεγαλύτερες του 1.35 οι τροχιές παρουσιάζουν απεριοδι-

κότητα, δηλαδή χαοτική συμπεριφορά, ωστόσο διακρίνονται κάποιες “λευκές” κάθετες λωρίδες στις οποίες παρατηρείται περιοδικότητα. Όσο περισσότερο μεγενθύνουμε το διάγραμμα τόσο περισσότερες “λευκές” κάθετες λωρίδες εντοπίζονται.



Σχήμα 4.2: Διάγραμμα διακλάδωσης της λογιστικής απεικόνισης με αρχική συνθήκη $x_0 = 0.5$. Με κόκκινο χρώμα δίνεται η τιμή της παραμέτρου ελέγχου $\vartheta = 1.71$ που χρησιμοποιήθηκε στα προηγούμενα παραδείγματα.

4.1.4 Στοχαστικά Δυναμικά Συστήματα

Στις προηγούμενες ενότητες του κεφαλαίου παρουσιάστηκαν τα ντετερμινιστικά δυναμικά συστήματα, ωστόσο στην ενότητα 4.1.2 αναφέρθηκαν οι κατηγορίες τυχαίου θορύβου και κάποια παραδείγματα δυναμικών συστημάτων τα οποία διαταράσσονται από κάποια πηγή τυχαίου θορύβου. Στην παρούσα ενότητα, μελετώνται δυναμικά συστήματα της μορφής

$$x_i = g(\vartheta, x_{i-1}) + e_i, \quad i \geq 1 \quad (4.9)$$

όπου e_i είναι μία εργοδική διαδικασία και κάθε κατάσταση της τροχιάς x_i εξαρτάται μόνο από την προηγούμενη κατάσταση x_{i-1} .

Για τον ορισμό ενός στοχαστικού δυναμικού συστήματος είναι απαραίτητη η αναφορά κάποιων ορισμών και εννοιών, που θα δοθούν σύμφωνα με τον Klenke (2013), για την κατανόηση του. Για τους ορισμούς που δίνονται παρακάτω θεωρείται ότι $(\Omega, \mathcal{F}, \mathbb{P})$ είναι ένας χώρος πιθανότητας και $\varphi : \Omega \rightarrow \Omega$ μία μετρήσιμη απεικόνιση.

Ορισμός 4.1.8 (Μετρήσιμη απεικόνιση, Folland 1999). Έστω (X, \mathcal{N}, μ) ένας μετρήσι-

μος χώρος². Μία απεικόνιση $g : \mathbf{X} \rightarrow \mathbf{X}$ ονομάζεται μετρήσιμη όταν

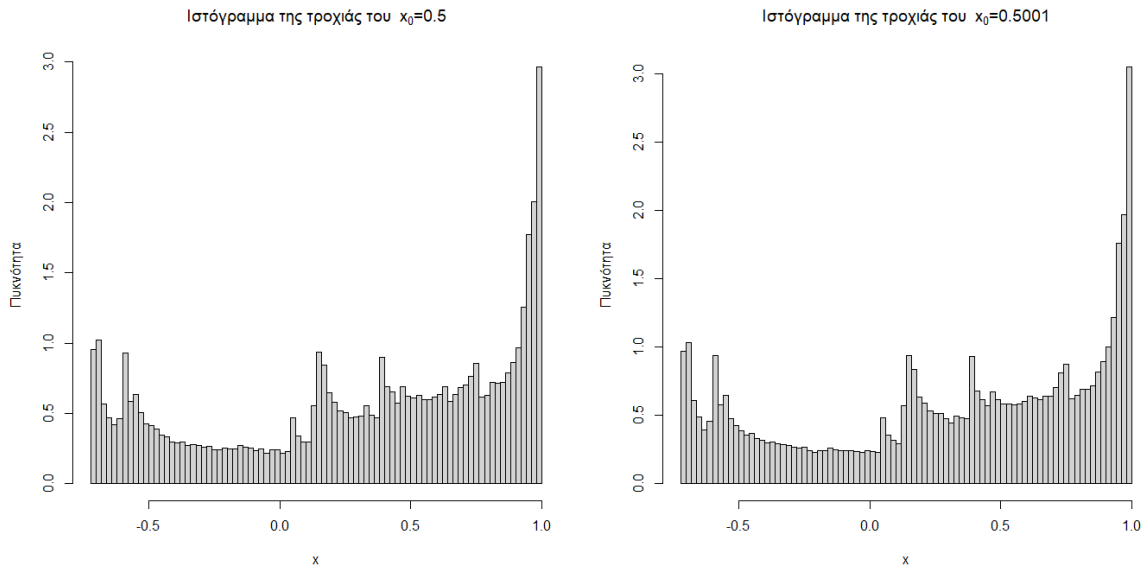
$$g^{-1}(E) \in \mathcal{N}, \quad \forall E \in \mathcal{N}$$

όπου $g^{-1}(E) = \{x \in \mathbf{X} : g(x) \in E\}$.

Ορισμός 4.1.9 (Αναλλοίωτο μέτρο, Day and Pianigiani 1991). Έστω $g : \mathbf{X} \rightarrow \mathbf{X}$ μία απεικόνιση. Ένα μέτρο μ καλείται αναλλοίωτο υπό την g ή αλλιώς g -αναλλοίωτο εάν

$$\mu(g^{-1}(E)) = \mu(E), \quad \forall E \in \mathcal{N}.$$

Διαισθητικά, όταν ένα μέτρο είναι αναλλοίωτο, αυτό σημαίνει ότι διατηρεί τις πιθανοτικές του ιδιότητες, δηλαδή η κατανομή παραμένει ίδια. Για παράδειγμα, στο Σχήμα 4.3 δίνονται τα ιστογράμματα 200,000 τιμών της λογιστικής απεικόνισης για $x_0 = 0.5$ και $x_0 = 0.5001$ αντίστοιχα. Οι διαφορές στα δύο διαγράμματα δεν είναι εύκολα αντιληπτές, αυτό σημαίνει ότι το μέτρο είναι αναλλοίωτο. Επομένως, παρόλο που οι δύο τροχιές αποκλίνουν εκθετικά, όπως είδαμε στην προηγούμενη ενότητα, η κατανομή τους παραμένει ίδια. Για περισσότερες λεπτομέρειες σχετικά με τα αναλλοίωτα μέτρα, ο αναγνώστης παραπέμπεται στην εξής βιβλιογραφία (Ding and Zhou, 2009).



Σχήμα 4.3: Ιστογράμματα των τροχιών της λογιστικής απεικόνισης (4.8) για $x_0 = 0.5$ (αριστερά) και $x_0 = 0.5001$ (δεξιά).

Ορισμός 4.1.10. (Αναλλοίωτο γεγονός). Ένα γεγονός $A \in \mathcal{F}$ ονομάζεται αναλλοίωτο αν $\varphi^{-1}(A) = A$ και η σ -άλγεβρα των αναλλοίωτων συνόλων δίνεται από

$$\mathcal{I} = \{A \in \mathcal{F} : \varphi^{-1}(A) = A\}.$$

²Ονομάζεται η τριάδα $(\mathbf{X}, \mathcal{N}, \mu)$, όπου \mathbf{X} ένα σύνολο, $\mathcal{N} \subseteq \mathbb{P}(\mathbf{X})$ μία σ -άλγεβρα και μ ένα μέτρο πιθανότητας στον $(\mathbf{X}, \mathcal{N})$.

Σημειώνεται ότι αν $\mathbb{P}[A] \in \{0, 1\}$, για κάθε $A \in \mathcal{I}$, τότε η σ -άλγεβρα \mathcal{I} λέγεται \mathbb{P} -τετριμμένη.

Ορισμός 4.1.11. (Διατήρηση μέτρου και εργοδικά δυναμικά συστήματα). Μία απεικόνιση φ διατηρεί το μέτρο εάν

$$\mathbb{P}[\varphi^{-1}(A)] = \mathbb{P}[A] \quad \text{για κάθε } A \in \mathcal{F}.$$

Επιπλέον, η τετράδα $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ ορίζεται να είναι το δυναμικό σύστημα που διατηρεί το μέτρο. Στην περίπτωση που η απεικόνιση φ διατηρεί το μέτρο και \mathcal{I} είναι \mathbb{P} -τετριμμένη, τότε το δυναμικό σύστημα $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ ονομάζεται εργοδικό.

Έστω $(X_n)_{n \in \mathbb{N}}$ μία στοχαστική διαδικασία και φ ο τελεστής μετατόπισης που ορίζεται ως

$$\varphi : \Omega \rightarrow \Omega \quad (\omega_n)_{n \in \mathbb{N}} \rightarrow (\omega_{n+1})_{n \in \mathbb{N}}.$$

Τότε η στοχαστική διαδικασία $X_n(\omega) = X_0(\varphi^n(\omega))$ είναι στάσιμη αν και μόνο αν η τετράδα $(\Omega, \mathcal{F}, \mathbb{P}, \varphi)$ είναι δυναμικό σύστημα που διατηρεί το μέτρο. Επιπλέον, αν το παραπάνω δυναμικό σύστημα είναι εργοδικό, τότε η στοχαστική διαδικασία είναι και αυτή εργοδική.

Θεώρημα 4.1.2. (Εργοδικό θεώρημα, Birkhoff 1931). Έστω $X_0 \in \mathcal{L}^1(\mathbb{P})$. Τότε

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k = \frac{1}{n} \sum_{k=0}^{n-1} g \circ \varphi^k \xrightarrow{n \rightarrow \infty} \mathbb{E}[X_0 | \mathcal{I}]. \quad \mathbb{P} - \sigma. \beta.$$

Συγκεκριμένα, αν η απεικόνιση φ είναι εργοδική, τότε

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{n \rightarrow \infty} \mathbb{E}[X_0]. \quad \mathbb{P} - \sigma. \beta.$$

Για τον ορισμό του δυναμικού συστήματος, το σύστημα της σχέσης (4.9) γράφεται ως

$$x_i = g(\vartheta, x_{i-1}) + e(\varphi^i \omega), \quad i \geq 1 \quad (4.10)$$

όπου $e(\varphi^i \omega)$ είναι μια εργοδική διαδικασία. Ένα στοχαστικό δυναμικό σύστημα ορίζεται από μία μετρήσιμη απεικόνιση ψ ως εξής:

$$\psi(\omega, x) := g(\vartheta, x_{i-1}) + e(\omega). \quad (4.11)$$

Ορισμός 4.1.12. (Στοχαστικό Δυναμικό Σύστημα, Arnold 1998). Ένα στοχαστικό δυναμικό σύστημα είναι μία μετρήσιμη απεικόνιση

$$\begin{aligned} \psi : T \times \Omega \times \mathbf{X} &\rightarrow \mathbf{X} \\ (t, \omega, x) &\mapsto \psi(t, \omega, x) \end{aligned}$$

τέτοια ώστε για όλα τα $\omega \in \Omega$ να ισχύουν τα εξής:

1. $\psi(0, \omega) = \mathbf{1}_X$
2. $\psi(t + s, \omega) = \psi(t, \varphi^s \omega) \circ \psi(s, \omega)$ για όλα τα $t, s \in T$

Όπως αναφέρθηκε, στην παρούσα εργασία μελετώνται τα στοχαστικά δυναμικά συστήματα με προσθετικό δυναμικό θόρυβο που δίνονται από τη σχέση (4.9) στην οποία ισχύει ότι $e(\varphi^i \omega) = e_i$. Επομένως, το δυναμικό σύστημα αποτελείται από δύο μέρη, το νετερμινιστικό και το τυχαίο. Επιπλέον, αξίζει να σημειωθεί ότι οι καταστάσεις του δυναμικού συστήματος (4.9) είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές της διαδικασίας $\{X_n = \psi(n, \omega, x_0), n \in \mathbb{N}, \omega \in \Omega\}$, η οποία πρόκειται για μια μαρκοβιανή αλυσίδα με πυρήνα μετάβασης (Arnold, 1998)

$$Q(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x). \quad (4.12)$$

Ημι-αναλλοίωτο μέτρο

Συνήθως, η μακροπρόθεσμη συμπεριφορά ενός στοχαστικού δυναμικού συστήματος οφείλεται στη συμπεριφορά του νετερμινιστικού του μέρους. Το ημι-αναλλοίωτο μέτρο ενός στοχαστικού συστήματος θεωρείται μία εξομαλυμένη παραμόρφωση του αντίστοιχου αναλλοίωτου μέτρου ενός νετερμινιστικού συστήματος. Συγκεκριμένα, ορίζεται ως

$$\tilde{\mu}_g(E) = \lim_{n \rightarrow \infty} \mathbb{P}\{x_n \in E | n < \tau_{X^c}\},$$

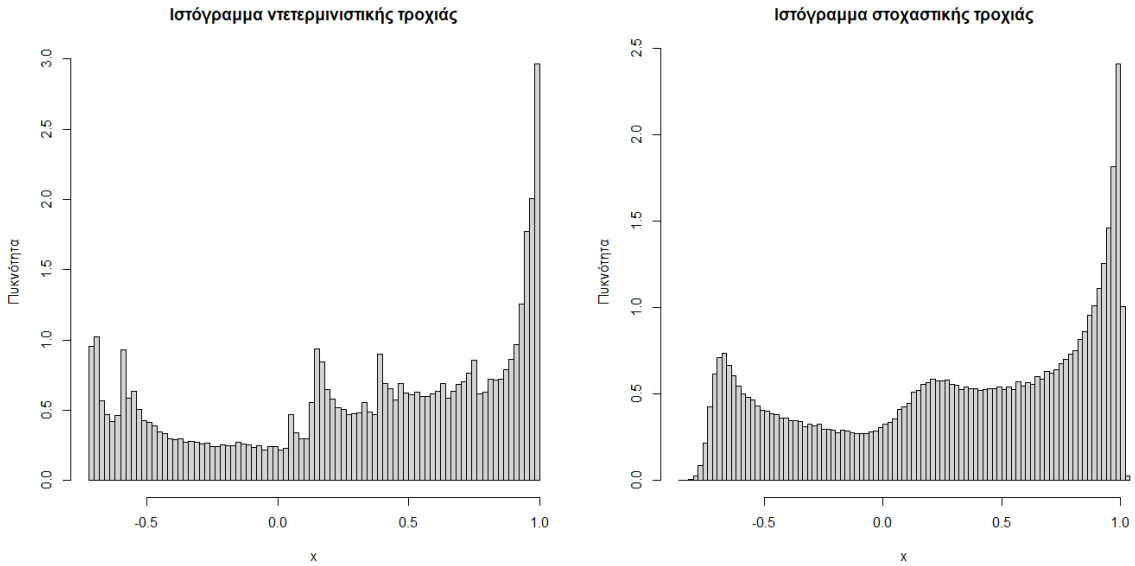
όπου τ_{X^c} είναι ο χρόνος κατά τον οποίο το σύστημα εισέρχεται στο σύνολο X^c όπου και παγιδεύεται (Collet et al. 2012). Για την καλύτερη κατανόηση του ημι-αναλλοίωτου μέτρου, έπεται ένα παράδειγμα, στο οποίο παρουσιάζονται τα ιστογράμματα μίας νετερμινιστικής τροχιάς και μίας στοχαστικής.

Έστω το νετερμινιστικό δυναμικό σύστημα της σχέσης (4.8) και το στοχαστικό δυναμικό σύστημα

$$x_i = 1 - \vartheta x_{i-1}^2 + e_i, \quad e_i \sim N(0, \sigma^2) \quad i = 1, \dots, 200,000 \quad (4.13)$$

όπου η παράμετρος ελέγχου $\vartheta = 1.71$, η αρχική συνθήκη $x_0 = 0.5$, και $\sigma = 0.01$.

Στο Σχήμα 4.4, αριστερά παρουσιάζεται το ιστόγραμμα μίας νετερμινιστικής τροχιάς 200,000 καταστάσεων από το δυναμικό σύστημα (4.8) με παράμετρο ελέγχου $\vartheta = 1.71$, ενώ αριστερά φαίνεται το ιστόγραμμα της στοχαστικής τροχιάς του δυναμικού συστήματος (4.13). Συγκρίνοντας τα δύο ιστογράμματα, με την εισαγωγή του θορύβου στο σύστημα, το ημι-αναλλοίωτο μέτρο φαίνεται ότι είναι μια εξομαλυμένη παραμόρφωση του αναλλοίωτου μέτρου του νετερμινιστικού συστήματος.



Σχήμα 4.4: Ιστογράμματα των τροχιών ενός ντετερμινιστικού (αριστερά) και ενός στοχαστικού (δεξιά) δυναμικού συστήματος με παράμετρο ελέγχου $\theta = 1.71$ και $\sigma = 0.01$.

Στόχος της επόμενης ενότητας είναι η ανακατασκευή των δυναμικών εξισώσεων με στόχο την εκτίμηση της παραμέτρου ελέγχου, της αρχικής συνθήκης x_0 , καθώς και της άγνωστης πυκνότητας του θορύβου.

4.2 Μοντέλο Ανακατασκευής Διαδικασίας Dirichlet

Στην ενότητα αυτή παρουσιάζεται ένα μη παραμετρικό Μπεϋζιανό μοντέλο για την ανακατασκευή μη γραμμικών δυναμικών συστημάτων με βάση τη διαδικασία Dirichlet. Στη διεθνή βιβλιογραφία επικρατεί η υπόθεση της κανονικότητας του θορύβου στα στοχαστικά δυναμικά συστήματα, ωστόσο η υπόθεση αυτή δεν είναι πάντα αληθής και μπορεί να δημιουργήσει προβλήματα στην εκτίμηση των παραμέτρων αλλά και στην πρόβλεψη. Για παράδειγμα, μπορεί να οδηγήσει σε υπερεκτίμηση της διασποράς του θορύβου λόγω ακραίων τιμών, όταν η κατανομή του έχει παχύτερες ουρές. Για το λόγο αυτό, η μοντελοποίηση του θορύβου γίνεται υπό την υπόθεση ότι ο θόρυβος προέρχεται από μία άπειρη μίξη κατανομών, που οι απλές παραμετρικές MCMC μέθοδοι αδυνατούν να χειριστούν και στη συνέχεια τα μοντέλα προσαρμόζουν την πολυπλοκότητά τους σύμφωνα με τα δεδομένα διατηρώντας μόνο τον αριθμό των ενεργών στοιχείων της μίξης. Το πλεονέκτημα των μη παραμετρικών Μπεϋζιανών μεθόδων έγκειται στο γεγονός ότι προσδίδουν μία πιο λεπτομερή αναπαράσταση της υποκείμενης πυκνότητας του δυναμικού θορύβου.

Έστω το στοχαστικό δυναμικό μοντέλο

$$x_i = T(\vartheta, x_{i-1}, e_i) = g(\vartheta, x_{i-1}) + e_i, \quad i \geq 1 \quad (4.14)$$

όπου $g : \Theta \times \mathbb{X} \rightarrow \mathbb{X}$, με $\mathbb{X} \subseteq \mathbb{R}$ και $(x_i)_{i \geq 0}, (e_i)_{i \geq 0}$ τυχαίες μεταβλητές από έναν χώρο πιθανότητας (Ω, \mathcal{F}, P) . Επιπλέον, Θ είναι ο παραμετρικός χώρος, g είναι μία μη γραμμική συνεχής απεικόνιση στο x_{i-1} και τα e_i είναι ανεξάρτητα μεταξύ τους και ανεξάρτητα από τα x_i .

Σύμφωνα με τους Hatjispyros et al. (2009), το μοντέλο ανακατασκευής υποθέτει ότι ο θόρυβος προέρχεται από μία μίξη Κανονικών κατανομών με μηδενική μέση τιμή $N(x|0, \tau^{-1})$ όπου τ είναι η ακρίβεια για την οποία ισχύει $\mathbb{P} = \sum_{j \geq 1} w_j \delta_{\tau_j}$. Επομένως,

$$f_{w,\tau}(x) = \int_{\tau > 0} N(x|0, \tau^{-1}) \mathbb{P}(d\tau) = \sum_{j=1}^{\infty} w_j N(x|0, \tau_j^{-1}). \quad (4.15)$$

Ο πυρήνας μετάβασης για τα δεδομένα $(x^{(n)}|x_0)$ είναι

$$f_{w,\tau}(x_i|x_{i-1}, \vartheta) = \sum_{j=1}^{\infty} w_j N(x_i|g(\vartheta, x_{i-1}), \tau_j^{-1}), \quad 1 \leq i \leq n \quad (4.16)$$

ενώ η πιθανοφάνεια του μοντέλου είναι

$$f_{w,\tau}(x_1, \dots, x_n|x_0, \vartheta) = \prod_{i=1}^n \sum_{j=1}^{\infty} w_j N(x_i|g(\vartheta, x_{i-1}), \tau_j^{-1}). \quad (4.17)$$

Για την εκτίμηση των παραμέτρων ενδιαφέροντος χρησιμοποιούνται τεχνικές δειγματοληψίας slice, όπου εισάγονται οι τυχαίες μεταβλητές d_i και κατασκευάζονται τα τυχαία σύνολα \mathbb{A}_i . Όπως έχει αναφερθεί και στο προηγούμενο κεφάλαιο, η μεταβλητή d_i έχει το ρόλο της δείκτριας μεταβλητής, καθώς προσδιορίζει από ποιο στοιχείο της μίξης προήλθε κάθε παρατήρηση x_i , και πραγματοποιείται δειγματοληψία από τα slice σύνολα \mathbb{A}_i . Θεωρώντας ότι η $d_i|\mathbb{A}_i$ είναι μία διακριτή ομοιόμορφη κατανομή, τότε

$$f(d_i|\mathbb{A}_i) = \frac{1}{|\mathbb{A}_i|} I(j \in \mathbb{A}_i)$$

και άρα

$$\begin{aligned} f_{\tau}(x_i|\mathbb{A}_i) &= \sum_{j=1}^{\infty} f(x_i, d_i = j|\mathbb{A}_i) \\ &= \sum_{j=1}^{\infty} f(d_i = j|\mathbb{A}_i) f_{\tau}(x_i|d_i = j) \end{aligned}$$

$$= \sum_{j \in \mathbb{A}_i} \frac{1}{|\mathbb{A}_i|} N(x_i|0, \tau_j^{-1}) \quad (4.18)$$

όπου $|\mathbb{A}_i|$ είναι ο αριθμός των στοιχείων του συνόλου. Επομένως, κάθε παρατήρηση $x_i|\tau, \mathbb{A}_i$ προέρχεται από μία πεπερασμένη μίξη κανονικών κατανομών με ίσα βάρη.

Για το μοντέλο ανακατασκευής της διαδικασίας Dirichlet, ανατίθεται σε κάθε παρατήρηση x_i ένα slice σύνολο \mathbb{A}_i το οποίο εξαρτάται από τα βάρη w μέσω των μεταβλητών u_i . Δηλαδή,

$$f_w(d_i = j|u_i) = \frac{I(j \in \mathbb{A}_i)}{\sum_{k=1}^{\infty} I(k \in \mathbb{A}_i)} = \frac{I(u_i < w_j)}{\sum_{k=1}^{\infty} I(u_i < w_k)} = \frac{w_j \mathcal{U}(u_i|0, w_j)}{\sum_{k=1}^{\infty} w_k \mathcal{U}(u_i|0, w_k)}.$$

Για τη μεταβλητή u_i ισχύει

$$u_i|w \sim \sum_{j=1}^{\infty} w_j \mathcal{U}(0, w_j) \quad \text{και} \quad u_i|w, d_i = j \sim \mathcal{U}(0, w_j).$$

Η από κοινού πυκνότητα των u_i, d_i είναι $(u_i, d_i = j) = w_j \mathcal{U}(0, w_j)$, ενώ για δοθέν $d_i = j$ ισχύει $f(x_i|d_i = j) = N(x_i|0, \tau_j)$. Επομένως,

$$f_{w,\tau}(x_i, u_i, d_i = j) = w_j \mathcal{U}(u_i|0, w_j) N(x_i|0, \tau_j^{-1}). \quad (4.19)$$

Από τις σχέσεις (4.16) και (4.17) και κατασκευάζοντας τα βάρη σύμφωνα με τη σχέση (3.7) προκύπτει ότι

$$f_{w,\tau}(x_i, u_i, d_i = j|x_{i-1}, \vartheta) = w_j \mathcal{U}(u_i|0, w_j) N(x_i|g(\vartheta, x_{i-1}), \tau_j^{-1}). \quad (4.20)$$

Παράλληλα, η ιεραρχική αναπαράσταση του μοντέλου είναι η εξής

$$\begin{aligned} (x_i|x_{i-1}, d_i = j, \vartheta, \tau) &\stackrel{ind}{\sim} N(x_i|g(\vartheta, x_{i-1}), \tau_j^{-1}) \\ (u_i|d_i = j, w) &\stackrel{ind}{\sim} \mathcal{U}(0, w_j) \\ P(d_i = j|w) &= w_j \\ w_j &= v_j \prod_{s < j} (1 - v_s) \\ v_j &\stackrel{iid}{\sim} \mathcal{Be}(1, c) \\ c &\sim \mathcal{G}(\alpha, \beta) \\ \tau_j &\stackrel{iid}{\sim} G_0 \\ \vartheta &\sim \pi(\vartheta) \\ x_0 &\sim \pi(x_0) \end{aligned}$$

Επιπλέον, η πιθανοφάνεια του μοντέλου είναι

$$f_{w,\tau}(x_i, u_i, d_i; 1 \leq i \leq n | \vartheta, x_0, c) \propto \prod_{i=1}^n I(u_i < w_{d_i}) \tau_{d_i}^{1/2} \exp \left\{ -\frac{\tau_{d_i}}{2} h_{\vartheta}(x_i, x_{i-1}) \right\} \quad (4.21)$$

όπου $h_{\vartheta}(x_i, x_{i-1}) = (x_i - g(\vartheta, x_{i-1}))^2$.

Πριν τη δειγματοληψία των μεταβλητών ενδιαφέροντος, είναι απαραίτητη η αρχικοποίηση των d_i για $i = 1, \dots, n$ καθώς και των c, x_0 και ϑ .

1. Πρώτο βήμα είναι η δημιουργία των v_j από την

$$f(v_j | \cdot) = \mathcal{B}e \left(1 + \sum_{i=1}^n I(d_i = j), c + \sum_{i=1}^n I(d_i > j) \right) \quad (4.22)$$

για $1 \leq j \leq \max_i d_i$. Στη συνέχεια, κατασκευάζονται τα stick-breaking βάρη $(w_j)_{j \geq 1}$ σύμφωνα με τη σχέση (3.7).

2. Δεύτερο βήμα είναι η δημιουργία των μεταβλητών u_i

$$f(u_i | \cdot) = \mathcal{U}(u_i | 0, w_{d_i}) \quad (4.23)$$

για $i = 1, \dots, n$.

3. Επόμενο βήμα είναι η δειγματοληψία των τ_j από την

$$f(\tau_j | \cdot) = \mathcal{G} \left(\alpha + \frac{1}{2} \sum_{i=1}^n I(d_i = j), \beta + \frac{1}{2} I(d_i = j) h_{\vartheta}(x_i, x_{i-1}) \right) \quad (4.24)$$

για $1 \leq j \leq \max_i d_i$. Στην περίπτωση που χρειαστούν περισσότερα τ_j η δειγματοληψία γίνεται από την prior $\mathcal{G}(\alpha, \beta)$ μαζί με τα επιπρόσθετα βάρη που λαμβάνονται από την αντίστοιχη prior.

4. Στη συνέχεια, για τον προσδιορισμό του αριθμού των επιπλέον βαρών και ακριβειών που χρειάζονται είναι απαραίτητος ο υπολογισμός του N^* για το οποίο ισχύει ότι

$$\sum_{j=1}^{N^*} w_j > 1 - u^* \quad (4.25)$$

όπου $u^* = \min_{1 \leq i \leq n} \{u_i\}$.

5. Επομένως τώρα μπορεί να γίνει η δειγματοληψία των δείκτριων μεταβλητών d_i

$$P(d_i = j | \cdot) \propto \tau_j^{\frac{1}{2}} \exp \left\{ -\frac{\tau_j}{2} h_{\vartheta}(x_i, x_{i-1}) \right\} I(j \in \mathbb{A}_i). \quad (4.26)$$

6. Για τον προσδιορισμό της παραμέτρου συγκέντρωσης c , η δειγματοληψία σύμφωνα με τον West (1992) πραγματοποιείται ως εξής:

i. Παραγωγή $s \sim \mathcal{B}e(c + 1, n)$

ii. Παραγωγή $c|s, \kappa \sim p_c \mathcal{G}(\alpha + \kappa, \beta - \log s) + (1 - p_c) \mathcal{G}(\alpha + \kappa - 1, \beta - \log s)$

όπου κ είναι το πλήθος των μοναδικών τιμών της μεταβλητής d_i . Δηλαδή, η παράμετρος c προέρχεται από μία μίξη Γάμμα κατανομών με $p_c = \frac{\alpha + \kappa - 1}{n(\beta - \log s) + \alpha + \kappa - 1}$.

7. Για την εκτίμηση της πυκνότητας του θορύβου η δειγματοληψία γίνεται από την $f(e_{n+1}|x_1, \dots, x_n)$. Κάθε επανάληψη του αλγορίθμου περιέχει τα ανανεωμένα βάρη $(w_j)_{1 \leq j \leq N^*}$ και τις ακρίβειες $(\tau_j)_{1 \leq j \leq \max N^*}$. Η επιλογή των $\tau_j, 1 \leq j \leq N^*$ πρέπει να ικανοποιεί τη σχέση

$$\sum_{i=0}^{j-1} w_i < \rho \leq \sum_{i=1}^j w_i, \quad w_0 = 0$$

όπου $\rho \sim \mathcal{U}(0, 1)$. Στην περίπτωση που $\rho > \sum_{i=0}^{N^*} w_i$ τότε η επιλογή των τ_j γίνεται από την prior $\mathcal{G}(\alpha, \beta)$. Σε κάθε περίπτωση, η δειγματοληψία γίνεται από μία Κανονική κατανομή $N(0, \tau_j^{-1})$.

8. Η δεσμευμένη κατανομή των παραμέτρων ϑ είναι

$$f(\vartheta|\cdot) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \tau_{d_i} h_{\vartheta}(x_i, x_{i-1}) \right\} I(\vartheta \in \Theta) \quad (4.27)$$

9. Τέλος, η δεσμευμένη κατανομή για την αρχική τιμή x_0 είναι

$$f(x_0|\cdot) \propto \exp \left\{ -\frac{\tau_{d_1}}{2} h_{\vartheta}(x_1, x_0) \right\} I(x_0 \in \mathbb{X}). \quad (4.28)$$

Στην ενότητα που ακολουθεί, παρατίθενται τα αποτελέσματα που προέκυψαν από την εφαρμογή του αλγορίθμου για την εκτίμηση της παραμέτρου ελέγχου, της αρχικής συνθήκης και της πυκνότητας του θορύβου ενός стоχαστικού δυναμικού συστήματος.

4.3 Προσομείωση

Με χρήση του στατιστικού πακέτου R, εφαρμόστηκε ο αλγόριθμος που παρουσιάστηκε στην προηγούμενη ενότητα για την εκτίμηση της παραμέτρου ϑ , της αρχικής συνθήκης x_0 και της πυκνότητας του θορύβου.

Οι prior που χρησιμοποιήθηκαν στο μοντέλο είναι οι εξής

$$\begin{aligned} G_0 &\sim \mathcal{G}(0.01, 0.01) \\ c &\sim \mathcal{G}(0.1, 0.1) \\ \vartheta &\sim \mathcal{U}(0, 2) \end{aligned}$$

$$x_0 \sim \mathcal{U}(-2, 2)$$

Επιπλέον, το μοντέλο εφαρμόστηκε σε 500 παρατηρήσεις από το стоχαστικό δυναμικό σύστημα της λογιστικής απεικόνισης

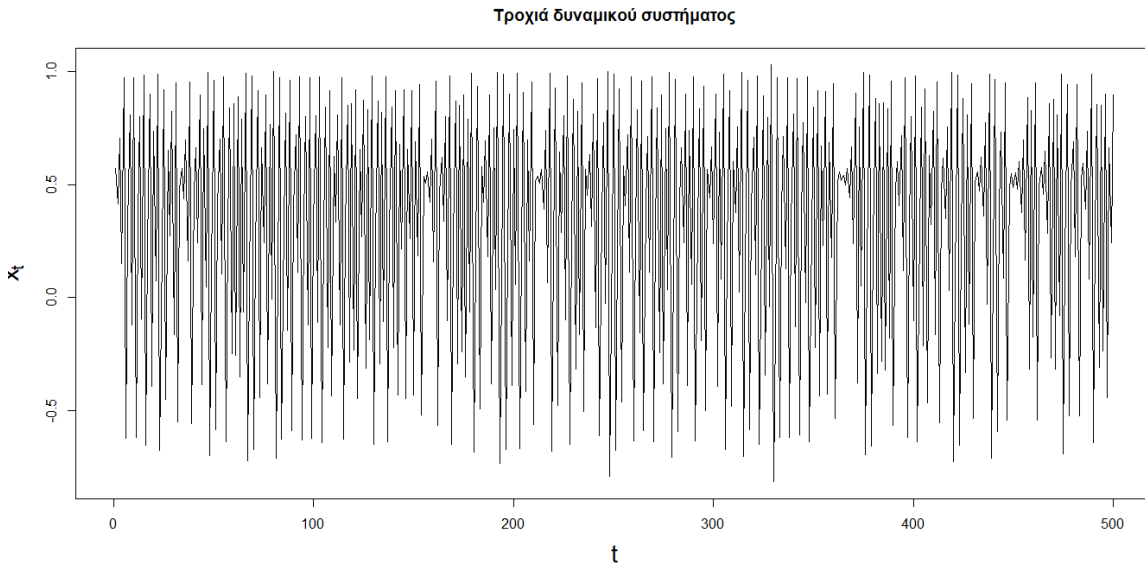
$$x_i = g(\vartheta, x_{i-1}) + e_i = 1 - \vartheta x_{i-1}^2 + e_i, \quad (4.29)$$

όπου η παράμετρος ελέγχου είναι $\vartheta = 1.71$ για την οποία το δυναμικό σύστημα παρουσιάζει χαοτική συμπεριφορά, η αρχική συνθήκη $x_0 = 0.5$, και $(e_i)_{i=1}^n$ ο τυχαίος θόρυβος ο οποίος προέρχεται από μία μίξη Κανονικών κατανομών. Συγκεκριμένα, η μίξη είναι

$$\frac{3}{4}N(0, \sigma_1^2) + \frac{1}{4}N(0, \sigma_2^2). \quad (4.30)$$

όπου $\sigma_1^2 = 10^{-6}$ και $\sigma_2^2 = 10^{-3}$.

Στο Σχήμα 4.5 φαίνεται η τροχιά του δυναμικού συστήματος (4.29) για $n = 500$.



Σχήμα 4.5: Τροχιά του δυναμικού συστήματος (4.29) για $\vartheta = 1.71$, $x_0 = 0.5$, και θόρυβο από τη μίξη (4.30).

Η δειγματοληψία από τις πλήρως δεσμευμένες κατανομές των ϑ και x_0 έγινε σύμφωνα με τους Hatjispyros et al. (2009), όπου εισάγεται η μεταβλητή z_i για $i = 1, \dots, n$. Η πλήρως δεσμευμένη κατανομή των z_i είναι

$$f(z_i|\cdot) \propto \exp\left\{-\frac{z_i \tau_{d_i}}{2}\right\} I(z_i > h_\vartheta(x_i, x_{i-1}))$$

η οποία πρόκειται για μία περικομμένη εκθετική κατανομή στο διάστημα $(h_{\vartheta}(x_i, x_{i-1}), \infty)$. Η δειγματοληψία γίνεται από

$$z_i = \frac{2}{\tau_{d_i}} \log(q_i) + h_{\vartheta}(x_i, x_{i-1})$$

όπου $q_i \sim \mathcal{U}(0, 1)$.

Οι τιμές που προκύπτουν από τις επαναλήψεις της αναδρομικής σχέσης $x_i = g(\vartheta, x_{i-1})$ είναι φραγμένες στο διάστημα $[-\lambda(\vartheta), \lambda(\vartheta)]$, όπου $\lambda(\vartheta) = \frac{1 + \sqrt{1 + 4\vartheta}}{2\vartheta}$ για κάθε $\vartheta \in [0, 2]$. Οι πλήρως δεσμευμένες κατανομές των ϑ και x_0 είναι

$$f(x_0|\cdot) \propto f(x_0)I(z_i > h_{\vartheta}(x_i, x_{i-1}))$$

και

$$f(\vartheta|\cdot) \propto f(\vartheta) \prod_{i=1}^n I(z_i > h_{\vartheta}(x_i, x_{i-1})).$$

Πιο αναλυτικά, για το x_0 κατασκευάζονται οι αριθμοί

$$\rho_1 = \frac{1 - x_1 - \sqrt{z_1}}{\vartheta}$$

$$\rho_2 = \frac{1 - x_1 + \sqrt{z_1}}{\vartheta}$$

επομένως η posterior παίρνει τη μορφή

$$f(x_0|\cdot) \propto f(x_0)I(\rho_1 < x_0^2 < \rho_2).$$

Χρησιμοποιώντας ως prior για το x_0 μία ομοιόμορφη $\mathcal{U}(-2, 2)$ και για το ϑ μία ομοιόμορφη $\mathcal{U}(0, 2)$, η δειγματοληψία για το x_0 γίνεται ως εξής:

$$\text{Αν } \rho_1 < 0 \text{ τότε } f(x_0|\cdot) = \begin{cases} \mathcal{U}(-2, 2) & \text{αν } \rho_2 \geq 4 \\ \mathcal{U}(-\sqrt{\rho_2}, \sqrt{\rho_2}) & \text{αν } \rho_2 < 4. \end{cases}$$

$$\text{Ενώ αν } \rho_1 > 0 \text{ τότε } f(x_0|\cdot) = \begin{cases} \frac{1}{2}\mathcal{U}(-2, -\sqrt{\rho_1}) + \frac{1}{2}\mathcal{U}(\sqrt{\rho_1}, 2) & \text{αν } \rho_2 \geq 4 \\ \frac{1}{2}\mathcal{U}(-\sqrt{\rho_2}, -\sqrt{\rho_1}) + \frac{1}{2}\mathcal{U}(\sqrt{\rho_1}, \sqrt{\rho_2}) & \text{αν } \rho < 4. \end{cases}$$

Ενώ, η πλήρως δεσμευμένη κατανομή του ϑ μετατρέπεται σε

$$f(\vartheta|\cdot) \propto \prod_{i=1}^n I(z_i > h_{\vartheta}(x_i, x_{i-1})) \propto I(\alpha < \vartheta < \beta)$$

όπου

$$\alpha = \max_i \left\{ \frac{1 - x_i - \sqrt{z_i}}{x_{i-1}^2} \right\}$$

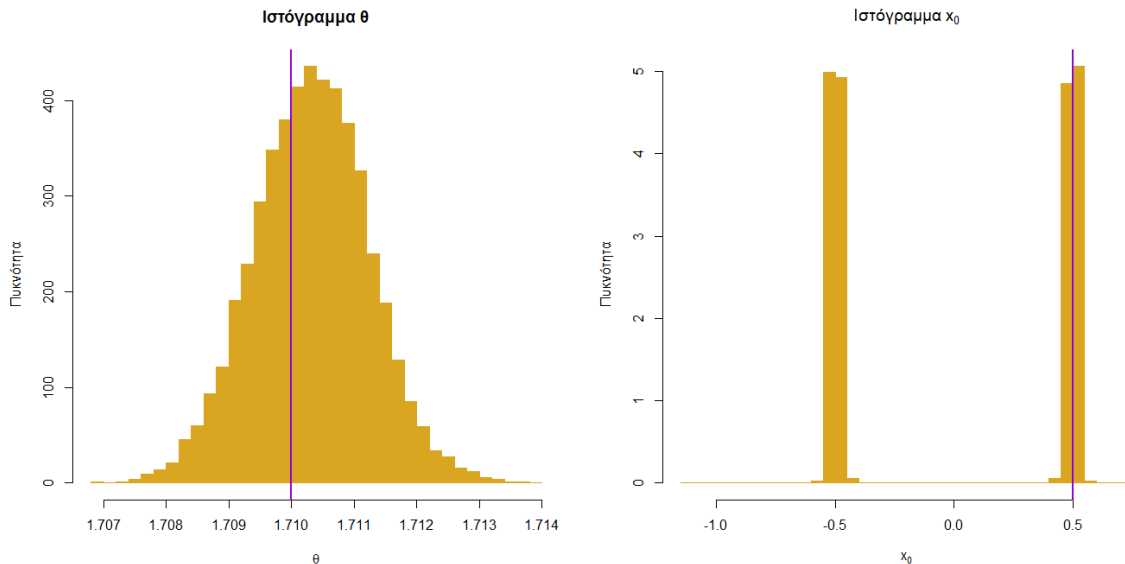
και

$$\beta = \max_i \left\{ \frac{1 - x_i + \sqrt{z_i}}{x_{i-1}^2} \right\}.$$

Επομένως,

$$f(\vartheta|\cdot) \propto I(\max\{\alpha, 0\} < \vartheta < \min\{\beta, 2\})$$

Στο Σχήμα 4.6 παρουσιάζονται τα ιστογράμματα των posterior δειγμάτων για το ϑ και το x_0 αντίστοιχα. Οι κάθετες μωβ γραμμές αποτελούν την πραγματική τιμή των ϑ και x_0 . Τα δείγματα που παράχθηκαν προέκυψαν ύστερα από 40,000 επαναλήψεις του αλγορίθμου με περίοδο burn-in 20,000.



Σχήμα 4.6: Ιστογράμματα για τα posterior δείγματα των ϑ (αριστερά) και x_0 (δεξιά).

Τα αποτελέσματα για τις εκτιμήσεις των ϑ και x_0 που προέκυψαν δίνονται παρακάτω.

- Για την εκτίμηση της παραμέτρου ϑ , εφαρμόστηκε η μέση τιμή του δείγματος που παράχθηκε από τις επαναλήψεις του αλγορίθμου. Η τιμή που δόθηκε είναι **1.71031**.
- Λόγω του ότι η κατανομή της αρχικής συνθήκης x_0 είναι δικόρυφη, για την εκτίμησή της χρησιμοποιήθηκε η επικρατούσα τιμή του δείγματος που παράχθηκε από τις επαναλήψεις του αλγορίθμου. Η τιμή που δόθηκε είναι **0.50486**.

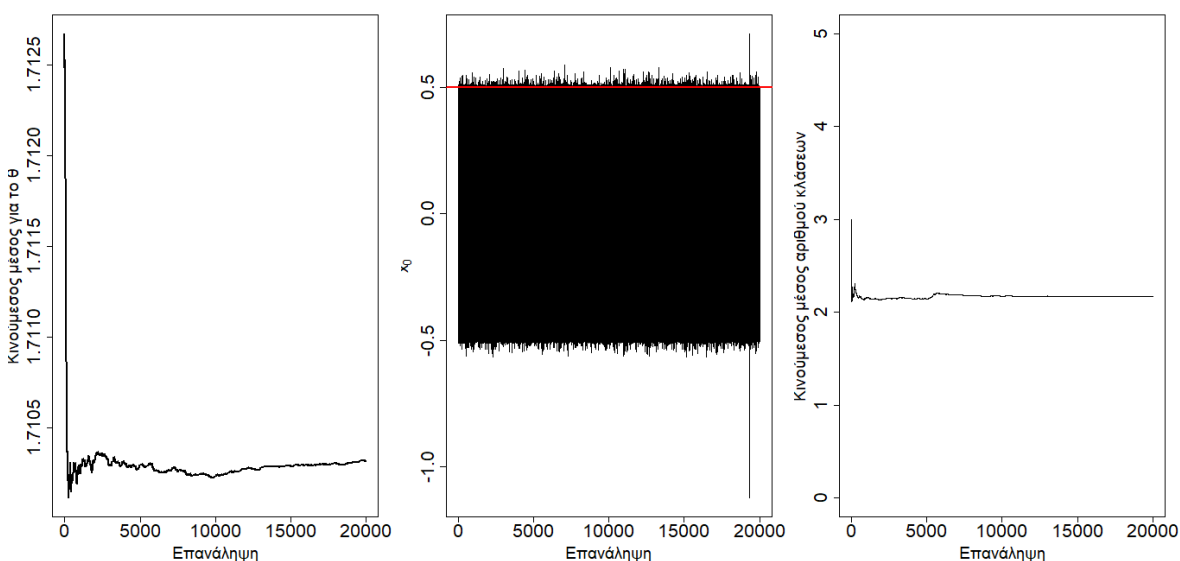
Για την απόδοση των εκτιμήσεων υπολογίστηκε το ποσοστό PARE (Percentage Absolute Relative Error) το οποίο είναι το ποσοστό του απόλυτου σχετικού σφάλματος της εκτίμησης και υπολογίζεται ως εξής

$$\text{PARE} = 100 \times \frac{|x - \hat{x}|}{|x|} \%$$

όπου x η πραγματική τιμή και \hat{x} η εκτιμώμενη τιμή της παραμέτρου ενδιαφέροντος.

Για την εκτίμηση του x_0 το ποσοστό PARE είναι **0.9721%** ενώ για την εκτίμηση της παραμέτρου ελέγχου θ είναι **0.0185%**. Τα ποσοστά είναι πολύ μικρά, επομένως φαίνεται να υπάρχει μεγάλη εκτιμητική ακριβεία στις εκτιμήσεις των θ και x_0 .

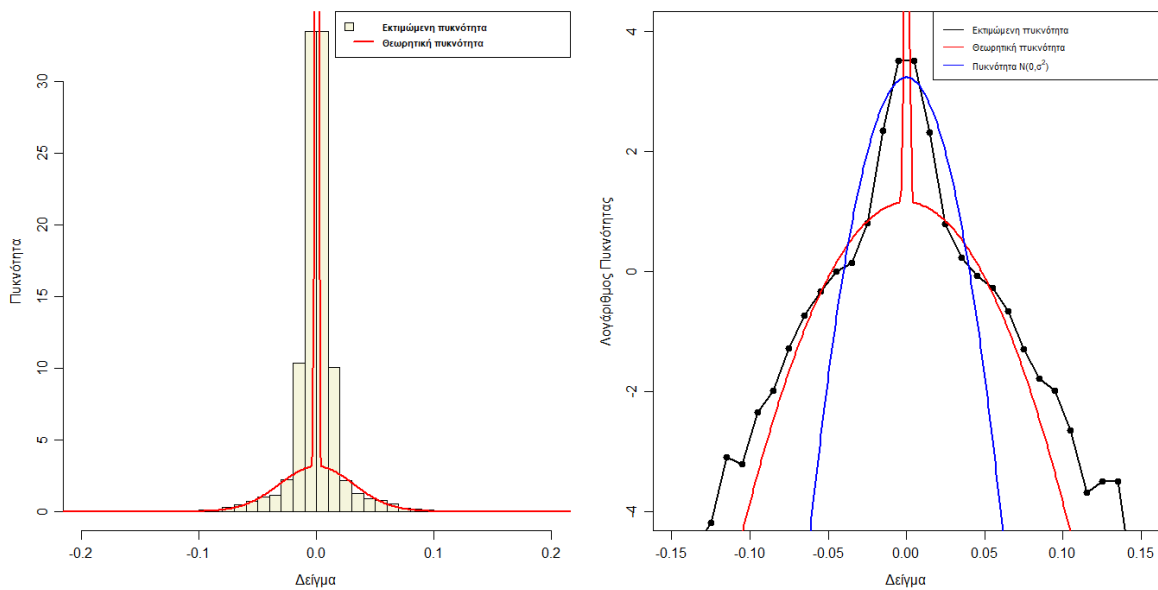
Στο Σχήμα 4.7 δίνονται τρία διαγράμματα. Στο πρώτο διάγραμμα απεικονίζεται το διάγραμμα κινούμενου μέσου του θ , στο δεύτερο διάγραμμα δίνεται το διάγραμμα ίχνους του x_0 , ενώ στο τρίτο φαίνεται το διάγραμμα κινούμενου μέσου των στοιχείων της μίξης (κλάσεων) που χρησιμοποιήθηκαν στον αλγόριθμο για την εκτίμηση της πυκνότητας του θορύβου.



Σχήμα 4.7: Διάγραμμα κινούμενου μέσου για το θ (αριστερά), διάγραμμα ίχνους για το x_0 (κέντρο) και διάγραμμα κινούμενου μέσου για τον αριθμό των στοιχείων μίξης (δεξιά).

Παρακάτω, στο Σχήμα 4.8 δίνονται δύο διαγράμματα. Το πρώτο διάγραμμα είναι το ιστόγραμμα του δείγματος που προέκυψε για την εκτίμηση της πυκνότητας του θορύβου, στο οποίο με κόκκινο χρώμα δίνεται η θεωρητική πυκνότητα του θορύβου. Το δεύτερο διάγραμμα που είναι ένα διάγραμμα σε λογαριθμική κλίμακα, έχει στόχο να

δειξει τη διαφορά που παρατηρείται στις ουρές των κατανομών. Συγκεκριμένα, με μαύρο χρώμα είναι η εκτιμώμενη πυκνότητα του θορύβου, με κόκκινο χρώμα η θεωρητική πυκνότητα του θορύβου και με μπλε χρώμα δίνεται πυκνότητα μιας κανονικής κατανομής με διασπορά ίση με $\sigma^2 = w_1\sigma_1^2 + w_2\sigma_2^2$, η πυκνότητα που θα εκτιμούσε δηλαδή ένα απλό παραμετρικό μοντέλο. Φαίνεται ότι η μαύρη γραμμή, δηλαδή η εκτιμώμενη πυκνότητα είναι πιο κοντά στην κόκκινη της θεωρητικής συγκριτικά με την μπλε, γεγονός που επιβεβαιώνει την εσφαλμένη υπόθεση της κανονικότητας του θορύβου. Με αυτό το διάγραμμα εύκολα συνειδητοποιεί κανείς τη σπουδαιότητα της χρήσης των μη παραμετρικών Μπεϋζιανών μοντέλων.



Σχήμα 4.8: Ιστόγραμμα του δείγματος από την posterior predictive του θορύβου (αριστερά) και λογαριθμικό διάγραμμα πυκνοτήτων (δεξιά).

Κεφάλαιο 5

Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκαν μη παραμετρικές Μπεϋζιανές μέθοδοι για την εκτίμηση παραμέτρων και την εκτίμηση πυκνότητας. Επιπλέον, έγινε εφαρμογή των μοντέλων που παρουσιάστηκαν στο στοχαστικό δυναμικό σύστημα της λογιστικής απεικόνισης. Πιο αναλυτικά, στο Κεφάλαιο 2 συζητήθηκαν τα παραμετρικά μοντέλα μίξης και δόθηκε ένα παράδειγμα για την εκτίμηση των παραμέτρων ενός μοντέλου μίξης κανονικών κατανομών. Έπειτα, στο Κεφάλαιο 3, δόθηκε ο ορισμός της διαδικασίας Dirichlet, η οποία αποτελεί βασικό εργαλείο των αλγορίθμων που αναπτύχθηκαν στη συνέχεια, καθώς χρησιμοποιείται ως prior για το μέτρο μίξης. Συγκεκριμένα, έγινε μία εισαγωγή στα μη παραμετρικά Μπεϋζιανά μοντέλα μίξης και δόθηκε ο αλγόριθμος που χρησιμοποιείται για την εκτίμηση πυκνότητας σε συνδυασμό με μία εφαρμογή του. Στο Κεφάλαιο 4, αναφέρθηκαν κάποιες βασικές έννοιες για τα δυναμικά συστήματα και παρουσιάστηκε ο αλγόριθμος για την εκτίμηση των παραμέτρων ενός στοχαστικού δυναμικού συστήματος και την εκτίμηση της πυκνότητας του θορύβου που το διαταράσσει.

Ο λόγος που η μη παραμετρική Μπεϋζιανή συμπερασματολογία είναι τόσο δημοφιλής, είναι επειδή μπορεί να εφαρμοστεί σε προβλήματα όπου ο αριθμός των παραμέτρων είναι άγνωστος. Τα μη παραμετρικά Μπεϋζιανά μοντέλα υποθέτουν ότι ο αριθμός των παραμέτρων είναι δυνητικά άπειρος και προσαρμόζουν την πολυπλοκότητά τους σύμφωνα με τα δεδομένα. Αυτός είναι και ο λόγος που τα καθιστά αποτελεσματικά και ευέλικτα. Η επέκταση αυτών των μοντέλων σε προβλήματα εκτίμησης παραμέτρων και εκτίμησης πυκνότητας μπορεί να εφαρμοστεί σε στοχαστικά δυναμικά συστήματα. Λόγω του ότι στη διεθνή βιβλιογραφία κυριαρχεί η υπόθεση της κανονικότητας του θορύβου στα στοχαστικά δυναμικά συστήματα, στις περιπτώσεις που η υπόθεση αυτή είναι λανθασμένη, δημιουργεί προβλήματα. Συγκεκριμένα, παρατηρείται υπερεκτίμηση στη διασπορά του θορύβου, η οποία οδηγεί σε μεγάλα εκτιμητικά σφάλματα για τα x_0 και θ . Το γεγονός αυτό, αποδεικνύει τη σπουδαιότητα του μοντέλου κατά την εφαρμογή του σε προβλήματα εκτίμησης πυκνότητας.

Βιβλιογραφία

- Afanas'ev, V.N., Kolmanovskii, V.B., Nosov, V.R. (1996). Continuous and Discrete Deterministic Systems. In: *Mathematical Theory of Control Systems Design. Mathematics and Its Applications*, vol 341. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-2203-2_1.
- Alligood K.T., Sauer T.D. and Yorke J.A. (1996). *Chaos: An Introduction to Dynamical Systems*. Springer New York, NY.
- Arnold, L. (1998). Basic Definitions. *Invariant Measures. Random Dynamical Systems*, 3-47. doi:10.1007/978-3-662-12878-7_1.
- Banks, J., Brooks, J., Cairns, G., Davis, G., and Stacey, P. (1992). On Devaney's Definition of Chaos. *The American Mathematical Monthly*, 99(4), 332. doi:10.2307/2324899
- Berger, J. O. (1985). Bayesian Analysis. *Statistical Decision Theory and Bayesian Analysis*, 118-307. doi:10.1007/978-1-4757-4286-2_4.
- Bernardo, J. M. DeGroot M.H. Lindley D.V. and Smith A.F.M. (1988). Recent Progress on de Finetti's Notions of Exchangeability. *Bayesian Statistics 3*. pp. 111-125. Available from: https://statweb.stanford.edu/~cgates/PERSI/papers/recent_progress.pdf
- Bernardo, J.M. and Smith A.F.M. (1994). *Bayesian Theory*. John Wiley & Sons.
- Billingsley, P. (1995) *Probability and Measure*. 3rd Edition, John Wiley and Sons, New York.
- Birkhoff, G. D. (1931). Proof of the Ergodic Theorem. *Proceedings of the National Academy of Sciences*, 17(12), 656-660. doi:10.1073/pnas.17.2.656.
- Blei D. M. Kucukelbir A. and McAuliffe J. D. (2017) Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*. 112:518. 859-877. doi:10.1080/01621459.2017.1285773.
- Bolstad, W.M. (2011). *Understanding Computational Bayesian Statistics*. John Wiley & Sons.

- Briggs, W. (2016) *Uncertainty : The Soul of Modeling, Probability & Statistics*. Springer.
- Brin, M. and Stuck, G. (2002). *Introduction to Dynamical Systems*. Cambridge: Cambridge University Press.
- Casella, G. and Berger, R.L. (2002) *Statistical Inference*. 2nd Edition, Duxbury Press, Pacific Grove
- Cinlar, E. (1975) *Introduction to Stochastic Processes*. Prentice Hall, Englewood Cliffs, NJ.
- Collet, P., Martínez S. and San Martín J. (2012). *Quasi-Stationary Distributions*. Springer Science & Business Media.
- Damien, P. and Walker, S.G. (2001). Sampling Truncated Normal, Beta, and Gamma Densities., *Journal of Computational and Graphical Statistics*, 10:2, 206-215, doi: 10.1198/10618600152627906.
- Damien, P., Wakefield, J. and Walker, S. (1999). Gibbs Sampling for Bayesian non-Conjugate and Hierarchical Models by using Auxiliary Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2), pp.331-344. doi:<https://doi.org/10.1111/1467-9868.00179>.
- Day, R. H., and Pianigiani, G. (1991). Statistical Dynamics and Economics. *Journal of Economic Behavior & Organization*, 16(1-2), 37-83. doi:10.1016/0167-2681(91)90043-w.
- Devaney R.L. (1989). *An Introduction to Chaotic Dynamical Systems*. Addison Wesley, New York and Reading.
- Diebolt, J., and Robert, C. P. (1994). Estimation of Finite Mixture Distributions Through Bayesian Sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2), 363-375. doi:10.1111/j.2517-6161.1994.tb01985.x.
- Ding, J. and Zhou, A. (2009). *Statistical Properties of Deterministic Systems*. Tsinghua University Texts. doi:10.1007/978-3-540-85367-1.
- Dong, Y. and Peng, C.-Y.J. (2013). *Principled Missing Data Methods for Researchers*. SpringerPlus, [online] 2(1). doi:<https://doi.org/10.1186/2193-1801-2-222>.
- Ferguson, T.S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, [online] 1(2), pp.209-230. doi:<https://doi.org/10.1214/aos/1176342360>.
- Folland, G.B. (1999). *Real Analysis : Modern Techniques and their Applications*. New York: John Wiley And Sons.

- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Science & Business Media.
- Fuentes-García, R., Mena, R.H., and Walker, S.G. (2010). A New Bayesian Nonparametric Mixture Model. *Communications in Statistics - Simulation and Computation*, 39(4), 669-682. doi:10.1080/03610910903580963.
- Ganguly, A. and Earp, S. W. F. (2021). *An Introduction to Variational Inference*. arXiv (Cornell University). doi:<https://doi.org/10.48550/arxiv.2108.13083>.
- Gelman, A. and Rubin, D.B. (1996). Markov Chain Monte Carlo Methods in Biostatistics. *Statistical Methods in Medical Research*, 5(4), pp.339-355. doi: <https://doi.org/10.1177/096228029600500402>.
- Gershman, S.J. and Blei, D.M. (2012). A Tutorial on Bayesian Nonparametric Models. *Journal of Mathematical Psychology*, [online] 56(1), pp.1-12. doi:<https://doi.org/10.1016/j.jmp.2011.08.004>.
- Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Applied Statistics*, 44(4), 455. doi:10.2307/2986138.
- Gleick, J. (1987), *Chaos: Making a New Science*, Viking, New York.
- Green P.J. (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, Volume 82, Issue 4, pp 711-732, <https://doi.org/10.1093/biomet/82.4.711>.
- Hasselblatt, B. and Katok A. (2003). *A First Course in Dynamics: with a Panorama of Recent Developments*. Cambridge University Press.
- Hastings, W.K. (1970). Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, 57(1), pp.97-109. doi:<https://doi.org/10.1093/biomet/57.1.97>.
- Hatjispyros, S.J., Theodoros Nicolieris and Walker, S.G. (2009). A Bayesian Nonparametric Study of a Dynamic Nonlinear Model. *Computational Statistics & Data Analysis*, 53(12), pp.3948-3956. doi:<https://doi.org/10.1016/j.csda.2009.05.009>.
- Heath, D. and Sudderth, W. (1976). De Finetti's Theorem on Exchangeable Variables, *The American Statistician*, 30:4, 188-189, doi: 10.1080/00031305.1976.10479175.
- Hershey, J. R. and Olsen, P. A. (2007). Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models, *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 4, pp. 317-320. doi: 10.1109/ICASSP.2007.366913.

- Hjort N.L, Holmes, C., Müller, P. and Walker, S.G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.
- Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods*. Springer eBooks. Springer Nature. doi:<https://doi.org/10.1007/978-0-387-92407-6>.
- Hoffman, M.J., Blei, D.M., Wang, C. and Paisley, J. (2012). *Stochastic Variational Inference*. arXiv (Cornell University). doi:<https://doi.org/10.48550/arxiv.1206.7051>.
- Ishwaran, H. and James, L.F. (2001). Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, 96(453), pp.161-173. doi:<https://doi.org/10.1198/016214501750332758>.
- Jara, A. (2017). *Theory and Computations for the Dirichlet Process and Related Models: An Overview*. *International Journal of Approximate Reasoning*, 81, 128-146. doi:[10.1016/j.ijar.2016.11.008](https://doi.org/10.1016/j.ijar.2016.11.008)
- 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*, 20(1), 50-67. doi:[10.1214/088342305000000016](https://doi.org/10.1214/088342305000000016).
- Joe, K., and Gooyabadi, M. (2021). A Bayesian Nonparametric Mixture Model for Studying Universal Patterns in Color Naming. *Applied Mathematics and Computation*, 395, 125868. doi:[10.1016/j.amc.2020.125868](https://doi.org/10.1016/j.amc.2020.125868).
- Kaloudis, K. (2019). *Bayesian Methods and Estimation of Nonlinear Dynamical Systems*. Ph. D. Thesis. University of the Aegean. Available at: <https://hellanicus.lib.aegean.gr/handle/11610/21791> (Accessed: 15 November 2023).
- Klenke, A. (2013). *Probability Theory: a Comprehensive Course*. Springer Science & Business Media.
- Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. and Blei, D.M. (2016). Automatic Differentiation Variational Inference. arXiv (Cornell University). doi:<https://doi.org/10.48550/arxiv.1603.00788>.
- Li, Y., Schofield, E., and Gönen, M. (2019). A Tutorial on Dirichlet Process Mixture Modeling. *Journal of Mathematical Psychology*, 91, 128-144. doi:[10.1016/j.jmp.2019.04.004](https://doi.org/10.1016/j.jmp.2019.04.004).
- Marin, J.M. and Robert, C.P. (2014). *Bayesian Essentials with R*. New York, Ny Springer.
- McAuliffe, J.D., Blei, D.M. and Jordan, M.I. (2006). Nonparametric Empirical Bayes for the Dirichlet Process Mixture Model. *Statistics and Computing*, 16(1), pp.5-14. doi:<https://doi.org/10.1007/s11222-006-5196-2>.

McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker

Merkatas, C. (2018). *Bayesian Nonparametrics and Applications*. Ph. D. Thesis. University of the Aegean. Available at: <https://hellanicus.lib.aegean.gr/handle/11610/18560> (Accessed: 3 December 2023).

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6), pp.1087-1092. doi:<https://doi.org/10.1063/1.1699114>.

Müller, P., Quintana, F.A. Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. Springer.

Neal, R. M. (2003). Slice Sampling. *The Annals of Statistics*, 31(3), 705-767. doi: 10.1214/aos/1056562461.

Ng, K.W., Tian, G.-L. and Tang, M.-L. (2011). *Dirichlet and Related Distributions*. John Wiley & Sons.

Nusse, H.E., Yorke, J.A., Kostelich, E.J. (1994). Bifurcation Diagrams. In: *Dynamics: Numerical Explorations*. Applied Mathematical Sciences, vol 101. Springer, New York, NY. https://doi.org/10.1007/978-1-4684-0231-5_6.

Ott, E. (2002). *Chaos in Dynamical Systems*. New York: Cambridge University Press.

Papaspiliopoulos, O. and Roberts, G.O. (2008). Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models. *Biometrika*, 95(1), pp.169-186. doi:<https://doi.org/10.1093/biomet/asm086>.

Papastamoulis, P. and Iliopoulos, G. (2010). An Artificial Allocations Based Solution to the Label Switching Problem in Bayesian Analysis of Mixtures of Distributions. *Journal of Computational and Graphical Statistics*, 19(2), pp.313-331. doi:<https://doi.org/10.1198/jcgs.2010.09008>.

Poirier, D.J. (2011). Exchangeability, Representation Theorems, and Subjectivity. *Oxford Handbooks Online*. doi: <https://doi.org/10.1093/oxfordhb/9780199559084.013.0003>.

Rizzo, M.L. (2007). *Statistical Computing with R*. CRC Press.

Robert, C.P. (2007). *The Bayesian Choice : From Decision Theoretic Foundations to Computational Implementation*. New York: Springer Science Bussines.

Robert, C.P. and Casella, G. (2010). *Monte Carlo Statistical Methods*. New York: Springer.

Robert, C.P. Marin, J.M. and Rousseau, J. (2010). *Bayesian Inference*. [online] [arXiv.org](https://arxiv.org). doi:<https://doi.org/10.48550/arXiv.1002.2080>.

- Sethuraman, J. (1994). A Constructive Definition of the Dirichlet Prior. *Statistica Sinica*, 4, 639-650.
- Stephens, M. (2000). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 795-809. doi:10.1111/1467-9868.00265.
- Strumik, M., and Macek, W.M. (2008). Influence of Dynamical Noise on Time Series Generated by Nonlinear Maps. *Physica D: Nonlinear Phenomena*, 237(5), 613-618. doi:10.1016/j.physd.2007.10.002.
- Teh, Y.W. (2017). Dirichlet Process. In: Sammut, C., Webb, G.I. (eds) *Encyclopedia of Machine Learning and Data Mining*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_219.
- Theodoridis, S. (2020). Bayesian Learning: Approximate Inference and Nonparametric Models. *Machine Learning*, 647-730. doi:10.1016/b978-0-12-818803-3.00025-8.
- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, [online] 22(4), pp.1701-1728. doi:<https://doi.org/10.1214/aos/1176325750>.
- Walker, S.G. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics: Simulation and Computation*, 36(1), pp.45-54. doi:<https://doi.org/10.1080/03610910601096262>.
- Wasserman, L. (2000). Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology*, 44(1), pp.92-107. doi:<https://doi.org/10.1006/jmps.1999.1278>.
- West, M. (1992). Hyperparameter Estimation in Dirichlet Process Mixture Models. Duke University ISDS Discussion Paper 92-A03.
- Zhang, C. Bütepage, J. Kjellström H. and Mandt, S. (2019) Advances in Variational Inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 2008-2026. doi: 10.1109/TPAMI.2018.2889774.

Παράρτημα

Οι αλγόριθμοι για την κατασκευή των μοντέλων και την αναπαραγωγή των αποτελεσμάτων στα Κεφάλαια 2,3 και 4 έγιναν με χρήση του στατιστικού πακέτου R.

Οι αλγόριθμοι είναι διαθέσιμοι στην ιστοσελίδα:

<https://github.com/DimitraMorali/Thesis-Codes>.