**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ**
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**
**ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ**
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ-ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

# Στατιστικη Συμπερασματολογια με Προσεγγιστικες Μπεϋζιανες Μεθοδους

**Βασιλική Γκορτζή**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

**Ιωάννινα, 2021**

*Αφιερώνεται στους γονείς μου*
*Σταύρο και Ειρήνη.*

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στον κλάδο Στατιστική και Επιχειρησιακή Έρευνα που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 15/11/2021 από την εξεταστική επιτροπή:

| **Ονοματεπώνυμο** | **Βαθμίδα** |
|---|---|
| Κωνσταντίνος Ζωγράφος | Καθηγητής |
| Δημήτριος Μπάγκαβος | Επ. Καθηγητής |
| Απόστολος Μπατσίδης | Επ. Καθηγητής **(Επιβλέπων)** |

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

"Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή."

Βασιλική Γκορτζή

# Ευχαριστιες

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στον Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου Ιωαννίνων, με Επιβλέποντα τον Επίκουρο Καθηγητή κ. Απόστολο Μπατσίδη και μέλη της Τριμελούς Εξεταστικής Επιτροπής τους κ.κ. Κωνσταντίνο Ζωγράφο και Δημήτριο Μπάγκαβο.

Στο σημείο αυτό, επιθυμώ να εκφράσω την ευγνωμοσύνη και τις ευχαριστίες μου σε όσους ενίσχυσαν την προσπάθειά μου και συνέβαλλαν στην ολοκλήρωση της διατριβής αυτής.

Πρωτίστως, αισθάνομαι την ανάγκη να ευχαριστήσω θερμά τον Επιβλέποντά μου κ. Απόστολο Μπατσίδη, για τη βοήθειά του, την επιστημονική καθοδήγησή του κατά τη συγγραφή και εκπόνηση της διατριβής και τον πολύτιμο χρόνο που μου διέθεσε, ώστε να επιλύει κάθε απορία και πρόβλημα που συναντούσα. Τον ευχαριστώ ιδιαίτερα για το αμέριστο ενδιαφέρον του, τη συνεχή ηθική του υποστήριξη, την κατανόησή του, καθώς και για την προθυμία του να κατευθύνει τα βήματά μου τόσο στις προπτυχιακές όσο και στις μεταπτυχιακές μου σπουδές. Η αδιάκοπη ενθάρρυνσή του ήταν πάντοτε καθοριστική.

Επιθυμώ, επίσης, να εκφράσω τις ειλικρινείς μου ευχαριστίες προς τους κ.κ. Κωνσταντίνο Ζωγράφο, Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, και Δημήτριο Μπάγκαβο, Επίκουρο Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για τη συμμετοχή τους στην Τριμελή Εξεταστική Επιτροπή, τις χρήσιμες παρατηρήσεις τους και τις υποδείξεις τους.

Θα ήθελα, επιπλέον, να ευχαριστήσω όλους τους καθηγητές του Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών για τις γνώσεις που μου προσέφεραν και την πολύτιμη βοήθειά τους όλα αυτά τα χρόνια. Επιπρόσθετα, ευχαριστώ πολύ τον Ομότιμο Καθηγητή του Τμήματος Μαθηματικών κ. Σωτήριο Λουκά, για τις συμβουλές του και τη συμπαράστασή του.

Τέλος, οφείλω ένα μεγάλο ευχαριστώ στους γονείς μου, Σταύρο και Ειρήνη, και τα αδέλφια μου, Χρήστο και Μαρία, για την πνευματική, ψυχολογική και συναισθηματική τους ενίσχυση, την υπομονή τους, και που δεν έπαψαν ποτέ να με στηρίζουν. Χωρίς τη συμβολή τους θα ήταν αδύνατη η πραγματοποίηση της διατριβής αυτής.

Ιωάννινα, Νοέμβριος 2021.

Βασιλική Γκορτζή

# Περιεχομενα

**Περίληψη** **1**

**Abstract** **5**

**1 Εισαγωγή** **9**

**2 Βασικές μέθοδοι προσομοίωσης Monte Carlo** **15**

2.1 Ψευδοτυχαίοι αριθμοί . . . . . 20

2.2 Μέθοδος αντιστροφής . . . . . 27

2.3 Μέθοδος αποδοχής-απόρριψης . . . . . 36

2.4 Δειγματοληψία σπουδαιότητας . . . . . 46

**3 Βασικές μέθοδοι προσομοίωσης Markov Chain Monte Carlo** **61**

3.1 Αλγόριθμος Metropolis-Hastings . . . . . 62

3.2 Δειγματοληψία κατά Gibbs . . . . . 76

**4 Εκτιμητική με ABC μεθόδους** **85**

4.1 Χαρακτηριστικά Μπεϋζιανής διαδικασίας . . . . . 86

4.2 Παρουσίαση βασικών ABC αλγορίθμων . . . . . 88

4.3 Διευκρινήσεις για την υλοποίηση του ABC αλγορίθμου . . . . . 91

4.3.1 Η σταθερά ανεκτικότητας . . . . . 91

4.3.2 Το μέτρο εγγύτητας . . . . . 92

4.3.3 Τα περιγραφικά στατιστικά μέτρα . . . . . 95

4.3.4 Οι εκ των προτέρων κατανομές . . . . . 97
4.4 Παραδείγματα κατανόησης . . . . . 98
4.5 Τροποποιήσεις του αλγορίθμου ABC . . . . . 106
4.5.1 MCMC-ABC αλγόριθμος . . . . . 107
4.5.2 Regression ABC . . . . . 108
4.5.3 Άλλες τροποποιήσεις . . . . . 110

**5 Επιλογή μοντέλου 115**
5.1 Μπεϋζιανές μέθοδοι . . . . . 115
5.2 ABC μέθοδοι . . . . . 120

**6 Εφαρμογή 129**
6.1 Πληθυσμός και δείγμα υπό μελέτη . . . . . 129
6.2 Μεροληπτικό διδιάστατο δείγμα και σταθμισμένες κατανομές . . . 130
6.3 Εκτιμητική και επιλογή μοντέλου . . . . . 135

**7 Επίλογος 141**

**Βιβλιογραφία 145**

# Περιληψη

Πολύ συχνά, με στόχο την καλύτερη περιγραφή και μοντελοποίηση πραγματικών φαινομένων, προκύπτουν πολύπλοκα στατιστικά υποδείγματα και μοντέλα. Ωστόσο, αυτό που συνήθως αντιμετωπίζουμε σε τέτοιου είδους μοντέλα είναι η αδυναμία να τα διαχειριστούμε αναλυτικά, καθώς έχουν δύσχρηστη (intractable) συνάρτηση πιθανοφάνειας. Με τον όρο intractable εννοούμε μοντέλα, όπου η συνάρτηση πιθανοφάνειας είναι πλήρως ορισμένη από το πιθανοθεωρητικό μοντέλο, δεν είναι διαθέσιμη όμως, σε κλειστή αναλυτική μορφή ως συνάρτηση των άγνωστων πληθυσμιακών παραμέτρων ή το κόστος για τον υπολογισμό της είναι αρκετά μεγάλο, δηλαδή ο υπολογισμός της είναι χρονοβόρος. Το παραπάνω πρόβλημα έχει αντίκτυπο τόσο στην κλασική στατιστική (εύρεση εκτιμητών μέγιστης πιθανοφάνειας) όσο και στη Μπεϋζιανή συμπερασματολογία, όπου η εξαγωγή των συμπερασμάτων προκύπτει από την εκ των υστέρων κατανομή των παραμέτρων, η οποία είναι ανάλογη του γινομένου με όρους της πιθανοφάνειας και της σ.π. ή σ.π.π. της εκ των προτέρων κατανομής των παραμέτρων. Επομένως, υπήρξε απαραίτητη η ανάγκη για εισαγωγή και ανάπτυξη μεθόδων οι οποίες από τη μια μεριά θα αποφεύγουν να υπολογίζουν την πιθανοφάνεια αλλά από την άλλη μεριά θα καθιστούν εφικτή την εξαγωγή στατιστικών συμπερασμάτων. Έχοντας την παραπάνω στόχευση, σχετικά πρόσφατα παρουσιάστηκαν στα πλαίσια επίλυσης προβλήματος πληθυσμιακής γενετικής οι Προσεγγιστικές Μπεϋζιανές Υπολογιστικές μέθοδοι (Approximate Bayesian Computational Methods ή ABC), οι οποίες πλέον αποτελούν μια από τις πιο χρήσιμες και διαδεδομένες τεχνικές για την ανάλυση σύνθετων στοχαστικών μοντέλων.

Σκοπός αυτής της μεταπτυχιακής διατριβής είναι η παρουσίαση και η εφαρμογή σε ένα πραγματικό σύνολο δεδομένων των Προσεγγιστικών Μπεϋζιανών Υπολογιστικών μεθόδων που θα παρουσιαστούν. Στο πλαίσιο αυτό, στο ΚΕΦΑΛΑΙΟ 1 (Εισαγωγή) δίνεται μια εισαγωγή στην κεντρική ιδέα τόσο των υπολογιστικών μεθόδων μελέτης των στοχαστικών φαινομένων, που είναι γνωστές ως μέθοδοι προσομοίωσης, όσο και στην κεντρική ιδέα της ABC μεθόδου. Καθώς για την υλοποίηση της ABC μεθόδου είναι απαραίτητη η προσομοίωση από την εκ των προτέρων κατανομή των άγνωστων πληθυσμιακών παραμέτρων στο ΚΕΦΑΛΑΙΟ

2 (Βασικές Μέθοδοι Προσομοίωσης Monte Carlo) και στο ΚΕΦΑΛΑΙΟ 3 (Βασικές Μέθοδοι Προσομοίωσης Markov Chain Monte Carlo) παρουσιάζονται οι βασικές μέθοδοι προσομοίωσης Monte Carlo και Markov Chain Monte Carlo, αντίστοιχα. Στο ΚΕΦΑΛΑΙΟ 4 (Εκτιμητική με ABC Μεθόδους) το ενδιαφέρον επικεντρώνεται στην αναλυτική παρουσίαση της ABC μεθόδου και στην παράθεση όλων των απαραίτητων διευκρινήσεων για την υλοποίηση στην πράξη του αλγορίθμου ABC, καθώς και των βασικών τροποποιήσεών του. Στο ΚΕΦΑΛΑΙΟ 5 (Επιλογή Μοντέλου) αντικείμενο μελέτης αποτελεί η επιλογή μοντέλου χρησιμοποιώντας ABC μεθόδους. Στο ΚΕΦΑΛΑΙΟ 6 (Εφαρμογή) η εκτίμηση και η επιλογή μοντέλου με Προσεγγιστικές Μπεϋζιανές Υπολογιστικές μεθόδους εφαρμόζεται σε ένα πραγματικό σύνολο δεδομένων. Τέλος, η μεταπτυχιακή διατριβή ολοκληρώνεται με το ΚΕΦΑΛΑΙΟ 7 (Επίλογος), όπου δίνονται κάποιες πιθανές επεκτάσεις, και τις βιβλιογραφικές αναφορές.

# Abstract

Often, in order to describe better some real phenomena, complex statistical models emerge. However, what we typically face with such models is the inability to perform statistical inference exactly, as they have an intractable likelihood function. By "intractable" we mean models where the likelihood function is completely defined by the probabilistic model, but it is not available in closed analytical form as a function of the unknown population parameters or the computational cost for its calculation is quite high, i.e. its calculation is time consuming. This problem has an impact on both classical and Bayesian statistical inference. Therefore, there was a need for the introduction and development of methods which on the one hand will avoid calculating the likelihood but on the other hand they will make it possible to draw statistical conclusions. Aiming at the above, the Approximate Bayesian Computational Methods, which are now one of the most useful and widespread techniques for the analysis of complex stochastic models, were recently introduced in the context of population genetics problem solving.

The goal of this MSc Thesis is the presentation and application in a real data set of the Approximate Bayesian Computational Methods (hereinafter referred to as ABC). In this context, Chapter 1 (Introduction) introduces the central idea of both computational methods for the study of stochastic phenomena, known as simulation methods, and to the central idea of the ABC method. Since the simulation from the prior distribution of the unknown population parameters is necessary for the implementation of the ABC method, in Chapter 2 (Basic Monte Carlo simulation methods) and Chapter 3 (Basic Markov Chain Monte Carlo simulation methods) we present the basic Monte Carlo and Markov Chain Monte Carlo simulation methods, respectively. In Chapter 4 (Estimation with ABC methods), we focus on the presentation of the ABC method and we provide all the necessary clarifications for the implementation of the ABC algorithm and its basic modifications. Chapter 5 (Model selection) refers to the model selection problem in the frame of ABC methods. In Chapter 6 (Application), the estimation and model selection with

Approximate Bayesian Methods is applied to a real data set. Finally, Chapter 7 (Summary) summarizes the MSc Thesis and discusses some possible extensions, and the Thesis is completed with its bibliography.

# ΚΕΦΑΛΑΙΟ 1

# Εισαγωγη

Στις μέρες μας οι ερευνητές φιλοδοξώντας να μοντελοποιήσουν πραγματικά τυχαία φαινόμενα, για παράδειγμα στην Αστρονομία, στην Οικολογία, στη Γενετική, καθώς και στα Οικονομικά και τα Χρηματοοικονομικά (βλέπε, μεταξύ άλλων, Frazier et al. (2020)), οδηγούνται σε ολοένα και πιο περίπλοκα μοντέλα. Όταν έχουμε ένα περίπλοκο στοχαστικό μοντέλο τότε είναι σύνηθες αυτό να έχει intractable (δύσχρηστη, βλέπε Cameron and Pettitt (2012) και Robert (2016)) συνάρτηση πιθανοφάνειας. Η συνάρτηση πιθανοφάνειας είναι ένα από τα σημαντικότερα εργαλεία σε πολλές περιοχές της Στατιστικής και ο ορισμός της έπεται για λόγους πληρότητας.

**Ορισμός 1.0.1.** *Η πυκνότητα πιθανότητας $f(\underset{\sim}{x};\theta)$ στο παρατηρηθέν σημείο $\underset{\sim}{x}$ θεωρούμενη ως συνάρτηση του $\theta$ λέγεται συνάρτηση πιθανοφάνειας.*

Καταστάσεις δύσχρηστων συναρτήσεων πιθανοφανειών μπορούν να προκύψουν, για παράδειγμα, σε συναρτήσεις πιθανοφάνειας που περιλαμβάνουν πολύπλοκα ολοκληρώματα, όπως στη Γενετική, όπου υπάρχει μία ολοκλήρωση πάνω από τα δέντρα συνύπαρξης (coalescence trees, βλέπε Cornuet et al. (2008)) ή ακόμα και σε στοχαστικά μοντέλα μεταβλητότητας (stochastic volatility models) όπου υπάρχει η ολοκλήρωση σε όλη τη χρονική περίοδο παρατήρησης (βλέπε, για παράδειγμα, Creel and Kristensen (2015)). Περιπτώσεις όπου η συνάρτηση πιθανοφάνειας για πλήρη δεδομένα δεν είναι διαθέσιμη σε κλειστή αναλυτική μορφή μπορούν επίσης να εμφανιστούν όταν η σ.π.π. (συνάρτηση πυκνότητας πιθανότητας) του μοντέλου προσδιορίζεται μόνο με έμμεσο τρόπο. Τέτοιες περιπτώσεις προκύπτουν για παράδειγμα, όταν η σ.π.π. προσδιορίζεται μέσω ποσοστιαίων σημείων ή χαρακτηριστικών συναρτήσεων (βλέπε Drovandi et al. (2011) και Peters et al. (2012)). Σε αυτές τις περιπτώσεις κάποιος θα μπορούσε να ισχυριστεί ότι ένας ερευνητής μπορεί να εξάγει τα συμπεράσματα που θέλει παρακολουθώντας πολλές πραγματοποιήσεις του στοχαστικού φαινομένου και καταγράφοντας τα αποτελέσματα που τον ενδιαφέρουν. Ωστόσο, μια τέτοια λύση είναι άλλοτε δαπα-

νηρή οικονομικά αλλά και χρονικά, ενώ σε κάποιες περιπτώσεις είναι ανέφικτη.

Μια εναλλακτική λύση στην αντιμετώπιση των παραπάνω προβλημάτων δόθηκε τις τελευταίες δεκαετίες με τη ραγδαία εξέλιξη στο χώρο της πληροφορικής και την ύπαρξη ηλεκτρονικών υπολογιστών μεγάλης ισχύος. Ειδικότερα, με τη βοήθεια του ηλεκτρονικού υπολογιστή (Η/Υ), το στοχαστικό φαινόμενο αναπαρίσταται εικονικά (είτε με ακρίβεια είτε με κάποιες αναγκαίες, λόγω πολυπλοκότητας, απλουστεύσεις) και αντί να παρατηρούμε ή να περιμένουμε να παρατηρήσουμε πραγματοποιήσεις του στον πραγματικό κόσμο, μπορούμε να πραγματοποιήσουμε έναν πολύ μεγάλο αριθμό φορών στον εικονικό κόσμο του ηλεκτρονικού υπολογιστή. Στη συνέχεια, καταγράφοντας τα αποτελέσματα που αφορούν τα ιδιαίτερα χαρακτηριστικά στα οποία επικεντρώνεται το ενδιαφέρον μας, μπορούμε να οδηγηθούμε στην εξαγωγή συμπερασμάτων. Αυτή ουσιαστικά είναι και η κεντρική ιδέα των εναλλακτικών υπολογιστικών μεθόδων μελέτης των στοχαστικών φαινομένων που είναι γνωστές με τον όρο προσομοίωση (simulation), καθώς μιμούνται τη λειτουργία ενός πραγματικού συστήματος με τη βοήθεια του ηλεκτρονικού υπολογιστή. Με αυτόν τον τρόπο αυτό που επιτυγχάνεται είναι η υλοποίηση ενός σχετικά μεγάλου αριθμού επαναλήψεων που αφορούν διαφορετικές περιπτώσεις των στοχαστικών χαρακτηριστικών, από τις οποίες εξάγονται οι ζητούμενες πληροφορίες που είναι επιθυμητές για τις ιδιότητες του στοχαστικού φαινομένου το οποίο μελετάται. Η έκφραση προσομοίωση είναι αρκετά γενική και αποτελείται από υπολογιστικές μεθόδους που χρησιμοποιούνται σε μια προσπάθεια μοντελοποίησης μιας πραγματικής κατάστασης σε έναν υπολογιστή ή ακόμη μπορεί να χρησιμοποιείται όταν το πραγματικό σύστημα μπορεί να μην είναι προσβάσιμο, ή απλά σχεδιάζεται χωρίς να έχει ακόμη κατασκευαστεί, ή μπορεί απλώς να μην υπάρχει (βλέπε, μεταξύ άλλων, Sokolowski and Banks (2009)). Καθώς δεν έχει δοθεί κάποιος αυστηρός ορισμός για τον όρο προσομοίωση, η παραπάνω περιγραφή θα μπορούσε να αποτελεί έναν μη αυστηρό ορισμό και θα μπορούσε να συμπληρωθεί λέγοντας ότι η προσομοίωση μπορεί να χαρακτηριστεί πρακτικά ως στατιστική δειγματοληψία, η οποία αντί να πραγματοποιείται στον πραγματικό πληθυσμό, πραγματοποιείται στον «ψευδοτυχαίο» κόσμο του ηλεκτρονικού υπολογιστή.

Η χρήση μεθόδων προσομοίωσης είναι απίστευτα ευρεία και το εύρος των εφαρμογών τους είναι δύσκολο να προσδιοριστεί με ακρίβεια. Η δυσκολία αυτή οφείλεται στη γενικότητα της μεθόδου, η οποία επιτρέπει την εφαρμογή της σε ένα τεράστιο φάσμα προβλημάτων. Ο αντίκτυπος των μεθόδων προσομοίωσης στη Στατιστική είναι κάτι που δε θα μπορούσε να γίνει αντιληπτό, όταν πρωτοπαρουσιάστηκαν τέτοιες μέθοδοι από τους von Neumann και Ulam (βλέπε Metropolis and Ulam (1949)). Η χρησιμότητα των μεθόδων προσομοίωσης στη Στατιστική δεν είχε γίνει αντιληπτή μέχρι και τις αρχές της δεκαετίας του 1980,

με τις δημοσιεύσεις των Rubinstein (1981) και Ripley (1987). Όμως έχουν γίνει πλέον πολύτιμες, καθώς μας δίνουν τη δυνατότητα να αντιμετωπίσουμε ένα μεγάλο φάσμα προβλημάτων βελτιστοποίησης και ολοκλήρωσης (που διαφορετικά δε θα ήταν εφικτό), ενώ επιπρόσθετα, η κατασκευή των περιοχών αποδοχής ή απόρριψης κάποιων μηδενικών υποθέσεων στη βάση κάποιων στατιστικών τεστ μπορεί να επιτευχθεί μόνο με τεχνικές προσομοίωσης. Έτσι, με μεθόδους προσομοίωσης είναι δυνατή η μελέτη κάποιων πολύπλοκων στατιστικών προβλημάτων, όπως για παράδειγμα:

- ο προσεγγιστικός υπολογισμός της μέσης τιμής μιας πολύπλοκης συνάρτησης μιας τυχαίας μεταβλητής,
- η προσεγγιστική εύρεση της κατανομής κάποιας πολύπλοκης στατιστικής συνάρτησης,
- ο προσεγγιστικός υπολογισμός της $p$-τιμής ενός στατιστικού ελέγχου,
- η κατασκευή περιοχής απόρριψης κάποιας μηδενική υπόθεσης,
- η προσέγγιση της ισχύος ενός στατιστικού ελέγχου.

Πέρα από τα παραπάνω, η προσομοίωση έχει αξιοποιηθεί και στην αντιμετώπιση των προβλημάτων που προκύπτουν όταν έχουμε δυσεπίλυτα (intractable) μοντέλα. Ειδικότερα, αποτελεί αναπόσπαστο τμήμα κάθε μεθόδου που έχει εισαχθεί στη βιβλιογραφία και η οποία για την εξαγωγή συμπερασμάτων δεν απαιτεί τον αναλυτικό προσδιορισμό της συνάρτησης πιθανοφάνειας, καθώς ουσιαστικά προσπαθεί να την εκτιμήσει μέσω προσομοίωσης. Οι μέθοδοι αυτές είναι γνωστές ως μέθοδοι στατιστικής συμπερασματολογίας που είναι απαλλαγμένες από τη συνάρτηση πιθανοφάνειας (likelihood-free inference methods). Πολλοί ερευνητές ισχυρίζονται ότι το όνομα αυτό δεν είναι αντιπροσωπευτικό και θεωρούν ότι ο όρος συμπερασματολογία βάσει προσομοίωσης (simulation-based inference) είναι πιο κατάλληλος (βλέπε, μεταξύ άλλων, Cranmer et al. (2020)). Τέτοιες μέθοδοι πρωτοπαρουσιάστηκαν από τους Diggle and Gratton (1984) και Rubin (1984). Από τότε πολλές τεχνικές έχουν εμφανιστεί στη στατιστική βιβλιογραφία που θα μπορούσαν να ταξινομηθούν, σύμφωνα με την παραπάνω περιγραφή, σε αυτήν την κατηγορία. Ενδεικτικά, αναφέρουμε ότι τέτοιες μέθοδοι είναι η έμμεση συμπερασματολογία (Indirect Inference, βλέπε Gourieroux et al. (1993)), η bootstrap filter (Gordon et al. (1993)) η Synthetic Likelihood μέθοδος (Wood (2010)) και η Approximate Bayesian Computation (βλέπε Robert et al. (2011b)).

Ανάμεσα στις μεθόδους βάσει προσομοίωσης, οι Προσεγγιστικές Μπεϋζιανές Υπολογιστικές μέθοδοι (Approximate Bayesian Computational Methods) ανα-

γνωρίζονται ως οι πλέον επικρατέστερες της κατηγορίας. Στη συνέχεια, αυτές οι μέθοδοι για λόγους απλότητας και συντομίας θα αναφέρονται ως ABC. Ο ABC αλγόριθμος αποτελεί μία κλάση (κατηγορία) υπολογιστικών μεθόδων, οι οποίες βασίζονται στη Μπεϋζιανή στατιστική και μπορούν να χρησιμοποιηθούν για την εκτίμηση της εκ των υστέρων κατανομής των παραμέτρων του μοντέλου. Το γεγονός αυτό αναδεικνύει και τη σπουδαιότητά του καθώς η εκ των υστέρων κατανομή των παραμέτρων του μοντέλου στη Μπεϋζιανή θεωρία παίζει θεμελιώδη ρόλο στην εξαγωγή συμπερασμάτων. Στη συνέχεια, αρχικά θα παρουσιαστούν ιστορικά στοιχεία και έπειτα η κεντρική ιδέα της μεθόδου (βλέπε, μεταξύ άλλων, Marin et al. (2012), Sunnåker et al. (2013) και Sisson et al. (2018)).

Οι ιδέες που σχετίζονται με τη μέθοδο ABC πρωτοπαρουσιάστηκαν στη βιβλιογραφία τη δεκαετία του 1980. Πιο συγκεκριμένα, ο Rubin (1984) περιέγραψε και ανέπτυξε έναν υποθετικό μηχανισμό δειγματοληψίας που παράγει ένα δείγμα από την εκ των υστέρων κατανομή. Ο μηχανισμός αυτός ήταν περισσότερο φιλοσοφικού περιεχομένου και είχε σκοπό να δείξει τι είδους χειρισμοί απαιτούνται κατά την εξαγωγή των εκ των υστέρων κατανομών των παραμέτρων ενός μοντέλου. Σε αυτό το πλαίσιο, επισημάνθηκε ότι στη Μπεϋζιανή στατιστική οι εφαρμοσμένοι ερευνητές δεν πρέπει να εστιάζουν το ενδιαφέρον τους μόνο στα αναλυτικά μοντέλα, αλλά να εξετάζουν τις υπολογιστικές μεθόδους που τους επιτρέπουν να εκτιμήσουν την εκ των υστέρων κατανομή. Κατά αυτόν τον τρόπο, μπορεί να εξεταστεί και να μελετηθεί ένα ευρύτερο φάσμα μοντέλων (βλέπε Sunnåker et al. (2013)).

Το ίδιο έτος οι Diggle and Gratton (1984), έχοντας ως απώτερο στόχο την προσέγγιση της συνάρτησης πιθανοφάνειας σε περιπτώσεις όπου το μοντέλο είναι περίπλοκο, ανέπτυξαν μια τεχνική η οποία στηρίχθηκε στον καθορισμό ενός πλέγματος στον παραμετρικό χώρο και στη χρήση του για την προσέγγιση της πιθανοφάνειας εκτελώντας αρκετές προσομοιώσεις για κάθε σημείο πλέγματος. Στη συνέχεια, η προσέγγιση αυτή βελτιώνεται εφαρμόζοντας τεχνικές για την εξασφάλιση της ομαλότητας (smoothing techniques) στα αποτελέσματα των προσομοιώσεων. Συμπερασματικά, ενώ η ιδέα της χρήσης προσομοιώσεων για ελέγχους υποθέσεων δεν ήταν καινούρια, οι Diggle and Gratton (1984) παρουσίασαν φαινομενικά την πρώτη διαδικασία με χρήση προσομοιώσεων ώστε να κάνουν στατιστική συμπερασματολογία σε μια περίπτωση που η πιθανοφάνεια είναι intractable (βλέπε Sunnåker et al. (2013) και τις εκεί αναφορές).

Η πρώτη εφαρμογή ABC αλγορίθμου παρουσιάστηκε από τους Tavaré et al. (1997) σε άρθρο πληθυσμιακής γενετικής, με αντικείμενο μελέτης την εκτίμηση του χρόνου συγκερασμού, δηλαδή του χρόνου από τον πιο πρόσφατο κοινό πρόγονο, ενός δείγματος αλληλουχιών DNA ενδοειδών. Οι Tavaré et al. (1997)

ουσιαστικά εισήγαγαν τις ABC μεθόδους ως μια τεχνική αποδοχής-απόρριψης αποφεύγοντας τον υπολογισμό της συνάρτησης πιθανοφάνειας, χρησιμοποιώντας προσομοιωμένες τιμές από την κατανομή ενδιαφέροντος. Στο παραπάνω πλαίσιο, ο αλγόριθμος που προτάθηκε από τους Tavaré et al. (1997) υποθέτει ότι το στήριγμα[1] της κατανομής είναι πεπερασμένο. Άρα δεν μπορεί να υλοποιηθεί όταν τα δεδομένα προέρχονται από ένα μη πεπερασμένο σύνολο. Για το λόγο αυτό ακολούθησε η γενίκευση του αλγορίθμου που προτάθηκε από τους Pritchard et al. (1999), στην εργασία τους με αντικείμενο μελέτης τη μοντελοποίηση της διακύμανσης των ανθρώπινων χρωμοσωμάτων. Για το λόγο αυτό οι Pritchard et al. (1999) θεωρούνται ότι είναι αυτοί που εισήγαγαν επίσημα τον πρώτο ABC αλγόριθμο.

Συγκεντρωτικά, ο ABC αλγόριθμος επίσημα πρωτοπαρουσιάστηκε από τους Pritchard et al. (1999) και καθιερώθηκε με τον όρο ABC από τους Beaumont et al. (2002). Προήλθε ουσιαστικά από την ανάγκη να αντιμετωπιστούν δύσκολα προβλήματα στατιστικής συμπερασματολογίας στη γενετική, όπου η πολυπλοκότητα ενός μοντέλου σήμαινε ότι η συνάρτηση πιθανοφάνειας δεν ήταν υπολογιστικά εφικτή και δεν μπορούσε να εκτιμηθεί αριθμητικά σε οποιοδήποτε χρονικό διάστημα. Πλέον, εκτός από τη Στατιστική, χρησιμοποιείται ευρέως καθώς απέκτησε ιδιαίτερη δημοτικότητα τα τελευταία χρόνια ιδίως για την ανάλυση σύνθετων προβλημάτων που προκύπτουν, για παράδειγμα, στην Επιδημιολογία, Γενετική, Οικολογία, Βιολογία αλλά και στην Κοσμολογία. Στη συνέχεια παρατίθεται η κεντρική ιδέα της ABC μεθόδου.

Σύμφωνα με την κεντρική ιδέα της ABC μεθόδου, αρχικά προσομοιώνεται μια τιμή $\theta^*$ της άγνωστης παραμέτρου $\theta$ από την εκ των προτέρων $\pi(\theta)$. Δοθείσης της τιμής $\theta^*$ προσομοιώνουμε ένα δείγμα, έστω $D^*$, ίδιου μεγέθους με το πραγματικό από το μοντέλο που προκύπτει για την τιμή της παραμέτρου $\theta^*$. Αν το προσομοιωμένο δείγμα $D^*$ είναι κοντά στο πραγματικό δείγμα $D$ τότε αποδεχόμαστε την τιμή $\theta^*$, ενώ διαφορετικά την απορρίπτουμε. Η εγγύτητα των $D^*$ και $D$ ποσοτικοποιείται μέσω ενός μέτρου απόκλισης ή μιας μετρικής και το πόσο κοντά ή μακριά είναι τα δύο δείγματα καθορίζεται με ένα διαχωριστικό όριο (tolerance threshold). Εναλλακτικά, αντί για την εγγύτητα των δύο συνόλων ελέγχεται η εγγύτητα στατιστικών περιγραφικών μέτρων αυτών.

Από την παράθεση της κεντρικής ιδέας της ABC μεθόδου, εύκολα προκύπτει ότι για την υλοποίησή της είναι απαραίτητη η προσομοίωση από την εκ των προτέρων κατανομή των άγνωστων παραμέτρων $\theta$. Στη βιβλιογραφία έχει εμφανιστεί μια πληθώρα τεχνικών για την προσομοίωση από μια κατανομή, η οποία ονο-

[1] Το στήριγμα (support) είναι το μικρότερο υποσύνολο του πεδίου ορισμού της πυκνότητας, για το οποίο η πυκνότητα είναι διάφορη του μηδενός.

μάζεται κατανομή στόχος (target distribution). Οι τεχνικές αυτές χωρίζονται στις τεχνικές Monte Carlo και στις τεχνικές Markov Chain Monte Carlo. Οι κυριότερες τεχνικές κάθε κατηγορίας παρουσιάζονται στο Κεφάλαιο 2 και στο Κεφάλαιο 3, αντίστοιχα, της παρούσας διατριβής. Ταυτόχρονα, παρατίθενται και τρόποι προσεγγιστικού υπολογισμού της μέσης τιμής μιας πολύπλοκης συνάρτησης μιας συνεχούς τυχαίας μεταβλητής. Έπειτα, στο Κεφάλαιο 4 παρουσιάζεται αναλυτικά η ABC μέθοδος. Το Κεφάλαιο 5 επικεντρώνεται στην επιλογή μοντέλου χρησιμοποιώντας ABC μεθόδους. Στο Κεφάλαιο 6, η εκτίμηση και η επιλογή μοντέλου με Προσεγγιστικές Μπεϋζιανές μεθόδους εφαρμόζεται σε ένα πραγματικό σύνολο δεδομένων. Αρχικά περιγράφεται ο υπό μελέτη πληθυσμός, ο τρόπος επιλογής του δείγματος και αιτιολογείται γιατί αυτό είναι ένα διδιάστατο μεροληπτικό και όχι ένα τυχαίο δείγμα. Στο πλαίσιο αυτό, χρησιμοποιείται η έννοια των σταθμισμένων κατανομών για τη μοντελοποίηση διδιάστατων μεροληπτικών δειγμάτων. Ολοκληρώνοντας, στο Κεφάλαιο 7 δίνονται κάποιες πιθανές επεκτάσεις.

# ΚΕΦΑΛΑΙΟ 2

# Βασικες μεθοδοι προσομοιωσης Monte Carlo

Για τη μελέτη, μέσω προσομοίωσης, κάθε στοχαστικού φαινομένου, δηλαδή κάθε φαινομένου που η έκβασή του εξαρτάται από τις τιμές που λαμβάνουν οι υπό μελέτη τυχαίες μεταβλητές, θα πρέπει και το εικονικό πειραματικό μοντέλο που θα κατασκευαστεί με τη βοήθεια ενός Η/Υ να εξαρτάται από τυχαίους αριθμούς, οι οποίοι προσομοιώνουν τους πραγματικούς. Επομένως, η βάση για τη μελέτη, μέσω προσομοίωσης, ενός στοχαστικού φαινομένου είναι μία διαδικασία η οποία προσομοιώνει τιμές τυχαίων μεταβλητών με χρήση του ηλεκτρονικού υπολογιστή. Στο πλαίσιο αυτό, ένα πρώτο σημαντικό πρόβλημα της Υπολογιστικής Στατιστικής είναι η προσομοίωση από μία $d$-διάστατη κατανομή, $d \geqq 1$, με συνάρτηση πιθανότητας (σ.π.) ή πυκνότητας πιθανότητας (σ.π.π.) $\pi(\cdot)$, ενώ ένα δεύτερο σημαντικό ζήτημα στη συνεχή περίπτωση αποτελεί ο υπολογισμός ολοκληρωμάτων[1] της μορφής: $\int_{\mathcal{X}} w(x)\,\mathrm{d}x$, όπου $\mathcal{X} \subseteqq \mathbb{R}^d$, $d \geqq 1$, είναι το στήριγμα της $\pi(\cdot)$. Γράφοντας την $w(x)$ ως το γινόμενο μιας συνάρτησης πυκνότητας πιθανότητας $\pi(x)$ με πεδίο ορισμού το $\mathcal{X}$ και μιας συνάρτησης $h(x)$ το παραπάνω πρόβλημα ανάγεται στον υπολογισμό του ολοκληρώματος:

$$I = \int_{\mathcal{X}} h(x)\pi(x)\,\mathrm{d}x = E_\pi\left(h\left(X\right)\right). \tag{2.1}$$

Δηλαδή ο υπολογισμός του ολοκληρώματος (2.1), ουσιαστικά ισοδυναμεί με τον υπολογισμό της αναμενόμενης τιμής μιας συνάρτησης $d$ το πλήθος τυχαίων μεταβλητών. Πολλές φορές σε περιπτώσεις που είτε η μορφή της συνάρτησης $h$ είναι περίπλοκη είτε η διάσταση $d$ είναι πολύ μεγάλη, ο υπολογισμός του ολοκληρώματος (2.1) σε κλειστή αναλυτική μορφή δεν είναι εφικτός με άμεσες τεχνικές.

[1] Όλα τα ολοκληρώματα που θα εμφανίζονται στη συνέχεια υποθέτουμε ότι είναι καλά ορισμένα και πεπερασμένα.

Τα δύο γενικά προβλήματα που αναφέρθηκαν πρωτύτερα (της προσομοίωσης από μια κατανομή και του υπολογισμού του ολοκληρώματος της μορφής (2.1)) παρότι αρχικά φαίνονται διαφορετικά, είναι άμεσα συνδεδεμένα μεταξύ τους. Αυτό εξηγείται καθώς ο προσεγγιστικός υπολογισμός του ολοκληρώματος ισοδυναμεί με την προσέγγιση αναμενόμενων τιμών. Η προσέγγιση των αναμενόμενων τιμών επιτυγχάνεται χρησιμοποιώντας τους νόμους των μεγάλων αριθμών, γεγονός που οδηγεί τελικά να απαιτείται για την προσέγγιση του ολοκληρώματος να δημιουργηθεί ένα τυχαίο δείγμα από την κατανομή $\pi(\cdot)$. Η παραπάνω σύνδεση θα εξηγηθεί στη συνέχεια αναλυτικά.

Έστω τυχαίο δείγμα (τ.δ.) $X_1, X_2, ..., X_n$ από τον πληθυσμό με σ.π.π. $\pi(x)$. Από τον Ασθενή Νόμο των Μεγάλων Αριθμών (ΑΝΜΑ) είναι γνωστό (βλέπε, μεταξύ άλλων, Van der Vaart (2000)) ότι η δειγματική μέση τιμή συγκλίνει κατά πιθανότητα στη μέση τιμή, υπό την υπόθεση ότι αυτή υπάρχει. Επομένως, υπό την υπόθεση ότι η $E_\pi(h(X))$ υπάρχει και είναι πεπερασμένη, εφαρμόζοντας τον ΑΝΜΑ έχουμε ότι:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i\right) \xrightarrow{p} E_\pi\left(h\left(X\right)\right) = I, \text{ καθώς } n \to \infty,$$

όπου συμβολίζεται με $\xrightarrow{p}$ η σύγκλιση κατά πιθανότητα, που ισοδύναμα σημαίνει ότι για κάθε θετικό αριθμό $\epsilon$:

$$\lim_{n\to\infty} P\left(|\hat{I} - I| > \epsilon\right) = 0.$$

Ερμηνεύοντας αυτό το αποτέλεσμα, συμπεραίνουμε ότι για οποιοδήποτε θετικό αριθμό $\epsilon$, ανεξάρτητα από το πόσο μικρός είναι, η δειγματική μέση τιμή $\hat{I}$ ενός επαρκώς (sufficiently) μεγάλου αριθμού παρατηρήσεων, $n$, έχει υψηλή πιθανότητα να είναι κοντά στην αναμενόμενη τιμή $I$ εντός του περιθωρίου $\epsilon$ που έχει καθοριστεί.

Επίσης από τον Ισχυρό Νόμο των Μεγάλων Αριθμών (ΙΝΜΑ) είναι γνωστό ότι η δειγματική μέση τιμή συγκλίνει σχεδόν βέβαια (almost surely) στην αναμενόμενη τιμή. Ειδικότερα, υπό την υπόθεση ότι η $E_\pi(h(X))$ υπάρχει και είναι πεπερασμένη, εφαρμόζοντας τον ΙΝΜΑ έχουμε ότι:

$$\hat{I}^{BMCI} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i\right) \xrightarrow{a.s.} E_\pi\left(h\left(X\right)\right) = I, \text{ καθώς } n \to \infty,$$

όπου συμβολίζεται με $\xrightarrow{a.s.}$ η σχεδόν βέβαιη σύγκλιση, που ισοδύναμα σημαίνει

ότι για κάθε θετικό αριθμό $\epsilon$:

$$P\left(\lim_{n\to\infty} |\hat{I}^{BMCI} - I| > \epsilon\right) = 0.$$

Ερμηνεύοντας αυτό το αποτέλεσμα, προκύπτει ότι με πιθανότητα 1 η ακολουθία των δειγματικών μέσων τιμών συγκλίνει, καθώς το $n$ μεγαλώνει, στη σταθερή τιμή $E_\pi\left(h\left(X\right)\right)$. Από τα παραπάνω συμπεραίνουμε ότι το ολοκλήρωμα $I$ μπορεί να εκτιμηθεί σε σημείο από το άθροισμα $\hat{I}^{BMCI}$.

Στη συνέχεια το ενδιαφέρον μας επικεντρώνεται στην κατασκευή ενός διαστήματος εμπιστοσύνης για το $I$, το οποίο μάλιστα θα είναι ένα ασυμπτωτικό διάστημα εμπιστοσύνης. Από το Κεντρικό Οριακό Θεώρημα (βλέπε Wasserman (2013)), υπό την υπόθεση ότι η $E_\pi(h(X))$ και η $Var_\pi(h(X))$ υπάρχουν και είναι πεπερασμένες, προκύπτει ότι καθώς το $n$ τείνει στο άπειρο:

$$\sqrt{n}\left(\hat{I}^{BMCI} - I\right) \xrightarrow{d} N\left(0, Var_\pi(h(X))\right),$$

όπου συμβολίζεται με $\xrightarrow{d}$ η σύγκλιση κατά κατανομή και με $N\left(\mu, \sigma^2\right)$ η κανονική κατανομή με μέση τιμή $\mu$ και διακύμανση $\sigma^2$. Καθώς συνήθως η διακύμανση $Var_\pi(h(X))$ είναι άγνωστη, προκύπτει με εφαρμογή του Θεωρήματος του Slutsky (βλέπε Van der Vaart (2000)) ότι:

$$\sqrt{n}\,\frac{\hat{I}^{BMCI} - I}{\hat{se}^{BMCI}} \xrightarrow{d} N\left(0, 1\right),$$

όπου με $\hat{se}^{BMCI}$ συμβολίζεται ένας συνεπής εκτιμητής της $\sqrt{Var_\pi(h(X))}$. Για παράδειγμα:

$$\hat{se}^{BMCI} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(h(X_i) - \hat{I}\right)^2}.$$

Από τα παραπάνω εύκολα προκύπτει ότι ένα $100(1-\alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το $I$ είναι το:

$$\left(\hat{I}^{BMCI} - z_{\alpha/2}\hat{se}^{BMCI}\sqrt{\frac{1}{n}},\ \hat{I}^{BMCI} + z_{\alpha/2}\hat{se}^{BMCI}\sqrt{\frac{1}{n}}\right),$$

όπου $z_\alpha$ είναι το σημείο εκείνο για το οποίο ισχύει $P(Z > z_\alpha) = \alpha$, με $Z \sim N(0,1)$.

Από το παραπάνω διάστημα εμπιστοσύνης είναι σαφές ότι η ακρίβεια της εκτίμησης του $I$ εξαρτάται από την ασυμπτωτική διασπορά και είναι γνωστό ότι

όσο μικρότερη ασυμπτωτική διασπορά έχει ένας εκτιμητής τόσο αποδοτικότερος είναι. Τα παραπάνω θα εξειδικευτούν μέσω ενός απλού παραδείγματος (βλέπε Wasserman (2013)).

**Παράδειγμα 2.0.1.** *Έστω $X_1, X_2, ..., X_n$ τ.δ. από έναν πληθυσμό που περιγράφεται ικανοποιητικά από την Ομοιόμορφη κατανομή $U(a, b)$. Το ενδιαφέρον μας επικεντρώνεται στην εκτίμηση σε σημείο και σε διάστημα του ολοκληρώματος $I = \int_a^b w(x)\,\mathrm{d}x$ (σχετικά με τη συνάρτηση $w(x)$ βλέπε εισαγωγή αυτού του κεφαλαίου).*

***Λύση***. Είναι $X_1, X_2, ..., X_n$ τ.δ. από τον πληθυσμό με σ.π.π. $\pi(x) = \dfrac{1}{b-a}$, $x \in (a, b)$. Θεωρούμε τη συνάρτηση $h(x) = w(x)(b-a)$, τότε έχουμε ότι $h(x)\pi(x) = w(x)$.

Επομένως, προκύπτει ότι $I = \int_a^b w(x)\,\mathrm{d}x = \int_a^b h(x)\pi(x)\,\mathrm{d}x$. Καθώς η $\pi(\cdot)$ είναι η συνάρτηση πυκνότητας πιθανότητας στο $(a, b)$, έχουμε ότι $I = E_\pi(h(X))$. Με εφαρμογή των όσων προηγήθηκαν προκύπτει ότι ένας σημειακός εκτιμητής του $I$ είναι ο:

$$\hat{I}^{BMCI} = \frac{1}{n}\sum_{i=1}^{n} h(X_i) = \frac{b-a}{n}\sum_{i=1}^{n} w(X_i) = (b-a)\bar{w},$$

με $\bar{w} = \dfrac{1}{n}\sum_{i=1}^{n} w(X_i)$. Επιπρόσθετα, ένα $100(1-\alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το $I$ είναι το:

$$\left(\hat{I}^{BMCI} - z_{\alpha/2}\frac{\hat{se}^{BMCI}}{\sqrt{n}},\ \hat{I}^{BMCI} + z_{\alpha/2}\frac{\hat{se}^{BMCI}}{\sqrt{n}}\right),$$

όπου

$$\hat{se}^{BMCI} = \sqrt{\frac{(b-a)^2}{n}\sum_{i=1}^{n}(w(X_i) - \bar{w})^2}.$$

□

Από τα παραπάνω και στη βάση του νόμου των μεγάλων αριθμών είναι προφανές ότι ο προσεγγιστικός υπολογισμός του αρχικού ολοκληρώματος $\int_{\mathcal{X}} w(x)\,\mathrm{d}x$ ή της

ισοδύναμης μέσης τιμής $I = \int_{\mathcal{X}} h(x)\pi(x)\,\mathrm{d}x$ οδηγεί σε μια μορφή στοχαστικής ολοκλήρωσης που απαιτεί και εμπλέκει τη δειγματοληψία, ήτοι την παραγωγή ενός τυχαίου δείγματος μεγέθους $n$, από τον πληθυσμό με σ.π.π. $\pi(\cdot)$. Τα παραπάνω αιτιολογούν πλήρως το γεγονός ότι τα δύο προβλήματα που αναφέρθηκαν στην αρχή είναι άμεσα συνδεδεμένα.

Για να αντιμετωπιστούν τα παραπάνω δύο προβλήματα έχει εμφανιστεί στη βιβλιογραφία μια πληθώρα τεχνικών. Οι τεχνικές αυτές χωρίζονται στις τεχνικές Monte Carlo και στις τεχνικές Markov Chain Monte Carlo. Οι πρώτες, όπως αναλυτικά θα δούμε στη συνέχεια, παράγουν ανεξάρτητες παρατηρήσεις είτε άμεσα από την κατανομή στόχο ή από κάποια διαφορετική κατανομή πρότασης, ενώ οι δεύτερες προσομοιώνουν Μαρκοβιανές αλυσίδες με οριακή κατανομή την κατανομή στόχο και επομένως οι παρατηρήσεις είναι εξαρτημένες. Στη συνέχεια αυτού του κεφαλαίου θα παρουσιαστούν οι πιο γνωστές τεχνικές Monte Carlo (MC), ενώ οι τεχνικές MCMC αποτελούν αντικείμενο μελέτης του επόμενου κεφαλαίου.

Η έκφραση μέθοδος Monte Carlo (MC) είναι πολύ γενική και περιλαμβάνει κυρίως υπολογιστικές τεχνικές για την επίλυση προβλημάτων μέσω της προσομοίωσης τυχαίων αριθμών. Ένας από τους πρώτους αλγορίθμους τύπου Monte Carlo αναπτύχθηκε από τον Buffon (1777) με σκοπό την προσέγγιση του αριθμού $\pi$. Όμως η συστηματική μελέτη των μεθόδων MC ξεκινά μετά την πρωτοπόρα εργασία των Metropolis and Ulam (1949) και στα πλαίσια της εργασίας των von Neumann, Ulam και Metropolis στο ερευνητικό πρόγραμμα Μανχάταν, στο Εθνικό Εργαστήριο του Los Alamos. Το πρόγραμμα αυτό ήταν σχετικό με τον σχεδιασμό πυρηνικών όπλων. Κατά μία εσφαλμένη εκδοχή οι μέθοδοι Monte Carlo οφείλουν το όνομά τους στην πόλη του Πριγκηπάτου του Μονακό, που είναι γνωστό για τα καζίνο του, μιας και μία από τις απλούστερες συσκευές παραγωγής τυχαίων αριθμών είναι η ρουλέτα. Στην πραγματικότητα, καθώς το έργο των von Neumann και Ulam στο Los Alamos απαιτούσε μυστικότητα, ο φίλος τους Metropolis πρότεινε τη χρήση του ονόματος Monte Carlo, το οποίο αναφέρεται στο καζίνο Monte Carlo του Μονακό, όπου ο θείος του Ulam δανείστηκε χρήματα από συγγενείς για να στοιχηματίσει (βλέπε Metropolis (1987)).

Τα τελευταία χρόνια η χρήση τέτοιων τεχνικών γνωρίζει ραγδαία εξέλιξη και έχουν γίνει πλέον πολύτιμες, καθώς δίνουν τη δυνατότητα να αντιμετωπιστεί ένα μεγάλο φάσμα προβλημάτων βελτιστοποίησης και ολοκλήρωσης. Επιπλέον, η κατασκευή των περιοχών αποδοχής ή απόρριψης κάποιων μηδενικών υποθέσεων στη βάση κάποιων στατιστικών τεστ μπορεί να επιτευχθεί μόνο με τεχνικές προσομοίωσης. Αυτό είχε ως αποτέλεσμα την επίλυση πληθώρας πραγματικών προβλημάτων και έπεισε τους επιστήμονες από άλλα πεδία, όπως για παράδειγμα

στη Φυσική, στη Θεωρία Παιγνίων, στα Οικονομικά και αλλού (βλέπε Gardiner (2009), Landau and Binder (2014)), ότι οι στατιστικές λύσεις ήταν πράγματι διαθέσιμες.

Παρότι στη βιβλιογραφία έχει εμφανιστεί πληθώρα ερευνητικών εργασιών, ενώ ταυτόχρονα η χρήση των μεθόδων Monte Carlo είναι απίστευτα ευρεία και έχει οδηγήσει σε μία σειρά πρωτοποριακών συμπερασμάτων σε διάφορους τομείς, δεν υπάρχει κοινή συναίνεση για έναν αυστηρό ορισμό των μεθόδων Monte Carlo. Αυτό ίσως οφείλεται και στη φύση και την ποικιλομορφία των μεθόδων αυτών που παράγουν ορισμούς μέσω παραδειγμάτων. Αν θα θέλαμε να δώσουμε έναν μη αυστηρό ορισμό θα λέγαμε ότι είναι μία μεγάλη κατηγορία υπολογιστικών μεθόδων, οι οποίες έχουν ως κοινό χαρακτηριστικό την επαναλαμβανόμενη τυχαία δειγματοληψία, τη χρήση τυχαίων αριθμών, ενώ επιπλέον διέπονται από συγκεκριμένες ιδιότητες (βλέπε Liu (2008) και Robert and Casella (2013)).

Στο πλαίσιο αυτού του κεφαλαίου σκοπός μας είναι να παραθέσουμε τις πιο σημαντικές και ευρέως χρησιμοποιούμενες μεθόδους Monte Carlo για την προσομοίωση από μια κατανομή με σ.π. ή σ.π.π. $\pi(x)$. Σε όσα ακολουθούν η κατανομή $\pi(\cdot)$ ονομάζεται **κατανομή στόχος** (target distribution). Καθώς οι μέθοδοι Monte Carlo προϋποθέτουν ότι μπορούμε να παράγουμε (ψευδο)-τυχαίους αριθμούς από την Ομοιόμορφη κατανομή στο διάστημα $(0, 1)$, προηγείται της παρουσίασης των μεθόδων Monte Carlo μια σύντομη παρουσίαση τεχνικών δημιουργίας ψευδοτυχαίων αριθμών από την $U(0, 1)$. Ο όρος ψευδο-τυχαίοι αριθμοί θα δικαιολογηθεί πλήρως στην επόμενη ενότητα.

## 2.1 Ψευδοτυχαίοι αριθμοί

Στην ενότητα αυτή το ενδιαφέρον επικεντρώνεται στον τρόπο δημιουργίας, γέννησης τυχαίων μεταβλητών που ακολουθούν ομοιόμορφη κατανομή στο διάστημα (0,1), καθώς όπως θα δούμε στο υπόλοιπο του κεφαλαίου, η δημιουργία τυχαίων μεταβλητών από όλες τις άλλες κατανομές προϋποθέτει μια τέτοια δημιουργία (βλέπε, μεταξύ άλλων, Ripley (1987)).

Η αναγκαιότητα ύπαρξης μεθόδων για τη δημιουργία μιας ακολουθίας από τυχαίους αριθμούς, δηλαδή από αριθμούς που είναι τέτοιοι ώστε όλοι οι αριθμοί της ακολουθίας να είναι ομοιόμορφα κατανεμημένοι στο πεδίο ορισμού τους και οι παραγόμενοι αριθμοί να είναι στατιστικά ανεξάρτητοι μεταξύ τους, έχει αναδειχθεί από τα πρώτα βήματα της ανάπτυξης της Στατιστικής. Για αυτό το λόγο έχουν προταθεί εδώ και πολλές δεκαετίες μερικές πολύπλοκες και ευφυείς μέθοδοι για την κατασκευή πινάκων τυχαίων αριθμών. Καθώς οι περισσότερες από

αυτές είναι πολύ αργές, αυτό οδήγησε στην εισαγωγή εναλλακτικών μεθόδων με τη βοήθεια του ηλεκτρονικού υπολογιστή, οι οποίες μπορούν να παράγουν μεγάλες σειρές φαινομενικά τυχαίων αποτελεσμάτων. Προηγούμενα χρησιμοποιήθηκε ο όρος φαινομενικά τυχαίων αποτελεσμάτων καθώς στην πραγματικότητα τα αποτελέσματα αυτά είναι πλήρως προκαθορισμένα από μια μικρότερη αρχική τιμή, που είναι γνωστή ως σπόρος (seed). Κάθε τέτοιος υπολογιστικός αλγόριθμος καλείται γεννήτρια ψευδοτυχαίων αριθμών ή ψευδοτυχαία γεννήτρια αριθμών (Pseudo-Random Number Generator, PRNG). Η ψευδοτυχαία γεννήτρια αριθμών έχει την ιδιότητα ότι παράγει αριθμούς που είναι όσο το δυνατόν πιο ανεξάρτητοι μεταξύ τους, είναι γρήγορη, δεν απαιτεί μεγάλη υπολογιστική ισχύ και τέλος έχει τη δυνατότητα να αναπαράγει την ίδια ακολουθία αριθμών, ώστε να μπορεί να επαναληφθεί η προσομοίωση (βλέπε Barker et al. (2012)). Συνοψίζοντας, οι ψευδοτυχαίοι αριθμοί παράγονται μέσω της υλοποίησης ορισμένων προσδιοριστικών επαναληπτικών διαδικασιών, γνωστές ως ντετερμινιστικές μέθοδοι, με αφετηρία κάποια αρχική τιμή. Στην περίπτωση των αλγορίθμων για την παραγωγή ψευδοτυχαίων αριθμών από την ομοιόμορφη στο (0,1) έχουμε τον ακόλουθο ορισμό (βλέπε, για παράδειγμα, Robert and Casella (1999)).

**Ορισμός 2.1.1.** *Μια γεννήτρια ομοιόμορφων ψευδοτυχαίων αριθμών είναι ένας αλγόριθμος ο οποίος ξεκινώντας από την αρχική τιμή $u_0$ και έναν μετασχηματισμό $D$ παράγει μια ακολουθία τιμών $(u_i) = (D_i(u_0))$, $i = 1, ..., n$, στο διάστημα $[0, 1]$, με τις τιμές $(u_1, ..., u_n)$ να προσομοιώνουν τη συμπεριφορά ενός τυχαίου δείγματος $(V_1, ..., V_n)$ ομοιόμορφων τυχαίων μεταβλητών.*

Είναι προφανές ότι η εγκυρότητα μιας γεννήτριας ομοιόμορφων ψευδοτυχαίων αριθμών ανάγεται στον έλεγχο της υπόθεσης ότι οι τιμές $u_1, ..., u_n$ αποτελούν ένα τυχαίο δείγμα από έναν πληθυσμό που περιγράφεται ικανοποιητικά από την ομοιόμορφη στο (0,1). Επομένως ανάγεται σε μια ομάδα ελέγχων που περιλαμβάνει, μεταξύ άλλων, έλεγχο καλής προσαρμογής, τυχαιότητας, ανεξαρτησίας και ασυσχέτιστου. Για περισσότερες πληροφορίες για το σύνολο των ελέγχων παραπέμπουμε στους Robert and Casella (1999) και τις εκεί αναφορές στις μεθόδους Kolmogorov-Smirnov, Die Hard Marsaglia καθώς και σε μεθόδους μη παραμετρικής στατιστικής και χρονικών σειρών.

Στη συνέχεια θα παρουσιαστούν κάποιες βασικές γεννήτριες ομοιόμορφων ψευδοτυχαίων αριθμών που παρότι δεν είναι ιδιαίτερα ισχυρές θα βοηθήσουν στην κατανόηση της φιλοσοφίας των ψευδοτυχαίων αριθμών, αλλά κυρίως αποτελούν ένα αναπόσπαστο κομμάτι ισχυρότερων. Όλες οι μέθοδοι που θα παρουσιαστούν σε αυτήν την ενότητα ανήκουν στην κατηγορία των αποκαλούμενων **συμπτωτικών μεθόδων** (congruential methods). Σε αυτές τις μεθόδους ο κάθε επόμενος αριθμός αποφασίζεται μέσω μίας απλής γραμμικής συνάρτησης, στην έξοδο

της οποίας τελικώς εφαρμόζεται μία αναγωγή στο διάστημα (0,1). Η αναγωγή αυτή επιτυγχάνεται με τη βοήθεια του τελεστή modulo και απλών μαθηματικών πράξεων.

Η πρώτη τέτοια μέθοδος που θα παρουσιαστεί είναι η λεγόμενη **πολλαπλασιαστική μέθοδος** (multiplicative congruential method). Η μέθοδος αυτή κατέχει εξέχουσα θέση μεταξύ των μεθόδων παραγωγής ψευδοτυχαίων αριθμών και χρησιμοποιείται πολύ συχνά στην πράξη. Το πρώτο βήμα είναι να θεωρήσουμε μια αρχική τιμή $x_0 \in \mathbb{N}$, η οποία θα μπορούσε να λαμβάνεται ας πούμε από τα εκατοστά του δευτερολέπτου που δείχνει το ρολόι του υπολογιστή. Τότε, θεωρώντας κάποιους προεπιλεγμένους φυσικούς αριθμούς $a$, $m$ (ο τρόπος επιλογής των οποίων θα σχολιαστεί αργότερα) η νέα ψευδοτυχαία τιμή υπολογίζεται από τη μαθηματική έκφραση:

$$x_1 = (ax_0)\, mod\, m,$$

όπου $x\, mod\, m$ ορίζεται να είναι το υπόλοιπο της διαίρεσης του $x$ δια του $m$. Με ανάλογο τρόπο προκύπτουν οι επόμενες ψευδοτυχαίες τιμές, δηλαδή $x_2 = (ax_1)\, mod\, m$, $x_3 = (ax_2)\, mod\, m$, κ.ο.κ..

Είναι προφανές ότι κάθε νέα ψευδοτυχαία τιμή $x_i$ λαμβάνει τιμές στο σύνολο $\{0, 1, ..., m-1\}$. Αυτό συνεπάγεται ότι ο αριθμός

$$U_i = \frac{x_i}{m} \in \left\{0, \frac{1}{m}, ..., \frac{m-1}{m}\right\} \subseteq [0, 1),\ i = 1, ..., n,$$

θεωρείται ένας ψευδοτυχαίος αριθμός και επίσης λαμβάνει τιμές μεταξύ του 0 και του 1.

Συνοψίζοντας, τα βήματα του πολλαπλασιαστικού συμπτωτικού αλγορίθμου (multiplicative congruential algorithm) είναι:

- Βήμα 0: Θέτουμε $x_0 = \text{seed}$.
- Βήμα $i$: Η ψευδοτυχαία τιμή $x_i$ υπολογίζεται από τη σχέση:

  $$x_i = (ax_{i-1})\, mod\, m,\ i = 1, 2, ..., n$$

  και έπειτα θέτουμε $U_i = \dfrac{x_i}{m}$, $i = 1, 2, ..., n$.

Άξιο επισήμανσης είναι ότι δίνεται ιδιαίτερη βαρύτητα στην επιλογή των $a$, $m$. Οι αριθμοί αυτοί επιλέγονται έτσι ώστε για κάθε αρχική τιμή $x_0$ το πλήθος των βημάτων που απαιτούνται μέχρι να φτάσουμε σε σημείο που η διαδικασία από εκεί και πέρα να επαναλαμβάνεται να είναι πολύ μεγάλο. Το γεγονός αυτό αιτιολογείται

από την παρατήρηση ότι οι αριθμοί $U_1, U_2, ...$ που προκύπτουν χρησιμοποιώντας τον παραπάνω αλγόριθμο δεν είναι στην πραγματικότητα τυχαίοι. Ειδικότερα, μετά από έναν πεπερασμένο αριθμό επαναλήψεων, που είναι μικρότερος ή ίσος του $m$, θα εμφανιστεί ένας αριθμός $x_k$ ο οποίος έχει ήδη εμφανιστεί. Δηλαδή, για $r < k$ θα ισχύει $x_k = x_r$ και επομένως από εκεί και πέρα η αρχική ακολουθία θα επαναληφθεί και έτσι $x_{k+1} = x_{r+1}$, $x_{k+2} = x_{r+2}$, κ.ο.κ..

Επιπλέον, η επιλογή των $a$ και $m$ θα πρέπει να είναι τέτοια ώστε οι αριθμοί $x_1, x_2, ..., x_n$ να παράγονται γρήγορα και εύκολα (βλέπε Cemgil (2012)). Τέλος, είναι αυτονόητο ότι για κάθε αρχική τιμή $x_0$, η ακολουθία των $x_1/m$, $x_2/m, ...$ πρέπει να μπορεί να θεωρηθεί ως μία πραγματοποίηση ακολουθίας ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν $U(0, 1)$.

Στο παραπάνω πλαίσιο, προτείνεται τα $x_0$ και $m$ να είναι πρώτοι αριθμοί μεταξύ τους, με το $m$ να είναι δύναμη του 2. Για περισσότερες προτάσεις σχετικά με την επιλογή των $a$, $m$ και $x_0$ που οδηγούν σε μεγαλύτερο κύκλο επανάληψης παραπέμπουμε, μεταξύ άλλων, στο σύγγραμμα του Knuth (1997) και τις εκεί αναφορές.

Μια άλλη μέθοδος παραγωγής ψευδοτυχαίων αριθμών είναι η λεγόμενη **μεικτή συμπτωτική μέθοδος** (mixed congruential method), η οποία είναι επίσης μια από τις συνηθέστερες μεθόδους. Ο αντίστοιχος αλγόριθμος είναι:

- Βήμα 0: Θέτουμε την αρχική τιμή $x_0 = \text{seed}$.
- Βήμα $i$: Η νέα ψευδοτυχαία τιμή $x_i$ υπολογίζεται από τη σχέση:

$$x_i = (ax_{i-1} + c) \, mod \, m, \; i = 1, 2, ..., n, \; \text{όπου } c \in \mathbb{N}_0$$

  και έπειτα θέτουμε $U_i = \dfrac{x_i}{m}$, $i = 1, 2, ..., n$.

Στα προηγούμενα $x_0 < m$ είναι ακέραιος, με $m > 0$, $a < m$, και $c < m$. Στο σημείο αυτό, και παρακινούμενοι από τους Johansen et al. (2007), αξίζει να επισημανθεί ότι η ακολουθία ψευδοτυχαίων αριθμών εξαρτάται από το σπόρο $x_0$ και τις τιμές των $a$, $c$ και $m$, αντίστοιχα. Παρατηρούμε ότι για $c = 0$ από τη μεικτή συμπτωτική μέθοδο προκύπτει η πολλαπλασιαστική. Για το λόγο αυτό στο επόμενο παράδειγμα που ακολουθεί, με στόχο την κατανόηση όσων προηγήθηκαν, περιοριζόμαστε στη μεικτή συμπτωτική μέθοδο, με σκοπό τη διευκρίνηση και την εφαρμογή όσων παρουσιάστηκαν.

**Παράδειγμα 2.1.1.** *Να υλοποιηθούν τα τέσσερα πρώτα βήματα της μεικτής συμπτωτικής μεθόδου θεωρώντας τις τιμές (βλέπε Johansen et al. (2007)):*

$$x_0 = 4, \; \alpha = 81, \; c = 35 \; \textit{και} \; m = 256.$$

***Λύση***. Αρχικά θέτουμε $x_0 = 4$ (από υπόθεση). Τότε, οι νέες ψευδοτυχαίες τιμές που προκύπτουν ακολουθώντας τα βήματα της μεικτής συμπτωτικής μεθόδου είναι:

$$\begin{aligned} x_1 = (\alpha x_0 + c)\, mod\, m &= (81 \cdot 4 + 35)\, mod\, 256 = 359\, mod\, 256 = 103, \\ x_2 = (\alpha x_1 + c)\, mod\, m &= (81 \cdot 103 + 35)\, mod\, 256 = 8378\, mod\, 256 = 186, \\ x_3 = (\alpha x_2 + c)\, mod\, m &= (81 \cdot 186 + 35)\, mod\, 256 = 15101\, mod\, 256 = 253. \end{aligned}$$

Αντίστοιχα, οι τιμές $U_i = \dfrac{x_i}{m}$, $i = 1, 2, 3$, που προκύπτουν είναι οι εξής:

$$U_1 = \frac{x_1}{m} = \frac{103}{256} = 0.402,\ U_2 = \frac{x_2}{m} = \frac{186}{256} = 0.726,\ U_3 = \frac{x_3}{m} = \frac{253}{256} = 0.988.$$

Τα παραπάνω υλοποιούνται στην R ορίζοντας την ακόλουθη συνάρτηση:

```
mcg = function(a,c,m,run.length,seed) {
  x = rep(0,run.length)
  x[1] = seed
  for (i in 1:(run.length-1)) {
    x[i+1] = (a * x[i] + c) %% m
  }
U = x/m
  return(list(x=x,U=U))
}
```

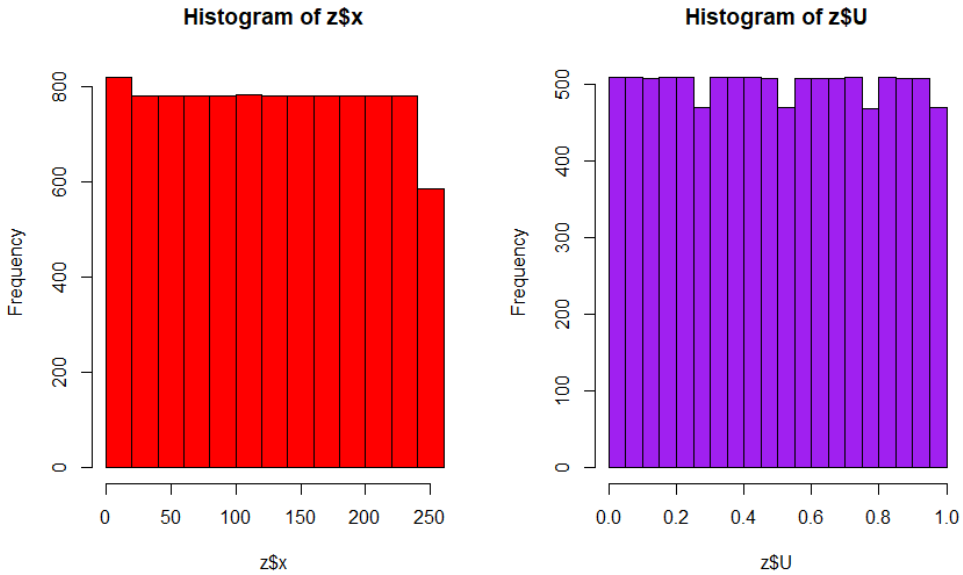
και εκτελώντας την παρακάτω εντολή.

```
z = mcg(81,35,256,20,4)
z
```

Οι τιμές που περιέχονται στα $z\$x$ είναι οι αριθμοί που παράγει η μεικτή συμπτωτική μέθοδος και οι τιμές που περιέχονται στα $z\$U$ είναι οι τιμές $z\$x$ που τυποποιούνται ώστε να εμπίπτουν στο διάστημα $[0, 1)$. Τα $z\$x$ και τα $z\$U$ κατανέμονται ομοιόμορφα, αν και το μοτίβο γίνεται πιο ξεκάθαρο και εμφανές όσο δημιουργούνται όλο και περισσότεροι τυχαίοι αριθμοί στην ακολουθία. Στο Σχήμα 2.1 βλέπουμε πώς μοιάζουν τα ιστογράμματα συχνοτήτων αν παράγουμε 10.000 τυχαίους αριθμούς. Πιο συγκεκριμένα, η βάση των ορθογωνίων παριστάνει το μήκος των ομάδων και το ύψος των ορθογωνίων την αντίστοιχη συχνότητα των ομάδων. Ακολούθως παρατίθενται οι εντολές που χρησιμοποιήθηκαν:

```
z = lcg(81,35,256,10000,4)
par(mfrow=c(1,2))
hist(z$x,col="red")
hist(z$U,col="purple")
```

Σχήμα 2.1: Ιστόγραμμα 10.000 τιμών της μεικτής συμπτωτικής μεθόδου και των αντίστοιχων ψευδοτυχαίων $U(0,1)$, με $\alpha = 81$, $c = 35$, $m = 256$ και $x_0 = 4$.




Προφανώς, εκτελώντας τη ψευδοτυχαία γεννήτρια αριθμών δύο φορές με τον ίδιο σπόρο δημιουργείται ακριβώς η ίδια ακολουθία ψευδοτυχαίων αριθμών. Επίσης, για $c = 0$, όπως έχει ήδη αναφερθεί, προκύπτει η πολλαπλασιαστική μέθοδος.

Τέλος, έχει προταθεί και η **προσθετική συμπτωτική μέθοδος** (additive congruential method), η οποία αποτελείται από τα ακόλουθα βήματα:

- Βήμα 0: Θέτουμε την αρχική τιμή $x_0 = \text{seed}$.
- Βήμα $i$: Έπειτα, $x_1 = x_0 \, mod \, m$ και

$$x_{i+1} = (x_i + x_{i-1}) \, mod \, m, \; i = 2, ..., n-1$$

και έπειτα θέτουμε $U_i = \dfrac{x_i}{m}$, $i = 1, 2, ..., n$.

Τα παραπάνω γίνονται πιο κατανοητά μέσω ενός απλού παραδείγματος που ακολουθεί.

**Παράδειγμα 2.1.2.** *Να εφαρμοσθούν πέντε βήματα της προσθετικής συμπτωτικής μεθόδου θεωρώντας τις τιμές* $x_0 = 100$ *και* $m = 128$.

***Λύση***. Από την υπόθεση ισχύει ότι ο σπόρος $x_0$ ισούται με 100. Όπως γίνεται εύκολα αντιληπτό, εφαρμόζοντας την προσθετική συμπτωτική μέθοδο, η αμέσως επόμενη τιμή δίνεται ως:

$$x_1 = x_0 \, mod \, m = 100 \, mod \, 128 \Rightarrow x_1 = 100.$$

Τότε, οι νέες ψευδοτυχαίες τιμές που προκύπτουν ακολουθώντας τα βήματα του παραπάνω αλγορίθμου είναι:

$$\begin{aligned} x_2 = (x_1 + x_0) \, mod \, m &= (100 + 100) \, mod \, 128 = 200 \, mod \, 128 = 72, \\ x_3 = (x_2 + x_1) \, mod \, m &= (72 + 100) \, mod \, 128 = 172 \, mod \, 128 = 44, \\ x_4 = (x_3 + x_2) \, mod \, m &= (44 + 72) \, mod \, 128 = 116 \, mod \, 128 = 116. \end{aligned}$$

Επομένως, οι τυποποιημένες τιμές $U_i = \dfrac{x_i}{m}$ που προκύπτουν είναι οι ακόλουθες:

$$U_1 = \frac{x_1}{m} = \frac{100}{128} = 0.781, \; U_2 = \frac{x_2}{m} = \frac{72}{128} = 0.562, \; U_3 = \frac{x_3}{m} = \frac{44}{128} = 0.343$$

και $U_4 = \dfrac{x_3}{m} = \dfrac{44}{128} = 0.906$, αντίστοιχα.

Τα παραπάνω υλοποιούνται στην R ορίζοντας την ακόλουθη συνάρτηση:

```
acg = function(m,run.length,seed) {
  x = rep(0,run.length)
  x[1] = seed
  x[2] = x[1] %% m
  for (i in 2:(run.length-1)) {
    x[i+1] = ( x[i] + x[i-1] ) %% m
  }
  U = x/m
  return(list(x=x,U=U))
}
```

και εκτελώντας την εντολή:

```
z = acg(128,50,100)
z
```

Στο σημείο αυτό επισημαίνεται ότι ο προεπιλεγμένος φυσικός αριθμός $m$ προτείνεται να είναι δύναμη του 2. Το γεγονός ότι το $m = 128 = 2^7$ αποτελεί και έναν από τους λόγους της επιλογής του αριθμού αυτού για το παράδειγμα. Ας σημειωθεί ακόμη πως οι ψευδοτυχαίες τιμές $x_i$ λαμβάνουν τιμές στο σύνολο $\{0, 1, ..., m - 1 = 127\}$. Αυτό σημαίνει πως επαληθεύεται ότι οι αντίστοιχες τυποποιημένες ψευδοτυχαίες τιμές κυμαίνονται μεταξύ του 0 και του 1. Δηλαδή:

$$U_i = \frac{x_i}{128} \in \left\{0, \frac{1}{128}, \frac{2}{128}, ..., \frac{127}{128}\right\} \subseteq [0, 1),\ i = 1, ..., n.$$

Με αφορμή το παραπάνω παράδειγμα αξίζει να αναφερθεί ότι στην προσθετική συμπτωτική μέθοδο όταν $x_0 < m$ η αρχική τιμή $x_0$ ταυτίζεται με την αμέσως επόμενη τιμή $x_1$, καθώς τότε ισχύει ότι $x_0 \, mod \, m = x_0$.

Συνοψίζοντας το περιεχόμενο αυτής της ενότητας μπορούμε να πούμε ότι παρόλο που οι ψευδοτυχαίες γεννήτριες αριθμών δε μας δίνουν ακολουθίες που είναι ακολουθίες πραγματικά τυχαίων αριθμών, μας εφοδιάζουν με ακολουθίες που μπορούν να θεωρηθούν ως τέτοιες και επιπλέον δημιουργούνται σε σύντομο χρονικό διάστημα. Επίσης, μπορούν να αναπαραχθούν αργότερα αν είναι γνωστό το αρχικό σημείο της ακολουθίας. Οι μέθοδοι παραγωγής ψευδοτυχαίων αριθμών και η σύγκρισή τους θα μπορούσε να αποτελούν ένα αυτοτελές αντικείμενο μελέτης. Στην ενότητα αυτή έγινε μια μικρή μνεία, καθώς, όπως ήδη αναφέρθηκε, η παραγωγή (ψευδο)τυχαίων αριθμών από την $U(0, 1)$ είναι αναπόσπαστο κομμάτι της προσομοίωσης από κάθε κατανομή, κάτι που θα γίνει σαφές στις επόμενες ενότητες.

## 2.2 Μέθοδος αντιστροφής

Στην προηγούμενη ενότητα αυτού του κεφαλαίου παρουσιάστηκαν απλοί τρόποι με τους οποίους μπορούμε να δημιουργήσουμε, να παράγουμε (ψευδο)τυχαίους αριθμούς από την ομοιόμορφη κατανομή στο $(0, 1)$. Στην ενότητα αυτή θα παρουσιαστεί η μέθοδος της **αντιστροφής** (Inversion Sampling), η οποία είναι

και η πιο απλή μέθοδος δημιουργίας αριθμών από μια κατανομή με αθροιστική συνάρτηση κατανομής (α.σ.κ.) $F(\cdot)$.

Ειδικότερα, έστω $X$ μια τυχαία μεταβλητή με α.σ.κ. $F(\cdot)$, δηλαδή $F(x) = P(X \leq x)$, για κάθε $x \in \mathbb{R}$. Η μέθοδος της αντιστροφής βασίζεται στην αντιστροφή της α.σ.κ., γεγονός που δικαιολογεί και την ονομασία της. Εύλογα προκύπτει το ερώτημα πώς μπορεί να χρησιμοποιηθεί η έννοια της αντίστροφης συνάρτησης, όταν αυτή, όπως καλά γνωρίζουμε, έχει νόημα μόνο όταν η συνάρτηση είναι $1-1$. Θυμηθείτε ότι εξ' ορισμού η α.σ.κ. είναι πάντοτε συνεχής από δεξιά, αλλά είναι συνεχής παντού μόνο στην περίπτωση των συνεχών τυχαίων μεταβλητών (βλέπε, μεταξύ άλλων, Casella and Berger (2002)), ενώ είναι αύξουσα αλλά όχι απαραίτητα γνησίως αύξουσα. Το πρόβλημα αυτό μπορεί να ξεπεραστεί καθώς μας αρκεί η έννοια της γενικευμένης αντίστροφης συνάρτησης (βλέπε, για παράδειγμα, Casella and Berger (2002)), ο ορισμός της οποίας ακολουθεί.

**Ορισμός 2.2.1.** *Έστω $F : \mathbb{R} \to \mathbb{R}$ μία αύξουσα συνάρτηση. Η γενικευμένη αντίστροφη συνάρτηση (generalized inverse function) της $F(\cdot)$ συμβολίζεται με $F^{-}(u)$ και ορίζεται ως:*

$$F^{-}(u) = inf\left\{x : F(x) \geq u\right\} = inf\left\{x : F(x) \in [u,1]\right\}, u \in [0,1]. \tag{2.2}$$

Παρατηρούμε ότι όταν η α.σ.κ. $F(\cdot)$ είναι αντιστρέψιμη (για παράδειγμα συνεχής και γνησίως αύξουσα) τότε η γενικευμένη αντίστροφη συνάρτηση ταυτίζεται με την κλασική αντίστροφη συνάρτηση, δηλαδή $F^{-}(u) = F^{-1}(u)$. Επιπλέον, ο ορισμός της γενικευμένης αντίστροφης συνάρτησης καλύπτει την περίπτωση όπου η α.σ.κ. $F(\cdot)$ δεν είναι αντιστρέψιμη. Κάτι τέτοιο συμβαίνει για παράδειγμα όταν η τ.μ. $X$ είναι διακριτή ή όταν είναι συνεχής αλλά η $F(\cdot)$ δεν είναι γνησίως αύξουσα. Τέλος, καθώς η $F(\cdot)$ είναι πάντοτε δεξιά συνεχής, το infimum θα επιτυγχάνεται εντός του συνόλου $\{x : F(x) \in [u,1]\}$. Έτσι (βλέπε, για παράδειγμα, Sigman (2010)) το infimum ταυτίζεται με το minimum. Επομένως, η σχέση (2.2) ισοδύναμα γράφεται ως:

$$F^{-}(u) = min\left\{x : F(x) \geq u\right\} = min\left\{x : F(x) \in [u,1]\right\}, u \in [0,1].$$

Στην πρόταση που ακολουθεί δίνεται η κεντρική ιδέα της μεθόδου της αντιστροφής (βλέπε, μεταξύ άλλων, Robert and Casella (1999)).

**Πρόταση 2.2.1.** *Αν $U$ είναι μία τ.μ. τέτοια ώστε $U \sim U(0,1)$, τότε η τ.μ. $X = F^{-}(U)$, όπου $F^{-}$ είναι η γενικευμένη αντίστροφη συνάρτηση οποιασδήποτε αθροιστικής συνάρτησης κατανομής, έχει συνάρτηση κατανομής την $F(\cdot)$.*

**Απόδειξη.** Η α.σ.κ. της τ.μ. $X = F^{-}(U)$ δίνεται ως

$$P(X \leq x) = P(F^{-}(U) \leq x) = P\left(inf\left\{t : F(t) \geq U\right\} \leq x\right).$$

Καθώς όμως η $F(\cdot)$ είναι αύξουσα προκύπτει ισοδύναμα ότι:

$$P\left(F^{-}(U) \leq x\right) = P\left(min\left\{t : F(t) \geq U\right\} \leq x\right).$$

Επομένως, αρκεί να δείξουμε ότι:

$$P\left(min\left\{t : F(t) \geq U\right\} \leq x\right) = F(x),$$

ή ισοδύναμα, λαμβάνοντας υπόψη ότι $U \sim U(0,1)$ και $F(x) \in [0,1]$, ότι:

$$P\left(min\left\{t : F(t) \geq U\right\} \leq x\right) = P(U \leq F(x)).$$

Δηλαδή αρκεί να δείξουμε ότι:

$$min\left\{t : F(t) \geq U\right\} \leq x \Leftrightarrow U \leq F(x).$$

Παρατηρούμε ότι αν

$$U \leq F(x) \Rightarrow x \in \left\{t : F(t) \geq U\right\} \Rightarrow x \geq min\left\{t : F(t) \geq U\right\}.$$

Από την άλλη μεριά, αν $min\left\{t : F(t) \geq U\right\} \leq x$ τότε υπάρχει $t_0 \leq x$ τέτοιο ώστε $F(t_0) \geq U$. Καθώς όμως η $F(\cdot)$ είναι αύξουσα έχουμε ότι $t_0 \leq x$ συνεπάγεται ότι $F(t_0) \leq F(x)$. Συνδυάζοντας τα παραπάνω έχουμε, λοιπόν, ότι $U \leq F(t_0) \leq F(x)$.

Συνεπώς, καταλήγουμε ότι $P\left(X \leq x\right) = P\left(U \leq F(x)\right) = F(x)$, δηλαδή, η τ.μ. $X = F^{-}(U)$ έχει α.σ.κ. την επιθυμητή $F$ και έτσι η απόδειξη ολοκληρώθηκε. □

Από όσα προαναφέρθηκαν γίνεται αντιληπτό ότι για να προσομοιωθεί μία τιμή, έστω $x$, από την τ.μ. $X$ με α.σ.κ. $F(\cdot)$ αρκεί να παραχθεί αρχικά μια τιμή $u \in (0,1)$ από την τ.μ. $U \sim U(0,1)$ και έπειτα να χρησιμοποιηθεί ο μετασχηματισμός $x = F^{-}(u)$, όπου $F^{-}$ είναι η γενικευμένη αντίστροφη συνάρτηση της $F$ που δίνεται από τη σχέση (2.2). Επομένως, έχουμε τα ακόλουθα (βλέπε, μεταξύ άλλων, Liang et al. (2011)):

### Αλγόριθμος αντιστροφής (συνεχής περίπτωση)

1. Έστω $U$ από την ομοιόμορφη στο $(0,1)$.
2. Υπολόγισε το $X = F^{-1}(U)$.
3. Η $X$ είναι πραγματοποίηση από την κατανομή με α.σ.κ. $F(\cdot)$.

**Αλγόριθμος αντιστροφής (διακριτή περίπτωση)**

1. Έστω $U$ από την ομοιόμορφη στο $(0,1)$.
2. Προσδιόρισε την τιμή $X$ που είναι τέτοια ώστε $F(X-1) < U \leq F(X)$.
3. Η $X$ είναι πραγματοποίηση από την κατανομή με α.σ.κ. $F(\cdot)$.

Εναλλακτικά όταν θέλουμε να δημιουργήσουμε τ.δ. από έναν πληθυσμό που περιγράφεται από τη συνάρτηση πιθανότητας (σ.π.) $p_i = P(X = x_i)$, $i = 0, 1, 2...$, με $\sum_i p_i = 1$, έχουμε ότι:

1. Έστω $U$ από την ομοιόμορφη στο (0,1).
2. Τότε

$$X = \begin{cases} x_0, & \text{αν } 0 \leq U < p_0 \\ x_1, & \text{αν } p_0 \leq U < p_0 + p_1 \\ . & \\ . & \\ . & \\ x_k, & \text{αν } \sum_{i=0}^{k-1} p_i \leq U < \sum_{i=0}^{k} p_i. \\ . & \\ . & \\ . & \end{cases}$$

Στη συνέχεια, για την καλύτερη κατανόηση όσων προηγήθηκαν, θα δούμε πώς μπορούμε χρησιμοποιώντας τη μέθοδο της αντιστροφής να παράγουμε τυχαίους αριθμούς από ορισμένες γνωστές συνεχείς και διακριτές κατανομές με μόνη προϋπόθεση την παραγωγή (ψευδο)τυχαίων αριθμών από την $U(0,1)$ (βλέπε, μεταξύ άλλων, Sigman (2010) και Cemgil (2012)).

**Παράδειγμα 2.2.2.** *Χρησιμοποιώντας τη μέθοδο της αντιστροφής να προταθεί ένας τρόπος παραγωγής μιας παρατήρησης από την ομοιόμορφη κατανομή στο διάστημα* $(a, b)$.

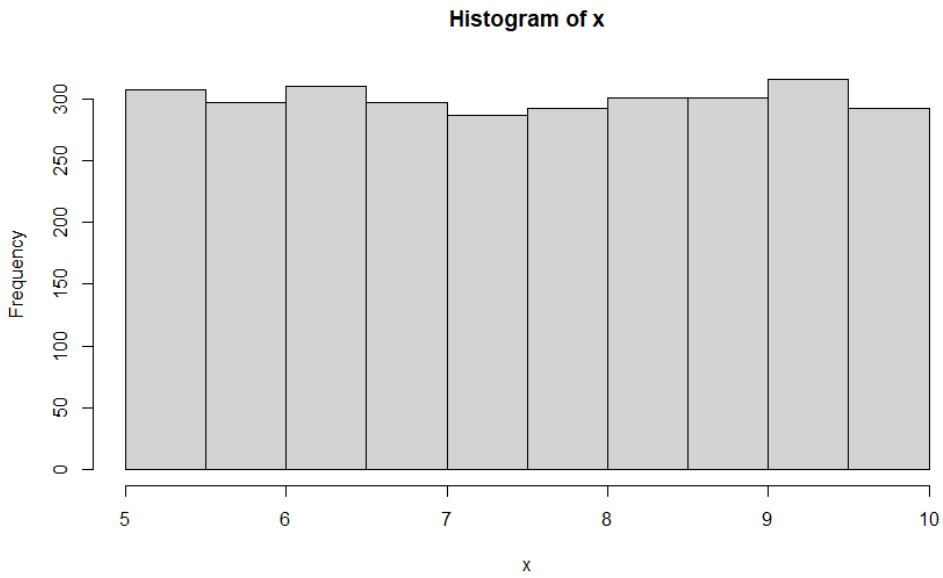
***Λύση***. Η α.σ.κ. της τ.μ. $X$ είναι: $F_X(x) = \dfrac{x-a}{b-a}$, για $x \in [a, b)$, $F_X(x) = 0$, $x < a$ και $F_X(x) = 1$, $x \geq b$. Λύνοντας την εξίσωση $u = \dfrac{x-a}{b-a}$ ως προς $x$

συνεπάγεται ότι $x = F_X^{-1}(u) = (b-a)u + a$, $u \in (0,1)$. Συμπερασματικά αν $U \sim U(0,1)$, τότε η:

$$X = a + U(b-a) \text{ ακολουθεί } U(a,b).$$

Στη συνέχεια τα παραπάνω υλοποιούνται στην $R$ για την ειδική περίπτωση της δημιουργίας 3.000 τυχαίων αριθμών από την $U(5,10)$. Στο Σχήμα 2.2 δίνεται το ιστόγραμμα αυτών των τυχαίων αριθμών.

Σχήμα 2.2: Ιστόγραμμα 3.000 τυχαίων αριθμών από την $U(5,10)$. Οι τυχαίοι αριθμοί δημιουργήθηκαν με τη μέθοδο της αντιστροφής.




```
u = runif(3000)
x = 5 + u*(10-5)
x
hist(x)
```

**Παράδειγμα 2.2.3.** *Χρησιμοποιώντας τη μέθοδο της αντιστροφής να παράγετε μια παρατήρηση από την Εκθετική κατανομή με παράμετρο* $\lambda > 0$.

***Λύση***. Στην περίπτωση αυτή έχουμε ότι η α.σ.κ. της τ.μ. $X$ δίνεται ως $F_X(x) = 1 - e^{-\lambda x}$, $x \geq 0$, ενώ $F_X(x) = 0$, $x < 0$. Λύνοντας την εξίσωση $u = 1 - e^{-\lambda x}$ ως προς $x$ και μετά από απλές αλγεβρικές πράξεις συνεπάγεται ότι $x = F_X^{-1}(u) = -\frac{1}{\lambda} ln(1-u)$, $u \in (0,1)$. Συνεπώς,

$$X = -\frac{1}{\lambda} ln(1-U),\ U \in (0,1). \tag{2.3}$$

Γνωρίζουμε όμως (βλέπε, για παράδειγμα, Casella and Berger (2002)) ότι αν η τ.μ. $U \sim U(0,1)$ τότε και η τ.μ. $1-U \sim U(0,1)$. Αυτό σημαίνει ότι ο αλγόριθμος για την παραγωγή τυχαίων αριθμών από την Εκθετική κατανομή με παράμετρο $\lambda$ μπορεί να απλοποιηθεί ως εξής:

- Βήμα 1: Παράγουμε την τ.μ. $U \sim U(0,1)$.
- Βήμα 2: Θέτουμε $X = -\frac{1}{\lambda} ln(U)$.

Στη συνέχεια τα παραπάνω υλοποιούνται στην $R$ για την ειδική περίπτωση της δημιουργίας 1.000 τυχαίων αριθμών από την Εκθετική κατανομή με $\lambda = 5$, ενώ στο Σχήμα 2.3 δίνεται το ιστόγραμμα αυτών των τυχαίων αριθμών.

```
u = runif(1000)
x = -log(1-u)/5
x
hist(x)
```

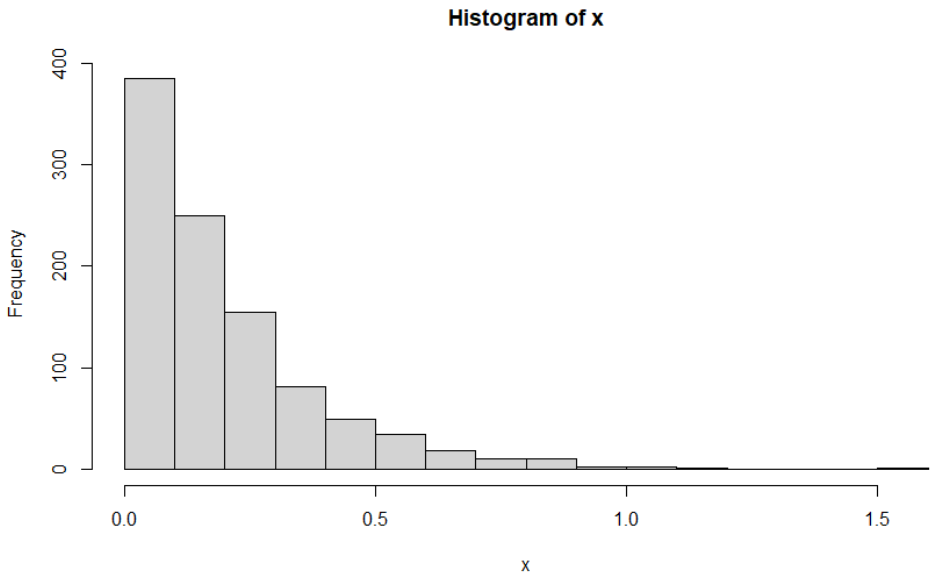
**Παράδειγμα 2.2.4.** *Να παράγετε μια παρατήρηση $X$ από την κατανομή Bernoulli($p$), με συνάρτηση πιθανότητας $P_X(X=0) = 1-p$ και $P_X(X=1) = p$, για κάποιο $p \in (0,1)$.*

***Λύση***. Εφαρμόζοντας τη διακριτή μέθοδο της αντιστροφής προκύπτουν τα ακόλουθα βήματα:

- Βήμα 1: Παράγουμε την τ.μ. $U \sim U(0,1)$.
- Βήμα 2: Έπειτα, θέτουμε

$$X = \begin{cases} 0, & \text{αν } 0 \leq U < 1-p, \\ 1, & \text{αν } 1-p \leq U. \end{cases}$$

Σχήμα 2.3: Ιστόγραμμα 1.000 τυχαίων αριθμών από την Εκθετική κατανομή με $\lambda = 5$. Οι τυχαίοι αριθμοί δημιουργήθηκαν με τη μέθοδο της αντιστροφής.




Στη συνέχεια τα παραπάνω υλοποιούνται στην $R$ για την ειδική περίπτωση της δημιουργίας 7.000 τυχαίων αριθμών από την Bernoulli με $p = 0.3$.

```
u = runif(7000)
p = 0.3
X = ifelse((1-p<=u), 1, 0)
X
```

Στα προηγούμενα παραδείγματα η μέθοδος της αντιστροφής ήταν ένα πολύ βολικό εργαλείο για την παραγωγή αριθμών από τις συγκεκριμένες κατανομές, καθώς η (γενικευμένη) αντίστροφη της αθροιστικής συνάρτησης κατανομής προσδιορίστηκε εύκολα. Με παρόμοιο τρόπο, η παραγωγή τυχαίων αριθμών με τη μέθοδο της αντιστροφής είναι εφικτή για τις Weibull (α, β), Pareto και Cauchy. Ωστόσο, σε πολλές περιπτώσεις δεν είναι εύκολος ή ακόμη και εφικτός ο

προσδιορισμός της αντίστροφης $F^{-}$. Αυτή είναι η κύρια αιτία που η μέθοδος της αντιστροφής εφαρμόζεται σε λίγες περιπτώσεις, ενώ ακόμα και σε περιπτώσεις που η μέθοδος δύναται να εφαρμοστεί μπορεί να επιλέγεται κάποια άλλη λόγω υπολογιστικού χρόνου, κόστους. Τέτοιες περιπτώσεις αποτελούν για παράδειγμα η προσομοίωση από την κανονική, τη Βήτα, τη Γάμμα και την κατανομή F. Ειδικότερα, στην περίπτωση που η τ.μ. $X$ ακολουθεί την κατανομή Γάμμα με παραμέτρους $a, b > 0$, τότε έχει α.σ.κ.:

$$F_X(x) = \int_0^x \frac{e^{-y/b}\, y^{a-1}}{b^a\, \Gamma(a)}\, \mathrm{d}y,\ x > 0.$$

Παρόλο που η $F_X(\cdot)$ είναι αντιστρέψιμη, η αντίστροφή της δεν μπορεί να παρασταθεί σε μια κλειστή αναλυτική μορφή, διότι το παραπάνω ολοκλήρωμα γενικά δεν υπολογίζεται αναλυτικά. Ένα ερώτημα που προκύπτει είναι πώς θα αντιμετωπιστούν τέτοιες περιπτώσεις. Την απάντηση μας την δίνει εν μέρει η ίδια η μέθοδος της αντιστροφής. Η ιδέα είναι να προσομοιώσουμε τυχαία δείγματα από μια διαφορετική κατανομή (την ομοιόμορφη στη μέθοδο της αντιστροφής) για την οποία είναι εφικτή η προσομοίωση και χρησιμοποιώντας κάποιο μετασχηματισμό (τη γενικευμένη αντίστροφη της α.σ.κ. στο πλαίσιο της μεθόδου της αντιστροφής) ή κάποιες γνωστές ιδιότητες να οδηγηθούμε στην κατανομή στόχο. Το παραπάνω σκεπτικό θα γίνει κατανοητό στα επόμενα παραδείγματα (βλέπε, μεταξύ άλλων, Robert and Casella (1999) και Sigman (2010)).

**Παράδειγμα 2.2.5.** *Χρησιμοποιώντας τη μέθοδο της αντιστροφής να προσομοιωθούν τυχαίοι αριθμοί από την κατανομή Γάμμα, με παραμέτρους a και b, με το a να είναι φυσικός αριθμός.*

***Λύση***. Γνωρίζουμε ότι αν $X \sim G(a, b)$, με $a$ ακέραιο τότε $X = \sum_{i=1}^{a} Y_i$, όπου $Y_i$ ακολουθούν Εκθετική κατανομή με παράμετρο $\frac{1}{b}$. Επομένως αφού μπορούμε να προσομοιώσουμε από την Εκθετική μπορούμε και από αυτήν την ειδική περίπτωση της Γάμμα. Επομένως, αν $U_i \sim U(0,1), i = 1, 2, ..., a$, είναι ανεξάρτητες τ.μ., τότε (βλέπε σχέση (2.3)) οι τ.μ. $Y_i = -b\, ln(U_i) \sim Exp\left(\frac{1}{b}\right)$, $i = 1, 2, ..., a$ και τελικά η τ.μ.

$$X = \sum_{i=1}^{a} Y_i = -\sum_{i=1}^{a} b\, ln(U_i) = -b\, ln\left(\prod_{i=1}^{a} U_i\right) \sim G(a, b),\ \text{με } a \text{ ακέραιο.}$$

Συνοψίζοντας, ο αλγόριθμος για την παραγωγή ενός τυχαίου αριθμού από την κατανομή $G(a, b)$, με $a$ ακέραιο είναι:

- Βήμα 1: Παράγουμε $a$ τυχαίους αριθμούς $U_1, U_2, ..., U_a \sim U(0,1)$.
- Βήμα 2: Θέτουμε $X = -b\,ln(U_1 \cdot U_2 \cdot ... \cdot U_a)$.

**Παράδειγμα 2.2.6.** *Χρησιμοποιώντας τη μέθοδο της αντιστροφής να προταθεί ένας τρόπος προσομοίωσης τυχαίων αριθμών από τη διωνυμική κατανομή $B(n,p)$.*

***Λύση***. Μπορούμε να χρησιμοποιήσουμε τη σύνδεση της $B(n,p)$ με την Bernoulli κατανομή. Ειδικότερα, η τ.μ. $X \sim B(n,p)$ μπορεί να αναπαρασταθεί ως $X = \sum_{i=1}^{n} Y_i$ όπου $Y_1, Y_2, ..., Y_n$ είναι $n$ το πλήθος ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με $Y_i$ να ακολουθούν $Bernoulli(p)$. Επομένως, αντί να χρησιμοποιηθεί άμεσα η μέθοδος της αντιστροφής θα έχουμε τον ακόλουθο αλγόριθμο:

- Βήμα 1: Παράγουμε $n$ ανεξάρτητες και ισόνομες τ.μ. $U_1, U_2, ..., U_n \sim U(0,1)$.
- Βήμα 2: Στη συνέχεια, για κάθε $1 \leq i \leq n$ δημιουργούμε $n$ ανεξάρτητες και ισόνομες τ.μ. $Bernoulli(p)$ θέτοντας

$$Y_i = \begin{cases} 0, & \text{αν } 0 \leq U_i < 1-p, \\ 1, & \text{αν } 1-p \leq U_i \leq 1. \end{cases}$$

- Βήμα 3: Τέλος, η $X = \sum_{i=1}^{n} Y_i$.

Το πλεονέκτημα αυτού του αλγορίθμου είναι η ευκολία και η απλότητά του, καθώς δε χρειάζεται να κάνουμε τους διάφορους υπολογισμούς που αφορούν τη συνάρτηση πιθανότητας της διωνυμικής. Από την άλλη πλευρά, αυτός ο αλγόριθμος απαιτεί την παραγωγή $n$ το πλήθος ομοιόμορφων τ.μ. για κάθε επανάληψη του $X$, έναντι μόνο μιας ομοιόμορφης τ.μ. όταν χρησιμοποιείται άμεσα η μέθοδος της αντιστροφής.

**Παράδειγμα 2.2.7.** *Να παράγετε μια παρατήρηση από την κατανομή $N(0,1)$, χρησιμοποιώντας ιδιότητές της που τη συνδέουν με την ομοιόμορφη κατανομή.*

***Λύση***. Οι κανονικές τυχαίες μεταβλητές μπορούν να δημιουργηθούν χρησιμοποιώντας τη μέθοδο της αντιστροφής, αλλά αυτό απαιτεί τον υπολογισμό της

αντίστροφης α.σ.κ. της τυπικής κανονικής, που είναι ένας μη τετριμμένος υπολογισμός. Εναλλακτικά μπορεί κάποιος να αξιοποιήσει ιδιότητες της κανονικής κατανομής. Ειδικότερα, είναι γνωστό ότι (βλέπε, μεταξύ άλλων, Robert and Casella (1999)) αν $U_i \sim U(0,1)$, $i = 1,2$, τότε οι τυχαίες μεταβλητές $X_1 = \sqrt{-2ln(U_1)}\,cos(2\pi U_2)$ και $X_2 = \sqrt{-2ln(U_1)}\,sin(2\pi U_2)$ ακολουθούν τυπική κανονική κατανομή και μάλιστα είναι ανεξάρτητες τ.μ.. Η παραπάνω μέθοδος είναι γνωστή ως Box-Müller μετασχηματισμός. Το μόνο μειονέκτημα της μεθόδου από άποψης ταχύτητας είναι ότι απαιτείται ο υπολογισμός συναρτήσεων, όπως είναι ο λογάριθμος, το ημίτονο και το συνημίτονο.

Από τα παραπάνω παραδείγματα ίσως γίνεται αντιληπτό ότι μπορούν να χρησιμοποιηθούν σχέσεις μεταξύ μεταβλητών και γνωστές ιδιότητές τους για την προσομοίωση. Έτσι, η προσομοίωση από την κανονική κατανομή μπορεί να επιτευχθεί μέσω του μετασχηματισμού Box-Müller, από τη Βήτα κατανομή μέσω της σχέσης με τα διατεταγμένα στατιστικά της ομοιόμορφης κατανομής κ.ο.κ.. Παρότι αυτές οι σχέσεις μπορούν να χρησιμοποιηθούν, συχνά δεν είναι βέλτιστες. Για εκείνες τις περιπτώσεις που δεν υπάρχουν σχέσεις με άλλες κατανομές ή με στόχο να έχουμε στη διάθεσή μας πιο γρήγορες σε υπολογιστικό χρόνο μεθόδους καταφεύγουμε σε άλλες τεχνικές προσομοίωσης που θα παρουσιαστούν στις επόμενες ενότητες και οι οποίες απαιτούν να γνωρίζουμε μόνο ότι η συναρτησιακή μορφή της πυκνότητας της κατανομής που μας ενδιαφέρει είναι ανάλογη μιας άλλης συνάρτησης.

## 2.3 Μέθοδος αποδοχής-απόρριψης

Η άμεση εφαρμογή της μεθόδου της αντιστροφής σε πολλές περιπτώσεις, όπως έχει ήδη αναφερθεί στην προηγούμενη ενότητα, είναι δύσκολη ή ακόμη και ανέφικτη. Επιπλέον, υπάρχουν περιπτώσεις που η κατανομή που θέλουμε να προσομοιώσουμε δεν μπορεί να γραφεί είτε ως μείξη κατανομών για τις οποίες είναι εφικτή η προσομοίωση είτε ως ένας άμεσος μετασχηματισμός κάποιας τέτοιας. Τα παραπάνω προβλήματα έχουν ως αποτέλεσμα να πρέπει να αναζητηθούν άλλες εναλλακτικές μέθοδοι για την αντιμετώπισή τους. Μεταξύ αυτών έχει εισαχθεί στη βιβλιογραφία (βλέπε Von Neumann (1951)) η αποκαλούμενη μέθοδος **αποδοχής-απόρριψης** (Accept-Reject method), η οποία αποτελεί και το αντικείμενο μελέτης αυτής της ενότητας.

Επομένως στην ενότητα αυτή το ενδιαφέρον επικεντρώνεται σε μια άλλη κατηγορία μεθόδων προσομοίωσης, πιο γενική και πιο εφαρμόσιμη από τη μέθοδο της αντιστροφής, η οποία όπως αναλυτικά θα δούμε, απαιτεί μόνο τη γνώση της

συναρτησιακής μορφής της πυκνότητας της κατανομής στόχου $\pi(\cdot)$, εκτός ίσως από κάποια πολλαπλασιαστική σταθερά.

**Παρατήρηση 2.3.1.** *Η ιδέα της γνώσης της πυκνότητας της κατανομής στόχου, εκτός ίσως από κάποια πολλαπλασιαστική σταθερά, είναι, όπως θα επισημανθεί αναλυτικότερα στο υπόλοιπο αυτού του κεφαλαίου καθώς και στο επόμενο κεφάλαιο, κεντρική για τις περισσότερες μεθόδους προσομοίωσης.*

Η βασική και θεμελιώδης ιδέα αυτής της μεθόδου περιγράφεται στη συνέχεια για τη συνεχή περίπτωση. Έστω ότι θέλουμε να προσομοιώσουμε από μια τ.μ. με σ.π.π. $\pi(\cdot)$, η οποία ονομάζεται κατανομή στόχος. Τότε ισχύει ότι:

$$\pi(x) = \int_0^{\pi(x)} 1\, du = \int_0^1 I_{(0,\pi(x))}(u)\, du = \int_0^1 \pi(x,u)\, du,$$

όπου $I_A(\cdot)$ η δείκτρια συνάρτηση που λαμβάνει την τιμή 1, για $x \in A$ και την τιμή μηδέν, για $x \notin A$. Από τα παραπάνω άμεσα προκύπτει ότι αρκεί να προσομοιώσουμε ομοιόμορφα από την περιοχή $\{(x,u) : 0 < u < \pi(x)\}$ και οι τιμές του $x$ που θα προκύψουν θα προέρχονται από την κατανομή στόχο. Το παραπάνω ουσιαστικά αποτελεί το Θεμελιώδες Θεώρημα της Προσομοίωσης (βλέπε, μεταξύ άλλων, Robert and Casella (2004)), η διατύπωση του οποίου ακολουθεί.

**Θεώρημα 2.3.2.** *Η προσομοίωση της $X \sim \pi(x)$ ισοδυναμεί με την προσομοίωση του διανύσματος $(X,U) \sim U\{(x,u) : 0 < u < \pi(x)\}$, το οποίο έχει την από κοινού σ.π.π. $I_{(0,\pi(x))}(u) I_{\mathcal{X}}(x)$.*

Ωστόσο, κάποιος δεν μπορεί συνήθως να παράγει ομοιόμορφες παρατηρήσεις από αυτήν την περιοχή με άμεσες μεθόδους. Για παράδειγμα, δεν μπορεί να πει κάποιος ότι θα παράγει $X \sim \pi$ και έπειτα $U|X = x \sim U(0, \pi(x))$, καθώς ο στόχος είναι η παραγωγή από την $\pi(\cdot)$ έτσι κι αλλιώς. Επιπλέον, η συμμετρική προσέγγιση της παραγωγής από την περιθώρια της $U$ και έπειτα από τη δεσμευμένη της $U|X = x$ δεν οδηγεί συνήθως σε εφικτό υπολογισμό. Η λύση που δόθηκε περιγράφεται στην επόμενη πρόταση και έγκειται στην προσομοίωση του ζεύγους $(X,U)$ σε ένα μεγαλύτερο σύνολο και η αποδοχή ή όχι της παρατήρησης $X$ έχοντας ως κριτήριο αν ανήκει ή όχι στην περιοχή ενδιαφέροντος. Η απόδειξη που θα δοθεί ακολουθεί την πορεία, μεταξύ άλλων, των Sigman (2007) και Haft (2014).

**Πρόταση 2.3.3.** *Έστω ότι η κατανομή στόχος είναι απόλυτα συνεχής ως προς το μέτρο πιθανότητας $\mu$, με πυκνότητα $\pi(\cdot)$, ενώ $\mathcal{X} = \{x : \pi(x) > 0\}$. Έστω μια άλλη κατανομή επίσης απόλυτα συνεχής ως προς το μέτρο πιθανότητας $\mu$, με πυκνότητα $g(\cdot)$. Υποθέτουμε ότι $Y \sim g$ και $U \sim U(0,1)$ ανεξάρτητες τυχαίες*

*μεταβλητές. Τότε η δεσμευμένη κατανομή της* $Y$ *δοθέντος* $U \leq \dfrac{\pi(Y)}{Mg(Y)}$*, όπου* $M$ *θετικός αριθμός*[2] *ο οποίος πληροί τη συνθήκη* $\dfrac{\pi(y)}{Mg(y)} \leq 1, \forall\, y \in \mathcal{X}$*, έχει σ.π.π.* $\pi(\cdot)$*.*

**Απόδειξη.** Από τον ορισμό της δεσμευμένης πιθανότητας έχουμε ότι για κάθε μετρήσιμο σύνολο $A$:

$$
\begin{aligned}
P\left(Y \in A \mid U \leq \frac{\pi(Y)}{Mg(Y)}\right) &= \frac{P\left(Y \in A,\, U \leq \dfrac{\pi(Y)}{Mg(Y)}\right)}{P\left(U \leq \dfrac{\pi(Y)}{Mg(Y)}\right)} \\
&= \frac{\displaystyle\int_A \int_0^{\frac{\pi(y)}{Mg(y)}} \mathrm{d}u\, g(y)\, \mu(\mathrm{d}y)}{\displaystyle\int_{\mathcal{X}} \int_0^{\frac{\pi(y)}{Mg(y)}} \mathrm{d}u\, g(y)\, \mu(\mathrm{d}y)} \\
&= \frac{\displaystyle\int_A \frac{\pi(y)}{Mg(y)} g(y)\, \mu(\mathrm{d}y)}{\displaystyle\int_{\mathcal{X}} \frac{\pi(y)}{Mg(y)} g(y)\, \mu(\mathrm{d}y)} \\
&= \int_A \pi(y)\, \mu(\mathrm{d}y).
\end{aligned}
$$

Άρα, η δεσμευμένη κατανομή της $Y$ δοθέντος $U \leq \dfrac{\pi(Y)}{Mg(Y)}$ έχει σ.π.π. την $\pi(\cdot)$. □

Ειδικότερα, η κεντρική ιδέα της μεθόδου, που εισήχθη στη βιβλιογραφία από τον Von Neumann (1951), είναι να χρησιμοποιηθεί κάποια άλλη απλούστερη σ.π.π. $g(\cdot)$, που να είναι κοντά στην $\pi(\cdot)$, και από την οποία να είναι πραγματικά πιο εύκολη η διαδικασία της προσομοίωσης. Η κατανομή $g(\cdot)$ ονομάζεται συνηθέστερα **κατανομή πρότασης** (trial or proposal distribution), ενώ έχει εμφανιστεί στη βιβλιογραφία και ως **καθοριστική** (instrumental) ή **υποψήφια** (candidate density). Στην παρούσα διατριβή θα αναφερόμαστε σε αυτήν την

[2] Ο ρόλος και το σύνολο των δυνατών τιμών αυτού του αριθμού θα σχολιασθεί στη συνέχεια.

κατανομή ως κατανομή πρότασης, εκτός αν διευκρινίζεται διαφορετικά. Επίσης, αρχικά υποθέτουμε ότι η $g(\cdot)$ είναι τέτοια ώστε αν $g(x) = 0 \Rightarrow \pi(x) = 0$. Τέλος, η ονομασία μέθοδος αποδοχής-απόρριψης οφείλεται στο γεγονός ότι κατά την υλοποίησή της οι τιμές που προσομοιώνονται από την $g(\cdot)$ στην ουσία προτείνονται ως παρατηρήσεις από την $\pi(\cdot)$ και γίνονται αποδεκτές ή απορρίπτονται στη βάση μιας ανισότητας, με μία πιθανότητα που θα προσδιοριστεί παρακάτω.

Πρακτικά η παραπάνω πρόταση μας οδηγεί στον ακόλουθο αλγόριθμο για τη δημιουργία μιας παρατήρησης από την $\pi(\cdot)$.

**Αλγόριθμος αποδοχής-απόρριψης (Accept-Reject Algorithm)**

- **Βήμα 1:** Αρχικά, παράγεται η τ.μ. $Y \sim g$.
- **Βήμα 2:** Ομοίως, προσομοιώνεται μία τ.μ. $U$ (ανεξάρτητη από την $Y$) τέτοια ώστε $U \sim U(0,1)$.
- **Βήμα 3:** Αν $U \leq \dfrac{\pi(Y)}{Mg(Y)}$ τότε η τιμή $X = Y$ γίνεται δεκτή (άρα ισχύει ότι $X \sim \pi$), και ο αλγόριθμος τερματίζει. Διαφορετικά, δηλαδή στην περίπτωση που $U > \dfrac{\pi(Y)}{Mg(Y)}$, απορρίπτεται η $Y$ και επιστρέφουμε πάλι στο Βήμα 1.

Εύλογα ερωτήματα που ίσως προκύπτουν είναι: ποια είναι η πιθανότητα αποδοχής μιας παρατήρησης που έχει παραχθεί από την $g$ και ποιος είναι ο ρόλος της σταθεράς $M$ και της πυκνότητας $g(\cdot)$. Στη συνέχεια θα απαντηθούν τα παραπάνω (βλέπε, μεταξύ άλλων, Johansen et al. (2007)).

**Πρόταση 2.3.4.** *Η πιθανότητα αποδοχής της $Y$ χρησιμοποιώντας τη μέθοδο αποδοχής-απόρριψης ισούται με $1/M$.*

**Απόδειξη.** Η $Y$ γίνεται αποδεκτή όταν $U \leq \dfrac{\pi(Y)}{Mg(Y)}$, ενώ η πιθανότητα αποδοχής της $Y$ είναι το ποσοστό των προτεινόμενων δειγμάτων που γίνονται αποδεκτά. Λαμβάνοντας υπόψη ότι $U \sim U(0,1)$ και $\dfrac{\pi(Y)}{Mg(Y)} \leq 1$ προκύπτει ότι:

$$\begin{aligned} P\left(U \leq \frac{\pi(Y)}{Mg(Y)}\right) &= \int_{\mathcal{X}} P\left(U \leq \frac{\pi(y)}{Mg(y)}\right) g(y)\,\mathrm{d}y \\ &= \int_{\mathcal{X}} \frac{\pi(y)}{Mg(y)}\, g(y)\,\mathrm{d}y \\ &= \frac{1}{M} \underbrace{\int_{\mathcal{X}} \pi(y)\,\mathrm{d}y}_{=1} = \frac{1}{M}. \end{aligned}$$

□

Από την προηγούμενη πρόταση άμεσα προκύπτει ότι:

$$M = \frac{1}{P\left(U \leq \frac{\pi(Y)}{Mg(Y)}\right)}.$$

Καθώς $0 \leq P\left(U \leq \frac{\pi(Y)}{Mg(Y)}\right) \leq 1$ έχουμε ότι $1 \leq M < +\infty$. Επομένως η σταθερά $M$ όχι μόνο είναι θετική αλλά είναι και μεγαλύτερη ή ίση του 1. Στο σημείο αυτό, επισημαίνεται ότι όταν $M = 1$ τότε, λόγω της Πρότασης 2.3.4, ισχύει ότι $P\left(U \leq \frac{\pi(y)}{Mg(y)}\right) = P\left(U \leq \frac{\pi(y)}{g(y)}\right) = 1$. Από αυτήν τη σχέση συμπεραίνουμε ότι όταν $M = 1$ η κατανομή στόχος ταυτίζεται με την κατανομή πρότασης.

Επιπρόσθετα, παρατηρείστε ότι οι υποψήφιες παρατηρήσεις που προτείνονται στη μέθοδο αποδοχής-απόρριψης παράγονται ανεξάρτητα και γίνονται αποδεκτές με πιθανότητα $1/M$. Ουσιαστικά, για να προσομοιωθεί μια παρατήρηση από την κατανομή στόχο πραγματοποιούνται ανεξάρτητα πειράματα Bernoulli με πιθανότητα επιτυχίας $1/M$ μέχρι να εμφανιστεί η πρώτη επιτυχία. Συνεπώς, ο αριθμός των δειγμάτων που απαιτούνται από την $Y$ για να ληφθεί μια αποδεκτή τιμή (μέχρι να έχουμε την πρώτη αποδοχή) ακολουθεί μια γεωμετρική κατανομή με πιθανότητα επιτυχίας $1/M$, και μέση τιμή $M$ (βλέπε, μεταξύ άλλων, Sigman (2007)). Διαισθητικά, το $M$ είναι ο αναμενόμενος αριθμός των επαναλήψεων που απαιτούνται μέχρι να γίνει η πρώτη αποδοχή και αποτελεί ένα μέτρο της υπολογιστικής πολυπλοκότητας του αλγορίθμου. Άρα, όσο μικρότερο το $M$ τόσο καλύτερη θα είναι η επιλογή της κατανομής πρότασης. Τα παραπάνω οδηγούν να προτείνεται στην πράξη η σταθερά $M$ να είναι όσο το δυνατόν πιο κοντά στη μονάδα καθώς αυτό συνεπάγεται ότι απορρίπτονται, κατά μέσο όρο, λιγότερα δείγματα και κατά συνέπεια προκύπτουν λιγότερες επαναλήψεις του αλγορίθμου.

**Παρατήρηση 2.3.5.** *Η συνθήκη $\pi(x) \leq Mg(x)$ ισοδυναμεί με το να είναι άνω φραγμένο το πηλίκο $\dfrac{\pi(x)}{g(x)}$. Προχωρώντας ένα βήμα πιο πέρα, αν για κάποια σταθερά $M$ ισχύει ότι η $\pi(x) \leq Mg(x)$, $\forall\, x$, τότε θα ισχύει και για οποιαδήποτε άλλη σταθερά $M^{'} > M$. Άρα, η μέθοδος της αποδοχής-απόρριψης μπορεί να εφαρμοστεί και για οποιοδήποτε σταθερά $M^{'}$ στη θέση του $M$ με $M^{'} > M$. Αυτό πρακτικά σημαίνει ότι αρκεί να βρεθεί ένα άνω φράγμα του πηλίκου $\dfrac{\pi(x)}{g(x)}$. Αξίζει να σημειωθεί ότι ακόμη και στις περιπτώσεις που η κατανομή στόχος ή/και η κατανομή πρότασης είναι γνωστές, εκτός από κάποιες πολλαπλασιαστικές σταθερές, πάλι μπορεί να εφαρμοστεί ο αλγόριθμος αποδοχής-απόρριψης με τη διαφορά ότι θα αποδέχεται πιο δύσκολα τις τιμές της $Y$.*

**Παρατήρηση 2.3.6.** *Συνοψίζοντας για τη μέθοδο αποδοχής-απόρριψης, τα βασικότερα πλεονεκτήματα είναι η απλότητα, η γενικότητα και η επεκτασιμότητά της σε ένα ευρύ φάσμα επιστημονικών κλάδων. Από την άλλη πλευρά όμως, μπορεί να οδηγήσει στη λήψη πολλών ανεπιθύμητων δειγμάτων αν η συνάρτηση της δειγματοληψίας είναι πολύ συγκεντρωμένη σε μια συγκεκριμένη περιοχή.*

Στη συνέχεια εφαρμόζονται, μέσω παραδειγμάτων (βλέπε Robert and Casella (1999)), μερικές από τις διαδικασίες που αναπτύχθηκαν σε αυτήν την παράγραφο με σκοπό τη διευκρίνηση των πιο βασικών παραπάνω αποτελεσμάτων.

**Παράδειγμα 2.3.7.** *Χρησιμοποιώντας τη μέθοδο αποδοχής-απόρριψης να προταθεί ένας τρόπος προσομοίωσης από την $N(0,1)$.*

***Λύση***. Έστω ότι η $\pi \sim N(0,1)$, δηλαδή $\pi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$. Επιλέγεται ως κατανομή πρότασης $g$ η κατανομή Cauchy, $g(x) = \dfrac{1}{\pi\,(1+x^2)}, x \in \mathbb{R}$, που αποτελεί μια απλούστερη σ.π.π., κοντά στην $\pi(x)$. Η επιλογή της κατανομής Cauchy γίνεται για τον επιπλέον λόγο ότι είναι πιο εύκολη η διαδικασία της προσομοίωσης από αυτήν (Robert and Casella (1999)). Πιο συγκεκριμένα, ακολουθώντας το Βήμα 1 της μεθόδου, προσομοιώνονται $n = 3.000$ τυχαίες τιμές από την $g(x)$. Κατά αυτόν τον τρόπο παράγεται ένα τ.δ. $X_1, X_2, ..., X_{3.000}$ από την κατανομή Cauchy. Ας αναφερθεί ότι εφόσον θα εφαρμοσθεί η μέθοδος της απόρριψης, κάποιες από αυτές τις τιμές δε θα γίνουν δεκτές με αποτέλεσμα το πλήθος των αποδεκτών τυχαίων παρατηρήσεων να είναι μικρότερο του 3.000. Στη συνέχεια, εκτελώντας το Βήμα 2 της μεθόδου, δημιουργούνται $n = 3.000$ τυχαίες τιμές $U$ από την ομοιόμορφη $U(0,1)$.

Θέλοντας να προσδιορίσουμε ένα άνω φράγμα της συνάρτησης $h(x) = \dfrac{\pi(x)}{g(x)} = \sqrt{\dfrac{\pi}{2}}\,(1+x^2)\,e^{-\frac{x^2}{2}}, x \in \mathbb{R}$, μελετούμε τη μονοτονία της. Παρατηρούμε ότι η πρώτη παράγωγος της $h(x)$ μηδενίζεται στα σημεία $-1, 0, 1$, με $h''(1) < 0$, $h''(-1) < 0$ και $h''(0) > 0$. Άρα τα σημεία $-1$ και $1$ είναι τα σημεία που μεγιστοποιείται η $h(x)$ με $h(1) = h(-1) = \sqrt{2\pi}e^{-1/2}$. Επομένως $\dfrac{\pi(x)}{g(x)} = \sqrt{\dfrac{\pi}{2}}\,(1+x^2)\,e^{-\frac{x^2}{2}} \leq \sqrt{2\pi}e^{-1/2} = 1.520347 = M$.

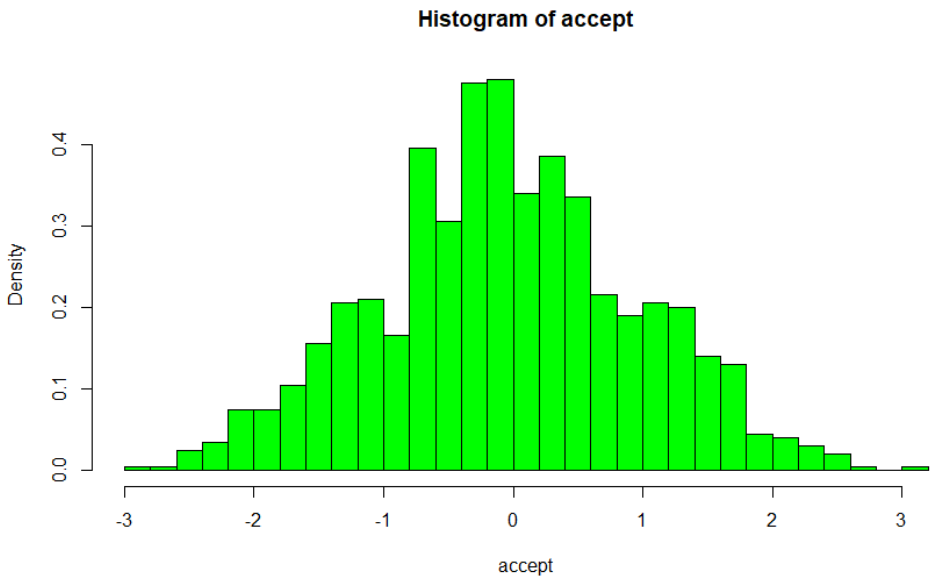
Από την παραπάνω σχέση προκύπτει ότι ο μέσος αριθμός των επαναλήψεων που απαιτούνται από τον αλγόριθμο μέχρι την πρώτη αποδοχή είναι $M = 1.520347$, ενώ η πιθανότητα αποδοχής, όπως έχει ήδη αναφερθεί, ισούται με $\dfrac{1}{M} = \dfrac{1}{1.520347} \simeq 0.66$.

Με άλλα λόγια αυτό σημαίνει ότι οι υποψήφιες τιμές που προτείνονται από τον αλγόριθμο αποδοχής-απόρριψης γίνονται δεκτές με πιθανότητα 66%. Οι εντολές της R που χρησιμοποιήθηκαν για την υλοποίηση του παραδείγματος παρατίθενται ακολούθως. Επίσης, στο Σχήμα 2.4 δίνεται το ιστόγραμμα 1.000 τιμών από την κατανομή στόχο. Αξίζει να επισημανθεί ότι για να επιτευχθεί αυτό χρησιμοποιήθηκαν 2.516 τιμές από τις 3.000 τιμές που δημιουργήθηκαν από την κατανομή πρότασης.

```
set.seed(3000)
X = rcauchy(3000,0,1)
U = runif(3000,0,1)
pi_x = function(x){
  new_x = (1/sqrt(2*pi))*exp(-x^2/2)
  return(new_x)
}
count=1
accept=c()
while(count<=3000 & length(accept)<1000){
  test_u = U[count]
  test_x = pi_x(X[count])/(1.520347 * dcauchy(X[count
     ],0,1))
  if(test_u <= test_x){
    accept = rbind(accept,X[count])
    count = count+1
  }
```

```
  count = count+1
}
hist(accept,breaks=30,col="green",prob = TRUE)
```

Σχήμα 2.4: Ιστόγραμμα των 1.000 τυχαίων αριθμών από την τυπική κανονική κατανομή. Οι τυχαίοι αριθμοί δημιουργήθηκαν με τη μέθοδο της αποδοχής-απόρριψης, με ποσοστό αποδοχής 1.000/2.516.




**Παράδειγμα 2.3.8.** *Να προταθεί, εφαρμόζοντας τη μέθοδο αποδοχής-απόρριψης, ένας τρόπος προσομοίωσης από τη συνάρτηση* $\pi(x) = \frac{3}{2}x^3 + \frac{11}{8}x^2 + \frac{1}{6}x + \frac{1}{12}$, $x \in [0,1]$.

***Λύση.*** Η κατανομή στόχος δίνεται από τη σχέση:

$$\pi(x) = \frac{3}{2}x^3 + \frac{11}{8}x^2 + \frac{1}{6}x + \frac{1}{12},\ x \in [0,1]. \tag{2.4}$$

Καθώς υποθέτουμε ότι η προσομοίωση από την $\pi(x)$ είναι δύσκολη, θα χρησιμοποιήσουμε τη μέθοδο αποδοχής-απόρριψης. Για το λόγο αυτό ο στόχος είναι η

εύρεση μιας άλλης προτεινόμενης κατανομής, κοντά στην $\pi(x)$, από την οποία η διαδικασία παραγωγής μιας παρατήρησης είναι απλούστερη. Έτσι, θεωρούμε ότι η συνάρτηση $g$ ακολουθεί ομοιόμορφη κατανομή στο διάστημα $[0,1]$. Η επιλογή αυτή αιτιολογείται από το γεγονός ότι η κατανομή πρότασης έχει κοινό πεδίο ορισμού με την κατανομή στόχο, αλλά είναι πιο εύκολη η προσομοίωση. Είναι:

$$g(x) = \begin{cases} 1, & \text{αν } x \in [0,1], \\ 0, & \text{διαφορετικά.} \end{cases}$$

Αρχικά (βλέπε Βήμα 1 της μεθόδου) προσομοιώνονται $n = 5.000$ τυχαίες τιμές από την $g$ (εδώ την $U(0,1)$). Κατά αυτόν τον τρόπο δημιουργείται ένα τ.δ. $X_1, X_2, ..., X_{5.000}$ από την $U(0,1)$. Σημειώνουμε ότι καθώς θα εφαρμοστεί η μέθοδος της απόρριψης, κάποιες από αυτές τις τιμές θα απορριφθούν και έτσι το πλήθος των αποδεκτών τυχαίων παρατηρήσεων θα είναι μικρότερο του 5.000. Έπειτα (βλέπε Βήμα 2 της μεθόδου) παράγονται $n = 5.000$ τυχαίες τιμές $U$ από την ομοιόμορφη $U(0,1)$ (εδώ απλά η κατανομή πρότασης $g$ έτυχε να είναι και αυτή ομοιόμορφη στο $[0,1]$). Στη συνέχεια (βλέπε Βήμα 3 της μεθόδου) θέλουμε να προσδιορίσουμε ένα άνω φράγμα $M$ του πηλίκου $\dfrac{\pi(x)}{g(x)} = \pi(x)$, $0 < x < 1$. Καθώς η συνάρτηση $\pi(x)$ είναι αύξουσα στο πεδίο ορισμού της, γίνεται εύκολα αντιληπτό ότι μεγιστοποιείται στο άνω άκρο του πεδίου ορισμού της. Άρα:

$$\frac{\pi(x)}{g(x)} = \pi(x) \leq \pi(1) = 3.125 = \frac{75}{24}.$$

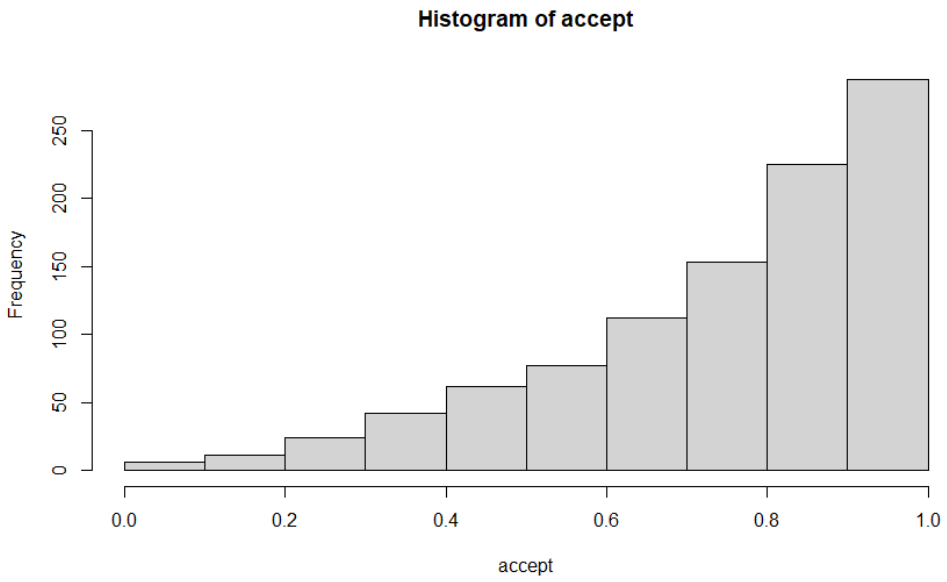
Αυτό σημαίνει ότι θα μπορούσε να χρησιμοποιηθεί η μέγιστη αυτή τιμή ως το άνω φράγμα $M$. Επομένως, αν $U \leq \dfrac{\pi(Y)}{M\, g(Y)}$, όπου $M = 3.125$, τότε η τιμή $X = Y$ γίνεται δεκτή από τον αλγόριθμο και το σύνολο των αποδεκτών τιμών αποτελεί το τ.δ. από την κατανομή στόχο $\pi$.

Τα παραπάνω υλοποιούνται στην R εκτελώντας τις εντολές που ακολουθούν, ενώ στο Σχήμα 2.5 δίνεται το ιστόγραμμα 1.000 τιμών από την κατανομή στόχο. Σημειώνεται ότι για να επιτευχθεί αυτό χρησιμοποιήθηκαν 4.141 τιμές από τις 5.000 τιμές που δημιουργήθηκαν από την κατανομή πρότασης.

```
X = runif(5000,0,1)
U = runif(5000,0,1)
pi_x = function(x){
  new_x = (3/2)*(x^3)+(11/8)*(x^2)+(1/6)*(x)+(1/12)
  return(new_x)
```

```
}
count=1
accept=c()
while(count<=5000 & length(accept)<1000){
  test_u=U[count]
  test_x=pi_x(X[count])/(3.125 * dunif(X[count],0,1))
  if(test_u <= test_x){
    accept=rbind(accept,X[count])
    count=count+1
  }
  count=count+1
 }
hist(accept)
```

Σχήμα 2.5: Ιστόγραμμα των 1.000 τυχαίων αριθμών από την κατανομή με συνάρτηση πυκνότητας πιθανότητας της σχέσης (2.4). Οι τυχαίοι αριθμοί δημιουργήθηκαν με τη μέθοδο της αποδοχής-απόρριψης, με ποσοστό αποδοχής 1.000/4.141.




**Παρατήρηση 2.3.9.** *Η δειγματοληψία αποδοχής-απόρριψης μπορεί να οδηγήσει*

*στη λήψη πολλών ανεπιθύμητων δειγμάτων αν η συνάρτηση δειγματοληψίας είναι πολύ συγκεντρωμένη σε μια συγκεκριμένη περιοχή. Επιπλέον, καθώς οι διαστάσεις του προβλήματος μεγαλώνουν, σύμφωνα με τον Liu (2008), πολλές απορρίψεις μπορούν να πραγματοποιηθούν πριν δημιουργηθεί ένα χρήσιμο δείγμα, καθιστώντας έτσι τον αλγόριθμο μη αποτελεσματικό και μη πρακτικό. Σε υψηλές διαστάσεις είναι απαραίτητο να χρησιμοποιηθεί μια διαφορετική προσέγγιση, συνήθως μια μέθοδος Markov Chain Monte Carlo, όπως η δειγματοληψία Metropolis ή η δειγματοληψία Gibbs (που θα παρουσιαστούν στο επόμενο κεφάλαιο). Ωστόσο, η δειγματοληψία Gibbs, η οποία αναλύει ένα πολυδιάστατο πρόβλημα δειγματοληψίας σε μια σειρά δειγμάτων χαμηλών διαστάσεων, μπορεί να χρησιμοποιήσει τη δειγματοληψία αποδοχής-απόρριψης ως ένα από τα βήματά της (βλέπε, μεταξύ άλλων, Robert and Casella (2004)).*

Κλείνοντας αυτήν την παράγραφο, θα ήταν παράλειψη να μην επισημανθεί ότι η τεχνική της αποδοχής-απόρριψης είναι μια πολύ ισχυρή μέθοδος, διότι παρέχει τη δυνατότητα να παράγονται τυχαίοι αριθμοί από κατανομές των οποίων η $F^{-}(\cdot)$ δεν μπορεί να δοθεί σε κλειστή μορφή. Η μέθοδος αποδοχής-απόρριψης είναι γενικά πολύ πιο εφαρμόσιμη από τη μέθοδο αντιστροφής, καθώς απαιτεί μόνο να γνωρίζουμε τη συναρτησιακή μορφή της $\pi(\cdot)$ και να μπορούμε να προσδιορίσουμε μια κατανομή πιθανότητας που είναι κατάλληλη για χρήση στον αλγόριθμο. Δυστυχώς, ο προσδιορισμός μιας τέτοιας κατανομής, έστω $Y$, μπορεί να είναι δύσκολος και μάλιστα αν το $M$ είναι μεγάλο τότε η δημιουργία ενός δείγματος μεγέθους $n$ από την $X$ μπορεί να είναι υπολογιστικά δύσκολη διαδικασία καθώς θα πρέπει να δημιουργήσουμε ένα δείγμα μεγέθους $M \cdot n$ από την από κοινού $(Y, U)$. Για το λόγο αυτό έχουν προταθεί και άλλες μέθοδοι. Μία από αυτές είναι η δειγματοληψία σπουδαιότητας που παρουσιάζεται στην επόμενη ενότητα.

## 2.4 Δειγματοληψία σπουδαιότητας

Στο πλαίσιο του υπολογισμού του ολοκληρώματος $\int_{\mathcal{X}} w(x)\,\mathrm{d}x$ ή του ισοδύναμου $I = \int_{\mathcal{X}} h(x)\pi(x)\,\mathrm{d}x$ και για την αντιμετώπιση του προβλήματος προσομοίωσης από μια κατανομή έχει εισαχθεί στη βιβλιογραφία (βλέπε Marshall (1956)) η αποκαλούμενη **δειγματοληψία σπουδαιότητας** (Importance Sampling, IS). Η μέθοδος αυτή, η οποία χρησιμοποιείται αντί της μεθόδου αποδοχής-απόρριψης στις περιπτώσεις που δεν είναι εφικτό να βρεθεί μια κατανομή $g$ και μια σταθερά $M$ ώστε να ισχύει $\pi(x) \leq Mg(x), \forall x$, ή όταν το $M$ είναι πολύ μεγάλο, είναι αποδοτικότερη και αποτελεί αντικείμενο μελέτης αυτής της ενότητας.

Η κεντρική ιδέα της δειγματοληψίας σπουδαιότητας είναι ο υπολογισμός του ολοκληρώματος $I$ μέσω της προσομοίωσης από μια διαφορετική συνάρτηση πυκνότητας πιθανότητας, έστω $g(\cdot)$, για την οποία η προσομοίωση είναι εφικτή. Επιθυμητό είναι η προσοχή μας να στραφεί στην περιοχή της μεγαλύτερης «σπουδαιότητας» (importance) έτσι ώστε να εξοικονομήσουμε υπολογιστικό χρόνο. Βασική προϋπόθεση είναι η $\pi(\cdot)$ και η $g(\cdot)$ να είναι απολύτως συνεχείς ως προς κάποιο μέτρο $\mu$, με το στήριγμα της $g(\cdot)$ να περιέχει το στήριγμα της $\pi(\cdot)$. Δηλαδή υποθέτουμε ότι:

$$\mathcal{X} = \{x : \pi(x) > 0\} \subseteq \{x : g(x) > 0\}.$$

Η παραπάνω υπόθεση ουσιαστικά σημαίνει ότι η συνάρτηση $g$ πληροί την ακόλουθη ιδιότητα: αν $g(x) = 0 \Rightarrow \pi(x)\, h(x) = 0$, ενώ δε σημαίνει ότι $\pi(x) = 0$ για όλα τα $x$ όπου $g(x) = 0$.

Εύκολα τότε προκύπτει ότι:

$$I = E_\pi\left(h(X)\right) = E_g\left(\frac{h(X)\,\pi(X)}{g(X)}\right). \tag{2.5}$$

Πράγματι, $E_\pi\left(h(X)\right) = \displaystyle\int_{x\in A} h(x)\pi(x)\,\mathrm{d}x$, όπου $A$ το πεδίο ορισμού της $\pi(x)$. Δηλαδή $\pi(x) > 0$, αν $x \in A$ και $\pi(x) = 0$, διαφορετικά. Επίσης ισχύει ότι:

$$E_g\left(\frac{h(X)\pi(X)}{g(X)}\right) = \int_{-\infty}^{+\infty} \frac{h(x)\pi(x)}{g(x)}\, g(x)\,\mathrm{d}x = \int_{x\in A} \frac{h(x)\pi(x)}{g(x)}\, g(x)\,\mathrm{d}x,$$

διότι η $\pi(x) = 0$ για τα $x$ που δεν ανήκουν στο $A$. Επομένως προκύπτει ότι $E_g\left(\dfrac{h(X)\,\pi(X)}{g(X)}\right) = E_\pi\left(h(X)\right)$.

Σε όσα ακολουθούν σε αυτήν την ενότητα η σ.π.π. $g(x)$ ονομάζεται **κατανομή σπουδαιότητας** (importance sampling distribution), η σ.π.π. $\pi(x)$ ονομάζεται **κατανομή στόχος** (target distribution), ενώ το πηλίκο $\dfrac{\pi(x)}{g(x)} = w(x)$, λέγεται **πηλίκο πιθανοφάνειας** (likelihood ratio) των $\pi$, $g$.

Επομένως, αν $X_1, X_2, ..., X_n$ είναι ένα τυχαίο δείγμα από τον πληθυσμό με σ.π.π. $g(x)$, ο εκτιμητής που προτείνεται με τη μέθοδο της δειγματοληψίας σπουδαιότητας για το ολοκλήρωμα $I = \displaystyle\int h(x)\pi(x)\,\mathrm{d}x$ είναι ο:

$$\hat{I}^{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)\pi(X_i)}{g(X_i)} = \frac{1}{n}\sum_{i=1}^{n} w(X_i)h(X_i), \tag{2.6}$$

όπου $w(X_i) = \dfrac{\pi(X_i)}{g(X_i)}$ είναι τα λεγόμενα **βάρη σπουδαιότητας** (importance weights), ενώ τα ζεύγη παρατηρήσεων $(X_i, w(X_i))$, $i = 1, ..., n$, αποτελούν ουσιαστικά ένα σταθμισμένο δείγμα από την $g$.

Χρησιμοποιώντας τα επιχειρήματα που παρουσιάστηκαν στην εισαγωγή αυτού του κεφαλαίου προκύπτει, υπό κατάλληλες υποθέσεις, ότι $\hat{I}^{IS} \xrightarrow{a.s.} I$, ενώ υπό την προϋπόθεση ότι η διακύμανση $\sigma_g^2 = Var_g\left(\dfrac{h(X)\pi(X)}{g(X)}\right)$ είναι πεπερασμένη, από το Κεντρικό Οριακό Θεώρημα έχουμε ότι:

$$\sqrt{n}\left(\hat{I}^{IS} - I\right) \xrightarrow{d} N\left(0, \sigma_g^2\right),$$

όπου $\sigma_g^2 = Var_g\left(\dfrac{h(X)\pi(X)}{g(X)}\right)$.

Η παραπάνω ασυμπτωτική κατανομή μπορεί να χρησιμοποιηθεί (βλέπε εισαγωγή αυτού του κεφαλαίου) για την κατασκευή ενός ασυμπτωτικού διαστήματος εμπιστοσύνης για το $I$. Με εφαρμογή του Θεωρήματος του Slutsky (βλέπε Van der Vaart (2000)) προκύπτει ότι:

$$\sqrt{n}\,\frac{\hat{I}^{IS} - I}{\hat{se}^{IS}} \xrightarrow{d} N(0,1),$$

με $\hat{se}^{IS}$ να είναι ένας συνεπής εκτιμητής της $\sqrt{Var_g(h(X)w(X))}$. Για παράδειγμα:

$$\hat{se}^{IS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(h(X_i)w(X_i) - \hat{I}\right)^2}.$$

Τότε, εύκολα προκύπτει ότι ένα $100(1-\alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το $I$ είναι το:

$$\left(\hat{I}^{IS} - z_{\alpha/2}\hat{se}^{IS}\sqrt{\frac{1}{n}},\ \hat{I}^{IS} + z_{\alpha/2}\hat{se}^{IS}\sqrt{\frac{1}{n}}\right).$$

Γίνεται αντιληπτό ότι, η επιλογή της συνάρτησης $g(x)$ διαδραματίζει κυρίαρχο ρόλο στην τεχνική αυτή και θα πρέπει να γίνεται με ιδιαίτερη προσοχή. Αυτό αιτιολογείται από το γεγονός ότι αρχικά απαιτείται η $\sigma_g^2$ να είναι πεπερασμένη. Αυτό ουσιαστικά σημαίνει ότι απαιτείται η διαφορά:

$$E_g\left(\frac{h^2(X)\pi^2(X)}{g^2(X)}\right) - \left(E_g\left(\frac{h(X)\pi(X)}{g(X)}\right)\right)^2 \tag{2.7}$$

ή ισοδύναμα η διαφορά

$$\int \frac{h^2(x)\,\pi^2(x)}{g^2(x)} g(x)\,\mathrm{d}x - \underbrace{\left(\int h(x)\,\pi(x)\,\mathrm{d}x\right)^2}_{\text{ανεξάρτητο της } g(x)}.$$

να είναι πεπερασμένη. Καθώς ο δεύτερος όρος είναι ισοδύναμος με το $I$, που έχουμε εξαρχής υποθέσει ότι είναι πεπερασμένο, άμεσα προκύπτει ότι αρκεί να είναι πεπερασμένος ο πρώτος όρος της διαφοράς. Το παραπάνω ισοδύναμα σημαίνει ότι αρκεί να είναι πεπερασμένο το ολοκλήρωμα:

$$\int \frac{h^2(x)\,\pi^2(x)}{g^2(x)} g(x)\,\mathrm{d}x. \tag{2.8}$$

Ποιες είναι εκείνες οι επιλογές της $g$ για τις οποίες επιτυγχάνεται κάτι τέτοιο θα αναφερθεί μετά την επόμενη πρόταση, στην οποία προσδιορίζεται η κατανομή $g$ για την οποία ελαχιστοποιείται η διασπορά $\sigma_g^2$ του εκτιμητή $\hat{I}^{IS}$ (για τη διατύπωση και απόδειξη βλέπε, μεταξύ άλλων, Robert and Casella (2004)).

**Πρόταση 2.4.1.** *Η $g(x)$ που ελαχιστοποιεί την ασυμπτωτική διασπορά του εκτιμητή $\hat{I}^{IS}$ δίνεται από τη σχέση:*

$$g^*(X) = \frac{|h(X)|\pi(X)}{\int |h(x)|\pi(x)\,\mathrm{d}x} \tag{2.9}$$

*και τότε έχουμε:*

$$\sigma_{g^*}^2 = Var_{g^*}\left(\frac{h(X)\pi(X)}{g^*(X)}\right) = (E_\pi|h(X)|)^2 - \left(I^{IS}\right)^2.$$

**Απόδειξη.** Στόχος είναι η εύρεση της $g$ για την οποία η $\sigma_g^2 = Var_g\left(\frac{h(X)\pi(X)}{g(X)}\right)$ ελαχιστοποιείται. Εύκολα προκύπτει ότι η ελαχιστοποίηση αυτή είναι ισοδύναμη με την ελαχιστοποίηση ως προς $g$ της διαφοράς που δόθηκε στη σχέση (2.7) ή ισοδύναμα του ολοκληρώματος που δόθηκε στη σχέση (2.8). Παρατηρείστε ότι το ολοκλήρωμα της σχέσης (2.8) είναι ίσο με την $E_g\left[\left(\frac{h(X)\pi(X)}{g(X)}\right)^2\right]$.

Καθώς η συνάρτηση $\phi(x) = x^2$ είναι κυρτή στο σύνολο των πραγματικών αριθμών, με εφαρμογή της ανισότητας Jensen (1906), προκύπτει ότι:

$$E_g\left(\frac{h^2(X)\,\pi^2(X)}{g^2(X)}\right) \geq \left(E_g\left(\left|\frac{h(X)\,\pi(X)}{g(X)}\right|\right)\right)^2.$$

Επομένως προσδιορίστηκε ένα κάτω φράγμα για την $E_g\left(\frac{h^2(X)\,\pi^2(X)}{g^2(X)}\right)$, το οποίο είναι ανεξάρτητο από την επιλογή της $g$. Το κάτω φράγμα επιτυγχάνεται για $g^*(X) = \frac{1}{c}\,|h(X)\,|\pi(X)$, όπου καθώς η $g^*(\cdot)$ πρέπει να είναι σ.π.π. έπεται ότι $c = \int |h(x)\,|\pi(x)\,\mathrm{d}x$. Όντως είναι:

$$\begin{aligned} E_{g^*}\left(\frac{h^2(X)\,\pi^2(X)}{(g^*)^2(X)}\right) &= E_{g^*}(c^2) = c^2 = \left(\int |h(x)\,|\pi(x)\,\mathrm{d}x\right)^2 \\ &= \left(\int \frac{|h(x)\,|\pi(x)}{g(x)} g(x)\,\mathrm{d}x\right)^2 \\ &= \left(E_g\left(\left|\frac{h(X)\,\pi(X)}{g(X)}\right|\right)\right)^2 \end{aligned}$$

και η απόδειξη ολοκληρώθηκε. □

Η Πρόταση 2.4.1 παρότι παρουσιάζει θεωρητικό ενδιαφέρον, στην πράξη δε βοηθά ιδιαίτερα, καθώς ο υπολογισμός του παρονομαστή της $g^*(x)$ ισοδυναμεί ουσιαστικά με τον υπολογισμό της $E_\pi(|\ h(X)\ |)$. Ο υπολογισμός αυτός όταν η $h(x) > 0$ ισοδυναμεί με τον υπολογισμό του αρχικού ολοκληρώματος $I$. Επομένως, κάποια άλλη κατανομή δειγματοληψίας σπουδαιότητας πρέπει να χρησιμοποιηθεί. Στο πλαίσιο αυτό αξιοποιώντας το αποτέλεσμα της Πρότασης 2.4.1 η $g$ πρέπει να επιλεγεί έτσι ώστε να μην απέχει πολύ από την βέλτιστη επιλογή $g^*$ και ταυτόχρονα να είναι τέτοια ώστε να είναι πεπερασμένο το ολοκλήρωμα της σχέσης (2.8), γεγονός που εξασφαλίζει ότι είναι πεπερασμένη η $\sigma_g$.

Ειδικότερα, όπως επισημαίνεται από τους Robert and Casella (2004), κατανομές δειγματοληψίας σπουδαιότητας με λεπτότερες ουρές από αυτές της $\pi$ (δηλαδή αυτές που είναι τέτοιες ώστε το πηλίκο $\pi/g$ να μην είναι φραγμένο) δεν είναι κατάλληλες για δειγματοληψία σπουδαιότητας. Στην πραγματικότητα σε αυτές τις περιπτώσεις η διακύμανση $\sigma_g^2$ θα είναι μη πεπερασμένη για πολλές συναρτήσεις $h$. Αντιστρόφως, επισημαίνουν ότι, κατανομές δειγματοληψίας με πιο βαριές ουρές από την $\pi$ εξασφαλίζουν ότι το πηλίκο $\pi/g$ δεν προκαλεί τη μη σύγκλιση του ολοκληρώματος της σχέσης (2.8). Συνοψίζοντας, επιθυμητό είναι να βρεθεί μια $g(x)$ η οποία έχει παρόμοιο σχήμα (ανάλογη μορφή) με την $\pi(x)\ |\ h(x)\ |$, υπό την προϋπόθεση ότι η $g(x)$ έχει πιο βαριές ουρές από την $\pi(x)\ |\ h(x)\ |$.

Σε όσα ακολουθούν υποθέτουμε ότι η $h$ είναι σ.π.π. (άρα $|h(x)| = h(x)$). Έστω τώρα $G, \Pi$ και $H$ οι α.σ.κ των $g, \pi$ και $h$, αντίστοιχα. Ουσιαστικά επιθυμητό είναι η γραφική παράσταση της συνάρτησης $g$ στις ουρές να είναι πιο πάνω από τη

γραφική παράσταση της συνάρτησης $\pi \cdot h$. Δηλαδή είναι επιθυμητό να υπάρχει ένα $x_0$ τέτοιο ώστε $g(x) > \pi(x)\, h(x), \forall\, x > x_0$ (για τη δεξιά ουρά). Αυτό σημαίνει ότι $P_G(X > x) > P_{\Pi H}(X > x), \forall\, x > x_0$, όπου με $\Pi H$ συμβολίζεται η α.σ.κ. της συνάρτησης $\pi \cdot h$. Από τη σχέση $P_G(X > x) > P_{\Pi H}(X > x)$ και υπό την ανεξαρτησία των $\Pi, H$ προκύπτει μετά από λίγη άλγεβρα ότι πρέπει να ισχύει ότι: $G_X(x) < \Pi_X(x)\, H_X(x)$, $\forall\, x \geq x_0$. Με ακριβώς ανάλογο τρόπο επιθυμητό είναι να υπάρχει $x_0$ τέτοιο ώστε $g(x) > \pi(x)\, h(x)$, $\forall\, x < x_0$.

Για να αντιμετωπιστεί το ζήτημα της πεπερασμένης διακύμανσης που αναφέρθηκε παραπάνω, εναλλακτικά έχει προταθεί στη βιβλιογραφία (ενδεικτικά παραπέμπουμε στους Hesterberg (1991), Robert and Casella (2004) και Shimkin (2015)) ο ακόλουθος κανονικοποιημένος εκτιμητής δειγματοληψίας σπουδαιότητας:

$$\tilde{I}^{IS} = \frac{\sum_{i=1}^{n} (h(X_i)\pi(X_i)/g(X_i))}{\sum_{i=1}^{n} (\pi(X_i)/g(X_i))} = \frac{\sum_{i=1}^{n} h(X_i)w(X_i)}{\sum_{i=1}^{n} w(X_i)}. \tag{2.10}$$

Δηλαδή, στην έκφραση του εκτιμητή που δίνεται από τη σχέση (2.6) αντικαθίσταται το $n$ με το άθροισμα των βαρών σπουδαιότητας. Υπό την υπόθεση ότι η $E_g\left(\frac{\pi(X)h(X)}{g(X)}\right)$ υπάρχει και είναι πεπερασμένη, εφαρμόζοντας τον ΙΝΜΑ έχουμε ότι:

$$\tilde{I}^{IS} = \frac{\sum_{i=1}^{n} (h(X_i)\pi(X_i)/g(X_i))}{\sum_{i=1}^{n} (\pi(X_i)/g(X_i))} = \frac{\frac{1}{n}\sum_{i=1}^{n} (h(X_i)\pi(X_i)/g(X_i))}{\frac{1}{n}\sum_{i=1}^{n} (\pi(X_i)/g(X_i))} \xrightarrow{a.s.} \frac{I}{1} = I,$$

καθώς $n \to \infty$. Αυτό το αποτέλεσμα προκύπτει διότι:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\pi(X_i)}{g(X_i)} \xrightarrow{a.s.} E_g\left(\frac{\pi(X)}{g(X)}\right) = \int \frac{\pi(x)}{g(x)}\, g(x)\, \mathrm{d}x = \underbrace{\int \pi(x)\, \mathrm{d}x}_{\pi(x)\ \text{σ.π.π.}} = 1,$$

για $n \to \infty$ και όπως εξηγήσαμε αρχικά, γνωρίζουμε ότι:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\pi(X_i)\, h(X_i)}{g(X_i)} \xrightarrow{a.s.} I = E_g\left(\frac{h(X)\, \pi(X)}{g(X)}\right),\ n \to \infty.$$

Παρατηρείστε βέβαια ότι ο κανονικοποιημένος εκτιμητής δεν είναι αμερόληπτος της $E_\pi(h(x))$. Όμως, η μεροληψία του είναι αρκετά μικρή, ενώ αποδίδει έναν πιο ευσταθή (stable) εκτιμητή, βελτιώνοντας τη διακύμανση. Η διακύμανση βελτιώνεται με την έννοια ότι όταν η $h$ είναι σταθερή συνάρτηση ο εναλλακτικός εκτιμητής (2.10) είναι κοντά σε αυτήν την τιμή, ενώ ο εκτιμητής της σχέσης (2.6) έχει μεγαλύτερη διακύμανση αφού το άθροισμα των βαρών σπουδαιότητας είναι διαφορετικό από τη μονάδα. Αυτή είναι και η αιτία που ο εκτιμητής αυτός είναι προτιμότερος από τον εκτιμητή της σχέσης (2.6). Πράγματι, οι Casella and Robert (1998) απέδειξαν ότι όταν η εκτίμηση γίνει με βάση το τετραγωνικό σφάλμα, τότε ο εναλλακτικός εκτιμητής μπορεί να είναι πιο αποδοτικός σε πολλές περιπτώσεις (βλέπε και Van Dijk and Kloek (1984)). Προχωρώντας ένα βήμα ακόμη, χρησιμοποιώντας τη Δέλτα Μέθοδο (βλέπε, για παράδειγμα, Van der Vaart (2000)) προκύπτει ότι για τον κανονικοποιημένο εκτιμητή ισχύει ότι:

$$\sqrt{n}\left(\tilde{I}^{IS} - I\right) \xrightarrow{d} N\left(0, \tau_h^2\right), n \to \infty, \tag{2.11}$$

όπου συμβολίζουμε με $\tau_h^2$ την ασυμπτωτική διακύμανση του $\tilde{I}^{IS}$. Για την $\tau_h^2$ προκύπτει ότι δίνεται από τη σχέση:

$$\tau_h^2 = E_g(w^2(X)h^2(X)) - 2E_\pi(h(X))E_g(w^2(X)h(X)) + (E_\pi(h(X)))^2 E_g(w^2(X))$$

ή ισοδύναμα από τη σχέση:

$$\tau_h^2 = E_\pi\left(w(X)h^2(X)\right) - 2E_\pi(h(X))E_\pi\left(w(X)h(X)\right) + (E_\pi(h(X)))^2 E_\pi\left(w(X)\right).$$

Επιπρόσθετα, ένας συνεπής εκτιμητής της $\tau_h^2$ είναι ο:

$$\hat{\tau}_h^2 = \frac{1}{n}\sum_{i=1}^{n}\left(h^2(X_i)\, w^2(X_i)\right) - 2\,\tilde{I}^{IS}\frac{1}{n}\sum_{i=1}^{n}\left(h(X_i)\, w^2(X_i)\right) + \left(\tilde{I}^{IS}\right)^2\frac{1}{n}\sum_{i=1}^{n} w^2(X_i)$$

και τελικά έχουμε ότι:

$$\hat{\tau}_h^2 = \frac{1}{n}\sum_{i=1}^{n} w^2(X_i)\left\{h(X_i) - \tilde{I}^{IS}\right\}^2. \tag{2.12}$$

Στην πράξη φυσικά, αρκετές φορές δε ξέρουμε ακριβώς τα βάρη σπουδαιότητας $w(x_i)$. Βέβαια, εξακολουθεί να ισχύει το Κεντρικό Οριακό Θεώρημα και ειδικότερα, το μόνο που πρέπει να προσέξουμε είναι ότι αλλάζει η εκτίμηση της ασυμπτωτικής διακύμανσης. Σε αυτές τις περιπτώσεις, λοιπόν, η ποσότητα που

εκτιμά την ασυμπτωτική διακύμανση (βλέπε, μεταξύ άλλων, Malefaki and Iliopoulos (2008)) δίνεται από τη σχέση:

$$\tilde{\tau}_h^2 = \frac{\sum_{i=1}^{n} w^2(X_i)\left\{h(X_i) - \tilde{I}^{IS}\right\}^2}{n\,\bar{w}^2},\ \bar{w} = \frac{1}{n}\sum_{i=1}^{n} w(X_i).$$

Επομένως, $\bar{w}^2\,\tilde{\tau}_h^2 = \frac{1}{n}\sum_{i=1}^{n} w^2(X_i)\left\{h(X_i) - \tilde{I}^{IS}\right\}^2 = \hat{\tau}_h^2 \Rightarrow \tilde{\tau}_h^2 = \frac{\hat{\tau}_h^2}{\bar{w}^2}$.

Έτσι αν αξιοποιήσουμε την Πρόταση 2.4.1, μία πιο πρακτική εναλλακτική λύση σύμφωνα με τους Robert and Casella (2004) είναι να χρησιμοποιηθεί ο κανονικοποιημένος εκτιμητής δειγματοληψίας σπουδαιότητας της έκφρασης (2.10) ως:

$$\tilde{I}^{IS} = \frac{\sum_{i=1}^{n}(h(X_i)\pi(X_i)/g(X_i))}{\sum_{i=1}^{n}(\pi(X_i)/g(X_i))} = \frac{\sum_{i=1}^{n} h(X_i) \mid h(X_i) \mid^{-1}}{\sum_{i=1}^{n} \mid h(X_i) \mid^{-1}},\ X_i \sim g,\ g \propto \mid h \mid \pi.$$

Δηλαδή, αντικαθίσταται το πηλίκο $\frac{\pi(X_i)}{g(X_i)}$ από την $\mid h(X_i) \mid^{-1} = \frac{1}{\mid h(X_i) \mid}$.

**Παρατήρηση 2.4.2.** *Στη βιβλιογραφία έχουν παρουσιαστεί διάφορες επεκτάσεις της δειγματοληψίας σπουδαιότητας που δε θα αποτελέσουν αντικείμενο μελέτης στα πλαίσια αυτής της διατριβής. Ανάμεσα σε αυτές εξέχουσα θέση κατέχει (i) η Διαδοχική Δειγματοληψία Σπουδαιότητας (sequential importance sampling), (ii) η Πληθυσμιακή MC μέθοδος (Population Monte Carlo) και (iii) η Προσαρμοσμένη Δειγματοληψία Σπουδαιότητας (adaptive importance sampling). Για σχετικές πληροφορίες παραπέμπουμε στους (i) Kong et al. (1994), Liu et al. (1998) και τις εκεί αναφορές, (ii) Iba (2000), Robert and Casella (2004) και Cappé et al. (2004) και τις εκεί αναφορές και (iii) Oh and Berger (1992) και Cornuet et al. (2012)) και τις εκεί αναφορές, αντίστοιχα.*

**Παρατήρηση 2.4.3.** *Επιγραμματικά αναφέρουμε στη συνέχεια, τα πιο σημαντικά πλεονεκτήματα και μειονεκτήματα της δειγματοληψίας σπουδαιότητας. Αρχικά, όταν δεν μπορούμε να προσομοιώσουμε από την $\pi$, προσομοιώνουμε από την $g$. Για $\pi h < g$ εξασφαλίζονται μικρές διακυμάνσεις, ταχύτερη σύγκλιση και μεγαλύτερη ακρίβεια. Επιπλέον, περιλαμβάνει αναλυτικές λύσεις σε πολλά προβλήματα και είναι ευσταθής για προβλήματα βελτιστοποίησης. Όσον αφορά τα μειονεκτήματα έπεται ότι το τυπικό σφάλμα του $\hat{I}$ μπορεί να πηγαίνει στο άπειρο. Σε πολλές περιπτώσεις είναι δύσκολο να βρεθεί μία $g$, που έχει παρόμοιο σχήμα*

*με την $\pi|h|$ και η $g$ να έχει πιο βαριές ουρές. Τέλος, για προβλήματα υψηλότερων διαστάσεων δε δίνει εξίσου καλά αποτελέσματα (ως προς τη διακύμανση του εκτιμητή).*

Εν συνεχεία, θα δούμε μέσα από απλά παραδείγματα πώς μπορούμε χρησιμοποιώντας τη μέθοδο της δειγματοληψίας σπουδαιότητας να υπολογίσουμε ολοκληρώματα της μορφής $I = \int_{\mathcal{X}} \pi(x)\, h(x)\, \mathrm{d}x$ μέσω της προσομοίωσης από μια διαφορετική σ.π.π. $g(x)$.

**Παράδειγμα 2.4.4.** *Να προταθούν, χρησιμοποιώντας τη δειγματοληψία σπουδαιότητας, τρόποι εκτίμησης του ολοκληρώματος:*

$$I = \int_0^{10} e^{-2\,|x-5|}\, \mathrm{d}x.$$

***Λύση***. Αρχικά επισημαίνεται ότι η αληθινή τιμή του ολοκληρώματος $I$ εύκολα προκύπτει ότι είναι ίση με $1 - e^{-10} = 0.9999546$, δηλαδή περίπου ίση με 1. Αν θεωρήσουμε ως $\pi(x)$ τη σ.π.π. της ομοιόμορφης $U(0, 10)$ και ως $h(x)$ την $h(x) = 10\, e^{-2\,|x-5|}$, $x \in \mathbb{R}$, τότε προφανώς ισχύει ότι: $E_\pi(h(X)) = \int_0^{10} \pi(x)h(x)\, \mathrm{d}x = I$. Επομένως, μπορούμε να εκτιμήσουμε το ολοκλήρωμα δημιουργώντας τυχαίο δείγμα παρατηρήσεων $X_1,...,X_n$ από την $U(0, 10)$ και σύμφωνα με τον ΙΝΜΑ θα έχουμε ότι η ποσότητα $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$ θα συγκλίνει σχεδόν βέβαια στο ζητούμενο ολοκλήρωμα (βλέπε σχετικά για τον εκτιμητή $\hat{I}^{BMCI}$ στην εισαγωγή). Από μια άλλη θεώρηση αυτός ο εκτιμητής είναι ένας εκτιμητής δειγματοληψίας σπουδαιότητας κατά την οποία ισχύει ότι $\pi(x) = g(x)$. Η εκτίμηση που προκύπτει με 100.000 παρατηρήσεις από την $U(0, 10)$ είναι ίση με 0.9942569 (με εκτιμώμενη διακύμανση 3.984059).

Στη συνέχεια θα χρησιμοποιηθεί η δειγματοληψία σπουδαιότητας έχοντας ως $g(x)$ τη συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής με μέση τιμή 5 και διακύμανση 1. Υποθέτουμε δηλαδή ότι $g(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x-5)^2/2}$, $x \in \mathbb{R}$.

Επομένως αν $Y_1,...,Y_n$ είναι ένα δείγμα από την $N(5, 1)$ ο εκτιμητής του ολοκληρώματος με τη μέθοδο δειγματοληψίας σπουδαιότητας είναι ο (βλέπε σχέση (2.6)):

$$\hat{I}^{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)\pi(X_i)}{g(X_i)} = \frac{1}{n}\sum_{i=1}^{n} w(X_i)h(X_i),$$

όπου $w(X_i) = \dfrac{\pi(X_i)}{g(X_i)}$ τα βάρη σπουδαιότητας.

Η εκτίμηση που προκύπτει με 100.000 παρατηρήσεις από την $N(5,1)$ είναι ίση με 0.9991394 (με εκτιμώμενη διακύμανση 0.3572763) και είναι πλησιέστερη στην αληθινή τιμή του ολοκληρώματος.

Τέλος, θα χρησιμοποιηθεί και ο κανονικοποιημένος εκτιμητής δειγματοληψίας σπουδαιότητας (βλέπε σχέση (2.10)):

$$\tilde{I}^{IS} = \frac{\sum_{i=1}^{n} (h(X_i)\pi(X_i)/g(X_i))}{\sum_{i=1}^{n} (\pi(X_i)/g(X_i))} = \frac{\sum_{i=1}^{n} h(X_i)w(X_i)}{\sum_{i=1}^{n} w(X_i)},$$

όπου $w(X_i) = \dfrac{\pi(X_i)}{g(X_i)}$ αποτελούν τα βάρη σπουδαιότητας. Στο πλαίσιο αυτό, η εκτίμηση που προκύπτει με 100.000 παρατηρήσεις από την $N(5,1)$ είναι ίση με 1.019856, αλλά δεν είναι η πλησιέστερη στην πραγματική τιμή του ολοκληρώματος.

Οι εντολές της $R$ για την υλοποίηση του παραδείγματος ακολουθούν.

```
set.seed(1)
X = runif(100000, 0, 10)
h = 10*exp(-2*abs(X-5))
ihat = mean(h)
ihat
0.9942569
var(h)
3.984059
dif1 = 1-exp(-10)-ihat
dif1
0.005697733
Y = rnorm(100000,mean=5,sd=1)
w = dunif(Y, 0, 10)/dnorm(Y, mean=5, sd=1)
h1 = 10*exp(-2*abs(Y-5))
ihatIS = mean (w*h1)
ihatIS
0.9991394
var(w*h1)
0.3572763
```

```
dif2 = 1-exp(-10)-ihatIS
dif2
0.0008151834
itilde=sum(w*h1)/sum(w)
itilde
1.019856
dif3=1-exp(-10)-itilde
dif3
[1] -0.01990131
```

**Παράδειγμα 2.4.5.** *Χρησιμοποιώντας τη μέθοδο της δειγματοληψίας σπουδαιότητας να εκτιμηθεί το ολοκλήρωμα:*

$$I = \int_{-\infty}^{+\infty} \frac{1}{2}\, x^2\, e^{-|x|}\, \mathrm{d}x.$$

***Λύση***. Πρωτίστως, από την παραγοντική ολοκλήρωση εύκολα προκύπτει ότι η αληθινή τιμή του ολοκληρώματος $I$ είναι ίση με 2. Θεωρώντας ως κατανομή στόχο $\pi(x)$ την $\frac{1}{2}\, e^{-|x|}$, $x \in \mathbb{R}$, η οποία είναι η γνωστή κατανομή Laplace, με παράμετρο θέσης 0 και παράμετρο κλίμακας 1 και ως $h(x) = x^2$, $x \in \mathbb{R}$, τότε προκύπτει ότι: $E_\pi(h(X)) = \int_{-\infty}^{+\infty} \pi(x)h(x)\,\mathrm{d}x = I$. Επομένως, μπορούμε να εκτιμήσουμε το ολοκλήρωμα δημιουργώντας τυχαίο δείγμα παρατηρήσεων $X_1,\ldots,X_n$ από την $\pi(x)$ και σύμφωνα με τον ΙΝΜΑ θα έχουμε ότι η ποσότητα $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$ θα συγκλίνει σχεδόν βέβαια στο ζητούμενο ολοκλήρωμα. Από μια άλλη θεώρηση αυτός ο εκτιμητής είναι ένας εκτιμητής δειγματοληψίας σπουδαιότητας κατά την οποία ισχύει ότι $\pi(x) = g(x)$. Η εκτίμηση που προκύπτει με 80.000 παρατηρήσεις από την $\frac{1}{2}\, e^{-|x|}$ είναι ίση με 2.013361 (με εκτιμώμενη διακύμανση 19.72315).

Προχωρώντας θα χρησιμοποιηθεί η δειγματοληψία σπουδαιότητας έχοντας ως $g(x)$ τη συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής με μέση τιμή 0 και διακύμανση 4, δηλαδή την $N(0,4)$. Υποθέτουμε ουσιαστικά ότι:

$$g(x) = \frac{1}{2\sqrt{2\pi}}\, e^{-(1/8)\,x^2},\ x \in \mathbb{R}.$$

Άρα, αν $Y_1,...,Y_n$ είναι ένα δείγμα από την $N(0,4)$ ο εκτιμητής του ολοκληρώματος με τη μέθοδο δειγματοληψίας σπουδαιότητας είναι ο (βλέπε σχέση (2.6)):

$$\hat{I}^{IS} = \frac{1}{n}\sum_{i=1}^{n}\frac{h(X_i)\pi(X_i)}{g(X_i)} = \frac{1}{n}\sum_{i=1}^{n} w(X_i)h(X_i),$$

όπου $w(X_i) = \dfrac{\pi(X_i)}{g(X_i)}$ αποτελούν τα βάρη σπουδαιότητας.

Η εκτίμηση που προκύπτει με 80.000 παρατηρήσεις από την $N(0,4)$ είναι ίση με 2.000568 (με εκτιμώμενη διακύμανση 19.0912) και είναι πλησιέστερη στην πραγματική τιμή του ολοκληρώματος. Τέλος, θα χρησιμοποιηθεί και ο κανονικοποιημένος εκτιμητής δειγματοληψίας σπουδαιότητας (βλέπε σχέση (2.10)):

$$\tilde{I}^{IS} = \frac{\sum_{i=1}^{n}(h(X_i)\pi(X_i)/g(X_i))}{\sum_{i=1}^{n}(\pi(X_i)/g(X_i))} = \frac{\sum_{i=1}^{n} h(X_i)w(X_i)}{\sum_{i=1}^{n} w(X_i)},$$

όπου $w(X_i) = \dfrac{\pi(X_i)}{g(X_i)}$ αποτελούν τα βάρη σπουδαιότητας. Στην περίπτωση αυτή, η εκτίμηση που προκύπτει με 80.000 παρατηρήσεις από την $N(0,4)$ είναι ίση με 2.001339, αλλά δεν είναι η πλησιέστερη στην πραγματική τιμή του ολοκληρώματος.

Τα παραπάνω υλοποιούνται στην $R$ με τις ακόλουθες εντολές:

```
library(rmutil)
set.seed(1)
X = rlaplace(80000)
h = X^2
ihat = mean(h)
ihat
[1] 2.013361
var(h)
[1] 19.72315
dif1 = 2-ihat    #dif1=I-ihat
dif1
[1] -0.01336148
Y = rnorm(80000, mean=0, sd=2)   # g=N(0,4)
h1 = Y^2
```

```
pi = function(Y) 0.5 *exp(-abs(Y))    # double
   exponential density
h1pi = function(Y) 0.5*(Y^2) *exp(-abs(Y))
integrate(h1pi, -Inf, +Inf)  # I=E_g(h1w)=E_g(h1pi/g)
2
w = 0.5*exp(-abs(Y))/dnorm(Y, mean=0, sd=2)
ihat2 = mean(h1*w)
ihat2
[1] 2.000568
var(h1*w)
[1] 19.0912
dif2 = 2-ihat2  #dif2=I-ihat2
dif2
[1] -0.0005677424
itilde = sum(h1*w)/sum(w)
itilde
[1] 2.001339
dif3=2-itilde
dif3
[1] -0.001339387
```

### Επιλογή κατανομής σπουδαιότητας

Υπάρχουν περιπτώσεις που κάποιος επιλέγει την κατανομή δειγματοληψίας σπουδαιότητας ανάμεσα σε ένα σύνολο πιθανών $\mathcal{G}$. Δηλαδή υποθέτουμε ότι η $g$ ανήκει σε μια (πιθανότατα) παραμετρική οικογένεια κατανομών και θέλουμε να επιλέξουμε τη βέλτιστη τιμή της παραμέτρου $\theta$. Για παράδειγμα στη μονοδιάστατη περίπτωση, όπως αναφέρει ο Shimkin (2015), είναι σύνηθες να υποθέτουμε ότι η συνάρτηση πυκνότητας πιθανότητας $g$ ανήκει στην εκθετική οικογένεια και γνωρίζοντας ότι κάθε μέλος της εκθετικής οικογένειας μπορεί με επανα-παραμετρικοποίηση να γραφεί στην κανονική μορφή της Natural Exponential Family (NEF) υποθέτουμε ότι $g(x;\theta) = c(\theta)\, e^{\theta\, t(x)}\, m(x)$, για γνωστές συναρτήσεις $t(x)$ και $m(x)$, $m(x) \geq 0$.

Από όσα προαναφέρθηκαν θα ήταν επιθυμητό η κατανομή δειγματοληψίας πιθανότητας $g(x;\theta)$ να είναι κοντά στην $g^*(x) = \dfrac{1}{c}\pi(x)|h\,(x)\,|$, όπου $c = E_\pi\left(|h\,(x)\,|\right)$ η σταθερά κανονικοποίησης. Ένα πολύ γνωστό μέτρο της εγγύτητας, απόκλισης δύο κατανομών είναι αυτό που προτάθηκε από τους Kullback and Leibler

(1951) και αναφέρεται στη βιβλιογραφία ως Kullback-Leibler απόκλιση. Επομένως, προκύπτει ότι θα πρέπει να προσδιοριστεί η παράμετρος $\theta$ έτσι ώστε να ελαχιστοποιείται ως προς $\theta$ η $D_{KL}\left(g^*, g\left(x;\theta\right)\right)$. Από τον ορισμό της K-L απόκλισης προκύπτει ότι:

$$\begin{aligned} D_{KL}\left(g^*(x), g\left(x;\theta\right)\right) &= \int g^*\left(x\right)\left(\ln\frac{g^*\left(x\right)}{g\left(x;\theta\right)}\right)\mathrm{d}x \\ &= \underbrace{\int g^*\left(x\right)\left(\ln g^*\left(x\right)\right)\mathrm{d}x}_{-H(g^*)} - \underbrace{\int g^*\left(x\right)\left(\ln g\left(x;\theta\right)\right)\mathrm{d}x}_{-H(g^*,g)}, \end{aligned}$$

όπου $H\left(g^*\right)$ είναι η εντροπία του Shannon (1948) της $g^*$, ενώ η ποσότητα $H\left(g^*, g\right)$ εκφράζει την cross entropy μεταξύ των $g$ και $g^*$. Επομένως, ο προσδιορισμός της παραμέτρου $\theta$ έτσι ώστε η $g(x;\theta)$ να είναι εγγύτερα στην $g^*(x)$ ανάγεται στο πρόβλημα (βλέπε Geyer et al. (2019)) που δίνεται συνοπτικά ως:

$$\min_{\theta\in\Theta}\{-H\left(g^*(x)\right) + H\left(g^*(x), g\left(x;\theta\right)\right)\} \tag{2.13}$$

ή ισοδύναμα

$$\min_{\theta\in\Theta}\{D_{KL}\left(g^*(x), g\left(x;\theta\right)\right)\} = \max_{\theta\in\Theta}\{-H\left(g^*(x), g\left(x;\theta\right)\right)\}.$$

Λαμβάνοντας υπόψη τη μορφή της $g^*$ έπειτα από λίγη άλγεβρα προκύπτει ότι:

$$\min_{\theta\in\Theta}\{D_{KL}\left(g^*(x), g\left(x;\theta\right)\right)\} = \max_{\theta\in\Theta} E_\pi\left(\left|h\left(X\right)\right| \ln g\left(X;\theta\right)\right).$$

# ΚΕΦΑΛΑΙΟ 3

# Βασικες μεθοδοι προσομοιωσης Markov Chain Monte Carlo

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν τεχνικές Monte Carlo, οι οποίες προσομοιώνουν μια τ.μ. $X$ από μία συγκεκριμένη κατανομή που λέγεται κατανομή στόχος. Αυτό μπορεί να συμβεί είτε απευθείας από την κατανομή στόχο είτε, όταν αυτό δεν είναι εφικτό, χρησιμοποιώντας μία βοηθητική κατανομή $g$. Στην ενότητα αυτή θα παρουσιαστούν μέθοδοι που ανήκουν στην κατηγορία των Markov Chain Monte Carlo (MCMC) και έχουν ως απώτερο στόχο την εκτίμηση του ολοκληρώματος $I = \int h(x)\pi(x)\,\mathrm{d}x$. Η κεντρική ιδέα αυτών είναι η αντικατάσταση των ανεξάρτητων και ισόνομων δειγμάτων των βασικών MC μεθόδων με εξαρτημένα δείγματα από μια εργοδική (ergodic) Μαρκοβιανή αλυσίδα, της οποίας η στάσιμη ή αλλιώς στατική (limiting, stationary) κατανομή είναι η κατανομή στόχος. Ουσιαστικά κάθε τεχνική MCMC στοχεύει στην παραγωγή εξαρτημένων δειγμάτων από μια κατανομή στόχο.

Στο πλαίσιο αυτό θα δοθούν τα βασικά χαρακτηριστικά της μεθόδου MCMC και έπειτα θα περιγραφούν κάποιοι γνωστοί MCMC αλγόριθμοι. Για την αναλυτική παρουσίαση και εφαρμογή τέτοιων μεθόδων υπάρχει εκτενής βιβλιογραφία και ενδεικτικά παραπέμπουμε στους Gelfand et al. (1996), Chen et al. (2000), Berg (2004), Gamerman and Lopes (2006), Wasserman (2013) και Rubinstein and Kroese (2016).

Στη συνέχεια δίνεται ένας γενικός ορισμός για τις τεχνικές MCMC (βλέπε, μεταξύ άλλων, Malefaki and Iliopoulos (2008)).

**Ορισμός 3.0.1.** *Μία τεχνική Markov Chain Monte Carlo (MCMC) για την προσομοίωση από την κατανομή $\pi$ είναι οποιαδήποτε μέθοδος παράγει μία εργοδική Μαρκοβιανή αλυσίδα $X_t$, της οποίας η στάσιμη κατανομή είναι η $\pi$.*

Η βασική ιδέα, λοιπόν, αυτών των μεθόδων είναι η κατασκευή μιας Μαρκοβια-

νής αλυσίδας $\{X_t\}$, $t \geq 0$, με στάσιμη κατανομή $\pi$. Είναι γνωστό ότι η στάσιμη κατανομή υπάρχει αν όλες οι καταστάσεις είναι θετικώς επαναληπτικές και αντίστροφα. Κατά αυτόν τον τρόπο θα μπορούν να υπολογιστούν τα χαρακτηριστικά που επιθυμούμε (π.χ. μέση τιμή, διακύμανση κ.ο.κ.) μέσω των δειγματικών τιμών της Μαρκοβιανής αλυσίδας που κατασκευάστηκε. Για παράδειγμα, μέσω του Ασθενή Νόμου των Μεγάλων Αριθμών για Μαρκοβιανές αλυσίδες (βλέπε Wasserman (2013)) θα έχουμε ότι:

$$\hat{I}^{MCMC} = \frac{1}{n}\sum_{t=0}^{n-1} h\left(X_t\right) \xrightarrow{p} E_\pi\left(h\left(X\right)\right) = I, \text{ καθώς } n \to \infty. \tag{3.1}$$

Στο σημείο αυτό επισημαίνεται ότι εφόσον είμαστε στο πλαίσιο των Μαρκοβιανών αλυσίδων, ο εκτιμητής της σχέσης (3.1) λέγεται εργοδικός μέσος (και όχι δειγματικός μέσος). Επιπλέον, υπό κατάλληλες υποθέσεις (βλέπε, για παράδειγμα, Malefaki and Iliopoulos (2008)), για μια μη διαχωρίσιμη Μαρκοβιανή αλυσίδα $\{X_t\}$ με στατική κατανομή $\pi$ ισχύει ότι:

$$\sqrt{n}\left(\hat{I}^{MCMC} - I\right) \xrightarrow{d} N\left(0, {\sigma_h}^2\right),$$

όπου ${\sigma_h}^2 = \sigma^2\left(1 + 2\sum_{i=1}^{+\infty}\rho_i\right)$, με $\sigma^2 = Var_\pi(h(X))$, $\rho_i = Corr_\pi(h(X_0), h(X_i))$, $i = 1, 2, \ldots$.

Ακολούθως θα παρουσιαστούν οι πιο ευρέως χρησιμοποιούμενοι MCMC αλγόριθμοι, ο αλγόριθμος Metropolis-Hastings και η δειγματοληψία κατά Gibbs.

## 3.1 Αλγόριθμος Metropolis-Hastings

Στην ενότητα αυτή παρουσιάζεται ο αλγόριθμος Metropolis-Hastings (MH), που είναι όχι μόνο ο πιο γνωστός αλγόριθμος Monte Carlo βασισμένος σε Μαρκοβιανές αλυσίδες, αλλά ίσως και γενικότερα ο πιο γνωστός αλγόριθμος Monte Carlo. Ο αλγόριθμος αυτός αρχικά προτάθηκε από τους Metropolis et al. (1953) για συμμετρικές κατανομές πρότασης, δηλαδή για κατανομές πρότασης που, όπως θα δούμε και παρακάτω, πληρούν τη σχέση $q\left(y \mid x_t\right) = q\left(x_t \mid y\right)$.. Παρόλα αυτά ο αλγόριθμος γίνεται γνωστός στο χώρο της στατιστικής μετά την αναπαραγωγή του στο σύγγραμμα των Hammersley and Handscomb (1964), όπου μελετήθηκε ως μία τεχνική υπολογισμού πολύπλοκων ολοκληρωμάτων. Στη συνέχεια ο αλγόριθμος επεκτάθηκε από τον Hastings (1970) ως μια πιο γενική διαδικασία

προσομοίωσης και για μη συμμετρικές περιπτώσεις κατανομής πρότασης. Ο τελευταίος αλγόριθμος είναι και αυτός που έχει καθιερωθεί στη βιβλιογραφία ως MH αλγόριθμος. Παρότι μέχρι και τη δεκαετία του 1990 ο αλγόριθμος MH δεν ήταν ευρέως διαδεδομένος, οι Gelfand and Smith (1990), εκμεταλλευόμενοι τη διαρκώς αυξανόμενη υπολογιστική ισχύ του αλγορίθμου, έδωσαν την πρώτη ώθηση ώστε σήμερα ο MH να θεωρείται ένα από τα σπουδαιότερα εργαλεία της Υπολογιστικής Στατιστικής, που χρησιμοποιείται σε ένα τεράστιο εύρος επιστημονικών πεδίων, όπως η Στατιστική, η Φυσική, η Χημεία, η Βιολογία και η Οικονομία. Για περισσότερες σχετικές λεπτομέρειες παραπέμπουμε, μεταξύ άλλων, στον Kass (1997).

Κύριο ρόλο στη μελέτη του MH αλγορίθμου διαδραματίζει η κεντρική του ιδέα, η οποία είναι ότι γεννά μια ακολουθία δειγματικών τιμών με τέτοιο τρόπο ώστε όσο περισσότερες δειγματικές τιμές παράγονται τόσο καλύτερα προσεγγίζεται η κατανομή στόχος $\pi$ από την κατανομή των τιμών. Επίσης, οι δειγματικές αυτές τιμές προσομοιώνονται με βάση μια συγκεκριμένη επαναληπτική διαδικασία (όπως θα εξηγηθεί και ακολούθως) με την ιδιότητα ότι η επόμενη δειγματική τιμή εξαρτάται μόνο από την τρέχουσα δειγματική τιμή και όχι από τις προηγούμενες. Ακολουθώντας αυτήν τη δομή, εύκολα διαπιστώνεται ότι η ακολουθία τιμών που παράγεται αποτελεί μια Μαρκοβιανή αλυσίδα. Πιο συγκεκριμένα, σε κάθε επανάληψη του αλγορίθμου λαμβάνοντας υπόψη μόνο την τρέχουσα δειγματική τιμή επιλέγεται από κάποια κατανομή πρότασης μία τιμή ως υποψήφια για την επόμενη δειγματική τιμή. Έπειτα για την υποψήφια αυτή τιμή υπάρχουν δύο επιλογές: είτε να γίνει δεκτή είτε να απορριφθεί, με κάποια πιθανότητα αποδοχής ή απόρριψης, αντίστοιχα. Στην πρώτη περίπτωση η τιμή που μόλις έγινε δεκτή χρησιμοποιείται στην αμέσως επόμενη επανάληψη του αλγορίθμου, ενώ στη δεύτερη περίπτωση η υποψήφια τιμή δεν είναι δεκτή και έτσι η τρέχουσα τιμή επαναχρησιμοποιείται για την επόμενη επανάληψη. Ο αλγόριθμος περιγράφεται αναλυτικά στη συνέχεια.

Αρχικά τη χρονική στιγμή $t = 0$ επιλέγεται τυχαία και αυθαίρετα μια αρχική κατάσταση $X_0 = x_0$, έτσι ώστε $\pi(x_0) > 0$. Θεωρώντας ότι έχουν παραχθεί τα $X_1, X_2, ..., X_t$ και δεδομένου ότι τη χρονική στιγμή $t$ βρισκόμαστε στην κατάσταση $X_t = x_t$ περιγράφονται στη συνέχεια, δίνοντας όλες τις λεπτομέρειες και επεξηγήσεις, τα βήματα του αλγορίθμου (βλέπε, μεταξύ άλλων, Wasserman (2013)).

**Metropolis-Hastings Algorithm (MHA)**

- **Βήμα 1:** Με δεδομένη την τρέχουσα κατάσταση $X_t = x_t$ παράγεται μια παρατήρηση $y$ από την τ.μ. $Y \sim q(y \mid X_t)$.
  Η $q(y \mid x_t)$ είναι μια κατανομή από την οποία είναι εφικτή και πιο εύκολη

η διαδικασία της προσομοίωσης. Η δεσμευμένη αυτή συνάρτηση πυκνότητας πιθανότητας $q(y \mid x_t)$ αποτελεί τη λεγόμενη **κατανομή πρότασης** (proposal distribution) και είναι αυτής της μορφής διότι τις περισσότερες φορές εξαρτάται από την προηγούμενη κατάσταση $x_t$.

- **Βήμα 2:** Στη συνέχεια υπολογίζεται η ποσότητα:

$$r(x_t, y) = \min\left\{\frac{\pi(y)\, q(x_t \mid y)}{\pi(x_t)\, q(y \mid x_t)}, 1\right\}. \tag{3.2}$$

με $\pi(x_t) > 0$, $\forall\, t \in \mathbb{N}$, αφού $\pi(x_0) > 0$ (από υπόθεση).
Η ποσότητα $r(x_t, y)$ που δίνεται στη σχέση (3.2) ονομάζεται **πιθανότητα αποδοχής** (acceptance probability) ή αλλιώς **συνάρτηση αποδοχής** (acceptance function ή acceptance ratio) και όπως θα γίνει αντιληπτό στο επόμενο βήμα είναι η πιθανότητα αποδοχής της προτεινόμενης, υποψήφιας κατάστασης $y$.

- **Βήμα 3:** Προσομοιώνεται ένα τυχαίο u από την ομοιόμορφη κατανομή $U(0,1)$. Τότε, η νέα τιμή της Μαρκοβιανής αλυσίδας είναι:

$$x_{t+1} = \begin{cases} y, & \text{αν } u \leq r(x_t, y) \\ x_t, & \text{αν } u > r(x_t, y). \end{cases} \tag{3.3}$$

Επομένως, αν η νέα υποψήφια τιμή $y$ της Μαρκοβιανής αλυσίδας γίνει δεκτή τότε θέτουμε $x_{t+1} = y$ και συνεχίζουμε στην επόμενη επανάληψη, ενώ αν απορριφθεί τότε θέτουμε $x_{t+1} = x_t$ και συνεχίζουμε στην επόμενη επανάληψη. Οι δειγματικές τιμές οι οποίες απορρίπτονται κατά τη διάρκεια του αλγόριθμου είναι γνωστές στη βιβλιογραφία και ως **burn-in δειγματικές τιμές** (βλέπε Gelman et al. (2004)).

- **Βήμα 4:** Θέτουμε $t = t + 1$ και επιστρέφουμε στο Βήμα 1.

**Παρατήρηση 3.1.1.** *Από τον αλγόριθμο που δόθηκε αναγνωρίζεται εύκολα η δομή της μεθόδου αποδοχής-απόρριψης που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Η κύρια διαφορά των δύο τεχνικών είναι ότι ο αλγόριθμος MH δεν επαναλαμβάνεται αν η δειγματική τιμή απορριφθεί, αλλά αποδέχεται την τωρινή κατάσταση ως μια νέα τιμή. Στο πλαίσιο αυτό αξίζει να σημειωθεί ότι η πιθανότητα η Μαρκοβιανή αλυσίδα να παραμείνει στην ίδια κατάσταση (δηλαδή $x_{t+1} = x_t$) είναι $1 - r(x_t, y)$ . Αυτό σημαίνει ότι η σχέση (3.3) ισοδύναμα μπορεί να γραφεί ως:*

$$x_{t+1} = \begin{cases} y, & \textit{με πιθανότητα } r(x_t, y) \\ x_t, & \textit{με πιθανότητα } 1 - r(x_t, y). \end{cases} \tag{3.4}$$

**Παρατήρηση 3.1.2.** *Στον αλγόριθμο Metropolis-Hastings η επιλογή της κατανομής πρότασης είναι σημαντική, καθώς θα πρέπει να είναι εύκολη η προσομοίωση από αυτήν. Επιπλέον η επιλογή της προφανώς επηρεάζει και την ταχύτητα σύγκλισης του αλγορίθμου. Επομένως, σε ένα γενικό πρόβλημα, δεν είναι ξεκάθαρο ποια κατανομή πρότασης $q$ πρέπει να χρησιμοποιηθεί. Ο αλγόριθμος είναι αποδοτικότερος όταν είναι πιο κοντά στην κατανομή στόχο, όταν δηλαδή το σχήμα της κατανομής πρότασης ταιριάζει με το σχήμα της κατανομής στόχου. Κάποιες φορές επιλέγεται ως κατανομή πρότασης η κανονική κατανομή με μέση τιμή $x_t$ και διακύμανση $\sigma^2$. Στην περίπτωση που η διακύμανση $\sigma^2$ είναι πολύ μικρή, η Μαρκοβιανή αλυσίδα θα αναμιχθεί αργά (mix slowly), δηλαδή παρότι το ποσοστό αποδοχής θα είναι υψηλό τα διαδοχικά δείγματα θα μετακινηθούν αργά γύρω από το χώρο, και έτσι η αλυσίδα θα συγκλίνει πολύ αργά στην $\pi$. Από την άλλη μεριά, αν η $\sigma^2$ είναι πολύ μεγάλη, τότε το ποσοστό αποδοχής θα είναι πολύ χαμηλό, επομένως και πάλι η αλυσίδα θα συγκλίνει πολύ αργά. Στην πράξη οι παράμετροι επιλέγονται έτσι ώστε ο αλγόριθμος να δέχεται περίπου το 30% όλων των δειγματικών τιμών, παρότι σύμφωνα με τους Roberts et al. (1997), έχει αποδειχθεί θεωρητικά ότι το ιδανικό ποσοστό αποδοχής για μια μονοδιάστατη κανονική κατανομή είναι περίπου 50%.*

**Παρατήρηση 3.1.3.** *Αν η κατανομή πρότασης δεν εξαρτάται από την κατάσταση $x_t$, δηλαδή αν $q(y \mid x_t)$ ανεξάρτητη από τη $x_t$, τότε προκύπτει η ειδική περίπτωση που είναι γνωστή στη βιβλιογραφία ως* ***ανεξάρτητος δειγματολήπτης MH*** *(Independent Sampler MH). Σ' αυτήν την περίπτωση η τεχνική που προκύπτει είναι ουσιαστικά μία παραλλαγή δειγματοληψίας σπουδαιότητας των MCMC μεθόδων. Παρά το γεγονός ότι η $X_t$ παράγεται ανεξάρτητα από την $Y$, η ακολουθία που προκύπτει δεν αποτελείται από ανεξάρτητες παρατηρήσεις, αφού η πιθανότητα αποδοχής της $Y$ εξαρτάται από την τιμή της $X_t$. Ο ανεξάρτητος δειγματολήπτης MH μπορεί να συγκλίνει πολύ γρήγορα, μπορεί όμως να οδηγήσει και σε αλυσίδες με υπερβολικά αργή σύγκλιση. Για να είναι αποτελεσματική αυτή η διαδικασία θα πρέπει η κατανομή πρότασης να έχει πιο βαριές ουρές σε σχέση με την κατανομή στόχο. Λαμβάνοντας υπόψη όσα προαναφέρθηκαν, ο ανεξάρτητος δειγματολήπτης MH είναι ένας εναλλακτικός αλγόριθμος δύο τεχνικών Monte Carlo που μελετήθηκαν στο προηγούμενο κεφάλαιο, της μεθόδου αποδοχής-απόρριψης αλλά και της δειγματοληψίας σπουδαιότητας. Ειδικότερα, σύμφωνα με τον Liu (1996), η συνολική σύγκριση των δύο αυτών τεχνικών με τον ανεξάρτητο δειγματολήπτη MH επιφέρει χρήσιμα αποτελέσματα. Ο Liu (1996) απέδειξε ότι η ασυμπτωτική διασπορά της μεθόδου αποδοχής-απόρριψης είναι μικρότερη από αυτήν του ανεξάρτητου δειγματολήπτη MH, αλλά είναι ασυμπτωτικά συγκρίσιμες. Έδειξε επίσης ότι ανάλογα με την υπό ολοκλήρωση συνάρτηση, η μέθοδος της δειγματοληψίας σπουδαιότητας μπορεί να είναι ασυμπτωτικά αποδοτικότερη από τις άλλες*

*δυο τεχνικές. Επιπρόσθετα, ο λεγόμενος* **τυχαίος περίπατος MH** *(Random Walk MH) αποτελεί μια άλλη ειδική περίπτωση του αλγορίθμου MH, η οποία προκύπτει θεωρώντας συμμετρική κατανομή πρότασης. Πιο συγκεκριμένα, στο Βήμα 2 του MH αλγορίθμου επιλέγεται ως κατανομή πρότασης κάποια που πληροί τη σχέση* $q(y \mid x_t) = q(x_t \mid y)$*. Αυτό έχει ως συνέπεια το πηλίκο στο Βήμα 2 να είναι διαφορετικό και να δίνεται, σύμφωνα με τους Metropolis et al. (1953), από τη σχέση* $r(x_t, y) = \min\left\{\frac{\pi(y)}{\pi(x_t)}, 1\right\}$*. Συνήθως, οι κατανομές σπουδαιότητας που χρησιμοποιούνται είναι συμμετρικές με κέντρο κάθε φορά την προηγούμενη κατάσταση (θα γίνει αντιληπτό στη συνέχεια μέσω ενός απλού παραδείγματος). Κλείνοντας, αξίζει να επισημανθεί ότι, σε σύγκριση με τον ανεξάρτητο δειγματολήπτη MH, ο τυχαίος περίπατος MH έχει ορισμένα μειονεκτήματα τα οποία πρέπει να αναφερθούν. Το πρώτο μειονέκτημα είναι ότι η συμμετρία της κατανομής πρότασης του τυχαίου περίπατου έχει ως συνέπεια να "ξοδεύει" αρκετό χρόνο προσομοίωσης επανεξετάζοντας περιοχές τις οποίες έχει ήδη επισκεφτεί. Για την αντιμετώπιση τέτοιων προβλημάτων έχουν προταθεί διάφορες εναλλακτικές που αφενός εξασφαλίζουν σε λιγότερο χρόνο την αποτελεσματικότητα του αλγορίθμου, αφετέρου είναι δύσκολες στην εφαρμογή τους. Για περισσότερες πληροφορίες σχετικά παραπέμπουμε ενδεικτικά στους Roberts and Rosenthal (1998). Το δεύτερο, αλλά εξίσου σημαντικό, μειονέκτημα είναι ότι ο αλγόριθμος απαιτεί μεγάλο πλήθος επαναλήψεων για να παρακάμψει κάποιες δυσκολίες που προκύπτουν, όπως για παράδειγμα διάφορες εκτελέσιμες περιοχές του αλγορίθμου με μικρή πιθανότητα. Συμπερασματικά, τα επιχειρήματα αυτά αιτιολογούν το γεγονός ότι ο τυχαίος περίπατος MH δεν αποτελεί πάντοτε την πιο αποτελεσματική επιλογή παρά την απλότητα και την επεκτασιμότητα της μεθόδου.*

**Πρόταση 3.1.4.** *Για την πιθανότητα αποδοχής Metropolis ισχύει ότι είτε* $r(x_t, y) = 1$ *ή* $r(y, x_t) = 1$ *(βλέπε, μεταξύ άλλων, Wasserman (2013)).*

**Απόδειξη.** Για τις καταστάσεις $x_t, y$ προφανώς ισχύει ότι είτε $\pi(x_t)\, q(y \mid x_t) > \pi(y)\, q(x_t \mid y)$ είτε $\pi(x_t)\, q(y \mid x_t) < \pi(y)\, q(x_t \mid y)$, καθώς η περίπτωση της ισότητας έχει αποκλειστεί αφού η πιθανότητα εμφάνισής της είναι μηδενική στις συνεχείς κατανομές.

Χωρίς βλάβη της γενικότητας υποθέτουμε ότι $\pi(x_t)\, q(y \mid x_t) > \pi(y)\, q(x_t \mid y)$, που συνεπάγεται ότι $\frac{\pi(y)\, q(x_t \mid y)}{\pi(x_t)\, q(y \mid x_t)} < 1$. Συνεπώς, από τη σχέση (3.2) προκύπτει ότι $r(x_t, y) = \frac{\pi(y)\, q(x_t \mid y)}{\pi(x_t)\, q(y \mid x_t)}$. Δηλαδή $r(y, x_t) = 1$.

Με παρόμοιο τρόπο αν $\pi(x_t)\, q(y \mid x_t) < \pi(y)\, q(x_t \mid y)$ προκύπτει ότι $r(x_t, y) = 1$. □

Γίνεται εύκολα αντιληπτό από τη σχέση (3.3) ότι η τελική τιμή $x_{t+1}$ εξαρτάται μόνο από την τιμή $x_t$ και όχι από τις προηγούμενες τιμές. Αυτό έχει ως αποτέλεσμα να εξασφαλίζεται η Μαρκοβιανή ιδιότητα, που σημαίνει ότι η ακολουθία που παράγεται από τον αλγόριθμο MH ουσιαστικά είναι προσομοίωση μίας Μαρκοβιανής αλυσίδας. Μένει στη συνέχεια να αποδειχθεί ότι ο αλγόριθμος MH προσομοιώνει πράγματι μια Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την κατανομή στόχο $\pi$. Μια επαρκής αλλά όχι αναγκαία συνθήκη για να είναι η Μαρκοβιανή αλυσίδα με πίνακα μετάβασης $P$ αντιστρέψιμη (reversible) και να έχει στάσιμη κατανομή την $\pi$ είναι η:

$$\pi\left(x_t\right) P\left(x_t, y\right) = \pi\left(y\right) P\left(y, x_t\right), \forall\, x_t \neq y. \tag{3.5}$$

Η παραπάνω σχέση, η οποία είναι γνωστή ως **λεπτομερής συνθήκη ισορροπίας** (detailed balance condition), ισοδύναμα γράφεται:

$$\frac{P\left(x_t, y\right)}{P\left(y, x_t\right)} = \frac{\pi\left(y\right)}{\pi\left(x_t\right)}, \forall\, x_t \neq y.$$

**Παρατήρηση 3.1.5.** *Υπενθυμίζεται ότι η $\pi$ είναι στάσιμη κατανομή της Μαρκοβιανής αλυσίδας με πίνακα μετάβαση $P$ αν ισχύει $\pi(x_t) = \int \pi(y) P\left(y, x_t\right)\, \mathrm{d}y$. Υπό τη λεπτομερή συνθήκη ισορροπίας διαπιστώνεται ότι:*

$$\int \pi(y) P\left(y, x_t\right)\, \mathrm{d}y = \int \pi(x_t) P\left(x_t, y\right)\, \mathrm{d}y = \pi(x_t) \underbrace{\int P\left(x_t, y\right)\, \mathrm{d}y}_{=1} = \pi(x_t).$$

*Πράγματι λοιπόν, η λεπτομερής συνθήκη ισορροπίας είναι μια επαρκής συνθήκη για να είναι η Μαρκοβιανή αλυσίδα με πίνακα μετάβασης $P$ αντιστρέψιμη (reversible) και να έχει στάσιμη κατανομή την $\pi$.*

Στην επόμενη πρόταση αποδεικνύεται ότι τα παραπάνω βήματα του αλγορίθμου MH οδηγούν σε μια Μαρκοβιανή αλυσίδα με στάσιμη κατανομή την κατανομή στόχο (βλέπε, μεταξύ άλλων, Robert and Casella (2004)).

**Πρόταση 3.1.6.** *Η Μαρκοβιανή αλυσίδα που παράγεται από τον αλγόριθμο MH έχει μοναδική στάσιμη κατανομή την κατανομή στόχο $\pi$.*

**Απόδειξη.** Από την Πρόταση 3.1.4 προκύπτει το συμπέρασμα ότι είτε $r(x_t, y) = 1$ είτε $r(y, x_t) = 1$. Σε όσα ακολουθούν χωρίς βλάβη της γενικότητας υποθέτουμε ότι $r(y, x_t) = 1$. Αρχικά θα δείξουμε ότι ισχύει η λεπτομερής συνθήκη ισορροπίας, δηλαδή ότι ικανοποιείται η σχέση (3.5). Είναι $P\left(x_t, y\right)$ η πιθανότητα μετάβασης

ενός βήματος από την κατάσταση $x_t$ στην κατάσταση $y$. Παρατηρούμε ότι στη σχέση (3.5) έχουμε ότι $x_t \neq y$, γεγονός που σημαίνει ότι η τιμή $y$ έχει παραχθεί από την κατανομή $q(y|x_t)$ και έχει γίνει αποδεκτή. Άρα είναι:

$$P\left(x_t, y\right) = q(y \mid x_t)\, r\left(x_t, y\right), \tag{3.6}$$

Καθώς υποθέσαμε ότι $r(y, x_t) = 1$, αυτό σημαίνει ότι σε αυτήν την περίπτωση $r(x_t, y) = \dfrac{\pi\left(y\right) q\left(x_t \mid y\right)}{\pi\left(x_t\right) q\left(y \mid x_t\right)}$. Επομένως:

$$P\left(x_t, y\right) = q(y \mid x_t)\, r\left(x_t, y\right) = q(y \mid x_t) \frac{\pi\left(y\right) q\left(x_t \mid y\right)}{\pi\left(x_t\right) q\left(y \mid x_t\right)} = q\left(x_t \mid y\right) \frac{\pi\left(y\right)}{\pi\left(x_t\right)}.$$

Άρα, $\pi\left(x_t\right) P\left(x_t, y\right) = \pi\left(y\right) q\left(x_t \mid y\right)$.

Με ανάλογο τρόπο έπεται ότι:

$$P\left(y, x_t\right) = q(x_t \mid y) \underbrace{r\left(y, x_t\right)}_{=1} = q(x_t \mid y),$$

και άρα $\pi\left(y\right) P\left(y, x_t\right) = \pi\left(y\right) q\left(x_t \mid y\right)$.

Από τα παραπάνω συνεπάγεται ότι $\pi\left(x_t\right) P\left(x_t, y\right) = \pi\left(y\right) P\left(y, x_t\right)$, $\forall\, x_t \neq y$ (λεπτομερής συνθήκη ισορροπίας) και επομένως η στάσιμη κατανομή της Μαρκοβιανής αλυσίδας που παράγεται από τον αλγόριθμο είναι όντως η κατανομή στόχος $\pi$. Πρέπει, όμως, να αποδειχτεί και η μοναδικότητα της στάσιμης κατανομής $\pi$. Αυτό εξασφαλίζεται από την εργοδικότητα της Μαρκοβιανής αλυσίδας η οποία απαιτεί ότι κάθε κατάσταση πρέπει να είναι απεριοδική (το σύστημα δεν επιστρέφει στην ίδια κατάσταση σε σταθερά διαστήματα) και θετικώς επαναληπτική (ο αναμενόμενος αριθμός βημάτων για την επιστροφή στην ίδια κατάσταση είναι πεπερασμένος). □

**Παρατήρηση 3.1.7.** *Στη βιβλιογραφία έχουν παρουσιαστεί παραλλαγές του MH αλγορίθμου τροποποιώντας τη συνάρτηση αποδοχής. Ειδικότερα, ο Barker (1965) πρότεινε τη συνάρτηση αποδοχής*

$$r_{_B}\left(x_t, y\right) = \frac{\pi\left(y\right) q\left(x_t \mid y\right)}{\pi\left(x_t\right) q\left(y \mid x_t\right) + \pi\left(y\right) q\left(x_t \mid y\right)}, \tag{3.7}$$

*η οποία στην ειδική περίπτωση της συμμετρίας, δηλαδή όταν* $q\left(x_t \mid y\right) = q\left(y \mid x_t\right)$ *λαμβάνει τη μορφή*

$$r_{_B}^{s}\left(x_t, y\right) = \frac{\pi\left(y\right)}{\pi\left(x_t\right) + \pi\left(y\right)}.$$

*Λίγα χρόνια αργότερα ο Charles Stein (βλέπε, μεταξύ άλλων, Liu (1999) και Hitchcock (2003)) πρότεινε την εξής μορφή:*

$$r_{CS}\left(x_t, y\right) = \frac{\delta\left(x_t, y\right)}{\pi\left(x_t\right) q\left(y \mid x_t\right)}, \tag{3.8}$$

*με $\delta\left(x_t, y\right)$ να είναι οποιαδήποτε συμμετρική συνάρτηση που είναι τέτοια ώστε $r_{CS}\left(x_t, y\right) \leq 1$, $\forall\, x_t, y$.*

**Παρατήρηση 3.1.8.** *Συνοψίζοντας τις πιο σημαντικές ιδιότητες του αλγορίθμου MH έχουμε ότι η Μαρκοβιανή αλυσίδα αναπτύσσεται χωρίς να προσομοιώσουμε κατευθείαν από την $\pi$ και ότι ο αλγόριθμος MH είναι πολύ απλός, καθώς χρειάζεται μόνο η συναρτησιακή μορφή των $\pi$, $q$ και τέλος πετυχαίνουμε σύγκλιση της προσομοιωμένης αλυσίδας στην $\pi$. Το βασικότερο μειονέκτημά του είναι ότι για μονοδιάστατες κατανομές δεν έχει άμεσα αποτελέσματα, για απευθείας ανεξάρτητα δείγματα.*

Στη συνέχεια μέσω παραδειγμάτων θα δούμε πώς μπορεί να χρησιμοποιηθεί ο αλγόριθμος MH στην πράξη (βλέπε Robert and Casella (2010), Wasserman (2013) και Stephens (2017)).

**Παράδειγμα 3.1.9.** *Να δημιουργηθεί μια ακολουθία 800 τιμών από την Cauchy κατανομή, χρησιμοποιώντας τον αλγόριθμο MH.*

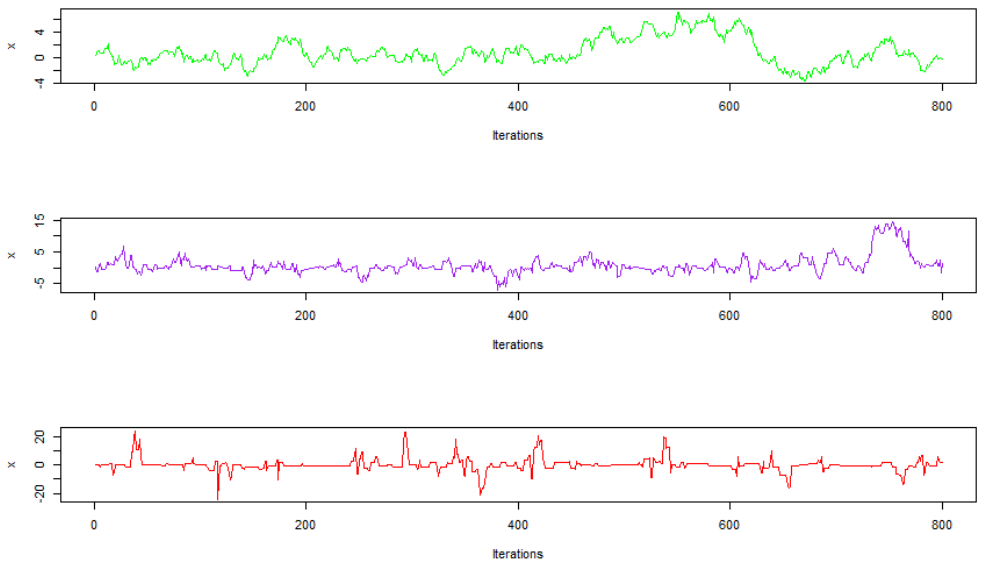
***Λύση***. Η σ.π.π. της κατανομής στόχου είναι η $\pi(x) = \dfrac{1}{\pi\left(1+x^2\right)}$, $x \in \mathbb{R}$.

Πρώτα απ΄ όλα επιλέγεται αυθαίρετα μια τρέχουσα κατάσταση $X_0 = x_0$ για την οποία ισχύει $\pi(x_0) > 0$. Εμείς επιλέγουμε αυθαίρετα την τιμή $x_0 = 0.4$. Προχωρώντας, θα πρέπει να επιλέξουμε και μια κατανομή πρότασης (βλέπε Βήμα 1 της μεθόδου). Στο παράδειγμα αυτό, θα επιλέξουμε την κανονική κατανομή με μέση τιμή την τωρινή κατάσταση και διακύμανση $\sigma^2$ για τιμές του $\sigma = 0.5$, 2 και 10. Η επιλογή αυτή έχει ως αποτέλεσμα, καθώς η κατανομή πρότασης είναι συμμετρική, δηλαδή $q\left(x_t \mid y\right) = q\left(y \mid x_t\right)$, να προκύπτει (βλέπε Βήμα 2 της μεθόδου) ότι: $r(x_t, y) = min\left\{\dfrac{\pi(y)}{\pi(x_t)}, 1\right\} = min\left\{\dfrac{1+x_t^2}{1+y^2}, 1\right\}$.

Τα παραπάνω υλοποιούνται στην R με τον κώδικα που δίνεται στη συνέχεια (βλέπε για παρόμοιο αλγόριθμο στο σύγγραμμα των Robert and Casella (2010)). Στο Σχήμα 3.1 απεικονίζονται οι προσομοιωμένες τιμές των τριών Μαρκοβιανών αλυσίδων με στάσιμη κατανομή την Cauchy και κατανομές πρότασης τις $N(x_t, \sigma^2)$ για τις διάφορες τιμές του $\sigma$ ($\sigma = 0.5$, $\sigma = 2$ και $\sigma = 10$). Παρατηρούμε ότι για μεγάλες τιμές του $\sigma$ (εδώ $\sigma = 10$) ο αλγόριθμος απορρίπτει πολλές από τις

προτεινόμενες τιμές με αποτέλεσμα να παραμείνει στην ίδια κατάσταση για αρκετό χρόνο. Συνεπώς, προκύπτουν μεγάλα άλματα μέσα στο χώρο καταστάσεων. Από την άλλη πλευρά μικρές τιμές της παραμέτρου κλίμακας (εδώ $\sigma = 0.5$) μας δίνουν μικρά άλματα στο χώρο καταστάσεων (η αλυσίδα δεν καλύπτει μεγάλο εύρος εξερεύνησης στο χώρο, βλέπε Wasserman (2013)), όμως ο αλγόριθμος θα δέχεται σχεδόν όλες τις προτεινόμενες τιμές.

Σχήμα 3.1: Γραφικές παραστάσεις 800 επαναλήψεων από τον τυχαίο περίπατο MH με κατανομή στόχο την Cauchy και κατανομές πρότασης τις $N(x_t, (0.5)^2)$, $N(x_t, 2^2)$ και $N(x_t, 10^2)$ αντίστοιχα.




```
n=800
x=rep(0,n)
x[1]=0.4
for (i in 2:n){
  proposedx=rnorm(1,mean=x[i-1],sd=0.5)
  currentx = x[i-1]
  r = min (dcauchy(proposedx,0,1)/dcauchy(currentx,0,
     1),1)
  if(runif(1)<r){
```

```
    x[i] = proposedx
  } else {
    x[i] = currentx
  }
}
par(mfrow=c(3,1))
plot(x,type="l",xlab = "Iterations",col="green")
for (i in 2:n){
  proposedx=rnorm(1,mean=x[i-1],sd=2)
  currentx = x[i-1]
  r = min (dcauchy(proposedx,0,1)/dcauchy(currentx,0,
     1),1)
  if(runif(1)<r){
    x[i] = proposedx
  } else {
    x[i] = currentx
  }
}
plot(x,type="l",xlab = "Iterations",col="purple")
for (i in 2:n){
  proposedx=rnorm(1,mean=x[i-1],sd=10)
  currentx = x[i-1]
  r = min (dcauchy(proposedx,0,1)/dcauchy(currentx,0,
     1),1)
  if(runif(1)<r){
    x[i] = proposedx
  } else {
    x[i] = currentx
  }
}
plot(x,type="l",xlab = "Iterations",col="red")
```

**Παράδειγμα 3.1.10.** *Χρησιμοποιώντας τον αλγόριθμο MH να προταθεί ένας τρόπος προσομοίωσης 9.000 δειγματικών τιμών από τη Βήτα κατανομή, με παραμέτρους* $a = 3$ *και* $b = 4$.

***Λύση***. Η κατανομή στόχος έχει συνάρτηση πυκνότητας πιθανότητας

$$\pi(x) = \frac{x^{3-1}(1-x)^{4-1}}{B(3,4)},\ x \in [0,1]. \tag{3.9}$$

Αρχικά επιλέγουμε αυθαίρετα μια τρέχουσα κατάσταση $X_0 = x_0$ που είναι τέτοια ώστε $\pi(x_0) > 0$. Εμείς επιλέγουμε αυθαίρετα την κατάσταση $x_0 = 0.3$. Στη συνέχεια θα πρέπει να επιλέξουμε μια κατανομή πρότασης. Στο παράδειγμα αυτό, για την ευκολότερη κατανόηση της μεθόδου, επιλέγουμε ως κατανομή πρότασης την $U(0,1)$. Δηλαδή επιλέγουμε μια κατανομή πρότασης ανεξάρτητη από την τωρινή κατάσταση. Η επιλογή αυτή έχει ως αποτέλεσμα (βλέπε Βήμα 2 της μεθόδου) να είναι: $r(x_t, y) = min\left\{\frac{\pi(y)}{\pi(x_t)}, 1\right\}$.

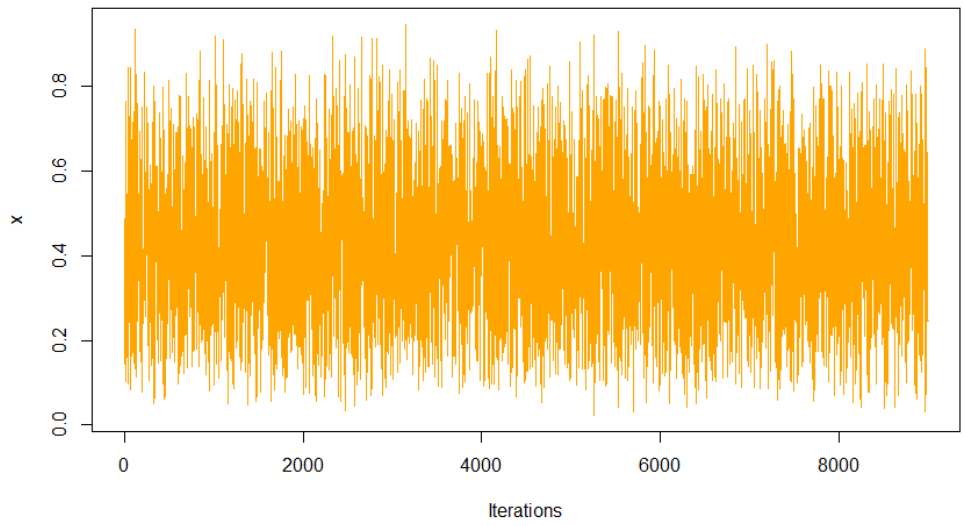
Τα παραπάνω υλοποιούνται με τις εντολές που δίνονται παρακάτω. Ταυτόχρονα, στο Σχήμα 3.2 απεικονίζονται οι 9.000 δειγματικές τιμές έτσι όπως προήλθαν ανά επανάληψη, ενώ στο Σχήμα 3.3 δίνεται το ιστόγραμμα αυτών των τιμών. Εύκολα διαπιστώνεται ότι ο αλγόριθμος δεν παράγει κάποιο μοτίβο στην προσομοίωση, αφού η αλυσίδα εξερευνά το ίδιο εύρος σε διαφορετικές περιόδους (βλέπε Robert and Casella (2010)). Επιπρόσθετα, το Σχήμα 3.2 δείχνει μερικά από τα χαρακτηριστικά γνωρίσματα των ακολουθιών Metropolis-Hastings. Ένα βασικό χαρακτηριστικό είναι ότι, για ορισμένα χρονικά διαστήματα, η ακολουθία των τιμών $x_t$ δεν αλλάζει διότι όλα τα αντίστοιχα $y_t$ απορρίπτονται. Ας σημειωθεί ακόμη ότι αυτές οι πολλαπλές εμφανίσεις της ίδιας αριθμητικής τιμής πρέπει να διατηρούνται στο δείγμα ως έχουν. Σε διαφορετική περίπτωση η εγκυρότητα της προσέγγισης της κατανομής στόχου $\pi$ χάνεται. Τέλος, από το ιστόγραμμα συχνοτήτων είναι φανερό ότι η αλυσίδα προσεγγίζει σωστά την κατανομή $Beta(3,4)$.

Τα παραπάνω υλοποιήθηκαν ως εξής:

```
a=3
b=4
n=9000
x=rep(0,n)
x[1]=0.3
for (i in 2:n){
  proposedx=runif(1)
  currentx = x[i-1]
  r=min(dbeta(proposedx,a,b)/dbeta(currentx,a,b),1)
  if(runif(1)<r){
    x[i] = proposedx
```

```
  } else {
    x[i] = currentx
}
x
hist(x,breaks=30,col="yellow")
plot(x,type="l",xlab = "Iterations",col="orange")
```

Σχήμα 3.2: Γραφική παράσταση της ακολουθίας των τιμών $X_t$, με $0 \leq t \leq 9.000$. Η ακολουθία των τιμών δημιουργήθηκε με τον αλγόριθμο MH έχοντας ως στάσιμη κατανομή τη Βήτα (3, 4).

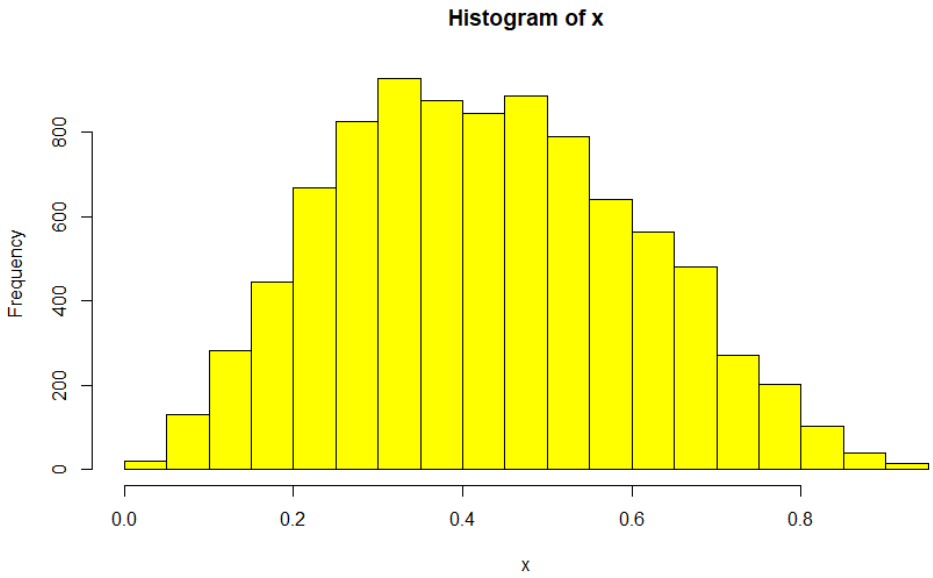



**Παράδειγμα 3.1.11.** *Να παράγετε, χρησιμοποιώντας τον αλγόριθμο MH, μια ακολουθία 1.000 τιμών από την εκθετική κατανομή με* $\lambda = 1$.

***Λύση.*** Η κατανομή στόχος έχει συνάρτηση πυκνότητας πιθανότητας

$$\pi(x) = \begin{cases} e^{-x}, & \text{αν } x \geq 0, \\ 0, & \text{διαφορετικά.} \end{cases} \tag{3.10}$$

Σχήμα 3.3: Ιστόγραμμα 9.000 τυχαίων αριθμών (με τον αλγόριθμο MH) από τη Βήτα κατανομή της σχέσης (3.9)




Αρχικά επιλέγουμε αυθαίρετα μια τρέχουσα κατάσταση $X_0 = x_0$ που είναι τέτοια ώστε $\pi(x_0) > 0$. Εμείς επιλέγουμε αυθαίρετα την κατάσταση $x_0 = 0.3$. Στη συνέχεια θα πρέπει να επιλέξουμε μια κατανομή πρότασης. Στο παράδειγμα αυτό, θα επιλέξουμε την κανονική κατανομή με μέση τιμή την τωρινή κατάσταση και διακύμανση 1. Η επιλογή αυτή έχει ως αποτέλεσμα, καθώς η κατανομή πρότασης είναι συμμετρική, να είναι: $r(x_t, y) = min\left\{\frac{\pi(y)}{\pi(x_t)}, 1\right\}$.

Τα παραπάνω υλοποιούνται με τις εντολές που δίνονται παρακάτω (βλέπε για παρόμοιο αλγόριθμο στην εργασία του Stephens (2017)). Ταυτόχρονα, στο Σχήμα 3.4 απεικονίζονται οι 1.000 δειγματικές τιμές έτσι όπως προήλθαν ανά επανάληψη, ενώ στο Σχήμα 3.5 δίνεται το ιστόγραμμα αυτών των τιμών. Εύκολα διαπιστώνεται ότι ο αλγόριθμος δεν παράγει κάποιο μοτίβο στην προσομοίωση, ενώ από το ιστόγραμμα συχνοτήτων είναι φανερό ότι η αλυσίδα προσεγγίζει σωστά την Εκθετική κατανομή.

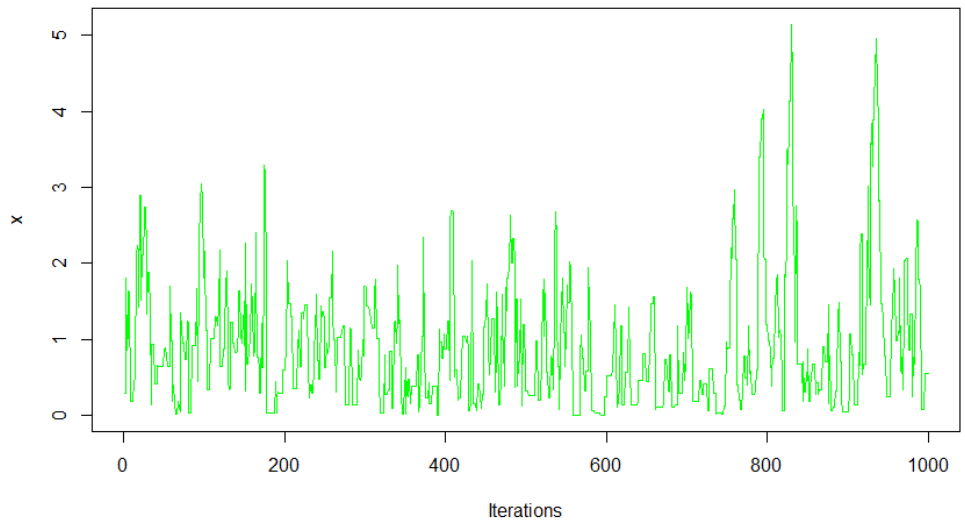
Τα παραπάνω υλοποιήθηκαν με τις εντολές:

```
target = function(x){
```

```
  if(x<0){
    return(0)}
  else {
    return( exp(-x))
  }
}
n=1000
x=rep(0,n)
x[1]=0.3
for (i in 2:n){
  proposedx=rnorm(1,x[i-1],1)
  currentx = x[i-1]
  r=min(target(proposedx)/target(currentx),1)
  if(runif(1)<r){
    x[i] = proposedx
  } else {
    x[i] = currentx
  }
}
x
hist(x,breaks=30,col="red")
plot(x,type="l",xlab = "Iterations",col="green")
```

**Παρατήρηση 3.1.12.** *Οι περισσότερες από τις απλές μεθόδους δειγματοληψίας αποδοχής-απόρριψης δε δίνουν ακριβή αποτελέσματα σε προβλήματα μεγαλύτερων διαστάσεων. Αυτή η συμπεριφορά αιτιολογείται διότι η πιθανότητα απόρριψης αυξάνεται εκθετικά ως συνάρτηση του αριθμού των διαστάσεων του προβλήματος (βλέπε, για παράδειγμα, Robert and Casella (2004), Johansen et al. (2007) και τις εκεί αναφορές). Από την άλλη πλευρά, ο Metropolis-Hastings και οι τεχνικές MCMC δεν έχουν αυτό το πρόβλημα σε τέτοιο βαθμό, και ως εκ τούτου είναι συχνά οι μόνες διαθέσιμες λύσεις όταν ο αριθμός των διαστάσεων της κατανομής που πρέπει να προσομοιωθεί είναι υψηλός. Το γεγονός αυτό έχει ως αποτέλεσμα να επιλέγονται για την παραγωγή δειγμάτων όταν έχουμε πολλές διαστάσεις. Ωστόσο σ΄ αυτές τις περιπτώσεις ο αλγόριθμος MH έχει συχνά αργή σύγκλιση και μια εναλλακτική προσέγγιση που συχνά λειτουργεί πολύ καλύτερα σε τέτοιες καταστάσεις είναι η αποκαλούμενη δειγματοληψία κατά Gibbs, η οποία θα αναλυθεί διεξοδικότερα στην επόμενη παράγραφο.*

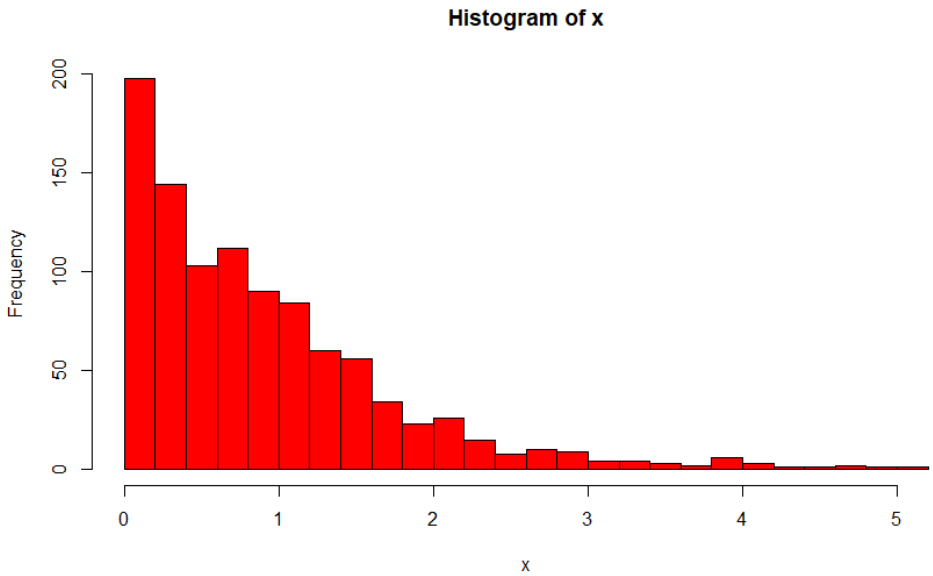
Σχήμα 3.4: Γραφική παράσταση της ακολουθίας των τιμών $X_t$, με $0 \leq t \leq 1.000$, με τον αλγόριθμο MH και στάσιμη κατανομή την Εκθ (λ=1).




## 3.2 Δειγματοληψία κατά Gibbs

Η **δειγματοληψία κατά Gibbs** (Gibbs Sampling ή αλλιώς Gibbs Sampler) είναι μία ευρέως διαδεδομένη MCMC μέθοδος για να παραχθεί, χωρίς την ανάγκη υπολογισμού της συνάρτησης πυκνότητας πιθανότητας, μια ακολουθία παρατηρήσεων από μία συγκεκριμένη πολυδιάστατη κατανομή πιθανότητας, όταν η άμεση προσομοίωση από αυτήν δεν είναι εφικτή. Ο δειγματολήπτης Gibbs παρότι πήρε το όνομά του από το φυσικό J.W.Gibbs, χάρη σε μια ανάλογη τεχνική μεταξύ δειγματοληψίας και στατιστικής φυσικής, οφείλει τη δημοφιλία του στην πρωτοπόρα εργασία των αδελφών Geman (βλέπε Geman and Geman (1984)), ενώ οι Gelfand and Smith (1990) ανέδειξαν τη σπουδαιότητα της μεθόδου σε ευρεία κατηγορία προβλημάτων κυρίως της Μπεϋζιανής στατιστικής ανάλυσης, και έκαναν, έτσι, τη δειγματοληψία Gibbs ένα δημοφιλές υπολογιστικό εργαλείο. Πλέον, η δειγματοληψία κατά Gibbs χρησιμοποιείται σε πολλούς επιστημονικούς κλάδους, όπως η Βιοστατιστική, η Χημεία, η Φυσική αλλά και η Φαρμακευτική (βλέπε, μεταξύ άλλων, Walsh (2004) και τις εκεί αναφορές). Όπως κάθε MCMC μέθοδος δημιουργεί μια Μαρκοβιανή αλυσίδα δειγματικών τιμών, καθεμία εκ των

Σχήμα 3.5: Ιστόγραμμα των τιμών $X_t$, με $0 \leq t \leq 1.000$, με τον αλγόριθμο MH και στάσιμη κατανομή την Εκθ (λ=1).




οποίων είναι συσχετισμένη με τις γειτονικές. Στην πραγματικότητα, όπως περιγράφουν αναλυτικά οι Robert and Casella (1999), η δειγματοληψία κατά Gibbs στην αρχική της μορφή, καθώς έχουν εμφανιστεί και επεκτάσεις της, αποτελεί μια ειδική περίπτωση του αλγορίθμου Metropolis-Hastings.

Όπως αναφέρθηκε ο στόχος της δειγματοληψίας κατά Gibbs είναι η κατασκευή μιας Μαρκοβιανής αλυσίδας της οποίας οι τιμές συγκλίνουν στην πολυδιάστατη κατανομή στόχο $\pi(x)$. Σύμφωνα με τον Walsh (2004), η βασική ιδέα για να επιτευχθεί ο στόχος αυτός είναι να προσομοιώνουμε διαδοχικά από δεσμευμένες κατανομές αντί από την από κοινού κατανομή. Κατά αυτόν τον τρόπο μετατρέπουμε ουσιαστικά ένα υψηλής διάστασης πρόβλημα σε πολλά απλούστερα μονοδιάστατα προβλήματα. Επομένως, το ιδιαίτερο γνώρισμά της είναι η διάσπαση ενός πολύπλοκου προβλήματος σε μια σειρά από απλούστερα προβλήματα. Επιπλέον, παρότι ελλοχεύει ο κίνδυνος η ακολουθία κατανομών μικρής διάστασης να απαιτεί μεγάλο χρόνο σύγκλισης, ο δειγματολήπτης Gibbs παράγει τιμές από τη δεσμευμένη κατανομή με τέτοιον τρόπο που να προσεγγίζει την από κοινού κατανομή μέσα σ΄ ένα συγκεκριμένο χρόνο. Η παραπάνω ιδιότητα ίσως αιτιολογεί και εξηγεί

και τη δημοφιλία της μεθόδου. Στη συνέχεια, και αφού δόθηκε η κεντρική ιδέα της μεθόδου, θα παρουσιαστεί η μέθοδος για την $k$-διάστατη περίπτωση.

Αρχικά, τη χρονική στιγμή $t = 0$ επιλέγεται τυχαία και αυθαίρετα μια αρχική $k$-διάστατη κατάσταση $X^0 = x^0$, με $x^0 = (x_1^0, ..., x_k^0)$, έτσι ώστε $\pi(x^0) > 0$, όπου $\pi$ είναι η συνάρτηση πυκνότητας πιθανότητας της $k-$διάστατης κατανομής στόχου. Θεωρώντας ότι έχουν παραχθεί τα $k$-διάστατα διανύσματα $X^1, X^2, ..., X^{t-1}$ και δεδομένου ότι τη χρονική στιγμή $t-1$ βρισκόμαστε στην κατάσταση $X^{t-1} = (x_1^{t-1}, ..., x_k^{t-1})$, περιγράφονται στη συνέχεια, δίνοντας όλες τις λεπτομέρειες και επεξηγήσεις, τα βήματα του αλγορίθμου (βλέπε, μεταξύ άλλων, Walsh (2004), Malefaki and Iliopoulos (2008), Jespersen (2010) και Liang et al. (2011)).

**Gibbs Sampling Algorithm (GSA)**

- **Βήμα 1:** Επίλεξε μια αυθαίρετη αρχική κατάσταση $X^0 = (x_1^0, ..., x_k^0)$, έτσι ώστε $\pi(x^0) > 0$, όπου $\pi$ είναι η συνάρτηση πυκνότητας πιθανότητας της $k-$διάστατης κατανομής στόχου.

- **Βήμα 2:** Με δεδομένη την τρέχουσα κατάσταση $X^{t-1} = (x_1^{t-1}, ..., x_k^{t-1})$, προσομοίωσε:
  $X_1^t \sim \pi_1(x_1|x_2^{t-1}, x_3^{t-1}, ..., x_k^{t-1})$
  $X_2^t \sim \pi_2(x_2|x_1^t, x_3^{t-1}, ..., x_k^{t-1})$
  $X_3^t \sim \pi_3(x_3|x_1^t, x_2^t, ..., x_k^{t-1})$
  ...
  ...
  $X_k^t \sim \pi_k(x_k|x_1^t, x_2^t, ..., x_{k-1}^t)$.
  Δηλαδή στο βήμα αυτό προσομοιώνεται η συνιστώσα $X_i^t$, για $i = 1, ..., k$, από τη δεσμευμένη κατανομή της $X_i$ δοθέντος των υπόλοιπων τυχαίων μεταβλητών πλην αυτής που είναι διαθέσιμη μέχρι εκείνο το στάδιο. Οι κατανομές $\pi_i$, $i = 1, ..., k$ ονομάζονται **πλήρως δεσμευμένες κατανομές** (full conditional distributions). Προφανώς υποθέτουμε ότι η προσομοίωση από τις δεσμευμένες κατανομές είναι εφικτή και εύκολη καθώς σε κάθε άλλη περίπτωση δε θα είναι χρήσιμη αυτή η μέθοδος (βλέπε Gelman et al. (2014)).

- **Βήμα 3:** Επαναλαμβάνουμε την ίδια διαδικασία πηγαίνοντας στο Βήμα 1 και εφαρμόζοντας για $t = t + 1$. Αν επαναληφθεί $n$ φορές η διαδικασία που αναφέρθηκε παραπάνω, τότε παράγεται μία ακολουθία $n$ το πλήθος $k$-διάστατων τυχαίων διανυσμάτων (βλέπε Walsh (2004)), που ονομάζεται **ακολουθία Gibbs μήκους $n$** (Gibbs sequence of length $n$).

Παρατηρείστε ότι ο αλγόριθμος δειγματοληψίας Gibbs δημιουργεί μια ακολουθία από την κατανομή κάθε μεταβλητής με τη σειρά, έχοντας τη γνώση από τις τρέχουσες τιμές των άλλων μεταβλητών και έτσι εξασφαλίζεται η Μαρκοβιανή ιδιότητα. Ο πυρήνας ή πίνακας μετάβασης της αλυσίδας είναι:

$$\begin{aligned} K(X^{t-1}, X^t) &= K\left((x_1^{t-1}, ..., x_k^{t-1}), (x_1^t, ..., x_k^t)\right) \\ &= \pi_1(x_1^t | x_2^{t-1}, x_3^{t-1}, ..., x_k^{t-1}) \times \pi_2(x_2^t | x_1^t, x_3^{t-1}, ..., x_k^{t-1}) \\ &\quad \times \cdots \times \pi_k\left(x_k^t | x_1^t, x_2^t, ..., x_{k-1}^t\right). \end{aligned}$$

Επιπλέον αποδεικνύεται ότι τα παραπάνω έχουν ως αποτέλεσμα η στάσιμη κατανομή της Μαρκοβιανής αλυσίδας να είναι η $\pi$ (βλέπε, μεταξύ άλλων, Johansen et al. (2007)).

**Παρατήρηση 3.2.1.** *Η δειγματοληψία Gibbs εφαρμόζεται πολύ συχνά στη Μπεϋζιανή στατιστική συμπερασματολογία, αφού εκεί έχουν βασικό ρόλο οι δεσμευμένες κατανομές, και χρησιμοποιείται ιδιαίτερα για την προσομοίωση από την εκ των υστέρων κατανομή. Το σημαντικότερο πλεονέκτημα αυτής της μεθόδου είναι ότι ακόμη και σε προβλήματα μεγαλύτερων διαστάσεων όλες οι προσομοιώσεις μπορούν να πραγματοποιηθούν μόνο από μονοδιάστατες κατανομές. Τέλος, στη βασική της εκδοχή η δειγματοληψία Gibbs μπορεί να θεωρηθεί ως μια ειδική περίπτωση του αλγορίθμου Metropolis-Hastings, καθώς για ένα πρόβλημα $m$ διαστάσεων μπορεί να ερμηνευτεί με $m$ αλγόριθμους MH με πιθανότητες αποδοχής ίσες με τη μονάδα. Αυτή η ιδιότητα αποτελεί και μια σημαντική διαφορά μεταξύ των δύο τεχνικών, επειδή όλα τα δείγματα γίνονται αποδεκτά στη δειγματοληψία Gibbs σε αντίθεση με αυτήν του αλγορίθμου Metropolis-Hastings. Επίσης, η δειγματοληψία Gibbs απαιτεί γνώση από όλες τις δεσμευμένες κατανομές για τις παραμέτρους του μοντέλου. Σε περιπτώσεις όπου αυτές δεν είναι διαθέσιμες σε απλή μορφή, μπορεί να χρησιμοποιηθεί ο αλγόριθμος MH που απαιτεί μόνο την πλήρη γνώση για την από κοινού κατανομή. Όπως αναφέρθηκε και προηγουμένως, μια άλλη ενδιαφέρουσα παρατήρηση είναι ότι ο δειγματολήπτης Gibbs εφαρμόζεται όταν η κατανομή στόχος είναι τουλάχιστον διδιάστατη, ενώ ο αλγόριθμος MH μπορεί να εφαρμοστεί και σε μονοδιάστατα προβλήματα (βλέπε Jespersen (2010)). Στις πιο εκτεταμένες εκδοχές της η δειγματοληψία Gibbs μπορεί να θεωρηθεί ως ένα γενικό πλαίσιο δειγματοληψίας από ένα μεγάλο σύνολο μεταβλητών προσομοιώνοντας κάθε μεταβλητή (ή σε ορισμένες περιπτώσεις κάθε σύνολο μεταβλητών) με τη σειρά, και μπορεί να ενσωματώσει τον αλγόριθμο MH για την εφαρμογή ενός ή περισσότερων από τα στάδια της προσομοίωσης.*

**Παρατήρηση 3.2.2.** *Συνοπτικά, τα βασικά χαρακτηριστικά του αλγορίθμου Gibbs είναι ότι μετατρέπει ένα υψηλής διάστασης πρόβλημα σε πολλά μονοδιάστατα προβλήματα και ότι εφαρμόζεται όταν η από κοινού κατανομή είναι άγνωστη,*

*αλλά η δεσμευμένη κατανομή κάθε μεταβλητής είναι γνωστή και η προσομοίωση από αυτήν εύκολη. Από την άλλη μεριά, μία κακή επιλογή του κανόνα μετάβασης έχει ως αποτέλεσμα να αυξάνει την πολυπλοκότητα του αλγορίθμου.*

Σε αυτό το σημείο θα δοθούν δύο απλά παραδείγματα (βλέπε Robert and Casella (2010)).

**Παράδειγμα 3.2.3.** *Χρησιμοποιώντας τον αλγόριθμο Gibbs να παράγετε 4.000 τιμές από το τυχαίο διάνυσμα* $(X_1, X_2)$ *με σ.π.π. την:*

$$\pi_{X_1,X_2}(x_1, x_2) = \binom{m}{x_1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x_2^{x_1+a-1}(1-x_2)^{m-x_1+b-1},$$

*όπου* $x_1 = 0, 1, ..., m$, $x_2 \in (0, 1)$, *για* $m = 17$, $a = 2$ *και* $b = 9$.

***Λύση***. Θα υλοποιηθεί ο αλγόριθμος του Gibbs στην $R$ επεξηγώντας αναλυτικά την υλοποίηση κάθε βήματος.

- **Βήμα 1:** Στο πρώτο βήμα θέλουμε να επιλέξουμε μια αυθαίρετη αρχική κατάσταση $X^0 = x^0$, με $x^0 = (x_1^0, x_2^0)$, έτσι ώστε $\pi_{X_1,X_2}(x^0) > 0$, όπου $\pi$ είναι η συνάρτηση πυκνότητας πιθανότητας της διδιάστατης κατανομής στόχου. Μπορούμε εύκολα να διαπιστώσουμε ότι η περιθώρια κατανομή της $X_2$ είναι Βήτα με παραμέτρους $(a, b)$, ενώ η δεσμευμένη κατανομή της $X_1$ δοθέντος ότι $X_2 = x_2$ είναι διωνυμική με παραμέτρους $m, x_2$. Για το λόγο αυτό επιλέγουμε ως αυθαίρετη τιμή αυτή που προκύπτει προσομοιώνοντας μια τιμή από τις παραπάνω κατανομές.

- **Βήμα 2:** Προκύπτει μετά από λίγη άλγεβρα ότι η δεσμευμένη κατανομή της $X_2$ δοθέντος της $X_1 = x_1$ είναι Βήτα με παραμέτρους $(x_1 + a, m - x_1 + b)$. Επομένως, με δεδομένη την τρέχουσα κατάσταση $X^{t-1} = x^{t-1}$, με $x^{t-1} = (x_1^{t-1}, x_2^{t-1})$ προσομοιώνουμε:
  $X_1^t \sim B(m, x_2^{t-1})$
  $X_2^t \sim Be(x_1^t + a, m - x_1^t + b)$.

- **Βήμα 3:** Επαναλαμβάνουμε την ίδια διαδικασία πηγαίνοντας στο Βήμα 1 και εφαρμόζοντας για $t = t + 1$.

Τα παραπάνω υλοποιούνται στην R χρησιμοποιώντας τον εξής κώδικα:

```
n=4000
m=17
```

```
a=2
b=9
X=rep(0,n)
Y=rep(0,n)
Y[1]=rbeta(1,a,b)
X[1]=rbinom(1,m,Y[1])
for (i in 2:n){
  X[i]=rbinom(1,m,Y[i-1])
  Y[i]=rbeta(1,a+X[i],m-X[i]+b)
}
```

Στο παράδειγμα που προηγήθηκε οι πλήρως δεσμευμένες κατανομές ήταν διαθέσιμες και ήταν εφικτή η προσομοίωση από αυτές. Σε περίπτωση που κάτι τέτοιο δεν μπορεί να γίνει άμεσα τότε κάποιος μπορεί να χρησιμοποιήσει έναν υβριδικό αλγόριθμο που προκύπτει όταν για την παραγωγή τιμών από τις δεσμευμένες κατανομές χρησιμοποιείται ο αλγόριθμος Metropolis-Hastings. Σε μια τέτοια περίπτωση ο αλγόριθμος που προκύπτει αναφέρεται στη βιβλιογραφία ως Metropolis within Gibbs. Για περισσότερες πληροφορίες παραπέμπουμε στους Liang et al. (2011) και Wasserman (2013).

**Παράδειγμα 3.2.4.** *Να παράγετε 8.500 τιμές από το τυχαίο διάνυσμα* $(X_1, X_2)'$ *με σ.π.π. τη διδιάστατη κανονική κατανομή* $N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.97\\0.97 & 1\end{pmatrix}\right)$ *εφαρμόζοντας τον αλγόριθμο Gibbs.*

***Λύση***. Από υπόθεση έχουμε ότι το τυχαίο διάνυσμα

$$\begin{pmatrix}X_1\\X_2\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\,\sigma_X\sigma_Y\\\rho\,\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right), \tag{3.11}$$

όπου $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ και $\rho = 0.97$.

Αρχικά, βλέπε Βήμα 1 της μεθόδου, είναι επιθυμητό να επιλεγεί μια αυθαίρετη αρχική κατάσταση $X^0 = x^0$ με $x^0 = (x_1^0, x_2^0)$, έτσι ώστε $\pi_{X,Y}(x^0) > 0$, με $\pi$ να είναι η σ.π.π. της διδιάστατης κατανομής στόχου. Μπορούμε εύκολα να διαπιστώσουμε ότι η περιθώρια κατανομή της $X_2$ είναι κανονική με παραμέτρους $(0, 1)$.

Προχωρώντας στο Βήμα 2 του αλγορίθμου, και λαμβάνοντας υπόψη ότι η δεσμευμένη κατανομή της $X_2$ δοθέντος της $X_1 = x_1$ είναι κανονική με μέση τιμή $\rho x_1$

και τυπική απόκλιση $\sqrt{1-\rho^2}$. Με ανάλογο τρόπο η δεσμευμένη κατανομή της $X_1$ δοθέντος της $X_2 = x_2$ είναι κανονική με μέση τιμή $\rho x_2$ και τυπική απόκλιση $\sqrt{1-\rho^2}$. Συνεπώς, με δεδομένη την τρέχουσα κατάσταση $X^{t-1} = (x_1^{t-1}, x_2^{t-1})$ προσομοιώνουμε:
$X_1^t \sim \pi_1(x_1|x_2^{t-1}) \Rightarrow X_1^t \sim N(\rho x_2^{t-1}, 1-\rho^2)$
$X_2^t \sim \pi_2(x_2|x_1^t) \Rightarrow X_2^t \sim N(\rho x_1^t, 1-\rho^2)$.

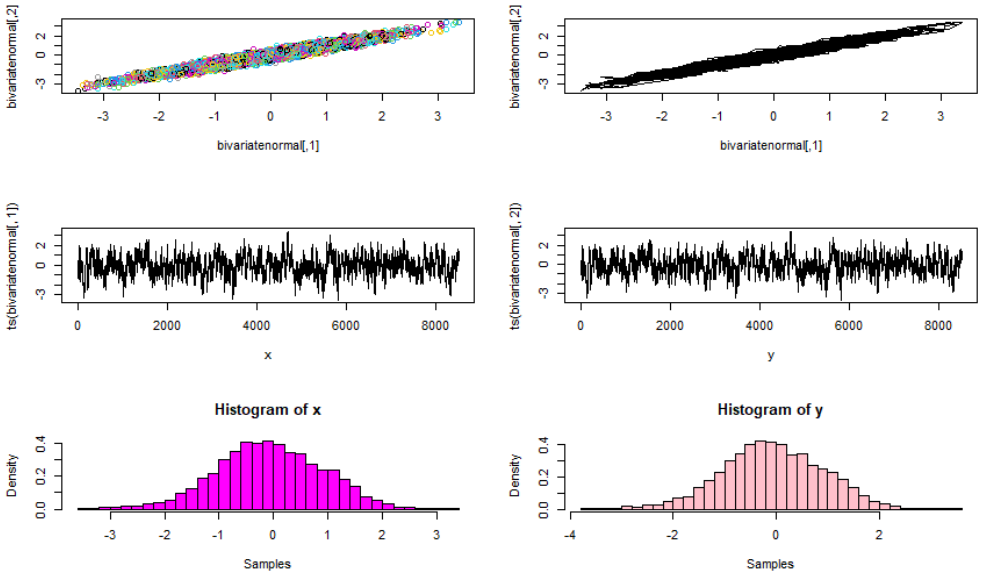
Τα παραπάνω υλοποιούνται χρησιμοποιώντας τις ακόλουθες εντολές (βλέπε Burke (2018)).

```
gibbs.sampling = function(n, rho)
{
     table = matrix(ncol = 2, nrow = n)
     x_{1} = 0
     x_{2} = 0
     table[1, ] = c(x_{1}, x_{2})
     for (i in 2:n) {
       x_{1} = rnorm(1, rho * x_{2}, sqrt(1 - rho^2))
       x_{2} = rnorm(1, rho * x_{1}, sqrt(1 - rho^2))
       table[i, ] = c(x_{1}, x_{2})
       }
     table
     }
```

Στο Σχήμα 3.6 δίνονται τα διαγράμματα διασποράς και τα ιστογράμματα για τις 8.500 τιμές από το τυχαίο διάνυσμα $(X_1, X_2)$ με σ.π.π. τη διδιάστατη κανονική κατανομή (βλέπε σχέση (3.11)).

```
bivariatenormal = gibbs.sampling(8500,0.97)
par(mfrow=c(3,2))
plot(bivariatenormal,col=1:8500)
plot(bivariatenormal,type="l")
plot(ts(bivariatenormal[,1]), xlab='x')
plot(ts(bivariatenormal[,2]), xlab='y')
hist(bivariatenormal[,1],breaks=30,prob = TRUE, col="
   magenta", xlab='Samples', main='Histogram  of x')
hist(bivariatenormal[,2],breaks=30,prob = TRUE, col="
   pink", xlab='Samples', main='Histogram of y')
```

Σχήμα 3.6: Διαγράμματα για τις 8.500 τιμές από το διάνυσμα $(X_1, X_2)$ με σ.π.π. τη διδιάστατη κανονική κατανομή της σχέσης (3.11).




**Παρατήρηση 3.2.5.** *Στη βιβλιογραφία έχουν παρουσιαστεί διάφορες άλλες μέθοδοι που ανήκουν στην κατηγορία των MCMC μεθόδων, που δε θα αποτελέσουν αντικείμενο μελέτης στα πλαίσια αυτής της διατριβής. Ενδεικτικά αναφέρουμε τον αποκαλούμενο Hit-and-run Δειγματολήπτη (βλέπε Turchin (1971), Smith (1984) και Chen and Schmeiser (1996)), τον Δειγματολήπτη Τμημάτων (Slice Sampler, βλέπε Neal (2003), Robert and Casella (2004) και Liu (2008)), την Προσαρμοσμένη Δειγματοληψία Κατεύθυνσης (Adaptive Direction Sampling (ADS), βλέπε Gilks et al. (1994) Liu (1999) και Liu (2008)) και τον Αλγόριθμο Snooker (βλέπε Liu et al. (2000)).*

# ΚΕΦΑΛΑΙΟ 4

# Εκτιμητικη με ABC μεθοδους

Σε πολλές περιπτώσεις μοντελοποίησης πραγματικών τυχαίων φαινομένων οδηγούμαστε σε περίπλοκα στοχαστικά μοντέλα των οποίων η συνάρτηση πιθανοφάνειας παρότι είναι πλήρως ορισμένη από το πιθανοθεωρητικό μοντέλο, δεν είναι διαθέσιμη σε κλειστή αναλυτική μορφή ως συνάρτηση της παραμέτρου $\theta$ ή το υπολογιστικό κόστος για τον υπολογισμό της είναι αρκετά μεγάλο. Το γεγονός αυτό έχει αντίκτυπο στην υλοποίηση τόσο των κλασικών όσο και των Μπεϋζιανών μεθοδολογιών στατιστικής συμπερασματολογίας, καθώς και οι δύο βασίζονται στη συνάρτηση πιθανοφάνειας. Επιπρόσθετα, όπως επισημαίνει και ο Robert (2016), τα παραπάνω προβλήματα καθιστούν απαγορευτική την άμεση εφαρμογή ενός γενικού MCMC αλγορίθμου, όπως ο Gibbs ή ο Metropolis-Hastings. Για την αντιμετώπιση των παραπάνω προβλημάτων, μεταξύ άλλων, παρουσιάστηκε στη βιβλιογραφία η κλάση των Προσεγγιστικά Μπεϋζιανών Υπολογιστικών Μεθόδων (Approximate Bayesian Computational Methods) ή αλλιώς η κλάση των ABC μεθόδων. Η κεντρική ιδέα αυτής της μεθόδου παρουσιάστηκε στο Κεφάλαιο 1 αυτής της μεταπτυχιακής διατριβής, όπου παρατέθηκαν και κάποια ιστορικά στοιχεία. Αντικείμενο μελέτης αυτού του κεφαλαίου αποτελεί η αναλυτική παρουσίαση και επεξήγηση των βημάτων της ABC μεθόδου. Καθώς οι ABC αλγόριθμοι χαρακτηρίζονται ως μία Μπεϋζιανή διαδικασία, αρχικά, παρουσιάζονται εν συντομία τα χαρακτηριστικά της Μπεϋζιανής προσέγγισης και έπεται η αναλυτική παρουσίαση της ABC μεθόδου. Στη συνέχεια παρατίθενται με λεπτομέρειες όλες οι απαραίτητες διευκρινήσεις για την υλοποίηση του αλγορίθμου ABC και έπειτα, μελετώνται κάποια απλά παραδείγματα κατανόησης. Τέλος, το ενδιαφέρον επικεντρώνεται στις κυριότερες τροποποιήσεις του αλγορίθμου, με βασικό κριτήριο τη συχνότητα χρησιμοποίησής τους σε πρακτικές εφαρμογές.

## 4.1 Χαρακτηριστικά Μπεϋζιανής διαδικασίας

Καθώς η ABC μέθοδος αποτελεί μια διαδικασία της Μπεϋζιανής στατιστικής στην ενότητα αυτή θα αναφερθούν εν συντομία σημαντικές έννοιές της. Στη Μπεϋζιανή προσέγγιση θεωρούμε ότι έχουμε ένα τυχαίο δείγμα $x = (x_1, ..., x_n)$ από την κατανομή $f(x|\theta)$, όπου η άγνωστη παράμετρος $\theta$ θεωρείται ότι είναι η τιμή μιας τυχαίας μεταβλητής. Αυτή είναι ουσιαστικά και η θεμελιώδης διαφορά μεταξύ Μπεϋζιανής και κλασικής στατιστικής. Στις κλασικές μεθοδολογίες οι παράμετροι είναι άγνωστες αλλά σταθερές ποσότητες, ενώ στις Μπεϋζιανές μεθοδολογίες το $\theta$ παριστά τις τιμές κάποιας τυχαίας μεταβλητής. Επομένως, ο συμβολισμός $f(\cdot|\cdot)$ δηλώνει τη δεσμευμένη κατανομή πιθανότητας η οποία καθορίζεται κάθε φορά από τις αντίστοιχες τυχαίες μεταβλητές.

Στο παραπάνω πλαίσιο, για την εξαγωγή συμπερασμάτων στη Μπεϋζιανή προσέγγιση ακολουθούμε τα ακόλουθα βήματα.

**Βήμα 1. Καθορισμός εκ των προτέρων κατανομής.** Αρχικά καθορίζουμε μια εκ των προτέρων κατανομή πιθανότητας για την παράμετρο $\theta$ που εκφράζει τις πρότερες πεποιθήσεις, την πρότερη γνώση που έχουμε για αυτήν, πριν τη συλλογή των δεδομένων. Η πρότερη κατανομή πιθανότητας της παραμέτρου $\theta$ συμβολίζεται $\pi(\theta)$.

**Βήμα 2. Συλλογή δεδομένων.** Το επόμενο βήμα είναι η συλλογή δεδομένων σχετικών με το υπό μελέτη τυχαίο φαινόμενο. Όλη η πληροφορία που φέρουν τα δεδομένα για την άγνωστη παράμετρο εμπεριέχεται στην από κοινού συνάρτηση κατανομής τους. Υπό την υπόθεση ότι τα δεδομένα αποτελούν τυχαίο δείγμα, δηλαδή ότι είναι ανεξάρτητες και ισόνομα κατανεμημένες παρατηρήσεις δοθείσης της παραμέτρου $\theta$, άμεσα έχουμε ότι η από κοινού κατανομή των $x_1, ..., x_n$ δίνεται από τη σχέση: $f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. Στο σημείο αυτό επισημαίνεται ότι όταν η $f(x|\theta)$ θεωρείται συνάρτηση του $\theta$ για δοθέν $x$ καλείται συνάρτηση πιθανοφάνειας και συμβολίζεται με $L(\theta)$, ενώ με $l(\theta)$ συμβολίζεται ο λογάριθμος της συνάρτησης πιθανοφάνειας, ήτοι $l(\theta) = log(L(\theta)) = log(f(x|\theta))$.

**Βήμα 3. Ανανέωση πληροφορίας για την $\theta$ και εξαγωγή συμπερασμάτων.** Τέλος, η διεξαγωγή συμπερασμάτων στη Μπεϋζιανή προσέγγιση στηρίζεται στο συνδυασμό της εκ των προτέρων γνώσης για την άγνωστη παράμετρο $\theta$ και της γνώσης που λαμβάνεται για την παράμετρο από τα δεδομένα. Ο συνδυασμός αυτός επιτυγχάνεται μέσω του προσδιορισμού της λεγόμενης εκ των υστέρων κατανομής για την παράμετρο $\theta$, που ουσιαστικά εκφράζει την ανανεωμένη γνώση που έχουμε για την παράμετρο $\theta$ μετά τη συλλογή των δεδομένων. Ειδικότερα, η εκ των υστέρων κατανομή $\pi(\theta|x)$ είναι το πηλίκο της από κοινού κατανομής των $(\theta, x)$ προς την περιθώρια των $x$, δηλαδή: $\pi(\theta|x) = \frac{g(\theta,x)}{h(x)}$. Όμως, η από κοινού συνάρτηση πιθανότητας των $(\theta, x)$ μπορεί να προκύψει ως το

γινόμενο της δεσμευμένης $x|\theta$ με την περιθώρια του $\theta$. Επομένως, καταλήγουμε στη σχέση:

$$\pi\left(\theta|x\right)=\frac{f\left(x|\theta\right)\pi\left(\theta\right)}{h\left(x\right)}, \tag{4.1}$$

όπου $\pi\left(\theta\right)$ είναι η εκ των προτέρων κατανομή της παραμέτρου $\theta$, και $h\left(x\right)$ η περιθώρια κατανομή των δεδομένων, ή ισοδύναμα (στη συνεχή περίπτωση)

$$\pi\left(\theta|x\right)=\frac{f\left(x|\theta\right)\pi\left(\theta\right)}{\int f\left(x|\theta\right)\pi\left(\theta\right)\,\mathrm{d}\theta}.$$

Πολύ συχνά καθώς ο παρονομαστής δεν εξαρτάται από το $\theta$ παραλείπεται και γράφουμε ότι $\pi\left(\theta|x\right)\propto f\left(x|\theta\right)\pi\left(\theta\right)$. Τότε λέμε ότι έχουμε τη μη κανονικοποιημένη εκ των υστέρων κατανομή.

Πριν ολοκληρώσουμε αυτήν τη σύντομη παρουσίαση των χαρακτηριστικών της Μπεϋζιανής διαδικασίας, καθώς ο ρόλος της εκ των προτέρων κατανομής είναι κομβικός στον προσδιορισμό της εκ των υστέρων κατανομής είναι απαραίτητος ο σχολιασμός για τον τρόπο επιλογής της. Όπως έχει ήδη αναφερθεί, η εκ των προτέρων κατανομή αντιπροσωπεύει τις πεποιθήσεις για την παράμετρο $\theta$ πριν το σύνολο των δεδομένων $x=(x_1,...,x_n)$ να είναι διαθέσιμο. Αν η γνώση αυτή είναι επαρκής τότε πρέπει να ενσωματωθεί στην επιλογή της εκ των προτέρων κατανομής, ενώ σε αντίθετη περίπτωση οφείλουμε να επιλέγουμε μία εκ των προτέρων κατανομή που αντανακλά την άγνοιά μας για την παράμετρο $\theta$. Στην πρώτη περίπτωση είναι σύνηθες να επιλέγονται εκ των προτέρων κατανομές που οδηγούν σε εκ των υστέρων κατανομές που ανήκουν στην ίδια οικογένεια κατανομών με την εκ των προτέρων. Σε αυτήν την περίπτωση λέμε ότι έχουμε συζυγείς εκ των προτέρων κατανομές και η επιλογή μιας τέτοιας έχει ως αποτέλεσμα τη διευκόλυνση στους υπολογισμούς (π.χ. για την εύρεση της σταθεράς κανονικοποίησης). Από την άλλη μεριά, όταν δεν έχουμε διαθέσιμη πληροφορία για την παράμετρο $\theta$ χρησιμοποιούνται οι λεγόμενες μη πληροφοριακές εκ των προτέρων κατανομές. Σε αυτήν την κατηγορία ανήκει η εκ των προτέρων κατανομή του Jeffreys, η οποία δίνεται από τη σχέση: $f(\theta)\propto(I(\theta))^{\frac{1}{2}}$, όπου η ποσότητα $I(\theta)$ είναι το μέτρο πληροφορίας του Fisher και ορίζεται ως:

$$I(\theta)=E\left(\left(\frac{dl(\theta)}{d\theta}\right)^2\right)=E\left(\left(\frac{dlog(f(x|\theta))}{d\theta}\right)^2\right).$$

Ας σημειωθεί επίσης ότι στην περίπτωση που ο λογάριθμος της πιθανοφάνειας, $l(\theta)$, είναι δύο φορές παραγωγίσιμη συνάρτηση ως προς $\theta$ τότε, υπό κατάλληλες

προϋποθέσεις, το μέτρο πληροφορίας του Fisher ισοδύναμα δίνεται από τη σχέση $I(\theta) = -E\left(\frac{d^2 l(\theta)}{d\theta^2}\right)$.

## 4.2 Παρουσίαση βασικών ABC αλγορίθμων

Από όσα αναφέρθηκαν στην προηγούμενη ενότητα είναι σαφές ότι κατά τον προσδιορισμό της εκ των υστέρων κατανομής της $\theta$ δοθέντος των δεδομένων $(x_1, ..., x_n)$ απαιτείται ο υπολογισμός της συνάρτησης πιθανοφάνειας. Σε πολλές περιπτώσεις, όπως αναλυτικά έχουν αναφερθεί στο Κεφάλαιο 1 αυτής της μεταπτυχιακής διατριβής, ο υπολογισμός αυτός δεν είναι εφικτός και έχει ως αντίκτυπο να μην μπορούν να χρησιμοποιηθούν τα κλασικά βήματα της Μπεϋζιανής προσέγγισης, τα οποία εν συντομία περιγράφηκαν στην προηγούμενη ενότητα. Επιπρόσθετα, δεν είναι εφικτή η προσομοίωση από την εκ των υστέρων κατανομή εφαρμόζοντας τεχνικές Monte Carlo. Για να ξεπεραστεί το πρόβλημα που δημιουργείται λόγω του μη υπολογισμού της συνάρτησης πιθανοφάνειας εισήχθησαν στη βιβλιογραφία οι Προσεγγιστικές Μπεϋζιανές Υπολογιστικές μέθοδοι ή ABC μέθοδοι, οι οποίες αναλυτικά θα παρουσιαστούν σε αυτήν την ενότητα.

Η πρώτη εφαρμογή ABC αλγορίθμου παρουσιάστηκε από τους Tavaré et al. (1997) σε άρθρο πληθυσμιακής γενετικής, με αντικείμενο μελέτης την εκτίμηση του χρόνου συγκερασμού, δηλαδή του χρόνου από τον πιο πρόσφατο κοινό πρόγονο, ενός δείγματος αλληλουχιών DNA ενδοειδών. Ειδικότερα, ο αλγόριθμος που προτάθηκε από τους Tavaré et al. (1997) υποθέτει ότι το στήριγμα της κατανομής είναι πεπερασμένο και η παράμετρος ενδιαφέροντος είναι η $\theta$. Στο πλαίσιο αυτό ουσιαστικά εισήγαγαν τις ABC μεθόδους ως μια τεχνική αποδοχής-απόρριψης σύμφωνα με τα ακόλουθα βήματα.

**Αλγόριθμος των Tavaré et al. (1997)**:

1. Υιοθετούμε μια εκ των προτέρων κατανομή για την παράμετρο $\theta$.
2. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή που καθορίστηκε στο προηγούμενο βήμα.
3. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$, από το μοντέλο $f(\cdot|\theta^*)$.
4. Αν $D^* = D$, όπου $D = (x_i)$, με $i = 1, ..., n$, οι τιμές του πραγματικού δείγματος, τότε αποδεχόμαστε την τιμή $\theta^*$, διαφορετικά επαναλαμβάνουμε τα βήματα 2-4.

5. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $M$, μη απορριπτέων τιμών $\theta^*$.

**Παρατήρηση 4.2.1.** *Ο αλγόριθμος των Tavaré et al. (1997) είναι στην πραγματικότητα μια ειδική περίπτωση της μεθόδου αποδοχής-απόρριψης κατά την οποία η παράμετρος $\theta$ παράγεται από την εκ των προτέρων κατανομή και η αποδοχή είναι υπό τη δέσμευση ότι τα προσομοιωμένα δεδομένα ταυτίζονται με τα πραγματικά δεδομένα (βλέπε Marin et al. (2012)).*

Από τα παραπάνω γίνεται κατανοητό ότι ο αλγόριθμος δίνει ως αποτέλεσμα ένα δείγμα ανεξάρτητων και ισόνομων μεταβλητών από την εκ των υστέρων κατανομή $\pi(\theta|x)$. Όντως, έστω ότι το $(\theta^*, x^*)$ γίνεται δεκτό, με $x^* = (x_1^*, ..., x_n^*)$. Τότε, η από κοινού κατανομή του $(\theta^*, x^*)$ δίνεται ως $f(\theta^*, x^*) = f(x^* \mid \theta^*)\, \pi(\theta^*)\, I_x(x^*)$, όπου:

$$I_x(x^*) = \begin{cases} 1, & \text{αν } x^* = x \\ 0, & \text{διαφορετικά.} \end{cases}$$

Επομένως:

$$f(\theta^*) \propto \sum_{x^*} \pi(\theta^*) f(x^*|\theta^*)\, I_x(x^*) = \pi(\theta^*) f(x^*|\theta^*) \propto \pi(\theta^*|x).$$

Η μέθοδος που προτάθηκε από τους Tavaré et al. (1997) έχει σημαντικά μειονεκτήματα καθώς υποθέτει ότι το σύνολο $D$ είναι διακριτό και πεπερασμένο, ενώ επιπλέον αν το μέγεθος του δείγματος είναι μεγάλο είναι πολύ δύσκολο έως και σπάνιο να ικανοποιηθεί η σχέση $D^* = D$. Ως συνέπεια των παραπάνω έχουμε ότι είτε θα χρειάζονται πολλές επαναλήψεις των βημάτων 2-4 μέχρις ότου να γίνει δεκτή μία τιμή, ενώ επιπρόσθετα ο αλγόριθμος δεν μπορεί να εφαρμοστεί στη συνεχή περίπτωση. Οι παραπάνω λόγοι οδήγησαν τους Pritchard et al. (1999) να προτείνουν μια γενίκευση του αλγορίθμου που θα είναι τέτοια ώστε να αντιμετωπίζονται τα παραπάνω προβλήματα. Στη συνέχεια παρατίθενται τα βήματα του αλγορίθμου που πρότειναν, ενώ έπονται κάποιες απαραίτητες σχετικές διευκρινήσεις για την υλοποίησή του.

**Αλγόριθμος των Pritchard et al. (1999)**:

1. Υιοθετούμε μια εκ των προτέρων κατανομή για την παράμετρο $\theta$.

2. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή που καθορίστηκε στο προηγούμενο βήμα.

3. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$, από το μοντέλο $f(\cdot|\theta^*)$.

4. Υπολογίζουμε ένα μέτρο της απόκλισης ή της εγγύτητας (ή μια μετρική) μεταξύ των τιμών του πραγματικού δείγματος $D = (x_i)$, $i = 1, ..., n$, και των τιμών του προσομοιωμένου δείγματος $D^*$, έστω $d(D, D^*)$. Αποδεχόμαστε την τιμή $\theta^*$ αν $d(D, D^*) \leq \epsilon$, για κάποιο $\epsilon > 0$, διαφορετικά επαναλαμβάνουμε τα βήματα 2-4.

5. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $M$, μη απορριπτέων τιμών $\theta^*$.

Απ΄ όσα προηγήθηκαν είναι προφανές ότι είναι απαραίτητο να συγκριθούν τα παρατηρούμενα με τα προσομοιωμένα δεδομένα. Ωστόσο καθώς αυξάνεται η διάσταση των δεδομένων η πιθανότητα δημιουργίας ενός προσομοιωμένου συνόλου δεδομένων $D^*$ με μικρή απόσταση από το $D$ συνήθως μειώνεται, με αποτέλεσμα να μειώνεται και η υπολογιστική αποτελεσματικότητα του παραπάνω βασικού αλγορίθμου αποδοχής-απόρριψης ABC. Για να αντιμετωπισθεί η παραπάνω αδυναμία έχει προταθεί μια εναλλακτική μορφή του παραπάνω αλγορίθμου. Ειδικότερα, ο εναλλακτικός αυτός αλγόριθμος προκύπτει αν αντικατασταθεί το μέτρο απόκλισης μεταξύ των $D$ και $D^*$ με ένα μέτρο απόκλισης μεταξύ ενός κατάλληλου στατιστικού περιγραφικού μέτρου υπολογισμένο στα δύο σύνολα (βλέπε, μεταξύ άλλων, Malmberg et al. (2004), Csilléry et al. (2010) και Nosofsky et al. (2011)). Επομένως, αντί να συγκρίνουμε την απόσταση των $D$ και $D^*$ συγκρίνουμε την απόσταση δύο συναρτήσεών τους, έστω $s(D)$ και $s(D^*)$, αντίστοιχα, και προκύπτει η ακόλουθη παραλλαγή του αλγορίθμου των Pritchard et al. (1999):

**Παραλλαγή του αλγορίθμου των Pritchard et al. (1999)**:

1. Υιοθετούμε μια εκ των προτέρων κατανομή για την παράμετρο $\theta$.

2. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή που καθορίστηκε στο προηγούμενο βήμα.

3. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$, από το μοντέλο $f(\cdot|\theta^*)$.

4. Υπολογίζουμε ένα μέτρο της απόκλισης ή της εγγύτητας (ή μια μετρική) μεταξύ των τιμών ενός περιγραφικού συνοπτικού στατιστικού, έστω $s$, υπολογισμένο στο πραγματικό δείγμα $D = (x_i)$, $i = 1, ..., n$, και των τιμών αυτού του περιγραφικού μέτρου υπολογισμένων στο προσομοιωμένο δείγμα

$D^*$. Οι τιμές αυτές συμβολίζονται $s(D)$ και $s(D^*)$, αντίστοιχα, ενώ συμβολίζουμε με $d(s(D), s(D^*))$ την απόκλισή τους. Αποδεχόμαστε την τιμή $\theta^*$ αν $d(s(D), s(D^*)) \leq \epsilon$, για κάποιο $\epsilon > 0$, διαφορετικά επαναλαμβάνουμε τα βήματα 2-4.

5. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $M$, μη απορριπτέων τιμών $\theta^*$.

Μέχρι τώρα απλά παραθέσαμε τους βασικούς αλγορίθμους ABC χωρίς να υπεισέρθουμε σε λεπτομέρειες και διευκρινήσεις σχετικές με την επιλογή των εκ των προτέρων κατανομών, της παραμέτρου $\epsilon$, του μέτρου απόκλισης και των περιγραφικών στατιστικών.

## 4.3 Διευκρινήσεις για την υλοποίηση του ABC αλγορίθμου

Στην προηγούμενη ενότητα παρουσιάστηκαν οι βασικοί ABC αλγόριθμοι, χωρίς να δοθεί επεξήγηση για το ρόλο των παραμέτρων που εμπλέκονται στην υλοποίησή τους και τον τρόπο επιλογής τους. Τα παραπάνω αποτελούν αντικείμενο μελέτης αυτής της ενότητας.

### 4.3.1 Η σταθερά ανεκτικότητας

Στην υποενότητα αυτή θα διευκρινιστεί ο ρόλος της σταθεράς ή επιπέδου ανεκτικότητας (tolerance threshold) $\epsilon$ ($\epsilon > 0$) που εμφανίζεται στο βήμα 4 του αλγορίθμου των Pritchard et al. (1999), και εκφράζει το επιθυμητό επίπεδο συμφωνίας των προσομοιωμένων και των πραγματικών δεδομένων. Παρατηρούμε ότι ο αλγόριθμος προσομοιώνει στην πραγματικότητα τυχαία δείγματα από την εκ των υστέρων κατανομή $\pi_\epsilon(\theta|x)$, που ικανοποιεί τη σχέση:

$$\pi_\epsilon(\theta|x) \propto P(d(D, D^*) \leq \epsilon)\, \pi(\theta). \tag{4.2}$$

Η πιο πάνω εκ των υστέρων κατανομή ουσιαστικά είναι η εκ των υστέρων κατανομή του $\theta$ δοθέντος ότι $d(D, D^*) \leq \epsilon$. Προκύπτει, όπως θα δούμε αναλυτικά στη συνέχεια, ότι για μικρές τιμές του $\epsilon$ η εκ των υστέρων κατανομή $\pi_\epsilon(\theta|x)$

προσεγγίζει την εκ των υστέρων κατανομή $\pi(\theta|x)$. Όντως:

$$\begin{aligned}\pi_\epsilon(\theta|x) &= \frac{\int f(x^*|\theta)\,\pi(\theta)\,I_{d(x,x^*)\leq\epsilon}(x^*)\,\mathrm{d}x^*}{\pi(x^*)} \\ &\approx \frac{f(x|\theta)\,\pi(\theta)}{\pi(x)} = \frac{f(x|\theta)\,\pi(\theta)}{\int f(x|\theta)\,\pi(\theta)\,\mathrm{d}\theta} = \pi(\theta|x)\,,\end{aligned}$$

όπου με $I_{d(x,x^*)\leq\epsilon}(x*)$ συμβολίζεται η δείκτρια συνάρτηση, η οποία ορίζεται ως:

$$I_{d(x,x^*)\leq\epsilon}(x^*) = \begin{cases}1, & \text{αν } x^* \text{ τέτοιο ώστε } d(x,x^*) \leq \epsilon \\ 0, & \text{διαφορετικά.}\end{cases}$$

Από τα παραπάνω γίνεται σαφές ότι τιμές του $\epsilon$ κοντά στο μηδέν εγγυώνται ότι οι τιμές του $\theta^*$ αποτελούν ένα δείγμα κοντά στην εκ των υστέρων κατανομή του $\theta$. Ωστόσο, πολύ μικρές τιμές του $\epsilon$ οδηγούν σε υψηλό ποσοστό απόρριψης ή ισοδύναμα σε πολύ μικρό λόγο αποδοχής (acceptance rate). Επομένως, απαιτείται μεγαλύτερος αριθμός επαναλήψεων της διαδικασίας μέχρι να επιτευχθεί ένα δείγμα μεγέθους $M$, με το $M$ συνήθως να είναι ένας μεγάλος αριθμός. Ειδικότερα, όταν το $\epsilon \to 0$ η εκ των υστέρων $\pi_\epsilon(\theta|x)$ συγκλίνει κατά κατανομή στην $\pi(\theta|x)$, ενώ όταν το $\epsilon \to \infty$ η εκ των υστέρων $\pi_\epsilon(\theta|x)$ συγκλίνει κατά κατανομή στην εκ των προτέρων κατανομή $\pi(\theta)$. Αυτό σημαίνει ότι πολύ μεγάλες τιμές του επιπέδου ανεκτικότητας δεν προσφέρουν καμία επιπλέον πληροφορία. Η προσαρμογή της παραμέτρου ανεκτικότητας $\epsilon$ πολύ κοντά στο μηδέν διασφαλίζει ένα ακριβές αποτέλεσμα, αλλά συνήθως καθιστά τους υπολογισμούς υπερβολικά χρονοβόρους. Έτσι, τιμές του $\epsilon$ μεγαλύτερες από το μηδέν χρησιμοποιούνται στην πράξη, γεγονός που εισάγει μια μεροληψία. Συμπερασματικά, είναι σαφές ότι, η επιλογή της σταθεράς ανεκτικότητας είναι κρίσιμη. Περισσότερα στοιχεία για τον τρόπο επιλογής της σταθεράς ανεκτικότητας θα δοθούν όταν θα αναφερθούμε στην επιλογή του μέτρου απόκλισης $d$.

### 4.3.2 Το μέτρο εγγύτητας

Στο τέταρτο βήμα του αλγορίθμου των Pritchard et al. (1999) απαιτείται ο υπολογισμός ενός μέτρου της εγγύτητας ή απόκλισης μεταξύ του προσομοιωμένου δείγματος $D^*$ και των πραγματικών δεδομένων $D$. Ο αντίκτυπος της επιλογής του μέτρου εγγύτητας έχει μελετηθεί σε διάφορες πρακτικές εφαρμογές, ενώ έχει παρατηρηθεί αρκετές φορές ότι ακόμη και απλές μετρικές δίνουν καλά αποτελέσματα. Δύο από τα πιο συνηθισμένα μέτρα εγγύτητας που αξιοποιούνται είναι

αυτά που προκύπτουν διαιρώντας τη σταθμισμένη ευκλείδεια απόσταση και τη σταθμισμένη $L_1$-απόσταση με το πλήθος των δεδομένων $n$ και έτσι προκύπτουν (στη μονοδιάστατη περίπτωση) τα ακόλουθα μέτρα:

$$d_2 = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(\frac{x_i - x_i^*}{s}\right)^2} \text{ και } d_3 = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - x_i^*|}{s}, \text{ αντίστοιχα,}$$

όπου με $s$ συμβολίζεται η τυπική απόκλιση του προσομοιωμένου δείγματος $D^* = (x_i^*)$. Επισημαίνεται ότι αν επιλεγεί η κλασική ευκλείδεια νόρμα ως απόσταση των προσομοιωμένων και των παρατηρούμενων δεδομένων, τότε η παραπάνω επαναληπτική διαδικασία αποτελεί την πιο απλή και εύκολη εκδοχή του λεγόμενου ABC αλγόριθμου αποδοχής απόρριψης (ABC accept-reject algorithm). Αυτό πρακτικά σημαίνει, όπως αναλυτικά θα δούμε στη συνέχεια, ότι μπορεί να εμπλουτιστεί με επιπλέον βήματα ώστε να αυξηθεί η αποτελεσματικότητα της δειγματοληψίας (βλέπε, για παράδειγμα, Marjoram et al. (2003) και Sisson et al. (2007)). Για λόγους πληρότητας παρατίθενται (πάλι για τη μονοδιάστατη περίπτωση δεδομένων) και κάποια άλλα γνωστά μέτρα εγγύτητας (βλέπε, μεταξύ άλλων, Turner and Van Zandt (2012)):

$$d_4 = |\overline{X} - \overline{X^*}| \text{ και } d_5 = |median(x_i) - median(x_i^*)|,$$

όπου $median(x)$ είναι η διάμεσος των δεδομένων, και τέλος:

$$d_6 = |[F^{-1}(0.75, X) - F^{-1}(0.25, X)] - [F^{-1}(0.75, X^*) - F^{-1}(0.25, X^*)]|.$$

Για συμμετρικές ή σχεδόν συμμετρικές κατανομές, είναι εφικτή η απόκτηση σωστών εκ των υστέρων με ένα μέτρο εγγύτητας που βασίζεται μόνο στις κεντρικές τιμές των $X$ και $X^*$ (βλέπε, για παράδειγμα, τα μέτρα εγγύτητας $d_4$ και $d_5$). Από την άλλη μεριά, για μη συμμετρικές κατανομές, όπως η εκθετική, μόνο η κεντρική τιμή δεν αρκεί για να παρέχει τις κρίσιμες πληροφορίες για τη λοξότητα ή τη μεταβλητότητα (βλέπε, για παράδειγμα, το μέτρο εγγύτητας $d_6$), και μπορεί να μην παράγει σωστές εκτιμήσεις της εκ των υστέρων κατανομής.

Ακολουθώντας τους Economou et al. (2021), ένα άλλο μέτρο για τη σύγκριση της εγγύτητας του συνόλου του προσομοιωμένου δείγματος $D^*$ και των πραγματικών δεδομένων $D$, που μπορεί να χρησιμοποιηθεί ανεξάρτητα από τη διάσταση, αποτελεί η απόλυτη τιμή της τυποποιημένης εκδοχής της ποσότητας:

$$T = \int \left(f_D(x) - f_{D^*}(x)\right)^2 dx,$$

όπου $f_D(x)$ και $f_{D^*}(x)$ είναι εκτιμητές τύπου πυρήνα των πυκνοτήτων (kernel density estimates) με βάση το πραγματικό δείγμα $D$ και το προσομοιωμένο δείγμα

$D^*$, αντίστοιχα. Οι εκτιμητές τύπου πυρήνα των πυκνοτήτων προκύπτουν με ένα μη παραμετρικό τρόπο εκτίμησης της συνάρτησης πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής (βλέπε, μεταξύ άλλων, Terrell and Scott (1992)).

**Παρατήρηση 4.3.1.** *Ο καθορισμός της παραμέτρου $\epsilon$ μπορεί να γίνει λαμβάνοντας υπόψη τόσο τη φύση του μοντέλου που μελετάται κάθε φορά καθώς και το μέτρο της εγγύτητας, το οποίο υπολογίζεται στο τέταρτο βήμα του αλγορίθμου των Pritchard et al. (1999). Ειδικότερα, αν χρησιμοποιηθεί η απόλυτη τιμή της τυποποιημένης εκδοχής της ποσότητας $T$ τότε μια λογική τιμή για το $\epsilon$ θα μπορούσε να είναι η τιμή 1.96. Η συγκεκριμένη τιμή δεν είναι τυχαία καθώς η στατιστική συνάρτηση $T$ ακολουθεί ασυμπτωτικά, υπό τη μηδενική υπόθεση ότι τα δείγματα $D$ και $D^*$ έχουν την ίδια πυκνότητα, κανονική κατανομή (για περισσότερες λεπτομέρειες παραπέμπουμε στους Duong et al. (2012)). Αυτό συνεπάγεται ότι η επιλογή $\epsilon = 1.96$ αντιστοιχεί σε ένα ασυμπτωτικό επίπεδο σημαντικότητας 0.05 για τον έλεγχο της μηδενικής υπόθεσης ότι και τα δύο δείγματα έχουν την ίδια πυκνότητα. Αν τώρα χρησιμοποιηθεί μια άλλη συνάρτηση $d$, τότε μια προσέγγιση για τον καθορισμό της παραμέτρου $\epsilon$ είναι να δημιουργηθούν $M$ το πλήθος τιμές $\theta^*$ από την εκ των προτέρων κατανομή που έχει υιοθετηθεί και έπειτα, με βάση την επιλογή της απόστασης $d_i$, να ταξινομηθούν τα ζεύγη $\{\theta_i, d_i\}$, $i = 1, 2, ..., l$. Στο πλαίσιο αυτό, προτείνεται το $\epsilon$ να ισούται με την $k$-οστή μικρότερη τιμή των αποστάσεων που ταξινομήθηκαν προηγουμένως (βλέπε Biau et al. (2015)), όπου $k$ ένας φυσικός αριθμός. Αυτό έχει ουσιαστικά ως αποτέλεσμα να κρατάμε τις τιμές $\theta_1, ...., \theta_{kN}$.*

**Παρατήρηση 4.3.2.** *Εκτός από την ακρίβεια της προσέγγισης, η απόσταση $d$ και το $\epsilon$ επηρεάζουν επίσης τον απαιτούμενο χρόνο υπολογισμού για να προσομοιωθούν τα δείγματα. Ως εκ τούτου, σε ένα αρχικό επίπεδο, υπάρχει μια αντιστάθμιση μεταξύ της στατιστικής και της υπολογιστικής αποτελεσματικότητας του ABC αλγορίθμου (βλέπε Beaumont et al. (2002)). Μεγαλύτερες τιμές του $\epsilon$ οδηγούν σε μεροληπτικές εκτιμήσεις $\pi_{d,\epsilon}(\theta \mid x)$, αλλά το κέρδος είναι ένας ταχύτερος αλγόριθμος, που σημαίνει μειωμένο σφάλμα Monte Carlo, καθώς μπορεί κανείς να παράγει περισσότερα δείγματα ανά μονάδα χρόνου. Επομένως, κατά τον καθορισμό του $\epsilon$, ο στόχος είναι να βρεθεί μια καλή ισορροπία μεταξύ της μεροληψίας και του σφάλματος Monte Carlo.*

Στην εργασία τους οι Pritchard et al. (1999) θεώρησαν ως μέτρο της εγγύτητας $d$ μεταξύ των περιγραφικών μέτρων την απόσταση του Chebyshev που είναι της μορφής: $\max |S_j(x) - S_j(x^*)|$, για $j = 1, ..., s$, το πλήθος περιγραφικά στατιστικά μέτρα ή την κανονικοποιημένη μορφή της που δίνεται από τη σχέση $\max\left|\frac{S_j(x)}{S_j(x^*)} - 1\right|$ για $j = 1, ..., s$. Σχετικά με την επιλογή των περιγραφικών

στατιστικών μέτρων διευκρινήσεις δίνονται στην υποενότητα που ακολουθεί.

### 4.3.3 Τα περιγραφικά στατιστικά μέτρα

Τα περιγραφικά στατιστικά μέτρα χρησιμοποιούνται για την αύξηση του ποσοστού αποδοχής του ABC για δεδομένα υψηλών διαστάσεων. Σε τέτοιες περιπτώσεις η απόδοση του ABC αλγορίθμου εξαρτάται από την επιλογή των περιγραφικών στατιστικών μέτρων (βλέπε για παράδειγμα Prangle (2015)). Τα περιγραφικά στατιστικά μέτρα υπολογίζονται από τα δεδομένα και είναι τιμές που αντιπροσωπεύουν το μέγιστο ποσό πληροφορίας που είναι διαθέσιμο από αυτά, με την απλούστερη δυνατή μορφή. Γίνεται εύκολα αντιληπτό ότι τα επαρκή περιγραφικά στατιστικά μέτρα χαμηλής διάστασης, καθώς παρέχουν όλες τις πληροφορίες που υπάρχουν στα δεδομένα με την απλούστερη δυνατή μορφή, είναι η βέλτιστη επιλογή για το σκοπό αυτό, καθώς μια επαρκή στατιστική συνάρτηση μας παρέχει τόση πληροφορία για την παράμετρο $\theta$ όση και τα ίδια τα δεδομένα μας. Δηλαδή στην ιδανική περίπτωση το περιγραφικό στατιστικό μέτρο $S(\cdot)$, σύμφωνα με τους Turner and Van Zandt (2012), πρέπει να είναι επαρκές (sufficient) για την άγνωστη παράμετρο $\theta$. Τότε η εκ των υστέρων κατανομή παίρνει τη μορφή $\pi(\theta|X) = \pi(\theta|S(X))$. Για να προσδιοριστεί λοιπόν αν μια στατιστική συνάρτηση $S(\cdot)$ είναι επαρκής πρέπει σύμφωνα με το Παραγοντικό Θεώρημα (Factorization Theorem) των Neyman-Fisher (βλέπε, για παράδειγμα, Casella and Berger (2002)), η πιθανοφάνεια $f(x|\theta)$ να μπορεί να παραγοντοποιηθεί στη μορφή:

$$f(x|\theta) = g(S(x)|\theta)\, h(x).$$

Σε περιπτώσεις που δεν είναι διαθέσιμα επαρκή στατιστικά χρησιμοποιούνται άλλα περιγραφικά στατιστικά μέτρα, τα οποία εισάγουν μια επιπλέον μεροληψία λόγω της απώλειας πληροφοριών. Ένα εύλογο ερώτημα είναι αν θα μπορούσε κάποιος να αξιοποιήσει τα περιγραφικά στατιστικά συνδυάζοντάς τα με κατάλληλα ορισμένες συναρτήσεις βάρους. Η απάντηση στο ερώτημα αυτό δίνεται στην παρατήρηση που ακολουθεί.

**Παρατήρηση 4.3.3.** *Για την κατάλληλη επιλογή των περιγραφικών στατιστικών μέτρων θα μπορούσε να χρησιμοποιηθεί η στάθμιση διαφορετικών περιγραφικών στατιστικών μέτρων (weighting summary statistics) ανάλογα με το μέγεθος κάποιου μέτρου συσχέτισης με τις τιμές των άγνωστων παραμέτρων. Ακολουθώντας τους Beaumont (2010), αναφέρουμε επίσης ότι οι Hamilton et al. (2005) χρησιμοποίησαν συναρτήσεις βάρους βάσει των περιθωριακών συντελεστών παλινδρόμησης κάθε περιγραφικού στατιστικού με την παράμετρο ενδιαφέροντος και μάλιστα απέδειξαν ότι αυτή η προσέγγιση απέδωσε βελτίωση σε σχέση με την*

*τεχνική των Beaumont et al. (2002). Η βασική ιδέα σε αυτό το πλαίσιο είναι ότι στις περιπτώσεις που ένα περιγραφικό στατιστικό έχει μικρή συσχέτιση με μια τιμή της άγνωστης παραμέτρου, τότε θα πρέπει να ανεχτούμε μεγάλες αποκλίσεις του περιγραφικού στατιστικού μέτρου από την κατανομή στόχο δίνοντάς του ένα χαμηλότερο βάρος στον υπολογισμό της Ευκλείδειας απόστασης. Σε διαφορετικές περιπτώσεις, υπάρχει ο κίνδυνος απόρριψης σημείων όπου τα πληροφοριακά (informative) περιγραφικά στατιστικά βρίσκονται κοντά στην κατανομή στόχο.*

Οι Sisson et al. (2007) απέδειξαν ότι η χρήση ενός μόνο διατεταγμένου στατιστικού μεγάλης τάξης, όπως το μέγιστο ή το ελάχιστο, δεν οδηγεί σε καλές εκτιμήσεις της εκ των υστέρων κατανομής. Στο πλαίσιο αυτό, επισημαίνεται ότι οι εκτιμητές μέγιστης πιθανοφάνειας στις περιπτώσεις που προσδιορίζονται σε κλειστή, αναλυτική, μορφή αποτελούν επίσης τη βάση για μια καλή επιλογή του μέτρου εγγύτητας. Ας σημειωθεί επίσης ότι δεν είναι απαραίτητο να περιοριζόμαστε στη χρήση μόνο ενός περιγραφικού στατιστικού μέτρου. Στη βιβλιογραφία, πολλές φορές, έχουν συνδυαστεί ορισμένα περιγραφικά στατιστικά μέτρα με στόχο την αποτελεσματική σύνδεση των παραμέτρων του μοντέλου με τα πραγματικά δεδομένα. Για περισσότερες πληροφορίες παραπέμπουμε στους Bazin et al. (2010) και τις εκεί αναφορές.

Αν αντικατασταθεί το πλήρες σύνολο δεδομένων από τα περιγραφικά στατιστικά μέτρα τότε με ανάλογο τρόπο θα αντικατασταθεί η σχέση $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$ από τη σχέση $\pi(\theta|s) \propto f(s|\theta)\pi(\theta)$, όπου $s = S(x)$ είναι το διάνυσμα των στατιστικών περιγραφικών μέτρων (βλέπε, για παράδειγμα, Sisson et al. (2018)). Ωστόσο, ο προσδιορισμός επαρκών στατιστικών συναρτήσεων χαμηλής διάστασης είναι συνήθως ανέφικτος για τα στατιστικά μοντέλα στα οποία χρησιμοποιείται η ABC, καθώς εμπεριέχεται στον προσδιορισμό τους η συνάρτηση πιθανοφάνειας. Κατά συνέπεια κάποια ευρετική μέθοδος είναι συνήθως απαραίτητη για τον προσδιορισμό χρήσιμων περιγραφικών στατιστικών μέτρων χαμηλών διαστάσεων, ενώ τις περισσότερες φορές στα πραγματικά προβλήματα τα περιγραφικά στατιστικά μέτρα καθορίζονται από το υπό μελέτη πρόβλημα (βλέπε Marin et al. (2012)). Προφανώς σε αυτήν την περίπτωση η πραγματοποίηση συμπερασμάτων βάσει των περιγραφικών στατιστικών μέτρων αντί του πλήρους συνόλου δεδομένων αναπόφευκτα συνεπάγεται την απόρριψη ενδεχομένως χρήσιμων πληροφοριών που υπάρχουν στα δεδομένα, ενώ αν ένα περιγραφικό στατιστικό μέτρο δεν είναι επαρκές για την παράμετρο ενδιαφέροντος, τότε η εκ των υστέρων κατανομή που υπολογίζεται με αυτήν τη στατιστική συνάρτηση δε θα είναι ανάλογη με την εκ των υστέρων κατανομή που υπολογίστηκε με το πλήρες σύνολο δεδομένων (βλέπε Marjoram and Tavaré (2006)). Επιπρόσθετα, οι Nunes and Balding (2010) παρατηρούν ότι η καλύτερη επιλογή των περιγραφικών στατιστικών μέτρων ποικίλει

στα σύνολα δεδομένων, ενώ η επιλογή περιγραφικών στατιστικών μέτρων που δεν έπρεπε να επιλεγούν, συχνά οδηγεί σε πληθώρα μη αξιόπιστων συμπερασμάτων λόγω απώλειας πληροφοριών (βλέπε, για παράδειγμα, Csilléry et al. (2010)).

Συνεπώς, επιθυμητή είναι μια ισορροπία του ποσού πληροφόρησης και της χαμηλής διάστασης. Στη βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι με στόχο την καλύτερη επιλογή περιγραφικών στατιστικών μέτρων (βλέπε, για παράδειγμα, Sisson et al. (2018)). Κάποιες από αυτές είναι οι ακόλουθες:

1. Ένα διαδοχικό σχήμα βασισμένο στην αρχή της κατά προσέγγισης επάρκειας. Οι Joyce and Marjoram (2008) αναπτύσσουν ένα διαδοχικό σχήμα (sequential scheme) των περιγραφικών στατιστικών με κριτήριο αν η ένταξη ενός τυχαία επιλεγμένου στατιστικού, σε κάθε βήμα, στην ανάλυση θα βελτιώσει ουσιαστικά την ποιότητα της στατιστικής συμπερασματολογίας. Η μέθοδος που πρότειναν σαφώς δεν είναι βέλτιστη, διότι η σειρά με την οποία προστίθενται τα στατιστικά στοιχεία θα έχει σημασία, αλλά μπορεί να εφαρμοστεί σε σύνολα δεδομένων υψηλότερων διαστάσεων, για τα οποία οι εξισώσεις πιθανοφάνειας δεν είναι διαθέσιμες σε κλειστή, αναλυτική μορφή.

2. Η επιλογή ενός υποσυνόλου των στατιστικών περιγραφικών μέτρων ελαχιστοποιώντας προκαθορισμένα κριτήρια όπως το Akaike κριτήριο πληροφορίας (βλέπε Blum et al. (2012)) ή η εντροπία μιας κατανομής (Nunes and Balding (2010)).

3. Οι Wegmann et al. (2009) εφαρμόζουν μια τεχνική μερικής παλινδρόμησης ελαχίστων τετραγώνων, η οποία βρίσκει γραμμικούς συνδυασμούς των αρχικών στατιστικών περιγραφικών μέτρων, για τα οποία χρησιμοποιείται οποιαδήποτε διαδικασία για τη μείωση της συσχέτισης, αλλά ταυτόχρονα είναι και ισχυρά συσχετισμένα με τις παραμέτρους.

4. Τέλος, θεωρώντας ένα στατιστικό μοντέλο μεταξύ των παραμέτρων και των μετασχηματισμένων στατιστικών των προσομοιωμένων δεδομένων, τα στατιστικά περιγραφικά μέτρα επιλέγονται ελαχιστοποιώντας μια συνάρτηση απώλειας (βλέπε, για παράδειγμα, Aeschbacher et al. (2012)).

### 4.3.4 Οι εκ των προτέρων κατανομές

Στο πρώτο βήμα του αλγορίθμου, όπως έχει ήδη αναφερθεί, θα πρέπει να επιλεγεί μια εκ των προτέρων κατανομή για κάθε συνιστώσα της παραμέτρου $\theta$. Οποιαδήποτε, λοιπόν, προηγούμενη γνώση πρέπει να εφαρμοσθεί σε αυτές τις εκ των προτέρων κατανομές. Σε περιπτώσεις που δεν υπάρχουν προηγούμενες

πληροφορίες, αρχικά συνιστώνται εκ των προτέρων κατανομές με μεγάλες τυπικές αποκλίσεις. Μετά από έναν προκαθορισμένο αριθμό μη απορριπτέων τιμών (βλέπε Βήμα 4 της μεθόδου), οι εκ των προτέρων κατανομές όλων των παραμέτρων μπορούν να επαναπροσδιοριστούν αξιοποιώντας τις πληροφορίες που αποκτήθηκαν από αυτά τα δείγματα για να αυξηθεί κατά αυτόν τον τρόπο το ποσοστό αποδοχής. Με απώτερο σκοπό να διατηρηθεί η επιλογή των εκ των προτέρων κατανομών της ABC μεθόδου όσο το δυνατόν απλούστερη, πολύ πρόσφατα, οι Economou et al. (2021) πρότειναν κάποιες εκ των προτέρων κατανομές βάσει του πεδίου τιμών της παραμέτρου. Ειδικότερα, πρότειναν τη χρήση της περικομμένης προς τα δεξιά στο μηδέν κανονικής κατανομής ως εκ των προτέρων κατανομή για οποιαδήποτε θετική παράμετρο, την κανονική κατανομή για οποιαδήποτε πραγματική παράμετρο και τέλος την ομοιόμορφη κατανομή για οποιαδήποτε παράμετρο με τιμές σε ένα πεπερασμένο διάστημα $[a, b]$. Αυτές οι εκ των προτέρων κατανομές είναι οι εκ των προτέρων κατανομές μέγιστης εντροπίας (maximum entropy priors), δεδομένης μόνο της συνέχειας και του στηρίγματος των παραμέτρων (βλέπε Schroeder (2010)). Ωστόσο, τέτοιες εκ των προτέρων κατανομές μπορεί να μην είναι βέλτιστες, καθώς οποιαδήποτε άλλη εκ των προτέρων κατανομή που ενσωματώνει με καλύτερο, πιο πληροφοριακό τρόπο, οποιαδήποτε προηγούμενη ειδική γνώση από το πρόβλημα είναι φυσικά προτιμότερη. Σε εφαρμογές που χρησιμοποιούνται μη βέλτιστες εκ των προτέρων κατανομές η μέθοδος μπορεί πάλι να επιφέρει λογικές εκτιμήσεις παραμέτρων, παρότι η υιοθέτηση τέτοιων εκ των προτέρων κατανομών μπορεί να οδηγήσει σε αυξημένο ποσοστό απόρριψης και συνεπώς σε μεγαλύτερο αριθμό προσομοιώσεων (βλέπε Βήμα 4).

## 4.4 Παραδείγματα κατανόησης

Στην ενότητα αυτή μέσω απλών παραδειγμάτων, που έχουν παρατεθεί και παρουσιασθεί, μεταξύ άλλων, από τους Turner and Van Zandt (2012), Sunnåker et al. (2013) και Hartig (2014), θα διευκρινιστούν και θα εφαρμοσθούν τα βασικά βήματα της μεθόδου ABC.

**Παράδειγμα 4.4.1** (**Toy example**). *Να υλοποιηθεί η παραλλαγή του αλγορίθμου ABC των Pritchard et al. (1999) για την κατανομή Weibull με παραμέτρους $a = 3$ και $b = 6$, όταν το σύνολο των πραγματικών δεδομένων είναι μεγέθους 40.*

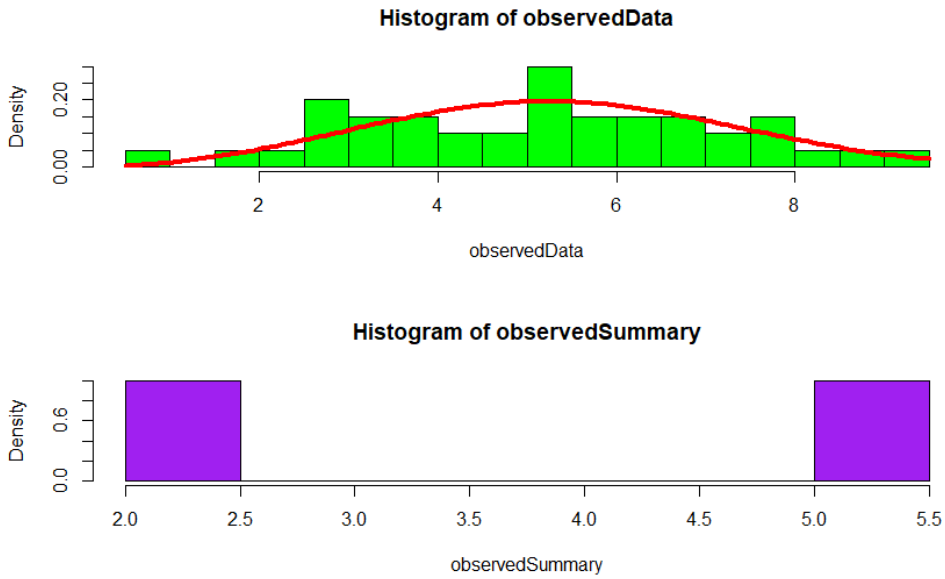
***Λύση***. Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Weibull δίνεται

από τη σχέση:

$$f\left(x|a,b\right)=\begin{cases}\dfrac{b}{a}\left(\dfrac{x}{a}\right)^{b-1}e^{-(x/a)^b}, & \text{αν } x\geq 0\\ 0, & \text{διαφορετικά.}\end{cases}$$

όπου για το συγκεκριμένο παράδειγμα $a=3, b=6$. Καθώς δε μας δίνεται το σύνολο δεδομένων, παράγουμε ένα τυχαίο δείγμα μεγέθους $n=40$ από αυτήν την κατανομή. Το σύνολο αυτό των δεδομένων αποτελεί το σύνολο των πραγματικών, παρατηρούμενων δεδομένων και στο Σχήμα 4.1 δίνεται το ιστόγραμμα αυτών.

Σχήμα 4.1: Ιστογράμματα των 40 παρατηρούμενων τιμών από την κατανομή $Weibull(3,6)$ και των παρατηρούμενων στατιστικών περιγραφικών μέτρων αντίστοιχα




Σύμφωνα με τα βήματα της μεθόδου θα πρέπει αρχικά να ορίσουμε μια εκ των προτέρων κατανομή για κάθε άγνωστη πληθυσμιακή παράμετρο. Στο πλαίσιο αυτό, υποθέτουμε ότι $a\sim U(0.01,7)$ και $b\sim U(0.01,12)$. Στο δεύτερο βήμα της μεθόδου παράγουμε ένα μεγάλο πλήθος διαφορετικών τιμών των παραμέτρων (έστω ότι είναι 80.000) από τις προαναφερθείσες εκ των προτέρων κατανομές, ενώ στο τρίτο βήμα προσομοιώνουμε ένα δείγμα $D^*=(x_i^*)$, $i=1,...,80.000$,

από το μοντέλο $f(\cdot|\theta_i)$, με $\theta_i$ να είναι οι τιμές που προσομοιώθηκαν στο προηγούμενο βήμα. Επιλέγουμε τη μέση τιμή και την τυπική απόκλιση ως στατιστικά περιγραφικά μέτρα, ενώ όσον αφορά την τιμή της παραμέτρου $\epsilon$ θα δοκιμάσουμε τρεις τιμές, τις 1, 0.5 και 0.35. Ας επισημανθεί ότι καθώς το $\epsilon$ γίνεται όλο και πιο μικρό ο αριθμός των παραμέτρων που γίνονται αποδεκτές μειώνεται. Στην πράξη, η εξασφάλιση περισσότερων αποδεκτών δειγμάτων είναι συνήθως το πιο σημαντικό ζήτημα. Τέλος, ως μέτρο της εγγύτητας των περιγραφικών στατιστικών μέτρων μπορεί να επιλεγεί είτε η Ευκλείδεια απόσταση είτε η απόσταση Manhattan, η οποία ορίζεται ως το άθροισμα των απόλυτων διαφορών των αντίστοιχων συνιστωσών των δύο σημείων.

Τα παραπάνω, χρησιμοποιώντας την Ευκλείδεια απόσταση, υλοποιούνται στο περιβάλλον της R χρησιμοποιώντας τις ακόλουθες εντολές (βλέπε, σχετικά παρόμοιο κώδικα, Hartig (2014)):

```
observedData = rweibull(40, 3, 6)
observedSummary = c(mean(observedData), sd(
   observedData))
par(mfrow=c(2,1))
hist(observedData,breaks=30,col="green",prob = TRUE)
curve(dweibull(x,3,6), add =TRUE, col = "red",lwd=4)
hist(observedSummary,breaks=10,col="purple",prob =
   TRUE)
model = function(par){
  simulatedData = rweibull(40, par[1,1], par[1,2])
  simulatedSummary = c(mean(simulatedData), sd(
     simulatedData))
  return(simulatedSummary)
}
n = 80000
fit = data.frame(shape = runif(n, 0.01, 7), scale =
   runif(n, 0.01,12), summary1 = rep(NA, n), summary2
    = rep(NA, n), distance = rep(NA, n))
for (i in 1:n){
  prediction = model(fit[i,1:2])
  deviation = sqrt(sum(( prediction - observedSummary
     )^2))
  fit[i,3:5] = c(prediction, deviation)
}
par(mfrow=c(1,2))
```

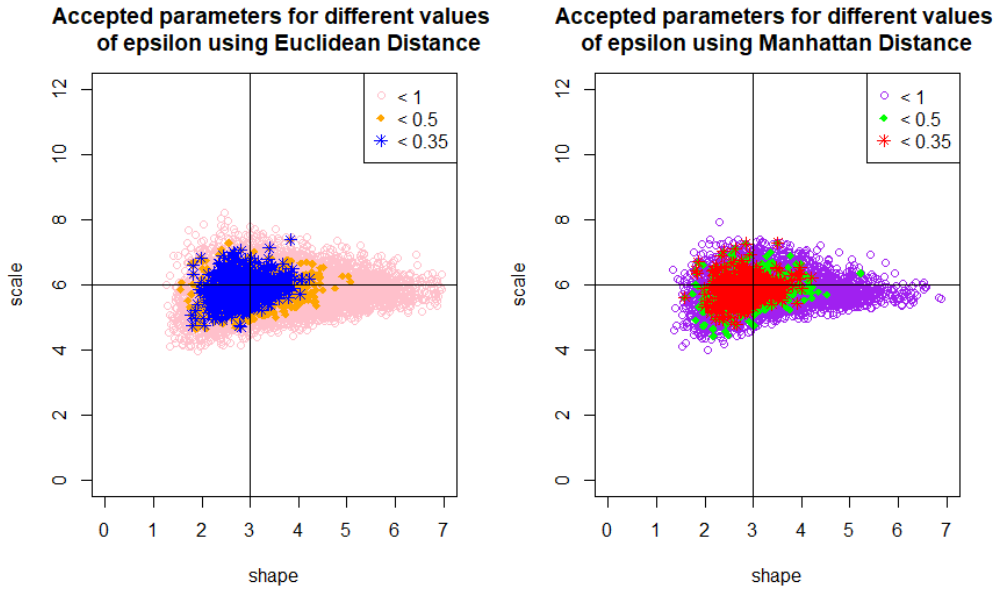
```
plot(fit[fit[,5] < 1, 1:2], xlim = c(0.01,7), ylim = c(0.01,12), col = "pink",
     main = "Accepted parameters for different values
     of epsilon using Euclidean Distance")
points(fit[fit[,5] < 0.5, 1:2],  pch = 18, col = "orange")
points(fit[fit[,5] < 0.35, 1:2],  pch = 8, col = "blue")
legend("topright", c("< 1", "< 0.5", "< 0.35"), pch = c(1,18,8), col = c("pink", "orange", "blue"))
abline(v = 3)
abline(h = 6)
```

Παρατηρείστε ότι τα περισσότερα σημεία συγκεντρώνονται στην περιοχή των πραγματικών τιμών των παραμέτρων, που απεικονίζονται από τις δύο γραμμές κατά μήκος του γραφήματος. Αν τώρα θέλαμε να χρησιμοποιήσουμε την απόσταση Manhattan θα έπρεπε να εκτελέσουμε τις ακόλουθες εντολές:

```
for (i in 1:n){
  prediction = model(fit[i,1:2])
  deviation = sum(abs(prediction - observedSummary))
  fit[i,3:5] = c(prediction, deviation)
}
plot(fit[fit[,5] < 1, 1:2], xlim = c(0.01,7), ylim = c(0.01,12), col = "purple",
     main = "Accepted parameters for different values
of epsilon using Manhattan Distance")
points(fit[fit[,5] < 0.5, 1:2],  pch = 18, col = "green")
points(fit[fit[,5] < 0.35, 1:2],  pch = 8, col = "red")
legend("topright", c("< 1", "< 0.5", "< 0.35"), pch = c(1,18,8), col = c("purple", "green", "red"))
abline(v = 3)
abline(h = 6)
```

Πριν προχωρήσουμε στο επόμενο παράδειγμα θα δώσουμε κάποιες σχετικές

Σχήμα 4.2: Γραφική παράσταση των αποδεκτών τιμών για τις διάφορες τιμές του $\epsilon$, με την παραλλαγή του αλγορίθμου ABC των Pritchard et al. (1999), χρησιμοποιώντας την Ευκλείδεια και την Manhattan απόσταση.




πληροφορίες. Η μέθοδος ABC μπορεί να εφαρμοστεί σε μια κατηγορία μοντέλων, που συναντώνται σε διάφορα επιστημονικά πεδία (Θερμοδυναμική, Βιολογία) και που έχουν τη δομή που θα περιγραφεί παρακάτω (βλέπε Sunnåker et al. (2013)). Το μοντέλο είναι στην πραγματικότητα ένα σύστημα δύο καταστάσεων (bistable), έστω A και B και μπορεί να χαρακτηριστεί από ένα κρυφό Μαρκοβιανό μοντέλο (hidden Markov model, HMM) που οι πιθανότητες μετάβασης καθορίζονται ως εξής. Η πιθανότητα μετάβασης από τη μία κατάσταση στην άλλη ισούται με $\theta$ και στις δύο κατευθύνσεις, δηλαδή, για το χώρο καταστάσεων $\{A, B\}$, ισχύουν τα ακόλουθα:

$$P(A \to B) = P(B \to A) = \theta \text{ και } P(A \to A) = P(B \to B) = 1 - \theta.$$

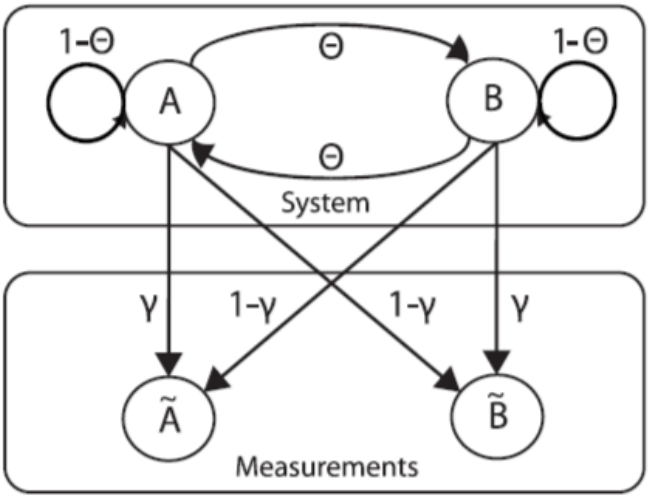
Επιπρόσθετα, η πιθανότητα σωστής μέτρησης μιας κατάστασης ισούται με $\gamma$ και, αντίθετα, η πιθανότητα λανθασμένης μέτρησης (εσφαλμένη ανάγνωση μιας κατάστασης του συστήματος) είναι $1 - \gamma$. Αυτό σημαίνει αν $\tilde{A}$ και $\tilde{B}$ είναι οι

μετρήσεις τότε ισχύει ότι:

$$P\left(A \to \tilde{A}\right) = P\left(B \to \tilde{B}\right) = \gamma \text{ και } P\left(A \to \tilde{B}\right) = P\left(B \to \tilde{A}\right) = 1 - \gamma.$$

Τα παραπάνω απεικονίζονται στο Σχήμα 4.3.

Σχήμα 4.3: Κρυφό Μαρκοβιανό δυναμικό μοντέλο δύο καταστάσεων. Πηγή σχήματος: Sunnåker et al. (2013).




Λόγω των εξαρτήσεων που υπάρχουν μεταξύ των καταστάσεων σε διαφορετικά χρονικά σημεία, ο υπολογισμός της συνάρτησης πιθανοφάνειας των δεδομένων, που στην ουσία αποτελούν χρονοσειρές, είναι μια χρονοβόρα και απαιτητική διαδικασία. Το γεγονός αυτό δίνει και το κίνητρο για να εφαρμοστεί η μέθοδος ABC, αποτελώντας το δεύτερο παράδειγμα κατανόησης της μεθόδου.

**Παράδειγμα 4.4.2.** *Για το κρυφό Μαρκοβιανό μοντέλο δύο καταστάσεων που περιγράφηκε να εφαρμόσετε την παραλλαγή του αλγορίθμου ABC των Pritchard et al. (1999) υπό την υπόθεση ότι τα πραγματικά δεδομένα είναι η σταθερή ακολουθία καταστάσεων AAAABAABBAAAAAABAAAA (αντιστοιχούν στην τιμή* $\theta = 0.25$).

***Λύση***. Σύμφωνα με τα βήματα της μεθόδου θα πρέπει αρχικά να ορίσουμε μια εκ των προτέρων κατανομή για κάθε άγνωστη πληθυσμιακή παράμετρο. Καθώς

δεν έχουμε καμία πληροφορία για την άγνωστη παράμετρο $\theta$ θα πρέπει να χρησιμοποιήσουμε μία μη πληροφοριακή εκ των προτέρων κατανομή. Για το λόγο αυτό επιλέγεται η ομοιόμορφη στο διάστημα $[0, 1]$. Η επιλογή αυτή γίνεται και για λόγους ευκολίας της προσομοίωσης των $\theta$ από την $U[0, 1]$. Στο δεύτερο βήμα της μεθόδου παράγουμε ένα πλήθος διαφορετικών τιμών της παραμέτρου $\theta$. Εδώ καθώς πρόκειται για παράδειγμα κατανόησης επιλέγουμε την τιμή $n = 5$. Ας σημειωθεί ότι το $n$ θα πρέπει να είναι πολύ μεγάλο για να προκύψει μια καλή προσέγγιση. Οι πέντε τιμές της παραμέτρου $\theta$ που παράγονται είναι: $\theta_1 = 0.08$, $\theta_2 = 0.68$, $\theta_3 = 0.87$, $\theta_4 = 0.43$ και $\theta_5 = 0.53$, αντίστοιχα. Στο τρίτο βήμα και για κάθε μία από τις τιμές $\theta_i$, $i = 1, ..., 5$, που προέκυψαν στο προηγούμενο βήμα παράγουμε μια ακολουθία (συνολικά πέντε το πλήθος) 20 καταστάσεων. Η τιμή 20 δεν επιλέγεται τυχαία αλλά τόσο είναι το πλήθος της ακολουθίας καταστάσεων που παρατηρήθηκε. Έτσι παράγονται τα ακόλουθα προσομοιωμένα σύνολα δεδομένων:

$$
\begin{aligned}
AABAAAABAABAAABAAAAA, &\ \text{για } \theta_1 = 0.08,\\
AABBABABAAABBABABBAB, &\ \text{για } \theta_2 = 0.68,\\
BBBABBABBBBABABBBBBA, &\ \text{για } \theta_3 = 0.87,\\
AABAAAAABBABBBBBBBBA, &\ \text{για } \theta_4 = 0.43,\\
ABBBBBAABBABBABAABBB, &\ \text{για } \theta_5 = 0.53
\end{aligned}
$$

όπου για διευκόλυνση με κόκκινο χρώμα επισημαίνεται οποιαδήποτε εναλλαγή μεταξύ των δύο καταστάσεων και αυτό γίνεται καθώς οι Sunnåker et al. (2013) επιλέγουν ως συνοπτικό περιγραφικό μέτρο τη στατιστική συνάρτηση $\omega(Y)$ που παριστάνει τη συχνότητα των εναλλαγών μεταξύ των δύο καταστάσεων στην ακολουθία καταστάσεων $Y$. Με άλλα λόγια η στατιστική συνάρτηση $\omega(Y)$ καταγράφει τον αριθμό των φορών που το σύστημα που μελετούμε θα μεταβεί από την κατάσταση $A$ στην $B$, και αντίστοιχα από την κατάσταση $B$ στην $A$ στην ακολουθία καταστάσεων $Y$. Με βάση αυτόν τον ορισμό προκύπτει ότι στην πραγματική ακολουθία καταστάσεων $S$ ισούται με $\omega(S) = 6$, ενώ για κάθε μία από τις προσομοιωμένες ακολουθίες καταστάσεων, έστω $S_i$, $i = 1, ..., n$, είναι

$$\omega(S_1) = 8,\ \omega(S_2) = 13,\ \omega(S_3) = 9,\ \omega(S_4) = 6,\ \omega(S_5) = 9.$$

Αφού επιλέχθηκε και το συνοπτικό περιγραφικό στατιστικό μέτρο, στο επόμενο βήμα πρέπει να επιλεγεί το μέτρο εγγύτητας των περιγραφικών μέτρων που θα χρησιμοποιηθεί και η σταθερά ανεκτικότητας $\epsilon$. Ως μέτρο της εγγύτητας επιλέγεται η απόλυτη διαφορά, ενώ ως τιμή της σταθεράς ανεκτικότητας η τιμή $\epsilon = 2$ (βλέπε Sunnåker et al. (2013)). Υπολογίζεται στη συνέχεια η απόλυτη διαφορά των $\omega(S_i) - \omega(S)$ για $i = 1, ..., 5$ και γίνονται αποδεκτές οι τιμές της παραμέτρου

$\theta$ που αντιστοιχούν σε προσομοιωμένα σύνολα δεδομένα με τιμή της απόλυτης διαφοράς μικρότερη ή ίση του $\epsilon = 2$. Οι τιμές των απόλυτων διαφορών είναι:

$$\begin{aligned}
d(\omega_{S,1}, \omega_S) &= \mid \omega_{S,1} - \omega_S \mid = \mid 8 - 6 \mid &= 2 \leq 2,\\
d(\omega_{S,2}, \omega_S) &= \mid \omega_{S,2} - \omega_S \mid = \mid 13 - 6 \mid &= 7 > 2,\\
d(\omega_{S,3}, \omega_S) &= \mid \omega_{S,3} - \omega_S \mid = \mid 9 - 6 \mid &= 3 > 2,\\
d(\omega_{S,4}, \omega_S) &= \mid \omega_{S,4} - \omega_S \mid = \mid 6 - 6 \mid &= 0 \leq 2,\\
d(\omega_{S,5}, \omega_S) &= \mid \omega_{S,5} - \omega_S \mid = \mid 9 - 6 \mid &= 3 > 2.
\end{aligned}$$

Τα παραπάνω σημαίνουν ότι μόνο οι τιμές $\theta_1 = 0.08$ και $\theta_4 = 0.43$ (με μέση τιμή 0.255) γίνονται δεκτές από τον αλγόριθμο αποδοχής-απόρριψης ABC, ενώ οι υπόλοιπες απορρίπτονται. Παρατηρείστε ότι αν το επίπεδο ανεκτικότητας ήταν $\epsilon = 0$, τότε από τον αλγόριθμο αποδοχής-απόρριψης ABC θα γινόταν δεκτή μία τιμή της παραμέτρου $\theta$, ενώ αν $\epsilon = 3$ θα γινόντουσαν δεκτές οι τέσσερις εκ των πέντε τιμών.

**Παράδειγμα 4.4.3.** *Να εφαρμόσετε την παραλλαγή του αλγορίθμου ABC των Pritchard et al. (1999) για τη διωνυμική κατανομή B(n, p) με πιθανότητα επιτυχίας $p = 0.7$, όταν το σύνολο των πραγματικών δεδομένων είναι μεγέθους $n = 25, 100$ και 1.000, αντίστοιχα.*

***Λύση.*** Η συνάρτηση πιθανότητας της διωνυμικής κατανομής δίνεται από τη σχέση:

$$p_X(x) = p(x|p) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, 2, ..., n,$$

όπου για το συγκεκριμένο παράδειγμα η πιθανότητα επιτυχίας ισούται με $p = 0.7$. Εφόσον δε μας δίνεται το σύνολο δεδομένων, παράγουμε ένα τυχαίο δείγμα μεγέθους $n = 25$ από αυτήν την κατανομή. Το σύνολο αυτό των δεδομένων αποτελεί το σύνολο των πραγματικών, παρατηρούμενων δεδομένων. Στη συνέχεια, με ανάλογο τρόπο, θα παράγουμε ένα τυχαίο δείγμα μεγέθους $n = 100$ και ένα τυχαίο δείγμα μεγέθους $n = 1.000$ κρατώντας σταθερή την πιθανότητα επιτυχίας $p = 0.7$.

Σύμφωνα με τα βήματα της μεθόδου, θα πρέπει αρχικά να ορίσουμε μια εκ των προτέρων κατανομή για κάθε άγνωστη πληθυσμιακή παράμετρο. Στην κλάση των διωνυμικών μοντέλων, στη Μπεϋζιανή ανάλυση, η υπόθεση μιας Βήτα εκ των προτέρων κατανομής, με παραμέτρους $a, b > 0$, αποτελεί μια καλή επιλογή. Το συμπέρασμα αυτό αιτιολογείται άμεσα επειδή γνωρίζουμε ότι η άγνωστη παράμετρος κυμαίνεται μεταξύ 0 και 1. Έπειτα, ο προσδιορισμός μιας μη πληροφοριακής εκ των προτέρων κατανομής για την παράμετρο $\theta = p$ είναι, σε πολλές περιπτώσεις,

απαραίτητος. Παρατηρείστε, επίσης, ότι για $p \sim Beta(1,1) \equiv U(0,1)$ (δηλαδή θέτοντας $a = 1$ και $b = 1$), εύκολα προκύπτει ότι η κατανομή $Beta(1,1)$ αποτελεί μια μη πληροφοριακή εκ των προτέρων κατανομή για την παράμετρο $p$. Άρα $\pi(\theta) = 1$, $\theta \in (0,1)$. Συνδυάζοντας τα παραπάνω και εφαρμόζοντας τη σχέση $\pi(\theta|x) = \dfrac{f(x|\theta)\,\pi(\theta)}{\displaystyle\int f(x|\theta)\,\pi(\theta)\,\mathrm{d}\theta}$, έπεται ότι η εκ των υστέρων κατανομή, δηλαδή η $\pi(\theta|x)$, είναι η $B(a+x, n+b-x)$, με $a = 1$ και $b = 1$.

Προχωρώντας, και παρακινούμενοι από τους Turner and Van Zandt (2012), προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n = 25$, από την κατανομή Bernoulli. Στο σημείο αυτό πρέπει να καθορισθεί ένα μέτρο της εγγύτητας ώστε να συγκριθούν οι πραγματικές τιμές με τις αντίστοιχες προσομοιωμένες που παράγονται. Σε αυτό το παράδειγμα, η σύγκριση των συνόλων $D$ και $D^*$ επιτυγχάνεται αξιοποιώντας την απόσταση $d = \dfrac{1}{n}|D - D^*|$ και θέτοντας το επίπεδο ανεκτικότητας $\epsilon$ να ισούται με 0.

Επιπλέον, αξίζει να επισημανθεί ότι όσο αυξάνεται το μέγεθος δείγματος $n$ τόσο αυξάνεται το ποσό της πληροφορίας για την παράμετρο $\theta = p$. Συνεπώς, αυξάνεται η ακρίβεια του αλγορίθμου και η εκ των υστέρων κατανομή παρουσιάζει σημείο μεγίστου. Το βασικό χαρακτηριστικό αυτού του παραδείγματος είναι η απλότητά του. Αυτό σημαίνει ότι η προσομοίωση μεγάλου πλήθους τιμών (στη συγκεκριμένη περίπτωση $N = 1.000$) για την παράμετρο $p$ είναι εφικτή και μάλιστα με αμελητέο κόστος (βλέπε Turner and Van Zandt (2012)).

Για την υλοποίηση των παραπάνω χρησιμοποιήθηκαν οι εντολές που ήταν διαθέσιμες στον ιστότοπο `https://github.com/elpidapatta/approximate-bayesian-computation`, στα πλαίσια της μεταπτυχιακής διατριβής της Πάττα (2020).

## 4.5 Τροποποιήσεις του αλγορίθμου ABC

Στην Ενότητα 4.2 παρουσιάστηκαν οι δύο βασικές μορφές του αλγορίθμου ABC, ήτοι η αρχική μορφή του αλγορίθμου για τη διακριτή περίπτωση και για πεπερασμένο σύνολο και η επέκτασή της στη συνεχή περίπτωση. Στη βιβλιογραφία έχουν εμφανιστεί πλήθος άλλων τροποποιήσεων του αλγορίθμου, σε μια προσπάθεια να μειωθεί ο υπολογιστικός χρόνος και να αυξηθεί η απόδοση του αλγορίθμου. Σκοπός αυτής της ενότητας δεν είναι η εξαντλητική παρουσίαση και παράθεση αυτών των τροποποιήσεων, αλλά η παρουσίαση των κυριότερων εξ αυτών, με κύριο κριτήριο τη συχνότητα χρησιμοποίησή τους σε πρακτικές εφαρ-

μογές.

### 4.5.1 MCMC-ABC αλγόριθμος

Μία σημαντική δυσκολία που προκύπτει κατά την υλοποίηση της τεχνικής που προτάθηκε από τους Pritchard et al. (1999) είναι ότι η χρήση μη πληροφοριακών εκ των προτέρων κατανομών οδηγεί σε προτεινόμενες τιμές που ανήκουν σε περιοχές χαμηλής εκ των υστέρων πιθανότητας. Το αποτέλεσμα αυτό αιτιολογείται από την παρατήρηση ότι τη στιγμή της προσομοίωσης τιμών από την εκ των προτέρων δεν αξιοποιούνται τα δεδομένα. Το γεγονός αυτό σημαίνει ότι η προσομοίωση τιμών $\theta^*$ από την εκ των προτέρων κατανομή (βλέπε Βήμα 2 της μεθόδου) δεν είναι αποδοτική. Τα παραπάνω αποτέλεσαν το βασικότερο κίνητρο των Marjoram et al. (2003) ώστε να εισάγουν μια εναλλακτική μορφή της παραλλαγής του αλγορίθμου των Pritchard et al. (1999), η οποία θα έχει την ιδιότητα να ξεπερνά τα ζητήματα που προαναφέρθηκαν. Ο στόχος επιτυγχάνεται αν ο προτεινόμενος μηχανισμός εμφυτευθεί σε έναν MCMC δειγματολήπτη, οπότε ο προτεινόμενος μηχανισμός γίνεται πιο προσαρμοστικός καθώς λαμβάνουμε πιο αποτελεσματικά δείγματα από την εκ των υστέρων (Beaumont (2010)). Στη συνέχεια παρουσιάζεται αναλυτικά η επαναληπτική διαδικασία που χρησιμοποίησαν. Ο αλγόριθμος που προκύπτει είναι γνωστός ως Αλγόριθμος MCMC-ABC.

**Αλγόριθμος MCMC-ABC των Marjoram et al. (2003)**:

1. Υιοθετούμε τις αρχικές τιμές $(\theta_0^*, x_0^*)$ χρησιμοποιώντας την παραλλαγή του Pritchard et al. (1999) αλγορίθμου.

2. Για $i = 1, 2, ..., N$:

3. Προσομοιώνουμε την τιμή $\theta^*$ από την κατανομή πρότασης $q(\cdot|\theta^{(i-1)})$.

4. Προσομοιώνουμε την τιμή $x^*$ από το μοντέλο $f(\cdot|\theta^*)$.

5. Προσομοιώνουμε την τιμή $u$ από την ομοιόμορφη κατανομή στο (0, 1).

6. Αν $u \leq \dfrac{f(\theta^*)q(\theta^{(i-1)}|\theta^*)}{f(\theta^{(i-1)})q(\theta^*|\theta^{(i-1)})}$ και $d(s(x^*), s(x)) \leq \epsilon$ τότε θέσε $(\theta^{(i)}, x^{(i)}) = (\theta^*, x^*)$, αλλιώς θέσε $(\theta^{(i)}, x^{(i)}) = (\theta^{(i-1)}, x^{(i-1)})$.

7. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $M$, μη απορριπτέων τιμών $\theta^*$.

**Παρατήρηση 4.5.1.** *Η υιοθέτηση των αρχικών τιμών αξιοποιώντας την παραλλαγή του Pritchard et al. (1999) αλγορίθμου (βλέπε Βήμα 1 του παραπάνω αλγορίθμου) αποτελεί ένα σημείο που μπορεί να παραλειφθεί, διότι η Μαρκοβιανή αλυσίδα ξεχνάει την αρχική της κατάσταση (βλέπε Πάττα (2020)). Συνεπώς, δεν υπάρχει υπολογιστικό κόστος που να οφείλεται στην αρχικοποίηση του αλγορίθμου. Μια τέτοια διαδικασία έχει σημαντικό υπολογιστικό κόστος καθώς έχει ως αποτέλεσμα ο αλγόριθμος MCMC-ABC των Marjoram et al. (2003) να χρειάζεται περισσότερο χρόνο ώστε να επιτευχθεί η σύγκλιση της αλυσίδας.*

### 4.5.2 Regression ABC

Η πιο διαδεδομένη τροποποίηση του αλγορίθμου ABC που στοχεύει στην αύξηση της υπολογιστικής αποδοτικότητας του αλγορίθμου είναι γνωστή ως Regression ABC, η οποία ονομάζεται έτσι καθώς όπως θα δούμε εμπεριέχει την προσαρμογή ενός μοντέλου παλινδρόμησης. Η τροποποίηση αυτή προτάθηκε από τους Beaumont et al. (2002), οι οποίοι παρατήρησαν ότι η διαδικασία που είχε προταθεί από τους Pritchard et al. (1999) μπορεί να θεωρηθεί ως μια μέθοδος εκτίμησης της δεσμευμένης πυκνότητας. Ειδικότερα, οι Beaumont et al. (2002) πρότειναν μια μέθοδο που βασίζεται στην τοπική γραμμική παλινδρόμηση (local linear regression) και, επιπρόσθετα, επέδειξαν ότι σε αυτήν την κατηγορία μεθόδων μπορεί να χρησιμοποιηθεί ένα πολύ μεγαλύτερο επίπεδο ανεκτικότητας $\epsilon$ (εφόσον στην πορεία χρησιμοποιούνται διάφοροι μετασχηματισμοί), που σημαίνει ότι θα προκύψει και ένα μεγαλύτερο ποσοστό αποδεκτών προσομοιωμένων σημείων, σε έναν κλασικό αλγόριθμο απόρριψης. Στη συνέχεια θα παρουσιαστεί εν συντομία ο αλγόριθμος που προκύπτει. Για περισσότερες πληροφορίες παραπέμπουμε στους Beaumont et al. (2002), Marin et al. (2012), Gutmann (2016), Blum (2018) και Frazier et al. (2020), ενώ για μια επισκόπηση των τεχνικών παλινδρόμησης στον ABC ενδεικτικά παραπέμπουμε στον Blum (2018).

**Αλγόριθμος Regression ABC των Beaumont et al. (2002)**:

1. Δοθείσης μιας παρατήρησης $y$, προσομοιώνουμε την τιμή $\theta_i$ από την εκ των προτέρων κατανομή και προσομοιώνουμε μια παρατήρηση $x_i$ από το μοντέλο $f(\cdot|\theta_i)$, μέχρι να παραχθούν $M$ το πλήθος σημεία.

2. Υπολογίζουμε τη δειγματική τυπική απόκλιση, έστω $k_j$ των $S_j(x)$, όπου $S_j$ είναι το $j$-οστό περιγραφικό στατιστικό μέτρο, για $j = 1, ..., s$.

3. Ορίζουμε το μέτρο της απόκλισης $d(S(x), S(y))$ να δίνεται από τη σχέση:

$$d(S(x), S(y)) = \sqrt{\sum_{j=1}^{s} \{(S_j(x) - S_j(y))/k_j\}^2}$$

4. Επιλέγουμε τη σταθερά ανεκτικότητας $\epsilon$ έτσι ώστε το ποσοστό των αποδεκτών τιμών να είναι $P_\epsilon = N/M$.

5. Σταθμίζουμε τα προσομοιωμένα σημεία $S(x_i)$ χρησιμοποιώντας τη συνάρτηση $K_\delta(d(S(x_i), S(y)))$, όπου:

$$K_\delta(t) = \begin{cases} \delta^{-1}(1 - (t/\delta)^2), & \text{αν } t \leq \delta \\ 0, & \text{αν } t > \delta, \end{cases}$$

   με $\delta$ να παριστά το εύρος ζώνης (bandwidth) της συνάρτησης πυρήνα. Δηλαδή σταθμίζουμε τα προσομοιωμένα σημεία χρησιμοποιώντας τη συνάρτηση Epanechnikov. Στο σημείο αυτό επισημαίνεται ότι η συνάρτηση Epanechnikov δεν είναι η μοναδική επιλογή για τη μορφή του πυρήνα, δηλαδή θα μπορούσε να χρησιμοποιηθεί και ένας Gaussian πυρήνας.

6. Εφαρμόζουμε σταθμισμένη γραμμική παλινδρόμηση στα σημεία εκείνα που έχουν μη μηδενικές σταθμίσεις (βάρη) για να προσδιορίσουμε έναν εκτιμητή $\hat{E}(\theta|S(x))$.

7. Υπολογίζουμε $\theta_i^* = \theta_i - \hat{E}(\theta|S(x_i)) + \hat{E}(\theta|S(y))$.

8. Οι τιμές $\theta_i^*$ είναι ένα τυχαίο δείγμα από μια προσέγγιση της εκ των υστέρων κατανομής $\pi(\theta|y)$.

Παρατηρείστε ότι στο Βήμα 6 εφαρμόζουμε μια σταθμισμένη γραμμική παλινδρόμηση (weighted linear regression) για τα $\theta_i$ με τα βάρη να προέρχονται από έναν Epanechnikov πυρήνα (βλέπε Βήμα 5). Ας σημειωθεί επίσης ότι όταν λέμε σταθμισμένη γραμμική παλινδρόμηση πρόκειται για μια προσαρμογή της κλασικής γραμμικής παλινδρόμησης με τη διαφορά ότι οι εκτιμητές ελαχίστων τετραγώνων προκύπτουν ελαχιστοποιώντας το άθροισμα του γινομένου των τετραγώνων των σφαλμάτων με τα αντίστοιχα βάρη.

**Παρατήρηση 4.5.2.** *Αξίζει να επισημανθεί ότι με βάση την παλινδρόμηση που χρησιμοποιείται, η κλάση της Regression ABC μεθόδου χωρίζεται σε τρεις βασικές κατηγορίες, οι οποίες είναι η τοπική γραμμική παλινδρόμηση (local linear regression), η μη γραμμική παλινδρόμηση (nonlinear regression) και η αντίστροφη παλινδρόμηση (inverse regression). Για περισσότερες λεπτομέρειες παραπέμπουμε στους Blum and François (2010) και Marin et al. (2012).*

### 4.5.3 Άλλες τροποποιήσεις

Ένα βασικό μειονέκτημα όλων των MCMC αλγορίθμων είναι ότι μπορεί να μην τερματίσουν ή να απαιτηθεί μεγάλος υπολογιστικός χρόνος. Όπως επισημαίνεται από τους Turner and Van Zandt (2012) το παραπάνω πρόβλημα είναι ιδιαίτερα έντονο με τον MCMC-ABC καθώς δεν απαιτείται μόνο η πιθανότητα αποδοχής αλλά να δημιουργηθούν και προσομοιωμένα δεδομένα που είναι αρκετά κοντά στα παρατηρούμενα. Αυτό έχει ως συνέπεια το ποσοστό απόρριψης αυτού του αλγορίθμου να είναι εξαιρετικά υψηλό, απαιτώντας πολλές επαναλήψεις ακόμα και για σχετικά απλά προβλήματα. Για την αντιμετώπιση του παραπάνω προβλήματος και στοχεύοντας στη βελτίωση της αποτελεσματικότητας της κατανομής πρότασης, έχουν προταθεί μέθοδοι που είναι εμπνευσμένες από τις ακολουθιακές Monte Carlo μεθόδους (Sequential Monte Carlo, SMC, βλέπε Liu (2008)).

Η βασική διαφοροποίηση των ακολουθιακών Monte Carlo μεθόδων από την κλασική MCMC-ABC μέθοδο έγκειται στο γεγονός ότι αντί να προσομοιώνονται υποψήφια $\theta^*$ ένα κάθε φορά, χρησιμοποιούν, ταυτόχρονα, ένα μεγάλο σύνολο που αποτελείται από τυχαία επιλεγμένες παραμετρικές τιμές που ονομάζονται particles. Στη συνέχεια θα αποδοθεί ο όρος μέλος για την έννοια particle. Τα μέλη διαταράσσονται (perturbed) και φιλτράρονται (filtered) σε κάθε στάδιο του αλγορίθμου. Κατά αυτόν τρόπο η συλλογή των τιμών έρχεται ολοένα και εγγύτερα σε ένα δείγμα από την επιθυμητή εκ των υστέρων.

Ειδικότερα, όπως περιγράφουν Turner and Van Zandt (2012), οι μέθοδοι που ανήκουν σε αυτήν την κατηγορία ξεκινούν δημιουργώντας μια συλλογή, ένα σύνολο από $N$ το πλήθος υποψήφιες τιμές για την παράμετρο $\theta$. Συνηθέστερα η συλλογή αυτή δημιουργείται προσομοιώνοντας τιμές από την εκ των προτέρων κατανομή $\pi(\theta)$. Έπειτα, με διαδοχικές (ακολουθιακές) επαναλήψεις, μέλη επιλέγονται τυχαία από αυτή τη συλλογή και η πιθανότητα να επιλεγεί οποιοδήποτε μέλος εξαρτάται από το βάρος που έχει δοθεί σε κάθε ένα από αυτά. Για την πρώτη επανάληψη η πιθανότητα επιλογής οποιουδήποτε μέλους είναι $1/N$, δηλαδή είναι το ίδιο πιθανό οποιοδήποτε μέλος να επιλεγεί. Στη συνέχεια οι διάφορες μέθοδοι που ανήκουν σε αυτήν την κατηγορία μεθόδων διαφοροποιούνται ανάλογα με τον τρόπο ανάθεσης βαρών στις επόμενες επαναλήψεις σε κάθε μέλος τιμών της συλλογής.

Για τη διαδικασία της διαταραχής, σε κάθε στάδιο της μεθόδου, και του φιλτραρίσματος των μελών που αναφέρθηκε προηγουμένως είναι απαραίτητη η επιλογή ενός πυρήνα μετάβασης (transition kernel). Σε αυτήν την κατηγορία μεθόδων κάποιος θα μπορούσε να ισχυριστεί ότι ο πυρήνας μετάβασης παίζει τον ίδιο ρόλο με την αντίστοιχη κατανομή πρότασης στις τεχνικές MCMC, οι οποίες μελετήθη-

καν στο Κεφάλαιο 3 της διατριβής, και φυσικά έχει και τον ίδιο σκοπό. Για τον καθορισμό του πυρήνα μετάβασης απαιτείται η επιλογή της κατανομής μιας τυχαίας μεταβλητής, έστω $\eta$. Πιο συγκεκριμένα, όπως αναφέρουν και οι Turner and Van Zandt (2012), η τυχαία μεταβλητή $\eta$ θα προστεθεί σε κάθε τυχαία επιλεγμένη παραμετρική τιμή του συνόλου, δηλαδή σε κάθε μέλος, με την ιδιότητα να μπορεί να μετακινεί το αντίστοιχο μέλος στον παραμετρικό χώρο.

Τα παραπάνω θα γίνουν πιο κατανοητά μέσω ενός απλού παραδείγματος. Έστω λοιπόν $\theta^*$ το μέλος που προσομοιώνεται από τη συλλογή των υποψήφιων τιμών. Υπό την υπόθεση ότι το $\theta^*$ διαταράσσεται προσθέτοντας σε αυτό μια Gaussian απόκλιση $\eta$ με μέση τιμή 0 και διακύμανση $\sigma^2$, δηλαδή $\eta \sim N(0, \sigma^2)$, έπεται ότι η νέα υποψήφια τιμή για την παράμετρο $\theta$ θα είναι η $\theta^{**} = \theta^* + \eta$. Σε αυτήν την περίπτωση ο πυρήνας μετάβασης περιγράφει την κατανομή της νέας προτεινόμενης τιμής $\theta^{**}$ δοθέντος του μέλους $\theta^*$ που είναι μια κανονική κατανομή με μέση $\theta^*$ και διακύμανση $\sigma^2$. Συμπερασματικά $\theta^{**}|\theta^* \sim N(\theta^*, \sigma^2)$.

Οι αλγόριθμοι που εφαρμόζονται διαφοροποιούνται βάσει του πυρήνα μετάβασης που αξιοποιείται κάθε φορά αλλά και του τρόπου υπολογισμού των βαρών για τον έλεγχο της προσομοίωσης των μελών από το σύνολο, τη συλλογή των υποψήφιων τιμών για την παράμετρο. Στο εδάφιο αυτό θα μελετηθούν οι τρεις βασικότερες κατηγορίες της ακολουθιακής μεθόδου Sequential Monte Carlo (SMC) προσαρμοσμένες για τον αλγόριθμο ABC, οι οποίες είναι ο Μερικός έλεγχος απόρριψης (Partial rejection control), η Πληθυσμιακή Monte Carlo δειγματοληψία (Population Monte Carlo sampling) και η Ακολουθιακή Monte Carlo δειγματοληψία (Sequential Monte Carlo sampling).

### Μερικός έλεγχος απόρριψης (Partial rejection control)

Οι Sisson et al. (2007), ανέπτυξαν τον λεγόμενο ABC Μερικό έλεγχο απόρριψης ABC Partial rejection control (ABC PRC) αλγόριθμο, ο οποίος αποτελεί τον πρώτο ABC αλγόριθμο που εφαρμόζει το Φιλτράρισμα μελών.

Ο αλγόριθμος αυτός αρχικά απαιτεί την επιλογή ενός προς τα εμπρός και ενός προς τα πίσω πυρήνα μετάβασης. Για λόγους απλότητας ο προς τα εμπρός πυρήνας δηλώνεται ως συνάρτηση πυκνότητας $q_f(\cdot|\theta^*)$ και με παρόμοιο τρόπο ο προς τα πίσω πυρήνας ως $q_b(\cdot|\theta^{**})$. Στη συνέχεια, χρησιμοποιείται ο $q_f(\cdot|\theta^*)$ για τη διαταραχή του μέλους $\theta^*$ σε $\theta^{**}$ και αυτό έχει ως αποτέλεσμα να προσομοιώνονται τα δεδομένα, αξιοποιώντας το $\theta^{**}$, και να συγκρίνονται με τα πραγματικά υπολογίζοντας, όπως έχει ήδη περιγραφεί, το μέτρο εγγύτητας $d(x, x^*)$. Επομένως, αν για κάποιο $\epsilon_0$ έχουμε $d(x, x^*) < \epsilon_0$ τότε κρατάμε το $\theta^{**}$ μαζί με ένα βάρος που θα καθορίσει την πιθανότητα να προσομοιωθεί σε επόμενες επαναλήψεις, ενώ αν $d(x, x^*) > \epsilon_0$ τότε απορρίπτεται το $\theta^{**}$ και η διαδικασία επαναλαμβάνεται έως

ότου λάβουμε ένα μέλος από τη συλλογή που γίνεται δεκτό. Το βάρος που δίνεται στο νέο μέλος $\theta^{**}$ συμβολίζεται με $w$ και ορίζεται ως:

$$w = \frac{\pi(\theta^{**})\, q_b(\theta^*|\theta^{**})}{\pi(\theta^*)\, q_f(\theta^{**}|\theta^*)}.$$

Αυτή η διαδικασία επαναλαμβάνεται έως ότου το σύνολο να αποτελείται από $N$ νέα μέλη, που το καθένα πληροί την ιδιότητα ότι $d(x, x^*) < \epsilon_0$. Αν σταματήσουμε τώρα, αφού ξαναδημιουργήσουμε μία φορά το σύνολο, τότε ο αλγόριθμος ABC PRC είναι ισοδύναμος με τον ABC αλγόριθμο αποδοχής-απόρριψης. Ωστόσο, θα επαναλάβουμε τη διαδικασία αυτή αρκετές φορές. Σε κάθε επανάληψη προσομοιώνονται μέλη με πιθανότητες που βασίζονται στα βάρη που έχουν καθοριστεί στην προηγούμενη επανάληψη. Επιπλέον, τα βάρη αυτά επιτρέπουν την απόρριψη μελών από το σύνολο σε περιοχές χαμηλής πιθανότητας και συνεπώς αυξάνουν τον αριθμό των μελών σε περιοχές υψηλής πιθανότητας, με αποτέλεσμα τελικά να έχουμε ένα δείγμα μελών που αντιπροσωπεύουν ένα δείγμα από την επιθυμητή εκτίμηση της εκ των υστέρων κατανομής $\pi(\theta|d(x, x^*) < \epsilon_0)$.

Κλείνοντας, σημειώνεται ότι το παραπάνω σχήμα στάθμισης επιλύει πολλά από τα προβλήματα του αλγορίθμου MCMC-ABC, συμπεριλαμβανομένου του προβλήματος μιας αλυσίδας να "κολλήσει" σε μια περιοχή χαμηλής πιθανότητας. Από την άλλη πλευρά, η αποτελεσματικότητα του δείγματος στηρίζεται σε μεγάλο βαθμό στις επιλογές των δύο πυρήνων $q_b(\theta^*|\theta^{**})$ και $q_f(\theta^{**}|\theta^*)$ όπως και στην επιλογή της εκ των προτέρων κατανομής $\pi(\theta)$. Για παράδειγμα, θεωρείστε μια κατάσταση με μη πληροφορικές εκ των προτέρων κατανομές, δηλαδή η $\pi(\theta)$ παραμένει σταθερή, και με ίσους πυρήνες, $q_f = q_b$. Σε μία τέτοια περίπτωση, όπως αναφέρουν και οι Turner and Van Zandt (2012), τα βάρη που δίνονται στα μέλη ουσιαστικά δεν αλλάζουν ποτέ, και έτσι ο αλγόριθμος μετατρέπεται σε ABC αποδοχής-απόρριψης. Επιπλέον, η ABC PRC παράγει μεροληπτικές εκτιμήσεις της εκ των υστέρων κατανομής. Για περισσότερες λεπτομέρειες βλέπε Beaumont et al. (2009) και τις εκεί αναφορές. Τελικά, σε αυτήν την περίπτωση, η κατανομή που ορίζεται από το σύνολο των μελών και τα βάρη τους δε συγκλίνει στην πραγματική εκ των υστέρων κατανομή. Η παρατήρηση αυτή ήταν το κίνητρο των Beaumont et al. (2009) που τους ώθησε να εφαρμόσουν μια Πληθυσμιακή Monte Carlo δειγματοληψία, η οποία θα παρουσιαστεί ακολούθως, για να διορθώσουν το ζήτημα με τις μεροληπτικές εκτιμήσεις.

### Πληθυσμιακή Monte Carlo δειγματοληψία (Population Monte Carlo sampling, PMC)

Η τεχνική δειγματοληψίας ABC PMC διαφοροποιείται από την ABC PRC, η οποία αναπτύχθηκε προηγουμένως, κυρίως ως προς τον πυρήνα μετάβασης αλλά και τον υπολογισμό των βαρών. Ενώ λοιπόν ο αλγόριθμος ABC PRC έχει

ως απαραίτητη προϋπόθεση τον καθορισμό του προς τα εμπρός και του προς τα πίσω πυρήνα μετάβασης, ο αλγόριθμος ABC PMC χρησιμοποιεί έναν μόνο προσαρμοστικό πυρήνα μετάβασης $q(\cdot|\theta^*)$. Σ' αυτήν την κατηγορία μεθόδων, δίνεται ιδιαίτερη προσοχή στο γεγονός ότι ο προσαρμοστικός πυρήνας εξαρτάται από τη διακύμανση των αποδεκτών μελών στην προηγούμενη επανάληψη. Πιο συγκεκριμένα, δοσμένου του βάρους $w_{i,t-1}$ για το μέλος $\theta_{i,t-1}$ στην επανάληψη $t-1$, το νέο βάρος $w_{i,t}$ για το μέλος $\theta_{i,t}$ στην επανάληψη $t$ υπολογίζεται ως:

$$w_{i,t} = \frac{\pi(\theta_{i,t})}{\sum_{j=1}^{N} w_{j,t-1}\, q(\theta_{j,t-1}|\theta_{i,t}\, \sigma_{t-1})},$$

όπου $q(\cdot|\theta_{i,t}, \sigma_{t-1})$ είναι ένας Gaussian πυρήνας με μέση τιμή $\theta_{i,t}$ και τυπική απόκλιση $\sigma_{t-1}$. Ακολουθώντας τους Turner and Van Zandt (2012), η διακύμανση δίνεται ως:

$$\sigma_t^2 = 2\,\frac{1}{N}\sum_{i=1}^{N}\left\{\theta_{i,t} - \sum_{j=1}^{N}\theta_{j,t}/N\right\}^2 = 2\,Var(\theta_{1:N,t}).$$

Στο σημείο αυτό τονίζεται ότι μία σημαντική δυσκολία με πολλά σχήματα δειγματοληψίας αποτελεί η ταχύτητα με την οποία μπορούν να ληφθούν εκ των υστέρων εκτιμήσεις. Επιπλέον, η ταχύτητα αυτή καθορίζεται από το ποσοστό αποδοχής των μελών του συνόλου, της συλλογής, ή από την πιθανότητα αποδοχής μιας κατανομής πρότασης. Σημειώστε ότι πολύ χαμηλά ποσοστά αποδοχής, τα οποία προκύπτουν από μια κακή επιλογή της κατανομής πρότασης ή αντίστοιχα από μια κακή επιλογή για τους πυρήνες μετάβασης, έχουν ως αποτέλεσμα να σπαταληθεί μια πολύ μεγάλη ποσότητα υπολογισμού κατά τον προσδιορισμό των κατανομών πρότασης που δεν έχουν καμία πιθανότητα να επιλεγούν. Συμπερασματικά, η αξία των μεθόδων στάθμισης ABC PMC είναι ότι βελτιστοποιεί την πιθανότητα αποδοχής.

Τέλος, θα παρουσιαστεί η τρίτη μέθοδος της κλάσης των Ακουλουθιακών Monte Carlo αλγορίθμων, η αποκαλούμενη Ακολουθιακή Monte Carlo δειγματοληψία (Sequential Monte Carlo sampling), η οποία είναι εξίσου σημαντική με τις προηγούμενες.

### Ακολουθιακή Monte Carlo δειγματοληψία (Sequential Monte Carlo sampling, SMC)

Οι Sisson et al. (2007) πρότειναν μια μέθοδο για την επαναληπτική βελτίωση της προσέγγισης ABC. Η προσέγγισή τους αποτελούνταν από δύο κύρια χαρακτηριστικά: τη σταθμισμένη δειγματοληψία από το σύνολο των σημείων που

έχουν ήδη παραχθεί και τη διαδοχική μείωση του επιπέδου ανεκτικότητας $\epsilon$. Στον αρχικό τους αλγόριθμο έχει σημειωθεί μια μεροληψία, αλλά διορθώθηκε μετέπειτα από τους Beaumont et al. (2009) και Toni et al. (2009). Οι Toni et al. (2009), παρακινούμενοι από τους Del Moral et al. (2006) που εφάρμοσαν έναν Ακολουθιακό αλγόριθμο δειγματοληψίας σπουδαιότητας (Sequential importance sampling algorithm), ανέπτυξαν την τεχνική ABC SMC. Αναλυτικότερα, οι συναρτήσεις βάρους στην ABC SMC είναι πολύ παρόμοιες με τα αντίστοιχα βάρη στη μέθοδο ABC PMC που επεξηγήθηκε παραπάνω. Η διαφορά μεταξύ των βαρών στηρίζεται στην παρατήρηση ότι ο πυρήνας $q(\cdot|\theta^*)$ εδώ είναι μη προσαρμοστικός (nonadaptive) και επίσης δεν απαιτείται να είναι Gaussian. Επομένως, λαμβάνοντας υπόψη όσα προαναφέρθηκαν, τα βάρη που αντιστοιχούν στο $i$-οστό μέλος στην επανάληψη $t$ του αλγορίθμου ABC SMC ορίζονται ως:

$$w_{i,t} = \frac{\pi(\theta_{i,t})}{\sum_{j=1}^{N} w_{j,t-1}\, q(\theta_{j,t-1}|\theta_{i,t})}$$

Όπως επισημαίνουν και οι Turner and Van Zandt (2012), ο αλγόριθμος ABC SMC είναι ιδιαίτερα χρήσιμος στις περιπτώσεις που ο πυρήνας μετάβασης στη μέθοδο ABC PMC δεν μπορεί να είναι Gaussian. Κάτι τέτοιο μπορεί να συμβεί για συγκεκριμένα μοντέλα στα οποία η παράμετρος $\theta$ δε λαμβάνει αρνητικές τιμές, όπως, για παράδειγμα, η παράμετρος $p$ (πιθανότητα επιτυχίας) στη διωνυμική κατανομή.

# ΚΕΦΑΛΑΙΟ 5

# Επιλογη μοντελου

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν οι κυριότερες Προσεγγιστικές Μπεϋζιανές Υπολογιστικές μέθοδοι στο πλαίσιο της Μπεϋζιανής εκτιμητικής. Ειδικότερα, υποθέτοντας μια εκ των προτέρων κατανομή για την παράμετρο $\theta$ και ένα πιθανοθεωρητικό μοντέλο $f(x|\theta)$ για τη μοντελοποίηση των διαθέσιμων δεδομένων δοθέντος της παραμέτρου $\theta$, ο στόχος ήταν η εξαγωγή συμπερασμάτων για την παράμετρο $\theta$ μέσω της προσέγγισης της εκ των υστέρων κατανομής $\theta|x$. Σε πρακτικές εφαρμογές υπάρχουν συνήθως αρκετά διαφορετικά πιθανοθεωρητικά, στοχαστικά μοντέλα που μπορεί να είναι κατάλληλα για την περιγραφή του πληθυσμού από τον οποίο προέρχεται το δείγμα των παρατηρήσεων. Αυτό έχει ως συνέπεια το πρόβλημα της σωστής επιλογής του μοντέλου ανάμεσα σε μια λίστα από πιθανά για ένα συγκεκριμένο σύνολο δεδομένων να είναι ένα σημαντικό ζήτημα. Για το λόγο αυτό έχει αναπτυχθεί στη στατιστική βιβλιογραφία μια πληθώρα στατιστικών μεθόδων με τα συγγράμματα των Linhart and Zucchini (1986), Burnham and Anderson (2002) και Claeskens and Hjort (2008) να αποτελούν μονογραφίες αναφοράς. Αντικείμενο μελέτης αυτού του κεφαλαίου αποτελεί η επιλογή μοντέλου χρησιμοποιώντας Προσεγγιστικές Μπεϋζιανές Υπολογιστικές (ABC) μεθόδους. Καθώς οι ABC μέθοδοι αποτελούν μια Μπεϋζιανή διαδικασία, αρχικά παρουσιάζονται εν συντομία οι βασικές Μπεϋζιανές μέθοδοι επιλογής μοντέλου και έπεται η παρουσίαση αυτών με τη χρήση της ABC μεθόδου.

## 5.1 Μπεϋζιανές μέθοδοι

Ας θεωρήσουμε αρχικά ότι η συλλογή των πιθανών μοντέλων αποτελείται από δύο στοιχεία, δηλαδή υποθέτουμε ότι έχουμε μόνο δύο πιθανά, υποψήφια μοντέλα, έστω $M_0$ και $M_1$. Επιπλέον θεωρούμε ότι $\pi(M_0)$ και $\pi(M_1)$ είναι οι εκ των προτέρων πιθανότητες κάθε μοντέλου. Επιπρόσθετα, αν $x = (x_1, ..., x_n)$ είναι τα διαθέσιμα δεδομένα αυτά έχουν, δοθέντος του μοντέλου $M_i$ και της παραμέτρου

αυτού $\theta_i$, από κοινού κατανομή $f(x|\theta_i, M_i)$, για $i = 0, 1$, όπου $\theta_i$, $i = 0, 1$, είναι οι παράμετροι υπό τα δύο διαφορετικά μοντέλα διάστασης $p_i$, $i = 0, 1$. Στο παραπάνω πλαίσιο, αν $M$ είναι το αληθινό άγνωστο μοντέλο θέλουμε ουσιαστικά να ελέγξουμε τη μηδενική υπόθεση $H_0 : M = M_0$, δηλαδή την υπόθεση $X \sim M_0$ με πιθανοφάνεια $f(x|\theta_0, M_0)$ και εκ των προτέρων $\pi(\theta_0)$, έναντι της εναλλακτικής $H_1 : M = M_1$, δηλαδή της υπόθεσης $X \sim M_1$ με πιθανοφάνεια $f(x|\theta_1, M_1)$ και εκ των προτέρων $\pi(\theta_1)$. Σε όσα ακολουθούν για λόγους συντομίας και αν δεν κρίνεται απαραίτητο θα γράφουμε $f(x|M_i)$ αντί για $f(x|\theta_i, M_i)$. Υπό τη Μπεϋζιανή προσέγγιση είναι γνωστό ότι η εξαγωγή συμπερασμάτων βασίζεται στην εκ των υστέρων κατανομή. Ειδικότερα, προκύπτει ότι αν $\pi(M_1|x) > \pi(M_0|x)$ ή ισοδύναμα αν:

$$\frac{\pi(M_1|x)}{\pi(M_0|x)} > 1$$

τότε έχουμε λόγους να απορρίψουμε το μοντέλο $M_0$. Όμως:

$$\pi(M_1|x) = \frac{f(x|M_1)\pi(M_1)}{f(x|M_1)\pi(M_1) + f(x|M_0)\pi(M_0)}$$

και

$$\pi(M_0|x) = \frac{f(x|M_0)\pi(M_0)}{f(x|M_1)\pi(M_1) + f(x|M_0)\pi(M_0)}.$$

Εύκολα προκύπτει ότι:

$$PO_{10} = \frac{\pi(M_1|x)}{\pi(M_0|x)} = \frac{\pi(M_1)}{\pi(M_0)} \frac{f(x|M_1)}{f(x|M_0)}. \tag{5.1}$$

Το παραπάνω πηλίκο είναι γνωστό ως εκ των υστέρων λόγος μοντέλων (posterior models odds) και ουσιαστικά μας οδηγεί στο κριτήριο να απορρίπτουμε τη μηδενική υπόθεση αν το πηλίκο αυτό είναι μεγαλύτερο από 1, ενώ διαφορετικά να την αποδεχόμαστε.

Παρατηρούμε ότι στο δεξί μέλος της σχέσης (5.1) σχηματίζεται το γινόμενο δύο πηλίκων. Το πρώτο πηλίκο $\frac{\pi(M_1)}{\pi(M_0)}$ είναι το πηλίκο των εκ των προτέρων κατανομών των δύο μοντέλων, ενώ το δεύτερο πηλίκο $\frac{f(x|M_1)}{f(x|M_0)}$ είναι το πηλίκο των περιθωρίων πιθανοφανειών των δύο μοντέλων, με την ποσότητα $f(x|M_i)$, $i = 0, 1$, να εκφράζει την πιθανότητα ή την πυκνότητα πιθανότητας σε περίπτωση που έχουμε συνεχή δεδομένα να παρατηρήσουμε τα πραγματικά δεδομένα πριν τη συλλογή τους υπό την υιοθέτηση ότι το μοντέλο $M_i$ είναι το αληθινό μοντέλο.

Το πηλίκο $\dfrac{f(x|M_1)}{f(x|M_0)}$ είναι γνωστό ως **παράγοντας Bayes** και ορίζεται:

$$BF_{10} = \frac{f(x|M_1)}{f(x|M_0)}, \tag{5.2}$$

με την πρώτη αναφορά σε αυτό να γίνεται από τον Jeffreys (1939). Το πηλίκο της σχέσης (5.2) στην ουσία μας παρέχει ένα κριτήριο για τη σύγκριση μοντέλων (βλέπε, μεταξύ άλλων, Berger (1985), Robert (2001) και Marin and Robert (2010)). Ειδικότερα, σύμφωνα με τους Kass and Raftery (1995), έχουμε την ακόλουθη ερμηνεία των τιμών του παράγοντα Bayes.

Πίνακας 5.1: Ερμηνεία των τιμών του παράγοντα Bayes

| $2\,log(BF_{10})$ | $BF_{10}$ | Ένδειξη κατά της $H_0$ |
|---|---|---|
| 0-2 | 1-3 | Πολύ αδύναμη |
| 2-6 | 3-20 | Θετική |
| 6-10 | 20-150 | Ισχυρή |
| $> 10$ | $> 150$ | Πολύ ισχυρή |

Τα παραπάνω είναι πλήρως αιτιολογημένα καθώς όταν η ποσότητα $BF_{10}$ παίρνει μεγάλες τιμές, τότε και το πηλίκο $\dfrac{f(M_1 \mid x)}{f(M_2 \mid x)}$ παίρνει τιμές μεγαλύτερες της μονάδας και επομένως υπάρχει πολύ ισχυρή ένδειξη για απόρριψη της μηδενικής υπόθεσης.

Από όσα προηγήθηκαν γίνεται αντιληπτό ότι όταν η σύγκριση των δύο μοντέλων κατά τη Μπεϋζιανή προσέγγιση βασίζεται στην εκ των υστέρων πιθανότητά τους (επιλέγοντας ως καλύτερο εκείνο με τη μεγαλύτερη εκ των υστέρων πιθανότητα) απαιτείται ο καθορισμός των εκ των προτέρων κατανομών των δύο μοντέλων ($\pi(M_1)$ και $\pi(M_0)$) και ο υπολογισμός των περιθωρίων πιθανοφανειών των δύο μοντέλων. Από την άλλη μεριά όταν η σύγκριση των δύο μοντέλων γίνεται ερμηνεύοντας τις τιμές του παράγοντα Bayes απαιτείται μόνο ο υπολογισμός των περιθωρίων πιθανοφανειών των δύο μοντέλων. Παρατηρείστε ότι όταν έχουμε ίσες εκ των προτέρων $\pi(M_0) = \pi(M_1)$, δηλαδή όταν δεν έχουμε κάποια υπόνοια ότι ένα από τα δύο μοντέλα είναι προτιμότερο, τα δύο κριτήρια ταυτίζονται. Όσον αφορά τον καθορισμό των εκ των προτέρων κατανομών των δύο μοντέλων, είναι φυσικό ότι θα πρέπει να αντιπροσωπεύουν τις πεποιθήσεις μας για ποιο μοντέλο πρέπει να χρησιμοποιηθεί προτού μελετηθούν τα δεδομένα μας και με αυτήν την έννοια μπορεί να οδηγήσει σε καθαρά υποκειμενική ανάλυση.

Από την άλλη μεριά αν επιλεγούν ίσες εκ των προτέρων κατανομές ή βασιστούμε μόνο στον παράγοντα Bayes, τότε έχουμε να αντιμετωπίσουμε μόνο το

πρόβλημα του υπολογισμού των περιθωρίων πιθανοφανειών των δύο μοντέλων $f(x|M_i)$, $i = 0, 1$, μέσω της σχέσης:

$$f(x|M_i) = \int f(x|\theta_i, M_i)\, \pi(\theta_i|M_i)\, d\theta_i, \tag{5.3}$$

όπου $\theta_i$ είναι το διάνυσμα των άγνωστων παραμέτρων υπό το $M_i$ μοντέλο, ενώ $f(x|\theta_i, M_i)$ είναι η πιθανοφάνεια του μοντέλου και $\pi(\theta_i|M_i)$ η εκ των προτέρων κατανομή των άγνωστων παραμέτρων δεδομένου του μοντέλου. Το παραπάνω ολοκλήρωμα μπορεί να υπολογιστεί αναλυτικά όταν χρησιμοποιούνται συζυγείς εκ των προτέρων κατανομές και ταυτόχρονα η πιθανοφάνεια $f(x|\theta_i, M_i)$ δεν είναι δύσχρηστη, ενώ σε μια πληθώρα περιπτώσεων ο υπολογισμός της καθίσταται δύσκολος ή και ανέφικτος. Κατά συνέπεια, για τον υπολογισμό αυτόν, δημιουργείται η ανάγκη να αναπτυχθούν διάφορες μεθοδολογίες, όπως είναι για παράδειγμα η δειγματοληψία σπουδαιότητας, οι Laplace προσεγγίσεις και οι MCMC μέθοδοι που επιτυγχάνουν, προσομοιώνοντας τιμές, την κατασκευή μιας Μαρκοβιανής αλυσίδας που συγκλίνει στην εκ των υστέρων κατανομή. Στην παρούσα διατριβή δε θα αναλυθούν αυτές οι μεθοδολογίες και παραπέμπουμε, μεταξύ άλλων, στους Kass and Raftery (1995) και Dellaportas et al. (2002) και τις εκεί αναφορές, καθώς στην επόμενη ενότητα αντικείμενο μελέτης είναι ο τρόπος που μπορεί να χρησιμοποιηθεί ο αλγόριθμος ABC για τον υπολογισμό των εκ των υστέρων πιθανοτήτων των υποψήφιων μοντέλων.

Πριν κλείσουμε αυτήν την ενότητα θα ήταν παράλειψη να μη σημειώσουμε ότι έχουν αναπτυχθεί στη βιβλιογραφία μεθοδολογίες για την προσέγγιση του παράγοντα Bayes κρατώντας την ίδια εκ των προτέρων κατανομή, έτσι ώστε το κριτήριο που χρησιμοποιείται να μην επηρεάζεται σε κανένα από τα μοντέλα. Τέτοιες μεθοδολογίες αποτελούν για παράδειγμα τα κριτήρια BIC (Bayesian Information Criterion ή Swartz Criterion) και DIC (Deviance Information Criterion).

Το κριτήριο BIC έχει εισαχθεί στη βιβλιογραφία από τον Schwarz (1978), βασίζεται εν μέρει στη συνάρτηση πιθανοφάνειας, ενώ συνδέεται άμεσα με το πολύ γνωστό Akaike Information Criterion (AIC). Ειδικότερα, για το $i-$οστό μοντέλο δίνεται από τη σχέση:

$$\mathrm{BIC}_i = p_i\, ln(n) - 2\, ln\left(f(x|\hat{\theta}_i, M_i)\right),$$

με $\hat{\theta}_i$ να είναι η εκτιμήτρια μέγιστης πιθανοφάνειας υπό το μοντέλο $M_i$ και $p_i$ η διάσταση της παραμέτρου $\theta_i$.

Προκύπτει ότι αν ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι δύο φορές παραγωγίσιμη συνάρτηση του $\theta$ τότε (βλέπε Konishi and Kitagawa (2008)):

$$f(M_i|x) \propto f(x|M_i)\, \pi(M_i) \simeq exp(-\mathrm{BIC}_i/2)\, \pi(M_i).$$

Επομένως, το κριτήριο BIC μπορεί να χρησιμοποιηθεί ως μια εναλλακτική προσέγγιση καθώς $\mathrm{BIC}_i - \mathrm{BIC}_j$ τείνει στο $-2ln(BF_{ij})$ και προτείνει ως καλύτερο το μοντέλο εκείνο με την ελάχιστη τιμή στο κριτήριο αυτό.

**Παρατήρηση 5.1.1.** *Το κριτήριο BIC είναι ανεξάρτητο της επιλογής της εκ των προτέρων κατανομής, μπορεί να ποσοτικοποιήσει την απόδοση του μοντέλου σε όρους πρόβλεψης των δεδομένων, ενώ περιέχει έναν όρο που εκφράζει την πολυπλοκότητα του υπό θεώρηση μοντέλου. Ο όρος αυτός συχνά αναφέρεται ως όρος ποινής (penalty term) και είναι εκείνος που διαφοροποιεί διάφορα κριτήρια επιλογής μοντέλου. Ειδικότερα, το Akaike Information Criterion (AIC) προκύπτει με όρο ποινής ίσο με* $2\,p_i$ *ενώ στο BIC είναι* $p_i\,ln(n)$*, με* $n$ *το μέγεθος του δείγματος.*

Μια γενίκευση του Akaike Information Criterion (AIC), που είναι ιδιαίτερα χρήσιμη στο πλαίσιο της Μπεϋζιανής επιλογής μοντέλου, είναι το γνωστό ως Deviance Information Criterion (DIC). Το κριτήριο αυτό παρουσιάστηκε από τους Spiegelhalter et al. (2002) και ορίζεται από τη σχέση (βλέπε, επιπλέον, και Celeux et al. (2006)):

$$\mathrm{DIC}_i = -4 \cdot E_{\theta_i}\left[logf(x|\theta_i)|x\right] + 2 \cdot logf(x|\tilde{\theta}_i) \tag{5.4}$$

όπου $\tilde{\theta}_i$ είναι ένας εκτιμητής του $\theta_i$ που εξαρτάται από τα δεδομένα $x$. Συνήθως $\tilde{\theta}_i$ είναι η εκ των υστέρων μέση τιμή του $\theta_i$, ενώ η εκ των υστέρων κορυφή ή η διάμεσος μπορούν επίσης να αποτελέσουν εναλλακτικές επιλογές. Ο όρος $E_{\theta_i}\left[logf(x|\theta_i)|x\right]$ μπορεί για παράδειγμα να εκτιμηθεί ως ο μέσος όρος των τιμών του λογαρίθμου της πιθανοφάνειας. Το μοντέλο με τη μικρότερη τιμή είναι εκείνο που θα πρέπει να προτιμηθεί.

**Παρατήρηση 5.1.2.** *Σε όσα προηγήθηκαν είχαμε υποθέσει ότι έχουμε να συγκρίνουμε δύο μοντέλα* $M_0$ *και* $M_1$*. Ωστόσο, πολλές φορές τα υποψήφια μοντέλα μπορεί να είναι περισσότερα και η προηγούμενη σύγκριση θα πρέπει να επεκταθεί και σε περιπτώσεις όπου το πλήθος των υποψήφιων μοντέλων είναι* $K > 2$*. Συγκεκριμένα υποθέτουμε ότι έχουμε ως υποψήφια τα μοντέλα* $M_0$*,* $M_1$*,...,*$M_{K-1}$*. Όσα έχουν ειπωθεί για τα κριτήρια BIC και DIC συνεχίζουν να ισχύουν χωρίς καμία τροποποίηση. Όσον αφορά τον παράγοντα Bayes για να είναι εφικτή η χρησιμοποίησή του ακολουθείται η διαδικασία που θα περιγραφεί στη συνέχεια. Ορίζουμε ένα μοντέλο, έστω το* $M_0$*, ως βάση για τη σύγκριση και υπολογίζουμε ξεχωριστά τους παράγοντες Bayes:* $BF_{j0} = \dfrac{f(x|M_j)}{f(x|M_0)}$*. Η ερμηνεία στη συνέχεια διεξάγεται με παρόμοιο τρόπο με αυτόν που αναφέρθηκε έχοντας ως αναφορά τον Πίνακα 5.1 (βλέπε Kass and Raftery (1995)).*

## 5.2 ABC μέθοδοι

Στο προηγούμενο κεφάλαιο είδαμε πως οι τεχνικές ABC μπορούν να χρησιμοποιηθούν στη Μπεϋζιανή στατιστική συμπερασματολογία και ειδικότερα στο πλαίσιο της εκτιμητικής (σε σημείο ή σε διάστημα). Εκτός όμως από την εκτιμητική οι μέθοδοι ABC μπορεί να χρησιμοποιηθούν και για τον υπολογισμό των εκ των υστέρων κατανομών των διαφορετικών υποψήφιων μοντέλων (βλέπε, για παράδειγμα, Wilkinson (2007), Grelaud et al. (2009), Toni and Stumpf (2010) και Didelot et al. (2011)). Για να επιτευχθεί αυτό εκτός από τις παραμέτρους που σχετίζονται με κάθε μοντέλο εισάγεται μία επιπλέον παράμετρος που ουσιαστικά αποτελεί το δείκτη του μοντέλου. Η παράμετρος αυτή συμβολίζεται με $m$ και υποθέτοντας ότι υπάρχουν $K$ το πλήθος υποψήφια μοντέλα είναι $m \in \{1, 2, ..., K\}$. Καθώς βρισκόμαστε στο πλαίσιο της Μπεϋζιανής προσέγγισης ο δείκτης αυτός συνδέεται με τη δική του εκ των προτέρων κατανομή $\pi(\mathcal{M} = m)$, $m = 1, ..., K$, η οποία εκφράζει την πιθανότητα επιλογής του $m$-οστού μοντέλου, για $m = 1, ..., K$. Για το μοντέλο που επιλέγεται γίνεται προσομοίωση τιμών για τις άγνωστες παραμέτρους και η διαδικασία συνεχίζεται όπως στους αλγορίθμους που παρουσιάστηκαν στο προηγούμενο κεφάλαιο αυτής της διατριβής. Ειδικότερα, έχουμε τους ακόλουθους δύο αλγορίθμους ABC, που διαφοροποιούνται ουσιαστικά μόνο στο Βήμα 6, όπου καθορίζεται ο τρόπος αποδοχής ή απόρριψης μιας τιμής. Ο πρώτος εξ αυτών χρησιμοποιεί όλα τα δεδομένα (άρα δεν υπάρχει κάποια απώλεια πληροφορίας), ενώ ο δεύτερος εξ αυτών χρησιμοποιεί κάποιο μέτρο εγγύτητας μεταξύ συνοπτικών περιγραφικών μέτρων θέσης.

**Αλγόριθμος ABC επιλογής μοντέλου στη βάση όλων των δεδομένων.**

1. Υιοθετούμε μια εκ των προτέρων κατανομή, έστω $\pi_i(\theta_i)$, για κάθε παράμετρο $\theta_i$, $i = 1, ..., K$, με $\theta_i$ την παράμετρο του $i-$οστού μοντέλου.

2. Υιοθετούμε μια εκ των προτέρων κατανομή, έστω $\pi(\mathcal{M} = m)$, $m = 1, ..., K$, που εκφράζει την επιλογή κάθε υποψήφιου μοντέλου από το σύνολο των $K$ υποψήφιων.

3. Προσομοιώνουμε την τιμή $l$ από την εκ των προτέρων που καθορίστηκε στο Βήμα 2.

4. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή $\pi_l(\theta_l)$, όπου $l$ η τιμή που προσομοιώθηκε στο Βήμα 3.

5. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$, από το μοντέλο $f(\cdot|\theta^*, M_l)$, όπου $\theta^*$ είναι η τιμή της παραμέτρου του $l$-οστού μοντέλου που προσομοιώθηκε στο Βήμα 4 και $f(\cdot|\theta^*, M_l)$ η πιθανοφάνεια του $l$-οστού μοντέλου υπολογισμένη για αυτήν την τιμή.

6. Αποδεχόμαστε την τιμή $\theta^*$ αν $d(D, D^*) \leq \epsilon$, για κάποιο $\epsilon > 0$, διαφορετικά επαναλαμβάνουμε τα βήματα 3-6.

7. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $N$, μη απορριπτέων τιμών $\theta^*$, ενώ ταυτόχρονα αποθηκεύονται και οι δείκτες $l$ για τις μη απορριπτέες τιμές $\theta^*$.

**Αλγόριθμος ABC επιλογής μοντέλου στη βάση συνοπτικών περιγραφικών μέτρων.**

1. Υιοθετούμε μια εκ των προτέρων κατανομή, έστω $\pi_i(\theta_i)$, για κάθε παράμετρο $\theta_i$, $i = 1, ..., K$, με $\theta_i$ την παράμετρο του $i-$οστού μοντέλου.

2. Υιοθετούμε μια εκ των προτέρων κατανομή, έστω $\pi(\mathcal{M} = m)$, $m = 1, ..., K$, που εκφράζει την επιλογή κάθε υποψήφιου μοντέλου από το σύνολο των $K$ υποψήφιων.

3. Προσομοιώνουμε την τιμή $l$ από την εκ των προτέρων που καθορίστηκε στο Βήμα 2.

4. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή $\pi_l(\theta_l)$, όπου $l$ η τιμή που προσομοιώθηκε στο Βήμα 3.

5. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$ από το μοντέλο $f(\cdot|\theta^*, M_l)$, όπου $\theta^*$ είναι η τιμή της παραμέτρου του $l$-οστού μοντέλου που προσομοιώθηκε στο Βήμα 4 και $f(\cdot|\theta^*, M_l)$ η πιθανοφάνεια του $l$-οστού μοντέλου υπολογισμένη για αυτήν την τιμή.

6. Υπολογίζουμε ένα μέτρο της απόκλισης ή της εγγύτητας (ή μια μετρική) μεταξύ των τιμών ενός περιγραφικού συνοπτικού στατιστικού, έστω $s$, υπολογισμένο στο πραγματικό δείγμα $D = (x_i)$, $i = 1, ..., n$, και των τιμών αυτού του περιγραφικού μέτρου υπολογισμένων στο προσομοιωμένο δείγμα $D^*$. Συμβολίζουμε με $d(s(D), s(D^*))$ αυτήν την απόκλιση. Αποδεχόμαστε την τιμή $\theta^*$ αν $d(s(D), s(D^*)) \leq \epsilon$, για κάποιο $\epsilon > 0$, διαφορετικά επαναλαμβάνουμε τα βήματα 3-6.

7. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $N$, μη απορριπτέων τιμών $\theta^*$, ενώ ταυτόχρονα αποθηκεύονται και οι δείκτες $l$ για τις μη απορριπτέες τιμές $\theta^*$.

Η κοινή ιδέα των δύο παραπάνω αλγορίθμων είναι ότι αρχικά στο Βήμα 3 λαμβάνεται ένα μοντέλο (χρησιμοποιώντας την εκ των προτέρων κατανομή που προσδιορίστηκε στο Βήμα 2). Στη συνέχεια στο Βήμα 4, δοθέντος του μοντέλου που προσομοιώθηκε, οι παράμετροι αυτού του μοντέλου προσομοιώνονται από την εκ των προτέρων κατανομή που έχει προσδιοριστεί για αυτό το μοντέλο (Βήμα 1 του αλγορίθμου). Τέλος, μια προσομοίωση εκτελείται όπως στο βασικό μοντέλο ABC στο Βήμα 5 και το δείγμα που προκύπτει απορρίπτεται ή όχι με βάση το κριτήριο στο Βήμα 6. Η παραπάνω διαδικασία έχει ως αποτέλεσμα να προκύψουν εν τέλει $N$ το πλήθος μη απορριπτέες τιμές και οι σχετικές συχνότητες αποδοχής για τα διαφορετικά μοντέλα προσεγγίζουν τώρα την εκ των υστέρων κατανομή για αυτά τα μοντέλα. Προφανώς, υπολογιστικές βελτιώσεις του παραπάνω αλγορίθμου αντίστοιχες με αυτές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο έχουν προταθεί στη βιβλιογραφία (βλέπε, ενδεικτικά, τις εργασίες των Toni et al. (2009), Toni and Stumpf (2010), Lee et al. (2015), Sun et al. (2015), Abdessalem et al. (2018) και Everitt and Rowińska (2021)).

Στο σημείο αυτό εφαρμόζεται η διαδικασία του αλγορίθμου ABC επιλογής μοντέλου που αναπτύχθηκε σε αυτήν την παράγραφο, με σκοπό τη διευκρίνηση των πιο βασικών παραπάνω αποτελεσμάτων, σε ένα απλό παράδειγμα (βλέπε Turner and Van Zandt (2012)).

**Παράδειγμα 5.2.1.** *Να συγκριθούν, χρησιμοποιώντας τον αλγόριθμο ABC επιλογής μοντέλου στη βάση όλων των δεδομένων, τα διωνυμικά μοντέλα $M_0$ και $M_1$, που ορίζονται ως:*

$$M_0 : x_i \sim B(p_0),\ \ \mu\epsilon\ p_0 = \theta_0 = 0.6$$

*και*

$$M_1 : x_i \sim B(p_1),\ \ \mu\epsilon\ p_1 = \theta_1 \sim Beta(1,1)\ \ \kappa\alpha\iota\ i = 1, 2, ..., n,$$

*όπου $x = (x_1, x_2, ..., x_n)$ είναι τα διαθέσιμα δεδομένα.*

***Λύση***. Αρχικά, προσομοιώνουμε ένα δείγμα $D = (x_i)$, $i = 1, ..., n$, με $n = 100$, από την κατανομή Bernoulli με πιθανότητα επιτυχίας $p = 0.7$. Η συλλογή αυτών των τιμών αποτελεί και το πραγματικό μας δείγμα. Για να γίνει πιο κατανοητή η διαδικασία, θα υλοποιηθούν ένα-ένα τα βήματα της μεθόδου όπως παρουσιάστηκαν παραπάνω. Αυτό σημαίνει ότι πρώτα πρέπει να υιοθετηθεί μια εκ των προτέρων κατανομή για κάθε άγνωστη παράμετρο των μοντέλων $M_0$ και $M_1$.

Στην κλάση των διωνυμικών μοντέλων, στο Μπεϋζιανό πλαίσιο, η υπόθεση μιας Βήτα εκ των προτέρων κατανομής, με παραμέτρους $a, b > 0$, αποτελεί συνήθως μια καλή επιλογή. Το γεγονός αυτό αιτιολογείται άμεσα διότι κατά αυτόν τον τρόπο η άγνωστη παράμετρος κυμαίνεται μεταξύ 0 και 1. Όπως έχει ήδη αναφερθεί, ο προσδιορισμός μιας μη πληροφοριακής εκ των προτέρων κατανομής για την παράμετρο $p$ είναι σε πολλές περιπτώσεις απαραίτητος. Παρατηρείστε, επίσης, ότι $p_1 \sim Beta(1,1) \equiv U(0,1)$, που σημαίνει ότι $\pi(p_1|M_1) = 1$, $p_1 \in (0,1)$. Δηλαδή, θέτοντας $a = 1$ και $b = 1$, εύκολα προκύπτει ότι η $Beta(1,1)$ αποτελεί μια μη πληροφοριακή εκ των προτέρων κατανομή για την παράμετρο $p_1$.

Στη συνέχεια, ακολουθώντας το Βήμα 2 του αλγορίθμου, θεωρούμε ότι $\pi(M_0)$ και $\pi(M_1)$ είναι οι εκ των προτέρων πιθανότητες επιλογής κάθε μοντέλου. Υπό την προϋπόθεση πως δεν υπάρχει κάποια υπόνοια ότι ένα από τα δύο μοντέλα είναι προτιμότερο τότε έχουμε ίσες εκ των προτέρων πιθανότητες, $\pi(M_0) = \pi(M_1) = 1/2$. Στο παραπάνω εδάφιο, αν υποθέσουμε ότι $M$ είναι το αληθινό άγνωστο μοντέλο επιθυμούμε ουσιαστικά να ελέγξουμε τη μηδενική υπόθεση $H_0 : M = M_0$ (δηλαδή την $X \sim M_0$, με πιθανοφάνεια $f(x|p_0, M_0)$ και εκ των προτέρων κατανομή $\pi(p_0)$) έναντι της εναλλακτικής υπόθεσης $H_1 : M = M_1$ ($X \sim M_1$, με πιθανοφάνεια $f(x|p_1, M_1)$ και εκ των προτέρων κατανομή $\pi(p_1)$, αντίστοιχα).

Προχωρώντας στα επόμενα τρία βήματα της διαδικασίας, και παρακινούμενοι από τους Turner and Van Zandt (2012), προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., 100$, από την κατανομή Bernoulli. Στο σημείο αυτό πρέπει να καθορισθεί ένα μέτρο της εγγύτητας ώστε να συγκριθούν οι πραγματικές τιμές με τις αντίστοιχες προσομοιωμένες που παράγονται χρησιμοποιώντας είτε το μοντέλο $M_0$ είτε το μοντέλο $M_1$ (βλέπε Βήμα 6 της μεθόδου). Σε αυτό το παράδειγμα, η σύγκριση των συνόλων $D$ και $D^*$ επιτυγχάνεται αξιοποιώντας την απόσταση $d = \frac{1}{n}|D - D^*|$ και θέτοντας το επίπεδο ανεκτικότητας $\epsilon$ να ισούται με 0.

Σε αυτό το σημείο, ας σημειωθεί ότι αφού εκτελέσουμε τον αλγόριθμο προσομοιώνονται 2.000 τιμές για την παράμετρο $p$ και παράλληλα γίνεται και η επιλογή του μοντέλου σε κάθε περίπτωση. Όπως αναφέρουν οι Turner and Van Zandt (2012), αξίζει να παρατηρήσουμε ότι καθώς αυξάνεται το $n$, αυξάνεται και το ποσό της πληροφορίας για την παράμετρο $p$ με αποτέλεσμα να αυξάνεται και η ακρίβεια του αλγορίθμου. Σύμφωνα με τη διαδικασία που αναπτύχθηκε προηγουμένως, και για να γίνει αντιληπτή η παραπάνω παρατήρηση, ο αλγόριθμος υλοποιείται ξανά από την αρχή αλλά αυτήν τη φορά θέτοντας $n = 1.000$.

Συγκεντρωτικά, στην πρώτη φορά που εκτελείται ο αλγόριθμος (δηλαδή για $n = 100$ τιμές) το πλήθος των φορών που γίνεται αποδεκτό το μοντέλο $M_0$ ισούται με 1.765 (από τις 2.000 συνολικά), ενώ όσον αφορά το μοντέλο $M_1$ το

αποδεχόμαστε 235 (από τις 2.000, αντίστοιχα) το πλήθος φορές. Τονίζουμε ότι ο αλγόριθμος αποδέχεται περισσότερες τιμές του μοντέλου $M_0$. Απ΄ όσα προαναφέρθηκαν στην ενότητα αυτή, για την επιλογή του κατάλληλου μοντέλου, καταλήγουμε ότι η εκτίμηση της εκ των υστέρων κατανομής των δύο υποψήφιων μοντέλων βασίζεται στη σχετική συχνότητα αποδοχής για τα δύο μοντέλα. Αυτό σημαίνει ότι αρκεί να εξετάσουμε τις τιμές που παίρνει η προσέγγιση του παράγοντα Bayes, η οποία υπολογίζεται από τη σχέση:

$$\frac{f(M_1|x)}{f(M_0|x)} = \frac{235/2.000}{1765/2.000} \approx 0,133.$$

Παρατηρείστε ότι καθώς το παραπάνω πηλίκο είναι μικρότερο της μονάδας, τότε και ο παράγοντας Bayes θα λαμβάνει τιμές μικρότερες της μονάδας. Επομένως, το συμπέρασμά μας είναι ότι δεν έχουμε ισχυρές ενδείξεις για να απορρίψουμε τη μηδενική υπόθεση ή ισοδύναμα δεν έχουμε ισχυρές ενδείξεις για να απορρίψουμε το μοντέλο $M_0$ και να θεωρήσουμε κατάλληλο το $M_1$.

Στη δεύτερη φορά εκτέλεσης του αλγορίθμου, για $n = 1.000$, τα αποτελέσματα διαφοροποιούνται σε πολύ μεγάλο βαθμό εφόσον καταλήγουμε πως το μοντέλο $M_1$ γίνεται αποδεκτό και στις 2.000 προσομοιωμένες τιμές και από την άλλη μεριά το μοντέλο $M_0$ δε γίνεται αποδεκτό καμία φορά. Ουσιαστικά κάθε τιμή της άγνωστης παραμέτρου $p$ προέρχεται από το μοντέλο $M_1$. Με παρόμοιο τρόπο όπως εξηγήθηκε προηγουμένως, και με βάση τη σχετική συχνότητα αποδοχής για τα δύο μοντέλα, υπολογίζεται η προσέγγιση του παράγοντα Bayes. Καθώς το αντίστοιχο πηλίκο είναι πολύ μεγαλύτερο της μονάδας (τείνει στο άπειρο), τότε και ο παράγοντας Bayes θα λαμβάνει τιμές πολύ μεγαλύτερες της μονάδας. Άρα, στη δεύτερη εκτέλεση του αλγορίθμου, το συμπέρασμά μας είναι ότι έχουμε πολύ ισχυρές ενδείξεις για να απορρίψουμε τη μηδενική υπόθεση ή ισοδύναμα έχουμε πολύ ισχυρές ενδείξεις για να απορρίψουμε το μοντέλο $M_0$ και να θεωρήσουμε πιο κατάλληλο το $M_1$.

Κλείνοντας, θα ήταν παράλειψη να μην επισημανθεί ότι στο συγκεκριμένο παράδειγμα, στο πλαίσιο του διωνυμικού μοντέλου, ο αναλυτικός υπολογισμός του παράγοντα Bayes είναι εφικτός. Υπενθυμίζεται ότι η κλειστή μορφή του εκφράζεται από τη σχέση (5.2) και συνεπώς αρκεί να υπολογίζουμε τις ποσότητες $f(x|M_0)$ και $f(x|M_1)$.

Αξιοποιώντας τη σχέση (5.3), η περιθώρια συνάρτηση πιθανοφάνειας του μοντέλου $M_1$, $f(x|M_1)$, δίνεται από τη σχέση:

$$f(x|M_1) = \int_0^1 f(x|p_1, M_1)\, \pi(p_1|M_1)\, dp_1, \tag{5.5}$$

όπου $p_1$ είναι το διάνυσμα των άγνωστων παραμέτρων υπό το $M_1$ μοντέλο, $f(x|p_1, M_1) = \binom{n}{x} p_1^x (1-p_1)^{n-x}$, $x = 0, 1, \ldots, n$, εκφράζει την πιθανοφάνεια του μοντέλου $M_1$ και με $\pi(p_1|M_1)$ συμβολίζεται η εκ των προτέρων κατανομή των άγνωστων παραμέτρων δεδομένου του μοντέλου $M_1$. Συνδυάζοντας τις παραπάνω παρατηρήσεις, η σχέση (5.5) ισοδύναμα μπορεί να γραφεί ως:

$$\begin{aligned} f(x|M_1) &= \binom{n}{x} \int_0^1 p_1^{(x+1)-1} (1-p_1)^{(n-x+1)-1} \, dp_1 \\ &= \binom{n}{x} B(x+1, n-x+1). \end{aligned}$$

Επιπλέον, καθώς $f(x|M_0) = \binom{n}{x} p_0^x (1-p_0)^{n-x}$, είναι:

$$BF_{10} = \frac{f(x|M_1)}{f(x|M_0)} = \frac{\binom{n}{x} B(x+1, n-x+1)}{\binom{n}{x} p_0^x (1-p_0)^{n-x}}.$$

Αυτό σημαίνει ότι αν μετά το πέρας του αλγορίθμου γνωρίζουμε το $x = \sum_{i=1}^{n} x_i$ μπορεί να υπολογιστεί η πραγματική τιμή του παράγοντα Bayes.

Για την υλοποίηση του παραδείγματος χρησιμοποιήθηκε, με κατάλληλες τροποποιήσεις, ο αλγόριθμος που είναι διαθέσιμος στην ιστοσελίδα `https://github.com/elpidapatta/approximate-bayesian-computation`.

Στο σημείο αυτό θα πρέπει να επισημανθούν κάποια σημαντικά προβλήματα που προκύπτουν χρησιμοποιώντας τον δεύτερο από τους παραπάνω αλγορίθμους, καθώς εκτός από μερικές ειδικές περιπτώσεις, μπορεί να δώσει εξαιρετικά φτωχά αποτελέσματα όσον αφορά την επιλογή μοντέλου. Ένα σημαντικό ζήτημα είναι ότι τα στατιστικά περιγραφικά μέτρα, τα οποία είναι καλά για την παραμετρική συμπερασματολογία των μοντέλων, δεν είναι και απαραίτητα πληροφοριακά για την επιλογή μεταξύ των μοντέλων. Ειδικότερα, έχει αποδειχθεί ότι ο συνδυασμός μη επαρκών στατιστικών περιγραφικών μέτρων και του ABC μπορεί να είναι προβληματικός για την επιλογή μοντέλου (βλέπε Robert et al. (2011b)). Ουσιαστικά, παρά το γεγονός ότι έχει προταθεί στη βιβλιογραφία η χρήση του ABC για την επιλογή κατάλληλου μοντέλου σε ένα ευρύτερο φάσμα εφαρμογών, οι Robert et al. (2011a) απέδειξαν ότι η χρήση του ABC αλγορίθμου με συνοπτικά περιγραφικά

μέτρα για την επιλογή μοντέλου είναι γεμάτη κινδύνους, με την έννοια ότι μπορεί να μην προσεγγίζονται ορθά οι εκ των υστέρων κατανομές των μοντέλων.

Πράγματι, αν υποθέσουμε για τον παράγοντα Bayes που βασίζεται στα στατιστικά περιγραφικά μέτρα $S(x)$ ότι συμβολίζεται με $BF^s_{1,2}$, τότε η σχέση μεταξύ των $BF_{1,2}$ και $BF^s_{1,2}$ έχει τη μορφή (Didelot et al. (2011)):

$$BF_{1,2} = \frac{P(x \mid M_1)}{P(x \mid M_2)} = \frac{P(x \mid S(x), M_1)P(S(x) \mid M_1)}{P(x \mid S(x), M_2)P(S(x) \mid M_2)}.$$

Επομένως:

$$BF_{1,2} = BF^s_{1,2}\,\frac{P(x \mid S(x), M_1)}{P(x \mid S(x), M_2)}, \tag{5.6}$$

όπου ο παράγοντας Bayes που συμπεριλαμβάνει το στατιστικό περιγραφικό μέτρο $S$, είναι η λεγόμενη χωρίς σφάλματα Monte Carlo εκδοχή του ABC παράγοντα Bayes και μάλιστα ισούται με:

$$BF^s_{1,2} = \frac{P(S(x) \mid M_1)}{P(S(x) \mid M_2)}.$$

Στο πλαίσιο αυτό, επισημαίνεται ότι ο παράγοντας Bayes $BF^s_{1,2}$ ουσιαστικά λαμβάνεται όταν η επιλογή του μοντέλου βασίζεται στο στατιστικό περιγραφικό μέτρο $S$ και όχι στα δεδομένα. Στην πιο συνηθισμένη περίπτωση που κάποιος δεν μπορεί να καταφύγει σε επαρκή στατιστικά περιγραφικά μέτρα, ο ABC παράγοντας Bayes μπορεί να αποκλίνει με τον ένα ή τον άλλο τρόπο καθώς αυξάνεται ο αριθμός των παρατηρήσεων. Μια αξιοσημείωτη εξαίρεση είναι η περίπτωση των τυχαίων πεδίων Gibbs, όπου οι Grelaud et al. (2009) έχουν δείξει πώς να αντλούνται επαρκή στατιστικά περιγραφικά μέτρα μεταξύ των μοντέλων, πέρα από το πρωτογενές δείγμα. Έτσι, ένα στατιστικό περιγραφικό μέτρο $S(x)$ είναι επαρκές για τη σύγκριση δύο μοντέλων $M_1$ και $M_2$ αν και μόνο αν:

$$P(x \mid S(x), M_1) = P(x \mid S(x), M_2),$$

που οδηγεί στο συμπέρασμα ότι $BF_{1,2} = BF^s_{1,2}$. Είναι επίσης σαφές από την παραπάνω εξίσωση ότι στις περιπτώσεις που η συνθήκη δεν ικανοποιείται μπορεί να υπάρχει τεράστια διαφορά μεταξύ των $BF_{1,2}$ και $BF^s_{1,2}$, όπως ανέδειξαν μέσω παραδειγμάτων οι Grelaud et al. (2009) και Robert et al. (2011b). Βασικά, αποδείχθηκε ότι η επάρκεια για τα μοντέλα $M_1$ και $M_2$ μόνο, ή και για τα δύο μοντέλα, δεν εγγυάται την καλή επιλογή των μοντέλων (Didelot et al. (2011)).

Επομένως, ο υπολογισμός των παραγόντων Bayes στο $S(x)$ μπορεί να είναι παραπλανητικός για την επιλογή μοντέλου, εκτός αν το πηλίκο μεταξύ των παραγόντων Bayes στα $x$ και $S(x)$ θα ήταν διαθέσιμο, ή τουλάχιστον θα μπορούσε να

προσεγγιστεί αρκετά καλά. Πιο συγκεκριμένα, ο υπολογισμός των παραγόντων Bayes στα στατιστικά περιγραφικά μέτρα μπορεί να μη σχετίζεται με τους παράγοντες Bayes των αρχικών δεδομένων, με αποτέλεσμα τα συμπεράσματα να μην έχουν νόημα και ερμηνεία. Σε αυτές τις περιπτώσεις μια λύση, που μπορεί να παρέχει χρήσιμη καθοδήγηση, είναι να χρησιμοποιηθούν μόνο στατιστικά περιγραφικά μέτρα που πληρούν τις επαρκείς συνθήκες, οι οποίες παράγουν μια συνεπή επιλογή Μπεϋζιανού μοντέλου, σύμφωνα με όσα απέδειξαν οι Marin et al. (2014). Ειδικότερα, οι Marin et al. (2014) κατέληξαν ότι τα περιγραφικά μέτρα δε θα πρέπει να χρησιμοποιούνται στους αλγόριθμους ABC παρά μόνο αν οι αναμενόμενες τιμές τους μπορεί να δειχθεί ότι διαφοροποιούνται για τα διαφορετικά υποψήφια μοντέλα.

Παρόλα αυτά, τα παραπάνω ζητήματα αφορούν μόνο τον αλγόριθμο ABC επιλογής μοντέλου όπου η διάσταση των δεδομένων έχει μειωθεί μέσω της σύγκρισης περιγραφικών συνοπτικών μέτρων, ενώ ο ABC αλγόριθμος στον οποίο συγκρίνονται άμεσα τα πραγματικά σύνολα δεδομένων ξεπερνά τα προβλήματα αυτά.

Τέλος, παρατηρείστε ότι και οι δύο παραπάνω αλγόριθμοι έχουν ως προαπαιτούμενο τον καθορισμό εκ των προτέρων κατανομής για την επιλογή μοντέλου. Ένας τρόπος για να αποφευχθεί ο καθορισμός αυτός, είναι η εφαρμογή της ABC μεθόδου και ο υπολογισμός ενός κριτηρίου επιλογής μοντέλου, όπως το $DIC$. Αυτή η εναλλακτική προσέγγιση εφαρμόστηκε πρόσφατα στην εργασία των Economou et al. (2021) και έγκειται στην υλοποίηση των παρακάτω.

1. Για $m = 1, ..., k$ υιοθετούμε μια εκ των προτέρων κατανομή, έστω $\pi_m(\theta_m)$, για κάθε παράμετρο $\theta_m$, $m = 1, ..., K$, με $\theta_m$ την παράμετρο του $m-$οστού μοντέλου.

2. Προσομοιώνουμε την τιμή $\theta^*$ από την εκ των προτέρων κατανομή $\pi_m(\theta_m)$.

3. Προσομοιώνουμε ένα δείγμα $D^* = (x_i^*)$, $i = 1, ..., n$ από το μοντέλο $f(\cdot|\theta^*, M_m)$, όπου $\theta^*$ είναι η τιμή της παραμέτρου του $m$-οστού μοντέλου που προσομοιώθηκε στο Βήμα 2 και $f(\cdot|\theta^*, M_m)$ η πιθανοφάνεια του $m$-οστού μοντέλου υπολογισμένη για αυτήν την τιμή.

4. Αποδεχόμαστε την τιμή $\theta^*$ αν $d(D, D^*) \leq \epsilon$, για κάποιο $\epsilon > 0$, διαφορετικά επαναλαμβάνουμε τα βήματα 3-6.

5. Τα προηγούμενα βήματα επαναλαμβάνονται μέχρις ότου προκύψει ένας προκαθορισμένος αριθμός, έστω $N$, μη απορριπτέων τιμών $\theta^*$ για το $m$-οστό μοντέλο.

6. Η διαδικασία επαναλαμβάνεται για όλα τα $K$ το πλήθος μοντέλα.

Η παραπάνω διαδικασία δημιουργεί $M$ το πλήθος τιμές από καθένα εκ των $K$ το πλήθος μοντέλα για τη σύγκριση των οποίων επιλέγεται το μοντέλο με τη μικρότερη τιμή του $DIC$ κριτηρίου που εκτιμάται από τη σχέση:

$$\mathrm{DIC}_m = -4 \cdot \frac{\sum_{i=1}^{n} logf(x_i^*|\theta^*, M_m)}{n} + 2 \cdot logf(x|\tilde{\theta}_i),$$

όπου $\tilde{\theta}_i$ είναι η εκ των υστέρων μέση τιμή του $\theta_i$. Το μοντέλο με τη μικρότερη τιμή είναι εκείνο που θα πρέπει να προτιμηθεί.

# ΚΕΦΑΛΑΙΟ 6

# Εφαρμογη

Στο κεφάλαιο αυτό η εκτίμηση και η επιλογή μοντέλου με Προσεγγιστικές Μπεϋζιανές μεθόδους θα εφαρμοστεί σε ένα πραγματικό σύνολο δεδομένων. Ειδικότερα, παρακινούμενοι από τους Economou et al. (2021), το ενδιαφέρον μας θα επικεντρωθεί στις μετρήσεις που αφορούν το βάρος και το επιτόπιο κατακόρυφο άλμα (μετρημένα σε lb και ίντσες, αντίστοιχα) αθλητών καλαθοσφαίρισης που έχουν επιλεγεί στα ντραφτ του NBA το έτος 2017. Τα δεδομένα αυτά έχουν προέλθει από την ιστοσελίδα `https://data.world/achou/nba-draft-combine-measurements/workspace/file?filename=nba_draft_combine_all_Years.csv`. Στο πλαίσιο αυτό, η διάρθρωση του κεφαλαίου έχει ως εξής. Στην Ενότητα 6.1 περιγράφεται ο υπό μελέτη πληθυσμός, ο τρόπος επιλογής του δείγματος και αιτιολογείται γιατί αυτό είναι ένα διδιάστατο μεροληπτικό και όχι ένα τυχαίο δείγμα. Στην Ενότητα 6.2, ακολουθώντας τους Economou et al. (2020), χρησιμοποιείται η έννοια των σταθμισμένων κατανομών για τη μοντελοποίηση διδιάστατων μεροληπτικών δειγμάτων, που προκύπτουν λόγω της μεθόδου δειγματοληψίας που έχει ως αποτέλεσμα την υπο(υπερ)εκπροσώπηση σ΄ αυτά πειραματικών μονάδων με συγκεκριμένα χαρακτηριστικά. Τέλος, στην Ενότητα 6.3 χρησιμοποιώντας τη μέθοδο ABC εκτιμούμε τις άγνωστες πληθυσμιακές παραμέτρους και επιλέγουμε το καλύτερο μοντέλο, ως προς την προσαρμογή, στο παραπάνω σύνολο δεδομένων.

## 6.1 Πληθυσμός και δείγμα υπό μελέτη

Τα ντραφτ του NBA είναι ένα ετήσιο γεγονός κατά το οποίο οι ομάδες, τριάντα τον αριθμό, του NBA επιλέγουν μελλοντικούς παίκτες που πληρούν συγκεκριμένα τυπικά κριτήρια, τα κυριότερα εκ των οποίων αναλυτικά θα περιγραφούν στη συνέχεια. Από το 1989, η διαδικασία αποτελείται από δύο γύρους και εξήντα παίκτες επιλέγονται, οι οποίοι δεν μπορούν να υπογράψουν συμβόλαιο για να παίξουν για

οποιαδήποτε άλλη ομάδα του NBA πλην αυτής που τους επέλεξε. Τα τυπικά κριτήρια που πρέπει ένας παίκτης να πληροί για να είναι επιλέξιμος αλλάζουν με την πάροδο του χρόνου, αλλά τα βασικά κριτήρια είναι τα ακόλουθα:

- να είναι ο παίκτης ηλικίας τουλάχιστον 19 ετών στο τέλος του ημερολογιακού έτους επιλογής από τα ντραφτ και να μην έχει παίξει ποτέ πριν στο NBA, ή
- να έχει παίξει τουλάχιστον ένα χρόνο στο κολεγιακό πρωτάθλημα.

Από τα παραπάνω προκύπτει ότι οι δυνητικά υποψήφιοι για τα ντραφτ του NBA είναι οι παίκτες του κολεγιακού πρωταθλήματος που συμμετέχουν στο NCAA (National Collegiate Athletic Association), αλλοδαποί καλαθοσφαιριστές ηλικίας τουλάχιστον 19 ετών στο τέλος του ημερολογιακού έτους επιλογής από τα ντραφτ και παίκτες που έχουν αποφοιτήσει εδώ και ένα έτος από το high school και είναι τουλάχιστον 19 ετών στο τέλος του ημερολογιακού έτους επιλογής από τα ντραφτ. Η παραπάνω συλλογή ατόμων αποτελεί τον υπό μελέτη πληθυσμό.

Όσον αφορά την επιλογή του ντραφτ αυτή συνηθέστερα γίνεται στο τέλος Ιουνίου κατά τη διάρκεια της διακοπής των αγώνων για το NBA, ενώ έχει προηγηθεί για πολλούς από τους αθλητές που επιθυμούν να επιλεγούν μια περίοδο εντατικών προπονήσεων. Καθώς με την επιλογή αθλητών στα ντραφτ δίνεται η δυνατότητα οι ομάδες να αποκτήσουν καλαθοσφαιριστές που θα έχουν ιδιαίτερα χαρακτηριστικά και θα δυναμώσουν αγωνιστικά τις ομάδες, είναι προφανές ότι οι παίκτες που επιλέγονται στα ντραφτ αποτελούν ένα μεροληπτικό δείγμα και όχι ένα τυχαίο δείγμα από τον πληθυσμό των παικτών που είναι δυνητικά υποψήφιοι για τα ντραφτ.

## 6.2 Μεροληπτικό διδιάστατο δείγμα και σταθμισμένες κατανομές

Επεκτείνοντας και γενικεύοντας τις ιδέες του Fisher (1934), ο Rao (1965) ήταν ο πρώτος που χρησιμοποίησε την έννοια των σταθμισμένων κατανομών (weighted distributions) ως μια μέθοδο που μπορεί να εφαρμοστεί σε πολλές περιπτώσεις που έχουμε παρατηρήσεις που δεν αποτελούν τυχαίο δείγμα από τον υπό μελέτη πληθυσμό, αλλά η πιθανότητα επιλογής μιας πειραματικής μονάδας είναι ανάλογη μιας συνάρτησης, της λεγόμενης συνάρτησης βάρους. Για ιδιότητες των σταθμισμένων κατανομών παραπέμπουμε, μεταξύ άλλων, στις εργασίες των Arnold and

Nagaraja (1991), Jain and Nanda (1995), Navarro et al. (2006) και Sarabia and Gómez-Déniz (2008).

Οι ιδέες του Rao (1965) για τη μοντελοποίηση μεροληπτικών δειγμάτων με τη βοήθεια σταθμισμένων κατανομών μπορούν άμεσα να γενικευθούν στη διδιάστατη περίπτωση (βλέπε Economou et al. (2020)). Ειδικότερα, έστω $(X_1, Y_1)^t$, $(X_2, Y_2)^t$, ..., $(X_n, Y_n)^t$ είναι ένα διδιάστατο δείγμα από έναν πληθυσμό με συνάρτηση πυκνότητας πιθανότητας $f(x, y; \theta)$, όπου $\theta \in \Theta \subset R^s$, $s \geq 1$. Τότε αν η πιθανότητα επιλογής μιας πειραματικής μονάδας από τον πληθυσμό είναι ανάλογη της μη αρνητικής συνάρτησης $w(x, y; \theta, \gamma)$ το παρατηρούμενο δείγμα μπορεί να θεωρηθεί ως ένα τυχαίο δείγμα από το τυχαίο διάνυσμα των σταθμισμένων τυχαίων μεταβλητών $(X_w, Y_w)^t$ με συνάρτηση βάρους την $w(x, y; \theta, \gamma)$ και από κοινού συνάρτηση πυκνότητας πιθανότητας:

$$f_w(x, y; \theta, \gamma) = \frac{w(x, y; \theta, \gamma)}{E_f\{w(X, Y; \theta, \gamma)\}} f(x, y; \theta), \tag{6.1}$$

όπου $E_f\{w(X, Y; \theta, \gamma)\} < \infty$, με $\gamma$ ένα διάνυσμα πρόσθετων παραμέτρων που μπορεί να συμμετέχουν στον ορισμό της συνάρτησης βάρους.

**Παρατήρηση 6.2.1.** *Από τα παραπάνω προκύπτει ότι οι συναρτήσεις βάρους για τη μοντελοποίηση ενός διδιάστατου μεροληπτικού δείγματος θα πρέπει να επιλέγονται έτσι ώστε να πληρούν τις ακόλουθες ιδιότητες:*

- ***Ιδιότητα 1.*** *Να είναι μη αρνητικές συναρτήσεις $w(x, y; \theta, \gamma)$.*
- ***Ιδιότητα 2.*** *Να έχουν πεπερασμένη μέση τιμή $E_f\{w(X, Y; \theta, \gamma)\} < \infty$.*
- ***Ιδιότητα 3.*** *Να είναι τέτοιες ώστε να αντανακλούν την πιθανότητα επιλογής ενός ζεύγους $(x, y)$ στο δείγμα.*

Μεροληπτικά διδιάστατα δείγματα προκύπτουν πολλές φορές λόγω της μεθόδου δειγματοληψίας που έχει ως αποτέλεσμα την υπο(υπερ)εκπροσώπηση σε αυτά πειραματικών μονάδων με συγκεκριμένα χαρακτηριστικά. Ειδικότερα, μεροληπτικά διδιάστατα δείγματα προκύπτουν υπό τα ακόλουθα τέσσερα διαφορετικά σενάρια.

- **Περίπτωση Α:** πειραματικές μονάδες με μεγάλες τιμές στην τυχαία μεταβλητή $X$ ή/και μεγάλες τιμές στην τυχαία μεταβλητή $Y$ είναι πιο πιθανό να επιλεγούν, ενώ από την άλλη μεριά ο μη τυχαίος μηχανισμός δειγματοληψίας δίνει μικρή πιθανότητα επιλογής στα ζευγάρια $(x, y)$ με μικρές τιμές και στις δύο μεταβλητές $X$ και $Y$.

- **Περίπτωση B:** πειραματικές μονάδες με μικρές τιμές στην τυχαία μεταβλητή $X$ ή/και μικρές τιμές στην τυχαία μεταβλητή $Y$ είναι πιο πιθανό να επιλεγούν, ενώ από την άλλη μεριά ο μη τυχαίος μηχανισμός δειγματοληψίας δίνει μικρή πιθανότητα επιλογής στα ζευγάρια $(x, y)$ με μεγάλες τιμές και στις δύο μεταβλητές $X$ και $Y$.

- **Περίπτωση C:** πειραματικές μονάδες με μεγάλες τιμές στην τυχαία μεταβλητή $X$ ή/και μικρές τιμές στην τυχαία μεταβλητή $Y$ είναι πιο πιθανό να επιλεγούν, ενώ από την άλλη μεριά ο μη τυχαίος μηχανισμός δειγματοληψίας δίνει μικρή πιθανότητα επιλογής στα ζευγάρια $(x, y)$ με μικρές τιμές στην $X$ και μεγάλες τιμές στην $Y$.

- **Περίπτωση D:** πειραματικές μονάδες με μικρές τιμές στην τυχαία μεταβλητή $X$ ή/και μεγάλες τιμές στην τυχαία μεταβλητή $Y$ είναι πιο πιθανό να επιλεγούν, ενώ από την άλλη μεριά ο μη τυχαίος μηχανισμός δειγματοληψίας δίνει μικρή πιθανότητα επιλογής στα ζευγάρια $(x, y)$ με μεγάλες τιμές στην $X$ και μικρές τιμές στην $Y$.

Για τη μοντελοποίηση των παραπάνω περιπτώσεων υπάρχει μη πεπερασμένος αριθμός συναρτήσεων βάρους που μπορούν να προταθούν. Στην πρόσφατη εργασία των Economou et al. (2021) προτάθηκε μια διαφορετική συνάρτηση βάρους για καθεμία από τις παραπάνω περιπτώσεις με κριτήριο να είναι λογικές και μαθηματικά βολικές. Πιο συγκεκριμένα προτάθηκαν αντίστοιχα οι συναρτήσεις βάρους:

$$w_1(x, y; \theta, \gamma_x, \gamma_y) = 1 - \{1 - F_X(x; \theta_x)\}^{\gamma_x}\{1 - F_Y(y; \theta_y)\}^{\gamma_y}, \tag{6.2}$$

$$w_2(x, y; \theta, \gamma_x, \gamma_y) = 1 - \{F_X(x; \theta_x)\}^{\gamma_x}\{F_Y(y; \theta_y)\}^{\gamma_y}, \tag{6.3}$$

$$w_3(x, y; \theta, \gamma_x, \gamma_y) = 1 - \{1 - F_X(x; \theta_x)\}^{\gamma_x}\{F_Y(y; \theta_y)\}^{\gamma_y}, \tag{6.4}$$

και

$$w_4(x, y; \theta, \gamma_x, \gamma_y) = 1 - \{F_X(x; \theta_x)\}^{\gamma_x}\{1 - F_Y(y; \theta_y)\}^{\gamma_y} \tag{6.5}$$

όπου $\theta_x$, $\theta_y$ είναι συναρτήσεις της άγνωστης παραμέτρου $\theta$ (υπενθύμιση $(X, Y) \sim f(x, y; \theta)$), $F_X(x; \theta_x)$ και $F_Y(y; \theta_y)$ είναι οι περιθώριες αθροιστικές συναρτήσεις κατανομής των τυχαίων μεταβλητών $X$ και $Y$, αντίστοιχα, ενώ $\gamma_x, \gamma_y \geq 0$ είναι πρόσθετες παράμετροι που συμμετέχουν στον ορισμό της συνάρτησης βάρους. Σε όσα ακολουθούν θα συμβολίζουμε με $\zeta$ το διάνυσμα $\zeta = (\theta, \gamma_x, \gamma_y)$.

Σχετικά με το ρόλο των παραμέτρων $\gamma_x$ και $\gamma_y$ επισημαίνουμε ότι δεν μπορούν να είναι ταυτόχρονα ίσες με μηδέν, καθώς σε αυτήν την περίπτωση η συνάρτηση βάρους μηδενίζεται, γεγονός που παραβιάζει μία από τις κύριες ιδιότητές της. Επιπλέον, αν η παράμετρος $\gamma_x$ ή/και η $\gamma_y$ τείνει στο άπειρο, τότε η συνάρτηση βάρους είναι η μοναδιαία συνάρτηση και αντιστοιχεί στο τυχαίο δείγμα. Τέλος, θέτοντας την παράμετρο $\gamma_x = 0$ ($\gamma_y = 0$) μοντελοποιούμε εκείνη την περίπτωση όπου η μεροληψία στο διαθέσιμο μεροληπτικό διδιάστατο δείγμα οφείλεται μόνο στη μεταβλητή $Y$ ($X$, αντίστοιχα). Από τα παραπάνω γίνεται αντιληπτό ότι οι παραπάνω τέσσερις το πλήθος συναρτήσεις βάρους ορίζουν σε κάθε περίπτωση τρία διαφορετικά μοντέλα (συνολικά δηλαδή 12) και πιο συγκεκριμένα έχουμε:

- το μοντέλο $M_i$, $i = 1, ..., 4$, που προκύπτει χρησιμοποιώντας τη συνάρτηση βάρους $w_i$, $i = 1, ..., 4$, με τις τιμές των παραμέτρων $\gamma_x$ και $\gamma_y$ να είναι αυστηρά θετικές (σε αυτήν την περίπτωση η μεροληψία οφείλεται και στις δύο μεταβλητές),
- το μοντέλο $\mathcal{M}_{iy}$, $i = 1, ..., 4$, που προκύπτει χρησιμοποιώντας τη συνάρτηση βάρους $w_i$, $i = 1, ..., 4$, με $\gamma_x = 0$ και $\gamma_y > 0$ (σε αυτήν την περίπτωση η μεροληψία οφείλεται στη μεταβλητή $Y$),
- το μοντέλο $\mathcal{M}_{ix}$, $i = 1, ..., 4$, που προκύπτει χρησιμοποιώντας τη συνάρτηση βάρους $w_i$, $i = 1, ..., 4$, με $\gamma_y = 0$ και $\gamma_x > 0$ (σε αυτήν την περίπτωση η μεροληψία οφείλεται στη μεταβλητή $X$).

Χρησιμοποιώντας τα παραπάνω έχουμε (βλέπε Economou et al. (2021)) ότι η συνάρτηση πιθανοφάνειας ενός διδιάστατου δείγματος $D = (x_j, y_j), j = 1, \ldots, n$, από τον πληθυσμό με συνάρτηση πυκνότητας πιθανότητας $f(x, y; \theta)$ όταν η μεροληψία στο δείγμα περιγράφεται από τη συνάρτηση βάρους $w_i(x, y; \theta, \gamma_x, \gamma_y)$ δίνεται από τη σχέση:

$$\prod_{j=1}^{n} f_{w_i}(x_j, y_j; \theta, \gamma_x, \gamma_y) = \prod_{j=1}^{n} f_{w_i}(x_j, y_j; \zeta) = \frac{\prod_{j=1}^{n} w_i(x_j, y_j; \zeta) f(x_j, y_j; \theta)}{E_f^n[w_i(X, Y; \zeta)]},$$

καθώς όπως ήδη αναφέραμε θεωρείται τυχαίο δείγμα από τη διδιάστατη σταθμισμένη κατανομή. Προφανώς αυτή η συνάρτηση πιθανοφάνειας είναι δύσχρηστη καθώς περιέχει υπολογισμό δυνάμεων αθροιστικών συναρτήσεων κατανομών. Για το λόγο αυτό οι Economou et al. (2021) πρότειναν τη χρήση προσεγγιστικών υπολογιστικών μεθόδων, για την εκτίμηση των άγνωστων πληθυσμιακών παραμέτρων, ήτοι του διανύσματος $\theta$. Επιπλέον, η διερεύνηση αν η παρατηρούμενη διδιάστατη μεροληψία προέρχεται από μια εκ των δύο τυχαίων μεταβλητών είτε

και από τις δύο, διεξάγεται μέσω της επιλογής μοντέλου. Επομένως πρότειναν την υλοποίηση του παρακάτω αλγορίθμου για κάθε ένα μοντέλο εκ των παραπάνω 12 (χωρίς τις διευκρινήσεις, σχόλια που παρεμβάλλουμε για πληρότητα).

1. Έστω $D = (x_j, y_j)$, $j = 1, \ldots, n$, είναι το διαθέσιμο μεροληπτικό διδιάστατο δείγμα μεγέθους $n$.

2. Για κάθε παράμετρο που συμμετέχει στην έκφραση της σταθμισμένης κατανομής υιοθετούμε μια εκ των προτέρων κατανομή.
   Σχόλιο: Το παραπάνω ουσιαστικά σημαίνει ότι για τα μοντέλα $M_i$, $i = 1, ..., 4$, καθορίζουμε μια εκ των προτέρων κατανομή για κάθε συνιστώσα του διανύσματος $\theta$ καθώς επίσης και μια εκ των προτέρων κατανομή για τις παραμέτρους $\gamma_x$ και $\gamma_y$. Καθώς συνήθως είναι διαθέσιμη πληροφορία για το διάνυσμα $\theta$ επιλέγονται για τις συνιστώσες του πληροφοριακές εκ των προτέρων κατανομές (βλέπε Ενότητα 4.3.4 για εκ των προτέρων κατανομές μέγιστης εντροπίας), ενώ για τις παραμέτρους $\gamma_x$ και $\gamma_y$ επιλέγονται μη πληροφοριακές εκ των προτέρων κατανομές.

3. Προσομοιώνουμε τις τιμές $\theta^*$, $\gamma_x^*$ και $\gamma_y^*$ από τις εκ των προτέρων κατανομές που καθορίστηκαν στο προηγούμενο βήμα. Θέτουμε $\zeta^* = (\theta^*, \gamma_x^*, \gamma_y^*)$.
   Στο σημείο αυτό σημειώνεται ότι στην περίπτωση των μοντέλων $\mathcal{M}_{iy}$ ($\mathcal{M}_{ix}$), $i = 1, ..., 4$, δεν απαιτείται προσομοίωση των $\gamma_x$ ($\gamma_y$, αντίστοιχα) καθώς οι τιμές αυτών είναι σταθερές και ίσες με μηδέν.

4. Αντικαθιστούμε την τιμή $\zeta^*$ στην προτεινόμενη συνάρτηση βάρους και προσομοιώνουμε ένα διδιάστατο δείγμα μεγέθους $n$, έστω $D^* = (x_j^*, y_j^*)$, $j = 1, \ldots, n$, από το μοντέλο με συνάρτηση πυκνότητας πιθανότητας $f_{w_i}(x, y, \zeta^*)$.
   Σχόλιο: Σε κάποιες περιπτώσεις η δημιουργία τυχαίου δείγματος από την $f_{w_i}(x, y, \zeta^*)$ δεν είναι τόσο εύκολη διαδικασία. Το πρόβλημα αυτό μπορεί να ξεπεραστεί ενθυμούμενοι ότι ένα τυχαίο δείγμα από την $f_{w_i}(x, y, \zeta^*)$ είναι ένα μεροληπτικό δείγμα από την $f(x, y; \theta^*)$ όπου η πιθανότητα επιλογής μιας πληθυσμιακής πειραματικής μονάδας είναι ανάλογη της συνάρτησης βάρους $w_i$. Επομένως μπορούμε να παράγουμε έναν πολύ μεγάλο αριθμό παρατηρήσεων (έστω $N = 10^3 \cdot n$) από την $f(x, y; \theta^*)$ που θα αποτελέσουν τις πληθυσμιακές παρατηρήσεις και να επιλέξουμε $n$ το πλήθος από αυτές με πιθανότητα ανάλογη της συνάρτησης βάρους. Προφανώς η παραπάνω διαδικασία προσθέτει μία ακόμη προσέγγιση στον προσδιορισμό της εκ των υστέρων κατανομής των παραμέτρων, ωστόσο η διαφοροποίηση δε θα είναι σημαντική αν το μέγεθος του πληθυσμού $N$ είναι αρκετά μεγάλο σε σχέση με το μέγεθος του δείγματος $n$.

5. Υπολογίζουμε ως ένα μέτρο της εγγύτητας $d(D, D^*)$ μεταξύ των προσομοιωμένων $D^*$ και των παρατηρούμενων πραγματικών δεδομένων την απόλυτη τιμή της κανονικοποιημένης μορφής της στατιστικής συνάρτησης

$$T = \int_X \int_Y (f_D(x, y) - f_{D^*}(x, y))^2 \, dy \, dx,$$

όπου $f_D(x, y)$ και $f_{D^*}(x, y)$ είναι οι εκτιμητές τύπου πυρήνα της πυκνότητας με βάση τα πραγματικά και προσομοιωμένα δεδομένα, αντίστοιχα. Αποδεχόμαστε την τιμή $\zeta^*$ αν $d(D, D^*) < 1.96$.
Σχόλιο: Η τιμή 1.96 όπως αναφέρθηκε στην Ενότητα 4.3.2 για το συγκεκριμένο μέτρο εγγύτητας ουσιαστικά αντιστοιχεί σε ένα ασυμπτωτικό επίπεδο σημαντικότητας 0.05 για τον έλεγχο ότι τα δύο δείγματα (προσομοιωμένο και παρατηρούμενο) έχουν την ίδια πυκνότητα. Σ′ αυτό το πλαίσιο, κάποιος μπορεί να εφαρμόσει τη συνάρτηση `kde.test` της R και να αποδεχθεί την τιμή $\zeta^*$ αν η $p$-τιμή του ελέγχου είναι μικρότερη από 0.05.

Τα παραπάνω Βήματα 3-5 επαναλαμβάνονται μέχρις ότου πάρουμε έναν προκαθορισμένο αριθμό $M$ μη απορριπτέων δειγμάτων, έστω $M = 5.000$, για κάθε υποψήφιο μοντέλο. Στη συνέχεια με βάση αυτές τις μη απορριπτέες τιμές για τα υποψήφια μοντέλα αξιοποιείται το κριτήριο DIC (βλέπε σχετικά Ενότητα 5.1).

Η εφαρμογή των παραπάνω αποτελεί αντικείμενο μελέτης της επόμενης ενότητας.

## 6.3 Εκτιμητική και επιλογή μοντέλου

Για τους αθλητές που επιλέχθηκαν στα ντραφτ του NBA το έτος 2017 επικεντρώνουμε το ενδιαφέρον μας στο βάρος τους $X$ (μετρημένο σε λίβρες) και στο επιτόπιο κατακόρυφο άλμα $Y$ (μετρημένο σε ίντσες). Όπως προηγούμενα επεξηγήθηκε πρόκειται για ένα μεροληπτικό και όχι τυχαίο διδιάστατο δείγμα από τον πληθυσμό ενδιαφέροντος που αποτελείται από όλους τους καλαθοσφαιριστές που πληρούν τα τυπικά κριτήρια επιλεξιμότητας από κάποια ομάδα του NBA στα ντραφτ. Στον Πίνακα 6.1 δίνονται κάποια περιγραφικά συνοπτικά μέτρα για το διαθέσιμο διδιάστατο μεροληπτικό δείγμα, που ίσως μας βοηθήσουν για να επιλέξουμε κατάλληλη συνάρτηση βάρους για τη μοντελοποίησή του.

Ο συντελεστής συσχέτισης του Spearman μεταξύ των δύο μεταβλητών υπολογισμένος στους αθλητές που επιλέχθηκαν στα ντραφτ του 2017 είναι ίσος με $-0.2867968$, ενώ η $p$-τιμή του ελέγχου της μη ύπαρξης γραμμικής συσχέτισης

Πίνακας 6.1: Περιγραφικά στατιστικά μέτρα για το βάρος (σε lb) και το επιτόπιο κατακόρυφο άλμα (σε ίντσες) για τους 60 αθλητές του ντραφτ του 2017.

| | Πλήθος ελλιπουσών τιμών | Μέση τιμή | Τυπική Απόκλιση | Ελάχιστη τιμή | $Q_1$ | Διάμεσος | $Q_3$ | Μέγιστη τιμή |
|---|---|---|---|---|---|---|---|---|
| Βάρος | 0 | 214.2 | 24.27049 | 166.0 | 192.8 | 219.5 | 232.0 | 277.0 |
| Επιτόπιο άλμα | 10 | 35.380 | 3.675012 | 27.500 | 32.500 | 35.250 | 38.000 | 44.500 |

απορρίπτεται σε επίπεδο σημαντικότητας 5% καθώς είναι ίση με 0.04346. Επομένως, προκύπτει ότι υπάρχει μια αρνητική συσχέτιση μεταξύ του βάρους και του επιτόπιου άλματος, κάτι που προφανώς είναι απόλυτα λογικό και που ίσως αναμέναμε σε μεγαλύτερο βαθμό.

Για να εφαρμοστεί η μοντελοποίηση που παρουσιάστηκε στην προηγούμενη ενότητα αρχικά πρέπει να καθοριστεί η από κοινού κατανομή του βάρους και του επιτόπιου κατακόρυφου άλματος στον πληθυσμό ενδιαφέροντος. Υποθέτουμε ότι η από κοινού κατανομή των $(X, Y)'$ είναι διδιάστατη κανονική με παράμετρο θέσης $(\mu_X, \mu_Y)'$, παραμέτρους κλίμακας $\sigma_X > 0$ και $\sigma_Y > 0$, ενώ ο πληθυσμιακός συντελεστής συσχέτισης των $X$ και $Y$ συμβολίζεται με $\rho$. Επιπλέον, με βάση την πρότερη γνώση ότι οι πιο εύσωμοι αθλητές έχουν μικρότερο επιτόπιο κατακόρυφο άλμα, ενώ οι λιγότερο εύσωμοι, που συνήθως είναι και πιο κοντοί έχουν μεγαλύτερο επιτόπιο κατακόρυφο άλμα, αναμένουμε να είναι $\rho \leq 0$.

Στη συνέχεια θα πρέπει να επιλεγεί μία από τις συναρτήσεις βάρους που ορίστηκαν στην προηγούμενη ενότητα και να διερευνηθεί αν κάποιο από τα δύο υπομοντέλα που αυτή περικλείει οδηγεί σε καλύτερα αποτελέσματα, ενώ θα πρέπει να καθορίσουμε εκ των προτέρων κατανομές για όλες τις παραμέτρους που υπεισέρχονται στη συνάρτηση βάρους. Για τον προσδιορισμό και καθορισμό των παραπάνω είναι πολύ χρήσιμη η εκ των προτέρων γνώση που έχει κάποιος για τις δύο υπό μελέτη τυχαίες μεταβλητές στον πληθυσμό ενδιαφέροντος. Δυστυχώς, μελέτες που αφορούν το βάρος και το επιτόπιο κατακόρυφο άλμα είναι διαθέσιμες μόνο για τους παίκτες που αγωνίζονται στο πρωτάθλημα του NCAA και όχι για τους αθλητές από άλλες χώρες που πληρούν τα τυπικά κριτήρια που έχουν αναφερθεί. Ωστόσο, το γεγονός ότι το ποσοστό των αθλητών από άλλες χώρες που αποτελεί μέλος του ρόστερ μιας ομάδας του NBA αυξάνεται τα τελευταία χρόνια, ενώ ορισμένοι από αυτούς έχουν σημαντικό ρόλο σε αυτές, μας κάνει να δεχθούμε ότι οι υποψήφιοι διεθνείς παίκτες έχουν τα ίδια ή αν όχι καλύτερα χαρακτηριστικά συγκριτικά με τους νέους παίκτες που αγωνίζονται στο NCAA. Επομένως, η πρότερη γνώση για τον υπό μελέτη πληθυσμό θα προέλθει από τις μελέτες για το βάρος και το επιτόπιο κατακόρυφο ύψος των αθλητών που αγωνίζονται στο NCAA. Στην εργασία των Richard et al. (1994) καταγράφεται ότι το μέσο βάρος (σε lb) και το μέσο κατακόρυφο άλμα (σε ίντσες) των αθλη-

τών που αγωνίζονται στο NCAA είναι περίπου 201.28 (εύρος: 140.21- 300.71, τυπική απόκλιση 24.47) και 28 (εύρος: 10 έως 41.5, τυπική απόκλιση 4.09), αντίστοιχα. Ωστόσο, σε πιο πρόσφατες ανεπίσημες μελέτες το μέσο κατακόρυφο άλμα αναφέρεται ότι είναι μεταξύ 27 και 30 ιντσών, με ένα εύρος τιμών από 18-42 (`http://www.homeexerciseequipmenthq.com/average-vertical-jump/`), ενώ το μέσο βάρος τους αναφέρεται ότι είναι περίπου 18 pounds λιγότερο από το μέσο βάρος ενός αθλητή του NBA που σύμφωνα με την ίδια έρευνα είναι ίσο με 222.04 (`https://www.psychguides.com/interact/male-body-image-and-the-average-athlete/`). Επομένως μπορούμε να θεωρήσουμε ότι το μέσο βάρος είναι περίπου 204 pounds.

Από τα προηγούμενα και λαμβάνοντας υπόψη τον Πίνακα 6.1 προκύπτει ότι στο διαθέσιμο μεροληπτικό διδιάστατο δείγμα έχουμε καλαθοσφαιριστές με ικανότητα στο άλμα μεγαλύτερη από την αντίστοιχη μέση τιμή που αναφέρεται για τους αθλητές που αγωνίζονται στο NCAA. Όσον αφορά το βάρος παρατηρούμε ότι και εδώ προτιμώνται αθλητές με βαριά κορμιά. Σημειώνουμε ότι το τελευταίο δε σημαίνει ότι προτιμώνται αθλητές με περιττά κιλά, καθώς από τα διαθέσιμα δεδομένα που αφορούν το ύψος και το δείκτη μάζας σώματος, προκύπτει ότι είναι αθλητές γεροδεμένοι, ψηλοί με πολύ καλά αθλητικά χαρακτηριστικά. Σε κάθε περίπτωση η επιλογή των μεταβλητών έγινε για να αποτελέσει μια ξέχωρη εφαρμογή από αυτήν των Economou et al. (2021). Από τα παραπάνω προκύπτει ότι πρέπει να υιοθετήσουμε τη συνάρτηση βάρους $w_1(x, y; \theta, \gamma_x, \gamma_y)$.

Το επόμενο βήμα είναι να επιλεγούν για τις παραμέτρους $\mu_X$, $\mu_Y$, $\sigma_X^2$, $\sigma_Y^2$ και $\rho$ πληροφοριακές εκ των προτέρων κατανομές. Ειδικότερα, ακολουθώντας τους Economou et al. (2021), ως εκ των προτέρων κατανομές για τις $\mu_X$ και $\mu_Y$ επιλέγουμε τις τιμές $N\left(201.28, 24.47^2\right)$ και $N\left(30, 4^2\right)$, αντίστοιχα. Η αντίστροφη γάμμα κατανομή (inverse gamma) χρησιμοποιήθηκε ως εκ των προτέρων κατανομή για τις παραμέτρους $\sigma_X^2$, $\sigma_Y^2$ με παράμετρο σχήματος ίση με 2 και παραμέτρους κλίμακας $24.47^2$ και $4^2$ έτσι ώστε η μέση τιμή της εκ των προτέρων να είναι ίση με τη διακύμανση που θεωρήσαμε στην εκ των προτέρων κατανομή των $\mu_X$ και $\mu_Y$. Για την παράμετρο $\rho$ θεωρήθηκε ο γραμμικός μετασχηματισμός $2W - 1$ της $W \sim beta(25, 30)$, έτσι ώστε η μέση τιμή και η κορυφή να είναι αρνητικές ($-0.09$ και $-0.094$, αντίστοιχα), καθώς αναμένουμε αρνητική συσχέτιση μεταξύ των δύο μεταβλητών. Τέλος, για τις παραμέτρους $\gamma_x$ και $\gamma_y$ επιλέχθηκαν μη πληροφοριακές εκ των προτέρων κατανομές. Καθώς λαμβάνουν θετικές τιμές, περιοριζόμαστε σε αποκομμένες κανονικές κατανομές, που θα είναι τέτοιες ώστε να διερευνηθεί όλος ο παραμετρικός χώρος. Με αυτό ως στόχο χρησιμοποιείται η $TN(1, 10^2)$ και για τις δύο παραμέτρους (όταν αυτές συμμετέχουν στον ορισμό της συνάρτησης βάρους). Με αυτές τις διευκρινήσεις και ακολουθώντας τον αλγόριθμο που δόθηκε στην προηγούμενη ενότητα προσαρμόζονται τα μοντέλα

$\mathcal{M}_1$, $\mathcal{M}_{1y}$ και $\mathcal{M}_{1x}$.

Στον Πίνακα 6.2 παρουσιάζονται τα περιγραφικά στατιστικά των εκ των υστέρων περιθώριων κατανομών των παραμέτρων των τριών μοντέλων χρησιμοποιώντας 5.000 μη απορριπτέες τιμές (δείγματα). Στον ίδιο πίνακα δίνεται η τιμή του κριτηρίου DIC για τα τρία υποψήφια μοντέλα. Από αυτές τις τιμές προκύπτει ότι το μοντέλο $\mathcal{M}_{1y}$ είναι εκείνο με τη μικρότερη τιμή στο κριτήριο DIC. Αυτό ουσιαστικά σημαίνει ότι το μοντέλο αυτό παρουσιάζει την καλύτερη προσαρμογή σε αυτά τα διδιάστατα μεροληπτικά δεδομένα και επιπλέον υποδηλώνει ότι η μεροληψία στην επιλογή των αθλητών του ντραφτ όταν αξιολογείται το βάρος και το επιτόπιο κατακόρυφο άλμα προέρχεται κυρίως από το επιτόπιο κατακόρυφο άλμα.

Πίνακας 6.2: Περιγραφικά στατιστικά μέτρα για τις εκ των υστέρων κατανομές των παραμέτρων της διδιάστατης κανονικής κατανομής υπό τα τρία διαφορετικά υποψήφια μοντέλα. Τα αποτελέσματα για κάθε μοντέλο βασίστηκαν σε 5.000 μη απορριπτέες τιμές.

| Μοντέλο | Περ. Στατ. | $\gamma_x$ | $\gamma_y$ | $\mu_X$ | $\sigma_X$ | $\mu_Y$ | $\sigma_Y$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_1$ | 2.5% | 0.3282 | 0.3808 | 192.3080 | 13.2936 | 30.9953 | 2.0609 | -0.3464 |
| $DIC = -3817.642$ | 25% | 3.2525 | 3.2235 | 204.3937 | 18.3045 | 32.8168 | 2.8474 | -0.1842 |
| | 50% | 6.7167 | 6.9001 | 211.7107 | 22.2489 | 33.9078 | 3.5012 | -0.0975 |
| | 75% | 11.7942 | 11.7966 | 219.1239 | 27.5502 | 35.1050 | 4.3615 | -0.0076 |
| | 97.5% | 22.5216 | 22.7049 | 231.8018 | 41.6440 | 37.2227 | 6.6802 | 0.1545 |
| | Μέση τιμή | 8.1424 | 8.1761 | 211.8147 | 23.6805 | 33.9814 | 3.7311 | -0.0963 |
| | Τυπική Απόκλιση | 6.1573 | 6.1674 | 10.5080 | 7.4272 | 1.6215 | 1.2072 | 0.1290 |
| | Κορυφή | 2.8726 | 2.3896 | 210.8683 | 19.1121 | 33.5476 | 3.1102 | -0.1054 |

| Μοντέλο | Περ. Στατ. | $\gamma_x$ | $\gamma_y$ | $\mu_X$ | $\sigma_X$ | $\mu_Y$ | $\sigma_Y$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_{1x}$ | 2.5% | 0.3169 | | 186.1862 | 14.2155 | 31.2830 | 2.1047 | -0.3403 |
| $DIC = -2261.115$ | 25% | 3.7495 | | 200.1344 | 19.4612 | 33.0134 | 2.8938 | -0.1804 |
| | 50% | 7.8140 | | 208.0372 | 23.7598 | 34.0901 | 3.5454 | -0.0937 |
| | 75% | 12.6134 | | 215.8773 | 29.3664 | 35.2798 | 4.4094 | -0.0055 |
| | 97.5% | 23.8826 | | 228.5236 | 45.0876 | 37.6719 | 6.7311 | 0.1634 |
| | Μέση τιμή | 8.8494 | | 207.9215 | 25.2588 | 34.1934 | 3.7746 | -0.0922 |
| | Τυπική Απόκλιση | 6.3600 | | 11.1563 | 8.0101 | 1.6548 | 1.2073 | 0.1294 |
| | Κορυφή | 2.6535 | | 207.1531 | 21.1893 | 33.9454 | 3.0055 | -0.0929 |

| Μοντέλο | Περ. Στατ. | $\gamma_x$ | $\gamma_y$ | $\mu_X$ | $\sigma_X$ | $\mu_Y$ | $\sigma_Y$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| $\mathcal{M}_{1y}$ | 2.5% | | 0.2434 | 192.0420 | 13.1261 | 29.7894 | 2.2052 | -0.3290 |
| $DIC = -3843.159$ | 25% | | 2.9964 | 204.6234 | 18.1708 | 32.1913 | 3.0623 | -0.1786 |
| | 50% | | 6.7838 | 212.5825 | 22.0679 | 33.3084 | 3.7501 | -0.0948 |
| | 75% | | 11.6878 | 219.8925 | 27.4525 | 34.4521 | 4.6684 | -0.0028 |
| | 97.5% | | 22.9867 | 233.1018 | 41.4593 | 36.7531 | 7.2891 | 0.1676 |
| | Μέση τιμή | | 8.0466 | 212.4031 | 23.5575 | 33.3199 | 4.0157 | -0.0897 |
| | Τυπική Απόκλιση | | 6.2309 | 10.7077 | 7.4366 | 1.7409 | 1.3307 | 0.1292 |
| | Κορυφή | | 2.0761 | 215.1876 | 20.1423 | 33.3951 | 3.2747 | -0.1169 |

Για την υλοποίηση της παραπάνω εφαρμογής χρησιμοποιήθηκε, με κατάλληλες τροποποιήσεις, ο αλγόριθμος που είναι διαθέσιμος από την εργασία των Economou et al. (2021).

# ΚΕΦΑΛΑΙΟ 7

# Επιλογος

Πολλές φορές κατά τη μοντελοποίηση πραγματικών τυχαίων φαινομένων οδηγούμαστε σε περίπλοκα στοχαστικά μοντέλα των οποίων η συνάρτηση πιθανοφάνειας παρότι είναι πλήρως ορισμένη από το πιθανοθεωρητικό μοντέλο, δεν είναι διαθέσιμη σε κλειστή, αναλυτική μορφή ως συνάρτηση της παραμέτρου $\theta$ ή το κόστος για τον υπολογισμό της είναι αρκετά μεγάλο. Οι ABC μέθοδοι που παρουσιάστηκαν στο Κεφάλαιο 4 αντιμετωπίζουν τις παραπάνω, ας μας επιτραπεί η έκφραση, προβληματικές καταστάσεις, καθώς δεν απαιτούν τη γνώση της πιθανοφάνειας, αλλά βασίζονται στη σύγκριση παρατηρούμενων και προσομοιωμένων τιμών. Επομένως, όπως αναφέρουν και οι Turner and Van Zandt (2012), οι ABC μέθοδοι είναι ιδιαίτερα χρήσιμες σε περιπτώσεις περίπλοκων μοντέλων καθώς είναι πολύ εύχρηστες και μπορούν να εφαρμοστούν εύκολα μόλις αναπτυχθεί ο βασικός αλγόριθμος, υπό τη μόνη υπόθεση ότι μπορούμε να παράγουμε δεδομένα από το υπό μελέτη μοντέλο, χωρίς να απαιτείται καμία επιπλέον υπόθεση για τις πιθανοθεωρητικές ιδιότητες των συνιστωσών του μοντέλου. Κατά αυτόν τον τρόπο, οι ABC μέθοδοι μας δίνουν ένα σύνολο τιμών της παραμέτρου $\theta$ που αποτελούν μια προσέγγιση της εκ των υστέρων κατανομής. Το γεγονός ότι είναι προσεγγιστικές μέθοδοι προβληματίζει ορισμένους ερευνητές αλλά θα πρέπει να θυμηθούμε την προτροπή του Tukey (1962), ο οποίος λέει ότι είναι πολύ καλύτερη μια προσεγγιστική απάντηση στη σωστή ερώτηση, παρά μια ακριβής απάντηση στη λάθος ερώτηση. Με άλλα λόγια και στο παρόν πλαίσιο, αν ένα περίπλοκο μοντέλο είναι αυτό που επιθυμούμε για να απαντήσουμε στη σωστή ερώτηση, τότε είναι προτιμότερη μια προσεγγιστική απάντηση από το να ξεγελάσουμε τον εαυτό μας υιοθετώντας ένα απλούστερο μοντέλο που θα επιτρέπει τη χρήση κλασικών (μη προσεγγιστικών) μεθόδων αλλά θα είναι εσφαλμένο.

Οι παραπάνω ιδιότητες δίνουν την ελευθερία στους ερευνητές από διάφορα επιστημονικά πεδία να μπορούν να χρησιμοποιούν σε πρακτικές εφαρμογές περίπλοκα μοντέλα, καθώς πλέον είναι εφικτή η εξαγωγή συμπερασμάτων χωρίς τη χρήση της πιθανοφάνειας. Το πλεονέκτημα αυτό, ταυτόχρονα με τις εξελίξεις στο

χώρο των υπολογιστών και των δυνατοτήτων που αυτοί έχουν, είχε ως συνέπεια να έχει εμφανιστεί στη βιβλιογραφία πληθώρα ερευνητικών εργασιών και να έχει εφαρμοστεί με μεγάλη επιτυχία η ABC μέθοδος σε ένα ευρύ φάσμα πραγματικών προβλημάτων, οδηγώντας σε μια σειρά πρωτοποριακών συμπερασμάτων. Για εργασίες ανασκόπησης σχετικά με τις μεθόδους ABC και τις εφαρμογές τους παραπέμπουμε, μεταξύ άλλων, στους Grelaud et al. (2009), Toni et al. (2009), Beaumont (2010), Bertorelle et al. (2010), Csilléry et al. (2010), Blum et al. (2012), Fearnhead and Prangle (2012), Robert (2016) και Grazian and Fan (2020).

Ενδεικτικά, οι ABC μέθοδοι, εκτός από την πληθυσμιακή γενετική όπου πρωτοεφαρμόσθηκαν, έχουν βρει εφαρμογές στην οικολογία (βλέπε Jabot and Chave (2009)), στη βιολογία-ιατρική-επιδημιολογία (Tanaka et al. (2006), Luciani et al. (2009), Blum and Tran (2009)), στην αρχαιολογία (Wilkinson and Tavaré (2009)). Για περισσότερες αναφορές παραπέμπουμε, μεταξύ άλλων, στους Csillery et al. (2012) και Sisson et al. (2018) και τις εκεί αναφορές.

Στο σημείο αυτό θα πρέπει να επισημανθεί ότι σε αυτούς τους τομείς έχουν ήδη αναπτυχθεί εφαρμογές λογισμικού για συγκεκριμένα προβλήματα. Ενδεικτικά αναφέρουμε τα λογισμικά DIY-ABC και ABCtoolbox, ενώ για περισσότερες πληροφορίες για παρόμοια λογισμικά παραπέμπουμε στις εργασίες των Csillery et al. (2012) και Sunnåker et al. (2013) και τις εκεί αναφορές. Όσον αφορά τη γλώσσα προγραμματισμού R είναι διαθέσιμα τα ακόλουθα πακέτα: EasyABC, abc, abctools, ABCanalysis, abcADM, abc.data και networkABC.

Από όσα προηγήθηκαν σε αυτό αλλά και στα προηγούμενα κεφάλαια είναι προφανές ότι οι ABC μέθοδοι διευρύνουν το πεδίο των μοντέλων για τα οποία μπορεί να διεξαχθεί στατιστική συμπερασματολογία, όντας μαθηματικά θεμελιωμένες και εφαρμόσιμες. Ωστόσο, παρά τη χρησιμότητα τους, είναι τέτοια η φύση των μεθόδων που φέρνουν στην επιφάνεια υπολογιστικές και στατιστικές δυσκολίες και μειονεκτήματα. Ειδικότερα,

- οι ABC μέθοδοι συγκριτικά με τις κλασικές Μπεϋζιανές μεθόδους είναι υπολογιστικά πιο χρονοβόρες. Πιστεύουμε όμως ότι, με τις εξελίξεις στην υπολογιστική δύναμη των υπολογιστών και με τη χρήση ακολουθιακών ABC μεθόδων το πρόβλημα αυτό μπορεί να ξεπεραστεί.
- Συνηθέστερα χρησιμοποιούνται αλγόριθμοι που χρησιμοποιούν συνοπτικά περιγραφικά μέτρα για τον έλεγχο αν τα προσομοιωμένα δεδομένα είναι κοντά στα αληθινά. Η αναγωγή πολυδιάστατων δεδομένων σε δεδομένα μικρότερων διαστάσεων συνεπάγεται ότι χάνουμε ορισμένες πληροφορίες. Επιπρόσθετα, προκύπτουν δύο προβλήματα που αφορούν τόσο την επιλογή

των περιγραφικών μέτρων όσο και τη βαρύτητα που θα δοθεί σε καθένα, καθώς κάποια από αυτά μπορούν να δίνουν περισσότερες πληροφορίες για τις άγνωστες παραμέτρους. Για την αντιμετώπιση αυτού του προβλήματος πρόσφατα δημοσιεύθηκε ένας αυτοματοποιημένος αλγόριθμος επιλογής των βαρών (10.1371/journal.pone.0236954), η απόδοση του οποίου χρήζει περαιτέρω μελέτης.

- Δύο άλλες κύριες δυσκολίες είναι ότι η προσέγγιση της εκ των υστέρων κατανομής επηρεάζεται από την επιλογή της σταθεράς ανεκτικότητας, ενώ επιπλέον η μέθοδος ABC είναι ιδιαίτερα ευαίσθητη στην επιλογή της απόστασης $d$. Τέλος, στις μεθόδους Regression ABC κρίσιμο ρόλο διαδραματίζει η στάθμιση που θα χρησιμοποιηθεί.

- Συμπερασματικά, ως Μπεϋζιανή διαδικασία απαιτεί δεξιότητες ώστε να μετατρέπονται όλες οι πρότερες πεποιθήσεις (γνώσεις), που είναι διαθέσιμες για την άγνωστη παράμετρο $\theta$, σε μια μαθηματικά διατυπωμένη έκφραση της εκ των προτέρων κατανομής.

Οι Προσεγγιστικές Μπεϋζιανές Υπολογιστικές Μέθοδοι δεν είναι οι μόνες που αντιμετωπίζουν προβλήματα στατιστικής συμπερασματολογίας όταν έχουμε intractable μοντέλα, καθώς υπάρχουν και άλλες ιδιαίτερα δημοφιλείς. Ενδεικτικά παραπέμπουμε στις τεχνικές Συνθετικής Πιθανοφάνειας (Synthetic Likelihood, SL) και Εμπειρικής Πιθανοφάνειας (Empirical Likelihood, EL) που αφενός παρουσιάζουν ιδιαίτερο θεωρητικό ενδιαφέρον και μελετώνται στη βιβλιογραφία, αφετέρου είναι ευρέως χρησιμοποιούμενες σε πρακτικές εφαρμογές, διότι μοντελοποιούν και επιλύουν προβλήματα ποικίλων ερευνητικών περιοχών. Για περισσότερες πληροφορίες για αυτές τις τεχνικές παραπέμπουμε, μεταξύ άλλων, στους Wood (2010), Mengersen et al. (2013), Sisson et al. (2018), Drovandi et al. (2018), Price et al. (2018) και Priddle et al. (2019).

# Βιβλιογραφια


[1] Abdessalem, A. B., Dervilis, N., Wagg, D., and Worden, K. Model selection and parameter estimation in structural dynamics using approximate bayesian computation. *Mechanical Systems and Signal Processing 99* (2018), 306–325.

[2] Aeschbacher, S., Beaumont, M. A., and Futschik, A. A novel approach for choosing summary statistics in approximate bayesian computation. *Genetics 192*, 3 (2012), 1027–1047.

[3] Arnold, B. C., and Nagaraja, H. On some properties of bivariate weighted distributions. *Communications in Statistics-Theory and Methods 20*, 5-6 (1991), 1853–1860.

[4] Barker, A. Monte carlo calculations of the radial distribution functions for a proton? electron plasma. *Australian Journal of Physics 18*, 2 (1965), 119–134.

[5] Barker, E., Barker, W., Burr, W., Polk, W., and Smid, M. Recommendation for key management-part 1: General (revision 3). *NIST Special Publication 800–57* (2012).

[6] Bazin, E., Dawson, K. J., and Beaumont, M. A. Likelihood-free inference of population structure and local adaptation in a bayesian hierarchical model. *Genetics 185*, 2 (2010), 587–602.

[7] Beaumont, M. A. Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics 41* (2010), 379–406.

[8] Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. Adaptive approximate bayesian computation. *Biometrika 96*, 4 (2009), 983–990.

[9] Beaumont, M. A., Zhang, W., and Balding, D. J. Approximate bayesian computation in population genetics. *Genetics 162*, 4 (2002), 2025–2035.

[10] Berg, B. A. *Markov chain Monte Carlo simulations and their statistical analysis*. World Scientific, 2004.

[11] Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 1985.

[12] Bertorelle, G., Benazzo, A., and Mona, S. Abc as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular ecology 19*, 13 (2010), 2609–2625.

[13] Biau, G., Cérou, F., and Guyader, A. New insights into approximate bayesian computation. In *Annales de l'IHP Probabilités et statistiques* (2015), vol. 51, pp. 376–403.

[14] Blum, M. Regression approaches for abc. *Handbook of Approximate Bayesian Computation* (2018), 71–85.

[15] Blum, M., Nunes, M., Prangle, D., and Sisson, S. A comparative review of dimension reduction methods in approximate bayesian computation. *arXiv preprint arXiv:1202.3819* (2012).

[16] Blum, M. G., and François, O. Non-linear regression models for approximate bayesian computation. *Statistics and computing 20*, 1 (2010), 63–73.

[17] Blum, M. G., and Tran, V. C. Hiv with contact tracing: a case study in approximate bayesian computation. *Biostatistics 11*, 4 (2009), 644–660.

[18] Buffon, G. Essai darithmétique morale. *Supplémenta lHistoire naturelle 4* (1777).

[19] Burke, N. Metropolis, metropolis-hastings and gibbs sampling algorithms. `https://www.lakeheadu.ca/sites/default/files/uploads/77/Burke.pdf`, 2018.

[20] Burnham, P. K., and Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach (2nd edition)*. Springer-Verlag, New York., 2002.

[21] Cameron, E., and Pettitt, A. Approximate bayesian computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift. *Monthly Notices of the Royal Astronomical Society 425*, 1 (2012), 44–65.

[22] Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. Population monte carlo. *Journal of Computational and Graphical Statistics 13*, 4 (2004), 907–929.

[23] Casella, G., and Berger, R. L. *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA, 2002.

[24] Casella, G., and Robert, C. P. Post-processing accept-reject samples: recycling and rescaling. *Journal of Computational and Graphical Statistics 7*, 2 (1998), 139–157.

[25] Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. Deviance information criteria for missing data models. *Bayesian Anal. 1*, 4 (12 2006), 651–673.

[26] Cemgil, A. T. A tutorial introduction to monte carlo methods, markov chain monte carlo and particle filtering, 2012.

[27] Chen, M.-H., and Schmeiser, B. W. General hit-and-run monte carlo sampling for evaluating multidimensional integrals. *Operations Research Letters 19*, 4 (1996), 161–169.

[28] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. *Monte Carlo methods in Bayesian computation*. Springer-Verlag, 2000.

[29] Claeskens, G., and Hjort, N. L. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge., 2008.

[30] Cornuet, J.-M., Marin, J.-M., Mira, A., and Robert, C. Adaptive multiple importance sampling. *Scandinavian Journal of Statistics 39*, 4 (2012), 798–812.

[31] Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., and Estoup, A. Inferring population history with diy abc: a user-friendly approach to

approximate bayesian computation. *Bioinformatics 24*, 23 (2008), 2713–2719.

[32] Cranmer, K., Brehmer, J., and Louppe, G. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences* (2020).

[33] Creel, M., and Kristensen, D. Abc of sv: Limited information likelihood inference in stochastic volatility jump-diffusion models. *Journal of Empirical Finance 31* (2015), 85–108.

[34] Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution 25*, 7 (2010), 410–418.

[35] Csillery, K., Francois, O., and Blum, M. G. B. abc: an r package for approximate bayesian computation (abc). *Methods in Ecology and Evolution 3*, 3 (2012), 475–479.

[36] Del Moral, P., Doucet, A., and Jasra, A. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*, 3 (2006), 411–436.

[37] Dellaportas, P., Forster, J., and Ntzoufras, I. On bayesian model and variable selection using mcmc. *Statistics and Computing 12* (2002), 27–36.

[38] Didelot, X., Everitt, R. G., Johansen, A. M., and Lawson, D. J. Likelihood-free estimation of model evidence. *Bayesian Analysis 6*, 1 (2011), 49–76.

[39] Diggle, P. J., and Gratton, R. J. Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B (Methodological) 46*, 2 (1984), 193–212.

[40] Drovandi, C. C., Grazian, C., Mengersen, K., and Robert, C. Approximating the likelihood in abc. *Handbook of Approximate Bayesian Computation* (2018), 321–368.

[41] Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. Approximate bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 60*, 3 (2011), 317–337.

[42] Duong, T., Goud, B., and Schauer, K. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences 109*, 22 (2012), 8382–8387.

[43] Economou, P., Batsidis, A., Tzavelas, G., Alexopoulos, P., and Initiative, A. D. N. Berkson's paradox and weighted distributions: An application to alzheimer's disease. *Biometrical Journal 62*, 1 (2020), 238–249.

[44] Economou, P., Batsidis, A., Tzavelas, G., and Malefaki, S. Understanding the sampling bias: A case study on nba drafts. *Journal of Statistical Theory and Practice 15*, 2 (2021), 1–20.

[45] Everitt, R. G., and Rowińska, P. A. Delayed acceptance abc-smc. *Journal of Computational and Graphical Statistics 30*, 1 (2021), 55–66.

[46] Fearnhead, P., and Prangle, D. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*, 3 (2012), 419–474.

[47] Fisher, R. A. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of eugenics 6*, 1 (1934), 13–25.

[48] Frazier, D. T., Robert, C. P., and Rousseau, J. Model misspecification in approximate bayesian computation: consequences and diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2020).

[49] Gamerman, D., and Lopes, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

[50] Gardiner, C. *Stochastic methods*, vol. 4. Springer Berlin, 2009.

[51] Gelfand, A., Gilks, W., Richardson, S., and Spiegelhalter, D. Markov chain monte carlo in practice. *Boca Rotan, FL: Chapman and Hall* (1996), 145–161.

[52] Gelfand, A. E., and Smith, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association 85*, 410 (1990), 398–409.

[53] Gelman, A., Carlin, J. B., Rubin, D. B., and Stern, H. S. *Bayesian data analysis*. Chapman and Hall, 2004.

[54] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. *Bayesian data analysis Third Edition.* CRC press, 2014.

[55] Geman, S., and Geman, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (1984), 721–741.

[56] Geyer, S., Papaioannou, I., and Straub, D. Cross entropy-based importance sampling using gaussian densities revisited. *Structural Safety 76* (2019), 15–27.

[57] Gilks, W. R., Roberts, G. O., and George, E. I. Adaptive direction sampling. *Journal of the Royal Statistical Society: Series D (The Statistician) 43*, 1 (1994), 179–189.

[58] Gordon, N. J., Salmond, D. J., and Smith, A. F. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE proceedings F (radar and signal processing)* (1993), vol. 140, IET, pp. 107–113.

[59] Gourieroux, C., Monfort, A., and Renault, E. Indirect inference. *Journal of applied econometrics 8*, S1 (1993), S85–S118.

[60] Grazian, C., and Fan, Y. A review of approximate bayesian computation methods via density estimation: Inference for simulator-models. *Wiley Interdisciplinary Reviews: Computational Statistics 12*, 4 (2020), e1486.

[61] Grelaud, A., Robert, C. P., Marin, J.-M., Rodolphe, F., Taly, J.-F., et al. Abc likelihood-free methods for model choice in gibbs random fields. *Bayesian Analysis 4*, 2 (2009), 317–335.

[62] Gutmann, M. Tutorial on approximate bayesian computation. `https://bit.ly/3p0F5wQ`, 2016.

[63] Haft, J. A review of basic monte carlo methods. `http://www.reed.edu/economics/parker/s14/312~`, 2014.

[64] Hamilton, G., Currat, M., Ray, N., Heckel, G., Beaumont, M., and Excoffier, L. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics 170*, 1 (2005), 409–417.

[65] Hammersley, J., and Handscomb, D. *Monte Carlo methods.* Chapman and Hall, 1964.

[66] Hartig, F. Explaining the abc-rejection algorithm in r. `https://theoreticalecology.wordpress.com/2014/06/02/explaining-the-abc-rejection-algorithm-in-r/`, 2014.

[67] Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika* (1970).

[68] Hesterberg, T. *Weighted average importance sampling and defensive mixture distributions.* Stanford University. Division of Biostatistics, 1991.

[69] Hitchcock, D. B. A history of the metropolis–hastings algorithm. *The American Statistician 57*, 4 (2003), 254–257.

[70] Iba, Y. Population-based monte carlo algorithms. *Trans. Jpn. Soc. Artif. Intell. 16*, cond-mat/0008226 (2000), 279–286.

[71] Jabot, F., and Chave, J. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecology letters 12*, 3 (2009), 239–248.

[72] Jain, K., and Nanda, A. K. On multivariate weighted distributions. *Communications in Statistics-Theory and Methods 24*, 10 (1995), 2517–2539.

[73] Jeffreys, H. Theory of probability, clarendon press, 1939.

[74] Jensen, J. On the convex functions and inequalities between mean values. *Acta Math 30* (1906), 175–193.

[75] Jespersen, N. S. An introduction to markov chain monte carlo. *Available at SSRN 1594971* (2010).

[76] Johansen, A. M., Evers, L., and Whiteley, N. *Monte carlo methods.* Lecture notes, 2007.

[77] Joyce, P., and Marjoram, P. Approximately sufficient statistics and bayesian computation. *Statistical Applications in Genetics and Molecular Biology 7*, 1 (2008).

[78] Kass, R., and Raftery, A. Bayes factor. *Journal of the American Statistical Association 90* (1995), 773–795.

[79] Kass, R. E. Markov chain monte carlo in practice. *Journal of the American Statistical Association 92*, 440 (1997), 1645–1647.

[80] Knuth, D. E. *The Art of Computer Programming, Vol. 2.* Addison-Wesley, Reading, Mass, 1997.

[81] Kong, A., Liu, J. S., and Wong, W. H. Sequential imputations and bayesian missing data problems. *Journal of the American statistical association 89*, 425 (1994), 278–288.

[82] Konishi, S., and Kitagawa, G. *Information criteria and statistical modeling.* Springer., 2008.

[83] Kullback, S., and Leibler, R. A. On information and sufficiency. *The annals of mathematical statistics 22*, 1 (1951), 79–86.

[84] Landau, D. P., and Binder, K. *A guide to Monte Carlo Simulations in Statistical Physics.* Cambridge University Press, 2014.

[85] Lee, X., Drovandi, C., and Pettitt, A. Model choice problems using approximate bayesian computation with applications to pathogen transmission data sets. *Biometrics 71* (2015), 198207.

[86] Liang, F., Liu, C., and Carroll, R. *Advanced Markov chain Monte Carlo methods: learning from past samples*, vol. 714. John Wiley & Sons, 2011.

[87] Linhart, H., and Zucchini, W. *Model Selection.* Wiley, New York., 1986.

[88] Liu, J. S. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and computing 6*, 2 (1996), 113–119.

[89] Liu, J. S. Markov chain monte carlo and related topics. *Dept. Statist 1348* (1999).

[90] Liu, J. S. *Monte Carlo strategies in scientific computing.* Springer Science & Business Media, 2008.

[91] Liu, J. S., Chen, R., and Wong, W. H. Rejection control and sequential importance sampling. *Journal of the American Statistical Association 93*, 443 (1998), 1022–1031.

[92] Liu, J. S., Liang, F., and Wong, W. H. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association 95*, 449 (2000), 121–134.

[93] Luciani, F., Sisson, S. A., Jiang, H., Francis, A. R., and Tanaka, M. M. The epidemiological fitness cost of drug resistance in mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences 106*, 34 (2009), 14711–14715.

[94] Malefaki, S., and Iliopoulos, G. On convergence of properly weighted samples to the target distribution. *Journal of Statistical Planning and Inference 138*, 4 (2008), 1210–1225.

[95] Malmberg, K. J., Zeelenberg, R., and Shiffrin, R. M. Turning up the noise or turning down the volume? on the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30*, 2 (2004), 540.

[96] Marin, J., Pillai, N., Robert, C., and Rosseau, J. Relevant statistics for bayesian model choice. *J. R. Statist. Soc. B 79* (2014), 833–859.

[97] Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. Approximate bayesian computational methods. *Statistics and Computing 22*, 6 (2012), 1167–1180.

[98] Marin, J.-M., and Robert, C. Importance sampling methods for bayesian discrimination between embedded models. *Frontiers of Statistical Decision Making and Bayesian Analysis* (2010).

[99] Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences 100*, 26 (2003), 15324–15328.

[100] Marjoram, P., and Tavaré, S. Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics 7*, 10 (2006), 759–770.

[101] Marshall, A. W. The use of multistage sampling schemes in monte carlo computations. Tech. rep., RAND CORP SANTA MONICA CALIF, 1956.

[102] Mengersen, K. L., Pudlo, P., and Robert, C. P. Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences 110*, 4 (2013), 1321–1326.

[103] Metropolis, N. The beginning of the monte carlo method. *Los Alamos Science 15* (1987), 125–130.

[104] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The journal of chemical physics 21*, 6 (1953), 1087–1092.

[105] Metropolis, N., and Ulam, S. The monte carlo method. *Journal of the American statistical association 44*, 247 (1949), 335–341.

[106] Navarro, J., Ruiz, J. M., and Del Aguila, Y. Multivariate weighted distributions: a review and some extensions. *Statistics 40*, 1 (2006), 51–64.

[107] Neal, R. M. Slice sampling. *Annals of statistics* (2003), 705–741.

[108] Nosofsky, R. M., Little, D. R., Donkin, C., and Fific, M. Short-term memory scanning viewed as exemplar-based categorization. *Psychological review 118*, 2 (2011), 280.

[109] Nunes, M. A., and Balding, D. J. On optimal selection of summary statistics for approximate bayesian computation. *Statistical applications in genetics and molecular biology 9*, 1 (2010).

[110] Oh, M.-S., and Berger, J. O. Adaptive importance sampling in monte carlo integration. *Journal of Statistical Computation and Simulation 41*, 3-4 (1992), 143–168.

[111] Peters, G. W., Fan, Y., and Sisson, S. A. On sequential monte carlo, partial rejection control and approximate bayesian computation. *Statistics and Computing 22*, 6 (2012), 1209–1222.

[112] Πάττα, Ε. *Μπεϋζιανή επιλογή μεταβλητών με χρήση προσεγγιστικών Μπεϋζιανών μεθόδων*. 2020.

[113] Prangle, D. Summary statistics in approximate bayesian computation. *arXiv preprint arXiv:1512.05633* (2015).

[114] Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics 27*, 1 (2018), 1–11.

[115] Priddle, J. W., Sisson, S. A., Frazier, D. T., and Drovandi, C. Efficient bayesian synthetic likelihood with whitening transformations. *arXiv preprint arXiv:1909.04857* (2019).

[116] Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution 16*, 12 (1999), 1791–1798.

[117] Rao, C. R. On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A* (1965), 311–324.

[118] Richard, L. W., Berg, K., and Baechle, T. Physical and performance characteristics of ncaa division i male basketball players. *The Journal of Strength & Conditioning Research 8*, 4 (1994), 214–218.

[119] Ripley, B. *Stochastic Simulation.* John Wiley, New York, 1987.

[120] Robert, C. The bayesian choice. *Springer-Verlag, New York* (2001).

[121] Robert, C. Approximate bayesian computation: A survey on recent results, 2016.

[122] Robert, C., and Casella, G. *Monte Carlo statistical methods.* Springer Science & Business Media, 2013.

[123] Robert, C., Marin, J.-M., and Pillai, N. S. Why approximate bayesian computational (abc) methods cannot handle model choice problems. *arXiv preprint arXiv:1101.5091* (2011).

[124] Robert, C. P., and Casella, G. *Monte Carlo Statistical Methods.* Springer, 1999.

[125] Robert, C. P., and Casella, G. *Monte Carlo statistical methods.* Springer, 2004.

[126] Robert, C. P., and Casella, G. *Introducing monte carlo methods with R*, vol. 18. Springer, 2010.

[127] Robert, C. P., Cornuet, J.-M., Marin, J.-M., and Pillai, N. S. Lack of confidence in approximate bayesian computation model choice. *Proceedings of the National Academy of Sciences 108*, 37 (2011), 15112–15117.

[128] Roberts, G. O., Gelman, A., Gilks, W. R., et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied probability 7*, 1 (1997), 110–120.

[129] Roberts, G. O., and Rosenthal, J. S. Markov-chain monte carlo: some practical implications of theoretical results. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* (1998), 5–20.

[130] Rubin, D. B. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics* (1984), 1151–1172.

[131] Rubinstein, R. *Simulation and the Monte Carlo Method.* John Wiley, New York, 1981.

[132] Rubinstein, R. Y., and Kroese, D. P. *Simulation and the Monte Carlo method*, vol. 10. John Wiley, 2016.

[133] Sarabia, J. M., and Gómez-Déniz, E. Construction of multivariate distributions: a review of some recent results. *SORT 32*, 1 (2008), 3–36.

[134] Schroeder, D. *Accounting and Causal Effects: Econometric Challenges.* Springer New York, 2010.

[135] Schwarz, G. E. Estimating the dimension of a model. *Annals of Statistics 6* (1978), 461–464.

[136] Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal 27*, 3 (1948), 379–423.

[137] Shimkin, N. 4 importance sampling. `https://rb.gy/azymd0`, 2015.

[138] Sigman, K. Acceptance-rejection method. *Columbia University* (2007).

[139] Sigman, K. Inverse transform method. *Columbia University* (2010).

[140] Sisson, S. A., Fan, Y., and Beaumont, M. *Handbook of approximate Bayesian computation.* CRC Press, 2018.

[141] Sisson, S. A., Fan, Y., and Tanaka, M. M. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences 104*, 6 (2007), 1760–1765.

[142] Smith, R. L. Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research 32*, 6 (1984), 1296–1308.

[143] Sokolowski, J. A., and Banks, C. M. *Principles of modeling and simulation: a multidisciplinary approach.* John Wiley & Sons, 2009.

[144] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*, 4 (2002), 583–639.

[145] Stephens, M. Simple examples of metropolis hastings algorithm. `https://stephens999.github.io/fiveMinuteStats/MH-examples1.html`, 2017.

[146] Sun, L., Lee, C., and Hoeting, J. A. Parameter inference and model selection in deterministic and stochastic dynamical models via approximate bayesian computation: modeling a wildlife epidemic. *Environmetrics 26*, 7 (2015), 451–462.

[147] Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. Approximate bayesian computation. *PLoS Comput Biol 9*, 1 (2013), e1002803.

[148] Tanaka, M. M., Francis, A. R., Luciani, F., and Sisson, S. Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics 173*, 3 (2006), 1511–1520.

[149] Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. Inferring coalescence times from dna sequence data. *Genetics 145*, 2 (1997), 505–518.

[150] Terrell, G. R., and Scott, D. W. Variable kernel density estimation. *The Annals of Statistics* (1992), 1236–1265.

[151] Toni, T., and Stumpf, M. P. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics 26*, 1 (2010), 104–110.

[152] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface 6*, 31 (2009), 187–202.

[153] Tukey, J. W. The future of data analysis. *The Annals of Mathematical Statistics 33*, 1 (1962), 1–67.

[154] Turchin, V. F. On the computation of multidimensional integrals by the monte-carlo method. *Theory of Probability & Its Applications 16*, 4 (1971), 720–724.

[155] Turner, B. M., and Van Zandt, T. A tutorial on approximate bayesian computation. *Journal of Mathematical Psychology 56*, 2 (2012), 69–85.

[156] Van der Vaart, A. W. *Asymptotic statistics*, vol. 3. Cambridge university press, 2000.

[157] Van Dijk, H., and Kloek, T. Experiments with some alternatives for simple importance sampling in monte carlo integration. Tech. rep., Erasmus University Rotterdam, 1984.

[158] Von Neumann, J. Various techniques used in connection with random digits. *Appl. Math Ser 12*, 36-38 (1951), 3.

[159] Walsh, B. Markov chain monte carlo and gibbs sampling. *Lecture notes for EEB 581* (2004).

[160] Wasserman, L. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

[161] Wegmann, D., Leuenberger, C., and Excoffier, L. Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics 182*, 4 (2009), 1207–1218.

[162] Wilkinson, R. Bayesian estimation of primate divergence times. *PhD thesis, University of Cambridge* (2007).

[163] Wilkinson, R. D., and Tavaré, S. Estimating primate divergence times by using conditioned birth-and-death processes. *Theoretical population biology 75*, 4 (2009), 278–285.

[164] Wood, S. N. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature 466*, 7310 (2010), 1102–1104.