



# Εθνικό Μετσόβιο Πολυτεχνείο

*Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών  
Επιστημών*

**Διπλωματική εργασία**

**Στοχαστική προσομοίωση με μεθόδους Markov Chain Monte  
Carlo (MCMC)**

Ταπίρη Ευανθία

**Επιβλέπων Καθηγητής:**  
Φουσκάκης Δημήτριος

Αθήνα  
Ιούλιος 2013



## Ευχαριστίες:

Αρχικά, θα ήθελα να ευχαριστήσω τον κύριο Δημήτρη Φουσκάκη, Επίκουρο Καθηγητή του Ε.Μ.Π., τον οποίο από τα πρώτα εξάμηνα της προπτυχιακής μου πορείας ξεχώρισα ως καθηγητή και άνθρωπο, για τη βοήθεια του και τις πολύτιμες γνώσεις του κατά τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας καθώς και για την καθοδήγηση του όσον αφορά τις μεταπτυχιακές μου σπουδές.

Ακόμη, θα ήθελα να ευχαριστήσω το Θεό για την αγάπη και τη ζωή που μου προσφέρει.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την συνεχή υποστήριξη τους όλα αυτά τα χρόνια και να αφιερώσω αυτήν την εργασία στη μητέρα μου για τη συνεχή της παρουσία και υπομονή.



# Πίνακας Περιεχομένων

---

<b>Περίληψη.....</b>	<b>7</b>
<b>Abstract .....</b>	<b>8</b>
<b>Κεφάλαιο 1 Εισαγωγή Στις Στοχαστικές Ανελίξεις .....</b>	<b>10</b>
1.Βασικές Έννοιες.....	10
2.Βασικές Κατηγορίες Στοχαστικών Ανελίξεων .....	10
3. Μαρκοβιανές Αλυσίδες.....	11
3.1. Ιδιότητες Μαρκοβιανών Αλυσίδων .....	13
4. Στάσιμη Κατανομή Της Αλυσίδας .....	13
4.1 Τυχαίος Περίπατος .....	14
<b>Κεφάλαιο 2 Markov Chain Monte Carlo .....</b>	<b>19</b>
1.Μέθοδοι Δειγματοληψίας .....	19
2.Μέθοδοι Monte Carlo .....	22
2.1 Υπολογισμός ολοκληρωμάτων με τη μέθοδο Monte Carlo.....	22
3.Εισαγωγή στους MCMC .....	24
3.1 Πιθανά προβλήματα των MCMC μεθόδων.....	26
4. Διαγνωστικές μέθοδοι για τις μεθόδους MCMC .....	26
5.Προσαρμοστικές μέθοδοι MCMC .....	28
<b>Κεφάλαιο 3 Ο Αλγόριθμος Metropolis-Hastings.....</b>	<b>30</b>
1.Ο Αλγόριθμος Metropolis .....	30
2.Ο Γενικός Αλγόριθμος Metropolis-Hastings .....	31
3.Σύγκριση Αλγόριθμου Metropolis-Hastings Με Μία Δειγματοληψία Από Ένα Ανεξάρτητο Και Ισόνομο Δείγμα Τυχαίων Μεταβλητών .....	38
4.Σημαντικά χαρακτηριστικά του αλγόριθμου Metropolis-Hastings .....	42
5.Διάφορες Μορφές Του Αλγόριθμου Metropolis-Hastings.....	43
5.1.Ανεξάρτητος Metropolis-Hastings.....	43
5.2.Τυχαίος Περίπατος Metropolis-Hastings .....	46
5.2.1 Μειονεκτήματα Του Τυχαίου Περίπατου .....	52

6.Ποσοστό αποδοχής .....	52
7.Εφαρμογές στην R με τη μέθοδο Metropolis-Hastings .....	57
Εφαρμογή 1: Ανορεξία .....	57
<b>Κεφάλαιο 4 Ο Δειγματολήπτης Gibbs .....</b>	<b>65</b>
1.Εισαγωγή .....	65
2.Ο Δειγματολήπτης Gibbs Δύο Σταδίων .....	66
3.Ο Δειγματολήπτης Gibbs Πολλαπλών Σταδίων .....	77
4.Παράγοντες Που Επηρεάζουν Την Εφαρμογή Του Δειγματολήπτη Gibbs.....	80
1.Εκ Νέου Ρύθμιση Παραμέτρων .....	80
2.Rao-Blackwellization.....	84
3.Metropolis Εντός Του Gibbs Και Υβριδικές Στρατηγικές.....	85
5.Εφαρμογές Στον Δειγματολήπτη Gibbs .....	86
Εφαρμογή 1: Θερμοκρασία Σώματος.....	86
<b>Κεφάλαιο 5 Άλλοι Μαρκοβιανοί Δειγματολήπτες .....</b>	<b>91</b>
1.Ο Δειγματολήπτης Hit-and-Run .....	91
2.Δειγματολήπτης Τμημάτων (Slice Sampler) .....	93
<b>Βιβλιογραφία .....</b>	<b>102</b>



# Περίληψη

---

Σκοπός της διπλωματικής αυτής εργασίας είναι η παρουσίαση των ισχυρών και πλέον γνωστών μεθόδων Markov Chain Monte Carlo (MCMC), που παράγουν κατά προσέγγιση δείγματα από μια κατανομή, με παραδείγματα και εφαρμογές στο στατιστικό πακέτο R.

Στο πρώτο κεφάλαιο γίνεται μια περιληπτική ανάλυση στη θεωρία των στοχαστικών ανελίξεων με αναφορά στις βασικές κατηγορίες στοχαστικών ανελίξεων ενώ δίνεται ιδιαίτερη έμφαση στις Μαρκοβιανές αλυσίδες και στις διάφορες ιδιότητες τους. Ακόμη, αναλύεται η έννοια της στάσιμης συνάρτησης κατανομής ενώ περιλαμβάνονται και κάποια παραδείγματα στην R με σκοπό την πλήρη κατανόηση της έννοιας αυτής.

Στο δεύτερο κεφάλαιο γίνεται μια επιγραμματική αναφορά σε διάφορες μεθόδους δειγματοληψίας ενώ παρουσιάζεται περιληπτικά και η μέθοδος εκτιμητριών Monte Carlo. Ακόμη, γίνεται μία εισαγωγή στις μεθόδους MCMC ενώ παράλληλα παραθέτονται και κάποια πιθανά προβλήματα των μεθόδων αυτών καθώς και ο εντοπισμός αυτών των προβλημάτων με τα κατάλληλα διαγνωστικά εργαλεία.

Στο τρίτο κεφάλαιο πραγματοποιείται η ανάλυση της μεθόδου Metropolis-Hastings που είναι ένα από τα κυριότερα μέλη των αλγόριθμων Markov Chain Monte Carlo. Αρχικά, παρουσιάζονται οι πιο κοινές μορφές του αλγόριθμου Metropolis-Hastings, ανάλογα με την κατανομή εισήγησης που χρησιμοποιείται κάθε φορά, με αρκετά παραδείγματα, ενώ γίνεται και σύγκριση αυτών των μεθόδων με απλούς δειγματολήπτες που παράγουν ανεξάρτητες και ισόνομες τυχαίες τιμές. Ακόμη, σε αυτό το κεφάλαιο συζητείται η βαθμονόμηση του αλγόριθμου μέσω του ποσοστού αποδοχής ενώ τέλος παραθέτονται και κάποιες εφαρμογές του Metropolis-Hastings στην R.

Στο τέταρτο κεφάλαιο αναλύεται επίσης μια ευρεία διαδεδομένη μέθοδος των MCMC, ο δειγματολήπτης Gibbs. Παρουσιάζεται αναλυτικά ο απλός δειγματολήπτης Gibbs δύο σταδίων αλλά και εκείνος με τα πολλαπλά στάδια συνοδευόμενοι από παραδείγματα στην R. Τέλος, δίνεται μια ιδιαίτερη έμφαση στους παράγοντες που εμποδίζουν την αποτελεσματικότητα των αλγόριθμων Gibbs με αντίστοιχα παραδείγματα και εφαρμογές στο στατιστικό πακέτο R.

Τέλος, στο πέμπτο κεφάλαιο γίνεται μια αναφορά σε κάποιους άλλους Μαρκοβιανούς δειγματολήπτες, όπως ο hit-and-run και ο slice sampler, ενώ παραθέτονται οι αντίστοιχοι αλγόριθμοι αυτών των δειγματοληπτών και κάποια παραδείγματα για την καλύτερη κατανόηση τους.



# Abstract

---

The purpose of this dissertation is the presentation of the powerful and well-known methods Markov Chain Monte Carlo (MCMC), which produce samples from an arbitrary distribution, with a number of examples and applications in the statistical package, R.

The first chapter is the summary of the theory of stochastic processes with a reference to the basic categories of stochastic processes with particular emphasis on Markov Chains and their various properties. Moreover, theorems and definitions on the stationary distribution are given including some examples in R in order to become fully comprehensible.

The second chapter seeks a general introduction to various sampling methods while it presents Monte Carlo estimation methods. Furthermore, in this chapter takes place an introduction in MCMC methods and the identification of some potential problems of MCMC algorithms with the appropriate diagnostic tools.

The third chapter includes the analysis of Metropolis-Hastings algorithm which is one of the main members of the Markov Chain Monte Carlo methods. Firstly, the most common versions of Metropolis-Hastings, according to the proposal distribution, are presented with several examples and comparisons with other samplers. Secondly, there is a discussion about the calibration of the algorithm via its acceptance rate. Finally, further applications of various types of Metropolis-Hastings are made.

The fourth chapter presents Gibbs sampler, which might be considered the workhorse of the MCMC world. This chapter covers both the two-stage and the multistage Gibbs samplers with corresponding examples in R. Furthermore, a special emphasis is given on several factors that might hinder the implementation of a Gibbs algorithm while a number of applications are cited in the last section.

Finally, the fifth chapter summarizes two other basic Markov samplers, hit-and-run and slice sampler, with the corresponding algorithms and relevant examples for better understanding.



# Κεφάλαιο 1 Εισαγωγή Στις Στοχαστικές Ανελίξεις

---

Οι στοχαστικές ανελίξεις μελετούν πιθανοθεωρητικά μοντέλα που περιγράφουν τυχαία φαινόμενα τα οποία εξαρτώνται από το χρόνο. Η αντιμετώπιση αυτών των εξαρτήσεων είναι απαραίτητη με σκοπό την κατάληξη σε ένα σωστό αποτέλεσμα. Γι' αυτό το λόγο παρουσιάζουν μεγάλο ενδιαφέρον όχι μόνο από θεωρητική άποψη αλλά και από το ευρύ φάσμα των εφαρμογών τους. Η θεωρία των στοχαστικών ανελίξεων είναι πολύ χρήσιμη στην περιγραφή φυσικών, βιολογικών αλλά και οικονομικών προβλημάτων. Έτσι, η κατανόηση των βασικών εννοιών τους είναι απαραίτητη.

## 1.Βασικές Έννοιες

---

**Ορισμός 1.1:** Στοχαστική ανέλιξη είναι μια συλλογή τυχαίων μεταβλητών  $\{X_t : t \in T\}$ , όπου ο χώρος  $T$ , που συνήθως είναι χρόνος, είναι είτε διακριτός (π.χ.  $T = \{0, 1, 2, \dots\}$ ) είτε συνεχής (π.χ.  $T = [0, \infty)$ ) και καλείται *παραμετρικός χώρος*.

**Ορισμός 1.2:** Οι τιμές που παίρνουν οι μεταβλητές  $X_t$  καλούνται *καταστάσεις* της ανέλιξης και το σύνολο όλων των καταστάσεων λέγεται **χώρος καταστάσεων** και θα τον συμβολίζουμε με  $\mathcal{S}$ , ο οποίος και αυτός μπορεί να είναι είτε διακριτός (π.χ.  $\mathcal{S} = \{0, 1, 2, \dots\}$ ) είτε συνεχής (π.χ.  $\mathcal{S} = \{-\infty, +\infty\}$ ).

**Ορισμός 1.3:** Αν  $X_1 = i$  και  $X_2 = j$  τότε λέμε ότι η ανέλιξη έχει κάνει μια **μετάβαση** από την κατάσταση  $i$  στο βήμα 1 στην κατάσταση  $j$  στο βήμα 2.

Ωστόσο, για να δώσουμε έναν ορισμό μιας στοχαστικής ανέλιξης δεν πρέπει να ξεχνάμε ότι είναι μια πολύ ευρεία έννοια. Δεν υπάρχουν περιορισμοί στο πλήθος των συναρτήσεων κατανομών πιθανότητας που τα διάφορα  $X_t$  μπορούν να ακολουθήσουν. Στις διάφορες εφαρμογές τους η ιδανική επιλογή είναι εκείνη που συνδυάζει μαθηματική απλότητα και πρακτικότητα.

## 2.Βασικές Κατηγορίες Στοχαστικών Ανελίξεων

---

**Ορισμός 1.4:** Μία στοχαστική ανέλιξη (σ.α.)  $\{X_t : t \in T\}$  ονομάζεται **στάσιμη υπό τη στενή έννοια** όταν οι κατανομές πεπερασμένης διάστασης (δηλαδή οι από κοινού συναρτήσεις κατανομών πιθανότητας των τυχαίων μεταβλητών  $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ ) είναι αναλλοίωτες σε χρονικές μεταθέσεις. Δηλαδή όταν για κάθε πεπερασμένο

σύνολο χρονικών στιγμών, έστω  $T_n = \{t_1, \dots, t_n\}$ , η από κοινού κατανομή πιθανότητας των  $X_{t_1}, \dots, X_{t_n}$  συμπίπτει με την από κοινού κατανομή πιθανότητας των  $X_{t_1+s}, \dots, X_{t_n+s} \forall s > 0$ .

**Ορισμός 1.5:** Μία σ.α.  $\{X_t : t \in T\}$  λέγεται **στάσιμη υπό την ευρεία έννοια** όταν η συνάρτηση του μέσου είναι  $\mu(t) = \mu$  και η συνάρτηση της αυτοσυσχέτισης είναι  $g(t, s) = h(|t - s|) \forall s \in T$ .

**Ορισμός 1.6:** Μια στοχαστική ανέλιξη  $\{X_t : t \in T\}$  ονομάζεται **ανεξαρτήτων προσαυξήσεων** εάν για κάθε  $n \in \mathbb{N}$  και κάθε  $t_0 < t_1 < t_2 < \dots < t_n$  οι διαφορές  $Y_{t_j} = X_{t_j} - X_{t_{j-1}}, j = 1, 2, \dots, n$  είναι ανεξάρτητες τυχαίες μεταβλητές.

**Ορισμός 1.7:** **Μαρκοβιανή ανέλιξη** είναι μια στοχαστική ανέλιξη στην οποία από όλο το παρελθόν της μόνο η πιο πρόσφατη κατάσταση καθορίζει το μέλλον της. Δηλαδή μία στοχαστική ανέλιξη  $\{X_t : t \in T\}$  καλείται Μαρκοβιανή αν  $\forall n \in \mathbb{N}$  και  $\forall \{t_1 < t_2, \dots, < t_n\} \subset T$  με  $t_n < t$  και  $A \subset \mathcal{R}$  ισχύει ότι

$$P(X_t \in A | X_{t_1}, X_{t_2}, \dots, X_{t_n}) = P(X_t \in A | X_{t_n}) \quad [1.1]$$

Η σχέση [1.1] καλείται **Μαρκοβιανή ιδιότητα** και αναφέρει στην ουσία ότι για να προβλέψουμε μελλοντικές τιμές αρκεί μόνο να γνωρίζουμε την τρέχουσα τιμή.

### 3. Μαρκοβιανές Αλυσίδες

**Ορισμός 1.8:** Μία στοχαστική ανέλιξη  $\{X_t : t \in T\}$  για την οποία ισχύει η Μαρκοβιανή ιδιότητα [1.1] και έχει χώρο καταστάσεων  $\mathcal{S}$  διακριτό καλείται **Μαρκοβιανή αλυσίδα**.

Θα επικεντρωθούμε κυρίως στις **ομογενείς Μαρκοβιανές αλυσίδες**, στις οποίες η πιθανότητα οποιασδήποτε συγκεκριμένης μετάβασης από το ένα βήμα στο άλλο παραμένει σταθερή στο χρόνο, δηλαδή η πιθανότητα  $P(X_{t+1} = j | X_t = i)$ , όπου  $i, j \in \mathcal{S}$ , να είναι ανεξάρτητη από το χρόνο.

Άρα σύμφωνα με τον πολλαπλασιαστικό κανόνα και με τη σχέση [1.1] θα ισχύει ότι:

$$\begin{aligned} P(X_{t_0}, \dots, X_{t_n}) &= P(X_{t_0})P(X_{t_1}|X_{t_0}), \dots, P(X_{t_n}|X_{t_0}, \dots, X_{t_{n-1}}) \\ &= P(X_{t_0})P(X_{t_1}|X_{t_0}), \dots, P(X_{t_n}|X_{t_{n-1}}) \end{aligned} \quad [1.2]$$

Εφόσον το  $\mathcal{S}$  είναι αριθμήσιμο, αυτές οι δεσμευμένες πιθανότητες ολοκληρώνουν τον λεγόμενο **πυρήνα μετάβασης ή Μαρκοβιανό πίνακα  $P$** .

Π.χ. αν ο χώρος καταστάσεων είναι  $\mathcal{S} = \{0, 1, 2, \dots\}$  τότε ο πίνακας μετάβασης έχει την παρακάτω μορφή:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \dots \\ p_{10} & p_{11} & p_{12} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad [1.3]$$

όπου  $p_{ij} = P(X_{t+1} = j | X_t = i)$  είναι η πιθανότητα μετάβασης από την κατάσταση  $i$  στην κατάσταση  $j$ .

Έτσι, μπορούμε να υπολογίσουμε εύκολα την πιθανότητα μετάβασης 2-βημάτων σύμφωνα με τις λεγόμενες εξισώσεις *Charpman-Kolmogorov*:

$$p_{ij}(2) = \sum_{k \in S} p_{ik} p_{kj} \quad [1.4]$$

για  $k, j \in S$ .

Αυτό σημαίνει ότι μπορούμε να βρούμε τις πιθανότητες 2-βημάτων με τον απλό πολλαπλασιασμό του πίνακα  $P$  και τις πιθανότητες μετάβασης του  $r$  βήματος  $p_{ij}(r)$  από τα στοιχεία του πίνακα  $P^r$ .

Στην περίπτωση των συνεχών τυχαίων μεταβλητών  $X_t$  επειδή δεν μπορούμε να δείξουμε την μεταβατική συμπεριφορά μέσα από ένα πίνακα, εδώ η προσέγγισή μας είναι να δείξουμε πως μια παρατηρούμενη τιμή  $X_{t-1} = s \in S$  στο βήμα  $t-1$  καθορίζει τη δεσμευμένη συνάρτηση πυκνότητας πιθανότητας (σ.π.π)  $f(t|s)$  της  $X_t | X_{t-1} = s$ . Τότε η πιθανότητα μετάβασης ενός βήματος από το σημείο  $s$  στο διάστημα  $T = (t_L, t_U)$  με  $t_L < t_U$  εκφράζεται ως:

$$P\{t_L < X_t < t_U | X_{t-1} = s\} = \int_{t_L}^{t_U} f(t|s) dt.$$

Πολλές φορές αυτή η Μαρκοβιανή εξάρτηση παρουσιάζει μια καλή προσαρμογή πραγματικών δεδομένων και κάποιες φορές όχι. Για παράδειγμα:

- Οι διάφορες δοκιμές σε πραγματικά δεδομένα έχουν αποδείξει ότι η Μαρκοβιανή εξάρτηση δουλεύει καλά στη μοντελοποίηση κάποιων μοτίβων καιρού. Η εξάρτηση του σημερινού καιρού από τον χτεσινό μπορεί να είναι ελάχιστη ή και ανύπαρκτη σε κάποια κλίματα. Σε άλλα, η εξάρτηση είναι μεγαλύτερη αλλά κατάλληλη για να προσαρμοστεί με ένα Μαρκοβιανό μοντέλο.
- Μία Μαρκοβιανή αλυσίδα θα ήταν προτιμότερη από μια αυστηρά ανεξάρτητη διαδικασία για το μοντέλο: ο εργάτης έχει μια χειρονακτική εργασία (0) ή δουλειά γραφείου (1) στο τέλος κάθε μήνα. Αλλά η προσαρμογή στα πραγματικά δεδομένα δε θα ήταν ακριβής, γιατί η ανθρώπινη μνήμη εκτείνεται πολύ περισσότερο από ένα μήνα πίσω. Έτσι για την πρόβλεψη της επιλογής της επόμενης δουλειάς ενός λογιστή (1) που μόλις απολύθηκε, μία Μαρκοβιανή αλυσίδα δε θα μπορούσε να ξεχωρίσει έναν λογιστή που παλιότερα έκανε μόνο δουλειές γραφείου και κάποιον που δούλεψε σαν ηλεκτρολόγος πριν από αρκετά χρόνια.

### 3.1. Ιδιότητες Μαρκοβιανών Αλυσίδων

---

Μια Μαρκοβιανή αλυσίδα ονομάζεται:

- **μη υποβιβάσιμη**, αν ανεξαρτήτως της κατάστασης που ξεκινάει θα επισκεφτεί κάθε άλλη κατάσταση σε πεπερασμένο αριθμό επαναλήψεων με θετική πιθανότητα.
- **απεριοδική**, όταν δεν υπάρχει ακέραιος  $d > 1$  :  
 $p_{ii}^n = P(X_n = i | X_0 = i) = 0$  όταν  $n \neq kd$  με  $k = 1, 2, \dots$
- **περιοδική**, όταν η Μαρκοβιανή αλυσίδα, δεδομένου ότι είναι σε μια συγκεκριμένη κατάσταση, μπορεί να επιστρέψει σε αυτή την κατάσταση μόνο σε συγκεκριμένα τακτά χρονικά διαστήματα.
- **γνήσια επαναληπτική**, όταν για κάθε κατάσταση  $i$  από την οποία ξεκινά η αλυσίδα, επανέρχεται σε αυτή με πιθανότητα 1 σε πεπερασμένο αναμενόμενο χρόνο.
- **ανάγωγη**, αν μπορούμε να πάμε από οποιαδήποτε κατάσταση σε μια άλλη μέσα σε ένα ή περισσότερα βήματα.
- **εργοδική**, αν είναι γνήσια επαναληπτική και απεριοδική. Κάθε εργοδική αλυσίδα έχει μια μοναδική στάσιμη κατανομή  $\pi$ .

### 4. Στάσιμη Κατανομή Της Αλυσίδας

---

Η οριακή συμπεριφορά μιας Μαρκοβιανής αλυσίδας καθώς  $t \rightarrow \infty$  έχει ιδιαίτερη σημασία και είναι σχεδόν πάντα πιο εύκολο να περιγραφεί και να αναλυθεί από ότι η «μεταβατική» συμπεριφορά της αλυσίδας για ένα συγκεκριμένο  $t$ .

**Ορισμός 1.9:** Το διάνυσμα  $\pi$  αποτελεί στάσιμη κατανομή για την αλυσίδα αν  $\pi_j \geq 0$ ,

$$\sum_j \pi_j = 1 \text{ και } \pi = \pi P, \quad [1.5]$$

όπου  $P$  είναι ο πίνακας μετάβασης [1.3].

**Θεώρημα 1.1:** Έστω μια Μαρκοβιανή αλυσίδα με πίνακα μετάβασης  $P$ . Αν υπάρχει στάσιμη κατανομή  $\pi$ , τότε είναι μοναδική και ισχύει:

$$\lim_{t \rightarrow \infty} P^t(i, j) = \pi_j \text{ για } 0 \leq \pi_j \leq 1 \text{ και } \forall i, j \in \mathcal{S} \text{ και } \forall t \geq 0. \quad [1.6]$$

Αν δεν υπάρχει στάσιμη κατανομή, τότε:

$$\lim_{t \rightarrow \infty} P^t(i, j) = 0, \quad \forall i, j \in \mathcal{S} \text{ και } \forall t \geq 0. \quad [1.7]$$

Δηλαδή μία στοχαστική ανέλιξη  $\{X_t : t \in \mathbb{N}\}$  λέγεται *στάσιμη* αν για κάθε θετικό  $\tau, t_0, t_1, \dots, t_n$  το διάνυσμα  $(X_{t_1}, \dots, X_{t_n})$  έχει την ίδια κατανομή με το διάνυσμα  $(X_{t_1+\tau}, \dots, X_{t_n+\tau})$ . Αντίστοιχα ορίζουμε και τη στάσιμη κατανομή όταν το σύνολο δεικτών είναι το  $\mathbb{Z}, \mathbb{R}_+, \mathbb{R}$ .

Αν ισχύει  $\sum_j \pi_j = 1$  τότε μπορούμε να ξαναγράψουμε την σχέση [1.5] σαν ένα σύστημα εξισώσεων, γνωστό και σαν *εξισώσεις ισορροπίας*:

$$\sum_j \pi_i p_{ij} = \sum_j \pi_j p_{ji} \text{ για } \forall i \in \mathcal{S}. \quad [1.8]$$

Η στάσιμη κατανομή χαρακτηρίζει πλήρως τη συμπεριφορά την οποία η αλυσίδα θα αποκτήσει τελικά, αφού έχει τρέξει για αρκετό διάστημα, ανεξαρτήτως της αρχικής κατάστασης.

Άρα, η ιδιότητα της σταθερότητας και η ύπαρξη μιας στάσιμης κατανομής είναι μεγίστης σημασίας στην διαδικασία προσομοίωσης, αφού αν ένας δοσμένος πυρήνας  $K$  δημιουργήσει μία εργοδική Μαρκοβιανή αλυσίδα με στάσιμη κατανομή  $f$  τότε δημιουργώντας μια αλυσίδα από αυτόν τον πυρήνα θα έχω σαν αποτέλεσμα προσομοιωμένες τιμές από την  $f$ .

## 4.1 Τυχαίος Περίπατος

Στη συνέχεια θα παρουσιάσουμε μια στοχαστική ανέλιξη διακριτού χώρου καταστάσεων, τον *τυχαίο περίπατο*, ο οποίος είναι ευρεία διαδεδομένος λόγω των πολλαπλών εφαρμογών του.

Έστω ένα σωματίδιο που κινείται πάνω σε ακεραίους. Ξεκινάει από το 0 και κάνει ένα βήμα δεξιά με πιθανότητα  $p$  ή ένα βήμα αριστερά με πιθανότητα  $q = 1 - p$  δηλαδή αν  $Y_n, n = 1, 2, \dots$  είναι η μετατόπιση του σωματιδίου τη χρονική στιγμή  $n$ , τότε:

$$P(Y_n = 1) = p, \quad P(Y_n = -1) = 1 - p = q.$$

Έστω  $X_n$  η θέση του σωματιδίου αμέσως μετά τη χρονική στιγμή  $n$ . Τα σημεία  $(n, X_n)$ ,  $n = 1, 2, \dots$  παριστάνουν τον τυχαίο περίπατο του σωματιδίου. Δηλαδή μια τέτοια υλοποίηση με  $X_0 = 0$  θα μπορούσε να είναι η εξής:

$$X_n = Y_1 + Y_2 + \dots + Y_n,$$

όπου οι μετατοπίσεις  $Y_i$ ,  $i = 1, 2, \dots, n$  είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές.

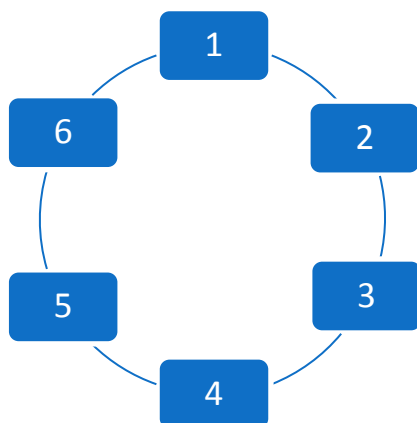
Αν θεωρήσουμε το  $n$  σταθερό, τότε η  $X_n$  είναι μια τυχαία μεταβλητή για την οποία εύκολα μπορούμε να προσδιορίσουμε την κατανομή της. Για παράδειγμα, αν μέχρι και τη χρονική στιγμή  $n$  το σωματίδιο έκανε  $j$  αρνητικές μετατοπίσεις  $j = 0, 1, 2, \dots, n$  τότε η θέση του είναι  $n - 2j$  και αυτό το ενδεχόμενο, δηλαδή οι  $j$  από τις  $n$  μετατοπίσεις να είναι αρνητικές, έχει πιθανότητα:

$$P(X_n = n - 2j) = \binom{n}{j} p^{n-j} q^j.$$

Βλέπουμε, λοιπόν, πως η  $X_n$  ακολουθεί τη Διωνυμική κατανομή. Έτσι, η ακολουθία  $\{X_n, n = 1, 2, \dots\}$  αποτελεί Στοχαστική ανέλιξη διακριτού χώρου καταστάσεων σε διακριτό χρόνο και καλείται *Τυχαίος Περίπατος*.

**Παράδειγμα 1.1:** Έχουμε ήδη αναφέρει πως μία από τις πιο γνωστές διακριτές Μαρκοβιανές αλυσίδες είναι η εφαρμογή του τυχαίου περιπάτου. Έστω, λοιπόν, ότι ένας άνθρωπος κάνει ένα ασυνήθιστο τυχαίο περίπατο στις τιμές 1, 2, 3, 4, 5 και 6 που βρίσκονται πάνω σε ένα κύκλο (Σχήμα 1.1). Αν ο άνθρωπος αυτός βρίσκεται τώρα σε μια συγκεκριμένη τοποθεσία, στο επόμενο δευτερόλεπτο είναι εξίσου πιθανό να μείνει σε αυτή την τοποθεσία ή να μετακινηθεί σε μια γειτονική.

Αν όντως μετακινηθεί, τότε είναι εξίσου πιθανό η μετακίνηση του να είναι είτε δεξιά είτε αριστερά. Αυτό είναι ένα από τα πιο απλά παραδείγματα των διακριτών Μαρκοβιανών αλυσίδων.



**Σχήμα 1.1:** Έξι τοποθεσίες σε έναν κύκλο για το παράδειγμα του τυχαίου περιπάτου.



Έτσι, ο πίνακας μετάβασης  $P$  θα πάρει σε αυτό το παράδειγμα τη μορφή:

$$P = \begin{Bmatrix} 0.50 & 0.25 & 0 & 0 & 0 & 0.25 \\ 0.25 & 0.50 & 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0.50 & 0.25 & 0 & 0 \\ 0 & 0 & 0.25 & 0.50 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 & 0.50 & 0.25 \\ 0.25 & 0 & 0 & 0 & 0.25 & 0.50 \end{Bmatrix}$$

Η πρώτη γραμμή στον πίνακα  $P$  μας δίνει τις πιθανότητες μετάβασης από τη τοποθεσία 1 σε οποιοσδήποτε από τις τοποθεσίες 1 μέχρι 6 με ένα μόνο βήμα, η δεύτερη γραμμή μας δίνει τις πιθανότητες μετάβασης από τη τοποθεσία 2 σε οποιαδήποτε άλλη με ένα βήμα κ.ο.κ. Έτσι, η τρέχουσα τοποθεσία αναπαρίσταται από το διάνυσμα:

$$p = (p_1, p_2, p_3, p_4, p_5, p_6)$$

Για να βεβαιωθούμε, λοιπόν, για την ύπαρξη της στάσιμης κατανομής της Μαρκοβιανής αλυσίδας ξεκινάμε τον τυχαίο μας περίπατο από μία συγκεκριμένη κατάσταση, έστω ότι είναι η 3, και στη συνέχεια προσομοιώνουμε πολλαπλά βήματα της Μαρκοβιανής αλυσίδας χρησιμοποιώντας τον πίνακα μετάβασης  $P$ . Οι σχετικές συχνότητες του ταξιδιώτη στις 6 θέσεις μετά από πολλά βήματα τελικά θα πλησιάσουν τη στατική κατανομή που μας ενδιαφέρει.


Θα χρησιμοποιήσουμε το στατιστικό πακέτο  $R$  για την εφαρμογή όλων των παραπάνω. Στην αρχή γράφουμε μια μικρή συνάρτηση για να προσομοιώσουμε μια διακριτή Μαρκοβιανή αλυσίδα. Οι παράμετροι μας είναι: ο πίνακας μετάβασης  $P$ , η αρχική θέση της αλυσίδας *starting.state* και *steps* ο αριθμός των προσομοιωμένων βημάτων μέσα στην αλυσίδα. Στη συνάρτηση αυτή χρησιμοποιείται και η *sample* συνάρτηση μαζί με την κατάλληλη γραμμή του πίνακα  $P$  για να κάνει τη δειγματοληψία.

```
>simulate.markov.chain=function(P,starting.state,steps){
+ n.states=dim(P)[1]
+ state=starting.state
+ STATE=rep(0,steps)
+ for(j in 1:steps){
+ state=sample(n.states,size=1,prob=P[state, ])
+ STATE[j]=state }
+ return(STATE)
+ }
> P=matrix(c(0.50, 0.25, 0, 0, 0,0.25,
+ 0.25, 0.50, 0.25, 0, 0, 0,
+ 0, 0.25, 0.50, 0.25, 0, 0,
+ 0, 0, 0.25, 0.50, 0.25, 0,
+ 0, 0, 0, 0.25, 0.50, 0.25,
+ 0.25, 0, 0, 0, 0.25, 0.50),
+ nrow=6,ncol=6,byrow=TRUE)
```



Στη συνέχεια προσομοιώνουμε 10.000 τιμές από μία Μαρκοβιανή αλυσίδα μέσω της συνάρτησης `simulate.markov.chain` με τον πίνακα μετάβασης  $P$ , με αρχική κατάσταση 3 και με 10.000 βήματα και τα καταχωρούμε σε ένα πίνακα. Τέλος, τα μετατρέπουμε και σε σχετικές συχνότητες διαιρώντας το αποτέλεσμα του πίνακα με τον αριθμό των επαναλήψεων.

Αυτές, λοιπόν, οι σχετικές συχνότητες προσεγγιστικά αναπαριστούν τη στάσιμη κατανομή της Μαρκοβιανής αλυσίδας. Μπορεί να αποδειχτεί, χρησιμοποιώντας ένα ξεχωριστό υπολογισμό, ότι η στάσιμη κατανομή αυτής της αλυσίδας είναι ομοιόμορφη και στις 6 καταστάσεις. Τώρα, αν θέλουμε να επιβεβαιώσουμε ότι όντως το διάνυσμα αυτό είναι η στάσιμη κατανομή μπορούμε απλά να πολλαπλασιάσουμε τον πίνακα μετάβασης  $P$  με το διάνυσμα της στάσιμης κατανομής μας. Όλα τα παραπάνω αναπαρίστανται στην R ως εξής:



```
> s=simulate.markov.chain(P,3,10000)
> table(s)
s
 1  2  3  4  5  6
1728 1642 1530 1532 1701 1867
> table(s)/10000
s
 1  2  3  4  5  6
0.1728 0.1642 0.1530 0.1532 0.1701 0.1867
> w=c(1,1,1,1,1,1)/6
> w
[1] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
> w%*%P
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
```

■

Στην περίπτωση του απείρου αριθμήσιμου χώρου καταστάσεων, όπως  $\mathcal{S} = \{0,1,2, \dots\}$  ή  $\mathcal{S} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ , μπορούμε και πάλι να χρησιμοποιήσουμε τις εξισώσεις Chapman-Kolmogorov [1.4] για τον υπολογισμό των πιθανοτήτων μετάβασης αλλά το άθροισμα τώρα θα είναι μία άπειρη σειρά.

Η διαφορά των άπειρων Μαρκοβιανών αλυσίδων με τις πεπερασμένες έγκειται κυρίως στο γεγονός ότι ενώ μια άπειρη αλυσίδα μπορεί να κινηθεί και να περάσει από όλες τις καταστάσεις της με πιθανότητα θετική αυτό δεν σημαίνει απαραίτητα και την ύπαρξη μιας στάσιμης κατανομής. Παρόλα αυτά, αν στις άπειρες συγκλίνουσες αλυσίδες οι ακραίες καταστάσεις τους είναι σχετικά απίθανο να πραγματοποιηθούν τότε η αλυσίδα δεν θα μπορεί να διαφύγει προς το «άπειρο» σε

καμία από τις 2 κατευθύνσεις και έτσι αυτό συνεπάγεται την ύπαρξη της στάσιμης κατανομής.

Στην περίπτωση του *συνεχούς χώρου καταστάσεων*  $\mathcal{S}$ , αν η Μαρκοβιανή αλυσίδα είναι ένα πεπερασμένο διάστημα τότε η κατανομή των  $X_t$  δεν μπορεί να αποδράσει στο άπειρο μακροπρόθεσμα. Ωστόσο, ένας πεπερασμένος χώρος καταστάσεων δεν συνεπάγεται και μία χρήσιμη για το εκάστοτε ζητούμενο πρόβλημα στάσιμη κατανομή. Για την προσομοίωση μιας Μαρκοβιανής αλυσίδας στην πράξη, πρέπει να βασίζομαι σε περιγραφικές διαγνωστικές μεθόδους έτσι ώστε να ξέρω με σιγουριά ότι οι εμφανείς στάσιμες κατανομές είναι όντως οι ζητούμενες.

Άρα, απαραίτητη προϋπόθεση όλων των παραπάνω είναι η Μαρκοβιανή αλυσίδα να συγκλίνει σε μια χρήσιμη για το εκάστοτε πρόβλημα στάσιμη κατανομή. Παρόλα αυτά, ακόμα κι αν η αλυσίδα συγκλίνει σε μια χρήσιμη στατική κατανομή υπάρχουν κάποια ερωτήματα στα οποία πρέπει να δώσω απάντηση όπως το πόσο πρέπει να περιμένω για να επιτευχθεί η στασιμότητα, το χρόνο που πρέπει να τρέξουμε την αλυσίδα μετά τη στασιμότητα και ποια θα είναι η τιμή εκκίνησης. Όλα αυτά τα ερωτήματα απαντώνται με τη βοήθεια των διαγνωστικών εργαλείων των μεθόδων MCMC, μεθόδους τις οποίες θα αναλύσουμε στο επόμενο κεφάλαιο.

## Κεφάλαιο 2 Markov Chain Monte Carlo

### 1. Μέθοδοι Δειγματοληψίας

Στα Μαθηματικά, και πιο συγκεκριμένα στη Στατιστική, όλη την πληροφορία την αποκτούμε από το δείγμα μας. Έτσι, οι μέθοδοι δειγματοληψίας είναι πολύ σημαντικοί για την κατάληξη σε μία σωστή και έγκυρη εκτίμηση. Η καλύτερη δυνατή περίπτωση είναι να έχουμε ανεξάρτητες πηγές πληροφοριών, δηλαδή ανεξάρτητες και ισόνομες τυχαίες τιμές. Αυτό το επιτυγχάνουμε μέσω της παραγωγής τιμών από μια γνωστή κατανομή με απλή μορφή.

Ένας εναλλακτικός τρόπος δειγματοληψίας για ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα είναι η μέθοδος Αποδοχής-Απόρριψης. Δηλαδή, όταν αποτυγχάνουμε να παράγουμε ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με άμεσο τρόπο από μια κατανομή, όπως περιγράψαμε παραπάνω, παράγουμε μια υποψήφια τυχαία μεταβλητή και την αποδεχόμαστε μόνο αν ικανοποιεί έναν κανόνα. Σε αυτές τις μεθόδους το μόνο που χρειάζεται να γνωρίζουμε είναι η μορφή της συνάρτησης. Χρησιμοποιούμε μια πιο απλή συνάρτηση  $G$  (φάκελος) και αφού την κανονικοποιήσουμε προκύπτει η  $g$  (που την ονομάζουμε κατανομή εισήγησης).

Οι μόνες προϋποθέσεις που έχουμε για την κατανομή εισήγησης  $g$  είναι:

1. Οι  $f$  και  $g$  να έχουν συμβατά στηρίγματα (π.χ.  $g(x) > 0$  όταν  $f(x) > 0$ ). Με τον όρο *στήριγμα* εννοούμε την περιοχή στην οποία βρίσκεται συσσωρευμένη η μεγαλύτερη πιθανότητα της συνάρτησης μας.
2. Υπάρχει μια σταθερά  $M$  με  $\frac{f(x)}{g(x)} \leq M \quad \forall x$ .

Σε αυτήν την περίπτωση το  $x$  προσομοιώνεται με τον ακόλουθο τρόπο:

Πρώτα παράγουμε  $Y \sim g$  και, ανεξάρτητα, παράγουμε  $U \sim \text{Uniform}_{[0,1]}$ . Αν

$$U \leq \frac{1}{M} \frac{f(Y)}{g(Y)},$$

τότε θέτουμε  $X = Y$ . Αν δεν ικανοποιείται η ανισότητα τότε απορρίπτουμε την  $Y$  και την  $U$  και ξεκινάμε από την αρχή. Επιγραμματικά, ο αλγόριθμος Αποδοχής-Απόρριψης παρουσιάζεται παρακάτω:

**Αλγόριθμος 1: Μέθοδος Αποδοχής-Απόρριψης**

1. Παρήγαγε  $Y \sim g$ ,  $U \sim \mathcal{U}_{[0,1]}$ ;
2. Αποδέξου  $X = Y$  αν  $U \leq f(Y)/Mg(Y)$ ;
3. Αλλιώς, γύρισε στο βήμα 1.

Άρα, για να παράγουμε τυχαίους αριθμούς από την  $f$ , παράγουμε τυχαίους αριθμούς από την κατανομή εισήγησης  $g$  και τους δεχόμαστε ή τους απορρίπτουμε (μέχρι να αποδεχτούμε κάποιον άλλον αριθμό) με μια συγκεκριμένη πιθανότητα αποδοχής  $f(y|x) / G(y|x)$ .

**Παράδειγμα 2.1:** Χρησιμοποιούμε έναν αλγόριθμο Αποδοχής-Απόρριψης για να παράγουμε ένα δείγμα από τη *Beta* (2.7,6.3) χρησιμοποιώντας για κατανομή εισήγησης την ομοιόμορφη κατανομή  $Uniform_{[0,1]}$ .

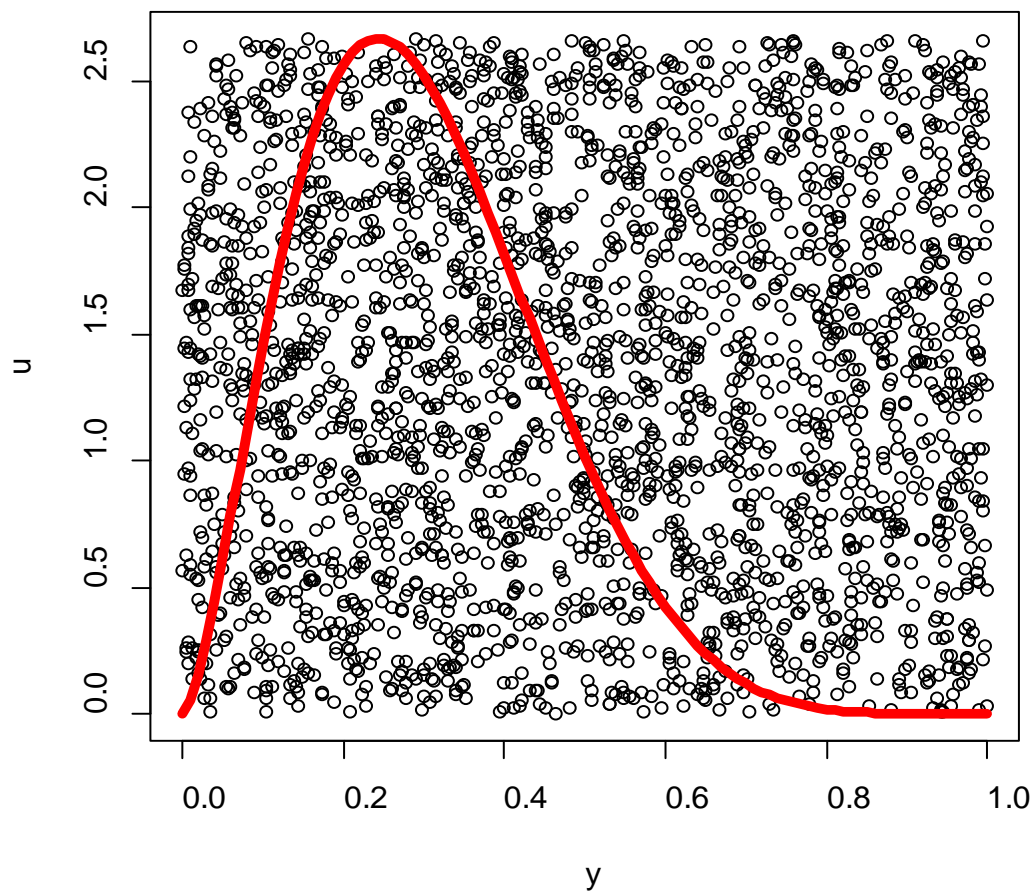
Το άνω φράγμα  $M$  είναι τότε το μέγιστο που λαμβάνεται με την εντολή της βελτιστοποίησης στην R *optimize*. Αφού η κατανομή εισήγησης  $g$  είναι ίση με 1 για κάθε  $x$ , η υποψήφια τιμή  $Y$  είναι αποδεκτή αν  $M \times U < f(Y)$ , δηλαδή αν  $M \times U$  είναι κάτω από τη συνάρτηση πυκνότητας πιθανότητας στο αντίστοιχο σχήμα αυτής της υλοποίησης.

Ας σημειώσουμε ότι η παραγωγή  $U \sim Uniform_{[0,1]}$  πολλαπλασιασμένη με το  $M$  είναι ίδια με την παραγωγή  $U \sim Uniform_{[0,M]}$ .

Για  $\alpha = 2.7$  και  $\beta = 6.3$  μια εφαρμογή της Μεθόδου Αποδοχής-Απόρριψης είναι:

```
> Nsim=2500
> optimize(f=function(x){dbeta(x,2.7,6.3)},
+ interval=c(0,1),maximum=TRUE)$objective
[1] 2.669744
> a=2.7
> b=6.3
> M=2.669
> u=runif(Nsim,max=M) #uniform over (0,M)
> y=runif(Nsim) #generation from g
> x=y[u<dbeta(y,a,b)] #accepted subsample
> plot(y,u)
> curve(dbeta(x,2.7,6.3), add=TRUE, lwd=5,col="red")
```





**Σχήμα 2.1:** Παραγωγή τυχαίων μεταβλητών από τη Βήτα κατανομή  $X \sim \text{Beta}(2.7, 6.3)$  χρησιμοποιώντας τη μέθοδο Αποδοχής-Απόρριψης με 2500 υποψήφιες τυχαίες μεταβλητές  $(Y, U)$  που δημιουργήθηκαν από την  $g$  και την  $U_{[0,M]}$  αντίστοιχα.

Τα σημεία κάτω από τη κόκκινη γραμμή της σ.π.π. είναι εκείνα που αποδεχόμαστε ενώ τα σημεία εκτός της σ.π.π. είναι εκείνα που απορρίπτουμε. Βλέπουμε και πάλι ξεκάθαρα από το Σχήμα 2.1 ότι τα σημεία είναι ομοιόμορφα κατανεμημένα και σύμφωνα με την πιθανότητα αποδοχής το

$$\frac{1}{M} = \frac{1}{2.67} = 37\%$$

των συνολικών τιμών είναι οι αποδεκτές τιμές μας.

■

Βέβαια, υπάρχει πάντα η περίπτωση να μην μπορούμε να χρησιμοποιήσουμε δειγματοληψία που να παράγει ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα. Στις περιπτώσεις με πολυδιάστατες συναρτήσεις και περίπλοκες συννοριακές τιμές είναι

αδύνατη η εφαρμογή των παραπάνω μεθόδων και έτσι είναι απαραίτητη η χρήση των Markov Chain Monte Carlo. Πριν εξετάσουμε, όμως, τους αλγόριθμους MCMC, θα αναφερθούμε περιληπτικά στις μεθόδους Monte Carlo.

## 2. Μέθοδοι Monte Carlo

---

Δύο είναι οι κύριες κλάσεις αριθμητικών προβλημάτων που προκύπτουν στη στατιστική συμπερασματολογία: τα προβλήματα βελτιστοποίησης και τα ολοκληρωτικά προβλήματα. Όντως, ένας μεγάλος αριθμός παραδειγμάτων μπορεί να αποδείξει πως δεν είναι πάντα εφικτός ο υπολογισμός διαφόρων εκτιμητριών. Έτσι, αναπόφευκτα οδηγούμαστε στην εξέταση αριθμητικών λύσεων όπως η προσομοίωση που στοχεύει στον υπολογισμό ποσοτήτων που μας ενδιαφέρουν.

Οι μέθοδοι Monte Carlo είναι μια μεγάλη κλάση υπολογιστικών αλγόριθμων που βασίζονται στην επαναλαμβανόμενη τυχαία δειγματοληψία με σκοπό την ανάκτηση αριθμητικών αποτελεσμάτων όπως για παράδειγμα με το «τρέξιμο» πολλών προσομοιώσεων πολλές φορές έτσι ώστε να υπολογίσουμε τις ίδιες πιθανότητες επαγωγικά όπως ακριβώς παίζουμε και καταγράφουμε τα αποτελέσματα σε ένα καζίνο. Από αυτήν την διαδικασία πήραν και οι μέθοδοι αυτοί το όνομα τους.

Χρησιμοποιούνται συχνά σε προβλήματα φυσικής και μαθηματικών και εφαρμόζονται συνήθως σε περιπτώσεις που είναι δύσκολη η χρήση μιας δειγματοληψίας που παράγει ένα τυχαίο και ανεξάρτητο δείγμα. Οι μέθοδοι αυτοί, είναι ευρεία διαδεδομένοι στα μαθηματικά, ειδικά στον υπολογισμό πολυδιάστατων ορισμένων ολοκληρωμάτων αλλά και σε φυσικά προβλήματα, όπως για παράδειγμα στην εξερεύνηση του διαστήματος.

Η σύγχρονη εκδοχή της μεθόδου Monte Carlo εφευρέθηκε στα τέλη της δεκαετίας του 1940 από τον Stanislaw Ulam κατά τη διάρκεια του Β Παγκοσμίου πολέμου. Πήρε το όνομα της, από τον Nicholas Metropolis, μετά το καζίνο του Monte Carlo, εμπνευσμένος από την ιδιαίτερη αδυναμία του Ulam στο πόκερ. Αυτή η ιδέα εγκρίθηκε και τέθηκε σε εφαρμογή από τον John von Neumann ο οποίος τη χρησιμοποίησε και στον πρώτο ηλεκτρονικό υπολογιστή, ENIAC.

Γενικά, η κύρια αρχή των μεθόδων Monte Carlo είναι: **οτιδήποτε χρειάζεται να γνωρίζουμε για μια τυχαία μεταβλητή  $X$  μπορούμε να το μάθουμε μέσω της επαναλαμβανόμενης δειγματοληψίας από την  $f(X)$ , δηλαδή τη συνάρτηση κατανομής της τ.μ.  $X$ .**

### 2.1 Υπολογισμός ολοκληρωμάτων με τη μέθοδο Monte Carlo

---

Ένα γενικό πρόβλημα στις πιθανότητες αλλά και στις στατιστικές εφαρμογές είναι ο υπολογισμός της αναμενόμενης τιμής μιας τυχαίας μεταβλητής. Για αυτό θα

αναφερθούμε πολύ περιληπτικά στις μεθόδους υπολογισμού ολοκληρωμάτων με Monte Carlo.

Ας υποθέσουμε, λοιπόν, ότι η τυχαία μεταβλητή  $X$  έχει μια συνάρτηση πυκνότητας πιθανότητας  $f(x)$  και εμείς ενδιαφερόμαστε στο να υπολογίσουμε την αναμενόμενη τιμή της  $h(x)$  που δίνεται από τη σχέση:

$$E(h(X)) = \int h(x)f(x)dx. \quad [2.1]$$

Στη συνέχεια υπενθυμίζουμε ένα πολύ βασικό θεώρημα σύγκλισης το οποίο χρησιμοποιείται στις μεθόδους εκτιμήσεων Monte Carlo.

**Θεώρημα 2.1 (Ασθενής Νόμος των Μεγάλων Αριθμών):** Έστω  $\{X_n, n = 1, 2, \dots\}$  ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με μέση τιμή  $\mathbb{E}(X) = \mu$ . Τότε,

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \varepsilon\right) \rightarrow 0 \text{ καθώς } n \rightarrow \infty \quad \forall \varepsilon > 0, \quad [2.2]$$

δηλαδή η  $\bar{X}_n$  θα συγκλίνει κατά πιθανότητα στη μέση τιμή  $\mu$ .

Έτσι, αν υποθέσουμε ότι μπορούμε να προσομοιώσουμε ένα τυχαίο δείγμα τιμών  $x_1, \dots, x_n$  από την  $f(x)$ , τότε η εκτίμηση  $E(h(x))$  με τη μέθοδο Monte Carlo δίνεται από την ποσότητα

$$\bar{h} = \frac{\sum_{j=1}^m h(x_j)}{m}. \quad [2.3]$$

Σύμφωνα και με το Θεώρημα 2.1 συμπεραίνουμε από τον Ασθενή Νόμο των Μεγάλων Αριθμών ότι καθώς το  $m \rightarrow \infty$ ,  $\bar{h} \rightarrow E(h(x))$  κατά πιθανότητα. Το  $E(h(x))$  δεν είναι τυχαίο αλλά το  $h$  είναι και άρα το  $\bar{h}$  είναι ένας δειγματικός εκτιμητής της αναμενόμενης τιμής, έτσι θα υπάρχει κάποιο σφάλμα σε αυτόν το προσομοιωμένο υπολογισμό. Η διασπορά του  $h$  δίνεται από τη παρακάτω σχέση:

$$Var(\bar{h}) = \frac{Var(h(X))}{m}. \quad [2.4]$$

Για την εκτίμηση της  $Var(h(X))$  χρησιμοποιούμε τη δειγματική διασπορά των προσομοιωμένων τιμών  $(h(x_j))$ :

$$\widehat{Var(h(X))} = \frac{\sum_{j=1}^m (h(x_j) - \bar{h})^2}{m - 1}. \quad [2.5]$$

Το τυπικό σφάλμα της εκτίμησης Monte Carlo δίνεται τότε από την παρακάτω σχέση:



$$se_{\bar{h}} = \sqrt{Var(\bar{h})} \approx \sqrt{\frac{Var(\bar{h}(X))}{m}}. \quad [2.6]$$

Δύο από τα κυριότερα χαρακτηριστικά της μεθόδου Monte Carlo είναι:

1. Στις μεθόδους αυτές είναι δυνατόν να εκτιμηθεί η τάξη μεγέθους του στατιστικού σφάλματος, σφάλμα το οποίο προκύπτει στους περισσότερους υπολογισμούς με τη χρήση αυτών των μεθόδων.
2. Οι Monte Carlo βασίζονται στην ιδέα ότι οι απλές έξυπνες ιδέες μπορούν να οδηγήσουν πρακτικά σε μεγάλη βελτίωση ειδικά όσον αφορά την αποτελεσματικότητα και την ακρίβεια.

Λόγω των παραπάνω χαρακτηριστικών υπάρχουν κάποια συγκεκριμένα πεδία στα οποία είναι απαραίτητη έστω και η ελάχιστη γνώση αυτών των μεθόδων όπως:

- Εξόρυξη δεδομένων και πληροφορική
- Βάσεις δεδομένων μικρού και μεγάλου μεγέθους
- Μπεϋζιανές μεθόδους
- Βιολογία και γενετική έρευνα

### 3.Εισαγωγή στους MCMC

Όπως αναφέραμε και στο προηγούμενο κεφάλαιο οι εργοδικές Μαρκοβιανές αλυσίδες συχνά χρησιμοποιούνται σαν βοήθεια στην προσομοίωση περίπλοκων μοντέλων πιθανότητας. Τέτοιες μέθοδοι προσομοίωσης συχνά ονομάζονται Markov Chain Monte Carlo (MCMC), που στην ουσία είναι ο συνδυασμός των μεθόδων Monte Carlo με τις Μαρκοβιανές αλυσίδες.

Βασικό θεώρημα σε αυτή την κλάση των μεθόδων είναι το Εργοδικό Θεώρημα το οποίο παρατίθεται παρακάτω:

**Εργοδικό Θεώρημα 2.1:** Αν μια Μαρκοβιανή αλυσίδα είναι εργοδική με στάσιμη κατανομή  $\pi$  και  $f$  είναι μια οποιαδήποτε πραγματική συνάρτηση με  $E_{\pi}[|f(X)|] < \infty$  τότε με πιθανότητα 1 καθώς  $m \rightarrow \infty$

$$\frac{1}{m} \sum_{i=1}^m f(X_i) \rightarrow E_{\pi}[f(X)] = \sum_i f(i)\pi(i) \quad [2.7]$$

Άρα, μπορούμε να συμπεράνουμε ότι ο Ασθενής Νόμος των Μεγάλων Αριθμών που βρίσκεται στη βάση των μεθόδων Monte Carlo μπορεί να εφαρμοστεί στους MCMC και αυτό μας επιτρέπει να αγνοήσουμε την εξάρτηση μεταξύ των τιμών της Μαρκοβιανής Αλυσίδας όταν υπολογίζουμε από τις τιμές αυτές τις ποσότητες που μας ενδιαφέρουν. Δηλαδή αν γνωρίζουμε τη στάσιμη κατανομή της αλυσίδας, τότε μπορούμε να μάθουμε σημαντικές πληροφορίες όπως π.χ. ο πληθυσμιακός μέσος,

η τυπική απόκλιση κ.α. απλά περιμένοντας να συγκλίνει η αλυσίδα, δηλαδή να επιτευχθεί η στασιμότητα.

Στην ουσία, οι MCMC μέθοδοι εξερευνούν το χώρο καταστάσεων  $\mathcal{S}$  ενώ παράλληλα χρησιμοποιούν τις ιδιότητες της Μαρκοβιανής αλυσίδας με σκοπό την παραγωγή δείγματος. Το ζητούμενο μας είναι οι τιμές του δείγματος να «μιμούνται» τις τιμές από τη ζητούμενη συνάρτηση κατανομής.

Έτσι, οι MCMC μέθοδοι επιτυγχάνουν την προσομοίωση από περίπλοκες πολυμεταβλητές κατανομές. Η κύρια αρχή των μεθόδων αυτών είναι αρκετά εύκολο να περιγραφεί. Συνοπτικά, δεδομένου ότι ξέρουμε μια συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) ή μια συνάρτηση μάζας πιθανότητας (σ.μ.π.), έστω  $f$ , φτιάχνουμε ένα Μαρκοβιανό πυρήνα  $K$  με στάσιμη κατανομή  $f$  και μετά παράγουμε μια Μαρκοβιανή αλυσίδα  $\{X_t\}$  χρησιμοποιώντας αυτόν τον πυρήνα έτσι ώστε η στάσιμη κατανομή της αλυσίδας να είναι η  $f$ . Η όλη δυσκολία έγκειται στην κατασκευή του πυρήνα  $K$  που συνδέεται με την επιλογή της  $f$ . Παρόλα αυτά υπάρχουν μέθοδοι που προσδιορίζουν τέτοιους πυρήνες και ισχύουν θεωρητικά για οποιαδήποτε συνάρτηση πυκνότητας  $f$ .

Όπως περιγράφηκε ήδη, οι μέθοδοι Monte Carlo παράγουν δείγματα από τη ζητούμενη συνάρτηση και στη συνέχεια εκτιμούν διάφορες ποσότητες. Οι MCMC αλγόριθμοι παράγουν αυτά τα δείγματα με το «τρέξιμο» μιας κατάλληλης Μαρκοβιανής αλυσίδας για κάποιο χρονικό διάστημα. Υπάρχουν πολλοί τρόποι για την κατασκευή αυτής της αλυσίδας αλλά οι δύο κυριότεροι και πιο γνωστοί είναι: ο αλγόριθμος Metropolis-Hastings και ο δειγματολήπτης Gibbs, στους οποίους θα αναφερθούμε αναλυτικά στα επόμενα κεφάλαια. Ωστόσο, για την απλή κατανόηση όλων των παραπάνω, θα αναφερθούμε σε ένα παράδειγμα μιας συγκεκριμένης μεθόδου MCMC.

Η εταιρία Google χρησιμοποιεί έναν Τυχαίο Περίπατο Metropolis-Hastings (βλ. Κεφάλαιο 3.5) για να υπολογίσει μια τιμή που ονομάζεται PageRank (κατάταξη σελίδας) για κάθε είσοδο που πραγματοποιείται στη μηχανή αναζήτησης της. Η τιμή αυτή μετράει ουσιαστικά πόσο σημαντική είναι και πόση αξία έχει μια ιστοσελίδα και καθορίζει τη κατάταξη της κάθε ιστοσελίδας όταν πραγματοποιείται η αναζήτηση. Εφόσον, η επόμενη ιστοσελίδα που θα επισκεφθεί ο τυχαίος χρήστης είναι επίσης τυχαία και εξαρτάται μόνο από την προηγούμενη σελίδα που έχει ήδη επισκεφθεί, αυτή η διαδικασία υποδηλώνει μια Μαρκοβιανή αλυσίδα και εξαιτίας της αδυναμίας άλλων μεθόδων να παράγουν τυχαίο δείγμα σε αυτήν τη περίπτωση, έτσι είναι λογική η χρήση MCMC μεθόδων και συγκεκριμένα του αλγόριθμου του Τυχαίου Περίπατου Metropolis-Hastings. Δηλαδή, απλά πληκτρολογώντας μια τυχαία λέξη στη μηχανή αναζήτησης και επιλέγοντας τυχαία ένα από τα αποτελέσματα μεταφερόμαστε σε μία άλλη ιστοσελίδα από την οποία μέσω ενός συνδέσμου μεταφερόμαστε σε μια άλλη κ.ο.κ. Πραγματοποιώντας την παραπάνω διαδικασία για έναν αριθμό επαναλήψεων και αφού αποδεχόμαστε τη κάθε σελίδα με μία συγκεκριμένη πιθανότητα αποδοχής δημιουργούμε προσεγγιστικά ένα τυχαίο δείγμα.

### 3.1 Πιθανά προβλήματα των MCMC μεθόδων

---

Οι MCMC αλγόριθμοι μπορούν να χρησιμοποιηθούν για να παράγουν τυχαίες μεταβλητές από οποιαδήποτε κατανομή μας ενδιαφέρει, ανεξάρτητα από τη διάσταση και τη πολυπλοκότητα της. Ωστόσο, είναι πιθανά κάποια προβλήματα από την εφαρμογή τους. Κάποια ενδεικτικά προβλήματα αυτών των αλγόριθμων είναι:

1. Τα δείγματα που προκύπτουν έχουν μεγάλη συσχέτιση σε αντίθεση με το ανεξάρτητο και ισόνομο δείγμα.
2. Τυπικά, χρειάζεται ένα σημαντικό χρονικό διάστημα μέχρι η Μαρκοβιανή αλυσίδα να συγκλίνει στη στάσιμη κατανομή της.
3. Οι εκτιμήσεις που προκύπτουν από τα MCMC δείγματα συχνά έχουν μεγαλύτερη διασπορά σε σχέση με εκείνα τα δείγματα που προκύπτουν από μία δειγματοληψία που δημιουργεί ανεξάρτητα και ισόνομα τυχαία δείγματα.

## 4. Διαγνωστικές μέθοδοι για τις μεθόδους MCMC

---

Από όλα τα παραπάνω διαπιστώσαμε ότι μία προσομοίωση MCMC θα συγκλίνει τελικά στη συνάρτηση που θέλουμε, αλλά αυτό δε μας δίνει καμία πληροφορία για τη ταχύτητα της σύγκλισης ή για τη μίξη των προσομοιωμένων τιμών της.

Στην πράξη, λοιπόν, θα πρέπει να ενδιαφερθούμε για τον αριθμό των επαναλήψεων που χρειάζεται η αλυσίδα για να φτάσει στη στάσιμη κατανομή της αλλά και για το *burn-in*. Ο όρος *burn-in* περιγράφει τον αριθμό των επαναλήψεων που δεν συμπεριλαμβάνουμε σε μία MCMC μέθοδο. Δηλαδή, εφόσον η σύγκλιση πραγματοποιείται ανεξαρτήτως της αρχικής μας κατάστασης, συνήθως μπορούμε να επιλέξουμε για αυτή τη κατάσταση οποιαδήποτε τιμή (με την προϋπόθεση φυσικά ότι ανήκει στο παραμετρικό μας χώρο). Ωστόσο, ο χρόνος σύγκλισης διαφοροποιείται ανάλογα με την αρχική μας τιμή. Έτσι, είναι σύνηθες να απορρίπτουμε ένα συγκεκριμένο αριθμό των αρχικών προσομοιωμένων τιμών τον οποίο καλούμε *burn-in*. Αυτό συμβαίνει γιατί θέλουμε οι προσομοιώσεις μας να προσεγγίζουν όσο το δυνατόν περισσότερο τη στάσιμη κατανομή μας και να εξαρτώνται όσο το δυνατόν λιγότερο από την αρχική μας κατάσταση. Έτσι, διαπιστώνουμε πως είναι τελικά ασαφές πόσο πρέπει να είναι το *burn-in* σε κάθε περίπτωση αφού οι προσομοιώσεις μας τιμές είναι εξαρτημένες και δεν γνωρίζουμε ακριβώς πότε επιτυγχάνεται η σύγκλιση.

Όπως αναφέραμε δε θεωρείται ιδιαίτερα δύσκολο να κατασκευάσουμε μια Μαρκοβιανή αλυσίδα με τα χαρακτηριστικά και τις ιδιότητες που θέλουμε. Το πιο δύσκολο πρόβλημα είναι να απαντήσουμε σε πολύ συχνά ερωτήματα που αφορούν τη μίξη και της σύγκλιση της αλυσίδας όπως για παράδειγμα «πότε σταματάμε τον αλγόριθμο MCMC;».

Με μία πρώτη ματιά στους MCMC αλγόριθμους διαπιστώσαμε ότι κάτω από σχεδόν γενικές συνθήκες αυτοί οι αλγόριθμοι συγκλίνουν επειδή οι αλυσίδες που παράγουν είναι εργοδικές. Παρόλο που κάτι τέτοιο είναι θεωρητικά έγκυρο, είναι αναποτελεσματικό όσον αφορά την εφαρμογή των MCMC μεθόδων. Όντως, οι θεωρητικές εγγυήσεις δεν μας ενημερώνουν για το πότε πρέπει να σταματήσουμε αυτούς τους αλγόριθμους και πότε πρέπει να είμαστε βέβαιοι για τις εκτιμήσεις μας. Το πιο ιδανικό σενάριο θα ήταν να υπάρχουν ξεκάθαρα κριτήρια, τα οποία με την εφαρμογή τους στην  $R$  θα ήταν ικανά να μας απαλλάσσουν από οποιαδήποτε περαιτέρω διεργασία. Όμως, στην πράξη μια τέτοια γενική αντιμετώπιση είναι αδύνατη και απαιτούνται πολλαπλά «τρεξίματα» του εκάστοτε προγράμματος μας έτσι ώστε να είμαστε ικανοποιημένοι με το αποτέλεσμα.

Η πρώτη προϋπόθεση για τη σύγκλιση ενός MCMC αλγόριθμου είναι η κατανομή της αλυσίδας  $X_t$  να είναι η στάσιμη κατανομή  $f$ . Βέβαια, για έναν αλγόριθμο για τον οποίο έχουμε ήδη αναφέρει ότι θεωρείται σαν μια γεννήτρια της  $f$ , μια τέτοια προϋπόθεση μοιάζει να ικανοποιείται από την αρχή. Δυστυχώς, κάτι τέτοιο δεν είναι τόσο άμεσο όσο φαίνεται. Πράγματι, το να καταλήξουμε στο ότι η  $X_t$  κατανέμεται ακριβώς από την  $f$  είναι αρκετά δύσκολο όταν η μορφή της  $f$  είναι περίπλοκη και σχεδόν αδύνατο όταν θεωρήσουμε μόνο μια υλοποίηση της Μαρκοβιανής αλυσίδας  $X_t$ .

**Παρατήρηση 2.1:** Όταν τρέχουμε έναν MCMC αλγόριθμο, τα πιο σημαντικά θέματα είναι η ταχύτητα της εξερεύνησης του στηρίγματος της  $f$  και ο βαθμός της συσχέτισης μεταξύ των  $X_t$ . Ωστόσο, κάτι τέτοιο δεν συνεπάγεται ότι η στασιμότητα δεν θα πρέπει να ελέγχεται γιατί σε μερικές περιπτώσεις, ασχέτως της αρχικής κατανομής, η αλυσίδα μπορεί να αργήσει να εξερευνήσει το στήριγμα της  $f$ , μένοντας για αρκετό καιρό σε κάποιες υπο-περιοχές της  $f$ . Ένα τεστ στασιμότητας που συγκρίνει πολλές αλυσίδες μπορεί να αποδειχθεί αρκετά χρήσιμο έτσι ώστε να αναγνωριστούν τέτοιου είδους δυσκολίες.

Έτσι, το κυριότερο εργαλείο για να αξιολογήσουμε τη σύγκλιση του αλγόριθμου είναι να χρησιμοποιήσουμε παράλληλα πολλές αλυσίδες με σκοπό να συγκρίνουμε την απόδοσή τους. Αυτό σημαίνει ότι η πιο αργή αλυσίδα στην ομάδα αυτών των αλυσίδων και η επιλογή της αρχικής κατάστασης θα παίξουν πολύ σημαντικό ρόλο στη σύγκλιση. Γενικότερα, για έναν αλγόριθμο MCMC ισχύει ότι αν μια αλυσίδα δεν έχει εξερευνήσει μια περιοχή της  $f$  μέχρι ένα συγκεκριμένο χρόνο τότε είναι σχεδόν αδύνατο να εντοπίσει αυτήν την περιοχή στο μέλλον. Άρα, ο έλεγχος για τη σύγκλιση ενός MCMC αλγόριθμου με τη χρήση παράλληλων αλυσίδων δεν είναι τόσο δαπανηρός όσο η χρήση μιας μόνο αλυσίδας και τα αποτελέσματα των παράλληλων αλυσίδων μπορούν να χρησιμοποιηθούν στη συνέχεια και για περαιτέρω εκτιμήσεις.

Όσον αφορά τις *διαγνωστικές μεθόδους*, ένας έλεγχος σύγκλισης θα μπορούσε να είναι το γράφημα των προσομοιωμένων τιμών με σκοπό να εντοπίσουμε αποκλίνοσες συμπεριφορές. Βέβαια, ένα φανερό μειονέκτημα αυτής της μεθόδου είναι ότι αν οι αλυσίδες «κολλήσουν» σε μια συγκεκριμένη περιοχή του

παραμετρικού χώρου μακριά από το στήριγμα της ζητούμενης  $f$  τότε μπορούν και πάλι να εμφανίσουν στασιμότητα στο αντίστοιχο γράφημα.

Μία ακόμη γραφική διάγνωση που είναι κατάλληλη για τη σύγκλιση της στασιμότητας είναι ο έλεγχος της σύγκλισης της εμπειρικής αθροιστικής συνάρτησης κατανομής που προέρχεται από τις Μαρκοβιανές αλυσίδες.

Όταν αντιμετωπιστεί το θέμα της σύγκλισης της στάσιμης κατανομής τότε ερχόμαστε αντιμέτωποι με το θέμα της σύγκλισης του εμπειρικού μέσου

$$\frac{1}{T} \sum_{i=1}^T h(X_t)$$

στην  $\mathbb{E}_f[h(X)]$  για μια συνάρτηση  $h$ . Σε αυτό το πρόβλημα μπορούν να χρησιμοποιηθούν κάποια εργαλεία Monte Carlo, αλλά πρέπει να τονίσουμε πάλι πως υπάρχουν δύο χαρακτηριστικά που ξεχωρίζουν τα MCMC αποτελέσματα με αυτά του ανεξάρτητου και ισόνομου δείγματος: η εξάρτηση που υπάρχει ανάμεσα στις διάφορες τιμές του δείγματος και η μικτή συμπεριφορά της μετάβασης (δηλαδή πόσο γρήγορα η αλυσίδα εξερευνά το στήριγμα της  $f$ ). Για αυτό το λόγο, επιπλέον έλεγχοι είναι απαραίτητοι έτσι ώστε να μπορούν να χρησιμοποιηθούν οι διάφοροι μέθοδοι εκτιμήσεων Monte Carlo και στην περίπτωση του συσχετισμένου δείγματος.

Το αρχικό και πιο φυσικό διαγνωστικό εργαλείο είναι να σχεδιάσουμε την εξέλιξη της αντίστοιχης εκτιμήτριας καθώς το  $T$  αυξάνεται. Δηλαδή, και σε αυτήν περίπτωση μπορεί να εφαρμοστεί ο έλεγχος της εμπειρικής αθροιστικής κατανομής για τη διάγνωση της σύγκλισης. Αν η καμπύλη των αθροιστικών μέσων δεν έχει σταθεροποιηθεί μετά από  $T$  επαναλήψεις, τότε οι απαραίτητες επαναλήψεις της Μαρκοβιανής αλυσίδας πρέπει να αυξηθούν. Βέβαια, στις περισσότερες περιπτώσεις το γράφημα είτε των προσομοιωμένων τιμών είτε του αντίστοιχου αθροιστικού μέσου δεν μας βοηθά στην ανίχνευση στασιμότητας ή σύγκλισης. Παραδόξως, μόνο όταν η αλυσίδα έχει εξερευνήσει διαφορετικές περιοχές του χώρου καταστάσεων κατά τη διάρκεια του χρόνου παρατήρησης μπορεί να ανιχνευτεί μια έλλειψη στασιμότητας.

## 5. Προσαρμοστικές μέθοδοι MCMC

Πολλές φορές, η Μαρκοβιανή αλυσίδα μπορεί να μείνει σε μια κατάσταση για αρκετό καιρό, υποδηλώνοντας έτσι τη μείωση της διασποράς της κατανομής εισήγησης. Έτσι, μετά τη διάγνωση της σύγκλισης, το επόμενο βήμα στην κατασκευή των MCMC αλγόριθμων θα μπορούσε να είναι η βαθμονόμηση και έτσι η κατάταξη τους σύμφωνα με την απόδοση τους με βάση πάντα τα διαγνωστικά εργαλεία σύγκλισης.

Για αυτό, θα αναφερθούμε επιγραμματικά στους προσαρμοστικούς αλγόριθμους, μια πρόσφατα ανεπτυγμένη επέκταση των MCMC αλγόριθμων, όπου οι πυρήνες μετάβασης είναι συντονισμένοι απευθείας με τις επιδόσεις που παρατηρήθηκαν μέχρι στιγμής στον αντίστοιχο αλγόριθμο. Ο στόχος αυτών των μεθόδων είναι η χρήση «του καλύτερου» πυρήνα μέσα από μια συλλογή υποψήφιων πυρήνων ή «των καλύτερων» παραμέτρων μέσα από μια οικογένεια παραμέτρων. Αυτό, βέβαια, είναι πολύ δύσκολο να εφαρμοστεί στην πράξη γιατί μπορεί να δημιουργήσει αρνητικές επιπτώσεις στις ιδιότητες σύγκλισης του αντίστοιχου αλγόριθμου. Δηλαδή, αν συνεχίζουμε να συντονίζουμε τον αλγόριθμο σύμφωνα με το αποτέλεσμα του μέχρι τη παρούσα χρονική στιγμή, αυτό μετατρέπει αυτόματα τον αλγόριθμο σε μη Μαρκοβιανό αφού πλέον εξαρτάται από όλες τις προηγούμενες προσομοιωμένες τιμές. Για αυτό, υπάρχουν αυστηροί περιορισμοί που πρέπει να τεθούν σε αυτή τη διαδικασία αν θέλουμε να εξασφαλίσουμε τη σίγουρη σύγκλιση του αλγόριθμου.

Από τα παραπάνω διαπιστώνουμε πως δεν είναι ευρεία η χρήση και η προσαρμογή των προσαρμοστικών μεθόδων Markov Chain Monte Carlo, αφού η συνεχής ανανέωση που υπάρχει στη βάση αυτών των αλγόριθμων απαιτεί την πολύ προσεκτική εξέταση των συνθηκών και των προϋποθέσεων που είναι απαραίτητοι για τη χρησιμοποίησή τους.

## Κεφάλαιο 3 Ο Αλγόριθμος Metropolis-Hastings

---

### 1.Ο Αλγόριθμος Metropolis

---

Ο αλγόριθμος Metropolis θεωρείται ένα από τα πιο επιτυχημένα και γνωστά μέλη των μεθόδων Markov Chain Monte Carlo. Σήμερα προβλήματα που μπορούν να αντιμετωπιστούν με αυτόν τον αλγόριθμο συνιστούν ένα ολόκληρο πεδίο της υπολογιστικής επιστήμης και οι εφαρμογές αυτού του αλγόριθμου κυμαίνονται από τις φυσικές προσομοιώσεις μέχρι και τις βάσεις της δυσεπίλυτης υπολογιστικής. Στην ουσία αυτός ο αλγόριθμος είναι το κυριότερο μέλος της ομάδας MCMC και άρα παράγει συσχετισμένες μεταβλητές από μια Μαρκοβιανή αλυσίδα.

Ο αρχικός αλγόριθμος Metropolis αναπτύχθηκε από μια ομάδα ερευνητών, τους Metropolis, Rosenbluth, Rosenbluth, Teller και Teller (1954) με σκοπό να λύσουν κάποια υπολογιστικά προβλήματα στατιστικής φυσικής στις περιπτώσεις εκείνες όπου ήταν δύσκολο να πάρουν ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα. Συγκεκριμένα, αυτό που έκαναν ήταν να επιτρέψουν στην *κατανομή εισήγησης* (δηλαδή στην προτεινόμενη κατανομή, από την οποία θέλουμε να παράγουμε τιμές) να εξαρτάται από την τωρινή κατάσταση της αλυσίδας και όταν μια τιμή απορριπτόταν, ανάγκαζαν την αλυσίδα να μείνει για ακόμη μία επανάληψη στην ίδια κατάσταση. Δηλαδή, από όλες τις παρελθοντικές καταστάσεις της αλυσίδας μόνο η πιο πρόσφατη είχε σημασία. Έτσι, στο τέλος θα προκύψει μια Μαρκοβιανή αλυσίδα επειδή οι τιμές της αλυσίδας είναι συσχετισμένες και επειδή επιτυγχάνεται η Μαρκοβιανή ιδιότητα.

Ο αλγόριθμος αυτός είναι στην ουσία μια διαδικασία Αποδοχής- Απορρίψης. Στο βήμα 1 επιλέγεται ένα αυθαίρετο σημείο εκκίνησης  $x_1$ . Στη συνέχεια μετά από κάθε βήμα  $i$  η κατάσταση της αλυσίδας  $x_i$  προσομοιώνεται σύμφωνα με τα δύο επόμενα στάδια:

1. Ένα συμμετρικό τυχαίο βήμα παρέχει μια υποψήφια κατάσταση για το βήμα  $i$ .
2. Η υποψήφια κατάσταση αυτή είτε γίνεται αποδεκτή είτε απορριπτέα. Το κριτήριο αποδοχής χρησιμοποιεί τη συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της κατανομής εισήγησης. Εάν η προτεινόμενη κατάσταση  $x_p$ , έχει μεγαλύτερη σ.π.π. από τη προηγούμενη κατάσταση  $x_{i-1}$  τότε η προτεινόμενη κατάσταση γίνεται αποδεκτή.

Όταν μια προτεινόμενη κατάσταση απορρίπτεται τότε η προηγούμενη κατάσταση επαναλαμβάνεται. Γι αυτό υπάρχει ένας συμβιβασμός στην τυχαία επιλογή του βήματος. Δηλαδή αν είναι πολύ διασκορπισμένο, τότε η διασπορά των προσομοιωμένων τιμών θα είναι πολύ μεγάλη και έτσι το ποσοστό αποδοχής μπορεί να είναι πολύ μικρό, ενώ αν είναι πολύ συμπυκνωμένο, τότε η διασπορά θα είναι πολύ μικρή και έτσι η αλυσίδα δεν θα μπορεί να κινηθεί ελεύθερα ανάμεσα στις διάφορες τιμές της κατανομής εισήγησης.

## 2.Ο Γενικός Αλγόριθμος Metropolis-Hastings

Ο αλγόριθμος Metropolis πάντα προϋποθέτει να είναι συμμετρική η κατανομή εισήγησης. Έτσι οι επιλογές μας για την κατάλληλη κατανομή εισήγησης μειώνονται. Ο Hastings γενίκευσε αυτή την ιδέα για μη συμμετρικές κατανομές εισήγησης, δημιουργώντας έτσι τον αλγόριθμο Metropolis-Hastings (1970).

Βασίζόμενος στον αλγόριθμο Metropolis, ο Hastings απέδειξε ότι μπορούμε να καταλήξουμε στην κατάλληλη στάσιμη κατανομή αρκεί να χρησιμοποιήσουμε σαν πιθανότητα αποδοχής την ακόλουθη ποσότητα την οποία την ονομάζουμε *πιθανότητα αποδοχής του Metropolis-Hastings*:

$$\rho(x, y) = \min \left\{ \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right\}. \quad [3.1]$$

**Παρατήρηση 3.1:** Δεν πρέπει να συγχέουμε το ρυθμό αποδοχής με τη πιθανότητα αποδοχής [3.1]. Ο ρυθμός αποδοχής είναι ο μέσος όρος των πιθανοτήτων αποδοχής μετά από έναν συγκεκριμένο αριθμό επαναλήψεων και δίνεται από τη σχέση:

$$\bar{\rho} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \rho(X_t, Y^t) \rightarrow \int \rho(x, y) f(x) q(y|x) dy dx. \quad [3.2]$$

Ας υποθέσουμε, λοιπόν, ότι θέλουμε να προσομοιώσουμε από μια συγκεκριμένη συνάρτηση πυκνότητας πιθανότητας  $f(y)$  και επιλέγουμε για κατανομή εισήγησης την  $q(y|x)$ , από την οποία είναι σχετικά εύκολο να προσομοιώσουμε. Διαπιστώνουμε πως η κατανομή εισήγησης  $q$  μπορεί να πάρει οποιαδήποτε μορφή και ότι τελικά η στάσιμη κατανομή της αλυσίδας θα είναι η  $f$ . Έτσι, θα προσομοιώσουμε μια Μαρκοβιανή αλυσίδα και τις προσομοιωμένες τιμές θα τις ονομάσουμε  $y^1, y^2, y^3$ . Ο αλγόριθμος αυτός ξεκινάει με αρχική τιμή  $y^0$  και παράγει την  $y^{t+1}$  σύμφωνα με την  $y^t$ . Άρα, από τα παραπάνω προκύπτει ο κάτωθι αλγόριθμος:



### **Αλγόριθμος 2: Ο γενικός Metropolis-Hastings**

Με δοσμένες τις  $x_t$

1. Παρήγαγε  $Y^t \sim q(y|x_t)$
2. Πάρε

$$X_{t+1} = \begin{cases} Y^t \text{ με πιθανότητα } \rho(x_t, Y^t) \\ x_t \text{ με πιθανότητα } 1 - \rho(x_t, Y^t) \end{cases}$$

όπου

$$\rho(x, y) = \min \left\{ \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right\}$$

### **Απόδειξη του Metropolis-Hastings:**

Ο μεταβατικός πυρήνας για τον αλγόριθμο Metropolis-Hastings είναι ο:

$$P(X_{t+1}|X_t) = q(X_{t+1}|X_t)\rho(X_t, X_{t+1}) + \mathbb{I}(X_{t+1} = X_t) \left[ 1 - \int q(Y|X_t)\rho(X_t, Y)dY \right], \quad [3.3]$$

όπου  $\mathbb{I}(\cdot)$  είναι η δείκτρια συνάρτηση, δηλαδή παίρνει τη τιμή 1 όταν ισχύει η σχέση μέσα στη δείκτρια συνάρτηση και 0 διαφορετικά.

Ο πρώτος όρος του δεξιού μέλους της σχέσης [3.3] προκύπτει από την αποδοχή του υποψήφιου σημείου  $Y = X_{t+1}$  και ο δεύτερος όρος από τη απόρριψη όλων των υποψήφιων τιμών  $Y$ . Χρησιμοποιώντας τη σχέση [3.2] θα έχουμε ότι:

$$f(X_t)q(X_{t+1}|X_t)\rho(X_t, X_{t+1}) = f(X_{t+1})q(X_t|X_{t+1})\rho(X_{t+1}, X_t). \quad [3.4]$$

Έτσι, χρησιμοποιώντας τη σχέση [3.4] αλλά και την [3.3] θα έχω την εξίσωση ισορροπίας:

$$f(X_t)P(X_{t+1}|X_t) = f(X_{t+1})P(X_t|X_{t+1}). \quad [3.5]$$

Τώρα, αν ολοκληρώσουμε την [3.5] ως προς  $X_t$ , θα έχουμε τότε την ακόλουθη σχέση:

$$\int f(X_t)P(X_{t+1}|X_t)dX_t = f(X_{t+1}). \quad [3.6]$$

Το αριστερό μέλος της [3.6] μας δίνει την περιθώρια κατανομή της  $X_{t+1}$ , υπό την προϋπόθεση ότι η  $X_t$  προέρχεται από την  $f$ . Δηλαδή, η σχέση [3.6] υποδηλώνει πως αν η  $X_t$  προέρχεται από την  $f$ , τότε το ίδιο θα συμβαίνει και για τη  $X_{t+1}$ . Άρα, αν

λάβουμε ένα δείγμα από τη στάσιμη κατανομή  $f$ , τότε όλες οι υπόλοιπες τιμές του δείγματος θα προέρχονται και εκείνες από αυτή τη στάσιμη κατανομή και το αποτέλεσμα αυτό θα ισχύει για οποιαδήποτε κατανομή εισήγησης  $q$ .

Βέβαια, πρέπει να επισημάνουμε πως η παραπάνω απόδειξη μας αποσαφηνίζει μόνο το γεγονός ότι η στάσιμη κατανομή είναι η  $f$ , χωρίς να δείχνει αν όντως η αλυσίδα συγκλίνει σε μία στάσιμη κατανομή.

**Παράδειγμα 3.1:** Ας υποθέσουμε ότι παρατηρούμε μόνο τη δειγματική διάμεσο  $M = \text{median}(y_1, y_2, \dots, y_n)$  ενός τυχαίου δείγματος  $y_1, y_2, \dots, y_n$  από τη μετατοπισμένη κατά  $\theta$  εκθετική σ.π.π. με άγνωστη τη παράμετρο θέσης  $\theta$ :

$$f(y|\theta) = e^{-(y-\theta)}, \quad y \geq \theta,$$

δηλαδή οι πραγματικές παρατηρήσεις  $y_1, y_2, \dots, y_n$  δεν είναι διαθέσιμες και έτσι μπορούμε να πάρουμε πληροφορίες για το  $\theta$  χρησιμοποιώντας μόνο τις δειγματικές διαμέσους. Αν ορίσουμε μια καινούρια μεταβλητή  $W = M - \theta$  τότε η κατανομή που θα προκύψει δε θα εξαρτάται από την άγνωστη μεταβλητή  $\theta$  και έτσι θα μπορούμε να βρούμε και τα διαστήματα εμπιστοσύνης της παραμέτρου μας. Δηλαδή, ενδιαφερόμαστε στην ουσία για τη δειγματική κατανομή της διαμέσου  $W$  από ένα τυχαίο δείγμα μεγέθους  $n$  ( $n$  περιττός) από μία τυπική εκθετική κατανομή  $f(w) = e^{-w}, w \geq 0$ .

Όμως η δειγματική κατανομή του  $W$ , στην περίπτωση όπου  $n$  περιττός, έχει αρκετά περίπλοκη μορφή:

$$f(m) = \frac{n!}{\left(\frac{n-1}{2}\right)! \left(\frac{n-1}{2}\right)!} e^{-m} (1 - e^{-m})^{(n-1)/2} (e^{-m})^{(n-1)/2}, m \geq 0$$

και γι αυτό είναι απαραίτητη η χρήση μεθόδων MCMC για την προσομοίωση της. Ένα ιδιαίτερο χαρακτηριστικό του αλγόριθμου Metropolis-Hastings είναι ότι δε χρειαζόμαστε τη σταθερά κανονικοποίησης του  $f(y)$  αφού θα διαγραφεί στον υπολογισμό της πιθανότητας αποδοχής. Έτσι, αν αγνοήσουμε τη σταθερά κανονικοποίησης, μπορούμε να γράψουμε τη δειγματική συνάρτηση πυκνότητας πιθανότητας του  $W$  ως:

$$f(y) \propto e^{-y} (1 - e^{-y})^{(n-1)/2} (e^{-y})^{(n-1)/2}, y \geq 0.$$

Αρχικά, λοιπόν, γράφουμε μια μικρή συνάρτηση για να υπολογίσουμε το λογάριθμο της ζητούμενης σ.π.π.  $f(y)$ . Είναι προτιμότερο να υπολογίσουμε το λογάριθμο της συνάρτησης πυκνότητας πιθανότητας αφού με αυτόν τον τρόπο δε θα δημιουργηθούν προβλήματα υποχείλισης.



```
> log.samp.med = function(y, n){
+ (n - 1) / 2 * log(pexp(y)) + (n - 1) / 2 * log(1 - pexp(y)) +
+ log(dexp(y))
+ }
```

Στη συνέχεια, πρέπει να γίνει μια κατάλληλη επιλογή για την κατανομή εισήγησης  $q(y|x)$ . Σίγουρα, υπάρχουν πολλές επιλογές για την κατανομή εισήγησης  $q$ , αλλά το ιδιαίτερα ελκυστικό χαρακτηριστικό αυτής της μεθόδου είναι ότι μπορεί να δουλέψει για μια μεγάλη κλάση συναρτήσεων.

Στην περίπτωση αυτή αφού το στήριγμα της κατανομής αποτελείται από θετικές τιμές, θα δοκιμάσουμε μια εκθετική συνάρτηση με παράμετρο  $x$ :

$$q(y|x) = xe^{-xy}, y \geq 0$$

Έτσι, με αυτήν την κατανομή εισήγησης, η πιθανότητα αποδοχής της τιμής  $y$  θα είναι:

$$\rho(x, y) = \min\left(\frac{f(y)ye^{-yx}}{f(x)xe^{-xy}}, 1\right).$$

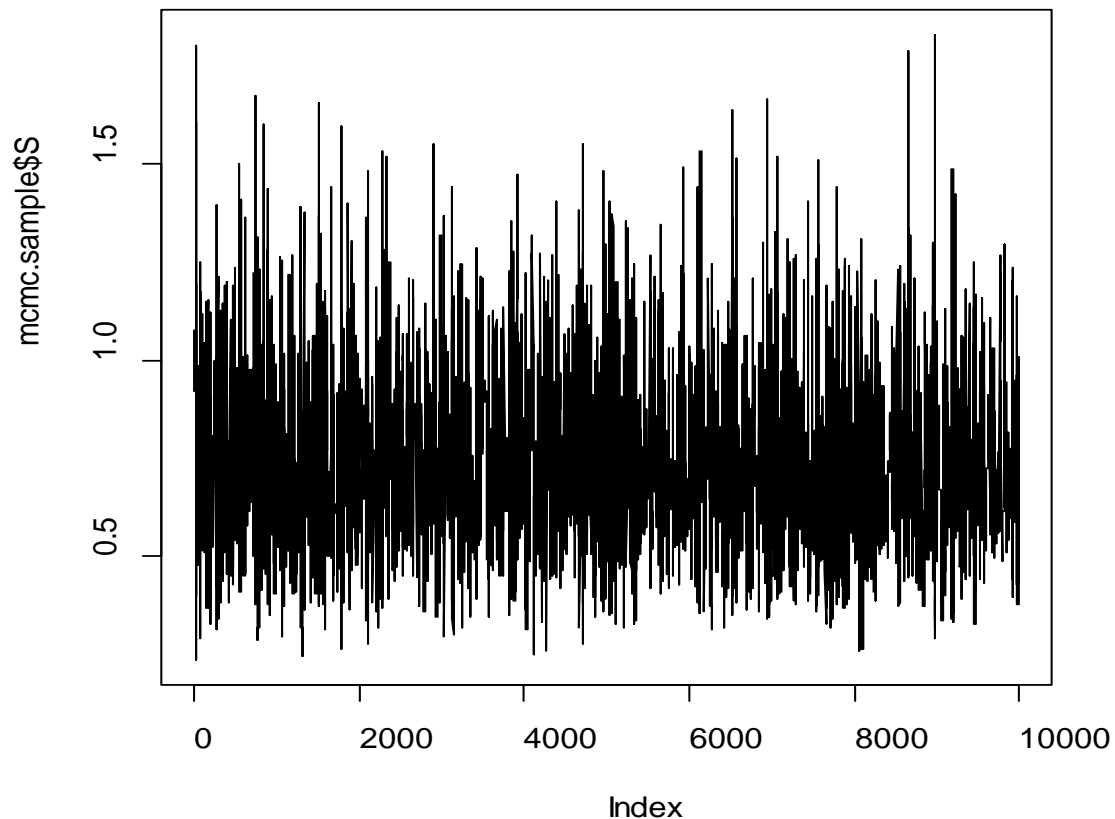
Άρα, θα κατασκευάσω στην R μια συνάρτηση *metrop.hasting.exp* για να εφαρμόσω όλα τα παραπάνω, χρησιμοποιώντας την εκθετική συνάρτηση κατανομής εισήγησης. Έχουμε 4 εισόδους σε αυτή τη συνάρτηση: *logf* είναι το όνομα της συνάρτησης που υπολογίζει το λογάριθμο της συνάρτησης πυκνότητας  $f$ , *current* είναι η αρχική κατάσταση της Μαρκοβιανής αλυσίδας, *iter* είναι ο αριθμός των βημάτων της αλυσίδας και  $n$  είναι το μέγεθος του δείγματος.



```
> metrop.hasting.exp = function(logf, current, iter, n){
+ S = rep(0, iter)
+ n.accept = 0
+ for(j in 1:iter){
+ candidate = rexp(1, current)
+ prob = exp(logf(candidate, n) - logf(current, n) +
+ dexp(current, candidate, log=TRUE) -
+ dexp(candidate, current, log=TRUE))
+ accept = ifelse(runif(1) < prob, "yes", "no")
+ current = ifelse(accept == "yes", candidate, current)
+ S[j] = current; n.accept = n.accept + (accept == "yes")}
+ list(S=S, accept.rate=n.accept / iter)
+ }
> mcmc.sample = metrop.hasting.exp(log.samp.med, 1, 10000, 21)
> mcmc.sample$accept.rate
[1] 0.2692
> plot(mcmc.sample$S, type="l")
```

Όπως βλέπουμε από τον παραπάνω κώδικα στην R:

- Χρησιμοποιώντας τη συνάρτηση *rep*, τοποθετούμε το διάνυσμα μήκους *iter* που θα αποθηκεύσει τις προσομοιωμένες τιμές. Ακόμη, αρχικοποιούμε τον αριθμό των αποδεχτών τιμών *n.accept* ίσο με 0.
- Γράφουμε ένα βρόγχο για να επαναλάβουμε το βασικό βήμα του Metropolis-Hastings *iter* φορές. Σε κάθε βήμα, χρησιμοποιούμε τη συνάρτηση *rexp* για να προσομοιώσουμε μια προτεινόμενη τιμή από μια εκθετική κατανομή με παράμετρο την τρέχουσα τιμή. Στη συνέχεια υπολογίζουμε τη πιθανότητα αποδοχής με τη βοήθεια των  $\log f(x)$ ,  $\log f(y)$ ,  $\log q(y|x)$ ,  $\log q(x|y)$  και αποθηκεύουμε την πιθανότητα αποδοχής στη μεταβλητή *prob*. Θέλουμε να αποδεχτούμε μια προτεινόμενη τιμή με πιθανότητα *prob*. Αυτό το επιτυγχάνουμε με την προσομοίωση μιας ομοιόμορφης μεταβλητής στο διάστημα (0,1) και αποδεχόμαστε την υποψήφια τιμή αν η τυχαία ομοιόμορφη μεταβλητή είναι μικρότερη από την *prob*. Έτσι, η μεταβλητή *accept* θα πάρει τη λογική τιμή “yes” ή “no” ανάλογα με την αποδοχή ή την απόρριψη της τιμής αντίστοιχα. Αν τελικά αποδεχτούμε την προτεινόμενη τιμή τότε η νέα τρέχουσα τιμή θα πάρει την τιμή της αποδεκτής μας πλέον τιμής, διαφορετικά θα κρατήσει την προηγούμενη τιμή της. Στη συνέχεια, αποθηκεύουμε τη νέα τρέχουσα τιμή και ανανεώνουμε το διάνυσμα *S* με τον καινούριο αριθμό των αποδεκτών τιμών.
- Η συνάρτηση επιστρέφει μια λίστα με 2 στοιχεία: *S* είναι το διάνυσμα των αποδεκτών τιμών και *accept.rate*, το ποσοστό των αποδεκτών τιμών.
- Στο τέλος προσομοιώσαμε τη δειγματική κατανομή μιας διαμέσου για ένα δείγμα μεγέθους  $n = 21$ , με τιμή εκκίνησης τη τιμή 1 και 10.000 επαναλήψεις της Μαρκοβιανής αλυσίδας. Όπως βλέπουμε, περίπου το 27% των υποψήφιων τιμών τελικά έγινε αποδεκτό.
- Για να διαπιστώσουμε αν το δείγμα μας τελικά έχει καλή μίξη, θα το αναπαραστήσουμε γραφικά σε συνδυασμό με τον αριθμό  $t_n$  επαναλήψεων. Από το Σχήμα 3.1 βλέπουμε ότι δεν υπάρχει κάποια συγκεκριμένη τάση στο δείγμα των προσομοιωμένων τιμών το οποίο συνεπάγεται ότι η περίοδος του burn-in είναι μικρή. Ακόμη, φαίνεται ότι οι προσομοιωμένες τιμές έχουν ένα ευρύ φάσμα κάτι οποίο δείχνει ότι η αλυσίδα έχει καλή μίξη και επισκέφτηκε το στήριγμα της  $f$ .

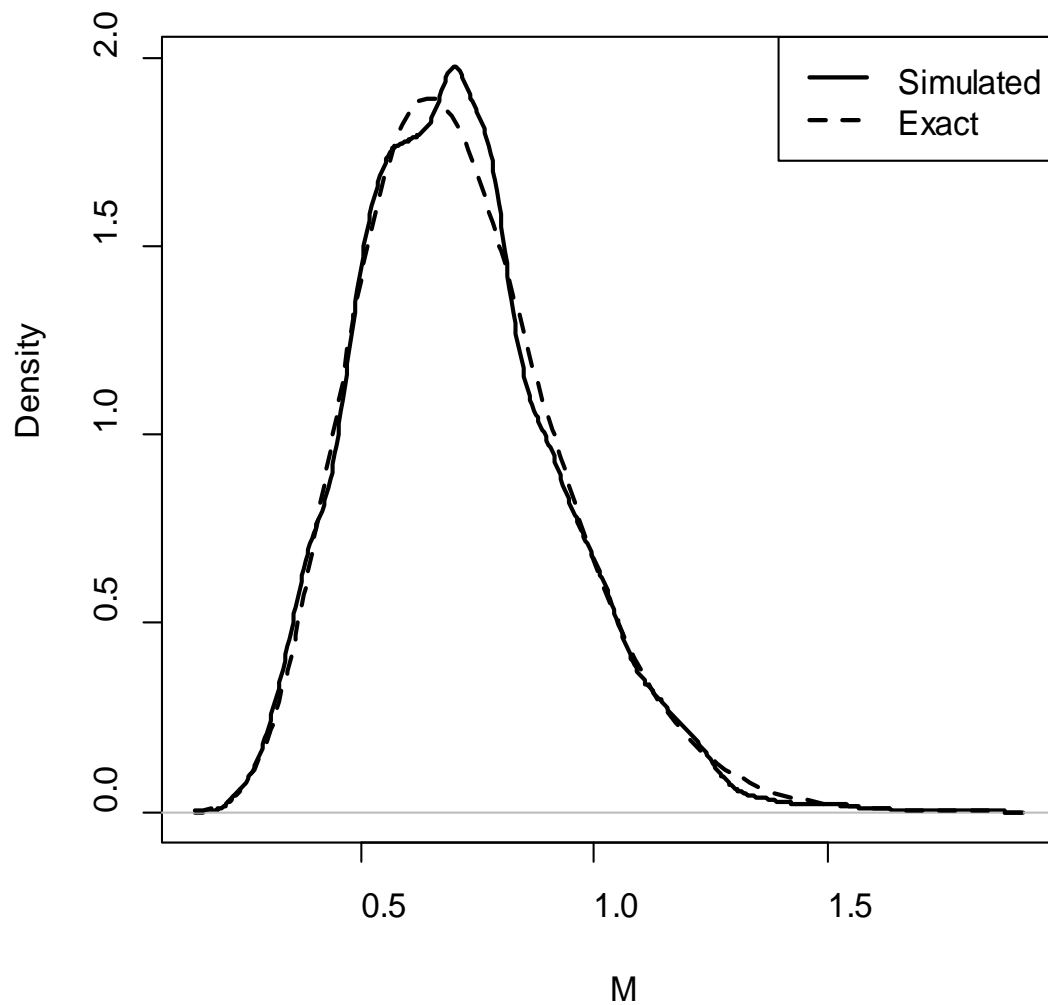


**Σχήμα 3.1:** Γράφημα των προσομοιωμένων τιμών από τον Metropolis-Hastings αλγόριθμο για το παράδειγμα της διαμέσου με εκθετική κατανομή εισήγησης.

Για να διαπιστώσουμε τελικά αν αυτή η μέθοδος όντως δουλεύει, μπορούμε να συγκρίνουμε τις προσομοιωμένες τιμές που προέκυψαν από τον αλγόριθμο Metropolis-Hastings με εκείνες από την ακριβή δειγματική συνάρτηση πυκνότητας πιθανότητας, την οποία τη βρίσκουμε από τον παρακάτω κώδικα της R.

```
> samp.med = function(m, n){
+   con = factorial(n) / factorial((- 1) / 2)^2
+   con * pexp(m)^((n - 1) / 2) * (1 - pexp(m))^((n - 1) / 2) * dexp(m)
+ }
> plot(density(mcmc.sample$S), lwd=2, main="", xlab="M")
> curve(samp.med(x, 21), add=TRUE, lwd=2, lty=2)
```






**Σχήμα 3.2:** 10.000 προσομοιωμένες τιμές από τον αλγόριθμο Metropolis-Hastings για το παράδειγμα της διαμέσου με εκθετική κατανομή. Η ακριβής δειγματική συνάρτηση πυκνότητας πιθανότητας είναι εκείνη με τη διακεκομένη γραμμή.

Όπως βλέπουμε από το Σχήμα 3.2 οι δύο συναρτήσεις πυκνότητας πιθανότητας είναι παρόμοιες, υποδεικνύοντας ότι αυτός ο αλγόριθμος Metropolis-Hastings συγκλίνει προσεγγιστικά στη ζητούμενη συνάρτηση κατανομής.

### 3.Σύγκριση Αλγόριθμου Metropolis-Hastings Με Μία Δειγματοληψία Από Ένα Ανεξάρτητο Και Ισόνομο Δείγμα Τυχαίων Μεταβλητών

Όπως έχουμε ήδη αναφέρει ο αλγόριθμος Metropolis-Hastings παράγει συσχετισμένες τιμές σε αντίθεση με μία δειγματοληψία η οποία παράγει ανεξάρτητες και ισόνομες τυχαίες τιμές. Στην πράξη, η απόδοση αυτού του αλγόριθμου θα εξαρτηθεί σε μεγάλο βαθμό από την επιλογή της κατανομής εισήγησης  $q$ , αλλά πρώτα θα εξετάσουμε ένα παράδειγμα στο οποίο συγκρίνονται οι προσομοιωμένες τιμές που προκύπτουν από τη μέθοδο Metropolis-Hastings με τις ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που προκύπτουν από μία απλή δειγματοληψία.

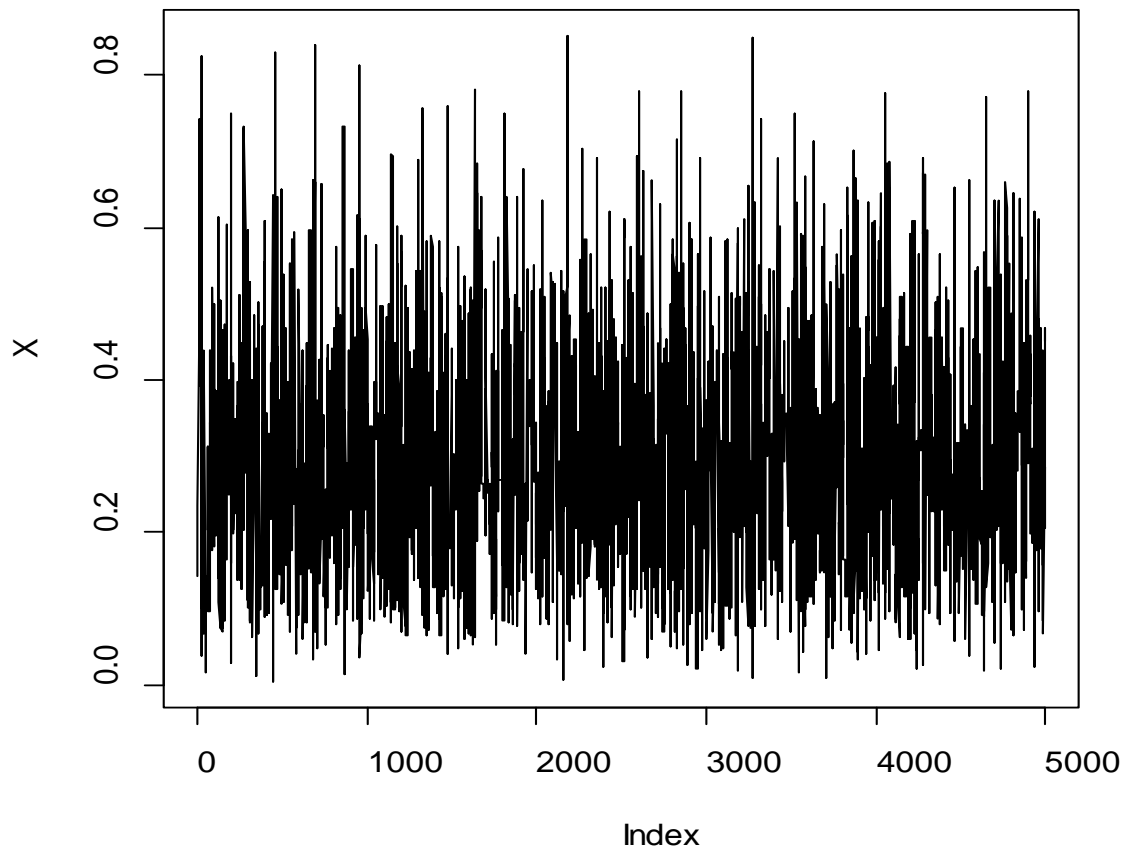
**Παράδειγμα 3.2:** Χρησιμοποιούμε ένα αλγόριθμο Metropolis-Hastings για να παράγουμε δείγμα από τη  $Beta(2.7,6.3)$  και με κατανομή εισήγησης  $q$  την ομοιόμορφη στο διάστημα  $[0,1]$ . Ένα δείγμα Metropolis-Hastings κατασκευάζεται λοιπόν από τον παρακάτω κώδικα της R:



```
> a=2.7
> b=6.3
> Nsim=5000
> X=runif(Nsim)
> last=X[1]
> for (i in 1:Nsim){
+ cand=runif(1)
+ alpha=(dbeta(cand,a,b)/dbeta(last,a,b))/
+ (dunif(cand)/dunif(last))
+ if (runif(1)<alpha)
+ last=cand
+ X[i]=last
+ }
> plot(X,type="l")
> plot(X,xlim=c(4500,4800),ylim=c(0.0,0.5),type="l",xlab="Iterations")
> par(mfrow=c(1,2))
> hist(X,freq=FALSE,col="wheat2")
> curve(dbeta(x,2.7,6.3),add=TRUE)
> Y<-rbeta(5000,2.7,6.3)
> hist(Y,freq=FALSE,col="wheat2")
> curve(dbeta(x,2.7,6.3),add=TRUE)
```

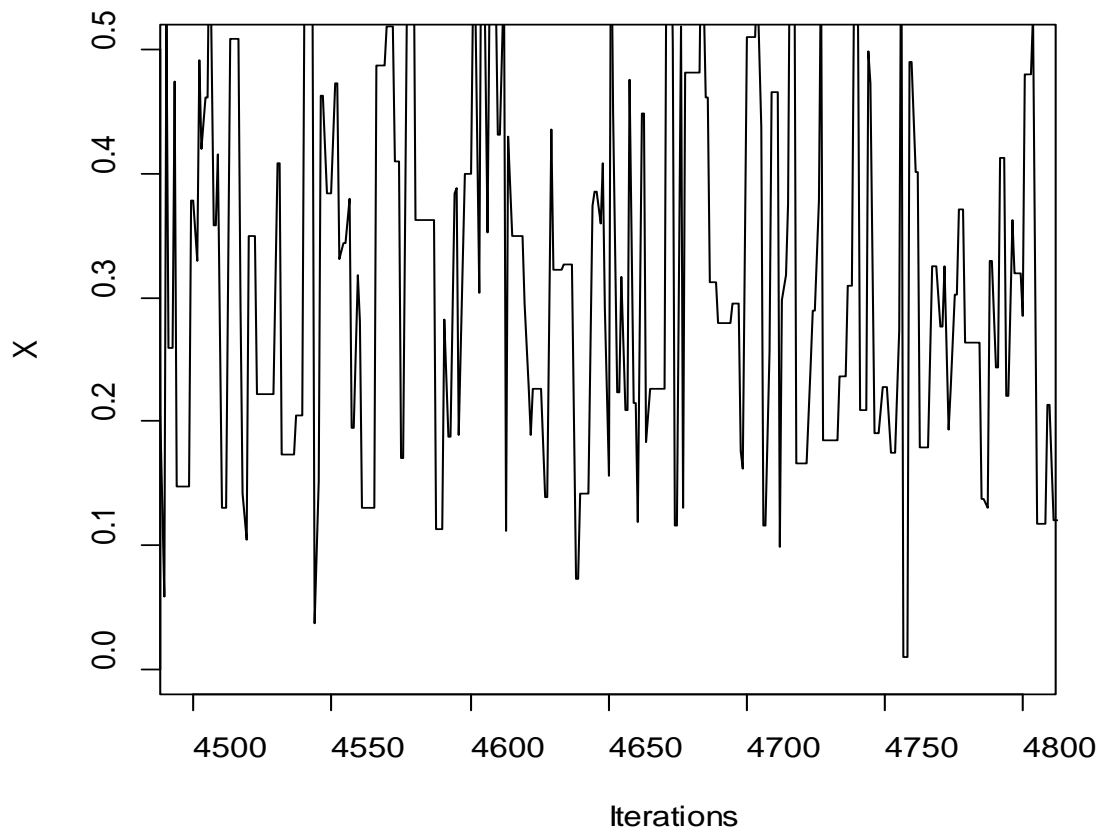
Μία αναπαράσταση της ακολουθίας  $X_t$  χρησιμοποιώντας τη συνάρτηση *plot* της R δεν φαίνεται να παράγει κάποιο συγκεκριμένο πρότυπο στην προσομοίωση καθώς η αλυσίδα εξερευνά το ίδιο εύρος σε διαφορετικές περιόδους (Σχήμα 3.3).

Αν όμως μεγενθύνουμε την τελευταία περιοχή, δηλαδή εκείνη για  $4500 \leq t \leq 4800$  το Σχήμα 3.4 παρουσιάζει κάποια χαρακτηριστικά στοιχεία των μεθόδων Metropolis-Hastings. Δηλαδή, βλέπουμε από το γράφημα των προσομοιωμένων τιμών πως για κάποια χρονικά διαστήματα η ακολουθία  $X_t$  δεν αλλάζει γιατί όλες οι αντίστοιχες υποψήφιες τιμές  $Y_t$  απορρίπτονται.



**Σχήμα 3.3:** Αναπαράσταση των σημείων της ακολουθίας  $X_t$  με τον αλγόριθμο Metropolis-Hastings για την προσομοίωση τυχαίων μεταβλητών από τη  $Be(2.7, 6.3)$  με ομοιόμορφη κατανομή εισήγησης.





**Σχήμα 3.4:** Ακολουθία  $X_t$  για  $t = 4500, \dots, 4800$  με προσομοίωση χρησιμοποιώντας τον αλγόριθμο Metropolis-Hastings με ομοιόμορφη κατανομή εισήγησης και ζητούμενη συνάρτηση  $Beta(2.7, 6.3)$ .

Το Σχήμα 3.5 αναπαριστά τα ιστογράμματα μαζί με τις συναρτήσεις πυκνότητας πιθανότητας για ένα δείγμα που προκύπτει από τη μέθοδο Metropolis-Hastings και για ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα που προκύπτει από τη συνάρτηση της R, την *rbeta*. Οι προσαρμογές είναι αρκετά παρόμοιες και αυτό μπορεί να αποδειχθεί χρησιμοποιώντας και το Kolmogorov-Smirnov τεστ μεταξύ των 2 αντίστοιχων δειγμάτων:

```
> ks.test(jitter(X),rbeta(5000,a,b))
```

Two-sample Kolmogorov-Smirnov test

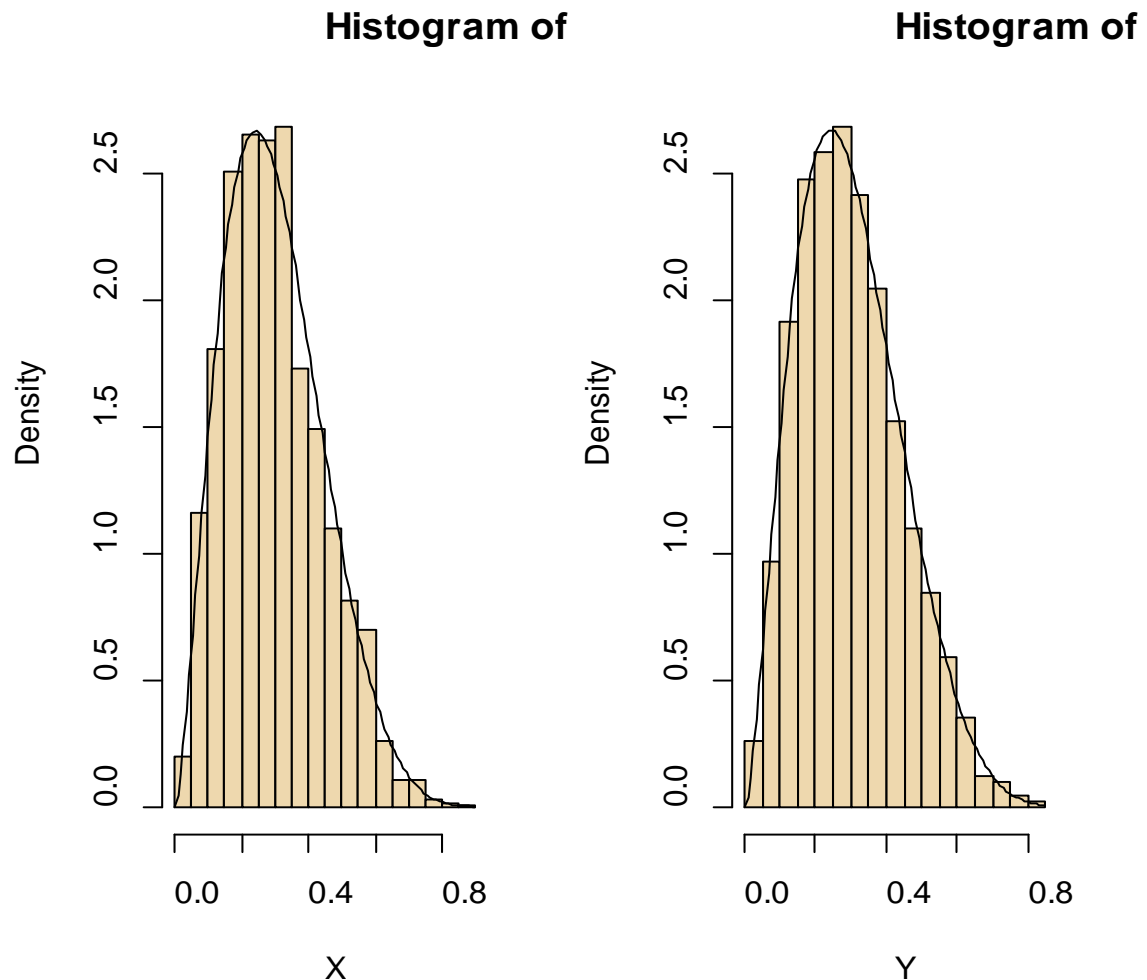
data: jitter(X) and rbeta(5000, a, b)

D = 0.0294, p-value = 0.02655

alternative hypothesis: two-sided



Τα παραπάνω αποτελέσματα από το τεστ ισότητας Kolmogorov-Smirnov δηλώνουν πως και τα δύο δείγματα ακολουθούν την ίδια κατανομή, αφού η  $p - τιμή > 0.05$  και έτσι αποδεχόμαστε τη μηδενική υπόθεση.



**Σχήμα 3.5:** Ιστογράμματα των τυχαίων μεταβλητών της Βήτα κατανομής  $Beta(2.7, 6.3)$  με τις επικαλυπτόμενες σ.π.π. Στο αριστερό ιστόγραμμα, οι τυχαίες μεταβλητές δημιουργήθηκαν από έναν Metropolis-Hastings αλγόριθμο με ομοιόμορφη κατανομή εισήγησης ενώ στο δεξί ιστόγραμμα οι παραγόμενες τυχαίες μεταβλητές προέρχονται από τη συνάρτηση  $rbeta(n, 2.7, 6.3)$  του στατιστικού πακέτου R.

Ένας τρίτος έλεγχος για την σύγκριση αυτών των δειγμάτων που παρήχθησαν με τις διαφορετικές μεθόδους είναι ο αναλυτικός υπολογισμός του μέσου και της διασποράς με τη βοήθεια των αντίστοιχων τύπων. Δηλαδή, εφόσον γνωρίζω πως η μέση τιμή και η διασπορά της Βήτα κατανομής είναι  $\alpha/(\alpha + b)$  και  $ab/(\alpha + b)^2(\alpha + b + 1)$  αντίστοιχα μπορώ να συγκρίνω τις εκτιμήσεις που βρήκα από τις προσομοιωμένες τιμές με τη μέθοδο Metropolis-Hastings:

$$\bar{X} = 0.31, \quad S^2 = 0.021$$

με τις θεωρητικές τιμές που ισχύουν για τη Βήτα κατανομή, δηλαδή 0.3 για τον μέσο και 0.021 για την τυπική απόκλιση.

**Παρατήρηση 3.2:** Παρόλα αυτά ενώ από το Σχήμα 3.5 φαίνεται ότι οι MCMC αλγόριθμοι και η δειγματοληψία που παράγει ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα μοιάζουν ίδιες, είναι σημαντικό να θυμόμαστε ότι το προσομοιωμένο δείγμα MCMC έχει συσχέτιση ενώ το ανεξάρτητο και ισόνομο τυχαίο δείγμα όχι. Αυτό σημαίνει ότι στην περίπτωση των συσχετισμένων μεταβλητών χρειαζόμαστε μεγαλύτερο αριθμό προσομοιώσεων για να επιτύχουμε την ίδια ακρίβεια.

## 4.Σημαντικά χαρακτηριστικά του αλγόριθμου Metropolis-Hastings

Σύμφωνα με τα παραπάνω καταλήξαμε σε κάποια συγκεκριμένα χαρακτηριστικά και ιδιότητες του γενικού αλγόριθμου Metropolis-Hastings :

- Ο αλγόριθμος αυτός αποδέχεται τις τιμές  $y$  όταν ο λόγος  $f(y)/q(y|x)$  μεγαλώνει σε σύγκριση με τη προηγούμενη τιμή του λόγου  $f(x)/q(x|y)$ .

Παρόλα αυτά κάποιες τιμές  $y$  στις οποίες ο αντίστοιχος λόγος μειώνεται μπορεί να αποδεχτούν αλλά αν η μείωση είναι πολύ μεγάλη τότε κατά βάση οι προτεινόμενες τιμές  $y$  απορρίπτονται. Άρα συμπεραίνουμε ότι η επιλογή της κατανομής εισήγησης είναι πολύ σημαντική καθώς επηρεάζει ολόκληρη την εκτέλεση του αλγόριθμου Metropolis-Hastings. Αν η περιοχή που εξερευνείται από τη  $q$  είναι πολύ μικρή σε σύγκριση με το εύρος της  $f$ , τότε η Μαρκοβιανή αλυσίδα θα αντιμετωπίσει δυσκολίες στην εξερεύνηση ολόκληρου του εύρους της  $f$  και έτσι θα συγκλίνει πολύ αργά.

- Ο Metropolis-Hastings, όπως αναφέραμε, εξαρτάται από την ποσότητα  $f(y)/f(x)$  και έτσι είναι ανεξάρτητος από τις σταθερές κανονικοποίησης που προκύπτουν από την ολοκλήρωση της.
- Άρα, αυτό που έχει τη μεγαλύτερη σημασία στο συγκεκριμένο αλγόριθμο είναι:
  - a) να μπορώ να προσομοιώσω από την κατανομή εισήγησης  $q$  και
  - b) να μπορώ να υπολογίσω το λόγο  $f(y)/q(y|x)$ .

Αυτό συνεπάγεται ότι η επιλογή της  $q$  θα είναι ικανή να εξαλείψει τα δύσχρηστα και δυσεπίλυτα κομμάτια της  $f$  στο παραπάνω λόγο. Γι αυτό το λόγο υπάρχει μία μεγάλη ελευθερία στην τελική επιλογή της κατανομής εισήγησης και έτσι το μόνο που τελικά

χρειαζόμαστε είναι να έχει η  $q$  αρκετά μεγάλη διασπορά ώστε να εξερευνήσει ολόκληρο το εύρος της περιοχής της  $f$ .

Άρα, είναι προτιμότερο το προσομοιωμένο δείγμα να έχει καλή μίξη, δηλαδή να εξερευνήσει γρήγορα ολόκληρη την περιοχή όπου η σ.π.π.  $f$  έχει τη μεγαλύτερη πιθανότητα της. Γι αυτό, όπως αναφέρθηκε, η επιλογή της κατανομής εισήγησης  $q$  είναι πολύ σημαντική.

Άρα, ο Metropolis-Hastings μπορεί να πάρει πολλές μορφές ανάλογα με τις διαφορετικές επιλογές της κατανομής εισήγησης  $q$ . Στο επόμενο κεφάλαιο παραθέτουμε μερικές από αυτές τις μορφές της.

## 5.Διάφορες Μορφές Του Αλγόριθμου Metropolis-Hastings

### 5.1.Ανεξάρτητος Metropolis-Hastings

Ο γενικός αλγόριθμος Metropolis-Hastings, που παρουσιάστηκε στην προηγούμενη παράγραφο, προϋποθέτει μια κατανομή εισήγησης  $q$  που να εξαρτάται από τη τρέχουσα κατάσταση της αλυσίδας. Αν επιτρέψουμε στη  $q$  να είναι ανεξάρτητη από την τρέχουσα κατάσταση της αλυσίδας, δηλαδή  $q(y|x) = g(y)$ , παίρνουμε μια ειδική περίπτωση του γενικού αλγορίθμου: τον *Ανεξάρτητο Metropolis-Hastings*.

#### **Αλγόριθμος 3: Ο Ανεξάρτητος Metropolis-Hastings**

Με δοσμένες τις  $x_t$

1. Παρήγαγε  $Y^t \sim g(y)$
2. Πάρε

$$X_{t+1} = \begin{cases} Y^t & \text{με πιθανότητα } \min \left\{ \frac{f(Y^t)g(x_t)}{f(x_t)g(Y^t)}, 1 \right\} \\ x_t & \text{διαφορετικά} \end{cases}$$


Γενικά, αυτή η μέθοδος μοιάζει σαν μια απλή γενίκευση της μεθόδου Αποδοχής-Απόρριψης, την οποία αναφέραμε στο Κεφάλαιο 2.1. Άρα, οι προτεινόμενες, αν όχι οι αποδεκτές, τιμές είναι οι ίδιες. Για αυτό το λόγο, είναι σημαντικό η κατανομή εισήγησης  $g$  να είναι πολύ κοντά στην ζητούμενη συνάρτηση  $f$ . Είναι φανερό, όμως, ότι η πρακτική εφαρμογή της είναι πιο προβληματική από τη μέθοδο Αποδοχής-

Απόρριψης επειδή η κατασκευή της κατανομής εισήγησης είναι περίπλοκη και η επιλογή της έχει μεγάλη επιρροή στην απόδοση του αλγορίθμου.

Πιο συγκεκριμένα, οι Metropolis-Hastings αλγόριθμοι και οι μέθοδοι Αποδοχής-Απόρριψης είναι και οι δύο γενικές μέθοδοι προσομοίωσης, έχουν αρκετές ομοιότητες μεταξύ τους που επιτρέπουν σύγκριση αν και γενικά είναι σπάνιο να σκεφτούμε τη λύση του Metropolis-Hastings όταν μπορούμε να χρησιμοποιήσουμε έναν αλγόριθμο Αποδοχής-Απόρριψης.

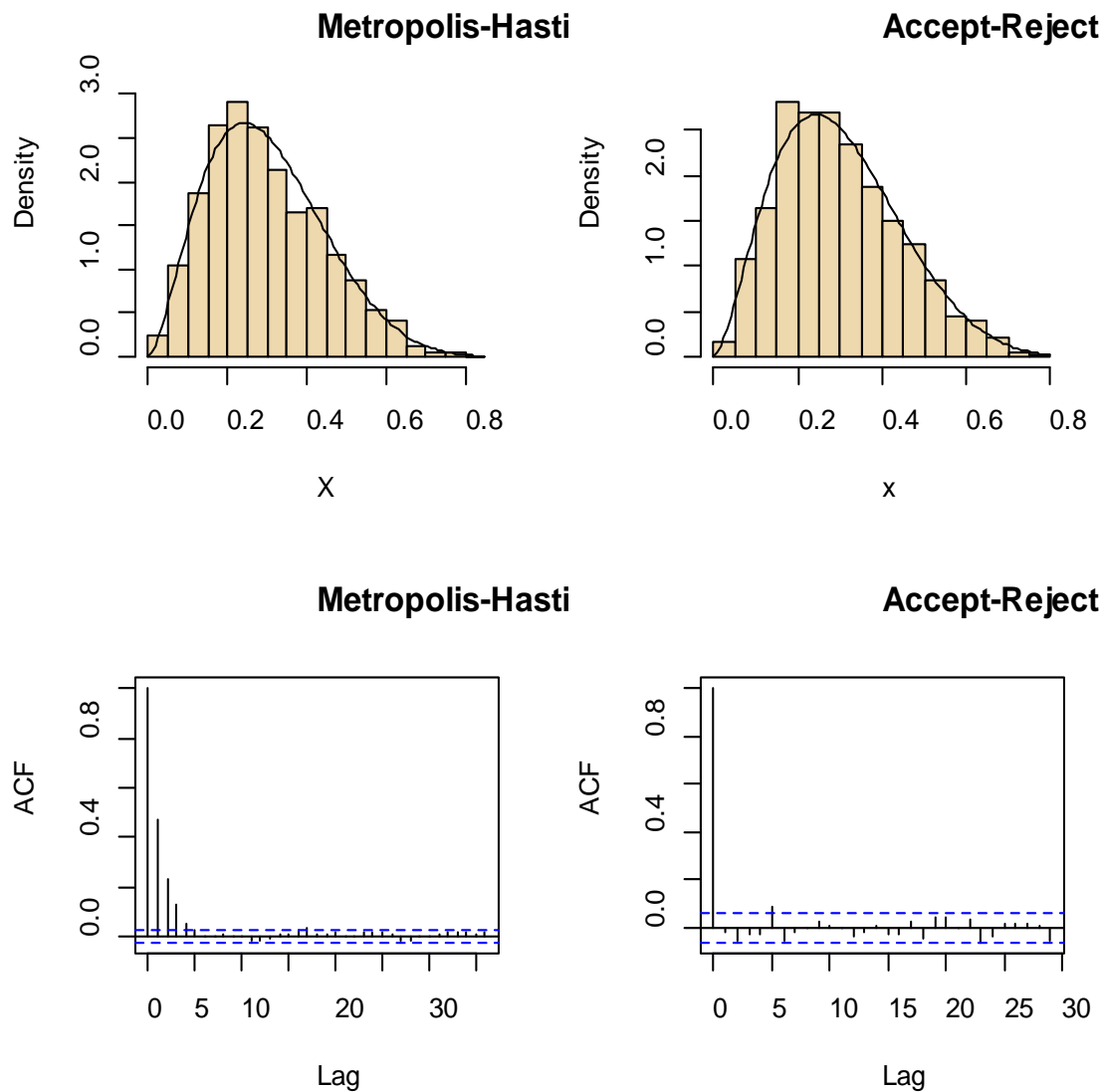
Μπορούμε να κατανοήσουμε καλύτερα τη σύγκριση αυτών των δύο μεθόδων με ένα παράδειγμα που στο οποίο θα συγκρίνουμε τις προσομοιωμένες τιμές τους.

**Παράδειγμα 3.3:** Στο προηγούμενο παράδειγμα εφαρμόσαμε έναν αλγόριθμο Metropolis-Hastings για τη παραγωγή τιμών από μία Βήτα κατανομή  $Beta(2.7, 6.3)$  ενώ στο Παράδειγμα 2.1 του Κεφαλαίου 2 εφαρμόσαμε μία μέθοδο Αποδοχής-Απόρριψης για την ίδια κατανομή. Σε αυτό το παράδειγμα θα συγκρίνουμε τις προσομοιωμένες τιμές που πήραμε από τη μέθοδο Αποδοχής-Απόρριψης με εκείνες από τη μέθοδο Metropolis-Hastings. Η σύγκριση αυτή γίνεται πιο κατανοητή και από τη γραφική αναπαράσταση των δύο αυτών δειγμάτων (Σχήμα 3.6).



```
> par(mfrow=c(2,2))
> hist(X,freq=FALSE,col="wheat2")
> curve(dbeta(x,2.7,6.3),add=TRUE)
> par(mfrow=c(2,2))
> hist(X,freq=FALSE,main="Metropolis-Hastings",col="wheat2")
> curve(dbeta(x,2.7,6.3),add=TRUE)
> hist(x,freq=FALSE,main="Accept-Reject",col="wheat2")
> curve(dbeta(x,2.7,6.3),add=TRUE)
> acf(X,main="Metropolis-Hastings")
> acf(x,main="Accept-Reject")
```

Άρα, σε αυτό το παράδειγμα, όπως φαίνεται και από το κάτωθι σχήμα, διαπιστώνουμε πως ο αλγόριθμος Metropolis-Hastings έχει παρόμοια προσαρμογή με τη μέθοδο Αποδοχής-Απόρριψης, αν και φαίνεται ξεκάθαρα η κυριότερη αναμενόμενη διαφορά τους: στις συναρτήσεις αυτοσυσχέτισης οι προσομοιωμένες τιμές του Metropolis-Hastings έχουν συσχέτιση ενώ εκείνες της μεθόδου Αποδοχής-Απόρριψης όχι.



**Σχήμα 3.6:** Ιστογράμματα με επικαλυπτόμενες τις σ.π.π. και συναρτήσεις αυτοσυσχέτισης για μία Βήτα κατανομή  $Beta(2.7, 6.3)$  για έναν αλγόριθμο Metropolis-Hastings (αριστερή στήλη) και για μία μέθοδο Αποδοχής-Απόρριψης (δεξιά στήλη).

Έτσι, παρατηρούμε κάποιες στοιχειώδεις διαφορές αυτών των 2 μεθόδων:

- a) Το τυχαίο δείγμα της μεθόδου Αποδοχής-Απόρριψης είναι ανεξάρτητο και ισόνομο σε αντίθεση με αυτό του Metropolis-Hastings. Παρόλο που τα  $Y^t$  παράγονται ανεξάρτητα, το δείγμα που προκύπτει δεν είναι ανεξάρτητο και ισόνομο αφού η πιθανότητα αποδοχής των  $Y^t$  εξαρτάται από τα  $X_t$ .
- b) Το δείγμα του Metropolis-Hastings θα περιλαμβάνει επαναλαμβανόμενες εμφανίσεις ίδιων τιμών αφού η απόρριψη των  $Y^t$  θα οδηγήσει στην επανάληψη του  $X_t$  σε χρόνο  $t + 1$ .

- c) Το βήμα της μεθόδου Αποδοχής-Απόρριψης απαιτεί τον υπολογισμό του άνω φράγματος  $M \geq \sup \frac{f(x)}{g(x)}$  που δεν είναι απαραίτητος στον αλγόριθμο

Metropolis-Hastings. Αυτό καθιστά τη μέθοδο Metropolis-Hastings πολύ πιο γρήγορη και αποτελεσματική στις περιπτώσεις εκείνες όπου ο υπολογισμός του  $M$  είναι χρονοβόρος ή το  $M$  είναι ανακριβές αφού έτσι παρατηρείται μια σπατάλη των προσομοιώσεων.

**Παρατήρηση 3.3:** Στην περίπτωση του ανεξάρτητου Metropolis-Hastings, δηλαδή όταν  $q(y|x) = g(y)$ , το ποσοστό αποδοχής είναι μεγαλύτερο με τη χρήση του Metropolis-Hastings αλγόριθμου από ότι με τη μέθοδο Αποδοχής-Απόρριψης. Αυτό συμβαίνει γιατί:

$$\rho(x, y) = \min \left( \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right) = \min \left( \frac{f(y) g(x)}{f(x) g(y)}, 1 \right)$$

Αφού η πιθανότητα αποδοχής ικανοποιεί τη σχέση:

$$\frac{f(y) g(x)}{f(x) g(y)} \geq \frac{f(y)/g(y)}{\max f(x)/g(x)},$$

άρα η πιθανότητα αποδοχής είναι μεγαλύτερη στον ανεξάρτητο Metropolis-Hastings από ότι στην μέθοδο Αποδοχής-Απόρριψης.

## 5.2. Τυχαίος Περίπατος Metropolis-Hastings

Η μελέτη του ανεξάρτητου Metropolis-Hastings είναι αρκετά ενδιαφέρουσα όμως από το να επιλέξουμε από την αρχή μια κατανομή εισήγησης είναι πολύ πιο ρεαλιστικό να συγκεντρώσουμε πληροφορίες σταδιακά με την ταυτόχρονη εξερεύνηση της περιοχής της τρέχουσας τιμής της αλυσίδας. Αν ο μηχανισμός εξερεύνησης έχει αρκετή ενέργεια για φτάσει μέχρι τα όρια της περιοχής της  $f$  τότε η μέθοδος τελικά θα αποκαλύψει την πολυπλοκότητα της  $f$ .

Μια πιο δημοφιλής προσέγγιση για την πρακτική κατασκευή μιας κατανομής εισήγησης του αλγόριθμου Metropolis-Hastings είναι να λάβουμε υπόψη μας την τιμή που προσομοιώθηκε προηγουμένως για να παράγουμε την επόμενη τιμή. Δηλαδή να θεωρήσουμε μια τοπική εξερεύνηση της «γειτονιάς» της τρέχουσας τιμής της Μαρκοβιανής αλυσίδας. Η εκτέλεση, δηλαδή, αυτής της ιδέας είναι να προσομοιώσουμε το  $Y^t$  σύμφωνα με τη σχέση  $Y^t = X_t + \varepsilon_t$ , όπου  $\varepsilon_t$  είναι μια τυχαία διαταραχή με κατανομή  $g$  ανεξάρτητη από την  $X_t$ . Η κατανομή εισήγησης του γενικού Metropolis-Hastings  $q(y|x)$  τώρα παίρνει την μορφή  $g(y - x)$ . Η Μαρκοβιανή αλυσίδα που συνδέεται με τη  $q$  είναι ένας τυχαίος περίπατος όταν η  $g$  είναι συμμετρική στο 0, δηλαδή όταν ικανοποιεί τη σχέση  $g(-t) = g(t)$ . Αλλά εξαιτίας του πρόσθετου βήματος αποδοχής του Metropolis-Hastings, η Μαρκοβιανή αλυσίδα  $X_t$

δεν είναι τυχαίος περίπατος. Αυτή η προσέγγιση οδηγεί στον παρακάτω Metropolis-Hastings αλγόριθμο που είναι και εκείνος που αρχικά πρότειναν οι Metropolis et al. (βλ. Κεφάλαιο 3.1 Αλγόριθμος Metropolis-Hastings)

#### **Αλγόριθμος 4: Τυχαίος περίπατος Metropolis-Hastings**

Με δοσμένες τις  $x_t$

1. Παρήγαγε  $Y^t \sim g(y - x_t)$

2. Πάρε

$$X_{t+1} = \begin{cases} Y^t & \text{με πιθανότητα } \min \left\{ 1, \frac{f(Y^t)}{f(x_t)} \right\} \\ x_t & \text{διαφορετικά} \end{cases}$$

**Παρατήρηση 3.4:** Όπως παρατηρούμε η πιθανότητα αποδοχής δεν εξαρτάται από τη  $g$ . Αυτό σημαίνει ότι για ένα δοσμένο ζευγάρι  $(x_t, y^t)$  η πιθανότητα αποδοχής είναι ίδια είτε π.χ. τα  $y^t$  παράγονται από μια Κανονική κατανομή είτε από μια Cauchy κατανομή. Βέβαια, δεν μπορούμε να πούμε ότι η επιλογή της  $g$  δεν έχει καμία επιρροή στη συμπεριφορά του αλγορίθμου, γιατί η αλλαγή της κατανομής εισήγησης θα καταλήξει σε διαφορετικές περιοχές τιμών για τα  $y^t$  και έτσι σε διαφορετικά ποσοστά αποδοχής.

**Παράδειγμα 3.4 (Συνέχεια Παραδείγματος 3.1):** Αν υποθέσουμε ότι η  $g$  είναι μια συμμετρική Κανονική σ.π.π. με μέσο ίσο με το 0 και τυπική απόκλιση  $C$ . Τότε η κατανομή εισήγησης θα έχει τη μορφή


$$g(y|x) = \frac{1}{C\sqrt{2\pi}} \exp \left\{ -\frac{1}{2C^2} (y - x)^2 \right\}$$

Σε αυτήν την περίπτωση το υποψήφιο  $y$  επιλέγεται από μια περιοχή της τρέχουσας τιμής  $x$  όπου το εύρος της περιοχής ελέγχεται από τη τυπική απόκλιση  $C$  της Κανονικής κατανομής. Αν επιλεγεί μια μικρή τιμή  $C$  τότε η υποψήφια κατανομή θα βρίσκεται σχετικά κοντά στη τρέχουσα τιμή και το ποσοστό αποδοχής του αλγορίθμου θα είναι υψηλό, σε αντίθεση με την επιλογή μιας μεγάλης τιμής του  $C$  που θα οδηγήσει σε ένα χαμηλό ποσοστό αποδοχής του αλγορίθμου.

Η εφαρμογή όλων των παραπάνω είναι σχετικά εύκολη στην R, απλά αλλάζοντας κάποιες γραμμές στον κώδικα του παραδείγματος 3.1. Οι μόνες αλλαγές που θα κάνουμε είναι στην κατανομή εισήγησης, όπου πλέον θα είναι μια Κανονική κατανομή με μέσο ίσο με τη τρέχουσα τιμή και τυπική απόκλιση  $C$ , και στην πιθανότητα αποδοχής, η οποία θα έχει τώρα μια πιο απλοποιημένη μορφή.

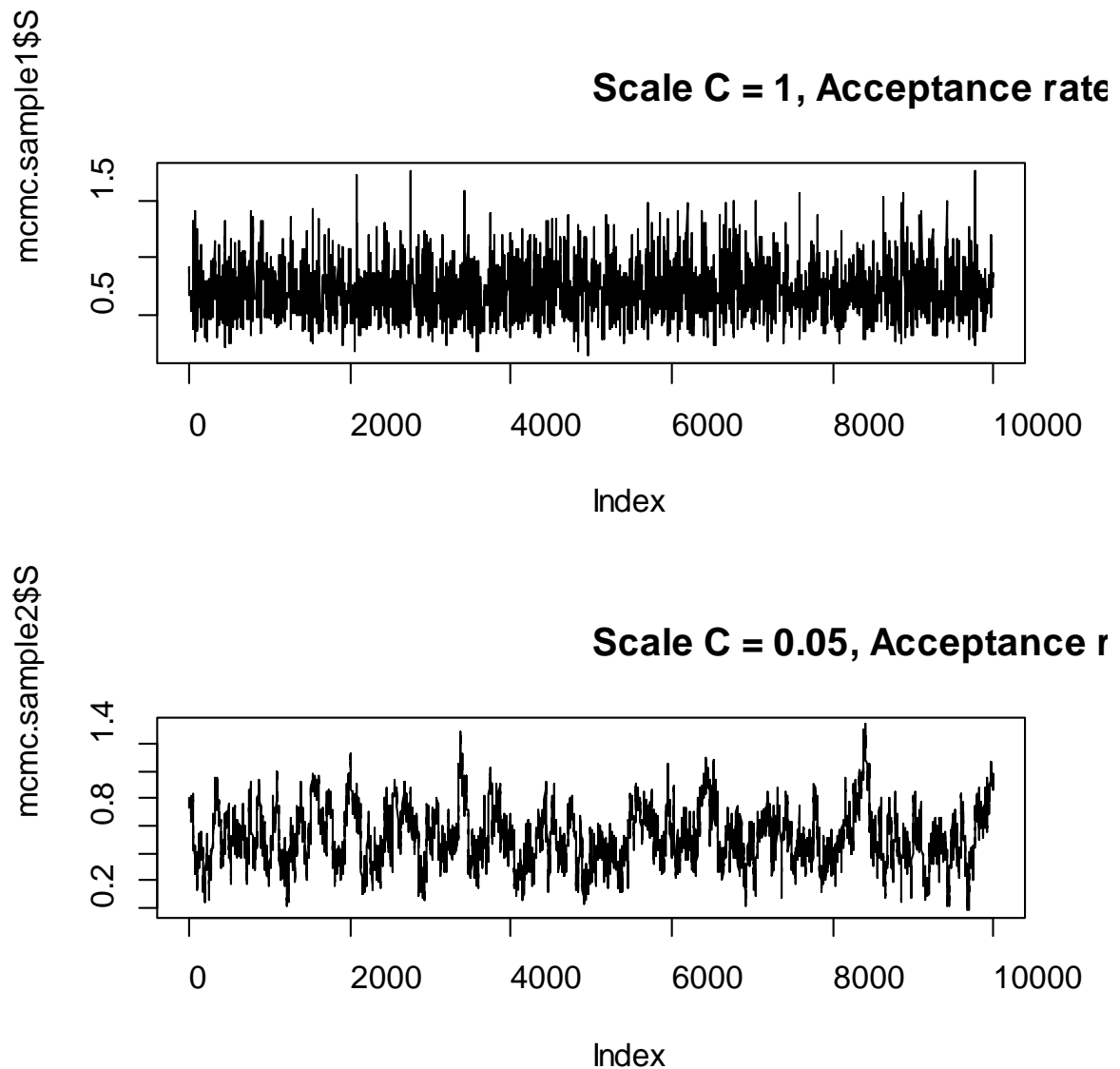


Έτσι, συνεχίζοντας το παράδειγμα της δειγματικής κατανομής μιας διάμεσου, για να χρησιμοποιήσουμε αυτόν τον τυχαίο περίπατο Metropolis-Hastings εισάγουμε μια τιμή της  $C$ . Παρακάτω έχουμε κάνει σύγκριση ανάμεσα σε 2 MCMC δείγματα για 10000 επαναλήψεις, ένα με τιμή  $C = 1$  και ένα δεύτερο με τιμή  $C = 0.05$ . Δηλαδή θα έχουμε τον παρακάτω κώδικα στην R:



```
> metrop.hasting.rw = function(logf, current, C, iter, n){
+ S = rep(0, iter)
+ n.accept = 0
+ for(j in 1:iter){
+ candidate = rnorm(1, mean=current, sd=C)
+ prob = exp(logf(candidate, n) - logf(current, n))
+ accept = ifelse(runif(1) < prob, "yes", "no")
+ current = ifelse(accept == "yes", candidate, current)
+ S[j] = current; n.accept = n.accept + (accept == "yes")
+ }
+ list(S=S, accept.rate=n.accept / iter)
+ }
> mcmc.sample1 = metrop.hasting.rw(log.samp.med, 1, 1, 10000, 21)
> mcmc.sample2 = metrop.hasting.rw(log.samp.med, 1, 0.05, 10000, 21)
> par(mfrow=c(1,2))
> plot(mcmc.sample1$S, type="l", main=paste(
+ "Scale C = 1, Acceptance rate =",
+ round(mcmc.sample1$accept.rate,2)))
> par(mfrow=c(2,1))
> plot(mcmc.sample1$S, type="l", main=paste(
+ "Scale C = 1, Acceptance rate =",
+ round(mcmc.sample1$accept.rate,2)))
> plot(mcmc.sample2$S, type="l", main=paste(
+ "Scale C = 0.05, Acceptance rate =",
+ round(mcmc.sample2$accept.rate,2)))
```

Όπως διαπιστώνουμε και από το Σχήμα 3.7 όταν επιλέγεται μια μικρή τιμή  $C = 0.05$ , το ποσοστό αποδοχής είναι υψηλό (92%) το οποίο συνεπάγεται ότι οι προσομοιωμένες τιμές είναι ισχυρά θετικά συσχετισμένες και ο αλγόριθμος εξερευνά αργά την περιοχή της ζητούμενης συνάρτησης. Αντίθετα, με την επιλογή της  $C = 1$ , το ποσοστό αποδοχής είναι μικρό (25%), οι προσομοιωμένες τιμές είναι ελάχιστα συσχετισμένες και ο αλγόριθμος μπορεί να κινηθεί στις περιοχές όπου η συνάρτηση έχει μεγαλύτερη πιθανότητα.



**Σχήμα 3.7:** Γραφήματα των προσομοιωμένων τιμών του τυχαίου περιπάτου Metropolis-Hastings χρησιμοποιώντας δύο επιλογές για τη παράμετρο  $C$ . Η επιλογή  $C = 0.05$  οδηγεί σε ένα υψηλό ποσοστό αποδοχής και πιο μικρή κίνηση μέσα στο χώρο των μεταβλητών σε αντίθεση με την επιλογή  $C = 1$  που οδηγεί σε μεγαλύτερη κίνηση του αλγορίθμου.

Άρα συμπεραίνουμε ότι η τιμή αυτού του μεγέθους  $C$  είναι ζωτικής σημασίας στο να επιτύχουμε μία καλή μίξη της αλυσίδας που θέλουμε σε ένα λογικό αριθμό επαναλήψεων.


■

**Παράδειγμα 3.5:** Το ιστορικό παράδειγμα του Hastings (1970) θεωρεί το τυπικό πρόβλημα της παραγωγής από μια Κανονική κατανομή  $\mathcal{N}(0,1)$  που βασίζεται στον τυχαίο περίπατο με κατανομή βήματος την ομοιόμορφη στο διάστημα  $[-\delta, \delta]$ . Δηλαδή, ας θεωρήσουμε το  $x$  να είναι η κατάσταση τη χρονική στιγμή  $t$  και επιλέγουμε  $y$  να είναι ομοιόμορφα κατανεμημένο στο διάστημα  $[x - \delta, x + \delta]$  όπου  $\delta > 0$  μία σταθερά.

Έτσι, σύμφωνα με τον αλγόριθμο 4, και εφόσον ξέρω ότι για τη σ.π.π. της Κανονικής Κατανομής ισχύει ότι  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ , άρα η πιθανότητα αποδοχής τότε θα είναι:

$$\rho(x, y) = \min\left\{\frac{f(y)}{f(x)}, 1\right\} = \min\left\{\frac{\exp(-y^2/2)}{\exp(-x^2/2)}, 1\right\} = \exp\left\{\frac{(x^2 - y^2)}{2}, 1\right\}.$$

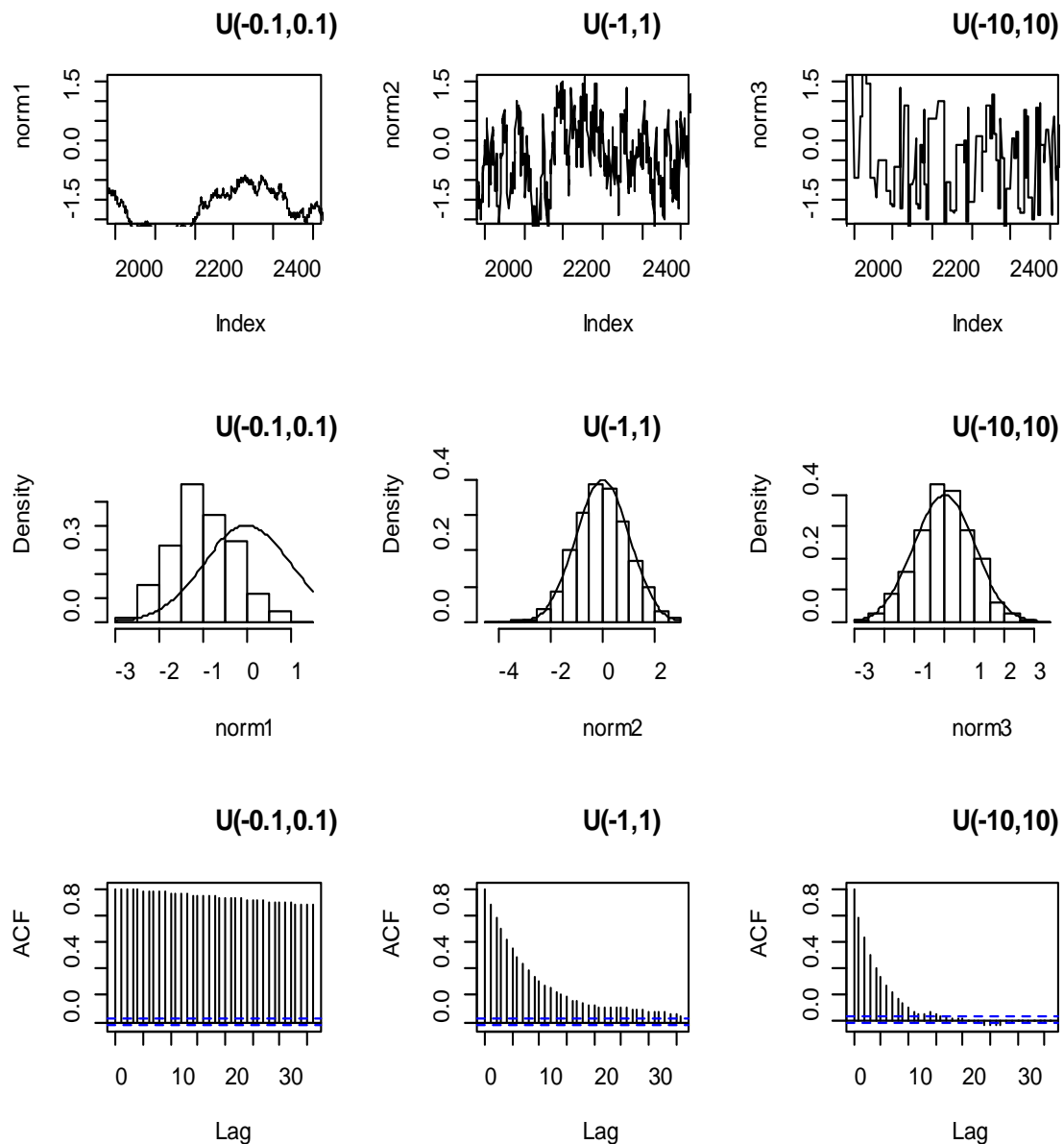
Το Σχήμα 3.8 περιγράφει 3 δείγματα 5000 τιμών που παρήχθησαν με αυτή τη μέθοδο για  $\delta = 0.1, 1, 10$  και δείχνει ξεκάθαρα τη διαφορά στις παραγόμενες αλυσίδες: αν η συνάρτηση βήματος είναι πολύ στενή ή πολύ ευρεία (δηλαδή αν το  $\delta$  είναι πολύ μικρό ή πολύ μεγάλο αντίστοιχα) τότε αυτό συνεπάγεται μεγαλύτερη αυτοσυσχέτιση και πιο αργή σύγκλιση. Ας παρατηρήσουμε ειδικά τα πρώτα γραφήματα για τις περιπτώσεις  $\delta = 0.1$  και  $\delta = 10$ : Στην πρώτη περίπτωση η Μαρκοβιανή αλυσίδα κινείται σε κάθε επανάληψη αλλά πολύ αργά, ενώ στην επόμενη περίπτωση παραμένει σταθερή για μεγάλα χρονικά διαστήματα. Σύμφωνα με όλα τα παραπάνω προκύπτει ο παρακάτω κώδικας στην R:



```
> norm<-function (n,delta){
+ vec <- vector("numeric", n)
+ x <- 0
+ vec[1] <- x
+ for (i in 2:n) {
+ innov <- runif(1, -delta,delta)
+ can <- x + innov
+ aprob <- min(1,(dnorm(can)/dnorm(x)))
+ u <- runif(1)
+ if (u < aprob)
+ x <- can
+ vec[i] <- x }
+ vec }
> norm1<-norm(5000, 0.1)
> norm2<-norm(5000, 1)
> norm3<-norm(5000, 10)
> par(mfrow=c(3,3))
> plot(norm1,xlim=c(2000,2500),ylim=c(-1.5,2),type="l",main="U(-0.1,0.1)")
> plot(norm2,xlim=c(2000,2500),ylim=c(-1.5,2),type="l",main="U(-1,1)")
> plot(norm3,xlim=c(2000,2500),ylim=c(-1.5,2),type="l",main="U(-10,10)")
> hist(norm1,freq=FALSE,main="U(-0.1,0.1)")
> curve(dnorm(x,0,1),add=TRUE)
```



```
> hist(norm2,freq=FALSE,main="U(-1,1)")
> curve(dnorm(x,0,1),add=TRUE)
> hist(norm3,freq=FALSE,main="U(-10,10)")
> curve(dnorm(x,0,1),add=TRUE)
> acf(norm1,main="U(-0.1,0.1)")
> acf(norm2,main="U(-1,1)")
> acf(norm3,main="U(-10,10)")
```



**Σχήμα 3.8:** Αποτελέσματα του τυχαίου περίπατου Metropolis-Hastings για παραγωγή δείγματος από την κανονική κατανομή  $\mathcal{N}(0,1)$ . Η αριστερή στήλη έχει κατανομή βήματος την  $\mathcal{U}(-0.1,0.1)$ , η μεσαία την  $\mathcal{U}(1,1)$  και η δεξιά την  $\mathcal{U}(-10,10)$ . Τα πάνω γραφήματα αναπαριστούν τις τελευταίες 500 επαναλήψεις της αλυσίδας, τα μεσαία γραφήματα δείχνουν πόσο καλά προσαρμόζονται τα ιστογράμματα στη ζητούμενη Κανονική κατανομή και τα τελευταία δείχνουν τις αντίστοιχες αυτοσυσχετίσεις.

Η ομοιόμορφη κατανομή με το πιο μικρό βήμα,  $\mathcal{U}(-0.1, 0.1)$ , κάνει μικρά βήματα και δεν φτάνει στις ουρές της Κανονικής κατανομής. Συγκριτικά, η ομοιόμορφη κατανομή με το μεγαλύτερο βήμα,  $\delta = 10$ , έχει φτάσει στις περιοχές των ουρών της κανονικής κατανομής αλλά από τη γραφική αναπαράσταση των τιμών της βλέπουμε ότι υπάρχουν περίοδοι όπου οι προτεινόμενες τιμές απορρίπτονται πολλές φορές. Αυτό σημαίνει ότι συγκεκριμένες επικρατούσες τιμές θα επανεμφανίζονται πολλές φορές, παρόλο που το ιστόγραμμα έχει εξομαλύνει τέτοιου είδους παρατηρήσεις.

■

### 5.2.1 Μειονεκτήματα Του Τυχαίου Περίπατου

Ο τυχαίος περίπατος Metropolis-Hastings, σε αντίθεση με τον ανεξάρτητο Metropolis-Hastings, είναι μια γενική μέθοδος και εφαρμόζεται σε πολλές περιπτώσεις. Παρόλα αυτά η λύση του τυχαίου περιπάτου δεν είναι πάντα και η πιο αποτελεσματική επιλογή γιατί:

1. Απαιτεί πολλές επαναλήψεις για να ξεπεράσει κάποιες δυσκολίες όπως για παράδειγμα διάφορες εκτελέσιμες περιοχές του αλγόριθμου με μικρή πιθανότητα.
2. Εξαιτίας των συμμετρικών του χαρακτηριστικών ξοδεύει σχεδόν το μισό χρόνο προσομοίωσης επανεξετάζοντας ξανά περιοχές που ήδη έχει επισκεφτεί. Υπάρχουν, βέβαια, εναλλακτικές που παρακάμπτουν αυτήν την τέλεια συμμετρία στην κατανομή εισήγησης του τυχαίου περιπάτου και πραγματοποιούν γρήγορη αποτελεσματικότητα αλλά είναι δύσκολες στην εφαρμογή τους. Μία από αυτές τις τεχνικές είναι ο *Langvin* αλγόριθμος των Roberts and Rosenthal (1998) που με την εισαγωγή μιας διαβάθμισης αναγκάζει τον αλγόριθμο να κινηθεί προς τις υψηλότερες τιμές της  $f$ . Ειδικά στην περίπτωση του τυχαίου περιπάτου, η τροποποίηση του μπορεί να εμποδίσει περαιτέρω την κινητικότητα της αλυσίδας ενισχύοντας έτσι την πόλωση των τοπικών επικρατουσών τιμών. Δηλαδή, η τεχνική αυτή, παρόλο που θεωρείται μια αρκετά εύκολη τροποποίηση του τυχαίου περιπάτου, απαιτεί πολύ λεπτές ρυθμίσεις και γι αυτό δεν είναι η ευρεία η εφαρμογή της.

## 6. Ποσοστό αποδοχής

Όπως είδαμε και από τα παραπάνω υπάρχουν πολλές δυνατότητες επιλογής της κατανομής εισήγησης  $q$  σε έναν Metropolis-Hastings αλγόριθμο και ο σκοπός μας είναι η βέλτιστη επιλογή. Εμφανώς σαν βέλτιστη επιλογή, ειδικά όσον αφορά την ταχύτητα σύγκλισης, θα έπρεπε να ορίσουμε  $q = f$  αλλά αυτό δεν έχει κανένα αντίκρισμα στην πράξη. Έτσι, είναι επιτακτική η ανάγκη ενός πρακτικού κριτηρίου που θα μας επιτρέψει τη σύγκριση διαφόρων κατανομών εισήγησης στις περιπτώσεις που γνωρίζουμε ελάχιστα για την κατανομή από την οποία θέλουμε να προσομοιώσουμε, δηλαδή την  $f$ . Ένα τέτοιο κριτήριο είναι το *ποσοστό αποδοχής* του Metropolis-Hastings αφού είναι εύκολα υπολογίσιμο. Ωστόσο, η απλή βελτιστοποίηση του ποσοστού

αποδοχής δεν συνεπάγεται και τον καλύτερο αλγόριθμο τουλάχιστον στο πλαίσιο της σύγκλισης και της μίξης.

Ο ανεξάρτητος Metropolis-Hastings όπως είδαμε μειώνει τον αριθμό των επαναλαμβανόμενων τιμών στην αλυσίδα  $X_t$  και έτσι επιτυγχάνει την εξάλειψη του συσχετισμού. Άρα, μπορεί να βελτιστοποιηθεί ή τουλάχιστον να συγκριθεί μέσω του ποσοστού αποδοχής. Παρόλα αυτά κάτι τέτοιο δεν ισχύει για τους υπόλοιπους τύπους των αλγόριθμων Metropolis-Hastings, κυρίως για τον τυχαίο περίπατο.

Όντως, ο τυχαίος περίπατος Metropolis-Hastings απαιτεί μια διαφορετική προσέγγιση όσον αφορά το ποσοστό αποδοχής, αφού η κατανομή εισήγησης εξαρτάται από την τρέχουσα τιμή της αλυσίδας. Στην πραγματικότητα, λοιπόν, ένα υψηλό ποσοστό αποδοχής δεν σημαίνει απαραίτητα ότι ο αλγόριθμος έχει ικανοποιητική συμπεριφορά καθώς αυτό μπορεί να σημαίνει ότι η αλυσίδα κινείται πολύ αργά στην επιφάνεια της  $f$ . Συγκεκριμένα όταν οι τιμές  $x_t$  και  $y^t$  είναι κοντά, υπό την έννοια ότι  $f(x_t) \approx f(y^t)$ , τότε ο τυχαίος περίπατος Metropolis-Hastings αποδέχεται την  $y^t$  με πιθανότητα :

$$\min\left(\frac{f(y^t)}{f(x_t)}, 1\right) \approx 1$$

Άρα ένα υψηλό ποσοστό αποδοχής μπορεί να σημαίνει ένα κακό πρότυπο σύγκλισης αφού οι κινήσεις στη περιοχή της  $f$  μπορεί να είναι πιο περιορισμένες. Παρόλα αυτά, αυτό δεν συμβαίνει πάντα. Για παράδειγμα, όταν η  $f$  είναι σχεδόν επίπεδη, τα υψηλά ποσοστά αποδοχής δεν είναι ενδεικτικά οποιασδήποτε λάθος συμπεριφοράς. Αλλά, εκτός αν η  $f$  είναι τελείως επίπεδη (κάτι το οποίο σημαίνει ότι η  $f$  είναι ομοιόμορφη), υπάρχουν τμήματα της περιοχής για να εξερευνηθούν όπου η συνάρτηση  $f$  παίρνει μικρότερες τιμές και άρα όπου το ποσοστό αποδοχής θα έπρεπε να είναι μικρό. Ένα μεγάλο ποσοστό αποδοχής τότε υποδεικνύει ότι ο αλγόριθμος Metropolis-Hastings επισκέπτεται σπάνια ή σχεδόν καθόλου αυτά τα τμήματα της περιοχής.

Συγκριτικά, αν το μέσο ποσοστό αποδοχής είναι χαμηλό, οι αποδεκτές τιμές  $f(y^t)$  είναι μικρές σε σύγκριση με τις  $f(x_t)$ . Αυτό αντιστοιχεί στην περίπτωση εκείνη όπου ο τυχαίος περίπατος κινείται γρήγορα στην επιφάνεια της  $f$ , αφού φτάνει στα «όρια» της  $f$  αρκετά συχνά. Βέβαια και σε αυτό το σενάριο ένα χαμηλό ποσοστό αποδοχής δεν σημαίνει ότι η αλυσίδα επισκέπτεται ολόκληρη την περιοχή της συνάρτησης. Ακόμα και με ένα μικρό ποσοστό αποδοχής, μπορεί να «χάσει» ένα μικρό μεν αλλά απομακρυσμένο τμήμα της  $f$ . Παρόλα αυτά, ένα μικρό ποσοστό αποδοχής είναι ένα σχετικά μικρό πρόβλημα, όχι από την πλευρά του υπολογιστικού χρόνου, γιατί δηλώνει ρητά την αναγκαιότητα ενός μεγάλου αριθμού προσομοιώσεων.

Ίσως τελικά να είναι αρκετά δελεαστική η πρόταση του συντονισμού της κατανομής εισήγησης με βάση το ποσοστό απόρριψης σε προηγούμενες επαναλήψεις της Μαρκοβιανής αλυσίδας, αλλά μία τέτοια «αναδρομική ρύθμιση» δεν είναι απόλυτα έγκυρη, αφού μπορεί να διαταράξει τη στάσιμη κατανομή στην οποία συγκλίνει η αλυσίδα. Ωστόσο, ο καθορισμός της κατανομής εισήγησης, με βάση μία προκαταρκτική εκτέλεση, είναι επιτρεπτός. Αλλά αν η αρχική κατανομή δεν ήταν

ιδιαίτερα καλή, τότε η δειγματοληψία μπορεί να μην προέρχεται από το σύνολο της κατανομής και έτσι να αποτελέσει μία κακή οδηγία για το συντονισμό αυτό.

**Παράδειγμα 3.6** (Συνέχεια παραδείγματος 3.5): Οι 3 τυχαίοι περίπατοι του παραδείγματος 3.5 έχουν ποσοστά αποδοχής ίσα με:

```
> length(unique(norm1))/5000  
[1] 0.98  
> length(unique(norm2))/5000  
[1] 0.8014  
> length(unique(norm3))/5000  
[1] 0.1546
```



Αν παρατηρήσουμε, όμως, τα ιστογράμματα από το Σχήμα 3.8 διαπιστώνουμε μια διαφορετική εικόνα. Μόνο στη μεσαία περίπτωση (δηλαδή για κατανομή εισήγησης  $\mathcal{U}(-1,1)$ ) συμφωνεί το ποσοστό αποδοχής με την εικόνα του γραφήματος. Βλέπουμε, όμως, ότι ακόμα και το χαμηλότερο ποσοστό αποδοχής έχει καλύτερη προσαρμογή από ότι το υψηλότερο.

■


Το ερώτημα λοιπόν είναι ποιο θα πρέπει να είναι το ποσοστό αποδοχής έτσι ώστε να αποφθεχθούν και τα πολύ μεγάλα αλλά και τα πολύ μικρά ποσοστά αποδοχής. Οι Roberts et al. (1997) πρότειναν τη χρήση κατανομών με ποσοστά αποδοχής κοντά στο  $1/4$  για υψηλής διάστασης προβλήματα και περίπου ίσο με  $1/2$  για μοντέλα με διάσταση 1 ή 2. Παρόλο που αυτός ο κανόνας δεν είναι γενικός, αφού σχεδιάστηκε κυρίως για ένα Γκαουσιανό περιβάλλον, ορίζεται σαν ένας προεπιλεγμένος στόχος εκεί όπου μπορεί να επιτευχθεί (κάτι βέβαια που είναι αρκετά σπάνιο).

**Παράδειγμα 3.7:** Θα συγκρίνουμε τα ποσοστά αποδοχής για μία κατανομή Γάμμα  $\text{Gamma}(4.3, 6.2)$  χρησιμοποιώντας

- a) Μια μέθοδο Αποδοχής-Απόρριψης με κατανομή εισήγησης  $\text{Gamma}(4, 7)$ .
- b) Έναν Ανεξάρτητο Metropolis-Hastings με κατανομή εισήγησης τη  $\text{Gamma}(4, 7)$ .
- c) Έναν Ανεξάρτητο Metropolis-Hastings με κατανομή εισήγησης τη  $\text{Gamma}(5, 6)$ .


Σε κάθε περίπτωση θα εξετάσουμε τη σύγκλιση κατά τη διάρκεια των επαναλήψεων.

Έτσι για την περίπτωση α) θα ισχύει ο παρακάτω κώδικας στην R:




```
>g47=rgamma(5000,4,7)
> u=runif(5000,max=dgamma(g47,4,7))
> x=g47[u<dgamma(g47,4.3,6.2)]
> par(mfrow=c(1,3),mar=c(4,4,1,1))
> hist(x,freq=FALSE,xlab="",ylab="",col="wheat2",
+ main="Accept-Reject G(4,7)")
> curve(dgamma(x,4.3,6.2),lwd=2,col="sienna",add=TRUE)
> length(x)/5000
[1] 0.8484
```

Για την περίπτωση β) θα έχουμε:



```
>X=rep(0,5000)
> X[1]=rgamma(1,4.3,6.2)
> for (t in 2:5000){
+ rho=(dgamma(X[t-1],4,7)*dgamma(g47[t],4.3,6.2))/
+ (dgamma(g47[t],4,7)*dgamma(X[t-1],4.3,6.2))
+ X[t]=X[t-1]+(g47[t]-X[t-1])*(runif(1)<rho)
+ }
> hist(X,freq=FALSE,xlab="",ylab="",col="wheat2",
+ main="Metropolis-Hastings G(4,7)")
> curve(dgamma(x,4.3,6.2),lwd=2,col="sienna",add=TRUE)
> length(unique(X))/5000
[1] 0.7822
```

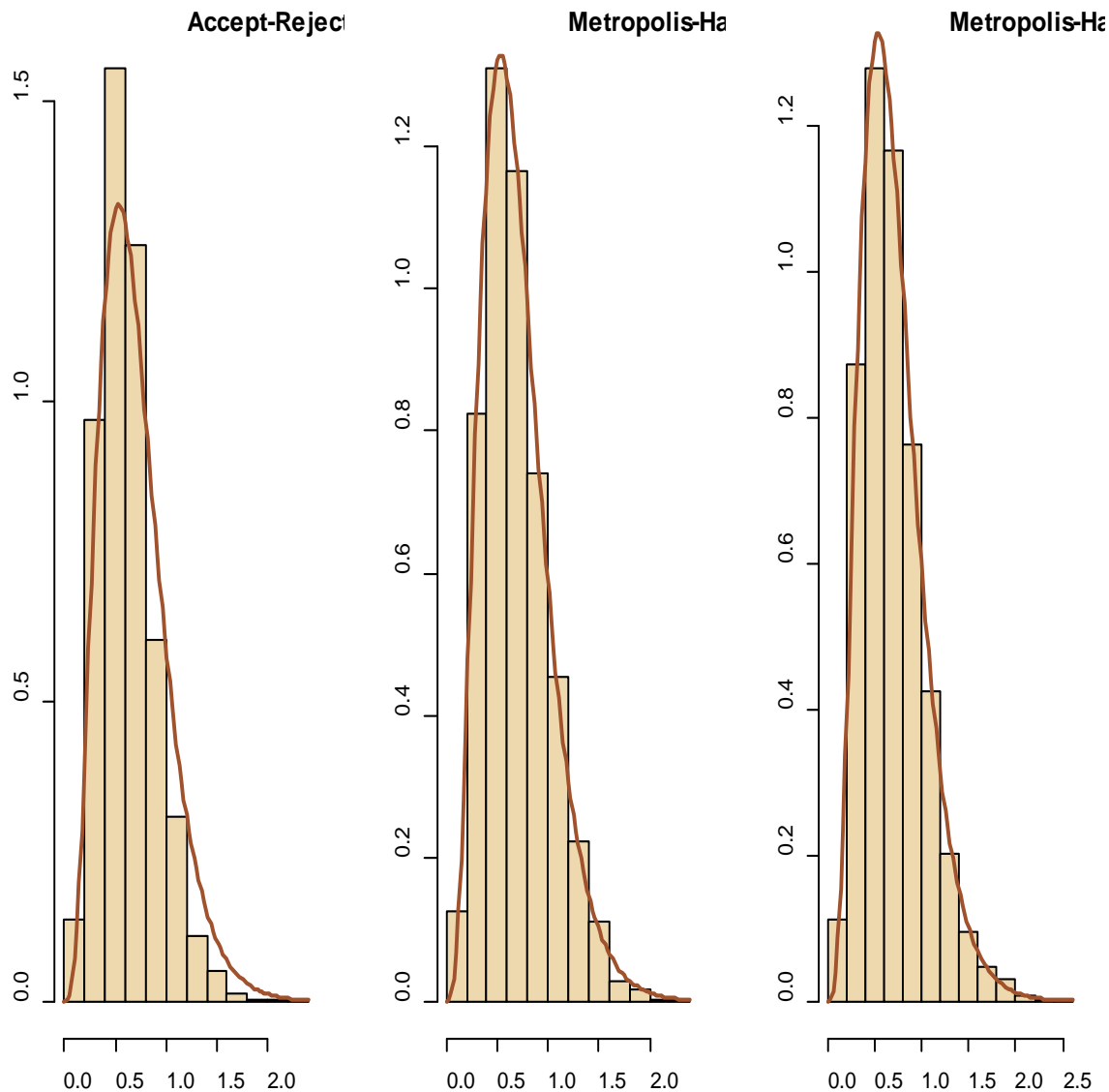
Τέλος, για την περίπτωση γ) θα ισχύει:



```
> g56=rgamma(5000,5,6)
> X[1]=rgamma(1,4.3,6.2)
> for (t in 2:5000){
+ rho=(dgamma(X[t-1],5,6)*dgamma(g56[t],4.3,6.2))/
+ (dgamma(g56[t],5,6)*dgamma(X[t-1],4.3,6.2))
+ X[t]=X[t-1]+(g56[t]-X[t-1])*(runif(1)<rho)
+ }
> hist(X,freq=FALSE,xlab="",ylab="",col="wheat2",
+ main="Metropolis-Hastings G((5,6))")
> curve(dgamma(x,4.3,6.2),lwd=2,col="sienna",add=TRUE)
> length(unique(X))/5000
[1] 0.772
```



Άρα, διαπιστώνουμε ότι με τον αλγόριθμο Metropolis-Hastings έχουμε περίπου τα ίδια ποσοστά αποδοχής ενώ από τα ιστογράμματα (Σχήμα 3.9) βλέπουμε ότι το μεσαίο ιστογράμμο (δηλαδή ο αλγόριθμος Metropolis-Hastings που χρησιμοποιεί για κατανομή εισήγησης την  $\mathcal{Gamma}(4,7)$ ) κάνει την καλύτερη προσαρμογή, σε αντίθεση με τη μέθοδο Αποδοχής-Απόρριψης, η οποία είναι η χειρότερη μέθοδος σε αυτήν την περίπτωση.



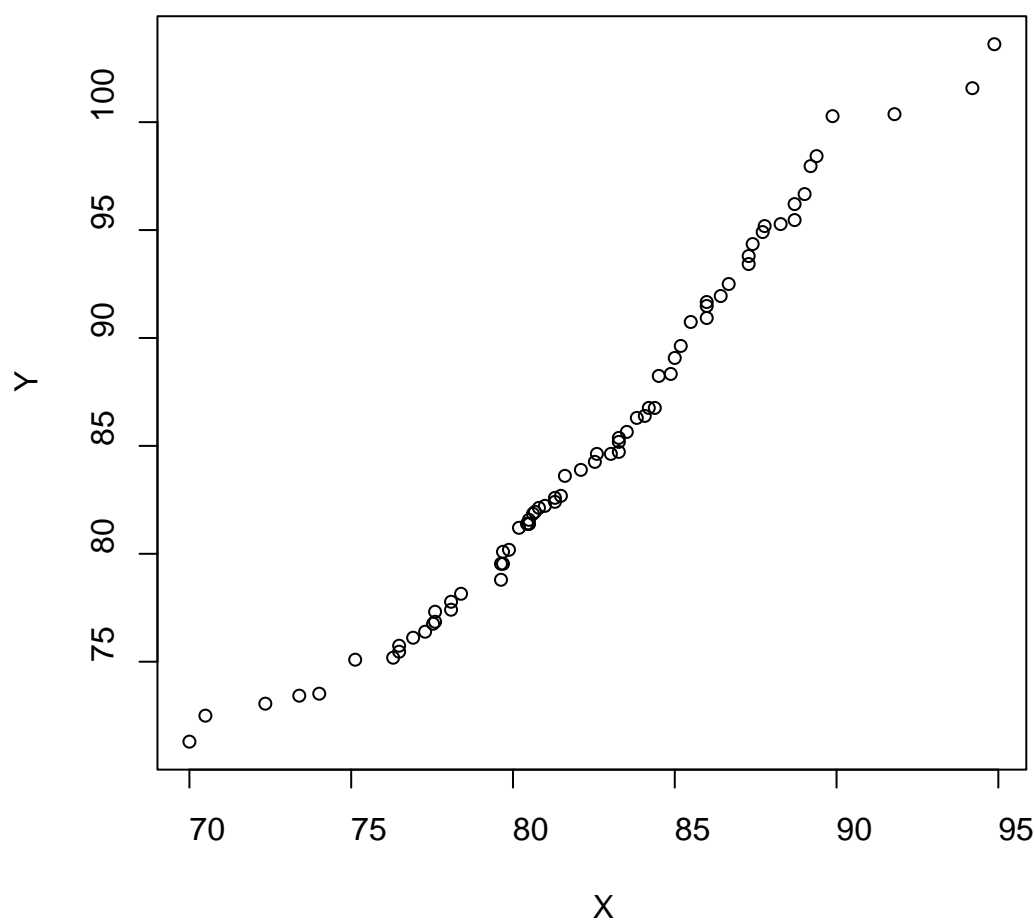
**Σχήμα 3.9:** Σύγκριση ποσοστών αποδοχής για την κατανομή  $\mathcal{G}(4.3, 6.2)$  με τις μεθόδους Αποδοχής-Απόρριψης και Metropolis-Hastings για διαφορετικές κατανομές εισήγησης.

## 7.Εφαρμογές στην R με τη μέθοδο Metropolis-Hastings

### Εφαρμογή 1: Ανορεξία

Η εφαρμογή αυτή αφορά τα δεδομένα μιας βιβλιοθήκης της R (MASS) την *anorexia*. Η βάση δεδομένων αναφέρεται σε 72 κοπέλες με ανορεξία που υποβλήθηκαν σε 3 θεραπείες και καταγράφηκε το βάρος τους πριν τη περίοδο της συγκεκριμένης μελέτης και μετά από αυτή. Θα ορίσουμε δύο συσχετισμένες μεταβλητές, τη μεταβλητή  $X$  για το βάρος τους πριν από τη μελέτη και τη μεταβλητή  $Y$  για το βάρος τους μετά τη μελέτη.

Αρχικά, διαπιστώνουμε από το Σχήμα 3.10 πως το μοντέλο μας είναι όντως Κανονικό αν εκτελέσουμε στην R τη συνάρτηση *qqplot*.



**Σχήμα 3.10:** QQplot για τα δεδομένα της βιβλιοθήκης *anorexia* στην εφαρμογή 1. Είναι φανερή η κανονικότητα των δεδομένων μας.

Τώρα, θα ήταν χρήσιμο να υπενθυμίσουμε τη συνάρτηση πυκνότητας πιθανότητας του διμεταβλητού μοντέλου σε μορφή πίνακα (αν αγνοήσουμε τη σταθερά κανονικοποίησης).

Έτσι, αν

$$x = \begin{pmatrix} x \\ y \end{pmatrix}, \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \text{ και } \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

τότε η σ.π.π. της  $x$  είναι ανάλογη με ποσότητα:

$$f(x) \propto \exp \left[ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]$$

Από το κάτωθι κώδικα της R βρίσκουμε τις διάφορες εκτιμήσεις από τα δεδομένα μας για αυτή τη κανονική διμεταβλητή κατανομή.

```
> library(MASS)
> X<-anorexia$Prewt
> Y<-anorexia$Postwt
> qqplot(X,Y)
> summary(X)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 70.00  79.60  82.30  82.41  86.00  94.90
> summary(Y)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 71.30  79.32  84.05  85.17  91.55 103.60
> cor(X,Y)
[1] 0.3324062
> sd(X)
[1] 5.182466
> sd(Y)
[1] 8.035173
```




Έτσι, στη συνέχεια θεωρώ το κάτωθι Κανονικό μοντέλο στην R με τις αντίστοιχες εκτιμήσεις από τα δεδομένα μας:

$$(X, Y) \sim \mathcal{N} \left( \begin{pmatrix} 82.41 \\ 85.14 \end{pmatrix}, \begin{pmatrix} 5.18^2 & 13.84 \\ 13.84 & 8.03^2 \end{pmatrix} \right)$$

Θα εκτελέσουμε σε αυτά τα δεδομένα αρχικά ένα τυχαίο περίπατο Metropolis-Hastings με κατανομές εισήγησης δύο ομοιόμορφες μεταβλητές με κέντρο το μέσο της διαφοράς (σε απόλυτη τιμή) των 2 μεταβλητών.

Αρχικά, λοιπόν, θα ορίσουμε δύο ομοιόμορφες κατανομές εισήγησης με κέντρο τη μέση απόλυτη τιμή της διαφοράς των  $X$  και  $Y$ , την οποία θα ονομάσουμε *width*. Αυτό γίνεται με σκοπό να έχουμε συμμετρικές κατανομές εισήγησης έτσι ώστε να μπορούμε

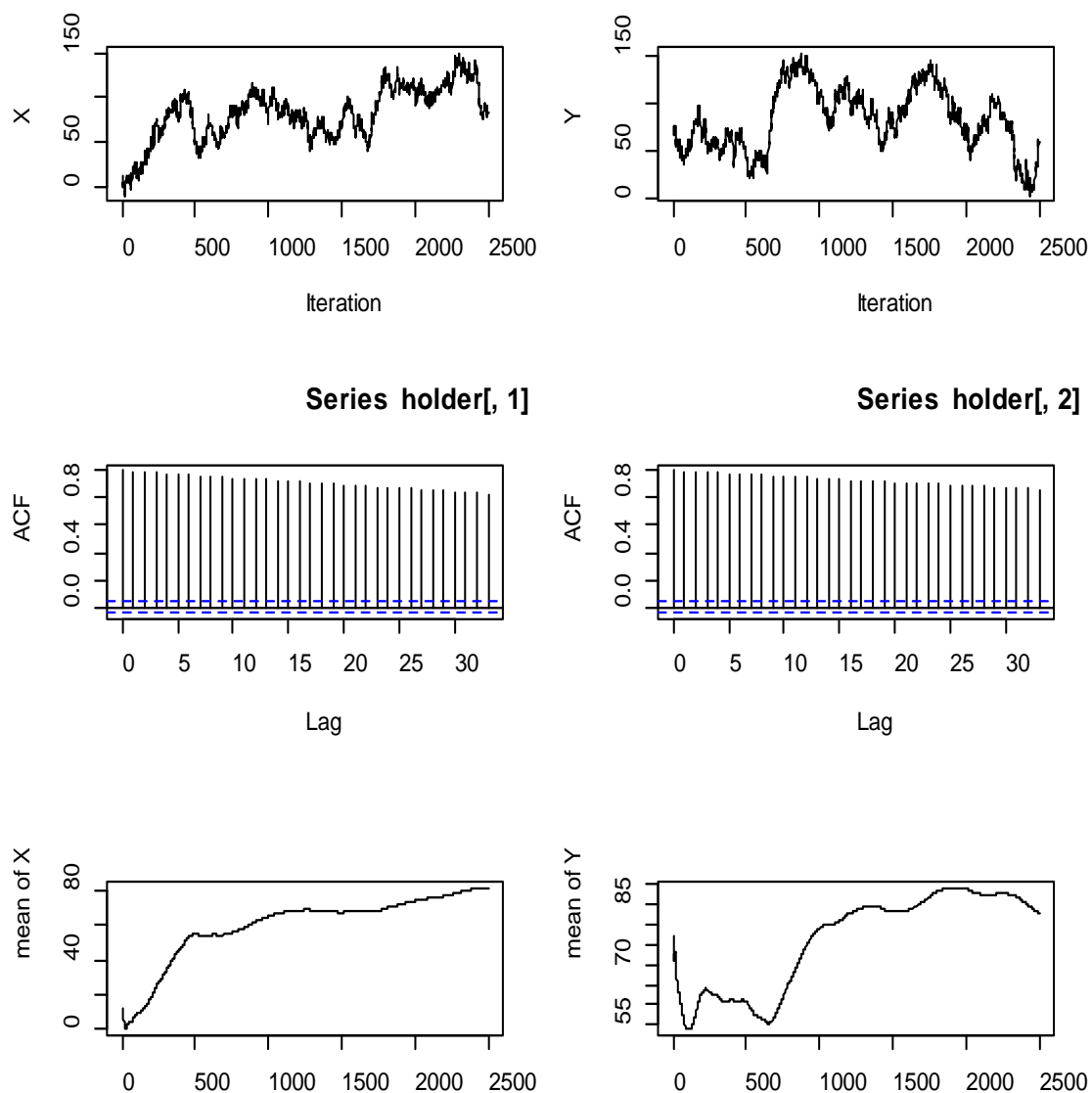
να εκτελέσουμε τον τυχαίο περίπατο Metropolis-Hastings. Σύμφωνα με όλα τα παραπάνω προκύπτει ο παρακάτω κώδικας στην R:



```
> mcmc <- 2500
> burnin <- 100
> count <- 1
> holder <- matrix(NA, mcmc, 2)
> w<-abs(Y-X)
> width<-mean(w)
> width
[1] 6.463889
> accept <- 0
> varmat <- matrix(c(var(X),cov(X,Y),cov(Y,X),var(Y)),2,2)
> invarmat <- solve(varmat)
> X.old<-X<-52
> Y.old<-Y<-85
>for(j in 1:(mcmc+burnin)) {
+ X.can <- runif(1, min=(X-width), max=(X+width))
+ Y.can <- runif(1, min=(Y-width), max=(Y+width))
+ logf.can <- -.5 * t(c(mean(X), mean(Y)) -
+ c(X.can,Y.can)) %*% invarmat %*% (c(mean(X), mean(Y)) - c(X.can,Y.can))
+ logf.cur <- -.5 * t(c(mean(X), mean(Y)) -
+ c(X,Y)) %*% invarmat %*% (c(mean(X), mean(Y)) - c(X,Y))
+ ratio <- exp(logf.can - logf.cur)
+ if (runif(1) < ratio){
+ X.old <- X
+ Y.old <- Y
+ X <- X.can
+ Y <- Y.can
+ accept <- accept+1}
+ if(j > burnin) {
+ holder[count,1] <- X
+ holder[count,2] <- Y
+ count <- count+1}
+ }
> print(apply(holder,2,mean))
[1] 109.88036 78.02561
> print(apply(holder,2,sd))
[1] 30.24964 32.78575
> print(cor(holder[,1],holder[,2]))
[1] 0.1543408
> print(accept/mcmc)
[1] 0.7408
> par(mfrow=c(3,2))
> par(mfrow=c(3,2))
> plot(1:mcmc, holder[,1], type="l", xlab="Iteration", ylab="X")
> plot(1:mcmc, holder[,2], type="l", xlab="Iteration", ylab="Y")
```



```
> acf(holder[,1],main="")
> acf(holder[,2],main="")
x1_cum<-rep(0,times=mcmc)
> x2_cum<-rep(0,times=mcmc)
> for(i in 1:mcmc){
+ x1_cum[i]<-mean(holder[1:i,1])
+ x2_cum[i]<-mean(holder[1:i,2])
+ }
> plot(x1_cum,type="l",xlab="",ylab="",main="Empirical mean of X")
> plot(x2_cum,type="l",xlab="",ylab="",main="Empirical mean of Y")
```



**Σχήμα 3.11:** Γραφική αναπαράσταση των τιμών X και Y και συναρτήσεις αυτοσυσχέτισης για το διμεταβλητό Κανονικό μοντέλο της Εφαρμογής 1 με ομοιόμορφες κατανομές εισήγησης.

Μπορούμε εύκολα να διαπιστώσουμε από το Σχήμα 3.11 ότι υπάρχει μικρή κίνηση ανάμεσα στις τιμές του αλγόριθμου, δηλαδή δεν εξερευνεί ολόκληρη την περιοχή της ζητούμενης Κανονικής συνάρτησης κατανομής, αφού οι προσομοιωμένες τιμές μας είναι ισχυρά συσχετισμένες. Διαπιστώνουμε και πάλι πόσο «παραπλανητικό» μπορεί να είναι το υψηλό ποσοστό αποδοχής. Βλέπουμε, ακόμα ότι ο εμπειρικός μέσος δεν έχει σταθεροποιηθεί και ίσως χρειάζονται περισσότερες επαναλήψεις. Ωστόσο, ακόμα και να αυξήσουμε τον αριθμό των επαναλήψεων οι τιμές της αλυσίδας μας πάλι θα παρουσιάζουν πολύ μεγάλη αυτοσυσχέτιση.

Άρα συμπεραίνουμε ότι ίσως ο εν λόγω τυχαίος περίπατος Metropolis-Hastings δεν είναι κατάλληλος και έτσι πρέπει να αλλάξουμε τη κατανομή εισήγησης. Αν θεωρήσουμε τη διμεταβλητή Κανονική κατανομή εισήγησης με

$$(X, Y) \sim \mathcal{N} \left( \begin{pmatrix} 81 \\ 84 \end{pmatrix}, \begin{pmatrix} 19 & 13 \\ 13 & 60 \end{pmatrix} \right),$$

τότε σύμφωνα με τις παραπάνω εκτιμήσεις θα εκτελέσουμε ένα παρόμοιο με τον προηγούμενο κώδικα στην R αλλά με διαφορετική τη κατανομή εισήγησης.

Επιλέξαμε την παραπάνω κατανομή εισήγησης μετά από πολλαπλές δοκιμές στις αντίστοιχες διασπορές των μεταβλητών μας και έτσι καταλήξαμε σε αυτές τις διασπορές που μας δίνουν ποσοστό αποδοχής μεγαλύτερο του 40%.



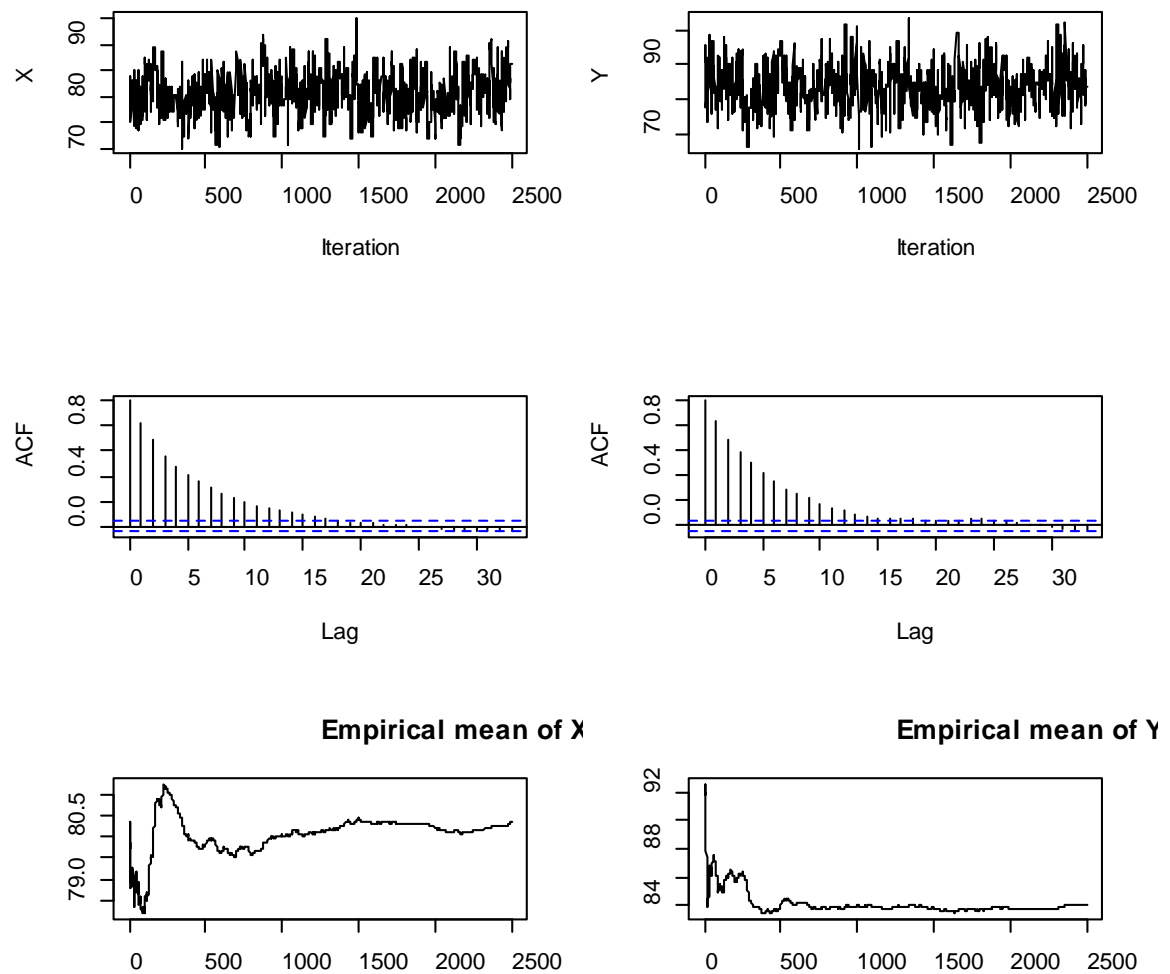
```
> mcmc <- 2500
> burnin <- 100
> count <- 1
> accept <- 0
> holder <- matrix(NA, mcmc, 2)
> varmat <- matrix(c(var(X), cov(X,Y), cov(Y,X), var(Y)), 2, 2)
> invarmat <- solve(varmat)
> mu <- c(81, 84)
> sigma <- matrix(c(19, 13, 13, 60), 2, 2)
> X.old <- X <- 80
> Y.old <- Y <- 83
> for(j in 1:(mcmc+burnin))
+ {
+   v <- mvrnorm(mu=mu, Sigma=sigma)
+   logf.can <- -.5 * t(c(mean(X), mean(Y)) - v) %*% invarmat %*% (c(mean(X),
+   mean(Y)) -
+   v)
+   logf.cur <- -.5 * t(c(mean(X), mean(Y)) -
+   c(X,Y)) %*% invarmat %*% (c(mean(X), mean(Y)) - c(X,Y))
+   ratio <- exp(logf.can - logf.cur)
+   if (runif(1) < ratio){
+     X.old <- X
+     Y.old <- Y
+     X <- v[1]
+     Y <- v[2]
+     accept <- accept+1
+   }
+   if(j > burnin)
+   {
+     holder[count,1] <- X
+     holder[count,2] <- Y
+     count <- count+1
+   }
+ }
> print(apply(holder, 2, mean))
[1] 80.83358 84.01896
> print(apply(holder, 2, sd))
[1] 4.402020 7.056169
> print(cor(holder[,1], holder[,2]))
[1] 0.3445883
> print(accept/mcmc)
[1] 0.4152
```



```
> par(mfrow=c(3,2))
> plot(1:mcmc, holder[,1], type="l", xlab="Iteration", ylab="X")
> plot(1:mcmc, holder[,2], type="l", xlab="Iteration", ylab="Y")
> acf(holder[,1],main="")
> acf(holder[,2],main="")
> x1_cum<-rep(0,times=mcmc)
> x2_cum<-rep(0,times=mcmc)
> for(i in 1:mcmc)
+ {
+ x1_cum[i]<-mean(holder[1:i,1])
+ x2_cum[i]<-mean(holder[1:i,2])
+ }
> plot(x1_cum,type="l",xlab="",ylab="",main="Empirical mean of X")
> plot(x2_cum,type="l",xlab="",ylab="",main="Empirical mean of Y")
```

Σε αυτήν την περίπτωση η αλυσίδα κινείται πιο εύκολα στις διάφορες τιμές της, αφού οι προσομοιωμένες τιμές δεν παρουσιάζουν τόσο μεγάλη αυτοσυσχέτιση όπως στην προηγούμενη περίπτωση. Ακόμη, όπως βλέπουμε και από το Σχήμα 3.12, ο εμπειρικός μέσος σταθεροποιείται στις αναμενόμενες τιμές σε ένα λογικό αριθμό επαναλήψεων. Από όλα τα παραπάνω, συμπεραίνουμε ότι αυτή η μέθοδος προσομοίωσης είναι τελικά αποτελεσματική.





**Σχήμα 3.12:** Γραφική αναπαράσταση των  $X$  και  $Y$  με Κανονική κατανομή εισήγησης για το διμεταβλητό μοντέλο της Εφαρμογής 1.

## Κεφάλαιο 4 Ο Δειγματολήπτης Gibbs

---

### 1.Εισαγωγή

---

Ο δειγματολήπτης Gibbs είναι μία ευρεία και διαδεδομένη μέθοδος MCMC για την παραγωγή τυχαίων μεταβλητών, χωρίς την ανάγκη υπολογισμού της συνάρτησης πυκνότητας πιθανότητας. Το ιδιαίτερο χαρακτηριστικό αυτών των αλγόριθμων είναι ότι πρώτα συγκεντρώνουν την περισσότερη πληροφορία από τη συνάρτηση που μας ενδιαφέρει και μετά επιτρέπουν τη διάσπαση πολύπλοκων προβλημάτων σε μια σειρά από απλούστερα προβλήματα, όπως για παράδειγμα μια ακολουθία κατανομών μικρής διάστασης. Βέβαια, δεν πρέπει να ξεχνάμε ότι υπάρχει πιθανότητα αυτή η ακολουθία κατανομών μικρής διάστασης να απαιτεί μεγάλο χρόνο σύγκλισης. Παρόλα αυτά, ο δειγματολήπτης Gibbs παράγει τιμές από τη δεσμευμένη κατανομή με τέτοιο τρόπο που να προσεγγίζει την από κοινού κατανομή μέσα σε ένα συγκεκριμένο χρόνο. Αυτό το χαρακτηριστικό αυτού του αλγόριθμου τον καθιστά ιδιαίτερα δημοφιλή στη Μπεϋζιανή στατιστική και γι αυτό ο μεγαλύτερος αριθμός των στατιστικών εφαρμογών των MCMC μεθόδων αφορά το δειγματολήπτη Gibbs.

Ο δειγματολήπτης Gibbs πήρε το όνομα του από το φυσικό J.W.Gibbs χάρη στην αναλογία που πρώτος εφάρμοσε ανάμεσα στον αλγόριθμο δειγματοληψίας και στη στατιστική φυσική. Το άρθρο σταθμός των Geman και Geman (1984), οι οποίοι πραγματοποίησαν πρώτοι έναν δειγματολήπτη Gibbs, ανέλυσε και περιέγραψε αυτόν τον αλγόριθμο. Είναι στην πραγματικότητα μια ειδική περίπτωση του αλγόριθμου Metropolis-Hastings, όπως περιγράφουν αναλυτικά οι Robert and Casella (2004). Αυτή η δουλειά των Geman και Geman, που βασίστηκε πάνω στην έρευνα των Metropolis et al. (1953), ενέπνευσε τους Gelfand και Smith (1990) να ερευνήσουν τις Μπεϋζιανές μεθόδους, τη στατιστική υπολογιστική, τους αλγόριθμους και τις στοχαστικές ανελίξεις μέσα από τη χρήση υπολογιστικών αλγόριθμων όπως ο δειγματολήπτης Gibbs και ο αλγόριθμος Metropolis-Hastings.

Συγκεκριμένα, ο δειγματολήπτης Gibbs είναι μια μέθοδος που δημιουργεί έναν αλγόριθμο παραγωγής τυχαίων μεταβλητών βασιζόμενο σε μία δειγματοληψία από ένα σύνολο δεσμευμένων συναρτήσεων κατανομών. Δηλαδή, η Μαρκοβιανή αλυσίδα κατασκευάζεται, με ένα ντετερμινιστικό ή τυχαίο τρόπο, από μια ακολουθία δεσμευμένων κατανομών. Είναι εμφανές ότι η μεθοδολογία του δειγματολήπτη Gibbs είναι αρκετά διαφορετική με αυτή του Metropolis-Hastings και ιδιαίτερα χρήσιμη για τη παραγωγή  $n$  –διάστατων τυχαίων διανυσμάτων. Αυτή η μέθοδος εφαρμόζεται όταν δεν γνωρίζουμε την από κοινού κατανομή ή είναι αδύνατη η δειγματοληψία που δημιουργεί ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα, αλλά η δεσμευμένη κατανομή της κάθε μεταβλητής δοθέντων των υπολοίπων είναι γνωστή και έχει απλή μορφή έτσι ώστε να μπορούμε εύκολα να προσομοιώσουμε τυχαίες τιμές. Αν δηλαδή γνωρίζουμε την πλήρη δεσμευμένη κατανομή, δηλαδή τη δεσμευμένη κατανομή μιας παραμέτρου

που είναι δεσμευμένη από όλες τις άλλες παραμέτρους, τότε μπορούμε να εφαρμόσουμε το δειγματολήπτη Gibbs.

Στην ουσία, ο δειγματολήπτης Gibbs είναι μία ειδική περίπτωση του αλγόριθμου Metropolis-Hastings με πιθανότητα αποδοχής ίση με 1. Αυτό σημαίνει πως οι υποψήφιος τιμές του δειγματολήπτη Gibbs είναι πάντα αποδεκτές. Έτσι, η συγκεκριμένη δειγματοληψία βασίζεται εξολοκλήρου στη δειγματοληψία από τις πλήρεις δεσμευμένες κατανομές.

Στη συνέχεια θα περιγράψουμε τον δειγματολήπτη Gibbs με δύο στάδια αλλά και αυτόν με τα πολλαπλά στάδια. Παρόλο που ο πρώτος είναι ειδική περίπτωση του δεύτερου, εκείνος με τα δύο στάδια έχει πολλές ιδιότητες σύγκλισης και εφαρμόζεται σε ένα ευρύ φάσμα στατιστικών μοντέλων που δεν απαιτούν το γενικό χαρακτήρα του δειγματολήπτη Gibbs πολλαπλών σταδίων. Παρόλα αυτά, ο τελευταίος διαθέτει και αυτός αρκετές ιδιότητες σύγκλισης και εξακολουθεί να θεωρείται μέχρι σήμερα η κινητήριος δύναμη του MCMC κόσμου.

## 2.Ο Δειγματολήπτης Gibbs Δύο Σταδίων

Ο Δειγματολήπτης Gibbs δύο σταδίων δημιουργεί μια Μαρκοβιανή Αλυσίδα από μια από κοινού κατανομή με τον παρακάτω τρόπο. Αν δύο τυχαίες μεταβλητές  $X$  και  $Y$  έχουν μια από κοινού συνάρτηση πυκνότητας πιθανότητας  $f(x, y)$  με αντίστοιχες δεσμευμένες σ.π.π.  $f_{Y|X}$  και  $f_{X|Y}$ , τότε ο δειγματολήπτης Gibbs είναι ο ακόλουθος:

### **Αλγόριθμος 5: Ο Δειγματολήπτης Gibbs 2-σταδίων**

Πάρε  $X_0 = x_0$

Για  $t = 1, 2, \dots$ , παρήγαγε

1.  $Y_t \sim f_{Y|X}(\cdot | x_{t-1})$  ;
2.  $X_t \sim f_{X|Y}(\cdot | y_t)$ .

Ο δειγματολήπτης Gibbs βασίζεται στην παρακάτω ιδέα: ανανεώνουμε το ένα μέρος του προηγούμενου στοιχείου ενώ κρατάμε τα άλλα μέρη σταθερά. Έτσι, όσο απλή είναι η προσομοίωση από τις δύο δεσμευμένες κατανομές τόσο απλή είναι και η εφαρμογή αυτού του αλγόριθμου. Επίσης είναι εύκολο να καταλάβουμε ότι αν η από κοινού σ.π.π. του  $(X_t, Y_t)$  είναι η  $f$  τότε είναι και του  $(X_{t+1}, Y_{t+1})$  γιατί και τα δύο βήματα της επανάληψης  $t$  παράγουν τιμές από τις πραγματικές δεσμευμένες κατανομές. Άρα, καθώς το  $t$  πλησιάζει το άπειρο η Μαρκοβιανή αλυσίδα θα συγκλίνει στην από κοινού κατανομή  $f$  που θέλουμε. Σε πολλά προβλήματα η σύγκλιση θα είναι

γρήγορη και έτσι οι παραγόμενες προσομοιωμένες τιμές  $(x_1, y_1), \dots, (x_t, y_t)$  θα μπορούν να θεωρηθούν σαν ένα κατά προσέγγιση δείγμα από την από κοινού κατανομή μας. Άρα, συμπεραίνουμε ότι η σύγκλιση του αλγόριθμου είναι εξασφαλισμένη εκτός και αν οι περιοχές των δεσμευμένων κατανομών δεν συνδέονται καθόλου μεταξύ τους. Στη συνέχεια, μπορούμε να χρησιμοποιήσουμε γραφικές μεθόδους για να επιβεβαιώσουμε ότι η αλυσίδα συγκλίνει σε μια στάσιμη κατανομή μετά από μια επαρκή περίοδο burn-in.

**Παράδειγμα 4.1:** Θα ξεκινήσουμε με ένα απλό παράδειγμα θεωρώντας ότι:

$$(X, Y) \sim \mathcal{N}_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Δηλαδή έχουμε το διμεταβλητό Κανονικό μοντέλο με  $\mu = 0$  και με αυτοσυσχέτιση  $\rho$ .

Εφόσον ισχύει ότι:

$$\begin{aligned} \mathbb{E}[Y|X = x] &= \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \\ \text{Var}[Y|X = x] &= \sigma_Y^2(1 - \rho^2) \end{aligned}$$

Άρα, για αυτήν τη διμεταβλητή Κανονική κατανομή ο δειγματολήπτης Gibbs θα είναι:

Δοσμένων των  $x_t$

$$Y_{t+1}|x_t \sim \mathcal{N}(\rho x_t, 1 - \rho^2)$$

$$X_{t+1}|y_{t+1} \sim \mathcal{N}(\rho y_{t+1}, 1 - \rho^2).$$

Τότε η υπο-αλυσίδα  $X_t$  ικανοποιεί

$$X_{t+1}|X_t = x_t \sim \mathcal{N}(\rho^2 x_t, 1 - \rho^4),$$

και μέσω της αναδρομικότητας αποδεικνύεται ότι:

$$X_t|X_0 = x_0 \sim \mathcal{N}(\rho^{2t} x_0, 1 - \rho^{4t}),$$

που όντως συγκλίνει στην  $\mathcal{N}(0,1)$  καθώς  $t \rightarrow \infty$ . Για την εφαρμογή αυτού του απλού δειγματολήπτη Gibbs δημιουργήσαμε τον παρακάτω κώδικα στην R όπου μέσα στο βρόγχο υπολογίζει τις απαραίτητες για αυτή τη δειγματοληψία δεσμευμένες σ.π.π. ενώ εφαρμόσαμε το δειγματολήπτη Gibbs για  $\rho = 0.98$  έτσι ώστε να παρατηρήσουμε τα αποτελέσματα του από μία Κανονική διμεταβλητή κατανομή.

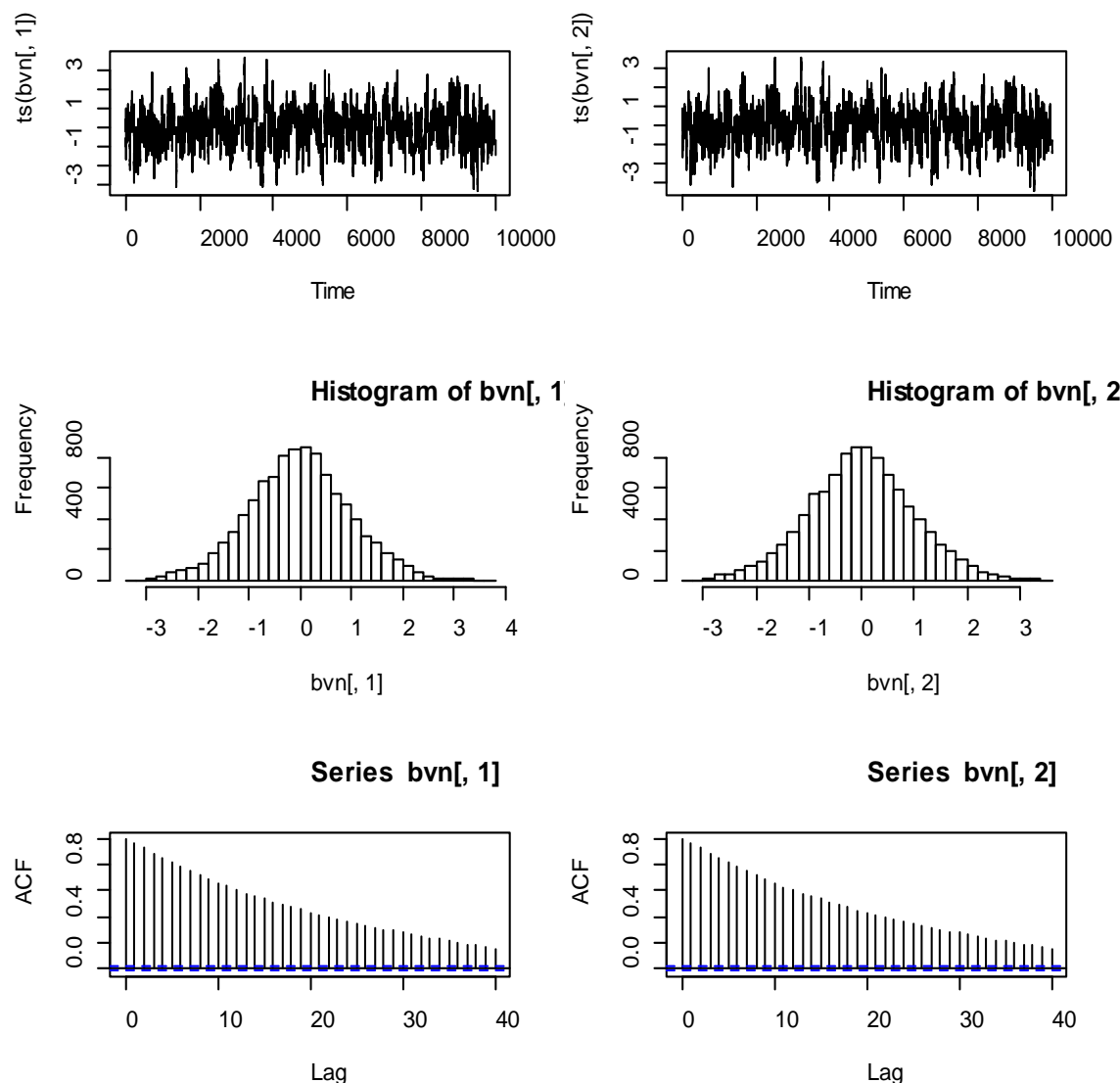


```

> simple.gibbs<-function (n, rho)
+ {
+ mat <- matrix(ncol = 2, nrow = n)
+ x <- 0
+ y <- 0
+ mat[1, ] <- c(x, y)
+ for (i in 2:n)
+ {
+ x <- rnorm(1, rho * y, sqrt(1 - rho^2))
+ y <- rnorm(1, rho * x, sqrt(1 - rho^2))
+ mat[i, ] <- c(x, y)
+ }
+ mat
+ }
> bvn<-gibbs(10000,0.98)
> summary(bvn)
      V1      V2
Min. :-3.42928 Min. :-3.62405
1st Qu.:-0.61616 1st Qu.:-0.62294
Median : 0.03409 Median : 0.02994
Mean : 0.01459 Mean : 0.01386
3rd Qu.: 0.64734 3rd Qu.: 0.65817
Max. : 3.36391 Max. : 3.29768
> par(mfrow=c(3,2))
> plot(ts(bvn[,1]))
> plot(ts(bvn[,2]))
> hist(bvn[,1],40)
> hist(bvn[,2],40)
> acf(bvn[,1])
> acf(bvn[,2])

```

Όπως διαπιστώνουμε από αυτό το παράδειγμα, η ακολουθία  $(X_t, Y_t)$  που παράγεται από μια δειγματοληψία Gibbs συγκλίνει στην από κοινού κατανομή  $f$  και σαν επακόλουθο και οι δύο ακολουθίες  $X_t$  και  $Y_t$  συγκλίνουν στην αντίστοιχη περιθώρια κατανομή τους.



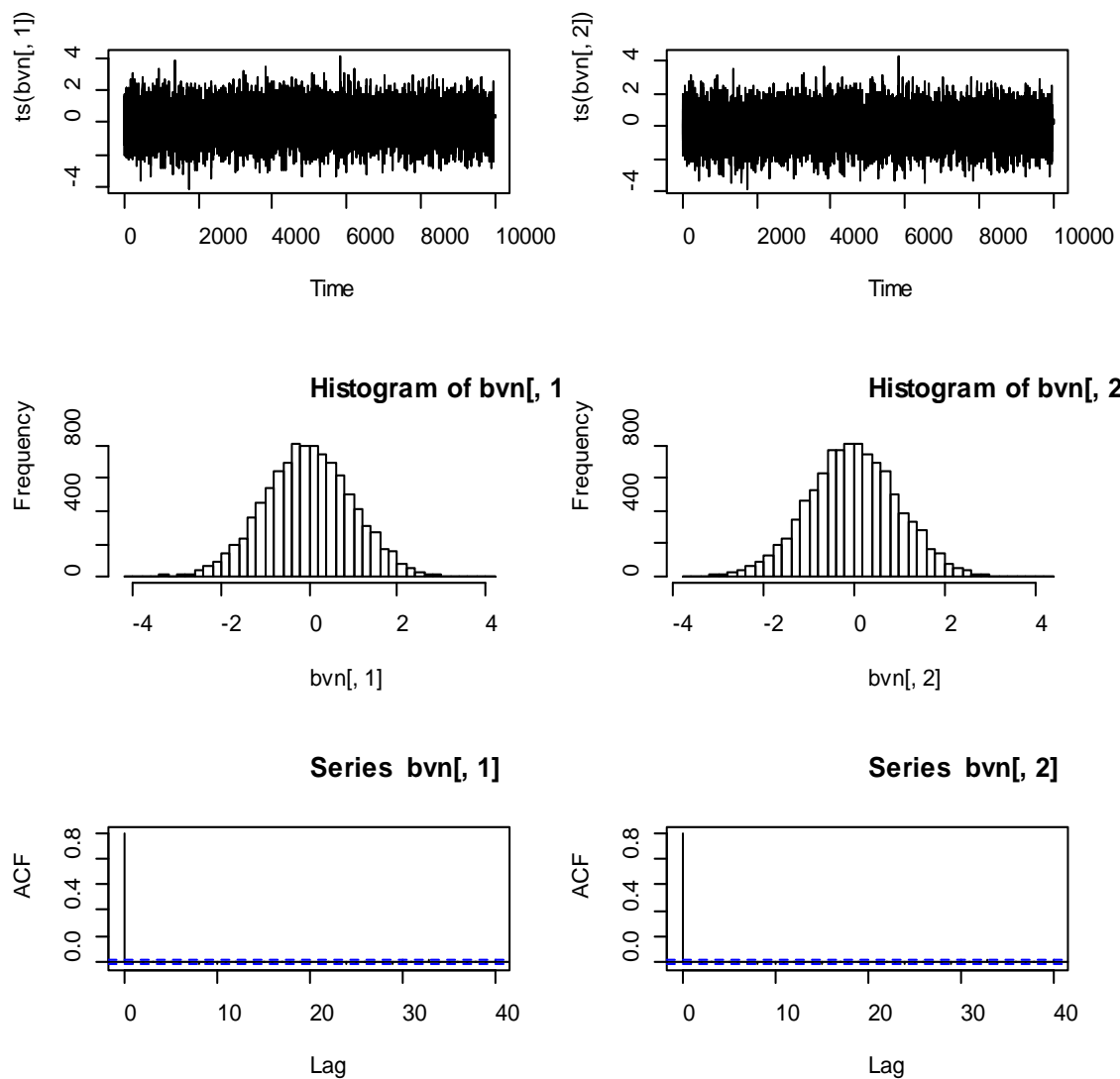
**Σχήμα 4.1:** Γραφήματα των παραγόμενων προσομοιωμένων τιμών, ιστογράμματα των περιθώριων κατανομών και συναρτήσεις αυτοσυσχέτισης με έναν δειγματολήπτη Gibbs από μία διμεταβλητή Κανονική κατανομή.

Φυσικά, για την προσομοίωση αυτή δε χρειαζόταν η εφαρμογή του δειγματολήπτη Gibbs αλλά θα μπορούσαμε να πραγματοποιήσουμε μια δειγματοληψία με πιο απλό τρόπο από ένα ανεξάρτητο και ισόνομο τυχαίο δείγμα, απλά προσομοιώνοντας από την περιθώρια σ.π.π. για το  $X$  και μετά από τη δεσμευμένη σ.π.π. για  $Y|X$ . Αυτός ο πιο απλός τρόπος αναπαρίσταται με τον παρακάτω κώδικα της R.



```
> rbvn<-function (n, rho)
+ {
+ x <- rnorm(n, 0, 1)
+ y <- rnorm(n, rho * x, sqrt(1 - rho^2))
+ cbind(x, y)
+ }
> bvn<-rbvn(10000,0.98)
> summary(bvn)
      x          y
Min. :-4.05651  Min. :-3.79949
1st Qu.: -0.69016 1st Qu.: -0.68003
Median :-0.02715 Median :-0.02662
Mean  :-0.02273  Mean  :-0.02361
3rd Qu.: 0.64257 3rd Qu.: 0.63708
Max.   : 4.07622  Max.   : 4.29324
> par(mfrow=c(3,2))
> plot(ts(bvn[,1]))
> plot(ts(bvn[,2]))
> hist(bvn[,1],40)
> hist(bvn[,2],40)
> acf(bvn[,1])
> acf(bvn[,2])
```

Αυτή η δειγματοληψία παράγει ένα ανεξάρτητο και ισόνομο δείγμα, όπως φαίνεται και από το Σχήμα 4.2 σε αντίθεση με τη δειγματοληψία που επιτυγχάνεται με τον αλγόριθμο Gibbs στην οποία υπάρχει αυτοσυσχέτιση. Βέβαια, μία τέτοια συμπεριφορά ήταν αναμενόμενη, καθώς ο δειγματολήπτης Gibbs είναι μία ειδική περίπτωση του αλγόριθμου Metropolis-Hastings και άρα παράγει συσχετισμένες τιμές.



**Σχήμα 4.2:** Γραφήματα από μία δειγματοληψία που παράγει ανεξάρτητες και ισόνομες τιμές από μία διμεταβλητή Κανονική κατανομή.

**Παράδειγμα 4.2:** (Ρίχνοντας ένα τυχαίο νόμισμα). Ας υποθέσουμε ότι ένα κουτί περιέχει ένα μεγάλο αριθμό ίδιων νομισμάτων. Κάθε νόμισμα όταν ριχτεί έχει μια διαφορετική πιθανότητα να δώσει «κεφαλή» και οι εν λόγω πιθανότητες ακολουθούν μια ομοιόμορφη κατανομή στο διάστημα  $(0,1)$ . Έστω ότι ρίχνω 10 νομίσματα, με 3 φορές να έχω «κεφαλή». Αν ρίξω ακόμη 12 νομίσματα ποια θα είναι η πιθανότητα να πάρω 4 φορές κεφαλή;

Σε αυτό το πρόβλημα, υπάρχουν 2 άγνωστες μεταβλητές: η πιθανότητα  $p$  που το νόμισμα θα δώσει κεφαλή και ο αριθμός των κεφαλών  $y$  που θα έχουμε στις 12 επιπλέον ρίψεις. Άρα, σύμφωνα με τα παραπάνω, η από κοινού κατανομή των  $(p, y)$  θα είναι:



$$f(p, y) = \binom{10}{3} p^3 (1-p)^7 \times \binom{12}{y} p^y (1-p)^{12-y}, \quad 0 < p < 1, y = 0, 1, 2, \dots, 12$$

Για να εφαρμόσουμε το Δειγματολήπτη Gibbs, πρέπει πρώτα από όλα να αναγνωρίσουμε τις δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας, που στην περίπτωση μας είναι η σ.π.π. της  $p$  δεδομένου ότι ισχύει η  $y$ ,  $f(p|y)$ , και η συνάρτηση πυκνότητας πιθανότητας του  $y$  δεδομένου ότι ισχύει η  $p$ ,  $f(y|p)$ . Άρα, για να βρούμε την  $f(p|y)$  συγκεντρώνουμε όλους τους όρους που περιέχουν τη μεταβλητή  $p$  και θεωρούμε το  $y$  σαν σταθερό όρο. Έτσι, θα έχουμε ότι:

$$f(p|y) \propto p^{y+3} (1-p)^{19-y}, \quad 0 < p < 1$$

Αν ξαναγράψουμε την συνάρτηση πυκνότητας πιθανότητας με τη παρακάτω μορφή:

$$f(p|y) \propto p^{y+4-1} (1-p)^{20-y-1}, \quad 0 < p < 1,$$

μπορούμε εύκολα να αναγνωρίσουμε ότι αυτή η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας είναι η Βήτα με παραμέτρους  $a = y + 4$  και  $b = 20 - y$ .

Για να πάρουμε τη δεύτερη δεσμευμένη σ.π.π., συγκεντρώνουμε όλους τους όρους που περιέχουν  $y$  και θεωρούμε την πιθανότητα  $p$  σαν μια σταθερά. Έτσι, με τον ίδιο τρόπο θα έχουμε ότι:

$$f(y|p) \propto \binom{12}{y} p^y (1-p)^{12-y}, \quad y = 0, 1, 2, \dots, 12.$$

Και σε αυτήν την περίπτωση μπορούμε εύκολα να αναγνωρίσουμε την παραπάνω συνάρτηση πυκνότητας πιθανότητας σαν Διωνυμική με μέγεθος δείγματος 12 και πιθανότητα επιτυχίας  $p$ .

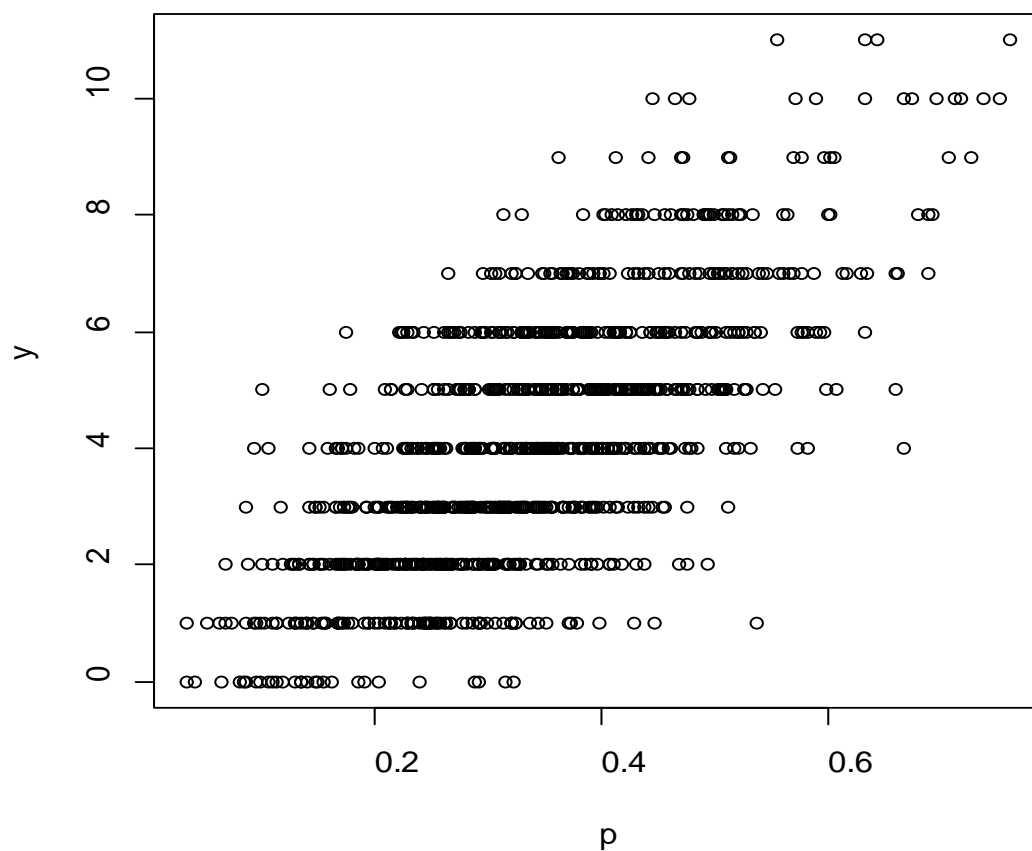
Για την εφαρμογή όλων των παραπάνω στο στατιστικό πακέτο R, αρχικά κατασκευάζουμε μια συνάρτηση που θα έχει δύο εισόδους:  $p$  είναι η αρχική τιμή της πιθανότητας  $p$  και  $m$  είναι ο αριθμός των επαναλήψεων που κάνει ο δειγματολήπτης Gibbs. Σε αυτή τη συνάρτηση, ο πίνακας  $S$  αποθηκεύει το προσομοιωμένο δείγμα και για μια ευκολία αναγνώρισης βάζουμε ετικέτες " $p$ " και " $y$ " στις δύο στήλες του πίνακα. Στη συνέχεια έχουμε ένα βρόγχο μέσα στον οποίο προσομοιώνουμε επιτυχώς δείγμα  $y$  από την Διωνυμική κατανομή (χρησιμοποιώντας τη συνάρτηση *rbinom*) και δείγμα  $p$  από την Βήτα κατανομή (*rbeta*).

Τέλος, τρέχουμε τη παραπάνω συνάρτηση για τις προεπιλεγμένες τιμές,  $p = 0.5$  και  $m = 1000$ , απλά πληκτρολογώντας *random.coins.gibbs* χωρίς παραμέτρους. Η μεταβλητή *sim.values* περιέχει τον πίνακα των προσομοιωμένων τιμών από την από κοινού κατανομή  $(p, y)$  και με την εντολή *plot(sim.values)* παίρνουμε το παρακάτω γράφημα σκέδασης αυτής της κατανομής (Σχήμα 4.3).



```
>random.coins.gibbs=function(p=0.5,m=1000){
+ S=matrix(0,m,2)
+ dimnames(S)[[2]]=c("p","y")
+ for(j in 1:m){
+ y=rbinom(1,size=12,prob=p)
+ p=rbeta(1,y+4,20-y)
+ S[j,]=c(p,y)
+ }
+ return(S)
+ }
> sim.values=random.coins.gibbs()
> plot(sim.values)
```

Σε αυτό το παράδειγμα, όπως βλέπουμε και από την απεικόνιση των προσομοιωμένων τιμών (Σχήμα 4.3), η περίοδος τρεξίματος (burn in) φαίνεται να είναι μικρή και να υπάρχει μια καλή μίξη της αλυσίδας.



**Σχήμα 4.3:** Γράφημα του προσομοιωμένου δείγματος του  $(p, y)$  από τον αλγόριθμο του δειγματολήπτη Gibbs για ένα παράδειγμα ρίψης τυχαίου νομίσματος.

Υποθέτοντας ότι ο πίνακας *sim.values* είναι ένα προσεγγιστικό τυχαίο δείγμα από την από κοινού κατανομή  $(p, y)$ , τα δείγματα από τις επιμέρους στήλες αναπαριστούν δείγματα από τις περιθώριες κατανομές του  $p$  και του  $y$ . Το ζητούμενο αυτής της άσκησης είναι η πιθανότητα  $P(y = 4)$ . Μέσω της συνάρτησης *table* στην R τοποθετώ τις τιμές της  $y$  σε έναν πίνακα και έτσι θα έχω:

```
>table(sim.values[, "y"])
```

```
 0  1  2  3  4  5  6  7  8  9 10 11
28 96 142 161 162 151 119 69 41 14 13  4
```



Στο δείγμα, λοιπόν, αυτών των 1000 τιμών του  $y$  παρατηρήσαμε 162 «4», άρα  $P(y = 4) \approx 162/1000 = 0.162$ . Με παρόμοιο τρόπο, μπορούν να βρεθούν και άλλα χαρακτηριστικά της περιθώριας συνάρτησης πυκνότητας πιθανότητας του  $y$ , όπως η τυπική απόκλιση και ο μέσος, μέσω της συνάρτησης *summary* στην R των προσομοιωμένων τιμών  $y$ .

**Παράδειγμα 4.3:** Ας υποθέσουμε ότι έχουμε μία διμεταβλητή συνάρτηση πυκνότητας πιθανότητας  $f(x, y)$  στο χώρο  $\mathbb{R}_+^2$  για την οποία για κάποιες παραμέτρους  $\lambda > 0$  και  $\mu > 0$  ισχύει:

$$f(x, y) \propto \exp(-[\lambda x + \lambda y + \mu xy]), \quad x \geq 0, y \geq 0$$

Για να εφαρμόσουμε το δειγματολήπτη Gibbs θα πρέπει πρώτα να βρούμε τις δεσμευμένες σ.π.π. του  $y$  και του  $x$ . Άρα, θα έχουμε την παρακάτω δεσμευμένη σ.π.π.:

$$f(y|x) \propto \exp(-[\lambda x + \lambda y + \mu xy]) \propto \exp(-(\lambda + \mu x)y),$$

που αντιστοιχεί σε μια  $\exp(\lambda + \mu x)$  συνάρτηση πυκνότητας πιθανότητας.

Με τον ίδιο τρόπο θα έχουμε ότι:

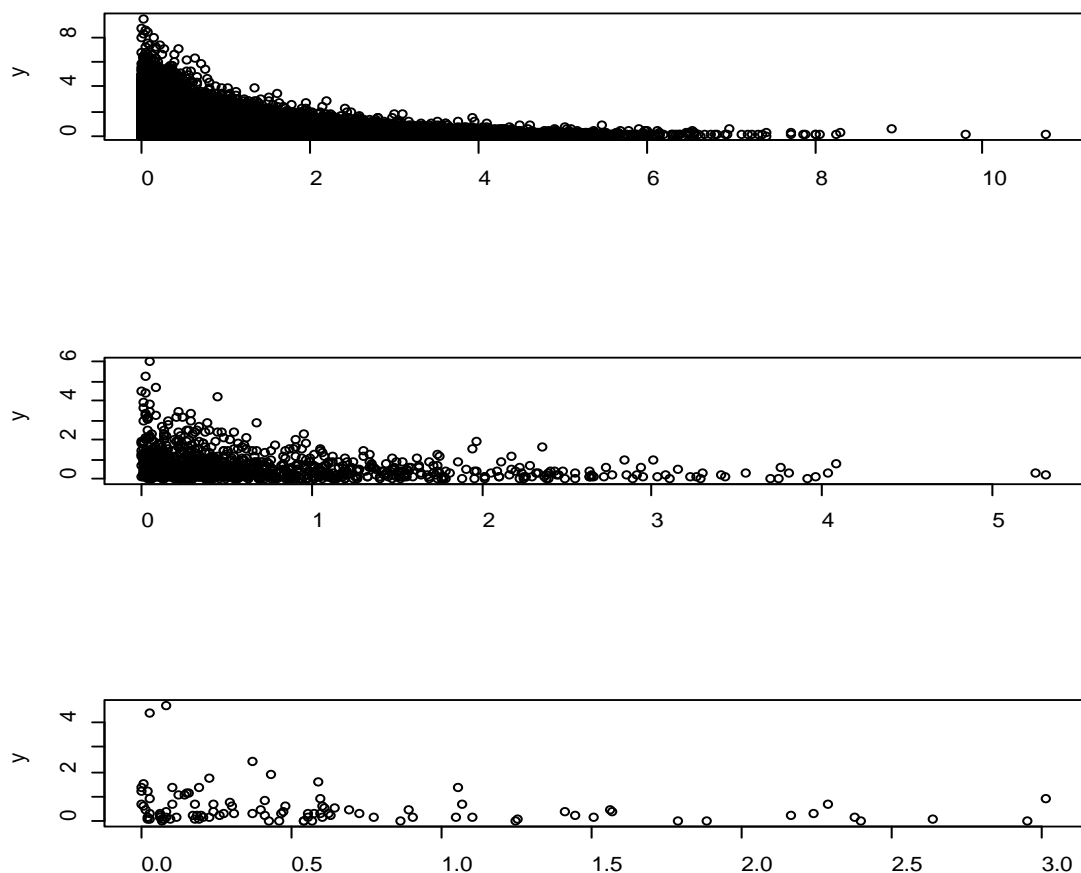
$$f(x|y) \propto \exp(-(\lambda + \mu y)x)$$

Στη συνέχεια θέλουμε να παράγουμε ένα δείγμα από τη συνάρτηση  $f$ , χρησιμοποιώντας αυτές τις δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας. Θα κατασκευάσω στην R μια συνάρτηση του δειγματολήπτη Gibbs έτσι ώστε να μπορούμε να προσομοιώσουμε από την από κοινού εκθετική κατανομή και έτσι θα ισχύει ο παρακάτω κώδικας στην R:



```
> gibbs.sampler.exp<-function(n,lambda,mu,x0=0){
+ #function to implement Gibbs sampler for bivariate density
+ x=rep(0,n+1)
+ y=rep(0,n)
+ x[1]=x0
+ for (i in 1:n){
+ y[i]=rexp(1,lambda+mu*x[i])
+ x[i+1]=rexp(1,lambda+mu*y[i])
+ }
+ sim.values<-cbind(x[-(n+1)],y)
+ return(sim.values)
+ }
> g.s=gibbs.sampler.exp(100000,1,1)
> par(mfrow=c(3,1))
> plot(g.s,main="Gibbs sampling from an exponential dens")
> plot(g.s[1:1000,])
> plot(g.s[1:100,])
```

Gibbs sampling from an exponential dens



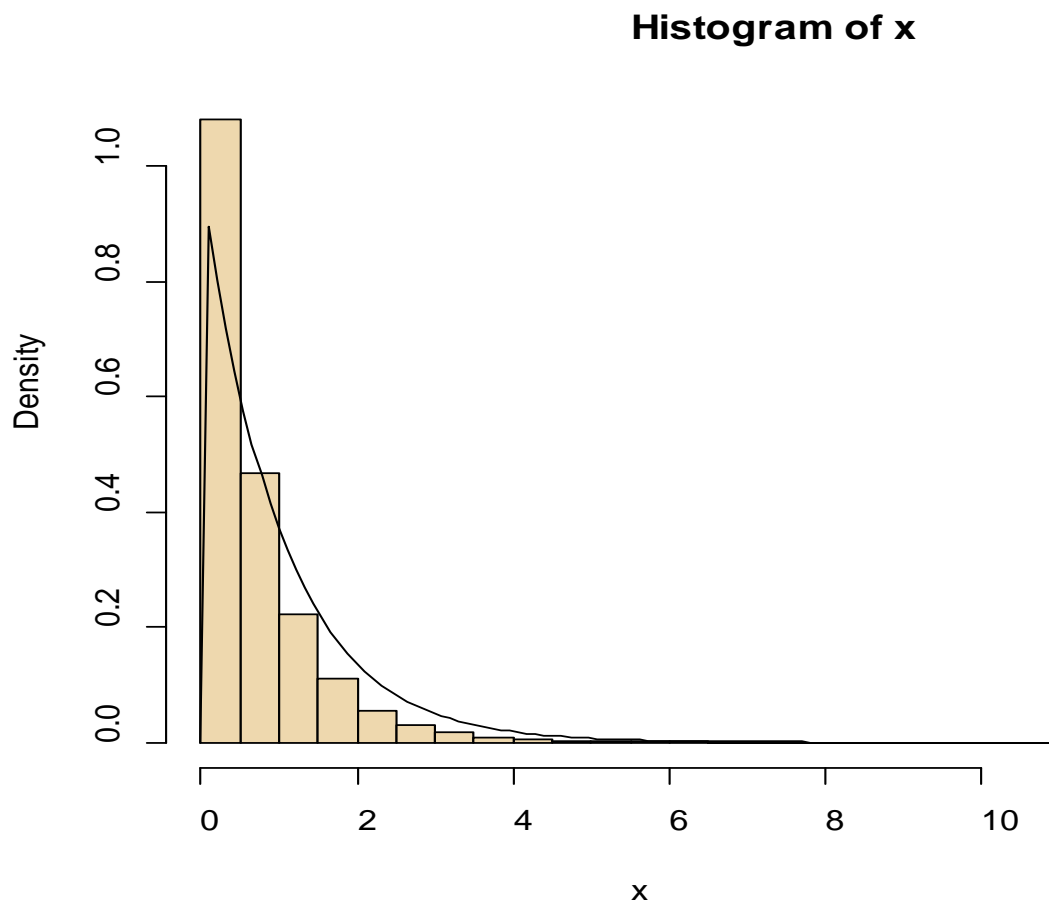
**Σχήμα 4.4:** Προσομοίωση δείγματος με το δειγματολήπτη Gibbs από μια εκθετική κατανομή  $f(x, y) \propto \exp(-[\lambda x + \lambda y + \mu xy])$  για  $\lambda=1$  και  $\mu=1$ .

Όπως διαπιστώνουμε από το παραπάνω σχήμα η μίξη είναι πολύ καλή όχι μόνο στις περισσότερες αλλά και στις λιγότερες (100) επαναλήψεις του αλγορίθμου αφού οι προσομοιωμένες τιμές κατανέμονται με τον ίδιο τρόπο και στα 3 γραφήματα. Δηλαδή, η Μαρκοβιανή αλυσίδα εξερευνά το ίδιο εύρος τιμών σε διαφορετικές περιόδους.

Τώρα μπορούμε να εκτιμήσουμε οποιαδήποτε ποσότητα θέλουμε της περιθώριας κατανομής του  $x$  αλλά και του  $y$ . Εδώ θα πρέπει να αναφέρουμε ότι η  $g.s[,1]$  αντιστοιχεί στην ακολουθία των προσομοιώσεων της μεταβλητής  $x$ , δηλαδή μπορούμε να θεωρήσουμε αυτή την ακολουθία σαν ένα δείγμα από την περιθώρια κατανομή της τ.μ.  $x$ . Έτσι, με μία μεγάλη προσομοίωση των 100.000 επαναλήψεων μπορούμε να βρούμε και τις διάφορες εκτιμήτριες για το μέσο και τη τυπική απόκλιση των διαφόρων περιθώριων κατανομών, αλλά και διάφορες πιθανότητες που μας ενδιαφέρουν όπως για παράδειγμα την πιθανότητα και το  $x$  και το  $y$  να πάρουν τιμές μικρότερες του 1. Δηλαδή, σύμφωνα με τα παραπάνω θα έχουμε τον ακόλουθο κώδικα στην R:

```
> x<-g.s[,1]
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.0000 0.1752  0.4399  0.6788  0.9190 10.0400
> mean(x)
[1] 0.6823054
> sd(x)
[1] 0.7400368
> hist(x,freq=FALSE,col="wheat2")
> curve(dexp(x),add=TRUE)
> sum(g.s[,1]<1 & g.s[,2]<1)/100000
[1] 0.57104
```





**Σχήμα 4.5:** Ιστόγραμμα της περιθώριας εκθετικής κατανομής του παραδείγματος 4.3.

### 3.Ο Δειγματολήπτης Gibbs Πολλαπλών Σταδίων

Υπάρχει μια φυσική προέκταση από τον δειγματολήπτη Gibbs 2 καταστάσεων σε εκείνον με τα πολλαπλά στάδια. Ας υποθέσουμε, λοιπόν, ότι για κάποιο  $p > 1$  η τυχαία μεταβλητή  $\mathbf{X} \in \mathcal{X}$  μπορεί να γραφτεί σαν  $\mathbf{X} = (X_1, \dots, X_p)$  όπου τα  $X_i$  είναι μονοδιάστατες ή πολυδιάστατες συνιστώσες. Ακόμη, ας υποθέσουμε ότι μπορούμε να προσομοιώσουμε από τις αντίστοιχες δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας  $f_1, \dots, f_p$ , δηλαδή μπορούμε να προσομοιώσουμε

$$X_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p \sim f_i(x_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p) \text{ με } i = 1, 2, \dots, p.$$

Έτσι, ο αλγόριθμος του δειγματολήπτη Gibbs πολλαπλών σταδίων παίρνει την παρακάτω μορφή:

**Αλγόριθμος 6: Ο δειγματολήπτης Gibbs πολλαπλών σταδίων**

Στην επανάληψη  $t = 1, 2, \dots$ , δοσμένων των  $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_p^t)$ , παρήγαγε

1.  $X_1^{t+1} \sim f_1(x_1 | x_2^t, \dots, x_p^t);$
2.  $X_2^{t+1} \sim f_2(x_2 | x_1^{t+1}, x_3^t, \dots, x_p^t);$
- $\vdots$
- $p$ .  $X_p^{t+1} \sim f_p(x_p | x_1^{t+1}, \dots, x_{p-1}^{t+1})$

Οι συναρτήσεις πυκνότητας  $f_1, \dots, f_p$  ονομάζονται *πλήρεις δεσμευμένες* συναρτήσεις πυκνότητας πιθανότητας και το ιδιαίτερο χαρακτηριστικό αυτής της δειγματοληψίας είναι ότι είναι οι μόνες σ.π.π. που χρησιμοποιούνται για αυτή τη προσομοίωση. Συνεπώς, ακόμα και σε ένα υψηλών διαστάσεων πρόβλημα, όλες οι προσομοιώσεις μπορεί να έχουν μόνο μια μεταβλητή, που συνήθως είναι ένα πλεονέκτημα.

Όπως ήδη αναφέραμε οι δειγματολήπτες Gibbs χρησιμοποιούνται για να προσεγγίσουν κατανομές που διαφορετικά θα ήταν δύσκολο έως ακατόρθωτο να βρεθούν με διαφορετικό τρόπο. Γι αυτό και χρησιμοποιούνται τόσο ευρεία στην Μπεϋζιανή εκτίμηση. Αυτοί οι δειγματολήπτες προσφέρουν τη μεγαλύτερη πρακτική χρησιμότητα σε Μαρκοβιανές αλυσίδες με μεγάλους χώρους καταστάσεων και ειδικά σε συνεχείς χώρους καταστάσεων.

**Παράδειγμα 4.4:** Αντίστοιχα με το παράδειγμα 4.1 στο οποίο εφαρμόσαμε ένα δειγματολήπτη Gibbs 2 σταδίων για προσομοίωση από μία διμεταβλητή Κανονική κατανομή, σε αυτό το παράδειγμα θεωρούμε το πολυμεταβλητό Κανονικό μοντέλο:

$$(X_1, X_2, \dots, X_p) \sim \mathcal{N}_p(0, (1 - \rho)I + \rho J),$$

όπου  $I$  είναι ο μοναδιαίος πίνακας διάστασης  $p$ ,  $J$  ένας πίνακας διάστασης  $p$  με όλα του τα στοιχεία μονάδες και  $\text{corr}(X_i, X_j) = \rho \ \forall i, j$  η αυτοσυσχέτιση των μεταβλητών μας.

Έτσι, οι δεσμευμένες συναρτήσεις κατανομών είναι:

$$X_i | \mathbf{x}_{-i} \sim \mathcal{N}\left(\frac{(\rho - 1)\rho}{1 + (\rho - 2)\rho} \bar{x}_{-i}, \frac{1 + (\rho - 2)\rho - (\rho - 1)\rho^2}{1 + (\rho - 2)\rho}\right)$$

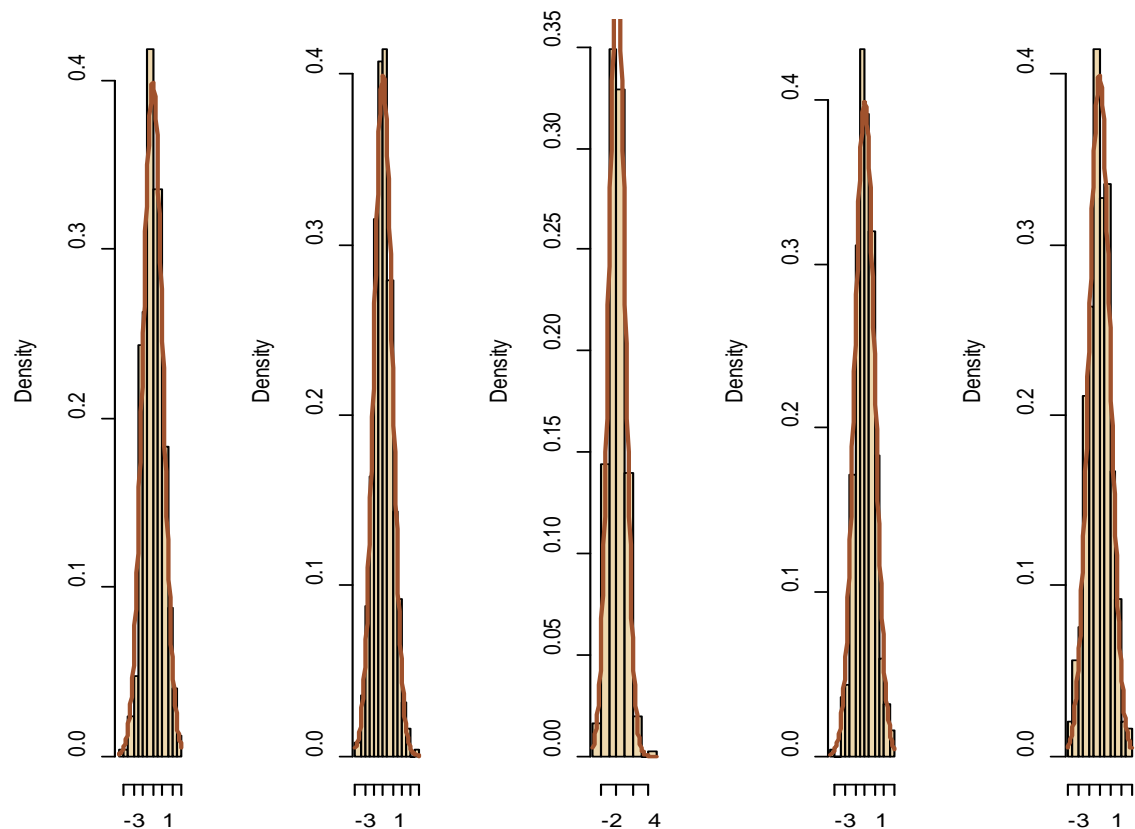
Όπου  $\mathbf{x}_{-i} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$  και το  $\bar{x}_{-i}$  είναι ο μέσος όρος αυτού του διανύσματος. Ο δειγματολήπτης Gibbs θα βασίζεται πάνω στις πλήρεις δεσμευμένες κατανομές και έτσι αν τρέξουμε τον αλγόριθμο στην R για  $p = 5$  και  $\rho = 0.25$  θα έχουμε τα παρακάτω αποτελέσματα:



```
> T=500
> p=5
> r=0.25
> X=cur=rnorm(p)
> for (t in 1 :T)
+ {
+ for (j in 1 :p)
+ {
+ m=sum(cur[-j])/(p-1)
+ cur[j]=rnorm(1,(p-1)*r*m/(1+(p-2)*r),
+ sqrt((1+(p-2)*r-(p-1)*r^2)/(1+(p-2)*r)))
+ }
+ X=cbind(X,cur)
+ }
> par(mfrow=c(1,5))
> for (i in 1:p){
+ hist(X[i,],prob=TRUE,col="wheat2",xlab="",main="")
+ curve(dnorm(x),add=TRUE,col="sienna",lwd=2)}
```

Όπως βλέπουμε και από το Σχήμα 4.6, οι πέντε περιθώριες κατανομές του πολυμεταβλητού Κανονικού μοντέλου του Παραδείγματος 4.5 ακολουθούν Κανονική κατανομή. Γενικά, ισχύει ότι όλες οι περιθώριες κατανομές ενός πολυμεταβλητού Κανονικού μοντέλου θα ακολουθούν Κανονική κατανομή.





**Σχήμα 4.6:** Ιστογράμματα του πολυμεταβλητού κανονικού μοντέλου του Παραδείγματος 4.5. Παρατηρούμε πως και οι πέντε περιθώριες κατανομές του κανονικού μοντέλου ακολουθούν Κανονικές κατανομές  $\mathcal{N}(0,1)$ .

## 4. Παράγοντες Που Επηρεάζουν Την Εφαρμογή Του Δειγματολήπτη Gibbs

Σε αυτό το κεφάλαιο θα αναλύσουμε μερικά προβλήματα που συχνά προκύπτουν στην εφαρμογή του δειγματολήπτη Gibbs.

### 1. Εκ Νέου Ρύθμιση Παραμέτρων

Πολλοί παράγοντες συμβάλλουν στις ιδιότητες σύγκλισης ενός δειγματολήπτη Gibbs. Για παράδειγμα, η σύγκλιση μπορεί να επηρεαστεί σε μεγάλο βαθμό από την επιλογή των συντεταγμένων (ή με άλλα λόγια την *παραμετροποίηση*). Αν η ο πίνακας συνδιακύμανσης της ζητούμενης συνάρτησης έχει ένα μεγάλο εύρος ιδιοτιμών, τότε ο δειγματολήπτης Gibbs μπορεί να αποδειχτεί πολύ αργός για να εξερευνήσει ολόκληρο το εύρος της περιοχής της ζητούμενης συνάρτησης.

**Παράδειγμα 4.5:** Ας θυμηθούμε το παράδειγμα 4.1 όπου είχαμε δει για πρώτη φορά τον δειγματολήπτη Gibbs για το διμεταβλητό Κανονικό μοντέλο

$$(X, Y) \sim \mathcal{N}_2 \left( 0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Για αυτή τη διμεταβλητή Κανονική κατανομή θα εκτελέσουμε τον κώδικα του παραδείγματος 4.1 για αυτοσυσχετίσεις  $\rho = 0.3, 0.6, 0.9$ . Η μεγαλύτερη αυτοσυσχέτιση θα μας αναγκάσει να χρησιμοποιήσουμε περισσότερες επαναλήψεις αφού ο αλγόριθμος δε θα μπορέσει διαφορετικά να εξερευνήσει ολόκληρη τη περιοχή της ζητούμενης συνάρτησης. Έτσι, για τις τρεις αυτές διαφορετικές τιμές αυτοσυσχέτισης θα ισχύουν τα γραφήματα των συναρτήσεων αυτοσυσχέτισης στο Σχήμα 4.7.

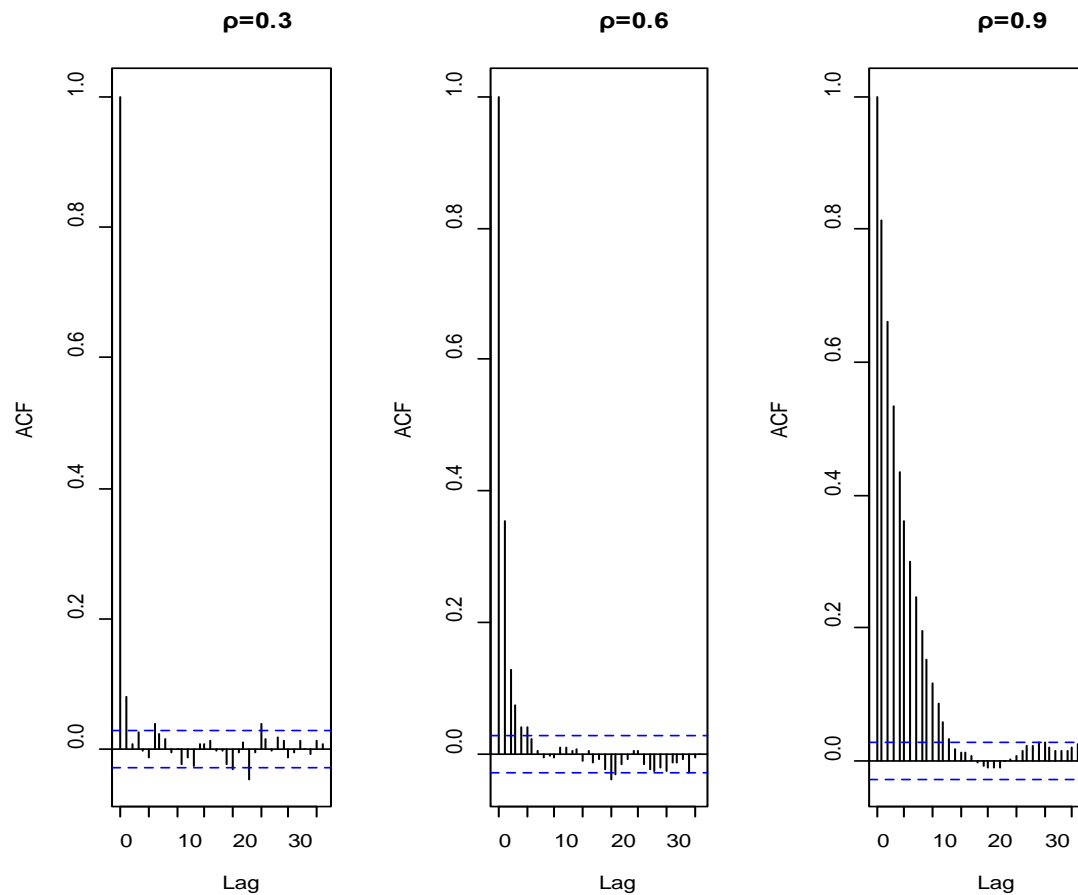
Τώρα, αν σε αυτό το παράδειγμα επιχειρήσουμε μια αλλαγή των μεταβλητών μας σε  $X + Y, X - Y$  τότε οι καινούριες αυτές μεταβλητές ακολουθούν πάλι Κανονική κατανομή με διασπορές  $2(1 + \rho)$  και  $2(1 - \rho)$  αντίστοιχα. Η διαφορά τους με τις προηγούμενες μεταβλητές μας είναι ότι είναι ανεξάρτητες αφού

$$\text{Cov}(X + Y, X - Y) = \mathbb{E}[(X + Y)(X - Y)] = \mathbb{E}(X^2 - Y^2) = 1 - 1 = 0.$$

Άρα, θα ισχύει:

$$X|Y \sim \mathcal{N} \left( \rho \left( \frac{\sigma_x}{\sigma_y} \right)^2 y, \sigma_x^2 (1 - \rho^2) \right)$$

και με παρόμοιο τρόπο βρίσκουμε και τη δεσμευμένη συνάρτηση κατανομής του  $Y|X$ . Έτσι, τώρα μπορούμε να εφαρμόσουμε τον αλγόριθμο της δειγματοληψίας Gibbs σε αυτό το μοντέλο με αυτές τις νέες μεταβλητές.

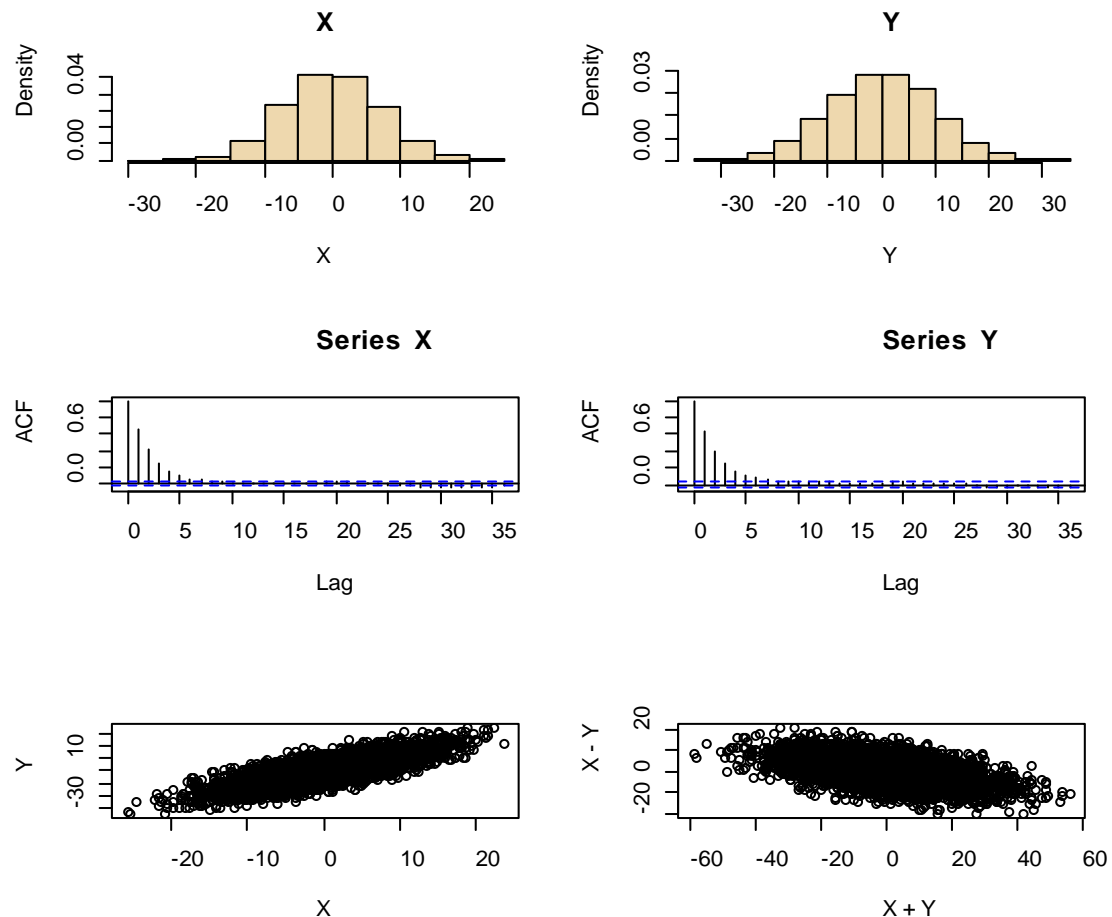


**Σχήμα 4.7:** Συσχετίσεις μιας περιθώριας κατανομής ενός διμεταβλητού κανονικού μοντέλου από τη δειγματοληψία Gibbs.

```
> T=5000
> rho=0.8
> sx=50
> sy=100
> X=rnorm(1)
> Y=rnorm(1)
> for (t in 1:T){
+ Yn=rnorm(1,r*sqrt(sy/sx)*X[t],sqrt(sy*(1-rho^2)))
+ Xn=rnorm(1,r*sqrt(sx/sy)*Yn,sqrt(sx*(1-rho^2)))
+ X=c(X,Xn)
+ Y=c(Y,Yn)
+ }
> par(mfrow=c(3,2))
> hist(X,prob=TRUE,main="X",col="wheat2")
> hist(Y,prob=TRUE,main="Y",col="wheat2")
> acf(X)
> acf(Y)
> plot(X,Y)
> plot(X+Y,X-Y)
```



Διαπιστώνουμε από τα παραπάνω ότι αυτή η αλλαγή παραμέτρων οδήγησε σε καλύτερη σύγκλιση του αλγόριθμου αφού βλέπουμε από το Σχήμα 4.8 πως οι αυτοσυσχετίσεις είναι μικρές και ότι με τη καινούρια ρύθμιση των παραμέτρων μας εξερευνούνται οι περιοχές της ζητούμενης συνάρτησης με τη μεγαλύτερη πιθανότητα.



**Σχήμα 4.8:** Ιστογράμματα και αυτοσυσχετίσεις των περιθώριων κατανομών του διμεταβλητού κανονικού μοντέλου με τις νέες μεταβλητές  $X+Y, X-Y$ .

■

Η σύγκλιση, λοιπόν, όχι μόνο του δειγματολήπτη Gibbs αλλά και του αλγόριθμου Metropolis-Hastings μπορούν να υποστούν κακή σύγκλιση από μία παραμετροποίηση. Για αυτό το λόγο, στην MCMC βιβλιογραφία αναφέρονται σαν λύση κάποιες αλλαγές στην παραμετροποίηση του μοντέλου έτσι ώστε να επιταχυνθεί η σύγκλιση σε μια δειγματοληψία Gibbs. Ωστόσο, οι περισσότερες προσπάθειες έγιναν για τη βελτίωση συγκεκριμένων μοντέλων καταλήγοντας έτσι στην απουσία μιας γενικής μεθοδολογίας που θα έχει σαν σκοπό την επιλογή της κατάλληλης παραμετροποίησης. Παρόλα αυτά, μια γενικότερη συμβουλή θα ήταν να κάνουμε τις συντεταγμένες όσο πιο πολύ

ανεξάρτητες μπορούμε και να χρησιμοποιήσουμε πολλές παραμετροποιήσεις ταυτόχρονα έτσι ώστε οι δεσμευμένες συναρτήσεις κατανομών να αναμιγνύονται.

## 2.Rao-Blackwellization

Στο γνωστό θεώρημα Rao-Blackwell, η δεσμευμένη εκτίμηση σε ένα υποσύνολο προσομοιωμένων μεταβλητών μπορεί να οδηγήσει σε μια στοιχειώδη βελτίωση στις τυπικές εκτιμήτριες, από την άποψη της διασποράς, μέσω μιας απλής «ανακύκλωσης» των απορριπτέων τιμών. Όμως, επειδή ο δειγματολήπτης Gibbs αποδέχεται κάθε προσομοιωμένη τιμή αυτή η «ανακύκλωση» δεν μπορεί να εφαρμοστεί σε αυτήν την περίπτωση. Έτσι, οι Gelfand και Smith (1990) πρότειναν το παραμετρικό Rao-Blackwellization για να το διαφοροποιήσουν από το παραπάνω γνωστό θεώρημα.

Δηλαδή, για  $(X, Y) \sim f(x, y)$  το παραμετρικό Rao-Blackwellization βασίζεται στη ταυτότητα περιθωριοποίησης

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]].$$

Ορίζοντας  $\delta(Y) = \mathbb{E}[X|Y]$ , έχουμε ότι  $\mathbb{E}[\delta(Y)] = \mathbb{E}[X]$  και  $\text{var}[\delta(Y)] \leq \text{var}(X)$  και έτσι αποδεικνύουμε ότι  $\delta(Y)$  είναι ο καλύτερος εκτιμητής (με την προϋπόθεση πάντα ότι μπορεί να υπολογιστεί).

**Παράδειγμα 4.6:** Στην περίπτωση όπου ζητούμενο μοντέλο είναι το απλό διμεταβλητό κανονικό μοντέλο, στο οποίο έχουμε ήδη αναφερθεί προηγουμένως (Παράδειγμα 4.1), έχουμε ότι οι πλήρεις δεσμευμένες κατανομές είναι οι:

$$X_{t+1}|y_t \sim \mathcal{N}(\rho y_t, 1 - \rho^2)$$

$$Y_{t+1}|x_{t+1} \sim \mathcal{N}(\rho x_{t+1}, 1 - \rho^2).$$

Άρα, από τα παραπάνω συνεπάγεται ότι  $\mathbb{E}[X|Y] = \rho Y$ . Αφού οι  $X$  και  $Y$  έχουν τη ίδια περιθώρια κατανομή, είναι φανερό πως η διασπορά του παραπάνω παραμετρικού Rao-Blackwellization μειώνεται από έναν παράγοντα  $\rho^2$ .

■

Δυστυχώς η μείωση διασποράς που επιτυγχάνεται από τη χρήση του  $\delta_Y$  δεν ισχύει γενικά εξαιτίας του συσχετισμού ανάμεσα στα MCMC δείγματα. Όμως οι Liu et al. (1994) απέδειξαν συγκεκριμένα ότι αυτή η βελτίωση ισχύει για τους δειγματολήπτες Gibbs 2 καταστάσεων.

### 3. Metropolis Εντός Του Gibbs Και Υβριδικές Στρατηγικές

Σε ένα σημείο που αξίζει να δώσουμε έμφαση σχετικά με την εφαρμογή του δειγματολήπτη Gibbs είναι ότι μπορεί εύκολα να επεκταθεί και σε ρυθμίσεις όπου κάποιες από τις δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας δεν μπορούν να προσομοιωθούν από τις τυπικές γεννήτριες τυχαίων τιμών. Δηλαδή, εάν μέσα σε ένα σύνολο πλήρων δεσμευμένων συναρτήσεων πυκνότητας πιθανότητας  $f_1, \dots, f_p$  κάποια  $f_i$  είναι ασυνήθιστη αυτό δεν θέτει σε κίνδυνο τον δειγματολήπτη Gibbs γιατί υιοθετείται η ακόλουθη *Metropolis-εντός του-Gibbs* στρατηγική. Όντως, αντί να προσομοιώσουμε:

$$x_i^{t+1} \sim f_i(x_i | x_1^{t+1}, \dots, x_{i-1}^{t+1}, x_{i+1}^t, \dots, x_p^t),$$

μπορούμε να τρέξουμε ένα μόνο βήμα μιας μεθόδου MCMC που να έχει για στάσιμη κατανομή της την:

$$f_i(x_i | x_1^{t+1}, \dots, x_{i-1}^{t+1}, x_{i+1}^t, \dots, x_p^t).$$

Για παράδειγμα, μία απλή λύση θα μπορούσε να είναι να χρησιμοποιήσουμε έναν Τυχαίο Περίπατο Metropolis με κέντρο το  $x_i^t$ . Παρόλο που στην αρχή αυτή η λύση φαίνεται σαν μια αργή προσέγγιση, αφού οι πλήρεις δεσμευμένες σ.π.π. δεν προσομοιώνονται ακριβώς, η εγκυρότητα του αλγόριθμου που προκύπτει είναι ακριβώς η ίδια με τον κανονικό δειγματολήπτη Gibbs, αφού η από κοινού κατανομή  $f$  παραμένει η στάσιμη κατανομή της αντίστοιχης Μαρκοβιανής αλυσίδας.


**Παρατήρηση 4.1:** Θα μπορούσε κάποιος να αναρωτηθεί ποιο είναι τελικά το νόημα να χρησιμοποιήσουμε το δειγματολήπτη Gibbs αν οι προσομοιώσεις των διαφόρων συνιστωσών θα πρέπει τελικά να αντικατασταθούν με βήματα του Metropolis-Hastings. Δηλαδή, σε αυτή την περίπτωση η χρήση του αλγόριθμου Metropolis-Hastings με ζητούμενη τη συνάρτηση  $f$  θα φαινόταν πιο λογική. Ενώ δεν υπάρχει καμία απαγόρευση από το να χρησιμοποιήσουμε έναν από κοινού Metropolis-Hastings, είναι αρκετά συχνό το φαινόμενο η κατασκευή αυτού του Metropolis-Hastings σε μία μεγάλων διαστάσεων  $f$  να είναι δύσκολη ή ακόμα και αδύνατη. Το θεμελιώδες κέρδος που αποκτούμε από τη χρήση μιας δομής σαν του Gibbs είναι ότι απλοποιεί ένα περίπλοκο μοντέλο σε έναν μεγάλο αριθμό μικρότερων και πιο απλών δομών της  $f$ , όπου οι «τοπικοί» Metropolis-Hastings αλγόριθμοι μπορούν να σχεδιαστούν και να εφαρμοστούν με μικρότερο κόστος.

## 5.Εφαρμογές Στον Δειγματολήπτη Gibbs

### Εφαρμογή 1: Θερμοκρασία Σώματος

Σε αυτή την εφαρμογή χρησιμοποιήσαμε δεδομένα από μία έρευνα της εφημερίδας της Αμερικανικής Ένωσης του 1992, στην οποία συμμετείχαν 130 άτομα. Ήταν κοινή τότε η άποψη πως η πραγματική θερμοκρασία του σώματος είναι 98,6°F (37°C). Αυτή η βάση δεδομένων μας οδηγεί στην πλέον κοινή άποψη όλων πως η πραγματική θερμοκρασία του σώματος είναι 98,2°F (36,78°C).

Ακόμη, περιέχονται και πληροφορίες για τον καρδιακό ρυθμό (σε παλμούς ανά λεπτό) αναφορικά με τη θερμοκρασία του σώματος. Με τη μεταβλητή *temper* παρουσιάζονται οι αντίστοιχες θερμοκρασίες και με τη *h.r.* ο καρδιακός παλμός. Έτσι, βλέπουμε κάποια χαρακτηριστικά των δύο αυτών μεταβλητών:



```
> summary(temper)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 96.30  97.80  98.30  98.25  98.70 100.80
> sd(temper)
[1] 0.7331832
> fitdistr(temper,"normal")
      mean      sd
98.24923077 0.73035778
( 0.06405661) ( 0.04529487)
> summary(h.r)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 57.00  69.00  74.00  73.76  79.00  89.00
> sd(h.r)
[1] 7.062077
> fitdistr(h.r,"normal")
      mean      sd
73.7615385  7.0348625
( 0.6169983) ( 0.4362836)
> lm(h.r~temper)

Call:
lm(formula = h.r ~ temper)

Coefficients:
(Intercept)    temper
   -166.285      2.443
```



```
> summary(lm(h.r~temper))
```

Call:

```
lm(formula = h.r ~ temper)
```

Residuals:

```
    Min     1Q  Median     3Q    Max
-16.6413 -4.6356  0.3247  4.8304 15.8474
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -166.2847   80.9123  -2.055  0.04190 *
temper       2.4432    0.8235   2.967  0.00359 **
---

```

Ακόμη, από τη συνάρτηση της R  $lm(h.r \sim temper)$  μπορούμε εύκολα να διαπιστώσουμε ότι αυτές οι μεταβλητές ορίζουν ένα γραμμικό μοντέλο:

$$h.r = 2.44 \times temper - 166.3.$$

Όπως φαίνεται και από το κάτωθι σχήμα (Σχήμα 4.9) και οι 2 μεταβλητές ακολουθούν Κανονική κατανομή,  $temper \sim \mathcal{N}(98.3, 0.73)$  και  $h.r \sim \mathcal{N}(73.76, 7.06)$ .

Στη συνέχεια θα εφαρμόσουμε το δειγματολήπτη Gibbs για αυτό το διμεταβλητό Κανονικό μοντέλο. Πριν από αυτό, θα αναφερθούμε αναλυτικά στο γενικό Κανονικό μοντέλο δύο μεταβλητών για να γίνει πιο κατανοητός ο αλγόριθμος Gibbs. Έχουμε ήδη αναφέρει σε προηγούμενα παραδείγματα (Παράδειγμα 4.1 και Παράδειγμα 4.4) περιπτώσεις αυτού του μοντέλου με μηδενική μέση τιμή ή ίδιες τυπικές αποκλίσεις. Το γενικό διμεταβλητό μοντέλο, παρόλα αυτά, χρειάζεται μια ιδιαίτερη διευκρίνιση γιατί οι δεσμευμένες συναρτήσεις κατανομών, που πρέπει να υπολογίσουμε για την εφαρμογή του δειγματολήπτη Gibbs, είναι αρκετά πιο περίπλοκες.

Αρχικά, ας θεωρήσουμε  $Z_1, Z_2 \sim \mathcal{N}(0, 1)$  τις οποίες θα χρησιμοποιήσουμε για την παρακάτω σ.π.π. μιας γενικής διμεταβλητής Κανονικής κατανομής:

$$f(z_1, z_2) = \frac{1}{2\pi} \exp \left[ -\frac{1}{2} (z_1^2 + z_2^2) \right].$$

Θέλουμε να μετατρέψουμε τις παραπάνω μεταβλητές ώστε να έχουν τις ακόλουθες αυθαίρετες παραμέτρους:

$$X = \sigma_X Z_1 + \mu_X,$$

$$Y = \sigma_Y \left[ \rho Z_1 + \sqrt{1 - \rho^2} Z_2 \right] + \mu_Y.$$

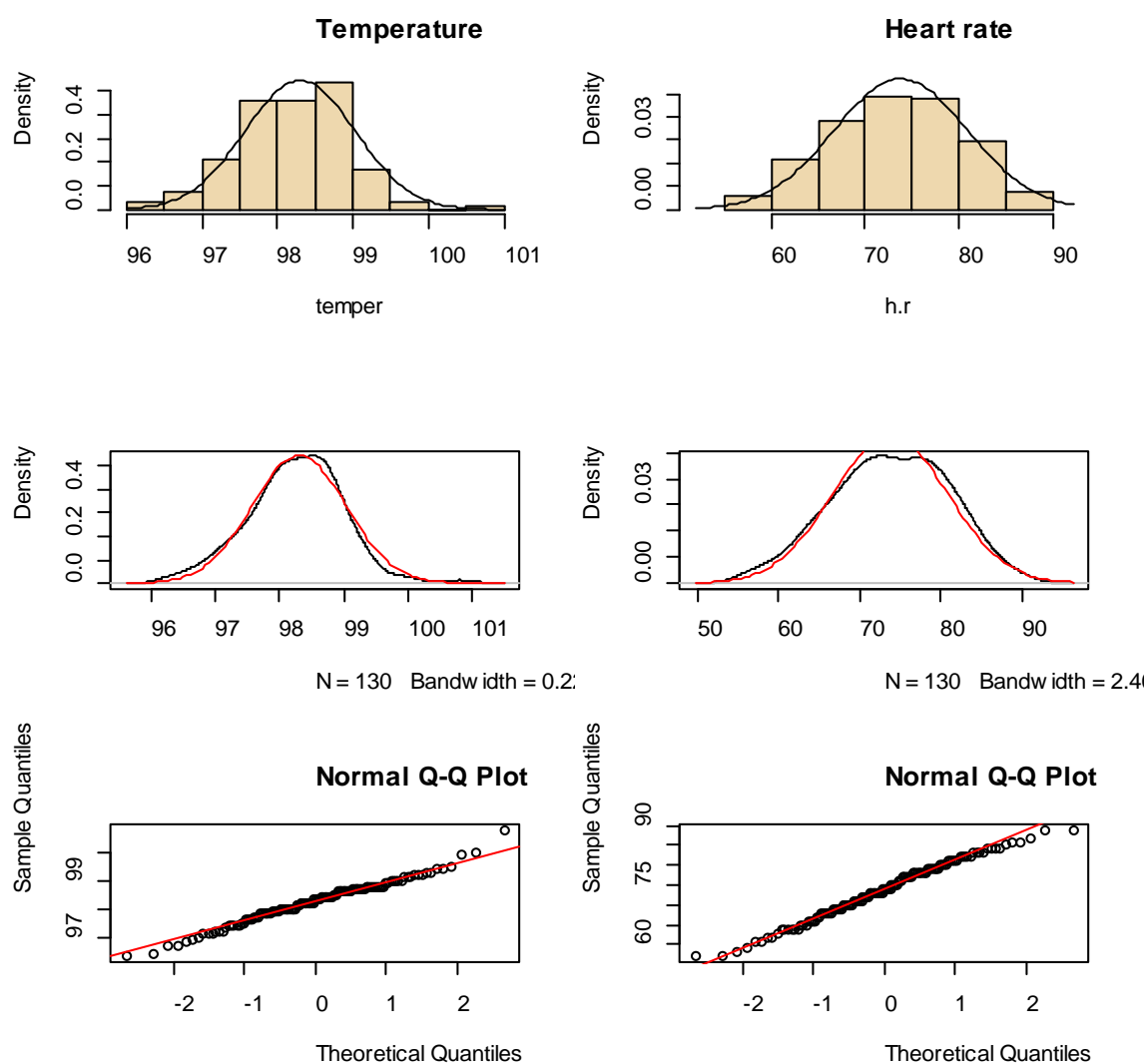


Είναι σχεδόν αυτονόητο πως οι περιθώριες κατανομές σε αυτό το μοντέλο ακολουθούν κανονικές κατανομές. Δηλαδή,  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  και  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ , ενώ η συνδιακύμανση είναι  $Cov(X, Y) = \sigma_X \sigma_Y \rho$  και άρα η αυτοσυσχέτιση θα είναι  $\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \rho$ .

Τώρα για τις δεσμευμένες κατανομές της κατανομής, χρησιμοποιώντας  $X = Z_1, Y = Z_2$ , για τη δεσμευμένη εκτίμηση  $\mathbb{E}[Y|X = x]$  θα ισχύουν τα παρακάτω:

$$\begin{aligned} \mathbb{E}[Y|X = x] &= \mathbb{E}\left[\mu_Y + \sigma_Y \left(\rho Z_1 + \sqrt{1 - \rho^2} Z_2\right) \middle| X = x\right] \\ &= \mathbb{E}\left[\mu_Y + \sigma_Y \left(\rho \frac{x - \mu_X}{\sigma_X} + \sqrt{1 - \rho^2} Z_2 \middle| X = x\right)\right] \\ &= \mu_Y + \sigma_Y \left(\rho \frac{x - \mu_X}{\sigma_X} + \sqrt{1 - \rho^2} \mathbb{E}[Z_2|X = x]\right) = \mu_Y + \sigma_Y \rho \left(\frac{x - \mu_X}{\sigma_X}\right). \end{aligned}$$

Λόγω συμμετρίας θα ισχύει ότι  $\mathbb{E}[Y|X = x] = \mu_X + \sigma_X \rho \left(\frac{y - \mu_Y}{\sigma_Y}\right)$ .



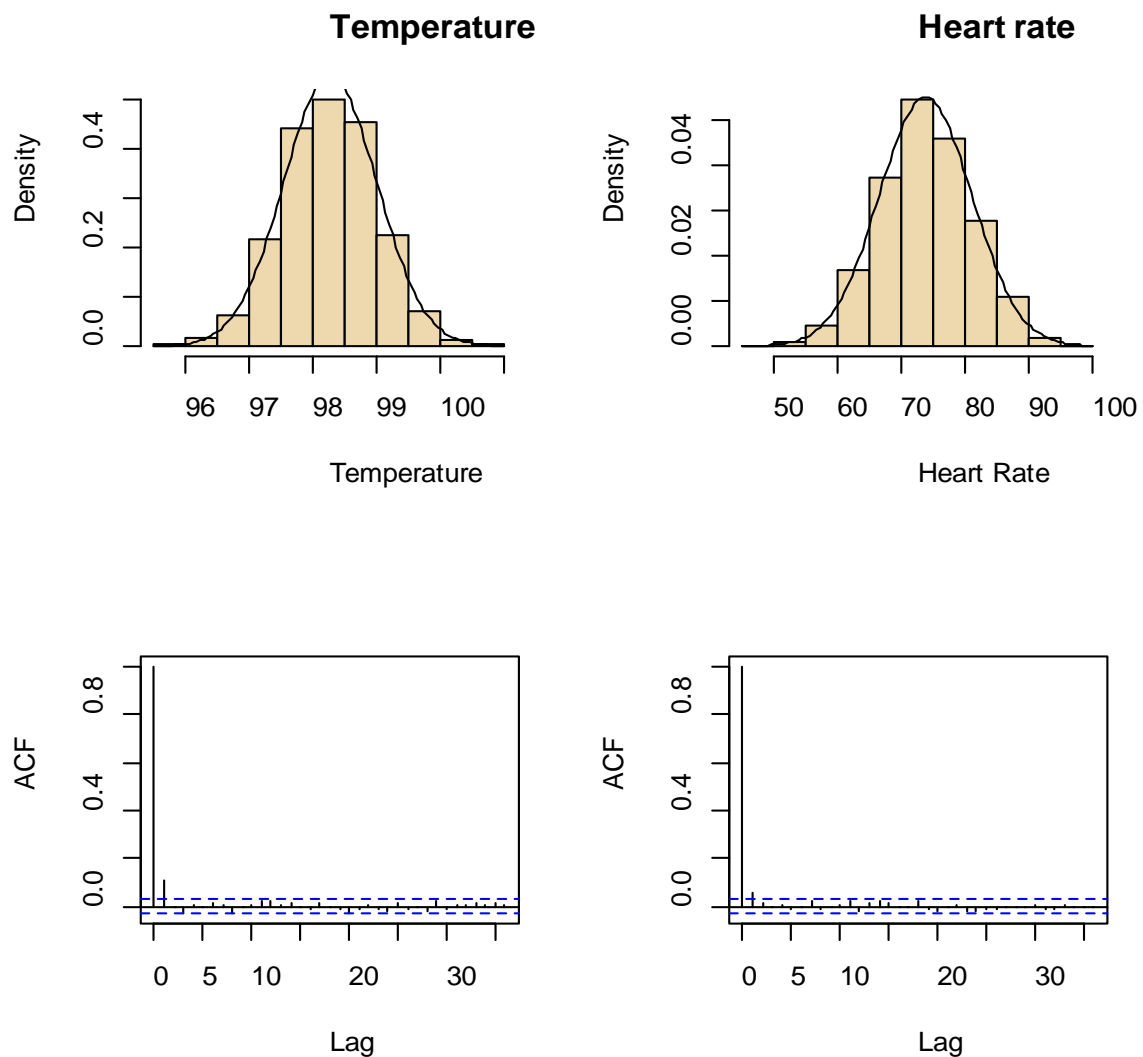
**Σχήμα 4.9:** Ιστογράμματα, συναρτήσεις πυκνότητας πιθανότητας και έλεγχοι κανονικότητας για τις 2 μεταβλητές: θερμοκρασία και καρδιακοί παλμοί της Εφαρμογής 1.

Άρα, θα κατασκευάσω τον ακόλουθο δειγματολήπτη Gibbs για το διμεταβλητό Κανονικό μοντέλο αυτής της εφαρμογής.



```
> N<-5000
> Burn<-1000
> X <- matrix(0, N, 2)
> rho<-0.25
> mean(temper)
[1] 98.24923
> mu1<-98.25
> mean(h.r)
[1] 73.76154
> mu2<-73.76
> sd(temper)
[1] 0.7331832
> sd(h.r)
[1] 7.062077
> sigma1<-0.73
> sigma2<-7.06
> s1<-sqrt(1-rho^2)*sigma1
> s2<-sqrt(1-rho^2)*sigma2
> X[1, ] <- c(mu1, mu2)
> for (i in 2:N) {
+ x2 <- X[i-1, 2]
+ m1 <- mu1 + rho * (x2 - mu2) * sigma1/sigma2
+ X[i, 1] <- rnorm(1, m1, s1)
+ x1 <- X[i, 1]
+ m2 <- mu2 + rho * (x1 - mu1) * sigma2/sigma1
+ X[i, 2] <- rnorm(1, m2, s2)
+ }
> b <- burn + 1
> x <- X[b:N, ]
> colMeans(x)
[1] 98.23859 73.67606
> cov(x)
      [,1] [,2]
[1,] 0.5365233 1.487357
[2,] 1.4873574 52.886383
> cor(x)
      [,1] [,2]
[1,] 1.0000000 0.2792217
[2,] 0.2792217 1.0000000
```

Όπως βλέπουμε και από το Σχήμα 4.10, όντως και οι 2 περιθώριες κατανομές ακολουθούν Κανονική κατανομή, κάτι το οποίο το περιμέναμε από τη παραπάνω θεωρία του μοντέλου. Όπως έχουμε ήδη αναφέρει, όλες οι περιθώριες κατανομές του πολυμεταβλητού κανονικού μοντέλου ακολουθούν Κανονική κατανομή. Ακόμη, διαπιστώνουμε πως δεν υπάρχει σημαντική αυτοσυσχέτιση των προσομοιωμένων τιμών μας.



**Σχήμα 4.10:** Ιστογράμματα και αυτοσυσχετίσεις για τις περιθώριες κατανομές και για τις προσομοιωμένες τιμές αντίστοιχα με το δειγματολήπτη Gibbs για το πολυμεταβλητό Κανονικό μοντέλο της Εφαρμογής 1.

## Κεφάλαιο 5 Άλλοι Μαρκοβιανοί Δειγματολήπτες

---

Ήδη έχουμε μελετήσει δύο από τις πιο σημαντικές Markov Chain Monte Carlo μεθόδους. Έτσι, διαπιστώσαμε τελικά πως για τη σωστή χρήση των Μαρκοβιανών αλυσίδων, είναι πολύ σημαντική η επιλογή της κατάλληλης MCMC μεθόδου με σκοπό τη μείωση των υπολογισμών αλλά και του απαραίτητου χρόνου για την εκάστοτε δειγματοληψία. Σε αυτό το κεφάλαιο θα κάνουμε μια σύντομη αναφορά σε κάποιους άλλους Μαρκοβιανούς δειγματολήπτες που χρίζουν ιδιαίτερης σημασίας.

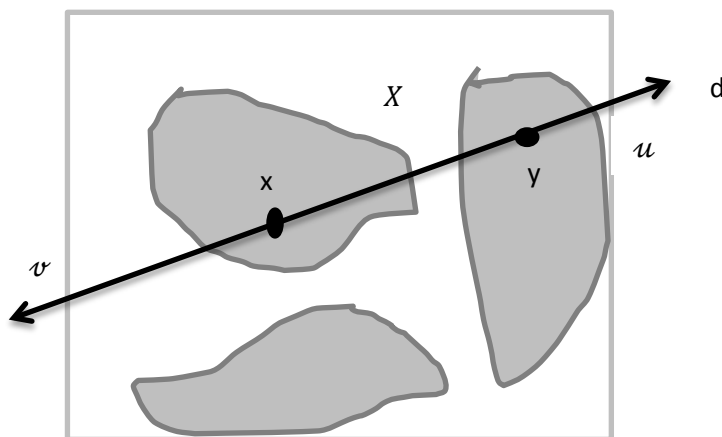
### 1.0 Δειγματολήπτης Hit-and-Run

---

Ο hit-and-run δειγματολήπτης από τον Robert Smith είναι ανάμεσα στους πρώτους MCMC δειγματολήπτες στην κατηγορία των διαδοχικών δειγματοληπτών. Και σε αυτόν τον δειγματολήπτη ο κύριος σκοπός μας είναι να παράγουμε δείγμα από τη ζητούμενη κατανομή  $f(x)$  σε ένα σύνολο  $X \subset \mathbb{R}^n$ . Το κύριο χαρακτηριστικό αυτού του είδους των δειγματοληπτών είναι ότι μπορούν να εξερευνήσουν ολόκληρη την περιοχή της ζητούμενης συνάρτησης σε ένα μόνο βήμα.

Πρώτα θα περιγράψουμε τον αρχικό hit-and-run δειγματολήπτη για την παραγωγή από μια ομοιόμορφη κατανομή σε μία οριοθετημένη ανοικτή περιοχή  $X$  του  $\mathbb{R}^n$ . Σε κάθε επανάληψη, ξεκινώντας από ένα τρέχον σημείο  $x$ , ένα διάνυσμα κατεύθυνσης  $d$  παράγεται ομοιόμορφα στην επιφάνεια μιας υπερσφαίρας  $n$ -διαστάσεων. Η τομή της γραμμής (που διέρχεται από το  $x$ ) με το κουτί που περικλείει το σύνολο  $X$  ορίζουν ένα ευθύγραμμο τμήμα  $\mathcal{L}$ . Το επόμενο σημείο  $y$  επιλέγεται ομοιόμορφα από τη τομή του  $\mathcal{L}$  και του  $X$ .

Το Σχήμα 5.1 απεικονίζει τον hit-and-run αλγόριθμο για την παραγωγή τιμών ομοιόμορφα από το σύνολο  $X$  (το οποίο απεικονίζεται με γκρι χρώμα) που οριοθετείται από ένα τετράγωνο. Δεδομένου ότι το σημείο  $x$  ανήκει στο σύνολο  $X$ , ένα τυχαίο διάνυσμα κατεύθυνσης  $d$  παράγεται, το οποίο ορίζει το ευθύγραμμο τμήμα  $\mathcal{L} = uv$ . Στη συνέχεια επιλέγεται με ομοιόμορφο τρόπο ένα σημείο  $y$  στην τομή  $\mathcal{M} = \mathcal{L} \cap X$ . Αυτή η επιλογή αυτή μπορεί να γίνει για παράδειγμα, μέσω της μεθόδου Αποδοχής-Απόρριψης (δηλαδή παράγεται ένα σημείο ομοιόμορφα πάνω στην  $\mathcal{L}$  και μετά γίνεται αποδοχή αυτού του σημείου μόνο αν βρίσκεται στο  $X$ ).



**Σχήμα 5.1:** Απεικόνιση του αλγόριθμου hit-and-run σε ένα τετράγωνο δύο διαστάσεων.

Ο Smith έδειξε ότι ο hit-and-run αλγόριθμος παράγει τιμές ομοιόμορφα κατανομημένες πάνω σε αυθαίρετες περιοχές συνόλων του  $\mathbb{R}^n$ . Όπως ήδη αναφέραμε, μία ιδιαίτερα σημαντική ιδιότητα αυτού του δειγματολήπτη είναι ότι μπορεί γενικά να φτάσει σε οποιοδήποτε σημείο που ανήκει στο σύνολο σε ένα μόνο βήμα, δηλαδή υπάρχει μια αυστηρά θετική πιθανότητα για τη δειγματοληψία από οποιαδήποτε συγκεκριμένη περιοχή του συνόλου. Αυτή η ιδιότητα σε συνδυασμό με μία συμμετρική ιδιότητα είναι ιδιαίτερα σημαντική για την εύρεση της οριακής κατανομής. Στην ουσία αποδεικνύεται ότι ο αλγόριθμος hit-and-run προσφέρει πρακτικά την πιο γρήγορη σύγκλιση σε μια ομοιόμορφη κατανομή.

Η μέθοδος αυτή που περιγράψαμε πιο πάνω είναι στην ουσία μια ειδική περίπτωση του αλγόριθμου Metropolis-Hastings όταν η κατανομή εισήγησης είναι συμμετρική και η ζητούμενη συνάρτηση  $f$  είναι σταθερή. Έτσι, συνεπάγεται ότι κάθε υποψήφια τιμή επιλέγεται με πιθανότητα 1.

Ο αλγόριθμος αυτός είναι ιδιαίτερα δημοφιλής στους συνεχείς χώρους. Δηλαδή, περιληπτικά η δειγματοληψία hit-and-run στους συνεχείς χώρους με  $X_1 \in X$  και  $t = 1$  παρατίθεται στον κάτωθι αλγόριθμο:

### **Αλγόριθμος 7: Δειγματολήπτης Hit-and-Run**

1. Παρήγαγε ένα τυχαίο διάνυσμα κατεύθυνσης  $d_t$  ομοιόμορφα κατανομημένο σε μία σφαίρα  $n$ -διαστάσεων.
2. Παρήγαγε  $Y = X_t + \lambda d_t$  το οποίο κατανέμεται ομοιόμορφα πάνω στο μη κενό σύνολο  $\{x: x \in X \text{ και } x = X_t + \lambda d_t, \lambda \in \mathbb{R}\}$
3. Πάρε  $X_{t+1} = \begin{cases} Y & \text{με πιθανότητα } \min\left\{\frac{f(y)}{f(x)}, 1\right\} \\ X_t & \text{διαφορετικά} \end{cases}$

## 2.Δειγματολήπτης Τμημάτων (Slice Sampler)

Όπως έχουμε ήδη αναφέρει οι μέθοδοι MCMC υιοθετούν τα χαρακτηριστικά της κατανομής από την οποία προσομοιώνονται χρησιμοποιώντας σαν κύρια ιδέα ότι κάποιος μπορεί να πάρει ένα δείγμα από μία κατανομή προσομοιώνοντας π.χ. από μία ομοιόμορφη κατανομή στην περιοχή κάτω από το διάγραμμα της σ.π.π.

Μία Μαρκοβιανή αλυσίδα που συγκλίνει σε αυτήν την ομοιόμορφη κατανομή μπορεί να κατασκευαστεί εναλλάσσοντας ομοιόμορφες δειγματοληψίες κατά την κατακόρυφη διεύθυνση με ομοιόμορφες δειγματοληψίες από τα οριζόντια τμήματα (slice) που ορίζονται από τη τρέχουσα κατακόρυφη θέση. Τέτοιες «δειγματοληψίες τμημάτων» (slice sampling) εφαρμόζονται πολύ εύκολα για κατανομές με μία μεταβλητή αλλά μπορούν να χρησιμοποιηθούν και στη δειγματοληψία πολυμεταβλητών κατανομών ανανεώνοντας με τη σειρά την κάθε μεταβλητή.

Αυτή η προσέγγιση είναι συχνά πολύ πιο εύκολη να εφαρμοστεί από το δειγματολήπτη Gibbs, στον οποίο πολλές φορές είναι απαραίτητη η ανάπτυξη μεθόδων για τη δειγματοληψία από όχι και τόσο συνηθισμένες μονοδιάστατες κατανομές, και πολύ πιο αποτελεσματική από τους αλγόριθμους Metropolis-Hastings, στους οποίους είναι απαραίτητη η εύρεση της κατάλληλης κατανομής εισήγησης που θα οδηγήσει σε μία αποτελεσματική δειγματοληψία. Ακόμη, οι περισσότεροι Μαρκοβιανοί δειγματολήπτες είναι αναποτελεσματικοί κυρίως για δύο λόγους:

1. Οι αλλαγές που μπορεί να επιχειρήσουν μπορεί να μην προσαρμόζονται καλά στις τοπικές ιδιότητες των σ.π.π. με αποτέλεσμα αυτές οι αλλαγές να πραγματοποιούνται τελικά σε πολύ μικρά βήματα.
2. Αυτά τα μικρά βήματα παίρνουν τη μορφή ενός τυχαίου περίπατου, στον οποίο απαιτούνται πολύ περισσότερα βήματα σε αντίθεση με εκείνα που θα μπορούσαν να συμβούν αν η κίνηση αυτών των βημάτων ήταν προς μία κατεύθυνση μόνο.

Παρόλα αυτά, η μέθοδος αυτή σε πολλές κατανομές δεν συνεπάγεται μια πιο αποτελεσματική δειγματοληψία από εκείνη των Metropolis-Hastings και Gibbs, αλλά αυτός ο *τμηματοποιημένος δειγματολήπτης* ίσως απαιτεί λιγότερη προσπάθεια και συντονισμό για την εφαρμογή του. Στην ουσία αυτή η μέθοδος είναι μια γενική εκδοχή του δειγματολήπτη Gibbs.

Πιο συγκεκριμένα, έστω ότι θέλουμε ένα δείγμα από μία κατανομή για μία τυχαία μεταβλητή  $x \in X \subset \mathbb{R}_n$ , της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι η συνάρτηση  $f(x)$ . Αυτό επιτυγχάνεται με τη δειγματοληψία από μία περιοχή  $(n+1)$ -διάστασης που βρίσκεται κάτω από το διάγραμμα της  $f(x)$ . Δηλαδή, με την εισαγωγή μιας βοηθητικής πραγματικής μεταβλητής  $y$  ορίζουμε μία από κοινού κατανομή πάνω στο  $x$  και  $y$  που είναι ομοιόμορφη στην περιοχή  $U = \{(x, y): 0 < y < f(x)\}$ , δηλαδή στην περιοχή κάτω από την καμπύλη ή την επιφάνεια που ορίζεται από την  $f(x)$ .

Δηλαδή, η από κοινού συνάρτηση πυκνότητας πιθανότητας θα είναι:

$$f(x, y) = \begin{cases} \frac{1}{\int f(x) dx}, & \text{αν } 0 < y < f(x) \\ 0, & \text{διαφορετικά} \end{cases}$$

και έτσι η περιθώρια σ.π.π. του  $X$  θα είναι:

$$f_X(x) = \int_0^{f(x)} \frac{1}{\int f(x) dx} dy = \frac{f(x)}{\int f(x) dx}.$$

Έτσι, για να πάρουμε μία τιμή του  $x$ , μπορούμε να πάρουμε ένα δείγμα από την από κοινού κατανομή στο  $U$  αγνοώντας τα  $y$ .

Η παραγωγή ανεξάρτητων σημείων ομοιόμορφα κατανεμημένων από το  $U$  μπορεί να μην είναι πάντα εύκολη, γι' αυτό μπορούμε να ορίσουμε μία Μαρκοβιανή αλυσίδα που να συγκλίνει στην ζητούμενη ομοιόμορφη κατανομή. Έτσι, μία λύση μπορεί να είναι η χρήση μίας δειγματοληψίας Gibbs, δηλαδή η παραγωγή τιμών από τη δεσμευμένη συνάρτηση πυκνότητας πιθανότητας  $f_{Y|X}(y|x)$ , που είναι ομοιόμορφη στο διάστημα  $(0, f(x))$ , και από την  $f_{X|Y}(x|y)$ , που είναι ομοιόμορφη πάνω στην περιοχή  $A = \{x: y < f(x)\}$ . Ωστόσο, η παραγωγή ανεξάρτητων σημείων από το σύνολο  $A$  μπορεί και πάλι να αποβεί και πάλι δύσκολη. Δηλαδή, περιληπτικά θα έχουμε τον παρακάτω αλγόριθμο.

### **Αλγόριθμος 8: Slice Sampler σε 2 διαστάσεις**

Στην επανάληψη  $t$  προσομοίωσε

1.  $Y^{t+1} \sim \mathcal{U}_{[0, f(x)]}$ ;
2.  $X^{t+1} \sim \mathcal{U}_{A^{t+1}}$

$$\text{με } A^{t+1} = \{x: f(x) \geq y^{t+1}\}$$

**Παράδειγμα 5.1:** Ας υποθέσουμε ότι έχουμε τη συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{2} e^{-\sqrt{x}} \quad \forall x > 0.$$

Αυτή η συνάρτηση μπορεί να προσομοιωθεί εύκολα και απευθείας αλλά και με τον δειγματολήπτη τμημάτων. Σύμφωνα με τον παραπάνω αλγόριθμο αρχικά θα βρω το

σύνολο  $A$  και τις αντίστοιχες δεσμευμένες σ.π.π. της παραπάνω συνάρτησης. Δηλαδή, εφόσον  $A = \left\{x: \frac{1}{2}e^{-\sqrt{x}} \geq y\right\}$  θα ισχύει:


$$A = \{x : 0 \leq x \leq \log(2y)\}.$$

Άρα:

$$X|Y \sim \mathcal{U}(0, (\log(2y))^2)$$

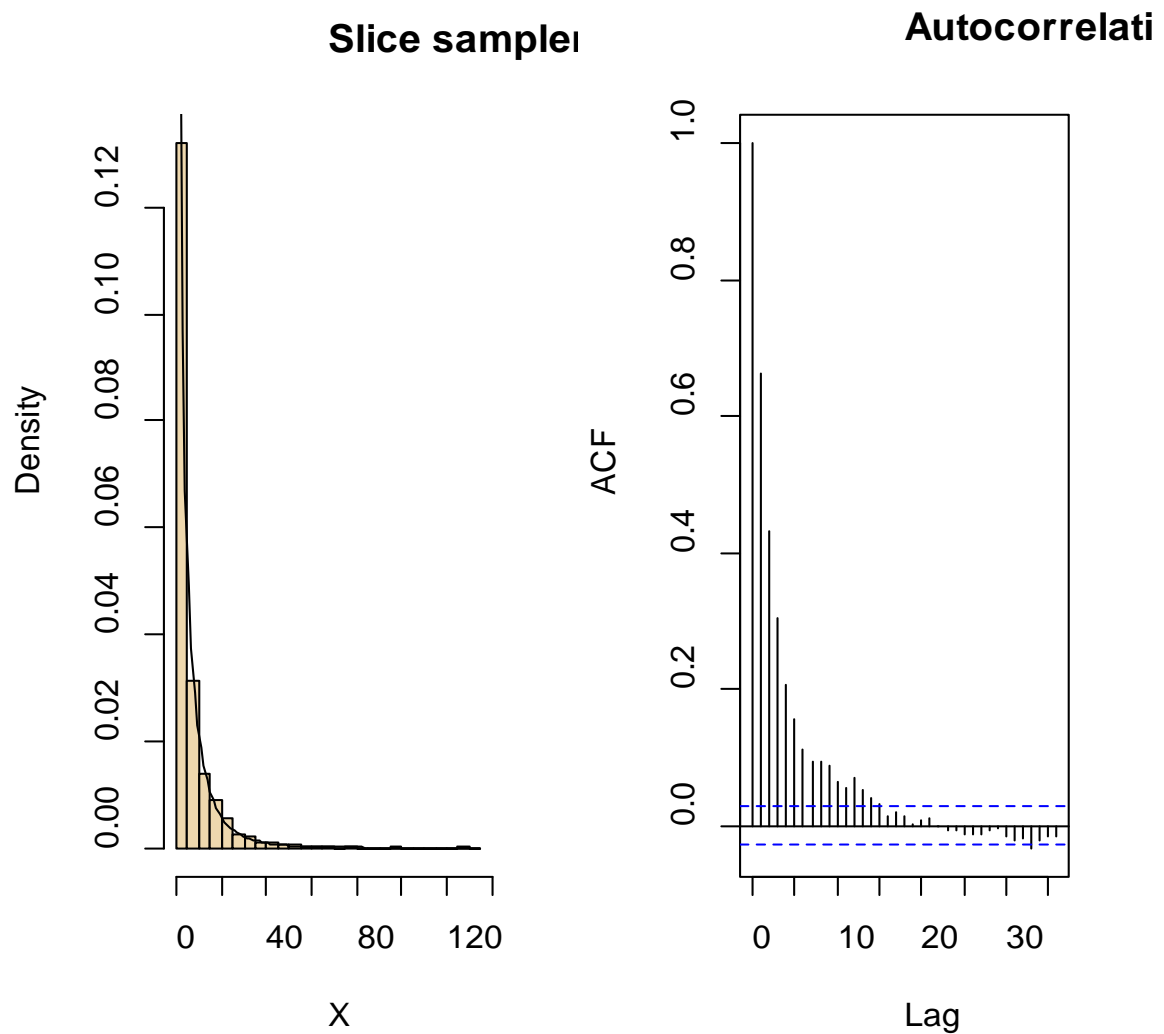
$$Y|X \sim \mathcal{U}\left(0, \frac{1}{2}e^{-\sqrt{x}}\right).$$

Έτσι, με την εφαρμογή της δειγματοληψίας Gibbs για την παραγωγή τυχαίων μεταβλητών από την  $f_X(x)$  θα έχω τον ακόλουθο κώδικα στην R:



```
> T=5000
> f=function(x){
+ 1/2*exp(-sqrt(x))}
> X=c(runif(1))
> Y=c(runif(1))
> for (t in 1:T){
+ Y=c(Y,runif(1,0,f(X[t])))
+ X=c(X,runif(1,0,(log(2*Y[t+1]))^2))
+ }
> par(mfrow=c(1,2))
> hist(X,prob=TRUE,col="wheat2",main="Slice sampler",breaks=20)
> curve(f(x), add=TRUE)
> acf(X,main="Autocorrelation")
```





**Σχήμα 5.2:** Ιστόγραμμα δειγματολήπτη τμημάτων με επικαλυπτόμενη τη σ.π.π. χρησιμοποιώντας 5000 επαναλήψεις και το αντίστοιχο γράφημα της συνάρτησης αυτοσυσχέτισης.

**Παράδειγμα 5.2:** Ας θεωρήσουμε τώρα την περικομμένη συνάρτηση  $TN(-3,1)$  περιορισμένη πάνω στο  $[0,1]$ :

$$f(x) \propto f_1(x) = \exp\left\{-\frac{(x+3)^2}{2}\right\} \mathbb{I}_{[0,1]}(x).$$

Γενικά, αν  $X \sim N(\mu, \sigma^2)$  και ορίσουμε σαν  $Y \sim TN(\mu, \sigma^2, a, b)$ , δηλαδή αν  $Y$  είναι μια περικομμένη τυχαία μεταβλητή, τότε η συνάρτηση πυκνότητας πιθανότητας της περικομμένης αυτής μεταβλητής θα είναι:

$$f_Y(y) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \mathbb{I}_{[a,b]}(y),$$

όπου  $\Phi$  είναι η τυπική σ.π.π. της Κανονικής Κατανομής και για τη δείκτρια συνάρτηση  $\mathbb{I}$  ισχύει:  $\mathbb{I}_{[a,b]}(y) = 1$  αν  $a \leq y \leq b$ , διαφορετικά είναι 0.

Σε αυτό το παράδειγμα και σύμφωνα με τα παραπάνω θα χρησιμοποιήσουμε ένα τμηματοποιημένο δειγματολήπτη. Έτσι θα έχουμε τα παρακάτω βήματα του αλγόριθμου:

1. Παράγω

$$Y^t \sim U[0, \exp\{-(x+3)^2/2\}]$$

2. Στη συνέχεια πρέπει να βρούμε το σύνολο  $A = \{x: f(x) \geq y\}$ . Τα σημεία για τα οποία ισχύει  $f(x) = y$  υπολογίζονται εύκολα από την παρακάτω εξίσωση:


$$\exp\left\{-\frac{(x+3)^2}{2}\right\} = y \Rightarrow \log(y) = -\frac{(x+3)^2}{2} \Rightarrow (x+3)^2 = -2\log(y)$$

και άρα θα έχουμε ότι:

$$A^{t+1} = \{y \in [0,1]; (y+3)^2 \leq -2\log(w^t)\},$$

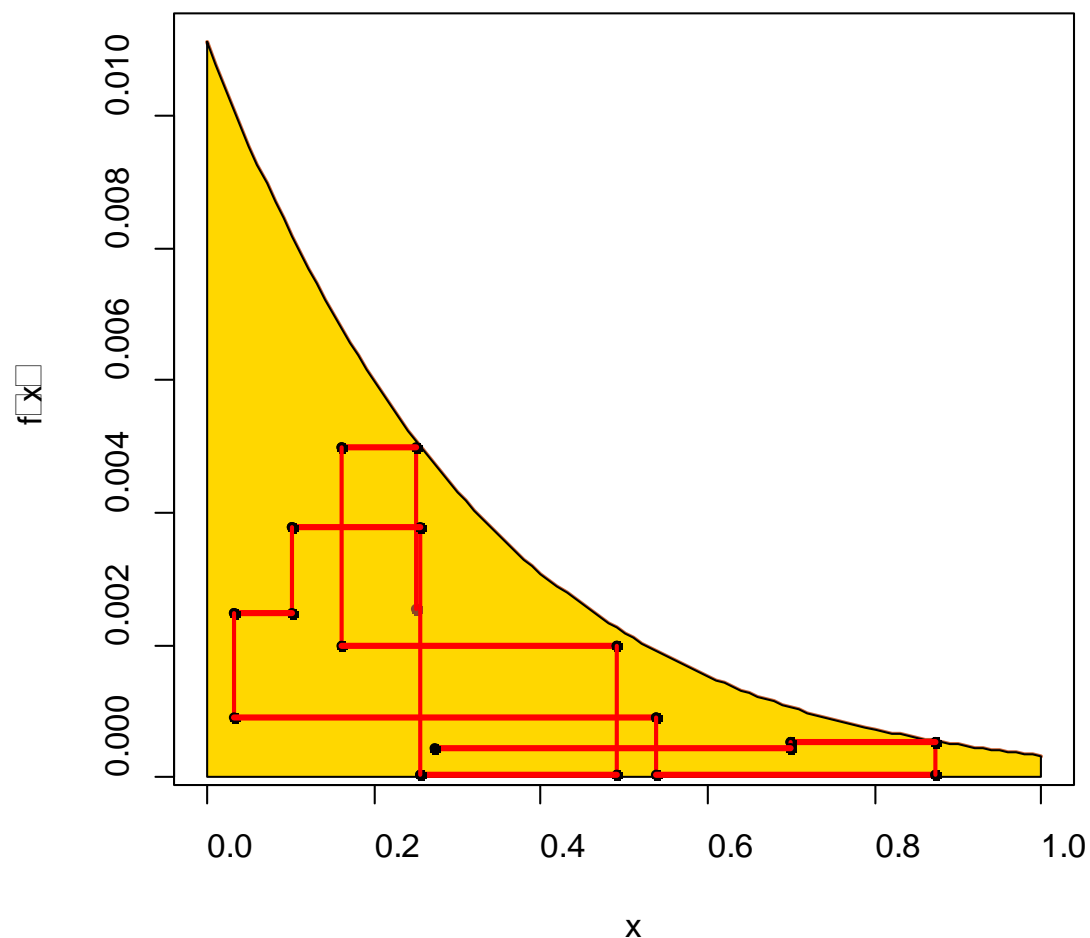
όπου  $w^t = u^t f_1(x^t)$ .

Σύμφωνα με όλα τα παραπάνω προκύπτει ο κάτωθι κώδικας στην R:



```
> xes <- seq(0,1,.01)
> xes <- c(xes,1-xes)
> yes <- c(exp(-(seq(0,1,.01)+3)^2/2),0*seq(0,1,.01))
> plot(xes[1:101],yes[1:101],type="l",lwd=2,col="sienna3",xlab="x",ylab="expression(f(x))")
> polygon(xes,yes,col="gold")
> obs <- matrix(0,ncol=2,nrow=100)
> obs[1,1] <- .25
> obs[1,2] <- 0.5*exp(-0.5*(obs[1,1]+3)^2)
> points(obs[1,1],obs[1,2],col="sienna4",pch=20)
> for (i in 2:100){
+   obs[i,2] <- exp(-0.5*(obs[i-1,1]+3)^2)*runif(1)
+   obs[i,1] <- runif(1)*min(1,-3+sqrt(-2*log(obs[i,2])))
+ }
> for (i in 2:10){
+   if (i>2) points(obs[i-1,1],obs[i-1,2],col="black",pch=20)
+   lines(c(obs[i-1,1],obs[i-1,1]),c(obs[i-1,2],obs[i,2]),col="red", lwd=2)
+   points(obs[i-1,1],obs[i,2],col="black",pch=20)
+   lines(c(obs[i-1,1],obs[i,1]),c(obs[i,2],obs[i,2]),col="red", lwd=3)
+   points(obs[i,1],obs[i,2],col="black",pch=20)
+ }
```

- Δηλαδή, σαν  $xes$  ορίσαμε μία ακολουθία από 0 μέχρι 1, αφού έχουμε ορίσει την περικομμένη Κανονική συνάρτηση μας πάνω στο  $[0,1]$  και τα αντίστοιχα  $y$  αυτής της περικομμένης συνάρτησης σαν  $yex$ .
- Στον πίνακα *obs* καταχωρούμε τα αποτελέσματα του δειγματολήπτη μας και δίνουμε αρχικές τιμές στις μεταβλητές μας ενώ στη συνέχεια εκτελούμε τα βήματα του δειγματολήπτη τμημάτων με την προϋπόθεση όμως οι τιμές του  $y$  να ανήκουν στο  $[0,1]$ .
- Στη συνέχεια με τις κατάλληλες συναρτήσεις της R σχεδιάζω το Σχήμα 5.3 που παρουσιάζει τα 10 πρώτα βήματα του αλγόριθμου μας.



Σχήμα 5.3: Τα πρώτα 10 βήματα του slice sampler του παραδείγματος 5.2.



```

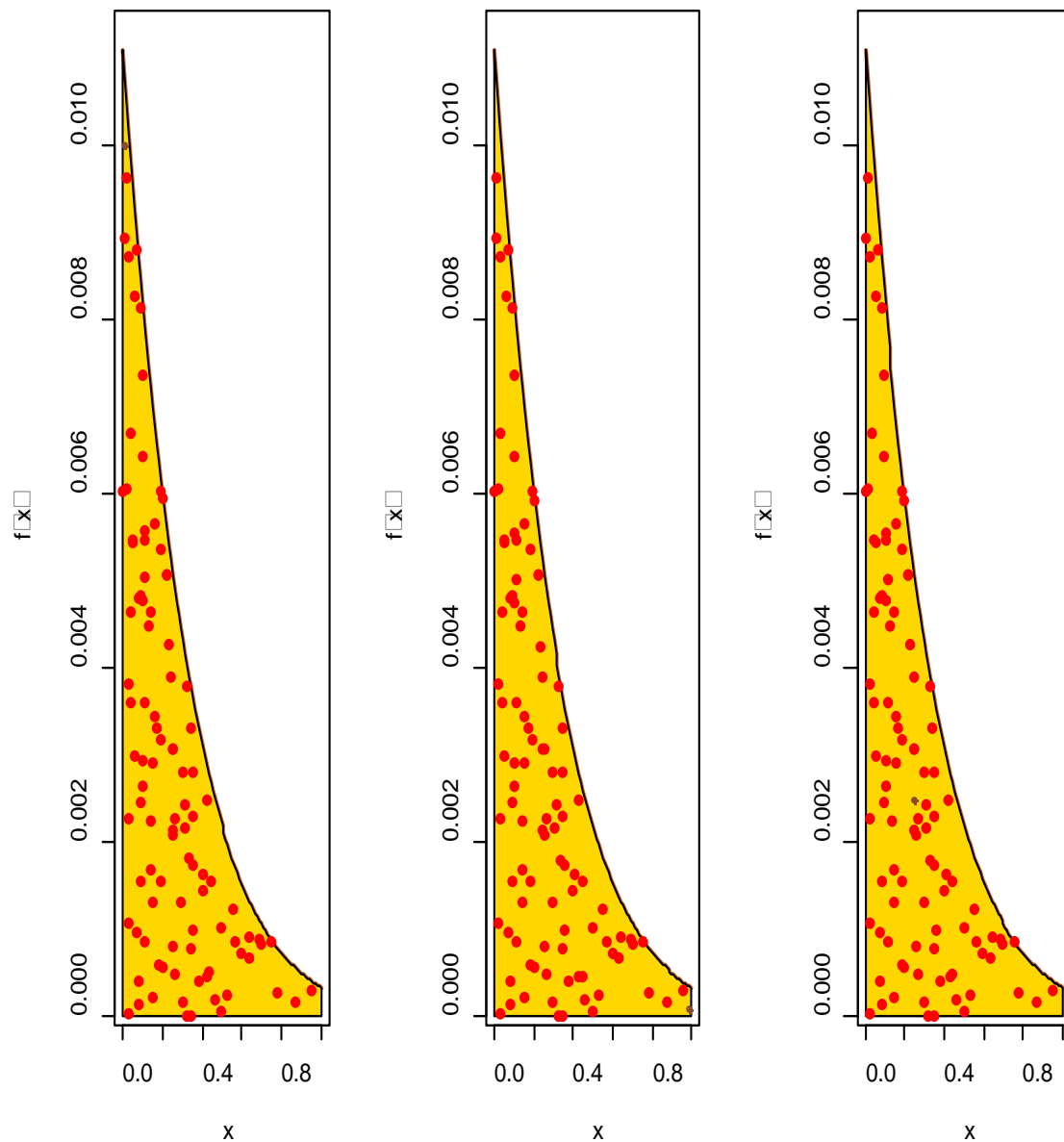
> X11()
> par(mfrow=c(1,3))
> losunif<-runif(2100)
> plot(xes[1:101],yes[1:101],type="l",lwd=2,col="sienna3",xlab="x",
+ ylab=expression(f(x)))
> polygon(xes,yes,col="gold")
> obs<-matrix(0,ncol=2,nrow=100)
> obs[1,1]<-.01
> obs[1,2]<-.01
> points(obs[1,1],obs[1,2],col="sienna4",pch=20)
> counte<-0
> for (j in 1:100){
+ for (i in 2:10){
+ counte<-counte+1
+ obs[i,2]<-exp(-0.5*(obs[i-1,1]+3)^2)*losunif[counte]
+ counte<-counte+1
+ obs[i,1]<-losunif[counte]*min(1,-3+sqrt(-2*log(obs[i,2]))) }
+ points(obs[i,1],obs[i,2],col="red",pch=20, lwd=3)
+ }
> plot(xes[1:101],yes[1:101],type="l",lwd=2,col="sienna3",xlab="x",
+ ylab=expression(f(x)))
> polygon(xes,yes,col="gold")
> obs<-matrix(0,ncol=2,nrow=100)
> obs[1,1]<-.99
> obs[1,2]<-.0001
> points(obs[1,1],obs[1,2],col="sienna4",pch=20)
> counte<-0
> for (j in 1:100){
+ for (i in 2:10){
+ counte<-counte+1
+ obs[i,2]<-exp(-0.5*(obs[i-1,1]+3)^2)*losunif[counte]
+ counte<-counte+1
+ obs[i,1]<-losunif[counte]*min(1,-3+sqrt(-2*log(obs[i,2])))}
+ points(obs[i,1],obs[i,2],col="red",pch=20, lwd=3)
+ }
> plot(xes[1:101],yes[1:101],type="l",lwd=2,col="sienna3",xlab="x",
+ ylab=expression(f(x)))
> polygon(xes,yes,col="gold")

```



```
> obs<-matrix(0,ncol=2,nrow=100)
> obs[1,1]<-.25
> obs[1,2]<-.0025
> points(obs[1,1],obs[1,2],col="sienna4",pch=20)
> counte<-0
> for (j in 1:100){
+   for (i in 2:10){
+     counte<-counte+1
+     obs[i,2]<-exp(-0.5*(obs[i-1,1]+3)^2)*losunif[counte]
+     counte<-counte+1
+     obs[i,1]<-losunif[counte]*min(1,-3+sqrt(-2*log(obs[i,2])))}
+     points(obs[i,1],obs[i,2],col="red",pch=20, lwd=3)
+   }
```

Το εντυπωσιακό σε αυτόν τον αλγόριθμο του τμηματοποιημένου δειγματολήπτη είναι πως αν δοκιμάσουμε διαφορετικές τιμές εκκίνησης θα διαπιστώσουμε από το Σχήμα 5.4 ότι δεν παίζει σημαντικό ρόλο η αρχική τιμή στον δειγματολήπτη τμημάτων σε αντίθεση με τον αλγόριθμο Metropolis-Hastings και το δειγματολήπτη Gibbs που αναλύσαμε στα προηγούμενα κεφάλαια.



**Σχήμα 5.4:** Σύγκριση των δειγμάτων για επαναλήψεις του Παραδείγματος 5.2 για τον τμηματοποιημένο δειγματολήπτη με αρχικές τιμές (0.01,0.01), (0.99,0.001) και (0.25,0.25).

# Βιβλιογραφία

---

## Ελληνική Βιβλιογραφία

1. Κοκολάκης, Γ.Ε. (2008) *Σημειώσεις Στοχαστικών Ανελίξεων*, Εκδόσεις ΕΜΠ, Αθήνα.
2. Κοκολάκης, Γ.Ε και Φουσκάκης, Δ.Σ. (2005) *Σημειώσεις Στατιστικής*, Εκδόσεις ΕΜΠ, Αθήνα.
3. Χρυσάφινου, Ο. (2010) *Εισαγωγή στις Στοχαστικές Ανελίξεις*, Εκδόσεις Σοφία, Θεσσαλονίκη.

## Ξενόγλωσση Βιβλιογραφία

4. Robert, C. and Casella, G. (2010) *Introducing Monte Carlo Methods with R (Use R)*, Springer, New York.
5. Rubinstein, Reuven Y. and Kroese, Dirk P. (2008) *Simulation and the Monte Carlo Method (Second Edition)*, Wiley, New Jersey.
6. Suess, Eric A. and Trumbo, Bruce E.(2010) *Introduction to Probability Simulation and Gibbs Sampling with R*, Springer, New York.
7. Albert, J. and Rizzo, M. (2012) *R by Example, Use R!*, Springer, New York.
8. Neal, R.M (2003) *Slice Sampling*, The Annals of Statistics, Vol.31, No.3, p. 705-767.
9. Andrieu, C., Nando De Freitas, Doucet, A. and Jordan, M.I (2003) *An Introduction to MCMC for Machine Learning*, Machine Learning,50,5-43, p. 12-13.
10. Dagpunar, J.S. (2007) *Simulation and Monte Carlo with applications in finance and MCMC*, Wiley, England.
11. Robert, C. and Casella, G. (2011) *A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data*, Statistical Science, Volume 26, No.1, p. 102-115.
12. Casella, G. and George, E.I. (1992) *Explaining the Gibbs Sampler*, The American Statistician, Volume 46, p. 167-174.

13. Congdon, P. (2006) *Bayesian Statistical Modelling (2<sup>nd</sup> Edition)*, Wiley, UK
14. Beichl, I. and Sullivan, F. (2000) *The Metropolis Algorithm*, IEEE Computing in Science and Engineering, p.65-69.
15. Diaconis, P. and Saloff-Coste, L. (1998) *What do we know about the Metropolis Algorithm?*, Journal of Computer and System Sciences, No. 57, p. 20-36.
16. Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The elements of Statistical Learning (2<sup>nd</sup> Edition)*, Springer-Verlag, New York.
17. Chib, S. and Greenberg, E. (1995) *Understanding the Metropolis-Hastings Algorithm*, The American Statistician, Vol.49, No.4, p. 327-33.
18. Robert, C. (1995) *Simulation of truncated Normal variables*, Statistics and Computing, No.5, p. 121-125.
19. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1995) *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.



