

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ



Μελετώντας τον αλγόριθμο Metropolis-Hastings

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΜΑΘΗΜΑΤΙΚΑ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΤΩΝ ΑΠΟΦΑΣΕΩΝ»

Νικόλαος Γιαννόπουλος

Επιβλέπουσα Καθηγήτρια : Λέκτορας Σόνια Μαλεφάκη

ΠΑΤΡΑ, Οκτώβριος 2012

ΜΕΛΗ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ

ΕΓΚΡΙΣΗ ΜΕ ΥΠΟΓΡΑΦΕΣ

Τα μέλη της
Τριμελούς Εξεταστικής Επιτροπής

Όνοματεπώνυμο

Όνοματεπώνυμο

Όνοματεπώνυμο

Σόνια Μαλεφάκη

Γεώργιος Ηλιόπουλος

Κωνσταντίνος Πετρόπουλος



Περιεχόμενα

1	Εισαγωγή	1
2	Μέθοδοι Προσομοίωσης	3
2.1	Μέθοδοι Monte Carlo	3
2.1.1	Ολοκλήρωση Monte Carlo	4
2.1.2	Μέθοδος Αντιστροφής	6
2.1.3	Μέθοδος Αποδοχής-Απόρριψης	7
2.1.4	Δειγματολήπτης Σπουδαιότητας (Importance Sampling).....	9
2.2	Αλγόριθμοι MCMC	10
2.2.1	Αλγόριθμος Metropolis-Hastings.....	12
2.2.2	Δειγματολήπτης Gibbs	13
2.2.3	Αλγόριθμος Metropolis Within Gibbs.....	15
3	Μελέτη του αλγόριθμου Metropolis - Hastings.....	17
3.1	Ιστορική αναδρομή	17
3.2	Περιγραφή του αλγόριθμου MH	20
3.3	Ειδικές μορφές του αλγορίθμου MH	24
3.3.1	Ανεξάρτητος δειγματολήπτης MH (Independent Sampler).....	24
3.3.2	Τυχαίος περίπατος MH (Random Walk MH).....	25
3.4	Παραδείγματα.....	26
3.4.1	Κανονικές κατανομές	26
3.4.2	Δισδιάστατες κατανομές.....	30
3.4.3	Μίξη κανονικών κατανομών	35
3.4.4	Σύνολο δεδομένων dugongs	37
4	Εκτίμηση Διασποράς.....	47
4.1	Μέθοδοι Εκτίμησης Διασποράς	47
4.1.1	Spectral Estimation	48
4.1.2	Batch means (BM)	51
4.1.3	Overlapping batch means (OBM)	51
4.2	Παράδειγμα	52

5 Προσδιορισμός κατάλληλης κατανομής πρότασης στον αλγόριθμο Metropolis-Hastings.....	55
5.1 Ρυθμός αποδοχής (Acceptance rate).....	55
5.1.1 Μονοδιάστατη περίπτωση.....	57
5.1.2 Πολυδιάστατη περίπτωση.....	58
5.2 Adaptive MCMC	62
5.2.1 Αλγόριθμος Adaptive Metropolis (AM).....	62
5.2.2 Adaptive Metropolis-Within-Gibbs	65
5.2.3 Single Component adaptive Metropolis algorithm (SCAM)	68
5.2.4 Robbins-Monro.....	70
6 Σύνοψη – Συμπεράσματα	75
ΠΑΡΑΡΤΗΜΑ	77
Βιβλιογραφία	87

Πίνακας Σχημάτων

<p>Σχήμα 3.1 500 επαναλήψεις από τον αλγόριθμο MH με κατανομή στόχο $\mathcal{N}(0,1)$ και κατανομή πρότασης (a) $g(\cdot X) = \mathcal{N}(X, 0.1^2)$, (b) $g(\cdot X) = \mathcal{N}(X, 0.5^2)$, (c) $g(\cdot X) = \mathcal{N}(X, 2^2)$ και (d) $g(\cdot X) = \mathcal{N}(X, 10^2)$.....</p>	27
<p>Σχήμα 3.2 Σύγκλιση εργοδικών μέσων πρώτης ροπής που προκύπτουν από τον αλγόριθμο MH με κατανομή-στόχο $\mathcal{N}(0,1)$ και κατανομή πρότασης (a) $g(\cdot X) = \mathcal{N}(X, 0.1^2)$, (b) $g(\cdot X) = \mathcal{N}(X, 0.5^2)$, (c) $g(\cdot X) = \mathcal{N}(X, 2^2)$ και (d) $g(\cdot X) = \mathcal{N}(X, 10^2)$.....</p>	28
<p>Σχήμα 3.3 Σύγκλιση εργοδικών μέσων δεύτερης ροπής που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο $\mathcal{N}(0,1)$ και κατανομή πρότασης (a) $g(\cdot X) = \mathcal{N}(X, 0.1^2)$, (b) $g(\cdot X) = \mathcal{N}(X, 0.5^2)$, (c) $g(\cdot X) = \mathcal{N}(X, 2^2)$ και (d) $g(\cdot X) = \mathcal{N}(X, 10^2)$.....</p>	29
<p>Σχήμα 3.4 Σύγκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH (—) με κατανομή στόχο $\pi(x, y) = \exp\{-(x^2 - \sqrt{2}xy + y^2)\}$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ και από το Δειγματολήπτη Gibbs (- - -).....</p>	32
<p>Σχήμα 3.5 Σύγκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH (—) με κατανομή στόχο $\pi(x, y) \propto y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I_{(0,\infty)}(y)$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ με $y > 0$, από τον MH (.....) με κατανομή πρότασης $g(x, y) \sim y^{\frac{3}{2}} e^{-2x - \frac{x^2}{2}}$ και από τον αντίστοιχο Δειγματολήπτη Gibbs (- - -).....</p>	34
<p>Σχήμα 3.6 Σύγκλιση εργοδικών μέσων των παραμέτρων α, β και γ που προκύπτουν από την πρώτη μέθοδο (—), τη δεύτερη μέθοδο (—) και την τρίτη μέθοδο (.....).....</p>	44

Πίνακας Πινάκων

Πίνακας 3.1 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο $\pi(x, y) = \exp\{-(x^2 - \sqrt{2}xy + y^2)\}$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ και από το αντίστοιχο Δειγματολήπτη Gibbs.	32
Πίνακας 3.2 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο $\pi(x, y) \propto y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I_{(0, \infty)}(y)$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$, από τον MH με κατανομή πρότασης $g(x, y) \sim y^{\frac{3}{2}} e^{-2x - \frac{x^2}{2}}$ και από τον αντίστοιχο Δειγματολήπτη Gibbs.	35
Πίνακας 3.3 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH, τη μέθοδο Αποδοχής-Απόρριψης και το Δειγματολήπτη Σπουδαιότητας με κατανομή στόχο $\pi(x) \sim \frac{1}{3} \mathcal{N}(0, 9) + \frac{1}{3} \mathcal{N}(5, 1) + \frac{1}{3} \mathcal{N}(15, 4)$ και κατανομή πρότασης $g(x, y) \sim \text{Cauchy}(0, 10)$	36
Πίνακας 3.4 Μετρήσεις της ηλικίας και του μήκους των 27 θαλάσσιων ελεφάντων που συλλέχθηκαν κοντά στο Townsville του Queensland.	37
Πίνακας 3.5 Μέση τιμή και τυπικά σφάλματα των εργοδικών μέσων των παραμέτρων α, β και γ για το πρόβλημα Dugongs.	43
Πίνακας 4.1 Εκτίμηση ασυμπτωτικής διασποράς εργοδικών μέσων πρώτης και δεύτερης ροπής από ένα δείγμα 10000 παρατηρήσεων που προσομοιώθηκε μέσω ενός αλγόριθμου MH με κατανομή-στόχο $\pi \sim \mathcal{N}(0, 1)$ και κατανομή πρότασης $g \sim \mathcal{N}(0, \sigma^2)$	53
Πίνακας 5.1 Μέσος ρυθμός αποδοχής, μέση εκτίμηση και τυπική απόκλιση της πρώτης και δεύτερης ροπής.	58
Πίνακας 5.2 Μέσος ρυθμός αποδοχής, μέσες εκτιμήσεις και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα.	59
Πίνακας 5.3 Μέσος ρυθμός αποδοχής, μέση εκτίμηση εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον AM αλγόριθμο στο μονοδιάστατο χώρο καταστάσεων.	64
Πίνακας 5.4 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον AM αλγόριθμο στο 10-διάστατο χώρο.	65

Πίνακας 5.5 Μέσος ρυθμός αποδοχής, μέση εκτίμηση εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον Adaptive Metropolis-within-Gibbs αλγόριθμο στο μονοδιάστατο χώρο καταστάσεων.....	66
Πίνακας 5.6 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον Adaptive Metropolis-within-Gibbs αλγόριθμο στον 10-διάστατο χώρο καταστάσεων.	67
Πίνακας 5.7 Μέσος ρυθμός αποδοχής, μέση τιμή εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον αλγόριθμο SCAM στο μονοδιάστατο χώρο καταστάσεων.	69
Πίνακας 5.8 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον αλγόριθμο SCAM στο 10-διάστατο χώρο καταστάσεων.....	70
Πίνακας 5.9 Μέσος ρυθμός αποδοχής, μέση τιμή εργοδικών μέσων και τυπική απόκλιση της πρώτης και της δεύτερης ροπής εφαρμόζοντας τη διαδικασία των Robbins-Monro στο μονοδιάστατο χώρο.....	71
Πίνακας 5.10 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τη διαδικασία των Robbins-Monro στον 10-διάστατο χώρο.....	73

1 Εισαγωγή

Τα σημαντικότερα προβλήματα που αντιμετωπίζει η Υπολογιστική Στατιστική είναι η προσομοίωση παρατηρήσεων από κάποια κατανομή $\pi(\cdot)$ (κατανομή-στόχο) και ο υπολογισμός ολοκληρωμάτων της μορφής

$$E_{\pi}(h) = \int_{\mathcal{X}} h(x)\pi(dx). \quad (1.1)$$

Σε πολλά πραγματικά προβλήματα η προσομοίωση από την π δεν μπορεί να γίνει με απλές τεχνικές. Επίσης πολύ συχνά ο υπολογισμός ολοκληρωμάτων της μορφής (1.1) είναι πολύ δύσκολο αν όχι αδύνατο να γίνει αναλυτικά. Κύριοι παράγοντες που δυσκολεύουν τον ακριβή υπολογισμό τους είναι η σύνθετη μορφή της υπό ολοκλήρωση ποσότητας ή/και η διάσταση του χώρου καταστάσεων. Σε αυτές τις περιπτώσεις καταφεύγουμε σε τεχνικές Monte Carlo (MC) και Markov Chain Monte Carlo (MCMC), οι οποίες προσομοιώνουν τιμές τυχαίων μεταβλητών και εκτιμούν τα ολοκληρώματα της μορφής (1.1) μέσω κατάλληλων συναρτήσεων των προσομοιωμένων τιμών.

Οι τεχνικές MC παράγουν ανεξάρτητες παρατηρήσεις είτε απ' ευθείας από την κατανομή-στόχο π είτε από κάποια διαφορετική κατανομή πρότασης g της οποίας η προσομοίωση είναι πολύ πιο απλή διαδικασία. Οι τεχνικές MCMC προσομοιώνουν αλυσίδες Markov με στάσιμη κατανομή την κατανομή-στόχο π και επομένως οι παρατηρήσεις είναι εξαρτημένες.

Στόχος της εργασίας αυτής είναι η παρουσίαση βασικών MC και MCMC μεθόδων και η διεξοδική μελέτη του αλγόριθμου Metropolis-Hastings (MH) που είναι ο σημαντικότερος MCMC αλγόριθμος.

Τα κεφάλαια της εργασίας έχουν οργανωθεί ως εξής:

Στο Κεφάλαιο 2 γίνεται μια σύντομη αναφορά σε γνωστές τεχνικές MC, όπως η μέθοδος Αποδοχής-Απόρριψης, η μέθοδος Αντιστροφής και η μέθοδος Δειγματοληψίας Σπουδαιότητας καθώς επίσης και σε τεχνικές MCMC, όπως ο αλγόριθμος MH, ο Δειγματολήπτης Gibbs και η μέθοδος Metropolis Within Gibbs.

Στο Κεφάλαιο 3 γίνεται αναλυτική αναφορά στον αλγόριθμο MH. Αρχικά, παραθέτουμε μια σύντομη ιστορική αναδρομή και στη συνέχεια δίνουμε την αναλυτική περιγραφή του. Παρουσιάζουμε κάποιες ειδικές μορφές του, καθώς και τις βασικές ιδιότητες που το χαρακτηρίζουν. Το κεφάλαιο ολοκληρώνεται με την παρουσίαση κάποιων εφαρμογών σε προσομοιωμένα καθώς και σε πραγματικά δεδομένα.

Το τέταρτο κεφάλαιο ασχολείται με μεθόδους εκτίμησης της διασποράς του εργοδικού μέσου ο οποίος προκύπτει από τις MCMC τεχνικές. Ιδιαίτερη αναφορά γίνεται στις μεθόδους Batch Means και Spectral Variance Estimators.

Τέλος, το Κεφάλαιο 5 ασχολείται με την εύρεση μιας κατάλληλης κατανομής πρότασης για τον αλγόριθμό MH. Παρόλο που ο αλγόριθμος MH μπορεί να συγκλίνει για οποιαδήποτε κατανομή πρότασης αρκεί να ικανοποιεί κάποιες βασικές υποθέσεις, είναι γνωστό ότι μία κατάλληλη επιλογή της κατανομής πρότασης βελτιώνει τη σύγκλιση του αλγόριθμου. Ο προσδιορισμός της βέλτιστης κατανομής πρότασης για μια συγκεκριμένη κατανομή-στόχο είναι ένα πολύ σημαντικό αλλά εξίσου δύσκολο πρόβλημα. Το πρόβλημα αυτό έχει προσεγγιστεί με απλοϊκές τεχνικές (trial-and-error τεχνικές) αλλά και με σύνθετους προσαρμοστικούς αλγόριθμους (adaptive algorithms) που βρίσκουν μια "καλή" κατανομή πρότασης αυτόματα.

Στο τελευταίο κεφάλαιο γίνεται μια σύντομη σύνοψη της εργασίας και παραθέτουμε τα βασικότερα συμπεράσματά μας.

Τέλος η εργασία αυτή ολοκληρώνεται με ένα παράρτημα στο οποίο δίνεται η βασική θεωρία των αλυσίδων Markov που χρειάζεται σε όλα τα προηγούμενα κεφάλαια.

2 Μέθοδοι Προσομοίωσης

2.1 Μέθοδοι Monte Carlo

Οι μέθοδοι MC είναι τεχνικές που προσομοιώνουν τη συμπεριφορά διαφόρων Φυσικών και Μαθηματικών συστημάτων κυρίως μέσω ηλεκτρονικού υπολογιστή. Η αρχή για τη μελέτη των μεθόδων αυτών έγινε από τον Πολωνό μαθηματικό Stanislaw Ulam το 1946 με αφορμή ένα παιχνίδι πόκερ στο καζίνο Monte Carlo. Για πρώτη φορά οι τεχνικές MC χρησιμοποιήθηκαν από τους Enrico Fermi, Stanislaw Ulam, John Von Neumann και Nicholas Metropolis τη δεκαετία του '40, σε εργασία τους για την εκτίμηση των ιδιοτήτων των νετρονίων, η οποία χρησιμοποιήθηκε αργότερα για την κατασκευή της βόμβας υδρογόνου. Οι μέθοδοι MC έπαιξαν καθοριστικό ρόλο στο περίφημο Manhattan project, που αφορούσε την κατασκευή των πρώτων πυρηνικών όπλων στο εθνικό εργοστάσιο του Los Alamos.

Πιο συγκεκριμένα λοιπόν, με τον όρο «Μέθοδος Monte Carlo» αναφερόμαστε σε κάθε τεχνική η οποία προσομοιώνει ανεξάρτητες παρατηρήσεις X_1, X_2, \dots, X_n από κάποια κατανομή και χρησιμοποιώντας κατάλληλους δειγματικούς μέσους προσεγγίζει τις μαθηματικές ελπίδες που μας ενδιαφέρουν.

Στη συνέχεια θα αναφερθούμε στις σημαντικότερες MC τεχνικές που είναι η ολοκλήρωση MC, η μέθοδος Αντιστροφής, η μέθοδος Αποδοχής-Απόρριψης και ο Δειγματολήπτης Σπυδαιότητας.

2.1.1 Ολοκλήρωση Monte Carlo

Θεωρούμε τη συνάρτηση $f(x): \mathbb{R}^d \rightarrow \mathbb{R}$, η οποία είναι ολοκληρώσιμη στο πεδίο ορισμού της. Θέλουμε να υπολογίσουμε το ολοκλήρωμα

$$I = \int_{\mathcal{X}} f(x) dx$$

όπου $\mathcal{X} \subseteq \mathbb{R}^d$, $d \geq 1$. Το παραπάνω ολοκλήρωμα μπορεί να γραφεί

$$I = \int_{\mathcal{X}} f(x) dx = \int_{\mathcal{X}} \frac{f(x)}{\pi(x)} \pi(x) dx = \int_{\mathcal{X}} h(x) \pi(x) dx = E_{\pi}(h)$$

όπου $h(x) = f(x)/\pi(x)$ και $\pi(x)$ οποιαδήποτε συνάρτηση πυκνότητας πιθανότητας με $\pi(x) \neq 0, \forall x \in \mathcal{X}$. Έτσι το ζητούμενο ολοκλήρωμα δίνεται πλέον ως μια μέση τιμή της τυχαίας μεταβλητής $h(x)$, όπου x ορίζεται στο διάστημα ολοκλήρωσης \mathcal{X} και έχει συνάρτηση πυκνότητας πιθανότητας την $\pi(x)$. Όταν η $E_{\pi}(h)$ υπάρχει και είναι πεπερασμένη θα γράφουμε $h \in \mathcal{L}(\pi)$.

Παίρνοντας ένα τυχαίο δείγμα (X_1, \dots, X_n) από την $\pi(x)$, μπορούμε να προσεγγίσουμε την $E_{\pi}(h)$ με το δειγματικό μέσο

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^n h(X_i).$$

Από τον Ισχυρό Νόμο των Μεγάλων αριθμών (INMA) έχουμε ότι καθώς το $n \rightarrow \infty$ ο δειγματικός μέσος \bar{h}_n συγκλίνει σχεδόν βεβαίως στην $E_{\pi}(h)$. Επίσης, όταν $\sigma^2 = \text{Var}_{\pi}(h) < \infty$, ισχύει το Κεντρικό Οριακό Θεώρημα (ΚΟΘ), δηλαδή έχουμε

$$\frac{\sqrt{n}\{\bar{h}_n - E_{\pi}(h)\}}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \text{ καθώς } n \rightarrow \infty,$$

όπου $\mathcal{N}(\mu, \sigma^2)$ συμβολίζει την κανονική κατανομή με μέση τιμή μ και διασπορά σ^2 . Επίσης πρέπει να παρατηρήσουμε ότι ο \bar{h}_n είναι αμερόληπτος εκτιμητής του $E_\pi(h)$:

$$E_\pi(\bar{h}_n(X)) = E\left(\frac{1}{n} \sum_{i=1}^n h(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(h(X_i)) = E_\pi(h).$$

Στην περίπτωση που το σ δεν είναι γνωστό, μπορούμε να χρησιμοποιήσουμε ένα συνεπή εκτιμητή του, όπως ο

$$\hat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{i=1}^n \{h(X_i) - \bar{h}_n\}^2}.$$

Τότε, λόγω του Θεωρήματος Slutsky, το ΚΟΘ εξακολουθεί να ισχύει, οπότε έχουμε

$$\frac{\sqrt{n}\{\bar{h}_n - E_\pi(h)\}}{\hat{\sigma}_n} \xrightarrow{d} \mathcal{N}(0,1) \text{ καθώς } n \rightarrow \infty.$$

Επομένως, η $E_\pi(h)$ μπορεί να εκτιμηθεί με ένα $100(1 - \alpha)\%$ διάστημα εμπιστοσύνης

$$\left[\bar{h}_n - z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{h}_n + z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \right].$$

Παρατηρούμε ότι η ακρίβεια της εκτίμησης της $E_\pi(h)$ εξαρτάται από την ασυμπτωτική διασπορά σ^2 , με αποτέλεσμα όσο μικρότερη ασυμπτωτική διασπορά έχει ένας εκτιμητής της $E_\pi(h)$ τόσο αποδοτικότερος να είναι.

2.1.2 Μέθοδος Αντιστροφής

Μια από τις πιο απλές μεθόδους προσομοίωσης τυχαίων μεταβλητών από μια κατανομή με αθροιστική συνάρτηση κατανομής $F(x) = P(X \leq x)$ βασίζεται στην αντιστροφή της $F(x)$. Η μέθοδος της αντιστροφής στηρίζεται στην προσομοίωση παρατηρήσεων από την ομοιόμορφη κατανομή $\mathcal{U}(0, 1)$ και στον κατάλληλο μετασχηματισμό τους έτσι ώστε οι μετασχηματισμένες πλέον παρατηρήσεις να προέρχονται από την επιθυμητή κατανομή.

Η αθροιστική συνάρτηση κατανομής είναι μια αύξουσα συνάρτηση αλλά όχι αναγκαία και συνεχής. Έτσι ορίζουμε τη γενικευμένη αντίστροφη συνάρτηση

$$F^-(u) = \inf \{x: F(x) \geq u\}.$$

Εάν η F είναι συνεχής τότε ισχύει ότι $F^-(u) = F^{-1}(u)$, δηλαδή η γενικευμένη αντίστροφη συνάρτηση F^- ταυτίζεται με την αντίστροφη συνάρτηση F^{-1} .

Θεώρημα: Θεωρούμε $U \sim \mathcal{U}(0, 1)$ και μια αθροιστική συνάρτηση κατανομής $F(x)$. Τότε η $X = F^-(U)$ έχει αθροιστική συνάρτηση κατανομής F .

Απόδειξη: Έστω $u \in (0, 1)$. Επειδή η $F(\cdot)$ είναι δεξιά συνεχής, ισχύει $F(F^-(u)) \geq u$. Επίσης από τον ορισμό της γενικευμένης αντίστροφης συνάρτησης, για κάθε $x \in \mathbb{R}$ ισχύει $F^-(F(x)) \leq x$. Επομένως,

$$\{(u, x): F^-(u) \leq x\} = \{(u, x): F(x) \geq u\},$$

άρα

$$P(X \leq x) = P\{F^-(U) \leq x\} = P\{F(F^-(U)) \leq F(x)\} = P\{U \leq F(x)\} = F(x).$$

Σε πολλές περιπτώσεις όμως η εύρεση κατάλληλων μετασχηματισμών είναι μια ιδιαίτερα δύσκολη, αν όχι αδύνατη διαδικασία, για αυτό καταφεύγουμε σε άλλες τεχνικές όπως είναι η μέθοδος της Αποδοχής-Απόρριψης και ο Δειγματολήπτης Σπουδαιότητας που θα περιγράψουμε στις επόμενες ενότητες.

2.1.3 Μέθοδος Αποδοχής-Απόρριψης

Μια άλλη μέθοδος προσομοίωσης τυχαίου δείγματος από μια κατανομή $\pi(\cdot)$, που προτάθηκε αρχικά από τον Von Neumann (1951), είναι η μέθοδος Αποδοχής-Απόρριψης (Accept-Reject method). Για την υλοποίηση της μεθόδου, χρειάζεται μόνο η γνώση της συναρτησιακής μορφής της κατανομής-στόχου $\pi(\cdot)$ καθώς η προσομοίωση του δείγματος γίνεται με τη βοήθεια κάποιας άλλης κατανομής πρότασης $g(\cdot)$.

Έστω ότι θέλουμε να παράγουμε δείγμα από μια κατανομή-στόχου με πυκνότητα πιθανότητα $\pi(\cdot)$ ορισμένη σε κάποιο σύνολο \mathcal{X} . Έστω επίσης μια κατανομή με πυκνότητα $g(\cdot)$ με πεδίο ορισμού τουλάχιστον το \mathcal{X} από την οποία μπορούμε να προσομοιώσουμε παρατηρήσεις πιο εύκολα. Βρίσκουμε $M > 0$ τέτοιο ώστε να ισχύει

$$\pi(x) \leq Mg(x), \quad \forall x \in \mathcal{X}.$$

Η μέθοδος Αποδοχής-Απόρριψης παράγει μια παρατήρηση x από τη $g(x)$ η οποία γίνεται αποδεκτή με πιθανότητα $\pi(x)/Mg(x)$, διαφορετικά απορρίπτεται και προσομοιώνεται νέα παρατήρηση x από τη $g(x)$. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να γίνει κάποια παρατήρηση x δεκτή. Τότε, η x είναι μια παρατήρηση από την κατανομή-στόχο $\pi(x)$.

Πιο συγκεκριμένα, προσομοιώνουμε μια $u \sim \mathcal{U}(0,1)$ και $Y \sim g(\cdot)$. Αν $u \leq \pi(Y)/Mg(Y)$, τότε αποδεχόμαστε τη Y , διαφορετικά την απορρίπτουμε. Η διαδικασία επαναλαμβάνεται έως ότου κάποια παρατήρηση Y γίνει αποδεκτή. Η διαδικασία προσομοίωσης μιας παρατήρησης από την $\pi(\cdot)$ μέσω της μεθόδου Αποδοχής-Απόρριψης περιγράφεται στον ακόλουθο αλγόριθμο.

Αλγόριθμος Αποδοχής-Απόρριψης

1. Προσομοίωσε $Y \sim g(\cdot)$
 2. Προσομοίωσε $u \sim \mathcal{U}(0,1)$
 3. Αν $u \leq \pi(Y)/Mg(Y)$, επέστρεψε $Y \sim \pi$, διαφορετικά επέστρεψε στο Βήμα 1.
-

Η μεθοδος Αποδοχής-Απόρριψης δίνει ανεξάρτητες παρατηρήσεις από την κατανομή-στόχο π αφού για κάθε μετρήσιμο σύνολο $A \subseteq \mathcal{X}$ ισχύει

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

Πρόταση: Η πιθανότητα αποδοχής της Y στον αλγόριθμο Αποδοχής-Απόρριψης είναι ίση με $1/M$.

Απόδειξη:

$$\begin{aligned} P(Y \text{ γίνεται δεκτή}) &= P\left(u \leq \frac{\pi(Y)}{Mg(Y)}\right) = \int_{\mathcal{X}} \int_0^{\pi(y)/Mg(y)} g(y) du \mu(dy) \\ &= \int_{\mathcal{X}} \frac{\pi(y)}{Mg(y)} g(y) \mu(dy) = \frac{1}{M} \int_{\mathcal{X}} \pi(y) \mu(dy) = \frac{1}{M}. \end{aligned}$$

Οι προτεινόμενες παρατηρήσεις στη μέθοδο Αποδοχής-Απόρριψης παράγονται ανεξάρτητα και γίνονται αποδεκτές με πιθανότητα $p = 1/M$. Επομένως, το πλήθος N των προτεινόμενων τιμών Y μέχρι να έχουμε την πρώτη αποδοχή, ακολουθεί γεωμετρική κατανομή με πιθανότητα επιτυχίας $p = 1/M$. Επομένως, ο αναμενόμενος αριθμός προτεινόμενων τιμών μέχρι να γίνει η πρώτη αποδοχή ισούται με $E(N) = 1/p = M$.

Παρατηρούμε ότι αν η ανισότητα $\pi(x) \leq Mg(x)$ ισχύει $\forall x \in \mathcal{X}$, θα ισχύει και για οποιοδήποτε $M' > M$. Άρα η μέθοδος Αποδοχής-Απόρριψης μπορεί να εφαρμοστεί για οποιοδήποτε M' στη θέση του M , άρα αρκεί να βρούμε ένα άνω φράγμα του. Προφανώς, όσο μεγαλύτερο είναι το M τόσο περισσότερες τιμές απορρίπτονται.

2.1.4 Δειγματολήπτης Σπουδαιότητας (Importance Sampling)

Ο Δειγματολήπτης Σπουδαιότητας (Importance Sampling - IS) εισήχθει από τον Marshall (1956) και χρησιμοποιείται για την προσέγγιση ολοκληρωμάτων της μορφής (1.1) μέσω της προσομοίωσης παρατηρήσεων από κάποια διαφορετική κατανομή πρότασης $g(\cdot)$.

Έστω μια κατανομή πρότασης $g(x)$, με στήριγμα τουλάχιστον το σύνολο για το οποίο ισχύει $\pi(x) \cdot h(x) \neq 0$. Έχουμε ότι

$$E_{\pi}[h(X)] = \int_{\mathcal{X}} \pi(x)h(x) dx = \int_{\mathcal{X}} g(x) \frac{h(x)\pi(x)}{g(x)} dx = E_g \left[\frac{h(X)\pi(X)}{g(X)} \right] = E_g[w(X)h(X)]$$

όπου $w(X) = \pi(X)/g(X)$. Τα $w(X)$ ονομάζονται βάρη σπουδαιότητας (Importance weights) και η $g(x)$ κατανομή σπουδαιότητας ή κατανομή πρότασης.

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n από τη $g(x)$, εκτιμούμε την $E_{\pi}(h)$ με την ποσότητα

$$\bar{h}_n^{IS} = \frac{1}{n} \sum_{i=1}^n h(X_i)w(X_i).$$

Το γεγονός ότι ο εκτιμητής \bar{h}_n^{IS} συγκλίνει ισχυρά στη ζητούμενη αναμενόμενη τιμή $E_{\pi}[h(X)]$ μας το εξασφαλίζει και πάλι ο INMA

$$\frac{1}{n} \sum_{i=1}^n h(X_i)w(X_i) \xrightarrow{\sigma.β.} E_g[h(X)w(X)] = E_{\pi}[h(X)].$$

Ο \bar{h}_n^{IS} μπορεί να χρησιμοποιηθεί μόνο όταν γνωρίζουμε τη σταθερά κανονικοποίησης των $\pi(x)$ και $g(x)$.

Ένας πιο γενικός εκτιμητής της $E_{\pi}(h)$, τον οποίο μπορούμε να χρησιμοποιήσουμε ακόμα και στην περίπτωση που γνωρίζουμε μόνο τη συναρτησιακή μορφή των βαρών σπουδαιότητας είναι ο

$$\hat{h}_n^{IS} = \frac{\sum_{i=1}^n h(X_i)w(X_i)}{\sum_{i=1}^n w(X_i)}.$$

Διαιρώντας με n αριθμητή και παρανομαστή θα έχουμε

$$\hat{h}_n^{IS} = \frac{\sum_{i=1}^n h(X_i)w(X_i)/n}{\sum_{i=1}^n w(X_i)/n},$$

παρατηρούμε ότι ο αριθμητής συγκλίνει στην ποσότητα $E_\pi(h)$ ενώ ο παρονομαστής συγκλίνει στη μονάδα. Επομένως, ο λόγος τους συγκλίνει στην ποσότητα $E_\pi(h)$. Στην περίπτωση όπου γνωρίζουμε τα βάρη εκτός από την πολλαπλασιαστική σταθερά τους δηλ. $w(x) = c\tilde{w}(x)$ τότε

$$\hat{h}_n^{IS} = \frac{\sum_{i=1}^n h(X_i)\tilde{w}(X_i)}{\sum_{i=1}^n \tilde{w}(X_i)} = \frac{c^{-1} \sum_{i=1}^n h(X_i)w(X_i)/n}{c^{-1} \sum_{i=1}^n w(X_i)/n}.$$

Σε αυτή την περίπτωση ο αριθμητής συγκλίνει στην ποσότητα $c^{-1}E_\pi(h)$ και ο παρονομαστής στο c^{-1} οπότε και πάλι ο λόγος τους συγκλίνει στην ποσότητα που θέλουμε να εκτιμήσουμε.

Ο Δειγματολήπτης Σπουδαιότητας μπορεί να χρησιμοποιηθεί αντί για τη μέθοδο Αποδοχής-Απόρριψης εάν δεν είναι εύκολο να βρεθεί μία $g(x)$ τέτοια ώστε $\pi(x) \leq Mg(x) \forall x \in \mathcal{X}$ ή το M είναι πολύ μεγάλο με αποτέλεσμα να χρειάζεται να προσομοιώσουμε ένα μεγάλο αριθμό παρατηρήσεων από τη $g(x)$ προκειμένου να έχουμε μία παρατήρηση από την $\pi(x)$.

2.2 Αλγόριθμοι MCMC

Η ιδέα πίσω από τις MCMC τεχνικές είναι η κατασκευή μιας στοχαστικής διαδικασίας της οποίας η στάσιμη κατανομή είναι η κατανομή-στόχος π . Έτσι

προσομοιώνουμε παρατηρήσεις οι οποίες ακολουθούν προσεγγιστικά την κατανομή-στόχο π και μας επιτρέπουν την εκτίμηση κάποιων μαθηματικών ελπίδων.

Έστω X_0, X_1, X_2, \dots μια ακολουθία που προκύπτει από κάποια μέθοδο MCMC με κατανομή-στόχο $\pi(\cdot)$. Λόγω της μορφής των μεθόδων MCMC, οι παρατηρήσεις που παράγονται δεν είναι ανεξάρτητες και οι περιθωριακές κατανομές τους γενικά διαφέρουν από την $\pi(\cdot)$. Εντούτοις κάτω από συγκεκριμένες συνθήκες το Εργοδικό Θεώρημα εξασφαλίζει την ισχυρή σύγκλιση του εργοδικού μέσου

$$\bar{h}_n = \frac{1}{n} \sum_{i=0}^{n-1} h(X_i)$$

στην $E_\pi(h)$ για κάθε $h \in \mathcal{L}(\pi)$.

Συνήθως η περιθώρια κατανομή των πρώτων όρων της αλυσίδας δεν είναι πολύ κοντά στην κατανομή-στόχο $\pi(\cdot)$ με αποτέλεσμα ο \bar{h}_k ο οποίος βασίζεται στις k πρώτες παρατηρήσεις να μην είναι ιδιαίτερα καλός εκτιμητής της $E_\pi(h)$. Γνωρίζουμε όμως ότι, υπό κατάλληλες συνθήκες, η απόσταση της περιθώριας κατανομής του k -οστού όρου της αλυσίδας Markov και της $\pi(\cdot)$ φθίνει ως προς k . Έτσι, συχνά δε λαμβάνουμε υπόψη τους πρώτους k όρους της ακολουθίας και χρησιμοποιούμε τον εκτιμητή

$$\bar{h}_{BI} = \frac{1}{n-k} \sum_{i=k}^n h(X_i).$$

Η περίοδος που αντιστοιχεί στις παρατηρήσεις που δε λαμβάνουμε υπόψη ονομάζεται περίοδος *burn-in*. Η πιο προφανής και κοινά αποδεκτή μέθοδος προσδιορισμού της *burn-in* περιόδου είναι το γράφημα των τιμών του αλγορίθμου, όπου φαίνεται πότε η αλυσίδα έχει απομακρυνθεί αρκετά από τις αρχικές τιμές (βλέπε επίσης Cowles and Carlin (1996)).

Στην συνέχεια θα αναφερθούμε στις σημαντικότερες MCMC τεχνικές.

2.2.1 Αλγόριθμος Metropolis-Hastings

Η σημαντικότερη MCMC τεχνική είναι ο αλγόριθμος Metropolis-Hastings (MH) τον οποίο εισήγαγαν οι Metropolis et al. (1953) και το γενίκευσε ο Hastings (1970). Ο αλγόριθμος MH προσομοιώνει παρατηρήσεις από κάποια κατανομή πρότασης $g(\cdot)$, διαφορετική από την κατανομή-στόχο π , τις οποίες αποδέχεται με κάποια πιθανότητα, διαφορετικά επαναλαμβάνει την προηγούμενη παρατήρηση.

Πιο συγκεκριμένα, έστω τη χρονική στιγμή t είμαστε στην κατάσταση $X_t = x$, για την επόμενη κατάσταση X_{t+1} προσομοιώνουμε μια υποψήφια τιμή Y από την κατανομή πρότασης $g(\cdot)$. Η κατανομή πρότασης θα μπορούσε να εξαρτάται από την προηγούμενη κατάσταση δηλαδή να είναι της μορφής $g(\cdot | X_t)$. Η υποψήφια τιμή Y γίνεται αποδεκτή με πιθανότητα $\alpha(X_t, Y)$ όπου

$$\alpha(X_t, Y) = \min\left(1, \frac{\pi(Y)g(X_t|Y)}{\pi(X_t)g(Y|X_t)}\right).$$

Εάν η υποψήφια τιμή γίνει αποδεκτή τότε $X_{t+1} = Y$, διαφορετικά $X_{t+1} = X_t$.

Αλγόριθμος Metropolis-Hasting

1. Αρχικοποίηση X_0
 2. Προσομοίωση υποψήφιας τιμής $Y \sim g(\cdot | X_t)$
 3. Προσομοίωση $u \sim \mathcal{U}(0,1)$
 4. Εάν $u \leq \alpha(X_t, Y)$ τότε θέτουμε $X_{t+1} = Y$
Διαφορετικά θέτουμε $X_{t+1} = X_t$
 5. Θέτουμε $t = t + 1$ και επιστροφή στο βήμα 2
-

Η αλυσίδα Markov που παράγεται από τον αλγόριθμο MH έχει την $\pi(\cdot)$ ως στάσιμη κατανομή ανεξάρτητα από την επιλογή της κατανομής πρότασης $g(\cdot)$ αρκεί να ικανοποιούνται κάποιες βασικές συνθήκες. Εντούτοις, η επιλογή της κατανομής πρότασης $g(\cdot)$ είναι κρίσιμη γιατί επηρεάζει την ταχύτητα σύγκλισης του αλγορίθμου.

2.2.2 Δειγματολήπτης Gibbs

Ο Δειγματολήπτης Gibbs προσομοιώνει παρατηρήσεις από πολυδιάστατες κατανομές-στόχου μέσω των πλήρως δεσμευμένων κατανομών τους, οι οποίες πρέπει να έχουν γνωστή μορφή. Αυτό είναι ένα βασικό πλεονέκτημα του Δειγματολήπτη Gibbs καθώς το πρόβλημα προσομοίωσης παρατηρήσεων από μια κατανομή-στόχο μεγάλης διάστασης μετατρέπεται σε πρόβλημα προσομοίωσης παρατηρήσεων από κατανομές μικρότερης διάστασης.

Οι Geman and Geman (1984) εισήγαγαν το Δειγματολήπτη Gibbs στη στατιστική κοινότητα ύστερα από ένα άρθρο τους όπου εφάρμοσαν αυτή τη μέθοδο για την μπεϋζιανή μελέτη των τυχαίων πεδίων Gibbs, από όπου πήρε και το όνομα του. Ο Δειγματολήπτης Gibbs εφαρμόστηκε αρχικά σε ένα διακριτό πρόβλημα επεξεργασίας εικόνας το οποίο δεν ολοκληρώθηκε. Εντούτοις η υλοποίηση αυτή είχε μεγάλη επίδραση στους Green, Smith, Spiegelhalter και άλλους. Οι Gelfand and Smith (1990) ανέπτυξαν το Δειγματολήπτη Gibbs δίνοντας νέα ώθηση στις μπεϋζιανές μεθόδους.

Περιγραφή του αλγορίθμου:

Έστω η πολυδιάστατη κατανομή-στόχος $\pi(\cdot)$ με χώρο καταστάσεων $\mathcal{X} \subseteq \mathbb{R}^k$. Θέλουμε να προσομοιώσουμε μια παρατήρηση $X = (X_1, X_2, \dots, X_d)$ με $d \leq k$ από την π . Θέτουμε $X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_d)$ το διάνυσμα που περιέχει όλα τα στοιχεία του X εκτός από το X_i . Ο Δειγματολήπτης Gibbs προσομοιώνει τα X_i , $i = 1, \dots, d$ με τη βοήθεια των πλήρως δεσμευμένων κατανομών των X_i δοθέντος των X_{-i} , $i = 1, \dots, d$.

$$X_i | X_{-i} = x_{-i} \sim \pi_i(x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_d = x_d)$$

Οι π_i καλούνται πλήρως δεσμευμένες κατανομές (full conditional distributions). Στο Δειγματολήπτη Gibbs για τη μετάβαση της αλυσίδας από την κατάσταση X^{t-1} στην κατάσταση X^t γίνεται προσομοιώνοντας διαδοχικά X_1^t, \dots, X_d^t ως εξής :

 Αλγόριθμος Gibbs Sampling

1. Αρχικοποίηση $X^0 = (x_1^0, \dots, x_d^0)$
 2. Στο χρόνο t δοθέντος $X^{t-1} = (x_1^{t-1}, \dots, x_d^{t-1})$ προσομοίωσε

$$X_1^t \sim \pi_1(x_1 | x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1})$$

$$X_2^t \sim \pi_2(x_2 | x_1^t, x_3^{t-1}, \dots, x_d^{t-1})$$

$$X_3^t \sim \pi_3(x_3 | x_1^t, x_2^t, \dots, x_d^{t-1})$$

$$\vdots$$

$$X_d^t \sim \pi_d(x_d | x_1^t, x_2^t, \dots, x_{d-1}^t)$$
 3. Θέσε $t = t + 1$ και επέστρεψε στο βήμα 2.
-

Η ακολουθία που προκύπτει είναι προσομοίωση μιας αλυσίδας Markov καθώς για την προσομοίωση της X^t έχουν χρησιμοποιηθεί μόνο οι τιμές της X^{t-1} επομένως η κατάσταση X^t δεν εξαρτάται από προηγούμενες καταστάσεις της ακολουθίας παρά μόνο από αυτές της X^{t-1} . Ο πυρήνας μετάβασης της αλυσίδας είναι

$$K(X^{t-1}, X^t) = K\left((x_1^{t-1}, x_2^{t-1}, \dots, x_d^{t-1}), (x_1^t, x_2^t, \dots, x_d^t)\right) =$$

$$\pi_1(x_1^t | x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1}) \times \pi_2(x_2^t | x_1^t, x_3^{t-1}, \dots, x_d^{t-1}) \times \dots \times \pi_d(x_d^t | x_1^t, x_2^t, \dots, x_{d-1}^t)$$

και η στάσιμη κατανομή της αλυσίδας Markov με αυτόν τον πυρήνα μετάβασης είναι η π . Για να το δούμε αυτό ας υποθέσουμε ότι $X^{t-1} \sim \pi(\cdot)$. Τότε,

$$(X_1^t, X_2^{t-1}, \dots, X_d^{t-1}) \sim \int_{x_1^{t-1}} \pi_1(x_1^t | x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1}) \pi(x_1^{t-1}, x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1}) dx_1^{t-1}$$

$$= \pi_1(x_1^t | x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1}) \pi(x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1})$$

$$= \pi(x_1^t, x_2^{t-1}, \dots, x_d^{t-1})$$

όπου $\pi(x_2^{t-1}, x_3^{t-1}, \dots, x_d^{t-1})$ συμβολίζει την πυκνότητα της περιθωριακής κατανομής των $X_2^{t-1}, \dots, X_d^{t-1}$. Επομένως το τυχαίο διάνυσμα $(X_1^t, X_2^{t-1}, \dots, X_d^{t-1})$ ακολουθεί την κατανομή π . Με την ίδια διαδικασία μπορούμε να δείξουμε ότι το διάνυσμα

$$(X_1^t, \dots, X_i^t, X_{i+1}^{t-1}, \dots, X_d^{t-1}) \sim \pi(\cdot) \text{ για κάθε } i = 1, \dots, d - 1.$$

Ολοκληρώνοντας τέλος και ως προς x_d^{t-1} , καταλήγουμε στο ότι $X^t = (X_1^t, X_2^t, \dots, X_d^t) \sim \pi$, πράγμα που σημαίνει ότι η π είναι η στάσιμη κατανομή της αλυσίδας.

2.2.3 Αλγόριθμος Metropolis Within Gibbs

Σε πολλά προβλήματα οι πλήρως δεσμευμένες κατανομές δεν είναι διαθέσιμες σε κλειστή μορφή, για αυτό συχνά χρειάζεται να χρησιμοποιήσουμε έναν υβριδικό Δειγματολήπτη Gibbs όπου τουλάχιστον μια προσομοίωση από την αντίστοιχη πλήρως δεσμευμένη κατανομή έχει αντικατασταθεί από ένα βήμα Metropolis Hastings.

Έστω ότι θέλουμε να προσομοιώσουμε ένα δείγμα από την πολυδιάστατη κατανομή-στόχο $\pi(\cdot)$ και έστω $\pi_i(x_i | X_{-i} = x_{-i})$, $i = 1, \dots, d$ οι αντίστοιχες πλήρως δεσμευμένες κατανομές. Παρακάτω περιγράφουμε επιγραμματικά τα βήματα που εκτελεί ο αλγόριθμος Metropolis within Gibbs.

Αλγόριθμος Metropolis Within Gibbs

1. Αρχικοποίηση $X_0 = (x_1^0, x_2^0, \dots, x_d^0)$
2. Στο χρόνο t δοθέντος $X^{t-1} = (x_1^{t-1}, x_2^{t-1}, \dots, x_d^{t-1})$
3. Προσομοίωσε $y_i \sim g_i(y_i | x_1^t, x_2^t, \dots, x_{i-1}^t, x_i^{t-1}, x_{i+1}^{t-1}, \dots, x_d^{t-1})$
4. Προσομοίωσε $u \sim \mathcal{U}(0,1)$
5. Υπολογισμός πιθανότητας αποδοχής

$$\alpha(x_i^{t-1} = y_i) = \min \left\{ 1, \frac{\pi_i(y_i | x_1^t, x_2^t, \dots, x_{i-1}^t, x_{i+1}^{t-1}, \dots, x_d^{t-1})}{\pi_i(x_i^{t-1} | x_1^t, x_2^t, \dots, x_{i-1}^t, x_{i+1}^{t-1}, \dots, x_d^{t-1})} \times \frac{g_i(x_i^{t-1} | x_1^t, x_2^t, \dots, x_{i-1}^t, y_i, x_{i+1}^{t-1}, \dots, x_d^{t-1})}{g_i(y_i | x_1^t, x_2^t, \dots, x_{i-1}^t, x_i^{t-1}, x_{i+1}^{t-1}, \dots, x_d^{t-1})} \right\}$$

6. Εάν $u \leq \alpha(x_i^{t-1}, y_i)$ τότε $x_i^{t-1} = y_i$
 7. Διαφορετικά $x_i^t = x_i^{t-1}$
 8. Θέσε $t = t + 1$ και επέστρεψε στο Βήμα 2.
-

Ο Δειγματολήπτης Gibbs είναι ειδική περίπτωση του παραπάνω αλγόριθμου με

$$g_i(y_i|x_1^t, x_2^t, \dots, x_{i-1}^t, x_i^{t-1}, x_{i+1}^{t-1}, \dots, x_d^{t-1}) = \pi_i(y_i|x_1^t, x_2^t, \dots, x_{i-1}^t, x_{i+1}^{t-1}, \dots, x_d^{t-1})$$

και έχει πιθανότητα αποδοχής $\alpha(x_i^{t-1} = y_i) = 1$.

Στη βιβλιογραφία υπάρχουν και άλλοι αλγόριθμοι με τη βοήθεια των οποίων μπορούμε να προσομοιώσουμε παρατηρήσεις από κάποια κατανομή-στόχο π ή/και να εκτιμήσουμε μέσες τιμές της μορφής $E_\pi(h)$. Οι αλγόριθμοι αυτοί μπορεί να είναι παραλλαγές κάποιων τεχνικών που παρουσιάσαμε νωρίτερα αλλά και αλγόριθμοι με διαφορετική λογική προσομοίωσης από την κατανομή π . Οι μέθοδοι Population Monte Carlo (Carpe et al. 2004), Sequential Monte Carlo, Sampling Importance Resample (Rubin 1987) είναι κάποιες από αυτές τις τεχνικές.

Στο επόμενο κεφάλαιο θα παρουσιάσουμε αναλυτικά τον αλγόριθμο MH ο οποίος είναι ο σημαντικότερος MCMC αλγόριθμος.

3

Μελέτη του αλγόριθμου Metropolis - Hastings

3.1 Ιστορική αναδρομή

Πριν από τον 20^ο αιώνα, οι επιστήμονες ξεκίνησαν να εφαρμόζουν τις τεχνικές Monte Carlo σε πειράματα στα οποία χρησιμοποιούνταν τυχαίοι αριθμοί. Ο George Louis Leclerc Comte de Buffon (1777) εφάρμοσε μια νέα μέθοδο για την προσέγγιση της τιμής του π με τη βοήθεια μιας επαναληπτικής διαδικασίας. Πραγματοποίησε τυχαίες ρίψεις μιας βελόνας σε ένα πλέγμα παράλληλων ευθειών και κατέγραφε τον αριθμό των φορών που η βελόνα έπεφτε πάνω σε μια ευθεία (Liu 2001, p. vii). Όμως, οι τεχνικές Monte Carlo, επί της ουσίας, άρχισαν να χρησιμοποιούνται τη δεκαετία του '40 από επιστήμονες του εργαστηρίου Los Alamos στο Νέο Μεξικό όπου προέκυψε και η ιδέα για τη δημιουργία του αλγορίθμου MH. Ο Nicholas C. Metropolis μαζί με τους Richard Feynman και John von Neumann (1943) συνεργάστηκαν για την κατασκευή της πρώτης ατομικής βόμβας στα πλαίσια του Manhattan Project. Ενοχλημένοι από τους αργούς και δυσκίνητους ηλεκτρομηχανικούς υπολογισμούς για την επίλυση μαθηματικών εξισώσεων που περιέγραφαν την τυχαία συμπεριφορά των νετρονίων, ενδιαφέρθηκαν για τη δημιουργία γρήγορων ηλεκτρονικών αριθμομηχανών. Μετά τον πόλεμο το 1948, ο Metropolis μαζί με μια πλειάδα επιστημόνων σχεδίασαν τον πρώτο υπερυπολογιστή που τον ονόμασαν MANIAC. Στο Los Alamos, οι Stanislaw Ulam and John von Neumann είχαν την ιδέα για την εκτέλεση υπολογισμών μέσω προσομοίωσης και ο Metropolis επινόησε το ελκυστικό όνομα "Monte Carlo τεχνικές" (Liu 2001, p. viii). Παρακινούμενοι από προβλήματα φυσικής οι Metropolis και Ulam (1949) εισήγαγαν την ιδέα τους στη

στατιστική κοινότητα με το άρθρο “The Monte Carlo Method”. Το 1953 ήταν μια χρονιά ορόσημο καθώς οι Metropolis, Marshall and Arianna Rosenblut, Edward and Augusta “Mici” Teller εισήγαγαν τον αλγόριθμο Metropolis ως μια μέθοδο βελτιστοποίησης σε ένα διακριτό χώρο καταστάσεων. Αν και το άρθρο αυτό δημοσιεύτηκε το 1953 στο περιοδικό Journal of Chemical Physics δεν έλαβε την απαιτούμενη προσοχή από τους στατιστικούς λόγω της οπτικής αλλά και του τρόπου παρουσίασης του. Πρώτοι, οι Hammersley and Handscomb (1964) στο βιβλίο τους “Monte Carlo Methods” παρουσιάζουν τον αλγόριθμο Metropolis ως μια τεχνική υπολογισμού ολοκληρωμάτων, αλλά όχι σαν μια γενική μέθοδο προσομοίωσης από κάποια κατανομή-στόχου.

Ο Barker (1965) δημοσίευσε ένα άρθρο που παρουσίαζε μια ανταγωνιστική μέθοδο προσομοίωσης από την κατανομή-στόχο $\pi(\cdot)$, με ελάχιστα διαφορετικό αλγόριθμο σε σχέση με αυτόν των Metropolis et al (1953). Θεώρησε ως κατανομή πρότασης τη $g(y, x)$ και η πιθανότητα να γίνει αποδεκτή η τιμή y ήταν

$$\alpha(x, y) = \frac{1}{1 + \frac{\pi(x)}{\pi(y)}}.$$

Πάλι όμως, δημοσιεύτηκε χρησιμοποιώντας ορολογία της φυσικής επιστήμης πράγμα που δυσκόλευε την κατανόηση του αλγόριθμου από τη στατιστική κοινότητα.

Ο αλγόριθμος Metropolis γενικεύτηκε, βελτιώθηκε και επαναδιατυπώθηκε από τον W.Keith Hastings (1970). Ο Hastings είδε τον αλγόριθμο των Metropolis et. al (1953) σα μια γενική διαδικασία προσομοίωσης από μια κατανομή-στόχο $\pi(\cdot)$. Το άρθρο του γράφτηκε υπό μια διαφορετική πιο στατιστική οπτική, επισημαίνοντας ότι στην πραγματικότητα η τεχνική Metropolis περιελάμβανε τον πίνακα μετάβασης μιας αλυσίδας Markov. Παρουσίασε την κατανομή-στόχο $\pi(\cdot)$ ως στάσιμη κατανομή της αλυσίδας Markov και γενίκευσε την κατανομή πρότασης σε μη συμμετρικές κατανομές αντίθετα με τους Metropolis και Barker που είχαν προτείνει συμμετρικές κατανομές πρότασης. Έτσι οι μέθοδοι του Metropolis και του Barker προκύπτουν ως ειδικές περιπτώσεις του αλγόριθμου του Hastings. Ο Hastings όρισε ως πιθανότητα αποδοχής για την υποψήφια τιμή y την

$$\alpha(x, y) = \frac{s(x, y)}{1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)}}, \quad (3.1)$$

όπου $s(x, y)$ μια οποιαδήποτε συμμετρική συνάρτηση, επιλεγμένη έτσι ώστε $0 \leq \alpha(x, y) \leq 1$ για όλα τα x και y . Ανάλογα με τη μορφή που παίρνει η $s(x, y)$ μπορούμε να καταλήξουμε σε γνωστούς αλγόριθμους. Πιο συγκεκριμένα για

$$s(x, y) = \begin{cases} 1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)} & \text{αν } \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} \geq 1 \\ 1 + \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} & \text{διαφορετικά} \end{cases} \quad (3.2)$$

καταλήγουμε στον αλγόριθμο ΜΗ. Αντικαθιστώντας την (3.2) στην (3.1) έχουμε

$$\begin{aligned} \alpha(x, y) &= \begin{cases} \frac{1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)}}{1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)}} & \text{όταν } \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} \geq 1 \\ \frac{1 + \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)}}{1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)}} & \text{διαφορετικά} \end{cases} \\ &= \begin{cases} 1 & \text{όταν } \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} \geq 1 \\ \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} & \text{διαφορετικά} \end{cases} = \min \left\{ 1, \frac{\pi(y)g(y, x)}{\pi(x)g(x, y)} \right\}. \end{aligned}$$

Όταν έχουμε $g(x, y) = g(y, x)$ τότε έχουμε τη μέθοδο που εισήγαγαν οι Metropolis et al. (1953) όπου

$$\alpha(x, y)^{(M)} = \begin{cases} 1 & \text{αν } \pi(y)/\pi(x) \geq 1 \\ \pi(y)/\pi(x) & \text{διαφορετικά.} \end{cases}$$

Από την άλλη αν $s(x, y) = 1$ έχουμε τον αλγόριθμο του Barker και αντίστοιχα θα έχουμε

$$\alpha(x, y)^{(B)} = \frac{1}{1 + \frac{\pi(x)g(x, y)}{\pi(y)g(y, x)}} = \frac{\pi(y)g(y, x)}{\pi(y)g(y, x) + \pi(x)g(x, y)}.$$

όπου λόγω συμμετρίας της $g(\cdot)$ παίρνει τη μορφή

$$\alpha(x, y)^{(B)} = \frac{1}{1 + \frac{\pi(x)}{\pi(y)}} = \frac{\pi(y)}{\pi(y) + \pi(x)}$$

Ο P.H. Peskun (1973) απέδειξε σε ένα άρθρο του στο περιοδικό *Biometrika*, ότι ο αλγόριθμος MH είναι βέλτιστος. Παρόλα αυτά, ο αλγόριθμος MH δεν ήταν ιδιαίτερα διαδεδομένος στη στατιστική κοινότητα μέχρι την δεκαετία του '90. Τότε, οι Gelfand and Smith (1990) με ένα άρθρο τους που βασιζόταν στο άρθρο των Geman and Geman (1984) για το δειγματολήπτη Gibbs έδωσαν νέα ώθηση στις μεθόδους προσομοίωσης. Επίσης, μια παρουσίαση του Luke Tierney σε συνέδριο στο Ohio State University το 1991 καθώς και η αυξανόμενη υπολογιστική ισχύ έκαναν τον αλγόριθμο MH δημοφιλή και έναν από τους σημαντικότερους αλγόριθμους της υπολογιστικής στατιστικής.

3.2 Περιγραφή του αλγόριθμου MH

Η βασική ιδέα του αλγόριθμου MH είναι η προσομοίωση μιας αλυσίδας Markov με στάσιμη κατανομή, την κατανομή-στόχο.

Έστω $\pi(\cdot)$ η κατανομή-στόχος με στήριγμα \mathcal{X} και $g(\cdot | \cdot)$ η κατανομή πρότασης μας. Επιλέγουμε τυχαία ένα αρχικό σημείο $X_0 \in \mathcal{X}$ τέτοιο ώστε $\pi(X_0) > 0$. Σε κάθε χρονική στιγμή t η επόμενη κατάσταση X_{t+1} επιλέγεται με τη βοήθεια μιας υποψήφιας τιμής Y που προέρχεται από την κατανομή πρότασης $g(\cdot | X_t)$.

Η Y γίνεται αποδεκτή με πιθανότητα

$$\alpha(X_t, Y) = \min\left(1, \frac{\pi(Y)g(X_t|Y)}{\pi(X_t)g(Y|X_t)}\right)$$

διαφορετικά απορρίπτεται και θέτουμε $X_{t+1} = X_t$. Ο αλγόριθμος τερματίζεται όταν συμπληρωθεί ένας προκαθορισμένος αριθμός επαναλήψεων.

Αλγόριθμος Metropolis Hastings

1. Αρχικοποίηση X_0
 2. Προσομοίωσε $Y \sim g(\cdot | X_t)$
 3. Προσομοίωσε $u \sim \mathcal{U}(0,1)$
 4. Εάν $u \leq \alpha(X_t, Y)$ τότε θέσε $X_{t+1} = Y$
Διαφορετικά $X_{t+1} = X_t$
 5. Θέσε $t = t + 1$ και επέστρεψε στο Βήμα 2.
-

Από τον αλγόριθμο MH προκύπτει μια αλυσίδα Markov καθώς η τιμή που παίρνει τελικά η X_{t+1} εξαρτάται μόνο από την τιμή της X_t και όχι από τις προηγούμενες τιμές X_{t-1}, X_{t-2}, \dots .

Ο πυρήνας μετάβασης της αλυσίδας είναι :

$$K(X_{t+1}|X_t) = g(X_{t+1}|X_t)\alpha(X_t, X_{t+1}) + I(X_{t+1} = X_t)[1 - \int g(Y|X_t)\alpha(X_t, Y)dY]$$

όπου $I(\cdot)$ η δείκτρια συνάρτηση που παίρνει την τιμή 1 όταν $X_{t+1} = X_t$ και 0 διαφορετικά. Έτσι ο πρώτος όρος προκύπτει από την αποδοχή της Y ενώ ο δεύτερος από την απόρριψη όλων των πιθανών υποψήφιων τιμών Y .

Για να αποδείξουμε ότι ο αλγόριθμος MH προσομοιώνει μια αλυσίδα Markov με στάσιμη κατανομή την κατανομή-στόχο $\pi(x)$ αρκεί να δείξουμε ότι ισχύει

$$\int \pi(X_t)K(X_t|X_{t+1})dX_t = \pi(X_{t+1}).$$

Ισοδύναμα μπορούμε να δείξουμε ότι ισχύει η λεπτομερής συνθήκη ισορροπίας (detailed balance condition)

$$\pi(X_t)K(X_t|X_{t+1}) = \pi(X_{t+1})K(X_{t+1}|X_t)$$

Πράγματι,

$$\begin{aligned} \pi(X_t)K(X_{t+1}|X_t) &= \\ &= \pi(X_t) \left\{ g(X_{t+1}|X_t)\alpha(X_t, X_{t+1}) + I(X_{t+1} = X_t) \left[1 - \int g(Y|X_t)\alpha(X_t, Y)dY \right] \right\} \\ &= \pi(X_t)g(X_{t+1}|X_t) \min \left(1, \frac{\pi(X_{t+1})g(X_t|X_{t+1})}{\pi(X_t)g(X_{t+1}|X_t)} \right) \\ &\quad + \pi(X_t)I(X_{t+1} = X_t) \left[1 - \int g(Y|X_t)\alpha(X_t, Y)dY \right] \\ &= \min(\pi(X_t)g(X_{t+1}|X_t), \pi(X_{t+1})g(X_t|X_{t+1})) \\ &\quad + \pi(X_{t+1})I(X_{t+1} = X_t) \left[1 - \int g(Y|X_{t+1})\alpha(X_{t+1}, Y)dY \right] \\ &= \pi(X_{t+1})g(X_t|X_{t+1}) \min \left(1, \frac{\pi(X_t)g(X_{t+1}|X_t)}{\pi(X_{t+1})g(X_t|X_{t+1})} \right) \\ &\quad + \pi(X_{t+1})I(X_{t+1} = X_t) \left[1 - \int g(Y|X_{t+1})\alpha(X_{t+1}, Y)dY \right] \\ &= \pi(X_{t+1})K(X_t|X_{t+1}). \end{aligned}$$

Επομένως η αλυσίδα Markov που παράγεται από την παραπάνω διαδικασία είναι αντιστρέψιμη και έχει την $\pi(x)$ ως στάσιμη κατανομή. Άρα αν προσομοιώσουμε μια παρατήρηση από τη στάσιμη κατανομή $\pi(x)$, όλες οι υπόλοιπες τιμές θα προέρχονται επίσης από την π .

Εφόσον η αλυσίδα που προσομοιώθηκε από τον αλγόριθμο MH έχει εκ κατασκευής την $\pi(x)$ ως στάσιμη κατανομή, εάν επιπλέον είναι Harris επαναληπτική και απεριοδική

μπορεί να εφαρμοστεί το Εργοδικό Θεώρημα, το οποίο εξασφαλίζει τη σύγκλιση κάθε εργοδικού μέσου στην αντίστοιχη μέση τιμή.

Για να είναι η αλυσίδα που προκύπτει από τον αλγόριθμο ΜΗ απεριοδική θα πρέπει τα ενδεχόμενα της μορφής $\{X_{t+1} = X_t\}$ να έχουν θετική πιθανότητα ή ισοδύναμα να ισχύει

$$P\{\pi(X_t)g(X_t|X_{t+1}) \leq \pi(X_{t+1})g(X_{t+1}|X_t)\} < 1.$$

Για να είναι Harris επαναληπτική, θα πρέπει να είναι π -μη διαχωρίσιμη. Μια ικανή συνθήκη για αυτό είναι

$$g(Y|X) > 0 \text{ για κάθε } (X, Y) \in \mathcal{X} \times \mathcal{X}.$$

Εφόσον ικανοποιούνται οι δυο παραπάνω σχέσεις ισχύουν τα εξής :

a) Αν $h \in \mathcal{L}(\pi)$, τότε

$$\frac{1}{n} \sum_{t=0}^{n-1} h(X_t) \rightarrow \int_{\mathbb{R}^n} h(x)\pi(dx) = E_{\pi}(h).$$

b)

$$\lim_{n \rightarrow \infty} \|P(X_n \in \cdot) - \pi(\cdot)\| = 0.$$

Ένα από τα σημαντικότερα πλεονεκτήματα της μεθόδου ΜΗ είναι ότι εξασφαλίζει υπό ελάχιστες συνθήκες τη σύγκλιση της ακολουθίας που προσομοιώνεται στην κατανομή-στόχο $\pi(\cdot)$. Επίσης, το γεγονός ότι για την υλοποίηση της χρειαζόμαστε μόνο τη συναρτησιακή μορφή των $\pi(\cdot)$ και $g(\cdot)$ την κάνει μια δημοφιλή μέθοδο αφού σε πολλά πραγματικά προβλήματα ο υπολογισμός αυτών των σταθερών είναι πάρα πολύ δύσκολος αν όχι αδύνατος. Ιδιαίτερα στην περίπτωση που η g είναι συμμετρική τότε ούτε η συναρτησιακή μορφή της μας είναι απαραίτητη.

3.3 Ειδικές μορφές του αλγορίθμου ΜΗ

Στη συνέχεια θα αναφερθούμε σε κάποιες ειδικές μορφές του αλγορίθμου ΜΗ που προκύπτουν ανάλογα με τη μορφή της κατανομής πρότασης.

3.3.1 Ανεξάρτητος δειγματολήπτης ΜΗ (Independent Sampler)

Μια ειδική μορφή του αλγορίθμου ΜΗ είναι ο ανεξάρτητος δειγματολήπτης ΜΗ. Σε αυτή την ειδική μορφή του αλγορίθμου ΜΗ, η κατανομή πρότασης είναι της μορφής

$$g(Y|X_t) = g(Y)$$

είναι δηλαδή ανεξάρτητη από την τρέχουσα κατάσταση με αποτέλεσμα η πιθανότητα αποδοχής να παίρνει τη μορφή

$$\alpha(X_t, Y) = \min\left(1, \frac{\pi(Y)g(X_t)}{\pi(X_t)g(Y)}\right)$$

Παρά το γεγονός ότι η Y παράγεται ανεξάρτητα από την X_t , η προκύπτουσα ακολουθία δεν αποτελείται από ανεξάρτητες παρατηρήσεις αφού η πιθανότητα αποδοχής της Y εξαρτάται από την τιμή της X_t .

Ο ανεξάρτητος δειγματολήπτης ΜΗ μπορεί να συγκλίνει πολύ γρήγορα, μπορεί όμως να οδηγήσει και σε αλυσίδες με υπερβολικά αργή σύγκλιση. Για να δουλεύει ο ανεξάρτητος δειγματολήπτης καλά, θα πρέπει η κατανομή πρότασης να είναι μια καλή προσέγγιση της κατανομής-στόχου ή ακόμα καλύτερα η κατανομή πρότασης να έχει πιο παχιές ουρές σε σχέση με την κατανομή-στόχο.

Ο ανεξάρτητος δειγματολήπτης ΜΗ θα μπορούσαμε να πούμε ότι είναι ένας εναλλακτικός αλγόριθμος των δύο τεχνικών Monte Carlo, της μεθόδου Αποδοχής-Απόρριψης και του Δειγματολήπτη Σπουδαιότητας. Ο Liu (1996) συνέκρινε τις τρεις αυτές τεχνικές και απέδειξε ότι η ασυμπτωτική διασπορά της μεθόδου Αποδοχής-

Απόρριψης είναι μικρότερη από αυτήν του ανεξάρτητου δειγματολήπτη MH αλλά είναι ασυμπτωτικά συγκρίσιμες. Έδειξε επίσης ότι ανάλογα με την υπό ολοκλήρωση συνάρτηση, η μέθοδος του Δειγματολήπτη Σπουδαιότητας μπορεί να είναι ασυμπτωτικά αποδοτικότερη από τις άλλες δυο μεθόδους.

3.3.2 Τυχαίος περίπατος MH (Random Walk MH)

Άλλη μια ειδική περίπτωση του αλγόριθμου MH είναι ο λεγόμενος τυχαίος περίπατος MH. Σε αυτή την περίπτωση η κατανομή πρότασης είναι της μορφής

$$g(Y - X) = g(X - Y)$$

δηλαδή η προτεινόμενη τιμή τη χρονική στιγμή t είναι της μορφής $Y = X_t + \epsilon_t$ όπου $\epsilon_t \sim g_\sigma(\cdot)$ ανεξάρτητες παρατηρήσεις για διαφορετικό t . Με σ συμβολίζουμε την παράμετρο κλίμακας της $g_\sigma(\cdot)$. Όπως παρατηρούμε οι κατανομές πρότασης που χρησιμοποιούνται είναι συμμετρικές με κέντρο κάθε φορά την προηγούμενη κατάσταση, για παράδειγμα θα μπορούσαν να χρησιμοποιηθούν ως κατανομές πρότασης οι $g_\sigma(x, \cdot) \sim \mathcal{N}(x, \sigma^2)$, $g_\sigma(x, \cdot) \sim \mathcal{U}(x - \sigma, x + \sigma)$ και άλλες συμμετρικές κατανομές.

Η τιμή που θα δώσουμε στην παράμετρο κλίμακας παίζει μεγάλο ρόλο στη σύγκλιση του αλγορίθμου. Έτσι, για μεγάλες τιμές του σ , ο αλγόριθμος θα κάνει μεγάλα άλματα αλλά τα περισσότερα από αυτά θα απορρίπτονται με αποτέλεσμα να υπάρχουν μεγάλες περιόδοι που δεν θα έχουμε μετακίνηση σε άλλη κατάσταση οπότε θα έχουμε αργή σύγκλιση με μικρά ποσοστά αποδοχής. Από την άλλη μεριά όταν η παράμετρος κλίμακας παίρνει μικρές τιμές ο αλγόριθμος θα δέχεται σχεδόν όλες τις προτεινόμενες τιμές αλλά οι μετακινήσεις θα είναι τόσο μικρές που και πάλι ο αλγόριθμος χρειάζεται πολύ χρόνο για να εξερευνήσει όλο το χώρο καταστάσεων. Σε αυτή την περίπτωση έχουμε και πάλι αργή σύγκλιση αλλά με μεγάλα ποσοστά αποδοχής. Σκοπός μας λοιπόν είναι να βρεθεί ένα σ που να μας εξασφαλίζει ένα βέλτιστο ποσοστό αποδοχής και μια γρήγορη σύγκλιση.

Η εύρεση του βέλτιστου ποσοστού αποδοχής είναι μια ιδιαίτερα δύσκολη διαδικασία. Υπάρχει μια αρκετά εκτενής βιβλιογραφία πάνω σε αυτό το θέμα (βλέπε Besag and Green 1993, Besag et al. 1995, Roberts et al. 1997, Breyer and Roberts 2000 και άλλοι) στο οποίο και θα αναφερθούμε εκτενέστερα στο Κεφάλαιο 5.

Ο τυχαίος περίπατος είναι μια ειδική περίπτωση του συμμετρικού MH αλγόριθμου όπου η κατανομή πρότασης είναι συμμετρική, δηλαδή ισχύει $g(X, Y) = g(Y, X)$. Σε αυτή την περίπτωση η πιθανότητα αποδοχής παίρνει τη μορφή

$$\alpha(Y, X) = \min\left(1, \frac{\pi(Y)}{\pi(X)}\right).$$

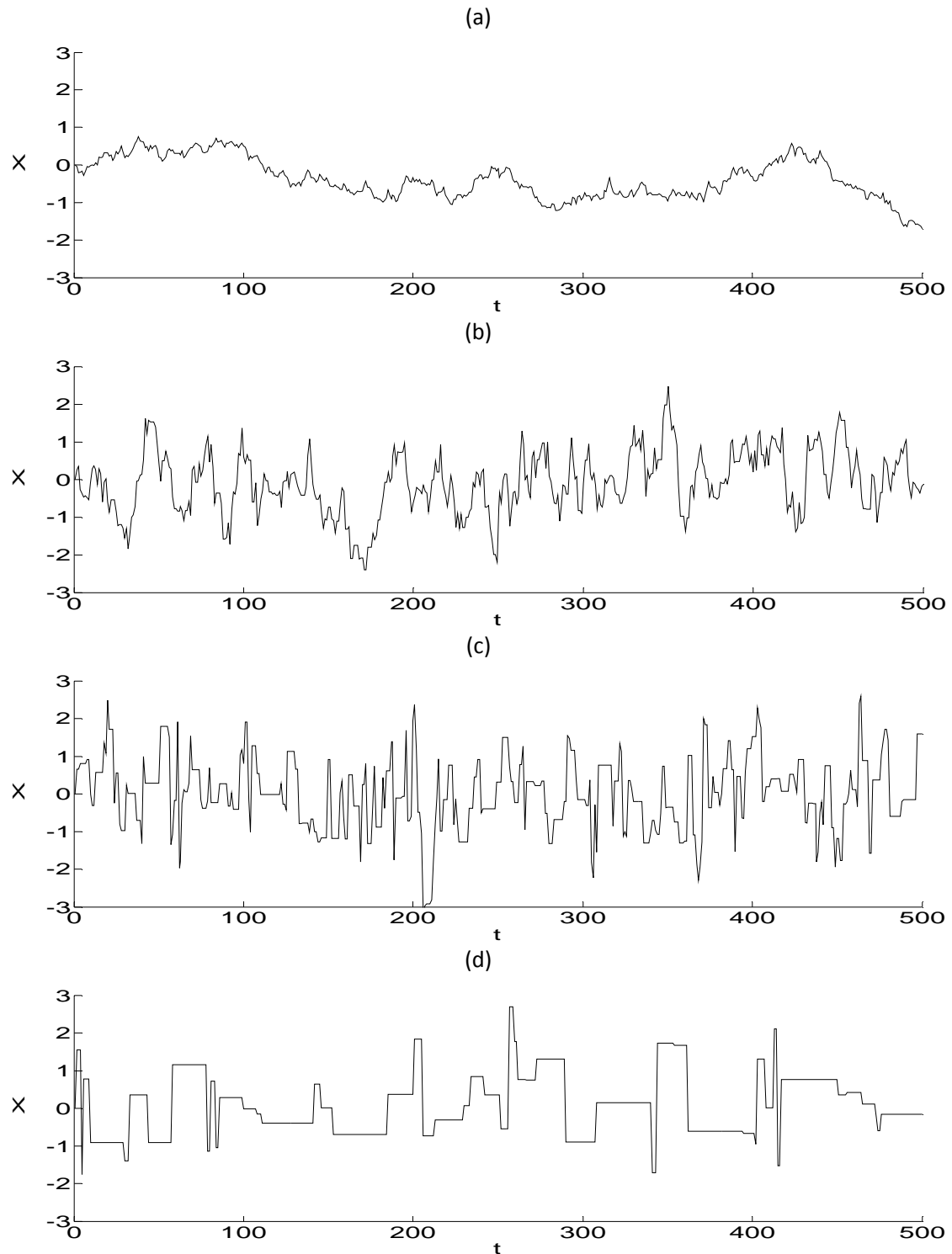
Οι Metropolis et al. (1953) δημοσίευσαν τον αλγόριθμο για πρώτη φορά σε αυτή τη μορφή.

3.4 Παραδείγματα

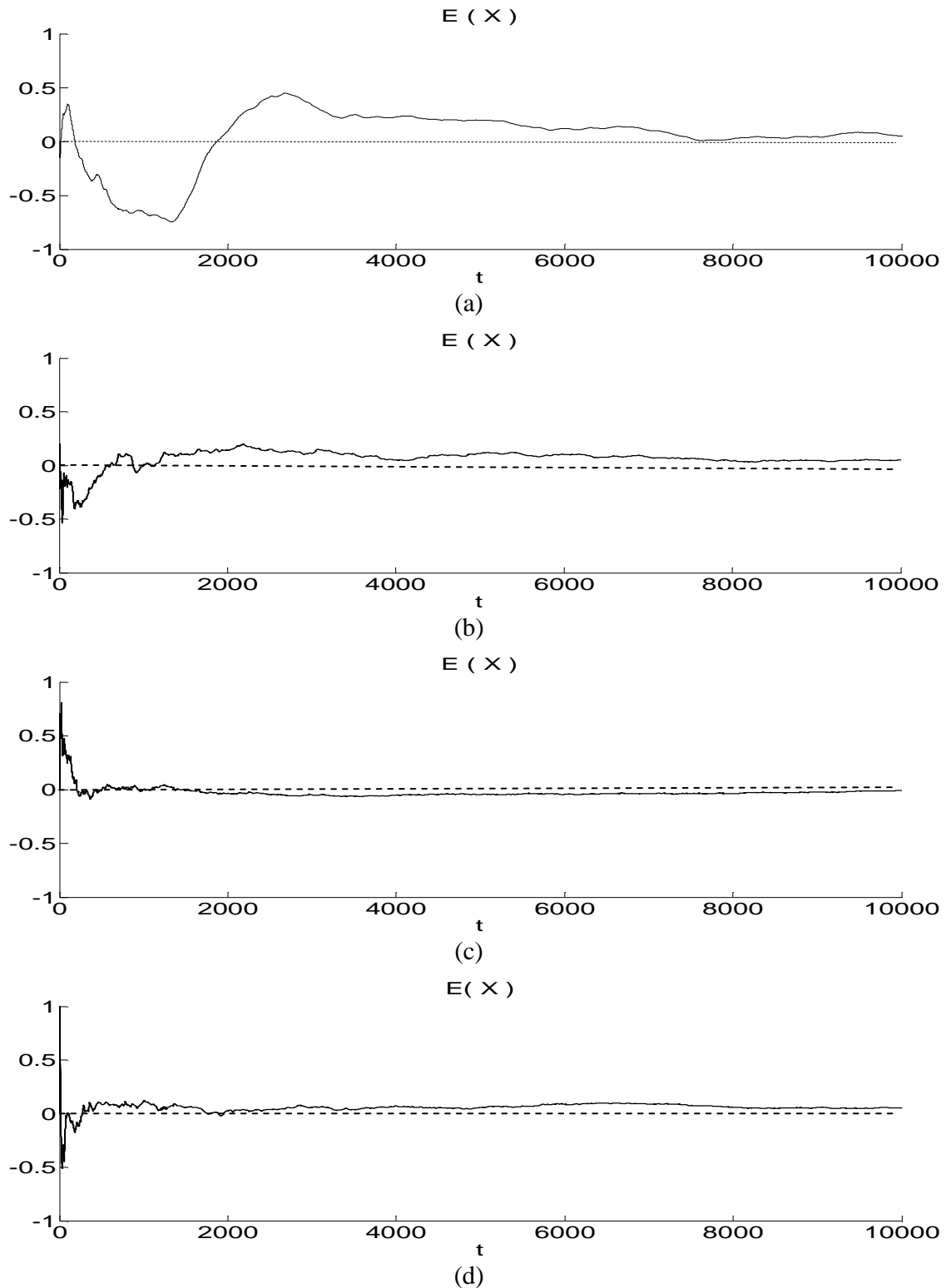
Στη συνέχεια θα υλοποιήσουμε τους παραπάνω αλγόριθμους σε κάποια απλά παραδείγματα καθώς επίσης και σε πραγματικά σύνολα δεδομένων (υποενότητα 3.4.4 σύνολο δεδομένων dugongs).

3.4.1 Κανονικές κατανομές

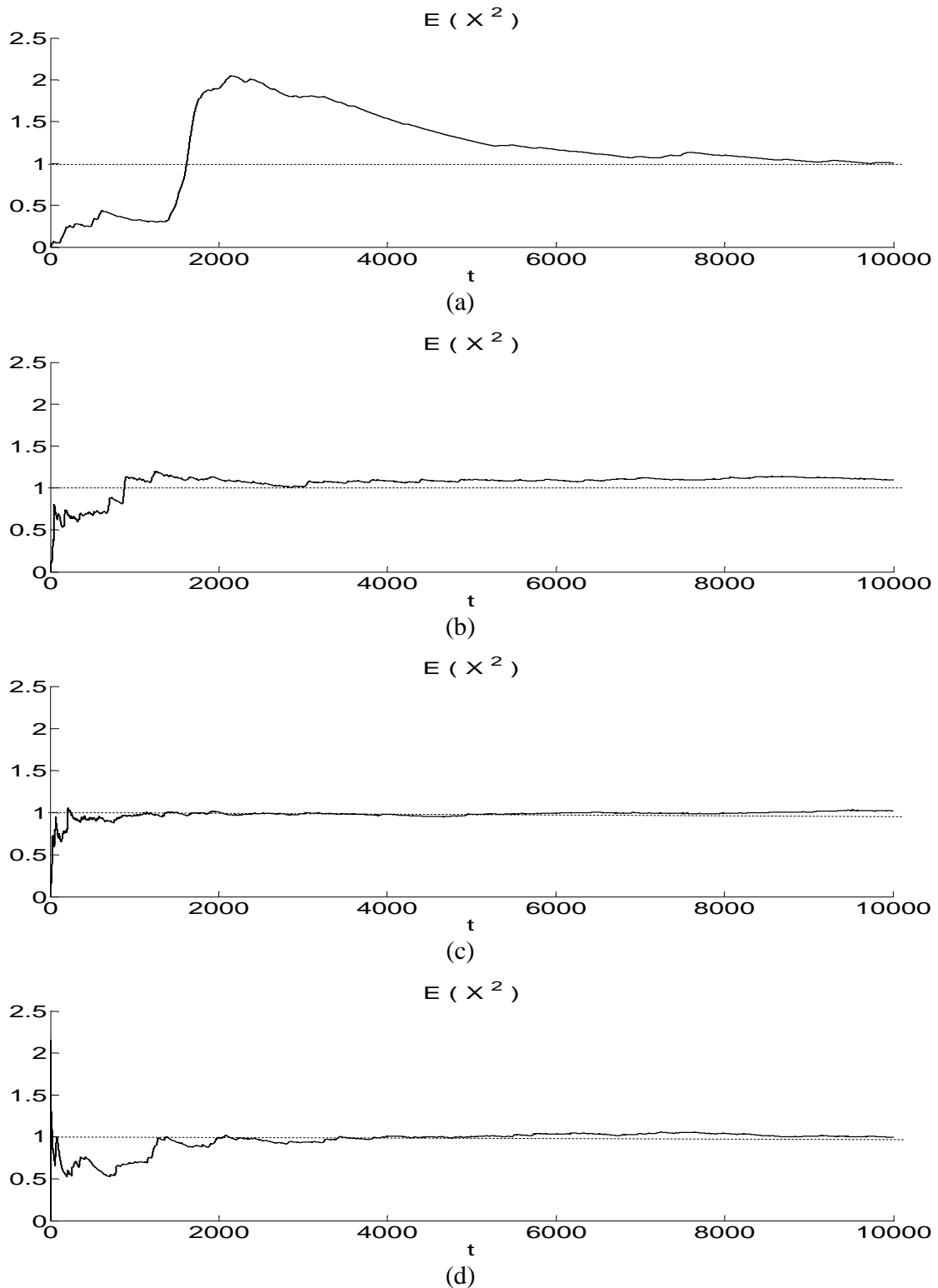
Έστω ότι η κατανομή-στόχος είναι η τυπική κανονική κατανομή $\pi(x) \sim \mathcal{N}(0, 1)$ ενώ η κατανομή πρότασης είναι της μορφής $g(\cdot | x) \sim \mathcal{N}(x, \sigma^2)$, είναι δηλαδή μια κανονική κατανομή με μέση τιμή την προηγούμενη κατάσταση της αλυσίδας Markov και τυπική απόκλιση σ . Προσομοιώνουμε μια αλυσίδα Markov 10000 παρατηρήσεων για διάφορες τιμές του σ , ($\sigma = 0.1, 0.5, 2$ και 10) και εκτιμάμε την πρώτη και τη δεύτερη ροπή της $\pi(x)$.



Σχήμα 3.1 500 επαναλήψεις από τον αλγόριθμο ΜΗ με κατανομή στόχο $\mathcal{N}(0, 1)$ και κατανομή πρότασης (a) $g(\cdot|X) = \mathcal{N}(X, 0.1^2)$, (b) $g(\cdot|X) = \mathcal{N}(X, 0.5^2)$, (c) $g(\cdot|X) = \mathcal{N}(X, 2^2)$ και (d) $g(\cdot|X) = \mathcal{N}(X, 10^2)$.



Σχήμα 3.2 Σύγκλιση εργοδικών μέσων πρώτης ροής που προκύπτουν από τον αλγόριθμο MH με κατανομή-στόχο και κατανομή πρότασης (a), (b), (c) και (d).



Σχήμα 3.3 Σύγκλιση εργοδικών μέσων δεύτερης ροπής που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο και κατανομή πρότασης (a) , (b) , (c) και (d) .

Στο Σχήμα 3.1 απεικονίζονται οι προσομοιωμένες τιμές της αλυσίδας Markov με στάσιμη κατανομή $\pi(x)$ για τις διάφορες τιμές του σ . Παρατηρούμε ότι για μεγάλες τιμές του σ ο αλγόριθμος απορρίπτει πολλές από τις προτεινόμενες τιμές με αποτέλεσμα να παραμείνει στην ίδια κατανομή για αρκετό χρόνο. Επίσης παρατηρούμε μεγάλα άλματα μέσα στο χώρο καταστάσεων (βλέπε Σχήμα 3.1(d)). Από την άλλη πλευρά μικρές τιμές της παραμέτρου κλίμακας μας δίνουν μικρές μετακινήσεις στο χώρο καταστάσεων, οι οποίες όμως σχεδόν όλες γίνονται αποδεκτές (βλέπε Σχήμα 3.1 (a) – (b)).

Στα Σχήματα 3.2 και 3.3 απεικονίζεται η σύγκλιση των εργοδικών μέσων της πρώτης και δεύτερης ροπής αντίστοιχα για τις διάφορες τιμές της παραμέτρου σ .

Είναι φανερό ότι για μικρές τιμές της τυπικής απόκλισης ($\sigma = 0.1$ και 0.5), δεν έχουμε γρήγορη σύγκλιση του εργοδικού μέσου. Όπως βλέπουμε στο Σχήμα 3.2(a) για $\sigma = 0.1$ δεν μπορούμε να ισχυριστούμε ότι μετά από 10000 επαναλήψεις ο εργοδικός μέσος έχει συγκλίνει στην προς εκτίμηση ποσότητα. Αντίθετα για $\sigma=2$ και 10 ο εργοδικός μέσος συγκλίνει πολύ γρήγορα (βλέπε Σχήμα 3.2 (c) – (d)). Όμοια, μετά από 10000 επαναλήψεις για τη δεύτερη ροπή παρατηρούμε ότι για $\sigma = 0.1$ ο εργοδικός μέσος δεν έχει συγκλίνει ακόμα. Για $\sigma = 0.5$ και $\sigma = 10$ έχουμε σύγκλιση από τις 3000 περίπου επαναλήψεις και μετά ενώ για $\sigma = 2$ παρατηρούμε γρηγορότερη σύγκλιση (περίπου 2000 επαναλήψεις και μετά).

3.4.2 Δισδιάστατες κατανομές

Έστω

$$\pi(x, y) \propto \exp\{-(x^2 - \sqrt{2}xy + y^2)\}$$

δισδιάστατη κατανομή-στόχος και θέλουμε να εκτιμήσουμε την $E(X)$ και την $E(Y)$. Η προσομοίωση παρατηρήσεων από την $\pi(x, y)$ μπορεί να γίνει είτε χρησιμοποιώντας τον αλγόριθμο MH είτε το Δειγματολήπτη Gibbs προσομοιώνοντας παρατηρήσεις από τις πλήρως δεσμευμένες κατανομές $\pi(x|y)$ και $\pi(y|x)$.

Στον αλγόριθμο MH χρησιμοποιήσαμε ως κατανομή πρότασης τη δισδιάστατη τυπική κανονική κατανομή

$$g(x, y) \sim \mathcal{N}(0, I_2).$$

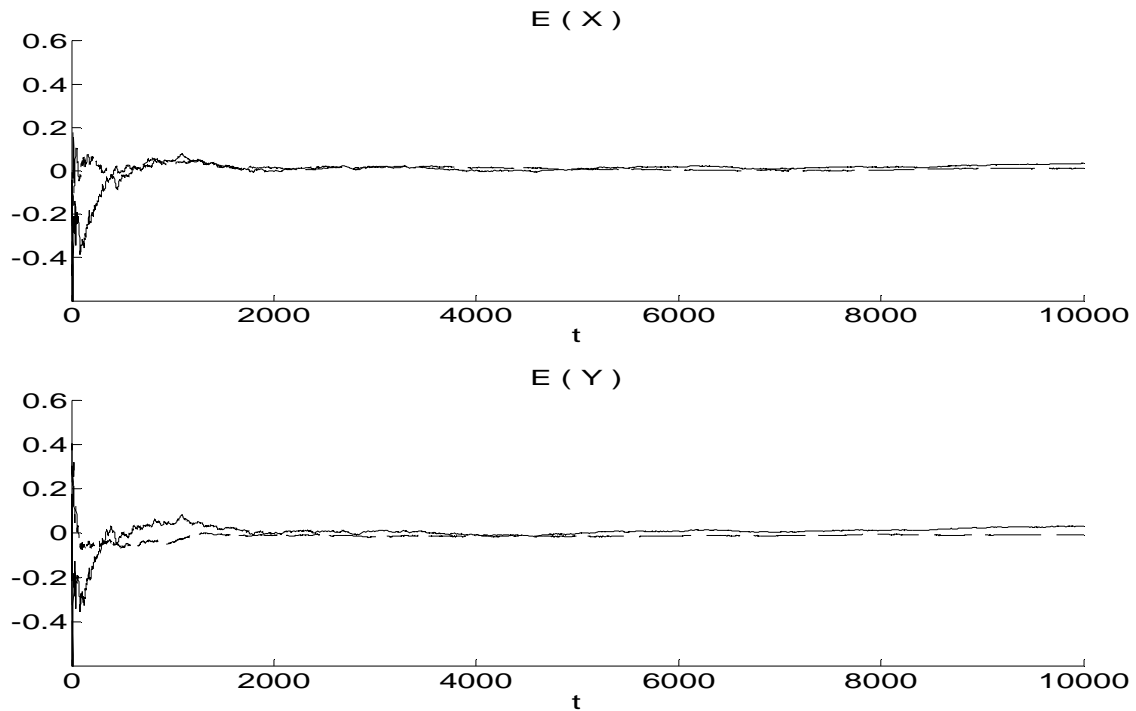
Η πλήρως δεσμευμένη κατανομής της X είναι η

$$\begin{aligned} \pi(x|y) &\propto \exp\{-(x^2 - \sqrt{2}xy)\} \\ &\propto \exp\left\{-\left(x^2 - 2\frac{\sqrt{2}xy}{2} + \left(\frac{\sqrt{2}}{2}y\right)^2 - \left(\frac{\sqrt{2}}{2}y\right)^2\right)\right\} \\ &\propto \exp\left\{-\left(x^2 - \sqrt{2}xy + \left(\frac{\sqrt{2}}{2}y\right)^2\right)\right\} \\ &\propto \exp\left\{-\left(x - \frac{\sqrt{2}}{2}y\right)^2\right\} \sim \mathcal{N}\left(\frac{\sqrt{2}}{2}y, \frac{1}{2}\right) \end{aligned}$$

Άρα η πλήρως δεσμευμένη κατανομή της X είναι μια κανονική κατανομή με μέση τιμή $\sqrt{2}y/2$ και διασπορά $1/2$. Αντίστοιχα βρήκαμε την πλήρως δεσμευμένη κατανομή της y :

$$\begin{aligned} \pi(y|x) &\propto \exp\{-(y^2 - \sqrt{2}xy)\} \\ &\propto \exp\left\{-\left(y^2 - 2\frac{\sqrt{2}xy}{2} + \left(\frac{\sqrt{2}}{2}x\right)^2 - \left(\frac{\sqrt{2}}{2}x\right)^2\right)\right\} \\ &\propto \exp\left\{-\left(y^2 - \sqrt{2}xy + \left(\frac{\sqrt{2}}{2}x\right)^2\right)\right\} \\ &\propto \exp\left\{-\left(y - \frac{\sqrt{2}}{2}x\right)^2\right\} \sim \mathcal{N}\left(\frac{\sqrt{2}}{2}x, \frac{1}{2}\right) \end{aligned}$$

Στο Σχήμα 3.4 απεικονίζεται η σύγκλιση των εργοδικών μέσων στις $E(X)$ και $E(Y)$ αντίστοιχα για τις δύο μεθόδους προσομοίωσης.



Σχήμα 3.4 Σύγκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH (—) με κατανομή στόχο $\pi(x, y) = \exp\{-(x^2 - \sqrt{2}xy + y^2)\}$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ και από τον Δειγματολήπτη Gibbs (- -).

Πίνακας 3.1 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο $\pi(x, y) = \exp\{-(x^2 - \sqrt{2}xy + y^2)\}$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ και από τον αντίστοιχο Δειγματολήπτη Gibbs.

	Metropolis-Hastings		Δειγματολήπτης Gibbs	
	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
x	-0,000506	0,0114	0,000068	0,0176
y	0,000307	0,0069	-0,000297	0,0176

Για να μπορέσουμε να συγκρίνουμε τις δύο μεθόδους προσομοίωσης ως προς τα τυπικά σφάλματα τρέξαμε 1000 ανεξάρτητες προσομοιωμένες αλυσίδες μήκους 10000 παρατηρήσεων η κάθε μια και εκτιμήσαμε τη μέση τιμή των εργοδικών μέσων καθώς και την τυπική απόκλιση τους. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.1.

Παρατηρούμε ότι ο αλγόριθμος ΜΗ είναι αποδοτικότερος από το Δειγματολήπτη Gibbs στο συγκεκριμένο παράδειγμα καθώς δίνει εκτιμητές με μικρότερες τυπικές αποκλίσεις.

Έστω τώρα η κατανομή-στόχος είναι η

$$\pi(x, y) \propto y^{3/2} \exp \left\{ -y \left(\frac{x^2}{2} + 2 \right) \right\} I_{(0, \infty)}(y)$$

της οποίας θέλουμε να εκτιμήσουμε την $E(X)$ και την $E(Y)$. Για την προσομοίωση παρατηρήσεων από την $\pi(x, y)$ θα χρησιμοποιήσουμε τον αλγόριθμο ΜΗ και το Δειγματολήπτη Gibbs. Για τον αλγόριθμο ΜΗ χρησιμοποιούμε δύο διαφορετικές κατανομές πρότασης. Τη διδιάστατη περικομμένη κανονική κατανομή (truncated normal distribution)

$$g(x, y) \sim \mathcal{N}(0, I_2) I_{(0, \infty)}(y)$$

και εναλλακτικά χρησιμοποιούμε ως κατανομή πρότασης την

$$\begin{aligned} g(x, y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \times \frac{2^{5/2}}{\Gamma(5/2)} y^{\frac{5}{2}-1} e^{-2y} \\ &\propto y^{\frac{3}{2}} e^{-2y - \frac{x^2}{2}}. \end{aligned}$$

Για το Δειγματολήπτη Gibbs χρησιμοποιήσαμε τις πλήρως δεσμευμένες κατανομές των X και Y . Η πλήρως δεσμευμένη κατανομή της X είναι η

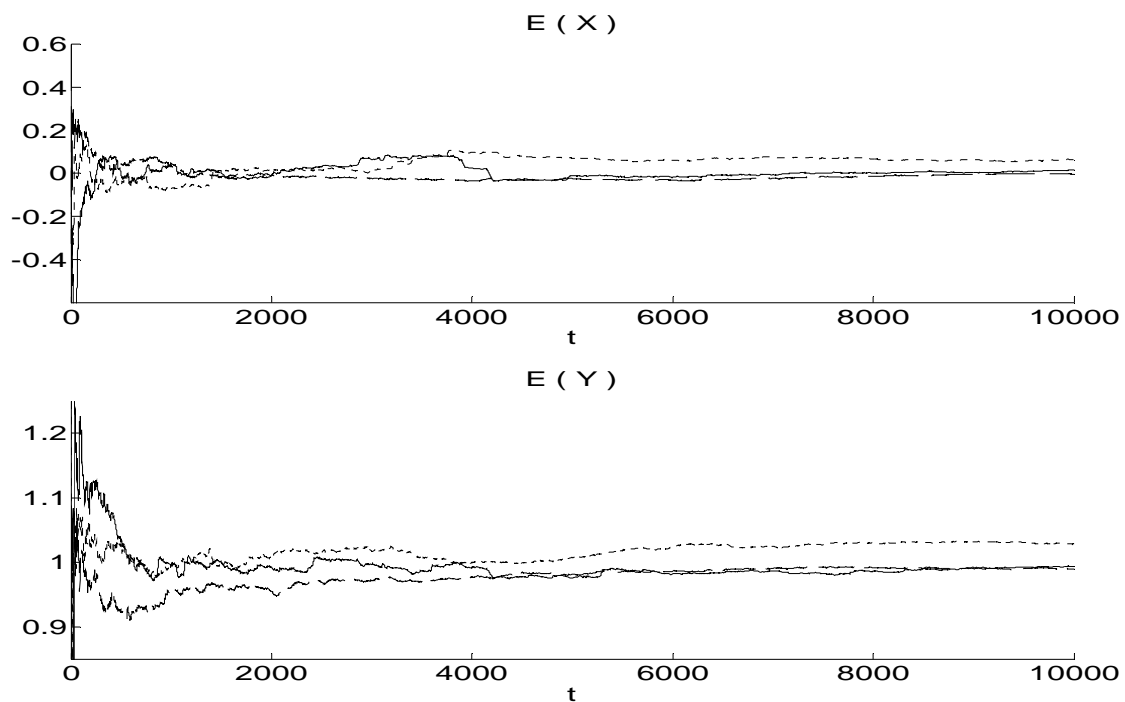
$$\pi(x|y) \propto \exp \left\{ -x^2 \cdot \frac{y}{2} \right\} \sim \mathcal{N} \left(0, \frac{1}{y} \right)$$

δηλαδή μια κανονική κατανομή με μέση τιμή μηδέν και διασπορά $1/y$ ενώ η πλήρως δεσμευμένη κατανομή της y είναι η

$$\pi(y|x) \propto y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I_{(0,\infty)}(y) \sim \mathcal{G}\left(\frac{5}{2}, \frac{x^2}{2} + 2\right)$$

δηλαδή είναι μια κατανομή γάμμα με παραμέτρους $5/2$ και $x^2/2 + 2$.

Στο Σχήμα 3.5 απεικονίζεται η σύγκλιση των εργοδικών μέσων για κάθε μεταβλητή X , Y ξεχωριστά και για τις τρεις μεθόδους προσομοίωσης.



Σχήμα 3.5 Σύγκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH (—) με κατανομή στόχο $\pi(x, y) \propto y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I_{(0,\infty)}(y)$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$ με $y > 0$, από τον MH (.....) με κατανομή πρότασης $g(x, y) \sim y^{3/2} e^{-2x - \frac{x^2}{2}}$ και από τον αντίστοιχο Δειγματολήπτη Gibbs (- - -).

Για να μπορέσουμε να συγκρίνουμε τις τρεις μεθόδους προσομοίωσης ως προς τα τυπικά σφάλματα τρέξαμε για κάθε μέθοδο 1000 ανεξάρτητες προσομοιωμένες αλυσίδες μήκους 10000 παρατηρήσεων η κάθε μια και εκτιμήσαμε το μέσο όρο των εργοδικών μέσων καθώς και τις τυπικές αποκλίσεις τους. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.2.

Πίνακας 3.2 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH με κατανομή στόχο $\pi(x, y) \propto y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I_{(0, \infty)}(y)$ και κατανομή πρότασης $g(x, y) \sim \mathcal{N}(0, I_2)$, από τον MH με κατανομή πρότασης $g(x, y) \sim y^{\frac{3}{2}} e^{-2x - \frac{x^2}{2}}$ και από τον αντίστοιχο Δειγματολήπτη Gibbs.

	Metropolis-Hastings		Metropolis-Hastings		Δειγματολήπτης Gibbs	
	$g(x, y) \sim \mathcal{N}(\mathbf{0}, I_2) I_{(0, \infty)}(y)$		$g(x, y) \propto y^{\frac{3}{2}} e^{-2x - \frac{x^2}{2}}$			
	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
x	0.0040	0.2011	0.0074	0.1725	-0.0004	0.0141
y	1.0045	0.1072	1.0086	0.0353	1.0000	0.0086

Παρατηρούμε ότι οι εκτιμητές που προκύπτουν από το Δειγματολήπτη Gibbs είναι καλύτεροι ως προς τα τυπικά σφάλματα από τους εκτιμητές που προκύπτουν από τον αλγόριθμο MH ανεξάρτητα από την κατανομή πρότασης.

3.4.3 Μίξη κανονικών κατανομών

Έστω ότι η κατανομή-στόχος είναι η

$$\pi(x) \sim \frac{1}{3} \mathcal{N}(0, 9) + \frac{1}{3} \mathcal{N}(5, 1) + \frac{1}{3} \mathcal{N}(15, 4)$$

της οποίας θέλουμε να εκτιμήσουμε την πρώτη ροπή. Σε αυτό το παράδειγμα θα χρησιμοποιήσουμε τη μέθοδο Αποδοχής-Απόρριψης, το Δειγματολήπτη Σπουδαιότητας και τον ανεξάρτητο MH αλγόριθμο. Στις τρεις αυτές μεθόδους ως κατανομή πρότασης χρησιμοποιείται η *Cauchy*(0, 10) με συνάρτηση πυκνότητας πιθανότητας

$$g(x) = \frac{10}{\pi \cdot (10^2 + x^2)} \propto \frac{1}{x^2 + 10^2}.$$

Για τη σύγκριση των παραπάνω μεθόδων, χρησιμοποιήθηκαν και στις τρεις τεχνικές οι ίδιες προτεινόμενες τιμές. Χρησιμοποιήσαμε δύο διαφορετικούς τρόπους για να πάρουμε συγκρίσιμα αποτελέσματα μεταξύ των αλγορίθμων. Στην πρώτη περίπτωση προσομοιώνουμε 1000 αλυσίδες των 10000 παρατηρήσεων για τον αλγόριθμο MH και το Δειγματολήπτη Σπουδαιότητας ενώ το πλήθος των παρατηρήσεων που έγιναν αποδεκτές από τη μέθοδο Αποδοχής-Απόρριψης ήταν τυχαίο. Κατά μέσο όρο περίπου 1460 παρατηρήσεις έγιναν αποδεκτές από τη μέθοδο Αποδοχής-Απόρριψης. Στη δεύτερη περίπτωση για να έχουμε μεγαλύτερη ακρίβεια τρέξαμε για τη μέθοδο Αποδοχής-Απόρριψης 1000 αλυσίδες μέχρι να αποδεχτούμε 10000 παρατηρήσεις για την κάθε μία. Σε αυτή την περίπτωση το πλήθος των παρατηρήσεων του αλγόριθμου MH και του Δειγματολήπτη Σπουδαιότητας ήταν τυχαίο. Στις 1000 αλυσίδες κατά μέσο όρο το μήκος των προσομοιωμένων παρατηρήσεων από αυτούς τους αλγόριθμους ήταν περίπου 70130 παρατηρήσεις. Στον Πίνακα 3.3 παρουσιάζονται οι μέσοι όροι και οι τυπικές αποκλίσεις των εργοδικών μέσων.

Πίνακας 3.3 Μέση τιμή και τυπική απόκλιση εργοδικών μέσων που προκύπτουν από τον αλγόριθμο MH, τη μέθοδο Αποδοχής-Απόρριψης και το Δειγματολήπτη Σπουδαιότητας με κατανομή στόχο $\pi(x) \sim \frac{1}{3}\mathcal{N}(0, 9) + \frac{1}{3}\mathcal{N}(5, 1) + \frac{1}{3}\mathcal{N}(15, 4)$ και κατανομή πρότασης $g(x, y) \sim \text{Cauchy}(0, 10)$.

	1 ^η Περίπτωση		2 ^η Περίπτωση	
	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
Metropolis-Hastings	6.6599	0.1867	6.6588	0.0703
Αποδοχής-Απόρριψης	6.6562	0.1678	6.6647	0.0656
Δειγματολήπτης Σπουδαιότητας	6.6667	0.1179	6.6619	0.0445

Παρατηρούμε ότι και με τις τρεις μεθόδους έχουμε βρει σχεδόν ίδιες μέσες τιμές των εργοδικών μέσων. Παρατηρούμε όμως ότι ο Δειγματολήπτης Σπουδαιότητας δίνει εκτιμητές με μικρότερα τυπικά σφάλματα σε σχέση με αυτούς που προκύπτουν από τις άλλες δύο μεθόδους κάτι που το περιμέναμε από τη βιβλιογραφία. Επίσης η μέθοδος

Αποδοχής-Απόρριψης δίνει εκτιμητές με μικρότερη τυπική απόκλιση από αυτούς που προκύπτουν από τον αλγόριθμο ΜΗ.

Στη συνέχεια θα ασχοληθούμε με ένα πρόβλημα πραγματικών δεδομένων.

3.4.4 Σύνολο δεδομένων dugongs

Το σύνολο δεδομένων dugongs αναλύθηκε αρχικά από τον Ratkowsky (1983) και έκτοτε έχει χρησιμοποιηθεί από πολλούς ερευνητές προκειμένου να εφαρμόσουν ή να συγκρίνουν διάφορες υπολογιστικές τεχνικές. Τα δεδομένα αφορούν $n=27$ θαλάσσιους ελέφαντες (dugongs) που συλλέχθηκαν κοντά στο Townsville του Queensland και αποτελούνται από μετρήσεις του μήκους (y) και της ηλικίας (x) τους. Στον Πίνακα 3.4 απεικονίζονται τα παραπάνω δεδομένα.

Πίνακας 3.4 Μετρήσεις της ηλικίας και του μήκους των 27 θαλάσσιων ελεφάντων που συλλέχθηκαν κοντά στο Townsville του Queensland.

Ηλικία (x_i)	1.0	1.5	1.5	1.5	2.5	4.0	5.0	5.0	7.0
Μήκος (y_i)	1.80	1.85	1.87	1.77	2.02	2.27	2.15	2.26	2.35
Ηλικία (x_i)	8.0	8.5	9.0	9.5	9.5	10.0	12.0	12.0	13.0
Μήκος (y_i)	2.47	2.19	2.26	2.40	2.39	2.41	2.50	2.32	2.43
Ηλικία (x_i)	13.0	14.5	15.5	15.5	16.5	17.0	22.5	29.0	31.5
Μήκος (y_i)	2.47	2.56	2.65	2.47	2.64	2.56	2.70	2.72	2.57

Οι Carlin and Gelfand (1991) υπέθεσαν ότι

$$y_i \sim N(\alpha - \beta \gamma^{x_i}, \tau^{-1}), \quad i = 1, \dots, n$$

όπου $\alpha, \beta, \tau > 0$ και $0 < \gamma < 1$. Ως εκ των προτέρων κατανομή (prior distribution) των παραμέτρων του μοντέλου θεώρησαν την

$$q(\alpha, \beta, \gamma, \tau) \propto \alpha^{-1} \tau^{1/2}, \quad \alpha > 0, \beta > 0, 0 < \gamma < 1, \tau > 0.$$

Αυτό σημαίνει ότι θεωρήσαν τις παραμέτρους του μοντέλου ανεξάρτητες με $\gamma \sim \mathcal{U}(0,1)$ και α, β, τ να έχουν κάποιες καταχρηστικές (γενικευμένες) εκ των προτέρων κατανομές. Εναλλακτικά θα μπορούσαμε να θεωρήσουμε ότι οι παράμετροι είναι εκ των προτέρων ανεξάρτητες και ακολουθούν τις κατανομές

$$\alpha \sim \mathcal{N}(0, \tau_\alpha^{-1}) I(\alpha > 0),$$

$$\beta \sim \mathcal{N}(0, \tau_\beta^{-1}) I(\beta > 0),$$

$$\gamma \sim \mathcal{U}(0,1),$$

$$\tau \sim \mathcal{G}(k, k),$$

με $\tau_\alpha = \tau_\beta = 10^{-4}$ και $k = 10^{-3}$.

Τότε, η εκ των υστέρων κατανομή (posterior distribution) του $\theta = (\alpha, \beta, \gamma, \tau)$ είναι η

$$\begin{aligned} \pi(\theta/data) &= f(data/\theta) f(\alpha) f(\beta) f(\gamma) f(\tau) \\ &\propto \prod_{i=1}^n \frac{1}{\tau^{-\frac{1}{2}}} \exp\left\{-\frac{\tau}{2}(y_i - \alpha + \beta\gamma^{x_i})^2\right\} \\ &\quad \times \exp\left\{-\frac{\tau_\alpha \alpha^2}{2}\right\} I(\alpha > 0) \times \exp\left\{-\frac{\tau_\beta \beta^2}{2}\right\} I(\beta > 0) \\ &\quad \times I_{(0,1)}(\gamma) \times \tau^{k-1} \exp\{-k\tau\} I_{(0,\infty)}(\tau) \\ &\propto \tau^{n/2} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2\right\} \\ &\quad \times \exp\left\{-\frac{\tau_\alpha \alpha^2}{2}\right\} \times \exp\left\{-\frac{\tau_\beta \beta^2}{2}\right\} \times \tau^{k-1} \exp\{-k\tau\} \\ &\quad \times I(\alpha > 0) \times I(\beta > 0) \times I_{(0,1)}(\gamma) \times I_{(0,\infty)}(\tau) \\ &\propto \tau^{n/2+k-1} \exp\left\{-\tau \left(\frac{1}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2 + k\right) - \frac{\tau_\alpha \alpha^2}{2} - \frac{\tau_\beta \beta^2}{2}\right\} \end{aligned}$$

με $\alpha > 0, \beta > 0, \tau > 0, 0 < \gamma < 1$.

Για να προσομοιώσουμε δείγμα από την $\pi(\theta/data)$ θα χρησιμοποιήσουμε τρεις διαφορετικές μεθόδους. Αρχικά θα χρησιμοποιήσουμε τον αλγόριθμο MH στον τετραδιάστατο χώρο καταστάσεων. Εναλλακτικά θα ολοκληρώσουμε ως προς τ και θα χρησιμοποιήσουμε τον αλγόριθμο MH στον τρισδιάστατο χώρο καταστάσεων αυτή τη φορά. Τέλος θα χρησιμοποιήσουμε το Δειγματολήπτη Gibbs ενσωματώνοντας ένα βήμα Metropolis για το γ (Metropolis within Gibbs).

1^η Μέθοδος

Για την παραγωγή παρατηρήσεων από την $\pi(\theta/data)$ με τη βοήθεια της μεθόδου MH χρησιμοποιούμε ως κατανομή πρότασης την περικομμένη κανονική κατανομή $\mathcal{N}_4(\mu, \Sigma)$ όπου

$$\mu = (2.6 \ 0.97 \ 0.87 \ 100)' \quad \text{και} \quad \Sigma = \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

με $\alpha, \beta, \tau > 0$ και $0 < \gamma < 1$.

Στο Σχήμα 3.7 απεικονίζεται η σύγκλιση των εργοδικών μέσων των παραμέτρων α, β και γ .

2^η Μέθοδος

Σε αυτή τη μέθοδο όπως είπαμε και παραπάνω ολοκληρώσαμε την εκ των υστέρων κατανομή της θ , $\pi(\theta/data)$ ως προς τ , οπότε έχουμε,

$$\begin{aligned} \pi(\alpha, \beta, \gamma/data) &\propto \\ &\propto \int_{-\infty}^{+\infty} \tau^{n/2+k-1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2\right\} \exp\left\{-\tau k - \frac{\tau\alpha^2}{2} - \frac{\tau\beta^2}{2}\right\} d\tau \\ &\propto \int_{-\infty}^{+\infty} \tau^{n/2+k-1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \tau k\right\} \exp\left\{-\frac{\tau\alpha^2}{2} - \frac{\tau\beta^2}{2}\right\} d\tau \end{aligned}$$

$$\begin{aligned}
& \propto \int_{-\infty}^{+\infty} \tau^{n/2+k-1} \exp \left\{ -\tau \left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right) \right\} \exp \left\{ -\frac{\tau \alpha^2}{2} - \frac{\tau \beta^2}{2} \right\} d\tau \\
& \propto \frac{\Gamma \left(\frac{n}{2} + k \right)}{\left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right)^{\frac{n}{2}+k}} \times \\
& \quad \int_{-\infty}^{+\infty} \frac{\tau^{n/2+k-1} \left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right)^{\frac{n}{2}+k}}{\Gamma \left(\frac{n}{2} + k \right)} \cdot \\
& \quad \exp \left\{ -\tau \left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right) \right\} d\tau \times \exp \left\{ -\frac{\tau \alpha^2}{2} - \frac{\tau \beta^2}{2} \right\} \\
& \propto \frac{\left(\frac{n}{2} + k \right)}{\left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right)^{\frac{n}{2}+k}} \exp \left\{ -\frac{\tau \alpha^2}{2} - \frac{\tau \beta^2}{2} \right\} \\
& \propto \left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta \gamma^{x_i})^2}{2} + k \right)^{-\left(\frac{n}{2}+k\right)} \exp \left\{ -\frac{\tau \alpha^2}{2} - \frac{\tau \beta^2}{2} \right\}
\end{aligned}$$

με $\alpha > 0, \beta > 0, 0 < \gamma < 1$.

Για την προσομοίωση παρατηρήσεων από την $\pi(\alpha, \beta, \gamma / data)$ χρησιμοποιούμε τον αλγόριθμο ΜΗ με κατανομή πρότασης την περικυμμένη κανονική κατανομή $\mathcal{N}_3(\mu, \Sigma)$ όπου

$$\mu = (2.6 \ 0.97 \ 0.87)' \quad \text{και} \quad \Sigma = \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}$$

με $\alpha, \beta, \tau > 0$ και $0 < \gamma < 1$.

Στο Σχήμα 3.6 απεικονίζεται η σύγκλιση των εργοδικών μέσων των παραμέτρων α, β και γ .

3^η Μέθοδος

Στην περίπτωση αυτή κάνουμε χρήση του δειγματολήπτη Gibbs ενσωματώνοντας ένα βήμα Metropolis για το γ , αφού η πλήρως δεσμευμένη κατανομή του γ δεν είναι κάποια γνωστή κατανομή, αντίθετα με τις πλήρως δεσμευμένες κατανομές των α, β που είναι περικομμένες κανονικές (truncated normal) και του τ που ακολουθεί γάμμα κατανομή. Παρακάτω υπολογίζουμε τις πλήρως δεσμευμένες κατανομές των α, β, γ και τ .

$$\pi(\theta/data) \propto \tau^{n/2+k-1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2\right\} \exp\left\{-\tau k - \frac{\tau\alpha^2}{2} - \frac{\tau\beta^2}{2}\right\}$$

Εύρεση πλήρως δεσμευμένης κατανομής της α :

$$\begin{aligned} \pi(\alpha/\beta, \gamma, \tau, data) &\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \frac{\tau\alpha^2}{2}\right\} \\ &\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [\alpha^2 - 2\alpha(y_i + \beta\gamma^{x_i}) + (y_i + \beta\gamma^{x_i})^2] - \frac{\tau\alpha^2}{2}\right\} \\ &\propto \exp\left\{-\frac{\tau}{2} \left[n\alpha^2 - 2\alpha \sum_{i=1}^n (y_i + \beta\gamma^{x_i}) \right] - \frac{\tau\alpha^2}{2}\right\} \\ &\propto \exp\left\{-\frac{\tau}{2} n\alpha^2 + 2\alpha \frac{\tau}{2} \sum_{i=1}^n (y_i + \beta\gamma^{x_i}) - \frac{\tau\alpha^2}{2}\right\} \\ &\propto \exp\left\{-\left(\frac{n\tau + \tau\alpha}{2}\right) \alpha^2 + 2\alpha \frac{\tau}{2} \sum_{i=1}^n (y_i + \beta\gamma^{x_i})\right\} \\ &\propto \exp\left\{-\left(\frac{n\tau + \tau\alpha}{2}\right) \left[\alpha^2 - 2\alpha \left(\frac{n\tau + \tau\alpha}{2}\right)^{-1} \frac{\tau}{2} \sum_{i=1}^n (y_i + \beta\gamma^{x_i}) \right. \right. \\ &\quad \left. \left. + \left[\left(\frac{n\tau + \tau\alpha}{2}\right)^{-1} \frac{\tau}{2} \sum_{i=1}^n (y_i + \beta\gamma^{x_i}) \right]^2 \right] \right\} \end{aligned}$$

Άρα τελικά η πλήρως δεσμευμένη κατανομή της α είναι:

$$\pi(\alpha/\beta, \gamma, \tau, data) \sim \mathcal{N}\left(\frac{\tau}{n\tau + \tau_\alpha} \sum_{i=1}^n (y_i + \beta\gamma^{x_i}), (n\tau + \tau_\alpha)^{-1}\right)$$

Εύρεση πλήρως δεσμευμένης κατανομής της β :

$$\begin{aligned} \pi(\beta/\alpha, \gamma, \tau, data) &\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \frac{\tau_\beta \beta^2}{2}\right\} \\ &\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [(y_i - \alpha)^2 + 2(y_i - \alpha)\beta\gamma^{x_i} + (\beta\gamma^{x_i})^2] - \frac{\tau_\beta \beta^2}{2}\right\} \\ &\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (\beta\gamma^{x_i})^2 - \frac{\tau}{2} 2 \sum_{i=1}^n (\beta\gamma^{x_i})(y_i - \alpha) - \frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha)^2 - \frac{\tau_\beta \beta^2}{2}\right\} \\ &\propto \exp\left\{\left(-\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 - \frac{\tau_\beta}{2}\right) \beta^2 - 2\beta \frac{\tau}{2} \sum_{i=1}^n \gamma^{x_i}(y_i - \alpha) - \frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha)^2\right\} \\ &\propto \exp\left\{\left(-\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 - \frac{\tau_\beta}{2}\right) \left[\beta^2 - 2\beta \frac{\tau}{2} \left(-\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 - \frac{\tau_\beta}{2}\right)^{-1} \sum_{i=1}^n \gamma^{x_i}(y_i - \alpha) - \frac{\tau}{2} \left(-\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 - \frac{\tau_\beta}{2}\right)^{-1} \sum_{i=1}^n (y_i - \alpha)^2\right]\right\} \end{aligned}$$

Άρα τελικά η πλήρως δεσμευμένη κατανομή της β είναι:

$$\begin{aligned} \pi(\beta/\alpha, \gamma, \tau, data) &\propto \\ &\exp\left\{-\left(\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 + \frac{\tau_\beta}{2}\right) \left(\beta - \frac{\tau}{2} \left(-\frac{\tau}{2} \sum_{i=1}^n (\gamma^{x_i})^2 - \frac{\tau_\beta}{2}\right)^{-1} \sum_{i=1}^n \gamma^{x_i}(y_i - \alpha)\right)^2\right\} \end{aligned}$$

$$\pi(\beta/\alpha, \gamma, \tau, data) \sim \mathcal{N}\left(\frac{-\tau \sum_{i=1}^n \gamma^{x_i}(y_i - \alpha)}{\tau \sum_{i=1}^n (\gamma^{x_i})^2 + \tau_\beta}, \frac{1}{\tau \sum_{i=1}^n (\gamma^{x_i})^2 + \tau_\beta}\right)$$

Εύρεση πλήρως δεσμευμένης κατανομής της γ :

$$\pi(\gamma/\alpha, \beta, \tau, data) \propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2\right\}$$

Εύρεση πλήρως δεσμευμένης κατανομής της τ :

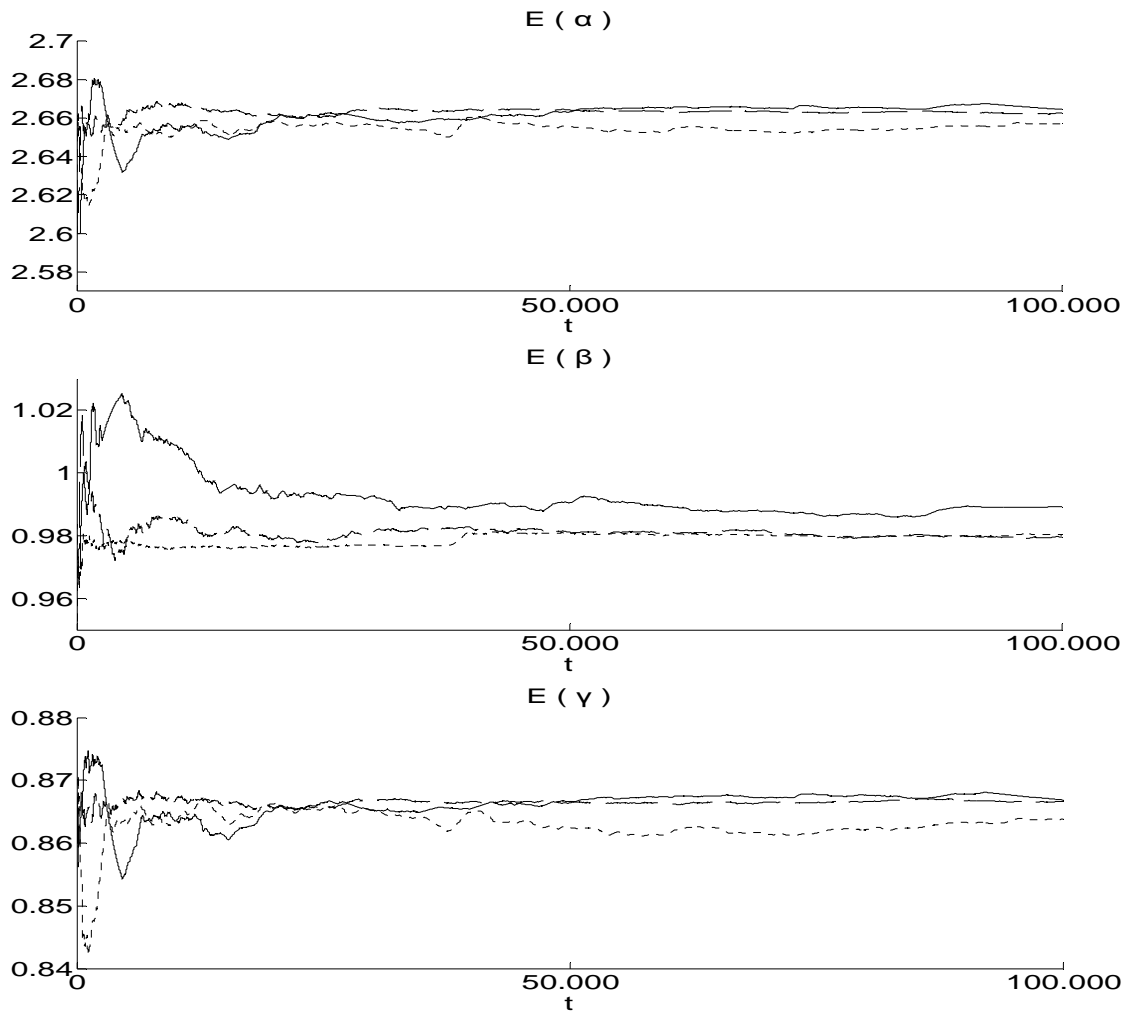
$$\begin{aligned} \pi(\tau/\alpha, \beta, \gamma, data) &\propto \tau^{\frac{n}{2}+k-1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2\right\} \exp\{-\tau k\} \\ &\propto \tau^{\frac{n}{2}+k-1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2 - \tau k\right\} \\ &\propto \tau^{\frac{n}{2}+k-1} \exp\left\{-\tau \left(\frac{\sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2}{2} + k\right)\right\} \end{aligned}$$

Άρα τελικά η πλήρως δεσμευμένη κατανομή της τ είναι:

$$\pi(\tau/\alpha, \beta, \gamma, data) \sim \mathcal{G}\left(\frac{n}{2} + k, \frac{\sum_{i=1}^n (y_i - \alpha + \beta\gamma^{x_i})^2}{2} + k\right).$$

Πίνακας 3.5 Μέση τιμή και τυπικά σφάλματα των εργοδικών μέσων των παραμέτρων α, β και γ για το πρόβλημα Dugongs.

1^η Μέθοδος	α	β	γ
Εκτίμηση μέσων	2.6634	0.9800	0.8668
Τυπική απόκλιση	0.0186	0.0181	0.0071
2^η Μέθοδος			
Εκτίμηση μέσων	2.6629	0.9802	0.8667
Τυπική απόκλιση	0.0017	0.0018	0.0007
3^η Μέθοδος			
Εκτίμηση μέσων	2.6628	0.9803	0.8667
Τυπική απόκλιση	0.0050	0.0015	0.0020



Σχήμα 3.6 Σύγκλιση εργοδικών μέσων των παραμέτρων α , β και γ που προκύπτουν από την πρώτη μέθοδο (—), τη δεύτερη μέθοδο (---) και την τρίτη μέθοδο (.....)

Οι παράμετροι α , β , τ παίρνουν νέα τιμή μέσω των πλήρως δεσμευμένων κατανομών τους $\pi(\alpha/\beta, \gamma, \tau, data)$, $\pi(\beta/\alpha, \gamma, \tau, data)$ και $\pi(\tau/\alpha, \beta, \gamma, data)$ αντίστοιχα ενώ για την παράμετρο γ θα κάνουμε χρήση ενός βήματος MH με κατανομή πρότασης την ομοιόμορφη $\mathcal{U}(0, 1)$ καθώς το γ παίρνει τιμές στο διάστημα $(0, 1)$ και με κατανομή-στόχο την πλήρως δεσμευμένη κατανομή της που είναι η $\pi(\gamma/\alpha, \beta, \tau, data)$.

Στο Σχήμα 3.7 απεικονίζεται η σύγκλιση των εργοδικών μέσων των παραμέτρων α , β και γ .

Για να εξετάσουμε ποια από τις τρεις μεθόδους που χρησιμοποιήσαμε δίνει εκτιμητές με μικρότερη διασπορά δημιουργήσαμε 500 ανεξάρτητες αλυσίδες των 100000

παρατηρήσεων η κάθε μια, από κάθε μέθοδο ξεχωριστά και εκτιμήσαμε τις μέσες τιμές των εργοδικών μέσων των παραμέτρων καθώς και τις αντίστοιχες τυπικές αποκλίσεις.

Παρατηρούμε ότι και οι τρεις μέθοδοι που χρησιμοποιήσαμε μας έδωσαν αντίστοιχες εκτιμήσεις για τις παραμέτρους α , β και γ του μοντέλου μας. Η δεύτερη μέθοδος (που έχουμε ολοκληρώσει την παράμετρο τ) δίνει εκτιμητές με τα μικρότερα τυπικά σφάλματα.

4 Εκτίμηση Διασποράς

4.1 Μέθοδοι Εκτίμησης Διασποράς

Υποθέτουμε ότι $X = \{X_1, X_2, X_3, \dots\}$ είναι μια απεριοδική (aperiodic), μη διαχωρίσιμη (irreducible), θετικά Harris επαναληπτική (positive Harris recurrent) αλυσίδα Markov με χώρο καταστάσεων \mathcal{X} και στάσιμη κατανομή π . Από το Εργοδικό θεώρημα έχουμε ότι

$$\bar{h}_n = \frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow E_\pi(h) \text{ καθώς το } n \rightarrow \infty.$$

Οι MCMC μέθοδοι κατασκευάζουν αλυσίδες Markov X που ικανοποιούν τις παραπάνω συνθήκες με αποτέλεσμα να ισχύει το εργοδικό θεώρημα και παίρνοντας ένα πεπερασμένο αριθμό n παρατηρήσεων, ο εκτιμητής \bar{h}_n συγκλίνει στην προς εκτίμηση μέση τιμή $E_\pi(h)$. Επιπλέον για τον εκτιμητή \bar{h}_n ισχύει το παρακάτω θεώρημα.

ΘΕΩΡΗΜΑ 4.1: Υποθέτουμε ότι $X = \{X_1, X_2, X_3, \dots\}$ είναι μια απεριοδική, μη διαχωρίσιμη, θετικά Harris επαναληπτική αλυσίδα Markov με στάσιμη κατανομή π . Εάν επιπλέον η X είναι γεωμετρικά εργοδική και $E_\pi|h|^{2+\varepsilon} < \infty$ για κάποιο $\varepsilon > 0$, τότε

$$\sqrt{n} \{\bar{h}_n - E_\pi(h)\} \xrightarrow{d} \mathcal{N}(0, \sigma_h^2) \quad (4.1)$$

όπου

$$\sigma_h^2 = \text{Var}_\pi[h(X_0)] + 2 \sum_{i=1}^{\infty} \text{Cov}_\pi[h(X_0), h(X_i)].$$

Παρατήρηση 4.1: Οι Roberts and Rosenthal (1997) έχουν δείξει ότι εάν η X είναι αναστρέψιμη ισχύει το παραπάνω θεώρημα και στην περίπτωση που έχουμε πεπερασμένη δεύτερη ροπή δηλ. $E_\pi(h)^2 < \infty$.

Η εκτίμηση της ασυμπτωτικής διασποράς σ_h^2 δεν είναι τυπική διαδικασία εξαιτίας της εξάρτησης των X_i . Για το σκοπό αυτό έχουν προταθεί διάφορες μέθοδοι στη βιβλιογραφία. Στα πλαίσια αυτού του κεφαλαίου θα ασχοληθούμε με τις σημαντικότερες από αυτές.

4.1.1 Spectral Estimation

Σε αυτή την ενότητα θα ορίσουμε μια ομάδα από εκτιμητές της σ_h^2 και θα αναφέρουμε συνθήκες που εγγυώνται τη συνέπεια τους. Πρώτα από όλα όμως θα ορίσουμε την αυτοσυσχέτιση με χρονική υστέρηση t ως

$$\gamma_t = \gamma_{-t} = E_\pi[Y_n Y_{n+t}]$$

όπου $Y_i := h(X_i) - E_\pi h$ για $i = 1, 2, 3, \dots$

Ένας λογικός εκτιμητής της αυτοσυσχέτισης με χρονική υστέρηση t , είναι η εμπειρική αυτοσυσχέτιση

$$\hat{\gamma}_{n,t} = \hat{\gamma}_{n,-t} = \frac{1}{n} \sum_{i=1}^{n-t} [h(X_i) - \hat{\mu}_n][h(X_{i+t}) - \hat{\mu}_n] \quad \text{για } t = 0, 1, 2, \dots, n-1$$

με

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

Ένας όχι και τόσο καλός εκτιμητής $\hat{\sigma}^2$ του σ_h^2 δίνεται από το άθροισμα των $\hat{\gamma}_{n,t}$ και λέγεται *Window estimator*

$$\hat{\sigma}_w^2 = \hat{\gamma}_{n,0} + 2 \sum_{t=1}^{n-1} \hat{\gamma}_{n,t}.$$

Ένας άλλος εκτιμητής της σ_h^2 είναι ο Spectral εκτιμητής που δίνεται από τον τύπο

$$\hat{\sigma}_s^2 = \sum_{t=-(b_n-1)}^{b_n-1} w_{n,t} \hat{\gamma}_{n,t}$$

όπου $w_{n,t}$ (lag window) είναι μια συνάρτηση βάρους και b_n είναι το σημείο όπου από εκεί και πέρα θεωρούμε όλες τις αυτοσυσχετίσεις μηδέν.

Υπόθεση 4.1: Η lag window $w_{n,t}$ είναι μια άρτια συνάρτηση που ορίζεται στους ακέραιους αριθμούς ως εξής

$$\begin{aligned} |w_{n,t}| &\leq 1 \text{ για όλα τα } n, t \\ w_{n,0} &= 1 \text{ για όλα τα } n \\ w_{n,t} &= 0 \text{ για όλα τα } |t| \geq b_n \end{aligned}$$

Υπόθεση 4.2: Υποθέτουμε ότι το b_n είναι μια ακέραια ακολουθία τέτοια ώστε $b_n \rightarrow \infty$ και $n/b_n \rightarrow \infty$ καθώς το $n \rightarrow \infty$ όπου b_n και n/b_n είναι μονότονα μη φθίνουσες ακολουθίες.

Υπόθεση 4.3: Υπάρχει μια συνάρτηση $s: \mathcal{X} \rightarrow [0,1]$ και ένα μέτρο πιθανότητας Q τέτοιο ώστε

$$P(x, \cdot) \geq s(x)Q(\cdot) \text{ για όλα τα } x \in \mathcal{X}.$$

Θεώρημα 4.2: Υποθέτουμε ότι X είναι μια γεωμετρικά εργοδική αλυσίδα με στάσιμη κατανομή π και $h: \mathcal{X} \rightarrow \mathbb{R}$ είναι μια Borel συνάρτηση με $E_\pi |h|^{4+\delta+\epsilon} < \infty$ για κάποιο $\delta > 0$ και $\epsilon > 0$. Έστω ότι οι υποθέσεις 4.1, 4.2 και 4.3 ισχύουν. Ορίζουμε

$$\Delta_1 w_n(k) = w_n(k-1) - w_n(k)$$

και

$$\Delta_2 w_n(k) = w_n(k-1) - 2w_n(k) + w_n(k+1).$$

Επίσης υποθέτουμε ότι (α) $b_n n^{-1} \sum_{k=1}^{b_n} k |\Delta_1 w_n(k)| \rightarrow 0$ καθώς το $n \rightarrow \infty$, (β) υπάρχει μια σταθερά $c \geq 1$ τέτοια ώστε $\sum_n (b_n/n)^c < \infty$, (γ) $b_n n^{-1} \log n \rightarrow 0$ καθώς το $n \rightarrow \infty$, (δ)

$$b_n n^{2a} (\log n)^n \left(\sum_{k=1}^{b_n} |\Delta_2 w_n(k)| \right)^2 \rightarrow 0 \text{ καθώς το } n \rightarrow \infty,$$

και

$$n^{2a} (\log n)^2 \sum_{k=1}^{b_n} |\Delta_2 w_n(k)| \rightarrow 0 \text{ καθώς το } n \rightarrow \infty,$$

όπου $a = 1/(4 + \delta)$ και (ε) $b_n^{-1} n^{2a} \log n \rightarrow 0$ καθώς το $n \rightarrow \infty$. Τότε για οποιαδήποτε στάσιμη κατανομή π , $\hat{\sigma}_s^2 \rightarrow \sigma_h^2$ καθώς το $n \rightarrow \infty$.

Ο Anderson (1994) πρότεινε μια σειρά lag window συναρτήσεων που ικανοποιούν τις υποθέσεις για τα $w_{n,t}$ κάποιες από τις οποίες δίνονται παρακάτω με $b_n = \lfloor n^\nu \rfloor$ και $0 < \nu < 1$.

- i. *Simple truncation* : $w_{n,t} = I(|t| < b_n)$
- ii. *Blackman-Tukey* : $w_{n,t} = [1 - 2\alpha + 2\alpha \cos(\pi|t|/b_n)] I(|t| < b_n)$ με $\alpha > 0$.
Όταν το $\alpha = 1/4$ τότε έχουμε την *Tukey-Hanning window*.
- iii. *Parzen* : $w_{n,t} = \left[1 - \frac{|t|^q}{b_n^q} \right] I(|t| < b_n)$ για $q \in \mathbb{Z}_+ = \{1, 2, 3, \dots\}$
Όταν $q = 1$ αυτή ονομάζεται *modified Bartlett window*.
- iv. *Scale-parameter modified Bartlett*: $w_{n,t} = \left[1 - \frac{\lambda|t|}{b_n} \right] I(|t| < b_n)$ όπου $\lambda \neq 1$ θετική σταθερά.

4.1.2 Batch means (BM)

Μια εναλλακτική μέθοδος για την εκτίμηση της διασποράς είναι η μέθοδος Batch means. Υποθέτουμε πως έχουμε μια αλυσίδα Markov $(X_n)_{n \in \mathbb{Z}_+}$. Διαιρούμε την αλυσίδα που αποτελείται από n παρατηρήσεις σε a κομμάτια που το καθένα περιέχει b παρατηρήσεις ($n = a \cdot b$) και βρίσκουμε τους μέσους κάθε κομματιού

$$\bar{Y}_k = \frac{\sum_{i=(k-1)b+1}^{kb} h(X_i)}{b}.$$

Τα \bar{Y}_k ασυμπτωτικά είναι ανεξάρτητα και ισόνομα και είναι η Monte Carlo εκτίμηση της $E_\pi(h)$ του k -οστού κομματιού. Έτσι εκτιμάται η διασπορά από την ποσότητα

$$\hat{\sigma}_{BM}^2 = \frac{b}{a-1} \sum_{k=1}^a (\bar{Y}_k - \hat{\mu})^2, \text{ όπου } \hat{\mu} \text{ ο συνολικός μέσος.}$$

Αυτός είναι ένας ασυνεπής εκτιμητής του σ_h^2 όμως έχει αποδειχθεί ότι αν ορίσουμε $a = b = \sqrt{n}$, τότε $\hat{\sigma}_{BM}^2 \rightarrow \sigma_h^2$ με πιθανότητα 1 καθώς το $n \rightarrow \infty$. (Glynn and Iglehart 1990 και Glynn and Whitt 1991)

4.1.3 Overlapping batch means (OBM)

Ο OBM εκτιμητής είναι μια γενίκευση του BM αλλά προκύπτει και από το modified Bartlett lag window εκτιμητή εκτός από κάποιους τελευταίους όρους. (Flegal and Jones (2010))

Η αλυσίδα Markov χωρίζεται σε $n - b + 1$ κομμάτια μεγέθους b παρατηρήσεων το κάθε ένα, συνήθως $b = \sqrt{n}$. Η X_k είναι η πρώτη παρατήρηση του k -οστού κομματιού, το οποίο αποτελείται από το σύνολο των διαδοχικών παρατηρήσεων (X_k, \dots, X_{k+b-1}) . Για το k -ιοστό κομμάτι ορίζουμε

$$\bar{Y}_k = \frac{1}{b} \sum_{i=1}^b h(X_{k+i-1})$$

και ο OBM εκτιμητής του σ_h^2 ορίζεται από

$$\hat{\sigma}_{OBM}^2 = \frac{nb}{(n-b)(n-b+1)} \sum_{k=1}^{n-b+1} (\bar{Y}_k - \hat{\mu})^2.$$

Θεώρημα 4.3: Υποθέτουμε ότι η X είναι μια γεωμετρικά εργοδική αλυσίδα Markov με στάσιμη κατανομή π και $h: X \rightarrow \mathbb{R}$ είναι μια συνάρτηση Borel με $E_\pi|h|^{2+\delta+\epsilon} < \infty$ για κάποιο $\delta > 0$ και $\epsilon > 0$. Θεωρούμε ότι ισχύουν οι Υποθέσεις 4.2 και 4.3. Επίσης, υποθέτουμε ότι (α) υπάρχει μια σταθερά $c \geq 1$ τέτοια ώστε $\sum_n (b_n/n)^c < \infty$, (β) $b_n n^{-1} \log n \rightarrow 0$ καθώς το $n \rightarrow \infty$, (γ) $n^{2a}(\log n)^3/b_n \rightarrow 0$ καθώς το $n \rightarrow \infty$ όπου το $a = 1/(2 + \delta)$, (δ) υπάρχει ένας ακέραιος αριθμός n_0 και μια σταθερά c_1 τέτοια ώστε για όλα τα $n \geq n_0$ έχουμε $\log n/b_n \leq c_1$, και (ε) $b_n^2 n^{-2} \log \log n \rightarrow 0$ και $b_n^4 n^{-3} \log \log n \rightarrow 0$ καθώς $n \rightarrow \infty$, τότε καθώς το $n \rightarrow \infty$, $\hat{\sigma}_{OBM}^2 \rightarrow \sigma_h^2$ με πιθανότητα 1.

Στη συνέχεια χρησιμοποιούμε ένα απλό παράδειγμα για να εκτιμήσουμε τη διασπορά των εργοδικών μέσων χρησιμοποιώντας τους παραπάνω εκτιμητές.

4.2 Παράδειγμα

Έστω ότι τρέχουμε έναν αλγόριθμο MH με κατανομή-στόχο την τυπική κανονική κατανομή $\pi \sim \mathcal{N}(0, 1)$ και με κατανομή πρότασης την $g \sim \mathcal{N}(0, \sigma^2)$. Παίρνουμε ένα δείγμα 10000 παρατηρήσεων για διάφορες τιμές του σ (π.χ. 2, 5 και 10). Εκτιμάμε την πρώτη και τη δεύτερη ροπή και με τη βοήθεια των μεθόδων που αναφέραμε προηγουμένως εκτιμάμε την ασυμπτωτική διασπορά τους. (βλέπε Πίνακα 4.1)

Πίνακας 4.1 Εκτίμηση ασυμπτωτικής διασποράς εργοδικών μέσων πρώτης και δεύτερης ροπής από ένα δείγμα 10000 παρατηρήσεων που προσομοιώθηκε μέσω ενός αλγόριθμου MH με κατανομή-στόχο $\pi \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$ και κατανομή πρότασης $g \sim \mathcal{N}(\mathbf{0}, \sigma^2)$.

	Εκτίμηση διασποράς 1 ^{ης} ροπής			Εκτίμηση διασποράς 2 ^{ης} ροπής		
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
	<i>Simple truncation</i>	1.9556	4.9540	14.3421	6.0970	12.1245
<i>Blackman -Tukey για $\delta = 1$</i>	1.8229	5.0216	13.8644	5.4318	13.0379	23.3005
<i>Blackman-Tukey για $\delta = 2$</i>	1.8451	5.0104	13.9441	5.5427	12.8857	23.1068
<i>Tukey Hanning Window για $\delta = 0$</i>	1.7898	5.0386	13.7451	5.2655	13.2663	23.5912
<i>Parzen για $q=2$</i>	1.8330	5.0626	14.3103	5.4214	12.8955	22.8747
<i>Modified Bartlett = Parzen για $q=1$</i>	1.7946	4.9870	13.1125	5.2220	12.9911	22.1851
<i>Scale-parameter Bartlett $\lambda=0.5$</i>	1.8751	4.9705	13.7273	5.6595	12.5578	22.1615
<i>Batch Mean</i>	1.8771	4.9835	12.0828	4.8159	13.8176	22.5375
<i>Overlapping Batch Mean</i>	1.7573	5.0596	13.2135	5.2918	12.7015	22.5214

Παρατηρούμε ότι όλες οι παραπάνω μέθοδοι δίνουν περίπου την ίδια εκτίμηση για την ασυμπτωτική διασπορά.

Παρά το γεγονός ότι οι μέθοδοι αυτές χρησιμοποιούνται ευρέως στη βιβλιογραφία, αντιμετωπίζουν κάποια προβλήματα. Ένα από τα σημαντικότερα προβλήματα τους είναι η ύπαρξη εξάρτησης μεταξύ των διαδοχικών παρατηρήσεων της αλυσίδας. Επίσης, σημαντικό πρόβλημα είναι ότι οι μέθοδοι αυτές τείνουν να είναι πιο αποτελεσματικές όταν η αλυσίδα είναι στάσιμη. Για την αποφυγή τέτοιων προβλημάτων χρησιμοποιείται η μέθοδος *Regenerative Simulation* (RS) η οποία δημιουργεί ένα εναλλακτικό ΚΟΘ που επιτρέπει την κατασκευή διαφορετικών ασυμπτωτικών διασπορών. Η μέθοδος αυτή δεν απαιτεί η αλυσίδα Markov να είναι στάσιμη. Με διάφορες τεχνικές, εντοπίζονται οι τυχαίοι χρόνοι στους οποίους η αλυσίδα αναγεννάται (regenerate). Τα κομμάτια της αλυσίδας μεταξύ των σημείων που γίνεται αναγέννηση (regeneration) είναι ανεξάρτητα. Οπότε, προσομοιώνεται ένας σταθερός αριθμός τέτοιων κομματιών και τελικά εκτιμάται η διασπορά από ανεξάρτητες παρατηρήσεις. Επίσης δεν χρειάζεται τα κομμάτια να

παράγονται από μια και μόνο διαδικασία αλλά μπορούν να παράγονται από διαφορετικές ανεξάρτητες διαδικασίες.

Στα πλαίσια αυτής της εργασίας δεν ασχοληθήκαμε με τη μέθοδο αυτή. Για περισσότερες πληροφορίες βλέπε Mykland et al. (1995), Robert (1995), Mengersen and Tweedie (1996), Jones and Hobert (2001), Brockwell and Kadane (2004).

5

Προσδιορισμός κατάλληλης κατανομής πρότασης στον αλγόριθμο Metropolis-Hastings

5.1 Ρυθμός αποδοχής (Acceptance rate)

Ο αλγόριθμος MH απαιτεί την επιλογή μιας κατανομής πρότασης και όπως είναι γνωστό κάποιες κατανομές πρότασης συμπεριφέρονται καλύτερα από άλλες. Στα πλαίσια αυτού του κεφαλαίου θα εστιάσουμε στον αλγόριθμο Random Walk Metropolis, στον οποίο οι προτεινόμενες τιμές έχουν τη μορφή $Y_{n+1} = Y_n + Z_n$ όπου τα Z_n είναι ανεξάρτητα και ισόνομα κατανομημένα από κάποια συμμετρική κατανομή με παράμετρο κλίμακας $\sigma > 0$ π.χ. $Z_i \sim \mathcal{N}(0, \sigma^2)$. Σε αυτή την περίπτωση η τιμή του σ παίζει καθοριστικό ρόλο για την ταχύτητα σύγκλισης του αλγόριθμου. Δίνοντας μια πολύ μικρή τιμή στο σ , τότε ο αλγόριθμος θα αποδέχεται όλες τις προτεινόμενες νέες τιμές αλλά η αλυσίδα θα κινείται πολύ αργά και δε θα έχουμε γρήγορη σύγκλιση. Με μια μεγάλη τιμή για το σ οι προτεινόμενες τιμές θα απορρίπτονται με αποτέλεσμα η αλυσίδα να παραμένει στο ίδιο σημείο για πολλές επαναλήψεις. Για να αποφύγουμε αυτές τις ακραίες περιπτώσεις χρειάζεται να αποδεχόμαστε την προτεινόμενη τιμή με ένα βέλτιστο ρυθμό. Αν ο ρυθμός αποδοχής είναι κοντά στο 0 τότε το σ είναι πολύ μεγάλο ενώ αν είναι κοντά στο 1 τότε είναι πολύ μικρό. Αν όμως βρίσκεται ανάμεσα στο 0 και 1 τότε θα έχουμε αποφύγει τις ακραίες τιμές και έτσι θα έχουμε μια καλή τιμή για το σ .

Οι πρώτοι που προσδιόρισαν μια συγκεκριμένη τιμή για το βέλτιστο ρυθμό αποδοχής ήταν οι Roberts, Gelman and Gilks (1997). Πιο συγκεκριμένα, θεώρησαν έναν RWM στον \mathbb{R}^d για πολύ ειδικές κατανομές-στόχου, της μορφής

$$\pi(x_1, x_2, \dots, x_d) = f(x_1)f(x_2) \dots f(x_d) \quad (5.1)$$

για κάποιες μονοδιάστατες κατανομές f . Έτσι η κατανομή-στόχος αποτελείται από ανεξάρτητες και ισόνομα κατανεμημένες συνιστώσες το οποίο όμως είναι μη ρεαλιστικό στις MCMC τεχνικές. Επίσης θεώρησαν ότι η κατανομή πρότασης είναι η $\mathcal{N}(x, \sigma^2 I_d)$. Για αυτή την περίπτωση απέδειξαν ότι καθώς το $d \rightarrow \infty$, ο βέλτιστος ρυθμός αποδοχής είναι ακριβώς ίσος με 0.234. Οι Roberts and Rosenthal (2001) απέδειξαν ότι στο μονοδιάστατο χώρο ο βέλτιστος ρυθμός αποδοχής είναι ίσος με 0.44.

Επίσης, οι Roberts and Rosenthal (2001) θώρησαν σαν κατανομή-στόχο την

$$\pi(x) = \prod_{i=1}^d C_i f(C_i x_i),$$

όπου $\{C_i\}$ είναι ανεξάρτητα και ισόνομα κατανεμημένα από κάποια κατανομή (η Σχέση (5.1) αντιστοιχεί στην περίπτωση όπου τα C_i είναι μια σταθερά). Απέδειξαν ότι και σε αυτή την περίπτωση ισχύουν τα αποτελέσματα των Roberts et al. (1997) για το βέλτιστο ρυθμό αποδοχής αν και στην πράξη όσο πιο ανομοιογενής είναι η κατανομή-στόχος, δηλαδή όσο μεγαλύτερη μεταβλητότητα έχουν τα C_i , τόσο πιο αργός είναι ο αλγόριθμος.

Προς την ίδια κατεύθυνση, η Bedard (2006,2007,2008) και οι Berard and Rosenthal (2008) θεώρησε την περίπτωση όπου η κατανομή-στόχος π αποτελείτε από ανεξάρτητες συνιστώσες με τελείως διαφορετικές παραμέτρους κλίμακας. Έτσι απέδειξαν ότι εάν κάθε συνιστώσα κυριαρχείται από το άθροισμα των άλλων τότε ο βέλτιστος ρυθμός αποδοχής παραμένει να είναι 0.234. Στην περίπτωση που μια συνιστώσα είναι συγκρίσιμη με το άθροισμα των υπολοίπων τότε ο βέλτιστος ρυθμός αποδοχής είναι γενικά μικρότερος από 0.234.

Οι Roberts and Tweedie (1996) και Roberts and Rosenthal (1998) μελέτησαν τον αλγόριθμο Metropolis-Adjusted Langevin ο οποίος είναι ένας αλγόριθμος παρόμοιος με το RWM με κατανομή-στόχο $\pi(x_1, x_2, \dots, x_d) = f(x_1)f(x_2) \dots f(x_d)$ αλλά με τη διαφορά ότι η κατανομή πρότασης είναι της μορφής $\mathcal{N}(X_n + \frac{\sigma_d^2}{2} \nabla \log \pi(X_n), \sigma_d^2 I_d)$. Για αυτή τη μορφή του αλγόριθμου απέδειξαν ότι ο βέλτιστος ρυθμός αποδοχής είναι ίσος με 0.574 αντί του 0.234 και ότι ο αλγόριθμος Metropolis-Adjusted Langevin συγκλίνει πιο γρήγορα από τον RWM. Ο υπολογισμός όμως της βάρμωσης (gradient) της π σε κάθε επανάληψη είναι δύσκολη και χρονοβόρα διαδικασία, πράγμα που κάνει τον αλγόριθμο RWM πιο δημοφιλή.

Παρακάτω παρουσιάζουμε μερικά παραδείγματα όπου φαίνεται ο τρόπος με τον οποίο επηρεάζεται η σύγκλιση της αλυσίδας ανάλογα με την παράμετρο κλίμακας της κατανομής πρότασης.

5.1.1 Μονοδιάστατη περίπτωση

Έστω ότι έχουμε για κατανομή-στόχο την τυπική κανονική κατανομή $\pi(x) \sim \mathcal{N}(0, 1)$ και κατανομή πρότασης τη $g(\cdot | x) \sim \mathcal{N}(x, \sigma^2)$, δηλαδή μια κανονική κατανομή με μέση τιμή την προηγούμενη κατάσταση της αλυσίδας και διασπορά σ^2 . Θέλουμε να επιλέξουμε κατάλληλο σ τέτοιο ώστε να έχουμε ρυθμό αποδοχής όσο πιο κοντά στη βέλτιστη τιμή 0.44 γίνεται (το οποίο είναι το βέλτιστο για μονοδιάστατες κατανομές πρότασης σύμφωνα με Roberts et al., 1997; Roberts and Rosenthal, 2001). Τρέχουμε 1000 ανεξάρτητες αλυσίδες 10000 προσομοιωμένων παρατηρήσεων η κάθε μια για διάφορες τιμές της τυπικής απόκλισης σ (π.χ. 0.5, 1.1, 1.5, 2, 2.4, 3, 5 και 10). Στον Πίνακα 5.1 απεικονίζονται ο μέσος ρυθμός αποδοχής καθώς επίσης και η μέση εκτίμηση και η τυπική απόκλιση της πρώτης και δεύτερης ροπής για τις διάφορες τιμές του σ .

Παρατηρούμε ότι για τυπική απόκλιση $\sigma = 2.4$ έχουμε καταφέρει να επιτύχουμε το βέλτιστο ρυθμό αποδοχής σύμφωνα με τη βιβλιογραφία. Επίσης παρατηρούμε ότι για αυτό το σ οι εκτιμητές έχουν τη μικρότερη ή πολύ κοντά στη μικρότερη διασπορά.

Πίνακας 5.1 Μέσος ρυθμός αποδοχής, μέση εκτίμηση και τυπική απόκλιση της πρώτης και δεύτερης ροπής.

σ	Ρυθμός αποδοχής	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής
0.5	0.84	-0.0022	0.0479	1.0019	0.0553
1.1	0.68	-0.0008	0.0287	1.0018	0.0341
1.5	0.59	-0.0011	0.0229	0.9994	0.0309
2.0	0.50	-0.0015	0.0219	0.9988	0.0304
2.4	0.44	0.0002	0.0207	1.0017	0.0305
3.0	0.37	-0.0008	0.0213	1.0028	0.0320
5.0	0.24	-0.0004	0.0249	1.0013	0.0386
10.0	0.13	-0.0006	0.0359	0.9970	0.0538

5.1.2 Πολυδιάστατη περίπτωση

Εφαρμόζουμε τον αλγόριθμο RWM με κατανομή-στόχο τη 10-διάστατη τυπική κανονική κατανομή $\mathcal{N}(0, I_{10})$ και κατανομή πρότασης τη 10-διάστατη κανονική κατανομή $\mathcal{N}(x, \sigma^2 I_{10})$ με x το 10-διάστατο διάνυσμα που δίνει την τρέχουσα κατάσταση. Θέλουμε να επιλέξουμε κατάλληλο σ^2 τέτοιο ώστε να έχουμε ρυθμό αποδοχής όσο πιο κοντά στο 0.234 γίνεται (το οποίο είναι το βέλτιστο για πολυδιάστατες κατανομές πρότασης σύμφωνα με Roberts et al. 1997, Roberts and Rosenthal 1998, 2001, Bedard 2006). Προσομοιώνουμε 1000 ανεξάρτητες αλυσίδες μήκους 10000 παρατηρήσεων η κάθε μία για διάφορες τιμές του σ^2 (π.χ. 0.1, 0.5, 0.65, 1.5, 2 και 2.5). Στον Πίνακα 5.2 απεικονίζονται ο μέσος ρυθμός αποδοχής καθώς επίσης και οι μέσες εκτιμήσεις και οι τυπικές αποκλίσεις της πρώτης και της δεύτερης ροπής για κάθε συνιστώσα της κατανομής-στόχου μας.

Πίνακας 5.2 Μέσος ρυθμός αποδοχής, μέσες εκτιμήσεις και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα.

$\sigma^2 = 0.1$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	-0.0015	0.0828	0.9975	0.0826	0.6279
x_2	-0.0028	0.0810	1.0037	0.0895	
x_3	0.0005	0.0805	1.0009	0.0855	
x_4	-0.0009	0.0842	0.9972	0.0830	
x_5	0.0054	0.0834	0.9963	0.0876	
x_6	-0.0017	0.0793	0.9953	0.0849	
x_7	-0.0005	0.0832	0.9979	0.0910	
x_8	-0.0030	0.0831	1.0001	0.0858	
x_9	0.0014	0.0819	0.9996	0.0836	
x_{10}	-0.0035	0.0839	0.9944	0.0849	

$\sigma^2 = 0.5$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	-0.0014	0.0570	1.0026	0.0633	0.2898
x_2	-0.0021	0.0580	1.0024	0.0620	
x_3	0.0000	0.0550	1.0006	0.0628	
x_4	-0.0019	0.0581	0.9998	0.0638	
x_5	0.0016	0.0558	1.0012	0.0617	
x_6	0.0002	0.0560	0.9988	0.0605	
x_7	0.0011	0.0565	1.0024	0.0653	
x_8	-0.0008	0.0578	1.0000	0.0627	
x_9	0.0006	0.0573	0.9993	0.0602	
x_{10}	-0.0006	0.0563	1.0006	0.0608	

$\sigma^2 = 0.65$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	-0.0027	0.0549	0.9986	0.0634	0.2315
x_2	0.0005	0.0586	0.9993	0.0616	
x_3	0.0022	0.0561	1.0022	0.0619	
x_4	0.0017	0.0537	1.0022	0.0640	
x_5	-0.0030	0.0567	0.9988	0.0621	
x_6	-0.0015	0.0560	0.9993	0.0605	
x_7	0.0005	0.0559	1.0012	0.0615	
x_8	-0.0002	0.0570	0.9996	0.0621	
x_9	-0.0040	0.0564	0.9987	0.0639	
x_{10}	-0.0043	0.0557	1.0011	0.0599	

$\sigma^2 = 1.5$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	-0.0017	0.0705	1.0022	0.0807	0.0816
x_2	0.0029	0.0687	0.9993	0.0822	
x_3	0.0003	0.0690	1.0013	0.0814	
x_4	-0.0019	0.0693	1.0022	0.0822	
x_5	0.0026	0.0711	1.0001	0.0812	
x_6	-0.0039	0.0677	1.0020	0.0792	
x_7	-0.0011	0.0708	1.0007	0.0796	
x_8	0.0002	0.0697	1.0020	0.0808	
x_9	0.0019	0.0716	0.9998	0.0812	
x_{10}	0.0040	0.0711	0.9975	0.0786	

$\sigma^2 = 2$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	0.0028	0.0873	0.9998	0.1027	0.0494
x_2	0.0005	0.0867	1.0035	0.1023	
x_3	0.0010	0.0845	1.0028	0.1059	
x_4	-0.0036	0.0841	0.9955	0.0993	
x_5	-0.0017	0.0849	1.0008	0.1005	
x_6	0.0001	0.0875	0.9928	0.0976	
x_7	0.0014	0.0856	1.0048	0.1042	
x_8	0.0018	0.0862	0.9975	0.1022	
x_9	0.0040	0.0846	1.0040	0.0991	
x_{10}	-0.0025	0.0837	1.0008	0.1006	

$\sigma^2 = 2.5$	Μέση εκτίμηση 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέση εκτίμηση 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής	Ρυθμός αποδοχής
x_1	0.0028	0.1064	0.9932	0.1290	0.0315
x_2	-0.0010	0.1028	1.0025	0.1233	
x_3	0.0005	0.1036	1.0012	0.1242	
x_4	0.0006	0.1073	1.0029	0.1249	
x_5	-0.0039	0.1068	0.9969	0.1283	
x_6	0.0016	0.1046	1.0066	0.1262	
x_7	0.0041	0.1048	1.0033	0.1336	
x_8	-0.0001	0.1040	1.0028	0.1269	
x_9	-0.0024	0.1097	0.9982	0.1270	
x_{10}	-0.0002	0.1038	1.0073	0.1237	

Παρατηρούμε ότι για διασπορά $\sigma^2 = 0.65$ έχουμε καταφέρει να φτάσουμε πολύ κοντά στο βέλτιστο ρυθμό αποδοχής σύμφωνα με τη βιβλιογραφία.

Όμως αυτό που κάναμε σε αυτά τα παραδείγματα δεν μπορούμε να το εφαρμόζουμε γενικά. Δεν είναι εύκολο με "trial and error" τεχνικές σε περίπλοκα προβλήματα να βρούμε ένα ικανοποιητικό σ . Για αυτό έχουν αναπτυχθεί στη βιβλιογραφία διάφορες τεχνικές για τον εντοπισμό ενός κατάλληλου σ για την κατανομή πρότασης. Στη συνέχεια θα αναφερθούμε σε κάποιες από αυτές τις τεχνικές.

5.2 Adaptive MCMC

Οι Adaptive MCMC τεχνικές επιλύουν το πρόβλημα της εύρεσης κατάλληλων τιμών των παραμέτρων της κατανομής πρότασης του αλγόριθμου με σκοπό την πιο γρήγορη σύγκλιση της αλυσίδας. Έτσι κατά τη διάρκεια της προσομοίωσης χρησιμοποιούνται πληροφορίες από την ιστορία της διαδικασίας ώστε να ενημερώνονται αυτόματα οι παράμετροι της κατανομής πρότασης με καταλληλότερες τιμές.

Είναι όμως γνωστό ότι οι Adaptive MCMC τεχνικές δε διατηρούν πάντα τη στασιμότητα της π (Rosenthal (2004), Roberts και Rosenthal (2005)). Παρόλα αυτά, αν οι ενημερώσεις σχεδιαστούν έτσι ώστε να ικανοποιούν κάποιες συγκεκριμένες συνθήκες, τότε εγγυώνται τη στασιμότητα και επιταχύνεται σημαντικά η σύγκλιση της αλυσίδας.

Στη συνέχεια θα παρουσιάσουμε μια σειρά από προσαρμοστικούς (adaptive) αλγόριθμους σε μονοδιάστατο αλλά και πολυδιάστατο χώρο καταστάσεων και θα δούμε ότι πράγματι συγκλίνουν και συχνά έχουν καλύτερες ιδιότητες μίξης από αντίστοιχους μη προσαρμοστικούς αλγόριθμους.

5.2.1 Αλγόριθμος Adaptive Metropolis (AM)

Ο αλγόριθμος Adaptive Metropolis (AM) εφαρμόστηκε από τους Haario et al. (2001) και ήταν ο πρώτος adaptive MCMC αλγόριθμος. Οι Roberts and Rosenthal (2006a)

εισήγαγαν έναν ελάχιστα διαφορετικό αλγόριθμο από αυτόν των Haario et al. (2001) στο d -διάστατο χώρο με μια κατανομή-στόχο π . Πρότειναν την εφαρμογή ενός Metropolis αλγόριθμου με κατανομή πρότασης στην n -οστή επανάληψη την

$$Q_n(x, \cdot) = \mathcal{N}(x, 0.1^2 I_d / d) \text{ για } n \leq 2d$$

και

$$Q_n(x, \cdot) = \begin{cases} (1 - \beta) \mathcal{N}\left(x, \frac{(2.38)^2 \hat{\Sigma}_n}{d}\right) + \beta \mathcal{N}\left(x, \frac{0.1^2}{d} I_d\right), & \text{με } \hat{\Sigma}_n \text{ θετικά ορισμένο} \\ \mathcal{N}(x, 0.1^2 I_d / d), & \text{με } \hat{\Sigma}_n \text{ μη θετικά ορισμένο} \end{cases} \quad (5.2)$$

για $n > 2d$, όπου β μια μικρή θετική σταθερά $0 < \beta < 1$ π.χ. $\beta = 0.05$ και $\hat{\Sigma}_n$ η εκτίμηση του πίνακα διασπορών-συνδιασπορών της κατανομής-στόχο που βρίσκεται από τον τύπο

$$\hat{\Sigma}_n = \frac{1}{n} \left(\sum_{i=0}^n X_i X_i^T - (n+1) \bar{X}_n \bar{X}_n^T \right)$$

όπου $\bar{X}_n = \frac{1}{n+1} \sum_{i=0}^n X_i$ η εκτίμηση της μέσης τιμής μέχρι εκείνη τη χρονική στιγμή. Ένας αναδρομικός τύπος για την εκτίμηση του πίνακα διασπορών-συνδιασπορών $\hat{\Sigma}_n$ είναι ο

$$\hat{\Sigma}_n = \frac{n-1}{n} \hat{\Sigma}_{n-1} + \frac{1}{n+1} (X_n - \bar{X}_{n-1})(X_n - \bar{X}_{n-1})^T.$$

Οι Haario et al. (2001) απέδειξαν ότι η διαδικασία AM διατηρεί τις εργοδικές της ιδιότητες υποθέτοντας ότι η κατανομή-στόχος είναι άνω φραγμένη και έχει φραγμένο πεδίο ορισμού.

Όπως παρατηρούμε παραπάνω στον τύπο της κατανομής πρότασης ενώ η μέση τιμή εξαρτάται μόνο από την προηγούμενη κατάσταση, ο πίνακας διασπορών-συνδιασπορών υπολογίζεται χρησιμοποιώντας όλες τις προηγούμενες καταστάσεις με αποτέλεσμα πλέον να χαθεί η μαρκοβιανή ιδιότητα της αλυσίδας μας. Έχει δειχθεί όμως ότι η ασυμπτωτική

εξάρτηση ανάμεσα στα στοιχεία της αλυσίδας δεν είναι τόσο ισχυρή με αποτέλεσμα να ισχύουν τα θεωρήματα των μεγάλων αριθμών (McLeish 1975 and Haario 2001).

5.2.1.1 Μονοδιάστατη περίπτωση

Εφαρμόζουμε τον AM αλγόριθμο στην περίπτωση που έχουμε μονοδιάστατο χώρο καταστάσεων με κατανομή-στόχο $\mathcal{N}(0, 1)$ και κατανομή πρότασης την (5.2) για $d = 1$. Τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 20000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων της πρώτης και δεύτερης ροπής, τις τυπικές αποκλίσεις τους καθώς και το μέσο ρυθμό αποδοχής.

Πίνακας 5.3 Μέσος ρυθμός αποδοχής, μέση εκτίμηση εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον AM αλγόριθμο στο μονοδιάστατο χώρο καταστάσεων.

$d = 1$	Μέσος	Μέση τιμή	Τυπική	Μέση τιμή	Τυπική
$\pi \sim \mathcal{N}(0, 1)$	ρυθμός	1 ^{ης} ροπής	απόκλιση	2 ^{ης} ροπής	απόκλιση
N=20000	αποδοχής		1 ^{ης} ροπής		2 ^{ης} ροπής
	0.4610	0.0006	0.0153	1.0002	0.0208

Παρατηρούμε ότι ο AM αλγόριθμος έχει καταφέρει να φτάσει πολύ κοντά στο βέλτιστο ρυθμό αποδοχής.

5.2.1.2 Πολυδιάστατη περίπτωση

Εφαρμόζουμε και πάλι τον αλγόριθμο AM στο 10-διάστατο χώρο καταστάσεων για να προσομοιάσουμε από την κατανομή-στόχο $\mathcal{N}_{10}(0, I_{10})$. Για κατανομή πρότασης χρησιμοποιούμε την $Q_n(x, \cdot)$ που δίνεται στην Σχέση (5.2) για $d = 10$ και τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 100000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση

τιμή των εργοδικών μέσων της πρώτης και δεύτερης ροπής, τις τυπικές αποκλίσεις τους για κάθε συνιστώσα καθώς και τον μέσο ρυθμό αποδοχής.

Πίνακας 5.4 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον AM αλγόριθμο στο 10-διάστατο χώρο.

$d = 10$	Μέσος	Μέση τιμή	Τυπική	Μέση τιμή	Τυπική
$\pi \sim \mathcal{N}(0, I_{10})$	ρυθμός	1 ^{ης} ροπής	απόκλιση	2 ^{ης} ροπής	απόκλιση
N=100000	αποδοχής		1 ^{ης} ροπής		2 ^{ης} ροπής
x_1	0.2921	-0.0016	0.0190	0.9946	0.0189
x_2		-0.0008	0.0191	0.9937	0.0186
x_3		0.0005	0.0192	0.9931	0.0199
x_4		0.0003	0.0182	0.9936	0.0192
x_5		0.0002	0.0183	0.9937	0.0198
x_6		-0.0003	0.0183	0.9940	0.0198
x_7		0.0005	0.0183	0.9926	0.0189
x_8		0.0006	0.0180	0.9930	0.0197
x_9		0.0013	0.0190	0.9945	0.0196
x_{10}		-0.0001	0.0189	0.9958	0.0193

Στον 10-διάστατο χώρο καταστάσεων ο αλγόριθμος AM μας έδωσε μέσο ρυθμό αποδοχής 0.2921.

5.2.2 Adaptive Metropolis-Within-Gibbs

Στη μέθοδο Adaptive Metropolis-Within-Gibbs η προτεινόμενη τιμή της i -οστής παραμέτρου προσομοιώνεται από την κατανομή πρότασης $\mathcal{N}(x_i, e^{2ls_i})$ όπου ls_i είναι ο λογάριθμος της τυπικής απόκλισης. Κάθε ls_i πρέπει να επιλέγεται έτσι ώστε ο ρυθμός αποδοχής να προσεγγίζει το 0.44. Όμως αυτό είναι δύσκολο να επιτευχθεί με δοκιμή.

Ένας τρόπος να προσαρμόσουμε κατάλληλα τις τιμές των ls_i είναι να χωρίσουμε τη διαδικασία της προσομοίωσης της αλυσίδας σε κομμάτια μήκους m παρατηρήσεων (π.χ. $m = 100$) το καθένα. Θεωρούμε αρχικά όλες τις τιμές των ls_i ίσες με μηδέν ($ls_i = 0$). Μετά από το n -οστό κομμάτι ενημερώνουμε κάθε ls_i προσθέτοντας ή αφαιρώντας ένα προσαρμοσμένο μέγεθος $\delta(n) = \min(0.01, n^{-1/2})$. Αυξάνουμε το ls_i με το $\delta(n)$ αν ο ρυθμός αποδοχής της i παραμέτρου είναι μεγαλύτερος από 0.44 στο n -οστό κομμάτι ή το μειώνουμε κατά $\delta(n)$ αν είναι μικρότερος.

5.2.2.1 Μονοδιάστατη περίπτωση

Εφαρμόζουμε τον Adaptive Metropolis-within-Gibbs αλγόριθμο στο μονοδιάστατο χώρο καταστάσεων για να προσομοιώσουμε από την κατανομή-στόχο $\mathcal{N}(0, 1)$. Για κατανομή πρότασης χρησιμοποιούμε τη $\mathcal{N}(x, e^{2ls})$ και τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 20000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων πρώτης και δεύτερης ροπής, τις τυπικές αποκλίσεις τους καθώς και το μέσο ρυθμό αποδοχής.

Πίνακας 5.5 Μέσος ρυθμός αποδοχής, μέση εκτίμηση εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον Adaptive Metropolis-within-Gibbs αλγόριθμο στο μονοδιάστατο χώρο καταστάσεων.

$d = 1$	Μέσος	Μέση τιμή	Τυπική	Μέση τιμή	Τυπική
$\pi \sim \mathcal{N}(0, 1)$	ρυθμός	1 ^{ης} ροπής	απόκλιση	2 ^{ης} ροπής	απόκλιση
N=20000	αποδοχής		1 ^{ης} ροπής		2 ^{ης} ροπής
	0.4446	-0.0003	0.0161	1.0005	0.0218

Παρατηρούμε ότι με τον αλγόριθμο Adaptive Metropolis-within-Gibbs έχουμε πετύχει ακριβώς τον βέλτιστο ρυθμό αποδοχής.

5.2.2.2 Πολυδιάστατη περίπτωση

Εφαρμόζουμε ακόμη μια φορά τον αλγόριθμο Adaptive Metropolis within Gibbs στο 10-διάστατο χώρο καταστάσεων για να προσομοιώσουμε από την πολυδιάστατη κανονική κατανομή $\mathcal{N}(0, I_{10})$. Για κατανομή πρότασης χρησιμοποιούμε τη $\mathcal{N}(x_i, e^{2ls_i})$ για κάθε μια συνιστώσα και δημιουργούμε 500 ανεξάρτητες αλυσίδες μήκους 100000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων πρώτης και δεύτερης ροπής, τις τυπικές αποκλίσεις τους για κάθε συνιστώσα καθώς και τον μέσο ρυθμό αποδοχής.

Πίνακας 5.6 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον Adaptive Metropolis-within-Gibbs αλγόριθμο στον 10-διάστατο χώρο καταστάσεων.

$d = 10$ $\pi \sim \mathcal{N}(0, I_{10})$ $N=100000$	Μέσος ρυθμός αποδοχής	Μέσος όρος 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέσος όρος 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής
x_1	0.4428	-0.00017	0.0069	1.0000	0.0094
x_2	0.4447	0.00001	0.0068	0.9993	0.0095
x_3	0.4367	0.00016	0.0071	0.9999	0.0102
x_4	0.4397	0.00032	0.0063	1.0008	0.0091
x_5	0.4452	-0.00006	0.0068	0.9999	0.0101
x_6	0.4402	-0.00010	0.0065	0.9999	0.0097
x_7	0.4403	0.00024	0.0066	0.9996	0.0103
x_8	0.4406	-0.00055	0.0067	1.0005	0.0098
x_9	0.4381	0.00015	0.0068	0.9999	0.0095
x_{10}	0.4422	-0.00001	0.0066	1.0000	0.0096

Παρατηρούμε ότι και στον 10-διάστατο χώρο καταστάσεων, όπως και στο μονοδιάστατο, καταφέραμε να πετύχουμε ρυθμό αποδοχής πολύ κοντά στο βέλτιστο για κάθε μία συνιστώσα ξεχωριστά.

5.2.3 Single Component adaptive Metropolis algorithm (SCAM)

Ο αλγόριθμος SCAM είναι ένας αλγόριθμος MH μιας συνιστώσας όπου η μονοδιάστατη κατανομή πρότασης ενημερώνεται κατά τη διάρκεια της προσομοίωσης όπως και στον AM αλγόριθμο. Έστω ότι βρισκόμαστε στο d -διάστατο ευκλείδειο χώρο και θέλουμε να προσομοιώσουμε από την κατανομή-στόχο π . Η κάθε συνιστώσα X_i $i = 1, \dots, d$ ενημερώνεται ξεχωριστά και θεωρούμε ότι μια νέα κατάσταση έχει ενημερωθεί όταν όλες οι συνιστώσες έχουν ενημερωθεί. Η ενημέρωση της i -οστής συνιστώσας X_i^t της κατάστασης X^t γίνεται σύμφωνα με τα ακόλουθα βήματα

1. Παίρνουμε μια υποψήφια τιμή Z_i από τη μονοδιάστατη κανονική κατανομή πρότασης $g_i^t \sim \mathcal{N}(X_i^{t-1}, v_i^t)$ (5.3)
2. Αποδεχόμαστε την υποψήφια τιμή με πιθανότητα

$$\alpha(x, z_i) = \min(1, \pi(Z_i)/\pi(X_i^{t-1}))$$
3. Αν την αποδεχτούμε τότε $X_i^t = Z_i$ διαφορετικά $X_i^t = X_i^{t-1}$

Η κύρια διαφορά με το βασικό αλγόριθμο MH μιας συνιστώσας είναι ότι η διασπορά της κατανομής πρότασης εξαρτάται από το t

$$v_i^t = \begin{cases} v_i^0, & t \leq t_0 \\ sVar(X_i^0, \dots, X_i^{t-1}) + s\varepsilon, & t > t_0 \end{cases}$$

με v_i^0 η αρχική διασπορά για την κατανομή πρότασης για τη συνιστώσα i και s μια θετική σταθερά (π.χ. $s = 2.4$ όπως δίνεται στο Gelman et al. 1996). Τέλος ε είναι μια μικρή σταθερά που αποτρέπει το μηδενισμό της διασποράς της κατανομής πρότασης (π.χ. $\varepsilon = 0.05$).

5.2.3.1 Μονοδιάστατη περίπτωση

Εφαρμόζουμε τον αλγόριθμο SCAM για να προσομοιώσουμε από την τυπική κανονική κατανομή $\mathcal{N}(0,1)$. Για κατανομή πρότασης χρησιμοποιούμε την $g_i^t \sim \mathcal{N}(X_i^{t-1}, v_i^t)$ μόνο που τώρα έχουμε μόνο μια συνιστώσα και τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 20000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων της πρώτης και της δεύτερης ροπής, τις τυπικές αποκλίσεις τους καθώς επίσης και το μέσο ρυθμό αποδοχής.

Πίνακας 5.7 Μέσος ρυθμός αποδοχής, μέση τιμή εργοδικών μέσων και τυπική απόκλιση της πρώτης και δεύτερης ροπής εφαρμόζοντας τον αλγόριθμο SCAM στο μονοδιάστατο χώρο καταστάσεων.

$d = 1$	Μέσος	Μέση τιμή	Τυπική	Μέση τιμή	Τυπική
$\pi \sim \mathcal{N}(0, 1)$	ρυθμός	1 ^{ης} ροπής	απόκλιση	2 ^{ης} ροπής	απόκλιση
N=20000	αποδοχής		1 ^{ης} ροπής		2 ^{ης} ροπής
	0.4273	-0.0007	0.0213	0.9995	0.0312

Όπως και στον AM αλγόριθμο έτσι και με την εφαρμογή του αλγόριθμου SCAM δεν έχουμε καταφέρει να πετύχουμε ρυθμό αποδοχής πολύ κοντά στο βέλτιστο.

5.2.3.2 Πολυδιάστατη περίπτωση

Εφαρμόζουμε τον αλγόριθμο SCAM για να προσομοιώσουμε από την 10-διάστατη κανονική κατανομή $\mathcal{N}(0, I_{10})$. Για κατανομή πρότασης χρησιμοποιούμε την $g_i^t \sim \mathcal{N}(X_i^{t-1}, v_i^t)$ για κάθε μια συνιστώσα ξεχωριστά και τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 100000 παρατηρήσεων η κάθε μια. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων της πρώτης και της δεύτερης ροπής, τις τυπικές αποκλίσεις τους για κάθε συνιστώσα καθώς και το μέσο ρυθμό αποδοχής.

Πίνακας 5.8 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τον αλγόριθμο SCAM στο 10-διάστατο χώρο καταστάσεων.

$d = 10$ $\pi \sim \mathcal{N}(0, I_{10})$ $N=100000$	Μέσος ρυθμός αποδοχής	Μέσος όρος 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέσος όρος 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής
x_1	0.4278	-0.00030	0.0068	0.9994	0.0097
x_2	0.4276	0.00003	0.0066	1.0003	0.0096
x_3	0.4274	0.00004	0.0066	0.9996	0.0096
x_4	0.4273	0.00026	0.0059	1.0009	0.0095
x_5	0.4277	-0.00015	0.0067	0.9999	0.0098
x_6	0.4275	-0.00028	0.0065	1.0000	0.0098
x_7	0.4275	0.00022	0.0067	0.9995	0.0103
x_8	0.4274	-0.00042	0.0065	1.0008	0.0103
x_9	0.4276	0.00007	0.0069	0.9998	0.0095
x_{10}	0.4275	-0.00005	0.0066	0.9999	0.0098

Παρατηρούμε ότι και στο 10-διάστατο χώρο καταστάσεων όπως και στο μονοδιάστατο ο αλγόριθμος δεν κατάφερε να πετύχει τιμή πολύ κοντά στο βέλτιστο ρυθμό αποδοχής για την κάθε συνιστώσα.

5.2.4 Robbins-Monro

5.2.4.1 Μονοδιάστατη περίπτωση

Για τη μονοδιάστατη κατανομή-στόχο π και κατανομή πρότασης $g \sim \mathcal{N}(x, \sigma^2)$ θα χρησιμοποιήσουμε τη διαδικασία των Robbins-Monro (1951) για να καθορίσουμε την τιμή του σ έτσι ώστε να πετύχουμε το βέλτιστο ρυθμό αποδοχής p^* ($p^* = 0.44$ σύμφωνα με τους Roberts και Rosenthal (2001)).

Σε αυτή την προσέγγιση η τιμή του σ προσαρμόζεται σε κάθε επανάληψη έτσι ώστε ο ρυθμός αποδοχής να πλησιάζει το βέλτιστο p^* . Εάν σ_i δηλώνει την τιμή του σ στην i -οστή επανάληψη τότε θέτουμε

$$\sigma_{i+1} = \begin{cases} \sigma_i + c(1 - p^*)/i, & \text{εάν αποδεχτούμε το } y \\ \sigma_i - cp^*/i, & \text{διαφορετικά} \end{cases} \quad (5.4)$$

όπου $c = \sigma_i / \{p^*(1 - p^*)\}$. Συνήθως παίρνουμε αυθαίρετα μια αρχική τιμή για το σ_0 (π.χ. $\sigma_0 = 1$).

Εφαρμόζουμε τη διαδικασία των Robbins-Monro στο μονοδιάστατο χώρο καταστάσεων για να προσομοιώσουμε παρατηρήσεις από την τυπική κανονική κατανομή $\mathcal{N}(0, 1)$. Για κατανομή πρότασης χρησιμοποιούμε τη $\mathcal{N}(x, \sigma_i^2)$ όπου το σ_i δίνεται από τη σχέση (5.4) και τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 20000 παρατηρήσεων η κάθε μια. Εκτιμούμε την μέση τιμή των εργοδικών μέσων πρώτης και δεύτερης ροπής, τις τυπικές αποκλίσεις τους καθώς και τον μέσο ρυθμό αποδοχής.

Πίνακας 5.9 Μέσος ρυθμός αποδοχής, μέση τιμή εργοδικών μέσων και τυπική απόκλιση της πρώτης και της δεύτερης ροπής εφαρμόζοντας τη διαδικασία των Robbins-Monro στο μονοδιάστατο χώρο.

$d = 1$	Μέσος	Μέση τιμή	Τυπική	Μέση τιμή	Τυπική
$\pi \sim \mathcal{N}(0, 1)$	ρυθμός	1 ^{ης} ροπής	απόκλιση	2 ^{ης} ροπής	απόκλιση
N=20000	αποδοχής		1 ^{ης} ροπής		2 ^{ης} ροπής
	0.4408	-0.0002	0.0155	1.0010	0.0215

Παρατηρούμε ότι όπως και στον Adaptive Metropolis-within-Gibbs αλγόριθμο έτσι και με τον αλγόριθμο Robbins-Monro έχουμε σχεδόν πετύχει το βέλτιστο ρυθμό αποδοχής.

5.2.4.2 Πολυδιάστατη περίπτωση

Για την πολυδιάστατη κατανομή-στόχου π , όλες οι προτεινόμενες παρατηρήσεις προσομοιώνονται από τη $\mathcal{N}(x, \sigma^2 A)$ όπου A είναι ένας θετικά ορισμένος πίνακας. Σκοπός της διαδικασίας των Robbins-Monro είναι να βρεθεί μια τιμή για το σ τέτοια ώστε ο ρυθμός αποδοχής να είναι κοντά στο βέλτιστο $p^* = 0.234$ (Roberts et al. 1997). Η διαδικασία αναζήτησης είναι η ίδια όπως και στη μονοδιάστατη περίπτωση μόνο που τώρα η σταθερά που καθορίζει το μήκος βήματος c ισούται με

$$c = \sigma_i \left\{ \left(1 - \frac{1}{m^*}\right) \cdot \frac{(2\pi)^{\frac{1}{2}} e^{-\alpha^2/2}}{2\alpha} + \frac{1}{m^* p^* (1 - p^*)} \right\}$$

όπου $\alpha = -\Phi^{-1}(p^*/2)$ και συνήθως το m^* ισούται με τη διάσταση του X .

Σε πολλές περιπτώσεις ο ρυθμός σύγκλισης της αλυσίδας Markov βελτιστοποιείται μόνο αν ο A είναι ανάλογος με τον πίνακα διασπορών-συνδιασπορών του X . Αυτός ο πίνακας είναι άγνωστος αλλά μπορεί να εκτιμηθεί κατά τη διάρκεια της προσομοίωσης της αλυσίδας. Μετά από κάθε επανάληψη της αλυσίδας, ο A μπορεί να τεθεί ίσος με την τρέχουσα εκτίμηση του Σ . Έτσι θα έχουμε

$$\hat{\Sigma}_i = \begin{cases} I_d & \text{για } i \leq 100 \\ \frac{1}{i-1} \sum_{j=1}^i (x_j - \bar{x}_i)(x_j - \bar{x}_i) & \text{για } i > 100 \end{cases}$$

όπου ο πίνακας διασπορών-συνδιασπορών $\hat{\Sigma}_i$ υπολογίζεται χρησιμοποιώντας τους αναδρομικούς τύπους

$$\bar{x}_i = \frac{1}{i} [(i-1)\bar{x}_{i-1} + x_i]$$

και

$$\hat{\Sigma}_i = \frac{i-2}{i-1} \hat{\Sigma}_{i-1} + \bar{x}_{i-1} \bar{x}'_{i-1} - \frac{i}{i-1} \bar{x}_i \bar{x}'_i + \frac{1}{i-1} x_i x'_i$$

Οι Haario et al. (2001) πρότειναν για A τον πίνακα $\hat{\Sigma}_i + \epsilon I_d$ με $\epsilon > 0$ σαν μια θετικά ορισμένη εκτίμηση του Σ . Εμείς χρησιμοποιούμε για $\epsilon = \sigma_i^2/i$, έτσι ώστε $A_i = \hat{\Sigma}_i + \sigma_i^2 I_d/i$, οπότε η κατανομή πρότασης μετά από i βήματα είναι η $\mathcal{N}_d(x, \sigma_i^2 A_i)$.

Εφαρμόζουμε τον αλγόριθμο Robbins-Monro στο 10-διάστατο χώρο καταστάσεων για να προσομοιώσουμε από τη $\mathcal{N}(0, I_{10})$. Τρέχουμε 500 ανεξάρτητες αλυσίδες μήκους 100000 παρατηρήσεων η κάθε μία. Εκτιμούμε τη μέση τιμή των εργοδικών μέσων της πρώτης και της δεύτερης ροπής, τις τυπικές αποκλίσεις τους για κάθε συνιστώσα καθώς και το μέσο ρυθμό αποδοχής.

Πίνακας 5.10 Μέσος ρυθμός αποδοχής, μέσες τιμές εργοδικών μέσων και τυπικές αποκλίσεις της πρώτης και δεύτερης ροπής για κάθε συνιστώσα εφαρμόζοντας τη διαδικασία των Robbins-Monro στον 10-διάστατο χώρο.

$d = 10$ $\pi \sim \mathcal{N}(0, I_{10})$ $N=100000$	Μέσος ρυθμός αποδοχής	Μέσος όρος 1 ^{ης} ροπής	Τυπική απόκλιση 1 ^{ης} ροπής	Μέσος όρος 2 ^{ης} ροπής	Τυπική απόκλιση 2 ^{ης} ροπής
x_1	0.2270	0.00038	0.0190	0.9960	0.0193
x_2		-0.00008	0.0182	0.9936	0.0188
x_3		0.00010	0.0185	0.9946	0.0202
x_4		0.00104	0.0182	0.9946	0.0190
x_5		0.00060	0.0183	0.9951	0.0190
x_6		-0.00048	0.0179	0.9972	0.0198
x_7		0.00005	0.0188	0.9957	0.0185
x_8		-0.00062	0.0186	0.9948	0.0205
x_9		-0.00118	0.0182	0.9935	0.0199
x_{10}		0.00010	0.0176	0.9956	0.0200

Εφαρμόζοντας τον αλγόριθμο Robbins-Monro στο 10-διάστατο χώρο καταστάσεων καταφέραμε να προσεγγίσουμε το βέλτιστο ρυθμός αποδοχής.

Γενικά παρατηρούμε ότι στο μονοδιάστατο χώρο καταστάσεων και οι τέσσερις προσαρμοστικοί αλγόριθμοι τείνουν να έχουν ρυθμό αποδοχής πολύ κοντά στο βέλτιστο με τον αλγόριθμο Adaptive Metropolis Within Gibbs και τον αλγόριθμο των Robbins-Monro να πετυχαίνουν ρυθμό αποδοχής σχεδόν ίσο με το βέλτιστο. Αντίστοιχα συμπεράσματα μπορούμε να βγάλουμε για την πολυδιάστατη περίπτωση όπου με εξαίρεση τον AM αλγόριθμο που απέχει αρκετά από το βέλτιστο ρυθμό αποδοχής οι άλλες μέθοδοι μας έδωσαν ικανοποιητικά αποτελέσματα. Έτσι στην περίπτωση που δεν μπορούμε να γνωρίζουμε ποια μπορεί να είναι η κατάλληλη κατανομή πρότασης για να έχουμε γρήγορη σύγκλιση της αλυσίδας μπορούμε να κάνουμε χρήση αυτών των προσαρμοστικών αλγόριθμων.

Στη βιβλιογραφία μπορούμε να βρούμε και άλλους προσαρμοστικούς αλγόριθμους που ενημερώνουν αυτόματα την κατανομή πρότασης για πιο γρήγορη σύγκλιση της αλυσίδας. Μια από αυτές είναι η μέθοδος *State Dependent Proposal Scaling* η οποία χρησιμοποιεί για κατανομή πρότασης τη $g(x, \cdot) = \mathcal{N}(x, \sigma_x^2)$, όπου $\sigma_x = e^a(1 + |x|)^b$ στην οποία μετά από κάθε k επαναλήψεις (π.χ. $k=100$) αυξομειώνονται οι τιμές των a και b έτσι ώστε ο ρυθμός αποδοχής να τείνει στο 0.234 ή 0.44 ανάλογα με τη διάσταση του χώρου καταστάσεων. Μία άλλη τεχνική είναι η *Regional Adaptive Metropolis Algorithm (RAMA)* η οποία χωρίζει το χώρο καταστάσεων σε m υποσύνολα και προτείνει τιμές από την κατανομή $\mathcal{N}(x, e^{2a_i})$, όπου το i αντιστοιχεί στο υποσύνολο που ανήκει το x . Το a_i ενημερώνεται μετά το τέλος κάθε k επαναλήψεων (π.χ. $k=100$) στο οποίο προστίθεται ή αφαιρείται κάποια σταθερά ανάλογα με το αν ο ρυθμός αποδοχής είναι μεγαλύτερος ή μικρότερος από το βέλτιστο. Περισσότερες πληροφορίες για τους δύο αλγόριθμους βλέπε Roberts and Rosenthal (2006a) και Rosenthal (2008).

6 Σύνοψη – Συμπεράσματα

Στα πλαίσια αυτής της εργασίας ασχοληθήκαμε με την προσομοίωση παρατηρήσεων από κάποια κατανομή-στόχο π μέσω MC και MCMC τεχνικών και κυρίως μέσω του αλγόριθμου MH και με την εκτίμηση ολοκληρωμάτων της μορφής (1.1).

Ο αλγόριθμος MH εφαρμόστηκε σε απλά παραδείγματα καθώς επίσης και σε πραγματικά σύνολα δεδομένων και συγκρίθηκε με αντίστοιχες τεχνικές προσομοίωσης και εκτίμησης μέσω των τιμών. Επίσης αναφερθήκαμε σε τεχνικές εκτίμησης της ασυμπτωτικής διασποράς των εργοδικών μέσων.

Τέλος, ασχοληθήκαμε με το πρόβλημα της εύρεσης μιας κατάλληλης κατανομής πρότασης για τον αλγόριθμο MH. Δείξαμε ότι η επιλογή της παραμέτρου κλίμακας παίζει καθοριστικό ρόλο στη σύγκλιση της αλυσίδας και στη συνέχεια παρουσιάσαμε μία σειρά από προσαρμοστικούς αλγόριθμους οι οποίοι μας βοηθούν στην εύρεση της καταλληλότερης κατανομής πρότασης για γρηγορότερη σύγκλιση του αλγόριθμου MH.

Συνοψίζοντας θα πρέπει να τονίσουμε τη σημαντικότητα του αλγόριθμου MH στην υπολογιστική στατιστική και να αναφέρουμε ότι δίκαια χαρακτηρίστηκε από τους εκδότες του περιοδικού *Computing in Science and Engineering* ως ένας από τους δέκα αλγόριθμους που επηρέασαν περισσότερο τις επιστήμες κατά τον εικοστό αιώνα.

ΠΑΡΑΡΤΗΜΑ

Αλυσίδες Markov

Στο κεφάλαιο αυτό έχουμε συγκεντρώσει κάποιους βασικούς ορισμούς και θεωρήματα σχετικά με τις αλυσίδες Markov, τα οποία είναι απαραίτητα για την κατανόηση βασικών ιδιοτήτων των μεθόδων προσομοίωσης που αναπτύσσονται στα προηγούμενα κεφάλαια.

Μια διαδικασία Markov $(X_t)_{t \in T}$ είναι μια στοχαστική διαδικασία όπου t παράμετρος με τιμές σε ένα διατεταγμένο σύνολο T . Το T μπορεί να είναι συνεχές ή διακριτό. Στα πλαίσια αυτής της εργασίας θα ασχοληθούμε μόνο με την περίπτωση όπου $T = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ δηλαδή θα ασχοληθούμε με διαδικασίες διακριτού χρόνου, τις οποίες καλούμε αλυσίδες Markov. Οι τιμές των τυχαίων μεταβλητών X_t ανήκουν σε ένα σύνολο \mathcal{X} . Το σύνολο \mathcal{X} λέγεται χώρος καταστάσεων της X_t και μπορεί να είναι πεπερασμένο ή αριθμήσιμο οπότε και η αλυσίδα λέγεται αλυσίδα Markov με διακριτό χώρο καταστάσεων ή συνεχής και τότε λέμε ότι έχουμε μια αλυσίδα Markov με γενικό χώρο καταστάσεων.

Επίσης, οι τυχαίες μεταβλητές X_t έχουν τη Μαρκοβιανή ιδιότητα δηλαδή το μέλλον είναι ανεξάρτητο από το παρελθόν δοθείσης της τιμής της κατάστασης την τρέχουσα χρονική στιγμή.

Ορισμός 1: Μια ακολουθία τυχαίων μεταβλητών X_0, X_1, X_2, \dots ορισμένων στον ίδιο χώρο πιθανότητας (Ω, \mathcal{F}, P) και με τιμές στο χώρο $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ λέμε ότι είναι αλυσίδα Markov αν

$$P(X_{n+1} \in A | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(X_{n+1} \in A | X_n = x_n)$$

$\forall A \in \mathcal{B}(\mathcal{X}), n \in \mathbb{N}$ και $x_n, x_{n-1}, \dots, x_0 \in \mathcal{X}$.

Ορισμός 2: Αν

$$P(X_{n+1} \in A | X_n = x_n) = P(X_1 \in A | X_0 = x_0)$$

τότε η αλυσίδα λέγεται *ομογενής*.

Με άλλα λόγια μια αλυσίδα Markov λέγεται ομογενής αν η δεσμευμένη κατανομή της X_{n+1} δοθέντος της X_n δεν εξαρτάται από τη χρονική στιγμή n για κάθε $x \in \mathcal{X}$.

Έστω \mathcal{X} ένας γενικός χώρος καταστάσεων και $\mathcal{B}(\mathcal{X})$ μια μετρήσιμη σ -άλγεβρα πάνω στο χώρο \mathcal{X} .

Ορισμός 3: Αν $\{K(x, A) : x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ είναι τέτοιο ώστε

1. για κάθε $A \in \mathcal{B}(\mathcal{X})$, η $K(\cdot, A)$ είναι μια μη αρνητική μετρήσιμη συνάρτηση ορισμένη πάνω στο \mathcal{X}
2. για κάθε $x \in \mathcal{X}$, η $K(x, \cdot)$ είναι ένα μέτρο πιθανότητας ορισμένο πάνω στο πεδίο $\mathcal{B}(\mathcal{X})$

τότε ονομάζουμε το K *πυρήνα μετάβασης (transition probability kernel)* ή *συνάρτηση μετάβασης (transition function)*.

Κάθε πυρήνας μετάβασης ορίζει μια αλυσίδα Markov και αντίστροφα, κάθε ομογενής αλυσίδα Markov ορίζει ένα πυρήνα μετάβασης

$$K(x, A) \equiv P(X_1 \in A | X_0 = x).$$

Μια αλυσίδα Markov καθορίζεται πλήρως από την κατανομή της X_0 η οποία καλείται αρχική κατανομή και από τον πυρήνα μετάβασης. Θέτοντας ένα σταθερό σημείο εκκίνησης $x \in \mathcal{X}$ έχουμε την παρακάτω διαδικασία.

$$P(X_1 \in A_1 | X_0 = x) = K(x, A_1)$$

$$P((X_1, X_2) \in A_1 \times A_2 | X_0 = x) = \int_{A_1} K(y_1, A_2) K(x, dy_1)$$

⋮

$$P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n | X_0 = x) = \int_{A_1} \dots \int_{A_{n-1}} K(y_{n-1}, A_n) \\ \times K(y_{n-2}, dy_{n-1}) \times \dots \times K(x, dy_1)$$

Ορισμός 4: Η συνάρτηση

$$K^n(x, B) = P(X_n \in B | X_0 = x_0)$$

λέγεται *πυρήνας μετάβασης n τάξης*.

Για τον πυρήνα μετάβασης n τάξης ισχύει

$$K^n(x, B) = \int_{\mathcal{X}} K^{n-1}(y, B)K(x, dy) = \int_{\mathcal{X}} K(y, B)K^{n-1}(x, dy)$$

για όλα τα $B \in \mathcal{B}(\mathcal{X})$, $x \in \mathcal{X}$, $n = 1, 2, \dots$. Για $n = 0$ θέτουμε

$$K^0(x, \cdot) = \delta_x(\cdot),$$

όπου $\delta_x(\cdot)$ το μέτρο Dirac στο $x \in \mathcal{X}$,

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

Εξισώσεις Charman-Kolmogorov

Θεώρημα 1: Για κάθε $m, n \geq 0$ ισχύει

$$K^{m+n}(x, A) = \int_{\mathcal{X}} K^n(y, A)K^m(x, dy)$$

που σημαίνει ότι καθώς η αλυσίδα κινείται από την αρχική κατάσταση x στο σύνολο A σε $m + n$ βήματα, σε οποιοδήποτε ενδιάμεσο βήμα n θα πάρει κάποια τιμή $y \in \mathcal{X}$. Και επειδή πρόκειται για αλυσίδα Markov ξεχνάει το παρελθόν της στο χρόνο n και συνεχίζει

στα επόμενα m βήματα με τον πιθανοθεωρητικό νόμο που διέπει την κίνηση της προσαρμοσμένο στο νέο σημείο εκκίνησης y .

Ορισμός 5: Αν υπάρχει ένα μέτρο πιθανότητας π που να ικανοποιεί την εξίσωση

$$\pi(A) = \int \pi(dx)K(x, A)$$

τότε λέμε ότι αποτελεί μια στάσιμη κατανομή της αλυσίδας.

Ορισμός 6: Για κάθε σύνολο $A \in \mathcal{B}(X)$, η τυχαία μεταβλητή

$$\eta_A = \sum_{n=1}^{\infty} I(X_n \in A)$$

είναι ο αριθμός των διελεύσεων της X από το σύνολο A έπειτα από τη χρονική στιγμή μηδέν και καλείται χρόνος κατοχής του A .

Ορισμός 7: Για κάθε σύνολο $A \in \mathcal{B}(X)$, η τυχαία μεταβλητή

$$\tau_A = \begin{cases} \min\{n \geq 1: X_n \in A\}, & \text{αν } X_n \in A \text{ για κάποιο } n < \infty \\ \infty, & \text{αν } X_n \notin A \text{ για κάθε } n, \end{cases}$$

λέγεται χρόνος πρώτης επανόδου στο A .

Μη διαχωρισιμότητα (Irreducibility)

Ορισμός 8: Έστω ένα μέτρο φ ορισμένο στη σ -άλγεβρα $\mathcal{B}(X)$. Αν για κάθε $x \in X$ και $A \in \mathcal{B}(X)$ με $\varphi(A) > 0$ υπάρχει ένα $n \geq 1$ τέτοιο ώστε

$$K^n(x, A) > 0 \text{ για όλα τα } x \in X,$$

ή ισοδύναμα, αν

$$P(\tau_A < \infty | X_0 = x) > 0 \text{ για όλα τα } x \in \mathcal{X},$$

τότε η αλυσίδα Markov $(X_n)_{n \in \mathbb{Z}_+}$ λέγεται φ -μη διαχωρίσιμη. Μια αλυσίδα λέγεται ισχυρά φ -μη διαχωρίσιμη αν $n = 1$ για όλα τα μετρήσιμα σύνολα A .

Αυτό σημαίνει ότι ανεξαρτήτως της αρχικής κατάστασης της αλυσίδας θα επισκεφτεί κάθε άλλη κατάσταση σε πεπερασμένο αριθμό επαναλήψεων με θετική πιθανότητα.

Πρόταση 1: Αν η αλυσίδα X είναι φ -μη διαχωρίσιμη για κάποιο μέτρο φ τότε υπάρχει ένα μέγιστο (maximal) μέτρο πιθανότητας ψ ορισμένο στο πεδίο $\mathcal{B}(\mathcal{X})$ τέτοιο ώστε

1. η X είναι ψ -μη διαχωρίσιμη,
2. για κάθε άλλο μέτρο φ' , η αλυσίδα X είναι φ' -μη διαχωρίσιμη αν και μόνο αν $\psi > \varphi'$, δηλαδή το μέτρο φ' είναι απολύτως συνεχές ως προς το μέτρο ψ . Αυτό σημαίνει ότι $\forall A \in \mathcal{B}(\mathcal{X})$ με $\psi(A) = 0$ ισχύει $\varphi'(A) = 0$.
3. αν $\psi(A) = 0$ τότε $\psi(\{x: K^n(x, A) > 0 \text{ για κάποιο } n \geq 1\}) = 0$.

Με απλά λόγια, αν η αλυσίδα είναι φ -μη διαχωρίσιμη για κάποιο μέτρο φ τότε υπάρχει κάποιο μέτρο ψ το οποίο καθορίζει όλα τα σύνολα στα οποία μπορεί αυτή να βρεθεί με θετική πιθανότητα. Γενικά η φ -μη διαχωρισιμότητα δεν καθορίζει όλα παρά μόνο κάποια από αυτά τα σύνολα, αυτά με $\varphi(A) > 0$. Δεν εξασφαλίζει ότι αν $\varphi(A) = 0$ τότε η αλυσίδα έχει μηδενική πιθανότητα να επισκεφθεί το A . Αντιθέτως από την τρίτη ιδιότητα, βλέπουμε ότι για το μέγιστο μέτρο ψ , αν κάποιο σύνολο A είναι ψ -μηδενικό τότε το σύνολο των $x \in \mathcal{X}$ από τα οποία αν ξεκινήσει η αλυσίδα θα βρεθεί κάποτε στο A με θετική πιθανότητα είναι επίσης ψ -μηδενικό.

Απεριοδικότητα (Aperiodicity)

Ορισμός 9: Η αλυσίδα λέγεται περιοδική αν υπάρχει μια διαμέριση $A_1, \dots, A_d \in \mathcal{B}(\mathcal{X})$ του \mathcal{X} τέτοια ώστε

$$P(x, A_{j+1}) = 1, \quad \forall x \in A_j \text{ με } (A_{d+1} \equiv A_1)$$

Αν δεν συμβαίνει κάτι τέτοιο η αλυσίδα λέγεται απεριοδική.

Μεταβατικότητα και Επαναληπτικότητα (Transience and Recurrence)

Ορισμός 10: Ένα σύνολο $A \in \mathcal{B}(\mathcal{X})$ ονομάζεται ομοιόμορφα μεταβατικό (uniformly transient) αν υπάρχει μια σταθερά $M < \infty$ τέτοια ώστε

$$E_x(\eta_A) \leq M \text{ για κάθε } x \in A.$$

Ορισμός 11: Ένα σύνολο $A \in \mathcal{B}(\mathcal{X})$ λέγεται επαναληπτικό αν

$$E_x(\eta_A) = \infty \text{ για κάθε } x \in A.$$

Ορισμός 12: Μια αλυσίδα λέγεται επαναληπτική αν κάθε σύνολο το οποίο αυτή μπορεί να επισκεφθεί με θετική πιθανότητα είναι επαναληπτικό.

Ορισμός 13: Ένα σύνολο $A \in \mathcal{B}(\mathcal{X})$ θα λέγεται Harris επαναληπτικό αν

$$P(\eta_A = \infty | X_0 = x) = 1 \text{ για κάθε } x \in A.$$

Αν η αλυσίδα είναι ψ -μη διαχωρίσιμη (με ψ μέγιστο) και κάθε ψ -θετικό σύνολο είναι Harris επαναληπτικό τότε η αλυσίδα λέγεται Harris επαναληπτική.

$$P(\eta_A = \infty | X_0 = x) = 1 \text{ για κάθε } x \in \mathcal{X}.$$

Με απλά λόγια ένα σύνολο λέγεται Harris επαναληπτικό αν η αλυσίδα επιστρέφει σε αυτό άπειρες φορές με πιθανότητα ένα.

Ορισμός 14: Αν μια ψ -μη διαχωρίσιμη αλυσίδα έχει μια στάσιμη κατανομή π τότε λέγεται θετική.

Ορισμός 15: Αν μια αλυσίδα είναι Harris επαναληπτική και θετική τότε λέγεται θετική Harris.

Αντιστρεψιμότητα (Reversibility)

Ορισμός 16: Μια αλυσίδα Markov με χώρο καταστάσεων X και πυρήνα μετάβασης K λέγεται αντιστρέψιμη ως προς μια κατανομή π ορισμένη στο X αν

$$\int_A \int_B \pi(dx)K(x, dy) = \int_B \int_A \pi(dy)K(y, dx) , \quad \forall A, B \in \mathcal{B}(X)$$

Η παραπάνω συνθήκη αντιστρεψιμότητας είναι γνωστή και ως *λεπτομερής συνθήκη ισορροπίας* (*detailed balance condition*). Η συνθήκη αυτή δεν είναι απαραίτητη για την εξασφάλιση της σύγκλισης, ωστόσο είναι επαρκής συνθήκη ώστε η π να είναι η στάσιμη κατανομή της αλυσίδας.

Η αλυσίδα είναι αντιστρέψιμη ως προς την π αν και μόνον αν, υπό στασιμότητα (δηλαδή αν $X_n \sim \pi$),

$$\begin{aligned} P\{X_n \in A, X_{n+1} \in B\} &= \int_{x \in A} \int_{y \in B} \pi(dx)K(x, dy) \\ &= \int_{x \in A} \int_{y \in B} \pi(dy)K(y, dx) \\ &= P\{X_n \in B, X_{n+1} \in A\}. \end{aligned}$$

Η απόσταση δύο μέτρων μ, ν ορισμένων πάνω σε μια σ -άλγεβρα \mathcal{B} μετριέται βάσει της νόρμας ολικής μεταβολής (total variation norm)

$$\|\mu - \nu\| = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|.$$

ΕΡΓΟΔΙΚΟ ΘΕΩΡΗΜΑ: Αν X_0, X_1, X_2, \dots μια Harris θετική αλυσίδα Markov με στάσιμη κατανομή π και $h \in \mathcal{L}(\pi)$, δηλαδή $E_\pi(|h|) < \infty$, τότε

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \xrightarrow{\sigma.β.} E_\pi(h) \text{ καθώς } n \rightarrow \infty$$

Ορισμός 17: Η αλυσίδα Markov $(X_n)_{n \in \mathbb{Z}_+}$ λέγεται f -γεωμετρικά εργοδική (f -geometrically ergodic) όπου $f \geq 1$, αν η $(X_n)_{n \in \mathbb{Z}_+}$ είναι Harris θετική με κάποια στάσιμη κατανομή $\pi(f) < \infty$ και υπάρχει μια σταθερά $r_f > 1$ τέτοια ώστε

$$\sum_{n=1}^{\infty} r_f^n \|K^n(x, \cdot) - \pi\| < \infty, \quad \forall x \in \mathcal{X}.$$

Αν έχουμε ότι $f \equiv 1$, τότε λέμε ότι η X είναι γεωμετρικά εργοδική.

Ορισμός 18: Μια εργοδική αλυσίδα $(X_n)_{n \in \mathbb{Z}_+}$ λέγεται ομοιόμορφα εργοδική αν

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} \|K^n(x, \cdot) - \pi\| = 0.$$

Κεντρικά οριακά θεωρήματα

Θεώρημα 2: Αν η αλυσίδα Markov $(X_n)_{n \in \mathbb{Z}_+}$ είναι απεριοδική, μη διαχωρίσιμη και αντιστρέψιμη με στάσιμη κατανομή π και

$$0 < \gamma_h^2 = \text{Var}_\pi[h(X_0)] + 2 \sum_{k=1}^{\infty} \text{Cov}_\pi[h(X_0)h(X_k)] < \infty$$

τότε ισχύει το Κεντρικό Οριακό Θεώρημα, δηλαδή

$$\sqrt{n}\{h_n - E_\pi(h)\} \xrightarrow{d} \mathcal{N}(0, \gamma_h^2).$$

Αν η αλυσίδα δεν είναι αντιστρέψιμη τότε δεν αρκεί να είναι πεπερασμένη μόνο η δεύτερη ροπή για να ισχύει το ΚΟΘ αλλά χρειάζεται και κάτι παραπάνω.

Θεώρημα 3: Αν η $(X_n)_{n \in \mathbb{Z}_+}$ είναι μια απεριοδική, μη διαχωρίσιμη, θετικά Harris επαναληπτική, με στάσιμη κατανομή την π και γεωμετρικά εργοδική και αν επιπλέον ισχύει

$$E_\pi[|h(X)^{2+\epsilon}|] < \infty$$

για κάποιο $\epsilon > 0$, τότε

$$\sqrt{n}\{h_n - E_\pi(h)\} \xrightarrow{d} \mathcal{N}(0, \gamma_h^2).$$

Θεώρημα Slutsky: Έστω X_n, Y_n ακολουθίες τυχαίων αριθμών. Εάν η X_n συγκλίνει στην κατανομή μιας τυχαίας ποσότητας X και η Y_n συγκλίνει με πιθανότητα σε μια σταθερά c , τότε

$$i. \quad X_n + Y_n \xrightarrow{d} X + c$$

$$ii. \quad Y_n X_n \xrightarrow{d} cX$$

$$iii. \quad Y_n^{-1} X_n \xrightarrow{d} c^{-1}X$$

με την προϋπόθεση ότι υπάρχει η c^{-1} .

Βιβλιογραφία

1. Anderson T.W., (1994), "The Statistical Analysis of Time Series", Wiley, New York.
2. Barker A.A., (1965), "Monte Carlo Calculations of the Radial Distribution Functions for a Proton-electron Plasma", Australian Journal of Physics, vol. 18, 119-133.
3. Bedard M., (2006), "Optimal Acceptance Rates for Metropolis Algorithms: Moving Beyond 0.234", Stochastic Processes and their Applications to appear.
4. Bedard M., (2007), "Weak Convergence of Metropolis Algorithms for Non-iid Target Distributions", The Annals of Applied Probability, vol. 17, 1222-1244.
5. Bedard M., (2008), "Efficient Sampling using Metropolis Algorithms: Applications of Optimal Scaling Results", Journal of Computational and Graphical Statistics, vol. 17, 1-21.
6. Bedard M. and Rosenthal J.S., (2008), "Optimal Scaling of Metropolis Algorithms: Heading Towards General Target Distributions", The Canadian Journal of Statistics, vol. 36, 483-503.
7. Besag J. and Green P. J., (1993), "Spatial statistics and bayesian computation", Journal of the Royal Statistical Society, Series B, vol. 55, 25-37.
8. Besag, J., Green P., Higdon D. and Mengersen K., (1995), "Bayesian computation and stochastic systems", Statistical Science, vol. 10, 3-41.
9. Breyer L. A. and Roberts G. O., (2000), "From Metropolis to diffusions: Gibbs states and optimal scaling", Stochastic Process, vol. 90, 181-206.
10. Brockwell A.E. and Kadane J.B., (2004), "Identification of regeneration times in MCMC simulation, with application to adaptive schemes", Journal of Computational and Graphical Statistics, vol. 14, 436-458.
11. Brooks S., Gelman A., Jones G.L. and Xiao-Li Meng, (2011), "Handbook of Markov Chain Monte Carlo" Chapman and Hall/CRC.
12. Cappe O., Guillin A., Marin J.M. and Robert C.P., (2004), "Population Monte Carlo", Journal of Computational and Graphical Statistics, vol. 13, 907-930.
13. Carlin B.P. and Gelfand A.E., (1991), "An iterative Monte Carlo method for nonconjugate Bayesian analysis", Statistics and Computing, vol. 1, 119-128.

14. Cowles M.K. and Carlin B.P., (1996), "Markov chain Monte Carlo convergence diagnostics: A Comparative Review". *Journal of the American Statistical Association*, Vol. 91, No. 434, 883-904.
15. Flegal J.M and Jones G.L., (2010), "Batch means and spectral variance estimators in Markov chain Monte Carlo", *Annals of Statistics*, vol. 38, No.2, 1034-1070.
16. Gelfand A.E. and Smith A.F.M., (1990), "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, vol. 85, 398-409.
17. Gelman A., Roberts G.O. and Gilks W.R., (1996), "Efficient Metropolis jumping rules", *Bayesian Statistics 5*, ed J. Bernardo et al., 599-607, Oxford University Press.
18. Geman S. and Geman D., (1984) "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, 721-741.
19. Glynn P.W. and Iglehart D. L., (1990), "Simulation output analysis using standardized time series", *Mathematics of Operations Research*, vol. 15, 1-16.
20. Glynn P. W. and Whitt W., (1991), "Estimating the asymptotic variance with batch means", *Operations Research Letters*, vol. 10, 431-435.
21. Haario H., Saksman E. and Tamminen J., (2001), "An adaptive Metropolis algorithm", *Bernoulli*, vol. 7, 223-242.
22. Hammersley J.M. and Handscomb D.C., (1964), "Monte Carlo Methods", London: Chapman and Hall.
23. Hastings W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, vol. 57, 97-109
24. Jones G.L. and Hobert, J.P., (2001), "Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo", *Statistical Science*, vol. 16, 312-334.
25. Liu J., (1996), "Metropolized independent sampling with comparisons to rejection sampling and importance sampling", *Statistics and Computing*, vol. 6, 113-119.
26. Liu J.S. (2001), "Monte Carlo Strategies in Scientific Computing", New York: Springer.

27. McLeish D.L. (1975) "A maximal inequality and dependent strong laws", *Annals of Probability*, vol. 3, 829-839.
28. Marshall A.W., (1956), "The use of multi-stage sampling schemes in Monte Carlo computations", In *Symposium on Monte Carlo Methods*, edited by Meyer M., Wiley, New York, 123–140.
29. Metropolis N. and Ulam S., (1949), "The Monte Carlo Method", *Journal of the American Statistical Association*, vol. 44, 335-341.
30. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E., (1953), "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, vol. 21, 1087-1092.
31. Mengersen K. and Tweedie R.L., (1996), "Rates of convergence of the Hastings and Metropolis algorithms", *The Annals of Statistics*, vol. 24, 101-121.
32. Mykland P., Tierney L. and Yu B., (1995), "Regeneration in Markov chain samplers", *Journal of the American Statistical Association*, vol. 90, 233-241.
33. Neumann J.V., (1951), "Various techniques used in connection with random digits.Monte Carlo Methods", *National Bureau of Standards Applied Mathematics Series*, vol. 12, 36-38.
34. Peskun P.H., (1973), "Optimum Monte Carlo Sampling Using Markov Chains", *Biometrika*, vol. 60, 607-612.
35. Ratkowsky D., (1983), "Nonlinear Regression Modeling", Marcel Dekker.
36. Robbins H. and Monro S., (1951), "A stochastic approximation method", *Annals of Mathematical Statistics*, vol. 22, 400-407.
37. Robert C.P., (1995), "Convergence Control Methods for Markov Chain Monte Carlo Algorithms", *Statistical Science*, vol. 10, 230–253.
38. Roberts G.O. and Tweedie R.L., (1996), "Exponential convergence of Langevin diffusions and their discrete approximations", *Bernoulli* 2, 341-363.
39. Roberts G.O. and Rosenthal J.S., (1997), "Shift-coupling and convergence rates of ergodic averages", *Stochastic Models*, vol. 13, 147-165.
40. Roberts G., Gelman A. and Gilks W., (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms", *The Annals of Applied Probability*, vol. 7, 110-120.

41. Roberts G.O. and Rosenthal J.S., (1998), "Optimal scaling of discrete approximations to Langevin diffusions", *Journal of the Royal Statistical Society, Series B*, vol. 60, 255-268.
42. Roberts, G.O. and Rosenthal J. S., (2001), "Optimal scaling for various Metropolis-Hastings algorithms", *Statistical Science*, vol. 16, 351-367.
43. Roberts G.O. and Rosenthal J.S., (2005), "Coupling and Ergodicity of Adaptive MCMC". Preprint.
44. Roberts G.O. and Rosenthal J.S., (2006a), "Examples of Adaptive MCMC", *Journal of Computational and Graphical Statistics*, to appear.
45. Rosenthal J.S., (2004), Adaptive MCMC Java Applet.
46. Rubin D.B., (1987), "The calculation of posterior distribution by Data Augmentation: A Noniterative Sampling/Importance Resampling alternative to the Data Augmentation algorithm for greating a few imputations when fractions of missing information are modest: The SIR algorithm", *Journal of the American Statistical Association*, vol. 82, 543-546.