

Υπολογιστικές Μέθοδοι  
στην  
Μπεϋζιανή Στατιστική

Λουκάς Ε. Λουκάς

24 Ιανουαρίου 2005



# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>7</b>
<b>2</b>	<b>Αλυσίδες Markov</b>	<b>13</b>
2.1	Ορισμός	13
2.2	Θεμελίωση σε Γενικό Χώρο Καταστάσεων	14
2.3	$\psi$ -Αναγωγιμότητα	19
2.4	Ψευδο-Άτομα και Μικρά Σύνολα	22
2.5	Κυκλική Συμπεριφορά	25
2.6	Παροδικότητα και Επαναληπτικότητα	26
2.7	Harris Επαναληπτικότητα	29
2.8	Η Ύπαρξη του Μέτρου $\pi$	31
2.8.1	Αναλλοίωτα Μέτρα	32
2.8.2	Υποαναλλοίωτα Μέτρα	34
2.8.3	Η Ύπαρξη του Μέτρου $\pi$ : Αλυσίδες με Άτομα	35
2.8.4	Η Ύπαρξη του Μέτρου $\pi$ : $\psi$ -Ανάγωγες Αλυσίδες	36
2.8.5	Η Δομή του Μέτρου $\pi$ για Επαναληπτικές Αλυσίδες	36
2.9	Εργοδικότητα	37
2.10	$f$ , Γεωμετρική και Ομοιόμορφη Εργοδικότητα	40
2.10.1	$f$ -Εργοδικότητα	40
2.10.2	Γεωμετρική Εργοδικότητα	41
2.10.3	$V$ -Ομοιόμορφη Εργοδικότητα	41
2.11	Οριακά Θεωρήματα	43
2.11.1	Αρμονικές Συναρτήσεις	44
2.11.2	Ο Νόμος των Μεγάλων Αριθμών	44
2.11.3	Κεντρικό Οριακό Θεώρημα	45
2.11.4	Αντιστρέψιμες Αλυσίδες Markov	46

<b>3</b>	<b>Ο Αλγόριθμος Metropolis - Hastings</b>	<b>49</b>
3.1	Εισαγωγή . . . . .	49
3.2	Ορισμός και Ιδιότητες . . . . .	51
3.3	Περιορισμοί για τη Συνάρτηση μετάβασης πρότασης . . . . .	57
3.4	Ιδιότητες Σύγκλισης . . . . .	59
3.5	Ειδικές Περιπτώσεις . . . . .	60
3.5.1	Ανεξάρτητες Αλυσίδες . . . . .	60
3.5.2	Αλυσίδες Τυχαίου Περιπάτου . . . . .	63
3.5.3	Αυτοπαλίνδρομες Αλυσίδες . . . . .	66
3.5.4	Αλυσίδες Πολλαπλασιαστικού Τυχαίου Περιπάτου . . . . .	66
3.6	Βελτίωση του Ποσοστού Αποδοχής . . . . .	68
3.6.1	Ανεξάρτητες Αλυσίδες . . . . .	68
3.6.2	Αλυσίδες Τυχαίου Περιπάτου . . . . .	70
3.7	Υβριδικοί Αλγόριθμοι . . . . .	71
3.7.1	Metropolis-Hastings κατά συντεταγμένη . . . . .	71
<b>4</b>	<b>Ο Δειγματολήπτης Gibbs</b>	<b>77</b>
4.1	Εισαγωγή . . . . .	77
4.2	Ορισμός και Ιδιότητες . . . . .	77
4.3	Gibbs και Metropolis - Hastings . . . . .	83
4.4	Συμπλήρωση . . . . .	84
4.5	Ιδιότητες Σύγκλισης . . . . .	86
4.6	Υβριδικοί Αλγόριθμοι . . . . .	88
4.6.1	Βοηθητικές Μεταβλητές . . . . .	88
4.6.2	Metropolis εντός του Gibbs . . . . .	91
4.6.3	Ομαδοποιημένος και Τροποποιημένος Εκφυλισμένος Gibbs . . . . .	95
4.7	Αύξηση Δεδομένων . . . . .	96
4.8	Δυικές Πιθανοτικές Δομές . . . . .	99
4.9	Rao - Blackwellization . . . . .	101
4.10	Η Πληροφοριακή Ικανότητα των πλήρως δεσμευμένων Κατανομών . . . . .	105
<b>5</b>	<b>Επιπρόσθετα Θέματα</b>	<b>107</b>
5.1	Ο Σχηματισμός του Δείγματος . . . . .	107
5.2	Η Burn In Περίοδος του Δείγματος . . . . .	109
5.3	Κατασκευή Εκτιμητών . . . . .	110
5.4	Αυτοσυσχέτιση . . . . .	113
5.5	Η Σύγκλιση των Αλυσίδων Markov . . . . .	114
5.5.1	Γραφικές Τεχνικές Ανίχνευσης της Μη Σύγκλισης . . . . .	116

5.6	Διαγνωστικά Κριτήρια Σύγκλισης . . . . .	120
5.6.1	Το διαγνωστικό Κριτήριο του Geweke . . . . .	120
5.6.2	Το Διαγνωστικό Κριτήριο των Gelman και Rubin . . . . .	122
5.6.3	Το διαγνωστικό Κριτήριο των Zellner και Min . . . . .	124
5.6.4	Μη Παραμετρικός Έλεγχος Στασιμότητας . . . . .	125
<b>6</b>	<b>Εφαρμογές</b>	<b>127</b>
6.1	Εισαγωγή . . . . .	127
6.2	Παράδειγμα 1.0.1 (Pump Failures) . . . . .	127
6.3	Παράδειγμα 1.0.2 (Pump Failures) . . . . .	128
6.4	Παράδειγμα 1.0.3 (Random Effects Model) . . . . .	129



# Κεφάλαιο 1

## Εισαγωγή

Το ανά χείρας αφιερώνεται στην παρουσίαση των ευρέως διαδεδομένων μεθόδων στοχαστικής προσομοίωσης, που είναι γνωστές με την ονομασία Markov Chain Monte Carlo (MCMC). Οι MCMC αλγόριθμοι χρησιμοποιούνται πολύ συχνά στη Μπεϋζιανή συμπερασματολογία. Σύμφωνα με το θεώρημα του Bayes, η εκ των υστέρων κατανομή (posterior distribution) ενός παραμετρικού διανύσματος  $\theta = (\theta_1, \dots, \theta_d)$ , δοθέντος ενός διανύσματος παρατηρήσεων  $y = (y_1, \dots, y_T)$  των τυχαίων μεταβλητών  $Y = (Y_1, \dots, Y_T)$ , δίνεται από τη σχέση

$$P(\theta|x) \propto P(x|\theta) P(\theta), \quad (1.1)$$

όπου  $P(\theta)$  είναι η εκ των προτέρων κατανομή (prior distribution) του διανύσματος  $\theta$  των παραμέτρων.

Πολλές φορές καλούμαστε να προσομοιώσουμε εκ των υστέρων κατανομές πολύπλοκης μορφής, στις οποίες εμπλέκονται πολλές παράμετροι. Μοντέλα με δυναμικές παραμέτρους (dynamic settings), ιεραρχικά μοντέλα (hierarchical models), μοντέλα τυχαίων επιδράσεων (random effects models), διαστημικά μοντέλα (spatial models) είναι παραδείγματα πολυπαραμετρικών μοντέλων. Ας εξετάσουμε κάποιες περιπτώσεις.

### Παράδειγμα 1.0.1 *Pump Failures*

Οι Gaver and O’Muirheartaigh (1987) εισήγαγαν ένα μοντέλο για να περιγράψουν τις αποτυχίες δέκα μηχανισμών άντλησης ιόντων από ένα πυρηνικό αντιδραστήρα. Τα δεδομένα δίνονται στον Πίνακα 1.1. Η τυχαία μεταβλητή  $X_i$  παριστάνει τον αριθμό των αποτυχιών του  $i$  μηχανισμού και η μεταβλητή  $t_i$  είναι το χρονικό διάστημα μέσα στο οποίο παρατηρούνται αυτές οι αποτυχίες.

Pump	1	2	3	4	5	6	7	8	9	10
Failures	5	1	5	14	3	19	1	1	4	22
Time	94.32	15.72	62.88	125.76	5.24	31.44	1.05	1.05	2.10	10.48

Πίνακας 1.1: Αριθμός αποτυχιών και χρόνοι παρατήρησης δέκα μηχανισμών άντλησης ιόντων.

Υποθέτουμε ότι  $X_i \stackrel{iid}{\sim} Poi(\lambda_i t_i)$ ,  $i = 1, \dots, 10$ . Επιπλέον, υποθέτουμε τις εκ των προτέρων κατανομές

$$\begin{aligned}\lambda_i &\stackrel{iid}{\sim} Ga(\alpha, \beta) \\ \beta &\sim Ga(\gamma, \delta),\end{aligned}$$

όπου  $\alpha = 1.8$ ,  $\gamma = 0.01$  και  $\delta = 1$ . Η από κοινού εκ των υστέρων κατανομή είναι

$$\begin{aligned}&\pi(\lambda_1, \dots, \lambda_{10}, \beta | x_1, \dots, x_{10}) \\ &\propto P(x_1, \dots, x_{10} | \lambda_1, \dots, \lambda_{10}, \beta) P(\lambda_1, \dots, \lambda_{10}, \beta) \\ &\propto P(x_1, \dots, x_{10} | \lambda_1, \dots, \lambda_{10}, \beta) P(\lambda_1, \dots, \lambda_{10} | \beta) P(\beta) \\ &\propto \left\{ \prod_{i=1}^{10} P(x_i | \lambda_i) \right\} \left\{ \prod_{i=1}^{10} P(\lambda_i | \beta) \right\} P(\beta) \\ &\propto \prod_{i=1}^{10} \{ e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i} \beta^\alpha \lambda_i^{\alpha-1} e^{-\beta \lambda_i} \} \beta^{\gamma-1} e^{-\delta \beta} \\ &\propto \prod_{i=1}^{10} \{ \lambda_i^{x_i + \alpha - 1} e^{-(t_i + \beta) \lambda_i} \} \beta^{10\alpha + \gamma - 1} e^{-\delta \beta}.\end{aligned}\tag{1.2}$$

### Παράδειγμα 1.0.2 Pump Failures (συνέχεια...)

Υποθέτουμε τις εκ των προτέρων κατανομές

$$\begin{aligned}\lambda_i &\stackrel{iid}{\sim} Ga(\alpha, \beta) \\ \beta &\sim Ga(\gamma, \delta) \\ \alpha &\sim Exp(1).\end{aligned}$$



Τώρα, η από κοινού εκ των υστέρων κατανομή των παραμέτρων  $\lambda_1, \dots, \lambda_{10}, \beta, \alpha$  είναι

$$\begin{aligned}
& \pi(\lambda_1, \dots, \lambda_{10}, \beta, \alpha | x_1, \dots, x_{10}) \\
& \propto P(x_1, \dots, x_{10} | \lambda_1, \dots, \lambda_{10}, \beta, \alpha) P(\lambda_1, \dots, \lambda_{10}, \beta, \alpha) \\
& \propto P(x_1, \dots, x_{10} | \lambda_1, \dots, \lambda_{10}, \beta, \alpha) \\
& \quad P(\lambda_1, \dots, \lambda_{10} | \beta, \alpha) P(\beta | \alpha) P(\alpha) \\
& \propto P(x_1, \dots, x_{10} | \lambda_1, \dots, \lambda_{10}, \beta, \alpha) \\
& \quad P(\lambda_1, \dots, \lambda_{10} | \beta, \alpha) P(\beta) P(\alpha) \\
& \propto \left\{ \prod_{i=1}^{10} P(x_i | \lambda_i) \right\} \left\{ \prod_{i=1}^{10} P(\lambda_i | \beta, \alpha) \right\} P(\beta) P(\alpha) \\
& \propto \prod_{i=1}^{10} \left\{ e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta \lambda_i} \right\} \beta^{\gamma-1} e^{-\delta \beta} e^{-\alpha} \\
& \propto \prod_{i=1}^{10} \left\{ \lambda_i^{x_i + \alpha - 1} e^{-(t_i + \beta) \lambda_i} \right\} \frac{\beta^{10\alpha + \gamma - 1}}{(\Gamma(\alpha))^{10}} e^{-\delta \beta - \alpha}. \tag{1.3}
\end{aligned}$$

### Παράδειγμα 1.0.3 *Random Effects Model*

Το παράδειγμα που ακολουθεί αναφέρεται σε ένα Μπεϋζιανό μοντέλο τυχαίων επιδράσεων με συζυγείς εκ των προτέρων κατανομές.

Ας θεωρήσουμε το μοντέλο ανάλυσης διασποράς κατά ένα παράγοντα με τυχαίες επιδράσεις.

$$Y_{ij} = \theta_i + \epsilon_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, m. \tag{1.4}$$

Για το μοντέλο αυτό υποθέτουμε ότι

$$Y_{ij} | \theta_i, \lambda_e \stackrel{iid}{\sim} N(\theta_i, \lambda_e^{-1}) \tag{1.5}$$

$$\theta_i | \mu, \lambda_\theta \stackrel{iid}{\sim} N(\mu, \lambda_\theta^{-1}) \tag{1.6}$$

$$\lambda_e \sim Ga(a_2, b_2) \tag{1.7}$$

$$\mu \sim N(\mu_0, \lambda_0^{-1}) \tag{1.8}$$

$$\lambda_\theta \sim Ga(a_1, b_1), \tag{1.9}$$

όπου  $(\theta_1, \dots, \theta_K) | \mu, \lambda_\theta$  και  $\lambda_e$  είναι ανεξάρτητες. Το ίδιο ισχύει για τις  $\mu$  και  $\lambda_\theta$ .

Η από κοινού εκ των υστέρων κατανομή των παραμέτρων αυτών είναι:

$$\begin{aligned} & \pi(\theta_1, \dots, \theta_K, \mu, \lambda_e, \lambda_\theta | \mathbf{y}) \\ & \propto P(\mathbf{y} | \theta_1, \dots, \theta_K, \mu, \lambda_e, \lambda_\theta) \pi(\theta_1, \dots, \theta_K, \mu, \lambda_e, \lambda_\theta) \\ & \propto P(\mathbf{y} | \theta_1, \dots, \theta_K, \lambda_e, \lambda_\theta) \pi(\theta_1, \dots, \theta_K | \mu, \lambda_\theta) \pi(\mu) \pi(\lambda_e) \pi(\lambda_\theta) \end{aligned} \quad (1.10)$$

Η από κοινού εκ των υστέρων κατανομή (1.10) αναλύεται ως εξής:

1.  $P(\mathbf{y} | \theta_1, \dots, \theta_K, \lambda_e) = \prod_{i=1}^K \prod_{j=1}^m (2\pi\lambda_e^{-1})^{-1/2} \exp\left\{-\frac{(y_{ij} - \theta_i)^2}{2\lambda_e^{-1}}\right\}$   
 $\propto \lambda_e^{\frac{mK}{2}} \exp\left\{-\frac{1}{2}\lambda_e \sum_{i=1}^K \sum_{j=1}^m (y_{ij} - \theta_i)^2\right\}.$
2.  $\pi(\theta_1, \dots, \theta_K | \mu, \lambda_\theta) = \prod_{i=1}^K (2\pi\lambda_\theta^{-1})^{-1/2} \exp\left\{-\frac{(\theta_i - \mu)^2}{2\lambda_\theta^{-1}}\right\}$   
 $\propto \lambda_\theta^{\frac{K}{2}} \exp\left\{-\frac{1}{2}\lambda_\theta \sum_{i=1}^K (\theta_i - \mu)^2\right\}.$
3.  $\pi(\mu) = (2\pi\lambda_0^{-1})^{-1/2} \exp\left\{-\frac{(\mu - \mu_0)^2}{2\lambda_0^{-1}}\right\} \propto \exp\left\{-\frac{1}{2}\lambda_0 (\mu - \mu_0)^2\right\}.$
4.  $\pi(\lambda_e) \propto \lambda_e^{a_2-1} e^{-b_2\lambda_e}.$
5.  $\pi(\lambda_\theta) \propto \lambda_\theta^{a_1-1} e^{-b_1\lambda_\theta}.$

Τώρα η σχέση (1.10) λαμβάνει τη μορφή:

$$\begin{aligned} & \pi(\theta_1, \dots, \theta_K, \mu, \lambda_e, \lambda_\theta | \mathbf{y}) \\ & \propto \lambda_e^{\frac{mK}{2} + a_2 - 1} \lambda_\theta^{\frac{K}{2} + a_1 - 1} \exp\left\{-\frac{1}{2}\lambda_e \sum_{i=1}^K \sum_{j=1}^m (y_{ij} - \theta_i)^2 - \frac{1}{2}\lambda_\theta \sum_{i=1}^K (\theta_i - \mu)^2\right\} \\ & \propto \exp\left\{-\frac{1}{2}\lambda_0 (\mu - \mu_0)^2\right\} e^{-b_1\lambda_\theta} e^{-b_2\lambda_e}. \end{aligned} \quad (1.11)$$

Η εφαρμογή των MCMC μεθόδων προσομοίωσης αποδεικνύεται πολύ βολική στην περίπτωση των πολυδιάστατων κατανομών άγνωστων παραμέτρων, όπως αυτές που παρουσιάστηκαν στα προηγούμενα παραδείγματα. Γενικά τέτοια προβλήματα συναντάμε στην Μπεύζιανή Στατιστική.

Οι MCMC αλγόριθμοι βασίζονται στην ανάπτυξη αλυσίδων Markov. Προσομοιώνουμε τιμές, οι οποίες θεωρούνται ως πραγματοποίηση μιας αλυσίδας

Markov. Η τιμή μιας τυχαίας μεταβλητής της αλυσίδας περιγράφεται από ένα πιθανοτικό νόμο που βασίζεται στην τιμή της αμέσως προηγούμενης τυχαίας μεταβλητής της αλυσίδας. Η διαδικασία της προσομοίωσης συνεχίζεται και οριακά ή αλυσίδα συγκλίνει στη στάσιμη κατανομή της. Αυτό σημαίνει ότι, οριακά, οι τιμές των τυχαίων μεταβλητών που προσομοιώνονται, θεωρούνται ότι προέρχονται από την ίδια (οριακή) κατανομή.

Πως όμως συνδέεται η παραπάνω διαδικασία με την Μπεϋζιανή Στατιστική; Η ιδέα είναι απλή: θεωρούμε την κατανομή που θέλουμε να προσομοιώσουμε (κατανομή στόχος) ως την οριακή (στάσιμη) κατανομή της αλυσίδας Markov που πρόκειται να προσομοιωθεί. Έτσι, όταν η αλυσίδα συγκλίνει, οι τιμές που προσομοιώνονται προέρχονται από την οριακή κατανομή της και επομένως από την κατανομή στόχο.

Η εισαγωγή των αλυσίδων Markov στα σχέδια προσομοίωσης είναι ουσιώδους σημασίας. Είπαμε ότι επιτρέπει το χειρισμό πολυδιάστατων κατανομών. Ωστόσο υπάρχουν κάποια ζητήματα που πρέπει να ληφθούν υπόψη κατά τη διαδικασία της προσομοίωσης. Ήδη αναφέραμε ότι οι προσομοιωμένες τιμές θεωρούνται ότι προέρχονται από την κατανομή στόχο, μόνο όταν η αλυσίδα έχει συγκλίνει στην κατανομή αυτή. Πότε όμως γίνεται αυτό; Πώς μπορεί κάποιος να είναι βέβαιος ότι η αλυσίδα έχει συγκλίνει και να άρει το δείγμα του με την πεποίθηση ότι πραγματικά είναι ένα δείγμα από την κατανομή στόχο; Ένα πλήθος τιμών που προσομοιώνονται είναι άχρηστες, αφού δεν προέρχονται από την στάσιμη κατανομή της αλυσίδας και το ερώτημα είναι πόσες τιμές χρειάζεται να πετάξουμε. Ευτυχώς, στη θεωρία των αλυσίδων Markov υπάρχουν ανάλογα του νόμου των μεγάλων αριθμών και του κεντρικού οριακού θεωρήματος, που εξασφαλίζουν ότι οι περισσότερες τιμές που προσομοιώνονται από μία αλυσίδα Markov μπορούν να χρησιμοποιηθούν ώστε να παράσχουν πληροφορίες σχετικά με την κατανομή του ενδιαφέροντός μας.

Υπάρχει ένα ερώτημα που προκύπτει ύστερα από αυτή την πανοραμική άποψη της MCMC θεωρίας. Πως μπορούμε να χτίσουμε μια αλυσίδα Markov της οποίας η οριακή κατανομή να είναι ακριβώς η κατανομή στόχος, δηλαδή η κατανομή των άγνωστων ποσοτήτων, που προκύπτει από τη Μπεϋζιανή ανάλυση; Είναι εκπληκτικό το ότι όχι μόνο είναι εφικτό κάτι τέτοιο, αλλά υπάρχουν και πολλοί τρόποι πραγματοποίησής του.

Ένας τρόπος είναι ο Metropolis - Hastings αλγόριθμος. Αυτός βασίζεται σε μια αλυσίδα Markov της οποίας η εξάρτηση από την προηγούμενη τιμή διακρίνεται σε δύο σημεία: μία τιμή πρότασης και την απόδοσή αυτής της τιμής. Στον αλγόριθμο λοιπόν, έχουμε ένα βήμα όπου προτείνεται αυθέραιτα μία τιμή, ως πιθανή νέα τιμή της αλυσίδας και ένα άλλο βήμα που εξασφαλίζει την απο-

δοχή μόνο εκείνων των τιμών που εξασφαλίζουν τη διατήρηση της κατάλληλης οριακής τροχιάς (trajectory) της αλυσίδας.

Ένας άλλος τρόπος είναι ο δειγματολήπτης Gibbs. Βασίζεται σε μια αλυσίδα Markov της οποίας η εξάρτηση από την προηγούμενη τιμή οφείλεται στις πλήρεις δεσμευμένες κατανομές που προκύπτουν από το μοντέλο. Είναι πολύ συχνό το φαινόμενο μοντέλα με πολύπλοκες από κοινού κατανομές να δίνουν πλήρεις δεσμευμένες κατανομές σχετικά απλής μορφής. Αυτό ακριβώς το σημείο εκμεταλλεύεται ο δειγματολήπτης Gibbs και μπορεί να δώσει απλές λύσεις σε πολλά προβλήματα.

Τέλος, ύστερα από αυτή τη συνοπτική παρουσίαση, μπορούμε να θέσουμε και ένα άλλο ερωτήμα: ποια είναι η αποδοτικότητα του MCMC αλγόριθμου που χρησιμοποιούμε; Την αποδοτικότητα μπορούμε να τη μετρήσουμε σε όρους ευκολίας με την οποία προσομοιώνουμε ένα δείγμα από την κατανομή στόχο. Αυτό όμως εξαρτάται από την επιλογή της συνάρτησης πρότασης που χρησιμοποιούμε, στην περίπτωση του Metropolis - Hastings αλγόριθμου, ή των πλήρεις δεσμευμένων κατανομών, στην περίπτωση του δειγματολήπτη Gibbs.

Η παρούσα εργασία διαπραγματεύεται όλα αυτά τα ζητήματα, καθώς και πλήθος άλλων σχετικών με τη θεωρία των MCMC μεθόδων προσομοίωσης.

# Κεφάλαιο 2

## Αλυσίδες Markov

### 2.1 Ορισμός

Μια αλυσίδα Markov  $(X_n)_{n \in \mathbb{Z}_+}$  είναι ένα συγκεκριμένο είδος στοχαστικής διαδικασίας λαμβάνοντας, στο χρόνο  $n \in \mathbb{Z}_+$ , τιμή  $X_n$  από ένα χώρο καταστάσεων  $\mathcal{X}$ . Οι τυχαίες μεταβλητές  $\{X_0, X_1, \dots\}$  λαμβάνουν τιμές από το σύνολο  $\mathcal{X}$  και θεωρούνται μετρήσιμες αναφορικά με κάποια δεδομένη σ-άλγεβρα  $\mathcal{B}(\mathcal{X})$ . Γενικά, συμβολίζουμε με  $x, y, z, \dots$  τα στοιχεία του  $\mathcal{X}$  και με  $A, B, C, \dots$  τα στοιχεία του  $\mathcal{B}(\mathcal{X})$ .

Αντιμετωπίζοντας τη στοχαστική διαδικασία σαν οντότητα, θεωρούμε τιμές (πραγματοποιήσεις) ολόκληρης της αλυσίδας  $\mathbf{X} = \{X_0, X_1, \dots\}$  προερχόμενες από το σύνολο που σχηματίζεται από το μετρήσιμο γινόμενο  $\Omega = \mathcal{X}^\infty = \prod_{i=0}^{\infty} \mathcal{X}_i$ , όπου κάθε  $\mathcal{X}_i$  είναι ένα αντίγραφο του  $\mathcal{X}$  στο οποίο αντιστοιχεί ένα αντίγραφο του  $\mathcal{B}(\mathcal{X})$ . Για να οριστεί η  $\mathbf{X}$  ως τυχαία μεταβλητή πρέπει να αντιστοιχίσουμε στο  $\Omega$  μια σ-άλγεβρα  $\mathcal{F}$  και για κάθε κατάσταση  $x \in \mathcal{X}$ , θεωρούμενη ως αρχική κατάσταση στην πραγματοποίηση (sample path) της στοχαστικής διαδικασίας, ένα μέτρο πιθανότητας  $P_x$  τέτοιο ώστε η πιθανότητα του ενδεχομένου  $\{\mathbf{X} \in A\}$  είναι καλά ορισμένη για κάθε σύνολο  $A \in \mathcal{F}$ . Η αρχική συνθήκη απαιτεί, φυσικά,  $P_x(X_0 = x) = 1$ . Η τριάδα  $(\Omega, \mathcal{F}, P_x)$  ορίζει μια στοχαστική διαδικασία.

Πριν προχωρήσουμε σε συγκεκριμένες λεπτομέριες σχετικά με τους νόμους πιθανοτήτων που διέπουν την κίνηση μιας αλυσίδας Markov  $\mathbf{X}$ , χρειάζεται να αποσαφηνίσουμε τη δομή χώρου καταστάσεων  $\mathcal{X}$  από τον οποίο προέρχονται οι τιμές της.

**Ορισμός 2.1.1** Ο χώρος καταστάσεων  $\mathcal{X}$  λέγεται διακριτός (countable space)

αν το σύνολο  $\mathcal{X}$  είναι διακριτό, με πεπερασμένο ή αριθμήσιμο πλήθος στοιχείων, και με το σύνολο  $\mathcal{B}(\mathcal{X})$  ως  $\sigma$ -άλγεβρα όλων των υποσυνόλων του  $\mathcal{X}$ .

Ο χώρος καταστάσεων  $\mathcal{X}$  λέγεται γενικός (*general space*) αν το σύνολο  $\mathcal{X}$  είναι συνεχές, με άπειρο ή μη αριθμήσιμο πλήθος στοιχείων, και με το σύνολο  $\mathcal{B}(\mathcal{X})$  ως  $\sigma$ -άλγεβρα όλων των υποσυνόλων του  $\mathcal{X}$ .

**Ορισμός 2.1.2** Μια ακολουθία τυχαίων μεταβλητών  $X_0, X_1, X_2, \dots$ , ορισμένων στον ίδιο χώρο πιθανότητας  $(\Omega, \mathcal{F}, P_x)$  και με τιμές στο χώρο  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$  λέμε ότι είναι αλυσίδα Markov αν

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

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

**Ορισμός 2.1.3** Αν

$$P(X_{n+1} \in A | X_n = x_n) = P(X_1 \in A | X_0 = x_0), \quad \forall A \in \mathcal{B}(\mathcal{X}), n \geq 1,$$

τότε η αλυσίδα λέγεται ομογενής.

## 2.2 Θεμελίωση σε Γενικό Χώρο Καταστάσεων

Έστω  $\mathcal{X}$  ένας γενικός χώρος καταστάσεων και  $\mathcal{B}(\mathcal{X})$  μια μετρήσιμα παραχθείσα  $\sigma$ -άλγεβρα πάνω στο  $\mathcal{X}$ . Ας ξεκινήσουμε με τις πιθανότητες μετάβασης.

**Ορισμός 2.2.1** Αν  $P = \{P(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$  είναι τέτοιο ώστε

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

<sup>1</sup>Έστω  $(\mathcal{X}_1, \mathcal{B}(\mathcal{X}_1))$  και  $(\mathcal{X}_2, \mathcal{B}(\mathcal{X}_2))$  δύο μετρήσιμοι χώροι (measurable spaces). Η απεικόνιση  $h : \mathcal{X}_1 \rightarrow \mathcal{X}_2$  λέγεται μετρήσιμη συνάρτηση (measurable function) αν

$$h^{-1}\{B\} := \{x : h(x) \in B\} \in \mathcal{B}(\mathcal{X}_1),$$

για κάθε σύνολο  $B \in \mathcal{B}(\mathcal{X}_2)$ .

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

Από τον πυρήνα μετάβασης  $P$  μπορούμε να ορίσουμε μια στοχαστική διαδικασία με τις κατάλληλες Μαρκοβιανές ιδιότητες, για την οποία η πιθανότητα  $P$  περιγράφει τον πιθανοθεωρητικό νόμο μετάβασης πρώτης τάξης.

Έστω μια πεπερασμένη ακολουθία  $\mathbf{X} = \{X_0, \dots, X_n\}$  τυχαίων μεταβλητών, ορισμένων στο σύνολο  $\mathcal{X}^{n+1} = \prod_{i=0}^n \mathcal{X}_i$ , στο οποίο αντιστοιχεί η σ-άλγεβρα  $\bigcup_{i=0}^n \mathcal{B}(\mathcal{X}_i)$ . Για όλα τα μετρήσιμα σύνολα  $A_i \subseteq \mathcal{X}_i$  αναπτύσσουμε ένα σύνολο συναρτήσεων  $P_x^n$  ορισμένων πάνω στο  $\mathcal{X}^{n+1}$  θέτοντας, για στεθερό σημείο εκκίνησης  $x \in \mathcal{X}$  και για τα σύνολα  $A_1 \times \dots \times A_n$

$$\begin{aligned} P_x^1(A_1) &= P(x, A_1) \\ P_x^2(A_1 \times A_2) &= \int_{A_1} P(x, dy_1) P(y_1, A_2) \\ &\vdots \\ P_x^n(A_1 \times \dots \times A_n) &= \int_{A_1} P(x, dy_1) \int_{A_2} P(y_1, dy_2) \dots \\ &\quad \int_{A_{n-1}} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A_n) \end{aligned}$$

Οι παραπάνω πιθανότητες είναι καλά ορισμένες. Αυτό εξασφαλίζεται από τη μετρησιμότητα της προς ολοκλήρωση έκφρασης  $P(\cdot, \cdot)$ , ως στην πρώτη μεταβλητή και το γεγονός ότι οι πυρήνες είναι μέτρα πιθανότητας ως προς τη δεύτερη μεταβλητή.

**Ορισμός 2.2.2** Η κατανομή  $\mu_0$  της  $X_0$ , δηλαδή το μέτρο πιθανότητας που ορίζεται ως

$$\mu_0 := P(X_0 \in A)$$

λέγεται **αρχική κατανομή της αλυσίδας**. Σε περίπτωση που η αλυσίδα ξεκινήσει από μια σταθερή τιμή  $x$  (που δεν έχει προκύψει τυχαία) τότε η  $\mu_0$  είναι η κατανομή *Dirac* στο  $x$  με μάζα πιθανότητας

$$\delta_x(y) := \begin{cases} 1, & y = x, \\ 0, & y \neq x. \end{cases}$$

**Ορισμός 2.2.3** Αν  $\mu_0$  είναι η αρχική κατανομή (που είναι δυνατόν να είναι και Dirac) τότε η κατανομή  $\mu_n$  της  $X_n$  δηλαδή το μέτρο πιθανότητας που ορίζεται ως

$$\mu_n(A) := P(X_n \in A),$$

δίνεται από τη σχέση

$$\mu_n(A) = \int \mu_0(dx) P^n(x, A).$$

Για δεδομένη αρχική κατανομή  $\mu_0$ , η  $\mu_n$  προκύπτει από τον αναδρομικό τύπο

$$\mu_n(A) = \int \mu_{n-1}(dx) P(x, A).$$

Χάριν απλότητας συμβολίζουμε την αρχική κατανομή μιας αλυσίδας ως  $\mu$ .

**Θεώρημα 2.2.1** Για κάθε αρχικό μέτρο πιθανότητας  $\mu$  πάνω στο  $\mathcal{B}(\mathcal{X})$  και κάθε πυρήνα μετάβασης  $P = \{P(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ , υπάρχει μια στοχαστική διαδικασία  $\mathbf{X} = \{X_0, X_1, \dots\}$  ορισμένη στο  $\Omega = \prod_{i=0}^{\infty} \mathcal{X}_i$ , μετρήσιμη με  $\sigma$ -άλγεβρα  $\mathcal{F} = \bigcup_{i=0}^{\infty} \mathcal{B}(\mathcal{X}_i)$  και ένα μέτρο πιθανότητας  $P_\mu$  πάνω στο  $\mathcal{F}$  έτσι ώστε  $P_\mu(B)$  είναι η πιθανότητα του ενδεχομένου  $\{\mathbf{X} \in B\}$  για  $B \in \mathcal{F}$ . Επίσης, για κάποιο μετρήσιμο σύνολο  $A_i \subseteq \mathcal{X}_i$ ,  $i = 0, \dots, n$  και για κάθε  $n$

$$\begin{aligned} & P_\mu(X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n) \\ &= \int_{y_0 \in A_0} \cdots \int_{y_{n-1} \in A_{n-1}} \mu(dy_0) P(y_0, dy_1) \cdots P(y_{n-1}, A_n) \end{aligned} \quad (2.1)$$

Μπορούμε τώρα να ορίσουμε τις αλυσίδες Markov σε γενικό χώρο καταστάσεων.

**Ορισμός 2.2.4** Η στοχαστική διαδικασία  $\mathbf{X}$ , ορισμένη στο χώρο  $(\Omega, \mathcal{F})$  λέγεται ομογενής αλυσίδα Markov με πυρήνα μετάβασης  $P(x, A)$  και αρχική κατανομή  $\mu$ , αν η από κοινού κατανομή της πεπερασμένης ακολουθίας τυχαίων μεταβλητών  $\mathbf{X}$  ικανοποιεί τη σχέση (2.1).

Ο πυρήνας μετάβασης  $n$ -τάξης (**n-step transition probability kernel**) ορίζεται επαναληπτικά. Θέτουμε  $P^0(x, A) = \delta_x(A)$ , το μέτρο Dirac που ορίζεται από τον τύπο

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



και, για  $n \geq 1$  ορίζουμε επαγωγικά

$$P^n(x, A) = \int_{\mathcal{X}} P(x, dy) P^{n-1}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}). \quad (2.2)$$

Γράφουμε  $P^n$  για τον πυρήνα μετάβασης  $n$ -τάξης και εννοούμε τις πιθανότητες  $\{P^n(x, A), x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ . Εναλλακτικά μπορούμε να τον ορίσουμε ως  $P^n(x, A) = P(X_n \in A | X_0 = x)$ . Ακολουθεί η περίφημη **εξίσωση Chapman-Kolmogorov**.

**Θεώρημα 2.2.2** Για κάθε  $m$  με  $0 \leq m \leq n$ ,

$$P^n(x, A) = \int_{\mathcal{X}} P^m(x, dy) P^{n-m}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}) \quad (2.3)$$

Η ερμηνεία της σχέσης (2.3) είναι η εξής: καθώς η αλυσίδα κινείται από την αρχική κατάσταση  $x$  στο σύνολο  $A$  σε  $n$  βήματα, σε οποιοδήποτε ενδιάμεσο βήμα  $m$  θα πάρει κάποια τιμή  $y \in \mathcal{X}$ . Και επειδή πρόκειται για αλυσίδα Markov, ξεχνάει το παρελθόν της στο χρόνο  $m$  και συνεχίζει στα επόμενα  $n - m$  βήματα με τον πιθανοθεωρητικό νόμο που διέπει την κίνησή της προσαρμοσμένο στο νέο σημείο εκκίνησης  $y$ . Εναλλακτικά, η εξίσωση (2.3) γράφεται

$$P_x(X_n \in A) = \int_{\mathcal{X}} P_x(X_m \in dy) P_y(X_{n-m} \in A). \quad (2.4)$$

Όπως ο πυρήνας μετάβασης πρώτης τάξης περιγράφει μια αλυσίδα  $\mathbf{X}$ , ο πυρήνας μετάβασης  $m$ -τάξης, απομονωμένος, ικανοποιεί τον ορισμό ενός πυρήνα μετάβασης και επομένως προσδιορίζει μια αλυσίδα Markov  $\mathbf{X}^m = \{X_n^m\}$ , με πιθανότητες μετάβασης

$$P_x(X_n^m \in A) = P^{mn}(x, A) \quad (2.5)$$

**Ορισμός 2.2.5** Η αλυσίδα  $\mathbf{X}^m$  με νόμο μετάβασης που δίνεται από τη σχέση (2.5) λέγεται  **$m$ -σκελετός ( $m$ -skeleton)** της αλυσίδας  $\mathbf{X}$ .

Στους μέχρι τώρα δοθέντες ορισμούς, χρησιμοποιήθηκαν πιθανότητες ενδεχομένων που αφορούν την αλυσίδα  $\mathbf{X}$ . Τώρα δίνουμε τον τελεστή της αναμενόμενης τιμής  $E_\mu$ , ο οποίος αντιστοιχεί στην πιθανότητα  $P_\mu$ . Ορίζουμε τον τελεστή  $E_\mu$  από τον τύπο

$$E_\mu [I_{A_0 \times \dots \times A_n}(X_0, \dots, X_n)] := P_\mu(\{X_0, \dots, X_n\} \in A_0 \times \dots \times A_n),$$

όπου  $I_B$  δηλώνει τη δείκτρια συνάρτηση του συνόλου  $B$ . Μπορούμε να επεκταθούμε στον ορισμό του  $E_\mu[h(X_0, X_1, \dots)]$ , για κάθε μετρήσιμη, φραγμένη και πραγματική συνάρτηση  $h$  που ορίζεται στο  $\Omega$ , με την προϋπόθεση ότι ο τελεστής  $E_\mu$  είναι γραμμικός.

Η από κοινού κατανομή και οι περιθωριακές κατανομές μιας αλυσίδας Markov  $\mathbf{X}$  στο χρόνο  $n$  αποτελούν τα βασικά εργαλεία για την περιγραφή της κίνησής της. Ωστόσο, η ανάλυση της συμπεριφοράς της  $\mathbf{X}$  συμπεριλαμβάνει τη μελέτη των κατανομών της σε συγκεκριμένες τυχαίες χρονικές στιγμές κατά την εξέλιξή της. Εισάγουμε αυτές τις χρονικές στιγμές τώρα.

**Ορισμός 2.2.6** 1. Για κάθε σύνολο  $A \in \mathcal{B}(\mathcal{X})$ , ο **αριθμός διελεύσεων (occupation time)**  $\eta_A$  είναι ο αριθμός των διελεύσεων της  $\mathbf{X}$  από το σύνολο  $A$ , έπειτα από τη χρονική στιγμή μηδέν και δίνεται από τον τύπο

$$\eta_A := \sum_{n=1}^{\infty} I(X_n \in A) \quad (2.6)$$

2. Για κάθε σύνολο  $A \in \mathcal{B}(\mathcal{X})$ , οι μεταβλητές

$$\tau_A := \min(n \geq 1 : X_n \in A) \quad (2.7)$$

$$\sigma_A := \min(n \geq 0 : X_n \in A) \quad (2.8)$$

λέγονται **χρόνος πρώτης επανόδου στο  $A$**  και **χρόνος πρώτης διέλευσης από το  $A$** , αντίστοιχα.

Για κάθε  $A \in \mathcal{B}(\mathcal{X})$ , οι  $\eta_A$ ,  $\tau_A$  και  $\sigma_A$  είναι προφανώς μετρήσιμες συναρτήσεις από το  $\Omega$  στο  $\mathbb{Z}_+ \cup \{\infty\}$ .

Αν δεν χρειάζεται να διακρίνουμε μεταξύ διαφορετικών διελεύσεων της  $\mathbf{X}$  από κάποιο σύνολο  $A$ , ονομάζουμε τις μεταβλητές  $\tau_A$  και  $\sigma_A$  **χρόνος πρώτης επιστροφής στο  $A$**  και **χρόνος πρώτης διέλευσης από το  $A$** , αντίστοιχα. Αν όμως επιθυμούμε τη διάκριση μεταξύ διαφορετικών διελεύσεων της  $\mathbf{X}$  από το σύνολο  $A$ , τότε γράφουμε  $\tau_A(k)$  για τον τυχαίο χρόνο της  $k$ -επανόδου της αλυσίδας από το  $A$ . Οι μεταβλητές αυτές ορίζονται επαγωγικά, για κάθε  $A$ , από τους τύπους

$$\begin{aligned} \tau_A(1) &:= \tau_A \\ \tau_A(k) &:= \min(n > \tau_A(k-1) : X_n \in A). \end{aligned} \quad (2.9)$$

**Πρόταση 2.2.1** 1. Για κάθε  $x \in \mathcal{X}$ ,  $A \in \mathcal{B}(\mathcal{X})$

$$P_x(\tau_A = 1) = P(x, A),$$

και επαγωγικά για  $n > 1$

$$\begin{aligned} P_x(\tau_A = n) &= \int_{A^c} P(x, dy) P_y(\tau_A = n - 1) \\ &= \int_{A^c} P(x, dy_1) \int_{A^c} P(y_1, dy_2) \cdots \int_{A^c} P(y_{n-2}, dy_{n-1}) P(y_{n-1}, A). \end{aligned}$$

2. Για κάθε  $x \in \mathcal{X}$ ,  $A \in \mathcal{B}(\mathcal{X})$

$$P_x(\sigma_A = 0) = I_A(x),$$

και για  $n \geq 1$ ,  $x \in A^c$

$$P_x(\sigma_A = n) = P_x(\tau_A = n).$$

Τέλος, η ανάλυση της αλυσίδας  $\mathbf{X}$  εμπλέκει τη συνάρτηση

$$U(x, A) := \sum_{n=1}^{\infty} P^n(x, A) = E_x[\eta_A] \quad (2.10)$$

η οποία είναι μια απεικόνιση (mapping) από το χώρο  $\mathcal{X} \times \mathcal{B}(\mathcal{X})$  στο χώρο  $\mathbb{R} \cup \{\infty\}$  και την πιθανότητα πρώτης επανόδου της αλυσίδας στην κατάσταση  $A$

$$L(x, A) := P_x(\tau_A < \infty) \quad (2.11)$$

η οποία δηλώνει την πιθανότητα η  $\mathbf{X}$  να επιστρέψει στην κατάσταση  $A$  σε κάποια χρονική στιγμή, ενώ αρχικά βρισκόταν στην κατάσταση  $x$ .

## 2.3 $\psi$ -Αναγωγιμότητα

Η μοντέρνα θεωρία πάνω στις αλυσίδες Markov με γενικό χώρο καταστάσεων  $\mathcal{X}$  εισάγει την έννοια της  $\varphi$ -αναγωγιμότητας.

**Ορισμός 2.3.1** Ονομάζουμε την αλυσίδα  $\mathbf{X}$   $\varphi$ -ανάγωγη ( $\varphi$ -irreducible) αν υπάρχει μέτρο  $\varphi$  (όχι απαραίτητα μέτρο πιθανότητας) ορισμένο στο πεδίο  $\mathcal{B}(\mathcal{X})$  τέτοιο ώστε, οποτεδήποτε  $\varphi(A) > 0$ , έχουμε και  $L(x, A) > 0$ ,  $\forall x \in \mathcal{X}$ .

**Πρόταση 2.3.1** Οι επόμενες προτάσεις είναι ισοδύναμες με  $\varphi$ -αναγωγιμότητα:

1. Για κάθε  $x \in \mathcal{X}$ , οποτεδήποτε  $\varphi(A) > 0$ ,  $U(x, A) > 0$ .
2. Για κάθε  $x \in \mathcal{X}$ , οποτεδήποτε  $\varphi(A) > 0$ , υπάρχει  $n > 0$ , πιθανώς εξαρτώμενο από τα  $A, x$ , τέτοιο ώστε  $P^n(x, A) > 0$ .

Ο ορισμός της  $\varphi$ -αναγωγιμότητας εγκυάται ότι «μεγάλα» σύνολα (μη μηδενικά σύνολα ως προς το μέτρο  $\varphi$ ), όπως μετρώνται από τη  $\varphi$ , είναι πάντοτε προσβάσιμα από την αλυσίδα με θετική πιθανότητα, ανεξάρτητα από το σημείο εκκίνησής της. Σε πολλές περιπτώσεις χρειάζεται να γνωρίζουμε το αντίστροφο συμπέρασμα: «αμαλητέα» σύνολα  $B$ , με την έννοια ότι  $\varphi(B) = 0$ , αποφεύγονται από την αλυσίδα με πιθανότητα ένα, για τα περισσότερα σημεία εκκίνησης.

Η έννοια της  $\varphi$ -αναγωγιμότητας επεκτείνεται κατά τέτοιο τρόπο, ώστε οδηγούμαστε στη έννοια του **μέγιστου ανάγωγου μέτρου** ("maximal" irreducibility measure). Το μέγιστο ανάγωγο μέτρο προσδιορίζει το εύρος της αλυσίδας πολύ πιο ολοκληρωμένα από άλλα, περισσότερο αυθαίρετα, ανάγωγα μέτρα.

**Πρόταση 2.3.2** Αν η αλυσίδα  $\mathbf{X}$  είναι  $\varphi$ -ανάγωγη για κάποιο μέτρο  $\varphi$ , τότε υπάρχει ένα μέτρο πιθανότητας  $\psi$ , ορισμένο στο πεδίο  $\mathcal{B}(\mathcal{X})$ , τέτοιο ώστε

1. Η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη ( $\psi$ -irreducible).
2. Για κάθε άλλο μέτρο  $\varphi'$ , η αλυσίδα  $\mathbf{X}$  είναι  $\varphi'$ -ανάγωγη αν και μόνο αν  $\psi \succ \varphi'$ , δηλαδή το μέτρο  $\varphi'$  είναι απολύτως συνεχές ως προς το μέτρο  $\psi$ . Αυτό σημαίνει ότι  $\forall A \in \mathcal{B}(\mathcal{X})$  με  $\psi(A) = 0$  ισχύει  $\varphi'(A) = 0$ .
3. Αν  $\psi(A) = 0$ , τότε  $\psi\{y : L(y, A) > 0\} = 0$ .

Με απλά λόγια, αν η αλυσίδα είναι  $\varphi$ -ανάγωγη για κάποιο μέτρο  $\varphi$  τότε υπάρχει κάποιο μέγιστο μέτρο  $\psi$  το οποίο καθορίζει όλα τα σύνολα στα οποία μπορεί να βρεθεί με θετική πιθανότητα η αλυσίδα. (Γενικά, η  $\varphi$ -αναγωγιμότητα 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 επισκεφτεί το  $A$ . Αντιθέτως, από την ιδιότητα (3) βλέπουμε ότι αν κάποιο σύνολο  $A$  είναι  $\psi$ -μηδενικό τότε το σύνολο των  $x \in \mathcal{X}$  από τα οποία αν ξεκινήσει θα βρεθεί κάποτε στο  $A$  με θετική πιθανότητα είναι επίσης  $\psi$ -μηδενικό.)

Υπάρχουν και άλλες προσεγγίσεις της έννοιας της αναγωγιμότητας, για τον γενικό χώρο καταστάσεων, θα επικεντρωθούμε όμως στην ιδέα της  $\varphi$ -αναγωγιμότητας και θα αναζητήσουμε συνθήκες οι οποίες εξασφαλίζουν την

ύπαρξή της. Επίσης, θα χρησιμοποιήσουμε το  $\psi$  ως ένα αυθέραιτο μέγιστο ανάγωγο μέτρο για την  $\mathbf{X}$ .

**Ορισμός 2.3.2** 1. Η αλυσίδα Markou λέγεται  $\psi$ -ανάγωγη, αν είναι  $\varphi$ -ανάγωγη για κάποιο μέτρο  $\varphi$  και το μέτρο  $\psi$  είναι μέγιστο ανάγωγο μέτρο το οποίο ικανοποιεί την Πρόταση 2.3.2.

2. Γράφουμε

$$\mathcal{B}^+(\mathcal{X}) := \{A \in \mathcal{B}(\mathcal{X}) : \psi(A) > 0\}$$

για τα σύνολα στα οποία το μέτρο  $\psi$  είναι θετικό. Η ύπαρξη μέγιστων ανάγωγων μέτρων ισοδυναμεί με το ότι το πεδίο  $\mathcal{B}^+(\mathcal{X})$  ορίζεται μοναδικά.

3. Ένα σύνολο  $A \in \mathcal{B}(\mathcal{X})$  ονομάζεται **πλήρες (full)**, αν  $\psi(A^c) = 0$ .

4. Ένα σύνολο  $A \in \mathcal{B}(\mathcal{X})$  ονομάζεται **απορροφητικό (absorbing)**, αν  $P(x, A) = 1$ , για  $x \in A$ .

Η ακόλουθη πρόταση υποδεικνύει τη σύνδεση μεταξύ απορροφητικών και πλήρων συνόλων.

**Πρόταση 2.3.3** Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη. Τότε

1. κάθε απορροφητικό σύνολο είναι πλήρες,
2. κάθε πλήρες σύνολο εμπεριέχει ένα μη κενό απορροφητικό σύνολο.

Αν κάποιο σύνολο  $C$  είναι απορροφητικό και υπάρχει μέτρο  $\psi$  για το οποίο

$$\psi(B) > 0 \Rightarrow L(x, B) > 0, \quad x \in C,$$

τότε το  $C$  λέγεται απορροφητικό  $\psi$ -ανάγωγο σύνολο.

**Πρόταση 2.3.4** Έστω  $A$  ένα απορροφητικό σύνολο.  $P_A$  είναι ο πυρήνας  $P$ , περιορισμένος στις καταστάσεις του συνόλου  $A$ . Τότε υπάρχει αλυσίδα Markou  $\mathbf{X}_A$ , με χώρο καταστάσεων το σύνολο  $A$  και πυρήνα μετάβασης τον  $P_A$ . Επιπλέον, αν η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη, τότε και η  $\mathbf{X}_A$  είναι  $\psi$ -ανάγωγη.

Οι Προτάσεις 2.3.3 και 2.3.4, εγχυώνται την αποτελεσματικότητα και λειτουργικότητα της ανάλυσης, όταν μελετούμε πλήρη σύνολα του χώρου καταστάσεων.

Εισάγουμε την έννοια της **προσιτότητας (accessibility)** και της **ομοιόμορφης προσιτότητας (uniform accessibility)**, οι οποίες ισχυροποιούν την ιδέα της επικοινωνίας (communication), πάνω στην οποία βασίζεται η έννοια της  $\psi$ -αναγωγής.

**Ορισμός 2.3.3** Λέμε ότι ένα σύνολο  $B \in \mathcal{B}(\mathcal{X})$  είναι **προσιτό (accessible)** από ένα άλλο σύνολο  $A \in \mathcal{B}(\mathcal{X})$ , αν  $L(x, B) > 0$ ,  $\forall x \in A$ .  
Λέμε ότι ένα σύνολο  $B \in \mathcal{B}(\mathcal{X})$  είναι **ομοιόμορφα προσιτό (uniformly accessible)** από ένα άλλο σύνολο  $A \in \mathcal{B}(\mathcal{X})$ , αν υπάρχει αριθμός  $\delta > 0$  έτσι ώστε

$$\inf_{x \in A} L(x, B) \geq \delta \quad (2.12)$$

και όταν ισχύει η σχέση (2.12), γράφουμε  $A \rightarrow B$ .

Η σχέση « $A \rightarrow B$ » ισχύει ομοιόμορφα για  $x \in A$ . Γενικότερα η σχέση « $\rightarrow$ » είναι μη ανακλαστική, πορόλο που μπορεί να υπάρχουν σύνολα  $A, B$  τέτοια ώστε το  $A$  να είναι ομοιόμορφα προσιτό από το  $B$  και το  $B$  να είναι ομοιόμορφα προσιτό από το  $A$ .

**Λήμμα 2.3.1** Αν  $A \rightarrow B$  και  $B \rightarrow C$ , τότε  $A \rightarrow C$ .

Εισάγουμε την παρακάτω σημειογραφία για να περιγράψουμε την επικοινωνιακή δομή μιας αλυσίδας.

**Ορισμός 2.3.4** Το σύνολο  $\bar{A} := \{x \in \mathcal{X} : L(x, A) > 0\}$  είναι το σύνολο των σημείων από τα οποία το σύνολο  $A$  είναι προσιτό.

Το σύνολο  $\bar{A}(m) := \{x \in \mathcal{X} : \sum_{n=1}^m P^n(x, A) \geq m^{-1}\}$

Το σύνολο  $A^0 := \{x \in \mathcal{X} : L(x, A) = 0\} = [\bar{A}]^c$  είναι το σύνολο των σημείων από τα οποία το σύνολο  $A$  δεν είναι προσιτό.

**Λήμμα 2.3.2** Το σύνολο  $\bar{A} = \cup_m \bar{A}(m)$ , και για κάθε  $m$  έχουμε  $\bar{A}(m) \rightarrow A$ .

Έπεται ότι, αν μια αλυσίδα είναι  $\psi$ -ανάγωγη, τότε μπορούμε να βρούμε ένα πεπερασμένο πλήθος συνόλων, τα οποία σαρώνουν το  $\mathcal{X}$  και από τα οποία κάθε άλλο σύνολο  $A \in \mathcal{B}^+(\mathcal{X})$  είναι ομοιόμορφα προσιτό, αφού σε αυτή την περίπτωση  $\bar{A} = \mathcal{X}$ .

## 2.4 Ψευδο-Άτομα και Μικρά Σύνολα

**Ορισμός 2.4.1** Ένα σύνολο  $\alpha \in \mathcal{B}(\mathcal{X})$  λέγεται **άτομο (atom)** για την  $\mathbf{X}$ , αν υπάρχει μέτρο  $\nu$ , ορισμένο στο πεδίο  $\mathcal{B}(\mathcal{X})$ , τέτοιο ώστε

$$P(x, A) = \nu(A), \quad x \in \alpha. \quad (2.13)$$

Αν η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και  $\psi(\alpha) > 0$ , τότε το  $\alpha$  λέγεται **προσιτό άτομο (accessible atom)**.

Όταν ο  $\mathcal{X}$  εμπεριέχει ένα άτομο, τότε η κατασκευή ενός μέγιστα ανάγωγου μέτρου  $\psi$  είναι εύκολη.

**Πρόταση 2.4.1** Έστω άτομο  $\alpha \in \mathcal{X}$ , τέτοιο ώστε  $\sum_n P^n(x, \alpha) > 0$ , για κάθε  $x \in \mathcal{X}$ . Τότε το  $\alpha$  είναι ένα προσιτό άτομο και η  $\mathbf{X}$  είναι  $\nu$ -ανάγωγη με  $\nu = P(\alpha, \cdot)$ .

Η ομοιόμορφη προσιτότητα  $\ll A \rightarrow B \gg$ , η οποία εισήχθει στην προηγούμενη ενότητα, απλοποιείται με την παρουσία ενός ατόμου στο χώρο καταστάσεων: αρκεί η ύπαρξη ενός συνόλου ατραπών (μονοπατιών), με θετική πιθανότητα, προς το  $A$  και η ομοιόμορφη επικοινωνία έπεται.

**Πρόταση 2.4.2** Αν  $L(x, A) > 0$  για κάποια κατάσταση  $x \in \alpha$ , όπου  $\alpha$  είναι ένα άτομο, τότε  $\alpha \rightarrow A$ .

Σε πολλές περιπτώσεις, τα άτομα του χώρου καταστάσεων είναι πραγματικά άτομα (real atoms): δηλαδή, απλά σημεία, προσβάσιμα με θετική πιθανότητα.

Ο λόγος για τον οποίο μελετάμε τα άτομα, είναι γιατί ακόμα και στην περίπτωση μιας  $\psi$ -ανάγωγης αλυσίδας στο γενικό χώρο καταστάσεων, με κατάλληλη επέκταση της πιθανοθεωρητικής δομής της αλυσίδας, μπορούμε να κατασκευάσουμε σύνολα που να παρουσιάζουν ατομική δομή. Το γεγονός αυτό επιτρέπει στην ανάλυση να ακολουθήσει τη μορφή της ανάλυσης σε διακριτό χώρο καταστάσεων, η οποία είναι, ομολογουμένως, πιο εύκολη.

Το ανέλπιστο αυτό αποτέλεσμα είναι ίσως η πιο σημαντική καινοτομία στη θεωρία των Μαρκοβιανών αλυσίδων σε γενικό χώρο καταστάσεων. Ανακαλύφθηκε σε κάπως διαφορετικές μορφές, ανεξάρτητα και σχεδόν ταυτόχρονα, από τους Nummelin και Athreya and Ney.

Ενώ στη διακριτή περίπτωση άτομο είναι κάθε πιθανή τιμή της αλυσίδας, στη συνεχή περίπτωση η ύπαρξη ατόμων είναι πιο σπανία. Αυτό συμβαίνει διότι απαιτείται η σταθερότητα του πυρήνα μετάβασης για κάποιο σύνολο θετικού μέτρου. Μια πιο ισχυρή γενίκευση είναι η λεγόμενη **συνθήκη minorizing (minorizing condition)**.

**Ορισμός 2.4.2** Για κάποιο  $\delta > 0$ , κάποιο  $C \in \mathcal{B}(\mathcal{X})$  και κάποιο μέτρο πιθανότητας  $\nu$  με  $\nu(C^c) = 0$  και  $\nu(C) = 1$

$$P(x, A) \geq \delta I_C(x) \nu(A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}) \quad (2.14)$$

Η σχέση (2.14) εξασφαλίζει ότι η αλυσίδα έχει πιθανότητες μετάβασης ομοιόμορφα κάτω φραγμένες, από πολλαπλάσια του μέτρου  $\nu$ , για κάθε  $x \in C$ . Το

ερώτημα που προκύπτει είναι αν ικανοποιείται ποτέ η συνθήκη minorizing από μια αλυσίδα. Η απάντηση είναι θετική και δίνεται στο Θεώρημα 2.4.1 παρακάτω: για κάποια σύνολα που ονομάζουμε μικρά σύνολα (small sets), τα οποία αναφέρονται σε  $\varphi$ -ανάγωγες αλυσίδες, η συνθήκη minorizing ισχύει, τουλάχιστον για τον  $m$ -σκελετό ( $m$ -skeleton) της αλυσίδας.

Τα μικρά σύνολα συμπεριφέρονται με τρόπο ανάλογο αυτού των ατόμων και συγκεκριμένα τα συμπεράσματα των προτάσεων 2.4.1 και 2.4.2 ισχύουν και για τα μικρά σύνολα. Είναι πολύ πρακτική η χρησιμοποίηση των «ψευδο-ατομικών» ιδιοτήτων των μικρών συνόλων.

**Ορισμός 2.4.3** Ένα σύνολο  $C \in \mathcal{B}(\mathcal{X})$  λέγεται **μικρό σύνολο (small set)** αν υπάρχει αριθμός  $m \in \mathbb{N}^*$  και μη μηδενικό μέτρο  $\nu_m$ , ορισμένο στο  $\mathcal{B}(\mathcal{X})$ , τέτοια ώστε για κάθε  $x \in C$ ,  $B \in \mathcal{B}(\mathcal{X})$

$$P^m(x, B) \geq \nu_m(B). \quad (2.15)$$

Όταν ισχύει η σχέση (2.15), τότε λέμε ότι το  $C$  είναι  $\nu_m$ -μικρό ( $\nu_m$ -small).

Το κύριο αποτέλεσμα της θεωρίας περί μικρών συνόλων (Θεώρημα 2.4.1), είναι ότι για μια  $\psi$ -ανάγωγη αλυσίδα, κάθε σύνολο  $A \in \mathcal{B}^+(\mathcal{X})$  εμπεριέχει κάποιο μικρό σύνολο στο  $\mathcal{B}^+(\mathcal{X})$ .

**Θεώρημα 2.4.1** Αν  $\mathbf{X}$  είναι μια  $\psi$ -ανάγωγη αλυσίδα, τότε για κάθε σύνολο  $A \in \mathcal{B}^+(\mathcal{X})$ , υπάρχει  $m \geq 1$  και ένα  $\nu_m$ -μικρό σύνολο  $C \subseteq A$ , τέτοιο ώστε  $C \in \mathcal{B}^+(\mathcal{X})$  και  $\nu_m(C) > 0$ .

Από το Θεώρημα 2.4.1 έπεται αμέσως το Θεώρημα 2.4.2.

**Θεώρημα 2.4.2** Αν  $\mathbf{X}$  είναι μια  $\psi$ -ανάγωγη αλυσίδα, τότε η minorizing συνθήκη ισχύει για κάποιο  $m$ -σκελετό της αλυσίδας.

**Πρόταση 2.4.3** 1. Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα. Τότε υπάρχει ένα πεπερασμένο πλήθος μικρών συνόλων  $C_i \in \mathcal{B}(\mathcal{X})$ , τέτοια ώστε

$$\mathcal{X} = \bigcup_{i=0}^{\infty} C_i. \quad (2.16)$$

2. Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα. Αν το σύνολο  $C \in \mathcal{B}^+(\mathcal{X})$  είναι  $\nu_n$ -μικρό, τότε μπορούμε να βρούμε  $M \in \mathbb{Z}_+$  και μέτρο  $\nu_M$ , τέτοια ώστε το σύνολο  $C$  να είναι  $\nu_M$ -μικρό και  $\nu_M(C) > 0$ .



## 2.5 Κυκλική Συμπεριφορά

Η ύπαρξη μικρών συνόλων μας επιτρέπει να δείξουμε ότι μπορούμε να πετύχουμε μια κερματοποίηση ενός γενικού χώρου καταστάσεων σε κυκλικά σύνολα, για  $\psi$ -ανάγωγες αλυσίδες.

Έστω  $C$  ένα  $\nu_M$ -μικρό σύνολο με  $\nu_M(C) > 0$ . Χωρίς βλάβη της γενικότητας, κάτι τέτοιο μπορεί να υποτεθεί από την Πρόταση 2.4.3.

Θα χρησιμοποιήσουμε το σύνολο  $C$  και το αντίστοιχο μέτρο  $\nu_M$  για να ορίσουμε ένα κύκλο για μια  $\psi$ -ανάγωγη αλυσίδα Markov. Προκειμένου να απλοποιήσουμε τη σημειογραφία, καταπνίγουμε το δείκτη  $M$  του μέτρου  $\nu$ . Επομένως, έχουμε  $P^M(x, \cdot) \geq \nu(\cdot)$ ,  $x \in C$ , και  $\nu(C) > 0$ , έτσι ώστε όταν η αλυσίδα ξεκινάει από το  $C$ , υπάρχει θετική πιθανότητα ότι θα επιστρέψει στο  $C$  τη χρονική στιγμή  $M$ .

**Θεώρημα 2.5.1** *Μια  $\psi$ -ανάγωγη αλυσίδα  $\mathbf{X}$  έχει κύκλο μήκους  $d$ , αν υπάρχει μικρό σύνολο  $C$ , ακέραιος αριθμός  $M$  και μέτρο πιθανότητας  $\nu_M$ , τέτοια ώστε το  $d$  να είναι ο μέγιστος κοινός διαιρέτης (μ.κ.δ.) του  $E_C = \{n \geq 1 : \text{το σύνολο } C \text{ είναι } \nu_n\text{-μικρό, με } \nu_n = \delta_n \nu_M, \text{ για κάποιο } \delta_n > 0\}$ .*

Η διάρκεια  $d$  του κύκλου, είναι ιδιότητα ολόκληρης της αλυσίδας  $\mathbf{X}$  και ανεξάρτητη από το μικρό σύνολο που επιλέγεται. Κάτι τέτοιο αναφέρεται στο επόμενο θεώρημα:

**Θεώρημα 2.5.2** *Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα Markov με χώρο καταστάσεων  $\mathcal{X}$ . Έστω  $C \in \mathcal{B}(\mathcal{X})$  είναι ένα  $\nu_M$ -μικρό σύνολο και  $d$  ο μέγιστος κοινός διαιρέτης του συνόλου  $E_C$ . Τότε υπάρχουν ξένα μεταξύ τους σύνολα  $D_1, \dots, D_d \in \mathcal{B}(\mathcal{X})$  (ένας  $d$ -κύκλος) τέτοια ώστε*

1. για κάθε  $x \in D_i$ ,  $P(x, D_{i+1}) = 1$ ,  $i = 0, \dots, d-1$ ,

2. Το σύνολο  $N = \left[ \bigcup_{i=1}^d D_i \right]^c$  είναι  $\psi$ -μηδενικό ( $\psi$ -null).

Ο  $d$ -κύκλος  $\{D_i\}$  είναι μέγιστος, με την έννοια ότι για κάθε άλλη συλλογή  $\{d', D'_k, k = 1, \dots, d'\}$  που ικανοποιεί τις (1)–(2), το  $d'$  διαιρεί το  $d$ . Αν  $d = d'$ , τότε, αναδιατάσσοντας τους δείκτες αν αυτο είναι απαραίτητο, έχουμε  $D_i = D'_i$ .

Από την απόδειξη του προηγούμενου θεωρήματος, της οποίας η παράθεση ξεφεύγει των σκοπών του παρόντος, γίνεται αντιληπτό ότι η διάρκεια  $d$  του κύκλου δεν εξαρτάται από το μικρό σύνολο που επιλέγεται αρχικά, εκτός ίσως από  $\psi$ -μηδενικά σύνολα. Κάθε μικρό σύνολο, πρέπει ουσιαστικά να εμπεριέχεται σε κάποιο από τα μέλη της κυκλικής κλάσης  $\{D_i\}$ .

**Ορισμός 2.5.1** Έστω  $\mathbf{X}$  μια  $\varphi$ -ανάγωγη αλυσίδα Markov. Το μεγαλύτερο  $d$  για το οποίο ένας  $d$ -κύκλος εμφανίζεται για την  $\mathbf{X}$ , λέγεται περίοδος της  $\mathbf{X}$ . Όταν  $d = 1$ , η αλυσίδα λέγεται απεριοδική (*aperiodic*). Όταν υπάρχει κάποιο  $\nu_1$ -μικρό σύνολο  $A$  με  $\nu_1(A) > 0$ , τότε η αλυσίδα λέγεται έντονα απεριοδική (*strongly aperiodic*).

Ως άμεσο επακόλουθο του Ορισμού 2.5.1 και του θεωρήματος 2.4.2 έχουμε την ακόλουθη πρόταση:

**Πρόταση 2.5.1** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα Markov.

1. Αν η  $\mathbf{X}$  είναι έντονα περιοδική, τότε η *minorizing* συνθήκη (2.14) ισχύει.
2. Αν η  $\mathbf{X}$  είναι έντονα απεριοδική, τότε κάθε σκελετός της είναι  $\psi$ -ανάγωγος και απεριοδικός, και κάποιος  $m$ -σκελετός είναι έντονα απεριοδικός.

Η Πρόταση 2.5.1 υποδηλώνει ότι είναι επιθυμητό να δουλεύουμε με έντονα απεριοδικές αλυσίδες. Δυστυχώς, κάτι τέτοιο δεν είναι εφικτό, στις περισσότερες των περιπτώσεων. Συχνά λοιπόν εξάγουμε αποτελέσματα για έντονα απεριοδικές αλυσίδες και μέσω ειδικών τεχνικών, που χρησιμοποιούν τους  $m$ -σκελετούς, επεκτείνουμε τα αποτελέσματα σε μη έντονα απεριοδικές αλυσίδες.

Η επικέντρωση της μελέτης μας σε απεριοδικές αλυσίδες, δεν είναι ουσιωδώς περιοριστική. Όπως και στη διακριτή περίπτωση, έχουμε

**Πρόταση 2.5.2** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα Markov με περίοδο  $d$  και  $d$ -κύκλο  $\{D_i, i = 1, \dots, d\}$ . Τότε κάθε ένα από τα σύνολα  $D_i$  είναι απορροφητικό  $\psi$ -ανάγωγο για την αλυσίδα  $\mathbf{X}_d$ , με πυρήνα μετάβασης  $P^d$ . Η  $\mathbf{X}_d$ , ορισμένη στο  $D_i$  είναι απεριοδική.

## 2.6 Παροδικότητα και Επαναληπτικότητα

Η ενότητα αυτή αφιερώνεται στην έννοια της στασιμότητας μιας αλυσίδας Markov. Θα μπορούσε κάποιος να πει ότι οι έννοιες που αναπτύσσονται εδώ, αφορούν ασθενείς μορφές στασιμότητας. Οι έννοιες αυτές, τυποποιούνται με τους όρους **παροδικότητα (transience)** και **επαναληπτικότητα (recurrence)**.

Γενικά, είναι πολύ πιο εύκολο να χαρακτηρίσουμε μια αλυσίδα Markov ως ασταθή, παρά ως στάσιμη: αποτυγχάνει να επιστρέψει στο σημείο εκκίνησής της, διαφεύγει τελικώς κάθε κλειστού συνόλου, με πιθανότητα ένα, πραγματοποιεί πεπερασμένο αριθμό επισκέψεων σε κάποιο σύνολο «ύψους μεγέθους». Στάσιμες αλυσίδες τότε, θεωρούνται αυτές οι οποίες δεν χάνονται από το σημείο

εκκίνησής τους, κατά έναν από τους προαναφερθέντες τρόπους. Υπάρχουν πολλές μορφές με τις οποίες εμφανίζεται η στασιμότητα, ποικίλοντας από ασθενή επιστροφή, κατά αναμενόμενη τιμή, στο σημείο εκκίνησής της, έως τη σύγκλιση όλων των πραγματοποιήσεων της στοχαστικής διαδικασίας σε κάποιο σημείο, όπως στην ολική ασυμπτωτική στασιμότητα (global asymptotic stability) ντετερμινιστικών διαδικασιών.

Εστιάζουμε την προσοχή μας στη συμπεριφορά του αριθμού των διελεύσεων της αλυσίδας σε κάποιο σύνολο  $A$ , ο οποίος είναι τυχαία μεταβλητή και δίνεται από τη σχέση  $\eta_A := \sum_{n=1}^{\infty} I_A(X_n)$ . Μελετούμε τη στασιμότητα της αλυσίδας σε όρους  $\eta_A$ , μέσω των εννοιών της παροδικότητας και επαναληπτικότητας των συνόλων της.

**Ορισμός 2.6.1** Το σύνολο  $A$  λέγεται **ομοιόμορφα παροδικό (uniformly transient)**, αν υπάρχει  $M < \infty$  τέτοιο ώστε  $E_x[\eta_A] \leq M$ , για κάθε  $x \in A$ . Το σύνολο  $A$  λέγεται **επαναληπτικό (recurrent)**, αν  $E_x[\eta_A] = \infty$ , για κάθε  $x \in A$ .

Το αποκορύφωμα της παραπάνω προσέγγισης είναι ένα διχοτομικό θεώρημα, εκπληκτικής ισχύος.

**Θεώρημα 2.6.1** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα Markov. Τότε είτε

1. κάθε σύνολο στο  $\mathcal{B}^+(\mathcal{X})$  είναι επαναληπτικό και στην περίπτωση αυτή καλούμε τη  $\mathbf{X}$  επαναληπτική, ή
2. υπάρχει μια πεπερασμένη επικάλυψη του  $\mathcal{X}$  από ομοιόμορφα παροδικά σύνολα και στην περίπτωση αυτή καλούμε τη  $\mathbf{X}$  παροδική.

Η επαναληπτικότητα συχνά ερμηνεύεται σε όρους της τυχαίας μεταβλητής που δίνει το χρόνο επανόδου  $\tau_A = \inf\{k \geq 1 : X_k \in A\}$  της αλυσίδας σε κάποιο σύνολο  $A$ . Η επαναληπτικότητα για το σύνολο  $A$  ορίζεται ως  $L(x, A) = P_x(\tau_A < \infty) = 1$ , για κάθε  $x \in A$ . Στην περίπτωση του γενικού χώρου καταστάσεων, η ισοδυναμία των ορισμών δεν είναι τόσο πλήρης. Η επαναληπτικότητα, σε όρους του  $\tau_A$  (την οποία καλούμε Harris επαναληπτικότητα) είναι βαθύτερη έννοια. Μια απλή σύνδεση είναι π.χ. όταν  $L(x, A) = 1$ , για κάθε  $x \in A$ , οπότε  $\eta_A = \infty$ , όταν  $X_0 \in A$ , και κατά συνέπεια το σύνολο  $A$  είναι επαναληπτικό.

Μπορούμε τώρα να περάσουμε στην ταξινόμηση μιας αλυσίδας  $\mathbf{X}$ , η οποία έχει ένα άτομο, ως επαναληπτικής ή παροδικής.

**Θεώρημα 2.6.2** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα, η οποία έχει ένα άτομο  $\alpha \in \mathcal{B}^+(\mathcal{X})$ . Τότε

1. Αν το  $\alpha$  είναι επαναληπτικό, τότε κάθε σύνολο στο  $\mathcal{B}^+(\mathcal{X})$  είναι επαναληπτικό.
2. Αν το  $\alpha$  είναι παροδικό, τότε υπάρχει επικάλυψη του  $\mathcal{X}$ , από πεπερασμένο αριθμό ομοιόμορφα παροδικών συνόλων.

Είναι δυνατό να βρούμε σύνολα τα οποία δεν είναι ομοιόμορφα παροδικά αφεαυτού, μπορούν όμως να επικαλυφθούν από ένα πεπερασμένο αριθμό ομοιόμορφα παροδικών συνόλων. Αυτό οδηγεί στον ακόλουθο ορισμό:

**Ορισμός 2.6.2** Αν το σύνολο  $A \in \mathcal{B}(\mathcal{X})$  μπορεί να επικαλυφθεί από ένα πεπερασμένο αριθμό ομοιόμορφα παροδικών συνόλων, τότε λέγεται παροδικό.

Ας θεωρήσουμε τώρα αλυσίδες, οι οποίες δεν έχουν άτομα, είναι όμως έντονα απεριοδικές.

**Ορισμός 2.6.3** Η αλυσίδα  $\mathbf{X}$  λέγεται επαναληπτική, αν είναι  $\psi$ -ανάγωγη και  $U(x, A) = \infty$ , για κάθε  $x \in \mathcal{X}$  και κάθε  $A \in \mathcal{B}^+(\mathcal{X})$ .

Η αλυσίδα  $\mathbf{X}$  λέγεται παροδική, αν είναι  $\psi$ -ανάγωγη και ο χώρος καταστάσεων  $\mathcal{X}$  είναι παροδικός.

**Θεώρημα 2.6.3** Αν η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη, τότε η  $\mathbf{X}$  είναι είτε επαναληπτική ή παροδική.

**Θεώρημα 2.6.4** Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και απεριοδική.

1. Η αλυσίδα  $\mathbf{X}$  είναι παροδική, αν και μόνο αν ένας, και στη συνέχεια όλοι οι  $m$ -σκελετοί  $\mathbf{X}^m$  είναι παροδικοί.
2. Η αλυσίδα  $\mathbf{X}$  είναι επαναληπτική, αν και μόνο αν ένας, και στη συνέχεια όλοι οι  $m$ -σκελετοί  $\mathbf{X}^m$  είναι επαναληπτικοί.

Είναι επιθυμητό να προάγουμε τον ορισμό της επαναληπτικότητας στη μορφή της Harris επαναληπτικότητας, σε όρους της ποσότητας  $L(x, A)$ . Τα προηγούμενα αποτελέσματα καθιστούν μια ξεκάθαρη διχοτόμηση μεταξύ παροδικότητας και επαναληπτικότητας, για  $\psi$ -ανάγωγες αλυσίδες. Παρόλα αυτά, δοθέντων των απαιτήσεων του Ορισμού 2.6.1, είναι χρήσιμο να εξετάσουμε απλούστερα κριτήρια για την ύπαρξη της επαναληπτικότητας.

**Θεώρημα 2.6.5** Μια  $\psi$ -ανάγωγη αλυσίδα  $\mathbf{X}$  είναι

1. επαναληπτική, αν υπάρχει κάποιο μικρό σύνολο  $C \in \mathcal{B}^+(\mathcal{X})$ , τέτοιο ώστε  $L(x, C) \equiv 1$ , για κάθε  $x \in C$ .
2. παροδική, αν και μόνο αν, υπάρχουν δύο σύνολα  $D, C \in \mathcal{B}^+(\mathcal{X})$ , με  $L(x, C) < 1$ , για κάθε  $x \in D$ .

## 2.7 Harris Επαναληπτικότητα

Στην ενότητα αυτή εισάγουμε ισχυρότερες μορφές επαναληπτικότητας και τις συνδέουμε με τη διχοτόμηση που παρουσιάστηκε στην προηγούμενη ενότητα.

Αναπτύσσοντας έννοιες επαναληπτικότητας ενός συνόλου  $A \in \mathcal{B}(\mathcal{X})$ , θα χρειαστούμε όχι μόνο το χρόνο πρώτης επανόδου  $\tau_A$ , ή την αναμενόμενη τιμή  $U(\cdot, A)$  της ποσότητας  $\eta_A$ , αλλά και το γεγονός ότι  $\mathbf{X} \in A$  απείρως συχνά (infinitely often, i.o.), ή αλλιώς  $\eta_A = \infty$ , που ορίζεται ως

$$\{\mathbf{X} \in A \text{ i.o.}\} := \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{X_k \in A\},$$

το οποίο είναι καλά ορισμένο ως  $\mathcal{F}$ -μετρήσιμο ενδεχόμενο στο  $\Omega$ . Για κάθε  $x \in \mathcal{X}$ ,  $A \in \mathcal{B}(\mathcal{X})$ , γράφουμε

$$Q(x, A) := P_x(\mathbf{X} \in A \text{ i.o.}). \quad (2.17)$$

Είναι φανερό ότι για κάθε  $x, A$  έχουμε  $Q(x, A) \leq L(x, A)$ .

**Ορισμός 2.7.1** Το σύνολο  $A$  λέγεται **Harris επαναληπτικό (Harris recurrent)**, αν

$$Q(x, A) := P_x(\eta_A = \infty) = 1, \quad x \in A.$$

Μια αλυσίδα  $\mathbf{X}$  λέγεται **Harris (επαναληπτική)**, αν είναι  $\psi$ -ανάγωγη και κάθε σύνολο στο  $\mathcal{B}^+(\mathcal{X})$  είναι Harris επαναληπτικό.

Θα δούμε παρακάτω ότι όταν  $A \in \mathcal{B}^+(\mathcal{X})$  και η  $\mathbf{X}$  είναι Harris επαναληπτική, τότε έχουμε τη φαινομενικά ισχυρότερη και ίσως περισσότερο συχνά χρησιμοποιούμενη ιδιότητα  $Q(x, A) = 1$ , για κάθε  $x \in \mathcal{X}$ .

Είναι φανερό από τους προηγούμενους ορισμούς, ότι αν κάποιο σύνολο είναι Harris επαναληπτικό, τότε είναι επαναληπτικό (recurrent). Πράγματι, στην

προηγούμενη διατύπωση, η ισχυροποίηση από την έννοια της επαναληπτικότητας σε αυτήν της Harris επαναληπτικότητας είναι σαφής, υποδεικνύοντας τη μετακίνηση από μια αναμενόμενη απειρότητα σε μια σχεδόν βέβαιη απειρότητα επανόδων σε κάποιο σύνολο.

Αυτός ο ορισμός της Harris επαναληπτικότητας εμφανίζεται να είναι ισχυρότερος από την απαίτηση  $L(x, A) \equiv 1$ , για  $x \in A$ , που είναι ένας τυπικός εναλλακτικός ορισμός της επαναληπτικότητας κατά Harris. Αποδεικνύεται ότι οι δύο ορισμοί είναι ισοδύναμοι.

**Πρόταση 2.7.1** Έστω ότι για κάποιο σύνολο  $A \in \mathcal{B}(\mathcal{X})$  έχουμε  $L(x, A) \equiv 1$ ,  $x \in A$ . Τότε  $Q(x, A) = L(x, A)$ , για κάθε  $x \in \mathcal{X}$  και συγκεκριμένα το σύνολο  $A$  είναι Harris επαναληπτικό.

Η παραπάνω πρόταση δηλώνει ότι ο ορισμός της Harris επαναληπτικότητας σε όρους της ποσότητας  $Q$  είναι ισοδύναμος με τον ορισμό σε όρους της ποσότητας  $L$ . Ο τελευταίος χρησιμοποιείται συχνά, όμως ο ορισμός που εμπλέκει την ποσότητα  $Q$  τονίζει τη διαφορά μεταξύ επαναληπτικότητας και Harris επαναληπτικότητας.

Το πιο δύσκολο, αλλά και ισχυρότερο, εκ των αποτελεσμάτων που παρουσιάζονται στην ενότητα αυτή δίνει ένα πιο εκλεπτυσμένο σύνδεσμο μεταξύ των πιθανοτήτων  $L(x, A)$  και  $Q(x, A)$  από αυτόν που δίνεται στην Πρόταση 2.7.1.

**Θεώρημα 2.7.1** 1. Έστω ότι το σύνολο  $A$  είναι ομοιόμορφα προσιτό (*uniformly accessible*) από ένα σύνολο  $D$  ( $D \rightarrow A$ ), για όλα τα σύνολα  $A, D \in \mathcal{B}(\mathcal{X})$ . Τότε

$$\{\mathbf{X} \in D \text{ i.o.}\} \subseteq \{\mathbf{X} \in A \text{ i.o.}\} \quad (2.18)$$

και για το λόγο αυτό  $Q(y, D) \leq Q(y, A)$ , για όλα τα  $y \in \mathcal{X}$ .

2. Αν το σύνολο  $A$  είναι ομοιόμορφα προσιτό από το σύνολο  $\mathcal{X}$  ( $\mathcal{X} \rightarrow A$ ), τότε το σύνολο  $A$  είναι Harris επαναληπτικό και ισχύει  $Q(x, A) \equiv 1$ , για κάθε  $x \in \mathcal{X}$ .

Ως συνέπεια του Θεώρηματος 2.7.1 έχουμε το ακόλουθο θεώρημα.

**Θεώρημα 2.7.2** Αν η  $\mathbf{X}$  είναι Harris επαναληπτική, τότε  $Q(x, B) = 1$ , για κάθε  $x \in \mathcal{X}$  και για κάθε  $B \in \mathcal{B}^+(\mathcal{X})$ .

Μπορούμε τώρα να ενισχύσουμε το σύνδεσμο μεταξύ των ιδιοτήτων της  $\mathbf{X}$  και των σκελετών της.

**Θεώρημα 2.7.3** Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και απεριοδική. Τότε η  $\mathbf{X}$  είναι Harris αν και μόνο αν κάθε σκελετός της είναι Harris.

Τα αποτελέσματα για την Harris επαναληπτικότητα, δίνουν μια χρήσιμη προέκταση του Θεώρηματος 2.6.5.

**Πρόταση 2.7.2** Αν υπάρχει κάποιο μικρό σύνολο  $C \in \mathcal{B}(\mathcal{X})$ , τέτοιο ώστε  $L(x, C) \equiv 1$ ,  $x \in \mathcal{X}$ , τότε η  $\mathbf{X}$  είναι Harris επαναληπτική.

## 2.8 Η Ύπαρξη του Μέτρου $\pi$

Μέχρι τώρα, στη διαπραγμάτευση εννοιών που αφορούν τη δομή και τη στασιμότητα ανάγωγων αλυσίδων, έχουμε εισάγει μόνο τη διχοτόμηση ανάμεσα σε παροδικές και επαναληπτικές αλυσίδες. Θα επικεντρωθούμε σε επαναληπτικές αλυσίδες και θα αναπτύξουμε τις συνέπειες της ιδιότητας της επαναληπτικότητας.

Στην ενότητα αυτή προχωρούμε σε περεταίρω διαχωρισμό των επαναληπτικών αλυσίδων σε **θετικές (positive)** και **μηδενικές (null)**. Θα δείξουμε ότι η πρώτη κλάση προσφέρει στοχαστική στασιμότητα πολύ μεγαλύτερης ισχύος από τη δεύτερη.

Για πλήθος λόγων η ισχυρότερη μορφή σταθερότητας που μπορεί να απαιτήσουμε, είναι η κατανομή της τυχαίας μεταβλητής  $X_n$  να μην μεταβάλεται καθώς το  $n$  παίρνει διαφορετικές τιμές. Στην περίπτωση αυτή, από τη Μαρκοβιανή ιδότητα προκύπτει ότι οι πεπερασμένες διάστασης κατανομές της αλυσίδας  $\mathbf{X}$  είναι στάσιμες ως προς το χρόνο. Τέτοιες σκέψεις, μας οδηγούν στην ιδέα των **αναλλοίωτων μέτρων (invariant measures)**.

**Ορισμός 2.8.1** Ένα  $\sigma$ -πεπερασμένο ( $\sigma$ -finite) μέτρο  $\pi$  ορισμένο στο πεδίο Borel  $\mathcal{B}(\mathcal{X})$  με την ιδιότητα

$$\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A), \quad A \in \mathcal{B}(\mathcal{X}) \quad (2.19)$$

λέγεται **αναλλοίωτο (invariant)**.

Παρόλο που θα αναπτύξουμε πλήθος αποτελεσμάτων, σχετικών με αναλλοίωτα μέτρα, το βασικό συμπέρασμα της παρούσας ενότητας είναι, αναμφίβολα, το παρακάτω θεώρημα:

**Θεώρημα 2.8.1** *Αν η αλυσίδα  $\mathbf{X}$  είναι επαναληπτική, τότε υπάρχει μοναδικό αναλλοίωτο μέτρο  $\pi$  το οποίο έχει για κάθε  $A \in \mathcal{B}^+(\mathcal{X})$ , τη μορφή*

$$\pi(B) = \int_A \pi(dw) E_w \left[ \sum_{n=1}^{\tau_A} I(X_n \in B) \right], \quad B \in \mathcal{B}(\mathcal{X}). \quad (2.20)$$

Ορίζουμε τις ακόλουθες κλάσεις αλυσίδων:

**Ορισμός 2.8.2** *Έστω μια  $\psi$ -ανάγωγη αλυσίδα  $\mathbf{X}$  η οποία έχει ένα αναλλοίωτο μέτρο πιθανότητας  $\pi$ . Τότε η  $\mathbf{X}$  λέγεται **θετική αλυσίδα (positive chain)**. Αν η  $\mathbf{X}$  δεν έχει τέτοιο μέτρο, τότε λέγεται **μηδενική (null chain)**.*

### 2.8.1 Αναλλοίωτα Μέτρα

Οι στοχαστικές διαδικασίες που έχουν την ιδιότητα οι περιθωριακές κατανομές των μεταβλητών  $\{X_n, \dots, X_{n+k}\}$ , για κάθε  $k$ , να μην μεταβάλλονται καθώς μεταβάλλεται το  $n$ , λέγονται **στάσιμες διαδικασίες (stationary processes)**. Γενικά, μια αλυσίδα Markov δεν είναι στάσιμη διαδικασία, αφού σε μια συγκεκριμένη δειγματική πραγματοποίηση μπορεί να έχουμε  $X_0 = x$  με πιθανότητα ένα, για κάποιο σταθερό  $x$  και να είναι δυνατό με την κατάλληλη επιλογή της αρχικής κατανομής για την  $X_0$  να παράξουμε μια στάσιμη διαδικασία  $\{X_n, n \in \mathbb{Z}_+\}$ .

Είναι άμεσο το γεγονός ότι χρειάζεται να θεωρήσουμε μια μορφή στασιμότητας πρώτης τάξης προκειμένου να δημιουργήσουμε μια ολόκληρη στάσιμη διαδικασία. Δοθέντος ενός αναλλοίωτου μέτρου πιθανότητας  $\pi$ , τέτοιου ώστε

$$\pi(A) = \int_{\mathcal{X}} \pi(dw) P(w, A),$$

μπορούμε επαγωγικά να πάρουμε

$$\begin{aligned} \pi(A) &= \int_{\mathcal{X}} \left[ \int_{\mathcal{X}} \pi(dx) P(x, dw) \right] P(w, A) \\ &= \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{X}} P(x, dw) P(w, A) \\ &= \int_{\mathcal{X}} \pi(dx) P^2(x, A) \\ &\vdots \\ &= \int_{\mathcal{X}} \pi(dx) P^n(x, A) = P_\pi(X_n \in A), \end{aligned} \quad (2.21)$$



για κάθε  $n$  και για κάθε  $A \in \mathcal{B}(\mathcal{X})$ .

Από τη Μαρκοβιανή ιδιότητα προκύπτει ότι η  $\mathbf{X}$  είναι στάσιμη αν και μόνο αν, η κατανομή της  $X_n$  δεν μεταβάλλεται με το χρόνο.

**Πρόταση 2.8.1** *Αν η αλυσίδα  $\mathbf{X}$  είναι θετική τότε είναι και επαναληπτική.*

Οι θετικές αλυσίδες συχνά αναφέρονται και ως «θετικά επαναληπτικές», τονίζοντας το γεγονός ότι είναι επαναληπτικές. Αυτό οδηγεί στον ορισμό

**Ορισμός 2.8.3** *Αν η  $\mathbf{X}$  είναι Harris επαναληπτική και θετική, τότε λέγεται **θετική Harris αλυσίδα (positive Harris chain)**.*

Δεν είναι ακόμα ξεκάθαρο αν ένα αναλλοίωτο μέτρο υπάρχει πάντα ή πότε αυτό είναι μοναδικό, όταν υπάρχει. Σε λίγο θα βρούμε κάποιες συνθήκες ύπαρξης του  $\pi$  και θα δείξουμε ότι για κάθε θετική (άρα και επαναληπτική) αλυσίδα, το μέτρο  $\pi$  είναι μοναδικό.

Τα αναλλοίωτα μέτρα είναι σημαντικά όχι μόνο γιατί προσδιορίζουν στάσιμες διαδικασίες. Είναι μέτρα που προσδιορίζουν τη μακροχρόνια ή εργοδική συμπεριφορά της αλυσίδας. Για να γίνει αντιληπτό κάτι τέτοιο ας θεωρήσουμε την πιθανότητα  $P_\mu(X_n \in A)$  για κάποια αρχική κατανομή  $\mu$ . Αν υπάρχει κάποιο οριακό μέτρο, σε κατάλληλο τόπο του χώρου των μέτρων πιθανότητας, τέτοιο ώστε

$$P_\mu(X_n \in A) \rightarrow \gamma_\mu(A),$$

για κάθε  $A \in \mathcal{B}(\mathcal{X})$ , τότε

$$\begin{aligned} \gamma_\mu(A) &= \lim_{n \rightarrow \infty} \int \mu(dx) P^n(x, A) \\ &= \lim_{n \rightarrow \infty} \int_{\mathcal{X}} \mu(dx) \int P^{n-1}(x, dw) P(w, A) \\ &= \int_{\mathcal{X}} \gamma_\mu(dw) P(w, A), \end{aligned} \quad (2.22)$$

αφού η σύγκλιση του  $\int \mu(dx) P^n(x, \cdot)$  υπονοεί τη σύγκλιση των ολοκληρωμάτων φραγμένων μετρήσιμων συναρτήσεων, όπως η  $P(w, A)$ .

Επομένως, αν υπάρχει μια οριακή κατανομή τότε αυτή είναι ένα αναλλοίωτο μέτρο πιθανότητας. Και προφανώς, αν υπάρχει ένα μοναδικό αναλλοίωτο μέτρο πιθανότητας, το όριό του, που είναι το  $\gamma_\mu$ , όταν υπάρχει θα είναι ανεξάρτητο από το  $\mu$ .

### 2.8.2 Υποαναλλοίωτα Μέτρα

Ο ευκολότερος τρόπος να διερευνήσουμε την ύπαρξη του  $\mu$ , είναι να εισάγουμε μια ακόμα ευρύτερη κλάση μέτρων, τα οποία ικανοποιούν ανισότητες σχετικές με την (2.19).

**Ορισμός 2.8.4** Αν το  $\mu$  είναι ένα  $\sigma$ -πεπερασμένο<sup>2</sup> μέτρο και ικανοποιεί τη σχέση

$$\mu(A) \geq \int_{\mathcal{X}} \mu(dx) P(x, A), \quad A \in \mathcal{B}(\mathcal{X}), \quad (2.23)$$

τότε το  $\mu$  λέγεται **υποαναλλοίωτο** (*subinvariant*).

Ας ξεκινήσουμε με κάποια δομικά αποτελέσματα που αφορούν αυθέραιτα υποαναλλοίωτα μέτρα.

**Πρόταση 2.8.2** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα. Αν το  $\mu$  είναι ένα μέτρο που ικανοποιεί την (2.23) με  $\mu(A) < \infty$ , για κάποιο  $A \in \mathcal{B}^+(\mathcal{X})$ , τότε

1. το  $\mu$  είναι  $\sigma$ -πεπερασμένο και άρα είναι υποαναλλοίωτο μέτρο,
2.  $\mu \succ \psi$ ,
3. αν το  $C$  είναι μικρό σύνολο, τότε  $\mu(C) < \infty$ ,
4. αν  $\mu(\mathcal{X}) < \infty$ , τότε το  $\mu$  είναι υποαναλλοίωτο.

Τα βασικά ζητήματα που μας ενδιαφέρουν κατά τη μελέτη υποαναλλοίωτων μέτρων αφορούν επαναληπτικές αλυσίδες.

**Πρόταση 2.8.3** Αν η αλυσίδα  $\mathbf{X}$  είναι παροδική, τότε υπάρχει ένα αυστηρώς υποαναλλοίωτο μέτρο για τη  $\mathbf{X}$ .

Ας προχωρήσουμε τώρα στη μελέτη επαναληπτικών αλυσίδων, όπου η ύπαρξη υποαναλλοίωτων μέτρων είναι λιγότερο προφανής.

---

<sup>2</sup>Ένα θετικό μέτρο  $\mu$  είναι  $\sigma$ -πεπερασμένο αν υπάρχει πεπερασμένο πλήθος συνόλων  $\{A_k\}$  τέτοια ώστε  $\mathcal{X} = \cup A_k$  και  $\mu(A_k) < \infty$ , για κάθε  $k$ .

### 2.8.3 Η Ύπαρξη του Μέτρου $\pi$ : Αλυσίδες με Άτομα

Στην υποενότητα αυτή προσφέρουμε ένα θεώρημα το οποίο μας βγάζει από το «κυνήγι» της ύπαρξης αναλλοίωτων και υποαναλλοίωτων μέτρων. Αποδεικνύεται ότι η ύπαρξη ενός ατόμου  $\alpha$  στο χώρο καταστάσεων, είναι αρκετή ώστε να περιγράψει την ύπαρξη και δομή τέτοιων μέτρων.

**Θεώρημα 2.8.2** Έστω  $\mathbf{X}$  μια  $\psi$ -ανάγωγη αλυσίδα με χώρο καταστάσεων  $\mathcal{X}$ , ο οποίος περιέχει ένα προσιτό άτομο (*accessible atom*)  $\alpha$ .

1. Υπάρχει πάντα κάποιο υποαναλλοίωτο μέτρο  $\mu_\alpha^o$  για την  $\mathbf{X}$ , το οποίο δίνεται από τη σχέση

$$\mu_\alpha^o(A) = U_\alpha(\alpha, A) = \sum_{n=1}^{\infty} {}_\alpha P^n(\alpha, A), \quad A \in \mathcal{B}(\mathcal{X}), \quad (2.24)$$

όπου  ${}_\alpha P^n(x, A) = P_x(X_n \in A, \tau_\alpha \geq n)$ ,  $x \in \mathcal{X}$ ,  $\alpha, A \in \mathcal{B}(\mathcal{X})$  και το μέτρο  $\mu_\alpha^o$  είναι αναλλοίωτο αν και μόνο αν η  $\mathbf{X}$  είναι επαναληπτική.

2. Το μέτρο  $\mu_\alpha^o$  είναι ελάχιστο (*minimal*), με την έννοια ότι αν  $\mu$  είναι ένα υποαναλλοίωτο μέτρο με  $\mu(\alpha) = 1$ , τότε

$$\mu(A) \geq \mu_\alpha^o(A), \quad A \in \mathcal{B}(\mathcal{X}).$$

Όταν η  $\mathbf{X}$  είναι επαναληπτική τότε το  $\mu_\alpha^o$  είναι το μοναδικό (υπο)αναλλοίωτο μέτρο με  $\mu_\alpha^o(\alpha) = 1$ .

3. Το υποαναλλοίωτο μέτρο  $\mu_\alpha^o$  είναι πεπερασμένο μέτρο αν και μόνο αν

$$E_\alpha[\tau_\alpha] < \infty,$$

και τότε είναι και αναλλοίωτο.

Θα χρησιμοποιήσουμε το  $\pi$  για να συμβολίσουμε το μοναδικό αναλλοίωτο μέτρο στην περίπτωση επαναληπτικών αλυσίδων. Επιπλέον, θεωρούμε ότι το  $\pi$  είναι μέτρο πιθανότητας όταν το  $\pi(\mathcal{X})$  είναι πεπερασμένο. Το αναλλοίωτο μέτρο  $\mu_\alpha^o$  έχει μια ισοδύναμη μορφή αναπαράστασης, για επαναληπτικές αλυσίδες:

$$\mu_\alpha^o(A) = E_\alpha \left[ \sum_{n=1}^{\tau_\alpha} I\{X_n \in A\} \right], \quad A \in \mathcal{B}(\mathcal{X}).$$

Ως άμεση συνέπεια της προηγούμενης αναπαράστασης, έχουμε το επόμενο κριτήριο για τη θετικότητα (*positivity*).

**Θεώρημα 2.8.3 (Kac's Theorem)** Αν  $\mathbf{X}$  είναι μια  $\psi$ -ανάγωγη αλυσίδα, η οποία έχει ένα άτομο  $\alpha \in \mathcal{B}^+(\mathcal{X})$ , τότε η  $\mathbf{X}$  είναι **θετική επαναλαμβανόμενη (positive recurrent)** αν και μόνο αν  $E_\alpha[\tau_\alpha] < \infty$ . Αν το  $\pi$  είναι αναλλοίωτο μέτρο πιθανότητας για τη  $\mathbf{X}$ , τότε

$$\pi(\alpha) = (E_\alpha[\tau_\alpha])^{-1}. \quad (2.25)$$

### 2.8.4 Η Ύπαρξη του Μέτρου $\pi$ : $\psi$ -Ανάγωγες Αλυσίδες

Θα δείξουμε στην ενότητα αυτή ότι μια επαναληπτική  $\psi$ -ανάγωγη αλυσίδα έχει ένα αναλλοίωτο μέτρο.

**Θεώρημα 2.8.4** Αν η  $\mathbf{X}$  είναι επαναληπτική τότε έχει ένα μοναδικό υποαναλλοίωτο μέτρο το οποίο είναι αναλλοίωτο.

Μπορούμε τώρα να αντιστοιχίσουμε τη θετικότητα για τη  $\mathbf{X}$  με τη θετικότητα για τους σκελετούς της.

**Θεώρημα 2.8.5** Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και απεριοδική. Τότε, για κάθε  $m$ , ένα μέτρο  $\pi$  είναι αναλλοίωτο για τον  $m$ -σκελετό της  $\mathbf{X}$  αν και μόνο αν είναι αναλλοίωτο για τη  $\mathbf{X}$ .

Για το λόγο αυτό, η  $\mathbf{X}$  είναι θετική αν και μόνο αν κάθε  $m$ -σκελετός της  $\mathbf{X}^m$  είναι θετικός.

### 2.8.5 Η Δομή του Μέτρου $\pi$ για Επαναληπτικές Αλυσίδες

**Θεώρημα 2.8.6** Έστω ότι η  $\mathbf{X}$  είναι επαναληπτική. Τότε το μοναδικό αναλλοίωτο μέτρο  $\pi$  για τη  $\mathbf{X}$  είναι ισοδύναμο με το  $\psi$  και ικανοποιεί, για κάθε  $A \in \mathcal{B}^+(\mathcal{X})$ ,  $B \in \mathcal{B}(\mathcal{X})$  τη σχέση

$$\begin{aligned} \pi(B) &= \int_A \pi(dy) U_A(y, b) \\ &= \int_A \pi(dy) E_y \left[ \sum_{k=1}^{\tau_A} I(X_k \in B) \right] \\ &= \int_A \pi(dy) E_y \left[ \sum_{k=0}^{\tau_A-1} I(X_k \in B) \right] \end{aligned} \quad (2.26)$$

Η ερμηνεία της (2.26) είναι η εξής: για κάθε σύνολο  $A \in \mathcal{B}^+(\mathcal{X})$ , το αναλλοίωτο μέτρο  $\pi(B)$  είναι ανάλογο του αριθμού των επισκέψεων στο σύνολο  $B$ , μεταξύ επανόδων στο σύνολο  $A$ , με την προϋπόθεση ότι η αλυσίδα ξεκινάει από το σύνολο  $A$ , με αρχική κατανομή την  $\pi_A$ , η οποία είναι αναλλοίωτη για την αλυσίδα  $\mathbf{X}^A$  που ορίζεται στο  $A$ .

**Θεώρημα 2.8.7** Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και  $\mu$  είναι ένα υποαναλλοίωτο μέτρο.

1. Η αλυσίδα  $\mathbf{X}$  είναι θετική αν και μόνο αν για ένα και έπειτα για όλα τα σύνολα  $\mu(A) > 0$

$$\int_A \mu(dy) E_y[\tau_A] < \infty. \quad (2.27)$$

2. Το μέτρο  $\mu$  είναι πεπερασμένο και επομένως η  $\mathbf{X}$  είναι θετική επαναληπτική αν για κάποιο μικρό σύνολο  $C \in \mathcal{B}^+(\mathcal{X})$ ,

$$\sup_{y \in C} E_y[\tau_C] < \infty. \quad (2.28)$$

Η αλυσίδα  $\mathbf{X}$  είναι θετική Harris (positive Harris) αν

$$E_x[\tau_C] < \infty, \quad x \in \mathcal{X}. \quad (2.29)$$

## 2.9 Εργοδικότητα

Στα προηγούμενα αναπτύξαμε ιδέες περί σταθερότητας, σε όρους δομών που είναι επαναληπτικές. Το ενδιαφέρον μας επικεντρώθηκε στον τρόπο με τον οποίο μια αλυσίδα επιστρέφει στο «κέντρο» του χώρου καταστάσεων, πόσο σίγουροι είμαστε ότι αυτό θα συμβεί και μάλιστα σε χρόνο πεπερασμένο.

Αργότερα, είδαμε κάποιες περσοσότερο σημαντικές και σίγουρα βαθύτερες έννοιες που έχουν να κάνουν με την «καταστάλαξη», ή σύγκλιση της αλυσίδας σε μια σταθερή ή στάσιμη περιοχή.

Στην υποενότητα αυτή, παρέχεται μια συστηματική προσέγγιση στο κρίσιμο ερώτημα: δεδομένης της ύπαρξης του μέτρου  $\pi$ , πότε οι πιθανότητες μετάβασης  $n$ -τάξης συγκλίνουν, με κατάλληλο τρόπο, στο  $\pi$ ;

Το Απεριοδικό Εργοδικό Θεώρημα (Aperiodic Ergodic Theorem), το οποίο ενοποιεί τους διάφορους ορισμούς περί θετικότητας, συνοψίζει την ασυμπτωτική θεωρία που εμπίπτει στο ενδιαφέρον μας.

**Θεώρημα 2.9.1 (Απεριοδικό Εργοδικό Θεώρημα)** Έστω  $X$  μια απεριοδική Harris επαναληπτική αλυσίδα με αναλλοίωτο μέτρο  $\pi$ . Τα επόμενα είναι ισοδύναμα:

1. Η αλυσίδα είναι θετική Harris: δηλαδή, το μοναδικό αναλλοίωτο μέτρο  $\pi$  είναι πεπερασμένο.
2. Υπάρχει ένα  $\nu$ -μικρό σύνολο  $C \in \mathcal{B}^+(\mathcal{X})$  και κάποιο  $P^\infty(C) > 0$ , τέτοιο ώστε, για κάθε  $x \in C$ ,

$$\lim_{n \rightarrow \infty} P^n(x, C) = P^\infty(C). \quad (2.30)$$

3. Υπάρχει ένα μικρό σύνολο  $C \in \mathcal{B}^+(\mathcal{X})$ , τέτοιο ώστε

$$\sup_{x \in C} E_x[\tau_C] < \infty. \quad (2.31)$$

Οι τρεις προηγούμενες προτάσεις είναι ισοδύναμες με την ύπαρξη ενός μοναδικού αναλλοίωτο μέτρου πιθανότητας  $\pi$ , τέτοιου ώστε για κάθε αρχική συνθήκη  $x \in \mathcal{X}$ ,

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{B}(\mathcal{X})} |P^n(x, A) - \pi(A)| = 0 \quad (2.32)$$

και επιπλέον, για όλες τις αρχικές κατανομές  $\lambda, \mu$ ,

$$\sum_{n=1}^{\infty} \int \int \lambda(dx) \mu(dy) \sup_{A \in \mathcal{B}(\mathcal{X})} |P^n(x, A) - P^n(y, A)| < \infty. \quad (2.33)$$

Είναι σκόπιμο στο σημείο αυτό να αναφέρουμε δυο λόγια για τον τρόπο της σύγκλισης που θα μας απασχολήσει. Εξατάζουμε λοιπόν τη σύγκλιση σε όρους των πιθανοτήτων μετάβασης. Τα αποτελέσματα που θα εξάγουμε, θα είναι σχετισμένα με το μέτρο  $(P^n - \pi)$  και θα αναφέρονται όχι απλώς στη σύγκλιση με βάση κάποιο σημείο ή ακόμα και κάποιο σύνολο, αλλά στην πιο γενική σύγκλιση, σε όρους του κανόνα ολικής απόστασης (total variation norm).

**Ορισμός 2.9.1** Αν  $\mu$  είναι ένα μέτρο ορισμένο στο  $\mathcal{B}(\mathcal{X})$ , τότε ο **κανόνας ολικής απόστασης (total variation norm)**  $\|\mu\|$  ορίζεται ως

$$\|\mu\| = \sup_{f: |f| \leq 1} |\mu(f)| = \sup_{A \in \mathcal{B}(\mathcal{X})} \mu(A) - \inf_{A \in \mathcal{B}(\mathcal{X})} \mu(A), \quad (2.34)$$

όπου το supremum λαμβάνεται ως προς όλες τις μετρήσιμες συναρτήσεις  $f$  που ορίζονται από το χώρο  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$  στο χώρο  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Η σχέση που θα μας απασχολήσει κυρίως στην ενότητα αυτή, είναι της μορφής

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi\| = \lim_{n \rightarrow \infty} \sup_A |P^n(x, A) - \pi(A)| = 0. \quad (2.35)$$

Αυτή η μετακίνηση προς τον κανόνα ολικής απόστασης, η οποία καθίσταται αναγκαία λόγω της έλλειψης δομής στις μεταβάσεις σε σημείο στο γενικό χώρο καταστάσεων, αποδεικνύεται εξαιρετικά καρποφόρα παρά περιοριστική.

Ένα δεύτερο σημείο που πρέπει να τονιστεί είναι η ανεξαρτησία μεταξύ της αρχικής και της οριακής κατανομής. Το όριο στη σχέση (2.35) ισχύει ανεξάρτητα από το σημείο εκκίνησης  $x$ .

Έχοντας διευκρινήσει κάτι τέτοιο, η αναγνώριση της κλάσης των αρχικών κατανομών για τις οποίες συγκεκριμένα ασυμπτωτικά αποτελέσματα εγείρουν, είναι ένα ενδιαφέρον ερώτημα, ενώ η απάντηση δεν είναι πάντα προφανής.

Πριν προχωρήσουμε στην παράθεση των αποτελεσμάτων, ένα μικρό σχόλιο για τον όρο «εργοδικός»: υιοθετούμε αυτό τον όρο για αλυσίδες όπου τό όριο στην (2.35) υπάρχει.

Γνωρίζουμε ότι γενικά οι αλυσίδες Markov είναι περιοδικές. Στην περίπτωση αυτή, το όριο της σχέσης (2.35) υπάρχει, στην καλύτερη περίπτωση μέσω κάποιας περιοδικής ακολουθίας  $nd$ , καθώς  $n \rightarrow \infty$ . Επομένως, εξόρισμού οι εργοδικές αλυσίδες είναι απεριοδικές και κάποια επουσιώδης, συχνά ενοχλητική αλλά πάντα ζωτική, αλλαγή στη δομή των αποτελεσμάτων είναι αναγκαία στην περίπτωση περιοδικών αλυσίδων.

Ας περάσουμε τώρα στην παρουσίαση κάποιων σημαντικών αποτελεσμάτων. Για γενικούς χώρους καταστάσεων, η Harris επαναληπτικότητα είναι απαραίτητη για τη σύγκλιση του πυρήνα μετάβασης  $P^n$  στο  $\pi$ .

**Θεώρημα 2.9.2** *Εστω  $X$  μια θετική Harris επαναληπτική και ισχυρά απεριοδική αλυσίδα Markov. Τότε για κάθε αρχικό μέτρο  $\lambda$*

$$\lim_{n \rightarrow \infty} \left\| \int \lambda(dx) P^n(x, \cdot) - \pi \right\| = 0. \quad (2.36)$$

Μπορούμε τώρα να μετακινηθούμε από τα αποτελέσματα για τις ισχυρά απεριοδικές αλυσίδες σε αυτά για αυθέραιτα απεριοδικές Harris επαναληπτικές αλυσίδες. Κάτι τέτοιο απλοποιείται ως αποτέλεσμα κάποιας πολύ χρήσιμης ιδιότητας του κανόνα ολικής απόστασης.

**Πρόταση 2.9.1** *Αν η  $\pi$  είναι αναλλοίωτη για τον πυρήνα μετάβασης  $P$ , τότε ο κανόνας ολικής απόστασης*

$$\left\| \int \lambda(dx) P^n(x, \cdot) - \pi \right\|$$

είναι μη αύξων ως προς  $n$ .

Μπορούμε τώρα να παραθέσουμε το επόμενο αποτέλεσμα που αφορά την περίπτωση της απεριοδικότητας.

**Θεώρημα 2.9.3** Έστω  $\mathbf{X}$  μια θετική Harris επαναληπτική και απεριοδική αλυσίδα Markov. Τότε για κάθε αρχική κατανομή  $\lambda$

$$\lim_{n \rightarrow \infty} \left\| \int \lambda(dx) P^n(x, \cdot) - \pi \right\| = 0. \quad (2.37)$$

## 2.10 $f$ , Γεωμετρική και Ομοιόμορφη Εργοδικότητα

### 2.10.1 $f$ -Εργοδικότητα

Στην ενότητα αυτή χαλαρώνουμε τις περιοριστικές συνθήκες, θεωρώντας πιο γενικές διατυπώσεις περί εργοδικότητας. Η πιο απλή μορφή εργοδικότητας είναι να θεωρήσουμε ταυτόχρονα όλες τις συναρτήσεις που είναι μικρότερες ή ίσες από τη συνάρτηση  $f$ , όπως ορίζονται από τον  $f$ -κανόνα

$$\| \nu \|_f = \sup_{g: |g| \leq f} |\nu(g)| = \sup_{g: |g| \leq f} \left| \int g(x) \nu(dx) \right|,$$

όπου  $\nu$  είναι κάποιο μέτρο. Δίνουμε λοιπόν τον παρακάτω ορισμό περί  $f$ -εργοδικότητας:

**Ορισμός 2.10.1** Λέμε ότι η αλυσίδα Markov  $\mathbf{X}$  είναι  $f$ -εργοδική, αν  $f \geq 1$  και

1. η  $\mathbf{X}$  είναι θετική Harris επαναληπτική με αναλλοίωτο μέτρο πιθανότητας  $\pi$ ,
2. η ποσότητα  $\pi(f)$  είναι πεπερασμένη,
3. για κάθε αρχική συνθήκη  $x$  της αλυσίδας,

$$\lim_{k \rightarrow \infty} \| P^k(x, \cdot) - \pi \|_f = 0.$$



### 2.10.2 Γεωμετρική Εργοδικότητα

Μια περισσότερο λεπτομερής περιγραφή των ιδιοτήτων της σύγκλισης εμπλέκει τη μελέτη της ταχύτητας σύγκλισης (speed of convergence)

$$\|P^n(x, \cdot) - \pi\|_f = o(\rho^n)$$

του πυρήνα μετάβασης  $P^n$  στο μέτρο  $\pi$ , όπου ο ρυθμός  $\rho$  είναι ανεξάρτητος από την αρχική συνθήκη  $x$ . Η εκτίμηση αυτής της ταχύτητας είναι σημαντική για τους Markov Chain Monte Carlo αλγόριθμους. Στη μελέτη του συγκεκριμένου θέματος θα μας βοηθήσει η ιδέα της **γεωμετρικής εργοδικότητας**.

**Ορισμός 2.10.2** Η αλυσίδα  $\mathbf{X}$  λέγεται  **$f$ -γεωμετρικά εργοδική** ( $f$  - *geometrically ergodic*), όπου  $f \geq 1$ , αν η  $\mathbf{X}$  είναι θετική Harris με  $\pi(f) < \infty$  και υπάρχει μια σταθερά  $r_f > 1$ , τέτοια ώστε

$$\sum_{n=1}^{\infty} r_f^n \|P^n(x, \cdot) - \pi\|_f < \infty, \quad (2.38)$$

για κάθε  $x \in \mathcal{X}$ . Αν η (2.38) ισχύει για  $f \equiv 1$  τότε λέμε ότι η  $\mathbf{X}$  είναι **γεωμετρικά εργοδική**.

Εξαιτίας της μεγάλης ισχύος των αποτελεσμάτων που αφορούν τη γεωμετρική εργοδικότητα, αλλά και του μεγάλου εύρους των στοχαστικών διαδικασιών για τις οποίες ισχύουν, δεν είναι υπερβολή να πούμε ότι η ιδέα της γεωμετρικής εργοδικότητας είναι η πιο σημαντική από όλες όσες έχουν παρουσιαστεί μέχρι τώρα.

### 2.10.3 $V$ -Ομοιόμορφη Εργοδικότητα

Προκειμένου να τυποποιήσουμε την ιδέα της  $V$ -ομοιόμορφης εργοδικότητας ( $V$ -uniform ergodicity), θεωρούμε  $P_1$  και  $P_2$ , δύο Μαρκοβιανές συναρτήσεις μετάβασης και μια θετική συνάρτηση  $1 \leq V < \infty$ . Ορίζουμε την **απόσταση κατά τον κανόνα  $V$**  ( $V$ -norm distance) μεταξύ  $P_1$  και  $P_2$  ως

$$\|P_1 - P_2\|_V := \sup_{x \in \mathcal{X}} \frac{\|P_1(x, \cdot) - P_2(x, \cdot)\|_V}{V(x)} \quad (2.39)$$

Συνήθως, θεωρούμε την απόσταση  $\|P^k - \pi\|_V$ , η οποία δεν ορίζεται από τη σχέση (2.39), αφού το  $\pi$  είναι μέτρο πιθανότητας και όχι πυρήνας μετάβασης.

Παρόλα αυτά, αν θεωρήσουμε την πιθανότητα  $\pi$  ως πυρήνα μετάβασης μέσω του ορισμού

$$\pi(x, A) := \pi(A), \quad A \in \mathcal{B}(\mathcal{X}), x \in \mathcal{X},$$

τότε ο κανόνας  $\|P^k - \pi\|_V$  είναι καλά ορισμένος.

**Ορισμός 2.10.3** *Μια εργοδική αλυσίδα  $\mathbf{X}$  λέγεται  $V$ -ομοιόμορφα εργοδική, αν*

$$\lim_{n \rightarrow \infty} \|P^n - \pi\|_V = 0. \quad (2.40)$$

**Θεώρημα 2.10.1** *Έστω ότι η  $\mathbf{X}$  είναι  $\psi$ -ανάγωγη και απεριοδική. Τότε τα επόμενα είναι ισοδύναμα για κάθε  $V \geq 1$ :*

1. Η  $\mathbf{X}$  είναι  $V$ -ομοιόμορφα εργοδική.
2. Υπάρχουν  $r > 1$  και  $R < \infty$ , τέτοια ώστε, για κάθε  $n \in \mathbb{Z}_+$

$$\|P^n - \pi\|_V \leq Rr^{-n}. \quad (2.41)$$

3. Υπάρχει κάποιο  $n > 0$  τέτοια ώστε  $\|P^i - \pi\|_V < \infty$ , για  $i \leq n$  και

$$\|P^n - \pi\|_V < 1. \quad (2.42)$$

Χρησιμοποιώντας τη σχέση (2.41), συνδέουμε την ιδέα της γεωμετρικής εργοδικότητας με μια από τις παλαιότερες, και ισχυρότερες, μορφές σύγκλισης στη θεωρία των αλυσίδων Markov, την ομοιόμορφη εργοδικότητα (uniform ergodicity), ή αλλιώς ισχυρή εργοδικότητα (strong ergodicity).

**Ορισμός 2.10.4** *Μια αλυσίδα  $\mathbf{X}$  λέγεται **ομοιόμορφα εργοδική**, αν είναι  $V$ -ομοιόμορφα εργοδική στην ειδική περίπτωση όπου  $V \equiv 1$ . Δηλαδή, αν*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} \|P^n(x, \cdot) - \pi\| = 0. \quad (2.43)$$

Επομένως, η ομοιόμορφη εργοδικότητα αποτελεί ειδική περίπτωση της  $V$ -ομοιόμορφης εργοδικότητας.

## 2.11 Οριακά Θεωρήματα

Η ενότητα αυτή αφιερώνεται στην ανάλυση της σειράς  $S_n(g)$ , την οποία ορίζουμε για κάθε συνάρτηση  $g$  ορισμένη στο  $\mathcal{X}$ ,

$$S_n(g) := \sum_{k=1}^n g(X_k). \quad (2.44)$$

Επικεντρωνόμαστε κυρίως σε δύο τύπους οριακών θεωρημάτων για θετικές επαναληπτικές αλυσίδες οι οποίες έχουν ένα αναλλοίωτο μέτρο πιθανότητας  $\pi$ :

1. τον **Ισχυρό Νόμο των Μεγάλων Αριθμών (Strong Law of Large Numbers LLN)**, ο οποίος δηλώνει ότι η ποσότητα  $n^{-1}S_n(g)$  συγκλίνει στο  $\pi(g) = E_\pi[g(X_0)]$ ,
2. το **Κεντρικό Οριακό Θεώρημα (Central Limit Theorem CLT)**, το οποίο δηλώνει ότι η συνέλιξη  $S_n(g - \pi(g))$ , κατάλληλα κανονικοποιημένη, κατανέμεται ασυμπτωτικά κανονικά.

Τα θεωρήματα LLN, CLT παρέχουν μέτρα τάσης παραμονής στο κέντρο (centrality) και μεταβλητότητας για τη  $X_n$ , καθώς το  $n$  γίνεται μεγάλο. Το μέγεθος της μεταβλητότητας μετριέται από τη διακύμανση, η οποία δίνεται στο CLT.

Τα δύο οριακά θεωρήματα που αναπτύσσουμε στην ενότητα αυτή, χρησιμοποιώντας τις ιδιότητες της πραγματοποίησης του δείγματος (sample path properties) για αλυσίδες που έχουν ένα μοναδικό αναλλοίωτο μέτρο πιθανότητας  $\pi$ , είναι:

1. **LLN** Λέμε ότι ο νόμος των μεγάλων αριθμών ισχύει για μια συνάρτηση  $g$ , αν

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n(g) = \pi(g). \quad (2.45)$$

2. **CLT** Λέμε ότι το Κεντρικό Οριακό Θεώρημα ισχύει για μια συνάρτηση  $g$ , αν υπάρχει μια σταθερά  $0 < \gamma_g^2 < \infty$ , τέτοια ώστε για κάθε αρχική συνθήκη  $x \in \mathcal{X}$ ,

$$\lim_{n \rightarrow \infty} P_x \left( (n\gamma_g^2)^{-1/2} S_n(\bar{g}) \leq t \right) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp^{-x^2/2} dx, \quad (2.46)$$

όπου  $\bar{g} = g - \pi(g)$ . Δηλαδή, καθώς  $n \rightarrow \infty$ ,

$$(n\gamma_g^2)^{-1/2} S_n(\bar{g}) \xrightarrow{d} N(0, 1).$$

### 2.11.1 Αρμονικές Συναρτήσεις

Για τη μελέτη των οριακών θεωρημάτων είναι απαραίτητη η εισαγωγή της έννοιας των **αρμονικών συναρτήσεων (harmonic functions)**, η οποία είναι συνδεδεμένη με την εργοδικότητα Harris επαναληπτικών Μαρκοβιανών αλυσίδων.

**Ορισμός 2.11.1** *Μια μετρήσιμη συνάρτηση  $h$  είναι αρμονική (harmonic) για την αλυσίδα  $\mathbf{X}$  αν*

$$E[h(X_{n+1}) | x_n] = h(x_n). \quad (2.47)$$

Οι συναρτήσεις αυτές χαρακτηρίζουν τη Harris επαναληπτικότητα ως εξής:

**Πρόταση 2.11.1** *Έστω μια θετική αλυσίδα Markov. Αν οι μοναδικές φραγμένες αρμονικές συναρτήσεις είναι οι σταθερές συναρτήσεις (constant functions), τότε η αλυσίδα είναι Harris επαναληπτική.*

Η Πρόταση 2.11.1 είναι πολύ χρήσιμη για την θεμελίωση της Harris επαναληπτικότητας στους Markov Chain Monte Carlo αλγόριθμους. Το αντίστροφο της πρότασης ισχύει.

**Λήμμα 2.11.1** *Για Harris επαναληπτικές αλυσίδες Markov, οι σταθερές είναι οι μόνες φραγμένες αρμονικές συναρτήσεις.*

Το παρακάτω θεώρημα προκύπτει ως συνέπεια του Λήμματος 2.11.1.

**Θεώρημα 2.11.1** *Έστω ότι η αλυσίδα  $\mathbf{X}$  είναι θετική Harris και ότι κάποιος από τους LLN, CLT ισχύει για κάποια συνάρτηση  $g$  και κάποια αρχική κατανομή. Τότε, τα ίδια οριακά θεωρήματα ισχύουν για κάθε αρχική κατανομή.*

Η Harris επαναληπτικότητα είναι μια ανώτερη μορφή στασιμότητας, με την έννοια ότι η σχεδόν βέβαιη σύγκλιση αντικαθίσταται από τη σύγκλιση παντού.

### 2.11.2 Ο Νόμος των Μεγάλων Αριθμών

Παρουσιάζουμε το Νόμο των Μεγάλων Αριθμών για θετικές Harris αλυσίδες. Αργότερα θα δούμε γενικότερα αποτελέσματα, τα οποία βασίζονται στην ύπαρξη ενός ατόμου στο χώρο καταστάσεων.

**Θεώρημα 2.11.2** *Τα επόμενα είναι ισοδύναμα όταν ένα αναλλοίωτο μέτρο πιθανότητας υπάρχει για την  $\mathbf{X}$ :*

1. Η  $\mathbf{X}$  είναι θετική Harris.
2. Για κάθε  $f \in L_1(\mathcal{X}, \mathcal{B}(\mathcal{X}), \pi)$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n(f) = \int f d\pi.$$

Παρακάτω δίνουμε μια εναλλακτική μορφή του Θεωρήματος 2.11.2, γνωστή και ως **Εργοδικό Θεώρημα (ergodic theorem)**.

**Θεώρημα 2.11.3** *Αν η αλυσίδα  $\mathbf{X}$  έχει ένα  $\sigma$ -πεπερασμένο αναλλοίωτο μέτρο  $\pi$ , οι επόμενες προτάσεις είναι ισodύναμες:*

1. Αν  $f, g \in L^1(\pi)$ , με  $\int g d\pi \neq 0$ ,

$$\lim_{n \rightarrow \infty} \frac{S_n(f)}{S_n(g)} = \frac{\pi(f)}{\pi(g)} = \frac{\int f d\pi}{\int g d\pi}. \quad (2.48)$$

2. Η  $\mathbf{X}$  είναι Harris επαναληπτική.

Μια ενδιαφέρουσα πτυχή του Θεωρήματος 2.11.3 είναι ότι το  $\pi$  δεν χρειάζεται να είναι μέτρο πιθανότητας και εξαιτίας αυτού μπορεί να υπάρξει κάποιου τύπου ισχυρής στασιμότητας, ακόμα και αν η αλυσίδα είναι μηδενική επαναληπτική (null recurrent).

### 2.11.3 Κεντρικό Οριακό Θεώρημα

**Θεώρημα 2.11.4** *Αν η αλυσίδα  $\mathbf{X}$  είναι θετική Harris και υπάρχει άτομο  $\alpha$ , τέτοιο ώστε*

$$E_\alpha[\tau_\alpha^2] < \infty, \quad E_\alpha \left[ \left( \sum_{n=1}^{\tau_\alpha} |h(X_n)| \right)^2 \right] < \infty, \\ \gamma_h^2 = \pi(\alpha) E_\alpha \left[ \left( \sum_{n=1}^{\tau_\alpha} \{h(X_n) - E_\pi[h]\} \right)^2 \right] > 0, \quad (2.49)$$

τότε το Κεντρικό Οριακό Θεώρημα ισχύει. Δηλαδή

$$\frac{1}{\sqrt{N}} \left( \sum_{n=1}^N (h(X_n) - E_\pi[h]) \right) \xrightarrow{d} N(0, \gamma_h^2). \quad (2.50)$$

Το παραπάνω αποτέλεσμα υποδηλώνει ότι μια επέκταση του Κεντρικού Οριακού Θεωρήματος στην περίπτωση της έλλειψης ατόμου είναι περισσότερο εύθραυστη από την αντίστοιχη στο Εργοδικό Θεώρημα. Παρακάτω δίνουμε μια εναλλακτική μορφή του Κεντρικού Οριακού Θεωρήματος, η οποία εμπλέκει αντιστρέψιμες αλυσίδες (reversible chains).

#### 2.11.4 Αντιστρέψιμες Αλυσίδες Markov

Έστω  $\mathbf{X}$  μια ομογενής αλυσίδα Markov σε γενικό χώρο καταστάσεων με πυρήνα μετάβασης  $P(x, \cdot)$  και στάσιμη κατανομή  $\pi$ . Ας υποθέσουμε ότι κάποιος επιθυμεί να μελετήσει την ακολουθία των καταστάσεων  $X_n, X_{n+1}, \dots$ . Αποδεικνύεται ότι η ακολουθία αυτή ικανοποιεί τη σχέση

$$\begin{aligned} P(X_n \in A | X_{n+1} = x, X_{n+2} = x_{n+2}, \dots) \\ = P(X_n \in A | X_{n+1} = x) = P_n^*(x, A) \end{aligned} \quad (2.51)$$

και για το λόγο αυτό ορίζει μια αλυσίδα Markov.

Στη γενική περίπτωση η αλυσίδα  $\mathbf{X}$  δεν είναι ομογενής. Όταν  $n \rightarrow \infty$ , ή εναλλακτικά όταν  $\mu = \pi$ , τότε  $P_n^*(x, A) = P^*(x, A)$  και η αλυσίδα γίνεται ομογενής.

Όταν  $P^*(x, A) = P(x, A)$ , για κάθε  $x \in \mathcal{X}$ ,  $A \in \mathcal{B}(\mathcal{X})$ , τότε η αντίστροφη αλυσίδα Markov έχει τις ίδιες πιθανότητες μετάβασης με την αυθεντική αλυσίδα. Αλυσίδες Markov με αυτή την ιδιότητα λέγονται αντιστρέψιμες και η συνθήκη αντιστρεψιμότητας συνήθως γράφεται ως

$$P(X_{n+1} \in A | X_n = x) = P(X_n \in A | X_{n+1} = x), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}). \quad (2.52)$$

**Ορισμός 2.11.2** Μια αλυσίδα Markov λέγεται **αντιστρέψιμη (reversible)** όταν η κατανομή της τυχαίας μεταβλητής  $X_n$ , δοθέντος ότι  $X_{n+1} = x$  είναι ίδια με την κατανομή της  $X_{n+1}$  δοθέντος ότι  $X_n = x$ .

Ένας εναλλακτικός ορισμός της αντιστρεψιμότητας είναι ο εξής:

**Ορισμός 2.11.3** Μια αλυσίδα Markov με χώρο καταστάσεων  $\mathcal{X}$  λέγεται **αντιστρέψιμη** ως προς μία κατανομή  $\pi$  ορισμένη στο  $\mathcal{X}$ , αν

$$\pi(dx) P(x, dy) = \pi(dy) P(y, dx), \quad \forall x, y \in \mathcal{X}. \quad (2.53)$$

Η συνθήκη αντιστρεψιμότητας (2.53) είναι γνωστή και ως **λεπτομερής εξίσωση ισορροπίας (detailed balance equation)**. Η συνθήκη αυτή δεν είναι απαραίτητη για την εξασφάλιση της σύγκλισης, ωστόσο είναι επαρκής συνθήκη ώστε η  $\pi$  να είναι η στάσιμη κατανομή της αλυσίδας.

**Θεώρημα 2.11.5** Έστω μια αλυσίδα Markov με πυρήνα μετάβασης  $p$  που ικανοποιεί την λεπτομερή εξίσωση ισορροπίας με  $\pi$  μια συνάρτηση πυκνότητας πιθανότητας. Τότε:

1. Η πυκνότητα  $\pi$  είναι η στάσιμη πυκνότητα της αλυσίδας.
2. Η αλυσίδα είναι αντιστρέψιμη.

Με την υπόθεση της αντιστρεψιμότητας μπορούμε να δώσουμε διαφορετικές προϋποθέσεις ισχύος του Κεντρικού Οριακού Θεωρήματος (Kipnis and Varadhan (1986)).

**Θεώρημα 2.11.6** Αν η αλυσίδα  $\mathbf{X}$  είναι αperiodική, ανάγωγη και επαναληπτική, με αναλλοίωτο μέτρο πιθανότητας  $\pi$ , το Κεντρικό Οριακό Θεώρημα ισχύει όταν

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

Το κύριο σημείο εδώ είναι ότι καίτοι η αντιστρεψιμότητα γενικά είναι πολύ περιοριστική υπόθεση, συχνά είναι πολύ εύκολο να επιβληθεί στους MCMC αλγόριθμους, εισάγοντας επιπρόσθετα βήματα προσομοίωσης (Geyer 1992, Tierney 1994 και Green 1995).





## Κεφάλαιο 3

# Ο Αλγόριθμος Metropolis - Hastings

### 3.1 Εισαγωγή

Το κεφάλαιο αυτό είναι αφιερωμένο στον αλγόριθμο που είναι γνωστός με την γενική ονομασία **αλγόριθμος Metropolis-Hastings**. Η ονομασία προέρχεται από τη δημοσίευση των Metropolis et al. (1953) και Hastings (1970). Αυτές θεωρούνται ως οι βασικές αναφορές για την ανάπτυξη της μεθόδου καιτοι και άλλες δημοσιεύσεις, συμπεριλαμβανομένων εκείνων των Barker (1965) και Peskun (1973), συνέβαλαν ουσιαστικά.

Η πρωτότυπη δημοσίευση από τους Metropolis et al. (1953) ασχολείται με τον υπολογισμό των ιδιοτήτων κάποιων χημικών ουσιών και δημοσιεύτηκε στο επιστημονικό περιοδικό *Journal of Chemical Physics*. Αργότερα διαπιστώθηκε η αξία της μεθόδου στη Στατιστική.

Ας θεωρήσουμε μια χημική ουσία αποτελούμενη από  $d$  μόρια (molecules). Ο πίνακας  $\theta = (\theta_1, \dots, \theta_d)$  δηλώνει τη θέση των μορίων. Πιο συγκεκριμένα το  $\theta_i$ ,  $i = 1, \dots, d$  είναι ένα διδιάστατο διάνυσμα με τις συντεταγμένες του  $i$ -μορίου στο επίπεδο. Από τη Στατιστική Μηχανική, η πυκνότητα πιθανότητας των θέσεων δίνεται από τον τύπο

$$f(\theta_1, \dots, \theta_d) \propto \exp \left\{ -\frac{1}{kT} E(\theta_1, \dots, \theta_d) \right\}, \quad (3.1)$$

όπου  $k$  είναι μια θετική σταθερά,  $T$  η θερμοκρασία του συστήματος και  $E(\cdot)$  μια θετική συνάρτηση που δηλώνει την ενέργεια του συστήματος.

Αν επιθυμούμε να υπολογίσουμε την ποσότητα ισορροπίας μιας οποιαδήποτε χημικής διεργασίας, τότε θα πρέπει να υπολογίσουμε την αναμενόμενη τιμή αυτής της ιδιότητας, ως προς την κατανομή του διανύσματος  $\theta$  των θέσεων των μορίων. Για μεγάλο  $d$ , ο άμεσος υπολογισμός αυτής της αναμενόμενης τιμής είναι ανέφικτος και αντικαθίσταται από μια Monte Carlo εκτίμηση. Οι Metropolis et al. (1953) πρότειναν την ακόλουθη μέθοδο δειγματοληψίας από την κατανομή του  $\theta$ :

1. Ξεκίνα με την αρχική τιμή  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$  και θέσε το μετρητή των επαναλήψεων (iteration counter)  $j = 1$ .
2. Μετακίνησε τα σωματίδια από τη θέση  $\theta^{(j-1)} = (\theta_1^{(j-1)}, \dots, \theta_d^{(j-1)})$ , σύμφωνα με μια ομοιόμορφη κατανομή «κεντραρισμένη» στις θέσεις αυτές, προκειμένου να πάρεις υποψήφιες νέες θέσεις  $\eta = (\eta_1, \dots, \eta_d)$ .
3. Υπολόγισε τη μεταβολή  $\Delta E$  που προκλήθηκε στην ενέργεια του συστήματος από τη μετακίνηση. Η μετακίνηση του βήματος 2 γίνεται αποδεκτή με πιθανότητα  $\min(1, e^{-c\Delta E})$ , όπου  $c$  είναι κάποια σταθερά. Αν η μετακίνηση γίνει αποδεκτή τότε θέσε  $\theta^{(j)} = \eta$ , διαφορετικά θέσε  $\theta^{(j)} = \theta^{(j-1)}$ .
4. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επέστρεψε στο βήμα 2 μέχρι να επέλθει κάποιου είδους σύγκλιση.

Μετά τη σύγκλιση, το διάνυσμα των θέσεων που γεννάται από τη μέθοδο αυτή έχει κατανομή με πυκνότητα που δίνεται από τη σχέση (3.1). Είναι προφανές ότι η παραπάνω μέθοδος ορίζει μια αλυσίδα Markov, καθώς οι μεταβάσεις εξαρτώνται από τις θέσεις των μορίων στο προηγούμενο βήμα. Παρόλα αυτά, δεν είναι προφανής η σύγκλιση της αλυσίδας σε κάποια στάσιμη κατανομή και η ταύτιση της τελευταίας με την (3.1). Οι Metropolis et al. (1953) παρουσιάζουν μια απόδειξη του παραπάνω αποτελέσματος. Η ίδια απόδειξη είναι έγκυρη στην περίπτωση που οι νέες μετακινήσεις, οι οποίες δίνονται από το διάνυσμα  $\eta$ , πραγματοποιούνται σύμφωνα με κάποια συμμετρική κατανομή «κεντραρισμένη» στο διάνυσμα των προηγούμενων θέσεων,  $\theta^{(j-1)}$ . Η μέθοδος αυτή ορίζει ένα πυρήνα μετάβασης  $q$  ο οποίος εξαρτάται από το ζεύγος  $(\theta, \eta)$  μέσω της σχέσης  $|\eta - \theta|$ .

Ο παραπάνω αλγόριθμος περιλαμβάνει ένα επιπρόσθετο βήμα σε σχέση με τον προηγούμενο: ο μηχανισμός μετάβασης εξαρτάται από ένα πυρήνα μετάβασης  $q$ , ο οποίος ονομάζεται πυρήνας μετάβασης πρότασης (proposal transition kernel).

Ένα σημείο άξιο προσοχής είναι ότι είναι δυνατό η προκύπτουσα αλυσίδα να παραμείνει σε μια κατάσταση χαμηλής ενέργειας (ή ισοδύναμα υψηλής πυκνότητας) για πολλές επαναλήψεις του αλγόριθμου. Στην περίπτωση αυτή, είναι πολύ πιθανό η νέα υποψήφια θέση των μορίων να είναι θέση υψηλής ενεργειακής κατάστασης (ή ισοδύναμα χαμηλής πυκνότητας) με  $\Delta E \gg 0$ . Η πιθανότητα αποδοχής της νέας αυτής τιμής είναι τώρα πολύ κοντά στο μηδέν. Υπολογιστικά, κάτι τέτοιο δεν είναι αποδεκτό και για το λόγο αυτό η επιλογή των πυρήνων μετάβασης πρέπει να γίνεται με προσοχή.

Η επόμενη ενότητα παρουσιάζει μια πιο κοινή και πλήρη έκδοση του αλγόριθμού, βασισμένη στη δουλιά του Hastings (1970).

## 3.2 Ορισμός και Ιδιότητες

Θεωρίστε μια κατανομή  $\pi$  από την οποία πρέπει να εξάγουμε ένα δείγμα μέσω αλυσίδων Markov. Πρέπει να κατασκευάσουμε ένα πυρήνα μετάβασης  $p(\theta, \eta)$  με τέτοιο τρόπο ώστε η κατανομή  $\pi$  να είναι η στάσιμη κατανομή της αλυσίδας. Επομένως, για επαρκώς μεγάλο  $N$ , η τυχαία μεταβλητή  $\theta^{(N)}$  μπορεί να θεωρηθεί ότι προέρχεται από την κατανομή  $\pi$ . Οι μέθοδοι που παρουσιάζονται στο κεφάλαιο αυτό παράγουν ένα μη ανεξάρτητο δείγμα από την  $\pi$ .

**Ορισμός 3.2.1** *Μια Markov Chain Monte Carlo (MCMC) μέθοδος για την προσομοίωση μιας κατανομής  $\pi$  είναι οποιαδήποτε μέθοδος παράγει μια εργοδική αλυσίδα Markov της οποίας στάσιμη κατανομή είναι η  $\pi$ .*

Ο Ορισμός 3.2.1 μας λέει ότι η χρήση μιας αλυσίδας τυχαίων μεταβλητών ( $\theta^{(t)}$ ), η οποία προκύπτει από ένα MCMC αλγόριθμο, με στάσιμη κατανομή την  $\pi$ , είναι ισοδύναμη με τη χρήση ενός δείγματος από ανεξάρτητες και ισόνομα κατανομημένες (iid) τυχαίες μεταβλητές από την  $\pi$ , με την εξής έννοια: το Εργοδικό Θεώρημα (Θεώρημα 2.9.1) εγγυάται τη σύγκλιση του δειγματικού μέσου

$$\frac{1}{N} \sum_{t=1}^N h(\theta^{(t)}) \quad (3.2)$$

στον θεωρητικό μέσο  $E_{\pi}[h(\theta)]$ , για κάθε ολοκληρώσιμη συνάρτηση  $h(\cdot)$ . Επομένως, μπορούμε να χρησιμοποιήσουμε την ακολουθία ( $\theta^{(t)}$ ), η οποία παράγεται από έναν MCMC αλγόριθμο, όπως και ένα δείγμα από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές.

Για τη γέννηση ενός δείγματος από την κατανομή  $\pi$  θεωρούμε μια αντιστρέψιμη εργοδική αλυσίδα Markov της οποίας ο πυρήνας μετάβασης  $p$  ικανοποιεί τη σχέση

$$\pi(\theta) p(\theta, \eta) = \pi(\eta) p(\eta, \theta), \quad \forall(\theta, \eta) \quad (3.3)$$

Ο πυρήνας  $p(\theta, \eta)$  αποτελείται από δύο στοιχεία: έναν αυθαίρετο πυρήνα μετάβασης  $q(\theta, \eta)$  και μία πιθανότητα  $\alpha(\theta, \eta)$ , τέτοια ώστε για  $\eta \neq \theta$ ,

$$\begin{aligned} p(\theta, \eta) &= p(\theta^{(n+1)} = \eta, I(\eta) | \theta^{(n)} = \theta) \\ &= p(\theta^{(n+1)} = \eta | \theta^{(n)} = \theta) p(I(\eta)) \\ &= q(\theta, \eta) \alpha(\theta, \eta), \quad \theta \neq \eta, \end{aligned} \quad (3.4)$$

όπου  $I(\delta)$  είναι η δείτρια συνάρτηση της αποδοχής του σημείου  $\delta$ :  $I(\delta) = 1$ , αν το σημείο  $\delta$  γίνεται αποδεκτό και  $I(\delta) = 0$  στην αντίθετη περίπτωση.

Ο υπολογισμός της πιθανότητας του ενδεχομένου να μεταβεί η αλυσίδα από την τιμή  $\theta^{(n)} = \theta$  στην ίδια (τρέχουσα) τιμή  $\theta$  είναι περισσότερο περίπλοκος, διότι κάτι τέτοιο μπορεί να συμβεί με δύο τρόπους: μια επιτυχημένη μετάβαση στην τρέχουσα κατάσταση  $\theta$  ή μια αποτυχημένη μετάβαση σε διαφορετική κατάσταση  $\eta$ , όπου  $\eta \neq \theta$ . Για τη διακριτή περίπτωση η πιθανότητα αυτή υπολογίζεται από τον τύπο

$$\begin{aligned} p(\theta, \theta) &= p(\theta^{(n+1)} = \theta, I(\theta) = 1 | \theta^{(n)} = \theta) \\ &\quad + p(\theta^{(n+1)} \neq \theta, I(\theta) = 0 | \theta^{(n)} = \theta) \\ &= q(\theta, \theta) + \sum_{\theta \neq \eta} q(\theta, \eta) [1 - \alpha(\theta, \eta)]. \end{aligned} \quad (3.5)$$

Για συνεχείς χώρους καταστάσεων, η πιθανότητα του πρώτου ενδεχομένου είναι μηδέν. Επομένως, μένει μόνο η πιθανότητα του ενδεχομένου μιας αποτυχημένης μετάβασης της αλυσίδας σε διαφορετική από το  $\theta$  κατάσταση:

$$p(\theta, \theta) = 1 - \int q(\theta, \eta) \alpha(\theta, \eta) d\eta. \quad (3.6)$$

Οι σχέσεις (3.4) και (3.6) ενσωματώνονται στον τύπο

$$p(\theta, A) = \int_A q(\theta, \eta) \alpha(\theta, \eta) d\eta + I(\theta \in A) \left[ 1 - \int q(\theta, \eta) \alpha(\theta, \eta) d\eta \right] \quad (3.7)$$

για κάθε μετρήσιμο υποσύνολο  $A$  του χώρου καταστάσεων της αλυσίδας.

Ο Hastings (1970) πρότεινε να ορίσουμε την **πιθανότητα αποδοχής (acceptance probability)**  $\alpha(\theta, \eta)$  με τέτοιο τρόπο, ώστε όταν συνδυάζεται με τον αυθαίρετο πυρήνα μετάβασης  $q(\theta, \eta)$  να ορίζει μια αντιστρέψιμη αλυσίδα. Η έκφραση που χρησιμοποιείται συχνότερα για την πιθανότητα αποδοχής είναι

$$\alpha(\theta, \eta) = \min \left\{ 1, \frac{\pi(\eta) q(\eta, \theta)}{\pi(\theta) q(\theta, \eta)} \right\} \quad (3.8)$$

Ας δούμε πως με βάση την παραπάνω τακτική ορίζεται μια αντιστρέψιμη αλυσίδα: για  $\theta = \eta$  είναι προφανές ενώ για  $\theta \neq \eta$  έχουμε

$$\begin{aligned} p(\theta, \eta) &= \alpha(\theta, \eta) q(\theta, \eta) \\ &= \min \left\{ 1, \frac{\pi(\eta) q(\eta, \theta)}{\pi(\theta) q(\theta, \eta)} \right\} q(\theta, \eta) \\ &= \min \left\{ q(\theta, \eta), \frac{\pi(\eta) q(\eta, \theta)}{\pi(\theta)} \right\} \\ &= \frac{1}{\pi(\theta)} \min(\pi(\theta) q(\theta, \eta), \pi(\eta) q(\eta, \theta)) \\ &\Rightarrow \pi(\theta) p(\theta, \eta) = \min(\pi(\theta) q(\theta, \eta), \pi(\eta) q(\eta, \theta)) \\ &= \pi(\eta) p(\eta, \theta). \end{aligned}$$

Είναι προφανές ότι η πιθανότητα  $\alpha(\theta, \eta)$  ορίζεται μόνο όταν  $\pi(\theta) > 0$ . Ωστόσο, αν η αρχική τιμή  $\theta^{(0)}$  της αλυσίδας είναι τέτοια ώστε  $\pi(\theta^{(0)}) > 0$ , τότε προκύπτει ότι  $\pi(\theta^{(n)}) > 0$ ,  $\forall n \in \mathbb{N}$ , αφού οι τιμές  $\eta^{(n)}$  για τις οποίες  $\pi(\eta^{(n)}) = 0$ , συνεπάγονται  $\alpha(\theta^{(n)}, \eta^{(n)}) = 0$  και απορρίπτονται από τον αλγόριθμο. Από σύμβαση, θεωρούμε ότι  $\alpha(\theta, \eta) = 0$  όταν  $\pi(\theta) = 0$  και  $\pi(\eta) = 0$ .

Οι αλγόριθμοι που βασίζονται σε αλυσίδες με πυρήνα μετάβασης της μορφής (3.7) και πιθανότητα αποδοχής της μορφής (3.8) αναφέρονται ως αλγόριθμοι Metropolis-Hastings. Αυτό αποτελεί αναγνώριση της συμβολής και των δύο δημοσιεύσεων. Ο Hastings (1970) αναφέρεται στο λόγο που εμφανίζεται στη σχέση (3.8) με τον όρο «test ratio». Η επιλογή της βέλτιστης μορφής της πιθανότητας αποδοχής οδηγεί στην ελαχιστοποίηση της ασυμπτωτικής διακύμανσης των εκτιμητών των ροπών της κατανομής  $\pi$ . Ο Peskun (1973) αποδεικνύει ότι για τη διακριτή περίπτωση η μορφή (3.8) είναι καλύτερη από ένα μεγάλο πλήθος επιλογών. Αποδεικνύει επίσης ότι με κατάλληλες επιλογές του πυρήνα μετάβασης  $q$ , η δειγματοληψία μέσω αλυσίδων Markov μπορεί να είναι περισσότερο ακριβής από τη δειγματοληψία μέσω ανεξάρτητων και ισόνομων τυχαίων μεταβλητών.

Ο λόγος της σχέσης (3.8) μπορεί να γραφτεί ως

$$\frac{\pi(\eta)/q(\theta, \eta)}{\pi(\theta)/q(\eta, \theta)} \quad (3.9)$$

Δηλαδή, η αποδοχή των προτεινόμενων τιμών βασίζεται στο λόγο της πυκνότητας της κατανομής στόχου (target distribution)  $\pi$  προς την πυκνότητα της κατανομής πρότασης (proposal distribution)  $q$ .

Σημειώστε ότι η συνθήκη αντιστρεψιμότητας (3.3) ικανοποιείται από τον πυρήνα  $p$  και όχι από τον  $q$ . Ο πυρήνας μετάβασης  $q$  παραμένει, μέχρι τώρα, αυθαίρετης μορφής και γι αυτό αποτελεί ένα εύκαμπτο εργαλείο για την κατασκευή του αλγόριθμου. Οι Roberts and Smith (1994) αποδεικνύουν ότι αν ο πυρήνας  $q$  είναι ανάγωγος (irreducible) και αperiοδικός (aperiodic) και  $\alpha(\theta, \eta) > 0$  για κάθε  $(\theta, \eta)$ , τότε ο αλγόριθμος παράγει μια ανάγωγη και αperiοδική αλυσίδα με πυρήνα μετάβασης  $p$  την (3.7) και οριακή κατανομή  $\pi$ .

Πρακτικά, η προσομοίωση ενός δείγματος από την κατανομή  $\pi$  χρησιμοποιώντας αλυσίδες Markov που ορίζονται από τον πυρήνα μετάβασης (3.7), μπορεί να γίνει ως εξής:

- Αλγόριθμος 3.2.1** 1. Θέσε το μετρητή των επαναλήψεων (iteration counter)  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)}$ .
2. προσομοίωσε μια νέα τιμή  $\eta$  από την κατανομή με πυκνότητα  $q(\theta^{(j-1)}, \cdot)$ .
3. Υπολόγισε την πιθανότητα αποδοχής της μετακίνησης  $\alpha(\theta^{(j-1)}, \eta)$ , η οποία δίνεται από τον τύπο (3.8). Αν η μετακίνηση γίνει αποδεκτή, τότε  $\theta^{(j)} = \eta$ , διαφορετικά  $\theta^{(j)} = \theta^{(j-1)}$  και η αλυσίδα δεν μετακινείται.
4. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανάλαβε τη διαδικασία από το Βήμα 2 μέχρι να επέλθει σύγκλιση.

Στο Βήμα 3 η αποδοχή ή απόρριψη της νέας τιμής πραγματοποιείται με τον εξής τρόπο: παράγουμε έναν τυχαίο αριθμό  $u$  από την ομοιόμορφη κατανομή  $U(0, 1)$ . Αν  $u \leq \alpha$  τότε η μετακίνηση γίνεται αποδεκτή. Διαφορετικά απορρίπτεται. Ο πυρήνας μετάβασης  $q$  ορίζει μια δυνατή μετακίνηση η οποία επικυρώνεται σύμφωνα με την τιμή της πιθανότητας  $\alpha$ . Για το λόγο αυτό, ο πυρήνας μετάβασης  $q$  αναφέρεται ως **πυρήνας μετάβασης πρότασης (proposal transition kernel)** ή **δεσμευμένη πυκνότητα πρότασης (proposal conditional density)** όταν θεωρηθεί ως δεσμευμένη πυκνότητα  $q(\theta, \cdot)$ .

Στην αρχική μορφή του αλγόριθμου Metropolis, [Metropolis et al. (1953)], ο πυρήνας  $q$  ορίζει μια συμμετρική μετάβαση γύρω από τις προηγούμενες θέσεις των σωματιδίων. Δηλαδή,  $q(\theta, \eta) = q(\eta, \theta)$ , για κάθε  $(\theta, \eta)$  και η πιθανότητα αποδοχής γίνεται

$$\alpha(\theta, \eta) = \min \left\{ 1, \frac{\pi(\eta)}{\pi(\theta)} \right\}, \quad (3.10)$$

εξαρτημένη μόνο από το λόγο  $\pi(\eta)/\pi(\theta)$ . Για τους Μπεϋζιανούς Στατιστικούς είναι ο λόγος των εκ των υστέρων πυκνοτήτων στην προτεινόμενη και την τρέχουσα θέση της αλυσίδας.

Πρέπει να τονίσουμε ότι η συμμετρία του πυρήνα μετάβασης  $q$  συνεπάγεται την αντιστρεψιμότητα του πυρήνα μετάβασης  $p$  γιατί

$$\begin{aligned} \pi(\theta) p(\theta, \eta) &= \pi(\theta) q(\theta, \eta) \alpha(\theta, \eta) \\ &= \pi(\theta) q(\theta, \eta) \min \left( 1, \frac{\pi(\eta)}{\pi(\theta)} \right) \\ &= q(\theta, \eta) \min(\pi(\theta), \pi(\eta)) \\ &= q(\eta, \theta) \min(\pi(\eta), \pi(\theta)) = \pi(\eta) p(\eta, \theta). \end{aligned}$$

Σημειώστε επίσης ότι είναι πιθανό η αλυσίδα να παραμείνει στην ίδια κατάσταση για πολλές επαναλήψεις του αλγόριθμου. Για την παρακολούθηση της λειτουργικότητας της μεθόδου είναι επιθυμητό να υπολογίζουμε το ποσοστό των επαναλήψεων του αλγορίθμου στις οποίες η μετακίνηση γίνεται αποδεκτή. Από το Εργοδικό Θεώρημα, για μεγάλα δείγματα, το ποσοστό αυτό είναι μια καλή εκτίμηση του θεωρητικού ποσοστού αποδοχής

$$E_{\pi} [1 - p(\theta, \theta)] = \int_{\mathcal{X}} [1 - p(\theta, \theta)] \pi(\theta) d\theta.$$

Ο Hastings προτείνει τον υπολογισμό του ποσοστού αποδοχής σε κάθε πρακτική εφαρμογή.

Μια τεχνική που εφαρμόζεται συχνά κατά τον υπολογισμό εργοδικών μέσων  $(1/N) \sum_t h(\theta^{(t)})$  είναι η επιβολή σταθμίσεων (weights) στις προτεινόμενες τιμές  $\eta^{(j)}$  από την πυκνότητα  $q(\theta^{(j-1)}, \cdot)$ . Οι σταθμίσεις αυτές είναι της μορφής  $m_j/N$ , ( $m_j = 0, 1, \dots$ ), όπου η ποσότητα  $m_j$  απαριθμεί το πλήθος των διαδοχικών επαναλήψεων του αλγόριθμου που καταλήγουν σε απόρριψη της προτεινόμενης τιμής:

$$m_j = \sum_{i < j: \theta^{(i)} = \eta^{(j-1)}}^{j-1} I(\theta^{(j+1)} \neq \eta^{(j)}).$$

Η επιτυχία της μεθόδου βασίζεται στην αποφυγή του πολύ μικρού ποσοστού αποδοχής. Μια απλοϊκή προσέγγιση του προβλήματος είναι να κάνουμε την αλυσίδα να κινείται πολύ αργά, δηλαδή οι μετακινήσεις να είναι πολύ μικρές. Υποθέτοντας για απλότητα ότι οι πυκνότητες  $q(\cdot, \cdot)$  και  $\pi(\cdot)$  είναι συνεχείς, παρόμοιες τιμές για τις προηγούμενες και προτεινόμενες καταστάσεις οδηγούν σε πιθανότητες αποδοχής πλησίον της μονάδας. Ακολουθώντας αυτή τη στρατηγική, ο αλγόριθμος έχει πολύ υψηλό ποσοστό αποδοχής και οι περισσότερες προτεινόμενες μετακινήσεις γίνονται αποδεκτές. Ωστόσο η αλυσίδα πρέπει να είναι ικανή να διαβαίνει ολοκληρω το χώρο καταστάσεων, ώστε να συγκλίνει στην στάσιμη κατανομή. Πολύ μικρές μετακινήσεις της αλυσίδας έχουν ως επακόλουθο την αύξηση του αριθμού των επαναλήψεων του αλγόριθμου προκειμένου να επιτευχθεί η επιθυμητή σύγκλιση. Από την άλλη μεριά, οι μεγάλες μετακινήσεις οδηγούν γρηγορότερα στη σύγκλιση, όμως είναι πιθανό να «πέσουν» στην ουρά της πυκνότητας  $\pi(\cdot)$  δίνοντας πολύ χαμηλή τιμή στο λόγο του τύπου (3.8). Επομένως, οι προτεινόμενες μετακινήσεις της αλυσίδας, όπως αυτές ορίζονται από τον πυρήνα  $q$ , πρέπει να είναι αξιοσημείωτου «εκτοπίσματος», αλλά με ουσιώδη πιθανότητα αποδοχής, όπως αυτή ορίζεται από τη συνάρτηση  $\alpha$ .

Μελέτες βελτιστοποίησης πάνω στο θέμα αυτό πραγματοποιούνται συνεχώς. Η ποικιλία των προς διαπραγμάτευση μοντέλων και των πυρήνων πρότασης  $q$  αποβαίνει αποτρεπτική για τη διαμόρφωση γενικότερων αποτελεσμάτων. Οι Bannett, Racine-Poon, and Wakefield (1995) και άλλοι συγγραφείς ισχυρίζονται ότι το ποσοστό αποδοχής πρέπει να είναι μεταξύ 20% και 50%. Κάτω από συγκεκριμένο θεωρητικό πλαίσιο, οι Gelman, Roberts and Gilks (1996) έδειξαν ότι το 24% είναι ένα βέλτιστο ποσοστό αποδοχής (οι προηγούμενες τιμές πρέπει να θεωρούνται ως ένας γενικός ενδεικτικός κανόνας και όχι ως επιτακτικός καθορισμός).

Η κατανομή στόχος  $\pi$  εισέρχεται στον Αλγόριθμο 3.2.1 μέσω του λόγου  $\pi(\eta) / \pi(\theta)$ . Επομένως, η πλήρης γνώση του τύπου της πυκνότητας  $\pi$  δεν είναι απαραίτητη. Συγκεκριμένα, δεν χρειάζεται να γνωρίζουμε τις σταθερές που εμπλέκονται στον τύπο αφού θα απλοποιηθούν όταν εισέλθουν στο λόγο που προαναφέραμε. Για παράδειγμα, όταν η πυκνότητα  $\pi$  είναι μια εκ των υστέρων κατανομή, καίτοι η συναρτησιακή της μορφή είναι γνωστή, η σταθερά κανονικοποίησης σπάνια είναι γνωστή. Επομένως, ο Αλγόριθμος 3.2.1 είναι ιδιαίτερα χρήσιμος για εφαρμογές στη Μπεϋζιανή Συμπερασματολογία.

Πολύ συζήτηση γίνεται σχετικά με τη λήψη του δείγματος από την πυκνότητα  $\pi$  μέσω μιας αλλά μεγάλης αλυσίδας ή μέσω πολλών μικρότερων αλυσίδων. Αναπτύσσοντας μία μεγάλη αλυσίδα, πρέπει να δώσουμε προσοχή στην «από-



σταση» (spacing) μεταξύ των τιμών της αλυσίδας που τελικά λαμβάνονται ως δείγμα. Η πιθανή επανάληψη της ίδιας κατάστασης σε πολλές διαδοχικές επαναλήψεις του αλγόριθμου δεν παρακωλύει την εκδήλωση των ιδιοτήτων της σύγκλισης, ωστόσο κάνει συχνότερη την εμφάνιση ακολουθιών επαναλαμβανόμενων τιμών, περιορίζοντας έτσι τη δυνατότητα διέλευσης της αλυσίδας από όσο περισσότερες τιμές του χώρου καταστάσεων γίνεται.

### 3.3 Περιορισμοί για τη Συνάρτηση μετάβασης πρότασης

Στην προηγούμενη ενότητα αναφέραμε ότι η κατανομή πρότασης είναι αυθαίρετης μορφής. Ωστόσο, είναι απαραίτητη η επιβολή κάποιων ελάχιστων περιορισμών τόσο στην πυκνότητα  $\pi$ , όσο και στην  $q$ , έτσι ώστε η  $\pi$  να είναι η οριακή κατανομή της αλυσίδας Markov  $\theta$  που παράγεται από τον Αλγόριθμο 3.2.1.

Ένα κρίσιμο ζήτημα είναι το γεγονός ότι η δειγματοληψία από την κατανομή με πυκνότητα  $q$  πρέπει να είναι εύκολη, αφού η μέθοδος αντικαθιστά τη δύσκολη δειγματοληψία από την  $\pi$  με αυτήν από την  $q$ .

Εξίσου σημαντική είναι η απαίτηση για σωστό «κούρδισμα» των μετακινήσεων που προτείνονται από την πυκνότητα  $q$ , ώστε να εξασφαλίζεται η πλήρης κάλυψη του παραμετρικού χώρου σε πραγματικό υπολογιστικό χρόνο (real computing time). Έστω ότι υπάρχει ένα σύνολο  $A \subset \mathcal{E}$  τέτοιο ώστε,

$$\int_A \pi(\theta) d\theta > 0 \quad \text{και} \quad \int_A q(\theta, \eta) d\eta = 0, \quad \forall \theta \in \mathcal{E}.$$

Τότε, η αλυσίδα που παράγεται από τον Αλγόριθμο 3.2.1 δεν έχει ως οριακή κατανομή την  $\pi$ , αφού για  $\theta^{(0)} \notin A$  η αλυσίδα  $(\theta^{(t)})$  δεν διέρχεται ποτέ από το σύνολο  $A$ . Επομένως, ένας ελάχιστος, αλλά απαραίτητος, περιορισμός είναι

$$\text{supp } \pi \subset \bigcup_{\theta \in \text{supp } \pi} \text{supp } q(\theta, \cdot). \quad (3.11)$$

Από το Θεώρημα 2.11.5 εξασφαλίζουμε το ότι η πυκνότητα  $\pi$  είναι η στάσιμη κατανομή μιας αντιστρέψιμης αλυσίδας η οποία έχει πυρήνα μετάβασης (3.7).

Το ακόλουθο αποτέλεσμα ουσιαστικά αναδεικνύει την γενικότητα και καθολικότητα του Metropolis-Hastings αλγόριθμου.

**Θεώρημα 3.3.1** *Για κάθε δεσμευμένη κατανομή  $q$  της οποίας το στήριγμα αποτελεί υπερσύνολο του στηρίγματος  $\mathcal{E}$  της συνάρτησης στόχου  $\pi$ , η  $\pi$  είναι*

η στάσιμη κατανομή της αλυσίδας Markov που παράγεται από τον Αλγόριθμο 3.2.1.

Ουσιαστικής σημασίας για την επιλογή της πυκνότητας μετάβασης πρότασης  $q$  είναι η διακύμανσή της. Σε πολυδιάστατα προβλήματα για γενικούς χώρους καταστάσεων, η πιο βολική επιλογή για την συνάρτηση  $q$  είναι η πυκνότητα της κατανομής  $N[\theta^{(j-1)}, \tilde{\Sigma}]$  λόγω της εύκολης δειγματοληψίας από αυτή και της συμμετρικότητας περί της προηγούμενης κατάστασης  $\theta^{(j-1)}$ .

Καίτοι στην θεωρία οποιοσδήποτε θετικά ορισμένος πίνακας συνδιακύμανσης  $\tilde{\Sigma}$  είναι μια αποδεκτή επιλογή, στην πράξη η επιλογή του χρειάζεται προσοχή. Σκεπτόμενοι προς στιγμή για μονοδιάστατες παραμέτρους, παρατηρήστε ότι αν η διακύμανση  $\tilde{\Sigma}$  της πυκνότητας πρότασης  $q$  είναι μεγάλη, τότε η κατανομή πρότασης θα είναι ευρύτερη σε σχέση με την κατανομή στόχο  $\pi$  και κάθε διαδοχικό βήμα θα είναι απομακρυσμένο, προς κάποια κεντεύθυνση, σε σχέση με το προηγούμενο, προσδίδοντας στην κίνηση της αλυσίδας μια αδεξιότητα. Επιπλέον, όταν το στήριγμα της  $q$  είναι ευρύτερο από το στήριγμα της  $\pi$ , τότε πολλές από τις προτεινόμενες τιμές θα βρίσκονται έξω από το στήριγμα  $\mathcal{E}$  της συνάρτησης  $\pi$  και θα απορρίπτονται από τον αλγόριθμο. Στην περίπτωση αυτή, η αλυσίδα έχει την τάση να «κολλάει» σε μία κατάσταση και να μην διαβαίνει το χώρο καταστάσεων σωστά.

Από την άλλη μεριά, επιλέγοντας μικρή διακύμανση  $\tilde{\Sigma}$  για την πυκνότητα  $q$  μπορεί να αποβεί επιζήμιο. Η αλυσίδα θα έχει μεγάλο ποσοστό αποδοχής των προτεινόμενων τιμών αλλά θα κινείται με πολύ μικρά βήματα μέσα στο χώρο των καταστάσεων. Έτσι, θα χρειαστούν περισσότερες επαναλήψεις του αλγόριθμου ώστε η αλυσίδα να «ταξιδέψει» σε ολόκληρο το χώρο καταστάσεων, μέχρι να βρει την περιοχή ισορροπίας της (equilibrium region). Παράλληλα, η αυτοσυσχέτιση μεταξύ των τυχαίων μεταβλητών της αλυσίδας αυξάνεται, με συνέπειες που θα συζητήσουμε σε επόμενο κεφάλαιο.

Ο Müller (1991) πρότεινε το ταίριασμα του πίνακα συνδιακυμάνσεων  $\tilde{\Sigma}$  της πυκνότητας πρότασης  $q$  με τον πραγματικό πίνακα συνδιακυμάνσεων  $\Sigma$  της πυκνότητας στόχου  $\pi$ , πιθανώς με τη βοήθεια κάποιου ακατέργαστου εκτιμητή του  $\Sigma$  από κάποια δειγματοληψία «πιλότο». Άλλοι συγγραφείς προτείνουν επιλογή  $\tilde{\Sigma}$  τέτοιου ώστε το ποσοστό αποδοχής στον αλγόριθμο να πλησιάζει το 50%. Μελέτες από τους Gelman, Roberts and Gilks (1996) δείχνουν ότι αυτή η δεύτερη συμβουλή είναι σχεδόν βέλτιστη όταν η κατανομή στόχος είναι η μονομεταβλητή κανονική: η αυτοσυσχέτιση πρώτης τάξης της αλυσίδας ελαχιστοποιείται θέτοντας  $\sqrt{\tilde{\Sigma}} = 2.4\sqrt{\Sigma}$ . Η πυκνότητα πρότασης με αυτή τη διακύμανση οδηγεί σε ποσοστό αποδοχής 44.1%. Ωστόσο, τα πράγματα εί-

ναι διαφορετικά σε πολυδιάστατα προβλήματα. Για πολυμεταβλητή πυκνότητα στόχο η οποία έχει τη μορφή

$$\pi(\boldsymbol{\theta}) = \prod_{i=1}^K g(\theta_i),$$

για κάποια μονοδιάστατη πυκνότητα  $g$ , το βέλτιστο ποσοστό αποδοχής στον Αλγόριθμο 3.2.1 προσεγγίζει το 23,4% καθώς η διάσταση (dimensionality)  $K$  τείνει στο άπειρο.

### 3.4 Ιδιότητες Σύγκλισης

Η θεωρία που αναπτύχθηκε στο Κεφάλαιο 1 μας βοηθάει να δείξουμε ότι η αλυσίδα Markov που παράγεται από τον Αλγόριθμο 3.2.1 συγκλίνει πράγματι και ότι εκτιμητές της μορφής (3.2) μπορούν να χρησιμοποιηθούν για την προσέγγιση ποσοτήτων της μορφής  $E_\pi[h(\boldsymbol{\theta})]$ .

Μια MCMC αλυσίδα έχει, από την κατασκευή της, ένα σταθερό μέτρο πιθανότητας  $\pi$ . Επιπλέον, αν η αλυσίδα είναι Harris απεριοδική, τότε το Εργοδικό Θεώρημα ισχύει και εκτιμήσεις σαν αυτές που περιγράψαμε είναι έγκυρες.

Η ιδιότητα της αναγωγής (irreducibility) μιας Metropolis-Hastings αλυσίδας  $(\theta^{(t)})$ , προκύπτει από συνθήκες όπως η θετικότητα της δεσμευμένης πυκνότητας  $q$ : δηλαδή,

$$q(\boldsymbol{\theta}, \boldsymbol{\eta}) > 0, \quad \forall (\boldsymbol{\theta}, \boldsymbol{\eta}) \in \mathcal{E} \times \mathcal{E}, \quad (3.12)$$

και τότε έπεται ότι κάθε μη μηδενικό υποσύνολο του στηρίγματος  $\mathcal{E}$  είναι προσιτό σε ένα απλό βήμα.

Από τον Ορισμό 2.8.2 προκύπτει ότι αν μια αλυσίδα έχει μια στάσιμη κατανομή με πυκνότητα  $\pi$ , τότε η αλυσίδα αυτή λέγεται θετική και επιπλέον, από την Πρόταση 2.8.1, είναι και επαναληπτική (recurrent). Το επόμενο αποτέλεσμα είναι ισχυρότερο.

**Λήμμα 3.4.1** *Αν η Metropolis-Hastings αλυσίδα  $(\theta^{(t)})$  είναι  $\pi$ -ανάγωγη, τότε είναι Harris επαναληπτική.*

Το επόμενο θεώρημα δηλώνει τη σύγκλιση για τις Metropolis-Hastings αλυσίδες Markov.

**Θεώρημα 3.4.1** *Έστω μια  $\pi$ -ανάγωγη Metropolis-Hastings αλυσίδα Markov  $(\theta^{(t)})$ .*

1. Αν  $h \in L^1(\pi)$ , τότε

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N h(\theta^{(t)}) = \int h(\theta) d\pi.$$

2. Αν, επιπλέον, η αλυσίδα  $(\theta^{(t)})$  είναι απεριοδική, τότε

$$\lim_{n \rightarrow \infty} \left\| \int p^n(\theta, \cdot) \mu(d\theta) - \pi \right\| = 0$$

για κάθε αρχική κατανομή  $\mu$ , όπου  $p^n(\theta, \cdot)$  είναι ο πυρήνας μετάβασης  $n$ -τάξης.

Όπως προείπαμε, το γεγονός ότι η δεμειμένη πυκνότητα  $q$  είναι θετική, συνεπάγεται ότι η Metropolis-Hastings αλυσίδα είναι  $\pi$ -ανάγωγη.

Παρόλο που η συνθήκη 3.12 φαίνεται να είναι περιοριστική, στην πράξη συχνά ικανοποιείται.

## 3.5 Ειδικές Περιπτώσεις

Όπως περιγράψαμε στην Ενότητα 3.3, υπάρχει μεγάλη ανοχή ως προς την επιλογή της συνάρτησης  $q$ , εκτός από κάποιους τεχνικούς περιορισμούς. Εκεί, κάναμε μια γενική θεώρηση και τώρα στρεφόμαστε προς τη μελέτη κάποιων ειδικών περιπτώσεων. Να επισημάνουμε ότι παρόλο που μια αλυσίδα προσδιορίζεται από τον πυρήνα μετάβασης  $p$  και όχι από τον πυρήνα μετάβασης πρότασης  $q$ , οι ονομασίες που χρησιμοποιούνται για την κατηγοριοποίηση του Αλγόριθμου 3.2.1 αναφέρονται στις ιδιότητες του  $q$  παρά του  $p$ .

### 3.5.1 Ανεξάρτητες Αλυσίδες

Στην περίπτωση αυτή, η προτεινόμενη μετάβαση γεννιέται ανεξάρτητα από την τρέχουσα κατάσταση  $\theta$  της αλυσίδας. Συνεπακολούθως, η πυκνότητα μετάβασης πρότασης  $q$  είναι ανεξάρτητη από την τρέχουσα κατάσταση  $\theta$  και σημαίνεται με  $g$ , δηλαδή  $q(\theta, \eta) = g(\eta)$ . Ο Αλγόριθμος 3.2.1 λαμβάνει την εξής μορφή:

**Αλγόριθμος 3.5.1** 1. Θέσε το μετρητή των επαναλήψεων (*iteration counter*)  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)}$ .

2. Προσομοίωσε μια νέα τιμή  $\eta \sim g(\cdot)$ .

3. Δέξου ως νέα τιμή για την αλυσίδα την

$$\theta^{(j)} = \begin{cases} \eta, & \text{με πιθανότητα } \alpha(\theta^{(j-1)}, \eta), \\ \theta^{(j-1)}, & \text{με πιθανότητα } 1 - \alpha(\theta^{(j-1)}, \eta). \end{cases}$$

4. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανάλαβε τη διαδικασία από το βήμα 2 μέχρι να επέλθει σύγκλιση.

$$\text{Η πιθανότητα αποδοχής είναι } \alpha(\theta^{(j-1)}, \eta) = \min \left( 1, \frac{\pi(\eta) g(\theta^{(j-1)})}{\pi(\theta^{(j-1)}) g(\eta)} \right).$$

Φαινομενικά, η ανεξαρτησία του πυρήνα μετάβασης πρότασης από την προηγούμενη κατάσταση της αλυσίδας έρχεται σε αντίθεση με τη Μαρκοβιανή ιδιότητα της αλυσίδας. Ωστόσο, θυμηθείτε ότι ο πυρήνας  $g$  είναι ένας πυρήνας πρότασης ο οποίος συνδυάζεται με κάποια πιθανότητα αποδοχής  $\alpha$  προκειμένου να δώσει τον πυρήνα μετάβασης  $p$  της αλυσίδας που παράγεται από τον αλγόριθμο. Όμως η πιθανότητα αποδοχής εξαρτάται από την προηγούμενη κατάσταση. Επομένως ο πυρήνας  $p$  εξαρτάται από την προηγούμενη κατάσταση της αλυσίδας διατηρώντας έτσι τη Μαρκοβιανή ιδιότητα. Με άλλα λόγια, οι προτεινόμενες τιμές  $\eta$  γεννώνται ανεξάρτητα, το προκύπτον δείγμα όμως δεν είναι ανεξάρτητα και ισόνομα κατανομημένο (iid), επειδή η πιθανότητα αποδοχής μιας τιμής  $\eta$  εξαρτάται από την προηγούμενη τιμή της αλυσίδας  $\theta^{(j-1)}$ .

Οι ιδιότητες σύγκλισης της αλυσίδας  $(\theta^{(t)})$  προκύπτουν από τις ιδιότητες της πυκνότητας  $g$ , με την εξής έννοια: η  $(\theta^{(t)})$  είναι ανάγωγη και αperiοδική αν και μόνο αν η πυκνότητα  $g$  είναι θετική σχεδόν παντού στο στήριγμα της  $\pi$ .

Το επόμενο αποτέλεσμα από τους Mengersen και Tweedie (1996) περιγράφει ισχυρότερες ιδιότητες σύγκλισης όπως είναι η γεωμετρική και η ομοιόμορφη εργοδικότητα.

**Θεώρημα 3.5.1** *Ο Αλγόριθμος 3.5.1 παράγει μια ομοιόμορφα εργοδική αλυσίδα αν υπάρχει σταθερά  $M$  τέτοια ώστε*

$$\pi(\theta) \leq Mg(\theta), \quad \theta \in \text{supp } \pi. \quad (3.13)$$

Στην περίπτωση αυτή,

$$\|p^n(\theta, \cdot) - \pi\| \leq 2 \left( 1 - \frac{1}{M} \right)^n, \quad (3.14)$$

όπου  $\|\cdot\|$  είναι η νόρμα ολικής μεταβολής (total variation norm). Από την άλλη μεριά, αν για κάθε  $M$  υπάρχει ένα μη μηδενικό υποσύνολο του στηρίγματος της

πυκνότητας  $\pi$  για το οποίο η σχέση (3.14) δεν ευσταθεί, τότε η αλυσίδα  $(\theta^{(t)})$  δεν είναι ούτε γεωμετρικά εργοδική.

Αυτή η συγκεκριμένη κλάση των Metropolis-Hastings αλγόριθμων προκαλεί σύγκριση με τους αλγόριθμους της μεθόδου **Αποδοχής - Απόρριψης (Accept - Reject algorithms)**, αφού κάθε ζεύγος  $(\pi, g)$  που ικανοποιεί τη σχέση (3.13) μπορεί να χρησιμοποιηθεί σε ένα αλγόριθμο Αποδοχής - Απόρριψης.

Σύμφωνα με τη μέθοδο αυτή, για την προσομοίωση ενός δείγματος από μια κατανομή  $\pi$ , προσομοιώνουμε τιμές  $\eta$  από μια βοηθητική κατανομή  $g$ , η οποία καλείται **συνάρτηση φάκελος (envelope function)**. Η συνάρτηση αυτή έχει το ίδιο στήριγμα με τη συνάρτηση στόχο  $\pi$  και είναι τέτοια ώστε να ικανοποιεί τη σχέση  $\pi(\eta) \leq cg(\eta)$ , για κάθε  $\eta \in \text{supp } g$  και κάποια σταθερά  $c \geq 1$ .

Παράγουμε μια ομοιόμορφη τυχαία μεταβλητή  $u \sim U(0, 1)$  και αν  $u \leq \frac{\pi(\eta)}{cg(\eta)}$  τότε αποδεχόμαστε την τιμή  $\eta$  ως κατανεμηθήσα από την κατανομή στόχο  $\pi$ .

Αποδεικνύεται ότι η βέλτιστη επιλογή για τη σταθερά είναι  $c = \max_{\text{supp } g} \left( \frac{\pi(\eta)}{g(\eta)} \right)$ .

Για περισσότερες λεπτομέρειες παραπέμπουμε στα συγγράμματα των Ripley (1987, section 3.2) και Devroye (1986, section II.3).

Είναι αξιοσημείωτο το γεγονός ότι η αναμενόμενη πιθανότητα αποδοχής, για την τυχαία μεταβλητή που προσομοιώνεται σύμφωνα με την κατανομή  $g$ , είναι υψηλότερη στον Αλγόριθμο 3.5.1 από ότι με τη μέθοδο της Αποδοχής - Απόρριψης. Αυτό οφείλεται στην αποτελεσματικότερη διαχείριση των τιμών που παράγονται από τη συνάρτηση φάκελο  $g$ , αφού λιγότερες τιμές απορρίπτονται από τον Αλγόριθμο 3.5.1.

**Λήμμα 3.5.1** *Αν η σχέση (3.13) ευσταθεί, τότε η αναμενόμενη πιθανότητα αποδοχής που συνδέεται με τον Αλγόριθμο 3.5.1 είναι τουλάχιστον  $\frac{1}{M}$  όταν η αλυσίδα είναι στάσιμη.*

Η σχέση (3.9), γνωστή και ως test ratio, υπό το πλαίσιο των ανεξάρτητων αλυσίδων παίρνει τη μορφή

$$\frac{\pi(\eta)/g(\eta)}{\pi(\theta)/g(\theta)} = \frac{\omega(\eta)}{\omega(\theta)}, \quad (3.15)$$

όπου η ποσότητα  $\omega = \pi/g$  αναφέρεται ως συνάρτηση στάθμισης (weighting function) και είναι η ίδια που χρησιμοποιείται στη δειγματοληψία σπουδαιότητας (importance sampling).

Για τους Μπεϋζιανούς Στατιστικούς, πολύ δημοφιλής επιλογή για τη συνάρτηση φάκελο  $g$  είναι η εκ των προτέρων κατανομή. Ο γενικός κανόνας για τις ανεξάρτητες αλυσίδες είναι η αποφυγή μεγάλης μεταβλητότητας για τη συνάρτηση στάθμισης, καθώς αυτό αυξάνει την πιθανότητα του ενδεχομένου να παραμείνει η αλυσίδα για πολλές επαναλήψεις του αλγόριθμου σε καταστάσεις με μεγάλη στάθμιση. Επομένως, συνίσταται η επιλογή συνάρτησης  $g$  τέτοιας ώστε η συνάρτηση στάθμισης  $\omega$  να είναι όσο πιο σταθερή γίνεται ή τουλάχιστον φραγμένη. Αφού οι συναρτήσεις  $\pi$  και  $g$  είναι και οι δύο πυκνότητες, η προηγούμενη σύσταση ισοδυναμεί με την επιλογή συνάρτησης  $g$  παρόμοιας, κατά το δυνατότερον, με τη συνάρτηση  $\pi$ .

Ο Tierney (1994) προτείνει την αποφυγή πυκνοτήτων  $g$  με ελαφριές ουρές, όπως είναι η κανονική κατανομή, και τη χρησιμοποίηση πυκνοτήτων της κατανομής του Student (t densities) με λίγους βαθμούς ελευθερίας. Με αυτή την τακτική, η συνάρτηση στάθμισης δεν επηρεάζεται σε μεγάλο βαθμό από τις ουρές της πυκνότητας  $g$  και είναι λιγότερο πιθανό να παρουσιάζει μεγάλη μεταβλητότητα.

Επίσης, είναι σημαντικό να εξασφαλίσουμε το γεγονός ότι οι ουρές της πυκνότητας πρότασης  $g$  επικρατούν αυτών της πυκνότητας στόχου  $\pi$ , προκειμένου να πετύχουμε ποσοστό αποδοχής των προτεινόμενων τιμών πλησίον του 50%, όπως επισημαίνεται από τους Gelman, Roberts και Gilks (1996). Στην αντίθετη περίπτωση, η  $g$  δεν μπορεί να εξερευνήσει τις περιοχές που αντιστοιχούν στην ουρά της  $\pi$ .

### 3.5.2 Αλυσίδες Τυχαίου Περιπάτου

Ο χαρακτηρισμός «αλυσίδες τυχαίου περιπάτου» (random walk chains) αναφέρεται στον πυρήνα μετάβασης πρότασης  $q$ . Τι είναι όμως ο τυχαίος περίπατος;

**Ορισμός 3.5.1** Υποθέστε ότι  $\mathbf{X} = (X_n : n \in \mathbb{Z}_+)$  είναι μια συλλογή τυχαίων μεταβλητών, η οποία ορίζεται με την επιλογή μιας αυθαίρετης κατανομής για την τυχαία μεταβλητή  $X_0$  και θέτοντας, για κάθε  $n \in \mathbb{Z}_+$ ,

$$X_{n+1} = X_n + W_{n+1}, \quad (3.16)$$

όπου οι τυχαίες μεταβλητές  $W_n$  είναι ανεξάρτητες και ισόνομες (iid) λαμβάνοντας τιμές στο σύνολο των πραγματικών αριθμών  $\mathbb{R}$ , με

$$P(W_n \leq y) = \Gamma(-\infty, y]. \quad (3.17)$$

Τότε η στοχαστική διαδικασία  $\mathbf{X}$  λέγεται τυχαίος περίπατος στο  $\mathbb{R}$ .

Η συνάρτηση πρότασης  $q$ , στον Αλγόριθμο 3.2.1, επιτρέπεται να εξαρτάται από την τρέχουσα τιμή  $\theta^{(j)}$ . Μια πολύ συνηθής επιλογή, για την προσομοίωση των προτεινόμενων τιμών  $\eta^{(j+1)}$ , είναι η ακόλουθη:

$$\eta^{(j+1)} = \theta^{(j)} + W^{(j+1)}, \quad (3.18)$$

όπου η τυχαία μεταβλητή  $W^{(j+1)}$ , έχει κατανομή  $\Gamma$ , όπως ορίστηκε παραπάνω και λέγεται τυχαίος θόρυβος (random perturbation or noise).

Αξίζει να παρακολουθήσουμε τη συγκεκριμένη μορφή που λαμβάνει η συνάρτηση πρότασης  $q$ , στην περίπτωση προσομοίωσης μέσω ενός τυχαίου περιπατού:

$$\begin{aligned} q(\theta, A) &= P(\eta^{(j+1)} \in A | \theta^{(j)} = \theta) \\ &= P(\theta^{(j)} + W^{(j+1)} \in A | \theta^{(j)} = \theta) \\ &= P(W^{(j+1)} = A - \theta) = \Gamma(A - \theta), \end{aligned} \quad (3.19)$$

όπου  $A - \theta := \{y - \theta; y \in A\}$ . Αν υποθέσουμε ότι η  $\Gamma$  έχει πυκνότητα  $\gamma$  τότε  $q(\theta, \eta) = \gamma(\eta - \theta)$ . Επιπλέον, αν η πυκνότητα  $\gamma$  είναι συμμετρική περί το μηδέν, δηλαδή  $\gamma(-\theta) = \gamma(\theta)$ , τότε η αλυσίδα είναι συμμετρική και λαμβάνουμε την ακόλουθη πρότυπη εκδοχή του Αλγόριθμου 3.2.1, όπως αυτή προτάθηκε από τους Metropolis et al. (1953):

**Αλγόριθμος 3.5.2** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)}$ .

2. Προσομοίωσε μια νέα τιμή  $\eta^{(j)} \sim \gamma(\eta - \theta^{(j-1)})$ .

3. Δέξου ως νέα τιμή για την αλυσίδα την

$$\theta^{(j)} = \begin{cases} \eta^{(j)}, & \text{με πιθανότητα } \alpha(\theta^{(j-1)}, \eta^{(j)}), \\ \theta^{(j-1)}, & \text{με πιθανότητα } 1 - \alpha(\theta^{(j-1)}, \eta^{(j)}). \end{cases}$$

4. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανάλαβε τη διαδικασία από το βήμα 2, μέχρι να εέλθει σύγκλιση.

$$\text{Η πιθανότητα αποδοχής είναι } \alpha(\theta^{(j-1)}, \eta^{(j)}) = \min \left( 1, \frac{\pi(\eta^{(j)})}{\pi(\theta^{(j-1)})} \right).$$

Παρατηρήστε ότι στην πιθανότητα αποδοχής, εξαιτίας της συμμετρικότητας της συνάρτησης  $\gamma$ , απαλοίφεται ο όρος

$$\frac{\gamma(\theta^{(j-1)} - \eta^{(j)})}{\gamma(\eta^{(j)} - \theta^{(j-1)})}.$$



Η παραπάνω εκδοχή του αλγόριθμου είναι πολύ συνηθισμένη, καθώς πολλές από τις πρακτικές υλοποιήσεις του Αλγόριθμου 3.2.1 χρησιμοποιούν αυτό το σχέδιο.

Οι πιο δημοφιλείς επιλογές για την κατανομή  $\gamma$  είναι η κανονική (Muller, 1991), η  $t$  του Student (1992) και η ομοιόμορφη. Συνήθως, οι κατανομές αυτές χρησιμοποιούνται στην τυποποιημένη μορφή τους.

Ένα σημαντικό ζήτημα που πρέπει να συζητηθεί είναι η διασπορά της πυκνότητας  $\gamma$ . Μεγάλη τιμή της διακύμανσης επιτρέπει τη μετακίνηση της αλυσίδας σε τιμή απομακρυσμένη από την προηγούμενη, υπό το κόστος του χαμηλού ποσοστού αποδοχής των προτεινόμενων τιμών από την πυκνότητα  $q$ . Από την άλλη πλευρά, μικρή τιμή της διακύμανσης επιτρέπει μόνο μικρές μετακινήσεις της αλυσίδας, όμως με υψηλό ποσοστό αποδοχής. Ο Tierney (1994) προτείνει να θέτουμε τον πίνακα διακυμάνσεων-συνδιακυμάνσεων της  $\gamma$  ως  $cV$ , όπου  $c$  είναι ένας θετικός αριθμός ο οποίος παίζει το ρόλο μιας σταθεράς συντονισμού (tuning constant) και  $V$  είναι κάποια προσέγγιση του πίνακα διακυμάνσεων-συνδιακυμάνσεων της κατανομής στόχου (π.χ. της εκ των υστέρων κατανομής). Με αυτή την τεχνική, οι κινήσεις της αλυσίδας είναι του ίδιου «εκτοπίσματος» σχετικά με το εύρος της κατανομής στόχου. Η επιλογή της σταθεράς συντονισμού εξαρτάται από τη μορφή της βελτιστοποίησης που επιθυμούμε (ποσοστό αποδοχής/μέγεθος μεταβάσεων). Ο Tierney (1994) προτείνει τιμές μεταξύ 1/2 και 1. Οι Gelman, Roberts και Gilks (1996) αποκόμισαν το δικό τους βέλτιστο ποσοστό αποδοχής για κανονικό τυχαίο περίπατο με τιμές της σταθεράς  $c$  μεταξύ 2 και 3.

Οι ιδιότητες σύγκλισης που παρουσιάστηκαν στην Ενότητα 3.4, ευσταθούν στην ειδική περίπτωση των αλυσίδων τυχαίου περιπάτου.

Με την επιβολή κάποιων αναγκαίων και επαρκών συνθηκών, είναι δυνατό να αποκομίσουμε γεωμετρική εργοδικότητα για την αλυσίδα που παράγεται από τον Αλγόριθμο 3.5.2. Οι Mengersen και Tweedie (1996) απέδειξαν μια συνθήκη, βασιζόμενοι στη λογαριθμική κοιλότητα (log-concavity) της συνάρτησης στόχου  $\pi$ , στις ουρές: δηλαδή, αν υπάρχει  $a > 0$  και  $x_1$  τέτοια ώστε,

$$\log \pi(x) - \log \pi(y) \geq a|y - x|, \quad (3.20)$$

για  $y < x < -x_1$  ή  $x_1 < x < y$ .

**Θεώρημα 3.5.2** Έστω μια συμμετρική πυκνότητα  $\pi$  της οποίας ο λογάριθμος είναι κοίλη συνάρτηση (log-concave) με σταθερά  $a$ , όπως αυτή αναφέρεται στον τύπο (3.20), για  $|x|$  αρκούντως μεγάλο. Αν η πυκνότητα  $\gamma$  είναι θετική και συμμετρική, η αλυσίδα  $(\theta^{(t)})$  που παράγεται από τον Αλγόριθμο 3.5.2 είναι

γεωμετρικά εργοδική. Αν η  $\pi$  δεν είναι συμμετρική, μια επαρκής συνθήκη για γεωμετρική εργοδικότητα είναι η πυκνότητα  $\gamma(t)$  να είναι φραγμένη από την ποσότητα  $b \exp\{-a|t|\}$ , για κάποια αρκετά μεγάλη σταθερά  $b$ .

Ωστόσο, ο Αλγόριθμος 3.5.2 δεν απολαμβάνει ιδιότητες ομοιόμορφης εργοδικότητας. Οι Mengersen και Tweedie (1996) απέδειξαν ότι στην περίπτωση όπου  $\text{supp } \pi = \mathbb{R}$ , ο Αλγόριθμος 3.5.2 δεν μπορεί να παραγάγει ομοιόμορφα εργοδική αλυσίδα στο  $\mathbb{R}$ .

### 3.5.3 Αυτοπαλίνδρομες Αλυσίδες

Μια άλλη μορφή αλγόριθμου, προέρχεται από την τροποποίηση των αλυσίδων τυχαίου περιπάτου σε αυτοπαλίνδρομες αλυσίδες (autoregressive chains) της μορφής

$$\eta^{(j)} = a + b(\theta^{(j-1)} - a) + W_j, \quad (3.21)$$

όπου η τυχαία προσαύξηση (τυχαίος θόρυβος)  $W_j$  κατανέμεται σύμφωνα με την κατανομή  $\Gamma$ . Η κατανομή πρότασης γίνεται  $q(\theta, \eta) = \gamma(\eta - a - b(\theta - a))$  και προκύπτει ακριβώς με τον ίδιο τρόπο που προέκυψε η σχέση (3.19). Αυτή η αυτοπαλίνδρομη αναπαράσταση μπορεί να θεωρηθεί ως μια ενδιάμεση κατάσταση μεταξύ της ανεξάρτητης εκδοχής ( $b = 0$ ) και του τυχαίου περιπάτου ( $b = 1$ ). Για  $b < 0$  οι παράμετροι  $\theta_j$  και  $\theta_{j-1}$  είναι αρνητικά συσχετισμένες, κάτι που μπορεί να επιτρέψει γρηγορότερη περιήγηση στην επιφάνεια (surface) της πυκνότητας  $\pi$ , αν το σημείο συμμετρίας  $a$  είναι σωστά επιλεγμένο.

Οι επιλογές  $b = 1$  και  $b = -1$  συγκρίθηκαν από τον Hastings (1970) για μονομεταβλητή κανονική κατανομή και από τους Chib και Greenberg (1995) για διμεταβλητή κανονική κατανομή με υψηλή συσχέτιση. Και στις δύο περιπτώσεις, η επιλογή  $b = -1$  παρήγαγε καλύτερους εκτιμητές της κατανομής στόχου. Οι Barone και Frigessi (1989) παρέχουν περαιτέρω θεωρητικό υπόβαθρο για την επιλογή του  $b$ . Το αποτέλεσμα των αρνητικών συσχετίσεων που επιτυγχάνεται με την επιλογή  $b = -1$  δεν είναι καινούρια ιδέα στην προσομοίωση, αφού είναι το βασικό αποτέλεσμα της μεθόδου των αντίθετων τυχαίων μεταβλητών (antithetic variables).

### 3.5.4 Αλυσίδες Πολλαπλασιαστικού Τυχαίου Περιπάτου

Μια εναλλακτική μορφή του αλγόριθμου του τυχαίου περιπάτου, είναι ο αλγόριθμος του λογαριθμικού τυχαίου περιπάτου, όπου η προτεινόμενη μετάβαση

είναι τυχαίο πολλαπλάσιο της τρέχουσας κατάστασης. Ο αλγόριθμος αυτός λέγεται αλγόριθμος πολλαπλασιαστικού τυχαίου περιπάτου (multiplicative random walk algorithm). Εφαρμόζεται στην περίπτωση που ο χώρος καταστάσεων της αλυσίδας είναι υποσύνολο του  $\mathbb{R}_+$ . Για παράδειγμα, μπορούμε να ορίσουμε μια ειδική μορφή του τυχαίου περιπάτου, ως εξής:

**Ορισμός 3.5.2** Υποθέστε ότι η στοχαστική διαδικασία  $\mathbf{X} = \{X_n : n \in \mathbb{Z}_+\}$  ορίζεται επιλέγοντας μια αυθαίρετη κατανομή για την τυχαία μεταβλητή  $X_0$  και θέτοντας, για κάθε  $n \in \mathbb{Z}_+$ ,

$$X_{n+1} = [X_n + W_{n+1}]^+, \quad (3.22)$$

όπου  $[X_n + W_{n+1}]^+ := \max(0, X_n + W_{n+1})$  και οι τυχαίες μεταβλητές  $W_n$  είναι ανεξάρτητες και ισόνομες (*iid*) με τιμές στο  $\mathbb{R}$ , και συνάρτηση κατανομής

$$P(W_n \leq y) = \Gamma(-\infty, y]. \quad (3.23)$$

Τότε η στοχαστική διαδικασία  $\mathbf{X}$  λέγεται τυχαίος περίπατος στο  $\mathbb{R}_+$ .

Οι προτεινόμενες τιμές είναι μη αρνητικές και επομένως ο χώρος καταστάσεων, της αλυσίδας που προκύπτει από τον Αλγόριθμο 3.5.2, είναι ο  $\mathbb{R}_+$ . Σε μια περίπτωση σαν κι' αυτή, ο αλγόριθμος του πολλαπλασιαστικού τυχαίου περιπάτου είναι μια καλή επιλογή.

Οι προτεινόμενες τιμές δίνονται από τον τύπο  $\eta^{(j)} = \theta^{(j-1)}e^{W^{(j)}}$ , όπου η τυχαία μεταβλητή  $W^{(j)}$  κατανέμεται σύμφωνα με κάποια συμμετρική πυκνότητα  $\gamma$ . Τυποποιημένα, ο αλγόριθμος δίνεται ως εξής:

**Αλγόριθμος 3.5.3** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)}$ .

2. Προσομοίωσε μια νέα τιμή  $\eta^{(j)} \sim \frac{1}{\eta^{(j)}\gamma} \left[ \ln \left( \frac{\eta^{(j)}}{\theta^{(j-1)}} \right) \right]$ .

3. Δέξου ως νέα τιμή για την αλυσίδα την

$$\theta^{(j)} = \begin{cases} \eta^{(j)}, & \text{με πιθανότητα } \alpha(\theta^{(j-1)}, \eta^{(j)}), \\ \theta^{(j-1)}, & \text{με πιθανότητα } 1 - \alpha(\theta^{(j-1)}, \eta^{(j)}). \end{cases}$$

4. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανέλαβε τη διαδικασία από το βήμα 2, μέχρι να επέλθει σύγκλιση.

$$\text{Δίνεται } \alpha(\theta^{(j-1)}, \eta^{(j)}) = \min \left( 1, \frac{\pi(\eta^{(j)})\eta^{(j)}}{\pi(\theta^{(j-1)})\theta^{(j-1)}} \right).$$

Μια επέκταση του αλγόριθμου που μόλις περιγράψαμε, είναι ο αλγόριθμος του τροποποιημένου πολλαπλασιαστικού τυχαίου περίπατου (modified multiplicative random walk algorithm). Χρησιμοποιεί το μετασχηματισμό  $\log(\theta + a)$ , για κάποια θετική σταθερά  $a$ , αντί του  $\log(\theta)$  και εφαρμόζεται όταν το στήριγμα της κατανομής στόχου  $\pi$  δεν είναι φραγμένο σε κάποια περιοχή μακριά από το μηδέν. Έτσι αποφεύγονται τα προβλήματα που προκαλούνται από αυθέραιτα μικρές μεταβάσεις της αλυσίδας στην περιοχή πλησίον του μηδένος.

### 3.6 Βελτίωση του Ποσοστού Αποδοχής

Στην ενότητα αυτή θα παρουσιάσουμε κάποιες σκέψεις σχετικά με τη βελτίωση του ποσοστού αποδοχής, σε δύο από τις πιο δημοφιλείς εκδοχές του αλγόριθμου Metropolis-Hastings: τις ανεξάρτητες αλυσίδες και τις αλυσίδες τυχαίου περίπατου. Ωστόσο, εκείνο που ενδιαφέρει είναι, όχι τόσο η μελέτη των ειδικών αυτών περιπτώσεων, αλλά η λογική που ακολουθείται ώστε να μπορεί ο αναγνώστης να γενικεύσει τα αποτελέσματα.

#### 3.6.1 Ανεξάρτητες Αλυσίδες

Για την περίπτωση των ανεξάρτητων αλυσίδων, επιλέγουμε πυκνότητα πρότασης  $g$  η οποία προσεγγίζει την πυκνότητα στόχο  $\pi$ , έτσι ώστε ο λόγος  $\pi/g$  να είναι φραγμένος. Στην περίπτωση αυτή, ο Αλγόριθμος 3.5.1 απολαμβάνει ιδιότητες ομοιόμορφης εργοδικότητας. Η ομοιότητά του με τον αλγόριθμο της μεθόδου Αποδοχής - Απόρριψης συνηγορεί στην επιλογή μιας συνάρτησης  $g$  η οποία μεγιστοποιεί το αναμενόμενο ποσοστό αποδοχής

$$\begin{aligned} E \left[ \min \left( 1, \frac{\pi(\eta)g(\theta)}{\pi(\theta)g(\eta)} \right) \right] \\ = 2P \left( \frac{\pi(\eta)}{g(\eta)} \geq \frac{\pi(\theta)}{g(\theta)} \right), \end{aligned} \quad (3.24)$$

όπου  $\theta \sim \pi$  και  $\eta \sim g$ . Στην πραγματικότητα, η βελτιστοποίηση που επιζητείται, με την επιλογή της πυκνότητας πρότασης  $g$ , μεταφράζεται σε όρους της ταχύτητας σύγκλισης του δειγματικού μέσου (3.2) στον πληθυσμιακό μέσο  $E_\pi[h(\theta)]$  και κατά συνέπεια στη ικανότητα του Αλγόριθμου 3.5.1 να εξερευνεί ταχέως κάθε πολυπλοκότητα της πυκνότητας  $\pi^1$ .

<sup>1</sup> Δείτε, για παράδειγμα, το Θεώρημα 3.5.1.

Αν θέλουμε αυτή η βελτιστοποίηση να είναι γενικότερη, δηλαδή η προαναφερθείσα ταχύτητα σύγκλισης να είναι ανεξάρτητη από τη συνάρτηση  $h$ , τότε πρέπει η πυκνότητα  $g$  να αναπαράγει την  $\pi$  όσο πιο πιστά γίνεται, γεγονός που υπονοεί τη μεγιστοποίηση του ποσοστού αποδοχής  $\alpha$ .

Το ποσοστό αποδοχής  $\alpha$  είναι δύσκολο να υπολογιστεί. Μια διέξοδος είναι η χρησιμοποίηση του αποτελέσματος του Λήμματος 3.5.1, δηλαδή  $\alpha \geq 1/M$ , με ελαχιστοποίηση της σταθεράς  $M$ .

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε μια πιο εμπειρική προσέγγιση: επιλέγουμε μια πυκνότητα πρότασης  $g$ , η οποία εξαρτάται από μία παράμετρο  $\lambda$ , δηλαδή  $g(\cdot|\lambda)$ . Στη συνέχεια, αναπροσαρμόζουμε τις τιμές της παραμέτρου, με οδηγό την εκτίμηση του ποσοστού αποδοχής, το οποίο τώρα μπορεί να εκφραστεί ως συνάρτηση του  $\lambda$ , δηλαδή  $\alpha(\lambda)$ . Πιο συγκεκριμένα, επιλέγουμε μια αρχική τιμή  $\lambda_0$  για την παράμετρο, εκτιμούμε το αντίστοιχο ποσοστό αποδοχής  $\hat{\alpha}(\lambda_0)$ , ύστερα από  $m$  επαναλήψεις του Αλγόριθμου 3.5.1 και τροποποιούμε το  $\lambda_0$  ώστε να αποκομίσουμε αύξηση του ποσοστού αποδοχής.

Στην απλούστερη περίπτωση, η παράμετρος  $\lambda$  είναι μια μονοδιάστατη παράμετρος θέσης, η οποία αυξάνεται ή μειώνεται ανάλογα με τη συμπεριφορά της εκτίμησης  $\hat{\alpha}(\lambda)$ . Σε σύνθετες περιπτώσεις, το  $\lambda$  είναι διάνυσμα που μπορεί να περιλαμβάνει μια παράμετρο θέσης ή ένα πίνακα στο ρόλο της παραμέτρου κλίμακας. Εδώ, η μεγιστοποίηση του ποσοστού αποδοχής  $\alpha(\lambda)$  είναι μια δύσκολη υπόθεση.

Για την εκτίμηση της ποσότητας  $\alpha(\lambda)$  μπορεί να χρησιμοποιηθεί ο απλός εκτιμητής που καταμετρά τις αποδοχές των προτεινόμενων τιμών ή ο

$$\frac{2}{m} \sum_{j=1}^m I(\pi(\eta^{(j)})g(\theta^{(j)}) > \pi(\theta^{(j)})g(\eta^{(j)})), \quad (3.25)$$

όπου  $\theta^{(1)}, \dots, \theta^{(m)}$  είναι μια δειγματική πραγματοποίηση από την πυκνότητα  $\pi$ , την οποία αποκτήσαμε από ένα αρχικό MCMC αλγόριθμο και  $\eta^{(1)}, \dots, \eta^{(m)}$  είναι μια δειγματική πραγματοποίηση ενός τυχαίου δείγματος ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από την κατανομή  $g$ . Επομένως, αν το  $\lambda$  αποτελείται από παραμέτρους θέσεως και κλίμακας, η δειγματική πραγματοποίηση  $((\theta^{(1)}, \eta^{(1)}), \dots, (\theta^{(m)}, \eta^{(m)}))$  που αντιστοιχεί στην αρχική τιμή  $\lambda_0$ , μπορεί να χρησιμοποιηθεί επαναληπτικά για να εκτιμήσει διαφορετικές τιμές του  $\lambda$  μέσω κάποιας νετερμινιστικής τροποποίησης των τιμών  $\eta^{(j)}$  η οποία διευκολύνει τη μεγιστοποίηση του ποσοστού αποδοχής  $\alpha(\lambda)$ .

### 3.6.2 Αλυσίδες Τυχαίου Περιπάτου

Η περίπτωση της προσομοίωσης μέσω ενός τυχαίου περιπάτου, όπως παρουσιάστηκε στην Παράγραφο 3.5.2, απαιτεί ένα διαφορετικό χειρισμό του ποσοστού αποδοχής, δεδομένης της εξάρτησης της κατανομής πρότασης  $\Gamma$  από την τρέχουσα κατάσταση της αλυσίδας. Στην πραγματικότητα, ένα υψηλό ποσοστό αποδοχής δεν σημαίνει απαραίτητα ότι ο Αλγόριθμος 3.5.2 εξελίσσεται σωστά. Μπορεί, για παράδειγμα, να υποδηλώνει ότι ο τυχαίος περίπατος κινείται πολύ αργά στην επιφάνεια (surface) της πυκνότητας στόχου  $\pi$ . Αν η τρέχουσα τιμή  $\theta^{(j-1)}$  και η προτεινόμενη  $\eta^{(j)}$  είναι κοντά, με την έννοια ότι  $\pi(\theta^{(j-1)}) \approx \pi(\eta^{(j)})$ , τότε ο Αλγόριθμος 3.5.2 αποδέχεται την προτεινόμενη τιμή  $\eta^{(j)}$  με πιθανότητα

$$\min \left( 1, \frac{\pi(\eta^{(j)})}{\pi(\theta^{(j-1)})} \right) \approx 1. \quad (3.26)$$

Ένα υψηλό ποσοστό αποδοχής είναι πιθανό να αντιστοιχεί σε μια πιο αργή σύγκλιση, καθώς οι μετακινήσεις στο στηρίγμα της κατανομής στόχου  $\pi$  είναι περιορισμένες. Στην περίπτωση πολυκόρυφων πυκνοτήτων (multimodal densities) στις οποίες οι κορυφές (modes) διαχωρίζονται από ζώνες εξαιρετικά χαμηλής πιθανότητας, η αρνητική επίδραση των περιορισμένων μετακινήσεων πάνω στην επιφάνεια της πυκνότητας  $\pi$  είναι ξεκάθαρη. Ενώ το ποσοστό αποδοχής είναι αρκετά υψηλό για μία κατανομή  $g$  με μικρή διακύμανση, η πιθανότητα της μετάβασης από τη μία κορυφή στην άλλη μπορεί να είναι πολύ χαμηλή. Το φαινόμενο αυτό εμφανίζεται, για παράδειγμα, στην περίπτωση της μείξης κατανομών (mixture of distributions) και στην υπερπαραμετροποίηση μοντέλων (overparameterized models). Παραπέμπουμε στους Tanner και Wong (1987) και στους Besag et al. (1995).

Στην αντίθετη περίπτωση, αν το αναμενόμενο ποσοστό αποδοχής είναι χαμηλό, οι διαδοχικές τιμές  $\pi(\eta^{(j)})$  τείνουν να είναι χαμηλές σε σχέση με τις τιμές  $\pi(\theta^{(j-1)})$ , που σημαίνει ότι ο τυχαίος περίπατος κινείται γρήγορα (με μεγάλα άλματα) στην επιφάνεια της πυκνότητας  $\pi$ , αφού συχνά αγγίζει τα όρια του στηρίγματος της  $\pi$  (ή τουλάχιστον ο τυχαίος περίπατος εξερευνεί περιοχές με χαμηλή πιθανότητα υπό την κατανομή  $\pi$ ).

Η παραπάνω ανάλυση δείχνει να απαιτεί την καλή γνώση της πυκνότητας στόχου  $\pi$ , αφού μια κατανομή πρότασης  $g$  με πολύ στενό εύρος, επιβραδύνει το ρυθμό σύγκλισης του αλγόριθμου. Από την άλλη μεριά, μια κατανομή  $g$  με μεγάλο εύρος «καταναλώνει» τον αλγόριθμο στην προσομοίωση άχρηστων σημείων τα οποία βρίσκονται εκτός του στηρίγματος της  $\pi$ , χωρίς αυτό

να σημαίνει τη βελτίωση της πιθανότητας του ενδεχομένου να επισκεφθεί η συνάρτηση πρότασης  $g$  όλες τις κορυφές της  $\pi$ .

## 3.7 Υβριδικοί Αλγόριθμοι

Οι υβριδικοί αλγόριθμοι (hybrid algorithms) προκύπτουν από τροποποιήσεις της βασικής μορφής του Metropolis-Hastings αλγόριθμου. Στην ενότητα αυτή θα παρουσιαστούν κάποιοι τέτοιοι αλγόριθμοι και θα δούμε ποια ανάγκη οδήγησε στη δημιουργία τους.

### 3.7.1 Metropolis-Hastings κατά συντεταγμένη

Ας υποθέσουμε ότι επιθυμούμε να προσομοιώσουμε από μια εκ των υστέρων κατανομή  $\pi(\theta)$ , όπου  $\theta = (\theta_1, \dots, \theta_d)$ . Αντί να προσομοιώνουμε όλο μαζί το διάνυσμα  $\theta$  σε κάθε βήμα του αλγόριθμου είναι πολλές φορές πιο βολικό και υπολογιστικά πιο εύκολο να προσομοιώνουμε ξεχωριστά κάθε συντεταγμένη του διανύσματος  $\theta$ . Οι συντεταγμένες  $\theta_1, \dots, \theta_d$  είναι δυνατό να αποτελούν και αυτές άλλα μικρότερα διανύσματα, διαφορετικών ίσως διαστάσεων. Ο αλγόριθμος που θα παρουσιάσουμε λέγεται Metropolis-Hastings κατά συντεταγμένη (single component or componentwise Metropolis-Hastings).

Οι Hastings (1970) και Tierney (1994) διαπραγματεύονται ένα τέτοιο σχέδιο προσομοίωσης. Πιο συγκεκριμένα, οι συντεταγμένες του  $\theta$  μπορούν να αναβαθμίζονται, σε κάθε βήμα του αλγόριθμου, με τους ακόλουθους τρόπους:

1. Σε κάθε επανάληψη, μία συντεταγμένη αναβαθμίζεται και η επιλογή της γίνεται τυχαία ανάμεσα στις  $d$  συντεταγμένες.
2. Σε κάθε επανάληψη, μία συντεταγμένη αναβαθμίζεται και η επιλογή της γίνεται με προκαθορισμένη σειρά ανάμεσα στις  $d$  συντεταγμένες. Για παράδειγμα, οι συντεταγμένες αναβαθμίζονται κατά σειρά, δηλαδή  $1 \rightarrow 2 \rightarrow \dots \rightarrow d$ .

Στην πρώτη περίπτωση λέμε ότι έχουμε μεταβάσεις μείξης (mixture transitions) και στη δεύτερη κυκλικές μεταβάσεις (cycle transitions).

Στις μεταβάσεις μείξης ορίζουμε πυρήνες μετάβασης  $p_m$ , οι οποίοι έχουν κοινή στάσιμη κατανομή  $\pi$  και πιθανότητες ή βάρη  $w_m$ ,  $m = 1, \dots, d$ , όπου  $w_m \geq 0$  και  $\sum_{m=1}^d w_m = 1$ . Ο πυρήνας  $p_m$  επιλέγεται με πιθανότητα  $w_m$  και αναβαθμίζει τη συντεταγμένη  $\theta_m$ . Δεν είναι απαραίτητο να αναβαθμίζονται όλες

οι συντεταγμένες του διανύσματος  $\theta$  σε κάθε επανάληψη του αλγόριθμου. Οι Zeger και Karim (1991) προτείνουν να αναβαθμίζονται οι υψηλά συσχετισμένες συντεταγμένες πιο συχνά από τις υπόλοιπες. Με αυτό τον τρόπο επιταχύνεται η σύγκλιση.

Μια μετάβαση μείξης  $p$  σχηματίζεται από τη σχέση  $p = \sum_{m=1}^d w_m p_m$ . Στην περίπτωση πλήρως τυχαίας επιλογής συντεταγμένης, τα βάρη είναι  $w_m = 1/d$ . Επαναλαμβάνουμε ότι κάθε πυρήνας μετάβασης  $p_m$  μετακινεί μόνο την  $m$  συντεταγμένη του διανύσματος  $\theta$ .

Οι ιδιότητες των πυρήνων μετάβασης  $p_m$  περνούν στον πυρήνα μείξης  $p$ . Ο πυρήνας αυτός ορίζει τις μεταβάσεις μιας αλυσίδας Markov με στάσιμη κατανομή την  $\pi$ . Επιπλέον, αν ένας από τους πυρήνες  $p_m$  είναι ανάγωγος και απεριοδικός, τότε ο πυρήνας μείξης  $p$  είναι ανάγωγος και απεριοδικός.

Στην περίπτωση των κυκλικών μεταβάσεων με πυρήνες κυκλικής μετάβασης  $p_c$ ,  $c = 1, \dots, d$ , μια επανάληψη του αλγόριθμου σημαίνει την αναβάθμιση όλων των συντεταγμένων του διανύσματος  $\theta$ , όπως αυτή υπαγορεύεται από τους πυρήνες  $p_c$ .

Πολλές από τις ιδιότητες των πυρήνων μετάβασης  $p_c$  περνούν στον κυκλικό πυρήνα  $p$ . Ο πυρήνας αυτός ορίζει μια αλυσίδα Markov με στάσιμη κατανομή την  $\pi$ . Αντίθετα με την περίπτωση των πυρήνων μείξης, η αναγωγικότητα και απεριοδικότητα ενός εκ των πυρήνων  $p_c$ , δεν είναι γενικά επαρκής για την αναγωγικότητα και απεριοδικότητα του κυκλικού πυρήνα  $p$ . Κάτι τέτοιο όμως συμβαίνει σίγουρα όταν όλοι οι πυρήνες  $p_c$  είναι ανάγωγοι και απεριοδικοί (Tierney (1994)).

Τα δύο παραπάνω δειγματοληπτικά σχέδια μπορούν να εφαρμοστούν με έναν αλγόριθμο Metropolis-Hastings. Ας θεωρήσουμε την περίπτωση των κυκλικών μεταβάσεων. Μια πλήρης επανάληψη του αλγόριθμου αποτελείται από  $d$  βήματα.

Ορίζουμε το διάνυσμα  $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ , το οποίο αποτελείται από όλες τις συντεταγμένες του διανύσματος  $\theta$  εκτός από τη  $\theta_i$ . Επίσης, ορίζουμε το διάνυσμα  $\theta^{(t)}$  που δηλώνει την κατάσταση του  $\theta$  στο τέλος της  $t$  επανάληψης του αλγόριθμου, καθώς και το διάνυσμα  $\theta_{-i}^{(t)} = (\theta_1^{(t)}, \dots, \theta_{i-1}^{(t)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})$ , που δηλώνει την τιμή του διανύσματος  $\theta_{-i}$  κατά την ολοκλήρωση του  $i - 1$  βήματος της  $t + 1$  επανάληψης του αλγόριθμου. Παρατηρήστε ότι έχουν αναβαθμισθεί μόνο οι  $i - 1$  συντεταγμένες του  $\theta_{-i}$ , στη συγκεκριμένη  $(t + 1)$  επανάληψη του αλγόριθμου.

Στο βήμα  $i$  της  $t + 1$  επανάληψης του αλγόριθμου, αναβαθμίζεται η συντε-



ταγμένη  $\theta_i$  με τη βοήθεια του Metropolis-Hastings αλγόριθμου. Η προτεινόμενη τιμή  $\eta_i$  προέρχεται από μια κατανομή πρότασης  $q_i(\eta_i|\theta_i^{(t)}, \theta_{-i}^{(t)})$ . Επομένως, η  $i$  κατανομή πρότασης  $q_i(\cdot|\cdot, \cdot)$  γεννάει μια υποψήφια τιμή μόνο για την  $i$  συντεταγμένη του διανύσματος  $\theta$  και μπορεί να εξαρτάται από τις τρέχουσες τιμές οποιωνδήποτε συντεταγμένων του  $\theta$ . Η υποψήφια τιμή γίνεται αποδεκτή με πιθανότητα  $\alpha(\theta_{-i}^{(t)}, \theta_i^{(t)}, \eta_i)$ , όπου

$$\alpha(\theta_{-i}, \theta_i, \eta_i) = \min\left(1, \frac{\pi(\eta_i|\theta_{-i}) q_i(\theta_i|\eta_i, \theta_{-i})}{\pi(\theta_i|\theta_{-i}) q_i(\eta_i|\theta_i, \theta_{-i})}\right). \quad (3.27)$$

Η κατανομή  $\pi(\theta_i|\theta_{-i}) = \pi(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$  λέγεται πλήρως δεσμευμένη κατανομή (full conditional distribution) της παραμέτρου  $\theta_i$ , υπό την κατανομή  $\pi$ . Δηλαδή, είναι η δεσμευμένη κατανομή της παραμέτρου  $\theta_i$ , δοθειςών όλων των υπόλοιπων συντεταγμένων του διανύσματος  $\theta$ , το οποίο κατανέμεται σύμφωνα με την  $\pi$ :

$$\pi(\theta_i|\theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}. \quad (3.28)$$

Αν η προτεινόμενη τιμή  $\eta_i$  γίνει αποδεκτή, τότε θέτουμε  $\theta_i^{(t+1)} = \eta_i$ , διαφορετικά  $\theta_i^{(t+1)} = \theta_i^{(t)}$ . Οι υπόλοιπες συντεταγμένες του διανύσματος  $\theta$  παραμένουν αμετάβλητες κατά το βήμα  $i$ .

**Αλγόριθμος 3.7.1** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)}$  για το διάνυσμα  $\theta$ .

2. Θέσε το μετρητή της συντεταγμένης που προσομοιώνεται (component counter)  $i = 1$ .

3. Για τη συντεταγμένη  $i$ , προσομοίωσε μια νέα τιμή  $\eta_i$  προερχόμενη από την κατανομή με πυκνότητα  $q_i(\eta_i|\theta_i^{(j-1)}, \theta_{-i}^{(j-1)})$ .

4. Υπολόγισε την πιθανότητα αποδοχής της μετακίνησης του βήματος 3  $\alpha(\theta_{-i}^{(j-1)}, \theta_i^{(j-1)}, \eta_i)$ , η οποία δίνεται από τον τύπο (3.27). Αν η μετακίνηση γίνει αποδεκτή, τότε  $\theta_i^{(j)} = \eta_i$ , διαφορετικά  $\theta_i^{(j)} = \theta_i^{(j-1)}$  και η αλυσίδα δεν μετακινείται ως προς τη συντεταγμένη  $i$ .

5. Άλλαξε το μετρητή από  $i$  σε  $i + 1$  και επανέλαβε τα βήματα 3 και 4, έως  $i = d$ .

6. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανέλαβε τη διαδικασία από το βήμα 2 μέχρι να επέλθει σύγκλιση.

Στην πραγματικότητα, αυτή ήταν η προτότυπη μορφή του αλγόριθμου που προτάθηκε από τους Metropolis et al. (1953). Τα σωματίδια, τα οποία αναφέρθηκαν στην Ενότητα 2.1, αλλάζουν θέσεις ένα προς ένα, σύμφωνα με μια συμμετρική κατανομή μετάβασης. Αυτή είναι μια ομοιόμορφη διμεταβλητή κατανομή, η οποία έχει ως παράμετρο θέσης την προηγούμενη θέση του σωματίδιου.

Η περιγραφή του παραπάνω αλγόριθμου επικεντρώθηκε στην περίπτωση κυκλικών μεταβάσεων, όπου οι συντεταγμένες του διανύσματος  $\theta$  αναβαθμίζονται μία προς μία, με τη σειρά που δίνονται στο διάνυσμα. Τα δειγματοληπτικά σχέδια μείξης και κύκλου υπονοούν την ύπαρξη και άλλων τρόπων εκσυγχρονισμού του διανύσματος  $\theta$ . Για παράδειγμα, είναι δυνατό κάποιες συντεταγμένες να ενημερώνονται πιο συχνά από άλλες, είτε επειδή οι τελευταίες είναι δύσκολο να προσομοιωθούν, ή λόγω της ικανότητας τους να μετακινούνται πιο ελεύθερα μέσα στο χώρο καταστάσεων. Στην περίπτωση των μεταβάσεων μείξης, αν υποθέσουμε ότι η πιθανότητα επιλογής της συντεταγμένης  $i$ ,  $i = 1, \dots, d$  είναι  $s(i)$  και επιτρέπεται να εξαρτάται από την τρέχουσα κατάσταση της  $i$  συντεταγμένης  $\theta_i^{(t)}$ , τότε η (3.27) πρέπει να τροποποιηθεί, αλλιώς η στάσιμη κατανομή της αλυσίδας δεν θα συμπίπτει με την κατανομή στόχο  $\pi$ . Συγκεκριμένα η πιθανότητα αποδοχής γίνεται

$$\alpha(\theta_{-i}, \theta_i, \eta_i) = \min \left( 1, \frac{\pi(\eta_i | \theta_{-i}) s(i | \eta_i, \theta_{-i}) q_i(\theta_i | \eta_i, \theta_{-i})}{\pi(\theta_i | \theta_{-i}) s(i | \theta_i, \theta_{-i}) q_i(\eta_i | \theta_i, \theta_{-i})} \right).$$

Μέχρι τώρα, κάθε συντεταγμένη του διανύσματος  $\theta$  αναβαθμιζόταν με τη βοήθεια του δικού της πυρήνα μετάβασης  $p_m$  ή  $p_c$  για τις μεταβάσεις μείξης και τις κυκλικές μεταβάσεις, αντίστοιχα. Μια άλλη παραλλαγή είναι η χρήση διαφορετικών πυρήνων μετάβασης για την αναβάθμιση της ίδιας συντεταγμένης ή συνόλου συντεταγμένων. Αυτό, γενικά, δεν εξασφαλίζει γρηγορότερη σύγκλιση. Οι Gelfand and Carlin (1995) παρουσιάζουν ένα όμορφο παράδειγμα όπου με τη μείξη διαφορετικών δειγματοληπτικών σχεδίων σε μια απλή αλυσίδα Markov, επιτυγχάνεται αξιοσημείωτη βελτίωση στην ταχύτητα σύγκλισης.

Οι πυρήνες μετάβασης πρότασης που παρουσιάστηκαν στην Ενότητα 3.5 μπορούν να χρησιμοποιηθούν και στον Metropolis-Hastings αλγόριθμο κατά συντεταγμένη, για την κατασκευή των πυρήνων  $q_i(\cdot | \cdot, \cdot)$ ,  $i = 1, \dots, d$ . Επίσης, για κάποιες συντεταγμένες μπορεί να είναι πιο λογικό να χρησιμοποιήσουμε ανεξάρτητες αλυσίδες, βασισμένες στην εκ των προτέρων κατανομή, δεδομένου

ότι υπάρχει αρκετή πληροφορία για τις συντεταγμένες αυτές. Άλλες συντεταγμένες μπορούν να προσομοιωθούν με τη βοήθεια των αλυσίδων τυχαίου περιπάτου.

Η ανάγκη να αντιμετωπίσουμε κάθε συντεταγμένη ξεχωριστά, είναι αυτή που οδήγησε στη δημιουργία της μεθόδου. Συνήθως, η πιθανότητα μετάβασης (3.27) απλοποιείται δίνοντας ένα σημαντικό υπολογιστικό πλεονέκτημα. Ένα άλλο πλεονέκτημα της μεθόδου εμφανίζεται όταν η κατανομή στόχος  $\pi$  προσδιορίζεται, εκ φύσεως, σε όρους των πλήρεις δεσμευμένων κατανομών της.

Κάθε ένας από τους πυρήνες μετάβασης  $p_i$ ,  $i = 1, \dots, d$  ορίζει μια αντιστρέψιμη (reversible) αλυσίδα με στάσιμη κατανομή την  $\pi_i(\theta_i|\theta_{-i})$ . Επομένως, κάθε πυρήνας  $p_i$  ικανοποιεί τη σχέση

$$\pi_i(\eta_i|\theta_{-i}) = \int \pi_i(\theta_i|\theta_{-i}) p_i(\theta_i, \eta_i) d\theta_i. \quad (3.29)$$

**Παράδειγμα 3.7.1** Έστω το διάνυσμα  $\theta = (\theta_1, \theta_2)$ . Μια μετακίνηση της αλυσίδας αυτής, από την κατάσταση  $\theta = (\theta_1, \theta_2)$  στην κατάσταση  $\eta = (\eta_1, \eta_2)$  πραγματοποιείται με τη μετακίνηση και των δύο συντεταγμένων, σύμφωνα με τους αντίστοιχους πυρήνες. Αν  $\pi$  είναι η στάσιμη κατανομή για ένα κυκλικό πυρήνα, τότε γράφουμε

$$\pi(\eta_1, \eta_2) = \int \int \pi(\theta_1, \theta_2) p_1(\theta_1, \eta_1) p_2(\theta_2, \eta_2) d\theta_1 d\theta_2. \quad (3.30)$$

Το δεξί μέλος της σχέσης (3.30) γράφεται ως εξής:

$$\begin{aligned} & \int \int \pi_1(\theta_1|\theta_2) \pi^{(2)}(\theta_2) p_1(\theta_1, \eta_1) p_2(\theta_2, \eta_2) d\theta_1 d\theta_2 \\ &= \int \pi^{(2)}(\theta_2) \left[ \int \pi_1(\theta_1|\theta_2) p_1(\theta_1, \eta_1) d\theta_1 \right] p_2(\theta_2, \eta_2) d\theta_2 \\ &= \int \pi^{(2)}(\theta_2) \pi_1(\eta_1|\theta_2) p_2(\theta_2, \eta_2) d\theta_2 \\ &= \int \pi^{(1)}(\eta_1) \pi_2(\theta_2|\eta_1) p_2(\theta_2, \eta_2) d\theta_2 \\ &= \pi^{(1)}(\eta_1) \int \pi_2(\theta_2|\eta_1) p_2(\theta_2, \eta_2) d\theta_2 \\ &= \pi^{(1)}(\eta_1) \pi_2(\eta_2|\eta_1) \\ &= \pi(\eta_1, \eta_2). \end{aligned}$$

Η δεύτερη και τρίτη ισότητα προκύπτουν από τη στασιμότητα των δεσμευμένων πυκνοτήτων, ενώ οι υπόλοιπες ισότητες από βασικούς πιθανοτικούς και ολοκληρωτικούς λογισμούς. Το αποτέλεσμα μπορεί να επεκταθεί σε μεγαλύτερο αριθμό συντεταγμένων και ευσταθεί και στην περίπτωση ενός Metropolis-Hastings πυρήνα μείξης, ίσως με κάποιες τεχνικές αλλαγές. Η παραπάνω απόδειξη βασίστηκε σε απολύτως συνεχείς πυρήνες μετάβασης, για σημειογραφική ευκολία.

Μπορεί να αποδειχτεί ότι παρά τη μη αναγωγιμότητα των πυρήνων μετάβασης  $p_i$  των συντεταγμένων, ο κυκλικός πυρήνας μετάβασης  $p$  είναι ανάγωγος και απεριοδικός (Tierney, 1994). Επομένως, η στάσιμη κατανομή  $\pi$  είναι μοναδική. Το ότι ο Αλγόριθμος 3.7.1 πράγματι παράγει δείγματα από την κατανομή στόχο  $\pi$ , προκύπτει από το γεγονός ότι η  $\pi$  είναι μοναδικά ορισμένη από το σύνολο των πλήρεις δεσμευμένων κατανομών της (full conditional distributions) (Besag, 1974).

# Κεφάλαιο 4

## Ο Δειγματολήπτης Gibbs

### 4.1 Εισαγωγή

Το κεφάλαιο αυτό αφιερώνεται στην παρουσίαση του πιο ευρέως χρησιμοποιούμενου σχεδίου στοχαστικής προσομοίωσης μέσω αλυσίδων Markov. Είναι γνωστό με την ονομασία **δειγματοληψία Gibbs (Gibbs sampling)**, λόγω της αρχικής χρησιμοποίησής του στην επεξεργασία εικόνας (image processing), όπου έπρεπε να δειγματοληπτήσουμε από μια κατανομή Gibbs.

Όπως αναφέρεται από τους Besag και Green (1993), ο δειγματολήπτης Gibbs απαντά στις ιδέες του Grenander (1983), ενώ η επίσημη ονομασία του εισάγεται από τους Geman και Geman (1984). Το βιβλιογραφικό ορόσημο της δειγματοληψίας Gibbs, σε προβλήματα Μπεϋζιανής συμπερασματολογίας, δίνεται από τους Gelfand και Smith (1990). Μια παρόμοια ιδέα, με την ονομασία **αύξηση δεδομένων (data augmentation)** εισήχθη από τους Tanner και Wong (1987). Οι Casella και George (1992) παρέχουν ένα πολύ επεξηγηματικό άρθρο για τον δειγματολήπτη Gibbs.

### 4.2 Ορισμός και Ιδιότητες

Ο πιο πρόδηλος τρόπος ώστε να απλοποιήσουμε ένα δύσκολο πολυδιάστατο πρόβλημα προσομοίωσης, είναι να το εκφυλίσουμε σε μια συλλογή από πιο εύκολες, μικρότερων διαστάσεων, προσομοιώσεις. Αυτό ακριβώς κάνει ο δειγματολήπτης Gibbs. Ας υποθέσουμε ότι η κατανομή στόχος είναι η  $\pi(\theta|y)$ , όπου  $\theta = (\theta_1, \dots, \theta_d)^T$ , με εκ των προτέρων κατανομή  $p(\theta)$  και δεδομένα  $y$ , με συνάρτηση πιθανοφάνειας  $L(\theta|y)$ . Σε πολλές περιπτώσεις είναι δύσκολο

να προσομοιώσουμε ένα δείγμα απευθείας από την εκ των υστέρων κατανομή  $\pi(\theta|y) = cp(\theta)L(\theta|y)$ .

Για την ανάπτυξη του δειγματολήπτη Gibbs χρησιμοποιούμε την  $i$  συντεταγμένη  $\theta_i$  του διάνυσματος  $\theta$  και το διάνυσμα  $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)^T$ . Κάθε μία από τις συντεταγμένες  $\theta_i, i = 1, \dots, d$  μπορεί να είναι βαθμωτός αριθμός, διάνυσμα ή πίνακας. Θεωρούμε επίσης ότι οι πλήρως δεσμευμένες κατανομές  $\pi(\theta_i|\theta_{-i})$  είναι διαθέσιμες, με την έννοια ότι είναι γνωστές και μπορούν άμεσα να προσομοιωθούν.

Το πρόβλημα που καλούμαστε να αντιμετωπίσουμε είναι να προσομοιώσουμε από την κατανομή στόχο (εκ των υστέρων κατανομή)  $\pi(\theta)$ , όταν η άμεση δειγματοληψία από αυτή είναι δαπανηρή, περίπλοκη ή απλά μη διαθέσιμη, ενώ η προσομοίωση από τις πυκνότητες  $\pi(\theta_i|\theta_{-i}), i = 1, \dots, d$  είναι διαθέσιμη. Ευτυχώς, στα πλαίσια της Μπεϋζιανής Στατιστικής, για μια μεγάλη κλάση ιεραρχικών μοντέλων (hierarchical models) με συζητείς εκ των προτέρων και υπέρ εκ των προτέρων κατανομές, οι πλήρως δεσμευμένες κατανομές, που απαιτούνται για την ανάπτυξη του δειγματολήπτη Gibbs, είναι διαθέσιμες σε κλειστή μορφή (closed form). Όπως αναφέρθηκε νωρίτερα, ο δειγματολήπτης Gibbs είναι κατασκευασμένος για την εξυπηρέτηση αυτής της διαδικασίας.

**Αλγόριθμος 4.2.1** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})^T$  για το διάνυσμα  $\theta$ .

2. Προσομοιώσε μια νέα τιμή  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})^T$ , μέσω διαδοχικών προσομοιώσεων

$$\begin{aligned}\theta_1^{(j)} &\sim \pi_1(\theta_1|\theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ \theta_2^{(j)} &\sim \pi_2(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ &\vdots \\ \theta_d^{(j)} &\sim \pi_d(\theta_d|\theta_1^{(j)}, \dots, \theta_{d-1}^{(j)})\end{aligned}$$

3. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανάλαβε τη διαδικασία από το βήμα 2 μέχρι να επέλθει σύγκλιση.

Για να πειστούμε ότι σε κάθε επανάληψη του αλγόριθμου λαμβάνουμε μία τιμή από τη στάσιμη κατανομή στόχο μπορούμε να σκεφτούμε ως εξής: κατά την ολοκλήρωση του βήματος  $j - 1$  έχουμε την τιμή  $(\theta_1^{(j-1)}, \dots, \theta_d^{(j-1)})$

η οποία προέρχεται από την πυκνότητα  $\pi(\theta_1, \dots, \theta_d)$ . Αυτή είναι η στάσιμη πυκνότητα στόχος. Στο πρώτο βήμα της  $j$  επανάληψης παίρνουμε το διάνυσμα των τιμών  $(\theta_1^{(j)}, \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)})$  που κατανέμεται σύμφωνα με την  $\pi(\theta_1^{(j)} | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) \pi(\theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) = \pi(\theta_1^{(j)}, \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)})$  που είναι και αυτή η στάσιμη κατανομή στόχος. Συνεχίζουμε αυτή τη λογική μέχρι που φτάνουμε στο  $d$  βήμα της  $j$  επανάληψης του αλγόριθμου όπου έχουμε το διάνυσμα  $(\theta_1^{(j)}, \dots, \theta_d^{(j)})$  που κατανέμεται σύμφωνα με την  $\pi(\theta_1^{(j)}, \dots, \theta_d^{(j)})$  που είναι η στάσιμη κατανομή στόχος.

Κάθε συντεταγμένη του διανύσματος  $\theta$  αναβαθμίζεται με κάποια φυσική σειρά και κάθε επανάληψη του αλγόριθμου απαιτεί την προσομοίωση  $d$  συντεταγμένων. Οι Gelfand και Smith (1990) δείχνουν ότι κάτω από συγκεκριμένες συνθήκες, όταν επέλθει η σύγκλιση, τότε η προκύπτουσα τιμή  $\theta^{(j)}$  κατανέμεται σύμφωνα με την κατανομή  $\pi$ . Καθώς ο αριθμός των επαναλήψεων του αλγόριθμου αυξάνεται, η αλυσίδα προσεγγίζει τις συνθήκες ισορροπίας της. Η σύγκλιση τότε θεωρείται ότι ισχύει κατά προσέγγιση:

**Θεώρημα 4.2.1** Για το δειγματολήπτη Gibbs 4.2.1:

1.  $\theta^{(j)} \xrightarrow{d} \theta \sim \pi(\theta)$ .
2. Η σύγκλιση στο 1 είναι εκθετική ως προς  $j$ , ως προς τη νόρμα  $\mathcal{L}_1$ .

Οι Schervish και Carlin (1992) παρέχουν μια επαρκή συνθήκη που εξασφαλίζει γεωμετρική εργοδικότητα.

Θα διαφωτίσουμε τον τρόπο λειτουργίας του δειγματολήπτη Gibbs με ένα απλό παράδειγμα.

**Παράδειγμα 4.2.1** Θεωρήστε μια παρατήρηση  $y = (y_1, y_2)$  από ένα διμεταβλητό κανονικό πληθυσμό με άγνωστη μέση τιμή  $\theta = (\theta_1, \theta_2)$  και γνωστό πίνακα συνδιακυμάνσεων  $\Sigma$ , όπου

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

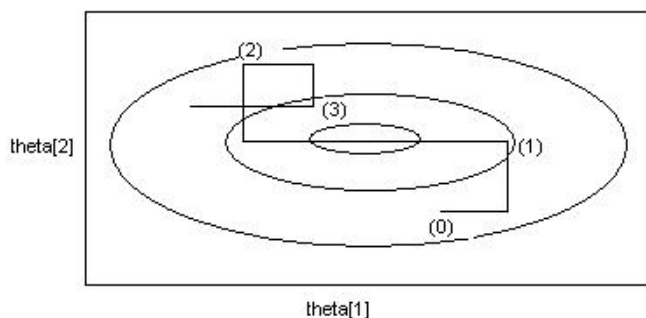
με  $-1 < \rho < 1$ . Με την  $\pi(\theta) = 1$  ως εκ των προτέρων κατανομή για το διάνυσμα  $\theta$ , η εκ των υστέρων κατανομή είναι η  $\theta|y \sim N(y, \Sigma)$ . Παρά το γεγονός ότι είναι πολύ εύκολο να λάβουμε απευθείας ένα δείγμα από την από κοινού εκ των υστέρων κατανομή των  $(\theta_1, \theta_2)$ , επιστρατεύουμε το δειγματολήπτη

Gibbs για τις ανάγκες της επίδειξης. Προκειμένου λοιπόν να εφαρμόσουμε το δειγματολήπτη Gibbs, χρειαζόμαστε τις πλήρως δεσμευμένες εκ των υστέρων κατανομές

$$\begin{aligned}\theta_1|\theta_2, y &\sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2) \\ \theta_2|\theta_1, y &\sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2).\end{aligned}$$

Ο δειγματολήπτης Gibbs εξελίσσεται με εναλλακτικές προσομοιώσεις από τις δύο παραπάνω κανονικές κατανομές, ξεκινώντας από μια αρχική τιμή  $\theta_2^{(0)}$ . Η ακολουθία  $(\theta_1, \theta_2)$ , είναι αλυσίδα Markov με στάσιμη κατανομή την  $\pi(\theta|y)$ . Αλλά και οι ακολουθίες  $\theta_1$  και  $\theta_2$  ξεχωριστά, είναι αλυσίδες Markov με στάσιμες κατανομές τις περιθωριακές (marginal distributions)  $\pi^{(1)}(\theta_1|y) = \int \pi(\theta_1, \theta_2|y) d\theta_2$  και  $\pi^{(2)}(\theta_2|y) = \int \pi(\theta_1, \theta_2|y) d\theta_1$ , αντίστοιχα. Ο πυρήνας μετάβασης για την αλυσίδα  $\theta_1$  είναι  $P(x, y) = \int \pi_2(x, w) \pi_1(w, y) dw$ , όπου  $\pi_2(x, w) = p_{\theta_2|\theta_1}(x, w)$  και  $\pi_1(w, y) = p_{\theta_1|\theta_2}(w, y)$ . Ανάλογα ορίζεται ο πυρήνας μετάβασης για την αλυσίδα  $\theta_2$ .

Μια τυπική τροχιά της δειγματοληψίας Gibbs, για ένα διδιάστατο παραμετρικό διάνυσμα  $\theta = (\theta_1, \theta_2)^T$ , δίνεται από το Σχήμα 4.1. Οι ομόκεντρες καμπύλες



Σχήμα 4.1: Υπόδειγμα της τροχιάς του δειγματολήπτη Gibbs σε διδιάστατο παραμετρικό χώρο.

παριστάνουν τις ισοϋψείς γραμμές (contour lines) της εκ των υστέρων πυκνότητας  $\pi(\theta|y)$ . Δοθείσης της αρχικής τιμής  $\theta_2^{(0)}$  παράγουμε μια τιμή για την παράμετρο  $\theta_1$ , την  $\theta_1^{(1)}$ . Αυτό συμβολίζεται με την πρώτη παράλληλη, στον οριζόντιο άξονα, ευθεία γραμμή. Με βάση αυτή την εκτίμηση, προσομοιώνουμε



μια νέα τιμή για την  $\theta_2$ , τη  $\theta_2^{(1)}$ , η οποία συμβολίζεται με την πρώτη παράλληλη, στον κάθετο άξονα, ευθεία γραμμή. Παρατηρήστε ότι κάθε μετακίνηση απεικονίζεται με μια ευθεία γραμμή, αφού στα ενδιάμεσα βήματα του δειγματολήπτη Gibbs αναβαθμίζεται μία μόνο συντεταγμένη, διατηρώντας όλες τις άλλες σταθερές. Στο σημείο αυτό ολοκληρώθηκε μια επανάληψη του αλγόριθμου, αφού αναβαθμίστηκαν και οι δύο συντεταγμένες του διανύσματος  $\theta$ . Η διαδικασία συνεχίζεται κατ' αυτόν τον τρόπο μέχρι να επέλθει η σύγκλιση.

Στην πρωτότυπη δημοσίευση των Geman και Geman (1984) μια επανάληψη του αλγόριθμου συνίσταται στην αναβάθμιση μιας μόνο συντεταγμένης του παραμετρικού διανύσματος  $\theta$ .

Ένας τρόπος ώστε να αποκτήσουμε ένα δείγμα μεγέθους  $n$  από την κατανομή  $\pi$  είναι να προσομοιώσουμε  $n$  αλυσίδες μέχρι να επέλθει σύγκλιση. Εναλλακτικά, μετά τη σύγκλιση όλες οι παραχθείσες τιμές προέρχονται από τη στάσιμη κατανομή στόχο  $\pi$ . Επομένως,  $n$  διαδοχικές τιμές από μία αλυσίδα, μετά τη burn in περίοδο, θεωρούνται ως δείγμα από την  $\pi$ . Το θέμα του σχηματισμού δείγματος θα μας απασχολήσει στο επόμενο κεφάλαιο.

Η δειγματοληψία Gibbs προσδιορίζει μια αλυσίδα Markov. Αυτό είναι ξεκάθαρο καθώς η κατάσταση μετά την  $j$  επανάληψη του αλγόριθμου εξαρτάται μόνο από τις τιμές της αλυσίδας στην επανάληψη  $j-1$ . Επίσης, η αλυσίδα είναι ομογενής, καθώς οι μεταβάσεις επηρεάζονται μόνο από τις τιμές της αλυσίδας και όχι από τον αύξοντα αριθμό της επανάληψης του αλγόριθμου.

Ο πυρήνας μετάβασης της αλυσίδας  $\theta = (\theta_1, \dots, \theta_d)$  είναι

$$p(\theta^{(t)}, \theta^{(t+1)}) = \prod_{i=1}^d \pi_i(\theta_i | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)}), \quad (4.1)$$

ο οποίος εξαρτάται από την επανάληψη του αλγόριθμου μόνο μέσω των τιμών  $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_d^{(t)})$  και  $\theta^{(t+1)} = (\theta_1^{(t+1)}, \dots, \theta_d^{(t+1)})$ . Αυτή η αλυσίδα, με συνολική σάρωση επί όλων των συντεταγμένων του διανύσματος  $\theta$ , δεν είναι αντιστρέψιμη (reversible), παρά το γεγονός ότι η αλυσίδα που σχηματίζεται από την αναβάθμιση κάθε συντεταγμένης ξεχωριστά, είναι αντιστρέψιμη. Ο Peter Green επισημαίνει ότι η αλυσίδα  $(\theta^{(t)})$  μπορεί να γίνει αντιστρέψιμη, αν κάθε επανάληψη του αλγόριθμου αποτελείται από μια σάρωση επί όλων των συνιστώσων του  $\theta$  ακολουθούμενη από μια σάρωση επί των ίδιων συντεταγμένων σε αντίστροφη σειρά (Besag, 1986):

**Αλγόριθμος 4.2.2** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$  για το διάνυσμα  $\theta$ .

2. Προσομοιώσε μια νέα τιμή  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})$ , μέσω διαδοχικών προσομοιώσεων

$$\begin{aligned} \theta_1^* &\sim \pi_1(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ \theta_2^* &\sim \pi_2(\theta_2 | \theta_1^*, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ &\vdots \\ \theta_{d-1}^* &\sim \pi_{d-1}(\theta_{d-1} | \theta_1^*, \dots, \theta_{d-2}^*, \theta_d^{(j-1)}) \\ \theta_d^{(j)} &\sim \pi_d(\theta_d | \theta_1^*, \dots, \theta_{d-1}^*) \\ \theta_{d-1}^{(j)} &\sim \pi_{d-1}(\theta_{d-1} | \theta_1^*, \dots, \theta_{d-2}^*, \theta_d^{(j)}) \\ &\vdots \\ \theta_1^{(j)} &\sim \pi_1(\theta_1 | \theta_2^{(j)}, \dots, \theta_d^{(j)}) \end{aligned}$$

3. Άλλαξε το μετρητή από  $j$  σε  $j+1$  και επανάλαβε τη διαδικασία από το βήμα 2 μέχρι να επέλθει σύγκλιση.

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε ένα απλούστερο, από τον 4.2.2, αλγόριθμο, ο οποίος προτάθηκε από τους Liu et al. (1995) και λέγεται **δειγματολήπτης Gibbs με τυχαία σάρωση (random sweep Gibbs sampler)**. Η προσομοίωση κάθε συντεταγμένης  $\theta_i$  του παραμετρικού διανύσματος  $\theta$ , γίνεται με τυχαία σειρά σε κάθε επανάληψη του αλγόριθμου. Η αλυσίδα που παράγεται είναι αντιστρέψιμη.

**Αλγόριθμος 4.2.3** 1. Προσομοιώσε μια τυχαία μετάθεση  $\sigma = (\sigma_1, \dots, \sigma_d)$  των στοιχείων του συνόλου  $\{1, \dots, d\}$ .

2. Προσομοιώσε

$$\begin{aligned} \theta_{\sigma_1}^{(j)} &\sim \pi_{\sigma_1}(\theta_{\sigma_1} | \theta_{-\sigma_1}^{(j-1)}) \\ &\vdots \\ \theta_{\sigma_d}^{(j)} &\sim \pi_{\sigma_d}(\theta_{\sigma_d} | \theta_{-\sigma_d}^{(j-1)}), \end{aligned}$$

όπου  $\theta_{-\sigma_i}^{(j-1)} = (\theta_{\sigma_1}^{(j-1)}, \dots, \theta_{\sigma_{i-1}}^{(j-1)}, \theta_{\sigma_{i+1}}^{(j-1)}, \dots, \theta_{\sigma_d}^{(j-1)})$ .

Ένα άλλο πολύ σημαντικό αποτέλεσμα είναι ότι η κατανομή ισορροπίας της αλυσίδας, με πυρήνα μετάβασης που δίνεται από τη σχέση (4.1), είναι η  $\pi$ . Με άλλα λόγια, αν μια αλυσίδα Markov με πυρήνα μετάβασης  $p(\theta^{(t)}, \theta^{(t+1)})$  έχει οριακή κατανομή την  $\pi^\infty$ , τότε η συνθήκη στασιμότητας (2.19) πρέπει να ικανοποιείται από το μέτρο  $\pi^\infty$  και τον πυρήνα μετάβασης  $p$  και η αλυσίδα πρέπει να είναι ανάγωγη (irreducible). Είναι εύκολο να διαπιστωθεί η αναγωγιμότητα σε κάποια παραχθείσα αλυσίδα, ελέγχοντας αν  $p(x, A) > 0$ , για όλα τα σύνολα  $A$  με θετική εκ των υστέρων πιθανότητα.

Μια παραλλαγή του Αλγόριθμου 4.2.1 είναι ο **τυχαίος δειγματολήπτης Gibbs (random Gibbs sampler)**, ο οποίος αναβαθμίζει μία συντεταγμένη  $\theta_\nu$ , του διανύσματος  $\theta$ , σε κάθε επανάληψη του.

**Αλγόριθμος 4.2.4** Δοθείσης της τιμής  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})$  και της πολωνυμικής κατανομής  $(\nu_1, \dots, \nu_d) \sim M(1; \xi_1, \dots, \xi_d)$ ,

1. αναβάθμισε τη συντεταγμένη  $\theta_\nu$ , όπου  $\nu \sim M(1; \xi_1, \dots, \xi_d)$ ,
2. προσομοίωσε  $\theta_\nu^{(j+1)} \sim \pi(\theta_\nu | \theta_{-\nu}^{(j)})$  και θέσε  $\theta_{-\nu}^{(j+1)} = \theta_{-\nu}^{(j)}$ .

Παρά το γεγονός ότι ο Αλγόριθμος 4.2.4 αναβαθμίζει μία συντεταγμένη κάθε φορά, η προκύπτουσα αλυσίδα  $(\theta^{(t)})$  είναι ισχυρά αναγώγιμη εξαιτίας της τυχαίας επιλογής της συντεταγμένης  $\nu$  που θα αναβαθμιστεί. Επίσης, απολαμβάνει μια επιπλέον ιδιότητα που θα συζητήσουμε στην Ενότητα 4.8.

### 4.3 Gibbs και Metropolis - Hastings

Τα  $d$  βήματα του Αλγόριθμου 4.2.1 μπορούν να θεωρηθούν ως μια συλλογή από  $d$  Metropolis - Hastings αλγόριθμους, στους οποίους η πιθανότητα αποδοχής των προτεινόμενων τιμών από τη συνάρτηση μετάβασης πρότασης είναι πάντα ίση με ένα. Ωστόσο, θα πρέπει να επισημάνουμε ότι κάθε ένας από τους  $d$  ξεχωριστούς Metropolis - Hastings αλγόριθμους δεν παράγει ανάγωγη αλυσίδα, αφού αναβαθμίζει μία μόνο συντεταγμένη του διανύσματος  $\theta$ .

Ο πυρήνας μετάβασης πρότασης για την  $i$  συντεταγμένη, στα πλαίσια ενός Metropolis - Hastings αλγόριθμου, είναι

$$q(\theta, \eta) = I(\theta_{-i} = \eta_{-i}) \pi_i(\eta_i | \theta_{-i}), \quad (4.2)$$

όπου  $I(\theta_{-i} = \eta_{-i})$  είναι η δείκτρια συνάρτηση του ενδεχομένου  $(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d) = (\eta_1, \dots, \eta_{i-1}, \eta_{i+1}, \dots, \eta_d)$ . Σημειώστε ότι η πυκνότητα στόχος, για τον  $i$  Metropolis - Hastings αλγόριθμο, είναι η  $\pi_i(\cdot|\theta_{-i})$ . Τότε, η πιθανότητα αποδοχής  $\alpha(\theta, \eta)$  είναι

$$\alpha(\theta, \eta) = \frac{\pi_i(\eta_i|\theta_{-i}) \pi_i(\theta_i|\eta_{-i})}{\pi_i(\theta_i|\theta_{-i}) \pi_i(\eta_i|\theta_{-i})} = \frac{\pi_i(\eta_i|\theta_{-i}) \pi_i(\theta_i|\theta_{-i})}{\pi_i(\theta_i|\theta_{-i}) \pi_i(\eta_i|\theta_{-i})} = 1. \quad (4.3)$$

Παρατηρήστε ότι χρησιμοποιήσαμε το γεγονός ότι  $I(\theta_{-i} = \eta_{-i}) = 1$ , οπότε  $\pi_i(\theta_i|\eta_{-i}) = \pi_i(\theta_i|\theta_{-i})$ . Αν  $I(\theta_{-i} = \eta_{-i}) = 0$ , τότε  $q(\theta, \eta) = 0$  και  $\alpha(\theta, \eta) = 0$ .

Η σύγκριση ανάμεσα στο δειγματολήπτη Gibbs και ένα τυπικό Metropolis - Hastings αλγόριθμο, φαίνεται να ευνοεί τον πρώτο. Στη δειγματοληψία Gibbs, οι προσομοιωμένες τιμές προέρχονται από τις πλήρως δεσμευμένες κατανομές που προκύπτουν απευθείας από την από κοινού πυκνότητα κατανομής των  $d$  συντεταγμένων του διανύσματος  $\theta$ , η οποία είναι και η κατανομή στόχος. Από την άλλη μεριά, ο πυρήνας Metropolis - Hastings βασίζεται σε μια πυκνότητα μετάβασης πρότασης, η οποία επιλέγεται έτσι ώστε να προσεγγίζει, όσο περισσότερο γίνεται, την κατανομή στόχο  $\pi$ . Από αυτή την άποψη, ο κίνδυνος μιας κακής επιλογής της πυκνότητας μετάβασης πρότασης, η οποία θα καταλήξει σε άσκοπες προσομοιώσεις τιμών που θα απορριφθούν, είναι υπαρκτός για την περίπτωση του αλγόριθμου Metropolis - Hastings.

## 4.4 Συμπλήρωση

Ο δειγματολήπτης Gibbs μπορεί να γενικευτεί με την ιδέα της συμπλήρωσης των πυκνοτήτων.

**Ορισμός 4.4.1** Δοθείσης μιας πυκνότητας πιθανότητας  $\pi$ , μια πυκνότητα  $g$  που ικανοποιεί τη σχέση

$$\int_{\eta} g(\theta, \eta) d\eta = \pi(\theta), \quad (4.4)$$

λέγεται συμπλήρωση (*completion*) της  $\pi$ .

Αντί να προσομοιώσουμε τιμές  $\theta^{(j)}$  του διανύσματος  $\theta$  από την κατανομή  $\pi$ , προσομοιώνουμε τιμές  $y^{(j)}$  του διανύσματος  $y = (\theta, \eta)$  από την κατανομή  $g$ .

Το  $\eta$  είναι και αυτό διάνυσμα παραμέτρων και λειτουργεί ως «αύξηση» του διανύσματος  $\theta$ , με την έννοια της διεύρυνσης του αρχικού χώρου καταστάσεων  $\Theta$  στο χώρο  $\Theta \times \eta$ . Άλλωστε, η συμπλήρωση μιας πυκνότητας  $\pi$  σε μια πυκνότητα  $g$ , έτσι ώστε  $\eta$  να είναι η περιθωριακή πυκνότητα της  $g$ , αποτελεί μια από τις πρώτες εμφανίσεις του δειγματολήπτη Gibbs στη Στατιστική, γνωστή ως αύξηση δεδομένων (data augmentation) (Tanner και Wong, 1987).

Η μετακίνηση από την τιμή  $y^{(j-1)}$  στην τιμή  $y^{(j)}$  ορίζεται με τον ακόλουθο τρόπο:

**Αλγόριθμος 4.4.1** 1. Θέσε το μετρητή των επαναλήψεων  $j = 1$  και μια αυθαίρετη αρχική τιμή  $y^{(0)} = (y_1^{(0)}, \dots, y_d^{(0)})$  για το διάνυσμα  $y$ .

2. Προσομοίωσε μια νέα τιμή  $y^{(j)} = (y_1^{(j)}, \dots, y_d^{(j)})$ , μέσω διαδοχικών προσομοιώσεων

$$\begin{aligned} y_1^{(j)} &\sim g_1(y_1 | y_2^{(j-1)}, \dots, y_d^{(j-1)}) \\ y_2^{(j)} &\sim g_2(y_2 | y_1^{(j)}, y_3^{(j-1)}, \dots, y_d^{(j-1)}) \\ &\vdots \\ y_d^{(j)} &\sim g_d(y_d | y_1^{(j)}, \dots, y_{d-1}^{(j)}) \end{aligned}$$

3. Άλλαξε το μετρητή από  $j$  σε  $j + 1$  και επανέλαβε τη διαδικασία από το βήμα 2 μέχρι να επέλθει σύγκλιση.

Η μέθοδος της συμπλήρωσης προτείνεται όταν οι για τις δεσμευμένες κατανομές που σχετίζονται με την  $\pi$  δεν είναι γνωστές οι σταθερές κανονικοποίησης ή όταν η κατανομή στόχος  $\pi$  είναι μονοδιάστατη και όπως ήδη έγινε αντιληπτό, ο δειγματολήπτης Gibbs εφαρμόζεται για την προσομοίωση κατανομών διαστάσεων  $d \geq 2$ .

**Παράδειγμα 4.4.1** Θέλουμε να προσομοιώσουμε την κατανομή  $\pi(\theta)$  όπου το  $\theta$  είναι μια μονοδιάστατη παράμετρος. Λόγω τεχνικών δυσκολιών, σαν και αυτές που αναφέραμε πιο πάνω, αποφασίζουμε να προσομοιώσουμε τη συμπλήρωσή της. Για κάθε συνάρτηση  $\pi(\theta)$ , υπάρχει η συνάρτηση

$$g(\theta, \eta) = I(0 \leq \eta \leq \pi(\theta)) \quad (4.5)$$

που αποτελεί φυσική συμπλήρωση της  $\pi$ . Στο παρόν παράδειγμα η  $\eta$  είναι και αυτή μια μονοδιάστατη παράμετρος. Παρατηρήστε ότι

$$\begin{aligned} \int g(\theta, \eta) d\eta &= \int I(0 \leq \eta \leq \pi(\theta)) d\eta \\ &= \int_0^{\pi(\theta)} 1 d\eta = [\eta]_0^{\pi(\theta)} = \pi(\theta). \end{aligned}$$

Η εφαρμογή του Αλγόριθμου 4.4.1 έχει ως εξής:

1.  $\theta^{(j)} | \eta^{(j-1)} \sim g_1(\theta | \eta^{(j-1)})$ .
2.  $\eta^{(j)} | \theta^{(j)} \sim g_2(\eta | \theta^{(j)})$ .

Στη συνέχεια, κρατάμε μόνο τις προσομοιωμένες τιμές  $\theta^{(j)}$  για την παράμετρο  $\theta$  που μας ενδιαφέρει και αγνοούμε τις προσομοιωμένες τιμές  $\eta^{(j)}$  για τη βοηθητική παράμετρο  $\eta$ .

Η συμπλήρωση της πυκνότητας  $\pi$  στην  $g$  και του διανύσματος  $\theta$  στο  $y = (\theta, \eta)$  δεν είναι υποχρεωτικά συνδεδεμένη με το υπό μελέτη πρόβλημα. Το διάνυσμα  $\eta$  είναι πιθανό να μην έχει θεωρητικό νόημα, αλλά μόνο πρακτικό ως ένα χρήσιμο επιπρόσθετο.

## 4.5 Ιδιότητες Σύγκλισης

Θα δουλέψουμε με τη γενική περίπτωση του Αλγόριθμου 4.4.1 που συναντήσαμε στην ενότητα 4.4. Η αλυσίδα Markov  $y = (\theta, \eta)$  συγκλίνει στην κατανομή  $g(\theta, \eta)$  και η υποαλυσίδα  $(\theta^{(t)})$  συγκλίνει στην κατανομή  $\pi$ . Είναι σημαντικό να σημειώσουμε ότι ενώ η  $y$  είναι αλυσίδα Markov, η υποαλυσίδα  $(\theta^{(t)})$  τυπικά δεν είναι αλυσίδα Markov. Μια περίπτωση κατά την οποία συμβαίνει αυτό είναι εκείνη της αύξησης δεδομένων (data augmentation).

**Θεώρημα 4.5.1** Για το δειγματολήπτη Gibbs του Αλγόριθμου 4.4.1, αν η αλυσίδα  $y$  είναι εργοδική, τότε η κατανομή  $g$  είναι η στάσιμη κατανομή για την αλυσίδα  $y$  και η κατανομή  $\pi$  είναι η οριακή κατανομή για την υποαλυσίδα  $(\theta^{(t)})$ .

Ο Besag (1974) παρέχει μια επαρκή συνθήκη ώστε η αλυσίδα Markov που παράγεται από το δειγματολήπτη Gibbs να είναι ανάγωγη.

**Ορισμός 4.5.1** Έστω  $(y_1, \dots, y_d) \sim g(y_1, \dots, y_d)$  και  $g^{(i)}$  είναι η περιθωριακή κατανομή της συντεταγμένης  $y_i$ . Αν οι σχέσεις  $g^{(i)}(y_i) > 0, i = 1, \dots, d$  συνεπάγονται ότι  $g(y_1, \dots, y_d) > 0$ , τότε η συνάρτηση  $g$  ικανοποιεί τη συνθήκη θετικότητας.

Απόρροια του παραπάνω ορισμού είναι, ότι το στήριγμα της συνάρτησης  $g$  είναι το καρτεσιανό γινόμενο των θεμέλιων των πυκνοτήτων  $g^{(i)}$ . Επιπλέον, το εύρος των δυνατών τιμών  $y_i$  που προσομοιώνουμε από τις πλήρως δεσμευμένες κατανομές δεν είναι μειωμένο σε σχέση με το εύρος των τιμών  $y_i$ , που προσομοιώνουμε απευθείας από την απο κοινού κατανομή  $g$ . Στην περίπτωση αυτή, δύο αυθαίρετα υποσύνολα Borel του στηρίγματος της  $g$ , είναι δυνατό να επικοινωνούν σε μία επανάληψη του αλγόριθμου.

**Θεώρημα 4.5.2** Αν η πυκνότητα  $g$  ικανοποιεί τη συνθήκη θετικότητας, τότε ο δειγματολήπτης Gibbs του Αλγόριθμου 4.4.1 ορίζει μια ανάγωση αλυσίδα Markov.

Δυστυχώς, η συνθήκη του Θεωρήματος 4.5.2 είναι συχνά δύσκολο να επιβεβαιωθεί και για το λόγο αυτό ο Tierney (1994) δίνει μια πιο βολική συνθήκη:

**Λήμμα 4.5.1** Αν ο πυρήνας μετάβασης της αλυσίδας που παράγεται από τον Αλγόριθμο 4.4.1 είναι απολύτως συνεχής, ως προς το κυρίαρχο μέτρο (dominating measure), τότε η αλυσίδα είναι Harris επαναληπτική (Harris recurrent).

Εδώ, ο περιορισμός που θέτουμε στον πυρήνα μετάβασης του δειγματολήπτη Gibbs<sup>1</sup>, αρκεί ώστε να παραχθεί μια ανάγωση αλυσίδα και η Harris επαναληπτικότητα έπεται. Αφού εξασφαλισθεί η αναγωγιμότητα της αλυσίδας που παράγεται από τον Αλγόριθμο 4.4.1, μπορούν να εδραιωθούν περισσότερο ισχυρές ιδιότητες σύγκλισης.

Ο περιορισμός της απόλυτης συνέχειας που επιβάλαμε στον πυρήνα μετάβασης, που δίνεται από την (4.1), ικανοποιείται στις περισσότερες περιπτώσεις. Ωστόσο, αν κάποιο από τα  $d$  βήματα, που απαιτεί μια επανάληψη του Αλγόριθμου 4.4.1, αντικατασταθεί από κάποιο Metropolis - Hastings υποαλγόριθμο, τότε η απόλυτη συνέχεια χάνεται και απαιτείται είτε η μελέτη των ιδιοτήτων επανόδου (recursion properties) της αλυσίδας ή η εισαγωγή ενός επιπλέον βήματος προσομοίωσης που θα εξασφαλίσει την Harris επαναληπτικότητα .

Από την Harris επαναληπτικότητα , μπορεί να εδραιωθεί το επόμενο αποτέλεσμα:

<sup>1</sup>Είναι αυτός που δίνεται από τη σχέση (4.1).

**Θεώρημα 4.5.3** Έστω ότι ο πυρήνας μετάβασης της αλυσίδας  $y = (\theta, \eta)$ , που παράγεται από το δειγματολήπτη Gibbs 4.4.1, είναι απολύτως συνεχής, ως προς το κυρίαρχο μέτρο (*dominating measure*). Τότε

1. Αν  $h_1, h_2 \in L^1(g)$  με  $\int h_2(y) g(dy) \neq 0$ , τότε

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^n h_1(y^{(j)})}{\sum_{j=1}^n h_2(y^{(j)})} = \frac{\int h_1(y) g(dy)}{\int h_2(y) g(dy)}. \quad (4.6)$$

2. Αν επιπλέον η αλυσίδα  $y$  είναι απεριοδική, τότε

$$\lim_{n \rightarrow \infty} \left\| \int p^n(y, \cdot) \mu(dx) - g \right\| = 0, \quad (4.7)$$

για κάθε αρχική κατανομή  $\mu$ .

## 4.6 Υβριδικοί Αλγόριθμοι

### 4.6.1 Βοηθητικές Μεταβλητές

Στο πνεύμα της Ενότητας 4.4, μπορούμε να κάνουμε μερικές επεκτάσεις ώστε να κατασκευάσουμε μια υβριδική μορφή του δειγματολήπτη Gibbs. Στην προσέγγιση αυτή, προκειμένου να αποκτήσουμε ένα δείγμα από τη στάσιμη κατανομή  $\pi(\theta)$ , με  $\theta \in \Theta$ , διευρύνουμε τον αρχικό χώρο καταστάσεων  $\Theta$  χρησιμοποιώντας μία ή περισσότερες βοηθητικές ή λανθάνουσες μεταβλητές (*auxiliary or latent variables*). Μια έξυπνη διεύρυνση του χώρου καταστάσεων  $\Theta$  μας δίνει τη δυνατότητα να διευρύνουμε την κλάση των πλήρεις δεσμευμένων κατανομών (*full conditional distribution*) από τις οποίες μπορεί να προσομοιώνει ο δειγματολήπτης Gibbs. Κατ' αυτό τον τρόπο, επιταχύνεται η σύγκλιση στην κατανομή στόχο της αλυσίδας που ορίζεται στο διευρυμένο χώρο καταστάσεων.

Η βασική ιδέα έχει ως εξής: έστω  $\pi(\theta)$  η κατανομή στόχος. Μια νέα παράμετρος  $\eta$  εισάγεται με δεσμευμένη κατανομή την  $p(\eta|\theta)$ , έτσι ώστε η από κοινού κατανομή των παραμέτρων (ή παραμετρικών διανυσμάτων)  $\theta$  και  $\eta$  να είναι η  $p(\theta, \eta) = \pi(\theta)p(\eta|\theta)$ . Μπορούμε να κατασκευάσουμε μια αλυσίδα Markov στο χώρο καταστάσεων  $\Theta \times \eta$  με εναλλακτικές προσομοιώσεις μεταξύ  $\theta$  και  $\eta$ , μέσω του δειγματολήπτη Gibbs. Η δεσμευμένη κατανομή του  $\eta|\theta$  επιλέγεται με «κατάλληλο» τρόπο, που σημαίνει είτε ότι η βοηθητική μεταβλητή εισάγεται για να επιτρέψει καλύτερη ή γρηγορότερη σύγκλιση της αλυσίδας



$(\theta, \eta)$  στην από κοινού κατανομή  $p(\theta, \eta)$ , είτε ότι ο νέος αλγόριθμος είναι πιο εύκολα υλοποιήσιμος ή πιο γρήγορος σε υπολογιστικό χρόνο.

Ας υποθέσουμε ότι για τη δεσμευμένη κατανομή του  $\eta|\theta$  επιλέγουμε  $\eta|\theta \sim U(0, \pi(\theta))$ . Αυτό οδηγεί σε μια ομοιόμορφη από κοινού κατανομή

$$p(\theta, \eta) \propto I(0 \leq \eta \leq \pi(\theta)), \quad (4.8)$$

της οποίας η περιθωριακή συνάρτηση πυκνότητας είναι η  $\pi(\theta)$ . Δηλαδή η  $p(\theta, \eta)$  είναι μια συμπλήρωση της  $\pi(\theta)$ . Ο δειγματολήπτης Gibbs, για την προσομοίωση της από κοινού κατανομής (4.8), απαιτεί δύο μόνο βήματα:

$$\begin{aligned} \eta|\theta &\sim U(0, \pi(\theta)) \\ \theta|\eta &\sim U(\theta : \pi(\theta) \geq \eta). \end{aligned}$$

Όταν αποκτήσουμε ένα δείγμα με τιμές  $(\theta^{(j)}, \eta^{(j)})$  για τις παραμέτρους  $(\theta, \eta)$ , αγνοούμε τις τιμές  $\eta^{(j)}$  και έτσι μας απομένει ένα δείγμα με τιμές  $(\theta^{(j)})$  που προέρχονται από την πυκνότητα στόχο  $\pi(\theta)$ , η οποία προκύπτει από την από κοινού πυκνότητα (4.8) με ολοκλήρωση ως προς  $\eta$ .

Στα πλαίσια της Μπεϋζιανής Στατιστικής, έχουμε  $\pi(\theta) \propto p(\theta) L(\theta)$ , όπου  $\pi(\theta)$  είναι η εκ των υστέρων κατανομή στόχος,  $p(\theta)$  είναι η εκ των προτέρων κατανομή για την παράμετρο  $\theta$  και  $L(\theta)$  είναι η συνάρτηση πιθανοφάνειας<sup>2</sup>. Τότε, εισάγοντας μια βοηθητική παράμετρο  $\eta$ , καταλήγουμε στην κατανομή στόχο  $p(\theta, \eta) = \pi(\theta) p(\eta|\theta) \propto p(\theta) L(\theta) p(\eta|\theta)$ . Μια κατάλληλη επιλογή για τη δεσμευμένη κατανομή της  $\eta$  ως προς  $\theta$ ,  $p(\eta|\theta)$ , είναι η  $U(0, L(\theta))$ . Η προσομοίωση της άγνωστης εκ των υστέρων κατανομής, με τη βοήθεια του δειγματολήπτη Gibbs και με την επιλογή της συγκεκριμένη δεσμευμένης κατανομής που αναφέραμε, λέγεται **slice sampling**.

Τέλος, υποθέστε ότι  $\pi(\theta) \propto p(\theta) \prod_{i=1}^k L_i(\theta)$ . Η συνάρτηση  $L_i(\theta)$  είναι μια θετική συνάρτηση που παριστάνει τη συμβολή του  $i$  δεδομένου ( $1 \leq i \leq k$ ) στη συνάρτηση πιθανοφάνειας  $L(\theta)$ . Μπορούμε να χρησιμοποιήσουμε μια συλλογή από βοηθητικές μεταβλητές  $\eta = (\eta_1, \dots, \eta_k)$ , όπου θέτουμε  $\eta_i|\theta \sim U(0, L_i(\theta))$ ,  $i = 1, \dots, k$ . Τώρα η από κοινού συνάρτηση πυκνότητας δίνεται από τον τύπο

$$p(\theta, \eta_1, \dots, \eta_k) \propto p(\theta) \prod_{i=1}^k I(0 \leq \eta_i \leq L_i(\theta)). \quad (4.9)$$

<sup>2</sup>Για λόγους σημειογραφικής ευκολίας συμβολίζουμε την εκ των υστέρων κατανομή του  $\theta$  με  $\pi(\theta)$  αντί για  $\pi(\theta|y)$  και τη συνάρτηση πιθανοφάνειας των δεδομένων με  $L(\theta)$  αντί για  $L(\theta|y)$ .

Ο δειγματολήπτης Gibbs εφαρμόζεται στο σχέδιο slice sampling με τον ακόλουθο τρόπο:

#### Αλγόριθμος 4.6.1

$$\begin{aligned}\eta_1^{(j+1)} &\sim U(0, L_1(\theta^{(j)})) \\ \eta_2^{(j+1)} &\sim U(0, L_2(\theta^{(j)})) \\ &\vdots \\ \eta_k^{(j+1)} &\sim U(0, L_k(\theta^{(j)})) \\ \theta^{(j+1)} &\sim U_{A^{(j+1)}},\end{aligned}$$

$$\mu \in A^{(j+1)} = \left\{ \theta : L_i(\theta) \geq \eta_i^{(j+1)}, i = 1, \dots, k \right\}.$$

Η περιθωριακή συνάρτηρηση πυκνότητας για την παράμετρο  $\theta$ , είναι η  $\pi(\theta)$ , όπως άλλωστε είναι το επιθυμητό. Οι πλήρως δεσμευμένες κατανομές για τις βοηθητικές παραμέτρους  $\eta_i$  είναι  $U(0, L_i(\theta))$ , ενώ η πλήρως δεσμευμένη κατανομή για τη  $\theta$  είναι η εκ των προτέρων κατανομή  $p(\theta)$  περιορισμένη στο σύνολο  $A = \{\theta : L_i(\theta) \geq \eta_i, i = 1, \dots, k\}$ .

Συχνά, στον τύπο (4.9) θέτουμε  $p(\theta) = c$ , όπου  $c$  είναι μια σταθερά. Στην περίπτωση αυτή η από κοινού συνάρτηση κατανομής  $p(\theta, \eta_1, \dots, \eta_k)$  έχει ομοιόμορφη πυκνότητα στο σύνολο  $\cap_{i=1}^k \{\theta : L_i(\theta) \geq \eta_i^{(j)}\}$ . Εδώ, η προσομοίωση μπορεί να είναι δύσκολη, αφού εμπλέκει την προσομοίωση τιμών από την ομοιόμορφη κατανομή σε μια, πιθανώς, πολυδιάστατη μη ομαλά σχηματισμένη περιοχή. Αυτό περιορίζει την εφαρμοσιμότητα του slice sampler. Οι Damien et al. (1999) και Neal (2002) παρέχουν ένα πλήθος τεχνικών για την πραγμάτωση αυτού του δύσκολου βήματος του αλγόριθμου.

Οι Damien et al. (1999) δείχνουν ότι αυτή η τελευταία παραγοντική μορφή του slice sampler, είναι πλεονεκτική σε πολλά ιεραρχικά και μη συζιγή Μπεϋζιανά μοντέλα, όπου οι συναρτήσεις  $L_i$  είναι αμέσως αντιστρέψιμες και το σύνολο  $A = \{\theta : L_i(\theta) \geq \eta_i, i = 1, \dots, k\}$  είναι εύκολο να υπολογιστεί. Ωστόσο, ο Neal (1997) επισημαίνει ότι εισάγοντας τόσες πολλές βοηθητικές μεταβλητές (μία για κάθε παράγοντα στη συνάρτηση πιθανοφάνειας), μπορεί να επιφέρει βραδείς ρυθμούς στη σύγκλιση.

Θα κλείσουμε την ενότητα των βοηθητικών μεταβλητών με ένα παράδειγμα για τον slice sampler.

**Παράδειγμα 4.6.1** Έστω ότι θέλουμε να προσομοιώσουμε τιμές από την τυπική κανονική κατανομή με πυκνότητα  $\pi(\theta) \propto \exp(-\theta^2/2)$ , όπου  $\theta$  είναι μια

μονοδιάστατη παράμετρος. Για να αναπτύξουμε τον slice sampler, θα χρησιμοποιήσουμε τη συμπλήρωση της  $\pi$ , δηλαδή την

$$p(\theta, \eta) = I(0 \leq \eta \leq \pi(\theta)).$$

Ο slice sampler προσομοιώνει από τις πλήρως δεσμευμένες κατανομές

$$\begin{aligned}\eta|\theta &\sim U(0, \pi(\theta)) \\ \theta|\eta &\sim U_A,\end{aligned}$$

όπου

$$\begin{aligned}A &= \{\theta : \pi(\theta) \geq \eta\} = \left\{\theta : e^{-\frac{\theta^2}{2}} \geq \eta\right\} \\ &= \left\{\theta : -\frac{\theta^2}{2} \geq \ln \eta\right\} = \{\theta : \theta^2 \leq -2 \ln \eta\} \\ &= \{\theta : \theta^2 + 2 \ln \eta \leq 0\} = \left\{\theta \in \left(-\sqrt{-2 \ln \eta}, \sqrt{-2 \ln \eta}\right)\right\}.\end{aligned}$$

#### 4.6.2 Metropolis εντός του Gibbs

Ο αλγόριθμος Metropolis εντός του Gibbs (Metropolis within Gibbs) είναι μια υβριδική μορφή του κλασσικού δειγματολήπτη Gibbs, η οποία πολλές φορές μας βγάζει από τη δύσκολη θέση.

Έστω το διάνυσμα των παραμέτρων  $\theta = (\theta_1, \dots, \theta_d)^T$ . Υποθέτουμε ότι όλες οι πλήρως δεσμευμένες πυκνότητες είναι διαθέσιμες, εκτός από την  $\pi_i(\theta_i|\theta_{-i})$ , όπου  $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ , πιθανώς λόγω μη συζυγούς (non conjugacy) εκ των προτέρων κατανομής για την παράμετρο  $\theta_i$ . Είναι λογικό να σκεφτούμε κάποιο τρόπο ώστε να αντικαταστήσουμε τη δύσκολη δειγματοληψία από την περιθωριακή κατανομή  $\pi_i(\theta_i|\theta_{-i})$  με κάποια πιο εύκολη. Ένας εύκολος τρόπος δειγματοληψίας, όπως αναφέραμε στο προηγούμενο κεφάλαιο, είναι ο αλγόριθμος Metropolis - Hastings. Η ιδέα λοιπόν είναι να χρησιμοποιήσουμε το βασικό κορμό του δειγματολήπτη Gibbs και να εισάγουμε στο  $i$  βήμα του έναν βήμα Metropolis - Hastings ώστε να προσομοιώσουμε την παράμετρο  $\theta_i$ . Για τον υποαλγόριθμο αυτό, μπορούμε να επιλέξουμε μια δεσμευμένη πυκνότητα μετάβασης πρότασης (conditional proposal density function)

$$q\left(\theta_i^{(j-1)}, \eta|\theta_{-i}^{(j-1)}\right), \quad (4.10)$$

όπου  $\theta_{-i}^{(j-1)} = (\theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})$  και να αναπτύξουμε τον υποαλγόριθμο Metropolis - Hastings για  $T$  επαναλήψεις, αποδεχόμενοι ή απορρίπτοντας τιμές, κατά τη συνήθη διαδικασία. Η τελευταία τιμή που δίνει ο υποαλγόριθμος (επανάληψη  $T$ ), είναι αυτή που χρησιμοποιούμε ως την  $j$  αναβάθμιση της συντεταγμένης  $\theta_i$ , είναι δηλαδή η τιμή  $\theta_i^{(j)}$ . Κατόπιν, συνεχίζουμε με τον εξωτερικό βρόγχο (loop) του δειγματολήπτη Gibbs. Σημειώστε ότι αυτή η προσέγγιση μπορεί να χρησιμοποιηθεί για οποιαδήποτε παράμετρο για την οποία στερούμαστε την κλειστής μορφής πλήρη δεσμευμένη πυκνότητα.

Για  $T = 1$ , ο αλγόριθμος έχει την εξής μορφή:

**Αλγόριθμος 4.6.2** Για  $i = 1, \dots, d$ , δίνεται  $\theta_{-i}^{(j-1)} = (\theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})$ :

1. Προσομοίωσε ένα νέο αριθμό  $\eta_i \sim q(\theta_i^{(j-1)}, \eta_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \dots, \theta_d^{(j-1)})$

2. Θέσε

$$\theta_i^{(j)} = \begin{cases} \eta_i, & \mu\epsilon \text{ πιθανότητα } \alpha, \\ \theta_i^{(j-1)}, & \mu\epsilon \text{ πιθανότητα } 1 - \alpha, \end{cases}$$

όπου

$$\alpha(\theta_i^{(j-1)}, \eta_i) = \min \left( 1, \frac{\frac{\pi_i(\eta_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})}{q_i(\theta_i^{(j-1)}, \eta_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})}}{\frac{\pi_i(\theta_i^{(j-1)} | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})}{q_i(\eta_i, \theta_i^{(j-1)} | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_d^{(j-1)})}} \right).$$

Αυτή η υβριδική μορφή του δειγματολήπτη Gibbs αναφέρεται συχνά ως Metropolis within Gibbs, καίτοι κάποιοι συγγραφείς, όπως οι Besag et al. (1995) και οι Chib και Greenberg (1995), διαφωνούν με αυτόν τον όρο ισχυριζόμενοι ότι οι Metropolis et al. (1953) και Hastings (1970) οραματίστηκαν τέτοιους αλγόριθμους στις πρωτότυπες δημοσιεύσεις τους πολύ πριν την εμφάνιση του δειγματολήπτη Gibbs. Αποτέλεσμα αυτής της ένστασης είναι η χρησιμοποίηση των όρων «μονομεταβλητός Αλγόριθμος Metropolis» (univariate Metropolis) και «βήματα Metropolis» (Metropolis steps).

Η χρήση του αλγόριθμου Metropolis within Gibbs αναδεικνύει δύο, μείζονως σημασίας, ζητήματα σύγκλισης. Το πρώτο είναι θεωρητικό: συγκλίνει

ο παραπάνω υβριδικός αλγόριθμος στη σωστή κατανομή στόχο; Οι μεταβάσεις της αλυσίδας ( $\theta^{(t)}$ ), που παράγει ο αλγόριθμος, δεν είναι κυκλικές<sup>3</sup> (cycle transitions), επειδή πρόκειται για μια ντετερμινιστική εναλλαγή μεταξύ Gibbs και Metropolis βημάτων, τα οποία αν αναπτυχθούν ξεχωριστά δεν συγκλίνουν, αφού τα πρώτα δεν αναβαθμίζουν τη συντεταγμένη  $\theta_i$ , ενώ τα τελευταία δεν αναβαθμίζουν τις υπόλοιπες συντεταγμένες. Ωστόσο, κάθε μία από τις συντεταγμένες του διανύσματος  $\theta$  διατηρεί τη στάσιμη κατανομή της αλυσίδας και επομένως, δεδομένου ότι η αλυσίδα του υβριδικού αλγόριθμου είναι απεριοδική και ανάγωγη, η σύγκλιση επιτυγχάνεται.

Το δεύτερο ζήτημα είναι πιο πρακτικό: πώς θα επιλέξουμε τον αριθμό  $T$  των επαναλήψεων του Metropolis υποαλγόριθμου; Η σύγκλιση στον εξωτερικό Gibbs βρόγχο επιτυγχάνεται για οποιαδήποτε τιμή του  $T$ . Βεβαίως, αυτό δεν σημαίνει ότι σε κάποιο συγκεκριμένο πρόβλημα όλες οι τιμές του  $T$  είναι το ίδιο καλές ως προς τη σύγκλιση. Για παράδειγμα, η τιμή  $T = 100000$  είναι μια κακή επιλογή, αφού το μόνο που αποκομίζουμε είναι η ισχυρή πεποίθηση ότι η προσομοιωμένη τιμή  $\theta_i^{(j)}$  είναι μια δειγματική πραγματοποίηση από τη σωστή πλήρη δεσμευμένη κατανομή. Ακόμα όμως και αυτή η πεποίθηση είναι άχρηστη στα αρχικά στάδια του εξωτερικού βρόγχου (δηλαδή όταν το  $j$  είναι μικρό), διότι ο τελευταίος είναι ακόμα μακριά από τη στάσιμη κατανομή στην οποία στοχεύει και επομένως το «ακριβό» δείγμα (με την έννοια του υπολογιστικού χρόνου) είναι πολύ πιθανό να μη διατηρηθεί για την συμπερασματολογία γύρω από την εκ των υστέρων κατανομή. Δηλαδή, μια πιο ακριβής προσέγγιση της πλήρεις περιθωριακής κατανομής  $\pi_i(\theta_i|\theta_{-i})$  δεν σημαίνει απαραίτητα καλύτερη προσέγγιση της κατανομής στόχου  $\pi(\theta)$ , όπου  $\theta = (\theta_1, \dots, \theta_d)$ .

Από την άλλη μεριά, η τιμή  $T = 1$  μπορεί να αντιστοιχεί σε μια πολύ κακή προσέγγιση της πλήρη δεσμευμένης κατανομής, ώστε να επιβραδύνεται η σύγκλιση του εξωτερικού Gibbs αλγόριθμου. Σε τριμεταβλητό κανονικό περιβάλλον, που μελετήθηκε από τους Gelfand και Lee (1993), οι τιμές  $T = 4$  και  $T = 8$  επέδειξαν καλή απόδοση, αλλά αυτό δεν αποτελεί γενική υπόδειξη. Στην πράξη η επιλογή  $T = 1$  υιοθετείται συχνά.

Η ευκαμψία του αλγόριθμου Metropolis within Gibbs είναι το ορόσημο της σπουδαιότητάς του. Από τη μία απολαμβάνουμε το πλεονέκτημα του δειγματολήπτη Gibbs να εκφυλίζει μια πολυδιάστατη προσομοίωση σε πολλές, μικρότερων διαστάσεων, προσομοιώσεις. Από την άλλη, η εισαγωγή του υποαλγόριθμου Metropolis - Hastings μας επιτρέπει να υπερβαίνουμε το εμπόδιο της δύσκολης δειγματοληψίας από μια στρυφνή πλήρεις περιθωριακή πυκνότητα

<sup>3</sup>Βλέπε και Υποενότητα 2.5.

$\pi_i(\theta_i|\theta_{-i})$ .

Ο Müller προτείνει μια μέθοδο για την περίπτωση όπου πολλοί Metropolis - Hastings υποαλγόριθμοι εμφανίζονται σε έναν εξωτερικό δειγματολήπτη Gibbs, είτε λόγω περίπλοκων δεσμευμένων κατανομών ή για λόγους επιτάχυνσης της σύγκλισης. Εισάγει ένα βήμα αποδοχής, ύστερα από τα  $d$  βήματα μιας επανάληψης του εξωτερικού Gibbs βρόγχου. Η μέθοδος αυτή είναι πολύ δαπανηρή υπολογιστικά, αφού μπορεί σε οποιαδήποτε βήμα  $j$  του εξωτερικού αλγόριθμου να απορρίπτει όλες τις προσομοιωμένες τιμές  $\theta_i^{(j)}$ ,  $i = 1, \dots, d$ , του διανύσματος  $\theta$ .

Κλείνουμε αυτήν την υποενότητα με την παρουσίαση ενός αποτελέσματος από τον Liu (1995), ο οποίος δείχνει ότι σε ένα διακριτό χώρο καταστάσεων, ο δειγματολήπτης Gibbs μπορεί να βελτιωθεί από Metropolis - Hastings βήματα. Η βελτίωση εκφράζεται σε όρους ελάττωσης της διακύμανσης του εργοδικού μέσου (3.2). Η τροποποίηση αυτή, που ονομάζεται Metropolization κατά Liu (1995) (Metropolization by Liu), βασίζεται στο επόμενο αποτέλεσμα, το οποίο οφείλεται στον Peskun (1973):

**Λήμμα 4.6.1** *Θεωρήστε δύο αντιστρέψιμες αλυσίδες Markou σε ένα διακριτό χώρο καταστάσεων, με πίνακες μετάβασης  $T_1$  και  $T_2$ , έτσι ώστε τα μη διαγώνια στοιχεία του πίνακα  $T_2$  να είναι μεγαλύτερα από τα αντίστοιχα του πίνακα  $T_1$ . Η αλυσίδα που προκύπτει από τον πίνακα μετάβασης  $T_2$  υπερέχει της άλλης, με την έννοια της μικρότερης διακύμανσης των εκτιμητών.*

Για κάποια πλήρη δεσμευμένη πυκνότητα  $\pi_i(\theta_i|\theta_{-i})$ , σε ένα διακριτό χώρο καταστάσεων, η τροποποίηση που προτάθηκε από τον Liu (1995) προβλέπει την εισαγωγή ενός Metropolis - Hastings βήματος:

**Αλγόριθμος 4.6.3** *Δοθείσης της τιμής  $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})$ ,*

1. *προσομοίωσε  $z_i \neq \theta_i^{(j)}$ , με πιθανότητα*

$$\frac{\pi_i(z_i|\theta_{-i}^{(j)})}{1 - \pi_i(\theta_i^{(j)}|\theta_{-i}^{(j)})},$$

2. *αποδέξου  $\theta_i^{(j+1)} = z_i$ , με πιθανότητα*

$$\frac{1 - \pi_i(\theta_i^{(j)}|\theta_{-i}^{(j)})}{1 - \pi_i(z_i|\theta_{-i}^{(j)})} \wedge 1,$$

όπου  $x \wedge y = \min(x, y)$ .

Στον Αλγόριθμο 4.6.3, για τη συντεταγμένη  $\theta_i$  του διανύσματος  $\theta$ , η πιθανότητα μετάβασης από την τιμή  $\theta_i^{(j)}$  σε κάποια διαφορετική τιμή, είναι μεγαλύτερη από την αντίστοιχη του δειγματολήπτη Gibbs, ο οποίος δίνεται από τον Αλγόριθμο 4.2.1. Το Λήμμα 4.6.1 εξασφαλίζει την υπεροχή του Αλγόριθμου 4.6.3 σε όρους της διακύμανσης των εκτιμητών του προσομοιωμένου πληθυσμού.

### 4.6.3 Ομαδοποιημένος και Τροποποιημένος Εκφυλισμένος Gibbs

Ο Liu (1994) προτείνει μια μέθοδο «ομαδοποίησης» (grouping) και «εκφυλισμού» (collapsing) στη χρήση του δειγματολήπτη Gibbs. Για να αντιληφθείτε τη σκέψη του Liu, θεωρήστε μια τριδιάστατη εκ των υστέρων πυκνότητα  $\pi(\theta)$ , όπου  $\theta = (\theta_1, \theta_2, \theta_3)$ . Προτείνει, λοιπόν, τις ακόλουθες αλγοριθμικές μορφές:

#### Αλγόριθμος 4.6.4 *Grouped Gibbs Sampler*

1.  $(\theta_1, \theta_2) \sim \pi_{12}(\theta_1, \theta_2 | \theta_3)$ ,
2.  $\theta_3 \sim \pi_3(\theta_3 | \theta_1, \theta_2)$ ,

Στον παραπάνω αλγόριθμο, ομαδοποιούμε τις παραμέτρους  $\theta_1$  και  $\theta_2$  και έπειτα τις προσομοιώνουμε ταυτόχρονα, από την από κοινού εκ των υστέρων δεσμευμένη πυκνότητά τους  $\pi(\theta_1, \theta_2 | \theta_3)$ . Μια γενική σύσταση είναι να ομαδοποιούμε θετικά συσχετισμένες παραμέτρους.

Ο επόμενος αλγόριθμος λέγεται **τροποποιημένος εκφυλισμένος δειγματολήπτης Gibbs**:

#### Αλγόριθμος 4.6.5 *Modified Collapsed Gibbs Sampler*

1.  $\theta_1 \sim p(\theta_1 | \theta_2)$ ,
2.  $\theta_2 \sim p(\theta_2 | \theta_1)$
3.  $\theta_3 \sim \pi_3(\theta_3 | \theta_1, \theta_2)$ ,

Μπορούμε να δείξουμε ότι ο τροποποιημένος εκφυλισμένος δειγματολήπτης Gibbs καταλήγει στην ίδια κατανομή στόχο με τον αυθεντικό δειγματολήπτη Gibbs. Θεωρήστε δύο διαδοχικές καταστάσεις της αλυσίδας  $(\theta^{(t)})$ , τις

$\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)})$  και  $\theta^{(j+1)} = (\theta_1^{(j+1)}, \theta_2^{(j+1)}, \theta_3^{(j+1)})$ . Ο πυρήνας μετάβασης που συνδέεται με τον Αλγόριθμο 4.6.5 είναι

$$\begin{aligned} p(\theta^{(j)}, \theta^{(j+1)}) &= p(\theta_1^{(j+1)} | \theta_2^{(j)}) p(\theta_2^{(j+1)} | \theta_1^{(j+1)}) \\ &\times p(\theta_3^{(j+1)} | \theta_1^{(j+1)}, \theta_2^{(j+1)}). \end{aligned} \quad (4.11)$$

Προκύπτει ότι

$$\int_{\mathbb{R}^3} p(\theta^{(j)}, \theta^{(j+1)}) \pi(\theta^{(j)}) d\theta^{(j)} = \pi(\theta^{(j+1)}). \quad (4.12)$$

Επομένως, η στάσιμη κατανομή της αλυσίδας που παράγει ο Αλγόριθμος 4.6.5 είναι η  $\pi(\theta)$  και είναι η ίδια με την αντίστοιχη του αυθεντικού Αλγόριθμου 4.2.1.

Αν θέλουμε να προβούμε σε μία σύγκριση των δύο προηγούμενων αλγόριθμων και του πρωτότυπου 4.2.1, έχουμε τα εξής: ο ομαδοποιημένος Gibbs είναι αποδοτικότερος από τον πρωτότυπο. Ο τροποποιημένος εκφυλισμένος Gibbs ενδέχεται να είναι αποδοτικότερος από τον πρωτότυπο. Μεταξύ του τροποποιημένου εκφυλισμένου και του ομαδοποιημένου, η εικόνα δεν είναι ξεκάθαρη. Η απόδοση των τελευταίων, πιθανώς εξαρτάται από τις συσχετίσεις μεταξύ των συντεταγμένων του διανύσματος  $\theta$ . Για παράδειγμα, αν οι συσχετίσεις μεταξύ των συντεταγμένων  $\theta_1$  και  $\theta_2$ , για τον τρισδιάστατο παραμετρικό χώρο του παραδείγματός μας, είναι υψηλές, τότε ο ομαδοποιημένος Gibbs αναμένεται να είναι αποδοτικότερος από τον τροποποιημένο εκφυλισμένο.

## 4.7 Αύξηση Δεδομένων

Οι Tanner και Wong (1987) εισήγαγαν μια μέθοδο με την ονομασία αύξηση δεδομένων (data augmentation). Η μέθοδος αυτή αποσκοπεί στο να προσομοιώσει τα δεδομένα που λείπουν από ένα σύνολο δεδομένων ή τις άγνωστες τιμές ενός συνόλου παραμέτρων. Βασίζεται στην αύξηση της ήδη υπάρχουσας πληροφορίας και επαναληπτικά βελτιώνει την ποιότητα της προσομοιωμένης χαμένης (ή αγνοούμενης) πληροφορίας. Η μέθοδος της αύξησης δεδομένων είναι μια MCMC τεχνική, η οποία συνεχώς παράγει βελτιωμένες εκτιμήσεις, δοθείσης της προηγούμενης εκτίμησης και για το λόγο αυτό η ακολουθία της προσομοιωμένης πληροφορίας σχηματίζει μια αλυσίδα Markov.



Ας υποθέσουμε ότι κατά την παρατήρηση κάποιων δεδομένων, αποτυγχάνουμε να παρατηρήσουμε το πλήρες σετ των δεδομένων:  $X = [X_{\text{obs}}, X_{\text{mis}}]$ , όπου  $X_{\text{obs}}$  είναι το παρατηρηθέν σύνολο των δεδομένων και  $X_{\text{mis}}$  το μη παρατηρηθέν. Και τα δύο υποσύνολα δεδομένων είναι δεσμευμένα ως προς την παράμετρο ενδιαφέροντος,  $\theta$ . Για σημειογραφική ευκολία θέτουμε  $X_{\text{obs}} = x$ ,  $X_{\text{mis}} = y$  και χρησιμοποιήσουμε το σύμβολο  $p$  ως ένα γενικό σύμβολο για τη συνάρτηση πυκνότητας. Η μέθοδος της αύξησης των δεδομένων απαιτεί τη γνώση της παραμετρικής μορφής της εκ των υστέρων κατανομής  $p(\theta|X)$ , που αντιστοιχεί στο πλήρες σετ των δεδομένων, καθώς και της πυκνότητας  $p(y|x)$ .

Ξεκινάμε την παρουσίαση της μεθόδου με τον ορισμό της εκ των υστέρων ταυτότητας

$$p(\theta|x) = \int_{X_{\text{mis}}} p(\theta|x, y) p(y|x) dy \quad (4.13)$$

Ορίζουμε επίσης την «ταυτότητα πρόβλεψης» (predictive identity), ισχυριζόμενοι ότι υπάρχει κάποια άγνωστη παράμετρος  $\eta$ , από τον παραμετρικό χώρο  $\Theta$  της παραμέτρου  $\theta$ , η οποία είναι καθοριστικής σημασίας για τη γέννηση των χαμένων δεδομένων:

$$p(y|x) = \int_{\Theta} p(y|\eta, x) p(\eta|x) d\eta, \quad (4.14)$$

όπου οι πυκνότητες  $p(\theta|x, y)$  και  $p(y|\eta, x)$  είναι γνωστές. Το σύστημα των εξισώσεων (4.13) και (4.14) θα μας βοηθήσει να προσδιορίσουμε την κατανομή στόχο,  $p(\theta|x)$ . Για να λύσουμε το σύστημα αυτό, αντικαθιστούμε τη εξίσωση (4.14) στην (4.13) και αλλάζουμε τη σειρά της ολοκλήρωσης:

$$\begin{aligned} p(\theta|x) &= \int_{X_{\text{mis}}} p(\theta|x, y) \left[ \int_{\Theta} p(y|\eta, x) p(\eta|x) d\eta \right] dy \\ &= \int_{\Theta} \left[ \int_{X_{\text{mis}}} p(\theta|x, y) p(y|\eta, x) dy \right] p(\eta|x) d\eta \\ &= \int_{\Theta} K(\theta, \eta) p(\eta|x) d\eta, \end{aligned} \quad (4.15)$$

όπου  $K(\theta, \eta) = \int_{X_{\text{mis}}} p(\theta|x, y) p(y|\eta, x) dy$ .

Τι συμβαίνει όμως αν στη σχέση (4.15), αντί για την πυκνότητα  $p(\eta|x)$ , θέσουμε μια προσέγγισή της; Δηλαδή, υποθέστε ότι η πυκνότητα  $p_0(\eta|x)$  είναι

κάποια αρχική εκτίμηση της  $p(\theta|x)$ . Τότε γράφουμε

$$p_1(\theta|x) = \int_{\Theta} K(\theta, \eta) p_0(\eta|x) d\eta \equiv I_K[p_0(\theta|x)], \quad (4.16)$$

όπου  $I_K$  υποδηλώνει τον τελεστή ολοκλήρωσης (integral operator) ως προς τη συνάρτηση  $K$ . Αν ορίσουμε  $p_{i+1}(\theta|x) = I_K[p_i(\theta|x)]$ ,  $i = 0, 1, \dots$ , τότε μήπως αυτή η επαναληπτική διαδικασία συγκλίνει;

**Θεώρημα 4.7.1** Δεδομένου ότι η από κοινού συνάρτηση κατανομής των  $\theta$  και  $y$ , έχει στήριγμα το χώρο  $\Theta \times \mathcal{Y}$ , τότε:

1. Η  $p(\theta|x)$ , η αληθινή περιθωριακή πυκνότητα, είναι η μοναδική λύση του συστήματος των εξισώσεων (4.13) και (4.14).
2. Η ακολουθία  $\{p_i(\theta|x)\}$  συγκλίνει μονότονα στην πυκνότητα  $p(\theta|x)$ .
3.  $\int |p_i(\theta|x) - p(\theta|x)| d\theta \rightarrow 0$ , εκθετικά ως προς  $i$ .

Τα αποτελέσματα αυτά αποδεικνύονται από τους Tanner και Wong (1987) στα πλαίσια των προβλημάτων ελλιπών δεδομένων (missing data problems).

Επομένως, έχουμε μια επαναληπτική διαδικασία για την εκτίμηση της περιθωριακής πυκνότητας  $p(\theta|x)$ . Ωστόσο, η διαδικασία αυτή απαιτεί τον αναλυτικό υπολογισμό του ολοκληρώματος, που εμφανίζεται στη σχέση (4.15), κάτι που γενικά δεν είναι εφικτό<sup>4</sup>. Αντί λοιπόν να υπολογίζουμε τη σχέση (4.15) αναλυτικά, χρησιμοποιούμε μια «δειγματικής φύσης» προσέγγιση: προσομοιώνουμε μια αρχική τιμή  $\theta^{(0)} \sim p_0(\theta|x)$ , έπειτα  $y^{(1)} \sim p(y|\theta^{(0)}, x)$ . Παρατηρήστε ότι η περιθωριακή πυκνότητα της  $y^{(1)}$  είναι η  $p^{(1)}(y|x) = \int p(y|\theta, x) p_0(\theta|x) d\theta$ . Στη συνέχεια, προσομοιώνουμε  $\theta^{(1)} \sim p(\theta|x, y^{(1)})$ , όπου η  $\theta^{(1)}$  έχει περιθωριακή πυκνότητα

$$\begin{aligned} p^{(1)}(\theta|x) &= \int p(\theta|y, x) p^{(1)}(y|x) dy \\ &= \int K(\theta, \eta) p_0(\eta|x) d\eta \\ &= I_K[p_0(\theta|x)]. \end{aligned}$$

<sup>4</sup>Αν ήταν εφικτός ο αναλυτικός υπολογισμός του ολοκληρώματος, που εμφανίζεται στη σχέση (4.15), τότε δεν θα υπήρχε η ανάγκη παρουσίασης της μεθόδου της αύξησης των δεδομένων.

Από την ιδιότητα 2 του Θεωρήματος 4.7.1, η επανάληψη αυτής της διαδικασίας παράγει ζεύγη  $(\theta^{(i)}, y^{(i)})$ , τέτοια ώστε

$$\theta^{(i)} \xrightarrow{d} \theta \sim p(\theta|x) \quad \text{και} \quad y^{(i)} \xrightarrow{d} y \sim p(y|x).$$

Επομένως, για αρκούντως μεγάλο  $j$ , η ποσότητα  $\theta^{(j)}$  μπορεί να θεωρηθεί ως δειγματική πραγματοποίηση από την περιθωριακή πυκνότητα  $p(\theta|x)$  και η ποσότητα  $y^{(j)}$  ως δειγματική πραγματοποίηση από την περιθωριακή πυκνότητα  $p(y|x)$ . Το αντίτιμο της αποφυγής του υπολογισμού του ολοκληρώματος της σχέσης (4.15) είναι ότι δεν αποκτούμε την πυκνότητα  $p(\theta|x)$  άλλα μια δειγματική πραγματοποίηση από αυτή.

Ο αλγόριθμος των Tanner and Wong (1987) είναι ο εξής:

**Αλγόριθμος 4.7.1** 1. προσομοίωσε  $\theta^{(j-1)} \sim \hat{p}^{(j-1)}(\theta|x) = p(\theta|x, y^{(j-1)})$ .

2. προσομοίωσε  $y^{(j)} = (y_1^{(j)}, \dots, y_m^{(j)}) \sim p(y|\theta^{(j-1)}, x)$ .

3. Αναβάθμισε την εκτίμηση της περιθωριακής πυκνότητας  $p(\theta|x)$  με τον τύπο

$$\hat{p}^{(j)}(\theta|x) = \frac{1}{m} \sum_{k=1}^m p(\theta|y_k^{(j)}, x).$$

Ο Tanner (1993) παρατηρεί ότι ακόμα και  $m = 1$ , στον Αλγόριθμο 4.7.1, είναι επαρκές για να επέλθει η σύγκλιση του αλγόριθμου και αναφέρεται σε αυτή την προσέγγιση με τη ονομασία «αλυσιδωτή αύξηση δεδομένων» (**chained data augmentation**). Επίσης για  $m = 1$  η μέθοδος της αύξησης δεδομένων συμπίπτει με το δειγματολήπτη Gibbs.

## 4.8 Δυικές Πιθανοτικές Δομές

Απλοποιώντας τη σημειογραφία του Αλγόριθμου 4.7.1 τον ξαναγράφουμε ως εξής:

**Αλγόριθμος 4.8.1** Δοθείσης της τιμής  $y^{(j)} = (y_1^{(j)}, y_2^{(j)})$ :

1. προσομοίωσε  $y_1^{(j+1)} \sim g_1(y_1|y_2^{(j)})$ ,

2. προσομοίωσε  $y_2^{(j+1)} \sim g_2(y_2|y_1^{(j+1)})$ ,

όπου  $(y_1, y_2) \sim g(\cdot, \cdot)$  και  $g$  είναι η από κοινού πυκνότητα των  $y_1$  και  $Y_2$ .

Στην προηγούμενη ενότητα μελετήσαμε τη μέθοδο αύξησης δεδομένων ως ειδική περίπτωση του δειγματολήπτη Gibbs. Αναφερθήκαμε στη αλληλοεξέλιξη δύο συντεταγμένων (ίσως παραμέτρων) ενός διανύσματος  $y = (y_1, y_2)$ . Αυτή η συγκεκριμένη μορφή του δειγματολήπτη Gibbs απλοποιεί τη μελέτη των πιθανοτικών ιδιοτήτων του Αλγόριθμου 4.8.1, αφού κάθε μία από τις υποαλυσίδες  $y_i$ ,  $i = 1, 2$  μπορεί να εξετασθεί ξεχωριστά. Αυτή η πολύ ενδιαφέρουσα ιδιότητα ισχύει μόνο για πλήθος συντεταγμένων  $d = 2$ . Για  $d \geq 3$  η ιδιότητα αυτή εξαφανίζεται διότι οι ακολουθίες  $y_1, \dots, y_d$  δεν είναι αλυσίδες Markov, παρόλο που το διάνυσμα  $y$ , συνολικά, είναι αλυσίδα Markov. Επομένως, δεν υπάρχει πυρήνας μετάβασης σχετικός με τις υποαλυσίδες  $y_i$ .

Τα όσα είπαμε παραπάνω εδραιώνονται με το ακόλουθο λήμμα:

**Λήμμα 4.8.1** *Κάθε μία από τις ακολουθίες  $y_1$  και  $y_2$ , που παράγονται από τον Αλγόριθμο 4.8.1, είναι αλυσίδα Markov με αντίστοιχη στάσιμη κατανομή στόχο*

$$g^{(1)}(y_1) = \int g(y_1, y_2) dy_2, \quad (4.17)$$

$$g^{(2)}(y_2) = \int g(y_1, y_2) dy_1. \quad (4.18)$$

Επιπλέον, αν ικανοποιείται η συνθήκη θετικότητας, κατά τον Ορισμό 4.5.1, τότε οι υποαλυσίδες  $y_1$  και  $y_2$  είναι ισχυρά ανάγωγες (*strongly irreducible*).

Οι Liu et al. (1994) έδειξαν ότι ο Αλγόριθμος 4.8.1 απολαμβάνει μια πολύ ισχυρή δομική ιδιότητα: οι δύο υποαλυσίδες  $y_1$  και  $y_2$  που παράγει, ικανοποιούν μια δεινή ιδιότητα, την οποία ονομάζουν **interleaving property**. Η ιδιότητα αυτή είναι χαρακτηριστική των σχεδίων αύξησης δεδομένων και οι μοναδικές παραχθείσες αλυσίδες Gibbs που ικανοποιούν αυτή την ιδιότητα, είναι αυτές που συνδέονται με τους αλγόριθμους αύξησης δεδομένων.

**Ορισμός 4.8.1** *Δύο αλυσίδες Markov  $y_1$  και  $y_2$  λέγονται συζυγείς (conjugate) μεταξύ τους, με την ιδιότητα interleaving αν*

1. οι  $y_1^{(j)}$  και  $y_1^{(j+1)}$  είναι ανεξάρτητες δοθείσης της  $y_2^{(j)}$ ,
2. οι  $y_2^{(j-1)}$  και  $y_2^{(j)}$  είναι ανεξάρτητες δοθείσης της  $y_1^{(j)}$  και
3. οι  $(y_1^{(j)}, y_2^{(j-1)})$  και  $(y_1^{(j)}, y_2^{(j)})$  κατανέμονται ισόνομα, υπό στασιμότητας.

Η ιδιότητα interleaving ικανοποιείται από τις υποαλυσίδες που παράγει η μέθοδος της αύξησης δεδομένων.

**Λήμμα 4.8.2** Κάθε μία εκ των αλυσίδων  $y_1$  και  $y_2$  που παράγονται από τον αλγόριθμο της αύξησης δεδομένων είναι αντιστρέψιμη (reversible) και η αλυσίδα  $y = (y_1, y_2)$  ικανοποιεί την ιδιότητα interleaving.

Γιατί όμως η ιδιότητα interleaving είναι σημαντική; Το παρακάτω λήμμα είναι διαφωτιστικό.

**Λήμμα 4.8.3** Αν η μία από τις δύο interleaved αλυσίδες Markov,  $y_1$  και  $y_2$  είναι εργοδική (γεωμετρικά εργοδική), τότε έπεται η εργοδικότητα (αντίστοιχα γεωμετρική εργοδικότητα) και για την άλλη.

Αξίζει να σημειώσουμε ότι η αλυσίδα  $y = (y_1, y_2)$ , που παράγεται από τον αλγόριθμο της αύξησης δεδομένων, δεν είναι πάντα αντιστρέψιμη. Για να εξασφαλίσουμε την αντιστρεψιμότητά της, είναι απαραίτητο να εισάγαγουμε ένα επιπλέον βήμα στον Αλγόριθμο 4.8.1:

**Αλγόριθμος 4.8.2** Δοθείσης της τιμής  $y^{(j)} = (y_1^{(j)}, y_2^{(j)})$ :

1. Προσομοίωσε  $W \sim g_1(w|y_2^{(j)})$ ,
2. προσομοίωσε  $y_2^{(j+1)} \sim g_2(y_2|w)$ ,
3. προσομοίωσε  $y_1^{(j+1)} \sim g_1(y_1|y_2^{(j+1)})$ .

## 4.9 Rao - Blackwellization

Όπως είπαμε και στο κεφάλαιο 3, ο βασικός σκοπός μας κατά την ανάπτυξη MCMC μεθόδων προσομοίωσης είναι ο υπολογισμός εργοδικών μέσων της μορφής (2.2), δηλαδή

$$\delta_0 = \frac{1}{T} \sum_{t=1}^T h(\theta_i^{(t)}).$$

Οι Gelfand και Smith (1990) προτείνουν έναν διαφορετικό εκτιμητή ο οποίος βασίζεται σε κάποια μορφή δέσμευσης (conditioning), την οποία ονομάζουν Rao - Blackwellization, λόγω του ότι βασίζεται στο θεώρημα Rao - Blackwell.

Ο Rao - Blackwellized εκτιμητής βασίζεται στην ταυτότητα

$$\pi^{(i)}(\theta_i) = \int \pi(\theta_1, \dots, \theta_d) d\theta_1, \dots, \theta_{i-1} d\theta_{i+1}, \dots, d\theta_d \quad (4.19)$$

και δίνεται από τη σχέση

$$\delta_{RB} = \frac{1}{T} \sum_{t=1}^T E \left[ h(\theta_i) | \theta_{-i}^{(t)} \right], \quad (4.20)$$

όπου  $\theta_{-i}^{(t)} = (\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})$ .

Και οι δύο εκτιμητές,  $\delta_0$  και  $\delta_{RB}$ , συγκλίνουν στο θεωρητικό μέσο  $E[h(\theta_i)]$ . Η σύγκλιση οφείλεται στο Εργοδικό Θεώρημα. Η σύγκλιση του δεύτερου προκύπτει από το γεγονός ότι  $E(\delta_0) = E(\delta_{RB})$ , αφού  $E[E(h(\theta_i) | \theta_{-i})] = E[h(\theta_i)] = E(\delta_0)$ ,  $\forall i$ . Υπό συνθηκών στασιμότητας και οι δύο εκτιμητές είναι αμερόληπτοι (unbiased).

Από τη γνωστή ταυτότητα

$$\text{Var}(X) = E_Y[\text{Var}(X|Y)] + \text{Var}_Y(E[X|Y]),$$

έχουμε

$$\text{Var}(h(\theta_i)) = \text{Var}\left(E\left[h(\theta_i) | \theta_{-i}^{(t)}\right]\right) + E\left[\text{Var}\left(h(\theta_i) | \theta_{-i}^{(t)}\right)\right],$$

που σημαίνει ότι

$$\text{Var}\left(E\left[h(\theta_i) | \theta_{-i}^{(t)}\right]\right) \leq \text{Var}(h(\theta_i)). \quad (4.21)$$

Το αποτέλεσμα αυτό οδήγησε τους Gelfand και Smith (1990) να προτείνουν τη χρήση του εκτιμητή  $\delta_{RB}$  αντί του  $\delta_0$ . Ωστόσο, πρέπει να παρατηρήσουμε κάτι σημαντικό. Για τη διακύμανση  $\text{Var}(\delta_{RB})$  έχουμε:

$$\begin{aligned} \text{Var}(\delta_{RB}) &= \frac{1}{T^2} \left\{ \sum_{t=1}^T \text{Var}\left(E\left[h(\theta_i) | \theta_{-i}^{(t)}\right]\right) \right. \\ &\quad \left. + \sum_{s \neq k}^T \text{Cov}\left(E\left[h(\theta_i) | \theta_{-i}^{(s)}\right], E\left[h(\theta_i) | \theta_{-i}^{(k)}\right]\right) \right\}. \quad (4.22) \end{aligned}$$

Επίσης, για τη διακύμανση  $Var(\delta_0)$  έχουμε:

$$Var(\delta_0) = \frac{1}{T^2} \left\{ \sum_{t=1}^T Var(h(\theta_i^{(t)})) + \sum_{s \neq k}^T Cov(h(\theta_i^{(s)}), h(\theta_i^{(k)})) \right\} \quad (4.23)$$

Η σχέση (4.21) ισχύει πάντα, παρόλα αυτά δεν μπορούμε να ισχυριστούμε ότι  $Var(\delta_{RB}) \leq Var(\delta_0)$ , διότι έτσι αγνοούμε τις αυτοσυνδιακυμάνσεις μεταξύ των τιμών της  $i$  συνιστώσας  $\theta_i^{(t)}$ , για  $t = 1, \dots, T$ . Υπάρχουν περιπτώσεις όπου ισχύει  $Var(\delta_{RB}) \geq Var(\delta_0)$ .

Ποια, λοιπόν, η αξία του Rao - Blackwellized εκτιμητή; Μήπως μπορούμε να βρούμε μια κλάση αλυσίδων Markov, για τις οποίες ο εκτιμητής  $\delta_{RB}$  να είναι καλύτερος, από άποψη διασποράς, από τον εκτιμητή  $\delta_0$ ; Η απάντηση είναι ναι, πρέπει όμως πρώτα να παρουσιάσουμε δύο προκαταρκτικά αποτελέσματα.

**Λήμμα 4.9.1** Αν  $h \in \mathcal{L}_2(\pi_2)$  και οι αλυσίδες Markov  $\theta$  και  $\eta$  είναι *interleaved*, τότε

$$Cov(h(\eta^{(1)}), h(\eta^{(2)})) = Var(E[h(\eta) | \theta]). \quad (4.24)$$

Το παραπάνω λήμμα μας εφοδιάζει με το πολύ ενδιαφέρον αποτέλεσμα ότι σε μία *interleaved* αλυσίδα, οι συνδιακυμάνσεις μεταξύ των διαδοχικών τυχαίων μεταβλητών της,  $\eta^{(t)}$ ,  $t = 1, 2, \dots$ , είναι θετικές. Μπορούμε να επεκτείνουμε το αποτέλεσμα αυτό:

**Πρόταση 4.9.1** Έστω  $\theta$  μια αλυσίδα Markov με την ιδιότητα *interleaving*. Τότε οι συνδιακυμάνσεις

$$Cov(h(\theta^{(1)}), h(\theta^{(t)})) \quad (4.25)$$

είναι θετικές και φθίνουσες ως προς  $t$ , για κάθε  $h \in \mathcal{L}_2(\pi_2)$ .

Τώρα μπορούμε να ορίσουμε την κλάση των Markovιανών αλυσίδων, για τις οποίες απολαμβάνουμε την υπεροχή του εκτιμητή  $\delta_{RB}$  έναντι του  $\delta_0$ .

**Θεώρημα 4.9.1** Αν οι αλυσίδες Markov  $\theta$  και  $\eta$  είναι *interleaved*, με στάσιμες κατανομές  $\pi$  και  $\tau$  αντίστοιχα, ο εκτιμητής

$$\delta_{RB} = \frac{1}{T} \sum_{t=1}^T E[h(\theta) | \eta^{(t)}] \quad (4.26)$$

υπερέχει, από άποψη διασποράς, του εκτιμητή

$$\delta_0 = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}), \quad (4.27)$$

για κάθε  $h \in \mathcal{L}_2(\pi)$  και  $h \in \mathcal{L}_2(\tau)$ .

Ας υποθέσουμε τώρα ότι ικανοποιούνται οι συνθήκες του Θεωρήματος (4.9.1). Είναι η βελτίωση αυτή συμφέρουσα, δεδομένης της αυξημένης υπολογιστικής δυσκολίας του εκτιμητή  $\delta_{RB}$ ; Παραμένει αξιόλογα ελαττωμένη η διακύμανση του εκτιμητή  $\delta_{RB}$  έναντι αυτής του  $\delta_0$ , καθώς το μέγεθος του δείγματος (ή αλλιώς των επαναλήψεων του αλγόριθμου) αυξάνει; Ομολογουμένως καυτά ερωτήματα!

Με το ζήτημα αυτό ασχολήθηκε ο Levine (1996), ο οποίος διατύπωσε το πρόβλημα σε όρους της Ασυμπτωτικής Σχετικής Αποτελεσματικότητας ARE (Asymptotic Relative Efficiency) του εκτιμητή  $\delta_0$  στην (4.27), σε σχέση με την Rao - Blackwellized εκδοχή του,  $\delta_{RB}$ , στην (4.26), όπου τα ζεύγη  $(\theta^{(t)}, \eta^{(t)})$  παράγονται από ένα διμεταβλητό δειγματολήπτη Gibbs. Η ARE είναι ο λόγος των διακυμάνσεων των οριακών κατανομών για τους δύο εκτιμητές. Οι διακυμάνσεις αυτές δίνονται από τους τύπους:

$$\sigma_{\delta_0}^2 = \text{Var}(h(\theta^{(0)})) + 2 \sum_{t=1}^{\infty} \text{Cov}(h(\theta^{(0)}), h(\theta^{(t)})) \quad (4.28)$$

$$\begin{aligned} \sigma_{\delta_{RB}}^2 &= \text{Var}(E[h(\theta) | \eta]) \\ &+ 2 \sum_{t=1}^{\infty} \text{Cov}(E[h(\theta^{(0)}) | \eta^{(0)}], E[h(\theta^{(t)}) | \eta^{(t)}]). \end{aligned} \quad (4.29)$$

Ο Levine (1996) έδειξε ότι  $\sigma_{\delta_0}^2 / \sigma_{\delta_{RB}}^2 \geq 1$ .

Οι Liu et al. (1995) επεκτείνουν την ισχύ της Πρότασης 4.9.1 στην περίπτωση του τυχαίου δειγματολήπτη Gibbs που δίνεται από τον Αλγόριθμο 4.2.4.

Ένα άλλο πλεονέκτημα της μεθόδου Rao - Blackwellization είναι η κατασκευή ενός αμερόληπτου εκτιμητή των περιθώριων πυκνοτήτων  $\pi_i(\theta_i)$ , ο οποίος δίνεται από τη σχέση

$$\hat{\pi}^{(i)}(\theta_i) = \frac{1}{T} \sum_{t=1}^T \pi_i(\theta_i | \theta_{-i}^{(t)}). \quad (4.30)$$

Βεβαίως, η  $\hat{\pi}_i(\theta_i)$  είναι υπολογίσιμη μόνο στην περίπτωση που οι πλήρως δεσμευμένες πυκνότητες,  $\pi_i(\theta_i | \theta_{-i}^{(t)})$ , δίνονται σε κλειστή μορφή.



## 4.10 Η Πληροφοριακή Ικανότητα των πλήρεις δεσμευμένων Κατανομών

Ένα πολύ σημαντικό χαρακτηριστικό του δειγματολήπτη Gibbs είναι η ικανότητα των πλήρεις δεσμευμένων πυκνοτήτων  $\pi_i(\theta_i|\theta_{-i})$  να περιέχουν επαρκή πληροφορία, ώστε να παράγουν ένα δείγμα από την απο κοινού κατανομή  $\pi(\theta)$  των συντεταγμένων του διανύσματος  $\theta$ . Ενώ οι πλήρεις δεσμευμένες πυκνότητες περιγράφουν τέλεια την από κοινού κατανομή, δεν συμβαίνει τι ίδιο με τις περιθωριακές πυκνότητες  $\pi^{(i)}(\theta_i)$ .

Ας θεωρήσουμε τον Αλγόριθμο 4.4.1, για την περίπτωση όπου  $d = 2$ . Ο αλγόριθμος αναπτύσσεται με τη βοήθεια δύο πλήρεις δεσμευμένων πυκνοτήτων, της  $g_1(y_1|y_2)$  και της  $g_2(y_2|y_1)$ . Η από κοινού πυκνότητα  $g(y_1, y_2)$  παράγεται άμεσα από τις πλήρεις δεσμευμένες πυκνότητες:

**Θεώρημα 4.10.1** Η από κοινού κατανομή που συνδέεται με τις πλήρεις δεσμευμένες πυκνότητες  $g_1$  και  $g_2$  έχει πυκνότητα

$$g(y_1, y_2) = \frac{g_2(y_2|y_1)}{\int \frac{g_2(w|y_1)}{g_1(y_1|w)} dw}. \quad (4.31)$$

Από το παραπάνω θεώρημα είναι ξεκάθαρο ότι για την παραγωγή της από κοινού πυκνότητας  $g(y_1, y_2)$  είναι απαραίτητη η ύπαρξη και ο υπολογισμός του ολοκληρώματος

$$\int \frac{g_2(w|y_1)}{g_1(y_1|w)} dw.$$

Ωστόσο, είναι σαφής η θεμελιώδης ιδιότητα των πλήρεις δεσμευμένων πυκνοτήτων να περιγράφουν επαρκώς την από κοινού πυκνότητα, βεβαίως όταν η τελευταία υπάρχει.

Η επέκταση του Θεωρήματος 4.10.1, για  $d > 2$ , είναι γνωστή με την ονομασία Θεώρημα Hammerslay - Clifford (Hammerslay - Clifford 1970, Besag 1974, Gelman και Speed 1993).

**Θεώρημα 4.10.2** Υπό τη συνθήκη της θετικότητας, η από κοινού κατανομή  $g(y_1, \dots, y_d)$  των συντεταγμένων του διανύσματος  $y = (y_1, \dots, y_d)$ , προκύπτει από τις πλήρεις δεσμευμένες πυκνότητες σύμφωνα με τον τύπο

$$g(y_1, \dots, y_d) \propto \prod_{i=1}^d \frac{g_{\sigma_i}(y_{\sigma_i}|y_{\sigma_1}, \dots, y_{\sigma_{i-1}}, y'_{\sigma_{i+1}}, \dots, y'_{\sigma_d})}{g_{\sigma_i}(y'_{\sigma_i}|y_{\sigma_1}, \dots, y_{\sigma_{i-1}}, y'_{\sigma_{i+1}}, \dots, y'_{\sigma_d})}, \quad (4.32)$$

για κάθε μετάθεση  $\sigma = \{\sigma_1, \dots, \sigma_d\}$  των στοιχείων του συνόλου  $\{1, \dots, d\}$  και κάθε  $y' \in \mathcal{Y}$ .

Η επέκταση του Θεωρήματος 4.10.2 σε συνθήκες μη θετικότητας είναι πιο εκλεπτυσμένη και απαιτεί επιπρόσθετες υποθέσεις. Ο Besag (1994) προτείνει μια γενίκευση, η οποία όμως δεν ευσταθεί πάντα στα πλαίσια του δειγματολήπτη Gibbs. Οι Hobert et al. (1997) τροποποιούν τη γενίκευση του Besag, ώστε να διατηρούνται οι ιδιότητες της σύγκλισης του δειγματολήπτη Gibbs και επιπλέον δείχνουν ότι η συνεκτικότητα (connectdness) του στηρίγματος της πυκνότητας στόχου  $g$ , είναι απαραίτητη προϋπόθεση ώστε ο αλγόριθμος να συγκλίνει υπό οποιαδήποτε επιτρεπτή παραμετροποίηση του μοντέλου στα πλαίσια της Μπεϋζιανής Στατιστικής.

# Κεφάλαιο 5

## Επιπρόσθετα Θέματα

### 5.1 Ο Σχηματισμός του Δείγματος

Υπάρχουν πολλοί τρόποι σχηματισμού ενός δείγματος μεγέθους  $n$  που λαμβάνουμε από ένα MCMC αλγόριθμο. Ο εμφανής τρόπος είναι να προσομοιώσουμε  $n$  παράλληλες αλυσίδες μέχρι σύγκλισης, η οποία υποθέστε ότι επέρχεται ύστερα από  $m$  επαναλήψεις του αλγόριθμου (burn in period). Τότε από κάθε αλυσίδα λαμβάνουμε την τελευταία τιμή, δηλαδή την  $m$  τιμή. Συνολικά προσομοιώνουμε  $mn$  τιμές. Αν τα σημεία εκκίνησης των  $n$  αλυσίδων έχουν ληφθεί ανεξάρτητα, τότε το τελικό δείγμα είναι ένα iid δείγμα από την κατανομή στόχο  $\pi$ . Οι αρχικές τιμές πρέπει να είναι διαφορετικές μεταξύ τους και με μεγαλύτερη διασπορά από εκείνη της κατανομής στόχου.

Ένας δεύτερος τρόπος για να σχηματίσουμε το δείγμα μας είναι να προσομοιώσουμε μία μόνο, αλλά μεγάλη αλυσίδα και να εκμεταλλευτούμε τις ιδιότητες του Εργοδικού Θεωρήματος. Ύστερα από την έλευση της σύγκλισης, όλες οι προσομοιωμένες τιμές θεωρούνται ότι προέρχονται από την κατανομή στόχο  $\pi$ . Επομένως, ένα δείγμα μεγέθους  $n$  μπορεί να σχηματιστεί με τη λήψη  $n$  διαδοχικών τιμών της αλυσίδας. Τώρα το δείγμα δεν είναι ανεξάρτητο, λόγω της αυτοσυσχέτισης που παρουσιάζουν οι όροι της ακολουθίας  $(\theta^{(t)})$ . Το Εργοδικό Θεώρημα μας εξασφαλίζει τη σύγκλιση των εργοδικών μέσων. Από πρακτικής πλευράς, είναι πιθανό να δημιουργηθούν προβλήματα αν η αυτοσυσχέτιση είναι υψηλή και το μέγεθος του δείγματος είναι μικρό. Στην περίπτωση αυτή, η αλυσίδα μπορεί να χρειαστεί πολύ χρόνο ώστε να διέλθει από ολόκληρο το στήριγμα της  $\pi$ .

Μια εναλλακτική προσέγγιση, που στεγάζει την έννοια της ανεξαρτησίας,

είναι η **υποδειγματοληψία (subsampling or batch sampling)**, που αναφέρεται από διάφορους συγγραφείς, όπως οι Geyer (1992), Raftery και Lewis (1992), Diebolt και Robert (1994). Σύμφωνα με τη μέθοδο αυτή, από την αλυσίδα που παράγει ο MCMC αλγόριθμος, διατηρούμε μία τιμή κάθε  $k$  ( $k > 1$ ) προσομοιωμένες τιμές, προκειμένου να ελαττώσουμε την αυτοσυσχέτιση μεταξύ των τιμών της υποαλυσίδας που τελικά προκύπτει. Η υποαλυσίδα αυτή αποτελείται από τις τιμές  $\eta^{(t)} = \theta^{(kt)}$ . Επομένως, για να αποκτήσουμε ένα δείγμα αποτελούμενο από  $n$  ψευδο - ανεξάρτητες τιμές, χρειάζεται να προσομοιώσουμε  $m + kn$  τιμές από την αλυσίδα ( $\theta^{(t)}$ ). Τυπικά, η τιμή του  $k$  είναι μικρότερη από αυτή του  $m$ . Αν οι συνδιακυμάνσεις  $Cov_{\pi}(\theta^{(0)}, \theta^{(t)})$  φθίνουν μονότονα ως προς  $t$ , τότε υπάρχει εμφανές κίνητρο για τη χρησιμοποίηση της μεθόδου της υποδειγματοληψίας<sup>1</sup>. Ωστόσο, το να ανιχνεύσει κάποιος το αν η συνδιακύμανση αυτή είναι μονότονα φθίνουσα είναι πολλές φορές δύσκολη υπόθεση. Άλλες πάλι φορές, η συνδιακύμανση ταλαντεύεται, έχοντας μια ασταθή σχέση ως προς το  $t$ , γεγονός που δυσχεραίνει την επιλογή του  $k$ .

Η εφαρμογή της μεθόδου της υποδειγματοληψίας επιφέρει απώλειες στην αποδοτικότητα του εκτιμητή

$$\frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) \quad (5.1)$$

της πληθυσμιακής ποσότητας  $E_{\pi}[h(\theta)]$ . Όπως έδειξαν οι MacEachern και Berliner (1994), είναι πάντα προτιμότερο να χρησιμοποιείται ολόκληρο το δείγμα ( $\theta^{(t)}$ ) που παράγεται από τον αλγόριθμο για την εκτίμηση της ποσότητας  $E_{\pi}[h(\theta)]$ , παρά το υποδείγμα ( $\eta^{(t)}$ ):

**Λήμμα 5.1.1** Θεωρήστε μια συνάρτηση  $h \in \mathcal{L}^2(\pi)$ . Για κάθε  $k > 1$ , αν η αλυσίδα Markov ( $\theta^{(t)}$ ) έχει στάσιμη κατανομή την  $\pi$  και αν

$$\delta_1 = \frac{1}{kT} \sum_{t=1}^{kT} h(\theta^{(t)}) \quad \text{και} \quad \delta_k = \frac{1}{T} \sum_{\ell=1}^T h(\theta^{(k\ell)}),$$

τότε ισχύει

$$Var(\delta_1) \leq Var(\delta_k).$$

Παρόλα αυτά, για θέματα ανίχνευσης της σύγκλισης και υπολογιστικών δυσκολιών (μνήμη υπολογιστή), η τεχνική της υποδειγματοληψίας μπορεί να είναι επωφελής.

<sup>1</sup>Ισχύει για τις Interleaved αλυσίδες που συζητήθηκαν στο προηγούμενο κεφάλαιο.

Μία άλλη τεχνική είναι η παράλληλη προσομοίωση  $l$  αλυσίδων Markov, όπου  $l \leq 10$  και η διατήρηση από κάθε αλυσίδα  $n/l$  διαδοχικών τιμών. Η απόκτηση ενός δείγματος μεγέθους  $n$  απαιτεί την προσομοίωση  $l[m + (n/l)] = lm + n$  τιμών. Επίσης, αντι να διατηρήσουμε  $n/l$  τιμές, από κάθε αλυσίδα, μπορούμε να εφαρμόσουμε την τεχνική της υποδειγματοληψίας, διατηρώντας μία τιμή κάθε  $k$ , επί συνόλου  $l[m + (n/l)k] = lm + kn$  προσομοιωμένων τιμών.

Οι υποδείξεις της βιβλιογραφίας είναι αντικρουόμενες μεταξύ τους. Οι Gelfand και Smith (1990) υποστήριζαν την ιδέα πολλών μικρών αλυσίδων, οι Gelman και Rubin (1992) αρκετές μεγάλες αλυσίδες, ενώ ο Geyer (1992) προσανατολίζεται σε μία πολύ μεγάλη αλυσίδα. Οι οπαδοί της σχολής των πολλών μικρών αλυσίδων ισχυρίζονται ότι η σύγκριση μεταξύ πολλών αλυσίδων που φαινομενικά έχουν συγκλίνει, καταλήγει στην επισήμανση διαφορών μεταξύ τους, γεγονός που σημαίνει ότι οι αλυσίδες δεν έχουν συγκλίνει ακόμα πραγματικά. Από την άλλη μεριά, οι οπαδοί της σχολής της μιας πολύ μεγάλης αλυσίδας, υπεραμύνονται της θέσης ότι μόνο έτσι διαβαίνεται ολόκληρος ο χώρος των καταστάσεων της αλυσίδας και ότι η σύγκριση μεταξύ πολλών μικρών αλυσίδων δεν μπορεί να αποδείξει τη σύγκλιση.

Σήμερα είναι καθομολογούμενο ότι η προσομοίωση πολλών μικρών αλυσίδων, παρακινήμενη από την επιθυμία μας να αποκτήσουμε ένα ανεξάρτητο δείγμα από την κατανομή  $\pi$ , είναι μια λάθος κατεύθυνση, εκτός και αν υπάρχει ειδικός λόγος για τον οποίο χρειαζόμαστε ανεξάρτητο δείγμα. Αν το ενδιαφέρον μας επικεντρώνεται όχι στην ανεξαρτησία αλλά στην εξερεύνηση των ιδιοτήτων της κατανομής στόχου  $\pi$ , τότε δεν υφίσταται η ανάγκη της γέννησης  $n$  ανεξάρτητων αλυσίδων  $\theta_k$ ,  $k = 1, \dots, n$  και διατήρησης μόνο της τελευταίας τιμής  $\theta_k^{(T)}$ . Με άλλα λόγια, μία και μόνο, αλλά μεγάλη, αλυσίδα Markov αρκεί για να εξασφαλίσει προσεγγίσεις ποσοτήτων της μορφής  $E_\pi[h(\theta)]$  μέσω εκτιμητών της μορφής (5.1). Είναι προφανές ότι ο χειρισμός τέτοιων ακολουθιών είναι δύσκολοτερος από την περίπτωση ενός iid δείγματος, εξαιτίας της εξάρτησης που υπάρχει στην αλυσίδα.

## 5.2 Η Burn In Περίοδος του Δείγματος

Ένα σπουδαίο ζήτημα κατά την ανάπτυξη ενός MCMC αλγόριθμου είναι ο καθορισμός της burn in περιόδου του δείγματος. Κατά την περίοδο αυτή, η αλυσίδα που προσομοιώνεται δεν έχει ακόμα συγκλίνει και έτσι οι τιμές που παράγονται δεν θεωρούνται ότι προέρχονται από την κατανομή στόχο  $\pi$ . Όσο πιο αργός είναι ο ρυθμός σύγκλισης της αλυσίδας, τόσο πιο προσεκτικοί πρέπει

να είμαστε ως προς την απόφασή μας για το μήκος της burn in περιόδου. Ακόμα και αν ξεκινήσουμε την αλυσίδα από την περιοχή της κορυφής της κατανομής στόχου (του σημείου με τη μεγαλύτερη πυκνότητα), ο καθορισμός της burn in περιόδου χρειάζεται προσοχή, αφού θα χρειαστεί χρόνος ώστε η αλυσίδα να «ξεχάσει» το αρχικό σημείο της και να εξερευνήσει ολόκληρο το στήριγμα της κατανομής στόχου  $\pi$ .

Το μήκος της burn in περιόδου εξαρτάται από το αρχικό σημείο  $\theta^{(0)}$ , το ρυθμό σύγκλισης του πυρήνα μετάβασης  $p(\theta, \eta)$  στη στάσιμη κατανομή στόχο  $\pi(\cdot)$  και στην ομοιότητα που απαιτείται να υπάρχει μεταξύ τους. Θεωρητικά, έχοντας διευκρινήσει τον όρο «ομοιότητα», μπορούμε να προσδιορίσουμε το μήκος της burn in περιόδου αναλυτικά. Πρακτικά όμως, κάτι τέτοιο είναι δύσκολο. Η οπτική εξέταση του διαγράμματος της προσομοιωμένης αλυσίδας (trace plot) είναι η πιο διαδεδομένη μέθοδος καθορισμού της burn in περιόδου. Πέρα από τη γραφική ανίχνευση, υπάρχουν και πιο τυποποιημένες μέθοδοι που λέγονται διαγνωστικά κριτήρια σύγκλισης.

Ο Geyer (1992) ισχυρίζεται ότι ο υπολογισμός της burn in περιόδου δεν είναι αναγκαίος, αφού το μήκος της θα είναι μικρότερο του 1% του μήκους ολόκληρης της αλυσίδας που χρειάζεται για την επίτευξη συγκεκριμένης ακρίβειας του εκτιμητή  $(1/T) \sum_{t=1}^T h(\theta^{(t)})$ . Αν αποφεύγονται τα ακραία σημεία εκκίνησης της αλυσίδας, ο Geyer (1992) προτείνει να θέτουμε το μήκος της burn in περιόδου μεταξύ του 1% και 2% του συνολικού μήκους της αλυσίδας.

### 5.3 Κατασκευή Εκτιμητών

Ύστερα από την ανάπτυξη ενός MCMC αλγόριθμου έχουμε στη διάθεσή μας το δείγμα  $(\theta^{(t)})$ ,  $t = 1, \dots, T$ , όπου  $\theta = (\theta_1, \dots, \theta_d)$ . Το δείγμα υποθέτουμε ότι προέρχεται από τη στάσιμη κατανομή στόχο  $\pi$ . Ένα δείγμα για την  $i$  συντεταγμένη του διανύσματος  $\theta$  είναι το  $(\theta_i^{(1)}, \dots, \theta_i^{(T)})$ . Μπορούμε να εκτιμήσουμε την εκ των υστέρων αναμενόμενη τιμή  $E_\pi[h(\theta)]$  ή  $E_\pi[h(\theta_i)]$  οποιασδήποτε συνάρτησης  $h$  του διανύσματος  $\theta$  ή της  $i$  συντεταγμένης του αντίστοιχα (ας θεωρήσουμε την πρώτη περίπτωση). Ένας εύκολος και άμεσος εκτιμητής είναι ο

$$\hat{E}_\pi[h(\theta)] = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}). \quad (5.2)$$

Η εκ των υστέρων διακύμανση της συνάρτησης  $h$  εκτιμάται με παρόμοιο τρόπο, παρατηρώντας ότι  $\sigma_h^2 = \text{Var}(h(\theta)) = E_\pi[h^2] - (E_\pi[h])^2$ . Κάθε αναμενόμενη

τιμή εκτιμάται με εφαρμογή του τύπου (5.2) και έτσι η διακύμανση  $\sigma_h^2$  εκτιμάται από τη δειγματική διακύμανση

$$\begin{aligned}\hat{\sigma}_h^2 &= \hat{E}_\pi [h^2] - \left(\hat{E}_\pi [h]\right)^2 \\ &= \frac{1}{T-1} \sum_{t=1}^T \left\{ h(\theta^{(t)}) - \hat{E}_\pi [h(\theta)] \right\}^2\end{aligned}\quad (5.3)$$

Ο παρανομαστής  $T-1$  μπορεί να αντικατασταθεί από το  $T$ , για δύο λόγους: το  $T$  είναι αρκετά μεγάλο ώστε η διαφορά είναι αμελητέα και επιπλέον η χρήση του παρανομαστή  $T-1$  δεν άρει τη μεροληψία του εκτιμητή, όπως συμβαίνει με την περίπτωση ενός ανεξάρτητου δείγματος.

Για τη διακύμανση του εκτιμητή  $\hat{E}_\pi [h(\theta)]$ , ακολουθώντας την πεπατημένη ενός iid δείγματος, έχουμε

$$\begin{aligned}\hat{Var}_{\text{iid}} \left( \hat{E}_\pi [h(\theta)] \right) &= \frac{\hat{\sigma}_h^2}{T} \\ &= \frac{1}{T(T-1)} \sum_{t=1}^T \left\{ h(\theta^{(t)}) - \hat{E}_\pi [h(\theta)] \right\}^2\end{aligned}\quad (5.4)$$

Ο εκτιμητής αυτός είναι εύκολος στον υπολογισμό, όμως τείνει να υποεκτιμά την ποσότητα  $Var \left( \hat{E}_\pi [h(\theta)] \right)$  λόγω των θετικών αυτοσυσχετίσεων που υπάρχουν σε ένα MCMC δείγμα. Το πρόβλημα αυτό μπορεί να διευθετηθεί είτε με τη χρήση παράλληλων αλυσίδων είτε με υποδειγματοληψία (batch sampling).

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε την ιδέα του **Αποτελεσματικού Μεγέθους Δείγματος (Effective Sample Size, ESS)**. Το ESS ορίζεται από τη σχέση

$$ESS = \frac{T}{\kappa(\theta)},\quad (5.5)$$

όπου  $\kappa(\theta)$  είναι ο **ολοκληρωμένος χρόνος αυτοσυσχέτισης (integrated autocorrelation time)** για την παράμετρο  $\theta$  και ορίζεται από τον τύπο

$$\kappa(\theta) = 1 + 2 \sum_{\nu=1}^{\infty} \rho_\nu(\theta),\quad (5.6)$$

με  $\rho_\nu(\theta)$  την αυτοσυσχέτιση υστέρησης  $\nu$ , για την ακολουθία  $(\theta^{(t)})$ . Οι αυτοσυσχετίσεις  $\rho_\nu(\theta)$ ,  $\nu = 1, 2, \dots$  εκτιμώνται από τις δειγματικές αυτοσυσχετίσεις<sup>2</sup>  $r_\nu(\theta)$ . Για την εκτίμηση του χρόνου αυτοσυσχέτισης  $\kappa(\theta)$ , χρησιμοποιούμε τον εκτιμητή  $\hat{\kappa}(\theta) = 1 + 2 \sum_{\nu=1}^k r_\nu(\theta)$  με  $k$  τέτοιο ώστε  $r_\nu(\theta) \leq 0.1$  για  $\nu > k$ . Τώρα, ο εκτιμητής της διακύμανσης  $Var(\hat{E}_\pi[h(\theta)])$  δίνεται από τον τύπο

$$\begin{aligned} \hat{Var}_{\text{ESS}}(\hat{E}_\pi[h(\theta)]) &= \frac{\hat{\sigma}_h^2}{\text{ESS}(\theta)} \\ &= \frac{\kappa(\theta)}{T(T-1)} \sum_{t=1}^T \left\{ h(\theta^{(t)}) - \hat{E}_\pi[h(\theta)] \right\}^2. \end{aligned} \quad (5.7)$$

Δείτε ότι  $\kappa(\theta) > 1$  και επομένως  $\text{ESS}(\theta) < T$  και  $\hat{Var}_{\text{ESS}}(\hat{E}_\pi[h(\theta)]) > \hat{Var}_{\text{iid}}(\hat{E}_\pi[h(\theta)])$ , εκτός αν οι παρατηρήσεις  $(\theta^{(t)})$ ,  $t = 1, \dots, T$  είναι ασυσχέτιστες, οπότε ευσταθεί η ισότητα.

Μια απλούστερη μέθοδος εκτίμησης της διακύμανσης  $Var(\hat{E}_\pi[h(\theta)])$  προκύπτει με τη χρήση της ακόλουθης τεχνικής: διαιρούμε το αρχικό δείγμα, μεγέθους  $T$ , σε  $m$  διαδοχικά τμήματα, μήκους  $k$ , δηλαδή  $T = mk$ . Η δειγματική μέση τιμή κάθε τμήματος είναι  $B_i$ ,  $i = 1, \dots, m$ . Επομένως, συνολικά

$$\hat{E}_\pi[h(\theta)] = \bar{B} = \frac{1}{m} \sum_{i=1}^m B_i. \quad (5.8)$$

Τότε, έχουμε τον εκτιμητή

$$\hat{Var}_{\text{batch}}(\hat{E}_\pi[h(\theta)]) = \frac{1}{m(m-1)} \sum_{i=1}^m \left( B_i - \hat{E}_\pi[h(\theta)] \right)^2, \quad (5.9)$$

με την προϋπόθεση ότι το  $k$  είναι αρκετά μεγάλο ώστε οι αυτοσυσχετίσεις μεταξύ των  $m$  τμημάτων να είναι αμελητέες και το  $m$  επίσης μεγάλο, ώστε οι εκτιμητές  $\hat{Var}(B_i)$  να είναι αξιόπιστοι. Είναι σημαντικό να επαληθευτεί

<sup>2</sup>Για μία ακολουθία τυχαίων μεταβλητών  $X_i$ ,  $i = 1, \dots, n$ , η δειγματική αυτοσυσχέτιση υστέρησης  $k$  δίνεται από τον τύπο

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$



ότι οι μέσοι  $B_i$ ,  $i = 1, \dots, m$  είναι σχεδόν ανεξάρτητοι, ελέγχοντας αν η αυτοσυσχέτισή τους, υστέρησης (lag) 1, είναι μικρότερη από 0.1. Στην αντίθετη περίπτωση, πρέπει να αυξήσουμε το  $k$  και άρα το  $T$ , εκτός αν το  $m$  είναι αρκετά μεγάλο.

Ανεξάρτητα από τον εκτιμητή που επιλέγουμε για την εκτίμηση της διακύμανσης  $Var(\hat{E}_\pi[h(\theta)])$ , ένα 95% διάστημα εμπιστοσύνης για την παράμετρο  $E_\pi[h(\theta)]$ , είναι το

$$\hat{E}_\pi[h(\theta)] \pm z_{0,025} \sqrt{\hat{V}}, \quad (5.10)$$

όπου  $z_{0,025} = 1.96$  το κάτω 0.025 ποσοστιαίο σημείο της τυπικής κανονικής κατανομής. Αν η μέθοδος αυτή χρησιμοποιείται με λιγότερα από 30 τμήματα διαχωρισμού του αρχικού δείγματος, τότε είναι καλή η ιδέα να αντικαταστήσουμε το  $z_{0,025}$  με  $t_{(m-1);0,025}$  το κάτω 0.025 ποσοστιαίο σημείο της κατανομής  $t$  με  $m - 1$  βαθμούς ελευθερίας.

Οι περιθωριακές κατανομές  $\pi^{(i)}(\theta_i)$ ,  $i = 1, \dots, d$ , εκτιμώνται από τα ιστογράμματα των προσομοιωμένων τιμών  $\theta_i^{(t)}$ . Καλύτεροι εκτιμητές είναι οι Rao - Blackwellized

$$\hat{\pi}^{(i)}(\theta_i) = \frac{1}{T} \sum_{t=1}^T \pi_i(\theta_i | \theta_{-i}^{(t)}), \quad (5.11)$$

όπου  $\theta_{-i}^{(t)} = (\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})$ . Το Εργοδικό Θεώρημα εξασφαλίζει τη συνέπεια του εκτιμητή  $\hat{\pi}^{(i)}(\theta_i)$ . Η ίδια ιδέα μπορεί να εφαρμοστεί για την εκτίμηση των ροπών της συνάρτησης  $h(\theta_i)$ :

$$\hat{E}(h(\theta_i)) = \frac{1}{T} \sum_{t=1}^T E(h(\theta_i) | \theta_{-i}^{(t)}). \quad (5.12)$$

## 5.4 Αυτοσυσχέτιση

Ας υποθέσουμε ότι επιθυμούμε να προσομοιώσουμε την εκ των υστέρων κατανομή  $\pi(\theta)$  του διανύσματος  $\theta = (\theta_1, \dots, \theta_d)$ . Η υψηλή συσχέτιση (correlation) μεταξύ των συντεταγμένων  $\theta_i, \theta_j$ ,  $i \neq j$ , τείνει να παρακωλύσει την ταχύτητα με την οποία συγκλίνει η αλυσίδα  $(\theta^{(t)})$  στη στάσιμη κατανομή της. Ωστόσο, πρέπει να ληφθεί υπόψη και η αυτοσυσχέτιση (autocorrelation) μεταξύ των

τυχαίων μεταβλητών  $\theta_i^{(t)}, \theta_i^{(s)}$ ,  $t \neq s$  που αναπαριστούν την αλγοριθμική εξέλιξη μιας συντεταγμένης  $\theta_i$ . Η υψηλή αυτοσυσχέτιση προκαλεί αργή κίνηση της αλυσίδας μέσα στο χώρο των καταστάσεων της και πιθανή μη σύγκλιση στην οριακή κατανομή, καθώς η αλυσίδα τείνει να «εξερευνήσει» μικρότερο μέρος του στηρίγματος της κατανομής στόχου  $\pi$ . Μελετώντας τις αυτοσυσχετίσεις μιας αλυσίδας Markov, είναι χρήσιμο να αναγνωρίσουμε τις μεγαλύτερες υστερήσεις (lags) της αλυσίδας για τις οποίες η παρουσία της αυτοσυσχέτισης είναι έντονη, καθώς και το αν οι αυτοσυσχετίσεις είναι φθίνουσες ως προς τη χρονική υστέρηση. Συνήθως δεν χρειάζεται να επεκταθούμε στη μελέτη των αυτοσυσχετίσεων χρονικής υστέρησης μεγαλύτερης από 50.

## 5.5 Η Σύγκλιση των Αλυσίδων Markov

Τα αποτελέσματα που προκύπτουν από την ανάπτυξη ενός MCMC αλγόριθμου δεν θεωρούνται αξιόπιστα αν η προσομοιωμένη αλυσίδα Markov ( $\theta^{(t)}$ ) δεν έχει προσεγγίσει τη στάσιμη κατανομή στόχο  $\pi(\cdot)$  και δεν έχει εξερευνήσει το μεγαλύτερο μέρος του στηρίγματος της  $\pi$ . Θεωρητικά, ισχύει  $\theta^{(t)} \sim \pi(\cdot)$  μόνο όταν ο αριθμός των επαναλήψεων  $t$  του αλγόριθμου τείνει στο άπειρο. Επειδή κάτι τέτοιο δεν είναι εφικτό, στην πραγματικότητα αποκτάμε μια παρατήρηση από μια κατανομή που προσεγγίζει την κατανομή στόχο, σε κάποια επαρκώς μεγάλη επανάληψη του αλγόριθμου. Το ερώτημα είναι το πόσο μεγάλος πρέπει να είναι ο αύξων αριθμός της επανάληψης του αλγόριθμου, ώστε η προσέγγιση αυτή να είναι ικανοποιητική και να μπορούμε να βασιστούμε στο Κεντρικό Οριακό Θεώρημα για περεταίρω συμπερασματολογία. Για το λόγο αυτό, ο πιο μεγάλος κίνδυνος που αντιμετωπίζει ο ερευνητής είναι να συμπεράνει σύγκλιση πριν αυτή πραγματοποιηθεί.

Υπάρχουν τρία βασικά συμπεράσματα που αφορούν την πρακτική εφαρμογή ενός MCMC αλγόριθμου:

1. Δεν υπάρχει γενική μεθοδολογία για τη διαπίστωση της σύγκλισης του δειγματολήπτη Gibbs και του γενικότερου Metropolis - Hastings αλγόριθμου.
2. Δεν υπάρχει τρόπος ώστε να εγγυηθούμε την καθολική εξερεύνηση του στηρίγματος της κατανομής στόχου  $\pi$ , από την παραχθείσα αλυσίδα ( $\theta^{(t)}$ ).
3. Δεν υπάρχει, γενικά, εγγύηση ότι για κάποια συγκεκριμένη χρονική στιγμή

μή  $t$ , η αλυσίδα κατανέμεται σύμφωνα με την επιζητούμενη στάσιμη κατανομή στόχο, δηλαδή  $\theta^{(t)} \sim \pi(\cdot)$ .

Τα άσχημα αυτά νέα μετριάζονται από ένα πολύ σημαντικό θετικό αποτέλεσμα: όλες οι εργοδικές αλυσίδες Markov, εγγυημένα, συγκλίνουν ασυμπτωτικά (Chan 1993, Polson 1995) και είναι σχετικά εύκολο να διαπιστωθεί η μη σύγκλιση.

Υπάρχουν τρεις κύριες προσεγγίσεις για τη διαπίστωση της σύγκλισης των Markovιανών αλυσίδων: η πρώτη βασίζεται στις θεωρητικές ιδιότητες μιας αλυσίδας και προσπαθεί να μετρήσει αποστάσεις και να θέσει φράγματα στις συναρτήσεις κατανομής που συνδέονται με την αλυσίδα. Συγκεκριμένα, κάποιος μπορεί να μετρήσει την απόσταση ολικής μεταβολής (total variation distance) μεταξύ της κατανομής της αλυσίδας στην επανάληψη  $t$  και της οριακής κατανομής  $\pi$ . Η δεύτερη προσέγγιση βασίζεται στην παρακολούθηση κάποιων περιγραφικών στατιστικών μέτρων κατά την ανάπτυξη του αλγόριθμου και η τρίτη στην εξολοκλήρου αποφυγή του θέματος με την εφαρμογή τέλειαις δειγματοληψίας, η οποία εκμεταλεύεται την ιδέα της σύζευξης (coupling) ώστε να παράγει δείγματα από την ακριβή κατανομή στόχο (Propp και Wilson, 1996).

Στο κεφάλαιο αυτό θα δώσουμε έμφαση στη δεύτερη προσέγγιση, που περιλαμβάνει τα **διαγνωστικά κριτήρια σύγκλισης (convergence diagnostics)**. Πρόκειται για την παρακολούθηση (monitoring) της ανάπτυξης του αλγόριθμου και τον υπολογισμό κάποιων στατιστικών ποσοτήτων που μας βοηθούν να αποφασίσουμε, υποκειμενικά πάντα, πότε να σταματήσουμε τη διαδικασία της προσομοίωσης. Είναι άμεσης προτεραιότητας να επισημάνουμε ότι τα διαγνωστικά κριτήρια είναι δείκτες μη σύγκλισης. Βεβαίως, αυτό δε σημαίνει ότι η έλλειψη ένδειξης μη σύγκλισης σημαίνει σύγκλιση. Ο έμπειρος ερευνητής πρέπει να αντιμετωπίζει τα αποτελέσματα ενός διαγνωστικού κριτηρίου με σκεπτικισμό. Συνίσταται, ή καλύτερα επιβάλλεται, η χρήση περισσότερων του ενός διαγνωστικών κριτηρίων σύγκλισης.

Πριν προχωρήσουμε στη μελέτη των τεχνικών ανίχνευσης της σύγκλισης, αξίζει να αναφέρουμε τρεις τύπους σύγκλισης που μας ενδιαφέρουν:

1. **Σύγκλιση στη στάσιμη κατανομή.** Πρόκειται για τη σύγκλιση της αλυσίδας ( $\theta^{(t)}$ ) στην στάσιμη κατανομή στόχο  $\pi$ . Είναι δύσκολο να ανιχνευτεί διότι η στάσιμη κατανομή στόχος δεν είναι άλλη από την οριακή κατανομή της αλυσίδας. Υπό αυτή την έννοια, η στασιμότητα επιτυγχάνεται μόνο ασυμπτωτικά. Ωστόσο, η διαδικασία της προσομοίωσης σταματάει σε κάποιο βήμα  $T$ . Στο σημείο εκείνο η αλυσίδα κατανέμεται σύμφωνα με την κατανομή  $\pi_\mu^T$ , όπου  $\mu$  είναι η αρχική κατανομή της.

2. **Σύγκλιση των εργοδικών μέσων.** Εδώ ενδιαφερόμαστε για τη σύγκλιση ποσοτήτων της μορφής

$$\frac{1}{T} \sum_{t=1}^T h(\theta^{(t)})$$

στην πληθυσμιακή ποσότητα  $E_{\pi}[h(\theta)]$ , για κάποια αυθαίρετη συνάρτηση  $h$ . Αυτός ο τύπος της σύγκλισης μας βοηθάει να διαπιστώσουμε την ταχύτητα με την οποία «ταξιδεύει» η αλυσίδα στο χώρο των καταστάσεών της. Παρόλο που το Εργοδικό Θεώρημα εξασφαλίζει αυτή τη σύγκλιση, το ερώτημα αφορά τον προσδιορισμό της τιμής  $T$  για την οποία η προσέγγιση στην ποσότητα  $E_{\pi}[h(\theta)]$  είναι ικανοποιητική.

3. **Σύγκλιση στην iid δειγματοληψία.** Οι έλεγχοι που αναπτύσσονται σε αυτή την κατηγορία αφορούν το αν ένα δείγμα  $(\theta_1^{(t)}, \dots, \theta_n^{(t)})$ , που προέρχεται από την προσομοίωση  $n$  παράλληλων αλυσίδων, μπορεί να θεωρηθεί ως iid δείγμα. Αντί για την προσομοίωση  $n$  παράλληλων αλυσίδων, προτιμότερη είναι η χρησιμοποίηση της μεθόδου της υποδειγματοληψίας (subsampling or batch sampling)<sup>3</sup>, με την οποία μειώνεται η αυτοσυσχέτιση μεταξύ των όρων μιας ακολουθίας  $(\theta^{(t)})$ .

### 5.5.1 Γραφικές Τεχνικές Ανίχνευσης της Μη Σύγκλισης

Πριν προχωρήσουμε στα διαγνωστικά κριτήρια σύγκλισης είναι χρήσιμο να αναφέρουμε κάποιες γραφικές μεθόδους που δείχνουν το αν η αλυσίδα δεν έχει ακόμα αποκτήσει στάσιμη συμπεριφορά. Ωστόσο, είναι πολύ εύκολο τα γραφήματα να μας παραπλανήσουν.

Μια από τις γραφικές τεχνικές που χρησιμοποιείται είναι το trace plot. Κατά την εξέλιξη του αλγόριθμου, παρακολουθούμε γραφικά την τροχιά της παραγόμενης αλυσίδας, ύστερα από τη burn in περίοδο. Συγκρίνουμε τα ποιοτικά χαρακτηριστικά του γραφήματος σε διάφορα τμήματά του, τα οποία επιλέγονται να είναι αρκετά απομακρυσμένα ώστε να περιορίζεται η επίδραση της αυτοσυσχέτισης, η οποία πιθανώς προκαλεί ομοιότητες. Το Στατιστικό πρόγραμμα WinBUGS παρέχει τη δυνατότητα της ανάπτυξης του αλγόριθμου με την ταυτόχρονη παρακολούθηση των trace plots.

<sup>3</sup>Βλέπε και ενότητα 5.1.

Ένα γνωστό πρόβλημα με τα trace plots είναι ότι μπορεί να οδηγήσουν σε εσφαλμένες εντυπώσεις αν η αλυσίδα παραμένει προσκολλημένη σε κάποια κορυφή (mode). Στην περίπτωση αυτή, δεν υπάρχει οπτική ένδειξη ότι δεν είναι αυτή η επιθυμητή πυκνότητα στόχος. Ο Geweke (1992) προτείνει να παρατηρήσουμε την ανάπτυξη του αλγόριθμου για μεγάλο χρονικό διάστημα (αριθμό επαναλήψεων), ώστε τελικά η αλυσίδα να «δραπετεύσει» από την κορυφή στην οποία «παγιδεύτηκε». Οι Gelman και Rubin (1992) προτείνουν την εκκίνηση της αλυσίδας από πολλαπλά, ευρέως διασπαρμένα ως προς το στήριγμα της κατανομής στόχου  $\pi$ , σημεία.

Στις περισσότερες περιπτώσεις, το trace plot της ακολουθίας  $(\theta^{(t)})$ , ως προς  $t$ , δεν είναι ιδιαίτερα χρήσιμο για την ανακάλυψη της σύγκλισης. Αν μας ενδιαφέρουν ποσότητες της μορφής  $E_{\pi}[h(\theta)]$ , τότε ένας πιο χρήσιμος δείκτης είναι η μελέτη της συμπεριφοράς του εργοδικού μέσου

$$S_T = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) \quad (5.13)$$

σε όρους του  $T$ . Απαραίτητη προϋπόθεση για τη σύγκλιση είναι η στασιμότητα της ακολουθίας  $S_T$ , παρόλο που η ένδειξη τέτοιας στασιμότητας μπορεί να οφείλεται στην επιρροή κάποιας κορυφής της κατανομής στόχου  $\pi$ .

Ο Robert (1995) προτείνει μια πιο σθεναρή μέθοδο γραφικής ανίχνευσης της σύγκλισης μιας αλυσίδας Markov: παρακολούθηση της συμπεριφοράς διάφορων εκτιμητών της ποσότητας  $E_{\pi}[h(\theta)]$ , οι οποίοι βασίζονται στην ίδια αλυσίδα  $(\theta^{(t)})$ . Οι εκτιμήσεις συγκλίνουν μέχρι που συμπίπτουν<sup>4</sup>, όταν έχει επέλθει η σύγκλιση.

Πέρα από τον εργοδικό μέσο του τύπου (5.13), χρησιμοποιείται επίσης η δεσμευμένη ή Rao Blackwellized εκδοχή του

$$S_T^C = \frac{1}{T} \sum_{t=1}^T E[h(\theta) | \eta^{(t)}], \quad (5.14)$$

σε περίπτωση που ο αλγόριθμος παράγει μια αλυσίδα  $(\eta^{(t)}, \theta^{(t)})$ .

Μια άλλη τεχνική που δίνει ένα συγκλίνοντα μέσο είναι η δειγματοληψία σπουδαιότητας (importance sampling). Αν είναι γνωστή η μορφή της κατανομής στόχου  $\pi$ , τότε ένας εκτιμητής είναι ο

$$S_T^P = \sum_{t=1}^T w_t h(\theta^{(t)}), \quad (5.15)$$

<sup>4</sup>Με κάποια δεδομένη ακρίβεια.

όπου  $w_t \propto \pi(\theta^{(t)}) / g_t(\theta^{(t)})$  και  $g_t$  είναι η πυκνότητα που χρησιμοποιείται για την προσομοίωση των τιμών  $\theta^{(t)}$ . Συγκεκριμένα, για την περίπτωση του δειγματολήπτη Gibbs,

$$g_t(\theta^{(t)}) = \prod_{i=1}^d \pi_i(\theta_i^{(t)} | \theta_{-i}^{(t-1)}). \quad (5.16)$$

Για την περίπτωση του αλγόριθμου Metropolis - Hastings, οι μεταβλητές που πραγματικά προσομοιώνονται είναι οι  $\eta^{(t)} \sim q(\theta^{(t-1)}, \cdot)$  και ο εκτιμητής (5.15) παίρνει τη μορφή

$$S_T^{MP} = \sum_{t=1}^T w_t h(\eta^{(t)}), \quad (5.17)$$

όπου  $w_t \propto \pi(\eta^{(t)}) / q(\theta^{(t-1)}, \eta^{(t)})$ .

Μια ενδιαφέρουσα ιδιότητα της δειγματοληψίας σπουδαιότητας είναι ότι αφαιρεί την αυτοσυσχέτιση μεταξύ των  $\theta^{(t)}$  (ή των  $\eta^{(t)}$ ):

**Λήμμα 5.5.1** Έστω  $(\theta^{(t)})$  μια αλυσίδα Markou που παράγεται από κάποιο Metropolis - Hastings αλγόριθμο με πυρήνα μετάβασης πρότασης  $q$ . Τότε

$$\text{Var} \left( \sum_{t=1}^T h(\theta^{(t)}) \frac{\pi(\theta^{(t)})}{q(\theta^{(t-1)}, \theta^{(t)})} \right) = \sum_{t=1}^T \text{Var} \left( h(\theta^{(t)}) \frac{\pi(\theta^{(t)})}{q(\theta^{(t-1)}, \theta^{(t)})} \right) \quad (5.18)$$

με την προϋπόθεση ότι αυτές οι ποσότητες είναι καλά ορισμένες.

Οι σταθμίσεις  $w_t$  είναι γνωστές ως προς κάποιο πολλαπλάσιό τους και για το λόγο αυτό τυποποιούνται πολλαπλασιαζόμενες με την ποσότητα  $\left(\sum_{t=1}^T w_t\right)^{-1}$ . Μπορεί όμως να υποθεθεί ότι η επίδραση αυτής της τυποποίησης, στις αυτοσυσχετίσεις των  $\theta^{(t)}$ , εξαφανίζεται καθώς το  $T$  αυξάνει.

Το Λήμμα 5.5.1 δηλώνει ότι οι εκτιμητές  $S_T^P$  και  $S_T^{MP}$  συμπεριφέρονται ως να επικρατούσαν συνθήκες ανεξαρτησίας και επομένως μπορούμε να ελέγξουμε πιο παραδοσιακά τη σύγκλιση της αλυσίδας  $(\theta^{(t)})$ . Επιπλέον, υπονοεί ότι η διακύμανση του  $S_T^P$  (ή του  $S_T^{MP}$ ) ελαττώνεται με ρυθμό  $1/T$ , υπό συνθήκες στασιμότητας. Επομένως, η μη στασιμότητα μπορεί να ανιχνευτεί αν η ελάττωση αυτή δεν προσαρμόζεται σε μία παραβολή τάξης  $1/\sqrt{T}$ . Σημειώστε ακόμα

ότι η πυκνότητα  $\pi$  που εμφανίζεται στις σταθμίσεις  $w_t$ , μπορεί να αντικατασταθεί από την εκτίμηση

$$\hat{\pi}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \pi_1(\theta | \eta^{(t)}), \quad (5.19)$$

στις περιπτώσεις όπου ο Rao - Blackwellized εκτιμητής βρίσκει εφαρμογή.

Οι Gelfand και Smith (1990) πρότειναν την ακόλουθη γραφική ανιχνευτική διαδικασία: προσομοιώνουν  $n$  παράλληλες αλυσίδες και ύστερα από  $m$  επαναλήψεις (burn in period) συλλέγουν τις  $n$ , παραχθείσες κατά το  $m + 1$  βήμα του αλγόριθμου, τιμές των  $n$  αλυσίδων και σχεδιάζουν το ιστογράμμα των συχνοτήτων τους ή κάποιας συνάρτησης αυτών. Η συνάρτηση αυτή μπορεί να είναι μία συντεταγμένη του διανύσματος  $\theta$ , δηλαδή από το χειρισμό των  $n$   $d$ -μεταβλητών<sup>5</sup> τιμών καταλήγουμε στο χειρισμό  $n$  μονομεταβλητών τιμών. Επαναλαμβάνουν την ίδια διαδικασία έπειτα από  $k$  επαναλήψεις του αλγόριθμου και συγκρίνουν τα δύο ιστογράμματα. Η έλλειψη μη σύγκλισης υιοθετείται όταν τα ιστογράμματα είναι μη διακεκριμένα. Το  $k$  επιλέγεται με γνώμονα την αφαίρεση της επίδρασης των αυτοσυσχετίσεων, οι οποίες δημιουργούν παραπλανητικές ομοιότητες μεταξύ των ιστογραμμάτων. Τυπικά, τιμές μεταξύ 10 και 50 κρίνονται λογικές.

Η ύπαρξη πολλών κορυφών (multimodality) στο διάγραμμα της εκτιμημένης συνάρτησης πυκνότητας, για το  $t$  βήμα του αλγόριθμου, είναι κλασσικό σημείο ένδειξης μη σύγκλισης.

Οι Gelman, Gilks και Roberts (1996) δίνουν ένα θεωρητικό αποτέλεσμα: ο Metropolis - Hastings αλγόριθμος με χαμηλά ποσοστά αποδοχής των προτεινόμενων τιμών είναι ασταθής και φτωχός στις μεταβάσεις του ως προς το εύρος του στηρίγματος της πυκνότητας στόχου  $\pi$ . Επομένως, ένα σειριακό διάγραμμα (serial plot) των ποσοστών αποδοχής, ως προς το χρόνο, είναι ενδεικτικό της εξέλιξης του αλγόριθμου.

Οι Yu και Mykland (1998) προτείνουν μία αμιγώς γραφική εκτίμηση της σύγκλισης του εργοδικού μέσου (5.1) στον πλυθησμιακό μέσο  $E_\pi[h(\theta)]$ . Η μέθοδος αυτή βασίζεται στα διαγράμματα των συσσωρευμένων αθροισμάτων CUSUM. Διαγραμματοποιούν τις μερικές διαφορές

$$D_T^i = \sum_{t=1}^i (h(\theta^{(t)}) - S_T), \quad i = 1, \dots, T \quad (5.20)$$

<sup>5</sup>Όσες είναι οι συντεταγμένες του διανύσματος  $\theta$ .

όπου

$$S_T = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}).$$

Οι Yu και Mykland αποκομίζουν μια ποιοτική εκτίμηση, τόσο για την ικανότητα της αλυσίδας να διέρχεται γρήγορα από ολόκληρο το χώρο των καταστάσεων της, όσο και για τη συσχέτιση μεταξύ των τιμών  $\theta^{(t)}$ : όταν η αλυσίδα διαβαίνει γρήγορα σε ολόκληρο το στήριγμα της κατανομής στόχου  $\pi$ , τότε το διάγραμμα των  $D_T^2$  είναι έντονα άτακτο και συγκεντρωμένο γύρω από το μηδέν. Στην αντίθετη περίπτωση το διάγραμμα είναι τακτικό, με μεγάλα «πετάγματα» μακριά από το μηδέν.

Το σημαντικότερο μειονέκτημα της μεθόδου είναι ότι βασίζεται σε μία μόνο αλυσίδα, με τα πλεονεκτήματα και τα μειονεκτήματα που αναφέραμε στην ενότητα 5.1.

## 5.6 Διαγνωστικά Κριτήρια Σύγκλισης

Όπως ήδη αναφέρθηκε, τα διαγνωστικά κριτήρια σύγκλισης είναι μέθοδοι που βασίζονται στις στατιστικές ιδιότητες των παρατηρηθεισών τιμών της αλυσίδας προσπαθώντας να ανιχνεύσουν ενδείξεις έλλειψης μη σύγκλισης. Υπάρχουν πολλές μέθοδοι που αναφέρονται στη βιβλιογραφία. Οι περισσότερες από αυτές συνοψίζονται στις δημοσιεύσεις των Brooks και Roberts (1995) και Cowles και Carlin (1996).

### 5.6.1 Το διαγνωστικό Κριτήριο του Geweke

Ο Geweke (1992) ανέπτυξε ένα κριτήριο ανίχνευσης της σύγκλισης, βασιζόμενος στη σύγκριση μεταξύ δύο μη αλληλοεπικαλυπτόμενων (overlapping) αναλογιών του δείγματος, αναφερόμενων σε δύο διαφορετικές περιόδους της εξέλιξης του αλγόριθμου, μετά τη burn in περίοδο.

Θεωρήστε μια πραγματική συνάρτηση  $\psi = h(\theta)$  και την τροχιά (trajectory) της  $\psi^{(1)}, \psi^{(2)}, \dots$ , όπου  $\psi^{(j)} = h(\theta^{(j)})$ ,  $j = 1, 2, \dots$ . Η τροχιά αυτή ορίζει μια χρονοσειρά και επομένως μπορούμε να υπολογίσουμε εργοδικούς μέσους. Ο Geweke (1992) προτείνει τη χρήση στατιστικών ελέγχων στους εργοδικούς μέσους ώστε να επιβεβαιωθεί η σύγκλιση.

Ας υποθέσουμε ότι αφήνουμε τον αλγόριθμο να εξελιχθεί και λαμβάνουμε



$m + n$  παρατηρήσεις της αλυσίδας. Σχηματίζουμε τους εργοδικούς μέσους

$$\bar{\psi}_1 = \frac{1}{n_1} \sum_{j=m+1}^{m+n_1} \psi^{(j)} \quad (5.21)$$

$$\bar{\psi}_2 = \frac{1}{n_2} \sum_{j=m+n-n_2+1}^{m+n} \psi^{(j)}, \quad (5.22)$$

όπου  $n_1 + n_2 < n$ . Αν  $m$  είναι το μήκος της burn in περιόδου, τότε  $\bar{\psi}_1$  και  $\bar{\psi}_2$  είναι οι εργοδικοί μέσοι στην αρχή και στο τέλος της περιόδου σύγκλισης και επομένως πρέπει να συμπεριφέρονται παρομοίως. Καθώς το  $n$  τείνει στο άπειρο και οι αναλογίες  $n_1/n$  και  $n_2/n$  παραμένουν σταθερές, τότε το Κεντρικό Οριακό Θεώρημα εφαρμόζεται και

$$z_G = \frac{\bar{\psi}_2 - \bar{\psi}_1}{\sqrt{\hat{V}ar(\psi_2) + \hat{V}ar(\psi_1)}} \xrightarrow{d} N(0, 1), \quad (5.23)$$

όπου  $\hat{V}ar(\psi_k) = s_k(0)/n_k$  και  $s_k(0)$ ,  $k = 1, 2$  είναι οι συμμετρικές φασματικές συναρτήσεις πυκνότητας (symmetric spectral density functions, Chatfield 1996, Backett and Diaconis 1994), δεδομένου ότι δεν υπάρχουν ασυνέχειες στη συχνότητα μηδέν. Αν η σύγκλιση έχει επέλθει, τότε η τυποποιημένη διαφορά μεταξύ των εργοδικών μέσων, υπολογισμένων στην αρχή και στο τέλος της περιόδου σύγκλισης, δεν πρέπει να είναι μεγάλη. Προσοχή, οι μεγάλες διαφορές υπονοούν έλλειψη σύγκλισης, αλλά οι μικρές διαφορές δεν σημαίνουν σύγκλιση. Επικαλούμενοι τη θεωρία περί κανονικότητας, τιμές  $|z_G| > 2$  είναι ενδεικτικές για την έλλειψη σύγκλισης. Ο Geweke προτείνει  $n_1/n = 0.1$  και  $n_2/n = 0.5$ , ωστόσο δεν είναι απαραίτητο να προσκολλούμαστε στις συστάσεις αυτές.

Η προσέγγιση του Geweke είναι, καταφανώς, μια προσέγγιση βασισμένη σε χρονοσειρές και αντανακλά την πεποίθησή του ότι η επεξεργασία μιας πολύ μεγάλης αλυσίδας, αντι πολλών μικρών, είναι πιο επικυδομητική, καθώς θα εξερευνήσει μεγαλύτερο μέρος του στηρίγματος της κατανομής στόχου  $\pi$ . Ένα μειονέκτημα του διαγνωστικού κριτηρίου του Geweke είναι ότι η τιμή του  $z_G$  επηρεάζεται από τις αναλογίες  $\bar{\psi}_1$  και  $\bar{\psi}_2$ . Καλό είναι να μη βασίζεται κάποιος σε ένα μόνο έλεγχο για τη διαπίστωση της σύγκλισης ενός αλγόριθμου.

### 5.6.2 Το Διαγνωστικό Κριτήριο των Gelman και Rubin

Οι Gelman και Rubin (1992) πρότειναν μια απλή μέθοδο για την ανίχνευση της μη σύγκλισης μιας αλυσίδας Markov που παράγεται από κάποιο MCMC αλγόριθμο. Η μέθοδος αυτή βασίζεται στην παράλληλη ανάπτυξη αλυσίδων που έχουν ξεκινήσει από διαφορετικά σημεία. Η χρήση πολλών αλυσίδων περιορίζει τον κίνδυνο της παγίδευσης σε περιοχές γύρω από κορυφές, κίνδυνος ο οποίος υπάρχει στην περίπτωση της ανάπτυξης μίας αλυσίδας. Επιπλέον, είναι πιο εύκολο να ανιχνευτούν οι «σχεδόν ασταθείς» (metastable) συμπεριφορές της αλυσίδας. Παραδείγματα τέτοιων συμπεριφορών δίνονται από τους Gelman και Rubin (1992) και Gelman (1996). Ύστερα από την επίτευξη της σύγκλισης, όλες οι παράλληλες αλυσίδες πρέπει να έχουν την ίδια ποσοτική και ποιοτική συμπεριφορά. Το κριτήριο βασίζεται στον έλεγχο της ισότητας μεταξύ των διασπορών εντός (within) και μεταξύ (between) των αλυσίδων. Αυτό ισοδυναμεί με την ομοιότητα μεταξύ του ιστογράμματος των προσομοιωμένων τιμών όλων των αλυσίδων και των τιμών κάθε μίας αλυσίδας ξεχωριστά.

Η διαδικασία προβλέπει την εκκίνηση των παράλληλων αλυσίδων από σημεία ευρέως διεσπαρμένα ως προς το στήριγμα της κατανομής στόχου  $\pi$ . Το πλήθος των παράλληλων αλυσίδων δεν είναι απαραίτητο να είναι μεγάλο προς αποφυγή υπολογιστικού φόρτου. Συνήθως προσομοιώνεται μονοψήφιος αριθμός αλυσίδων. Όταν υπάρχει ένδειξη ότι η κατανομή στόχος είναι πολυκόρυφη (multimodal), τότε συστήνεται η εκκίνηση τουλάχιστον μίας αλυσίδας από κάθε κορυφή.

Θεωρήστε  $m \geq 2$  παράλληλες αλυσίδες μήκους  $2n$ , οι οποίες ξεκινούν την ανάπτυξή τους από  $m$  διαφορετικά σημεία:

$$\begin{aligned} & \theta_1^{(1)}, \theta_1^{(2)}, \dots, \theta_1^{(2n-1)}, \theta_1^{(2n)} \\ & \theta_2^{(1)}, \theta_2^{(2)}, \dots, \theta_2^{(2n-1)}, \theta_2^{(2n)} \\ & \vdots \\ & \theta_m^{(1)}, \theta_m^{(2)}, \dots, \theta_m^{(2n-1)}, \theta_m^{(2n)}. \end{aligned}$$

Στη συνέχεια αγνοούμε τις πρώτες  $n$  τιμές όλων των αλυσίδων (burn in period) και υπολογίζουμε τις τιμές  $\psi_i^{(j)}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  κάποιας πραγματικής συνάρτησης ενδιαφέροντος  $\psi = h(\theta)$ . Η διακύμανση μεταξύ των αλυσίδων (between chain variance) δίνεται από τον τύπο

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{\psi}_i - \bar{\psi})^2 \quad (5.24)$$

και η διακύμανση εντός των αλυσίδων (within chain variance) από τον τύπο

$$W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\psi_i^{(j)} - \bar{\psi}_i)^2, \quad (5.25)$$

όπου  $\bar{\psi}_i = \frac{1}{n} \sum_{j=1}^n \psi_i^{(j)}$  και  $\bar{\psi} = \frac{1}{m} \sum_{i=1}^m \bar{\psi}_i$ . Ύστερα από τη σύγκλιση, όλες αυτές οι  $mn$  τιμές  $\psi_i^{(j)}$  θεωρούνται ότι κατανέμονται σύμφωνα με την κατανομή στόχο  $\pi$ . Η διακύμανση  $\sigma_\psi^2 = \text{Var}(h(\theta))$  εκτιμάται από τους συνεπείς εκτιμητές  $B, W$  και  $\hat{\sigma}_\psi^2 = (1 - 1/n)W + (1/n)B$ . Αν οι αλυσίδες δεν έχουν συγχλίνει ακόμα, οι αρχικές τιμές τους θα επηρεάζουν ακόμα τις τροχιές τους. Εξαιτίας της διασποράς των αρχικών τιμών, ως προς το στήριγμα της κατανομής στόχου  $\pi$ , ο εκτιμητής  $\hat{\sigma}_\psi^2$  υπερεκτιμά την πραγματική διακύμανση  $\sigma_\psi^2$ , μέχρι την επίτευξη της στασιμότητας. Από την άλλη μεριά, πριν να επέλθει σύγκλιση, ο εκτιμητής  $W$  τείνει να υποεκτιμά την πραγματική διακύμανση  $\sigma_\psi^2$ , διότι κάθε αλυσίδα δεν έχει διαβεί ολόκληρο το χώρο καταστάσεων και επομένως βασίζεται σε μικρότερες διαφορές  $\psi_i^{(j)} - \bar{\psi}_i$ . Με βάση αυτή τη συλλογιστική, ως δείκτης σύγκλισης μπορεί να ληφθεί ο εκτιμητής

$$\hat{R} = \sqrt{\frac{\hat{\sigma}_\psi^2}{W}}, \quad (5.26)$$

του οποίου η τιμή είναι πάντα μεγαλύτερη από 1. Ο εκτιμητής  $\hat{R}$  αναφέρεται ως potential scale reduction. Καθώς  $n \rightarrow \infty$ , οι εκτιμητές  $\hat{\sigma}_\psi^2$  και  $W$  συγχλίνουν, ως απόρροια του Εργοδικού Θεωρήματος, στη διακύμανση  $\sigma_\psi^2$  και τότε  $\hat{R} \rightarrow 1$ . Η σύγκλιση ανιχνεύεται από την εγγύτητα του εκτιμητή  $\hat{R}$  στην τιμή 1. Ο Gelman (1996) προτείνει να αποδεχόμαστε τη σύγκλιση όταν  $\hat{R} \leq 1.2$ .

Ο πρωτότυπος εκτιμητής που προτάθηκε από τους Gelman και Rubin (1992) είναι πολύ πιο περίπλοκος. Ο Gamerman σημειώνει ότι η περιπλοκότητα αυτή δεν είναι αναγκαία.

Ο εκτιμητής  $\hat{R}$  πρέπει να υπολογίζεται για όλες τις ποσότητες που εμπύπτουν στο ενδιαφέρον μας, ώστε να παρέχει τεκμηριωμένη πληροφορία για τη σύγκλιση της αλυσίδας. Σημειώστε ότι πρόκειται για μονομεταβλητή τεχνική, αλλά μπορεί να εφαρμοστεί και στην περίπτωση όπου αναπτύσουμε παράλληλα  $m$   $d$ -διάστατες αλυσίδες  $\theta_i = (\theta_{i1}, \dots, \theta_{id})$ ,  $i = 1, \dots, m$ , θέτοντας  $h(\theta) = -2 \log \pi(\theta)$ . Οπωσδήποτε, σε αυτή την περίπτωση η επιλογή των διεσπαρμένων αρχικών σημείων εκκίνησης είναι πιο δύσκολη.

Στα μειονεκτήματα της μεθόδου συγκαταλέγονται η εξάρτησή της από την κανονικότητα, η επιλογή ευρέως διεσπαρμένων αρχικών καταστάσεων για τις παράλληλες αλυσίδες, η οποία απαιτεί κάποια γνώση της κατανομής στόχου, η επιλογή του αριθμού των αλυσίδων που θα προσομοιωθούν και ο υπολογισμός των εκτιμήσεων της πραγματικής διακύμανσης  $\sigma_{\psi}^2$ . Εναλλακτικά, μπορούν να χρησιμοποιηθούν μη παραμετρικοί εκτιμητές της διακύμανσης. Επίσης, παραμετρικοί μετασχηματισμοί (reparametrization) μπορούν να επιστρατευτούν για τις παραμέτρους που δεν εμφανίζουν κανονική συμπεριφορά, αλλά αυτό αυξάνει την πολυπλοκότητα της εξακρίβωσης της σύγκλισης. Τέλος, η απόκλιση από την κανονικότητα δυσχεραίνει το έργο της επιλογής των αρχικών σημείων για την εκκίνηση των παράλληλων αλυσίδων.

### 5.6.3 Το διαγνωστικό Κριτήριο των Zellner και Min

Υποθέστε ότι το διάνυσμα των παραμέτρων  $\theta$  μπορεί να διαιρεθεί σε δύο μπλοκ,  $\theta_1$  και  $\theta_2$ . Τότε,  $\pi(\theta) = \pi_1(\theta_1|\theta_2)\pi^{(2)}(\theta_2) = \pi_2(\theta_2|\theta_1)\pi^{(1)}(\theta_1)$ , για κάθε  $\theta$ . Σε εφαρμογές όπου ο δειγματολήπτης Gibbs μπορεί να εφαρμοστεί, οι πλήρεις δεσμευμένες πυκνότητες  $\pi_i(\theta_i|\theta_{-i})$  είναι διαθέσιμες, αλλά οι περιθωριακές πυκνότητες  $\pi^{(i)}(\theta_i)$  όχι. Ωστόσο, μπορούν να εκτιμηθούν με τη βοήθεια του Rao - Blackwellized εκτιμητή

$$\hat{\pi}^{(i)}(\theta_i) = \frac{1}{n} \sum_{j=1}^n \pi_i(\theta_i|\theta_{-i}^{(j)}), \quad (5.27)$$

τον οποίο συναντήσαμε στην Ενότητα 4.7 για την αύξηση δεδομένων.

Οι Zellner and Min (1995) προτείνουν δύο κριτήρια για την επαλήθευση της σύγκλισης του δειγματολήπτη Gibbs. Το πρώτο κριτήριο βασίζεται στο στατιστικό

$$\hat{\eta} = \pi_1(\theta_1|\theta_2)\hat{\pi}^{(2)}(\theta_2) - \pi_2(\theta_2|\theta_1)\hat{\pi}^{(1)}(\theta_1). \quad (5.28)$$

Αν η αλυσίδα έχει συγκλίνει, τότε το  $\hat{\eta}$  είναι κοντά στο μηδέν, για κάθε  $\theta$ .

Το δεύτερο κριτήριο βασίζεται στους λόγους

$$\hat{\xi}_1 = \frac{\pi_2(\theta_2|\theta_1)\hat{\pi}^{(1)}(\theta_1)}{\pi_2(\theta_2^*|\theta_1^*)\hat{\pi}^{(1)}(\theta_1^*)} \quad (5.29)$$

$$\hat{\xi}_2 = \frac{\pi_1(\theta_1|\theta_2)\hat{\pi}^{(2)}(\theta_2)}{\pi_1(\theta_1^*|\theta_2^*)\hat{\pi}^{(2)}(\theta_2^*)}, \quad (5.30)$$

όπου  $\theta^* = (\theta_1^*, \theta_2^*)$  είναι μια άλλη τιμή από το χώρο των καταστάσεων τις αλυσίδας  $(\theta^{(t)})$ . Τόσο το στατιστικό  $\hat{\xi}_1$ , όσο και το  $\hat{\xi}_2$ , είναι εκτιμητές της ποσότητας  $\xi = \pi(\theta) / \pi(\theta^*)$ . Αν η αλυσίδα έχει συγκλίνει τότε οι εκτιμήσεις  $\hat{\xi}_1$  και  $\hat{\xi}_2$  είναι κοντά. Επιπλέον, αν είναι κοντά στην τιμή του  $\xi$ , τότε η αλυσίδα έχει συγκλίνει στη σωστή κατανομή ισορροπίας.

Οι Zellner and Min (1995) τυποποιούν την προσέγγισή τους υποθέτοντας κανονικές κατανομές για τις πλήρεις δεσμευμένες κατανομές με πυκνότητα  $\pi_i(\theta_i | \theta_{-i})$ .

Οι Ritter and Tanner (1992) προτείνουν να αποτιμούμε τη σύγκλιση, για την προσομοιωθείσα αλυσίδα, εξετάζοντας τους λόγους  $\hat{\xi}_1$  και  $\hat{\xi}_2$ . Εισηγούνται, επίσης τον υπολογισμό των λόγων στις προσομοιωμένες τιμές  $\theta^{(t)}$  και την ταυτόχρονη κατασκευή του ιστογράμματος των τιμών των λόγων. Καθώς  $n \rightarrow \infty$ , τα ιστογράμματα αυτά πρέπει να προσεγγίζουν κάποια εκφυλισμένη κατανομή με παράμετρο θέσης την τιμή ένα. Δείτε επίσης Gelfand (1992) για περαιτέρω συζήτηση επί των ιστογραμμάτων αυτών.

Η μεταστάσιμη (metastable) συμπεριφορά της αλυσίδας μπορεί να αποτελέσει πρόβλημα, το οποίο υπερπηδάται με τη χρήση παράλληλων αλυσίδων. Ένα άλλο μεονέκτημα της μεθόδου είναι η ανάγκη της διαθεσιμότητας των πλήρεις δεσμευμένων κατανομών, γεγονός που περιορίζει την εφαρμογή της στην περίπτωση του δειγματολήπτη Gibbs. Επίσης, δεν είναι σαφής ο τρόπος διαχωρισμού του παραμετρικού διανύσματος  $\theta$  σε δύο μπλοκ. Όταν η διάσταση  $d$  του διανύσματος  $\theta$  είναι μεγάλη, τότε υπάρχουν πολλοί τρόποι διαχωρισμού του τελευταίου και το κριτήριο μπορεί να εφαρμοστεί σε όλους.

#### 5.6.4 Μη Παραμετρικός Έλεγχος Στασιμότητας

Οι κλασσικοί μη παραμετρικοί έλεγχοι, όπως το τεστ των Kolmogorov - Smirnov ή το τεστ του Kuiper, μπορούν να εφαρμοστούν στο αποτέλεσμα ενός MCMC αλγόριθμου, δηλαδή στις προσομοιωμένες τιμές  $\theta^{(t)}$ , προκειμένου να εξαχθούν συμπεράσματα για τη στασιμότητα της αλυσίδας  $(\theta^{(t)})$ . Όταν η αλυσίδα είναι στάσιμη, οι τιμές  $\theta^{(t_1)}$  και  $\theta^{(t_2)}$  έχουν την ίδια περιθωριακή κατανομή, για αυθαίρετες τιμές  $t_1$  και  $t_2$ .

Ύστερα από την ανάπτυξη του αλγόριθμου, έχουμε αποκτήσει το δείγμα  $\theta^{(1)}, \dots, \theta^{(T)}$ . Αν χωρίσουμε το δείγμα σε δύο μπλοκ,  $\theta^{(1)}, \dots, \theta^{(\frac{T}{2})}$  και  $\theta^{(\frac{T}{2}+1)}, \dots, \theta^{(T)}$ , τότε είναι λογικό να συγκρίνουμε τις κατανομές των δύο δειγμάτων. Τα συνήθη μη παραμετρικά τεστ είναι κατασκευασμένα για τον έλεγχο της ταυτονομίας μεταξύ iid δειγμάτων. Επομένως, πρέπει να εισάγουμε κάποια

διόρθωση, εξαιτίας της συσχέτισης μεταξύ των  $\theta^{(k)}, \theta^{(s)}$ ,  $k \neq s$ .

Η διόρθωση επιτυγχάνεται με την εισαγωγή του μεγέθους  $G$ , που οδηγεί στην κατασκευή δύο ψευδο - ανεξάρτητων δειγμάτων. Για κάθε ένα από τα προηγούμενα δείγματα επιλέγουμε τα υποδείγματα  $\theta_1^{(G)}, \theta_1^{(2G)}, \dots$  και  $\theta_2^{(G)}, \theta_2^{(2G)}, \dots$ . Τότε η στατιστική συνάρτηση του τεστ των Kolmogorov - Smirnov γίνεται

$$K = \frac{1}{M} \sup_{\eta} \left| \sum_{g=1}^M I_{(0, \eta]} \left( \theta_1^{(gG)} \right) - \sum_{g=1}^M I_{(0, \eta]} \left( \theta_2^{(gG)} \right) \right|, \quad (5.31)$$

για την περίπτωση μιας μονοδιάστατης αλυσίδας. Για πολυδιάστατες αλυσίδες, ο τύπος (5.31) μπορεί να υπολογιστεί είτε σε κάποια συνάρτηση ενδιαφέροντος ή σε κάθε συντεταγμένη του διανύσματος  $\theta$  ξεχωριστά. Το στατιστικό  $K$  μπορεί να δώσει έναν κανόνα για την παύση του αλγόριθμου, όταν η σύγκλιση έχει επιτευχθεί. Υπό συνθήκες στασιμότητας, καθώς  $M \rightarrow \infty$ , η οριακή κατανομή της ποσότητας  $\sqrt{M}K$  έχει συνάρτηση κατανομής

$$R(x) = 1 - \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2}, \quad (5.32)$$

η οποία προσεγγίζεται από ένα πεπερασμένο άθροισμα. Το αντίστοιχο p - value μπορεί να υπολογιστεί για κάθε τιμή  $T$ , μέχρι να φτάσει κάτω από κάποιο προκαθορισμένο επίπεδο.

Η διαδικασία του κριτηρίου που παρουσιάζουμε μπορεί να βελτιωθεί λαμβάνοντας υπόψη την ακολουθιακή φύση του τεστ και το γεγονός ότι το  $K$  μπορεί να επιλεγεί ως η μικρότερη από όλες τις τιμές  $K$  που αντιστοιχούν στις συντεταγμένες του παραμετρικού διανύσματος  $\theta$ . Μια τρίτη χρήση της (5.31), είναι η αναπαράσταση των ποσοτήτων  $\sqrt{M}K$  ως προς  $T$  και ο οπτικός έλεγχος για στασιμότητα στις μικρές τιμές.

# Κεφάλαιο 6

## Εφαρμογές

### 6.1 Εισαγωγή

Το κεφάλαιο αυτό αφιερώνεται στην ανάλυση κάποιων υποθέσεων (case studies), που θα μας βοηθήσουν να κατανοήσουμε καλύτερα τη λειτουργία τόσο του δειγματολήπτη Gibbs όσο και του αλγόριθμου Metropolis - Hastings. Η πρώτη εφαρμογή είναι η πολύ συχνά συναντώμενη στη βιβλιογραφία περίπτωση των αποτυχιών ενός μηχανισμού άντλησης ιόντων από ένα πυρηνικό αντιδραστήρα (nuclear pump failures). Θα δούμε δύο παραλλαγές του μοντέλου και θα αναπτύξουμε αντίστοιχους τρόπους προσομοίωσης. Η δεύτερη εφαρμογή ασχολείται με την προσομοίωση των εκ των υστέρων κατανομών ενός μοντέλου τυχαίων επιδράσεων (random effects model). Ας δούμε πιο αναλυτικά κάθε περίπτωση.

### 6.2 Παράδειγμα 1.0.1 (Pump Failures)

Σύμφωνα με την από κοινού εκ των υστέρων κατανομή που δίνεται από τον τύπο (1.2), οι πλήρεις δεσμευμένες κατανομές είναι οι εξείς:

$$\lambda_i | \beta, t_i, x_i \sim Ga(x_i + \alpha, t_i + \beta), \quad i = 1, \dots, 10 \quad (6.1)$$

$$\beta | \lambda_1, \dots, \lambda_{10} \sim Ga\left(\gamma + 10\alpha, \delta + \sum_{i=1}^{10} \lambda_i\right). \quad (6.2)$$

Χρησιμοποιώντας το υπολογιστικό πακέτο Mathematica μπορούμε να κατασκευάσουμε ένα Gibbs αλγόριθμο. Σε κάθε βήμα του δειγματολήπτη, προσομοιώνεται πρώτα μία τιμή για τα  $\lambda_1, \dots, \lambda_{10}$  και έπειτα για το  $\beta$ . Ως αρχική τιμή

θέτουμε  $\beta = 1$ . Παρατηρήστε ότι δεν χρειάζεται να θέσουμε αρχικές τιμές για τα  $\lambda_i$ ,  $i = 1, \dots, 10$ , αφού οι πλήρεις δεσμευμένες κατανομές τους εξαρτώνται μόνο από το  $\beta$  και όχι από το διάνυσμα  $\lambda_{-i}$ . Ο δειγματολήπτης Gibbs που κατασκευάζουμε:

1. προσομοιώνει δείγματα από τις εκ των υστέρων περιθωριακές κατανομές των παραμέτρων  $\lambda_1, \dots, \lambda_{10}$  και  $\beta$ ,
2. υπολογίζει τις εκ των υστέρων αναμενόμενες τιμές και τυπικές αποκλίσεις των περιθώριων κατανομών,
3. εκτιμάει περιοχές αξιοπιστίας (credibility intervals) για τις παραμέτρους,
4. υπολογίζει τις αυτοσυσχετίσεις, υστέρησης  $k = 1, \dots, 20$ , των προσομοιωμένων τιμών σε κάθε ακολουθία,
5. κατασκευάζει trace plots, διαγράμματα των εργοδικών μέσων, ιστογράμματα των προσομοιωμένων τιμών των παραμέτρων και διαγράμματα αυτοσυσχετίσεων.

### 6.3 Παράδειγμα 1.0.2 (Pump Failures)

Σύμφωνα με τον τύπο (1.3), οι πλήρεις δεσμευμένες κατανομές είναι οι ακόλουθες:

$$\lambda_i | \beta, \alpha, t_i, x_i \sim Ga(x_i + \alpha, t_i + \beta), \quad i = 1, \dots, 10 \quad (6.3)$$

$$\beta | \lambda_1, \dots, \lambda_{10}, \alpha \sim Ga\left(\gamma + 10\alpha, \delta + \sum_{i=1}^{10} \lambda_i\right), \quad (6.4)$$

$$\alpha | \lambda_1, \dots, \lambda_{10}, \beta \propto \left(\frac{\prod_{i=1}^{10} \lambda_i}{e}\right)^\alpha \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{10}. \quad (6.5)$$

Με το Mathematica κατασκευάζουμε ένα Metropolis within Gibbs αλγόριθμο για την προσομοίωση τιμών από την εκ των υστέρων περιθωριακή κατανομή των παραμέτρων  $\lambda_1, \dots, \lambda_{10}, \beta, \alpha$ . Σε κάθε βήμα του, ο αλγόριθμος εξελίσσεται ως εξής: ο εξωτερικός Gibbs βρόγχος προσομοιώνει μία τιμή για κάθε παράμετρο  $\lambda_1, \dots, \lambda_{10}, \beta$ , από τις πλήρεις δεσμευμένες κατανομές τους (6.3) και (6.4), αντίστοιχα. Ακολουθεί η προσομοίωση μιας τιμής για την παράμετρο  $\alpha$ , με την εισαγωγή ενός βήματος Metropolis - Hastings.



Ο υποαλγόριθμος Metropolis - Hastings αναπτύσσεται σε μία μόνο επανάληψη (δηλαδή  $T = 1$ ), σε κάθε βήμα του εξωτερικού βρόγχου. Στην εφαρμογή που παραθέτουμε χρησιμοποιούμε δύο παραλλαγές του Metropolis - Hastings αλγόριθμου: τη γενική περίπτωση και την ανεξάρτητη περίπτωση.

Στη γενική περίπτωση χρησιμοποιούμε ως συνάρτηση πρότασης  $q(\alpha^{(t-1)}, \cdot)$  την εκθετική κατανομή  $Exp(1/\alpha^{t-1})$ . Ο υποαλγόριθμος εξελίσσεται σύμφωνα με τα ακόλουθα βήματα:

1. προσομοίωσε μία τιμή  $\eta \sim Exp(1/\alpha^{(t-1)})$ ,
2. προσομοίωσε μία τιμή  $u \sim U(0, 1)$ ,
3. αν  $u \leq \rho(\alpha^{(t-1)}, \eta)$ , τότε θέσε  $\alpha^{(t)} = \eta$ , διαφορετικά θέσε  $\alpha^{(t)} = \alpha^{(t-1)}$ ,

όπου  $\rho(\alpha^{(t-1)}, \eta)$  είναι η πιθανότητα αποδοχής τις τιμές που προτείνεται από τη συνάρτηση πρότασης και στη συγκεκριμένη περίπτωση, με εφαρμογή του τύπου (3.8), έχει τη μορφή

$$\rho(\alpha^{(t-1)}, \eta) = 1 \wedge \left( \frac{\prod_{i=1}^{10} \lambda_i}{e} \right)^{\eta - \alpha^{(t-1)}} \left( \beta^{\eta - \alpha^{(t-1)}} \frac{\Gamma(\alpha^{(t-1)})}{\Gamma(\eta)} \right)^{10} \frac{\eta}{\alpha^{(t-1)}}. \quad (6.6)$$

Στην ανεξάρτητη αλυσίδα η συνάρτηση πρότασης δεν εξαρτάται από την προηγούμενη τιμή της παραμέτρου που προσομοιώνεται. Επιλέγουμε λοιπόν ως συνάρτηση πρότασης την εκθετική  $Exp(1)$ . Τώρα η πιθανότητα αποδοχής υπολογίζεται σύμφωνα με τον τύπο που δίνεται στον Αλγόριθμο 3.5.1 και λαμβάνει τη μορφή

$$\rho(\alpha^{(t-1)}, \eta) = 1 \wedge \left( \frac{\prod_{i=1}^{10} \lambda_i}{e} \right)^{\eta - \alpha^{(t-1)}} \left( \beta^{\eta - \alpha^{(t-1)}} \frac{\Gamma(\alpha^{(t-1)})}{\Gamma(\eta)} \right)^{10} e^{\eta - \alpha^{(t-1)}}. \quad (6.7)$$

## 6.4 Παράδειγμα 1.0.3 (Random Effects Model)

Δίνονται  $K = 6, m = 8, \mu_0 = 0, a_1 = a_2 = b_1 = b_2 = \lambda_0 = 1$ . Επίσης, στον Πίνακα 6.1 δίνονται τα επαρκή στατιστικά δεδομένα που χρειάζονται για την προσομοίωση. Ως αρχικές τιμές για την προσομοίωση χρησιμοποιούμε τις  $\theta_i = \bar{y}_i, i = 1, \dots, 6$  και  $\mu = \bar{y}$ .

cell	1	2	3	4	5	6
$\bar{y}_i$	-0.22795	-1.1913	0.030547	0.48428	0.036639	-0.026581

mK=48

$$\bar{y} = (\text{mK})^{-1} \sum_{i=1}^6 \sum_{j=1}^8 y_{ij} = -0,14906$$

$$\text{sse} = \sum_{i=1}^6 \sum_{j=1}^8 (y_{ij} - \bar{y}_i)^2 = 23,251$$

Πίνακας 6.1: Random Effects Model Data

Για να προσομοιώσουμε τιμές από την εκ των υστέρων περιθωριακή κατανομή των παραμέτρων  $\lambda_e, \lambda_\theta, \mu, \theta_i, i = 1, \dots, 6$ , κατασκευάζουμε ένα δειγματολήπτη Gibbs. Η από κοινού εκ των υστέρων κατανομή των παραμέτρων αυτών δίνεται από τον τύπο (1.10). Οι πλήρεις περιθωριακές κατανομές προκύπτουν άμεσα από την από κοινού κατανομή (1.11):

$$1. \pi(\theta_i | \mu, \lambda_e, \lambda_\theta, \mathbf{y}) \propto \exp \left\{ (m\lambda_e \bar{y}_i + \lambda_\theta \mu) \theta_i - \frac{1}{2} (m\lambda_e + \lambda_\theta) \theta_i^2 \right\},$$

$$i = 1, \dots, K, \text{ δηλαδή}^1$$

$$\theta_i | \mu, \lambda_e, \lambda_\theta, \mathbf{y} \sim N \left( \frac{m\lambda_e \bar{y}_i + \lambda_\theta \mu}{m\lambda_e + \lambda_\theta}, (m\lambda_e + \lambda_\theta)^{-1} \right), i = 1, \dots, K \quad (6.8)$$

$$2. \pi(\mu | \theta_1, \dots, \theta_K, \lambda_\theta, \mathbf{y})$$

$$\propto \exp \left\{ \left( \lambda_\theta \sum_{i=1}^K \theta_i + \lambda_0 \mu_0 \right) \mu - \frac{1}{2} (\lambda_\theta K + \lambda_0) \mu^2 \right\}, \text{ δηλαδή}$$

$$\mu | \theta_1, \dots, \theta_K, \lambda_\theta, \mathbf{y} \sim N \left( \frac{\lambda_\theta \sum_{i=1}^K \theta_i + \lambda_0 \mu_0}{\lambda_\theta K + \lambda_0}, (\lambda_\theta K + \lambda_0)^{-1} \right). \quad (6.9)$$

$$3. \lambda_e | \theta_1, \dots, \theta_K, \mathbf{y} \sim Ga \left( \frac{mK}{2} + a_2, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^m (y_{ij} - \theta_i)^2 + b_2 \right), \text{ δηλαδή}$$

$$\lambda_e | \theta_1, \dots, \theta_K, \mathbf{y} \sim Ga \left( \frac{mK}{2} + a_2, \frac{\text{sse} + m \sum_{i=1}^K (\bar{y}_i - \theta_i)^2}{2} + b_2 \right) \quad (6.10)$$

$$\text{όπου } \text{sse} = \sum_{i=1}^K \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2.$$

<sup>1</sup>Αν  $f(x) \propto \exp \left\{ ax - \frac{1}{2} bx^2 \right\} = \exp \left\{ -\frac{b}{2} \left( x^2 - \frac{2}{b} ax \right) \right\} \propto \exp \left\{ -\frac{b}{2} \left( x^2 - \frac{2}{b} ax + \frac{a^2}{b^2} \right) \right\} \propto \exp \left\{ -\frac{b}{2} \left( x - \frac{a}{b} \right)^2 \right\}$ , δηλαδή  $X \sim N \left( \frac{a}{b}, \frac{1}{b} \right)$ .

4.

$$\lambda_\theta | \theta_1, \dots, \theta_K, \mu, \mathbf{y} \sim Ga \left( \frac{K}{2} + a_1, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + b_1 \right). \quad (6.11)$$

Με το υπολογιστικό πακέτο Mathematica κατασκευάζουμε το δειγματολήπτη Gibbs, ο οποίος σε κάθε επανάληψή του ακολουθεί την εξής διαδικασία:

1. προσομοιώνει μία τιμή για την παράμετρο  $\lambda_e$  από την (6.10),
2. προσομοιώνει μία τιμή για την παράμετρο  $\lambda_\theta$  από την (6.11),
3. προσομοιώνει μία τιμή για την παράμετρο  $\mu$  από την (6.9),
4. προσομοιώνει μία τιμή για κάθε παράμετρο  $\theta_i$ ,  $i = 1, \dots, K$ , από τις (6.8).

Πέρα από τη συνήθη μορφή του δειγματολήπτη Gibbs, αναπτύξαμε και τον ομαδοποιημένο δειγματολήπτη Gibbs<sup>2</sup>. Κάθε επανάληψη του αλγόριθμου:

1. προσομοιώνει μία τιμή για την παράμετρο  $\lambda_e$  από την (6.10),
2. προσομοιώνει μία τιμή για την παράμετρο  $\lambda_\theta$  από την (6.11),
3. προσομοιώνει ταυτόχρονα μία τιμή για κάθε παράμετρο  $\mu, \theta_1, \dots, \theta_K$  από την πυκνότητα  $\pi(\mu, \theta_1, \dots, \theta_K | \lambda_e, \lambda_\theta)$ .

Αν αναλογιστούμε ότι

$$\pi(\mu, \theta_1, \dots, \theta_K | \lambda_e, \lambda_\theta) = \pi(\mu | \lambda_e, \lambda_\theta) \pi(\theta_1, \dots, \theta_K | \mu, \lambda_e, \lambda_\theta), \quad (6.12)$$

τότε το βήμα 3 είναι ισοδύναμο με τις ακόλουθες ενέργειες:

1. προσομοιώνει μία τιμή για την παράμετρο  $\mu$  από την  $\pi(\mu | \lambda_e, \lambda_\theta)$ ,
2. προσομοιώνει μία τιμή για κάθε παράμετρο  $\theta_i$ ,  $i = 1, \dots, K$ , από τις  $\pi(\theta_i | \mu, \lambda_e, \lambda_\theta)$ .

---

<sup>2</sup>Βλέπε και Ενότητα 4.6.3, Αλγόριθμο 4.6.4.

Η πυκνότητα  $\pi(\theta_1, \dots, \theta_K | \mu, \lambda_e, \lambda_\theta)$  δίνεται από το γινόμενο των σχέσεων (6.8) και μπορεί να λάβει τη μορφή:

$$\begin{aligned} \pi(\theta_1, \dots, \theta_K | \mu, \lambda_e, \lambda_\theta) &\propto \prod_{i=1}^K \exp \left\{ -\frac{\left(\theta_i - \frac{m\lambda_e \bar{y}_i + \lambda_\theta \mu}{m\lambda_e + \lambda_\theta}\right)^2}{2(m\lambda_e + \lambda_\theta)^{-1}} \right\} \\ &= \prod_{i=1}^K \exp \left\{ -\frac{m\lambda_e + \lambda_\theta}{2} \theta_i^2 + (m\lambda_e \bar{y}_i + \lambda_\theta \mu) \theta_i \right. \\ &\quad \left. - \frac{(m\lambda_e \bar{y}_i + \lambda_\theta \mu)^2}{2(m\lambda_e + \lambda_\theta)} \right\}. \end{aligned} \quad (6.13)$$

Από την από κοινού εκ των υστέρων πυκνότητα (1.11) είναι εύκολο να δείξουμε ότι

$$\begin{aligned} &\pi(\mu, \theta_1, \dots, \theta_K | \lambda_e, \lambda_\theta) \\ &\propto \exp \left\{ -\frac{\lambda_e}{2} \sum_{i=1}^K \sum_{j=1}^m (y_{ij} - \theta_i)^2 - \frac{\lambda_\theta}{2} \sum_{i=1}^K (\theta_i - \mu)^2 - \frac{\lambda_0}{2} (\mu - \mu_0)^2 \right\} \\ &\propto \prod_{i=1}^K \exp \left\{ (m\lambda_e \bar{y}_i + \lambda_\theta \mu) \theta_i - \frac{1}{2} (m\lambda_e + \lambda_\theta) \theta_i^2 + \frac{\lambda_0 \mu_0}{K} \mu \right. \\ &\quad \left. - \frac{1}{2} \left( \lambda_\theta + \frac{\lambda_0}{K} \right) \mu^2 \right\}. \end{aligned} \quad (6.14)$$

Αντικαθιστώντας τις σχέσεις (6.13) και (6.14) στη σχέση (6.12) λαμβάνουμε την πυκνότητα

$$\begin{aligned} &\pi(\mu | \lambda_e, \lambda_\theta) \\ &\propto \prod_{i=1}^K \exp \left\{ \frac{\lambda_0 \mu_0}{K} \mu - \frac{1}{2} \left( \lambda_\theta + \frac{\lambda_0}{K} \right) \mu^2 + \frac{(m\lambda_e \bar{y}_i + \lambda_\theta \mu)^2}{2(m\lambda_e + \lambda_\theta)} \right\} \end{aligned} \quad (6.15)$$

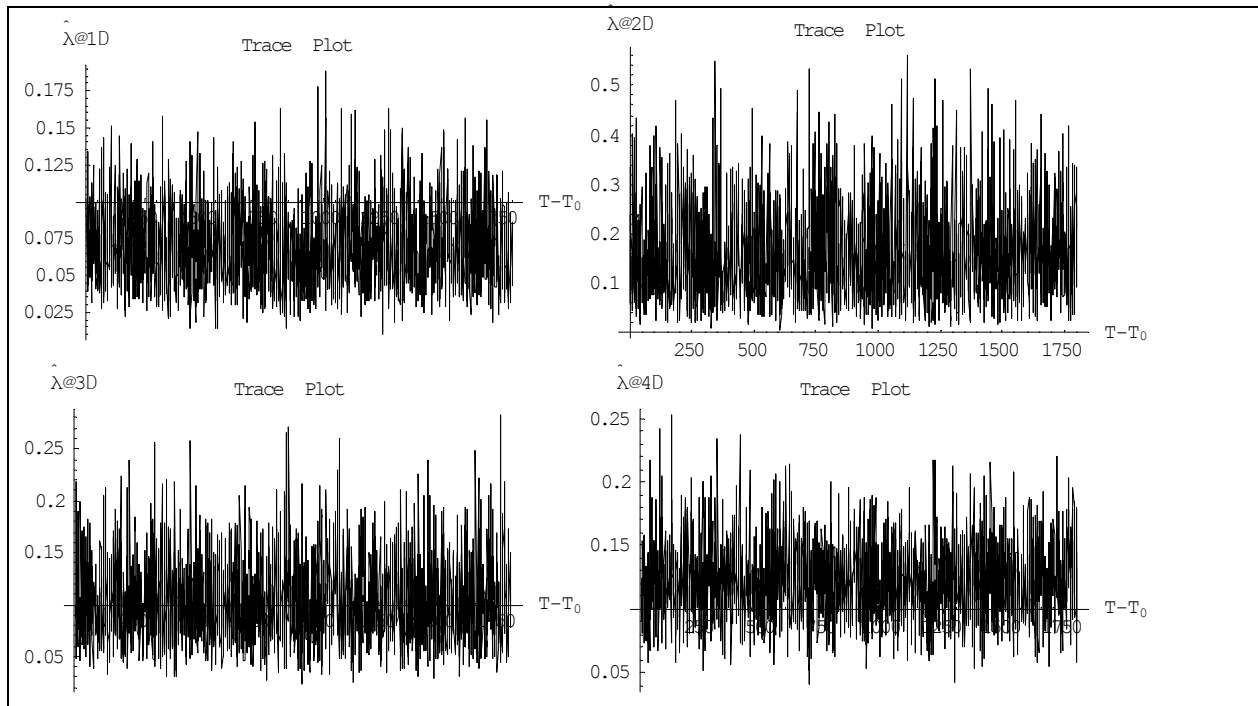
που απαιτείται για την ανάπτυξη του ομαδοποιημένου δειγματολήπτη Gibbs της παρούσας εφαρμογής.

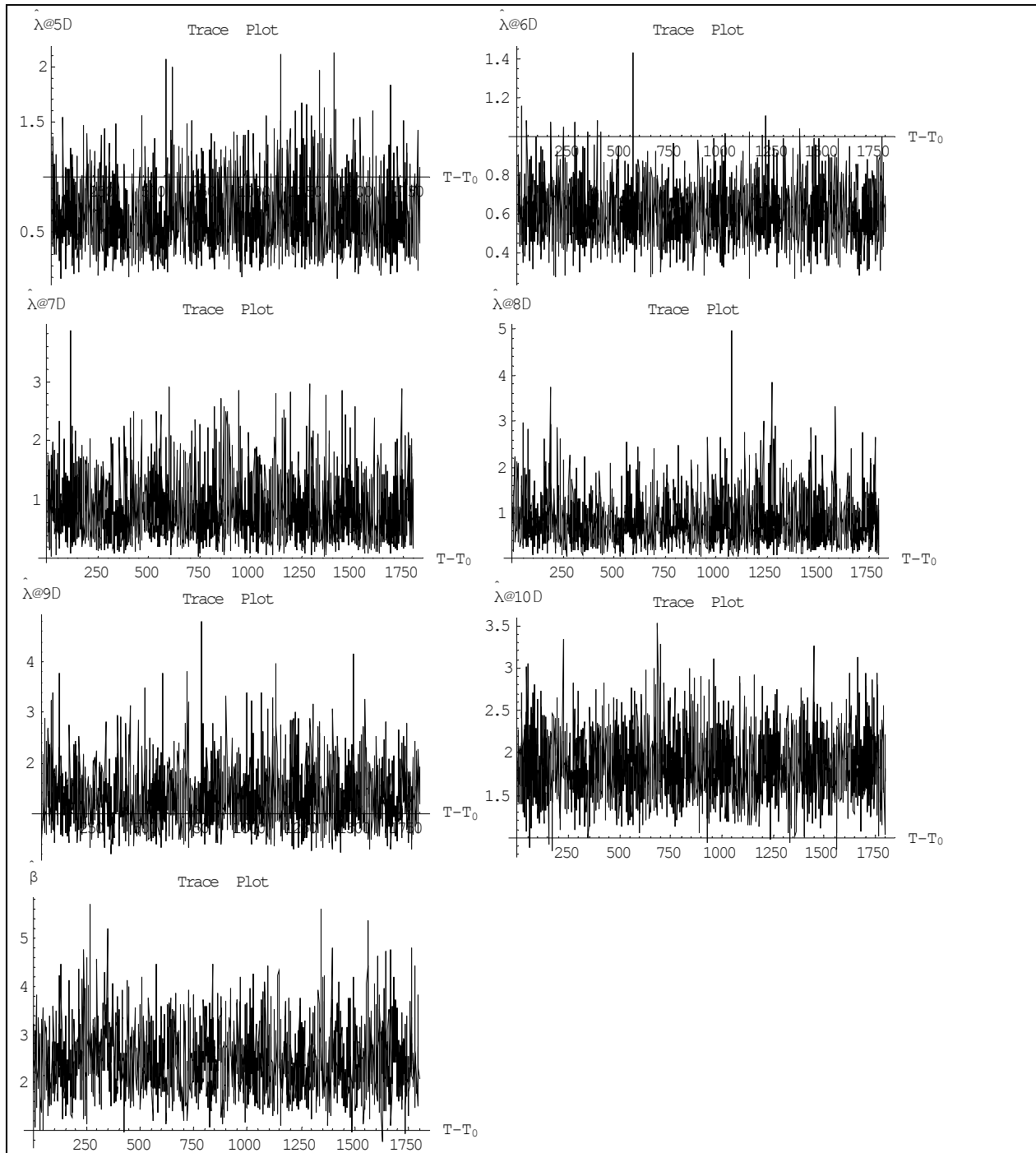
Ακολουθεί η ανάπτυξη των αποτελεσμάτων όλων των παραδειγμάτων που συζητήσαμε σε αυτή την ενότητα.

**Παράδειγμα 1.0.1 (Pump Failures) - Gibbs sampler**

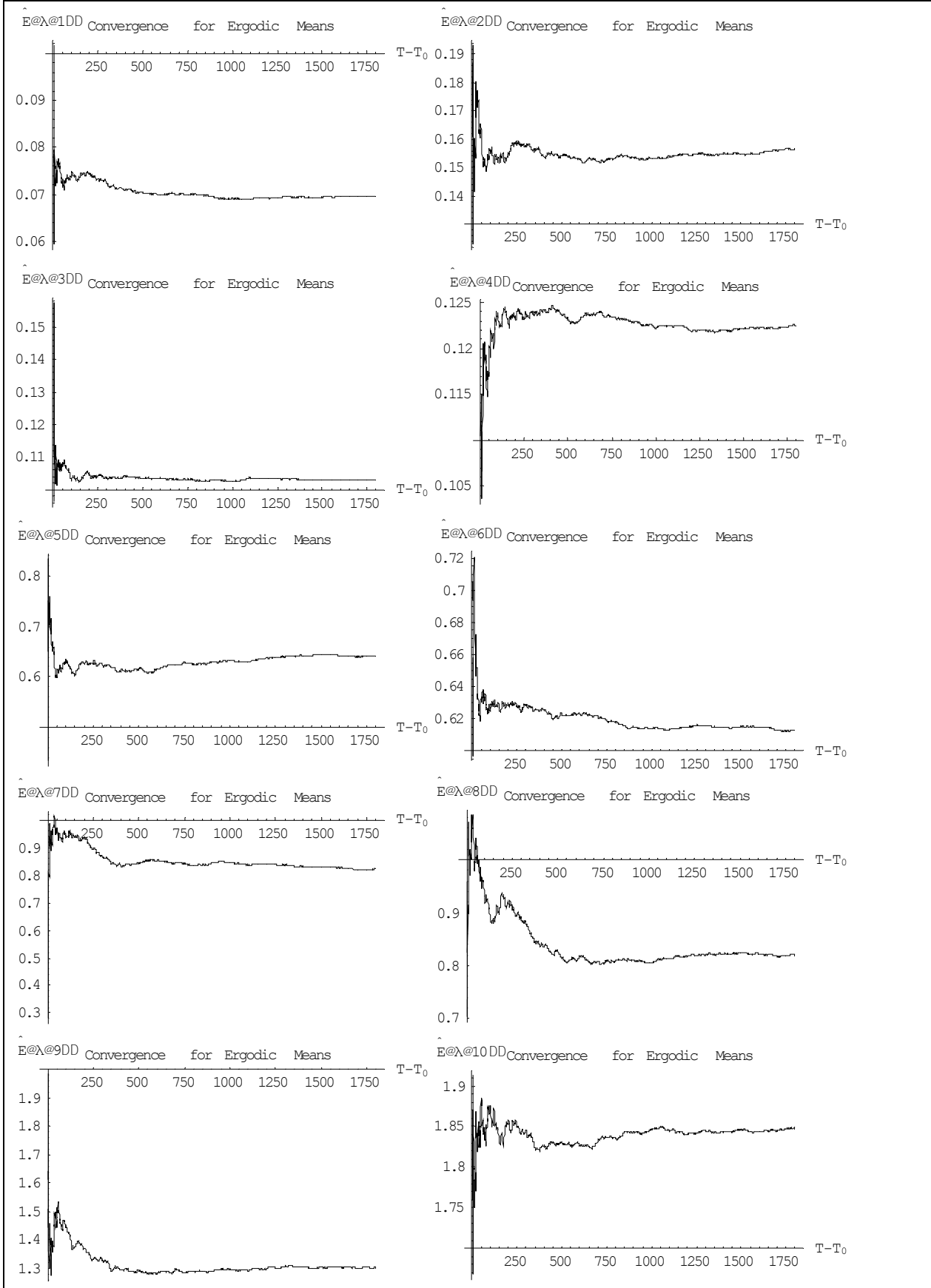
Parameter	Posterior Mean	Posterior St. Deviation	2.5% Quantile	97.5% Quantile
$\lambda_1$	0.0695896	0.0271127	0.0274984	0.129287
$\lambda_2$	0.156598	0.0953015	0.0264974	0.384042
$\lambda_3$	0.103118	0.0401571	0.0411483	0.193156
$\lambda_4$	0.122543	0.0305136	0.0668365	0.18808
$\lambda_5$	0.641567	0.302945	0.196703	1.34922
$\lambda_6$	0.61319	0.141258	0.360204	0.92344
$\lambda_7$	0.82191	0.519994	0.14811	2.12092
$\lambda_8$	0.820143	0.529824	0.13998	2.12097
$\lambda_9$	1.30309	0.581592	0.441649	2.72028
$\lambda_{10}$	1.848	0.388558	1.17935	2.68737
$\beta$	2.44052	0.689849	1.35445	3.96714

**TABLE 6.2.1 Descriptive statistics**





**TABLE 6.2.2 Trace Plots**



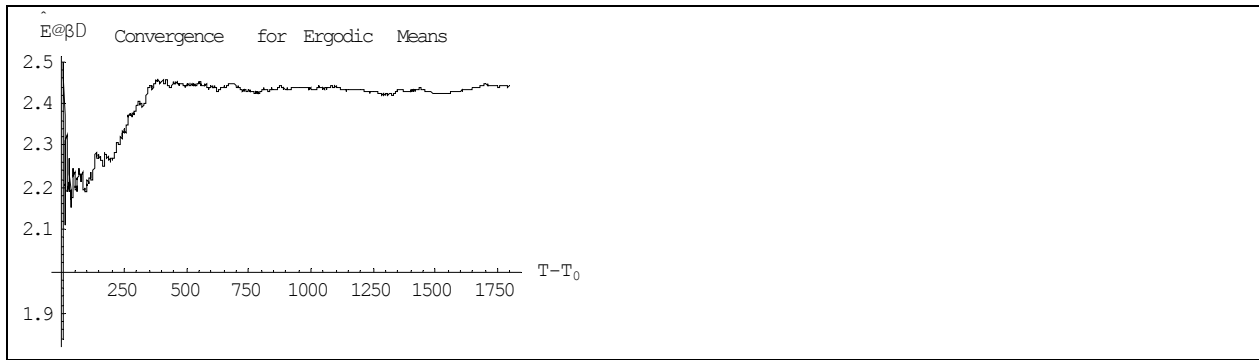
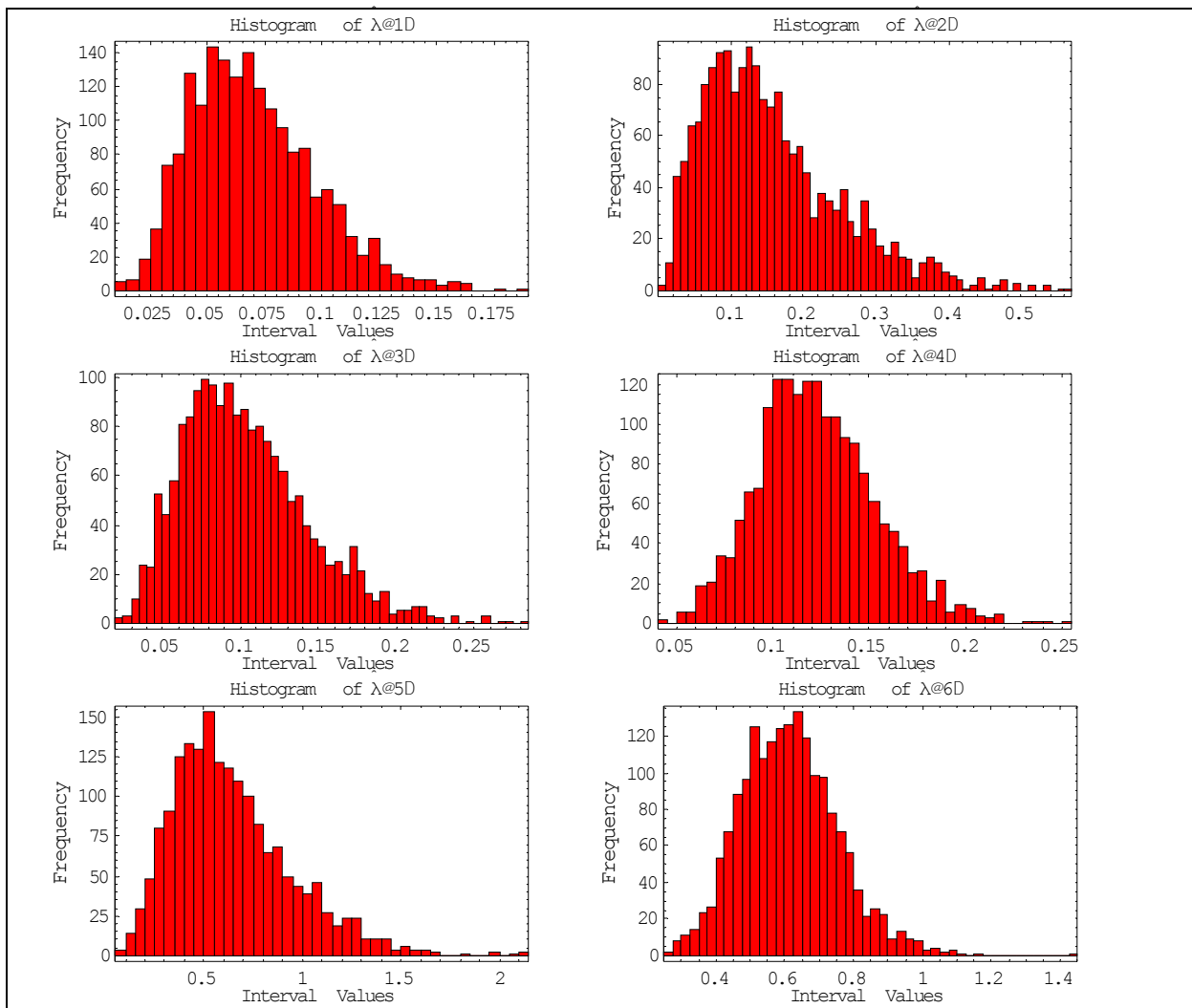
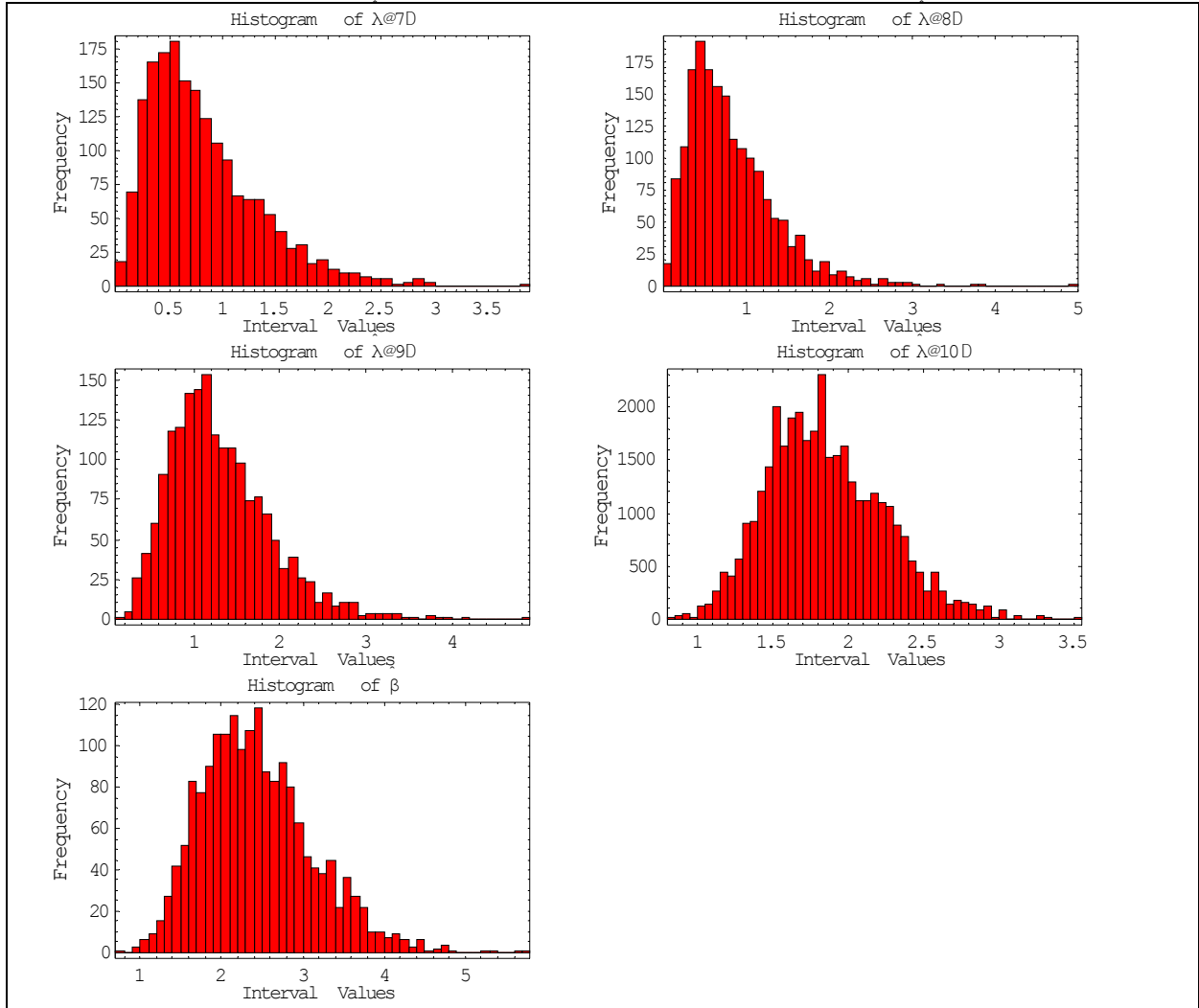


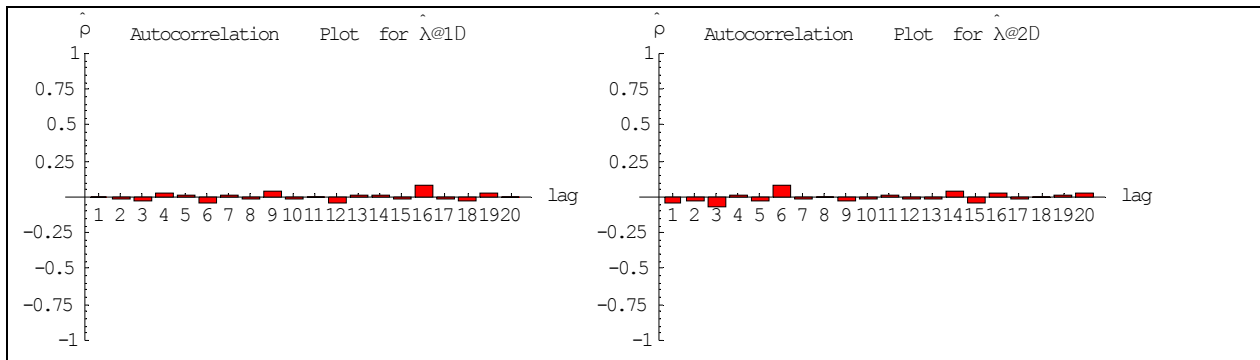
TABLE 6.2.3 Convergence for Ergodic Means

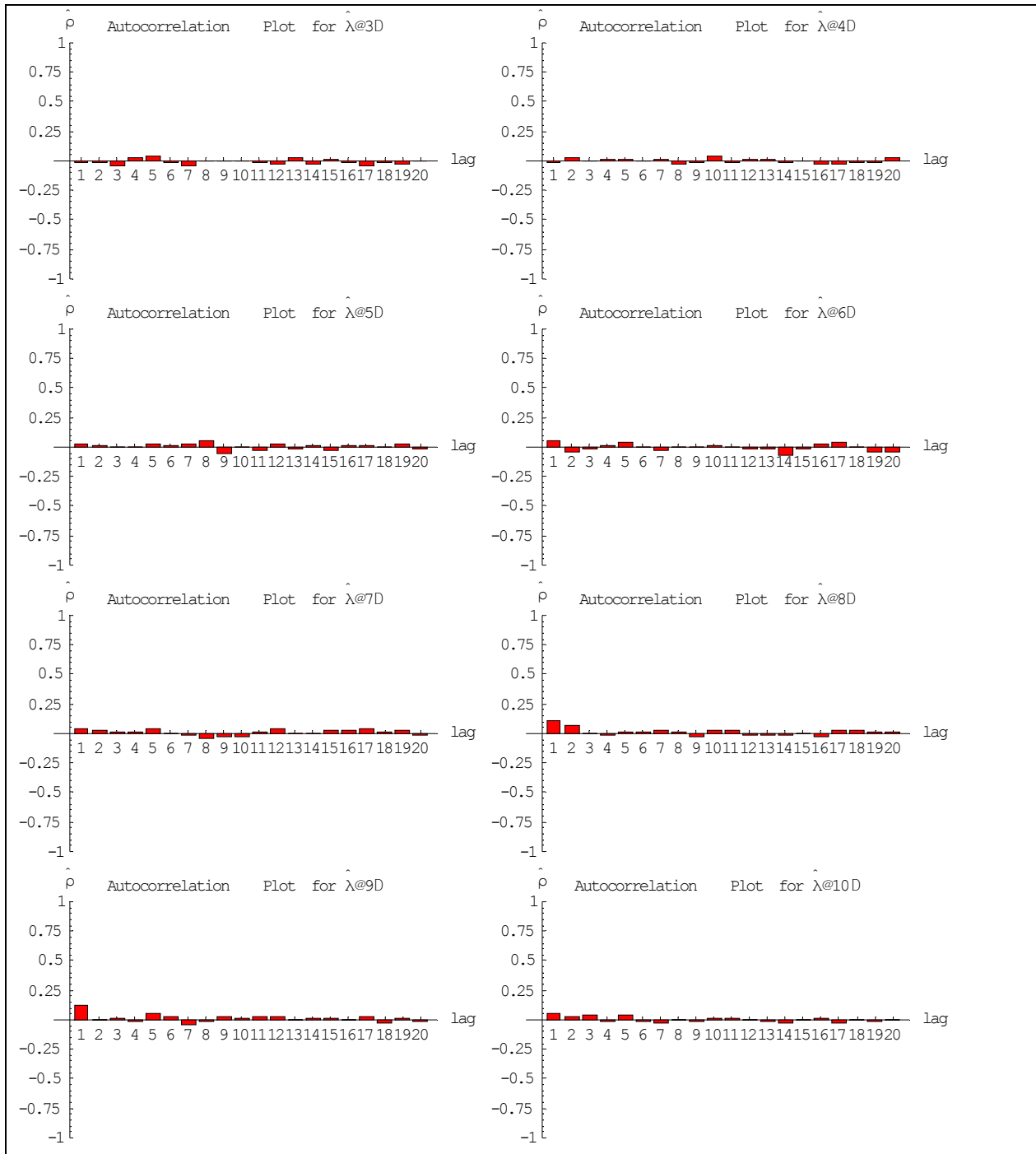






**TABLE 6.2.4 Histograms**





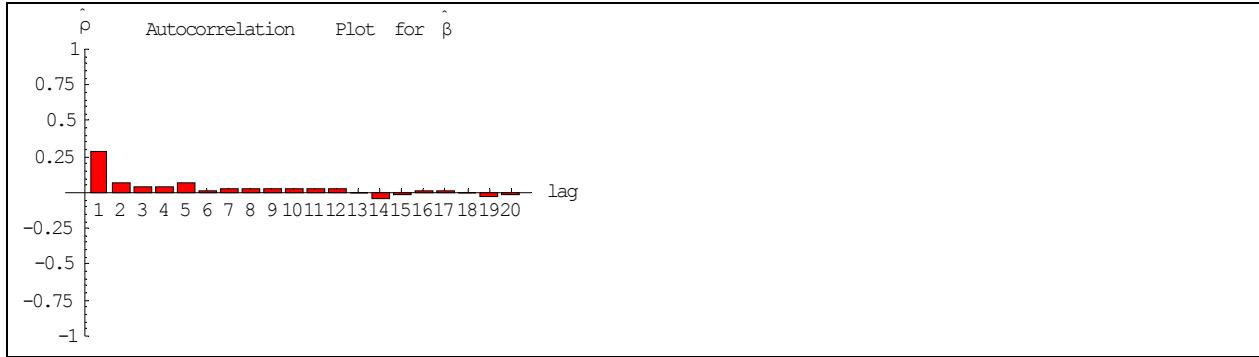
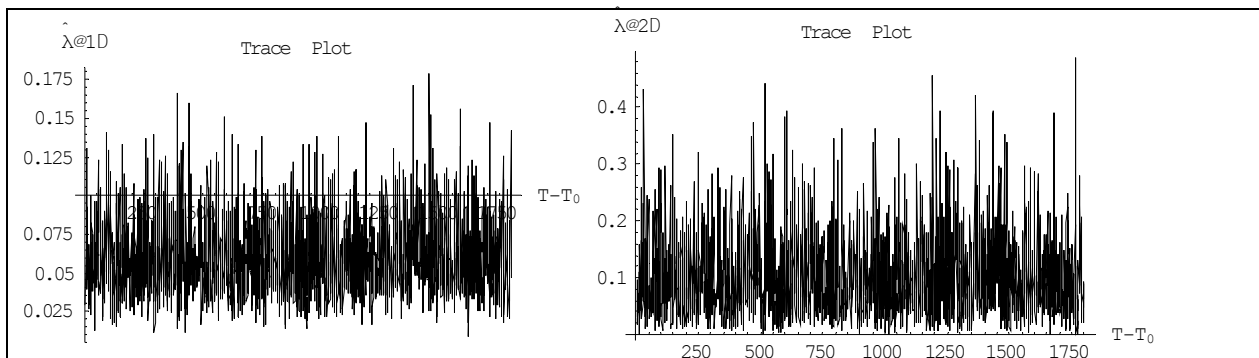


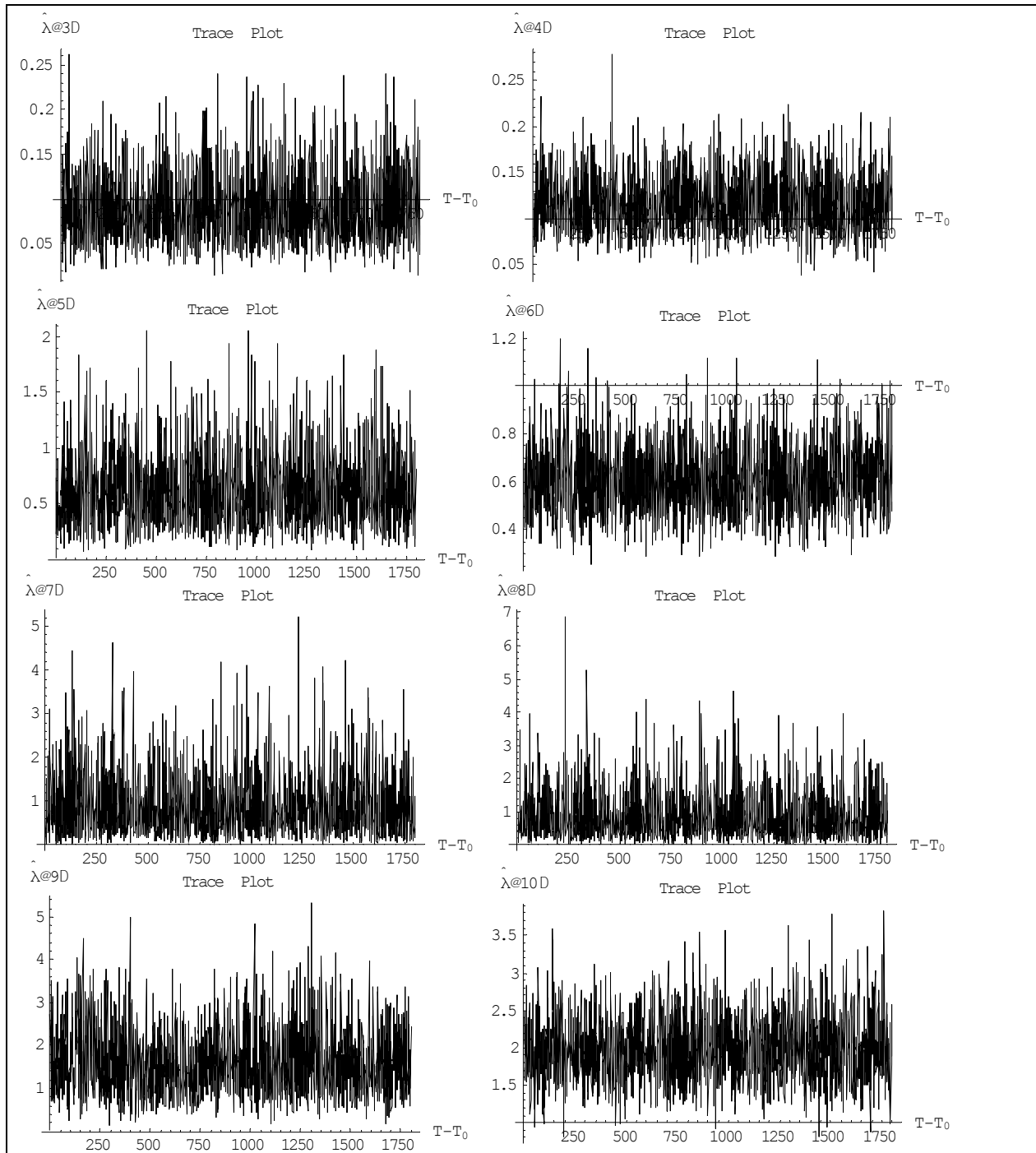
TABLE 6.2.5 Autocorrelation Plots

**Παράδειγμα 1.0.2 (Pump Failures) – Metropolis within Gibbs (general case)**

Parameter	Posterior Mean	Posterior St. Deviation	2.5% Quantile	97.5% Quantile
$\lambda_1$	0.0603452	0.0254019	0.0215955	0.120391
$\lambda_2$	0.1012	0.0747414	0.00946787	0.292178
$\lambda_3$	0.0893765	0.03796	0.0337345	0.177801
$\lambda_4$	0.1166	0.0299705	0.0673067	0.182184
$\lambda_5$	0.609973	0.321353	0.149014	1.38515
$\lambda_6$	0.608871	0.14161	0.36313	0.928121
$\lambda_7$	0.88494	0.712258	0.0754029	2.68321
$\lambda_8$	0.860765	0.718684	0.0627571	2.69672
$\lambda_9$	1.60884	0.757425	0.460273	3.39195
$\lambda_{10}$	1.97872	0.425748	1.22357	2.88205
$\beta$	0.92071	0.543284	0.155921	2.23406
$\alpha$	0.688115	0.269986	0.262344	1.2845

TABLE 6.3.1 Descriptive statistics





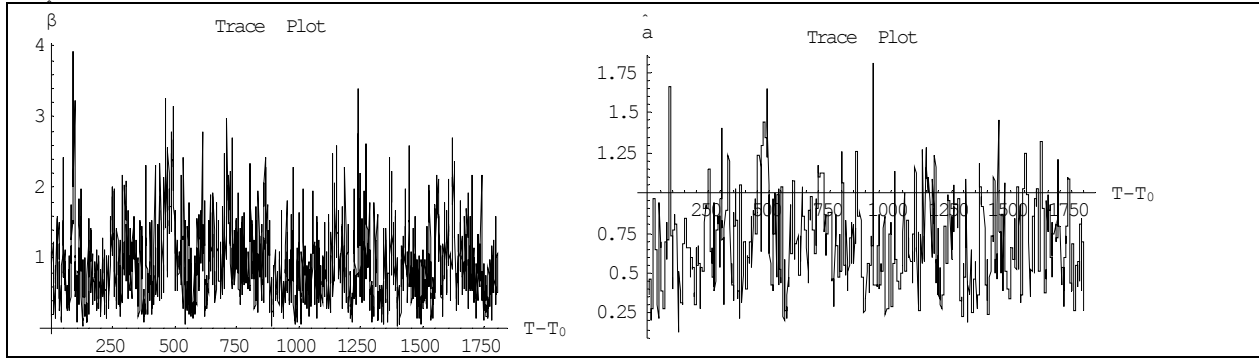
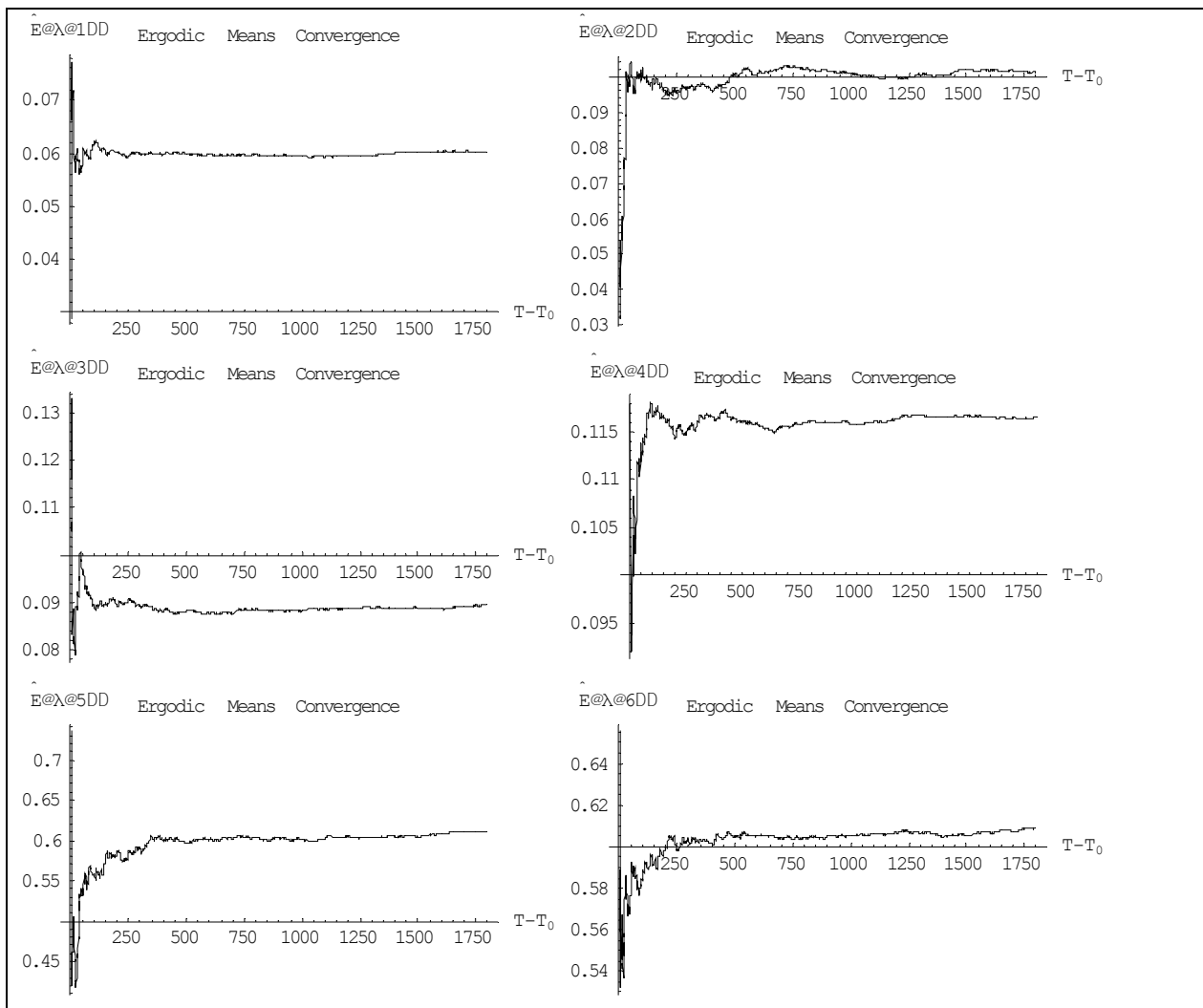
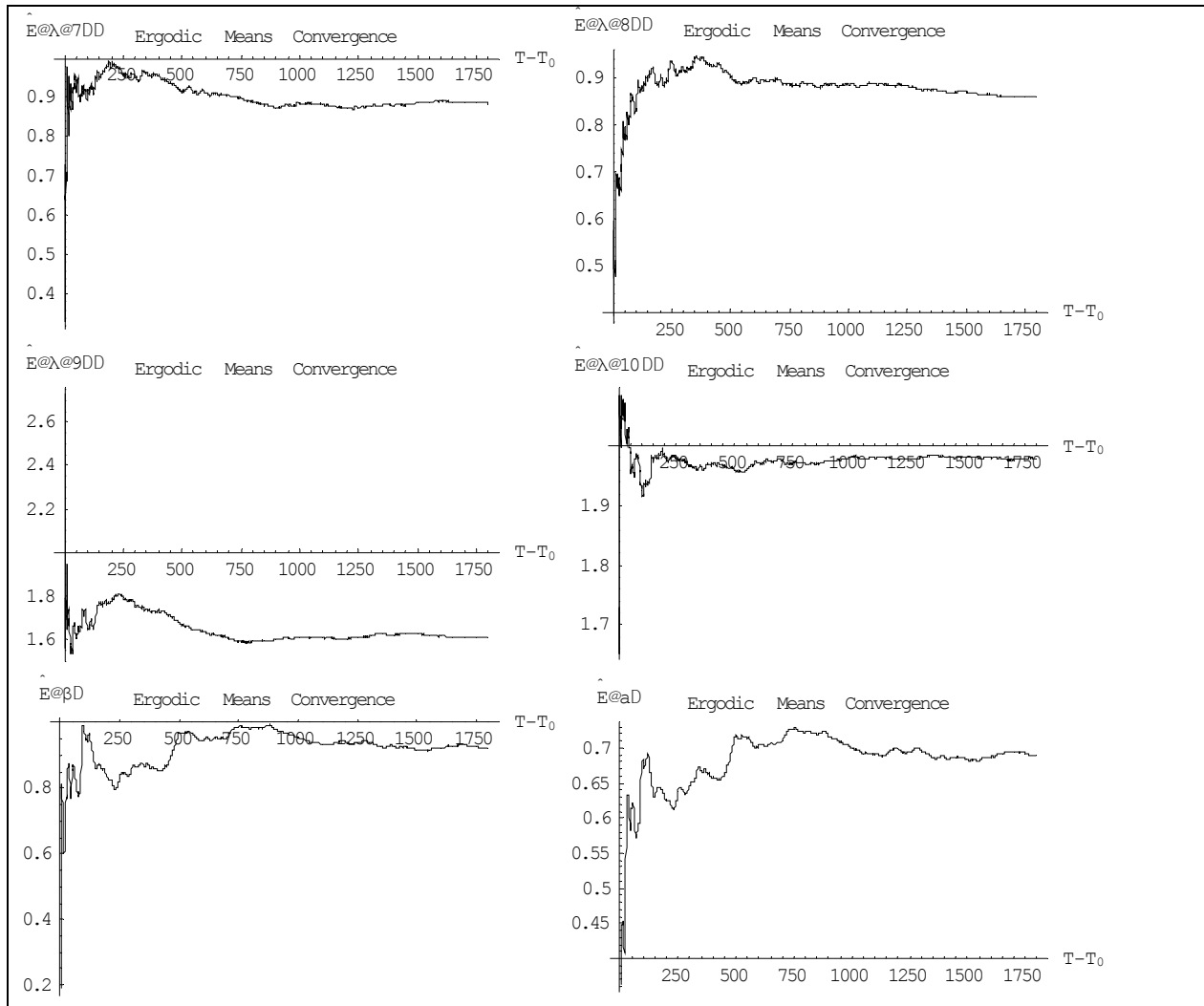
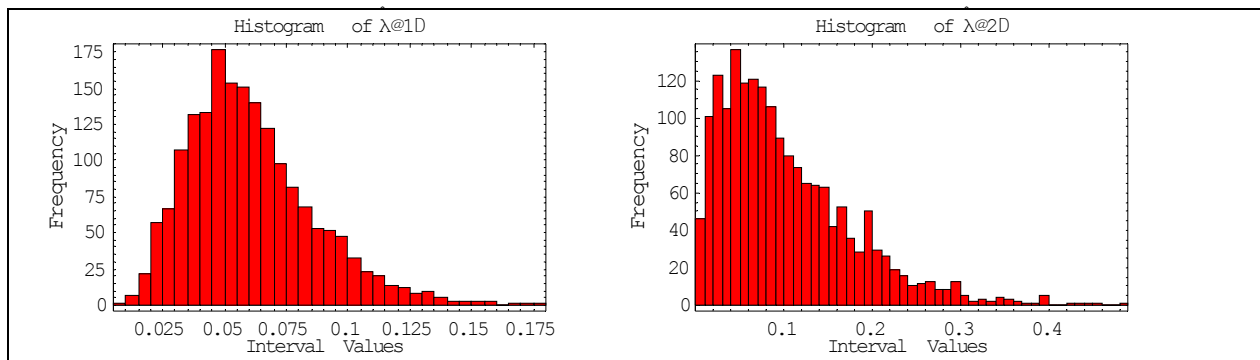


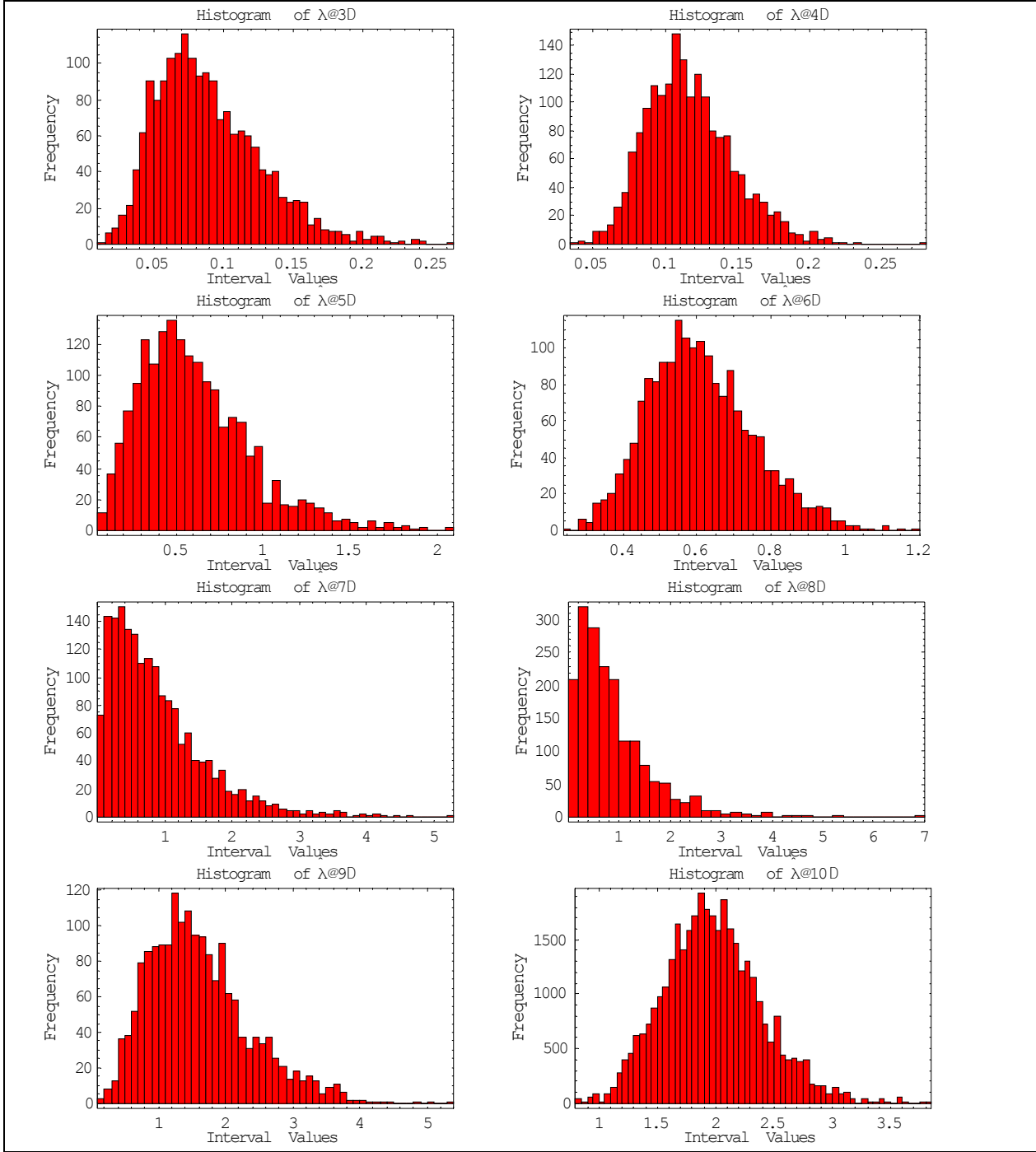
TABLE 6.3.2 Trace Plots





**TABLE 6.3.3 Convergence for Ergodic Means**





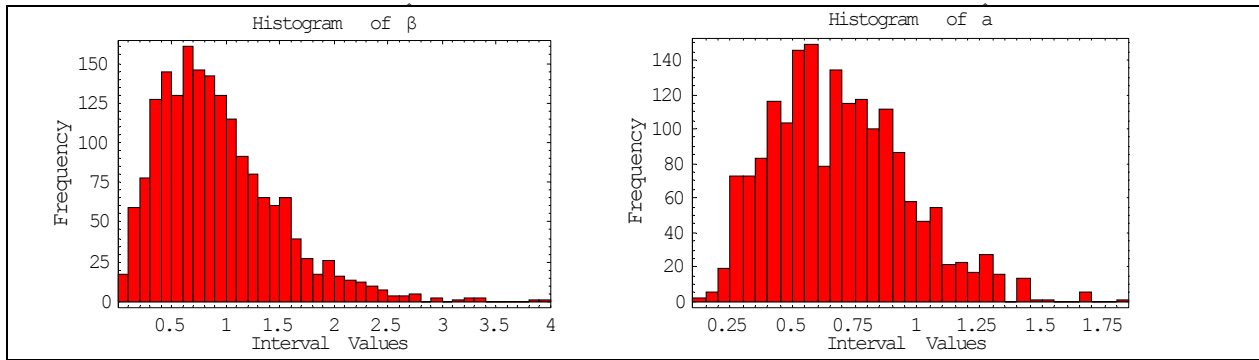
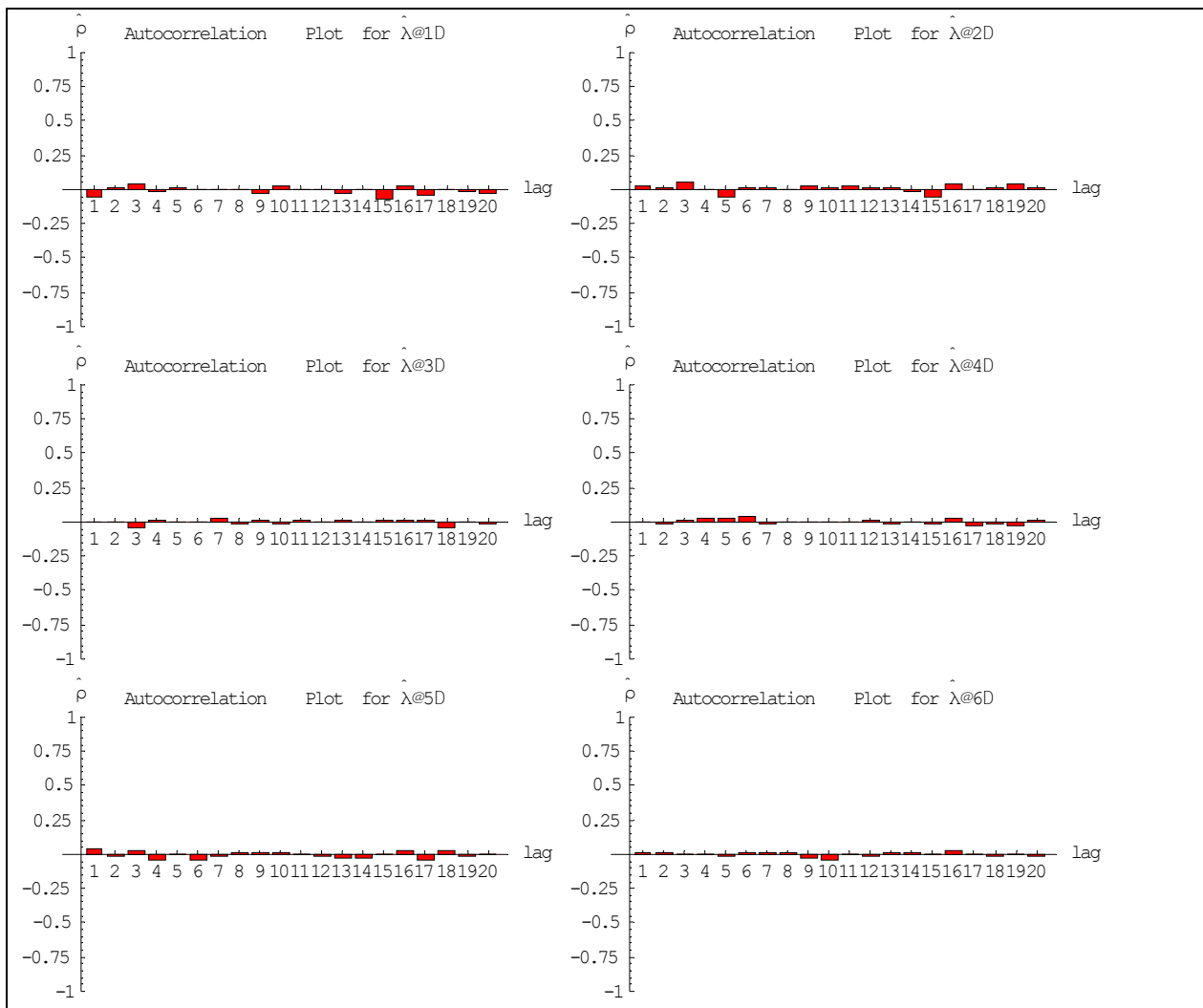
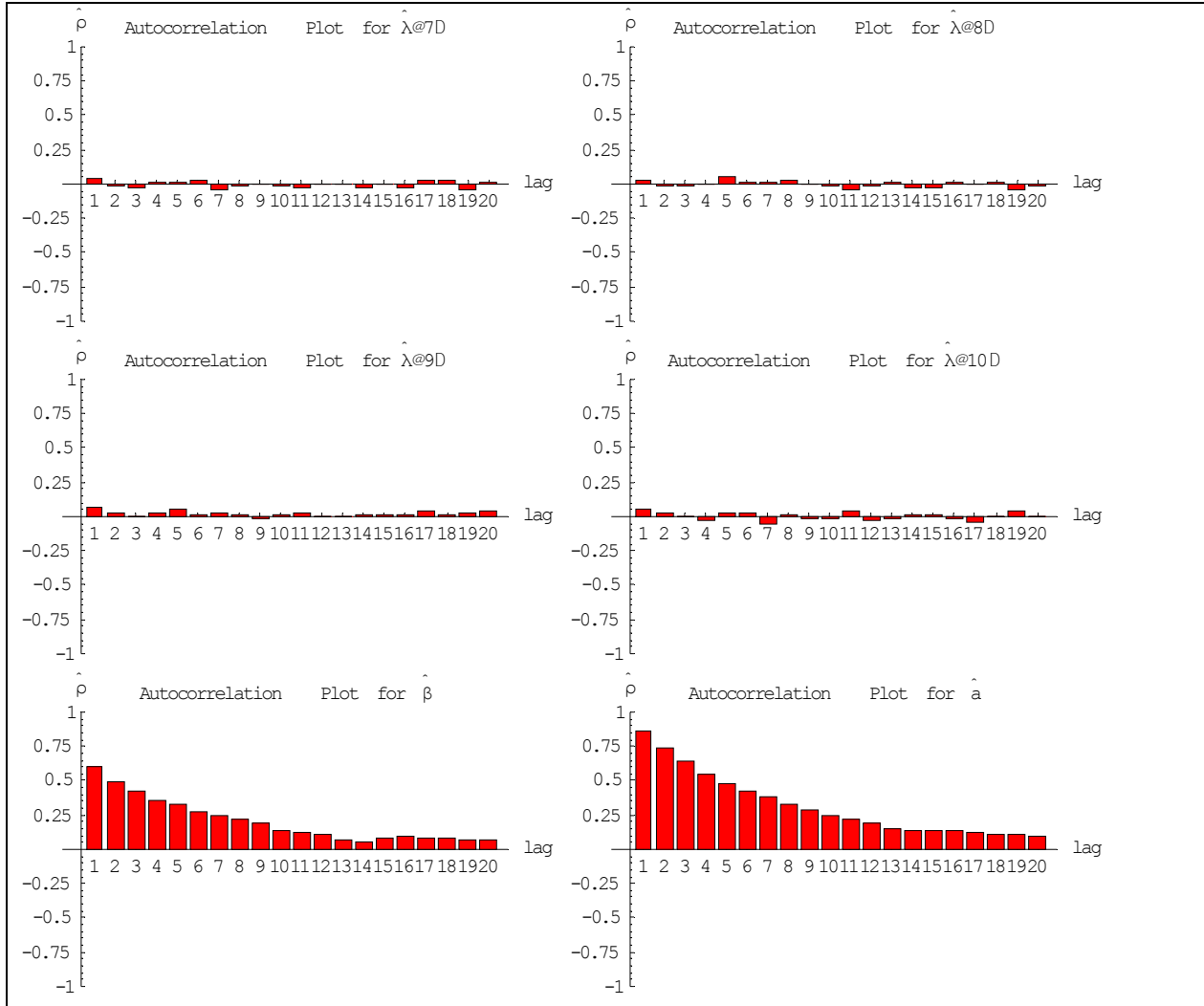


TABLE 6.3.4 Histograms





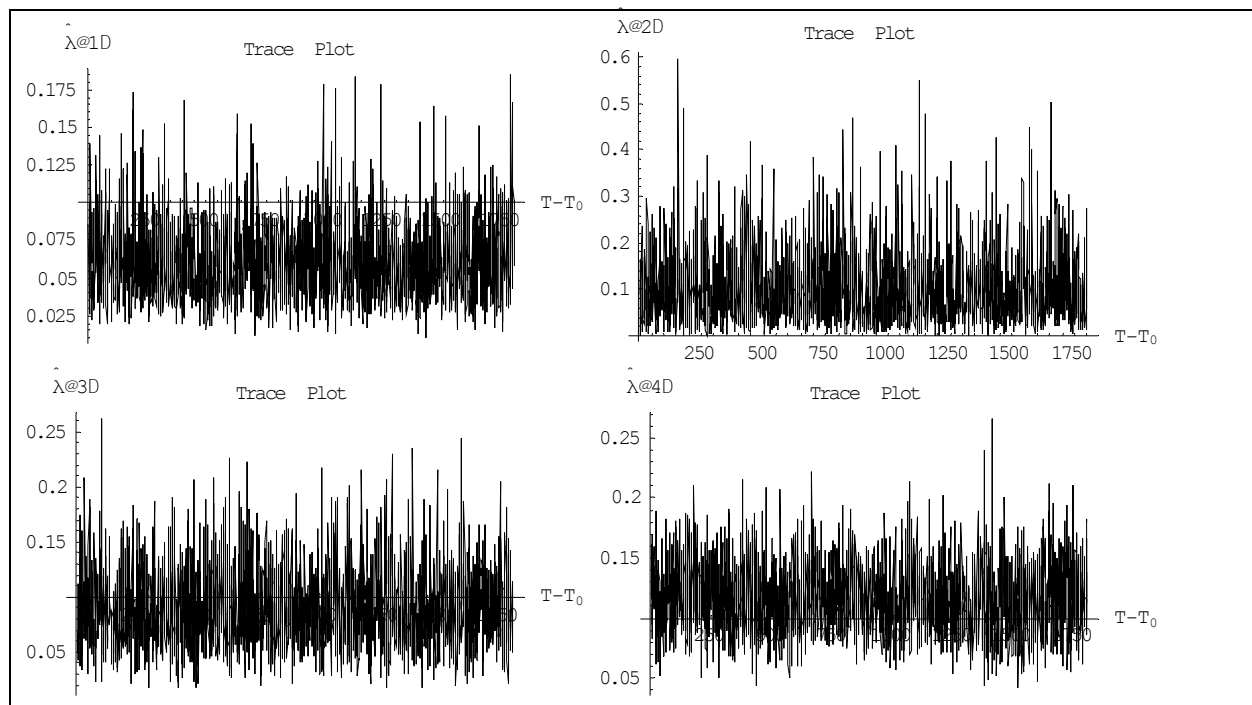


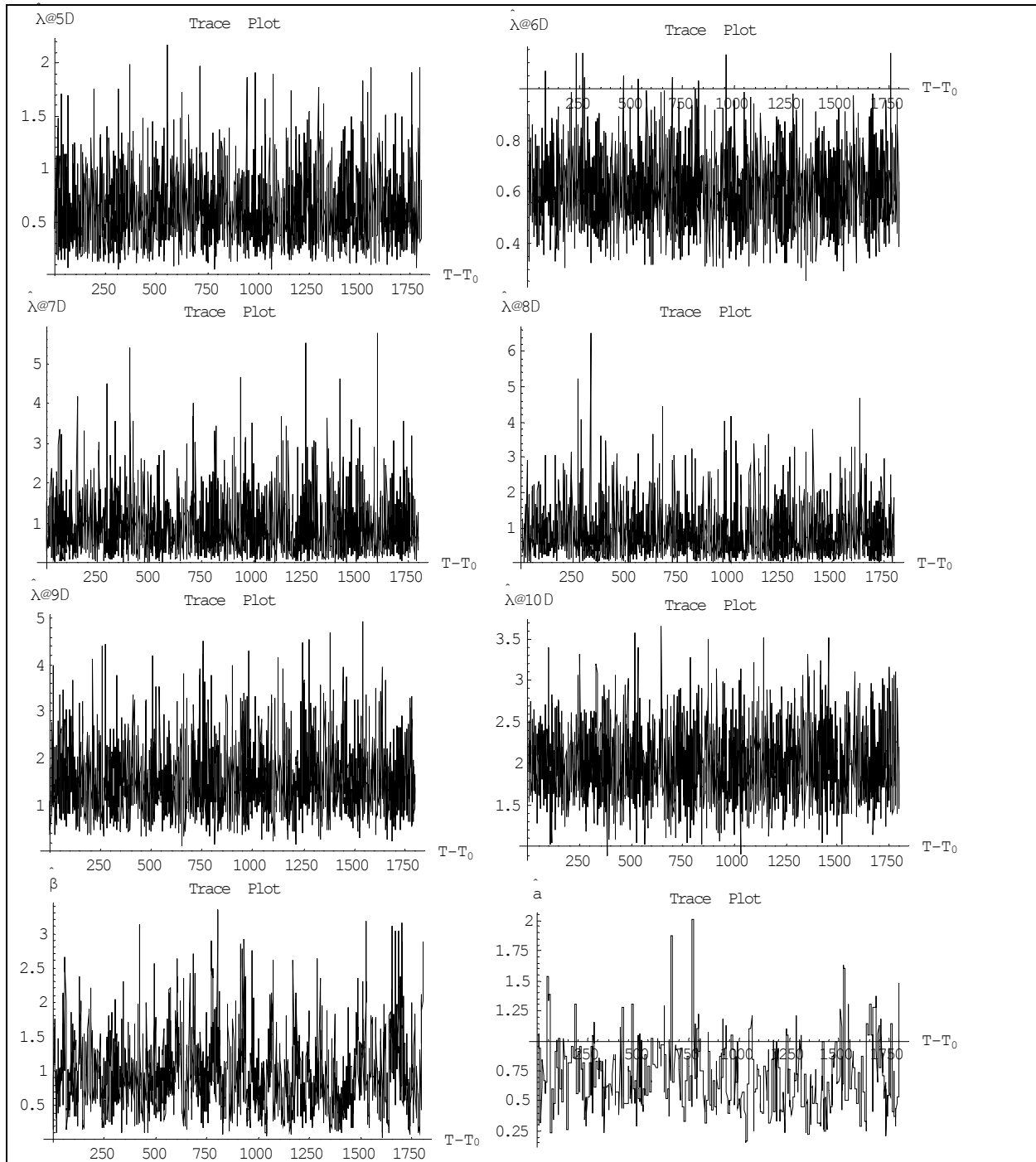
**TABLE 6.3.5 Autocorrelation Plots**

### Παράδειγμα 1.0.2 (Pump Failures) – Metropolis within Gibbs (independence case)

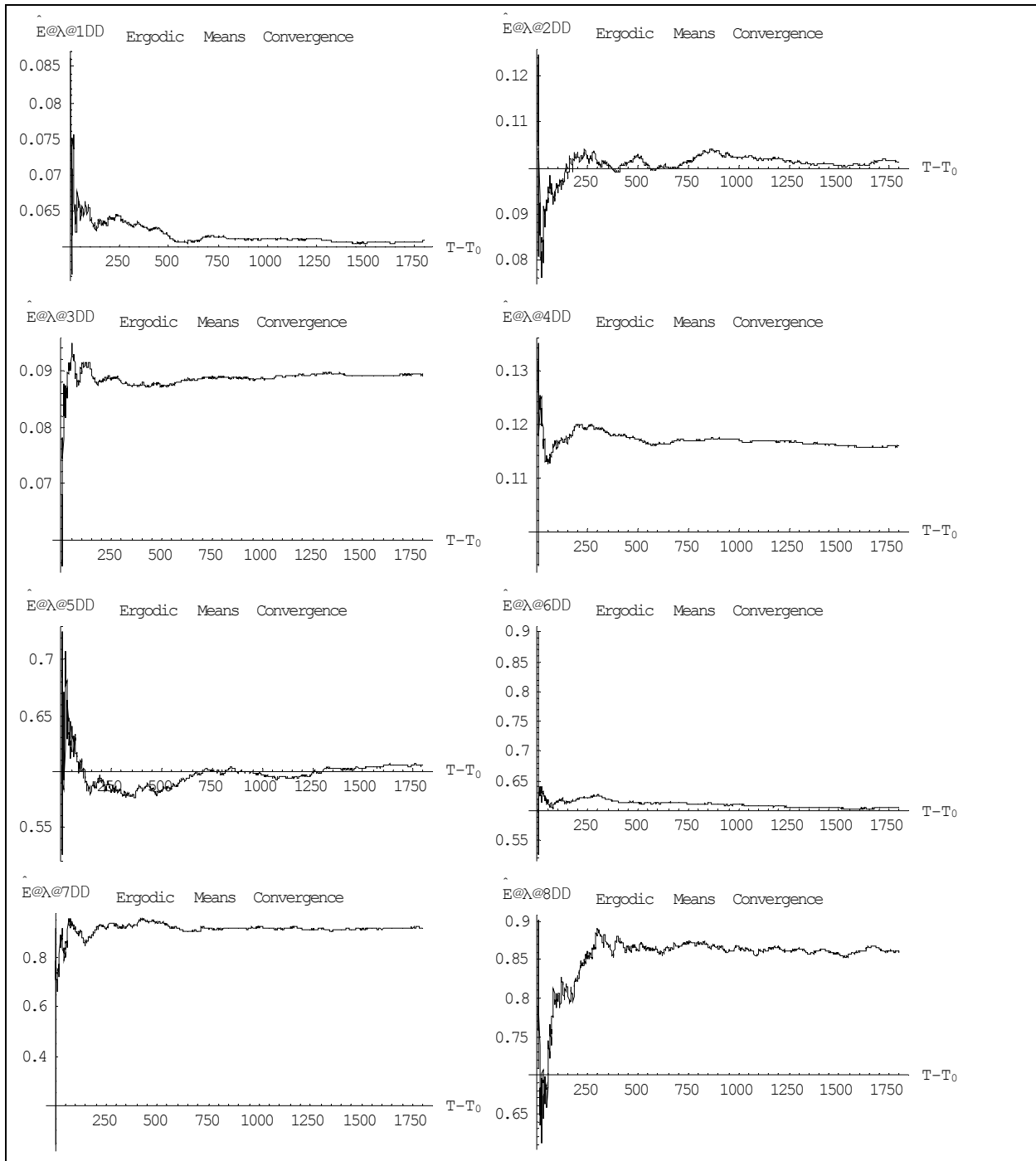
Parameter	Posterior Mean	Posterior St. Deviation	2.5% Quantile	97.5% Quantile
$\lambda_1$	0.0609741	0.0262254	0.0219618	0.119543
$\lambda_2$	0.10132	0.0808365	0.00688064	0.309028
$\lambda_3$	0.0891692	0.0373745	0.0305379	0.179608
$\lambda_4$	0.115915	0.0301816	0.0633982	0.181164
$\lambda_5$	0.605952	0.32061	0.151863	1.38188
$\lambda_6$	0.605668	0.139649	0.351441	0.916152
$\lambda_7$	0.917959	0.732204	0.0833064	2.85369
$\lambda_8$	0.860361	0.70612	0.0632053	2.77337
$\lambda_9$	1.54209	0.744426	0.44738	3.3216
$\lambda_{10}$	1.99	0.428243	1.24802	2.91793
$\beta$	0.947538	0.534508	0.18801	2.3068
$\alpha$	0.710264	0.274273	0.265939	1.30813

TABLE 6.3.6 Descriptive statistics





**TABLE 6.3.7 Trace Plots**



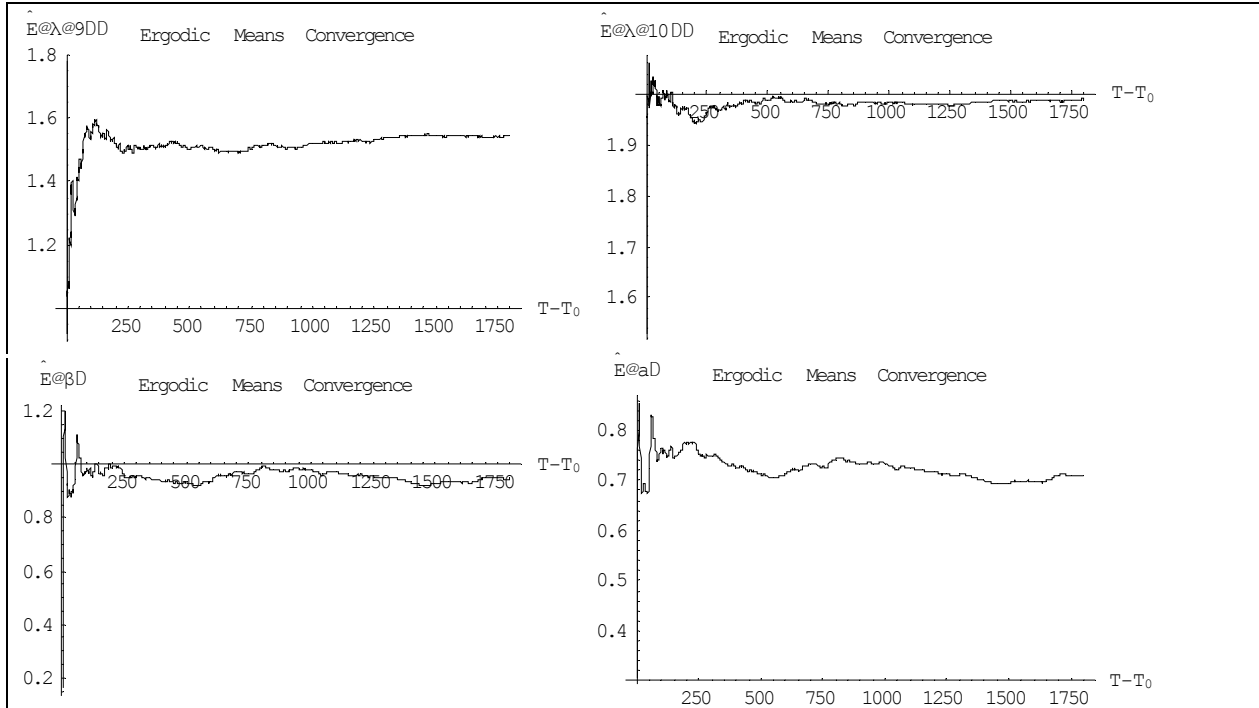
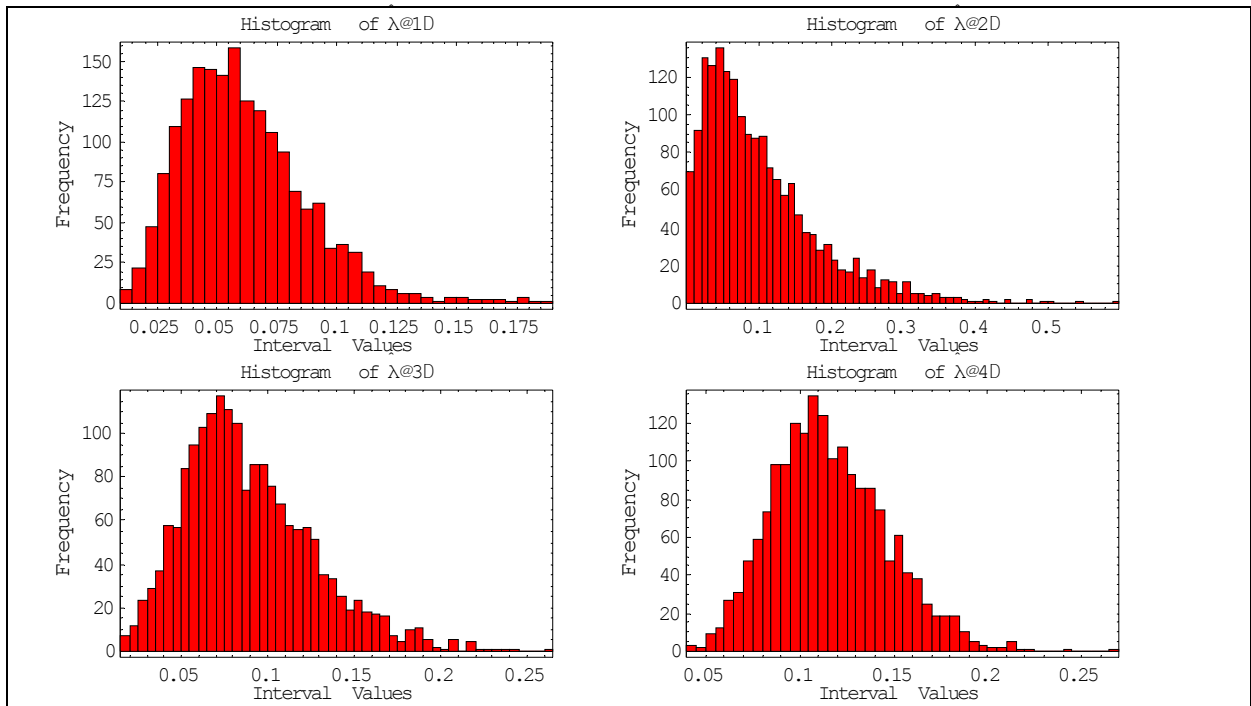
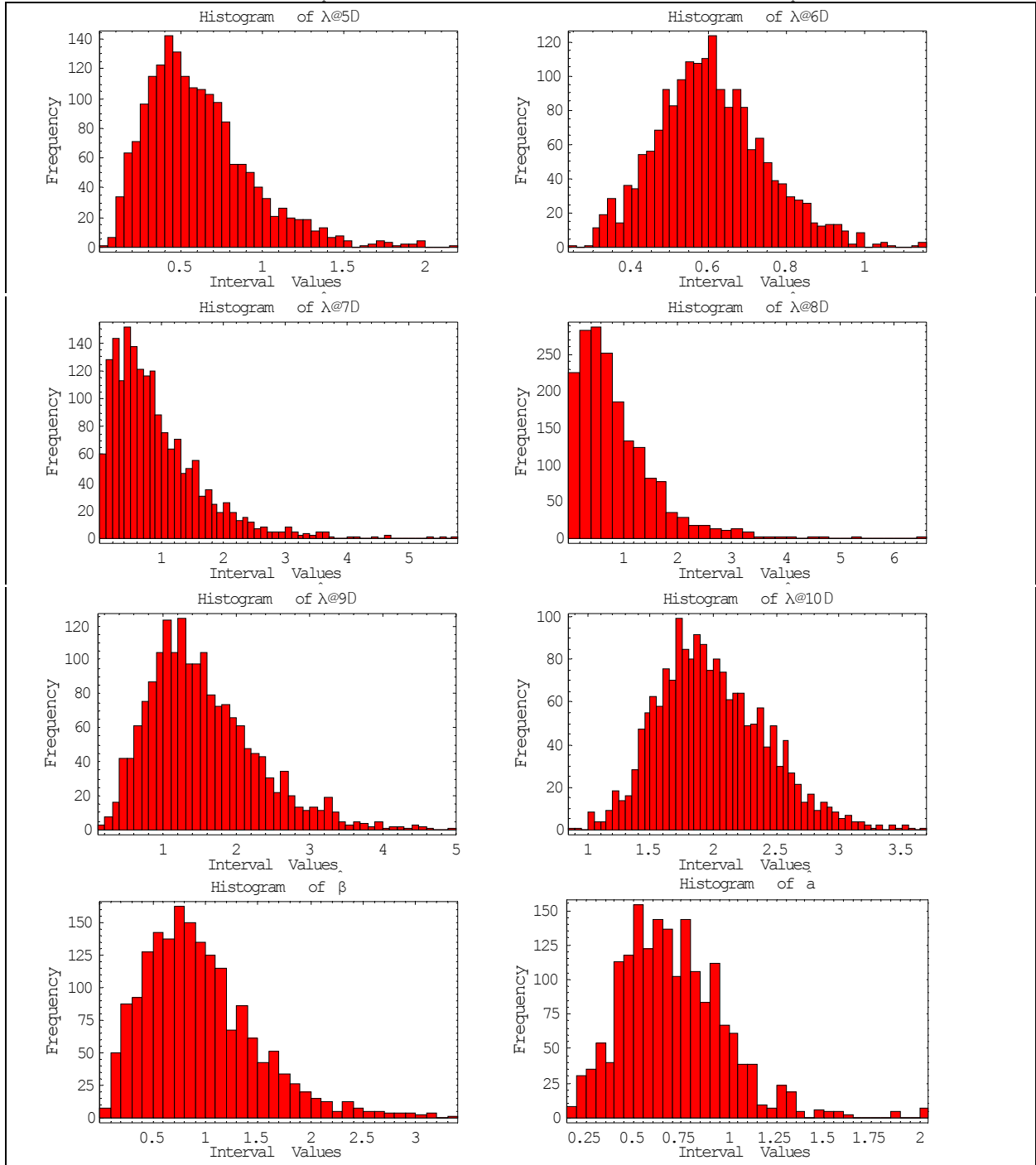
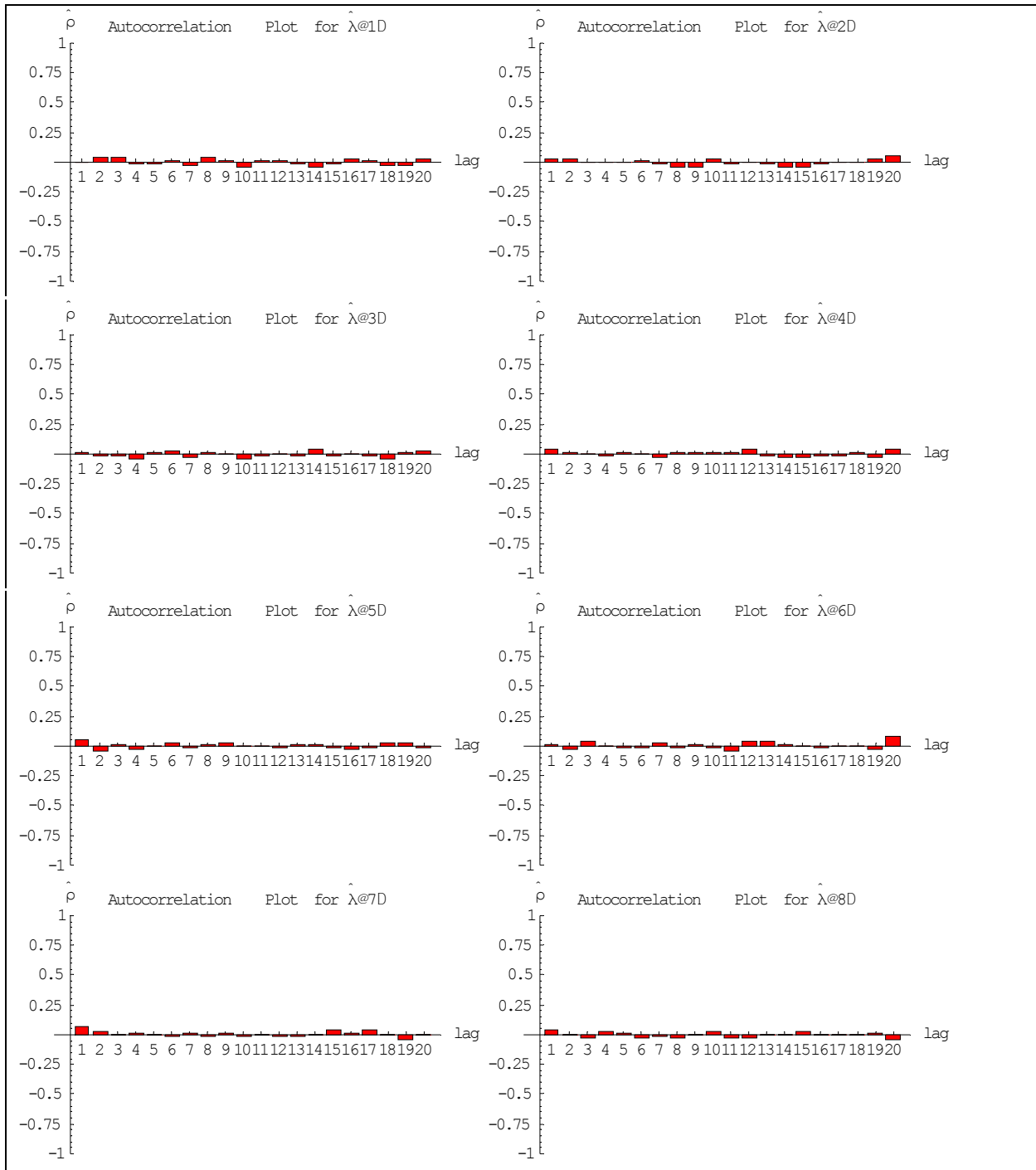


TABLE 6.3.8 Convergence for Ergodic Means





**TABLE 6.3.9 Histograms**



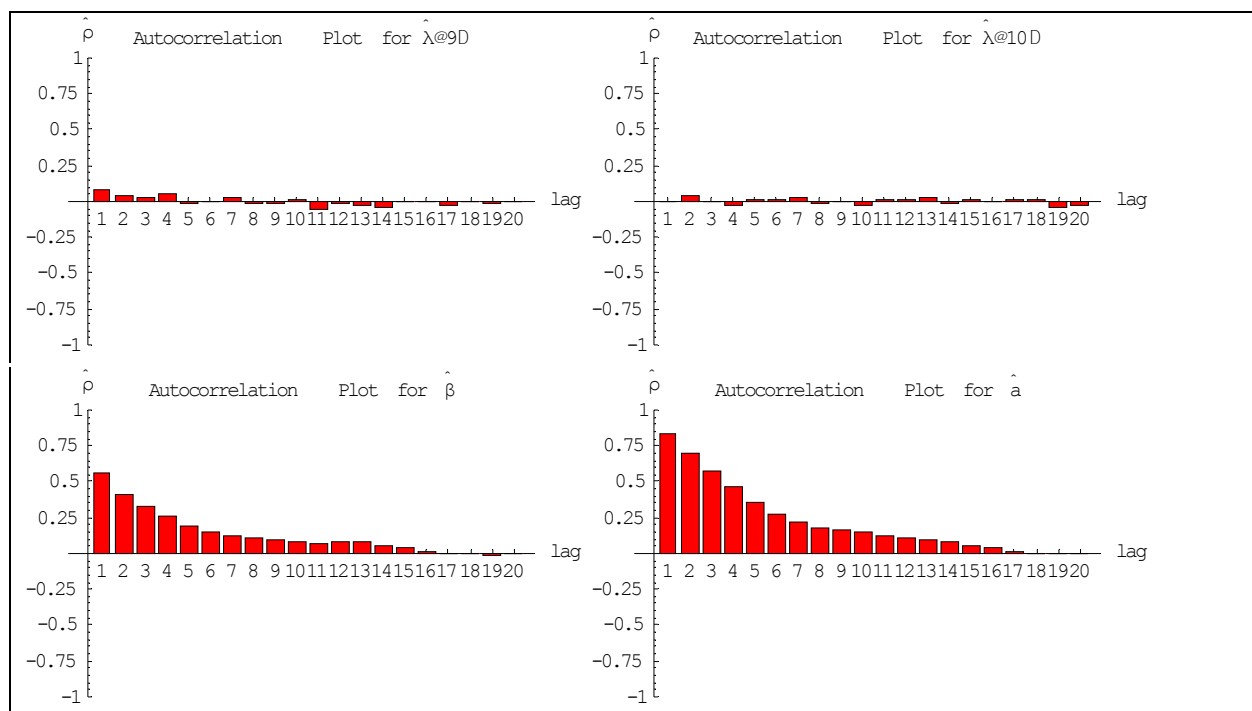


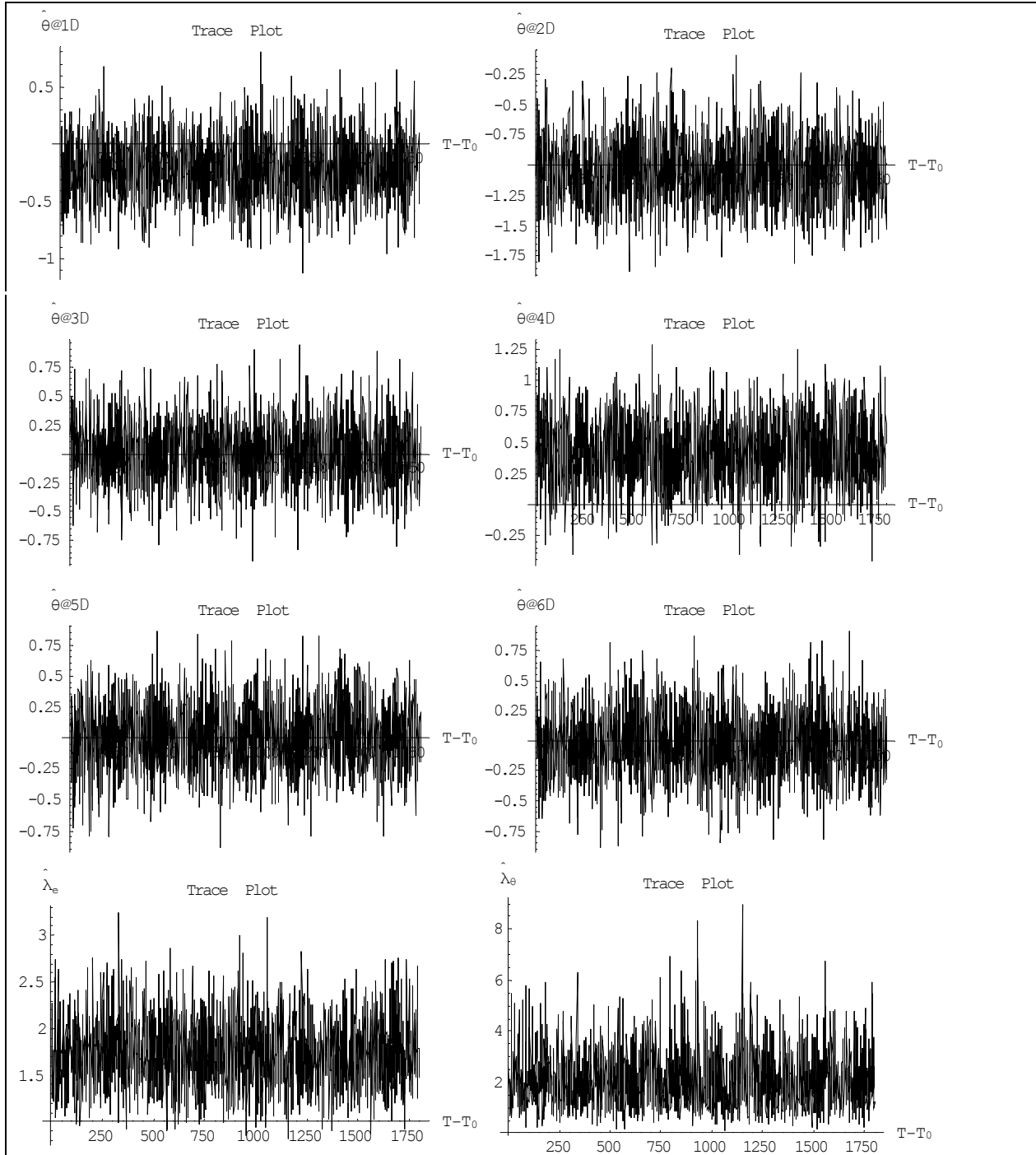
TABLE 6.3.10 Autocorrelation Plots

### Παράδειγμα 1.0.3 (Random Effect Model) –Gibbs sampler

Parameter	Posterior Mean	Posterior St. Deviation	2.5% Quantile	97.5% Quantile
$\theta_1$	-0.206835	0.263307	-0.739953	0.325852
$\theta_2$	-1.04437	0.265928	-1.5496	-0.504731
$\theta_3$	0.00867603	0.26324	-0.494242	0.527081
$\theta_4$	0.413772	0.271071	-0.125947	0.936238
$\theta_5$	0.0128829	0.259071	-0.511951	0.523845
$\theta_6$	-0.0343403	0.265229	-0.562985	0.505055
$\lambda_e$	1.73054	0.369822	1.07862	2.53002
$\lambda_\theta$	2.07683	1.12452	0.507927	4.84859
$\mu$	-0.130076	0.326398	-0.768145	0.508779

TABLE 6.4.1 Descriptive statistics





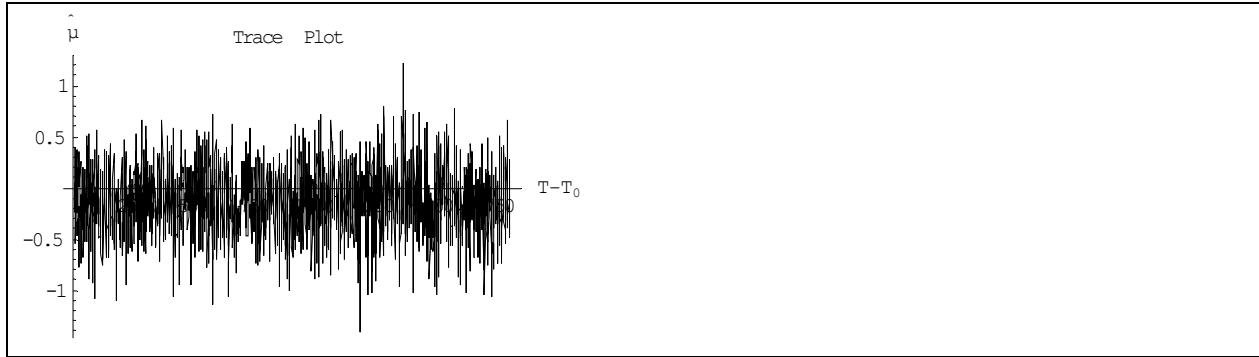
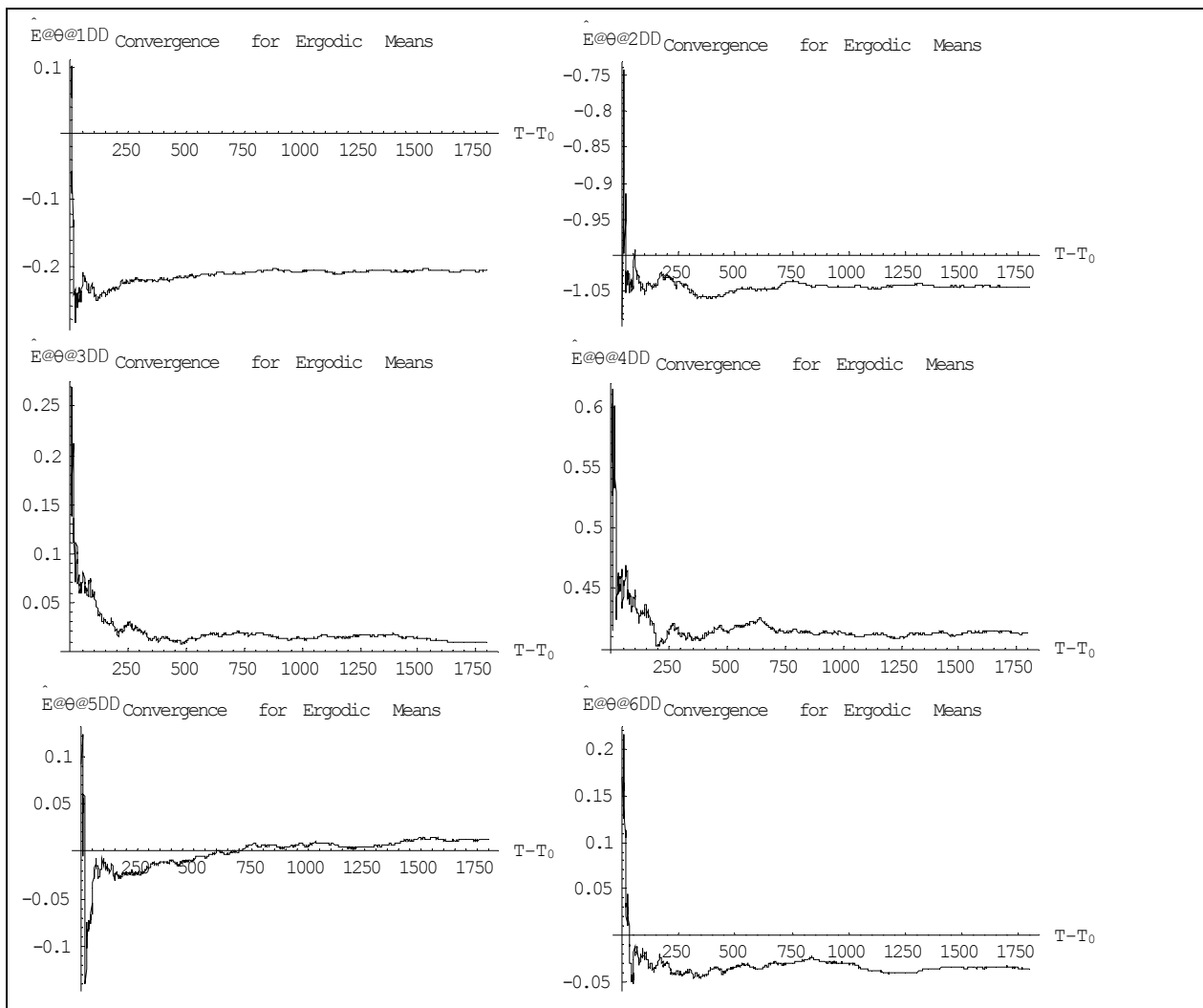


TABLE 6.4.2 Trace Plots



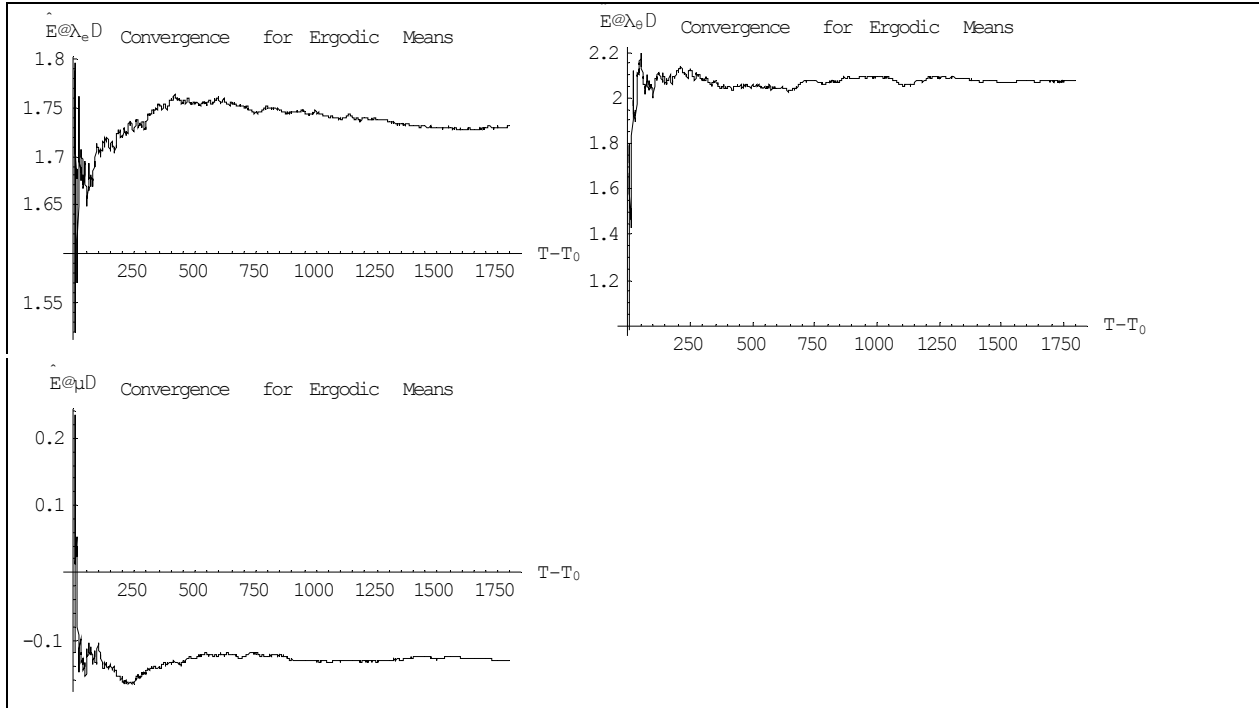
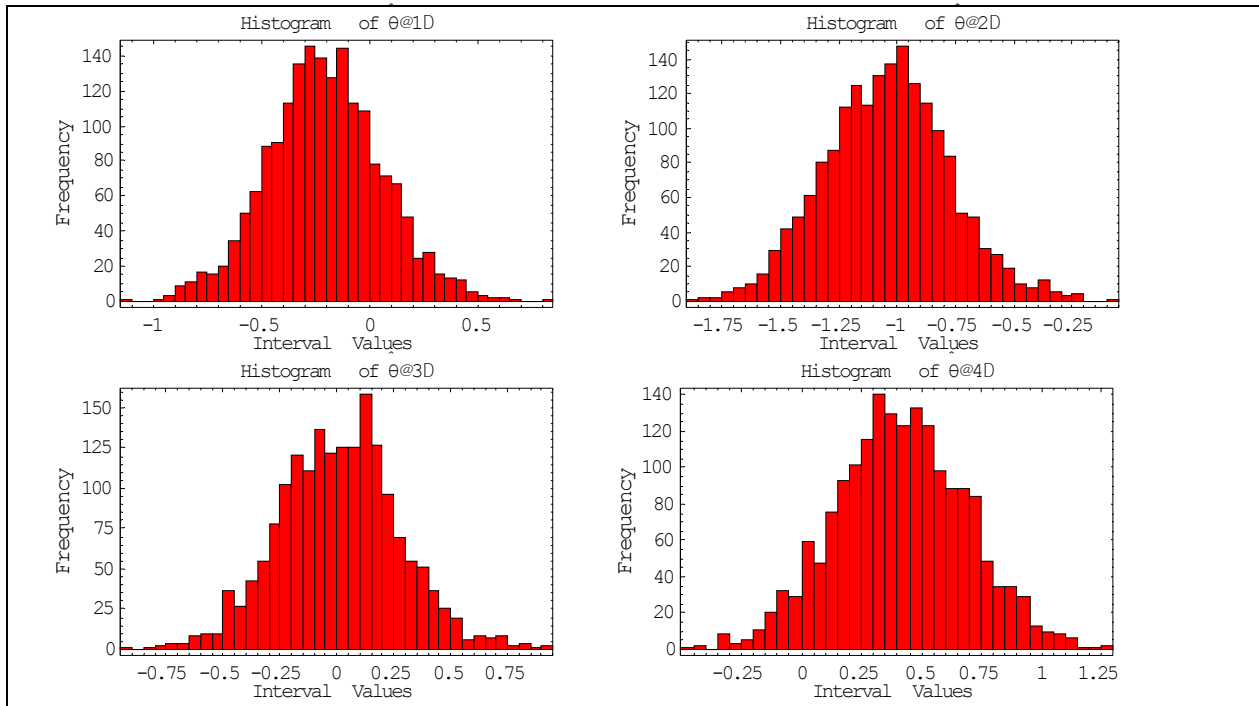
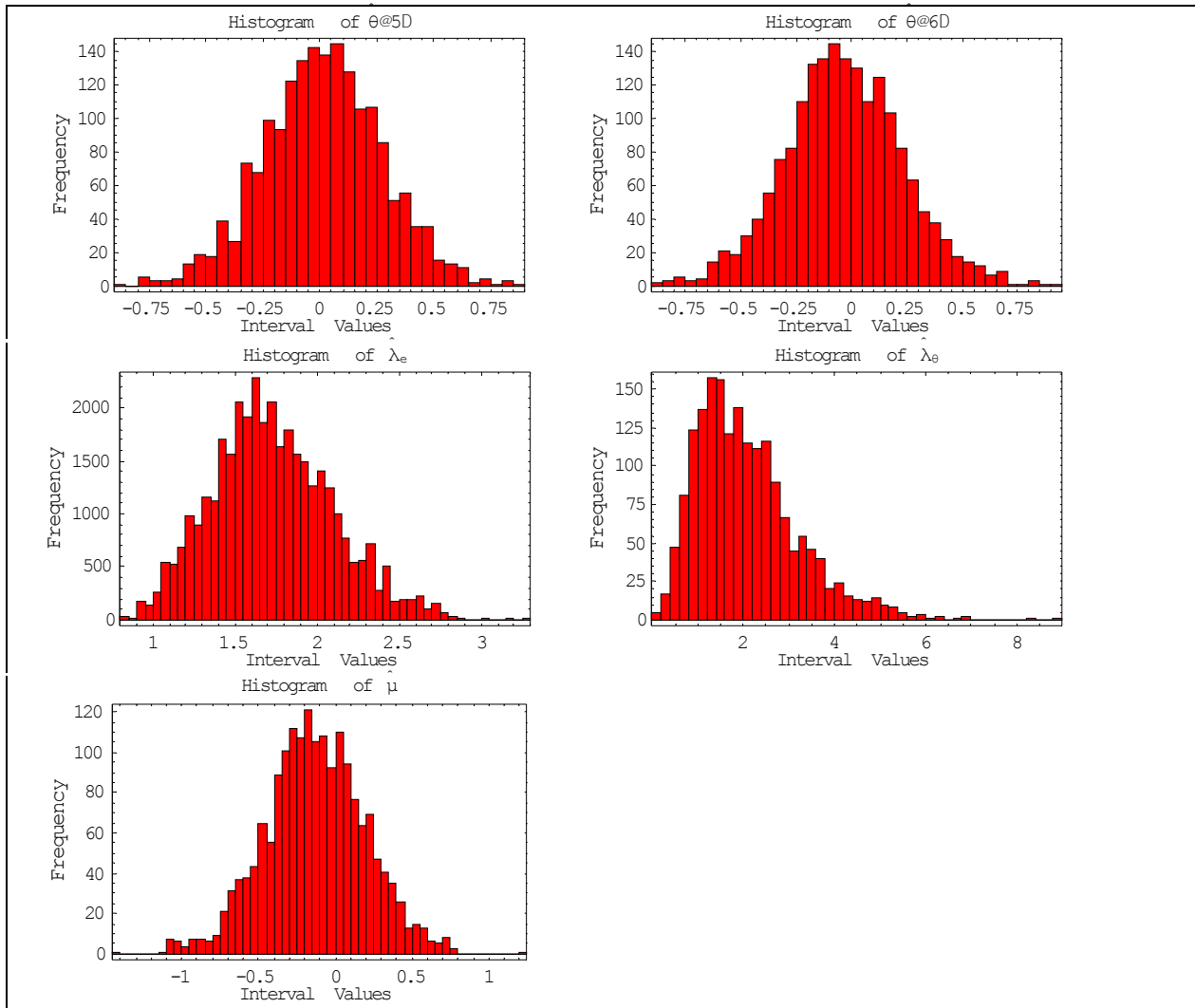
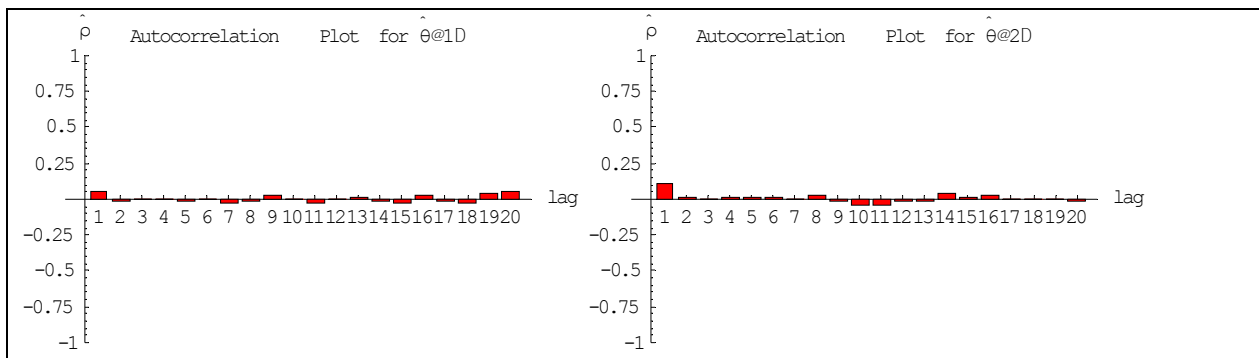


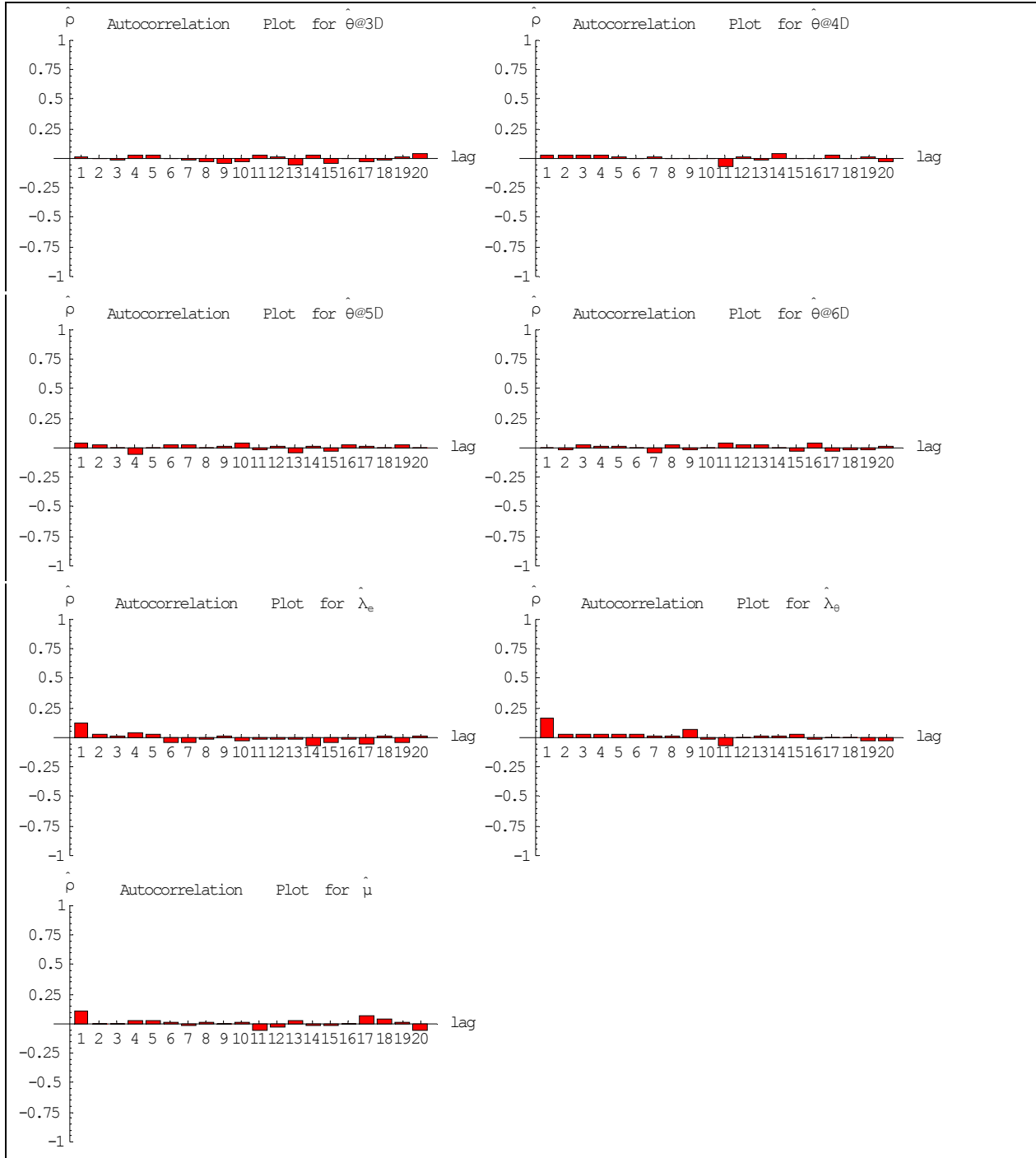
TABLE 6.4.3 Convergence for Ergodic Means





**TABLE 6.4.4 Histograms**

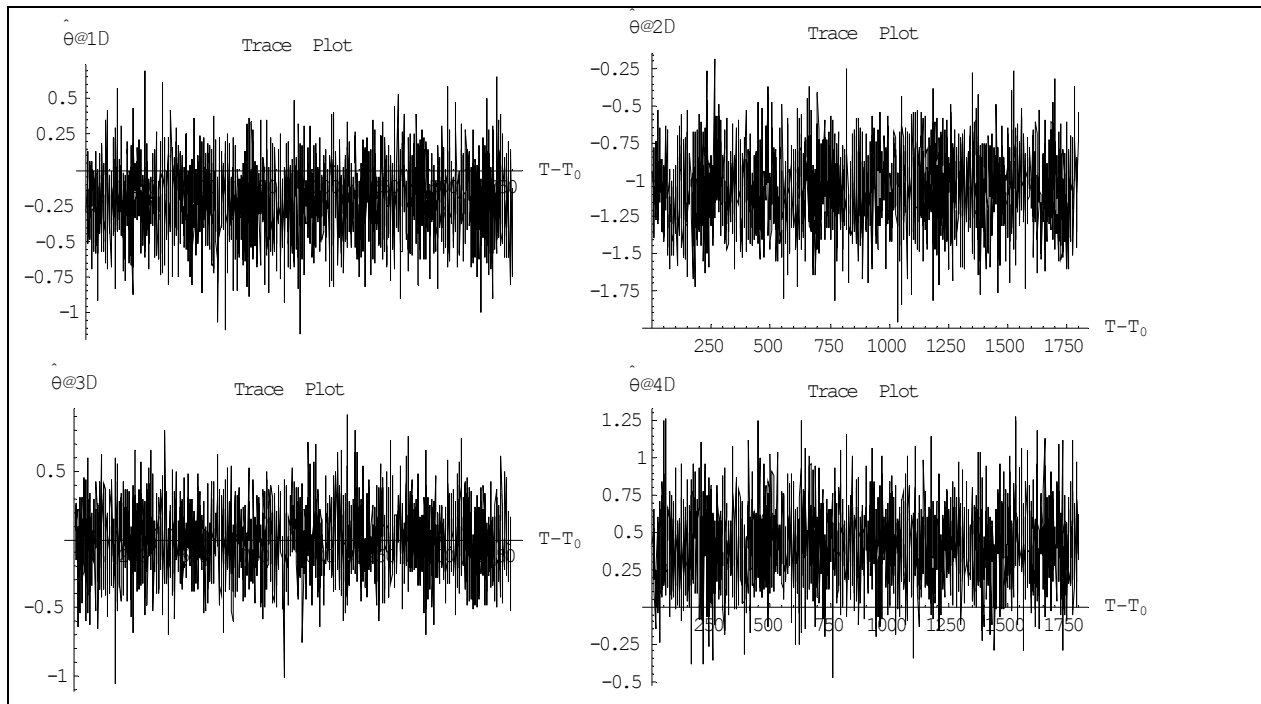


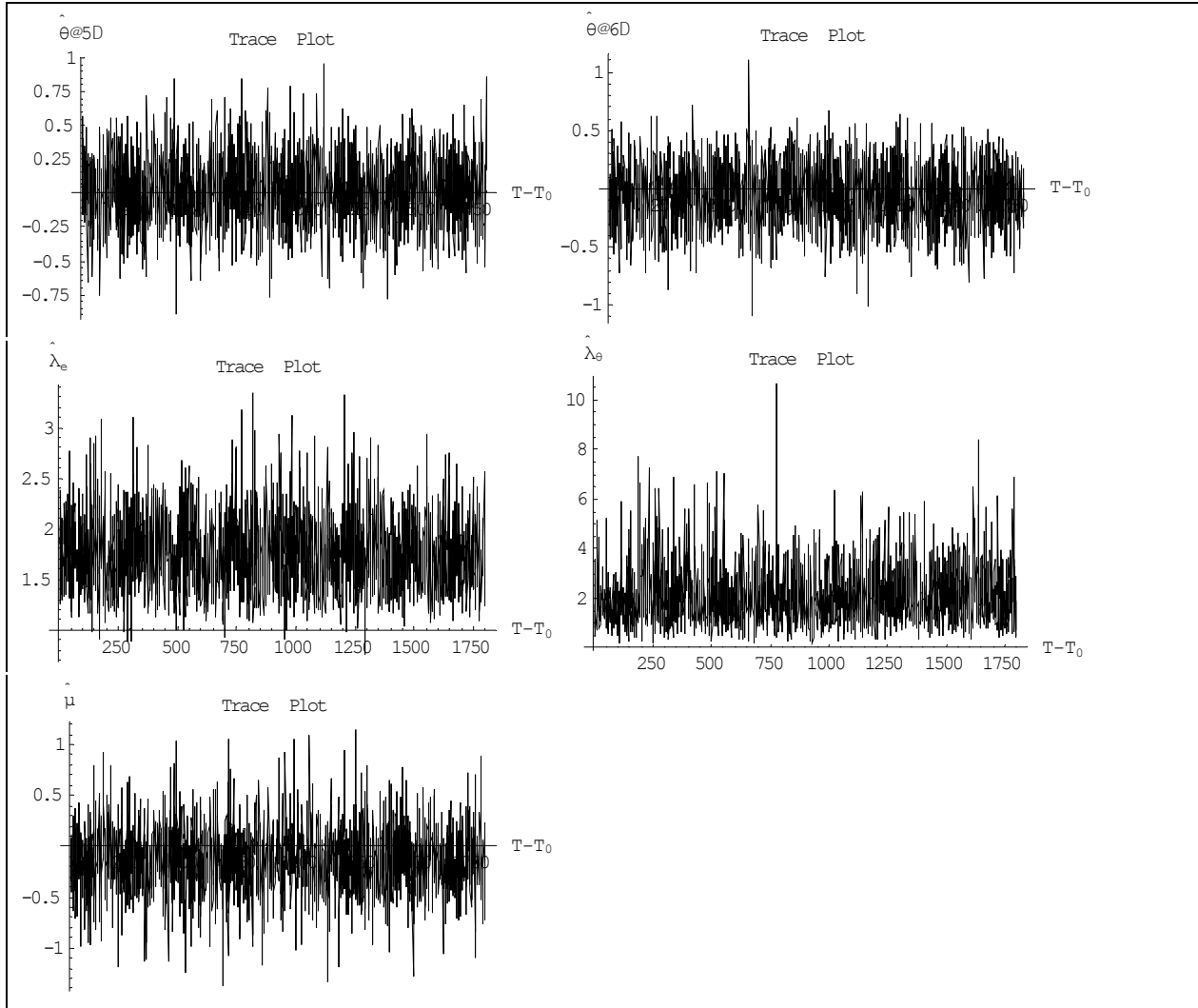


**TABLE 6.4.5 Autocorrelation Plots**

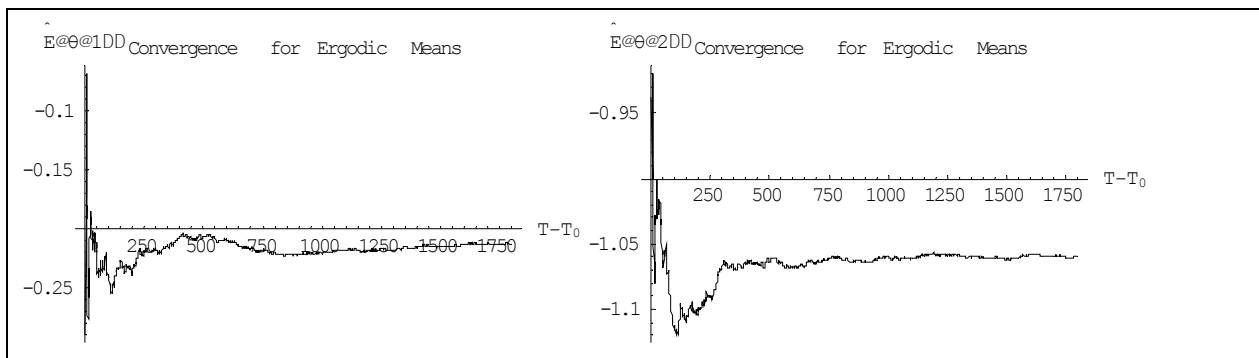
**Παράδειγμα 1.0.3 (Random Effect Model) – Blocked Gibbs sampler**

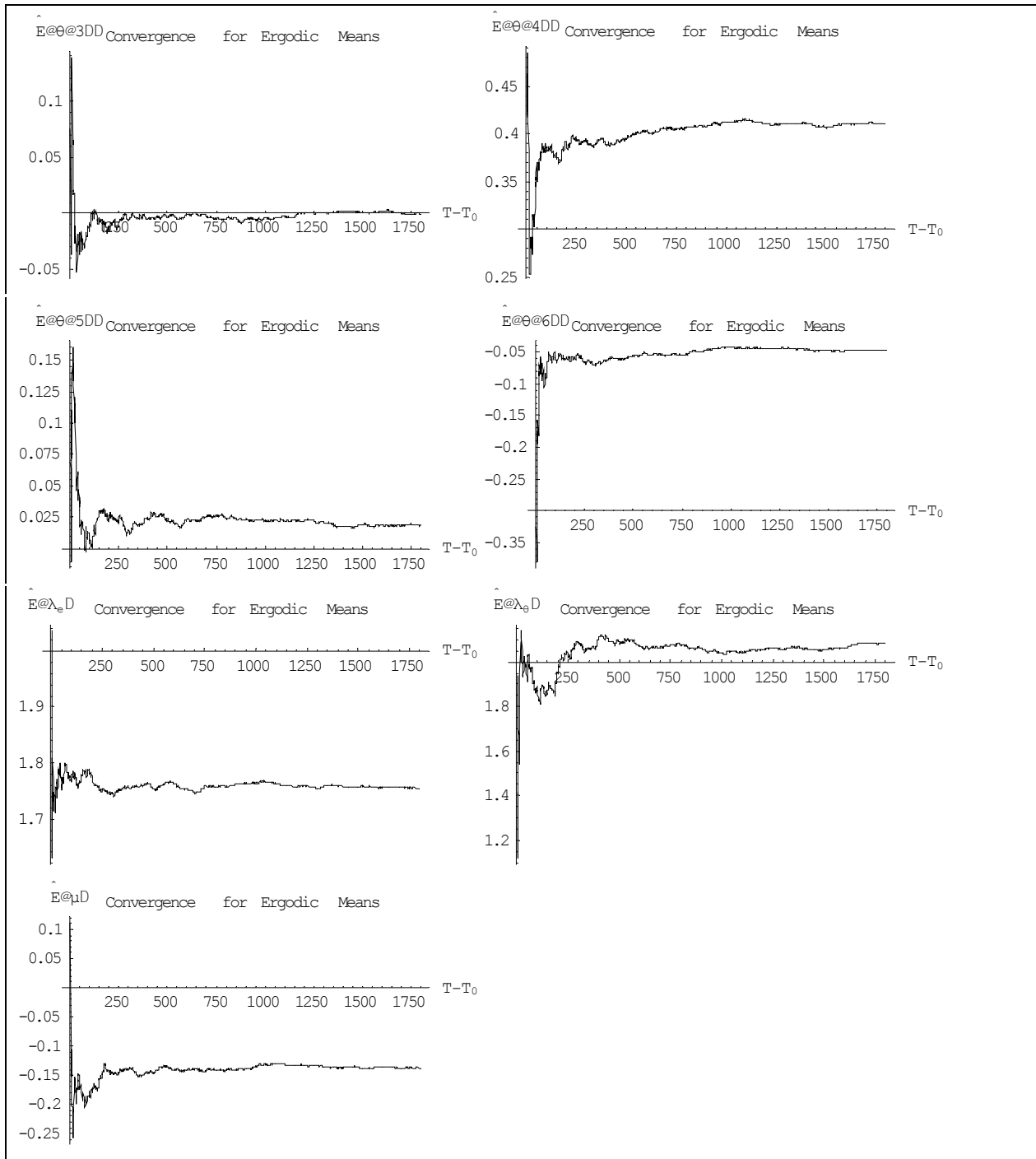
Parameter	Posterior Mean	Posterior St. Deviation	2.5% Quantile	97.5% Quantile
$\theta_1$	-0.212801	0.258344	-0.723503	0.2809
$\theta_2$	-1.05941	0.264546	-1.58235	-0.564029
$\theta_3$	-0.00019515	0.252425	-0.497621	0.488209
$\theta_4$	0.410767	0.263181	-0.0900171	0.944788
$\theta_5$	0.0187443	0.25406	-0.477989	0.514184
$\theta_6$	-0.0483057	0.26298	-0.575256	0.440069
$\lambda_e$	1.75432	0.374597	1.14602	2.57166
$\lambda_0$	2.08625	1.21363	0.4738	5.08761
$\mu$	-0.13739	0.328325	-0.808756	0.538903

**TABLE 6.4.6 Descriptive statistics**



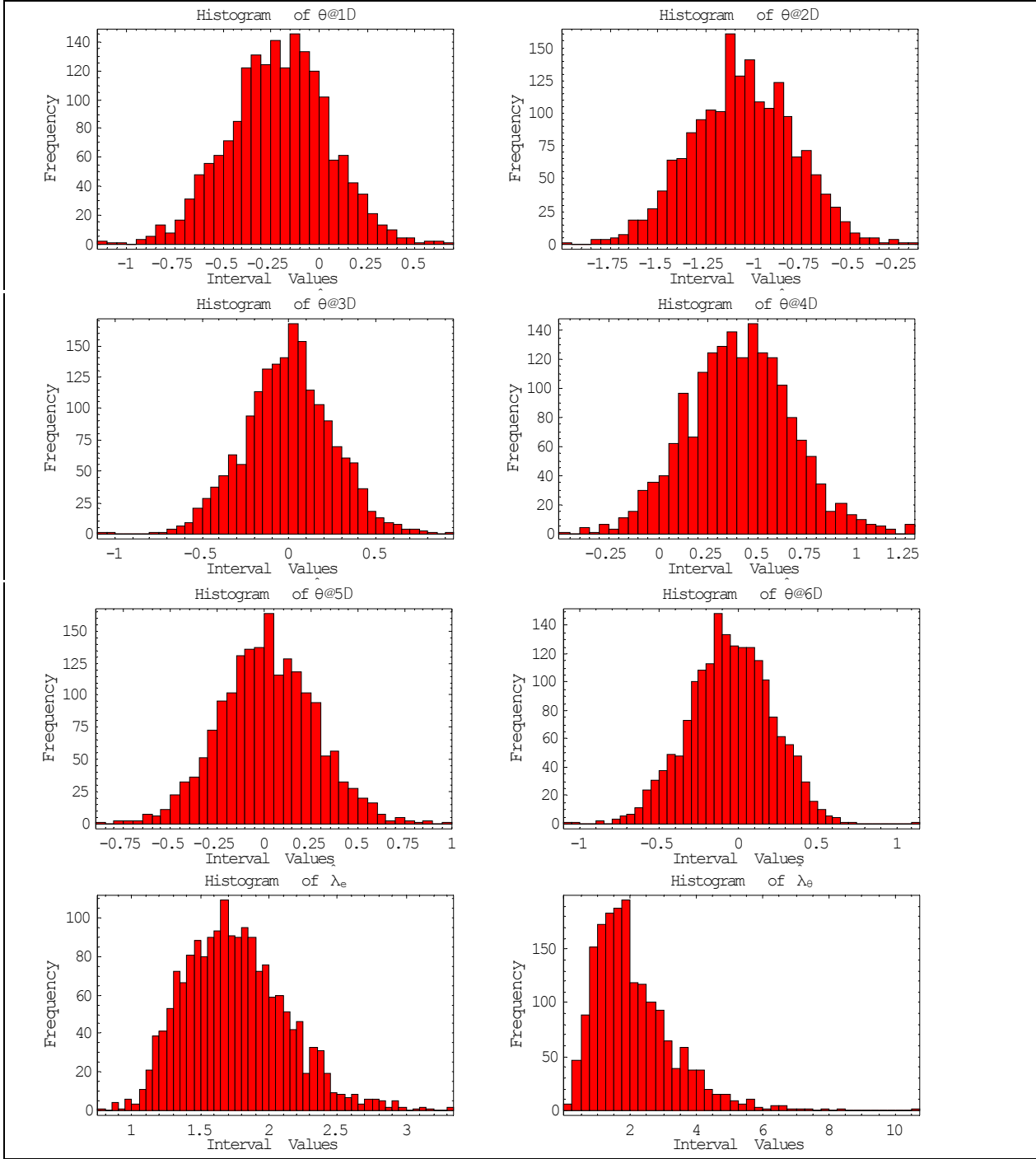
**TABLE 6.4.7 Trace Plots**





**TABLE 6.4.8 Convergence for Ergodic Means**





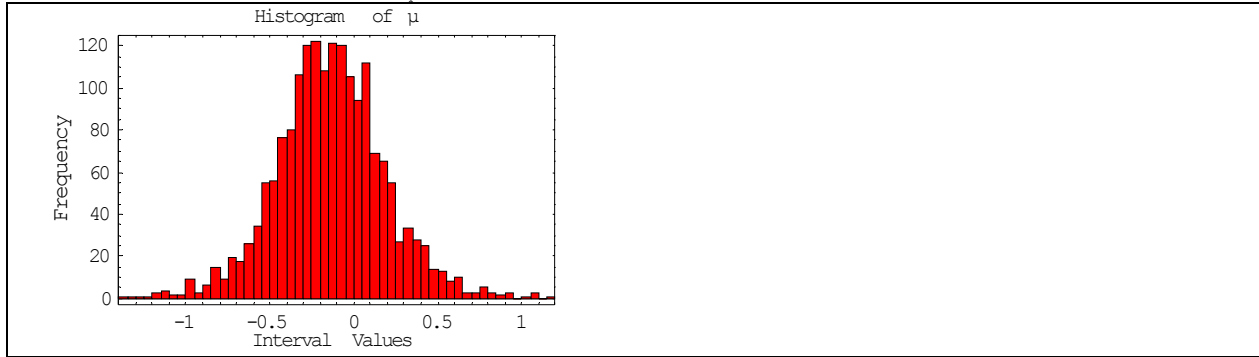
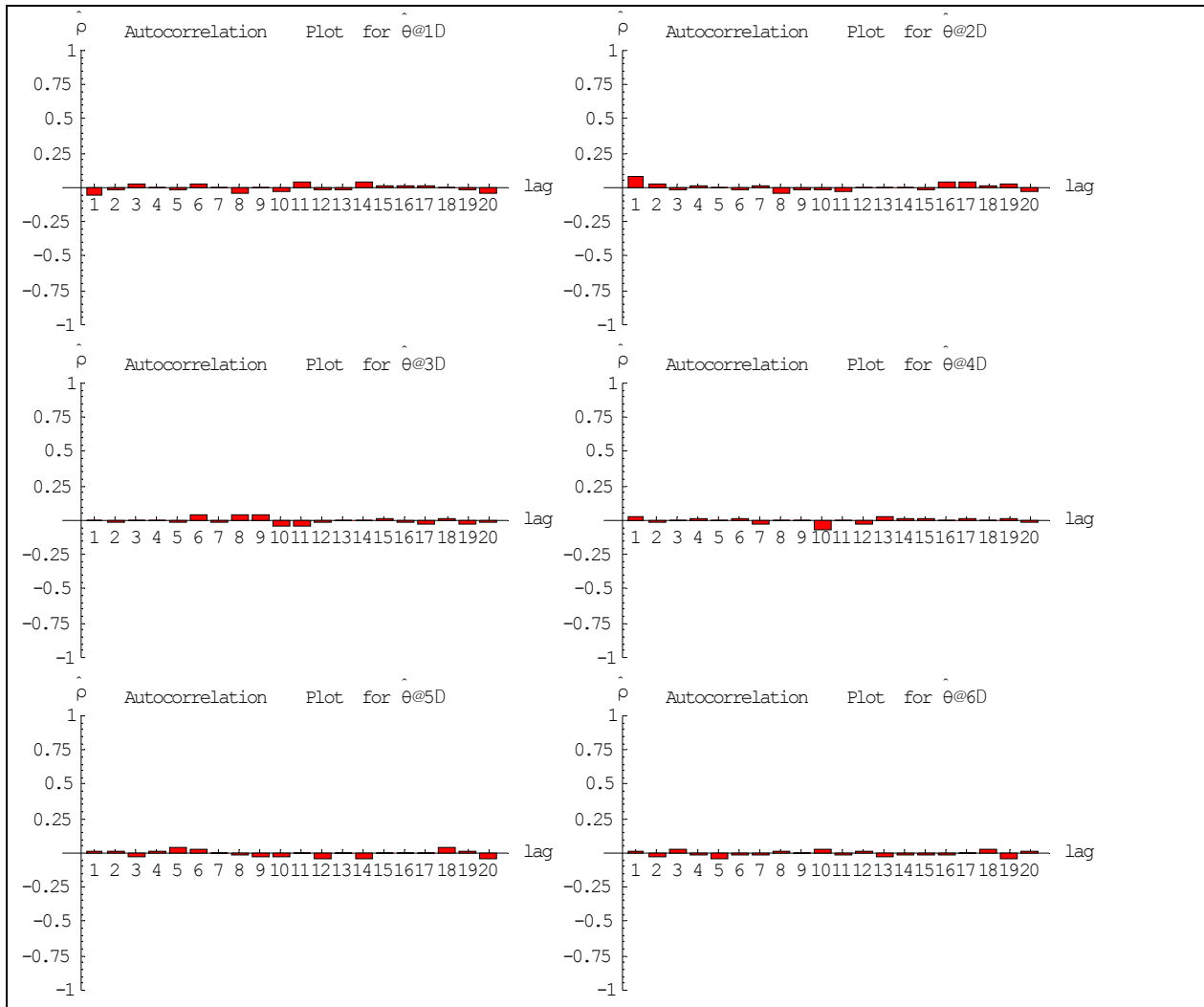
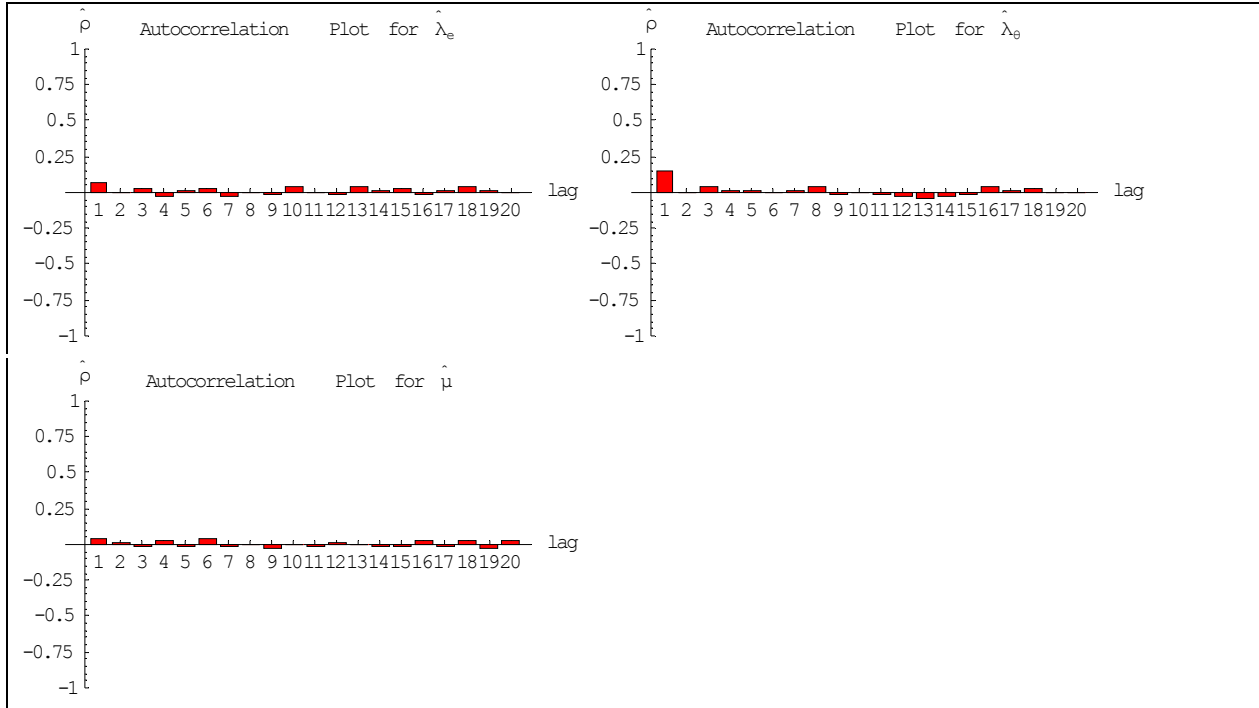


TABLE 6.4.9 Histograms





**TABLE 6.4.10 Autocorrelation Plots**