

Εθνικό και Καποδιστριακό Πανεπιστήμιο  
Αθηνών

Αλγοριθμικές τεχνικές κλασικής και μπεϋζιανής  
εκτίμησης σε μοντέλα ανάπτυξης φυτών και ζητήματα  
σύγκλισης στο σύνορο του παραμετρικού χώρου

Συγγραφέας:  
Δημήτριος Λογοθέτης

Επιβλέπων:  
Σάμης Τρέβεζας

Διπλωματική εργασία για την απόκτηση  
Μεταπτυχιακού Διπλώματος Ειδίκευσης

Τμήμα Μαθηματικών  
Τομέας Στατιστικής και Επιχειρησιακής Έρευνας

Αθήνα, 2016

Η παρούσα Διπλωματική εργασία εκπονήθηκε στα πλαίσια των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική και την Επιχειρησιακή Έρευνα που απονέμει το Τμήμα Μαθηματικών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών.

Εγκρίθηκε στις ..... από Εξεταστική Επιτροπή αποτελούμενη από τους:

Όνοματεπώνυμο	Βαθμίδα	Υπογραφή
Απόστολος Μπουρνέτας	Καθηγητής	.....
Λουκία Μελιγχοτσίδου	Επίκουρη Καθηγήτρια	.....
Σάμης Τρέβεζας (επιβλέπων)	Λέκτορας	.....

## Ευχαριστίες

Καταρχάς θα επιθυμούσα να ευχαριστήσω τον επιβλέποντα Δρ. Σάμη Τρέβεζα για την δυνατότητα που μου προσέφερε να συμμετάσχω στις ακαδημαϊκές του δραστηριότητες, για το χρόνο που αφιέρωσε ώστε να ενταχθώ σε αυτές αλλά και για την ουσιαστική καθοδήγηση που μου προσέφερε γενικότερα.

Θα ήθελα επίσης να ευχαριστήσω ιδιαίτερα τον αναπληρωτή καθηγητή Αντώνιο Οικονόμου καθώς υπήρξε καθοριστικός παράγοντας στην επιλογή μου να συμμετάσχω στο συγκεκριμένο μεταπτυχιακό πρόγραμμα και κατά τη διάρκεια του οποίου είχα την αμέριστη βοήθεια του και τη συνεχόμενη ενίσχυση του μαθηματικού μου ενδιαφέροντος.

Στο μεταπτυχιακό πρόγραμμα αυτό, θεωρώ ότι αποκόμισα όσα χρειαζόμουν για να επιδιώξω τους στόχους μου. Πάντα οι πόρτες των καθηγητών ήταν ανοιχτές και για το λόγο αυτό τους ευχαριστώ όλους. Κυρίως όμως τον καθηγητή Απόστολο Μπουρνέτα και την επίκουρη καθηγήτρια Λουκία Μελιγκοτσίδου που με τίμησαν με τη συμμετοχή τους στην εξεταστική επιτροπή αλλά και γενικότερα για την καθοδήγηση και την υποστήριξη που μου προσέφεραν απλόχερα. Ακόμη, θα ήθελα να ευχαριστήσω την επίκουρη καθηγήτρια του πανεπιστημίου Πατρών, Σόνια Μαλεφάκη, για την πολύτιμη συνεργασία που είχαμε.

Τέλος, ευχαριστώ την Παρασκευή για την υποστήριξη της όπως επίσης την οικογένεια και τους φίλους μου που ήταν πάντα στο πλευρό μου.

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>5</b>
<b>2</b>	<b>Δύο μοντέλα ανάπτυξης φυτών</b>	<b>8</b>
2.1	Εισαγωγή . . . . .	8
2.2	Το μοντέλο GreenLab . . . . .	8
2.2.1	Διανομή της Βιομάζας . . . . .	9
2.2.2	Παραγωγή της Βιομάζας . . . . .	10
2.3	Εκτίμηση των καταστάσεων και των παραμέτρων. . . . .	11
2.4	Το μοντέλο LNAS . . . . .	15
2.4.1	Παραγωγή Βιομάζας . . . . .	15
2.4.2	Διανομή της Βιομάζας . . . . .	15
2.4.3	Διαδικασία Γήρανσης . . . . .	16
2.4.4	Αναπαράσταση ως μοντέλο χώρου καταστάσεων . . . . .	17
<b>3</b>	<b>Στοιχεία θεωρίας μαρκοβιανών αλυσίδων σε γενικούς χώρους</b>	<b>19</b>
3.1	Εισαγωγή . . . . .	19
3.1.1	Ο Πυρήνας μιας Στοχαστικής Διαδικασίας και οι βασικότερες ιδιότητες του . . . . .	20
3.1.2	Πυρήνας και Τυχαίες Μεταβλητές . . . . .	22
3.1.3	Ορισμοί-Βασικές έννοιες μ.α. . . . .	25
3.2	Ασυμπτωτική Θεωρία M.A. . . . .	31
3.2.1	Στάσιμο μέτρο . . . . .	31
3.2.2	Βασικά χαρακτηριστικά-έννοιες M.A. . . . .	34
3.2.3	Οριακά Θεωρήματα . . . . .	39
3.3	Κρυμμένα Μαρκοβιανά μοντέλα (KMM) . . . . .	44
3.3.1	Εισαγωγή . . . . .	44
3.3.2	Ορισμός-Βασικές έννοιες . . . . .	45
3.3.3	Βασικές ιδιότητες . . . . .	48
3.3.4	Βασικά Προβλήματα στα KMM . . . . .	49
3.4	Σύνδεση με τη Στατιστική . . . . .	50
<b>4</b>	<b>Μέθοδοι Markov Chain Monte Carlo (MCMC)</b>	<b>53</b>
4.1	Εισαγωγή στις Μεθόδους MCMC . . . . .	53
4.1.1	Ο αλγόριθμος Metropolis-Hastings (MH) . . . . .	54
4.1.2	Ο δειγματολήπτης Gibbs . . . . .	62
4.1.3	Ο υβριδικός δειγματολήπτης Gibbs . . . . .	65
4.2	Ασυμπτωτική συμπεριφορά . . . . .	67

# Περιεχόμενα

<b>5</b>	<b>Στοιχεία Θεωρίας Αλγορίθμων</b>	<b>70</b>
5.1	Εισαγωγή . . . . .	70
5.2	Βασικές έννοιες και Κεντρικό Θεώρημα Σύγκλισης . . . . .	71
5.2.1	Θεώρημα Ολικής Σύγκλισης . . . . .	72
5.3	Αλγόριθμος EM . . . . .	75
5.3.1	Εισαγωγή . . . . .	75
5.3.2	Γενικό πλαίσιο-Ορισμοί . . . . .	75
5.3.3	Βασικές Ιδιότητες του αλγορίθμου EM . . . . .	81
5.4	Σύγκλιση του αλγορίθμου EM . . . . .	84
5.4.1	Εισαγωγή . . . . .	84
5.4.2	Σύγκλιση της ακολουθίας $L(\phi^{(k)})$ . . . . .	85
5.4.3	Σύγκλιση της $\phi^{(k)}$ . . . . .	88
5.5	Η ειδική περίπτωση της εκθετικής οικογένειας . . . . .	93
5.6	Η περίπτωση του Monte Carlo EM . . . . .	97
<b>6</b>	<b>Η μέθοδος της Γκαουσιανής τυχαιοποίησης</b>	<b>98</b>
6.1	Εισαγωγή . . . . .	98
6.2	Ορισμός και Βασικές Ιδιότητες . . . . .	98
6.3	Εφαρμογή του αλγορίθμου EM . . . . .	106
6.3.1	Εφαρμογή στο μοντέλο LNAS . . . . .	107
6.4	Ζητήματα σύγκλισης . . . . .	108
<b>7</b>	<b>Αλγοριθμικές τεχνικές για το μοντέλο GreenLab</b>	<b>114</b>
7.1	Εισαγωγή . . . . .	114
7.2	Αλγόριθμος MCMC-Bayes . . . . .	114
7.2.1	Η εκ των υστέρων κατανομή . . . . .	116
7.2.2	Σχεδιασμός του MCMC-Bayes . . . . .	117
7.2.3	Περιγραφή των μεταβάσεων . . . . .	118
7.2.4	Ανάλυση της αλυσίδας . . . . .	123
7.3	Άλλες αλγοριθμικές τεχνικές για το μοντέλο GreenLab . . . . .	124
7.4	Αριθμητικά Αποτελέσματα . . . . .	126
7.4.1	Ο MCMC-Bayes σε πραγματικά δεδομένα . . . . .	126
7.4.2	Ο MCMC-Bayes σε προσομοιωμένα δεδομένα . . . . .	133
7.4.3	Συμπεράσματα και μελλοντικοί στόχοι . . . . .	134
<b>8</b>	<b>Παράρτημα</b>	<b>135</b>
8.0.1	A.1 Πιθανοθεωρητικά . . . . .	135
8.0.2	A.2 Πλαίσιο στατιστικής . . . . .	138
8.0.3	A.3 Επιλεγμένες αποδείξεις θεωρημάτων και προτάσεων . . . . .	142

# Κεφάλαιο 1

## Εισαγωγή

Η παρούσα διπλωματική εργασία αφορά δύο μαθηματικά ζητήματα, διαφορετικής υφής, που προκύπτουν στα πλαίσια της μελέτης μοντέλων ανάπτυξης φυτών. Για τα μοντέλα αυτά έχει δοθεί τα τελευταία χρόνια ιδιαίτερη προσοχή, αφενός λόγω της ανάπτυξης των υπολογιστικών εργαλείων και αφετέρου για το μεγάλο εύρος των εφαρμογών τους σε πολλές επιστημονικές περιοχές, όπως στη βιολογία, στη γεωργία και γενικότερα στις περιβαλλοντολογικές επιστήμες.

Η μαθηματική μοντελοποίηση των μηχανισμών ανάπτυξης και λειτουργίας των φυτών, αποτελεί ένα πολύ ενδιαφέρον ζήτημα από μόνο του, καθώς είναι ένα πρόβλημα μοντελοποίησης της εξέλιξης ενός ζωντανού οργανισμού. Αν και τα φυτά γενικότερα, μπορούν να θεωρηθούν από τους πιο απλούς ζωντανούς οργανισμούς, η διαδικασία ανάπτυξής τους δεν παύει να διέπεται από πολύπλοκους μηχανισμούς. Κατά συνέπεια, η μαθηματική και στατιστική μοντελοποίησή τους, είναι ένα πολύ απαιτητικό πρόβλημα. Ωστόσο, οι τεχνικές και η μεθοδολογία που αναπτύσσονται στα πλαίσια της μοντελοποίησης αυτής, μπορούν να ρίξουν φως σε προβλήματα μοντελοποίησης της εξέλιξης σε πιο σύνθετους οργανισμούς. Ένα από τα πιο ελπιδοφόρα μοντέλα ανάπτυξης φυτών είναι το μοντέλο GreenLab. Στα πλαίσια αυτής της εργασίας θα ασχοληθούμε με ένα συγκεκριμένο τύπο του μοντέλου αυτού αλλά και με μία απλοποιημένη έκδοσή του, το μοντέλο LNAS.

Η παρούσα εργασία διακρίνεται σε δύο μέρη: Το πρώτο μέρος αφορά τη στατιστική μέθοδο της Γκαουσιανής τυχαιοποίησης (Gaussian Randomization) η οποία προτάθηκε στο [Chen, Trevezas and Cournéde, (2015)] [12] και κατασκευάστηκε με αφορμή την εκτίμηση των παραμέτρων του μοντέλου LNAS. Ωστόσο, η μέθοδος αυτή έχει γενικότερο εύρος εφαρμογών και έντονο θεωρητικό ενδιαφέρον. Ένα θεωρητικό ζήτημα που προκύπτει, αφορά μία ειδική έκδοση του αλγορίθμου EM, που κατασκευάζεται στα πλαίσια της μεθόδου αυτής, και πιο συγκεκριμένα, αφορά τη σύγκλισή του. Το επιθυμητό σημείο σύγκλισης βρίσκεται στο σύνορο του παραμετρικού χώρου, δημιουργώντας έτσι μία ιδιάζουσα περίπτωση η οποία απαιτεί αντιμετώπιση με νέες συνθήκες σύγκλισης καθώς τα αποτελέσματα που υπάρχουν στη βιβλιογραφία δεν καλύπτουν το συγκεκριμένο πλαίσιο. Στην παρούσα εργασία εστιάζουμε στο συγκεκριμένο αυτό πρόβλημα και διατυπώνουμε ορισμένες συνθήκες οι οποίες ορίζουν μία ερευνητική κατεύθυνση με απώτερο σκοπό να συμπληρωθούν πλήρως τα θεωρητικά αυτά ζητήματα.

Για το δεύτερο μέρος της παρούσας διπλωματικής εργασίας έχει σχεδιαστεί η πρώτη αμιγώς Μπεϋζιανή προσέγγιση εκτίμησης των παραμέτρων του μοντέλου GreenLab. Πιο συγκεκριμένα, θα παρουσιαστούν αναλυτικά όλα τα στάδια του σχεδιασμού της αλγοριθμικής αυτής μεθόδου η οποία φέρει το όνομα MCMC-Bayes. Επίσης, θα συζητηθούν διάφορα χαρακτηριστικά της, όπως η θεωρητική της σύγκλιση και η αποδοτικότητα της. Σχετικά με το τελευταίο ζήτημα, αντιπαραβάλλονται τα αποτελέσματα της

μεθόδου αυτής, με εκείνα που προκύπτουν από άλλες αλγοριθμικές τεχνικές, οι οποίες βασίζονται στη μέθοδο μέγιστης πιθανοφάνειας.

Στην παρούσα διπλωματική περιλαμβάνονται και τα βασικά στοιχεία της θεωρίας που συνοδεύουν τα παραπάνω ζητήματα. Η παραπάνω ανάγκη δεν προκύπτει μονάχα για λόγους πληρότητας, αλλά αποτελεί ουσιαστικό στόχο για την παρούσα εργασία. Η σχετική θεωρία αν και εξαιρετικά κομψή, είναι αρκετά προχωρημένη ώστε να διδάσκεται σε προπτυχιακό επίπεδο ή και ορισμένες φορές, σε μεταπτυχιακό επίπεδο. Πιο συγκεκριμένα, το πρώτο μέρος της διπλωματικής αφορά τη σύγκλιση του αλγορίθμου EM, ο οποίος έχει σπουδαίο ρόλο σε προβλήματα εκτιμητικής τόσο κλασικής όσο και μπεϋζιανής προσέγγισης. Η θεωρία της σύγκλισης του EM θεμελιώθηκε με τις ερευνητικές εργασίες των (Dempster Laird and Rubin (1977), Baum et al.(1970), Sundberg (1975)) και Wu (1983), με τα αποτελέσματα του τελευταίου να έχουν κεντρική θέση σήμερα. Επιπλέον, υπάρχουν και πιο πρόσφατα αποτελέσματα από τον Vaida (2005), τα οποία στηρίζονται στη θεωρία που ανέπτυξε ο Zangwill (1969). Οι εισαγωγικές έννοιες για τον αλγόριθμο EM καθώς και η πιο πρόσφατη θεωρία για τη σύγκλισή του, θα παρουσιαστούν στην παρούσα διπλωματική εργασία.

Το δεύτερο μέρος της διπλωματικής έχει ως κορμό τις μαρκοβιανές στοχαστικές διαδικασίες με τιμές σε γενικό χώρο και κυρίως την ασυμπτωτική θεωρία που τις συνοδεύει. Με τη βοήθεια εργαλείων που θα οριστούν στα πλαίσια της θεωρίας θα περάσουμε στα κρυμμένα μαρκοβιανά μοντέλα (KMM) και στις τεχνικές MCMC που θα χρειαστούν. Η βασική θεωρία πιθανοτήτων, μαρκοβιανών διαδικασιών πεπερασμένου χώρου καταστάσεων καθώς και η θεωρία που αφορά τη δεσμευμένη πιθανότητα θεωρείται γνωστή και μπορεί να βρεθεί στα [1], [17], [19].

Η παρούσα διπλωματική εργασία οργανώνεται ως εξής:

Στο δεύτερο κεφάλαιο παρουσιάζονται τα δύο μοντέλα ανάπτυξης φυτών γύρω από τα οποία πλαισιώνεται η παρούσα εργασία.

Τα κεφάλαια 3-5 αφορούν τη θεωρία στην οποία βασιζόμαστε στην εργασία αυτή. Συγκεκριμένα, τα κεφάλαια 3-4 περιέχουν τη θεωρία που απαιτείται για το δεύτερο μέρος της διπλωματικής, όπως οι μαρκοβιανές αλυσίδες γενικού χώρου, η ασυμπτωτική θεωρία, τα κρυμμένα μαρκοβιανά μοντέλα και οι τεχνικές MCMC. Το κεφάλαιο 5 αφορά τη θεωρία που σχετίζεται με το πρώτο μέρος και συγκεκριμένα με τη σύγκλιση της έκδοσης του αλγορίθμου EM που κατασκευάζεται στα πλαίσια της μεθόδου Γκαουσιανής τυχαιοποίησης. Αναλυτικότερα σε αυτό το κεφάλαιο γίνεται η εισαγωγή του αλγορίθμου EM και παρουσιάζονται η θεωρία σύγκλισης με τα βασικότερα αποτελέσματα μέχρι σήμερα.

Στο έκτο κεφάλαιο γίνεται η παρουσίαση και η μελέτη της στατιστικής μεθόδου εκτίμησης, Γκαουσιανή τυχαιοποίηση. Επίσης γίνεται και η εφαρμογή της στο μοντέλο LNAS.

Με το έβδομο κεφάλαιο ολοκληρώνεται η εργασία. Στο κεφάλαιο αυτό παρουσιάζεται αναλυτικά ο αλγόριθμος Μπεϋζιανής προσέγγισης που σχεδιάστηκε για το πρόβλημα της εκτίμησης στο μοντέλο GreenLab καθώς και τα αποτελέσματα του. Γίνεται επίσης

συζήτηση και για άλλες αλγοριθμικές τεχνικές που έχουν κατασκευαστεί για το μοντέλο αυτό ενώ αντιπαραβάλλονται τα αποτελέσματά τους με εκείνα του MCMC-Bayes.



## Κεφάλαιο 2

### Δύο μοντέλα ανάπτυξης φυτών

#### 2.1 Εισαγωγή

Στο [Cournéde et al. (2011)] [13], γίνεται μία πρώτη προσπάθεια εκτίμησης των παραμέτρων η οποία βασίζεται στην υπόθεση πως οι διαδικασίες παραγωγής της βιομάζας είναι ντετερμινιστικές και τα σφάλματα στις μετρήσεις είναι ασυσχέτιστα. Μία υπόθεση που είναι αρκετά περιοριστική. Στο [Trevezas and Cournéde (2013)] [41], οι συγγραφείς προτείνουν ένα γενικό πλαίσιο για στατιστική ανάλυση, κλασικής προσέγγισης, το οποίο μπορεί να εφαρμοστεί σε πολλές διαφορετικές κλάσεις φυτών με ντετερμινιστικό μηχανισμό οργανογένεσης (στα φυτά, είναι η διαδικασία κατά την οποία δημιουργούνται νέα όργανα). Το πλαίσιο αυτό που προτείνουν μπορεί να θεωρηθεί και ως βάση για στατιστική ανάλυση με στοχαστικό μηχανισμό οργανογένεσης. Βασική ιδέα αποτέλεσε η θεώρηση πως η εξέλιξη του συστήματος μπορεί να περιγραφεί από ένα μη-ομογενές κρυμμένο μαρκοβιανό μοντέλο. Ωστόσο, το συγκεκριμένο μοντέλο είναι σύνθετο με αποτέλεσμα οι κλασικές μέθοδοι εκτίμησης να μην εφαρμόζονται άμεσα. Έτσι, διάφορες αλγοριθμικές τεχνικές είναι δυνατόν να επινοηθούν προκειμένου να γίνει εφικτή η εκτίμηση των παραμέτρων. Γενικότερα, τα μοντέλα ανάπτυξης φυτών χαρακτηρίζονται από μεγάλο πλήθος παραμέτρων, μη-γραμμική δυναμική και με πολύ χαμηλό πλήθος δεδομένων καθώς οι διαδικασίες απόκτησης πειραματικών δεδομένων είναι αρκετά δαπανηρές. Τα παραπάνω συνθέτουν κάποιους από τους λόγους για τους οποίους η εκτίμηση των παραμέτρων είναι ένα απαιτητικό πρόβλημα και προκειμένου να γίνει εφικτή, χρησιμοποιούνται εκτενώς μεγάλο πλήθος επαναληπτικών διαδικασιών, στοχαστικών αλλά και ντετερμινιστικών.

#### 2.2 Το μοντέλο GreenLab

Το μοντέλο GreenLab είναι ένα μαθηματικό μοντέλο που προσομοιώνει τις αλληλεπιδράσεις μεταξύ της δομής και των λειτουργιών των φυτών συνδυάζοντας κανόνες για την παραγωγή και διανομή της βιομάζας (μάζα ζωντανών οργανισμών) με τη διαδικασία της ανάπτυξης των οργάνων του. Προτάθηκε από τους de Reffye και Hu (2003) [15] και σχεδιάστηκε ως ένα μη γραμμικό, (διακριτό) δυναμικό στοχαστικό μοντέλο στο [Cournéde et al (2006)] [14], το οποίο παρέχει μία δυναμική αναπαράσταση της εξέλιξης του φυτού. Η διακριτοποίηση αντιστοιχεί σε μία διαμέριση της ανάπτυξης, σε κύκλους (Growth Cycles, GC's) των οποίων η διάρκεια καθορίζεται με βάση τους θερμικούς χρόνους που χρειάζεται ώστε να δημιουργηθεί ένα καινούριο όργανο [Barthelemy and Caraglio (2007)] [6].

Το μοντέλο Greenlab μπορεί να ενταχθεί στο γενικότερο πλαίσιο των δυναμικών μοντέλων παραγωγής και διανομής, όπου τόσο η παραγωγή όσο και η ζήτηση εξαρτώνται κάθε φορά από την κατάσταση του συστήματος (κύκλος GC). Η βασική διαδικασία

είναι η  $(q_n)$  η οποία αντιπροσωπεύει τη βιομάζα που έχει παραχθεί στον κύκλο  $n$  και είναι διαθέσιμη προς διανομή στα διάφορα όργανα του φυτού. Κάθε όργανο έχει διαφορετική ζήτηση η οποία εξαρτάται από το στάδιο ανάπτυξης στο οποίο βρίσκεται. Το σύνολο διαφορετικών τύπων οργάνων των φυτών (έλασμα (blade), μίσχος (petiole), ρίζα (root), κόμβος (node), μεσογονάτιο διάστημα (internode), κλπ) συμβολίζεται με  $\mathcal{O}$  και εξαρτάται από το συγκεκριμένο υπό μελέτη φυτό. Κάθε κύκλος ( $GC$ ) καθορίζεται πλήρως από τους ακόλουθους μηχανισμούς-διαδικασίες: Οργανογένεση και λειτουργία. Ο πρώτος (οργανογένεση) αφορά τη δημιουργία νέων οργάνων για το φυτό και στην παρούσα εργασία υποθέτουμε πως αυτός ο μηχανισμός είναι γνωστός και ντετερμινιστικός. Έτσι, το ενδιαφέρον συγκεντρώνεται στη λειτουργία που αντιπροσωπεύεται μέσω ενός αναδρομικού μηχανισμού διανομής και παραγωγής της βιομάζας. Επομένως για όσα ακολουθούν θεωρούμε κλάσεις φυτών με γνωστό μηχανισμό οργανογένεσης. Για τους σκοπούς αυτής της διπλωματικής, το μοντέλο GreenLab εξειδικεύεται στο φυτό ζαχαρότευτλο το οποίο έχει μελετηθεί αρκετά λόγω της σημασίας του για την παραγωγή της ζάχαρης. Επίσης, για το συγκεκριμένο αυτό είδος, υπάρχουν διαθέσιμα πραγματικά δεδομένα τα οποία θα αξιοποιηθούν στις επόμενες ενότητες για στατιστική ανάλυση. Συχνά, η συλλογή του φυτού αυτού γίνεται όταν ακόμα βρίσκεται σε αρχικό στάδιο ανάπτυξης. Επομένως, η δομή του είναι σχετικά απλή καθώς στα στάδια αυτά, τα μόνα διαθέσιμα όργανα είναι τα φύλλα (τα οποία αποσυντίθενται στα ελάσματα και στον μίσχο) και η ρίζα. Προκύπτει λοιπόν ότι  $\mathcal{O} = \{b, p, r\}$ . Οι υπόλοιπες υποθέσεις που κάνουμε για το φυτό αυτό παρουσιάζονται στη συνέχεια.

### 2.2.1 Διανομή της Βιομάζας

Η διαδικασία διανομής της βιομάζας έχει ως εξής: Ξεκινάμε από μία αρχική τιμή  $q_0$  (αρχική βιομάζα σπόρου) και στον  $n$ -οστό κύκλο, θα είναι διαθέσιμη ποσότητα βιομάζας  $q_n$  (ανά μονάδα επιφάνειας:  $g \cdot m^{-2}$ ), για διανομή στα όργανα του φυτού που βρίσκονται σε διαδικασία ανάπτυξης. Προκειμένου να καθορίσουμε τον τρόπο με τον οποίο γίνεται η διανομή, κάνουμε τις ακόλουθες υποθέσεις:

Το  $t_o$  εκφράζει το χρόνο (σε κύκλους ανάπτυξης) που απαιτείται μέχρι το συγκεκριμένο όργανο  $o \in \{b, p, r\}$  να φτάσει το μέγιστο δυνατό μέγεθος από τη στιγμή της δημιουργίας του. Οι χρόνοι αυτοί αναφέρονται ως χρόνοι ανάπτυξης των οργάνων στην κλάση  $\mathcal{O}$ . Ακόμη, με  $p_{al}$  (ο δείκτης  $al$  από το allocation) συμβολίζουμε το (ευκλείδειο) διάνυσμα που περιέχει όλες τις παραμέτρους που υπεισέρχονται στο μηχανισμό διανομής της βιομάζας και γράφουμε  $p_{al} = (p_{al}^o)_{o \in \mathcal{O}}$ . Το  $p_{al}^o$ , εκφράζει το διάνυσμα των παραμέτρων που επηρεάζουν τη διανομή της βιομάζας στα όργανα τύπου  $o \in \mathcal{O}$ .

#### Υπόθεση 2.1

- (i) Στον κύκλο  $GC(n)$ , η βιομάζα  $q_n$  που έχει παραχθεί, είναι διαθέσιμη για διανομή προς όλα τα όργανα (αυτά που προϋπάρχουν καθώς και εκείνα που δημιουργήθηκαν κατά τη διάρκεια του προηγούμενου κύκλου) και διανέμεται ανάλογα με τις συναρτήσεις απορρόφησης (sink functions) που εξαρτώνται από την κλάση  $o$ :

$$s_o(i; p_{al}^o) = p_o c \left( \frac{i + 0.5}{t_o} \right)^{a_o - 1} \left( 1 - \frac{i + 0.5}{t_o} \right)^{b_o - 1}, \quad i \in \{0, 1, \dots, t_o - 1\}, \quad (2.2.1)$$

όπου,  $p_{al}^o = (p_o, a_o, b_o) \in \mathbb{R}_+^* \times [0, +\infty)^2$  και το  $(p_o)_{o \in \mathcal{O}}$  είναι ένα διάνυσμα που αντιπροσωπεύει το βαθμό απορρόφησης κάθε κλάσης. Το  $c = c(a_o, b_o)$  είναι η σταθερά κανονικοποίησης, δηλαδή  $c^{-1} = \sum_{o \in \mathcal{O}} \sum_{i=0}^{t_o-1} \left(\frac{i+0.5}{t_o}\right)^{a_o-1} \left(1 - \frac{i+0.5}{t_o}\right)^{b_o-1}$ . Η ποσότητα αυτή μέσα στο άθροισμα, αντιστοιχεί στη συνάρτησης μάζας πιθανότητας της διακριτής  $Beta(a_o, b_o)$ .

- (ii) Για όλα τα  $o \in \mathcal{O} \setminus \{r\}$ , υποθέτουμε πως  $t_o = T$ . (Δηλαδή για το συγκεκριμένο φυτό υποθέτουμε πως  $t_b = t_p = T$ ). Μπορούμε να χαλαρώσουμε την τελευταία υπόθεση, καθώς χρησιμοποιείται για ευκολία στο συμβολισμό. Επίσης, προς αποφυγή προβλημάτων αναγνωρισιμότητας (identifiability) (περιπτώσεις με χαμηλό πλήθος διαθέσιμων δεδομένων), ορισμένες από τις προαναφερθείσες παραμέτρους μπορούν να θεωρηθούν σταθερές, μειώνοντας έτσι τη διάσταση του παραμετρικού χώρου.

**Ορισμός 2.1** Η ολική ζήτηση βιομάζας στο κύκλο  $GC(n)$ , συμβολίζεται με  $d_n$ , και εκφράζει το άθροισμα των τιμών απορρόφησης (sink values) για όλα τα μέρη του φυτού στο κύκλο  $GC(n)$ . Έστω  $\{(N_n^o)_{o \in \mathcal{O} \setminus \{r\}}\}_{n \in \mathbb{N}_0}$ , ακολουθία (διανυσμάτων) που αντιπροσωπεύει το πλήθος των οργάνων που σχηματίστηκαν κατά τη διάρκεια του κάθε κύκλου (υποθέτουμε ότι είναι όλα γνωστά). Από την υπόθεση 5.1 (2.2.1), βλέπουμε ότι ένα οργανικό μέρος βρίσκεται στο  $i$ -οστό στάδιο ανάπτυξης αν έχει σχηματιστεί στο κύκλο  $GC(n - i)$ . Από το γεγονός αυτό και τη μοναδικότητα της ρίζας έχουμε:

$$d_n(p_{al}) = \sum_{o \in \mathcal{O} \setminus \{r\}} \sum_{i=0}^{n \wedge (T-1)} N_{n-i}^o s_o(i; p_{al}^o) + s_r(n; p_{al}^r). \quad (2.2.2)$$

Τέλος, σημειώνεται πως στην περίπτωση του ζαχαρότευτλου είναι  $N_n^o = 1$  για κάθε  $n \geq 0$  και για κάθε  $o \in \mathcal{O} \setminus \{r\}$ .

### 2.2.2 Παραγωγή της Βιομάζας

Εκτός από την αρχική τιμή του σπόρου  $q_0$ , οι υπόλοιπες βιομάζες  $(q_n)_{n \geq 1}$  παράγονται μέσω της φωτοσύνθεσης στην οποία το μοναδικό μέρος του φυτού που συμμετέχει είναι τα ελάσματα (πράσινο μέρος των φύλλων, blade). Έτσι, σε έναν κύκλο  $GC(n)$ , η παραγόμενη βιομάζα θα εξαρτάται άμεσα από την ολική επιφάνεια των ελασμάτων που είναι φωτοσυνθετικά ενεργή στον κύκλο  $n$ . Πιο συγκεκριμένα έχουμε τον ακόλουθο ορισμό:

**Ορισμός 2.2** (i) Η φωτοσυνθετικά ενεργή επιφάνεια στον κύκλο  $GC(n+1)$ , συμβολίζεται με  $s_n^{act}$ , και εκφράζει την ολική επιφάνεια των ελασμάτων που έχουν σχηματιστεί μέχρι και τον κύκλο  $GC(n)$  και θα είναι φωτοσυνθετικά ενεργά στον  $GC(n+1)$ . (ii) Ο λόγος (ποσοστό) συνεισφοράς της βιομάζας  $(q_l)$  στην  $s_n^{act}$  θα συμβολίζεται με  $\pi_{l,n}^{act}$  το οποίο εκφράζει το ποσοστό της παραγόμενης βιομάζας στον  $GC(l)$  κύκλο που είναι φωτοσυνθετικά στον  $GC(n)$ .

Επίσης, υποθέτουμε τα ακόλουθα:

### Υπόθεση 2.2

## Εκτίμηση των καταστάσεων και των παραμέτρων.

- (i) Η αρχική βιομάζα του σπόρου  $q_0$ , θεωρείται γνωστή.
- (ii) Τα ελάσματα έχουν κοινή περίοδο στην οποία είναι φωτοσυνθετικά ενεργά, ίση με  $T$  (θα μπορούσε να είναι διαφορετική από  $T$ ).
- (iii) Τα ελάσματα έχουν κοινή μάζα ανά μονάδα επιφάνειας (επιφανειακή μάζα ή ειδικό βάρος του ελάσματος) την οποία συμβολίζουμε με  $e_b$ .

Από την παραπάνω υπόθεση (ii) καθώς και από τους ορισμούς 1 και 2(i), έχουμε πως το έλασμα στο στάδιο ανάπτυξης  $j$ , στον κύκλο  $GC(l)$  συνεισφέρει στην  $s_n^{act}$ , αν  $l + T - j \geq n + 1$ . Έτσι, σε συνδυασμό με τον ορισμό 2(ii), παίρνουμε την ακόλουθη έκφραση:

$$\pi_{l,n}^{act} = \frac{1}{d_l(pal)} \sum_{j=0}^{l \wedge (l+T-n-1)} N_{l-j}^b s_b(j; p_{al}), \quad (n - T + 1)^+ \leq l \leq n. \quad (2.2.3)$$

Από τα (ii)-(iii) στην υπόθεση 2 και από τον ορισμό 2, μπορούμε να έχουμε μία έκφραση για το  $s_n^{act}$  διαιρώντας τη φωτοσυνθετικά ενεργή μάζα των ελασμάτων με την επιφανειακή μάζα:

$$s_n^{act}(q_{(n-T+1)^+:n}; pal) = e_b^{-1} \sum_{l=(n-T+1)^+}^n \pi_{l,n}^{act} q_l \quad (2.2.4)$$

Τώρα έχουμε όλα όσα χρειαζόμαστε για να ορίσουμε τη σχέση που θα μας δώσει την  $(q_n)_{n \geq 1}$ .

### Υπόθεση 2.3

Υποθέτοντας πως δεν υπάρχουν σφάλματα από τη μοντελοποίηση, η ακολουθία των βιομαζών  $(q_n)_{n \geq 1}$  καθορίζεται από το ακόλουθο αναδρομικό σχήμα, γνωστό ως εμπειρικός νόμος των Beer- Lambert [Guo et al. 2006] [21]:

$$q_{n+1} = F_n(q_{(n-T+1)^+:n}, u_n; p) = u_n \mu s^{pr} \left( 1 - \exp \left\{ -k_B \frac{s_n^{act}(q_{(n-T+1)^+:n}; pal)}{s^{pr}} \right\} \right), \quad (2.2.5)$$

όπου  $p = (\mu, s^{pr}, k_B, pal)$  διάνυσμα του οποίου η διάσταση, θα εξαρτάται από τις τρέχουσες ερευνητικές ανάγκες, καθώς ορισμένες από τις παραμέτρους που περιλαμβάνει μπορούν να θεωρηθούν σταθερές. Το  $s^{pr}$  είναι η χαρακτηριστική επιφάνεια που αντιπροσωπεύει τη διδιάστατη προβολή στο έδαφος, δηλαδή του χώρου που τελικά θα καταλαμβάνει το φυτό. Το  $k_B$  είναι ο συντελεστής στο νόμο εξαφάνισης του Beer-Lambert (B-L extinction law). Το διάνυσμα  $(u_n)_{n \geq 1}$  είναι σταθερές που εξαρτώνται από το περιβάλλον και καθορίζονται μεταγενέστερα (έπειτα από την απόκτηση των δεδομένων από μετεωρολογικά δεδομένα). Τέλος, το  $\mu$  συνδέεται με την αποδοτικότητα της χρήσης ακτινοβολίας.

## 2.3 Εκτίμηση των καταστάσεων και των παραμέτρων.

Στο Cournéde et al. (2011) έγινε μία πρώτη προσπάθεια εκτίμησης των παραμέτρων η οποία βασίζεται στην αρκετά περιοριστική υπόθεση πως η διαδικασία παραγωγής της

βιομάζας είναι ντετερμινιστική όπως επίσης και ότι τα σφάλματα των μετρήσεων για τα διαφορετικού τύπου όργανα, είναι ασυσχέτιστα. Στην εργασία αυτή θα παρουσιαστεί ένα ενιαίο πλαίσιο για στατιστική ανάλυση σε διάφορα είδη φυτών λαμβάνοντας υπόψη τις διαδικασίες και τα σφάλματα των μετρήσεων.

Προκειμένου να μοντελοποιηθεί η εξέλιξη, οι συγγραφείς στο Trevezas and Cournéde (2013), προτείνουν μία πιο ευέλικτη έκδοση του μοντέλου Greenlab με την οποία είναι δυνατή η αναπαράσταση ως ένα μη-ομογενές Κρυμμένο Μαρκοβιανό μοντέλο (HMM) (Κεφάλαιο 3) στο οποίο η κρυμμένη ακολουθία καταστάσεων αντιπροσωπεύει τις διαθέσιμες (άγνωστες) βιομάζες που παράγονται και γίνονται διαθέσιμες προς διανομή σε διαδοχικούς κύκλους ανάπτυξης. Συγκεκριμένα:

Τα δεδομένα αποτελούνται από το διάνυσμα των παρατηρήσεων  $\mathbf{Y}_N$  το οποίο καταγράφει τις μάζες όλων των διαθέσιμων οργάνων στον κύκλο ανάπτυξης  $GC(N)$  όπου επιλέγουμε να σταματήσουμε την εξέλιξη κόβοντας το φυτό. Τα υποδιανύσματα  $(Y_n)_{o \in O}$ , περιέχουν τις μάζες των οργάνων που σχηματίστηκαν στον κύκλο  $GC(n)$  και εμφανίστηκαν πρώτη φορά στον κύκλο  $GC(n+1)$ .

Έστω  $G_n = (G_{n,o})_{o \in O}$  η συνάρτηση που εκφράζει τις θεωρητικές τελικές μάζες όλων των οργάνων που ξεκίνησαν την ανάπτυξη τους στον κύκλο  $GC(n)$ . Σε κάθε κύκλο ανάπτυξης από το  $n$  και ύστερα, το ποσοστό της βιομάζας που διατίθεται σε ένα τέτοιο όργανο είναι:  $\frac{q_{j+n}}{d_{j+n}(p_{al})} s_o(j; p_{al}^o)$ . Τότε, αθροίζοντας τις βιομάζες που συνεισέφεραν στην ανάπτυξη τους αλλά και από την υπόθεση 5.1, παίρνουμε ότι:

$$G_n(q_{n:(n+T-1)}; p_{al}) = \left( \sum_{j=0}^{T-1} \frac{q_{j+n}}{d_{j+n}(p_{al})} s_o(j; p_{al}^o) \right)_{o \in O \setminus \{r\}} \quad (2.3.1)$$

$$G_r(q_{0:N}; p_{al}) = \sum_{j=0}^N \frac{q_j}{d_j(p_{al})} s_r(j; p_{al}^r) \quad (2.3.2)$$

Η στοχαστική φύση του μοντέλου καθορίζεται από τις ακόλουθες υποθέσεις:

#### Υπόθεση 2.4

Έστω  $(W_n)_{n \in \mathbb{N}}$  και  $(V_n)_{n \in \mathbb{N}}$ , δύο ανεξάρτητες ακολουθίες από ανεξάρτητες και ισόνομες, τυχαίες μεταβλητές και διανύσματα αντίστοιχα, και ανεξάρτητες από το  $Q_0$ , με  $W_n \sim \mathcal{N}(0, \sigma^2)$  και  $V_n \sim \mathcal{N}_d(0, \Sigma)$ . Όπου  $d$  η πληθικότητα του  $O \setminus \{r\}$  και  $\mathcal{N}_d$  η  $d$ -διάστατη κανονική κατανομή.

Υποθέτοντας πως  $N_n^o = 1$  για κάθε  $o \in O \setminus \{r\}$  (μοντέλο για ζαχαρότευτλα), τότε θεωρούμε ότι:

- (i) ένα μοντέλο με πολλαπλασιαστικά σφάλματα καθορίζει τις κρυμμένες μεταβλητές:

$$Q_{n+1} = F_n(Q_{(n-T+1)+:n}; p)(1 + W_n), \quad (2.3.3)$$

- (ii) ένα μοντέλο με αθροιστικά σφάλματα καθορίζει τα παρατηρούμενα δεδομένα:

$$Y_n = G_n(Q_{n:(n+T-1)}; p_{al}) + V_n, \quad (2.3.4)$$

όπου  $F_n, G_n$ , οι συναρτήσεις στις (2.2.5) και (2.3.1)

### Παρατηρήσεις.

- (i) Υποθέτουμε ότι τα σφάλματα στη διαδικασία καταστάσεων  $(Q_n)$  επιδρούν πολλαπλασιαστικά και όχι αθροιστικά καθώς στα πλαίσια των εφαρμογών, οι βιομάζες αλλάζουν τάξη μεγέθους, και
- (ii) οι καταστάσεις  $Q_n$  και οι παρατηρήσεις  $Y_n$ , αντιπροσωπεύουν μάζες και επομένως παίρνουν θετικές τιμές. Οι εξισώσεις καταστάσεων και παρατηρήσεων (2.3.3) και (2.3.4) αντίστοιχα, παραβιάζουν αυτή τη συνθήκη. Όμως η διασπορά  $\sigma^2$  που εμφανίζεται στη  $W_n$  καθώς και ο πίνακας συνδιακυμάνσεων  $\Sigma$  στο διάνυσμα  $V_n$  μπορούν να περιοριστούν με τέτοιο τρόπο ώστε η πιθανότητα να πάρουν αρνητικές τιμές να είναι αρκετά μικρή.

Η παραπάνω συναρτησιακή έκφραση είναι της μορφής ενός μοντέλου χώρου καταστάσεων (state-space), στο οποίο, η ακολουθία καταστάσεων είναι η  $\mathbf{Q} = (Q_n)_{n \geq 0}$  και η ακολουθία παρατηρήσεων δίνεται στην υπόθεση 5.4. Ακολουθεί η αναπαράσταση του παραπάνω, ως κρυμμένο μαρκοβιανό μοντέλο (KMM)(Hidden Markov Model (HMM)):

**Πρόταση 2.1** Κάτω από τις υποθέσεις (2.1-2.4), η διδιάστατη στοχαστική διαδικασία  $(\mathbf{Q}, \mathbf{Y})$  ορισμένη στον  $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ , όπου  $\theta = (p, \sigma^2, \Sigma)$ , μπορεί να αναπαρασταθεί ως KMM, ως εξής:

- (i) Η κρυμμένη διαδικασία  $\mathbf{Q}$ , με τιμές στον  $\mathbb{R}_+^*$ , εξελίσσεται ως μία χρονικά μη-ομογενής  $T$ -οστής τάξης μαρκοβιανή αλυσίδα με αρχική κατανομή  $\mathbb{P}_\theta(Q_0 \in \cdot) = \delta_{q_0}(\cdot)$  (μέτρο Dirac),  $q_0 \in \mathbb{R}_+^*$ , και οι κατανομές μετάβασης καθορίζονται από την:

$$\mathbb{P}_\theta(Q_{n+1} \in \cdot \mid Q_{(n-T+1)+:n}) \stackrel{D}{\approx} \mathcal{N}(F_n(Q_{(n-T+1)+:n}; p), \sigma^2 F_n^2(Q_{(n-T+1)+:n}; p)). \quad (2.3.5)$$

- (ii) Η παρατηρούμενη διαδικασία  $\mathbf{Y}(Y_n)_{n \geq 0}$ , με τιμές στον  $(\mathbb{R}_+)^d$ , δοθείσης της  $\mathbf{Q}$ , σχηματίζει μία ακολουθία δεσμευμένα ανεξάρτητων τυχαίων διανυσμάτων όπου κάθε  $Y_n$  δοθείσης της  $\mathbf{Q}$  εξαρτάται μόνο από το διάνυσμα  $Q_{n:(n+T-1)}$  με δεσμευμένη κατανομή:

$$\mathbb{P}_\theta(Y_n \in \cdot \mid Q_{n:(n+T-1)}) \stackrel{D}{\approx} \mathcal{N}_d(G_n(Q_{n:(n+T-1)}; pa), \Sigma), n \geq 0. \quad (2.3.6)$$

Όπως προαναφέρθηκε στην παρατήρηση (ii), οι παραπάνω ποσότητες είναι θετικές και έτσι η κανονικότητα είναι έγκυρη μόνο κατά προσέγγιση. Επίσης, παρατηρούμε ότι αν  $\tilde{\mathbf{Q}} \stackrel{def}{=} (Q_{n:n+T-1})_{n \geq 0}$ , τότε η διδιάστατη στοχαστική διαδικασία  $(\tilde{\mathbf{Q}}, \mathbf{Y})$  είναι μία μη ομογενής 1ης τάξης κρυμμένη Μαρκοβιανή αλυσίδα. Το γεγονός αυτό ωστόσο δεν έχει κάποιο πρακτικό όφελος και έτσι δεν θα χρησιμοποιηθεί αυτή η αναπαράσταση στη συνέχεια.

## Εκτίμηση των καταστάσεων και των παραμέτρων.

Είναι φανερό ότι τα όργανα που ξεκίνησαν την ανάπτυξη τους στους  $T - 1$  τελευταίους κύκλους ανάπτυξης, δεν θα έχουν προλάβει να ολοκληρώσουν την ανάπτυξη τους. Αν αυτά τα όργανα προστεθούν στο παρατηρούμενο δείγμα τότε:

$$G_n(Q_{n:(n+T-1)\wedge N}; p_{al}) = \left( \sum_{j=0}^{T-1\wedge N-n} \frac{q_{j+n}}{d_{j+n}(p_{al})} s_o(j; p_{al}^o) \right)_{o \in \mathcal{O} \setminus \{r\}}. \quad (2.3.7)$$

και έτσι η εξίσωση στη (2.3.6) θα αντικατασταθεί από την:

$$\mathbb{P}_\theta(Y_n \in \cdot \mid Q_{n:(n+T-1)\wedge N}) \stackrel{D}{\approx} N_d(G_n(Q_{n:(n+T-1)\wedge N}; p_{al}), \Sigma), n \geq 0. \quad (2.3.8)$$

Για το μοντέλο αυτό έχουν προταθεί μέχρι στιγμής στη βιβλιογραφία πολλές μέθοδοι για τον υπολογισμό της εκτιμήτριας μέγιστης πιθανοφάνειας (Ε.Μ.Π.). Ωστόσο, θα παρουσιαστεί μία νέα μέθοδος Μπεϋζιανής προσέγγισης, στην οποία, για την προσομοίωση από των εκ των υστέρων κατανομών που προκύπτουν, χρησιμοποιούνται τεχνικές MCMC. Η κατασκευή της καθώς και η παρουσίαση των σχετικών αποτελεσμάτων, γίνεται στο κεφάλαιο 7. Προχωράμε στην περιγραφή του μοντέλου LNAS.

## 2.4 Το μοντέλο LNAS

Το μοντέλο LNAS (Log Normal Allocation and Senescence) αποτελεί μία απλοποιημένη έκδοση του μοντέλου GreenLab που συζητήθηκε στην προηγούμενη ενότητα. Το μοντέλο αυτό δύναται να προσαρμόζεται καλύτερα σε περιπτώσεις όπου μόλις ένα τμήμα των δεδομένων είναι διαθέσιμα, ενώ η απλοποίηση αφορά τόσο στο πλήθος των διαθέσιμων οργάνων (συγκεκριμένα εδώ θεωρούμε 2 τύπους οργάνων, τα φύλλα και τη ρίζα) όσο και στη διάσταση του παραμετρικού χώρου. Όπως και στην περίπτωση του GreenLab, στόχος είναι η περιγραφή της ανάπτυξης του ζαχαρότευτλου ως προς την εξέλιξη της βιομάζας του ανά κύκλο ανάπτυξης ( $GC$ ), όπου εδώ, αντίθετα με την περίπτωση του μοντέλου GreenLab, ο κύκλος διαρκεί 1 ημερολογιακή ημέρα, χωρίς να υπεισέρχεται η έννοια του θερμικού χρόνου. Πάραυτα, καθορίζοντας τα οργανικά τμήματα, και με κατάλληλες τροποποιήσεις, το μοντέλο αυτό μπορεί να προσαρμοστεί και σε άλλα είδη φυτών. Οι βασικοί μηχανισμοί που περιγράφουν την ανάπτυξη είναι η διαδικασία παραγωγής της βιομάζας, η διανομή της στα οργανικά τμήματα (φύλλα και ρίζα) και η διαδικασία της γήρανσης.

### 2.4.1 Παραγωγή Βιομάζας

Όπως και στα προηγούμενα, θεωρούμε το διάνυσμα  $(q_n)_{n \geq 1}$  όπου  $q_n$ , η παραγόμενη βιομάζα στο  $n$ -οστό κύκλο ανάπτυξης ( $GC(n)$ ). Τροποποιώντας το συμβολισμό που είχαμε στο μοντέλο GreenLab, έστω  $q_n^f$  η ολική μάζα των φύλλων,  $q_n^g$ , η ολική μάζα των φωτοσυνθετικά ενεργών φύλλων και  $q_n^s$ , η ολική μάζα των μη φωτοσυνθετικά ενεργών φύλλων (φύλλα που είτε έχουν πέσει από το φυτό, είτε παραμένουν σε αυτό αλλά δεν συμμετέχουν στη διαδικασία της φωτοσύνθεσης) στο  $n$ -οστό κύκλο (προφανώς:  $q_n^g = q_n^f - q_n^s$ ). Τέλος, το  $q_n^r$  θα είναι η ολική μάζα της ρίζας στο  $n$ -οστό κύκλο. Με  $Q_n, Q_n^f, Q_n^g, Q_n^s, Q_n^r$  συμβολίζουμε τις αντίστοιχες στοχαστικές διαδικασίες.

Σύμφωνα με τον νόμο Beer-Lambert (σχέση 2.2.5, και χωρίς την παρουσία σφαλμάτων, η παραγόμενη βιομάζα δίνεται από τη σχέση:

$$q_n = F_n(q_n; \mu\lambda) = \mu(1 - \exp\{-\lambda q_n^g\}), \quad n \geq 1, \quad (2.4.1)$$

όπου  $\lambda = k_B/s^{spr}$ , και στο  $\mu$  έχουν ενσωματωθεί οι σταθερές  $u_n, s^{spr}$ .

Στο μοντέλο αυτό υποθέτουμε επίσης ότι οι παραγόμενες βιομάζες διέπονται από τις εξής εξισώσεις κατάστασης:

$$Q_{n+1} = F_n(Q_n^g; \mu, \lambda)e^{\xi_n^g}, \quad n \geq 1, \quad (2.4.2)$$

όπου  $\xi_n^g \sim \mathcal{N}(0, \sigma_q^2)$ .

### 2.4.2 Διανομή της Βιομάζας

Η διαθέσιμη βιομάζα,  $Q_n$ , στο  $n$ -οστό κύκλο, θα διανεμηθεί στα φύλλα και στη ρίζα, σε αντίθεση με το μοντέλο GreenLab που η διανομή υπολογίζεται σε οργανικό επίπεδο, στο LNAS ο υπολογισμός αυτός απλοποιείται:

$$Q_{n+1}^f = Q_n^f + \gamma_n Q_n \quad (2.4.3)$$

$$Q_{n+1}^r = Q_n^r + (1 - \gamma_n)Q_n, \quad (2.4.4)$$



με αρχική συνθήκη:  $Q_0^f = Q_0^r = 0$

Η συνάρτηση  $\gamma$  ορίζεται ως:

$$\gamma_n(\tau_n; \gamma_0, \gamma_f, \mu_\gamma, \sigma_\gamma^2) = (\gamma_0 + (\gamma_f - \gamma_0)H(\tau_n))(1 + \eta_n^\gamma), \quad (2.4.5)$$

όπου  $\tau_n > 0$  ο θερμικός χρόνος και  $H(x; \mu, \sigma^2)$  η σ.κ. της λογαριθμοκανονικής (log-normal) κατανομής, με παράμετρο θέσης  $\mu$  και παράμετρο κλίμακας  $\sigma^2$  και  $\eta_n^\gamma \sim N(0, \sigma_\gamma^2)$ , τα σφάλματα του μοντέλου. Η συγκεκριμένη συνάρτηση είναι κατάλληλη για την περιγραφή γενικών βιολογικών διαδικασιών και για το λόγο αυτό έχει επιλεγεί. Επίσης, είναι  $\gamma_0, \gamma_f \in (0, 1)$  καθώς οι ποσότητες αυτές αντιπροσωπεύουν το αρχικό και τελικό ποσοστό, αντίστοιχα, της βιομάζας που διανέμεται στο τμήμα των φύλλων.

Έστω  $\Gamma_n$  το ποσοστό της βιομάζας που διανέμεται στο τμήμα των φύλλων την ημέρα  $n$  το οποίο συνδέεται με θερμικό χρόνο  $\tau_n$ . Οι συγγραφείς στο (Chen, Trevezas, Courn'ede (2015) υποθέτουν πως το  $\Gamma_n$  είναι μοντέλο logit-normal με μέσο τον logit μετασχηματισμό του  $\gamma_n(\tau_n; \gamma_0, \gamma_f, \mu_\gamma, \sigma_\gamma^2)$  και διασπορά  $\sigma_{\gamma\gamma}^2$ . Δηλαδή:

$$\log\left(\frac{\Gamma_n}{1 - \Gamma_n}\right) = \log\left(\frac{\gamma_n(\tau_n; \mu_\gamma, \sigma_\gamma^2)}{1 - \gamma_n(\tau_n; \mu_\gamma, \sigma_\gamma^2)}\right) + \xi_n^\gamma, \quad (2.4.6)$$

όπου  $\xi_n^\gamma \sim \mathcal{N}(0, \sigma_{\gamma\gamma}^2)$ .

Με την υπόθεση αυτή, η αναδρομική σχέση (2.4.3) για τον υπολογισμό της ολικής μάζας των φύλλων καθώς και της ολικής μάζας της ρίζας, απλοποιείται στην:

$$Q_{n+1}^f = Q_n^f + \Gamma_n Q_n \quad (2.4.7)$$

$$Q_{n+1}^r = Q_n^r + (1 - \Gamma_n)Q_n, \quad (2.4.8)$$

με την ίδια αρχική συνθήκη:  $Q_0^f = Q_0^r = 0$

### 2.4.3 Διαδικασία Γήρανσης

Τέλος, η διαδικασία της γήρανσης μοντελοποιείται χρησιμοποιώντας την  $Q_n^s$  που όπως προαναφέρθηκε, αντιπροσωπεύει το μέρος της ολικής μάζας των φύλλων που δεν είναι πλέον φωτοσυνθετικά ενεργό στο  $n$ -οστό κύκλο. Υποθέτουμε ότι:

$$Q_n^s = H(\tau_n; \mu_s, \sigma_s^2)Q_n^f \quad (2.4.9)$$

Από την παραπάνω και την  $Q_n^g = Q_n^f - Q_n^s$  παίρνουμε ότι:

$$Q_n^g = H^c(\tau_n; \mu_s, \sigma_s^2)Q_n^f \quad (2.4.10)$$

όπου  $H^c(x) = 1 - H(x)$  η συνάρτηση επιβίωσης.

#### 2.4.4 Αναπαράσταση ως μοντέλο χώρου καταστάσεων

Για την εκτίμηση των παραμέτρων του μοντέλου, στο LNAS, οι συγγραφείς στο (Chen, Trevezas, Cournéde, (2015)) [12] προτείνουν μία νέα μέθοδο με το όνομα Γκαουσιανή Τυχαιοποίηση (Gaussian Randomization), κατά την οποία, το πρόβλημα της εκτίμησης μεταφέρεται σε ένα πλαίσιο στο οποίο μπορεί να εφαρμοστεί ο αλγόριθμος EM. Η αναλυτική περιγραφή της μεθόδου γίνεται στο κεφάλαιο 6. Οι συγγραφείς, προκειμένου να εφαρμόσουν τη μέθοδο, θεωρούν μία αναπαράσταση του μοντέλου ως μοντέλο χώρου καταστάσεων (μ.χ.κ.) (state-space model), η οποία περιγράφεται στη συνέχεια.

Τα δυνητικά διαθέσιμα δεδομένα αποτελούνται από μετρήσεις των βιομαζών:  $(Q_n^g, Q_n^r)_{n=1}^N$ , υπό την παρουσία σφαλμάτων (όπου  $N$  ο χρονικός ορίζοντας). Έστω  $Z_n = (Z_{n,1}, Z_{n,2})$  το διάνυσμα που αντιστοιχεί στις μετρήσεις αυτές με την παρουσία σφαλμάτων. Συγκεκριμένα, υποθέτουμε  $\epsilon_n^g \sim \mathcal{N}(0, \sigma_g^2)$ ,  $\epsilon_n^r \sim \mathcal{N}(0, \sigma_r^2)$ ,  $n \geq 1$ , δύο αμοιβαία ανεξάρτητες, ανεξάρτητες τ.μ. που αντιπροσωπεύουν τα σφάλματα των μετρήσεων, και  $Z_n \sim \mathcal{N}(0, \Sigma)$ , με πίνακα συνδιακύμανσης  $\Sigma = \text{diag}\{\sigma_g^2, \sigma_r^2\}$ . Στο μοντέλο αυτό, προκειμένου να λάβουμε υπόψη μας τον περιορισμό ότι οι μάζες είναι θετικές έχουμε χρησιμοποιήσει λογαριθμοκανονικές κατανομές σφαλμάτων. Συγκεκριμένα, για  $n = 1, 2, \dots, N$  υποθέτουμε:

$$Y_n = (\log Z_{n,j})_j = \begin{pmatrix} \log Q_n^g + \epsilon_n^g \\ \log Q_n^r + \epsilon_n^r \end{pmatrix} = \begin{pmatrix} \log(H^c(\tau_n; \mu_s, \sigma_s^2)Q_n^f) + \epsilon_n^g \\ \log Q_n^r + \epsilon_n^r \end{pmatrix}. \quad (2.4.11)$$

Για την αναπαράσταση ως μ.χ.κ. οι συγγραφείς σημειώνουν πως μία πιθανή επιλογή της διαδικασίας καταστάσεων είναι η  $X_n = (Q_n^f, Q_n^r)$ . Τότε:

$$X_{n+1} = \Phi_n(X_n, \xi_n; \mu), \quad (2.4.12)$$

$$Y_n = R_n(X_n, \epsilon_n; \mu), \quad (2.4.13)$$

όπου,  $X_n = (X_{n,1}, X_{n,2}) = (Q_n^f, Q_n^r)$  και τα διανύσματα  $\xi_n, \epsilon_n$ ,  $n \geq 1$  αντιπροσωπεύουν τα σφάλματα στο μοντέλο. Το διάνυσμα  $\mu$  περιλαμβάνει όλες τις παραμέτρους του μοντέλου ( $\mu_a, \lambda, \mu_\gamma, \sigma_\gamma^2, \gamma_0, \gamma_f$ ) μαζί με ορισμένες από τις διασπορές των σφαλμάτων ( $\sigma_\gamma^2, \sigma_q^2, \sigma_g^2, \sigma_r^2$ ). Οι συναρτήσεις  $\Phi_n(\cdot)$  και  $R_n(\cdot)$  δίνονται στο λήμμα που ακολουθεί:

**Λήμμα 2.1** Το μοντέλο που διαμορφώνεται από τις εξισώσεις (2.4.1 -2.4.11) μπορεί να αναπαρασταθεί ως ένα μοντέλο χώρου-καταστάσεων, με εξίσωση καταστάσεων που δίνεται από το σύστημα που ακολουθεί:

$$X_{n+1,1} = \Phi_{n,1}(X_n, \xi_n; \mu) = X_{n,1} + \frac{F_n(H^c(\tau_n; \mu_s, \sigma_s^2)X_{n,1}; \mu_a, \lambda)e^{\xi_{n,1}}}{1 + (\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1)e^{-\xi_{n,2}}} \quad (2.4.14)$$

$$X_{n+1,2} = \Phi_{n,2}(X_n, \xi_n; \mu) = X_{n,2} + \frac{F_n(H^c(\tau_n; \mu_s, \sigma_s^2)X_{n,1}; \mu_a, \lambda)e^{\xi_{n,1}}}{1 + (\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1)^{-1}e^{\xi_{n,2}}} \quad (2.4.15)$$

όπου η  $(\xi_n)_{n \geq 1}$  ακολουθία ανεξάρτητων διδιάστατων κεντροποιημένων κανονικών και ανεξάρτητη του  $X_0$ , με διαγώνιο πίνακα συνδιακυμάνσεων που περιέχει τις διασπορές  $\sigma_1^2$  και  $\sigma_2^2$  που αντιστοιχούν στις  $X_{n,1}$  και  $X_{n,2}$ . Η εξίσωση των παρατηρήσεων είναι:

$$Y_n = R_n(X_n, \epsilon_n; \mu) = \begin{pmatrix} \log H^c(\tau_n; \mu_s, \sigma_s^2) + \log X_{n,1} \\ \log X_{n,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{n,1} \\ \epsilon_{n,2} \end{pmatrix}, \quad (2.4.16)$$

όπου η  $(\epsilon_n)_{n \geq 1}$  είναι ακολουθία ανεξάρτητων και ισόνομων διδιάστατων κεντροποιημένων κανονικών με διαγώνιο πίνακα συνδιακυμάνσεων με διασπορές  $\sigma_i^2$  για τα  $\epsilon_{n,i}$ ,  $i = 1, 2$  και ανεξάρτητη της  $(\xi_n)_{n \geq 1}$ .

### Απόδειξη

Λόγω της επιλογής του διανύσματος καταστάσεων, είναι  $X_{n,1} = Q_n^f$  και  $X_{n,2} = Q_n^r$ . Από τις  $Q_{n+1}^f = Q_n^f + \Gamma_n Q_n$  και  $Q_{n+1}^r = Q_n^r + (1 - \Gamma_n)Q_n$ , για να έχουμε τις (2.4.14) και (2.4.15) αρκεί να δείξουμε ότι:

$$\Gamma_n Q_n = \frac{F_n(H^c(\tau_n; \mu_s, \sigma_s^2)X_{n,1}; \mu_a, \lambda)e^{\xi_{n,1}}}{1 + (\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1)e^{-\xi_{n,2}}} \quad (2.4.17)$$

και

$$(1 - \Gamma_n)Q_n = \frac{F_n(H^c(\tau_n; \mu_s, \sigma_s^2)X_{n,1}; \mu_a, \lambda)e^{\xi_{n,1}}}{1 + (\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1)^{-1}e^{\xi_{n,2}}} \quad (2.4.18)$$

Επίσης, από τις (2.4.1) και (2.4.10) βλέπουμε ότι ο αριθμητής, που είναι κοινός για τα παραπάνω κλάσματα, είναι ίσος με  $Q_n$ . Από την (2.4.6) έχουμε:

$$\log\left(\frac{\Gamma_n}{1 - \Gamma_n}\right) = \log\left(\frac{\gamma_n(\tau_n)}{1 - \gamma_n(\tau_n)}\right) + \xi_{n,2} \Leftrightarrow \Gamma_n = \frac{\gamma_n(\tau_n)e^{-\xi_{n,2}}}{1 + \gamma_n(\tau_n)(e^{-\xi_{n,2}} - 1)} \Leftrightarrow$$

$$\Gamma_n = \frac{1}{1 + (\gamma_n^{-1}(\tau_n) - 1)e^{-\xi_{n,2}}},$$

και

$$1 - \Gamma_n = \frac{1}{1 + (\gamma_n^{-1}(\tau_n) - 1)^{-1}e^{\xi_{n,2}}}.$$

Πολλαπλασιάζοντας τις τελευταίες με  $Q_n$ , προκύπτουν οι ζητούμενες σχέσεις: (2.4.17) και (2.4.18).  $\square$

## Κεφάλαιο 3

# Στοιχεία θεωρίας μαρκοβιανών αλυσίδων σε γενικούς χώρους

### 3.1 Εισαγωγή

Οι μαρκοβιανές στοχαστικές διαδικασίες έχουν κεντρικό ρόλο σε εφαρμογές τόσο της θεωρίας πιθανοτήτων, όσο και στη στατιστική καθώς αποτελούν ένα ιδανικό εργαλείο για τη μοντελοποίηση πολλών συστημάτων είτε σε συνεχή είτε σε διακριτό χρόνο. Οι διαδικασίες αυτές, όταν είναι διακριτές, αναφέρονται ως μαρκοβιανές αλυσίδες (μ.α.). Στο κεφάλαιο αυτό παρουσιάζονται κάποια βασικά στοιχεία της θεωρίας των μ.α. με τιμές σε γενικό χώρο (μετρήσιμο τοπολογικό χώρο ή μετρικό χώρο). Ο κύριος λόγος είναι πως τα οι τεχνικές MCMC που θα παρουσιαστούν (και χρησιμοποιούνται), βασίζονται στα αποτελέσματα (ασυμπτωτικά και μη) της θεωρίας για γενικούς χώρους. Επίσης, η στατιστική ανάλυση για τα μοντέλα ανάπτυξης φυτών που παρουσιάστηκαν στηρίζεται στην θεωρία των κρυμμένων μαρκοβιανων μοντέλων (KMM), τα οποία είναι στην ουσία μία μαρκοβιανή διαδικασία. Για την περίπτωση του αριθμήσιμου χώρου θα γίνουν ορισμένες βασικές αναφορές ώστε να είναι σαφής η σύνδεση αλλά η γενική θεωρία θεωρείται γνωστή.

Το κεφάλαιο οργανώνεται ως εξής: Αρχικά θα γίνει η εισαγωγή ενός εξαιρετικά χρήσιμου εργαλείου που λέγεται πυρήνας και με την εφαρμογή του στις μ.α. θα μπορέσουμε να πραγματοποιήσουμε μία ομαλή μετάβαση της θεωρίας των μ.α., από αριθμήσιμους, σε γενικούς χώρους καταστάσεων (πεδίο τιμών της σ.δ.). Στη συνέχεια θα παρουσιαστούν κάποιες βασικές έννοιες και ιδιότητες για τις μ.α., ενώ γίνεται και μία σύντομη αναφορά στο θεωρητικό ζήτημα της ύπαρξης και της κατασκευής. Πρωταρχικές έννοιες των σ.δ. υπάρχουν στο παράρτημα. Στη συνέχεια παρουσιάζονται τα βασικότερα ασυμπτωτικά αποτελέσματα που αφορούν μ.α. με αριθμήσιμο και γενικό χώρο. Με βάση τα αποτελέσματα αυτά μπορούμε να καλύψουμε την ασυμπτωτική συμπεριφορά των αλγορίθμων MCMC που παρουσιάζονται στο επόμενο κεφάλαιο. Τέλος, γίνεται μία σύντομη παρουσίαση της θεωρίας των κρυμμένων μαρκοβιανών μοντέλων, με ορισμένες εφαρμογές, και περιγράφεται η σύνδεση αλλά και ο ρόλος τους στη στατιστική.

### 3.1.1 Ο Πυρήνας μιας Στοχαστικής Διαδικασίας και οι βασικότερες ιδιότητες του

Ο πυρήνας, ο οποίος συχνά αναφέρεται και ως μεταβατικός πυρήνας ή πυρήνας μεταβάσεων (transition kernel), είναι ένα βασικό εργαλείο που θα μας επιτρέψει να διαχειριστούμε στοχαστικές διαδικασίες που παίρνουν τιμές σε υπεραριθμήσιμους χώρους. Ως μαθηματική έννοια, ο πυρήνας συνδέεται με την (μετροθεωρητική) έννοια της δεσμευμένης πιθανότητας και της δεσμευμένης κατανομής πιθανότητας. Για τις έννοιες αυτές που είναι ζωτικής σημασίας για την κατασκευή και την ανάλυση σχεδόν όλων των σ.δ., μία πολύ πλούσια περιγραφή και συζήτηση γίνεται στα: [19] (κεφάλαια 24-33), [9] και [1].

Το πλεονέκτημα με τη χρήση του πυρήνα είναι ότι μπορούμε να διατυπώσουμε πολύπλοκες εκφράσεις και έννοιες που αφορούν τις στοχαστικές διαδικασίες με έναν πολύ απλό τρόπο. Επίσης, μας παρέχει ένα ενιαίο πλαίσιο για τη μελέτη των στοχαστικών διαδικασιών με αριθμήσιμο ή υπεραριθμήσιμο χώρο καταστάσεων, διακριτού, είτε συνεχούς χρόνου. Ιδιαίτερα για τις μαρκοβιανές αλυσίδες, ο συμβολισμός είναι εξαιρετικά εύχρηστος.

**Ορισμός 3.1** (πυρήνας). Έστω  $(X, \mathcal{X})$  και  $(Y, \mathcal{Y})$  δύο μετρήσιμοι χώροι. Η συνάρτηση  $Q : X \times \mathcal{Y} \rightarrow [0, \infty]$  που ικανοποιεί τις:

- (i) για κάθε  $x \in X$ , το  $Q(x, \cdot)$  είναι θετικό μέτρο στον  $(Y, \mathcal{Y})$ , και
- (ii) για οποιοδήποτε  $A \in \mathcal{Y}$ , η συνάρτηση  $Q(\cdot, A)$ ,  $x \mapsto Q(x, A)$  είναι Borel-μετρήσιμη,

λέγεται μη κανονικοποιημένος πυρήνας (μ.κ.π.) από τον  $(X, \mathcal{X})$  στον  $(Y, \mathcal{Y})$ .

Επιπλέον, αν:

- (α)  $Q(x, Y) = 1$  για κάθε  $x \in X$ , τότε ο  $Q$  λέγεται (κανονικοποιημένος) πυρήνας. Στην περίπτωση αυτή, το  $Q(x, \cdot)$  είναι μέτρο πιθανότητας στον  $(Y, \mathcal{Y})$  για κάθε  $x \in X$
- (β)  $(X, \mathcal{X}) = (Y, \mathcal{Y})$  και  $Q(x, X) = 1$  για κάθε  $x \in X$ , τότε ο  $Q$  λέγεται μαρκοβιανός πυρήνας στον  $(X, \mathcal{X})$ .

Ένας (μ.κ.π.)  $Q : X \times \mathcal{Y} \rightarrow [0, \infty]$  έχει πυκνότητα, ως προς κάποιο μέτρο  $\nu$  στον  $(Y, \mathcal{Y})$ , αν υπάρχει μία μη-αρνητική συνάρτηση  $q : X \times Y \rightarrow [0, \infty]$ , μετρήσιμη ως προς τη σ-άλγεβρα γινόμενο  $\mathcal{X} \otimes \mathcal{Y}$ , τέτοια ώστε:

$$Q(x, A) = \int_A q(x, y) \nu(dy). \quad (3.1.1)$$

Συνήθως η πυκνότητα αυτή συμβολίζεται με  $q(x, y) = q(y|x)$  (χρησιμοποιώντας το συμβολισμό δεσμευμένων συναρτήσεων πυκνότητας-πιθανότητας)

Ειδικά στην περίπτωση όπου τα  $X, Y$  είναι αριθμήσιμα σύνολα, συνηθίζεται να γράφουμε  $Q(x, y)$  για τις πιθανότητες μετάβασης, αντί για  $Q(x, \{y\})$  και να λέμε τον  $Q$  πίνακα

## Εισαγωγή

(πιθανοτήτων) μετάβασης.

Συμβολισμός: Το ολοκλήρωμα μίας συνάρτησης  $f : Y \rightarrow \mathbb{R}$  ως προς το μέτρο  $Q(x, \cdot)$  θα συμβολίζεται με  $\int f(y)Q(x, dy)$  ή με  $Qf$ . Ο τελευταίος είναι ιδιαίτερα εύχρηστος. Συγκεκριμένα:

Αν  $f \in \mathcal{E}^+$ , ορίζουμε:

$$Qf(x) \stackrel{\text{def}}{=} \int_X f(y)Q(x, dy), \quad x \in X, \quad (3.1.2)$$

ενώ αν  $\nu \in \mathcal{M}^+$ ,

$$\nu Q(A) \stackrel{\text{def}}{=} \int_X Q(x, A)\nu(dx), \quad A \in \mathcal{Y}, \quad (3.1.3)$$

Τότε οι απεικονίσεις  $f \rightarrow Qf$  και  $\nu \rightarrow \nu Q$  είναι τελεστές με  $Qf : \mathcal{E}^+ \rightarrow (X, \mathcal{X})$  και  $\nu Q : \mathcal{M}^+ \rightarrow (Y, \mathcal{Y})$  αντίστοιχα, τέτοιοι ώστε:  $Q(f+g) = Qf + Qg$ ,  $Q(af) = aQf$  για  $f, g \in \mathcal{E}^+$ ,  $a \geq 0$  και ομοίως για τον  $\nu Q$ . Με  $\mathcal{E}^+$  συμβολίζεται το σύνολο όλων των μετρήσιμων συναρτήσεων  $f : Y \rightarrow \overline{\mathbb{R}}_+$  και με  $\mathcal{M}^+$  το σύνολο όλων των θετικών μέτρων του  $(X, \mathcal{X})$ .

Επομένως, μπορούμε να δούμε τον  $Q$  ως τελεστή που επιδρά στις συναρτήσεις από δεξιά και στα μέτρα από αριστερά.

Επαληθεύεται άμεσα πως το  $\nu Q$  είναι θετικό μέτρο στον  $(Y, \mathcal{Y})$  όπως επίσης και ότι η  $Qf$  είναι μετρήσιμη συνάρτηση από τον  $X$  στο  $[0, \infty]$ . Για το δεύτερο, αν  $f = \mathbb{1}_A$ , για  $A \in \mathcal{Y}$ , τότε η  $Qf(x) = Q(x, A)$  είναι μετρήσιμη για κάθε  $x \in X$ . Αυτό επεκτείνεται για απλές μετρήσιμες και στη συνέχεια, με το Θεώρημα μονότονης σύγκλισης, η μετρησιμότητα της  $Qf$  επεκτείνεται για θετικές μετρήσιμες  $f$ . Επίσης, για  $\nu = \delta_x$ , οι  $\delta_x Q(A)$  και  $Q\mathbb{1}_A$  εκφράζουν την ίδια ποσότητα. Όταν τα  $X, Y$  είναι πεπερασμένα σύνολα, το  $Qf$  αντιστοιχεί στον πολλαπλασιασμό του πίνακα  $Q$  με το διάνυσμα στήλη,  $f$ , και το  $\nu Q$  αντιστοιχεί στον πολλαπλασιασμό του διανύσματος γραμμή,  $\nu$ , με τον πίνακα  $Q$ .

**Ορισμός 3.2** Αν  $Q, R$  δύο μ.κ.π από τον  $(\mathcal{X}, \mathcal{X})$  στον  $(\mathcal{Y}, \mathcal{Y})$  και από τον  $(\mathcal{Y}, \mathcal{Y})$  στον  $(\mathcal{Z}, \mathcal{Z})$  αντίστοιχα, τότε ορίζεται ο πυρήνας γινόμενο,  $QR$  ως:

$$QR(x, A) \stackrel{\text{def}}{=} \int Q(x, dy)R(y, A), \quad x \in \mathcal{X}, A \in \mathcal{Z}. \quad (3.1.4)$$

Προκύπτει άμεσα ότι η συνάρτηση  $QR$  είναι μ.κ.π. από τον  $(\mathcal{X}, \mathcal{X})$  στον  $(\mathcal{Z}, \mathcal{Z})$  και αν οι δύο είναι κανονικοποιημένοι, τότε και ο  $QR$  είναι επίσης, δηλαδή  $QR(x, \mathcal{Z}) = 1$  για κάθε  $x \in \mathcal{X}$ .

Επίσης, ο πυρήνας που λειτουργεί ως μοναδιαίο στοιχείο στον παραπάνω πολλαπλασιασμό, είναι ο:

$$I(x, A) = \begin{cases} 1, & \text{αν } x \in A \\ 0, & \text{διαφορετικά} \end{cases} \quad (3.1.5)$$

Όπου η απεικόνιση  $x \mapsto I(x, A)$  αντιστοιχεί στη δείτρια του συνόλου  $A$ , ενώ η απεικόνιση  $A \mapsto I(x, A)$  αντιστοιχεί στο μέτρο  $\delta_x(A)$ . Επομένως, είναι  $IQ = Q = QI$ .

### 3.1.2 Πυρήνας και Τυχαίες Μεταβλητές

Η πιο σημαντική εφαρμογή των πυρήνων είναι στην περιγραφή των μεταβάσεων μιας μαρκοβιανής αλυσίδας. Πριν φτάσουμε όμως εκεί, θα δούμε πως γίνεται γενικά η σύνδεση του πυρήνα με τις τυχαίες μεταβλητές.

Έστω  $(\Omega, \mathcal{F}, \mathbb{P})$  χώρος πιθανότητας και  $X$  μία τ.μ. (δηλαδή μετρήσιμη συνάρτηση με τιμές στον  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$  και ορισμένη στον παραπάνω χώρο). Έστω επίσης  $\mathcal{G} \subset \mathcal{F}$ , μία σ-άλγεβρα. Ένα βασικό αποτέλεσμα είναι ότι υπάρχει πάντα πυρήνας  $Q : \Omega \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  που είναι κανονική έκδοση δεσμευμένης πιθανότητας της  $X$  δοθέντος της  $\mathcal{G}$  (Fristedt and Gray (1996)). Δηλαδή να ισχύει:

- (1)  $\mathbb{P}(X \in B | \mathcal{G}) = E[\mathbb{1}_{\{X \in B\}} | \mathcal{G}](\omega) = Q(\omega, B)(\omega)$   $\mathbb{P}$ -σ.π. για κάθε  $B \in \mathcal{B}(\mathbb{R})$ ,
- (2) Το  $Q(\omega, B)$  να είναι μέτρο πιθανότητας στο  $B$  για κάθε  $\omega \in \Omega$ .

Το αντίστροφο επίσης ισχύει. Δηλαδή, κάθε πυρήνας είναι μία κανονική έκδοση δεσμευμένης πιθανότητας για κάποια τ.μ.

Πιο απλά, ο πυρήνας περιγράφει τη δεσμευμένη κατανομή της τ.μ.  $Y | X = x$ . Δηλαδή,

$$Q(x, A) = \mathbb{P}(Y \in A | X = x), \quad (3.1.6)$$

Ωστόσο, το δεξί μέλος δεν είναι καλά ορισμένο όταν  $\mathbb{P}(X = x) = 0$  και επομένως η παραπάνω είναι ουσιαστικά συντομογραφία για την:

$$Q(\cdot, A) = \mathbb{P}(A | \sigma(X)), \quad \mathbb{P}\text{-σ.π.}, \quad (3.1.7)$$

στην οποία το δεξί μέλος είναι ο (μετροθεωρητικός) ορισμός της δεσμευμένης πιθανότητας. Η έννοια αυτή αποτελεί ένα εξαιρετικά σύνθετο και προχωρημένο ζήτημα της θεωρίας πιθανοτήτων το οποίο όμως έχει ουσιαστική σημασία και είναι απαραίτητο για τη κατασκευή και την ανάλυση των σ.δ. Στην εισαγωγή έχει προταθεί βιβλιογραφία για τα συγκεκριμένα ζητήματα.

Πιο γενικά, αν  $(X_n)_{n \geq 0}$  μία χρονικά ομογενής στοχαστική διαδικασία διακριτού χρόνου με τιμές στον  $\mathcal{X}$ , τότε μπορούμε να ορίσουμε τον πυρήνα  $Q : (\mathcal{X} \times \mathcal{X}) \mapsto [0, 1]$ , για τη στοχαστική διαδικασία, ώστε να έχουμε μία έκφραση για τις δεσμευμένες κατανομές των  $X_n | X_{n-1} = x$  για  $n \geq 1$ . Δηλαδή,  $Q(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x)$ . Αν επιπλέον έχουμε μία στοχαστική διαδικασία  $(Y_n)_{n \geq 0}$  με τιμές στον  $\mathcal{Y}$ . Τότε ο πυρήνας  $R : \mathcal{X} \times \mathcal{Y} \mapsto [0, 1]$  αντιστοιχεί στις δεσμευμένες κατανομές των  $Y_n | X_{n-1} = x$ .

Επομένως, η σχέση (3.1.2) είναι μία ισοδύναμη έκφραση για την:

$$E[f(X_n) | X_{n-1} = x] = \int_{\mathcal{X}} f(y) P(x, dy), \quad (3.1.8)$$

όταν  $Q = P$  και όπου το μέτρο πιθανότητας  $P(x, A)$ , αντιστοιχεί στην κατανομή της  $X_n | X_{n-1} = x$  υποθέτοντας ότι αυτή είναι ίδια για κάθε  $n$  αλλά και ότι η μέση τιμή υπάρχει. Μάλιστα, αν αυτή η κατανομή είναι απόλυτα συνεχής με πυκνότητα  $g(y|x)$ , τότε  $P(x, A) = \int_A g(y|x) dy$  και έχουμε και έτσι τη σχέση (3.1.1) (αν δούμε το τελευταίο ως ολοκλήρωμα Lebesgue με  $\nu$  ως μέτρο αναφοράς).

## Εισαγωγή

Επίσης, αν  $\nu$  η κατανομή της  $X_{n-1}$  τότε, η κατανομή της  $X_n$  δίνεται από την:

$$\mathbb{P}(X_n \in A) = \int_X Q(x, A)\nu(dx), \quad (3.1.9)$$

στην οποία αναγνωρίζουμε από τη σχέση (3.1.3) πως το δεξί μέλος είναι το  $\nu Q(A)$ .

Ακόμη, με τον πυρήνα μπορούμε να έχουμε εκφράσεις για τις δεσμευμένες κατανομές των  $X_{n+k} | X_n = x$  σε χρονικά ομογενείς σ.δ.. Θα χρειαστούμε τον ορισμό (3.1.4) όπως επίσης και τον παρακάτω:

**Ορισμός 3.3** Έστω  $Q$  ένας μ.κ.π. μαρκοβιανός πυρήνας στον  $X, \mathcal{X}$ . Τότε ορίζεται το αναδρομικό σχήμα:

$$Q^{(0)}(x, \cdot) = I(x, \cdot) \quad \text{για } x \in X \quad \text{και} \quad Q^{(k)} = QQ^{(k-1)} \quad \text{για } k \geq 1. \quad (3.1.10)$$

Με τον παραπάνω ορισμό βλέπουμε πως:

$$\mathbb{P}(X_{n+2} \in A | X_n = x) = \int Q(x, dy)Q(y, A). \quad (3.1.11)$$

Το δεύτερο μέλος της παραπάνω είναι το  $Q^{(2)}$  και επομένως, συνεχίζοντας την παραπάνω διαδικασία έχουμε αναδρομικά ότι:

$$\mathbb{P}(X_{n+k} \in A | X_n = x) = Q^{(k)}(x, A). \quad (3.1.12)$$

Μπορούμε να παρατηρήσουμε επίσης πως για οποιουδήποτε θετικούς ακεραίους  $n, m$  ισχύει:

$$Q^{(n)}Q^{(m)} = Q^{(n+m)}, \quad (3.1.13)$$

δηλαδή:

$$Q^{(n+m)}(x, A) = \int Q^{(n)}(x, dy)Q^{(m)}(y, A). \quad (3.1.14)$$

Οι εξισώσεις αυτές είναι οι γνωστές εξισώσεις Chapman-Kolmogorov.

**Παράδειγμα** (Διαδικασία Galton-Watson). Η διαδικασία Galton-Watson είναι μία από τις πιο γνωστές κλαδωτές ανελίξεις (Branching processes) και χρησιμοποιείται ευρέως ως μαθηματικό μοντέλο για τη μελέτη του μεγέθους ενός πληθυσμού. Συγκεκριμένα, είναι η στοχαστική διαδικασία που περιγράφεται στη συνέχεια. Έστω  $\xi_n^r$ ,  $n \geq 0$ ,  $r \geq 1$  ανεξάρτητες και ισόνομες, θετικές ακέραιες τ.μ. που αντιπροσωπεύουν τους απογόνους που θα αφήσει το  $r$ -οστό άτομο της γενιάς  $n$  ενός πληθυσμού (αν αυτό υπάρχει), στην αμέσως επόμενη γενιά. Οι απόγονοι όλων των ατόμων της  $n$ -οστής γενιάς (που θα υπάρχουν στη  $n+1$ ), αποτελούν τον ενεργό πληθυσμό της  $(n+1)$ -οστής γενιάς. Έστω  $Z_{n+1} = \xi_n^{(1)} + \dots + \xi_n^{(Z_n)}$ . Η διαδικασία  $\{Z_n, n \geq 0\}$  ονομάζεται διαδικασία Galton-Watson (GWP) και αντιπροσωπεύει το πλήθος ενεργών ατόμων στη  $n$ -οστή γενιά. Αυτή η στοχαστική διαδικασία έχει πυρήνα  $Q$  που ορίζεται για κάθε  $n \geq 1$  ως:

$$Q(x, A) = \begin{cases} \delta_0(A) & \text{αν } x = 0 \\ \mathbb{P}(\xi_n^{(1)} + \dots + \xi_n^{(x)} \in A) & \text{διαφορετικά} \end{cases} \quad (3.1.15)$$



## Εισαγωγή

Υποθέτουμε ότι οι α.ι.τ.μ  $\xi_n^{(r)}$  έχουν κάποια διακριτή κατανομή, έστω  $\mu$  (αφού αντιπροσωπεύουν πλήθος ατόμων). Δηλαδή  $\mu(j) = \mathbb{P}(\xi_n^{(r)} = j)$  για κάθε  $n \geq 0$  και κάθε  $r \geq 1$ . Υποθέτουμε πως το πλήθος παιδιών του  $r$ -οστού ατόμου στη γενιά  $n$  (που θα υπάρχουν στην επόμενη), δεν εξαρτάται από το πόσα άτομα υπάρχουν σε αυτή. Αν και αυτό δεν είναι ρεαλιστικό, μπορούμε να σκεφτόμαστε έναν πληθυσμό μονογονικών οργανισμών. Πχ τα φυτά, αλλά και γενικότερα μονοκύτταρους οργανισμούς. Αυτό δικαιολογεί την ανεξαρτησία. Για την ισονομία μπορούμε να σκεφτόμαστε ότι κάθε άτομο είναι ένα πιστό αντίγραφο του γεννήτορα. Υποθέτουμε χωρίς βλάβη της γενικότητας (όπως θα δούμε στη συνέχεια) ότι υπάρχει ένας γεννήτορας. Αυτό το δηλώνουμε με την αρχική κατανομή της διαδικασίας. Συγκεκριμένα, για  $n = 0$  είναι  $Z_0 = \xi_n^{(r)}$  και υποθέτουμε αρχική κατανομή την  $\delta_1$ .

Η πιο σημαντική εφαρμογή των πυρήνων (μετάβασης) είναι στην περιγραφή της δυναμικής μεταβάσεων μιας μαρκοβιανής αλυσίδας για τις οποίες γίνεται συζήτηση στην επόμενη παράγραφο. Ορισμένα εισαγωγικά στοιχεία στη θεωρία στοχαστικών διαδικασιών υπάρχουν στο παράρτημα A1.

### 3.1.3 Ορισμοί-Βασικές έννοιες μ.α.

**Ορισμός 3.4** (Ομογενής Μαρκοβιανή Αλυσίδα). Έστω  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$  χώρος πιθανότητας με φίλτρο και  $Q$  ένας μαρκοβιανός πυρήνας στον  $(\mathcal{X}, \mathcal{X})$ . Μία στοχαστική διαδικασία  $(X_n)_{n \geq 0}$  με τιμές στον  $(\mathcal{X}, \mathcal{X})$  λέγεται (ομογενής) μαρκοβιανή αλυσίδα (κάτω από το  $\mathbb{P}$ ), ως προς τη διήθηση  $(\mathcal{F}_n)_{n \geq 0}$  και με πυρήνα τον  $Q$ , αν:

- (i) είναι προσαρμοσμένη στην  $(\mathcal{F}_n)_{n \geq 0}$ ,
- (ii)  $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q(X_n, A)$  για κάθε  $n \geq 0$  και  $A \in \mathcal{X}$ .

Η κατανομή της  $X_0$ , έστω  $\mu$ ,  $(\mathbb{P}(X_0 \in A) = \mu(A))$  λέγεται αρχική κατανομή της αλυσίδας και το σύνολο  $\mathcal{X}$  λέγεται χώρος καταστάσεων της αλυσίδας και είναι ένας αυθαίρετος μετρήσιμος χώρος (όχι απαραίτητα το  $\mathbb{R}^d$ ). Οι πιθανότητες στο (ii) που καθορίζονται από τον πυρήνα, ονομάζονται πιθανότητες μετάβασης.

Διαισθητικά, η ποσότητα  $Q(x, A)$  αντιπροσωπεύει την πιθανότητα η αλυσίδα όντας στην κατάσταση  $x$  τη χρονική στιγμή  $n$ , να κάνει μία μετάβαση στο σύνολο  $A$  την επόμενη χρονική στιγμή, ενώ ο όρος ομογενής σημαίνει πως αυτή η πιθανότητα αυτή είναι ανεξάρτητη του χρόνου. Δηλαδή:

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mathbb{P}(X_n \in A | \mathcal{F}_{n-1}) \quad \text{για κάθε } n \geq 1, \quad (3.1.16)$$

το οποίο δηλώνεται έμμεσα μέσω της (ii), καθώς σε διαφορετική περίπτωση θα γράφαμε:

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q_n(X_n, A), \quad (3.1.17)$$

ώστε να δηλώσουμε πως ο πυρήνας πιθανώς εξαρτάται από το  $n$ .

Η ιδιότητα (ii) λέγεται μαρκοβιανή ιδιότητα και εκφράζει ότι αν (για κάποιο  $n \geq 1$ ) είναι γνωστή η τιμή της  $X_n$ , (δηλαδή  $X_n = x$ ) τη χρονική στιγμή  $n$ , η γνώση αυτή καθορίζει (πλήρως) την κατανομή της  $X_{n+1}$  (δηλαδή την αμέσως επόμενη χρονική στιγμή  $(n+1)$ ). Επομένως, η επιπρόσθετη πληροφορία που παρέχει η  $(\mathcal{F}_k)_{k \leq n-1}$  για τις τιμές που είχε η  $\{X_n, n \geq 0\}$  στο παρελθόν (πριν τη χρονική στιγμή  $n$ ) δεν μας είναι χρήσιμη. Αυτό εκφράζεται και μέσω της:

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A | \sigma(X_n)) \quad (3.1.18)$$

Η ιδιότητα αυτή λέγεται επίσης και ιδιότητα έλλειψης μνήμης. Χαλαρώνοντας την ιδιότητα αυτή στην απαίτηση (ii) του ορισμού της μ.α., θα λέμε ότι:

**Ορισμός 3.5** (Μαρκοβιανή αλυσίδα τάξης  $m$ ) Μία μ.α. είναι τάξης  $m$  για  $m \geq 1$  αν:

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q(X_{n-m+1:n}, A). \quad (3.1.19)$$

**Παρατήρηση** Με άλλα λόγια, η μελλοντική κατάσταση εξαρτάται από παρελθόν μέσω των προηγούμενων  $m$  καταστάσεων. Αν η  $X$  είναι μ.α. τάξης  $m$  τότε η  $Y_n = (X_{n-m+1}, X_{n-1}, \dots, X_n)$  είναι μ.α. τάξης 1, δηλαδή, μ.α. με τον ορισμό που δόθηκε. Επομένως, αρκεί να αναπτύξουμε τη θεωρία μ.α. τάξης 1.

## Εισαγωγή

Στο εξής, όταν θα χρησιμοποιούμε το σύμβολο  $\mathbf{X}$  για να αναφερθούμε σε μία μ.α., θα εννοείται πως  $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbb{P})$ .

Αν η  $\mathbf{X}$  ικανοποιεί την (i), τότε για κάθε  $n \geq 0$  ισχύει  $\mathcal{F}_n \subset \mathcal{F}_n^0$ . Οπότε μία μ.α. ως προς τη διήθηση  $\mathcal{F}_n$  είναι και μ.α. ως προς τη φυσική διήθηση. Στην τελευταία περίπτωση, η  $\{X_n, n \geq 0\}$  καλείται απλά μ.α. και γενικότερα δεν θα γίνεται αναφορά στο μέτρο  $\mathbb{P}$  αν αυτό δεν κρίνεται αναγκαίο.

Επιπλέον, αν δεν γίνεται αναφορά στο χώρο καταστάσεων ή στον πυρήνα θα εννοείται ο χώρος  $(\mathcal{X}, \mathcal{X})$  και ο μαρκοβιανός πυρήνας  $Q$  στον  $\mathcal{X}$ .

### Παραδείγματα

1. Ο μοναδιαίος πυρήνας  $I(x, A)$  όπως ορίστηκε στην (3.1.5) έχει τη μαρκοβιανή ιδιότητα διότι αφού ισχύει  $\delta_x I = \delta_x$ , θα έχουμε ότι αν  $X_n \sim \delta_x$ , τότε  $X_{n+1} \sim \delta_x$ . Επομένως η αλυσίδα δεν αλλάζει κατάσταση ποτέ. Έτσι, ο μοναδιαίος πυρήνας είναι πυρήνας μετάβασης για την πιο αδιάφορη αλυσίδα που είναι δυνατόν να κατασκευαστεί.
2. **(Τυχαία δυναμικά συστήματα).** Έστω  $(\xi_n)_{n \geq 1}$  ακολουθία από ανεξάρτητες και ισόνομες τ.μ. με τιμές στο  $\mathbb{R}$  και κατανομή  $\nu$ . Ορίζουμε αναδρομικά τις τ.μ. :

$$X_0 = a, \quad X_{n+1} = f(X_n, \xi_{n+1}), \quad n \geq 0,$$

όπου  $f : E \times \mathbb{R} \mapsto E$ , μία μετρήσιμη συνάρτηση και  $a \in E$ . Τότε η  $(X_n)_{n \geq 0}$  είναι μία ομογενής μ.α. στον  $(E, \mathcal{E})$  με πυρήνα:

$$Q(x, A) = \int \mathbb{1}_{\{f(x,s) \in A\}} \nu(ds), \quad x \in E, A \in \mathcal{E}$$

Πράγματι, επειδή η  $\xi_{n+1}$  είναι ανεξάρτητη των  $X_0, \dots, X_n$ , είναι:

$$\begin{aligned} \mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) &= E[\mathbb{1}_{\{X_{n+1} \in A\}} | \mathcal{F}_n] \\ &= E[\mathbb{1}_{\{f(X_n, \xi_{n+1}) \in A\}} | \mathcal{F}_n] \\ &= h(X_n), \end{aligned}$$

όπου  $h(x) = E[\mathbb{1}_{\{f(x, \xi_{n+1}) \in A\}}] = Q(x, A)$ . Άρα,  $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q(X_n, A)$ .

3. **(Διαδικασία Galton-Watson).** Η διαδικασία Galton-Watson είναι μ.α. Πράγματι, αν  $\mathcal{F}_n^0$  η φυσική διήθηση, τότε είναι φανερό πως η τιμή της  $Z_n$  εξαρτάται μονάχα από την τιμή της  $Z_{n-1}$  και άρα

$$\mathbb{P}(Z_{n+1} = j | \mathcal{F}_n) = \mathbb{P}(Z_{n+1} = j | Z_n). \quad (3.1.20)$$

Ο πίνακας πιθανοτήτων μετάβασης (καθώς εδώ έχουμε αριθμήσιμο χώρο καταστάσεων) είναι ο  $Q$  όπως αυτός έχει οριστεί στην (3.1.15).

3. **(Αυτοπαλίνδρομο μοντέλο) (AR(1))** Θεωρούμε το μοντέλο που ορίζεται από τη σχέση:

$$X_n = aX_{n-1} + \epsilon_n, \quad (3.1.21)$$

## Εισαγωγή

όπου  $\epsilon_n$  ανεξάρτητες και ισόνομες τ.μ. με πυκνότητα  $f$ . Η παραπάνω σχέση ορίζει μία μ.α. τέτοια ώστε:

$$q(x, y) = f(y - ax), \quad (3.1.22)$$

όπου  $q$  η πυκνότητα του πυρήνα. Συγκεκριμένα, αν  $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$ , τότε έχουμε πως

$$q(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - ax)^2\right\}. \quad (3.1.23)$$

Επίσης έχουμε ότι:

$$X_{n+2} = a(aX_n + \epsilon_{n+1}) + \epsilon_{n+2} = a^2X_n + a\epsilon_{n+1} + \epsilon_{n+2} \quad (3.1.24)$$

και έτσι:

$$X_{n+k} = a^k X_n + \sum_{i=1}^k a^{k-i} \epsilon_{n+i}. \quad (3.1.25)$$

Οπότε, όταν  $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$  είναι:

$$q^k(x, y) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2}(y - a^k x)^2\right\}, \quad (3.1.26)$$

όπου

$$\sigma_k^2 = \sigma^2 \sum_{i=1}^k (a^2)^{(k-i)} = \sigma^2 \frac{1 - a^{2k}}{1 - a^2}.$$

Ένα σενάριο στο οποίο ο πυρήνας δεν έχει πυκνότητα είναι να είχαμε ότι με πιθανότητα  $r(X_{n-1})$  να είναι  $X_n = aX_{n-1} + \epsilon_n$ , διαφορετικά, με πιθανότητα  $1 - r(X_{n-1})$  να είναι  $X_n = X_{n-1}$ . Στην περίπτωση αυτή, η διαδικασία εξακολουθεί να είναι μ.α., ωστόσο, πυκνότητα ως προς μέτρο Lebesgue δεν υπάρχει. Μπορούμε να γράψουμε:

$$Q(x, dy) = r(x)f(y - ax)dy + (1 - r(x))\delta_x(dy), \quad (3.1.27)$$

και τότε  $Q(x, A) = \int_A Q(x, dy)$ .

Αν ο  $\mathcal{X}$  είναι αριθμήσιμος και  $\mathcal{F}_n = \mathcal{F}_n^0 = \sigma(X_0, X_1, \dots, X_n)$  (η φυσική διήθηση), τότε ο προηγούμενος ορισμός παίρνει μία πιο απλή μορφή. Συγκεκριμένα, αφού ο  $\mathcal{X}$  είναι αριθμήσιμος, η  $\sigma$ -άλγεβρα  $\mathcal{F}_n^0$  παράγεται από τη διαμέριση του  $\Omega$  που σχηματίζεται από τα γεγονότα  $\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\}$ ,  $i_{0:n} \in \mathcal{X}^{(n+1)}$ , και έτσι η (ii) μπορεί να αντικατασταθεί από την:

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_{0:n} = i_{0:n}) = Q(i_n, i_{n+1}), \quad (3.1.28)$$

για κάθε  $n \geq 0$  και κάθε  $i_{0:n+1} \in \mathcal{X}^{(n+2)}$  τέτοιο ώστε  $\mathbb{P}(X_{0:n} = i_{0:n}) > 0$ .

Ως προς τη ύπαρξη και την κατασκευή των μ.α., παραθέτουμε τα σημαντικότερα αποτελέσματα χωρίς απόδειξη τα οποία δίνουν απάντηση σε 2 βασικά ερωτήματα: (1) Ποια είναι η πεπερασμένης διάστασης κατανομή μίας μ.α.; (2) Έστω ότι έχουμε έναν πυρήνα  $Q$  και μία αρχική κατανομή  $\mu$ . Υπάρχει μ.α. που έχει ως πυρήνα τον  $Q$  και

## Εισαγωγή

αρχική κατανομή το  $\mu$ .

Σχετικά με το (1): Προκειμένου να χαρακτηρίσουμε όλες τις μ.α. διακριτού χρόνου, αρκεί να γνωρίζουμε όλες τις πεπερασμένης-διάστασης κατανομές της διαδικασίας (δηλαδή την από κοινού κατανομή της διαδικασίας για οποιαδήποτε πεπερασμένη συλλογή δεικτών). Το γιατί αυτό αρκεί δεν είναι προφανές και μία συζήτηση σχετικά με αυτό γίνεται στο παράρτημα. Η απάντηση στο (1) είναι πως οι μ.α έχουν κατανομή πεπερασμένης διάστασης η οποία καθορίζεται πλήρως από την αρχική κατανομή και τον πυρήνα. Σχετικά με το (1) έχουμε το παρακάτω Θεώρημα του οποίου η απόδειξη είναι στο παράρτημα.

**Θεώρημα 3.1** (i) Έστω  $\mathbf{X}$  μία Μ.Α. με αρχική κατανομή  $\mu$  και πυρήνα  $Q$ . Τότε για κάθε  $B \in \mathcal{X}^{n+1}$ , της μορφής  $B = B_0 \times \dots \times B_n$  ισχύει:

$$\mathbb{P}(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \dots \int_{B_n} Q(x_{n-1}, dx_n), \quad (3.1.29)$$

(ii) Αντίστροφα, αν μία σ.δ.  $\mathbf{X} = (\Omega, \mathcal{F}, (X_n)_{n \geq 0}, \mathbb{P})$  με τιμές στον  $(\mathcal{X}, \mathcal{X})$ , ικανοποιεί την 3.1.3 για ένα μέτρο πιθανότητας  $\mu$  και ένα πυρήνα  $Q$ , τότε η  $\mathbf{X}$  είναι Μ.Α με αρχική κατανομή το  $\mu$  και πυρήνα το  $Q$  (ως προς τη φυσική διήθηση  $(\mathcal{F}_n^0)_{n \geq 0}$ )

Ορισμένες άμεσες συνέπειες του θεωρήματος 3.1.3 για μία μ.α  $\mathbf{X}$  με αρχική κατανομή  $\mu$ , πυρήνα  $Q$  και τιμές στον  $(\mathcal{X}, \mathcal{X})$  είναι οι παρακάτω:

1. Για κάθε  $n \geq 0$  και κάθε μετρήσιμη  $f : \mathcal{X}^{n+1} \mapsto \mathbb{R}$ ,  $\mathcal{X}^{\otimes(n+1)}$ , συνεχή και φραγμένη ( $f \in C_b(\mathcal{X}^{(k+1)})$ ) ισχύει:

$$E[f(X_0, \dots, X_n)] = \int f(x_0, \dots, x_n) \mu(dx_0) Q(x_0, dx_1) \dots Q(x_{n-1}, dx_n). \quad (3.1.30)$$

2. Αν στη 3.1.30 επιλέξουμε για  $k = n$ ,  $f(x_0, \dots, x_n) = \mathbb{1}_{\{x_0 \in B_0\}} \dots \mathbb{1}_{\{x_n \in B_n\}}$ ,  $B_i \subset X$ , τότε παίρνουμε:

$$\mathbb{P}(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0 \times \dots \times B_n} \mu(dx_0) Q(x_0, dx_1) \dots Q(x_{n-1}, dx_n). \quad (3.1.31)$$

που είναι η σχέση 3.1.3.

3. Χρησιμοποιώντας το συμβολισμό για τον πυρήνα θα είναι:

$$E[f(X_{n+1}) | \mathcal{F}_n] = (Qf)(X_n)$$

Ορίζοντας για  $n \geq 0$  το αναδρομικό σχήμα:  $Q^0 f = f$  και  $Q^n f = Q Q^{n-1} f$ , τότε έπεται πως:

$$\begin{aligned} E[f(X_{k+n}) | \mathcal{F}_k] &= E[E[f(X_{k+n}) | \mathcal{F}_{k+n-1}] | \mathcal{F}_k] = E[Qf(X_{k+n-1}) | \mathcal{F}_k] \\ &= E[E[Qf(X_{k+n-1}) | \mathcal{F}_{k+n-2}] | \mathcal{F}_k] = E[Q^2 f(X_{k+n-2}) | \mathcal{F}_k] \\ &= \dots = E[Q^n f(X_k) | \mathcal{F}_k] = Q^n f(X_k). \end{aligned}$$

## Εισαγωγή

4. Στην περίπτωση όπου ο  $\mathcal{X}$  είναι αριθμήσιμος, τότε οι 3.1.30, 3.1.31 γίνονται:

$$E[f(X_0, \dots, X_k)] = \sum_{x_0, \dots, x_k \in X} f(x_0, \dots, x_k) \mu(x_0) Q(x_0, x_1) \dots Q(x_{k-1}, x_k).$$

και

$$\mathbb{P}(X_j \in B_j, 0 \leq j \leq n) = \sum_{x_0 \in B_0, \dots, x_n \in B_n} \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n).$$

6. Επιπλέον, αν στη 3.1.31 επιλέξουμε  $B_i = x_i$  για κάποιους δείκτες και  $A_j = X$  για τους υπόλοιπους, για  $l_1, \dots, l_k \in \mathbb{N}$  θα είναι:

$$\mathbb{P}(X_0 = x_0, X_{l_1} = x_1, \dots, X_{l_1 + \dots + l_k} = x_k) = \mu(x_0) Q^{l_1}(x_0, x_1) \dots Q^{l_k}(x_{k-1}, x_k).$$

Οπότε για κάθε  $(x_{0:n}) \in \mathcal{X}^{(n+1)}$  τέτοια ώστε  $\mathbb{P}(X_{0:n} = x_{0:n}) > 0$  είναι:

$$\mathbb{P}(X_{n+1:n+k} = x_{n+1:n+k} | X_{0:n} = x_{0:n}) = Q(x_n, x_{n+1}) \dots Q(x_{n+k-1}, x_{n+k})$$

και

$$\mathbb{P}(X_{n+k} = y | X_{0:n} = x_{0:n}) = Q^k(x_n, y)$$

Η πεπερασμένη διάσταση κατανομή της μ.α. έστω  $(R_n)_{n \geq 0}$  (δηλαδή  $R_n = \mathbb{P}(X_j \in B_j, 0 \leq j \leq n)$ ) ικανοποιεί την ιδιότητα συνέπειας:  $R_n(B) = R_{n+1}(A \times \mathcal{X})$  (όπου  $\mathcal{X}$  ο χώρος καταστάσεων). Τότε, υπάρχει μοναδικό μέτρο πιθανότητας,  $R_\infty$ , για την ακολουθία  $X_0, X_1, \dots$  τέτοιο ώστε το  $R_\infty$  συμφωνεί με τις πεπερασμένες διάστασης κατανομές. Δηλαδή, αν  $B$  κάποιο μετρήσιμο σύνολο στον  $\mathcal{X}^n$ , και

$$B = \{(x_1, x_2, \dots) \in X^\infty : (x_1, \dots, x_n) \in B\}, \quad (3.1.32)$$

τότε  $R_n(B) = R_\infty(B)$ .

Αυτό έπεται από το Θεώρημα 3 του κεφαλαίου 22 στο Fristedt and Gray (1997).

Υποχρεωτική αναφορά στο μέτρο αυτό γίνεται μόνο σε οριακά θεωρήματα όπως στον ισχυρό νόμο των μεγάλων αριθμών (NMA) σύμφωνα με τον οποίο, η ακολουθία  $(X_n)_{n \geq 0}$  έχει την ιδιότητα:  $n^{-1} \sum_{k=1}^n X_k \rightarrow \mu$  με πιθανότητα 1. Η πιθανότητα εδώ αναφέρεται στο  $R_\infty$  καθώς αφορά πιθανότητες σε χώρο άπειρης διάστασης. Τις περισσότερες φορές ωστόσο, δουλεύουμε με τις κατανομές πεπερασμένης διάστασης. Το Κεντρικό Οριακό Θεώρημα για παράδειγμα αφορά μόνο τις κατανομές πεπερασμένης διάστασης.

Η απάντηση στο (2) δίνεται με το επόμενο Θεώρημα στο οποίο εισάγεται και ένας συμβολισμός που θα υιοθετήσουμε στην παρούσα εργασία.

**Θεώρημα 3.2** Έστω  $(\mathcal{X}, \mathcal{X})$  μετρήσιμος χώρος και  $\mu$  μέτρο πιθανότητας σε αυτόν, και  $Q$  πυρήνας στον  $(X, \mathcal{X})$ . Υπάρχει μοναδικό μέτρο πιθανότητας στον  $(X^\mathbb{N}, \mathcal{X}^{\otimes \mathbb{N}})$ , το οποίο συμβολίζεται με  $\mathbb{P}_\mu$ , τέτοιο ώστε η διαδικασία  $\{X_n : n \geq 0\}$  είναι Μ.Α (ως προς τη φυσική διήθηση) με αρχική κατανομή το  $\mu$  και πυρήνα το  $Q$

Συμβολισμός: Για  $x \in \mathcal{X}$  με  $\mathbb{P}_x$  συμβολίζουμε το  $\mathbb{P}_{\delta_x}$ , όπου  $\delta_x$  το μέτρο Dirac στο  $x$ :  $\delta_x(A) = \mathbb{1}_{\{x \in A\}}$ . Από την 3.1.3, για  $A = \{X_0 \in B_0, X_1 \in B_1, \dots, X_n \in B_n\}$  είναι:

$$\mathbb{P}_\mu(A) = \int \mu(dx) \mathbb{P}_x(A) \quad (3.1.33)$$

## Εισαγωγή

Η παραπάνω σχέση ισχύει για οποιοδήποτε  $A \in \mathcal{F}$ .

Πλήρης απόδειξη υπάρχει στα:

**Παράδειγμα** Έστω  $Z_n$  μία διαδικασία Galton-Watson με αρχική κατανομή  $Z_0 \sim \delta_i$ . Τότε,  $\mathbb{P}_i(Z_n = j) = \mathbb{P}(Z_n = j | Z_0 = i) = \mathbb{P}(\xi_{n-1}^{(1)} + \dots + \xi_{n-1}^{(i)} = \mu^{*i}(j))$ , όπου  $m$  η κατανομή των α.ι.τ.μ.  $\xi_n^{(r)}$ , δηλαδή  $\mu(j) = \mathbb{P}(\xi_n^{(r)} = j)$ , και με  $\mu^{*i}(j)$  συμβολίζουμε την  $i$ -οστή συνέλιξη του  $\mu$  με τον εαυτό του, υπολογισμένη στο σημείο  $j$ . Επειδή  $\mathbb{P}_1(Z_n = j) = \mu(j)$ , βλέπουμε ότι  $Q(i, j) = Q(1, j)^{*i}$  για κάθε  $j \geq 0$ . Αυτό μεταφράζεται στο ότι μία διαδικασία GW με  $i$  απογόνους είναι το άθροισμα  $i$  ανεξάρτητων διαδικασιών GW με έναν απόγονο. Το τελευταίο γεγονός είναι χρήσιμο διότι, για τη γενικότερη μελέτη της διαδικασίας GW μας επιτρέπει να υποθέτουμε, χωρίς βλάβη της γενικότητας, ότι  $Z_0 = 1$ .

Περισσότερα στοιχεία σχετικά με τη θεωρητική θεμελίωση των μ.α. γενικού χώρου υπάρχουν στα [30] και [17]. Θα κλείσουμε την ενότητα αυτή με τον παρακάτω ορισμό.

**Ορισμός 3.6** (Μη-Ομογενής Μαρκοβιανή Αλυσίδα). Έστω  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$  χώρος πιθανότητας με φίλτρο και  $\{Q_n\}_{n \geq 0}$  οικογένεια Μαρκοβιανών πυρήνων στον  $(X, \mathcal{X})$ . Μία στοχαστική διαδικασία  $\{X_n, n \geq 0\}$  με τιμές στον  $(X, \mathcal{X})$  λέγεται (μη-ομογενής) Μαρκοβιανή αλυσίδα κάτω από το  $\mathbb{P}$ , ως προς τη διήθηση  $(\mathcal{F}_n)_{n \geq 0}$  και με πυρήνες τους  $\{Q_n\}$ , αν:

(i) είναι προσαρμοσμένη στην  $(\mathcal{F}_n)_{n \geq 0}$ ,

και για κάθε  $n \geq 0$  και  $A \in \mathcal{X}$  ισχύει:

$$(ii) \mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q_n(X_n, A) \quad (3.1.34)$$

Αν  $\mu$  η κατανομή της  $X_0$  (αρχική κατανομή) τότε η κατανομή της  $X_n$  είναι:  $\mu Q_{0:n-1}$ , όπου  $Q_{i:j} = Q_i Q_{i+1} \dots Q_j$  για  $i \leq j$ .

## 3.2 Ασυμπτωτική Θεωρία Μ.Α.

### 3.2.1 Στάσιμο μέτρο

Όπως έχει τονιστεί, η βασικότερη διάκριση για τις μ.α. διακριτού χρόνου, όσον αφορά το πλαίσιο της θεωρίας, γίνεται με βάση το χώρο καταστάσεων, σε αριθμήσιμο και γενικό. Πριν ξεκινήσουμε την παρουσίαση της ασυμπτωτικής θεωρίας για γενικούς χώρους, αναφέρουμε τα βασικότερα ασυμπτωτικά αποτελέσματα σε αριθμήσιμους χώρους ώστε να είναι εφικτή σε κάποιο βαθμό μία σύνδεση με τα αντίστοιχα αποτελέσματα σε γενικούς χώρους.

Η θεωρία των μαρκοβιανών αλυσίδων διακριτού χρόνου και χώρου καταστάσεων επικεντρώνεται στην ασυμπτωτική συμπεριφορά της  $(Q^n(x, y))_{n \geq 0}$ . Τα βασικότερα αποτελέσματα αφορούν την ύπαρξη του ορίου:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Q^k(x, y), \quad \text{υπάρχει πάντα (για κάθε } x, y \in \mathcal{X}), \quad (3.2.1)$$

και κάτω από μία σχετικά ασθενή συνθήκη που ονομάζεται απεριοδικότητα:

$$\text{Το } \lim_{n \rightarrow \infty} Q^n(x, y) \text{ υπάρχει} \quad (3.2.2)$$

Ακόμη, σε ορισμένες βολικές περιπτώσεις, όταν η  $(X_n)_{n \geq 0}$  είναι αδιαχώριστη και θετικά επαναληπτική τα πρώτο από τα παραπάνω όρια είναι συνάρτηση πιθανότητας (για σταθερό  $x$ ) και μάλιστα ανεξάρτητη της αρχικής κατάστασης  $x$ . Συγκεκριμένα, σε αυτή την περίπτωση:  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Q^k(x, y) = \pi_y$  για κάθε  $x \in \mathcal{X}$ , όπου προκύπτει ότι το διάνυσμα  $(\pi_y)_{y \in \mathcal{X}}$  είναι το μοναδικό (στάσιμο) διάνυσμα πιθανοτήτων ( $\sum \pi_y = 1$ ) και το οποίο προκύπτει ως η (μοναδική) λύση του γραμμικού συστήματος:  $\pi_y = \sum_{x \in \mathcal{X}} \pi_x Q(x, y)$ . Επίσης είναι  $\pi_y > 0$  για κάθε  $y \in \mathcal{X}$ . Στην περίπτωση που έχουμε και την απεριοδικότητα, τα όρια (3.2.1) και (3.2.2) ταυτίζονται.

Αξίζει να τονισθεί ότι ένας από τους βασικότερος λόγος που εστιάζεται τόσο μεγάλο ενδιαφέρον για τις μ.α. είναι πως οι παραπάνω υποθέσεις είναι εύκολα επαληθεύσιμες και γενικότερα ικανοποιούνται σε πολύ μεγάλο εύρος εφαρμογών.

Όπως φαίνεται από τα παραπάνω, στην περίπτωση των Μ.Α. με αριθμήσιμο χώρο καταστάσεων, το στάσιμο μέτρο έχει το σπουδαιότερο ρόλο και μάλιστα, ο τρόπος υπολογισμού του είναι δύσκολος μόνο όταν το παραπάνω σύστημα είναι δύσκολο, αλλά ακόμα και τότε υπάρχουν στη μαθηματική βιβλιογραφία πολλές αριθμητικές και αλγοριθμικές τεχνικές για την επίλυση του. Όταν η Μ.Α. παίρνει τιμές σε γενικούς (υπεραριθμήσιμους εν γένει) χώρους, η κατάσταση είναι αρκετά διαφορετική. Το στάσιμο μέτρο εξακολουθεί να έχει κεντρικό ρόλο, ωστόσο, αυξάνει σημαντικά η δυσκολία επαλήθευσης των συνθηκών που απαιτούνται για την ύπαρξη του. Ακόμα και στις περιπτώσεις που είναι επαληθεύσιμες, ο υπολογισμός του στάσιμου μέτρου, εκτός τετριμμένων περιπτώσεων, είναι δύσκολος.

Αρχικά θα αναφέρουμε ποια διαδικασία καλείται στάσιμη και στη συνέχεια θα περάσουμε στον ορισμό και τις ιδιότητες του στάσιμου μέτρου για μ.α. σε γενικό χώρο.



Πριν φτάσουμε στα οριακά αποτελέσματα, θα γίνει η περιγραφή τω βασικών χαρακτηριστικών μιας μ.α. σε γενικό χώρο. Ουσιαστικά θα δούμε πως μεταφέρονται οι (γνωστές) έννοιες της απεριοδικότητας, επαναληπτικότητας κ.α. στην περίπτωση μετρήσιμου χώρου καταστάσεων.

**Ορισμός 3.7** (Στάσιμη Στοχαστική Διαδικασία). Μία στοχαστική διαδικασία  $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbb{P})$  λέγεται στάσιμη (υπό το  $\mathbb{P}$ ) αν οι πεπερασμένης διάστασης κατανομές της είναι αναλλοίωτες ως προς τη μεταφορά. Δηλαδή, αν για κάθε  $k, n \geq 1$  και όλα τα  $n_1, \dots, n_k$ , η κατανομή του τυχαίου διανύσματος  $(X_{n_1+n}, \dots, X_{n_k+n})$  δεν εξαρτάται από το  $n$ , δηλαδή τα τυχαία διανύσματα  $(X_{n_1+n}, \dots, X_{n_k+n})$  και  $(X_{n_1}, \dots, X_{n_k})$  είναι ισόνομα.

Δύο τυχαίες μεταβλητές μπορούν να είναι ανεξάρτητες με μόνο έναν τρόπο. Ωστόσο, πολλοί είναι οι τρόποι που μπορεί αυτές να είναι εξαρτημένες. Η στασιμότητα είναι μία ιδιότητα που μας επιτρέπει να χειριστούμε τη δομή της εξάρτησης. Το σημαντικότερο είναι πως πολλά από τα σπουδαία αποτελέσματα που έχουμε για ανεξάρτητες τ.μ. (Νόμοι Μεγάλων αριθμών, Κεντρικά Οριακά Θεωρήματα, κλπ) ισχύουν για στάσιμες στοχαστικές διαδικασίες. Πιο συγκεκριμένα, το ανάλογο του ισχυρού νόμου μεγάλων αριθμών, είναι το εργοδικό Θεώρημα του Birkhoff, σύμφωνα με το οποίο αν η  $(X_n)$  είναι στάσιμη στοχαστική διαδικασία, και η  $X_1$  έχει μέση τιμή (άρα και όλες, λόγω της στασιμότητας), τότε  $\bar{X}_n \xrightarrow{a.s.} Y$ , όπου η  $Y$  είναι μία τ.μ. τέτοια ώστε  $E(X_1) = E(Y)$ .

Ειδικά στην περίπτωση των χρονικά ομογενών μ.α., ο χαρακτηρισμός της στάσιμης στοχαστικής διαδικασίας απλουστεύεται σημαντικά επειδή οποιαδήποτε πολυδιάστατη κατανομή εκφράζεται μέσω της αρχικής κατανομής και του πυρήνα. Συγκεκριμένα:

**Ορισμός 3.8** (στάσιμη μ.α.). Μία μ.α. λέγεται στάσιμη, αν η αρχική της κατανομή  $\mu$  και ο πυρήνας  $Q$  ικανοποιούν την (εξίσωση ολικής ισορροπίας):

$$\mu Q = \mu$$

Για όσα ακολουθούν θα θεωρούμε την μ.α.  $\mathbf{X}$  με τιμές στον μετρήσιμο χώρο  $(\mathcal{X}, \mathcal{X}, \nu)$ . Δίνουμε αρχικά τον ορισμό του γενικότερου στάσιμου μέτρου (όχι απαραίτητα κατανομή).

**Ορισμός 3.9** (στάσιμο μέτρο). Οποιοδήποτε μέτρο  $\pi$  ικανοποιεί την:

$$\pi Q = \pi \tag{3.2.3}$$

λέγεται στάσιμο μέτρο (ή στάσιμη κατανομή όταν το  $\pi$  είναι μέτρο πιθανότητας) για τη μ.α. που έχει πυρήνα τον  $Q$ .

### Παρατηρήσεις

1. Η τελευταία εξίσωση είναι διατυπωμένη για την περίπτωση που έχουμε γενικό χώρο καταστάσεων της μ.α.. Διακρίνοντας στις δύο βασικές περιπτώσεις, η  $\pi Q = \pi$  γράφεται ως:

$$\sum_{x \in X} \pi(x) Q(x, y) = \pi(y), \tag{3.2.4}$$

όταν ο  $\mathcal{X}$  είναι διακριτός (τα  $(Q(x, y))_{x, y \in \mathcal{X}}$  αντιπροσωπεύουν τον πίνακα πιθανοτήτων μετάβασης 1ης τάξης όταν έχουμε μ.α.), και

$$\int_{\mathcal{X}} \pi(dx)Q(x, dy) = \pi(dy), \quad \text{για οποιοδήποτε } y \in \mathcal{X}, \quad (3.2.5)$$

για γενικό χώρο  $\mathcal{X}$ . Όταν  $\mathcal{X} \subset \mathbb{R}^d$  και υπάρχει πυκνότητα (πάλι συμβολίζεται με  $\pi$ ) τότε η τελευταία εξίσωση γράφεται:

$$\int_{\mathcal{X}} Q(x, A)\pi(x)dx = \pi(A).$$

Η συνάρτηση  $\pi$  λέγεται και στάσιμη πυκνότητα (stationary or invariant density) ή συνάρτηση ισορροπίας.

2. Επίσης, η εξίσωση 3.2.3 λέγεται και εξίσωση ολικής ισορροπίας (Balance equation) και λέει ότι:  $\mathbb{P}_{\pi}(X_1 \in A) = \pi(A)$ . Οπότε, για μία μ.α., χρησιμοποιώντας τη μακροβιανή ιδιότητα, έπεται:  $\mathbb{P}_{\pi}(X_n \in A) = \pi(A)$  για κάθε  $n \geq 1$ .

Πράγματι,

$$\begin{aligned} \mathbb{P}_{\pi}(X_n \in A) &= \int \pi(ds_0) \int Q(s_0, ds_1) \dots \int Q(s_{n-2}, ds_{n-1})Q(s_{n-1}, A) \\ &= \int \int \dots \left( \int Q(s_0, ds_1)\pi(ds_0) \right) Q(s_1, ds_2) \dots Q(s_{n-2}, ds_{n-1})Q(s_{n-1}, A) \\ &= \int \int \dots \left( \int Q(s_1, ds_2)\pi(ds_1) \right) \dots Q(s_{n-2}, ds_{n-1})Q(s_{n-1}, A) \\ &= \dots = \int \pi(ds_{n-1})Q(s_{n-1}, A) = \pi(A). \end{aligned}$$

Στις παραπάνω ισότητες εφαρμόζουμε διαδοχικά την (3.2.5).

3. Οι εξισώσεις Αν το  $\pi$  δεν είναι μέτρο πιθανότητας αλλά κάποιο πεπερασμένο μέτρο, τότε διαιρώντας με  $\pi(X)$  παίρνουμε μία στάσιμη κατανομή.
- 3 Με απλά λόγια, για την περίπτωση αριθμήσιμου χ.κ., η κατανομή  $(\mu)_{x \in \mathcal{X}}$  λέγεται στάσιμη για την μ.α.  $X_n$  αν ισχύει ότι: αν η  $X_0$  έχει κατανομή  $\pi$  τότε και η  $X_n$  έχει κατανομή  $\pi$ , για κάθε  $n \geq 1$ .

Επίσης, μία ενδιαφέρουσα χαρακτηριστική ιδιότητα είναι ότι μπορεί μία αδιαχώριστη μ.α. να γίνει στάσιμη, υπολογίζοντας τη στάσιμη κατανομή και στη συνέχεια, θεωρώντας το ως αρχική κατανομή. Όπως φαίνεται από την παραπάνω, στην περίπτωση αριθμήσιμου χ.κ., η παραπάνω εξίσωση ανάγεται σε ένα μη ομογενές γραμμικό σύστημα (ειδικής μορφής:  $Ax = x$ , όπου  $A = Q^t$ ,  $x = \pi^t$ ) με περισσότερες εξισώσεις από αγνώστους (καθώς απαιτούμε κατανομή: δηλαδή το  $x$  να αθροίζει στη μονάδα). Αρκετές φορές, ο υπολογισμός αυτός είναι πολύ απλός (επίλυση ενός απλού γραμμικού συστήματος που μπορεί να γίνει και με το χέρι).

### 3.2.2 Βασικά χαρακτηριστικά-έννοιες Μ.Α.

Πριν πάμε σε ζητήματα ύπαρξης στάσιμης κατανομής και γενικότερα ζητήματα σύγκλισης θα χρειαστούμε τις έννοιες της επαναληπτικότητας, της επικοινωνίας, της διαχωρισιμότητας, και της απεριοδικότητας μιας μ.α. ή πιο σωστά, τα ανάλογα των εννοιών αυτών στην περίπτωση γενικού χώρου, καθώς οι έννοιες αυτές έχουν ορισθεί για αριθμήσιμο χώρο. Σε αυτή την ενότητα παρουσιάζονται οι έννοιες αυτές όπως επίσης και η έννοια της αντιστρεψιμότητας. Προκειμένου να γίνει αυτό, θα χρειαστούμε τα παρακάτω βασικά εργαλεία:

Για ένα σύνολο  $A \in \mathcal{X}$ , ορίζεται ο χρόνος πρώτης εισόδου  $\sigma_A$  και ο χρόνος επανόδου  $\tau_A$  ως:

$$\sigma_A = \inf\{n \geq 0 : X_n \in A\} \quad (3.2.6)$$

$$\tau_A = \inf\{n \geq 1 : X_n \in A\}, \quad (3.2.7)$$

με τη σύμβαση  $\inf \emptyset = +\infty$ . Θέτοντας  $\tau_A^0 = 0$  και για  $k \geq 1$  ορίζεται ο  $k$ -οστός χρόνος επανόδου:

$$\tau_A^k = \inf\{n > \tau_A^{k-1} : X_n \in A\}$$

Προφανώς,  $\tau_A^1 = \tau_A$ .

Τέλος, ορίζεται ο χρόνος παραμονής (ή συνολικό πλήθος μεταβάσεων) στο  $A$ :

$$\eta_A = \sum_{k=1}^{\infty} \mathbb{1}_{X_k \in A}.$$

Σε όσα ακολουθούν, αλλά και γενικότερα στην ανάπτυξη της θεωρίας, οι ποσότητες  $\mathbb{P}_x(\tau_A < \infty)$ ,  $E_x[\eta_A]$  έχουν σημαντική θέση και αντιπροσωπεύουν το μέσο πλήθος μεταβάσεων στο  $A$  και την πιθανότητα επιστροφής στο  $A$  σε πεπερασμένο πλήθος βημάτων. Οι παραπάνω ορισμοί καλύπτουν και την περίπτωση αριθμήσιμων χώρων επιλέγοντας ως  $A$  κάποιο μονοσύνολο  $\{y\}$ .

Στην περίπτωση αυτή, λέμε επίσης, ότι δύο καταστάσεις,  $x, y \in \mathcal{X}$ , όχι απαραίτητα διαφορετικές, επικοινωνούν, αν υπάρχουν  $n_1, n_2 \geq 0$  ώστε  $Q^{n_1}(x, y) > 0$  και  $Q^{n_2}(y, x) > 0$  ή ισοδύναμα, αν  $\mathbb{P}_x(\tau_y < \infty) > 0$ . (όπου  $Q^0(x, \cdot) = \delta_x$ , για κάθε  $x \in \mathcal{X}$ ). Τότε, γράφουμε  $x \leftrightarrow y$ . Η σχέση επικοινωνίας αποδεικνύεται πως είναι μία σχέση ισοδυναμίας.

Τέλος, ορίζεται το σύνολο (ή κλάση) επικοινωνίας  $C \subset \mathcal{X}$  ως το σύνολο με την ιδιότητα:

- (i)  $x \in C, y \in C \implies x \leftrightarrow y$
- (ii)  $x \in C, x \leftrightarrow y \implies y \in C$

Δηλαδή το  $C$  είναι το μέγιστο σύνολο (από ιδιότητα (ii)) καταστάσεων που επικοινωνούν (δηλαδή δεν μπορεί να υπάρξει γνήσιο υποσύνολο μιας κλάσης επικοινωνίας που να είναι κλάση επικοινωνίας). Μπορούμε να παρατηρήσουμε πως επιτρέπονται οι μεταβάσεις από κάποια κατάσταση  $i$  στο  $C$  σε κάποια κατάσταση  $j$  έξω από το  $C$ , δηλαδή

μπορούμε να έχουμε  $i \in C, j \notin C, i \rightarrow j$ . Ωστόσο, στην περίπτωση αυτή,  $j \nrightarrow i$ . Επίσης, είναι δυνατόν να έχουμε  $i \notin C, j \in C, i \rightarrow j$ , αλλά πάλι θα είναι αναγκαστικά  $j \nrightarrow i$ . Αν στερήσουμε την πρώτη από τις 2 αυτές δυνατότητες, τότε έχουμε τον ακόλουθο ορισμό: Το  $C$  λέγεται κλειστό (για τον πυρήνα  $Q$ ), όταν η έξοδος από αυτό είναι αδύνατη, δηλαδή αν  $x \in C, y \notin C \implies Q^n(x, y) = 0$  για κάθε  $n$  ή ισοδύναμα, αν  $Q(x, C^c) = 0$  για κάθε  $x \in C$ .

Έτσι, μία μ.α. αν βρεθεί σε ένα κλειστό  $C$ , παγιδεύεται σε αυτό. Μπορεί επίσης ναδειχθεί ότι ο (αριθμήσιμος) χώρος καταστάσεων, διαμερίζεται μοναδικά ως εξής:

$$\mathcal{X} = \cup_{i \in I} C_i \cup T,$$

όπου το  $I$  είναι ένα αριθμήσιμο (συνήθως πεπερασμένο) σύνολο δεικτών, τα  $C_i$  είναι κλειστά σύνολα επικοινωνίας και το  $T$  είναι η ένωση όλων των άλλων (μη κλειστών) συνόλων επικοινωνίας). Τα σύνολα επικοινωνίας έχουν πολύ σημαντικό ρόλο στην ανάπτυξη της θεωρίας και το βασικό χαρακτηριστικό είναι πως τα μέλη ενός κλειστού συνόλου μοιράζονται τις ίδιες ιδιότητες όπως θα τονιστεί στη συνέχεια.

Για αλυσίδες σε γενικούς χώρους, το ανάλογο (ως κάποιο βαθμό) των συνόλων επικοινωνίας είναι τα προσβάσιμα (ή προσιτά) σύνολα. Ένα σύνολο  $A \in \mathcal{X}$  λέγεται προσιτό από το  $x$  (προσιτό για τον πυρήνα  $Q$ , αν  $\mathbb{P}_x(\tau_A < \infty) > 0$  (για κάθε  $x \in \mathcal{X}$ ). Πολλά αποτελέσματα σχετικά με τη δομή μιας Μ.Α. σε γενικούς χώρους (αλλά και για αριθμήσιμους χώρους), μπορούν να αναδειχθούν χρησιμοποιώντας σύνολα επικοινωνίας. (πχ στα:...).

Η κλασική έννοια της μη-διαχωρισιμότητας (irreducibility) είναι πως μία αλυσίδα έχει θετική πιθανότητα να επισκεφτεί, τελικά, οποιαδήποτε κατάσταση, από οποιαδήποτε κατάσταση. Αν ο χώρος καταστάσεων  $X$ , είναι υπεραριθμήσιμος, τότε αυτό είναι αδύνατον. Αντ' αυτού λοιπόν, απαιτούμε μία ασθενέστερη συνθήκη, αυτή της  $\phi$ -διαχωρισιμότητας ( $\phi$ -irreducible):

**Ορισμός 3.10** (Αδιαχώριστη μ.α.) Έστω μία μ.α. με τιμές στον  $\mathcal{X}$ . Η αλυσίδα λέγεται  $\phi$ -αδιαχώριστη, αν υπάρχει κάποιο  $\sigma$ -πεπερασμένο μέτρο  $\phi$  στον  $\mathcal{X}$ , τέτοιο ώστε για οποιοδήποτε  $A \in \mathcal{X}$  με  $\phi(A) > 0$ , και για κάθε  $x \in X$ , υπάρχει ένας θετικός ακέραιος  $n = n(x, A)$ , ώστε  $Q^n(x, A) > 0$  ή ισοδύναμα,  $\mathbb{P}_x(\tau_A < \infty) > 0$ . Ένα τέτοιο μέτρο λέγεται αδιαχώριστο (ή ανάγωγο) μέτρο για τον  $Q$ .

Ένας ισοδύναμος ορισμός στην περίπτωση μ.α. με αριθμήσιμο χ.κ. που συνδέεται με την έννοια του κλειστού συνόλου που ορίστηκε στα προηγούμενα είναι και ο εξής: Μία μ.α. είναι αδιαχώριστη αν όλες οι καταστάσεις επικοινωνούν (αν κανένα γνήσιο υποσύνολο του χ.κ. δεν είναι κλειστό). Τυπικά, αυτό μπορεί να εξακριβωθεί ελέγχοντας αν για κάθε  $x, y \in \mathcal{X}$  ισχύει  $\mathbb{P}_x(\tau_y < \infty) > 0$ .

### Παραδείγματα

1. Αν μία μ.α. είναι  $\delta_{x^*}$ -αδιαχώριστη, τότε το  $x^*$  είναι προσιτό από οποιαδήποτε κατάσταση. Τότε, αν μία αλυσίδα έχει μία κατάσταση που είναι προσιτή από κάθε άλλη, θα είναι  $\phi$ -αδιαχώριστη. Ωστόσο, αν ο  $\mathcal{X}$  είναι υπεραριθμήσιμος, τότε  $Q(x, y) = 0$  για κάθε  $x, y$  και επομένως το παραπάνω μέτρο δεν μας καλύπτει.

2. (GWP) Είναι εύκολο να διαπιστώσουμε πως η διαδικασία Galton-Watson δεν είναι αδιαχώριστη. Συγκεκριμένα, αν για κάποιο  $n$  προκύψει ότι  $Z_n = 0$ , τότε θα είναι  $Z_m = 0$  για κάθε  $m \geq n$ . Αυτό είναι άμεσο διότι η  $Z_n$  αντιπροσωπεύει τον ενεργό(εν ζωή) πληθυσμό στη  $n$ -οστή γενιά. Αν δεν υπάρχει κανένα άτομο σε αυτή, δεν θα υπάρχει και στις επόμενες. Οπότε η κατάσταση 0 είναι κλειστό σύνολο και μάλιστα και είναι και το μοναδικό (λέγεται και κατάσταση απορρόφησης). Έτσι, όλες οι υπόλοιπες καταστάσεις είναι μεταβατικές.
3. (Μοντέλο(AR(1))) Όταν  $X_{n+1} = \theta X_n + \epsilon_{n+1}$  και τα  $(\epsilon_n)_{n \geq 0}$  είναι ανεξάρτητες κεντροποιημένες κανονικές τ.μ., η αλυσίδα είναι αδιαχώριστη ως προς το μέτρο Lebesgue. Συγκεκριμένα, για οποιοδήποτε  $A$  τέτοιο ώστε  $\lambda(A) > 0$  και για κάθε  $x \in \mathbb{R}$  ισχύει  $Q(x, A) > 0$  (αφού  $Q(x, \cdot)$  η κατανομή της  $\mathcal{N}(\theta x, \sigma^2)$ ).
- Ωστόσο, αν  $\epsilon_n \sim U(-1, 1)$  και  $|\theta| > 1$ , τότε η αλυσίδα δεν είναι αδιαχώριστη:

Για παράδειγμα, όταν  $\theta > 1$ , θα είναι  $X_{n+1} - X_n = \theta X_n - X_n + \epsilon_{n+1} \geq X_n(\theta - 1) - 1 \geq 0$  όταν  $X_n \geq 1/(\theta - 1)$ , και τότε αλυσίδα είναι αύξουσα. Οπότε δεν μπορεί να επιστρέψει σε προηγούμενες καταστάσεις.

Συνεχίζουμε με την έννοια της επαναληπτικότητας για μία αλυσίδα σε αριθμήσιμο χ.κ.:

**Ορισμός 3.11** (Επαναληπτικότητα) Μία κατάσταση  $y$  λέγεται επαναληπτική αν  $\mathbb{P}_y(\tau_y < \infty) = 1$  ή αν  $E_x[\eta_x] = \infty$  ( $\implies \mathbb{P}_x(\eta_x = \infty) = 1$ , διαφορετικά λέγεται μεταβατική. Μία μ.α. λέγεται επαναληπτική, αν κάθε κατάσταση είναι επαναληπτική.

Αν η κατάσταση  $y$  είναι επαναληπτική, τότε  $\mathbb{P}_y(\tau_y^k < \infty) = 1$  για κάθε  $k$ , και άρα  $\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$ . Ακόμη, μία επαναληπτική κατάσταση  $y$ , λέγεται θετικά επαναληπτική αν:  $E_x[\tau_x < \infty] < \infty$ . Διαφορετικά λέγεται μηδενική.

Σύμφωνα λοιπόν με τον παραπάνω ορισμό, μία κατάσταση λέγεται επαναληπτική, αν αναχωρώντας από αυτή, (με πιθανότητα 1) η αλυσίδα θα επιστρέψει σε αυτή. Χρησιμοποιώντας την παρακάτω διαπίστωση, είναι άμεσο πως τότε θα επιστρέψει στην κατάσταση αυτή άπειρες φορές δικαιολογώντας έτσι το  $\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$

Η χρονική ομογένεια σε συνδυασμό με την ισχυρή μαρκοβιανή ιδιότητα (παράρτημα) μας οδηγούν στην εξής σημαντική διαπίστωση:

**Παρατήρηση** Με κάθε επίσκεψη στην κατάσταση  $y$ , η μ.α. μπορεί να θεωρηθεί ότι ανανεώνεται (δηλαδή ότι ξεκινά από την αρχή). Από το γεγονός αυτό γίνεται αντιληπτό οι μ.α. συνδέονται με τις ανανεωτικές διαδικασίες (cite).

Η επαναληπτικότητα συνδέεται φυσικά και με τη δομή του χ.κ. Για παράδειγμα, αν ο χ.κ. είναι πεπερασμένος και αδιαχώριστος τότε όλες οι καταστάσεις είναι επαναληπτικές. Μάλιστα, αποδεικνύεται πως σε αυτή την περίπτωση είναι όλες θετικά επαναληπτικές. Αν αφήσουμε την υπόθεση πως η μ.α. είναι αδιαχώριστη, τότε είναι διαισθητικά άμεσο πως είναι αδύνατο να είναι όλες οι καταστάσεις μεταβατικές.

Οι αποδείξεις για τους παραπάνω ισχυρισμούς βασίζονται στην έννοια της επικοινωνίας. Συγκεκριμένα, αφού η επικοινωνία, ως διμελής σχέση, είναι και σχέση ισο-

δυναμίας, έπεται πως αν 2 καταστάσεις επικοινωνούν τότε είναι του ίδιου τύπου (πχ και οι 2 επαναληπτικές κλπ.). Οπότε, τα κλειστά σύνολα (: όλες οι καταστάσεις που ανήκουν σε αυτά επικοινωνούν) θα έχουν πάντα μία επαναληπτική κατάσταση και μία μ.α. αν περιοριστεί σε κάποιο κλειστό σύνολο, μπορεί να θεωρηθεί αδιαχώριστη(αφού θα παραμείνει σε αυτό).

Επίσης, είναι άμεσο πως αν  $C$  ένα κλειστό σύνολο καταστάσεων, και κάποια κατάσταση στο  $C$  είναι προσιτή από μία άλλη κατάσταση  $y$ , κατάσταση που δεν βρίσκεται στο  $C$ , τότε η  $y$  είναι μεταβατική.

Με τα παραπάνω, το επόμενο αποτέλεσμα φαίνεται εντελώς φυσιολογικό:

**Πρόταση 3.1.** Μία αδιαχώριστη μ.α. είναι είτε επαναληπτική είτε μεταβατική.

Φυσικά το τελευταίο προϋποθέτει πως ο χ.κ. είναι άπειρος. Επίσης, η συνθήκη της θετικής επαναληπτικότητας συνδέεται άμεσα με την ύπαρξη στάσιμης κατανομής όπως φαίνεται στο παρακάτω Θεώρημα:

**Θεώρημα 3.3** Έστω μία αδιαχώριστη μ.α. με αριθμήσιμο χώρο καταστάσεων, τότε είναι θετικά επαναληπτική αν το σύστημα :

$$\sum_{x \in \mathcal{X}} \pi_x Q(x, y) = \pi_y$$

$$\sum_{x \in \mathcal{X}} \pi_x = 1.$$

έχει μοναδική λύση. Σε καταφατική περίπτωση, η λύση αυτή είναι μη αρνητική και είναι η στάσιμη κατανομή.

Τι γίνεται όμως στην περίπτωση γενικού χώρου καταστάσεων:

Φυσικά οι έννοιες αυτές δεν μπορούν να περάσουν άμεσα για την περίπτωση γενικού χώρου αφού η πιθανότητα η αλυσίδα να επισκεφτεί μία συγκεκριμένη κατάσταση είναι πάντα 0. Ωστόσο, οι έννοιες αυτές περνάνε στα σύνολα. Συγκεκριμένα έχουμε τον παρακάτω ορισμό:

**Ορισμός 3.12** Ένα σύνολο λέγεται επαναληπτικό αν  $E_x[\eta_A] = \infty$  για κάθε  $x \in A$ . Το σύνολο  $A$  λέγεται ομοιόμορφα μεταβατικό αν υπάρχει σταθερά  $M$  ώστε  $E_x[\eta_A] < M$  για κάθε  $x \in A$ . Το  $A$  λέγεται μεταβατικό αν υπάρχει κάλυψη του  $\mathcal{X}$  από ομοιόμορφα μεταβατικά σύνολα.

Η αλυσίδα λέγεται επαναληπτική αν είναι  $\phi$ -αδιαχώριστη και για κάθε  $A \in \mathcal{X}$  με  $\phi(A) > 0$ , ισχύει  $E_x[\eta_A] = \infty$  για κάθε  $x \in A$ . Τέλος, μία  $\phi$ -αδιαχώριστη μ.α. λέγεται θετική αν έχει κάποιο στάσιμο μέτρο, διαφορετικά λέγεται μηδενική.

Μία πιο ισχυρή έννοια επαναληπτικότητας είναι η Harris-επαναληπτικότητα. Η εισαγωγή της έννοιας αυτής έγινε από τον Harris (1956).

**Ορισμός 3.13** (Harris-επαναληπτικότητα). Ένα σύνολο  $A \in \mathcal{X}$  λέγεται Harris-επαναληπτικό αν  $\mathbb{P}_x(\tau_A < \infty) = 1$  για κάθε  $x \in X$ . Μία  $\phi$ -αδιαχώριστη Μ.Α. λέγεται Harris (επαναληπτική) αν κάθε προσιτό σύνολο είναι Harris-επαναληπτικό.

**Παράδειγμα** (Μοντέλο(AR(1))) Είναι  $Q(x_n, \cdot) \equiv \mathcal{N}(\theta x_n, \sigma^2)$  και μία κανονική κατανομή  $\mathcal{N}(\mu, \tau^2)$  είναι στάσιμη για το AR(1) μόνο στην περίπτωση που:

$$\mu = \theta\mu \text{ και } \tau^2 = \tau^2\theta^2 + \sigma^2.$$

Δηλαδή, όταν  $\mu = 0$  και  $\tau^2 = \sigma^2/(1 - \theta^2)$ . Η τελευταία, είναι λύση της πιο πάνω δευτεροβάθμιας εξίσωσης όταν  $|\theta| < 1$ . Στην περίπτωση αυτή, η αλυσίδα AR(1) έχει στάσιμη κατανομή την  $\mathcal{N}(0, \sigma^2/(1 - \theta^2))$ .

Μένει να περιγράψουμε και την έννοια της απериοδικότητας μιας αλυσίδας. Συγκεκρι-

κριμένα, ο πιο απλό ορισμός για την απериοδικότητα μιας μ.α. σε αριθμησιμο χώρο καταστάσεων είναι πως μία κατάσταση  $x$  είναι απериοδική, αν  $Q^n(x, x) > 0$  για οσοδήποτε μεγάλο  $n$ , ή ισοδύναμα, η περίοδος μιας κατάστασης είναι ίση με τη μονάδα. Η περίοδος της κατάστασης  $x$  ορίζεται ως μέγιστος κοινός διαιρέτης του συνόλου:  $\{n > 0 : Q^n(x, x) > 0\}$ . Προφανώς ο παραπάνω ορισμός δεν μπορεί να περάσει στη μορφή αυτή, σε γενικούς χώρους καταστάσεων. Το ανάλογο είναι το εξής:

**Ορισμός 3.14** (Απериοδικότητα). Μία  $\phi$ -αδιαχώριστη Μ.Α. λέγεται απериοδική, αν υπάρχει διαμέριση:  $\mathcal{X} = A_0 \cup \dots \cup A_{n-1} \cup A_n$  σε  $n + 1 \geq 3$  ξένα σύνολα  $A_0, \dots, A_n$  τέτοια ώστε  $\phi(A_n) > 0$  και:

$$\begin{aligned} x \in A_0 &\implies Q(x, A_1) = 1, x \in A_1 \implies Q(x, A_2) = 1, \dots, \\ x \in A_{n-2} &\implies Q(x, A_{n-1}) = 1, x \in A_{n-1} \implies Q(x, A_0) = 1 \end{aligned}$$

Διαφορετικά, η αλυσίδα λέγεται απериοδική.

Κλείνουμε την ενότητα αυτή με μία ακόμα βασική έννοια των Μ.Α., αυτή της αντιστρεψιμότητας:

**Ορισμός 3.15** Μία Μ.Α. με τιμές στον  $X$ , λέγεται αντιστρέψιμη ως προς τη  $\sigma$  κατανομή  $\pi$  στον  $X$  αν:

$$\pi(dx)Q(x, dy) = \pi(dy)Q(y, dx), \text{ για κάθε } x, y \in X \quad (3.2.8)$$

Μία πολύ βασική ιδιότητα των αντιστρέψιμων αλυσίδων είναι:

$$\int_{x \in X} \pi(dx)Q(x, dy) = \int_{x \in X} \pi(dy)Q(y, dx) = \pi(dy) \int_{x \in X} Q(y, dx) = \pi(dy).$$

**Ορισμός 3.16.** Μία μ.α. με πυρήνα  $Q$  και τιμές στον  $X$  ικανοποιεί τις εξισώσεις λεπτομερούς ισορροπίας (Detailed Balance equations) αν υπάρχει κάποια συνάρτηση  $f$  τέτοια ώστε, για κάθε  $x, y \in X$

$$f(x)Q(x, y) = f(y)Q(y, x). \quad (3.2.9)$$

**Θεώρημα 3.4** Έστω μία μ.α. με πυρήνα  $Q$  και τιμές στον  $\mathcal{X}$  που ικανοποιεί τις εξισώσεις λεπτομερούς ισορροπίας με την πυκνότητα  $f$ . Τότε:

(i) Η πυκνότητα  $f$  είναι στάσιμη πυκνότητα. (ii) Η αλυσίδα είναι αντιστρέψιμη.

### Απόδειξη

Για το (i): αν  $A \in \mathcal{X}$  τότε

$$\begin{aligned} \int_X Q(y, B)f(y)dy &= \int_X \int_B Q(y, x)f(y)dydx \\ &= \int_X \int_B Q(x, y)f(x)dx dy = \int_B f(x)dx, \end{aligned}$$

αφού  $\int_X P(x, dy) = 1$ .

Το (ii) είναι άμεσο, καθώς αν η  $f$  είναι πυκνότητα, οι εξισώσεις ισορροπίας και η αντιστρέψιμότητα εκφράζουν την ίδια ακριβώς ιδιότητα.  $\square$

**Παράδειγμα** Ας υποθέσουμε ένα δίκτυο στο οποίο η ακμή  $Q(x, y)$  εκφράζει την ποσοστό προϊόντων που μεταφέρονται από την κορυφή (πχ κάποια πόλη)  $x$  στη  $y$ . Αν συμβολίσουμε με  $f(x)$  το συνολικό διαθέσιμο προϊόν της  $x$ , τότε το  $f(x)Q(x, y)$  εκφράζει το συνολικό ποσοστό προϊόντων που εξάγει η  $x$  στην  $y$ . Τότε η εξίσωση:

$$f(x)Q(x, y) = f(y)Q(y, x), \quad \text{για κάθε } x, \quad (3.2.10)$$

λέει όλες οι εξαγωγές είναι ίσες με τις εισαγωγές. Δηλαδή ότι υπάρχει τέλεια ισορροπία στην αγορά. Επίσης, αυτό σημαίνει πως αν ολοκληρωθεί ένας κύκλος εισαγωγών-εξαγωγών θα υπάρχει ακριβώς το ίδιο συνολικό διαθέσιμο προϊόν σε κάθε πόλη. Αυτό σημαίνει πως η  $f$  αν τη δούμε ως κατανομή, είναι στάσιμη.

### 3.2.3 Οριακά Θεωρήματα

Κάποια βασικά ερωτήματα που θα μπορούσε να θέσει κάποιος σχετικά με την οριακή συμπεριφορά μιας μ.α. είναι τα ακόλουθα:

- (1) Η απεικόνιση:  $n \rightarrow X_n$  συγκλίνει καθώς  $n \rightarrow \infty$ .
- (2) Η κατανομή του  $X_n$  συγκλίνει σε κάποια άλλη κατανομή κάποια, ας πούμε  $X_\infty$  καθώς  $n \rightarrow \infty$

Το κλειδί για τις απαντήσεις είναι οι έννοιες που περιγράψαμε στην προηγούμενη ενότητα ενώ το βασικό εργαλείο είναι ο νόμος των μεγάλων αριθμών.

**Παράδειγμα** (Διαδικασία GW) Θεωρούμε την  $(Z_n)_{n \geq 0}$  όπως ορίστηκε στα προηγούμενα, με αρχική κατανομή:  $\mathbb{P}(Z_0 = 1)$ . Προκύπτει ότι:

Αν  $\mathbb{P}(\xi_n^{(r)} = 0) = p_0 = 0$  και  $p_1 < 1$ , τότε  $Z_n \rightarrow \infty$  με πιθανότητα 1. Επίσης, αν  $p_0 > 0$ , τότε  $\mathbb{P}_1(\lim Z_n = +\infty) + \mathbb{P}_1(\lim Z_n = 0) = 1$ .

Ξεκινάμε με ένα κλασικό αποτέλεσμα για αδιαχώριστες μ.α. (Feller 1970) με αριθμησιμο χώρο καταστάσεων:



**Θεώρημα 3.5** Έστω μία αδιαχώριστη και επαναληπτική μ.α. σε αριθμήσιμο χώρο  $\mathcal{X}$ . Τότε υπάρχει μοναδικό (μέχρι μία πολλαπλασιαστική σταθερά) στάσιμο μέτρο  $\pi$ . Επιπλέον,  $0 < \pi(x) < \infty$  για κάθε  $x \in \mathcal{X}$ . Ακόμη, το μέτρο είναι αθροίσιμο ( $\sum \pi(x) < \infty$ ) αν η μ.α. είναι θετικά επαναληπτική, δηλαδή υπάρχει  $x$  τέτοιο ώστε:

$$E_x[\tau_x] < \infty$$

Στην περίπτωση αυτή, η στάσιμη κατανομή, όταν υπάρχει δίνεται από την:

$$\pi_x = (E_x[\tau_x])^{-1}, \quad x \in X.$$

Από την παραπάνω, η  $E_x[\tau_x]$  μπορεί να ερμηνευθεί ως το μακροπρόθεσμο ποσοστό του χρόνου που η αλυσίδα ξοδεύει στην κατάσταση  $x$ .

Όταν επιπλέον η κατάσταση  $x$  είναι απεριοδική, τότε:

$$\lim_{n \rightarrow \infty} Q^n(x, x) = (E_x[\tau_x])^{-1}$$

ενώ αν η  $x$  έχει περίοδο  $d_x$ , τότε:

$$\lim_{n \rightarrow \infty} Q^n(x, x) = d_x (E_x[\tau_x])^{-1}$$

Για αδιαχώριστες, θετικά επαναληπτικές και απεριοδικές μ.α. αριθμήσιμου χώρου καταστάσεων, το  $\lim_{n \rightarrow \infty} Q^n(x, y) = \pi_x$  μπορεί να ερμηνευθεί ως:

- (i) Η οριακή πιθανότητα η αλυσίδα να βρεθεί στην κατάσταση  $x$ .
- (ii) Το μακροπρόθεσμο ποσοστό του χρόνου που η αλυσίδα ξοδεύει στην κατάσταση  $x$ .

Για περιοδικές μ.α., η δεύτερη ερμηνεία εξακολουθεί να ισχύει ενώ η πρώτη όχι. Ωστόσο, η  $\pi$  είναι στάσιμη κατανομή ασχέτως της περιόδου. Δηλαδή, αν η  $X_0$  έχει κατανομή την  $\pi$  τότε και η  $X_n$  έχει κατανομή την  $\pi$  για κάθε  $n$ .

Για τη γενική περίπτωση, θα χρειαστούμε τον ορισμό του atom.

Ως (atom), καλούμε ένα σύνολο το οποίο συμβολίζεται με  $a$  για το οποίο υπάρχει μέτρο πιθανότητας στον  $(\mathcal{X}, \mathcal{X})$ , έστω  $\nu$ , τέτοιο ώστε  $Q(a, A) = \nu(A)$  για κάθε  $x \in a$  και για κάθε  $A \in \mathcal{X}$ .

Τα σύνολα αυτά, συμπεριφέρονται όπως και οι καταστάσεις για μ.α. σε αριθμήσιμο χώρο. Αν και το  $a$  είναι atom για τον  $Q$  τότε για οποιοδήποτε  $n \geq 1$  είναι atom και για τον  $Q^n$ . Αυτό έχει ιδιαίτερη σημασία: αν μία αλυσίδα ξεκινήσει από κατάσταση μέσα σε κάποιο  $a$ , τότε η κατανομή όλης της αλυσίδας δεν εξαρτάται από το αρχικό σημείο. Ο συμβολισμός, επομένως,  $E_a$ , στην περίπτωση αυτή, έχει νόημα, καθώς ο πυρήνας είναι ίδιος για οποιοδήποτε  $x \in a$ .

Το ανάλογο λοιπόν του προηγούμενου αποτελέσματος είναι:

**Θεώρημα 3.6** Έστω μία  $\phi$ -αδιαχώριστη μ.α. για την οποία υπάρχει atom, έστω  $a$ . Η μ.α. λέγεται θετική αν  $(E_a[\tau_a])^{-1} < \infty$ . Τότε η στάσιμη κατανομή ικανοποιεί την:

$$\pi(a) = (E_a[\tau_a])^{-1} \quad (3.2.11)$$

Ολοκληρώνοντας την περίπτωση αριθμησιμου χώρου, το βασικό Θεώρημα σύγκλισης είναι το ακόλουθο:

**Θεώρημα 3.7** Σύγκλισης της Μ.Α. Έστω Μ.Α. σε αριθμησιμο χώρο καταστάσεων.

1. Αν είναι αδιαχώριστη και θετικά επαναληπτική με περίοδο  $d > 0$ , τότε για οποιαδήποτε  $x, y \in \mathcal{X}$ :

$$(i). \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Q^k(x, y) = \pi_y,$$

όπου  $\pi$  η στάσιμη κατανομή (η οποία υπάρχει, είναι μη μηδενική και μοναδική). Το παραπάνω όριο σε μορφή πινάκων είναι:

$$\lim_{n \rightarrow \infty} \frac{1}{n} (I + Q + \dots + Q^{n-1}) = \Pi,$$

$$(ii). \lim_{n \rightarrow \infty} Q^{nd+k}(x, y) = d\pi_y s_{x,y}(k), \quad k = 0, 1, \dots, d-1,$$

όπου  $s_{x,y} = \mathbb{P}_x(\tau_y(\text{mod } d) = k)$ .

Ειδικότερα, αν είναι απεριοδική ( $d = 1$ ) τότε:

$$\lim_{n \rightarrow \infty} Q^n(x, y) = \pi_y$$

2. Αν είναι αδιαχώριστη και μεταβατική, τότε τα παραπάνω δύο όρια υπάρχουν και είναι ίσα με 0 (με πιθανότητα 1) για οποιαδήποτε  $x, y \in \mathcal{X}$ . Επίσης ισχύει πως για οποιοδήποτε,  $A$ , πεπερασμένο υποσύνολο του  $\mathcal{X}$ :

$$\sum_{n=0}^{\infty} \mathbb{P}_x(X_n \in A) < \infty$$

και έτσι, (από Borel-Cantelli),  $\mathbb{P}_x(X_n \in A \text{ i.o.}) = 0$ .

3. Αν είναι αδιαχώριστη και μηδενικά επαναληπτική, τότε τα παραπάνω δύο όρια υπάρχουν και είναι ίσα με 0 (με πιθανότητα 1) για οποιαδήποτε  $x, y \in \mathcal{X}$ . Ωστόσο, στην περίπτωση αυτή, για οποιοδήποτε,  $A$ , πεπερασμένο υποσύνολο του  $\mathcal{X}$ :

$$\sum_{n=0}^{\infty} \mathbb{P}_x(X_n \in A) = \infty$$

Στην περίπτωση γενικού χώρου καταστάσεων, το πρώτο βασικό αποτέλεσμα είναι το Κεντρικό Θεώρημα Σύγκλισης μ.α.:

**Θεώρημα 3.8** (Κεντρικό Θεώρημα Σύγκλισης μαρκοβιανής αλυσίδας). Για μία φ-αδιαχώριστη και απεριοδική αλυσίδα, όπου  $\pi$  η στάσιμη κατανομή της, υπάρχει κάποιο  $C \subset X$ , τέτοιο ώστε  $\Pi(C) = 1$  και για κάθε  $x \in C$  και  $A \subset X$ ,

$$\mathbb{P}_x(X_n \in A) \rightarrow \Pi(A) \text{ καθώς } n \rightarrow \infty. \quad (3.2.12)$$

Αν επιπλέον είναι Harris-επαναληπτική, τότε μπορούμε να πάρουμε  $C = X$ . και άρα αν  $\mu$  η αρχική κατανομή, τότε:

$$\mathbb{P}_\mu(X_n \in A) \rightarrow \Pi(A) \text{ καθώς } n \rightarrow \infty. \quad (3.2.13)$$

Η παραπάνω σύγκλιση, όπως έχει αποδειχθεί σε πιο πρόσφατες εργασίες Athreya et al. (2012) [4] είναι και ως προς την Total variation distance  $\|\cdot\|_{TV}$ .

Διαισθητικά, το παραπάνω Θεώρημα (όπως και το ανάλογο για αριθμήσιμο χ.κ.) λέει πως σε μία μ.α., (ξεκινώντας από οπουδήποτε), η επιρροή της αρχικής θέσης 'χάνεται' καθώς  $n \rightarrow \infty$  και έτσι η οριακή κατανομή της αλυσίδας θα είναι κατά προσέγγιση στάσιμη.

Παρόλο που δεν παρέχει κάποια πληροφορία για την ταχύτητα σύγκλισης στη στάσιμη κατανομή, οι συνθήκες είναι οι ελάχιστες δυνατές. Συγκεκριμένα οι συνθήκες : Harris-επαναληπτική και απεριοδικότητα, είναι και αναγκαίες ώστε η σύγκλιση της αλυσίδας να ισχύει για οποιαδήποτε αρχική κατάσταση  $x \in X$ . (Nummelin (1986) [24]).

Επίσης, από τον χαρακτηρισμό της σύγκλισης κατά κατανομή, από το παραπάνω Θεώρημα, μπορούμε να συνάγουμε την:

$$|E_\mu[f(X_n)] - E_\pi[f(X)]| \rightarrow 0 \text{ καθώς } n \rightarrow \infty.$$

για οποιαδήποτε φραγμένη  $f$ .

Η μορφή αυτή του θεωρήματος είναι ιδιαίτερα χρήσιμη όταν η στάσιμη κατανομή είναι γνωστή, όπως στις περιπτώσεις των μεθόδων MCMC που θα παρουσιαστούν στη συνέχεια και παρέχει συνθήκες εύκολα επαληθεύσιμες και κάτω από τις οποίες ένας MCMC αλγόριθμος συγκλίνει στη στάσιμη κατανομή.

Ειδικά στους αλγορίθμους αυτούς, πολύ χρήσιμο οριακό αποτέλεσμα είναι η σύγκλιση του λεγόμενου εργοδικού μέσου:

**Θεώρημα 3.9** (Σύγκλιση Εργοδικού μέσου για μ.α.). Έστω μία π-αδιαχώριστη Μ.Α. με στάσιμη κατανομή το  $\pi$ . Έστω επίσης:  $f : \mathcal{X} \rightarrow \mathbb{R}$  μία συνάρτηση για την οποία υπάρχει η μέση τιμή  $\theta = \int f(x)\pi(x)dx$ . Για οποιοδήποτε  $m \geq 0$ , ορίζεται ο εργοδικός μέσος ως:

$$\hat{\theta}_N = \frac{1}{N+1} \sum_{i=m}^{m+N} f(X_i).$$

Τότε, υπάρχει ένα σύνολο  $A \subset \mathcal{X}$  με  $\pi(A) = 1$  για κάθε  $x \in A$  και:

$$\mathbb{P}_x(\lim_{n \rightarrow \infty} \hat{\theta}_N = \theta) = 1$$

Επιπλέον, αν είναι και Harris-επαναληπτική, τότε μπορούμε να πάρουμε  $A = \mathcal{X}$

Η εκτιμήτρια  $\hat{\theta}_N$  είναι λόγω της διαχωρισιμότητας, συνεπής για κάθε αρχική κατάσταση στο  $A$  ενώ η συνθήκη της Harris-επαναληπτικότητας, εξασφαλίζει τη συνέπεια για οποιαδήποτε κατάσταση. Τέλος, για το παραπάνω αποτέλεσμα παρατηρούμε πως δεν απαιτείται η απεριοδικότητα της αλυσίδας.

Ένα εύλογο ερώτημα είναι αν η σταθερά  $m$  παίζει κάποιο ρόλο στο όριο. Στην πράξη, το  $m$  επιλέγεται ως η Burn-in / Warm up περίοδος. Είναι ο χρόνος που θεωρούμε πως χρειάζεται μέχρι η αλυσίδα να φτάσει σε ισορροπία. Επομένως το  $m$  θα πρέπει να είναι αρκετά μεγάλο ώστε η κατανομή του  $X_m$  να είναι κοντά στο  $\pi$ .

### 3.3 Κρυμμένα Μαρκοβιανά μοντέλα (ΚΜΜ)

#### 3.3.1 Εισαγωγή

Αν και θεμελιωτής της μαρκοβιανής θεωρίας υπήρξε ο Andrei Markov στις αρχές του 20-ού αιώνα, η θεωρία των κρυμμένων μαρκοβιανών μοντέλων (ΚΜΜ) μελετήθηκε κυρίως από τον Leonard Baum (στα τέλη της δεκαετίας του 60), ο οποίος ανέπτυξε σημαντικό μέρος της θεωρίας μαζί με τους Eagon, Petrie, Soules και Weiss (Baum and Petrie 1966, Baum and Eagon 1967, Baum, et al. 1970, Baum 1972). Το όνομα ωστόσο αποδίδεται στον L.P. Neuwirth.

Ένα ΚΜΜ είναι στην ουσία μία μαρκοβιανή διαδικασία χωρισμένη σε 2 μέρη. Ένα παρατηρήσιμο και ένα μη παρατηρήσιμο ή κρυμμένο που εξηγεί και την ονομασία. Είναι λοιπόν μία διαδικασία  $(X_n, Y_n)_{n \geq 0}$  (που είναι μαρκοβιανή) για την οποία υποθέτουμε πως μπορούμε να παρατηρήσουμε την  $(Y_n)_{n \geq 0}$ , στην οποία αναφερόμαστε ως διαδικασία παρατηρήσεων, αλλά όχι τη  $(X_n)_{n \geq 0}$ , η οποία καλείται διαδικασία καταστάσεων και είναι ωστόσο έμμεσα παρατηρήσιμη μέσω της  $(Y_n)_{n \geq 0}$ .

Η θεωρία των ΚΜΜ είναι εξαιρετικά πλούσια ως προς τα εργαλεία που παρέχει για την επίλυση ρεαλιστικών σεναρίων αλλά ταυτόχρονα αποτελεί ένα αρκετά σύνθετο μαθηματικό αντικείμενο κυρίως λόγω της θεωρίας που απαιτεί. Στον επιστημονικό κλάδο γενικότερα, τα ΚΜΜ ξεχωρίζουν για το πολύ μεγάλο εύρος πραγματικών προβλημάτων που μπορούν να καλύψουν μέσα σε διαφορετικές μεταξύ τους περιοχές. Τέτοιες εφαρμογές μπορεί να συναντήσει κανείς :

- (1) Στην επεξεργασία σημάτων (Signal Processing), κυρίως στις Τηλεπικοινωνίες και στην αυτόματη αναγνώριση ομιλίας.

Για παράδειγμα, η  $X_n$  θα μπορούσε να αντιπροσωπεύει κάποιο σήμα που πρόκειται να μεταδοθεί μέσω κάποιου συστήματος τηλεπικοινωνίας. Το σύστημα αυτό ενδεχομένως να αλλοιώσει το σήμα και έτσι ο αποδέκτης να λάβει μία αλλοιωμένη έκδοση  $(Y_n)$  του αρχικού σήματος. Το πρόβλημα λοιπόν θα είναι να ανακατασκευάσει το σήμα αυτό προκειμένου να μοιάζει περισσότερο στο αρχικό.

- (2) Στα Χρηματοοικονομικά.

Για παράδειγμα, ένα μοντέλο που μας επιτρέπει να χειριστούμε χρηματοοικονομικά δεδομένα (δηλαδή μία ακολουθία οικονομικών τιμών πχ μετοχών σε ένα δεδομένο χρονικό ορίζοντα) που διακρίνεται για την απλή δομή του, είναι το μοντέλο Black-Scholes για χρονοσειρές:

$$S_n = \exp\left\{\mu - \frac{\sigma^2}{2} + \sigma \xi_n\right\} S_{n-1},$$

όπου  $\xi_n \sim \mathcal{N}(0, 1)$  α.ι.τ.μ. και το  $\sigma \in \mathbb{R}$  εκφράζει πόσο πολύ ή πόσο γρήγορα αλλάζει η τιμή μιας επένδυσης, ή γενικότερα πως μεταβάλλεται η αγορά. Το  $\mu \in \mathbb{R}$  είναι ο ρυθμός επιστροφής (δηλαδή το κέρδος μιας επένδυσης σε έναν δεδομένο χρονικό ορίζοντα). Το τελευταίο φαίνεται και από την  $E[S_n/S_{n-1}] = e^\mu$ . Ένα μοντέλο όπως και το παραπάνω μπορεί να είναι αποδοτικό όσο οι τιμές των μετοχών διατηρούν έναν μαρκοβιανό χαρακτήρα, δηλαδή για μικρές χρονικές περιόδους. Αυτό βέβαια μπορεί να το περιμένει κανείς αφού οι παράμετροι εκφράζουν

## Κρυμμένα Μαρκοβιανά μοντέλα (KMM)

ποσότητες που δεν μπορούν να είναι σταθερές για μεγάλα χρονικά διαστήματα. Μπορούμε ωστόσο να επιτρέψουμε στις παραμέτρους αυτές να μεταβάλλονται. Η ιδέα είναι να εισάγουμε μία Μαρκοβιανή διαδικασία  $X_n$  (ανεξάρτητη των  $\xi_n$ ) και να θέσουμε:

$$S_n = \exp\left\{\mu(X_n) - \frac{\sigma(X_n)^2}{2} + \sigma(X_n)\xi_n\right\}S_{n-1},$$

όπου τώρα τα  $\mu, \sigma$  είναι κατάλληλες συναρτήσεις. Αν επιλέξουμε ως παρατηρήσιμη διαδικασία την  $Y_n = \log(S_n/S_{n-1})$ , τότε η ακολουθία  $(X_n, Y_n)_{n \geq 0}$  αποτελεί ένα KMM. Ρυθμίζοντας κατάλληλα τη δυναμική της  $X_k$ , μπορούμε να αποκτήσουμε ένα μοντέλο, πολύ πιο ρεαλιστικό από το Black-Scholes. Μία επιλογή για τη δυναμική της  $(X_n)$ , είναι να έχει θέσουμε:

$$X_n = \alpha(X_{n-1}) + \beta(X_{n-1})\eta_n,$$

όπου οι  $\eta_n$  α.ι.τ.μ. (αντιπροσωπεύοντας τα σφάλματα του μοντέλου). Ειδικά το παραπάνω, είναι ένα αρκετά κοινό σενάριο στη βιβλιογραφία. Τυπικά, μόνο οι τιμές των μετοχών είναι παρατηρήσιμες στους επενδυτές. Ακόμα και αν ο οικονομικός παράγοντας ( $X_n$ ) έχει κάποια πραγματική σημασία (αντί να είναι απλά ένα βολικό μαθηματικό εργαλείο), τέτοιοι οικονομικοί παράγοντες τυπικά δεν είναι γνωστοί στους επενδυτές. Επομένως, οποιαδήποτε στατιστική συμπερασματολογία, αποφάσεις, εκτιμήσεις κτλ, θα πρέπει να βασίζονται στην ακολουθία παρατηρήσεων ( $Y_n$ ).

(3) (3) Στις βιοεπιστήμες (Bioinformatics ).

Ένα παράδειγμα είναι η μοντελοποίηση της ανάπτυξης των φυτών και των δένδρων. Συγκεκριμένα, τα μοντέλα GreenLab και LNAS, που παρουσιάστηκαν στο κεφάλαιο 2, αποτελούν ένα βασικό παράδειγμα κατασκευής KMM.

Άλλες περιοχές με πολύ σημαντικές εφαρμογές των KMM είναι: (4) Αναγνώριση μορφών (Pattern recognition) (πχ τρόπος γραφής) και (5) Βιοφυσική κ.α. Τέλος, τα KMM εφαρμόζονται και σε προβλήματα Δυναμικού προγραμματισμού (Μαρκοβιανές Διαδικασίες Αποφάσεων).

Έτσι λοιπόν, τα KMM προσφέρουν ένα πλαίσιο μελέτης πραγματικών προβλημάτων με πολύ υψηλό επίπεδο ρεαλισμού. Ο στόχος λοιπόν της θεωρίας των KMM είναι να παρέχει τα απαραίτητα εργαλεία ώστε να είναι εφικτή μία τόσο ρεαλιστική μοντελοποίηση.

### 3.3.2 Ορισμός-Βασικές έννοιες

Στις πιο απλές περιπτώσεις, όπως για παράδειγμα με αριθμήσιμο χώρο καταστάσεων για τις 2 διαδικασίες, ένα KMM μπορεί να οριστεί χρησιμοποιώντας την έννοια της δεσμευμένης ανεξαρτησίας (παράρτημα). Συγκεκριμένα:

Λέμε πως η ακολουθία  $(X_n, Y_n)_{n \geq 0}$  είναι KMM, αν:

1. Η  $(X_n)_{n \geq 0}$  είναι μία μη παρατηρήσιμη Μ.Α. με αριθμήσιμο χώρο καταστάσεων, αρχική κατανομή  $\mu$  και πυρήνα  $Q$ .

## Κρυμμένα Μαρκοβιανά μοντέλα (ΚΜΜ)

2. Η  $(Y_n)_{n \geq 0}$  δοθείσης της  $\mathbf{X} = (X_n)_{n \geq 0}$  σχηματίζει μία ακολουθία δεσμευμένα ανεξάρτητων τ.μ. Δηλαδή αν  $m$  θετικός ακέραιος και  $\{k_1 < \dots < k_m\}$  ένα οποιοδήποτε διατεταγμένο σύνολο δεικτών, οι τ.μ.  $Y_{k_1}, \dots, Y_{k_m}$  είναι δεσμευμένα ανεξάρτητες δοθείσης της  $\mathbf{X}$ .
3. Για οποιοδήποτε  $n \geq 0$ , η  $Y_n$  δοθείσης της  $\mathbf{X}$  εξαρτάται μόνο από τη  $X_n$ .

Η έννοια της δεσμευμένης ανεξαρτησίας είναι δυσκολότερο να ορισθεί σε γενικούς χώρους. Για το λόγο αυτό, τα ΚΜΜ ορίζονται ως διδιάστατες Μαρκοβιανές αλυσίδες, μερικώς παρατηρήσιμες, των οποίων οι μεταβάσεις περιγράφονται από έναν πυρήνα σε χώρο γινόμενο που έχει ειδική δομή. Συγκεκριμένα ο πυρήνας θα πρέπει να είναι τέτοιος ώστε η από κοινού διαδικασία,  $(X_n, Y_n)$ , καθώς και η (κρυμμένη) διαδικασία καταστάσεων,  $(X_n)$ , να είναι μαρκοβιανές. Από τον ορισμό αυτό, αποδεικνύονται και οι παραπάνω ιδιότητες της δεσμευμένης ανεξαρτησίας των ΚΜΜ για αριθμησιμο χώρο (πόρισμα).

**Ορισμός 3.17** (ΚΜΜ). Έστω  $(\mathcal{X}, \mathcal{X})$  και  $(\mathcal{Y}, \mathcal{Y})$  δύο μετρήσιμοι χώροι και έστω  $Q : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$  και  $G : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ , μαρκοβιανό πυρήνες. Θεωρούμε το μαρκοβιανό πυρήνα που ορίζεται στον χώρο γινόμενο:  $(\mathcal{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$  ως εξής:

$$T[(x, y), C] = \iint_C Q(x, dx')G(x', dy'), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}, C \in \mathcal{X} \otimes \mathcal{Y}. \quad (3.3.1)$$

Η μαρκοβιανή αλυσίδα  $(X_n, Y_n)_{n \geq 0}$  με μαρκοβιανό πυρήνα τον  $T$  και αρχική κατανομή  $\nu \otimes G$ , όπου  $\nu$  ένα μέτρο πιθανότητας στον  $(\mathcal{X}, \mathcal{X})$ , ονομάζεται Κρυμμένο Μαρκοβιανό Μοντέλο.

Ο πυρήνας  $T$  περιγράφει πως γίνονται οι μεταβάσεις από την  $(X_n, Y_n)$  στην  $(X_{n+1}, Y_{n+1})$ . Δηλαδή,  $\mathbb{P}((X_{n+1}, Y_{n+1}) \in C | \mathcal{F}_n) = T((X_n, Y_n), C)$ , για  $n \geq 0$ . Από τον παραπάνω ορισμό, έπεται:

$$E_\nu[f(Z_0, \dots, Z_n)] = \int f(z_0, \dots, z_n) \nu \otimes G(dz_0)T(z_0, dz_1) \dots T(z_{n-1}, dz_n) \quad (3.3.2)$$

$$= \int f(x_0, y_0, \dots, x_n, y_n) \nu(dx_0)G(x_0, dy_0)Q(x_0, dx_1)G(x_1, dy_1) \dots Q(x_{n-1}, dx_n)G(x_n, dy_n). \quad (3.3.3)$$

Ο όρος κρυμμένο, αναφέρεται μόνο στο γεγονός ότι η διαδικασία  $(X_n)$  είναι μη παρατηρήσιμη. Μπορούμε επομένως να βλέπουμε τη  $(X_n)$  ως μία εικονική ενδιάμεση διαδικασία, που χρησιμοποιείται γενικότερα στο να οριστεί η κατανομή της  $(Y_n)$ . Φυσικά, η τελευταία δεν είναι μ.α.

Με  $\mathbb{P}_\nu$  και  $E_\nu$  θα συμβολίζεται το μέτρο πιθανότητας και η μέση τιμή που αντιστοιχούν στη διαδικασία  $(X_n, Y_n)_{n \geq 0}$  που ορίζεται στον κανονικό χώρο  $((\mathcal{X} \times \mathcal{Y})^{\mathbb{N}}, (\mathcal{X} \otimes \mathcal{Y})^{\otimes \mathbb{N}})$  παρόλο που το μέτρο πιθανότητας  $\nu$  ορίζεται στον  $\mathcal{X}$  μόνο. Αυτό δικαιολογείται άμεσα από τον ορισμό καθώς οποιαδήποτε και αν είναι η αρχική κατάσταση  $(X_0, Y_0)$ , η κατανομή της  $(X_n, Y_n)_{n \geq 0}$  εξαρτάται μόνο από την περιθώρια κατανομή της  $X_0$ .

Η Πρόταση που ακολουθεί προσφέρει έναν εναλλακτικό, πιο διαισθητικό τρόπο να σκεφτόμαστε ένα ΚΜΜ.

**Πρόταση 3.2.** Έστω  $(X_n, Y_n)_{n \geq 0}$  μία δισδιάστατη σ.δ. ορισμένη στον  $((\mathcal{X} \times \mathcal{Y}))$  με πυρήνα τον  $T$  που δίνεται στην (3.3.1). Τότε, για οποιοδήποτε ακέραιο  $m$  και οποιοδήποτε διατεταγμένο σύνολο δεικτών  $\{k_1 < \dots < k_m\}$  και όλες τις μετρήσιμες και φραγμένες συναρτήσεις  $f_1, \dots, f_m$  ισχύει:

$$E_\nu \left[ \prod_{i=1}^m f_i(Y_{k_i}) | X_{k_1}, \dots, X_{k_m} \right] = \prod_{i=1}^m E_\nu [f_i(Y_{k_i}) | X_{k_i}]. \quad (3.3.4)$$

Η απόδειξη είναι στο παράρτημα.

Από την παραπάνω Πρόταση, έπεται άμεσα η ιδιότητα 2. στην εισαγωγή. Δηλαδή η  $(Y_n)_{n \geq 0}$  δοθέντος της  $\mathbf{X} = (X_n)_{n \geq 0}$  σχηματίζει μία ακολουθία δεσμευμένων ανεξάρτητων τ.μ.

Η ιδιότητα 3 ισχύει επίσης σύμφωνα με το παρακάτω:

**Λήμμα 3.1** Για οποιοσδήποτε ακεραίους  $l, m$  και οποιοδήποτε διατεταγμένο σύνολο δεικτών  $\{k_1 < \dots < k_m\}$  τέτοιο ώστε  $l \notin \{k_1 < \dots < k_m\}$ , οι τ.μ.  $Y_l$  και  $(X_{k_1}, \dots, X_{k_m})$  είναι δεσμευμένα ανεξάρτητες δοθέντος της  $X_l$ .

### Απόδειξη

Για οποιασδήποτε φραγμένες και μετρήσιμες  $f, h$  ισχύει:

$$\begin{aligned} E_\nu [f(Y_l)h(X_{k_1}, \dots, X_{k_m}) | X_k] &= E_\nu [E_\nu [f(Y_l) | X_{k_1}, \dots, X_{k_m}, X_l] h(X_{k_1}, \dots, X_{k_m}) | X_k] \\ &= E_\nu [f(Y_l) | X_l] E_\nu [h(X_{k_1}, \dots, X_{k_m}) | X_l] \end{aligned}$$

Για την πρώτη ισότητα χρησιμοποιήθηκε η ιδιότητα :  $E[X | \mathcal{G}] = E[E[X | \mathcal{F}] | \mathcal{G}] = E[E[X | \mathcal{G}] | \mathcal{F}]$  για δύο σ-άλγεβρες με  $\mathcal{G} \subset \mathcal{F}$  όπως επίσης και η  $E[f(T)Q | \mathcal{G}] = f(T)E[X | \mathcal{G}]$  για  $T$  ανεξάρτητη της  $\mathcal{G}$  και  $f$  μετρήσιμη, θετική ή ολοκληρώσιμη.  $\square$

Σημειώνεται επίσης πως οι παραπάνω ιδιότητες χαρακτηρίζουν τα ΚΜΜ. Επίσης, έπεται πως η πεπερασμένης διάστασης κατανομή της  $\mathbf{Y} | \mathbf{X} = \mathbf{x}$  είναι για  $A = (A_0 \times \dots \times A_n)$ :

$$\mathbb{P}_\nu(Y_0, \dots, Y_n \in A | X_0, \dots, X_n) = G(X_0, A_0) \dots G(X_n, A_n)$$

Τέλος, τα ΚΜΜ μπορούν να χαρακτηριστούν και από την ακόλουθη Πρόταση που έπεται από τα παραπάνω:

**Πρόταση 3.3** (Χαρακτηρισμός ΚΜΜ). Έστω  $(X_n, Y_n)_{n \geq 0}$  μία δισδιάστατη ορισμένη στον  $((\mathcal{X} \times \mathcal{Y}))$ . Η διαδικασία αυτή είναι ΚΜΜ αν υπάρχουν πυρήνες  $Q : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$  και  $G : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$  τέτοιο ώστε:

$$E_\nu [f(X_{n+1}, Y_{n+1}) | X_0, Y_0, \dots, X_n, Y_n] = \int f(x, y) Q(X_n, dx) G(x, dy), \quad (3.3.5)$$

όπου το  $\nu$  είναι μέτρο πιθανότητας στον  $\mathcal{X}$  και είναι τέτοιο ώστε

$$E_\nu [f(X_0, Y_0)] = \int f(x, y) \nu(dx) G(x, dy), \quad (3.3.6)$$



## Κρυμμένα Μαρκοβιανά μοντέλα (ΚΜΜ)

για οποιαδήποτε φραγμένη και μετρήσιμη συνάρτηση  $f$ .

### Παραδείγματα.

1. Έστω  $(\alpha_n)_{n \geq 1}$  και  $(\beta_n)_{n \geq 0}$  ακολουθίες από ανεξάρτητες και ισόνομες τ.μ με κατανομές  $\alpha$  και  $\beta$  αντίστοιχα. Ορίζουμε:  $X_0 = z$ ,  $Y_0 = R(X_0, \beta_0)$  και για  $n \geq 1$ :

$$\begin{aligned} X_n &= \Phi(X_{n-1}, \alpha_n), \\ Y_n &= R(X_n, \beta_n) \end{aligned}$$

όπου  $\Phi : (X) \times \mathbb{R} \mapsto \mathcal{X}$  και  $R : (X) \times \mathbb{R} \mapsto \mathcal{Y}$  και  $z \in \mathcal{X}$ . Τότε η στοχαστική διαδικασία  $(X_n, Y_n)_{n \geq 0}$  είναι ΚΜΜ με πυρήνες  $Q, G$  που δίνονται από τις:

$$Q(x, A) = \int \mathbb{1}_{\{\Phi(x,s) \in A\}} \alpha(ds),$$

και

$$G(x, B) = \int \mathbb{1}_{\{R(x,s) \in B\}} \beta(ds),$$

και αρχική κατανομή  $\delta_x$ .

Πράγματι, λόγω της ανεξαρτησίας της  $\beta_{n+1}$  από τις  $X_{0:k+1}$  και  $Y_{0:n}$ , είναι:

$$\begin{aligned} E_{\delta_x}(Q_{n+1}, Y_{n+1}) | X_0, Y_0, \dots, X_n, Y_n &= E_{\delta_x}[f(X_{n+1}), \dots] = E_{\delta_x}[E_{\delta_x}[f(X_{n+1}, R(X_{n+1}, \beta_{n+1})) | X_{n+1}]] \\ &= \int g(x, y) G(x, dy) Q(X_k, dx) \end{aligned}$$

2. Έστω το μοντέλο:

$$Y_n = R(X_n) + \xi_n$$

όπου η  $R$  είναι μία μετρήσιμη συνάρτηση και τα  $\xi_n \sim \mathcal{N}(0, 1)$  και η  $X_n$  είναι μία Μ.Α. με πυρήνα  $Q$  και αρχική κατανομή  $\nu$ . Τότε η  $(X_n, Y_n)_{n \geq 0}$  είναι ένα ΚΜΜ με:

$$G(x, B) = \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{(s-h(x))^2}{2}} ds$$

### 3.3.3 Βασικές ιδιότητες

Θεωρούμε τη Μ.Α.  $(X_n, Y_n)_{n \geq 0}$  με τιμές στον χώρο γινόμενο  $(\mathcal{X} \text{ times } \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$  και πυρήνα  $T$  που ικανοποιεί τον ορισμό (3.3.1). Έστω επίσης οι πυρήνες  $Q$  και  $G$  που αντιστοιχούν στις στοχαστικές διαδικασίες  $(X_n)_{n \geq 0}$  και  $(Y_n | X = x)_{n \geq 0}$ . Στην υποενότητα αυτή θα συζητηθούν ορισμένες βασικές έννοιες για την παραπάνω Μ.Α. (που λέγεται ΚΜΜ), όπως:

- 1  $\phi$ -αδιαχωρησιμότητα.
- 2 Περιοδικότητα.
- 3 Επαναληπτικότητα και ύπαρξη στάσιμου μέτρου.

## Κρυμμένα Μαρκοβιανά μοντέλα (KMM)

Θα εκφράσουμε τις ιδιότητες αυτές μέσω του  $T$ . Οι αποδείξεις των προτάσεων υπάρχουν στο (...).

Η  $\phi$  αδιαχωρησιμότητα της κοινής αλυσίδας  $T$  κληρονομείται από την αδιαχωρησιμότητα της κρυμμένης αλυσίδας  $(X_n)_{n \geq 0}$ . Συγκεκριμένα:

**Πρόταση 3.4.** Αν ο  $Q$  είναι  $\phi$ -αδιαχώριστος και έχει περίοδο  $d$ , τότε ο  $T$  είναι  $\phi$ -αδιαχώριστος με την ίδια περίοδο.

Επομένως, αν η αλυσίδα  $(X_n)_{n \geq 0}$  είναι απεριοδική, τότε θα είναι και η  $(X_n, Y_n)_{n \geq 0}$ . Σχετικά με την επαναληπτικότητα και την ύπαρξη στάσιμου μέτρου έχουμε το επόμενο αποτέλεσμα:

**Πρόταση 3.5.** Υποθέτουμε πως ο  $Q$  είναι  $\phi$ -αδιαχώριστος. Τότε ισχύουν τα ακόλουθα:

- 1 Η  $(X_n, Y_n)_{n \geq 0}$  είναι μεταβατική (επαναληπτική) αν η  $(X_n)_{n \geq 0}$  είναι μεταβατική (επαναληπτική).
- 2 Η  $(X_n, Y_n)_{n \geq 0}$  είναι θετική αν η  $(X_n)_{n \geq 0}$  είναι θετική. Συγκεκριμένα, αν η  $(X_n)_{n \geq 0}$  έχει στάσιμο μέτρο  $\pi$ , τότε η  $(X_n, Y_n)_{n \geq 0}$  έχει στάσιμο μέτρο  $\pi \otimes G$ .

### 3.3.4 Βασικά Προβλήματα στα KMM

Το κύριο πρόβλημα της μοντελοποίησης χρησιμοποιώντας KMM είναι το εξής: Έχοντας καθορίσει πλήρως το μοντέλο και δοθέντος των δεδομένων που έχουμε, έστω  $Y_0, \dots, Y_N$ , τι μπορεί να ειπωθεί για τις αντίστοιχες κρυμμένες καταστάσεις  $X_0, \dots, X_N$ . Ειδικότερα, με τα KMM συνήθως συνδέονται τα ακόλουθα 3 βασικά προβλήματα εκτιμητικής smoothing, prediction, filtering. Τα προβλήματα αυτά έχουν μελετηθεί σε μεγάλο βαθμό τα τελευταία χρόνια και μία πολύ αναλυτική συζήτηση για τις τεχνικές επίλυσης τους, γίνεται μεταξύ άλλων, στα:...

Αναφορικά, τα προβλήματα αυτά έχουν ως εξής:

Στο πρώτο πρόβλημα έχουμε πεπερασμένου πλήθους παρατηρήσεις  $Y_0, \dots, Y_N$  και ζητάμε να εκτιμήσουμε την ακολουθία  $X_1, \dots, X_N$  που αντιστοιχεί σε αυτά. Στην περίπτωση αυτή λοιπόν θα θέλαμε να υπολογίσουμε τις δεσμευμένες μέσες τιμές:

$$E(f(X_n)|Y_0, \dots, Y_N), \quad 0 \leq n \leq N,$$

όπου  $f$  μία οποιαδήποτε συνάρτηση.

Στο δεύτερο πρόβλημα, πάλι γνωρίζοντας τις παρατηρήσεις  $Y_0, \dots, Y_N$  θα θέλαμε να εκτιμήσουμε τη μελλοντική εξέλιξη της αλυσίδας  $(X_n)$ , δηλαδή να υπολογίσουμε τις :

$$E(f(X_n)|Y_0, \dots, Y_N), \quad n \geq N,$$

Τέλος, το πιο συχνό σενάριο, που αποτελεί και το τρίτο πρόβλημα, είναι να θέλουμε να εκτιμήσουμε την παρούσα κατάσταση  $X_n$  δοθέντος όλων παρατηρήσεων μέχρι την τρέχουσα χρονική στιγμή. Δηλαδή ζητάμε να υπολογίσουμε τις:

$$E(f(X_n)|Y_0, \dots, Y_n), \quad n \geq 0,$$

### 3.4 Σύνδεση με τη Στατιστική

Στα προβλήματα συμπερασματολογίας (δες παράρτημα) θεωρούμε παραμετρικές οικογένειες για την ακολουθία κρυμμένων καταστάσεων  $((X_n))$  (ή μεταβάσεων), έστω  $\mathbb{P}_\phi^1$ , για την ακολουθία παρατηρήσεων  $((Y_n))$ , έστω  $\mathbb{P}_\phi^2$  (όχι απαραίτητα διαφορετική από την  $\mathbb{P}^1$ , και για το αρχικό μέτρο  $\nu_\phi$ . Η παράμετρος παίρνει τιμές σε κάποιο σύνολο  $\Phi$  (συνήθως Ευκλείδειο). Ο στόχος είναι να επιλέξουμε ένα  $\phi^* \in \Phi$  ώστε κάτω από το  $\mathbb{P}_\phi^2$ , η  $((Y_n))$  να αναπαράγει τις στατιστικές ιδιότητες κάποιο δοσμένου συνόλου δεδομένων:  $y_0, \dots, y_N$ . Συχνά το πρόβλημα αντιμετωπίζεται μέσω της εκτιμήτριας μέγιστης πιθανοφάνειας αλλά είναι εξαιρετικά σπάνιο αυτή μπορεί να έχει αναλυτική λύση. Ο αλγόριθμος EM που παρουσιάζεται στο κεφάλαιο 5 αποτελεί μία από τις πιο κατάλληλες αλγοριθμικές τεχνικές με τις οποίες μπορούμε να αντιμετωπίσουμε το πρόβλημα. Γενικότερα τα προβλήματα στατιστικής συμπερασματολογίας που εμπλέκουν KMM λέγονται προβλήματα με μη πλήρη δεδομένα (Incomplete Data Problems) και γίνεται συζήτηση στο παράρτημα για αυτά τα προβλήματα.

Για παράδειγμα, όταν η κρυμμένη διαδικασία έχει πεπερασμένο χ.κ., ο μεταβατικός πυρήνας  $(Q)$  είναι πίνακας και η αρχική κατανομή είναι διάνυσμα. Στην περίπτωση αυτή είναι εφικτό να εκτιμήσουμε τον  $Q$  εξ' ολοκλήρου καθώς και την αρχική κατανομή με τον αλγόριθμο EM. Αρχικά θα πρέπει να δούμε τη μορφή της συνάρτησης πιθανοφάνειας των παρατηρούμενων δεδομένων. Συγκεκριμένα:

Έστω  $a = (a_1, \dots, a_s)$  οι αρχικές πιθανότητες της μ.α. και  $p = (p_{ij})_{i,j \in \mathcal{X}}$  οι πιθανότητες μετάβασης. (δηλαδή  $Q(i, j) = p_{ij}$ ). Έστω επίσης  $\phi = (\theta_1, \dots, \theta_s)$ , όπου το  $\theta_i$  χαρακτηρίζει την κατανομή  $Y_n | X_n = i$ . Θα γράψουμε μία έκφραση για συνάρτηση πιθανοφάνειας, έστω  $L(\phi) = L(a, p, \theta)$ . Διακρίνουμε τις περιπτώσεις:

- (i) Οι  $(Y_k | X_k = i)_{k \geq 0}$  είναι απόλυτα συνεχής στην ίδια παραμετρική οικογένεια για  $i = 1, 2, \dots, s$ . (Δηλαδή ο πυρήνας  $G$  έχει πυκνότητα  $g(i, \cdot; \phi)$ ). Τότε:

$$\begin{aligned} L_y(\phi) &\equiv L_{y_{0:n}}(\phi) = \\ &= \sum_{x_{0:n}} \mathbb{P}_\nu(X_{0:n} = x_{0:n}) g(y_{0:n} | x_{0:n}; \theta) \\ &= \sum_{x_{0:n}} \nu(x_0) \prod_{k=0}^n Q(x_k, x_{k+1}) \prod_{k=0}^n g(x_k, y_k; \theta) \\ &= \sum_{x_{0:n}} a_{x_0} \prod_{k=0}^n p_{x_k, x_{k+1}} \prod_{k=0}^n g(x_k, y_k; \theta) \end{aligned} \quad (3.4.1)$$

- (ii) Οι  $(Y_k | X_k = i)_{k \geq 0}$  είναι διακριτές στην ίδια παραμετρική οικογένεια για  $i = 1, 2, \dots, s$ . (Δηλαδή ο πυρήνας  $G$  είναι πίνακας, με στοιχεία ας πούμε  $r_{ij}$ ). Τότε από την προηγούμενη:

$$L_y(\phi) = \sum_{x_{0:n}} a_{x_0} \prod_{k=0}^n p_{x_k, x_{k+1}} \prod_{k=0}^n r_{x_k, y_k} \quad (3.4.2)$$

Από τις παραπάνω βλέπουμε πως εξαιτίας του αθροίσματος δεν είναι δυνατόν να έχουμε αναλυτική έκφραση για τη συνάρτηση πιθανοφάνειας.

Θα επιστρέψουμε στο παραπάνω πρόβλημα στο κεφάλαιο 5, αφού έχει οριστεί ο αλγόριθμος EM.

Επίσης, η πιο γενική περίπτωση είναι να έχουμε γενικό  $\chi.x$  για την  $(X_n, Y_n)$ . Τότε η πιθανοφάνεια έχει την ακόλουθη μορφή:

$$L_y(\phi) = \int \dots \int \nu(x_0)g(x_0, y_0; \phi)Q(x_0, dx_1)g(x_1, y_1; \phi) \dots Q(x_{n-1}, dx_n)g(x_n, y_n; \phi). \quad (3.4.3)$$

Η εκτίμηση του  $\phi$  εδώ δεν είναι ένα απλό πρόβλημα. Στις περισσότερες περιπτώσεις αντιμετωπίζεται με αλγοριθμικές τεχνικές αλλά και πάλι το πλαίσιο ορισμού του μοντέλου και οι υποθέσεις παίζουν σημαντικό ρόλο για το πόσο εφικτό μπορεί να είναι το πρόβλημα εκτίμησης. Ωστόσο, ένα πλαίσιο το οποίο έχουν μελετηθεί συστηματικά προβλήματα εκτίμησης γενικών χώρων καταστάσεων για τα KMM είναι το ακόλουθο:

Ένα Κρυμμένο Μαρκοβιανό μοντέλο (HMM) είναι ένα δυναμικό σύστημα που χαρακτηρίζεται από τις εξισώσεις:

$$\begin{aligned} X_{n+1} &= \Phi_{n+1}(X_n, W_{n+1}, \theta), \quad n \geq 0 \\ Y_n &= R_n(X_n, V_n, \theta), \quad n \geq 0, \end{aligned}$$

όπου, οι συναρτήσεις  $\Phi, R$  είναι Borel μετρήσιμες και οι  $(W_n)_{n \geq 0}, (V_n)_{n \geq 0}$  είναι ακολουθίες ανεξάρτητων και ισόνομων τ.μ. που αντιπροσωπεύουν σφάλματα του μοντέλου. Οι στοχαστικές διαδικασίες  $(X_n)_{n \geq 0}, (Y_n)_{n \geq 0}$  αντιπροσωπεύουν τις καταστάσεις και τα παρατηρούμενα δεδομένα αντίστοιχα. Τέλος το  $\theta$  είναι το διάνυσμα με τις παραμέτρους του μοντέλου το οποίο παίρνει τιμές σε κάποιο χώρο  $\Theta \subset \mathbb{R}^{d_\theta}$ . Στόχος είναι η ταυτόχρονη εκτίμηση της  $X_n, \theta_n$  από το διάνυσμα παρατηρήσεων  $Y_{0:N}$  για κάποιο σταθερό  $N$ .

Αξίζει να σημειωθεί ότι αυτό είναι και το πλαίσιο σε γενικές γραμμές που ακολουθήθηκε για την κατασκευή των μοντέλων GreenLab και LNAS που περιγράφηκαν στο κεφάλαιο 2.

Μία πολύ αναλυτική συζήτηση για τα προβλήματα εκτιμητικής καθώς και η αντιμετώπιση γίνεται από τους Carpe-Mouline-Ryden στο [11]. Επίσης, αξίζει να σημειωθεί ότι στα προηγούμενα προβλήματα εκτιμητικής, έχουμε θεωρήσει πως το KMM είναι ήδη γνωστό. Ωστόσο, στις περισσότερες εφαρμογές, δεν είναι καθόλου προφανές πως θα σχεδιαστούν οι πυρήνες μετάβασης και παρατηρήσεων, που αντιστοιχούν στις διαδικασίες  $(X_n)$  και  $(Y_n)$ .

Κλείνουμε την ενότητα αυτή αναφέροντας ορισμένες επεκτάσεις και γενικεύσεις των KMM:

1. KMM  $\{X_n, Y_n\}_{n \geq 0}$  όπου ο  $(Q, \mathcal{X})$  είναι υπεραριθμήσιμος χώρος καταστάσεων.
2. KMM συνεχούς χρόνου:  $\{X_t, Y_t\}_{t \geq 0}$
3. Πολυδιάστατα KMM.

4. Κρυμμένα Ημιμαρκοβιανά μοντέλα.

5. KMM τύπου  $M_k - M_l$

Σχετικά με το 5.: Η  $\{X_n, Y_n\}_{n \geq 0}$  είναι τύπου  $M_k - M_l$  αν η  $(X_n)_{n \geq 0}$  είναι μία  $k$ -τάξης Μαρκοβιανή αλυσίδα, δηλαδή:

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n^X) = Q(X_{n-k+1:n}, A)$$

και η  $(Y_n)_{n \geq 0} | (X_n)_{n \geq 0}$  είναι μία δεσμευμένη Μαρκοβιανή αλυσίδα  $l$ -τάξης, δηλαδή:

$$\mathbb{P}(Y_n \in A | \mathcal{F}_n^X, \mathcal{F}_{n-1}^Y) = T[(X_n, Y_{n-1:n-l}), A]$$

Ειδικά τα KMM τύπου  $M_1 - M_0$ , ( $M_0$ : δεσμευμένα ανεξάρτητα) λέγονται μοντέλα ανεξαρτήτων Μίξεων (IMM: Independent Mixture Models).

## Κεφάλαιο 4

### Μέθοδοι Markov Chain Monte Carlo (MCMC)

#### 4.1 Εισαγωγή στις Μεθόδους MCMC

Το πρόβλημα που αντιμετωπίζεται με τους αλγόριθμους MCMC είναι το ακόλουθο: Έστω  $\Pi^*$  μία κατανομή σε κάποιο χώρο  $(\mathcal{X}, \mathcal{X}, \nu)$ , με πυκνότητα ή σ.πιθανότητας ανάλογη της  $(\pi(x))_{x \in \mathcal{X}}$ , δηλαδή μη-κανονικοποιημένη, η οποία τουλάχιστον ικανοποιεί την  $0 < \int \pi(x)\nu(dx) < \infty$ . Τυπικά το  $\mathcal{X}$  είναι ανοικτό σύνολο στον  $\mathbb{R}^d$  (και  $\mathcal{X} = \mathcal{B}(\mathbb{R}^d)$ ) και οι πυκνότητες αντιστοιχούν σε παραγώγους Radon-Nikodym ως προς το μέτρο Lebesgue (σε άλλες (ρεαλιστικές εξίσου) περιπτώσεις, το πλαίσιο μπορεί να είναι σε αριθμήσιμο χώρο). Υποθέτουμε ότι η  $\pi(x)$  είναι μία γνωστή συνάρτηση.

Από τη συνάρτηση αυτή, μπορούμε να κατασκευάσουμε ένα μέτρο πιθανότητας  $\Pi^*$  στον  $\mathcal{X}$  θέτοντας:

$$C = \begin{cases} \sum_{x \in A} \pi(x), & \text{στη διακριτή περίπτωση} \\ \int \pi(x)\nu(dx), & \text{στη συνεχή περίπτωση} \end{cases}$$

Τότε, το  $\Pi^*$  είναι η κατανομή με πυκνότητα (ως προς το  $\nu$ ) (ή σ. πιθανότητας) την:  $\pi^* \stackrel{\text{def}}{=} \pi(x)/C$ . Για τις δύο βασικές περιπτώσεις (αριθμήσιμος χώρος με  $\nu = \mu_a$  (αριθμητικό),  $X \subset \mathbb{R}^d$  με  $\nu = \lambda_d$ :

$$\Pi^*(A) = \int_A \pi^*(x)dx = \frac{\int_A \pi(x)dx}{\int_X \pi(x)dx} \quad (4.1.1)$$

Το ερώτημα που τίθεται είναι πώς θα εκτιμήσουμε ολοκληρώματα συναρτήσεων  $f : X \rightarrow \mathbb{R}$  ως προς το μέτρο  $\Pi^*$ , δηλαδή:

$$\Pi^*(f) = E_{\pi^*}[f(X)] = \int f(x)\pi^*(x)dx = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \quad (4.1.2)$$

Ειδικά όταν ο  $Q$  έχει μεγάλη διάσταση, και η  $\pi$  έχει σχετικά πολύπλοκη έκφραση, τότε η αναλυτική (ή η αριθμητική) επίλυση των ολοκληρωμάτων είναι ανέφικτη.

Ο τρόπος που αντιμετωπίζεται το πρόβλημα με κλασικό Monte Carlo είναι να προσμοιωθούν α.ι.τ.μ.  $\xi_1, \dots, \xi_N \sim \Pi^*$  και στη συνέχεια να γίνει εκτίμηση του  $\Pi^*$  από την εμπειρική κατανομή  $\hat{\Pi}^*$  :

$$\hat{\Pi}_N^*(A) = \frac{1}{N+1} \sum_{i=0}^N \mathbb{1}_{\{\xi_i \in A\}}. \quad (4.1.3)$$

για οποιοδήποτε μετρήσιμο  $A \in \mathcal{X}$ . Αποδεικνύεται πως η εμπειρική κατανομή, κάτω από ασθενείς συνθήκες, συγκλίνει στο  $\Pi^*$  και τότε το  $\hat{\Pi}_N^*$  θα είναι κοντά στο  $\Pi^*$

για μεγάλο  $N$ . Επομένως, η ζητούμενη εκτιμήτρια του  $\Pi^*$  θα είναι:

$$\hat{\Pi}^*(f) = \frac{1}{N+1} \sum_{i=0}^N f(\xi_i). \quad (4.1.4)$$

Η παραπάνω εκτιμήτρια αποδεικνύεται πως είναι αμερόληπτη με τυπική απόκλιση της τάξης του  $O(1/\sqrt{N})$ . Επιπλέον, αν  $\pi^*(f^2) < \infty$ , τότε με το Κεντρικό Οριακό Θεώρημα παίρνουμε ότι το σφάλμα:  $\hat{\Pi}^*(f) - \Pi^*(f)$  έχει την κανονική ως οριακή κατανομή που είναι εξίσου χρήσιμο. Το πρόβλημα ωστόσο είναι πως αν η  $\pi^*$  είναι αρκετά περίπλοκη, υπάρχει μεγάλη δυσκολία στον να προσομοιωθούν ευθέως οι τ.μ. με από την  $\Pi^*$ .

Τέλος, αν και στο προηγούμενο κεφάλαιο παρουσιάστηκαν οι συνθήκες με τις οποίες έχουμε εγγύηση πως για μία Μ.Α. (με γνωστό) πυρήνα  $Q$ , υπάρχει (ένα άγνωστο αρχικά) στάσιμο μέτρο  $\Pi^*$  (με πυκνότητα ή σ.π.  $\pi^*$ ) και ισχύει:  $Q^n(x, y) \rightarrow \pi^*(y)$ , με τις τεχνικές των MCMC κάνουμε το αντίστροφο: Το στάσιμο μέτρο είναι γνωστό (όχι απαραίτητα κανονικοποιημένο) αλλά ο πυρήνας είναι άγνωστος. Για να πάρουμε δείγματα από την  $\pi^*$ , με τις μεθόδους αυτές κατασκευάζεται ένας πυρήνας  $P$  για τον οποίο  $P^n(x, y) \rightarrow \pi^*(y)$ .

Η αντιμετώπιση λοιπόν του προβλήματος με MCMC τεχνικές έχεις ως εξής: κατασκευάζεται μία Μαρκοβιανή αλυσίδα στον  $X$  (εύκολα υπολογίσιμη στον υπολογιστή), η οποία έχει στάσιμη κατανομή το  $\pi^*$ .

Ο Tierney (1994) ήταν από τους πρώτους που διατύπωσε συνθήκες με τις οποίες μπορούν να αναλυθούν οι μέθοδοι MCMC και οι ιδιότητες. Επίσης, στο Meyn, Tweedie, Hibe (1993), διατυπώνονται συνθήκες για τη σύγκλιση του εργοδικού μέσου και για κεντρικά οριακά θεωρήματα.

Στη συνέχεια θα δούμε συγκεκριμένους αλγόριθμους που βασίζονται στις τεχνικές αυτές και θα συζητηθεί το θέμα της σύγκλισης τους.

#### 4.1.1 Ο αλγόριθμος Metropolis-Hastings (MH)

Γενικά η προσομοίωση μιας Μαρκοβιανής αλυσίδας δεν έχει δυσκολίες καθώς η δεσμευμένη κατανομή της  $X_{n+1}|X_n = x$  είναι γνωστή από τις πιθανότητες μετάβασης της αλυσίδας που εκφράζονται μέσω του αντίστοιχου πυρήνα ( $Q(x, \cdot)$ ). Ο βασικός στόχος είναι να κατασκευαστούν κατάλληλες πιθανότητες μετάβασης οι οποίες θα εξασφαλίζουν ότι η  $(X_n)_{n \geq 0}$  θα έχει ως στάσιμη κατανομή την  $\pi$  (κατανομή-στόχος (target-distribution)).

Αν και υπάρχουν αρκετοί τρόποι κατασκευής των πιθανοτήτων αυτών, η Πρόταση που δόθηκε από τον Metropolis (Metropolis et al. 1953) και στη συνέχεια επεκτάθηκε από τον Hastings (Hastings 1970) είναι εξαιρετικά απλή:

Υποθέτουμε αρχικά ότι διαθέτουμε έναν πυρήνα  $Q$  (κατανομή-πρόταση (proposal-distribution)) που αντιστοιχεί σε μία οποιαδήποτε άλλη μ.α., της οποίας οι μεταβάσεις έχουν επίσης πυκνότητα (ενδεχομένως μη κανονικοποιημένη). Δηλαδή:  $Q(x, dy) \propto$

$q(x, y)dy$ .

Ο αλγόριθμος Metropolis-Hastings (MH) ορίζεται ως εξής:

### Αλγόριθμος MH.

Αρχικά (στο βήμα  $n = 0$ ) διαλέγουμε μία τιμή  $x_0$ .

Τότε, στο βήμα  $n$  για  $n \geq 1$ , δοθέντος της τιμής της  $X_n = x_n$ , προσομοιώνουμε μία (υποψήφια μετάβαση) τιμή  $y$  από τον  $Q(x_n, \cdot)$ .

Στη συνέχεια, 'δεχόμαστε' την τιμή της  $y$  και θέτουμε  $X_{n+1} = y$  με πιθανότητα:

$$a(x_n, y) \stackrel{def}{=} \min\{1, r(x_n, y)\}, \quad \text{όπου } r(x_n, y) \stackrel{def}{=} \frac{\pi(y)q(y, x_n)}{\pi(x_n)q(x_n, y)}. \quad (4.1.5)$$

Διαφορετικά η τιμή  $y$  απορρίπτεται (με πιθανότητα  $1 - a(x_n, y)$ ) και θέτουμε  $X_{n+1} = x_n$ .

Αντικαθιστούμε το  $n$  με  $n + 1$  και συνεχίζουμε.

Με την παραπάνω διαδικασία, κατασκευάζεται μία στοχαστική διαδικασία  $(X^{(n)})_{n \geq 0}$ , η οποία, κληρονομεί τη μαρκοβιανή ιδιότητα από το μαρκοβιανό πυρήνα  $Q$  και άρα είναι μία μ.α. Πράγματι, αν συμβολίσουμε με  $P$  τον πυρήνα της αλυσίδα που παράγεται από τον αλγόριθμο MH, για οποιοδήποτε μετρήσιμο σύνολο  $A$ , έπεται:

$$\mathbb{P}_{\mu_Q}(X^{(n+1)} \in A | \mathcal{F}_n) = P(X^{(n)}, A),$$

όπου,

$$P(x, A) = \int_A a(x, y)q(x, y)dy + \mathbb{1}_{\{x \in A\}} \int [1 - a(x, y)]r(x, y)dy \quad (4.1.6)$$

και η πυκνότητα:

$$p(x, y) = \begin{cases} a(x, y)q(x, y), & y \neq x \\ a(x, y)q(x, y) + \int [1 - a(x, y)]r(x, y)dy, & y = x \end{cases} \quad (4.1.7)$$

Επίσης, με  $\mu_Q$  συμβολίζεται η αρχική κατανομή της αλυσίδας που αντιστοιχεί στο Μαρκοβιανό πυρήνα  $Q$ . Μάλιστα, η μ.α. που παράγεται από τον αλγόριθμο MH είναι και κάτι περισσότερο:

**Πρόταση 4.1.** Ο αλγόριθμος Metropolis-Hastings παράγει μία μ.α.  $(X^{(n)})_{n \geq 0}$  που είναι αντιστρέψιμη ως προς το  $\pi^*$ .

### Απόδειξη

Αρκεί να δειχθεί ότι:

$$\pi^*(dx)P(x, dy) = \pi^*(dy)P(y, dx).$$



## Εισαγωγή στις Μεθόδους MCMC

Για  $x = y$  η παραπάνω έπεται άμεσα. Για  $x \neq y$ :

$$\begin{aligned} \pi(dx)P(x, dy) &= (C^{-1}p_i^*(x)dx)(q(x, y)a(x, y)dy) \\ &= C^{-1}p_i^*(x)q(x, y)\min\left\{1, \frac{\pi(y)q(y, x_n)}{\pi(x_n)q(x_n, y)}\right\}dxdy \\ &= C^{-1}\min\{\pi(x)q(x, y), \pi(y)q(y, x)\}dxdy \end{aligned}$$

Η τελευταία έκφραση είναι συμμετρική ως προς τα  $x$  και  $y$ . ( $C = \int \pi(x)dx$ ), στην αρχή της ενότητας).  $\square$

Ο παραπάνω είναι και ο λόγος για αυτή την ασυνήθιστη σχέση για τις πιθανότητες  $a(x, y)$ .

### Παρατηρήσεις.

1. Με τα παραπάνω υπονοείται πως αν ο πυρήνας  $Q$  δεν είναι Μαρκοβιανός (πχ στο βήμα  $n - 1$ , ο  $Q$  εξαρτάται από όλη την προηγούμενη ιστορία  $X^{(0)}, \dots, X^{(n-1)}$ ), τότε η αλυσίδα δεν θα είναι Μαρκοβιανή. Αυτά είναι συνήθως σεναρία Adaptive MH και σε αυτά, μπορεί η επιλογή του πυρήνα  $Q$  να είναι τέτοια ώστε να υπάρχει εξάρτηση, αλλά στο όριο να χαλαρώνει. Έτσι μπορεί να γίνει λόγος για ασυμπτωτικά ζητήματα.
2. Προφανώς, η πιθανότητα  $a(x_n, y)$  ορίζεται μονάχα αν  $\pi(x_n) > 0$ . Ωστόσο, αν η αρχική τιμή  $x_0$  επιλεγεί ώστε  $\pi(x_0) > 0$ , τότε  $\pi(x_n) > 0$  για κάθε  $n \geq 1$  διότι οι τιμές  $y = y_n$  για τις οποίες  $\pi(y_n) = 0$  οδηγούν σε  $a(x_n, y_n) = 0$  και έτσι απορρίπτονται από την αλγόριθμο. Θα θεωρούμε τη σύμβαση: ο λόγος  $a(x_n, y) = 0$  όταν οι  $\pi(x_n), \pi(y_n)$  είναι ταυτόχρονα 0 (προς αποφυγή ορισμένων θεωρητικών δυσκολιών).
3. Η πιθανότητα αποδοχής  $a$  είναι μία ποσότητα συνυφασμένη με την αποδοτικότητα του αλγορίθμου Roberts, Rosenthal (2001) και μία εναλλακτική ερμηνεία της πιθανότητας αυτής μας δίνει το εργοδικό Θεώρημα κατά το οποίο το  $a$  ισούται με το μακροπρόθεσμο ποσοστό προτεινόμενων μεταβάσεων που έγιναν αποδεκτές από την αλυσίδα:

$$a = \int a(x, y)\pi^*(x)q(x, y)dxdy$$

4. Ένα δείγμα που παράγεται από τον παραπάνω αλγόριθμο διαφέρει από ένα τυχαίο δείγμα καθώς για παράδειγμα, είναι δυνατόν να περιλαμβάνει επαναλαμβανόμενες πραγματοποιήσεις της ίδιας τιμής, καθώς η απόρριψη του  $y$  οδηγεί σε επανάληψη της  $X_n$  στο βήμα  $n$ . (πράγμα που είναι αδύνατο να συμβεί σε πλαίσιο με α.ι.τ.μ. από συνεχή κατανομή).
5. Ακόμα, ένα βασικό χαρακτηριστικό του αλγορίθμου MHόπως φαίνεται και από τη σχέση 4.1.5, είναι πως χρειάζεται να υπολογίσουμε μόνο λόγους:  $\frac{\pi(y)}{\pi(x)}$  και επομένως μπορούμε να αγνοήσουμε εντελώς τη σταθερά κανονικοποίησης ( $C = \int \pi(x)dx$ ) το οποίο είναι μεγάλο πλεονέκτημα.

Ο στόχος είναι να είναι να εξασφαλίσουμε ότι:

$$P^N(x, y) \rightarrow \pi^*(x, y) \quad \text{καθώς} \quad N \rightarrow \infty. \quad (4.1.8)$$

Έτσι, από ένα μεγάλο  $m$ , θα μπορούμε να θεωρήσουμε πως οι παρατηρήσεις  $x^{(m)}, x^{(m+1)}, \dots$  είναι τυχαίο δείγμα από την  $\pi^*$ .

Στην ενότητα 4.3 δίνονται συνθήκες ώστε η αλυσίδα  $((X^{(n)})_{n \geq 0})$  που παράγεται από τον αλγόριθμο, να ικανοποιεί την (4.1.8). Ωστόσο, μία απαραίτητη προϋπόθεση είναι η εξής:

Αν  $\mathcal{E}$  είναι το στήριγμα της  $\pi^*$ , τότε αν για παράδειγμα υπάρχει κάποιο  $A \subset \mathcal{E}$  για το οποίο:

$$\int_A \pi^*(dx) > 0 \quad \text{και} \quad \int_A q(y|x)dy = 0 \quad \text{για κάθε} \quad x \in \mathcal{E},$$

τότε ο παραπάνω αλγόριθμος δεν έχει ως οριακή κατανομή την  $\pi^*$  καθώς αν  $x_0 \notin A$  η αλυσίδα  $(X^{(n)})$  δεν θα επισκεφτεί ποτέ το  $A$ . Οι Robert, Casella στο [35] σημειώνουν πως η ελάχιστη αναγκαία συνθήκη είναι:

$$\text{supp} \pi \subset \cup_{x \in \text{supp} \pi} \text{supp} q(x, \cdot). \quad (4.1.9)$$

Ολοκληρώνοντας την εισαγωγή, μένει να δούμε πως μπορεί να γίνει η προσομοίωση σε έναν υπολογιστή όπως επίσης και να εξετάσουμε την επιλογή της συνάρτησης-πρότασης.

Όπως αναφέρθηκε, η προσομοίωση τιμών από τις δεσμευμένες κατανομές  $Q(x, \cdot)$  είναι μία σχετικά απλή διαδικασία. Αυτό διότι αν διαθέτουμε τις πιθανότητες μετάβασης ή τη σ. πυκνότητας (και η τελευταία είναι απλή), το μόνο που χρειάζεται επιπλέον είναι μία διαδικασία αποδοχής-απόρριψης (που είναι εξίσου απλή).

Συνεπώς, το να τρέξουμε τον αλγόριθμο Metropolis-Hastings είναι μία σχετικά απλή διαδικασία.

Υπάρχουν πολλοί τρόποι επιλογής της συνάρτησης-πρότασης. Η βασικότερη προϋπόθεση είναι να έχει στήριγμα που να περιλαμβάνει το στήριγμα (συνθήκη (4.1.9)) της συνάρτησης-πρότασης. Ωστόσο, μπορούμε να διακρίνουμε τις παρακάτω κατηγορίες:

1. Συμμετρικός Metropolis-Hastings. Εδώ,  $q(x, y) = q(y, x)$ , και άρα η πιθανότητα αποδοχής απλοποιείται σε:  $a(x, y) = \min\{1, \frac{\pi(y)}{\pi(x)}\}$ .
2. Metropolis-Hastings Τυχαίου Περιπάτου. Εδώ,  $q(x, y) = q(y-x)$ . Για παράδειγμα:  $Q(x, \cdot) = \mathcal{N}_d(x, \Sigma)$ , οπότε  $x_{n+1} \sim \mathcal{N}_d(x_n, \Sigma)$ .
3. Ανεξάρτητες προσομοιώσεις. Εδώ η  $q(x, y) = q(y)$ , δηλαδή ο  $Q(x, \cdot)$  είναι ανεξάρτητος του  $x$ . Αν και η επιλογή μιας τέτοιας συνάρτησης-πρότασης έχει πολλές ευκολίες, μπορεί να εφαρμοστεί μόνο σε συγκεκριμένες περιπτώσεις, καθώς θα πρέπει η κατανομή-πρόταση να αποτελεί μία σχετικά καλή προσέγγιση της συνάρτησης-στόχου.

**Παρατήρηση** Σχετικά με τον ΜΗ τυχαίου περιπάτου προκύπτει το εξής ζήτημα: Ας υποθέσουμε για απλότητα πως  $d = 1$ , και η  $q(x, y)$  είναι η πυκνότητα της  $\mathcal{N}(x, \sigma^2)$ . Δηλαδή, όταν  $X_n = x$  θα είναι  $Y_{n+1} \sim \mathcal{N}(x, \sigma^2)$ . Θα πρέπει στη συνέχεια να καθορίσουμε την κατανομή (διαλέγοντας κατάλληλα το  $\sigma^2$ ). Το πρόβλημα έγκειται στην επιλογή της διασποράς  $\sigma^2$ . Αρχικά θα πρέπει να έχουμε κάποιο κριτήριο αποδοτικότητας. Για παράδειγμα θα ήταν επιθυμητό ο αλγόριθμος να συγκλίνει ταχύτατα στη στάσιμη κατανομή (κατανομή-στόχο)  $\pi^*$ .

Είναι διαισθητικά άμεσο πως για πολύ μικρό  $\sigma^2$  ο αλγόριθμος θα προτείνει πολύ μικρά άλματα τα οποία θα είναι σχεδόν όλα αποδεκτά (καθώς η πιθανότητα αποδοχής  $a$  θα είναι κατά προσέγγιση κοντά στο 1 λόγω της συνέχειας των  $\pi$  και  $q$ ). Οπότε, η σύγκλιση θα είναι πολύ αργή καθώς θα απαιτείται πολύς χρόνος μέχρι ο αλγόριθμος να εξερευνήσει το χώρο. Από την άλλη πλευρά, αν το  $\sigma^2$  επιλεγεί πολύ μεγάλο, τότε ο αλγόριθμος θα πραγματοποιεί πολύ συχνά μεγάλα άλματα σε περιοχές όπου το  $\pi$  θα είναι πολύ μικρό. Επομένως θα παρουσιάσει μεγάλο ποσοστό απόρριψης και θα έτσι η αλυσίδα θα παραμείνει σταθερή για μεγάλα χρονικά διαστήματα. Εύλογο είναι λοιπόν το ερώτημα αν υπάρχει κάποια ας πούμε βέλτιστη επιλογή για το  $\sigma^2$ , τόσο σε αυτό το συγκεκριμένο παράδειγμα, όσο και γενικότερα.

Η βελτιστότητα-αποδοτικότητα είναι ένα δημοφιλές ζήτημα στις μέρες για το οποίο γίνεται λόγος στα [38], [3]. Η έννοια αποδοτικότητας για την περιοχή αυτή σχετίζεται με ταχύτητα σύγκλισης και εκφράζεται μέσω των διαφόρων ποσοτήτων-παραμέτρων του αλγορίθμου ΜΗ. Συγκεκριμένα, οι Roberts, Rosenthal (2001) στο [37] καταλήγουν πως η βέλτιστη πιθανότητα αποδοχής θα πρέπει να είναι κοντά στο 0.44 όταν  $d = 1$ . Ο κανόνας αυτός βασίζεται στην ασυμπτωτική συμπεριφορά ενός κριτηρίου αποδοτικότητας που θεώρησαν οι συγγραφείς. Επίσης, για πολυδιάστατα μοντέλα οι συγγραφείς προτείνουν να ρυθμίζεται η πιθανότητα αποδοχής κοντά στο 1/4 (συγκεκριμένα στο 0.234).

Τέλος, η επιλογή ανεξάρτητης συνάρτησης-πρότασης

Το παράδειγμα που ακολουθεί, είναι κατάλληλο ώστε να παρουσιαστούν τα διάφορα προβλήματα που θα πρέπει να αντιμετωπιστούν σε μία κατασκευή του αλγορίθμου ΜΗ.

### Παράδειγμα.

Ας υποθέσουμε πως έχουμε το παρακάτω ολοκλήρωμα:

$$C = \int (1 + \log(1 + \sin^2(2x)) + \sin(4x))e^{-x^2} dx \quad (4.1.10)$$

Το πρόβλημα μπορεί να είναι είτε να θέλουμε να υπολογίσουμε το παραπάνω, είτε να θέλουμε να προσομοιώσουμε δείγμα από την πυκνότητα:

$$\pi^*(x) = \frac{(1 + \log(1 + \sin^2(2x)) + \sin(4x))e^{-x^2}}{C} = \frac{\pi(x)}{C} \quad (4.1.11)$$

Φυσικά, το  $C$  εκφράζεται και ως:

$$E_\mu\left[\frac{1}{\sqrt{\pi}}(1 + \log(1 + \sin^2(2X)) + \sin(4X))\right] = E_\mu[h(X)],$$

όπου  $\mu \sim \mathcal{N}(0, 1/2)$ . Από την τελευταία, μπορεί πολύ εύκολα κανείς να προσομοιώσει δείγμα, και έπειτα, χρησιμοποιώντας το νόμο μεγάλων αριθμών, να πάρει:

$$1/N \sum_{i=1}^N h(X_i) \rightarrow C.$$

Ωστόσο, για τα παραπάνω προβλήματα θα χρησιμοποιήσουμε τον αλγόριθμο MH. Το στήριγμα της συνάρτησης  $\pi$  είναι όλη η ευθεία και επομένως επιλέγοντας αλγόριθμο τυχαίου περιπάτου πάνω στην ευθεία, θα έχουμε καλύψει την ελάχιστη απαίτηση (4.1.9). Για το συγκεκριμένο παράδειγμα έχει επιλεγθεί ο:

$$Q(x, y) \sim U(x - \delta, x + \delta)$$

Μία άλλη επιλογή θα ήταν ο:  $Q_2(x, y) \sim \mathcal{N}(x, \sigma^2)$ . Επίσης, μπορούμε να παρατηρήσουμε από τον τύπο της  $\pi$ , πως συγκεντρώνεται κοντά στο μηδέν ενώ χάνεται σχετικά γρήγορα. Συγκεκριμένα, η μάζα για τα  $x \geq 3$  και  $x \leq -3$  είναι σχεδόν 0. Είναι εύλογο λοιπόν να επιλέξουμε μία οικογένεια κατανομών και όχι απλά μία κατανομή, καθώς στην πρώτη μπορούμε να ρυθμίζουμε τα άλματα που θα κάνει κάθε φορά. (πχ ρυθμίζοντας το  $\delta$  στη  $Q$  ή το  $\sigma$  αντίστοιχα στη  $Q_2$ ). Θα δώσουμε έμφαση με το παράδειγμα αυτό, σε 4 παραμέτρους: Στην πιθανότητα αποδοχής, στο πλήθος επαναλήψεων, στο Burn in, και στην επιλογή του αρχικού σημείου. Θα εξετάσουμε τα εξής 2 χαρακτηριστικά:

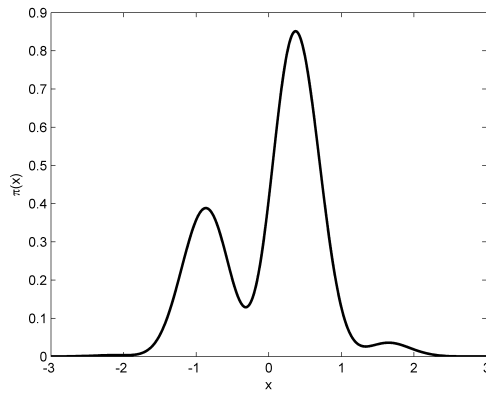
- 1 Πόσο καλή προσαρμογή έχει το ιστόγραμμα της αλυσίδας στην πραγματική πυκνότητα  $\pi$ , και
- 2 Τη σύγκλιση του εργοδικού μέσου:  $\frac{1}{N+1} \sum_{i=m}^{m+N} X_i \rightarrow \int x d\pi^*(x)$ , όπου το  $m$  είναι η περίοδος burn in.

Θα δούμε πως όλες οι παραπάνω παράμετροι επηρεάζουν αυτά τα αποτελέσματα.

Πρώτα από όλα, η πυκνότητα  $\pi^*(x)$  που θέλουμε να προσομοιώσουμε παρουσιάζεται στο γράφημα (;;). Με τον αλγόριθμο MH θα παραχθεί μία ακολουθία  $X^{(0)}, \dots, X^{(N)}$  όπου στο βήμα  $n$ , ( $0 \leq n \leq N$ ) προσομοιώνουμε ένα  $y$  από την  $Q(x^{(n)}, y) \sim U(x^{(n)} - \delta, x^{(n)} + \delta)$  ως η υποψήφια νέα κατάσταση  $x^{(n+1)}$  η οποία θα γίνει δεκτή με πιθανότητα:

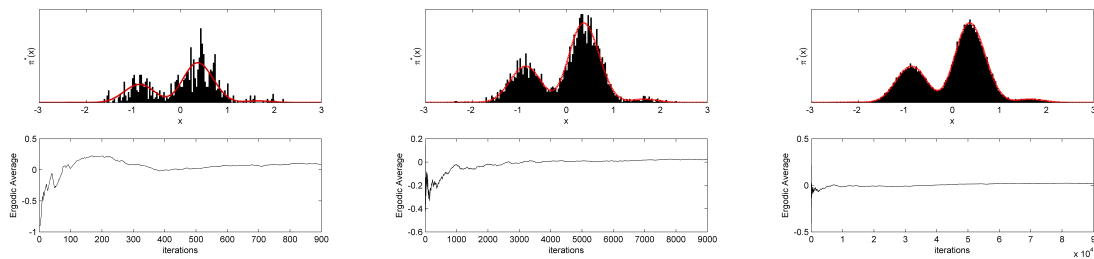
$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\},$$

(καθώς η  $Q$  είναι συμμετρική :  $Q(x, y) = Q(y, x)$ ) διαφορετικά θα απορριφθεί, και η αλυσίδα παραμένει στην ίδια κατάσταση (δηλ  $x^{(n+1)} = x^{(n)}$ ).



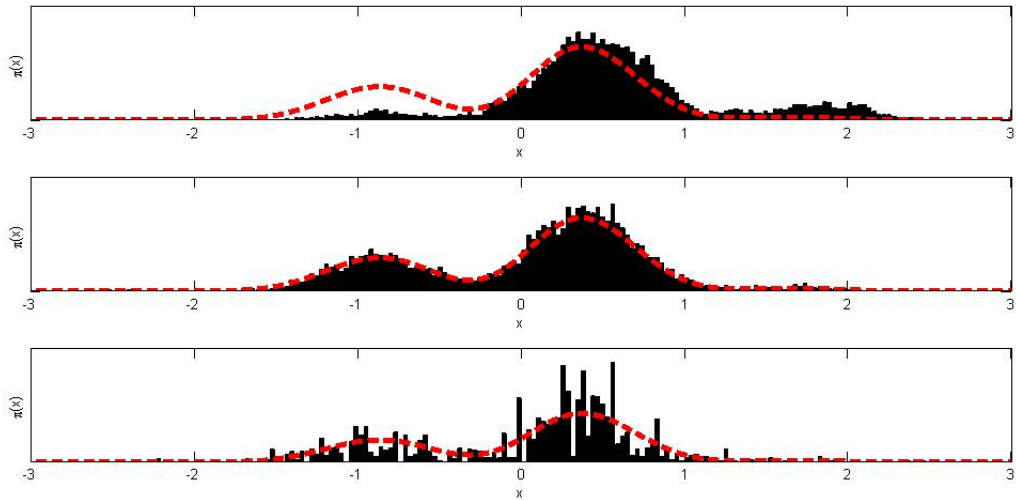
Σχήμα 4.1: Το γράφημα της πυκνότητας  $\pi^*$

Αρχικά θεωρούμε ένα λογικό σημείο εκκίνησης (εκεί που συγκεντρώνεται περισσότερη μάζα), όπως για παράδειγμα το  $x^{(0)} = 0$ . Στο γράφημα (4.2) έχουμε 3 διαφορετικές πραγματοποιήσεις του αλγορίθμου. Συγκεκριμένα, μεταβάλλουμε το πλήθος επαναλήψεων και κρατάμε ενώ προσαρμόζουμε το  $\delta$  κοντά στο 1.8, για το οποίο έχουμε πιθανότητα αποδοχής 0.44 και επιλέγουμε ως περίοδο burn in 10 %. Επίσης στο σχήμα αυτό βλέπουμε και τη σύγκλιση του εργοδικού μέσου για τις 3 αυτές περιπτώσεις.



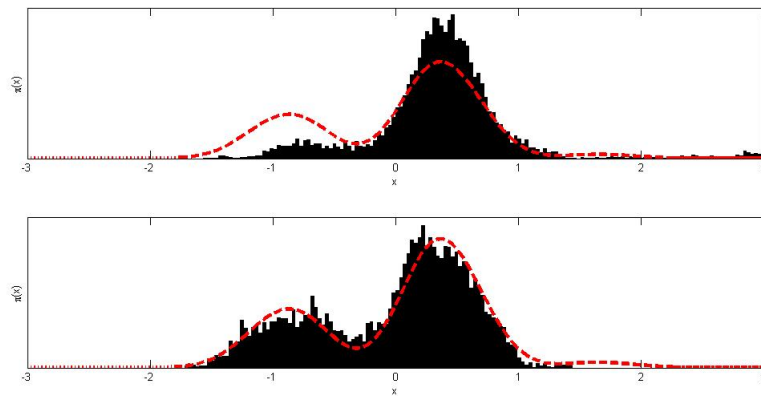
Σχήμα 4.2: Προσαρμογή της πυκνότητας που επάγει το στάσιμο μέτρο της αλυσίδας MH στην πυκνότητα  $\pi^*$  για διαφορετικά πλήθη επαναλήψεων:  $N=1000$ ,  $N=10000$ ,  $N=10000$  που αντιστοιχούν στα παραπάνω γραφήματα και στα κάτω γραφήματα έχουμε τη σύγκλιση του εργοδικού μέσου

Ένα κακό σημείο εκκίνησης (πχ  $x^{(0)} = 3$ ), σε συνδυασμό με μία κακή επιλογή για το  $\delta$  μπορεί να οδηγήσει σε μεροληπτικά σφάλματα ακόμα και για πολύ μεγάλα δείγματα ( $N = 10000$ ). Όπως επισημάνθηκε, με ένα πολύ μικρό  $\delta$ , οι προτεινόμενες μεταβάσεις θα βρίσκονται πολύ κοντά μεταξύ τους και επομένως θα έχουμε πολύ μεγάλη πιθανότητα αποδοχής. Το αποτέλεσμα θα είναι, η αλυσίδα να χρειάζεται πολύ χρόνο για να εξερευνησει όλο το χώρο (που στην προκειμένη είναι ουσιαστικά το  $(-3, 3)$ ). Από την άλλη, για μεγάλο  $\delta$ , θα έχουμε πολύ μεγάλα άλματα και έτσι, πολύ μικρή πιθανότητα αποδοχής. Το αποτέλεσμα θα είναι να μένει η αλυσίδα για πολύ χρόνο στην ίδια κατάσταση. Οι δύο αυτές προβληματικές περιπτώσεις παρουσιάζονται στο σχήμα (4.3), στο 1ο και 3ο γράφημα αντίστοιχα. Στο 2ο γράφημα του σχήματος (4.3), έχει επιλεγεί  $\delta = 2$ . Για την επιλογή αυτή έχουμε πιθανότητα αποδοχής  $\alpha = 0.44$ , ενώ για τις άλλες 2 περιπτώσεις είναι  $\alpha = 0.94$  και  $\alpha = 0.042$  αντίστοιχα.



Σχήμα 4.3: Προσαρμογή της πυκνότητας που επάγει το στάσιμο μέτρο της αλυσίδας MH στην πυκνότητα  $\pi^*$  όπου  $N = 10000$  ενώ  $\delta = 0.1, \delta = 2$  και  $\delta = 20$  αντίστοιχα.

Τέλος, στο σχήμα 4.4 βλέπουμε πως η περίοδος burn in ( $m$ ), επηρεάζει σε μεγάλο βαθμό τη σύγκλιση ακόμα και έχουμε ένα σχετικά μεγάλο πλήθος επαναλήψεων για το συγκεκριμένο πρόβλημα. Αυτό συμβαίνει διότι αν το σημείο εκκίνησης δεν είναι ιδανικό, και δεν επιτρέψουμε μεγάλα άλματα, τότε οι πρώτες επαναλήψεις, δεν θα δώσουν σωστά αποτελέσματα και έτσι, αν τις συμπεριλάβουμε θα παρουσιαστεί μεροληψία. Η κατάσταση βελτιώνεται όταν το  $X^{(m)}$  βρίσκεται κοντά στη  $\pi^*$ .



Σχήμα 4.4: Προσαρμογή της πυκνότητας που επάγει το στάσιμο μέτρο της αλυσίδας MH στην πυκνότητα  $\pi^*$  όπου  $N = 10000$  ενώ  $\delta = 0.1, \delta = 2$  και  $\delta = 20$  αντίστοιχα. Το αρχικό σημείο είναι το  $x^{(0)} = 3$ .

Οπότε, ολοκληρώνοντας το παράδειγμα αυτό, θεωρούμε αποδεκτό το δείγμα  $x^{(m)}, \dots, x^{(N)}$  που προσομοιώθηκε από την παραπάνω αλυσίδα με  $N = 10000$ ,  $\delta = 1.8$  και  $m = 1000$ , ως αντιπροσωπευτικό δείγμα από την  $\pi^*$  και υπολογίσαμε πως  $E_{\pi^*}[h(X)] = 1.061$ .

### 4.1.2 Ο δειγματολήπτης Gibbs

Η ονομασία οφείλεται στον J.W. Gibbs, αν και η περιγραφή του αλγορίθμου έγινε αργότερα από τους αδερφούς Stuard, Donald Geman (1984), (1990). Στη γενική του μορφή, ο αλγόριθμος αυτός είναι μία ειδική περίπτωση του MH και εφαρμόζεται σε χώρους μεγάλης διάστασης όταν η από κοινού κατανομή είναι άγνωστη ή όταν είναι ανέφικτη η προσομοίωση της, αλλά οι δεσμευμένες κατανομές είναι γνωστές και εύκολα προσομοιώσιμες.

Υποθέτουμε ότι ο χώρος  $E = E_1 \times E_2 \times \dots \times E_k =$  έχει διάσταση  $d$ , όπου  $E_i \subset \mathbb{R}^{d_i}$  με  $d_1 + \dots + d_p = d$ . Έστω  $\mathbf{X} \in \mathbf{E}$  ένα τυχαίο διάνυσμα με πυκνότητα  $\pi$ . Το πρόβλημα είναι να προσομοιώσουμε την πυκνότητα  $\pi$ .

Μπορούμε να γράψουμε  $\mathbf{X} = (X_1, \dots, X_k)$ , όπου το τ.δ.  $X_i$  είναι η προβολή του  $\mathbf{X}$  στον  $E_i$ , αλλά θα αναφερόμαστε σε αυτό ως η  $i$  συνιστώσα. (θα αποφεύγουμε τα έντονα γράμματα για τις συνιστώσες, παρόλο που είναι διανύσματα).

Έστω  $X^{(-i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$  (δηλαδή χωρίς την  $i$  συνιστώσα), και κρατάμε όμοιο συμβολισμό και για τα  $E^{(-i)}$ . Ακόμη, με  $\pi(x)_{(-i,s)}$  συμβολίζεται η από κοινού πυκνότητα με σταθερές όλες τις συνιστώσες εκτός από την  $i$ -οστή. Οπότε  $x = (x_i, x^{(-i)})$  και υποθέτουμε (για απλότητα) πως για κάθε  $x \in E$  είναι  $\pi(x) > 0$  Η υπόθεση αυτή είναι ουσιαστική όπως θα δούμε στη συνέχεια).

Η περιθώρια πυκνότητα του  $X^{(-i)}$  είναι:

$$\pi_{(-i)}(x^{(-i)}) = \int_{E_i} \pi(x)_{(-i,s)} ds, \quad x^{(-i)} \in E^{(-i)}$$

και η δεσμευμένη πυκνότητα του  $X_i | X^{(-i)} = x^{(-i)}$  είναι:

$$\pi_i(x_i | x^{(-i)}) = \pi(x) / \pi_{(-i)}(x^{(-i)}), \quad x_i \in E_i$$

Έστω:

$$S_{x,A}^{(-i)} = \{y \in X : y_j = x_j \text{ για } j \neq i, y_i \in A\}.$$

Τότε

$$P_i(x_i, S_{x,A}^{(-i)}) = \int_A \pi_i(x_i | x^{(-i)}) ds = \frac{\int_A \pi(x)_{(-i,s)} ds}{\int \pi(x)_{(-i,s)} ds}$$

όπου  $\pi(x)_{(-i,s)} = \pi(x_1, \dots, x_{i-1}, s, x_{i+1}, \dots, x_d)$

#### Παρατηρήσεις

1. Η  $P_i$  λέγεται πλήρης δεσμευμένη κατανομή full conditional για τη συνιστώσα  $i$  και ένα βασικό χαρακτηριστικό του Gibbs είναι πως οι παραπάνω πλήρης δεσμευμένες πυκνότητες ( $\pi_i(\cdot | x^{(-i)})$ ) είναι οι μόνες πυκνότητες που χρειάζονται για την προσομοίωση. Οπότε, ακόμα και σε χώρους μεγάλης διάστασης, η προσομοίωση μπορεί να αφορά χώρους χαμηλής διάστασης (πχ τα  $d_i$  είναι μικρά) το οποίο συχνά είναι πλεονέκτημα.

2. Αποδεικνύεται (από τον ορισμό της δεσμευμένης πυκνότητας) ότι η αλυσίδα με πυρήνα το  $P_i$  είναι αντιστρέψιμη ως προς το  $\pi$ . Επομένως, το  $P_i$  έχει την  $\pi$  ως στάσιμη κατανομή.
3. Ακόμη, μπορούμε να δούμε το  $P_i$  ως ειδική περίπτωση του  $MH$  με πιθανότητα  $a(x, y) = 1$ . Συγκεκριμένα, μόλις η  $i$  προς ανανέωση συνιστώσα σταθεροποιηθεί, την κίνηση που πραγματοποιεί ο Gibbs sampler μπορούμε να τη δούμε ως μία κίνηση του αλγορίθμου  $MH$  στο  $E_i$ , όπου η κατανομή-στόχος και η κατανομή-πρόταση είναι η  $\pi_i(\cdot|x^{(-i)})$  και έτσι  $a(x, y) = 1$ .

Ο Gibbs sampler παράγει μία ακολουθία  $\mathbf{X}^{(n)} = (X_1^{(n)}, \dots, X_d^{(n)}) \in E, n \geq 0$  όπου στο βήμα  $m$  μόνο μία συνιστώσα ανανεώνεται, έστω η  $i = i_m$ , όπου σύμφωνα με τον παραπάνω κανόνα η επόμενη κατάσταση για τη συνιστώσα  $i$ , δηλαδή η  $X_i^{(n+1)}$ , παράγεται από την  $\pi(-i, s)(\cdot)$  όταν  $Q_i^{(n)} = x$ .

Για την ακολουθία  $(i(n))_{n \geq 0}$  υπάρχουν 2 τρόποι ώστε αυτή να καθοριστεί. Ο συστηματικός (ή κυκλικός) και ο τυχαίος. Στον κυκλικό επιλέγουμε τις συνιστώσες κυκλικά:  $1, 2, \dots, d, 1, 2, \dots, d, \dots$  ενώ στο τυχαίο, κάθε συνιστώσα για το βήμα  $n$  επιλέγεται τυχαία (συχνά ομοιόμορφα από τις  $d$  συνιστώσες). Εν αντιθέσει με το συστηματικό τρόπο, στην περίπτωση της τυχαίας επιλογής, η ακολουθία  $(\mathbf{X}^{(n)})_{n \geq 0}$  είναι Μ.Α. Ωστόσο, με τον τρόπο αυτό, χάνεται ενδεχομένως σημαντικός χρόνος στην προσομοίωση (ειδικά όταν το  $d$  είναι μεγάλο και όταν υπάρχει σημαντική διαφορά στη μεταβλητότητα των συνιστωσών). Για το ζήτημα αυτό γίνεται συζήτηση στο LIU, J. S., WONG, W. H. and KONG, A. (1995).

Αν υποθέσουμε πως είναι δυνατή η προσομοίωση από όλες τις πλήρη-δεσμευμένες κατανομές, τότε οι αλγόριθμοι Gibbs που αντιστοιχούν σε αυτές τις προσομοιώσεις δίνονται από την ακόλουθη περιγραφή για τη μετάβαση από την κατάσταση  $\mathbf{X}^{(n)}$  στη  $\mathbf{X}^{(n+1)}$ :

### (Κυκλικός) Αλγόριθμος Gibbs Sampler.

Στο βήμα  $n$  του αλγορίθμου Gibbs, για  $i = 1, \dots, d$ , δοθέντος του  $(x_1^{(n)}, \dots, x_d^{(n)})$

1. Προσομοίωση της:

$$X_1^{(n+1)} \sim \pi_1(\cdot|x^{(-1)})$$

2. Προσομοίωση της:

$$X_2^{(n+1)} \sim \pi_2(\cdot|x^{(-2)})$$

...

- d. Προσομοίωση της:

$$X_d^{(n+1)} \sim \pi_d(\cdot|x^{(-d)})$$

**(Τυχαίος) Αλγόριθμος Gibbs Sampler.** Στο βήμα  $n$  του αλγορίθμου Gibbs, για  $i = 1, \dots, d$ , δοθέντος του  $(x_1^{(n)}, \dots, x_d^{(n)})$

1. Προσομοίωση μιας:

$$I \sim U(\{1, \dots, d\})$$



2. Προσομοίωση της:

$$X_I^{(n+1)} \sim \pi_I(\cdot | x^{(-I)})$$

### 4.1.3 Ο υβριδικός δειγματολήπτης Gibbs

Η περίπτωση αυτή αποτελεί επέκταση του Gibbs sampler. Στον τελευταίο, μία συνιστώσα σε κάθε βήμα ανανεώνεται και η κατανομή-στόχος σε κάθε βήμα είναι η πλήρης δεσμευμένη κατανομή της συνιστώσας που πρόκειται να ανανεωθεί. Επομένως, σε κάθε βήμα χρειάζεται να προσομοιωθούν οι πλήρεις-δεσμευμένες κατανομές  $\Pi_i$ . Ωστόσο, αν κάποια από αυτές έχει δυσκολίες στην προσομοίωση, τότε μία τεχνική προς την επίλυση του προβλήματος αυτού είναι να εφαρμοστεί αλγόριθμος MH για την προσομοίωση της. Δηλαδή, αντί της απευθείας προσομοίωσης της:

$$X_i^{(n+1)} \sim \pi_i(x_i | x_1^{(n+1)}, \dots, x_{i-1}^{(n+1)}, x_{i+1}^{(n)}, \dots, x_d^{(n)}),$$

εφαρμόζουμε ένα βήμα από τον MH και όποτε είναι δυνατόν, ο αλγόριθμος αυτός να είναι τυχαίου περιπάτου με κέντρο το  $x_i^{(n)}$ .

Επομένως, η διαφορά είναι πως η ανανέωση από  $x_i$  σε  $y_i$  γίνεται με μία γενική κατανομή-πρόταση:  $q_{(x_i, x^{(-i)})}(x_i, \cdot)$  αντί της  $\pi_i(\cdot | x^{(-i)})$  (όπου  $q_{(a, x^{(-i)})}(a, \cdot) = q_i((x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_d))$  ή με το συμβολισμό της δεσμευμένης πυκνότητας αντί της πυκνότητας πυρήνα,  $q_{(a, x^{(-i)})}(a, \cdot) = q_i(\cdot | (x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_d))$ ). Ο λόγος μέσα στην πιθανότητα αποδοχής γίνεται:

$$r_x^{(i)}(x_i, y_i) = \frac{\pi_i(y_i | x^{(-i)}) q_{(y_i, x^{(-i)})}(y_i, x_i)}{\pi_i(x_i | x^{(-i)}) q_{(x_i, x^{(-i)})}(x_i, y_i)} \quad (4.1.12)$$

Η πλήρης περιγραφή του αλγορίθμου είναι η εξής:

**Αλγόριθμος Hybrid MH.** Στο βήμα  $n$  του αλγορίθμου Gibbs, για  $i = 1, \dots, d$ , δοθέντος του  $(x_1^{(n+1)}, \dots, x_{i-1}^{(n+1)}, x_{i+1}^{(n)}, \dots, x_d^{(n)})$

1. Προσομοίωση της:

$$y_i \sim q_i(x_1^{(n+1)}, \dots, x_i^{(n)}, x_{i+1}^{(n)}, \dots, x_d^{(n)}, y_i)$$

2. Όπου:

$$x_i^{(n+1)} = \begin{cases} y_i & \text{με πιθανότητα } r_x^{(i)}(x_i, y_i) \\ x_i^{(n)} & \text{με πιθανότητα } 1 - r_x^{(i)}(x_i, y_i) \end{cases} \quad (4.1.13)$$

Στην περίπτωση τυχαίου περιπάτου:  $q_{(x_i, x^{(-i)})}(x_i, y_i) = f_i(y_i - x_i)$  έχουμε σημαντική απλοποίηση για τον λόγο που εμφανίζεται μέσα στην πιθανότητα αποδοχής. Συγκεκριμένα:

$$r_x^{(i)}(x_i, y_i) = \frac{\pi_i(y_i | x^{(-i)})}{\pi_i(x_i | x^{(-i)})} \quad (4.1.14)$$

όταν  $f_i(y_i - x_i) > 0$ .

Ο Gibbs Sampler μπορεί να θεωρηθεί ειδική περίπτωση του παραπάνω όταν  $q_{(x_i, x^{(-i)})}(x_i, y_i) = \pi_{(x_i, x^{(-i)})}(x_i, y_i)$  όπου προκύπτει:  $a(x, y) = 1$ . (όπου  $a(x, y) = \min\{1, r_x^{(i)}(x_i, y_i)\}$ .)

Όπως και στην περίπτωση του Gibbs Sampler, το  $\pi$  είναι στάσιμη κατανομή. Αυτό μπορούμε να το δούμε και ως εξής:

Θεωρούμε μία ανανέωση για την  $i$ -οστή συνιστώσα για σταθερό  $x^{(-i)}$ . Με παρόμοια επιχειρήματα όπως αυτά στην περίπτωση ( $\cdot$ ), η ανανέωση της  $i$ -οστή συνιστώσα ικανοποιεί τις εξισώσεις λεπτομερούς ισορροπίας και άρα η πλήρης δεσμευμένη πυκνότητα  $\pi_i(\cdot|x^{(-i)})$  είναι στάσιμη. Άρα και η  $\pi$  είναι στάσιμη.

Ιδιότητες όπως η διαχωρισιμότητα και η απεριοδικότητα θα πρέπει να αποδειχθούν μέσα από τη δομή του παραπάνω αλγόριθμου που έχει κατασκευαστεί για το συγκεκριμένο πρόβλημα που έχουμε κάθε φορά.

Τέλος, για τους αλγόριθμους MH δεν ισχύει ο κανόνας για τη βέλτιστη πιθανότητα αποδοχής που αναφέρθηκε στα προηγούμενα.

## 4.2 Ασυμπτωτική συμπεριφορά

Στην ενότητα αυτή θα παρουσιαστούν ικανές συνθήκες για τους αλγόριθμους MH και Gibbs.

### Για τον αλγόριθμο MH

Ξεκινάμε με την πιο βασική απαίτηση: η κατανομή-στόχος να θα πρέπει να είναι στάσιμη κατανομή για την αλυσίδα που παράγεται από τον αλγόριθμο MH.

Αυτό όμως έπεται άμεσα από την Πρόταση 4.1. αλλά και από το Θεώρημα 3.4. Οπότε:

**Θεώρημα 4.1.** Για οποιαδήποτε κατανομή (-πρόταση)  $q$  (με στήριγμα που περιλαμβάνει το στήριγμα της  $\pi^*$ ), το  $\pi^*$  είναι στάσιμο μέτρο για την αλυσίδα  $(X^{(n)})$  που παράγεται από τον αλγόριθμο MH.

Προκειμένου να εφαρμόσουμε το Κεντρικό Θεώρημα Σύγκλισης M.A. (Θ.3.8), για όλο το χώρο  $\mathcal{X}$ , θα πρέπει η αλυσίδα να είναι  $\phi$ -αδιαχώριστη, Harris-επαναληπτική, και απεριοδική. Υπάρχουν αρκετές ικανές συνθήκες προκειμένου ο αλγόριθμος MH να είναι  $\phi$ -αδιαχώριστος. (Roberts and Tweedie (1996), Mengersen and Tweedie (1996)). Ωστόσο, η μ.α. του αλγόριθμου MH, είναι σχεδόν πάντα  $\pi^*$ -αδιαχώριστη καθώς, για σχεδόν οποιαδήποτε κατανομή-πρότασης  $q$  το  $\pi^*$  είναι στάσιμη κατανομή. Θα δούμε πως αρκεί: αν για κάθε  $y \in X$  με  $\pi^*(y) > 0$  έπεται ότι  $q(x, y) > 0$  για οποιοδήποτε  $x \in X$ .

Για παράδειγμα, η διαχωρισιμότητα της αλυσίδας αυτής, εξασφαλίζεται αν η πυκνότητα  $q$  είναι θετική, δηλαδή αν:

$$q(x, y) > 0 \text{ για κάθε } x, y \in \mathcal{E}, \quad (4.2.1)$$

αφού αν ικανοποιείται η παραπάνω, έπεται πως κάθε υποσύνολο του στηρίγματος  $\mathcal{E}$ , με θετικό μέτρο Lebesgue, είναι προσιτό σε ένα βήμα. Επειδή τώρα η πυκνότητα  $\pi^*$  είναι στάσιμη, η αλυσίδα είναι θετική εξ-ορισμού. Επομένως είναι και επαναληπτική.

Με το παραπάνω εξασφαλίζεται η διαχωρισιμότητα της αλυσίδας που παράγεται από τον MH για σχεδόν οποιαδήποτε κατανομή-πρόταση  $q$  και όπως δούμε στη συνέχεια, για την Harris-επαναληπτικότητα, η διαχωρισιμότητα αρκεί. Επομένως, ζητάμε μόνο συνθήκες για την απεριοδικότητα.

Μία ικανή συνθήκη για την απεριοδικότητα της αλυσίδας του αλγόριθμου MH,  $(X^{(n)})$ , είναι να επιτρέπονται από τον αλγόριθμο γεγονότα όπως  $\{X_{n+1} = X_n\}$ , δηλαδή, η πιθανότητα τέτοιων ενδεχομένων να είναι μη μηδενική και άρα:

$$\mathbb{P}[\pi^*(X_n)q(X_n, Y) \leq \pi^*(Y)q(Y, X_n)] < 1. \quad (4.2.2)$$

Για την Harris-επαναληπτικότητα έχουμε το ακόλουθο:

**Λήμμα 4.1** Αν η αλυσίδα  $(X^{(n)})$  που παράγεται από τον αλγόριθμο MH είναι  $\pi^*$ -αδιαχώριστη, τότε είναι και Harris-επαναληπτική.

## Ασυμπτωτική συμπεριφορά

Επομένως, συνοψίζοντας τα παραπάνω, έχουμε ήδη δείξει το επόμενο Θεώρημα σύγκλισης:

**Θεώρημα 4.2** Θεώρημα Σύγκλισης του αλγορίθμου ΜΗ. Αν η αλυσίδα  $(X^n)$  που παράγεται από τον ΜΗ είναι  $\pi^*$ -αδιαχώριστη, τότε:

(i) Αν  $f \in L^1(\pi^*)$ , τότε:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(X^i) = \int f(x) \pi^*(x) dx \quad \text{σ.π.} - \pi^*$$

(ii) Αν επιπλέον είναι και απεριοδική, τότε:

$$\mathbb{P}_\mu(X^{(n)} \in A) \rightarrow \Pi^*(A) \quad \text{καθώς } n \rightarrow \infty.$$

για οποιαδήποτε αρχική κατανομή  $\mu$  και  $A \in \mathcal{X}$ .

Επίσης, από την προηγούμενη συζήτηση καταλήξαμε και στο:

**Πόρισμα.** Αν η αλυσίδα που παράγεται από τον ΜΗ με συνάρτηση-πρόταση τη  $g$  ικανοποιεί τις συνθήκες (;), (4.2.2) τότε ισχύει το συμπέρασμα του παραπάνω θεωρήματος.

### Για τον αλγόριθμο Gibbs sampler

Είδαμε πως ο Gibbs sampler παράγει μία ακολουθία  $\mathbf{X}^{(n)} = (X_1^{(n)}, \dots, X_d^{(n)}) \in E, n \geq 0$  όπου στο βήμα  $n$  μόνο μία συνιστώσα ανανεώνεται και υπάρχουν 2 τρόποι για να γίνει αυτό. Στις επόμενες γραμμές περιγράφεται πως κάτω από πολύ ασθενείς συνθήκες το Κεντρικό Θεώρημα Σύγκλισης Μ.Α. εφαρμόζεται και για τις 2 αυτές περιπτώσεις.

Για την περίπτωση του κυκλικού Gibbs Sampler, επειδή:

$$X_i^{(n+1)} | (X_1^{(0)}, X_2^{(0)}, \dots, X_{i-1}^{(n+1)}) \sim \pi_i(\cdot | X_1^{(n+1)}, \dots, X_{i-1}^{(n+1)}, X_{i+1}^{(n)}, \dots, X_d^{(n)})$$

(με τις προφανείς μετατροπές αν  $i = 1$ , έχουμε πως οι  $X_i^{(n+1)}$  και  $(X_1^{(0)}, X_2^{(0)}, \dots, X_{i-1}^{(n+1)})$  είναι ανεξάρτητες δοθέντος της  $(X_{i+1}^{(n)}, \dots, X_d^{(n)}, X_1^{(n+1)}, \dots, X_{i-1}^{(n+1)})$ ). Επομένως η:

$$(X_1^{(0)}, X_2^{(0)}, \dots, X_d^{(0)}, X_1^{(1)}, X_2^{(1)}, \dots, X_d^{(1)}, \dots)$$

είναι μία  $(d-1)$ -τάξης Μ.Α. (που μπορεί να είναι μη-ομογενής). Έπεται λοιπόν από τη θεωρία στη ενότητα 3.1, πως η  $(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \dots)$  είναι μία ομογενής Μ.Α.

Και στις παραπάνω δύο περιπτώσεις, έχουμε το ακόλουθο αποτέλεσμα:

**Θεώρημα.** Η  $\Pi^*$  είναι στάσιμη κατανομή για την  $(\mathbf{X}^{(n)})_{n \geq 0}$

Η αλυσίδα που παράγεται από τον κυκλικό αλγόριθμο δεν είναι αντιστρέψιμη (εν αντιθέσει με την αλυσίδα από τον τυχαίο). Ωστόσο, κάθε ανανέωση είναι κατά κάποιον τρόπο αντιστρέψιμη, καθώς για οποιοδήποτε  $(x_1, \dots, x_d)$  και  $y_i$ , θέτοντας  $y = (x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_d)$  παίρνουμε τις εξισώσεις λεπτομερούς ισορροπίας:

$$\pi(x) \pi_i(y_i | x^{(-i)}) = \pi(x) \frac{\pi(y)}{\pi_{(-i)}(x^{(-i)})} = \pi(y) \pi_i(x_i | x^{(-i)})$$

## Ασυμπτωτική συμπεριφορά

Από το παραπάνω, επιβεβαιώνεται πως η αλυσίδα αυτή έχει το πως στάσιμη κατανομή.

Η συνθήκη

$$\pi(x) > 0 \text{ για κάθε } x \in E \quad (4.2.3)$$

που διατυπώθηκε στην αρχή, είναι ικανή συνθήκη ώστε ο κυκλικός αλγόριθμος Gibbs να είναι Harris-επαναληπτικός και αperiοδικός. Επομένως, το Κεντρικό Θεώρημα σύγκλισης εφαρμόζεται.

## Κεφάλαιο 5

### Στοιχεία Θεωρίας Αλγορίθμων

#### 5.1 Εισαγωγή

Στο κεφάλαιο αυτό διατυπώνονται αρχικά τα βασικότερα εργαλεία στα οποία στηρίζεται η θεωρία σύγκλισης του αλγορίθμου EM όπως θεμελιώθηκε από τον Wu (1983). Το πλαίσιο αυτό θα χρησιμοποιηθεί και στο κεφάλαιο 7. Στην ενότητα 5.2 παρουσιάζεται ο αλγόριθμος EM και γίνεται μία αναλυτική συζήτηση των χαρακτηριστικών του. Στη συνέχεια, η ενότητα 5.3 αφορά τη σύγκλιση του EM. Σε αυτήν, θα παρουσιαστεί το πλαίσιο μελέτης που ανέπτυξε ο Wu (1983) το οποίο αποτελεί μέχρι σήμερα το βασικό πλαίσιο για τη μελέτη της σύγκλισης του αλγορίθμου EM.

## 5.2 Βασικές έννοιες και Κεντρικό Θεώρημα Σύγκλισης

Ένα βασικό μαθηματικό αντικείμενο που παρουσιάζεται στην ενότητα αυτή είναι η έννοια του αλγορίθμου (ή επαναληπτική διαδικασία). Πρόκειται για μία διαδικασία, κατά την οποία, από ένα σημείο  $x^{(k)}$ , παράγεται το επόμενο σημείο  $x^{(k+1)}$ . Έτσι, ξεκινώντας με ένα αρχικό σημείο  $x^{(0)}$ , μπορεί να υπολογιστεί μία ακολουθία  $\{x^{(k)}\}_{k \geq 0}$ . Γενικά, θα χρησιμοποιούμε το συμβολισμό  $\{x^{(k)}\}_{k \geq 0}$  για να εκφράσουμε ότι πρόκειται για μία επαναληπτική διαδικασία σε αντίθεση με τη συνήθη συμβολισμό για τις ακολουθίες πραγματικών αριθμών. Θα ασχοληθούμε με αλγόριθμους οι οποίοι 'σύγκλινουν' σε κάποιο 'βέλτιστο' σημείο.

Το βέλτιστο σημείο μπορεί να είναι για παράδειγμα κάποιο στάσιμο σημείο ή κάποιο ελάχιστο (ολικό ή τοπικό) της αντικειμενικής συνάρτησης (η συνάρτηση ενδιαφέροντος). Σύγκλιση με την έννοια, ότι ο αλγόριθμος είτε υπολογίζει το σημείο αυτό, είτε το προσεγγίζει με την έννοια του ορίου. Στην περίπτωση αυτή, το όριο θα συμβολίζεται με  $x^\infty$ . Ειδικότερα, σε πολύ λίγα σενάρια (κυρίως σε περιπτώσεις γραμμικού ή τετραγωνικού (quadratic) προγραμματισμού) υπάρχει η δυνατότητα υπολογισμού του βέλτιστου σημείου με ακρίβεια.

Σε όσα ακολουθούν, θα διασαφηνιστεί τι θεωρείται σύγκλιση ενός αλγορίθμου και θα παρουσιάσουμε έναν τρόπο με τον οποίο θα μπορέσουμε να την εξασφαλίσουμε. Αρχικά θα ορίσουμε την έννοια του αλγορίθμου.

**Ορισμός 4.1** (Αλγόριθμος). Έστω  $X$  ένα σύνολο και  $x^{(0)} \in X$  ένα σημείο. Ένας (επαναληπτικός) αλγόριθμος,  $\mathcal{M}$ , με αρχικό σημείο το  $x^{(0)}$ , είναι μία (σημείο-σε-σύνολο) απεικόνιση  $\mathcal{M} : X \rightarrow X$  η οποία παράγει την ακολουθία  $\{x^{(k)}\}_{k \geq 1}$  σύμφωνα με:

$$x^{(k+1)} \in \mathcal{M}(x^{(k)}), \quad k = 0, 1, \dots \quad (5.2.1)$$

### Παρατηρήσεις

1. Μία σημείο-σε-σύνολο απεικόνιση (point-to-set map (p-t-s map))  $\phi$  από τον  $Q$  στον  $Y$  είναι μία απεικόνιση που στέλνει το σημείο  $x \in X$  στο  $\phi(x) \subset Y$ . Οπότε  $\phi : X \rightarrow 2^Y$ . Ωστόσο, γίνεται συχνά κατάχρηση του συμβολισμού γράφοντας  $\phi : X \rightarrow Y$  όταν είναι σαφές ότι η  $\phi$  είναι σημείο-σε-σύνολο απεικόνιση.
2. Με αυτόν τον ορισμό είναι ξεκάθαρο πως η παραγόμενη ακολουθία  $\{x^{(k)}\}_{k \geq 0}$  δεν είναι δυνατό να προβλεφθεί απλά γνωρίζοντας το αρχικό σημείο.
3. Ένας αλγόριθμος συγκλίνει αν η ακολουθία  $\{x^{(k)}\}_{k \geq 0}$  που παράγεται από κάποιο σημείο  $x^0$ , έχει σημείο συσσώρευσης.

Οι p-t-s απεικονίσεις αποτελούν γενίκευση των πραγματικών συναρτήσεων. Επειδή η έννοια της συνέχειας είναι απαραίτητη για την θεωρία, θα χρειαστεί να την επεκτείνουμε στις p-t-s. Η αντίστοιχη έννοια της συνέχειας στις p-t-s αναφέρεται ως κλειστότητα (closedness).

Προκειμένου να γίνει σαφής η αναλογία μεταξύ κλειστότητας και συνέχειας, ο ορισμός της συνέχειας μιας πραγματικής συνάρτησης  $f : \mathbb{R} \rightarrow \mathbb{R}$ , στο σημείο  $x_\infty$  (αν



$x_n \rightarrow x_\infty$  τότε  $f(x_n) \rightarrow f(x_\infty)$ ) εκφράζεται και ως εξής:

αν  $x_n \rightarrow x_\infty$  και  $y_n = f(x_n)$

τότε  $y_n \rightarrow y_\infty$  και  $y_\infty = f(x_\infty)$

**Ορισμός 4.2** (Κλειστότητα p-t-s απεικόνισης). Μία p-t-s απεικόνιση  $\mathcal{M} : X \rightarrow X$  λέγεται κλειστή στο  $x^\infty$

αν  $x^k \rightarrow x^\infty$  και  $y_k \in \mathcal{M}(x^k)$  και  $y^k \rightarrow y^\infty$

τότε  $y^\infty \in \mathcal{M}(x_\infty)$ .

Η  $\mathcal{M}$  λέγεται κλειστή στο  $X$  αν είναι κλειστή σε κάθε σημείο  $x \in X$

Αντί για το βέλτιστο σημείο, θα χρησιμοποιούμε μία ευρύτερη έννοια, αυτή του σημείου λύσης:

**Ορισμός 4.3** (Σημείο και σύνολο λύσης). Ως σημείο λύσης καλείται ένα οποιοδήποτε σημείο κάποιου συνόλου, έστω  $\Gamma$ , το οποίο λέγεται σύνολο λύσης.

Ο αλγόριθμος, αντί να πραγματοποιεί αναζήτηση για κάποιο βέλτιστο σημείο, αναζητά κάποιο σημείο λύσης. Για παράδειγμα, το  $\Gamma$  μπορεί να είναι το σύνολο όλων των ελαχίστων μιας κυρτής αντικειμενικής συνάρτησης, ή εναλλακτικά (αν μας ενδιαφέρουν οι αναγκαίες συνθήκες), το σύνολο όλων των λύσεων της εξίσωσης  $\nabla f(x) = 0$ , το οποίο είναι το σύνολο όλων των κρίσιμων σημείων. Το πως θα προσδιορίσουμε το σύνολο  $\Gamma$ , εξαρτάται επομένως, από το ίδιο το πρόβλημα και το συγκεκριμένο αλγόριθμο. Για παράδειγμα, ως στόχος θα μπορούσε να είναι να φτάσει η αντικειμενική συνάρτηση μία επιθυμητή ποσότητα  $a$ . Οπότε,  $\Gamma = \{x : f(x) \geq a, x \in F\}$ , όπου  $F$  η εφικτή περιοχή του προβλήματος (δηλαδή το σύνολο των σημείων που ικανοποιούν όλους του περιορισμούς του προβλήματος). Άλλα σύνολα λύσεων θα μπορούσαν να είναι οι ρίζες κάποιου συστήματος εξισώσεων, σαγματικά σημεία, σημεία ευστάθειας εξισώσεων διαφορών κ.α. Θα υποθέτουμε ότι το σύνολο λύσεων θα δίνεται. Τότε ο στόχος του αλγορίθμου θα είναι να προσδιορίσει (με την έννοια του ορίου) ένα σημείο στο σύνολο αυτό.

### 5.2.1 Θεώρημα Ολικής Σύγκλισης

Το επόμενο κεντρικό αποτέλεσμα διατυπώθηκε από τον Zangwill (1969) και θα το επικαλεστούμε στη συνέχεια για τη σύγκλιση του αλγορίθμου EM. Η

**Θεώρημα Ολικής Σύγκλισης** (Global Convergence Theorem) Έστω  $\mathcal{M} : X \rightarrow X$  μία p-t-s απεικόνιση η οποία, ξεκινώντας από το αρχικό σημείο  $x^{(0)} \in X$ , παράγει την ακολουθία  $\{x^{(k)}\}_{k \geq 0}$  σύμφωνα με την  $x^{(k+1)} \in \mathcal{M}(x^{(k)})$ . Έστω επίσης πως το σύνολο λύσεων είναι το  $\Gamma$ . Υποθέτουμε:

## Βασικές έννοιες και Κεντρικό Θεώρημα Σύγκλισης

- (1) Όλα τα οριακά σημεία της  $x^{(k)}$  βρίσκονται στο συμπαγές σύνολο  $S \subset X$
- (2) Υπάρχει μία συνεχής συνάρτηση  $g : X \rightarrow \mathbb{R}$  τέτοια ώστε:
  - (i) Αν  $x \notin \Gamma$ , τότε  $g(y) < g(x) \forall y \in \mathcal{M}(x)$
  - (ii) Αν  $x \in \Gamma$ , τότε  $g(y) \leq g(x) \forall y \in \mathcal{M}(x)$
- (3) Η απεικόνιση  $\mathcal{M}$  είναι κλειστή στο  $X \setminus \Gamma$

Τότε, όλα τα οριακά σημεία οποιασδήποτε υπακολουθίας  $\{x^{(k)}\}_{k \geq 0}$  είναι στο  $\Gamma$ . Ακόμα,  $\lim_{x \rightarrow \infty} g(x) = g(x^*)$  για όλα τα οριακά σημεία  $x^*$ .

### Απόδειξη

Έστω  $x^*$  οριακό σημείο της  $(x^{(k)})_{k \geq 0}$ . Από τον ορισμό του οριακού σημείου και τη συνέχεια της  $g$  έχουμε ότι υπακολουθία  $(x^{(k_n)})_{n \geq 0} g(x^{(k_n)}) \rightarrow g(x^*)$ . Αρχικά θα δείξουμε ότι  $g(x^{(k)}) \rightarrow g(x^*)$ .

Από τα (i), (ii), η ακολουθία  $\{g(x^{(k)})\}_{k \geq 0}$  είναι αύξουσα. Οπότε,  $g(x^*) - g(x^{(k)}) \geq 0$  για κάθε  $k$ .

Έστω  $\epsilon > 0$ ). Επειδή  $g(x^{(k_n)}) \rightarrow g(x^*)$ , υπάρχει  $n_0$  τέτοιο ώστε, για  $n \geq n_0$  να ισχύει:

$$g(x^*) - g(x^{(k_n)}) < \epsilon, \quad \text{για κάθε } n \geq n_0.$$

Οπότε, για κάθε  $k \geq n_0$

$$g(x^*) - g(x^{(k)}) = g(x^*) - g(x^{(k_n)}) + g(x^{(k_n)}) - g(x^{(k)}) < \epsilon$$

Θα δείξουμε τώρα ότι το σημείο  $x^* \in \Gamma$ . Έστω ότι δεν είναι στο  $\Gamma$ . Θεωρούμε την ακολουθία  $(x^{(k_n+1)})_{n \geq 1}$ , η οποία έχει την ιδιότητα: για κάθε  $n, x^{(k_n+1)} \in \text{razocalM}(x^{(k_n)})$ . Η νέα αυτή ακολουθία βρίσκεται στο συμπαγές  $S$  και άρα έχει συγκλίνουσα υπακολουθία, έστω  $x^{(k_n+1)l} \rightarrow \bar{x}$ , καθώς  $l \rightarrow \infty$ . Επειδή η  $\mathcal{M}$  είναι κλειστή στο  $X \setminus \Gamma$  και  $x^* \notin \Gamma$  (από την υπόθεση) θα είναι:

$$\bar{x} \in \mathcal{M}(x^*)$$

Όμως  $g(x^{(k)}) \rightarrow g(x^*)$  και  $g(x^{(k)}) \rightarrow g(\bar{x})$ , οπότε,  $g(x^*) = g(\bar{x})$  το οποίο έρχεται σε αντίθεση με το (2) (i).  $\square$

### Παρατηρήσεις

1. Η συνθήκη (1) του θεωρήματος εξασφαλίζει πως όλες οι υπακολουθίες έχουν όριο στο  $X$ , όπως επίσης και ότι αυτό το όριο είναι πεπερασμένο. Ακόμη, η εφικτή περιοχή δεν είναι απαραίτητα συμπαγές σύνολο. Μόνο τα σημεία που σχηματίζονται από τον αλγόριθμο θα πρέπει να βρίσκονται σε συμπαγές σύνολο.

## Βασικές έννοιες και Κεντρικό Θεώρημα Σύγκλισης

2. Η συνθήκη 2 απαιτεί πως η αντικειμενική συνάρτηση  $g$  θα βελτιώνεται σε κάθε επανάληψη, μέχρι ο αλγόριθμος να φτάσει σε σημείο λύσης.
3. Η συνθήκη 3 έχει πολύ βασικό ρόλο καθώς έτσι αποφεύγονται σημεία ασυνέχειας τα οποία μπορεί να οδηγήσουν σε προβλήματα στη σύγκλιση. Αυτό χρειάζεται να συμβαίνει μόνο για τα σημεία λύσης καθώς για τα υπόλοιπα κυριαρχεί η συνθήκη 2(ii).

## 5.3 Αλγόριθμος EM

### 5.3.1 Εισαγωγή

Ο αλγόριθμος EM θεμελιώθηκε για πρώτη φορά από τους Demster, Laird and Rubin (1977) (DLR) ως μία επαναληπτική διαδικασία για τον υπολογισμό της εκτιμήτριας μέγιστης πιθανοφάνειας (E.M.Π.). Όπως και οι ίδιοι έχουν τονίσει, ο EM έχει προταθεί πολλές φορές στο παρελθόν από διάφορους συγγραφείς. Ωστόσο, εκείνοι κατάφεραν να γενικεύσουν τη μέθοδο και ξεκίνησαν μία θεωρία σύγκλισης σε ένα πολύ ευρύ φάσμα εφαρμογών, προσφέροντας έτσι στη στατιστική κοινότητα ένα πολύ χρήσιμο εργαλείο τόσο για προβλήματα με μη πλήρη (λογοκρινμένα ή κρυμμένα (censored ή μη παρατηρήσιμα unobserved) δεδομένα όσο και για προβλήματα όπου η αναλυτική έκφραση της E.M.Π. είναι αδύνατο να βρεθεί ή παρουσιάζει υπολογιστικές δυσκολίες. Επίσης, η σύνδεση με την εκθετική οικογένεια δίνει πολλά χρήσιμα αποτελέσματα. Μία λεπτομερέστατη ανάλυση πέτυχε ο Rolf Sundberg (1974)

Ο C.F. Jeff Wu (1983) συνέχισε την ανάλυση της θεωρίας σύγκλισης παρουσιάζοντας γενικότερα αποτελέσματα και διορθώνοντας ορισμένα προβλήματα στο άρθρο των DLR. Μεταξύ άλλων, έδειξε ότι τα οριακά σημεία του αλγορίθμου EM είναι στάσιμα σημεία της συνάρτησης πιθανοφάνειας ( $L$ ) και πως όταν η πιθανοφάνεια είναι μονοκόρυφη, οποιαδήποτε ακολουθία EM είναι συγκλίνουσα.

Ο Boyles (1983) κινούμενος πάνω στις ίδιες γραμμές διατύπωσε παρόμοια αποτελέσματα με κόστος όμως ισχυρότερες συνθήκες. Ακόμη, ο Dan Nettleton (1991) διατύπωσε αποτελέσματα για τη σύγκλιση σε προβλήματα σε παραμετρικούς χώρους με περιορισμούς.

Αρχικά θα αναφέρουμε το πλαίσιο στο οποίο στηρίζεται η θεωρία και θα ορίσουμε τον αλγόριθμο EM (ενότητα 5.3). Στη συνέχεια θα περάσουμε στα θεωρήματα σύγκλισης (ενότητα 5.4). Θα δούμε επίσης και πως απλοποιείται ο EM για τις εκθετικές οικογένειες (ενότητα 5.5) και θα αναφέρουμε ορισμένες τεχνικές MCMC που μπορούμε να εφαρμόσουμε σε ειδικές περιπτώσεις (ενότητα 5.5).

### 5.3.2 Γενικό πλαίσιο-Ορισμοί

Στο γενικό πλαίσιο με μη πλήρη δεδομένα υποθέτουμε την ύπαρξη των τυχαίων διανυσμάτων  $X, Y$  με τιμές στους  $(\mathcal{X}, \mathcal{Y})$  τα οποία αντιπροσωπεύουν τα δεδομένα  $x, y$  αντίστοιχα. Το διάνυσμα  $y$  υποθέτουμε ότι αποτελείται από τα δεδομένα που μπορούμε να παρατηρήσουμε ενώ το  $x$  αντιπροσωπεύει τα πλήρη δεδομένα. Μία περιγραφή για το πλαίσιο με πλήρη δεδομένα (incomplete data) γίνεται στο παράρτημα.

Αυστηρότερα, το  $\mathbf{y} = (y_1, \dots, y_p)$  αποτελεί μία πραγματοποίηση ενός τυχαίου δείγματος ( $p$ -διάστατου τυχαίου διανύσματος με α.ι.τ.μ ως συνιστώσες) από κάποιον (άπειρο) πληθυσμό με πυκνότητα  $g(y)$  ώστε  $\mathbb{P}(Y \in B) = \int_B g(y) d\mu(y)$  για οποιοδήποτε μετρήσιμο σύνολο  $B$  όπου  $\mu$  κάποιο  $\sigma$ -πεπερασμένο μέτρο που κυριαρχεί το  $\mathbb{P}^Y$ , ( $\mathbb{P}^Y \ll \mu$ ). Υποθέτουμε ότι η πραγματική πυκνότητα,  $g$ , είναι μέλος παραμετρικής οικογένειας πυκνοτήτων  $\mathcal{P} = \{g(y; \phi) : \phi \in \Phi\}$ ,  $\Phi \subset \mathbb{R}^{d_\phi}$ . Δηλαδή υπάρχει κάποιο  $\phi^* \in \Phi$  τέτοιο ώστε  $g(y; \phi^*) = g(y)$  για όλα τα  $y$ . Η από κοινού σ.π.π. του  $\mathbf{Y}$  θα είναι  $g(\mathbf{y}; \phi) = \prod_{i=1}^p g(y_i; \phi)$

## Αλγόριθμος EM

Το πρόβλημα είναι η μεγιστοποίηση της συνάρτησης πιθανοφάνειας  $L(\phi) \stackrel{def}{=} g(\mathbf{y}; \phi)$ . Δηλαδή η εύρεση σημείου  $\phi^*$  τέτοιο ώστε

$$\phi^* = \underset{\phi \in \bar{\Phi}}{\operatorname{argmax}} L(\phi) = \{\phi^* \in \bar{\Phi} : L(\phi^*) = \underset{\phi \in \bar{\Phi}}{\operatorname{max}} L(\phi)\}. \quad (5.3.1)$$

Συνήθως, για τα πλήρη δεδομένα έχουμε τη μορφή:  $x = (z, y)$ , όπου το διάνυσμα  $z$  αποτελείται από 'κρυμμένα' (μη παρατηρήσιμα) δεδομένα, κάποιας διάστασης  $N$ , ενώ η από κοινού σ.π.π.:  $f(x; \phi) = f(z, y; \phi) \equiv L_c(\phi)$  (ως προς το μέτρο γινόμενο  $\mu_z \times \mu_y$  στον  $(\mathcal{Z} \times \mathcal{Y}, \mathcal{Z} \otimes \mathcal{Y})$  αναφέρεται ως η πλήρης συνάρτηση πιθανοφάνειας η οποία υποθέτουμε ότι είναι σχετικά εύκολο να μεγιστοποιηθεί ως προς το  $\phi$ . Σε πολλές περιπτώσεις, τα κρυμμένα δεδομένα έχουν εισαχθεί τεχνητά ώστε να επιτρέπουν τη χρήση του αλγορίθμου EM. Επίσης, ο EM μπορεί να χρησιμοποιηθεί και σε περιπτώσεις όπου η  $g(y; \phi)$  παρουσιάζει δυσκολίες στη μεγιστοποίηση (ως προς το  $\phi$ ) και υπάρχει ένα εναλλακτικό μοντέλο:  $f(y, z; \phi)$  που μεγιστοποιείται ευκολότερα και είναι τέτοιο ώστε  $g(y; \phi) \propto f(y, z; \phi)$ .

Το πλαίσιο των παρατηρούμενων/πλήρων δεδομένων μπορεί επίσης να γενικευτεί: Τα παρατηρούμενα δεδομένα  $y$  δεν είναι απαραίτητο να αποτελούν μέρος στα πλήρη δεδομένα (όπως στην προηγούμενη παράγραφο) αλλά ένας μη-αντιστρέψιμος μετασχηματισμός των πλήρη δεδομένων, δηλαδή  $y = h(x)$ , όπου  $x$  τα πλήρη δεδομένα. Στην περίπτωση αυτή:

$$g(y; \phi) = \int_{h^{-1}(y)} f(x; \phi) d\mu(x),$$

$$\text{όπου } h^{-1}(y) = \{x : h(x) = y\}.$$

Επομένως το αρχικό πρόβλημα (5.3.1) ανάγεται στο πρόβλημα μεγιστοποίησης του ολοκληρώματος:

$$C(\phi) = \int_{h^{-1}(x)} f(x; \phi) d\mu(x). \quad (5.3.2)$$

Διαισθητικά, μπορούμε να σκεφτόμαστε την  $f(\cdot; \phi)$  ως μία μη-κανονικοποιημένη πυκνότητα ως προς το  $\mu$ . Έτσι, η  $C$  αποτελεί τη σταθερά κανονικοποίησης της  $f$  η οποία υποθέτουμε ότι είναι σχετικά απλή συνάρτηση ως προς το  $\phi$  (δηλαδή δεν είναι απαραίτητα η συνάρτηση πιθανοφάνειας της  $Y$ , παρά μόνο αν η  $f$  είναι η πυκνότητα της  $X$ ). Έτσι μπορούμε να αντιστοιχίσουμε σε κάθε συνάρτηση  $f(\cdot; \phi)$  την πυκνότητα  $k_y(\cdot; \phi)$  (ως προς το κυρίαρχο μέτρο  $\mu$ ) που ορίζεται από την:

$$k_y(x; \phi) \stackrel{def}{=} \frac{f(x; \phi)}{C(\phi)}. \quad (5.3.3)$$

Όταν η  $f$  είναι η πυκνότητα της  $X$ , η  $k_y(\cdot; \phi)$  είναι η δεσμευμένη πυκνότητα της  $X|Y = y$ .

Ο αλγόριθμος EM αντιμετωπίζει το (δύσκολο) πρόβλημα μεγιστοποίησης στην 5.3.1 επιλύοντας διαδοχικά (απλούστερα) προβλήματα μεγιστοποίησης. Συγκεκριμένα, στο

## Αλγόριθμος EM

βήμα M, έχουμε το πρόβλημα μεγιστοποίησης της  $f(x; \phi)$  στο  $\Phi$ . Επειδή το  $x$  είναι μη-παρατηρήσιμο, αντικαθιστούμε την  $f(x; \phi)$  με τη δεσμευμένη μέση τιμή δοθέντος των  $y$  και την τρέχουσα τιμή του  $\phi$  ( $\phi^{(k)}$ ) (βήμα-E). Πιο συγκεκριμένα, έστω  $\phi^{(0)}$  μία αρχική τιμή για το  $\phi$ . Τότε στην πρώτη επανάληψη, στο βήμα-E απαιτείται να υπολογιστεί η ποσότητα:

$$E_{\phi^{(0)}}[\log L_c(\phi)|Y = y]. \quad (5.3.4)$$

για όλα τα  $\phi \in \Phi$ , η οποία είναι ντετερμινιστική συνάρτηση. Στο βήμα M απαιτείται η μεγιστοποίηση της παραπάνω συνάρτησης (5.3.4) για όλα τα  $\phi \in \Phi$ . Δηλαδή θα επιλέξουμε το  $\phi^{(1)}$  ώστε:

$$E_{\phi^{(0)}}[\log L_c(\phi^{(1)})|Y = y] \geq E_{\phi^{(0)}}[\log L_c(\phi)|Y = y],$$

για όλα τα  $\phi \in \Phi$ . Στη συνέχεια τα βήματα E, M επαναλαμβάνονται αντικαθιστώντας το  $\phi^{(0)}$  με την τρέχουσα τιμή του  $\phi$ , δηλαδή το  $\phi^{(1)}$ .

Η συνάρτηση που εμφανίζεται στη 5.3.4 καλείται ενδιάμεση ποσότητα του αλγορίθμου EM και συμβολίζεται με:

$$Q(\phi; \phi') \stackrel{def}{=} E_{\phi'}[\log L_c(\phi)|Y = y] = \int \log f(x; \phi) k_y(x; \phi) \mu(dx). \quad (5.3.5)$$

Στη απλούστερη περίπτωση όπου  $x = (z, y)$ ,  $(f(x; \phi) = f(z, y; \phi)$ ,  $k_y(x; \phi) = k(z|y; \phi)$ ) θα είναι:

$$Q(\phi; \phi') = E_{\phi'}[\log f(Z, Y; \phi)|Y = y] = \int \log f(z, y; \phi) k(z|y; \phi') \mu(dz). \quad (5.3.6)$$

Με τον παραπάνω συμβολισμό, στη  $(k + 1)$ -οστή επανάληψη, τα βήματα E και M ορίζονται ως εξής:

**Βήμα-E.** Υπολογισμός της  $Q(\phi; \phi^k)$ , όπου

$$Q(\phi; \phi^k) = E_{\phi^{(k)}}[\log L_c(\phi)|Y = y]. \quad (5.3.7)$$

**Βήμα-M.** Επιλογή ενός  $\phi^{(k+1)}$  από τα  $\phi \in \Phi$  που μεγιστοποιεί τη  $Q(\phi; \phi^{(k)})$ :

$$Q(\phi^{(k+1)}; \phi^{(k)}) \geq Q(\phi; \phi^{(k)}) \quad (5.3.8)$$

για όλα τα  $\phi \in \Phi$ .

Τα βήματα E και M επαναλαμβάνονται έως ότου η διαφορά:  $L(\phi^{k+1}) - L(\phi^k)$  γίνει μικρή (μικρότερη από κάποιο  $\epsilon$ ).

Η 5.3.8 εκφράζεται λέγοντας ότι το  $\phi^{(k+1)}$  ανήκει στο  $\mathcal{M}(\phi^{(k)})$ , όπου:

$$\mathcal{M}(\phi^{(k)}) = \underset{\phi \in \Phi}{\operatorname{argmax}} Q(\phi, \phi^{(k)})$$

Ένα  $\phi' \in \mathcal{M}(\phi)$  είναι τέτοιο ώστε  $Q(\phi', \phi) \geq Q(\theta, \phi) \quad \forall \theta \in \Phi$

Σε όσα ακολουθούν θα χρησιμοποιήσουμε την τελευταία έκφραση καθώς μας συνδέει με τη θεωρία που αναπτύχθηκε στην ενότητα 5.1. Έτσι έχουμε τον εξής ορισμό:

## Αλγόριθμος EM

**Ορισμός 4.2** (Ακολουθία EM). Έστω  $\mathcal{M} : \Phi \rightarrow \Phi$ , μία απεικόνιση τέτοια ώστε:

$$\mathcal{M}(\phi) = \underset{\phi' \in \Phi}{\operatorname{argmax}} Q(\phi', \phi)$$

Για οποιαδήποτε αρχική τιμή  $\phi^{(0)}$ , μία ακολουθία EM:  $\{\phi^{(k)}\}_{k \geq 0}$ , είναι μία ακολουθία που παράγεται από την:

$$\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$$

Οπότε, μία ακολουθία EM,  $\{\phi^{(k)}\}_{k \geq 0}$  είναι τέτοια ώστε:

$$Q(\phi^{(k+1)}, \phi^{(k)}) \geq Q(\phi, \phi^{(k)}), \text{ για οποιοδήποτε } \phi \in \Phi \text{ και } k \geq 0. \quad (5.3.9)$$

**Ορισμός 4.3** (αλγόριθμος EM) Ο αλγόριθμος EM αποτελείται από τα 2 ακόλουθα βήματα:

**Βήμα-Ε.** Υπολογισμός της  $Q(\phi, \phi^k)$

**Βήμα-Μ.** Επιλογή  $\phi^{k+1}$  να είναι οποιοδήποτε  $\phi \in \Phi$  μεγιστοποιεί την  $Q(\phi, \phi^k)$

Σταματάμε όταν:  $\|\phi^{k+1} - \phi^k\| \leq \epsilon$

Στις περιπτώσεις όπου ο αναλυτικός υπολογισμός του βήματος Μ παρουσιάζει δυσκολίες, χρησιμοποιείται ο Γενικευμένος EM (GEM) που ορίζεται ως η ακόλουθη επαναληπτική διαδικασία:

**Ορισμός 4.4** (αλγόριθμος GEM) Έστω  $\mathcal{M} : \Phi \rightarrow \Phi$ , μία απεικόνιση τέτοια ώστε:

$$Q(\phi'; \phi) \geq Q(\phi; \phi) \text{ για κάθε } \phi' \in \mathcal{M}(\phi)$$

Για οποιαδήποτε αρχική τιμή  $\phi^0$  μία ακολουθία EM:  $\{\phi^{(k)}\}_{k \geq 0}$  είναι μία ακολουθία που παράγεται από την:

$$\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$$

Δηλαδή, επιλέγουμε το  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$  να αυξάνει την  $Q(\phi; \phi^{(k)})$  για την τιμή του  $\phi = \phi^{(k+1)}$  αντί να μεγιστοποιούμε για όλα τα  $\phi \in \Phi$ .

Είναι φανερό πως ο EM αποτελεί ειδική περίπτωση του GEM.

Παρατηρήσεις για τη συνάρτηση:

$$Q(\phi; \phi') = \int \log f(x; \phi) k(x; \phi') \mu(dx).$$

Αν συμβολίσουμε με  $h(x; \phi, \phi')$  το γινόμενο μέσα στο παραπάνω ολοκλήρωμα, τότε το ολοκλήρωμα υπάρχει αν τουλάχιστον μία από τις  $\int h^+$ ,  $\int h^-$  είναι  $< \infty$  και είναι πραγματικός αριθμός αν και τα 2 αυτά ολοκλήρωματα είναι πεπερασμένα. Προκειμένου λοιπόν η συνάρτηση  $Q$  να ορίζεται καλά για οποιεσδήποτε τιμές των  $\phi, \phi'$  θα χρειαστούμε τις παρακάτω συνθήκες ομαλότητας:

### Υπόθεση 5.1

- (i) Ο παραμετρικός χώρος  $\Phi$  είναι ανοικτό υποσύνολο του  $\mathbb{R}^{d_\phi}$ .
- (ii) Για οποιοδήποτε  $\phi \in \Phi$ , η συνάρτηση  $L$  είναι μη αρνητική και άνω φραγμένη.
- (iii) Για οποιαδήποτε  $\phi, \phi' \in \Phi$  ισχύει  $\int |\nabla_\phi \log k(x; \phi)| k(x; \phi') \lambda(dx) < \infty$

**Πρόταση 5.1** Κάτω από την υπόθεση 1, για οποιαδήποτε  $\phi, \phi' \in \Phi$  ισχύει:

$$Q(\phi; \phi') = \log L(\phi) - H(\phi; \phi'), \quad (5.3.10)$$

όπου

$$H(\phi; \phi') \stackrel{def}{=} \int -\log k_y(x; \phi) k_y(x; \phi') \mu(dx). \quad (5.3.11)$$

### Απόδειξη

Επειδή η  $k_y(x; \phi)$  είναι η πυκνότητα της  $X|Y = y$ , και  $L$  η πιθανοφάνεια της  $Y$ , έπεται πως:

$$\log L(\phi) = \log g(y; \phi) = \log f(x; \phi) - \log k_y(x; \phi).$$

Έστω  $\phi' \in \Phi$ . Ολοκληρώνοντας τα 2 μέλη της παραπάνω ως προς το μέτρο  $k_y(x; \phi') \mu(dx)$  θα πάρουμε:

$$\int \log L(\phi) k_y(x; \phi') \mu(dx) = \int \log f(\phi) k_y(x; \phi') \mu(dx) - \int \log k_y(x; \phi) k_y(x; \phi') \mu(dx)$$

Η συνάρτηση στο πρώτο ολοκλήρωμα είναι ανεξάρτητη του  $x$  ενώ τα άλλα ολοκληρώματα αντιστοιχούν στις ποσότητες που έχουν οριστεί. Επομένως,  $\log L(\phi) = Q(\phi; \phi') + H(\phi; \phi')$  και το ζητούμενο έπεται άμεσα.  $\square$

Το νόημα της παραπάνω σχέσης είναι ότι η  $Q(\phi, \phi')$  διαφέρει από την αντικειμενική συνάρτηση  $\log L(\phi)$  μιας ποσότητας ( $H(\phi; \phi')$ ) η οποία έχει γνωστές ιδιότητες. Πράγματι, η  $H(\theta'; \theta')$  είναι η εντροπία της πυκνότητας  $p(\cdot; \theta')$ . Ακόμη, η προσαύξηση της  $H(\phi; \phi')$ :

$$H(\phi; \phi') - H(\phi'; \phi') = - \int \log \frac{k_y(x; \phi)}{k_y(x; \phi')} k_y(x; \phi') \mu(dx), \quad (5.3.12)$$

είναι η απόκλιση Kullback-Leibler (ή σχετική εντροπία) μεταξύ των πυκνοτήτων  $k_y(x; \phi), k_y(x; \phi')$  όπου  $k_y(\cdot; \phi) d\mu \ll k_y(\cdot; \phi') d\mu$ .

Ειδικά το τελευταίο έπεται από την υπόθεση 1(iii) καθώς γενικότερα η υπόθεση αυτή εξασφαλίζει πως όλα τα μέτρα στην οικογένεια  $\{k_y(\cdot; \phi) d\mu : \phi \in \Phi\}$  κυριαρχούν το ένα το άλλο. (Με  $k_y(\cdot; \phi) d\mu$  συμβολίζεται η κατανομή της  $X|Y = y$  και όχι η πυκνότητα καθώς η τελευταία είναι συνάρτηση). Γενικά, ένα μέτρο  $k_y(\cdot; \phi) d\mu$  είναι μηδενικό μόνο σε σύνολα μέτρου (πιθανότητας) 0 από όλα τα υπόλοιπα μέτρα της οικογένειας. Οπότε οι συναρτήσεις  $H(\phi; \phi')$  και  $Q(\phi; \phi')$  ορίζονται καλά για οποιοδήποτε ζεύγος παραμέτρων.



## Αλγόριθμος EM

Συνήθως ο συμβολισμός της απόκλισης Kullback-Leibler για 2 μέτρα πιθανότητας  $\nu, \mu$  με  $\nu \ll \mu$  είναι:

$$D(\mu||\nu) = -E_\mu[\log \frac{d\nu}{d\mu}].$$

Με μία πρώτη ματιά δεν είναι άμεσα φανερό πως ο παραπάνω ορισμός δεν αντιμετωπίζει το κλασικό πρόβλημα  $\infty - \infty$ . Έστω  $p, q$ , οι πυκνότητες των  $\mu, \nu$  αντίστοιχα. Από το Θεώρημα Taylor, για  $x > -1$ :  $(1+x) \log(1+x) = x + R(x)$  με  $R(x) = \frac{1x^2}{1+x^*}$  για κάποιο  $x^*$  μεταξύ του 0 και του  $x$ . Όταν  $p > 0$  και  $q > 0$  παίρνουμε  $p \log(p/q) \geq p - q$ . Η ίδια ανισότητα ισχύει και για σημεία όπου  $p > 0$  και  $q = 0$  (όπου σε αυτή την περίπτωση το πρώτο μέλος ερμηνεύεται ως  $\infty$ ). Έπεται πως  $-E_\mu[(\log \frac{d\nu}{d\mu})^-] < \infty$ .

Αν το  $\nu$  δεν είναι απόλυτα συνεχές ως προς το  $\mu$  τότε  $D(\mu||\nu) = \infty$ .

**Λήμμα 5.1** Ισχύει  $D(\mu||\nu) \geq 0$  με ισότητα αν  $\nu = \mu$  σ.π. (ως προς το  $\mu$ ).

### Απόδειξη

Από την ανισότητα του Jensen:

$$E_\mu[\log \frac{d\nu}{d\mu}] \leq \log E_\mu[\frac{d\nu}{d\mu}] = \log 1 = 0.$$

□ Οπότε, από το Λήμμα 5.1:

**Πόρισμα 5.1** Για τη συνάρτηση  $H$  που ορίστηκε στην Πρόταση 5.1 ισχύει ότι:

$$H(\phi; \phi') - H(\phi'; \phi') > 0 \quad \text{για κάθε } \phi' \neq \phi, \quad (5.3.13)$$

με ισότητα αν  $\phi' = \phi$ . □

### 5.3.3 Βασικές Ιδιότητες του αλγορίθμου EM

Η βασικότερη ιδιότητα του αλγορίθμου EM είναι πως η ακολουθία των τιμών της πιθανοφάνειας:  $\{L(\phi^{(k)})\}_{k \geq 0}$  είναι μονότονη. Συγκεκριμένα:

**Πρόταση 5.2** (Μονοτονία αλγορίθμου EM). Κάτω από την Υπόθεση 5.1, για οποιαδήποτε ακολουθία  $\{\phi^{(k)}\}_{k \geq 0}$  που παράγεται από τον GEM, ισχύει:

$$L(\phi^{(k+1)}) \geq L(\phi^{(k)}) \quad \text{για κάθε } k \geq 0. \quad (5.3.14)$$

και η ανισότητα είναι γνήσια εκτός αν τα  $p(\cdot; \phi^{(k)})$ ,  $p(\cdot; \phi^{(k+1)})$  ταυτίζονται (σ.π. ως προς το μέτρο αναφοράς).

#### Απόδειξη

Έστω  $\phi^{(k)} \in \Phi$  η τρέχουσα τιμή και  $\phi^{(k+1)}$  η ανανέωση της, σύμφωνα με τον αλγόριθμο EM, δηλαδή

$$Q(\phi^{(k+1)}; \phi^{(k)}) \geq Q(\phi; \phi^{(k)}) \quad \text{για κάθε } \phi \in \Phi$$

Από την Πρόταση (5.1) είναι για κάθε  $\phi' \in \Phi$ :

$$Q(\phi; \phi') = \log L(\phi) - H(\phi; \phi').$$

Οπότε, για  $\phi' = \phi^{(k)}$ :

$$\begin{aligned} \log L(\phi^{(k+1)}) - \log L(\phi^{(k)}) &= Q(\phi^{(k+1)}; \phi^{(k)}) - Q(\phi^{(k)}; \phi^{(k)}) \\ &\quad + H(\phi^{(k+1)}; \phi^{(k)}) - H(\phi^{(k)}; \phi^{(k)}) \end{aligned}$$

Η διαφορά:  $Q(\phi^{(k+1)}; \phi^{(k)}) - Q(\phi^{(k)}; \phi^{(k)})$  επομένως είναι μη αρνητική, ενώ το γεγονός πως  $H(\phi^{(k+1)}; \phi^{(k)}) - H(\phi^{(k)}; \phi^{(k)}) \geq 0$  έπεται από το πόρισμα (5.1) στην προηγούμενη ενότητα.  $\square$

Από την παραπάνω, είναι σαφές πως για μπορούμε να μιλάμε για τη σύγκλιση της  $\{L(\phi^{(k)})\}_{k \geq 0}$  σε κάποιο σημείο, θα πρέπει η συνάρτηση πιθανοφάνειας να είναι άνω φραγμένη.

Από την ιδιότητα της μονοτονίας έχουμε και το εξής:

**Πόρισμα 5.2** Κάτω από την Υπόθεση 5.1, για οποιαδήποτε  $\phi, \phi' \in \Phi$  ισχύει:

$$\log L(\phi) - \log L(\phi') \geq Q(\phi; \phi') - Q(\phi'; \phi') \quad (5.3.15)$$

Για τον GEM η παραπάνω ιδιότητα εξακολουθεί να ισχύει. Συγκεκριμένα, για κάθε  $\phi \in \Phi$ :

$$\log L(\mathcal{M}(\phi)) \geq \log L(\phi). \quad (5.3.16)$$

#### Υπόθεση 5.2

## Αλγόριθμος EM

- (i) Η συνάρτηση  $L : \Phi \rightarrow \mathbb{R}$  είναι συνεχώς διαφορίσιμη στο  $\Phi$ .
- (ii) Για οποιοδήποτε  $\phi' \in \Phi$ , η συνάρτηση  $H(\cdot; \phi')$ , με  $\phi \mapsto H(\phi; \phi')$  είναι συνεχώς διαφορίσιμη στο  $\Phi$ .
- (iii) Η παράγωγος της  $H(\cdot; \phi')$ , για οποιοδήποτε  $\phi' \in \Phi$ , μπορεί να περάσει μέσα στο ολοκλήρωμα.

**Πρόταση 5.3** Κάτω από την Υπόθεση 5.2, για κάθε  $\phi' \in \Phi$ , συνάρτηση  $\phi \mapsto Q(\phi; \phi')$  είναι συνεχώς διαφορίσιμη στο  $\Phi$  και:

$$\nabla_{\phi} \log L(\phi)|_{\phi=\phi'} = \nabla_{\phi} Q(\phi; \phi')|_{\phi=\phi'} \quad (5.3.17)$$

### Απόδειξη

Για οποιοδήποτε  $\phi' \in \Phi$ , η  $Q(\cdot; \phi') = \log L(\cdot) - H(\cdot; \phi')$  είναι συνεχώς διαφορίσιμη στο  $\Phi$  ως διαφορά συνεχών διαφορίσιμων συναρτήσεων κάτω από την υπόθεση 5.2.

Η ποσότητα  $H(\phi; \phi')$  παρουσιάζει ελάχιστο στο σημείο  $\phi = \phi'$ , όπως διαπιστώθηκε στα προηγούμενα (Λήμμα 5.1). Επομένως  $\nabla_{\phi} H(\phi; \phi')|_{\phi=\phi'} = 0$  και έτσι έπεται άμεσα το ζητούμενο.  $\square$

Από την παραπάνω έπεται πως αν το  $\phi^*$  είναι στάσιμο σημείο της  $L$ , τότε

$$\nabla_{\phi} Q(\phi; \phi^*)|_{\phi=\phi^*} = 0. \quad (5.3.18)$$

δηλαδή είναι και στάσιμο σημείο της  $Q(\cdot; \phi')$ . Έτσι, είναι δυνατόν ο EM να συγκλίνει σε κάποιο σαγματικό σημείο. ( Ένα παράδειγμα με σύγκλιση σε σαγματικό σημείο μπορεί να βρεθεί στο [] σελ 85)

Κλείνουμε την ενότητα αυτή με ορισμένες παρατηρήσεις σχετικά με τον αλγόριθμο EM.

Το υπολογιστικό κόστος σε κάθε βήμα του αλγορίθμου είναι σε γενικές γραμμές μικρό και έτσι εξισορροπείται το μεγάλο πλήθος επαναλήψεων που απαιτείται σε σχέση με άλλες μεθόδους.

Από την άλλη, σε περιπτώσεις όπου δεν έχουμε μοναδικό μέγιστο, δεν έχουμε κάποια εγγύηση ότι ο EM θα συγκλίνει. Επίσης, στην περίπτωση αυτή, η εκτίμηση που θα πάρουμε εξαρτάται από την αρχική τιμή. Γενικά όμως δεν υπάρχει κάποιος αλγόριθμος που να εγγυάται σύγκλιση σε ολικό ή τοπικό μέγιστο κάτω από τις συνθήκες αυτές. Υπάρχουν ωστόσο τεχνικές όπως το simulated annealing για την επίλυση τέτοιων ζητημάτων με το κόστος όμως της δύσκολης εφαρμογής.

Στο μεγάλο πλήθος εφαρμογών του EM σε προβλήματα εύρεσης της πιθανοφάνειας μπορούν να ξεχωρίσουν 2 ειδικές ερμηνείες:

1. Πολλές φορές, ο αλγόριθμος EM προκύπτει φυσιολογικά μέσα από το πρόβλημα της μεγιστοποίησης με χρήση παραγώγων. Πολλοί συγγραφείς έχουν φτάσει στον αλγόριθμο EM καθώς προσπαθούσαν να πετύχουν μία κομψή μορφή για τις εξισώσεις

## Αλγόριθμος EM

πιθανοφάνειας προκειμένου να φτάσουν στην επίλυση τους.

2. Πολλά προβλήματα για τα οποία η αναλυτική έκφραση της Ε.Μ.Π είναι δύσκολη ή αδύνατη μπορούν να μεταφερθούν σε πλαίσιο με μη πλήρη δεδομένα. Σε τέτοια προβλήματα, είναι δυνατόν να κατασκευαστεί μία κατάλληλη έκδοση του EM για την εύρεση της Ε.Μ.Π.

## 5.4 Σύγκλιση του αλγορίθμου EM

### 5.4.1 Εισαγωγή

Στην ενότητα αυτή διατυπώνονται τα βασικά θεωρήματα σύγκλισης τα οποία βασίζονται στο Ολικό Θεώρημα σύγκλισης (Θ.5.1) που συζητήθηκε στην ενότητα 5.2. Στις επόμενες γραμμές, αναφέρονται ορισμένες βασικές έννοιες ώστε να καταστεί σαφές τι εννοούμε σύγκλιση του αλγορίθμου EM.

Το σημείο  $\phi$  είναι οριακό σημείο για την ακολουθία  $(\phi^{(k)})_{k \geq 0}$  αν υπάρχει υπακολουθία  $(\phi^{(k_n)})_{n \geq 0}$  η οποία συγκλίνει στο  $\phi$ . Ένα σημείο  $\phi^* \in \Phi$  λέγεται οριακό σημείο του EM αν υπάρχει μία ακολουθία EM ή GEM, έστω  $(\phi^{(k)})_{k \geq 0}$ , για την οποία, το  $\phi^*$  είναι οριακό σημείο. Στην περίπτωση που η  $(\phi^{(k)})_{k \geq 0}$  συγκλίνει στο  $\phi^*$ , το  $\phi^*$  λέγεται σημείο σύγκλισης του αλγορίθμου EM ή GEM.

Γενικά, ο αλγόριθμος EM λέμε ότι συγκλίνει αν κάποια ακολουθία EM συγκλίνει. Επομένως, για έναν EM που συγκλίνει, δεν είναι απαραίτητο όλες οι ακολουθίες να συγκλίνουν στο ίδιο όριο για όλες τις αρχικές τιμές. Αυτό που απαιτείται είναι για οποιαδήποτε αρχική τιμή, η ακολουθία EM που παράγεται, να συγκλίνει (ενδεχομένως σε διαφορετικό όριο).

Έστω  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία GEM που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ . Το πρόβλημα που θα ασχοληθούμε στην παράγραφο αυτή είναι να εξετάσουμε (1) τη σύγκλιση της ακολουθίας  $(L(\phi^{(k)}))_{k \geq 0}$  και στη συνέχεια, (2) τη σύγκλιση της ακολουθίας  $(\phi^{(k)})_{k \geq 0}$ . Το πιο ισχυρό αποτέλεσμα που θα μπορούσαμε να είχαμε θα ήταν η σύγκλιση στην εκτιμήτρια μέγιστης πιθανοφάνειας ( $\phi^*$ ):

Για οποιαδήποτε αρχική τιμή  $\phi_0 \in \Phi$  να ισχύει:

$$L(\phi^{(k)}) \rightarrow L(\phi^*)$$

καθώς  $k \rightarrow \infty$  με

$$\phi^* = \underset{\phi \in \Phi}{\operatorname{argmax}} L(\phi) = \{\phi^* \in \bar{\Phi} : L(\phi^*) = \underset{\phi \in \bar{\Phi}}{\operatorname{max}} L(\phi)\}.$$

Σχετικά με το ζήτημα (1), τα βασικά αποτελέσματα συνοψίζονται στις παραγράφους που ακολουθούν.

Το αποτέλεσμα με ίσως τις πιο εύκολα επαληθεύσιμες συνθήκες είναι η σύγκλιση σε στάσιμο σημείο. Από την προηγούμενη ενότητα, είδαμε πως αν η ακολουθία  $(L(\phi^{(k)}))_{k \geq 0}$  είναι άνω φραγμένη, με την ιδιότητα της μονοτονίας εξασφαλίζεται πως συγκλίνει σε κάποια τιμή  $L^*$  ( $L(\phi^{(k)}) \uparrow L^*$ ). Υπό ορισμένες προϋποθέσεις, το  $L^*$  είναι στάσιμη τιμή. Δηλαδή υπάρχει κάποιο στάσιμο σημείο της  $L$ :  $\phi^*$  με  $L(\phi^*) = L^*$ . Ένα σημείο δηλαδή στο οποίο μηδενίζεται η πρώτη παράγωγος της  $L$  και μπορεί να είναι τοπικό μέγιστο, τοπικό ελάχιστο ή σαγματικό σημείο.

Γενικότερα, αν η  $L$  έχει πολλά στάσιμα σημεία, τότε ο EM μπορεί να συγκλίνει σε κάποιο από αυτά. Η επιλογή του αρχικού σημείου μπορεί να καθορίσει σε ποιο από αυτά θα έχουμε τη σύγκλιση. Στο [27] σελίδες 85-90, υπάρχει παράδειγμα με σύγκλιση

σε τοπικό ελάχιστο. Γενικότερα παραδείγματα πάνω στο ζήτημα αυτό υπάρχουν στα: Hasselblad (1969), Wolfe (1970), Haberman (1974), Laird (1978), Rubin and Thayer (1982). Τεχνικές όπως να τρέξουμε τον αλγόριθμο πολλές φορές με διαφορετικά αρχικά σημεία, ή με τεχνικές όπως το simulated annealing (πχ Finch et al. (1989)) μπορούν να δώσουν κάποια εγγύηση ότι το ολικό μέγιστο έχει βρεθεί. Ο Wu (1983) διατύπωσε και ένα άλλο Θεώρημα (Θεώρημα Θ.3) στο οποίο εξασφαλίζεται η σύγκλιση σε τοπικό μέγιστο, ωστόσο, η συνθήκη που απαιτεί το Θεώρημα είναι δύσκολο να επαληθευθεί.

Υπό ορισμένες συνθήκες (Πόρισμα 5.3), στην περίπτωση που η  $L$  έχει μοναδικό μέγιστο (την Ε.Μ.Π.), ο EM συγκλίνει σε αυτό ανεξάρτητα της αρχικής τιμής.

Σχετικά με το ζήτημα (2), ο Wu επισημάνει πως είναι λιγότερα ουσιαστικό από τη σύγκλιση της  $(L(\phi^{(k)}))_{k \geq 0}$  σε κάποια στάσιμη τιμή και συγκεκριμένα σε μέγιστο. Ακόμη, όπως θα δούμε στη συνέχεια, για να έχουμε τη σύγκλιση αυτή, οι συνθήκες είναι πιο απαιτητικές.

Αρχικά, θα υποθέσουμε ορισμένες συνθήκες κανονικότητας (που περιέχουν τις συνθήκες του Wu (1983)), οι οποίες θα υποθέτουμε ότι ισχύουν σε όσα ακολουθούν στην ενότητα αυτή.

### Υπόθεση 5.3 (Συνθήκες κανονικότητας)

R1. Ο παραμετρικός χώρος  $\Phi$  είναι ανοικτό υποσύνολο του  $\mathbb{R}^{d_\phi}$ .

R2. Το σύνολο  $\Phi_{\phi_0} \stackrel{def}{=} \{\phi \in \Phi : L(\phi) \geq L(\phi_0)\}$  είναι συμπαγές για κάθε  $\phi_0$  με  $L(\phi_0) > -\infty$ .

R3. Η συνάρτηση  $L$  είναι συνεχώς διαφορίσιμη στο  $\Phi$ .

R4. Για οποιοδήποτε  $\phi' \in \Phi$ , η συνάρτηση  $H(\cdot; \phi')$ , με  $\phi \mapsto H(\phi; \phi')$  είναι συνεχώς διαφορίσιμη στο  $\Phi$ .

R5. Η παράγωγος της  $H(\cdot; \phi')$ , για οποιοδήποτε  $\phi' \in \Phi$ , μπορεί να περάσει μέσα στο ολοκλήρωμα.

Από τις R1-R3) έπεται πως (R6.) κάθε ακολουθία EM ή GEM είναι τέτοια ώστε:  $\phi^{(k)} \in \Phi$  για κάθε  $k \geq 1$  και η ακολουθία  $(L(\phi^{(k)}))_{k \geq 0}$  είναι άνω φραγμένη για οποιαδήποτε επιλογή αρχικής τιμής  $\phi_0$ . Στα επόμενα θεωρούμε ακολουθίες GEM, καθώς αποτελούν γενικότερη περίπτωση από ακολουθίες EM.

#### 5.4.2 Σύγκλιση της ακολουθίας $L(\phi^{(k)})$

Το βασικό εργαλείο που χρησιμοποιεί ο Wu (1983) είναι το Ολικό Θεώρημα Σύγκλισης (Θ.2). Προκειμένου να εφαρμοστεί στην περίπτωση του EM, θα κάνουμε τις ακόλουθες υποθέσεις:

Έστω  $M$  η απεικόνιση για τον ΓEM και  $a(x)$  η συνάρτηση λογαριθμοπιθανοφάνειας την οποία θα συμβολίσουμε με  $L$  και όχι με  $l$ . Θέτουμε το σύνολο λύσεων  $\Gamma$  να είναι

ένα από τα παρακάτω:

$$\mathcal{M} = \text{σύνολο τοπικών μεγίστων του } \Omega,$$

ή

$$\mathcal{S} = \text{σύνολο στάσιμων σημείων του } \Omega.$$

Τότε η συνθήκη (1) έπεται άμεσα από την Πρόταση 5.2 και την υπόθεση (R6) και η συνθήκη από (2)(ii) από την Πρόταση 5.2.

Οπότε, το επόμενο Θεώρημα είναι ειδική περίπτωση του Θ.5.1

**Θεώρημα 5.1** Έστω  $\phi^{(k)}$  μία ακολουθία GEM που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ , και υποθέτουμε ότι:

- (i) Η  $\mathcal{M}$  είναι κλειστή π-τ-ς απεικόνιση στο  $\Phi \setminus \mathcal{S}$  (αντίστοιχα:  $\Phi \setminus \mathcal{M}$ ),
- (ii) Αν  $\phi^{(k)} \notin \mathcal{S}$  (αντίστοιχα:  $\mathcal{M}$ ), τότε  $L(\phi^{(k+1)}) > L(\phi^{(k)})$ .

Τότε όλα τα οριακά σημεία της  $(\phi^{(k)})_{k \geq 0}$  είναι στάσιμα σημεία (αντίστοιχα: τοπικά μέγιστα) της  $L$ , και η  $(L(\phi^{(k)}))_{k \geq 0}$  συγκλίνει στο  $L^* = L(\phi^*)$  για κάποιο  $\phi^* \in \mathcal{S}$  (αντίστοιχα  $\mathcal{M}$ )

Στην περίπτωση όπου  $\Gamma = \mathcal{S}$ , η συνθήκη (ii) ικανοποιείται. Συγκεκριμένα:

Αν για κάποιο  $k \geq 0$ ,  $\phi^{(k)} \notin \mathcal{S}$ , τότε  $\nabla_{\phi} \log L(\phi)|_{\phi=\phi^{(k)}} \neq 0$ . Από την Πρόταση 5.2, έπεται πως  $\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k)}} \neq 0$  και άρα δεν μεγιστοποιείται στο σημείο  $\phi = \phi^{(k)}$ . Από τον ορισμό του βήματος M και το Πρόσχημα 5.2 είναι:  $L(\phi^{(k+1)}) > L(\phi^{(k)})$ .

Για την (i), μία ικανή συνθήκη σύμφωνα με τον Wu (1983) είναι η (C1.) συνέχεια της  $Q$  ως προς και τα δύο ορίσματα, η οποία είναι αρκετά ασθενής και ικανοποιείται σε μεγάλο εύρος εφαρμογών.

Μία ακόμη ικανή συνθήκη (Vaida (2005)) είναι:

(C2.) Για οποιοδήποτε  $\phi^* \in \mathcal{S}$ , η  $Q(\cdot; \phi^*)$  έχει μοναδικό ολικό μέγιστο.

Έτσι έχουμε το επόμενο βασικό Θεώρημα σύγκλισης για τη σύγκλιση σε στάσιμο σημείο.

**Θεώρημα 5.2** Έστω  $\phi^{(k)}$  μία ακολουθία GEM που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$  και. Αν ισχύει κάποια από τις C1, C2, τότε

(i) όλα τα οριακά σημεία οποιασδήποτε ακολουθίας  $(\phi^{(k)})_{k \geq 0}$  GEM είναι στάσιμα σημεία ( $\in \mathcal{S}$ ) και

(ii) Η ακολουθία  $(L(\phi^{(k)}))_{k \geq 0}$  συγκλίνει στο  $L^* = L(\phi^*)$  για κάποιο (στάσιμο σημείο)  $\phi^* \in \mathcal{S}$ .

Οπότε, μία ακολουθία GEM είτε συγκλίνει, είτε έχει σύνολο οριακών σημείων με τις τιμές της συνάρτησης πιθανοφάνειας σε αυτά να ταυτίζονται. Προφανώς, αν

Το Θεώρημα 5.1, σύμφωνα με τον Wu (1983) αποτελεί το γενικότερο αποτέλεσμα, ενώ το Θεώρημα 5.2 δίνει τη σύγκλιση σε στάσιμη τιμή με μόνη συνθήκη που χρειάζεται

έλεγχο τη (C1.) ή την (C2.) η οποίες στην πράξη ικανοποιούνται στις περισσότερες εφαρμογές. Συγκεκριμένα, μία κλάση πυκνοτήτων που τις ικανοποιούν είναι η εκθετική οικογένεια για την οποία γίνεται αναφορά στη συνέχεια.

Στην περίπτωση όπου  $\Gamma = \mathcal{M}$  η συνθήκη (ii) του θεωρήματος 1 δεν ικανοποιείτε με κάποιο επιχείρημα όπως στην προηγούμενη περίπτωση. Για παράδειγμα για ένα  $\phi^{(k)}$  που δεν ανήκει στο  $\mathcal{M}$  αλλά ανήκει στο  $\mathcal{S}$ , θα ισχύει:  $\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k)}} = 0$  και το  $\phi^{(k)}$  θα μπορούσε να μεγιστοποιεί την  $Q(\cdot; \phi^{(k)})$  στο  $\Phi$ . Τότε ο EM θα σταματήσει στο  $\phi^{(k)}$  το οποίο είναι στάσιμο σημείο αλλά όχι τοπικό μέγιστο και η  $L(\phi^{(k+1)}) > L(\phi^{(k)})$  δεν ισχύει για το  $\phi^{(k)}$ .

Για να έχουμε τη σύγκλιση σε τοπικό μέγιστο, ο Wu (1983) προτείνει τη συνθήκη:

$$C3. \sup_{\phi' \in \Phi} Q(\phi', \phi) \geq Q(\phi, \phi) \quad \forall \phi \in \mathcal{S} \setminus \mathcal{M}$$

Η παραπάνω συνθήκη δεν είναι εύκολα επαληθεύσιμη στην πράξη, ωστόσο, με αυτήν εξασφαλίζεται το επόμενο αποτέλεσμα:

**Θεώρημα 5.3** Έστω  $\phi^{(k)}$  μία ακολουθία GEM που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ . Αν ισχύει κάποια από τις (C1, C2) και η C3, τότε

(i) όλα τα οριακά σημεία οποιασδήποτε ακολουθίας  $(\phi^{(k)})_{k \geq 0}$  GEM είναι τοπικά μέγιστα ( $\in \mathcal{M}$ ) και

(ii) Η ακολουθία  $(L(\phi^{(k)}))_{k \geq 0}$  συγκλίνει στο  $L^* = L(\phi^*)$  για κάποιο (τοπικό μέγιστο)  $\phi^* \in \mathcal{M}$ .

### Απόδειξη

Αρκεί να δείξουμε πως  $L(\phi^{(k+1)}) \geq L(\phi^{(k)})$  για οποιοδήποτε  $\phi^{(k)}$  στο συμπλήρωμα του  $\mathcal{M}$ . Είναι:

$$\begin{aligned} L(\phi^{(k+1)}) &= Q(\phi^{(k+1)}; \phi^{(k)}) + H(\phi^{(k+1)}; \phi^{(k)}) = \sup_{\phi' \in \Phi} Q(\phi', \phi^{(k)}) + H(\phi^{(k+1)}; \phi^{(k)}) \\ &> Q(\phi^{(k)}; \phi^{(k)}) + H(\phi^{(k)}; \phi^{(k)}) = L(\phi^{(k)}) \end{aligned}$$

□



### 5.4.3 Σύγκλιση της $\phi^{(k)}$

Με τη σύγκλιση της  $(L(\phi^{(k)}))_{k \geq 0}$  σε κάποιο  $L^* = L(\phi^*)$ , δεν εξασφαλίζεται η σύγκλιση της  $(\phi^{(k)})_{k \geq 0}$  σε κάποιο  $\phi^*$ . Η αντίστροφη κατεύθυνση ισχύει (με τη συνθήκη R3). Γενικότερα, κάτω από την υπόθεση ότι η ακολουθία EM ή GEM συγκλίνει, μπορούμε να έχουμε ορισμένα βασικά αποτελέσματα. Κάποια από αυτά, θα χρησιμοποιηθούν στα θεωρήματα που ακολουθούν ενώ κάποια άλλα έχουν ξεχωριστή σημασία από μόνα τους. Με τα αποτελέσματα αυτά ξεκινάμε την υποενότητα αυτή.

Αρχικά, στην περίπτωση αυτή, το  $\phi^*$  είναι και στάσιμο σημείο της  $L$ , όπως φαίνεται στην επόμενη Πρόταση.

**Πρόταση 5.4** Κάτω από τις υποθέσεις 1 και 2, και την επιπρόσθετη υπόθεση ότι η  $\nabla Q$  είναι συνεχής ως προς και τα 2 ορίσματα, αν  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία EM (αλλά όχι GEM) και  $\phi^{(k)} \rightarrow \phi^*$ , τότε,

$$\nabla_{\phi} \log L(\phi)|_{\phi=\phi^*} = 0.$$

#### Απόδειξη

Για οποιοδήποτε  $\phi' \in \Phi$ , η  $Q(\cdot; \phi') = \log L(\cdot) - H(\cdot; \phi')$  είναι συνεχώς διαφορίσιμη στο  $\Phi$  ως διαφορά συνεχών διαφορίσιμων συναρτήσεων κάτω από την υπόθεση 5.2.

Επειδή  $\phi^{(k+1)} \in \mathcal{M}(\phi)$ , δηλαδή μεγιστοποιεί την  $Q$  για κάθε  $k$ , έπεται ότι  $\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k+1)}} = 0$  για κάθε  $k$ . Από τη συνέχεια της  $\nabla Q$  έπεται πως  $\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k+1)}} \rightarrow \nabla_{\phi} Q(\phi^*; \phi^*)$  καθώς  $k \rightarrow \infty$  και άρα  $\nabla_{\phi} Q(\phi; \phi^*)|_{\phi=\phi^*} = 0$ . Η  $\nabla_{\phi} H(\phi; \phi)$  είναι συνεχής, και άρα  $\nabla_{\phi} H(\phi; \phi^{(k)})|_{\phi=\phi^{(k+1)}} \rightarrow \nabla_{\phi} H(\phi^*; \phi^*)$  καθώς  $k \rightarrow \infty$ . Τέλος, για οποιοδήποτε  $\phi \in \Phi$ , επειδή η παράγωγος μπορεί να περάσει μέσα στο ολοκλήρωμα, θα είναι:

$$\begin{aligned} \nabla_{\phi} H(\phi; \phi) &= \int \nabla_{\phi} \log k(x; \phi) k(x; \phi) \mu(dx) \\ &= \int \frac{\frac{\partial}{\partial \phi} k(x; \phi)}{k(x; \phi)} k(x; \phi) \mu(dx) \\ &= \frac{\partial}{\partial \phi} \int k(x; \phi) \mu(dx) = 0. \end{aligned}$$

το ζητούμενο έπεται άμεσα. □

Για να έχουμε το ίδιο αποτέλεσμα για μία ακολουθία GEM θα πρέπει να υποθέσουμε επιπλέον πως:

$$\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k+1)}} = 0.$$

Η παραπάνω συνθήκη ικανοποιείται για οποιαδήποτε ακολουθία EM (δεδομένου των υποθέσεων 1 και 2).

Ακόμη, αν η ακολουθία EM ή GEM,  $(\phi^{(k)})_{k \geq 0}$ , συγκλίνει σε κάποιο  $\phi^*$  και η  $\mathcal{M}$  είναι συνεχής, τότε:

$$\phi^* = \mathcal{M}(\phi^*).$$

Δηλαδή το  $\phi^*$  είναι σταθερό σημείο της απεικόνισης  $\mathcal{M}$  (Για την απόδειξη εφαρμόζουμε την αρχή της μεταφοράς).

Σε συνδυασμό λοιπόν με την προηγούμενη Πρόταση έχουμε πως:

**Πόρισμα 5.3** Κάτω από τις υποθέσεις 1 και 2

(i) Αν  $\phi^*$  σταθερό σημείο της  $\mathcal{M}$ , τότε το  $\phi^*$  είναι σημείο σύγκλισης του EM και στάσιμο σημείο της  $L$ .

(ii) Υποθέτοντας επιπλέον πως η  $Q(\phi; \cdot)$  είναι συνεχής, αν  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία GEM και  $\phi^{(k)} \rightarrow \phi^*$ , τότε το  $\phi^*$  είναι σταθερό σημείο της  $\mathcal{M}$  και στάσιμο σημείο της  $L$ .

Αν βέβαια το  $\phi^*$  είναι στάσιμο σημείο της  $L$ , τότε δεν σημαίνει πως είναι και σταθερό σημείο της  $\mathcal{M}$ .

Αν λοιπόν υπάρχει μοναδική Ε.Μ.Π., έστω  $\hat{\phi}$ , τότε αυτή θα είναι σταθερό σημείο της  $\mathcal{M}$ :

$$\hat{\phi} = \mathcal{M}(\hat{\phi}) \quad (5.4.1)$$

### Απόδειξη

Επειδή για την Ε.Μ.Π. ισχύει:  $L(\hat{\phi}) > L(\phi)$ , για οποιοδήποτε  $\phi \in \Phi, \phi \neq \hat{\phi}$ , για  $\phi = \mathcal{M}(\hat{\phi})$ , θα είναι

$$L(\hat{\phi}) \geq L(\mathcal{M}(\hat{\phi})).$$

Από την άλλη, επειδή  $\hat{\phi} \in \mathcal{M}(\hat{\phi})$ , από τη μονοτονία του αλγορίθμου, θα είναι:

$$L(\mathcal{M}(\hat{\phi})) \geq L(\hat{\phi}).$$

και έχουμε το ζητούμενο. □

Ωστόσο, δεν υπάρχει κάποια εγγύηση ότι η  $\hat{\phi}$  θα είναι και μοναδικό σταθερό σημείο. Ακόμη, από την 5.3.9, η Ε.Μ.Π. μεγιστοποιεί την  $Q$  καθώς για κάθε  $\phi \in \Phi$  ισχύει:

$$Q(\hat{\phi}, \hat{\phi}) \geq Q(\phi, \hat{\phi})$$

Διαφορετικά, θα υπήρχε κάποιο  $\phi_0$  τέτοιο ώστε:

$$Q(\hat{\phi}, \hat{\phi}) \geq Q(\phi_0, \hat{\phi})$$

και έτσι θα ήταν:  $L(\hat{\phi}) < L(\phi_0)$  το οποίο δεν γίνεται.

Στα προηγούμενα είχαμε ως δεδομένο πως είτε η ακολουθία EM είτε η GEM συγκλίνει. Σε όσα ακολουθούν θα εξεταστούν συνθήκες με τις οποίες η σύγκλιση αυτή μπορεί να εξασφαλισθεί. Στο επόμενο Θεώρημα του Wu (1983), παρουσιάζονται ικανές συνθήκες για τη σύγκλιση αυτή. Αρχικά ορίζουμε το σύνολο:

$$\mathcal{L}(a) = \{\phi \in \Phi : L(\phi) = a\}$$

**Θεώρημα 5.4** Έστω  $\phi^{(k)}$  μία ακολουθία GEM που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ . Υποθέτουμε ότι:

- (i) Η  $\nabla Q$  είναι συνεχής ως προς και τα 2 ορίσματα, και
- (ii)  $\nabla_{\phi} Q(\phi; \phi^{(k)})|_{\phi=\phi^{(k+1)}} = 0$
- (iii) Είτε

$$\mathcal{L}(L^*) = \{\phi^*\}$$

είτε

Το  $\mathcal{L}(L^*)$  είναι πεπερασμένο και  $\|\phi^{(k+1)} - \phi^{(k)}\| \rightarrow 0$  καθώς  $k \rightarrow \infty$

Τότε,

$(\phi^{(k)}) \rightarrow \phi^*$  καθώς  $k \rightarrow \infty$  για κάποιο στάσιμο σημείο  $\phi^*$ , με  $L(\phi^*) = L^*$ , όπου  $L^*$ , το όριο της  $(L(\phi^{(k)}))_{k \geq 0}$

### Απόδειξη

Από την υπόθεση 5.3 έπεται πως  $L(\phi^{(k)}) \rightarrow L^*$ . Στην περίπτωση που  $\mathcal{L}(L^*) = \{\phi^*\}$ , συμπεραίνουμε ότι  $(\phi^{(k)}) \rightarrow \phi^*$ . Στην περίπτωση όπου το  $\mathcal{L}(L^*)$  είναι πεπερασμένο (αλλά όχι μονοσύνολο), είναι σαφές πως η συνθήκη  $\|\phi^{(k+1)} - \phi^{(k)}\| \rightarrow 0$  είναι ικανή για τη σύγκλιση της  $(\phi^{(k)})_{k \geq 0}$  στο  $\phi^*$ . Οπότε, σε κάθε περίπτωση:  $(\phi^{(k)}) \rightarrow \phi^*$ . Από την Πρόταση 5.4, το  $\phi^*$  είναι στάσιμο σημείο της  $L$ .  $\square$

Το πλεονέκτημα του παραπάνω θεωρήματος είναι πως δεν απαιτεί τις συνθήκες (i) και (ii) του θεωρήματος 1 και μία ειδική περίπτωση είναι η εξής:

**Πόρισμα 5.3** Αν η  $L$  έχει μοναδική κορυφή στον  $\Phi$  και το  $\phi^*$  είναι το μοναδικό στάσιμο σημείο και η  $\nabla Q$  είναι συνεχής ως προς και τα 2 ορίσματα, τότε, για οποιαδήποτε ακολουθία EM,  $(\phi^{(k)})_{k \geq 0}$ , ισχύει  $(\phi^{(k)}) \rightarrow \phi^*$ , όπου το  $\phi^*$  είναι το μοναδικό μέγιστο της  $L$ .

Ειδικά το παραπάνω πόρισμα παρέχει εύκολα επαληθεύσιμες συνθήκες που ικανοποιούνται σε ένα πολύ μεγάλο εύρος εφαρμογών. Για το Θεώρημα 5.5, στην περίπτωση του EM, η συνθήκη στο (ii) ικανοποιείται αυτόματα όπως δείχθηκε στην Πρόταση (5.2).

Για τη σύγκλιση της ακολουθίας GEM σε στάσιμο σημείο ή σημείο τοπικού μεγίστου, ικανές συνθήκες δίνουν και τα επόμενα 2 θεωρήματα του Wu (1983) τα οποία όμως υποθέτουν τις συνθήκες (i) και (ii) του θεωρήματος 1.

(Οι αποδείξεις τους είναι άμεσες αν ακολουθηθούν τις γραμμές της απόδειξης του θεωρήματος 5.4.) Έστω:

$$\mathcal{S}(a) = \{\phi \in \mathcal{S} : L(\phi) = a\} \text{ και } \mathcal{M}(a) = \{\phi \in \mathcal{M} : L(\phi) = a\}$$

**Θεώρημα 5.5** Έστω  $(\phi^{(k)})_{k \geq 0}$  ακολουθία GEM που ικανοποιεί τις (i) και (ii) του θεωρήματος 1. Αν  $\mathcal{S}(L^*) = \{\phi^*\}$  (αντίστοιχα  $\mathcal{M}(L^*) = \{\phi^*\}$ ), όπου  $L^*$  το οριακό σημείο της  $L(\phi_n)$  στο Θεώρημα 1, τότε  $\phi_n \rightarrow \phi^*$

Η συνθήκη  $\mathcal{S}(L^*) = \{\phi^*\}$  μπορεί να χαλαρώσει υποθέτοντας  $\|\phi^{(k+1)} - \phi^{(k)}\| \rightarrow 0$ , μία συνθήκη αναγκαία για το επιθυμητό αποτέλεσμα:  $(\phi^{(k)}) \rightarrow \phi^*$

**Θεώρημα 5.6** Έστω  $(\phi^{(k)})_{k \geq 0}$  ακολουθία GEM που ικανοποιεί τις (i) και (ii) του θεωρήματος 1. Αν  $\|\phi^{(k+1)} - \phi^{(k)}\| \rightarrow 0$ , τότε όλα τα οριακά σημεία της ακολουθίας  $(\phi^{(k)})_{k \geq 0}$ , βρίσκονται σε ένα συνεκτικό και συμπαγές υποσύνολο του  $\mathcal{S}(L^*)$

(αντίστοιχα:  $\mathcal{M}(L^*)$ , όπου το  $L^*$  είναι το όριο της  $(L(\phi^{(k)}))_{k \geq 0}$  του θεωρήματος 1. Συγκεκριμένα, αν το  $\mathcal{S}(L^*)$  (αντίστοιχα:  $\mathcal{M}(L^*)$  είναι πεπερασμένο, τότε  $(\phi^{(k)}) \rightarrow \phi^*$  με  $\phi^* \in \mathcal{S}(L^*)$  (αντίστοιχα:  $\mathcal{M}(L^*)$ ).

Επειδή τα  $\mathcal{M}(L)$  και  $\mathcal{S}(L)$  είναι υποσύνολα του  $\mathcal{L}(L)$ , οι συνθήκες που αφορούν τα σύνολα αυτά στα θεωρήματα 5.5, 5.6 είναι ασθενέστερες από τις αντίστοιχες συνθήκες στο Θεώρημα 5.4. Μάλιστα, την περίπτωση που το σύνολο στάσιμων σημείων της  $L$ ,  $\mathcal{S}$ , αποτελείται από μεμονωμένα σημεία, μελέτησε ο Vaida (2005).

Συγκεκριμένα, υπό την υπόθεση πως το  $\mathcal{S}$  αποτελείται από μεμονωμένα σημεία, για οποιοδήποτε πραγματικό αριθμό  $L^*$ , υπάρχουν το πολύ πεπερασμένα στάσιμα σημεία  $\phi^*$  τέτοια ώστε  $L(\phi^*) = L^*$ . Αυτό είναι μπορούμε να το δούμε υποθέτοντας πως το  $\mathcal{S}(L^*)$  είναι άπειρο και καταλήγοντας σε άτοπο:

Έστω  $L(\phi^*) = L^*$  για κάποιο στάσιμο σημείο  $\phi^*$ . Τότε το  $\mathcal{S}(L^*)$  είναι φραγμένο καθώς είναι υποσύνολο του  $\Phi_{\phi_0}$ . Από την υπόθεση πως είναι και άπειρο, υπάρχει ακολουθία σημείων του  $\mathcal{S}(L^*)$ , έστω  $(\phi_n^*)$  με  $\phi_n^* \rightarrow \phi_0 \in \Phi_{\phi_0}$ . Επειδή η συνάρτηση  $\nabla_{\phi} L(\phi)$  είναι συνεχής, έπεται πως  $\nabla_{\phi} L(\phi_n^*) \rightarrow \nabla_{\phi} L(\phi_0)$ . Επειδή  $\nabla_{\phi} L(\phi_n^*) = 0$  για κάθε  $n$ , θα είναι  $\nabla_{\phi} L(\phi_0) = 0$  και άρα  $\phi_0 \in \mathcal{S}$ . Από το τελευταίο, έπεται πως το  $\phi_0$  είναι σημείο συσσώρευσης, γεγονός που αντικρούει την υπόθεση.

Ένα σύνολο από τιμές  $\phi_1^*, \phi_2^*, \dots, \phi_m^*$  λέγεται κύκλος μήκους  $m \geq 2$  για την απεικόνιση  $\mathcal{M}$  αν  $\mathcal{M}(\phi_i^*) = \phi_{i+1}^*$  για  $i = 1, \dots, m-1$  και  $\mathcal{M}(\phi_m^*) = \phi_1^*$ .

Υπό την υπόθεση πως το  $\mathcal{S}$  αποτελείται μόνο από μεμονωμένα σημεία, ο Vaida (2005) διατύπωσε τα ακόλουθα 2 αποτελέσματα:

**Θεώρημα 5.7** Έστω  $(\phi^{(k)})_{k \geq 0}$  ακολουθία EM. Υποθέτουμε πως απεικόνιση  $\mathcal{M}$  είναι συνεχής στα στάσιμα σημεία του  $\mathcal{S}$ . Τότε είτε η  $(\phi^{(k)})_{k \geq 0}$  συγκλίνει σε κάποιο στάσιμο σημείο  $\phi^*$  με  $\mathcal{M}(\phi^*) = \phi^*$ , είτε υπάρχει ένα πεπερασμένο σύνολο  $C = \{\phi_1^*, \dots, \phi_m^*\}$ , τέτοιο ώστε:

- (i) Τα  $\phi_1^*, \dots, \phi_m^*$  είναι στάσιμα σημεία με  $L(\phi_i^*) = L^*$  για κάποιο  $L^*$  και για κάθε  $i = 1, \dots, m$  (δηλαδή με την ίδια τιμή πιθανοφάνειας).
- (ii) Τα  $\phi_1^*, \dots, \phi_m^*$  είναι κύκλος μήκους  $m$  για τη  $\mathcal{M}$ .
- (iii) Οι παράλληλες υπακολουθίες  $\{\phi^{mt+i} : t \geq 1\}$  είναι τέτοιες ώστε  $\phi^{mt+i} \rightarrow \phi_i^*$  καθώς  $t \rightarrow \infty$  για  $i = 1, \dots, m$ .

Είναι σαφές από το παραπάνω πως τα οριακά σημεία του EM συμπίπτουν με τα σταθερά σημεία και τους κύκλους της απεικόνισης  $\mathcal{M}$ .

Ο Vaida (2005) σημειώνει πως σπανίως συναντάμε στην πράξη αυτή την κυκλική συμπεριφορά που εμφανίζεται στο συμπέρασμα του θεωρήματος, ωστόσο στην εργασία του παραθέτει ένα παράδειγμα. Το παραπάνω Θεώρημα παίζει ουσιαστικό ρόλο για την απόκτηση του παρακάτω κεντρικού αποτελέσματος:

**Θεώρημα 5.8** Υποθέτουμε πως για κάθε στάσιμο σημείο  $\phi \in \mathcal{S}$ , υπάρχει ολικό μέγιστο για την  $Q(\cdot; \phi)$ . Τότε, για οποιαδήποτε αρχική τιμή  $\phi^{(0)}$ , η ακολουθία EM,  $(\phi^{(k)})_{k \geq 0}$ , που παράγεται από αυτή, συγκλίνει σε κάποιο στάσιμο σημείο  $\phi^*$ . Επιπλέον,

## Σύγκλιση του αλγορίθμου EM

$\mathcal{M}(\phi^*) = \phi^*$  και αν  $\phi^{(k)} \neq \phi^*$  για κάθε  $k$ , τότε για την ακολουθία τιμών πιθανοφάνειας ισχύει  $L(\phi^{(k)}) \rightarrow L(\phi^*)$ .

Το Θεώρημα 5.8 προσφέρει μία γενική, επαληθεύσιμη συνθήκη για τη σύγκλιση η οποία σχεδόν πάντα ικανοποιείται στην πράξη και ικανοποιείται, για παράδειγμα, ελέγχοντας το πρόσημο της παραγώγου της συνάρτησης που πρόκειται να μεγιστοποιηθεί στο Βήμα-M. Επίσης, η συνθήκη αυτή είναι και ικανή ώστε η  $\mathcal{M}$  να είναι συνεχής στα στάσιμα σημεία του  $\mathcal{S}$ , δηλαδή για τη συνθήκη του θεωρήματος 5.7.

## 5.5 Η ειδική περίπτωση της εκθετικής οικογένειας

Στην ενότητα αυτή περιγράφεται η σύνδεση της εκθετικής οικογένειας με τον αλγόριθμο EM.

Με το συμβολισμό αυτού του κεφαλαίου, το διάνυσμα  $X$  αντιπροσωπεύει τα πλήρη δεδομένα, και έχουμε υποθέσει ότι έχει πυκνότητα  $f(\cdot; \phi)$  (ως προς το μέτρο αναφοράς  $\mu$ ), όπου  $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$ . Υποθέτουμε επιπλέον πως η πυκνότητα αυτή είναι μέλος εκθετικής οικογένειας. Στην περίπτωση αυτή, η συνάρτηση πιθανοφάνειας των πλήρη δεδομένων,  $L_c$ , είναι:

$$L_c(\phi) = C(\phi)^{-1} e^{\phi \cdot t(x)},$$

όπου  $C(\phi) = \int e^{\phi \cdot t(x)} \mu(dx)$ , η σταθερά κανονικοποίησης και το  $\cdot$  συμβολίζει το εσωτερικό γινόμενο. Η συνάρτηση  $t : \mathcal{X} \mapsto \mathbb{R}^{d_\phi}$  είναι επαρκής στατιστική συνάρτηση, ίδιας διάστασης με το  $\phi$ .

Η έκφραση αυτή, εμπλέκει λιγότερες συναρτήσεις από τη συνηθισμένη μορφή, και καθιστά έτσι αρτιότερη την αναπαράσταση των αποτελεσμάτων. Για όσα ακολουθούν κάνουμε τις ακόλουθες υποθέσεις:

### Υπόθεση 5.4

(i) Η  $X$  παίρνει τιμές στον  $\mathcal{X}$  που είναι υποσύνολο Ευκλείδειου χώρου.

(ii) Υποθέτουμε πως το ολοκλήρωμα στη  $C(\phi)$  είναι πεπερασμένο (διαφορετικά η  $f$  δεν θα ήταν πυκνότητα) και το εσωτερικό του συνόλου  $N = \{\phi : C(\phi) < \infty\}$  είναι ο φυσικός παραμετρικός χώρος.

(iii) Υποθέτουμε επίσης πως ο παραμετρικός χώρος  $\Phi$  είναι ανοικτό σύνολο (η εκθετική οικογένεια σε αυτή την περίπτωση λέγεται κανονική (regular», καθώς τα περισσότερα αποτελέσματα δεν ισχύουν στο σύνορο (κυρίως αυτά που εμπλέκουν τις παραγώγους).

Αρχικά θα αναφέρουμε τα βασικά αποτελέσματα για την  $L_c$  (Sundberg (1974)). Από το Θεώρημα αλλαγής μέτρου η συνάρτηση  $C(\phi)$  εκφράζεται ως:

$$C(\phi) = \int e^{\phi \cdot t} \nu(dt),$$

όπου με  $\nu$  συμβολίζεται το μέτρο που επάγει το  $\mu$  μέσω του μετασχηματισμού  $x \mapsto t(x)$  και δίνεται από την:

$$\nu(A) = \mu(t^{-1}(A)) = \int_{t^{-1}(A)} \mu(dx), \quad A \in \mathcal{B}(\mathbb{R}^{d_\phi}).$$

Από την παραπάνω μορφή, παρατηρούμε πως η  $C(\phi)$  είναι ο μετασχηματισμός Laplace για το μέτρο  $\nu$ .

## Η ειδική περίπτωση της εκθετικής οικογένειας

Το βασικό αποτέλεσμα είναι πως η συνάρτηση  $C(\zeta)$ , είναι αναλυτική στο (ανοιχτό)  $\Phi$ , όπου  $\zeta$  μιγαδικός με  $Re\zeta = \phi$ . Οπότε είναι απείρως διαφορίσιμη και η παράγωγος υπολογίζεται μέσα από το ολοκλήρωμα. Επομένως:

$$\nabla_{\phi}^{(m)} C(\phi) = \int \nabla_{\phi}^{(m)}(e^{\phi \cdot t}) \nu(dt) = \int t^m e^{\phi \cdot t} \nu(dt) = C(\phi) E_{\phi}[t^m], \quad (5.5.1)$$

επειδή  $C(\phi) > 0$  και ανεξάρτητη της  $t$  (σ.π.). Οπότε,

$$E_{\phi}[t^m] = \frac{\nabla_{\phi}^{(m)} C(\phi)}{C(\phi)}$$

Για  $m = 1$  είναι:

$$E_{\phi}[t] = \nabla_{\phi} \log C(\phi)$$

Ακόμη, παραγωγίζοντας την  $\log C(\phi)$ ,  $m$  φορές, παίρνουμε τα cumulants  $k^{(m)}$ . Επειδή τα cumulants για  $m = 2, 3$  ταυτίζονται με τις αντίστοιχες κεντρικές ροπές ( $\mu_r = E_{\phi}[(X - \mu)^r]$ ) για  $r = 2, 3$ , η δεύτερη και τρίτη παράγωγος της  $\log C(\phi)$  θα είναι ίσες με τη διασπορά και τη 3η κεντρική ροπή της  $t$  αντίστοιχα. Ειδικότερα για τη διασπορά (ή πίνακας συνδιακυμάνσεων)

$$\nabla_{\phi}^2 \log C(\phi) = Var_{\phi}[t].$$

Επομένως, η εξίσωση πιθανοφάνειας ανάγεται στην:

$$\frac{1}{n} \sum_{i=1}^n t(x_i) = E_{\phi}[t], \quad (5.5.2)$$

αφού  $\nabla_{\phi} \log L_c(\phi) = \nabla_{\phi} \log(C(\phi)^{-1} e^{\phi \cdot t}) = t - \nabla_{\phi} \log C(\phi) = t - E_{\phi}[t]$ .

Επίσης,

$$\begin{aligned} Var_{\phi}[t] &= \nabla_{\phi}^2 \log C(\phi) = \nabla_{\phi}^2 (-\log(C(\phi)^{-1})) \\ &= -\nabla_{\phi}^2 (\log L_c(\phi) - \log e^{\phi \cdot t}) = \nabla_{\phi}^2 (\phi \cdot t) - \nabla_{\phi}^2 \log L_c(\phi) \\ &= -\nabla_{\phi}^2 \log L_c(\phi) \end{aligned}$$

Είναι λοιπόν:  $-\nabla_{\phi}^2 \log f(x; \phi) = \nabla_{\phi}^2 \log C(\phi)$  και επειδή το δεύτερο μέλος είναι σταθερά, παίρνοντας μέσες τιμές, καταλήγουμε στην:

$$I(\phi) = -E[\nabla_{\phi}^2 \log f(x; \phi)] = Var_{\phi}[t]$$

Από την τελευταία, ο  $Var_{\phi}[t]$  είναι ο πίνακας πληροφορίας του Fisher και είναι θετικά ορισμένος. Από το γεγονός αυτό, συμπεραίνουμε πως η  $\log C(\phi)$  είναι κυρτή και υπάρχει 1-1 αντιστοιχία μεταξύ του  $\phi$  και του  $E_{\phi}[t]$ . Έπεται πως υπάρχει το πολύ μία ρίζα,  $\hat{\phi}$ , της εξίσωσης (5.5.2) στο  $\Phi$ , και όταν αυτή υπάρχει, μεγιστοποιεί την  $L_c$  στο  $\Phi$  και επομένως είναι η Ε.Μ.Π.

Επίσης, τα μέλη της εκθετικής οικογένειας μοιράζονται και πολύ ισχυρές ασυμπτωτικές ιδιότητες. Περισσότερα λοιπόν υπάρχουν στα:

## Η ειδική περίπτωση της εκθετικής οικογένειας

Για την πυκνότητα των παρατηρούμενων δεδομένων  $k_y(x; \phi)$  τα βασικότερα αποτελέσματα είναι πως και αυτή αποτελεί μέλος εκθετικής οικογένειας με ίδιο φυσικό παραμετρικό χώρο και ίδια επαρκή στατιστική συνάρτηση αλλά ορισμένη σε διαφορετικούς χώρους. ( $\mathcal{X}$  και  $h^{-1}(y) = \{x \in \mathcal{X} : h(x) = y\}$  για τις  $f$  και  $k_y$  αντίστοιχα).

Έστω  $\nu$  η περιθώρια κατανομή του  $Y$  και για κάθε  $y \in \mathcal{Y}$ , συμβολίζουμε με  $k_y = k(x|y; \phi)$  τη δεσμευμένη κατανομή του  $X$  δοθέντος  $Y = y$ . Τότε για οποιοδήποτε  $\phi \in \Phi$ , και οποιοδήποτε Borel,  $A \in \mathcal{Y}$  είναι:

$$\mathbb{P}(h(x) \in A; \phi) = \int_{h^{-1}(A)} \frac{1}{C(\phi)} e^{\phi \cdot t(\theta)} dx \quad (5.5.3)$$

$$= \int_{h^{-1}(A)} \frac{1}{C(\phi)} \left( \int e^{\phi \cdot t(x)} dk_y(x) \right) d\nu(y). \quad (5.5.4)$$

Θέτοντας  $C_y(\phi) = \int e^{\phi \cdot t(x)} dk_y(x)$ , από την παραπάνω, για κάθε  $\phi \in \Phi$ , η πυκνότητα της  $Y$  (ως προς το  $\nu$ ) είναι:

$$g(y; \phi) = \frac{C_y(\phi)}{C(\phi)}$$

και η δεσμευμένη πυκνότητα του  $X|Y = y$  (ως προς το  $k_y$ ) είναι:

$$k(x|y; \phi) = \frac{f(x; \phi)}{g(y; \phi)} = \frac{1}{C_y(\phi)} e^{\phi \cdot t(x)}.$$

Η  $C_y(\phi)$ , για κάθε  $\phi$ , ως συνάρτηση του  $y$  είναι σχεδόν παντού πεπερασμένη, καθώς είναι ανάλογη της συνάρτησης πυκνότητας, ενώ, για σταθερό  $y$  ως συνάρτηση του  $\phi$  έχει τις ίδιες ιδιότητες με την  $C(\phi)$ . Συγκεκριμένα, για οποιοδήποτε  $y$ , η  $C_y(\zeta)$  είναι αναλυτική στο  $\{\phi \in \Phi\}$ , όπου  $\zeta$  μιγαδικός αριθμός με  $\phi = \text{Re}\zeta$ . Επομένως υπάρχουν οι παράγωγοι όλων των τάξεων και μπορούν να περάσουν μέσα στο ολοκλήρωμα. Οπότε:

$$\nabla_{\phi}^{(m)} C_y(\phi) = C_y(\phi) E_{\phi}[t^m|y] \quad (5.5.5)$$

$$\nabla_{\phi} \log C_y(\phi) = E_{\phi}[t|y] \quad (5.5.6)$$

$$\nabla_{\phi}^2 \log C_y(\phi) = \text{Var}_{\phi}[t|y] \quad (5.5.7)$$

Επιπλέον, η εξίσωση της πιθανοφάνειας ( $\tilde{L}$ ) για το δείγμα  $(y_1, \dots, y_n)$  είναι:

$$\frac{1}{n} \sum_{i=1}^n E_{\phi}[t|y_i] = E_{\phi}[t] \quad (5.5.8)$$

**Πρόταση 5.5** Συνδυάζοντας τις (5.5.1, 5.5.2) μπορούμε να πάρουμε πως:

$$\log L(\phi) = \log p(y; \phi) = \log \left( \frac{C_y(\phi)}{C(\phi)} \right) = \log C_y(\phi) - \log C(\phi)$$

και επομένως,

$$\nabla_{\phi} \log L(\phi) = E_{\phi}[t|y] - E_{\phi}[t] \quad (5.5.9)$$

Ο αλγόριθμος EM για την περίπτωση που τα πλήρη δεδομένα έχουν πυκνότητα μέλος της εκθετικής οικογένειας, έχει μία πιο κομψή μορφή. Συγκεκριμένα, για  $\phi, \phi' \in \Phi$ , η συνάρτηση  $Q(\phi; \phi')$  είναι:

$$Q(\phi; \phi') = E_{\phi'}[\log f(x; \phi)|y] = -\log C(\phi) + \phi \cdot E_{\phi'}[t|y].$$



## Η ειδική περίπτωση της εκθετικής οικογένειας

Για τη μεγιστοποίηση της  $Q$ , παίρνοντας την παράγωγο ως προς το  $\phi$  και παρατηρώντας την (5.5.2), καταλήγουμε στην εξίσωση:

$$E_{\phi}[t] = E_{\phi'}[t|y].$$

Οπότε έχουμε το εξής αποτέλεσμα:

**Πόρισμα 5.4** (Αλγόριθμος EM για εκθετική οικογένεια). Έστω  $(\phi^{(k)})$  η τρέχουσα τιμή για το  $\phi$  έπειτα από  $k$  επαναλήψεις του αλγορίθμου EM. Τότε για τον επόμενο κύκλο τα βήματα είναι:

Βήμα-E. Εκτίμηση της στατιστικής συνάρτησης  $t(x)$  υπολογίζοντας την:

$$t^{(k)} = E_{\phi^{(k)}}[t(x)|y]$$

Βήμα-M. Το  $(\phi^{(k+1)})$  καθορίζεται από την εξίσωση:

$$E_{\phi}[t(x)]$$

Μερικές παρατηρήσεις: Οι εξισώσεις στο Βήμα-M δεν έχουν πάντα λύση στο (ανοικτό)  $\Phi$ . Σε αυτές τις περιπτώσεις, η λύση βρίσκεται στο σύνορο του παραμετρικού χώρου. Από την άλλη, αν υπάρχει λύση, τότε αυτή είναι μοναδική εξαιτίας της κυρτότητας της πιθανοφάνειας που συζητήθηκε στα προηγούμενα.

## 5.6 Η περίπτωση του Monte Carlo EM

Μία δυσκολία στη εφαρμογή του EM είναι πως κάθε βήμα-E απαιτεί τον υπολογισμό της συνάρτησης  $Q(\phi; \phi')$ . Οι Wei and Tanner (1990a,b) πρότειναν μία λύση για να ξεπεραστεί η δυσκολία αυτή, υιοθετώντας τεχνικές Monte Carlo. Συγκεκριμένα, το πρόβλημα μπορεί να ξεπεραστεί προσομοιώνοντας  $Z_1, \dots, Z_N$  από την δεσμευμένη κατανομή  $k(x|y; \phi)$  της  $X|Y = y$  και στη συνέχεια μεγιστοποιώντας την κατά προσέγγιση συνάρτηση πιθανοφάνειας για τα πλήρη δεδομένα ( $L_c$ ):

$$\hat{Q}_N(\phi; \phi') = \frac{1}{N} \sum_{i=1}^N \log L_c(\phi)$$

Στο όριο, η παραπάνω ποσότητα συγκλίνει στη  $Q(\phi; \phi')$  και επομένως η οριακή μορφή του MC EM είναι ο κανονικός EM. Όταν  $X = (Z, Y)$ , τότε:

$$\hat{Q}_N(\phi; \phi') = \frac{1}{N} \sum_{i=1}^N \log f(z^{(k)}, y; \phi)$$

Από την άλλη, η μεγιστοποίηση του παραπάνω αθροίσματος μπορεί να έχει εξίσου δυσκολίες. Ωστόσο, στην περίπτωση της εκθετική οικογένειας μπορούμε να έχουμε αναλυτικές λύσεις.

Σε γενικές γραμμές, (αν αυτό δεν είναι προφανές) η μελέτη της οριακής συμπεριφοράς το MC-EM είναι πολύ πιο δύσκολη από αυτή για τον κανονικό EM και δεν είναι λίγοι εκείνοι που ασχολήθηκαν με τα ζητήματα αυτά. Τα βασικότερα αποτελέσματα διατυπώνονται στα:

Για να έχουμε μία εικόνα για το ζήτημα αυτό, αναφέρουμε μερικά από αυτά τα αποτελέσματα εν συντομία. Οι Chan and Ledolter (1995) έδειξαν πως για οποιαδήποτε αρχική τιμή, η ακολουθία που παράγεται από τον MC-EM, θα είναι οσοδήποτε κοντά στο μέγιστο της παρατηρούμενης πιθανοφάνειας,  $L$ , με μεγάλη πιθανότητα. Οι υποθέσεις ωστόσο είναι κάπως διαφορετικές. Η βασική απαίτηση είναι πως η σύγκλιση του παραπάνω στον κανονικό EM (καθώς  $N \rightarrow \infty$  θα πρέπει να είναι ομοιόμορφη στα συμπαγή υποσύνολα του  $\Theta$ . Μία υπόθεση αρκετά δεσμευτική και δύσκολα επαληθεύσιμη, ωστόσο στο οι Chan and Ledolter (1995), οι συγγραφείς διατύπωσαν ικανές συνθήκες που να την εγγυούνται. Στην πράξη όμως, το Θεώρημα αυτό μας λέει πως η ακολουθία, με μεγάλη πιθανότητα, κάποια στιγμή θα είναι κοντά στην Ε.Μ.Π. Αυτό όμως ακόμα και αν επιλεγεί ένα πολύ μεγάλο  $N$ , δεν υπάρχει τρόπος να γνωρίζουμε τότε συνέβη, ούτε αν έχει συμβεί. Ένα πιο ισχυρό αποτέλεσμα θα ήταν να διατυπωθούν συνθήκες κάτω από τις οποίες ο αλγόριθμος πλησιάζει την Ε.Μ.Π. και επιπλέον, παραμένει κοντά σε αυτή.

## Κεφάλαιο 6

### Η μέθοδος της Γκαουσιανής τυχαιοποίησης

#### 6.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η στατιστική μέθοδος της Γκαουσιανής τυχαιοποίησης (Gaussian Randomization) η οποία προτάθηκε στο [12]. Πρόκειται για μία νέα μέθοδο εκτίμησης παραμέτρων για σύνθετα μοντέλα, η οποία παρέχει τη δυνατότητα εύρεσης της Ε.Μ.Π. όταν η αναλυτική της έκφραση είτε είναι αδύνατη είτε παρουσιάζει σημαντικές δυσκολίες.

Η βασική ιδέα πίσω από τη μέθοδο αυτή είναι να μεταφερθεί το πρόβλημα της μεγιστοποίησης σε πλαίσιο στο οποίο να μπορεί να επιλυθεί με την εφαρμογή του αλγορίθμου EM. Συγκεκριμένα, αυτό επιτυγχάνεται τυχαιοποιώντας κατάλληλα το διάνυσμα των παραμέτρων (ή ένα μέρος του). Η τυχαιοποίηση γίνεται με τέτοιο τρόπο, ώστε το νέο πρόβλημα που θα προκύψει, θα μπορεί να αντιμετωπιστεί στο πλαίσιο των προβλημάτων με μη πλήρη δεδομένα (incomplete data problems) -ένα πλαίσιο στο οποίο είναι γνωστό από τη σχετική θεωρία (κεφάλαιο 5) ότι προσαρμόζεται ο αλγόριθμος EM. Τα δύο αυτά προβλήματα, δείχνεται πως είναι ισοδύναμα ως προς τις λύσεις τους και ως εκ τούτου αρκεί να μπορεί βρεθεί λύση για το δεύτερο, κατασκευάζοντας μία κατάλληλη έκδοση του αλγορίθμου EM.

Στην παρούσα εργασία θα ασχοληθούμε με τις θεωρητικές ιδιότητες της Γκαουσιανής τυχαιοποίησης και ο στόχος είναι να αναπτύξουμε ένα πλαίσιο για τη θεωρία σύγκλισης του EM που αφορά τη μέθοδο αυτή. Η δυσκολία που αντιμετωπίζουμε είναι ότι το επιθυμητό σημείο σύγκλισης του αλγορίθμου EM βρίσκεται στο σύνορο του παραμετρικού χώρου. Η γενική θεωρία σύγκλισης που παρουσιάστηκε στην ενότητα 5.4 αποτυγχάνει να εγγυηθεί τη σύγκλιση για την περίπτωση αυτή. Για το λόγο αυτό, με εφιαλτήριο τις ιδέες του Daniel Nettleton διατυπώνονται συνθήκες με τις οποίες μπορούμε να έχουμε εγγύηση για τη ζητούμενη σύγκλιση.

Η μέθοδος της Γκαουσιανής τυχαιοποίησης κατασκευάστηκε αρχικά για την εκτίμηση σε μοντέλα ανάπτυξης φυτών, αλλά όπως τονίζουν οι συγγραφείς στο [12], το πλαίσιο της εφαρμογής της είναι αρκετά ευρύτερο.

#### 6.2 Ορισμός και Βασικές Ιδιότητες

**Πλαίσιο.** Έστω  $m = \{(\Omega, \mathcal{F}, \mathbb{P}_\mu), \mu \in M\}$ , ένα στατιστικό μοντέλο όπου το  $M$  (παραμετρικός χώρος) είναι υποσύνολο του  $\mathbb{R}^{d_\mu}$ . Έστω ακόμη, το τυχαίο διάνυσμα  $Y$  (που αντιπροσωπεύει τα δεδομένα) ορισμένο στο παραπάνω χώρο με τιμές στον χ.μ.  $\{(\mathcal{Y}, \mathcal{Y}, \nu)\}$ , όπου το  $\nu$  είναι το μέτρο αναφοράς.

Υποθέτουμε ότι για κάθε  $\mu \in M$  το διάνυσμα  $Y$  έχει πυκνότητα (ως προς το  $\nu$ )  $p(y; \mu)$  για μία παρατήρηση  $Y = y$  και έστω  $L(\mu)$  η συνάρτηση πιθανοφάνειας.

Σε όσα ακολουθούν υποθέτουμε ότι:

**Υπόθεση 6.1**

- (i) Η συνάρτηση πιθανοφάνειας  $L(\mu)$  είναι συνεχής στο  $M$ .
- (ii) Το μοντέλο έχει μοναδική Ε.Μ.Π., έστω  $\hat{\mu}$ . Δηλαδή:

$$0 \leq L(\mu) \leq L(\hat{\mu}) \quad \text{για κάθε } \mu \in M,$$

και η τελευταία ανισότητα είναι γνήσια για  $\mu \neq \hat{\mu}$ .

Ο όρος τυχαιοποίηση αφορά τις παραμέτρους που είναι επιθυμητό να εκτιμηθούν. Υπό το πρίσμα αυτό, μπορούμε να θεωρήσουμε ότι  $\mu = (\mu_1, \mu_2)$  και εστιάζουμε στην εκτίμηση του  $\mu_1$ . Η περίπτωση αυτή θα αναφέρεται ως μερική τυχαιοποίηση όταν  $d_\mu > d_{\mu_1}$ . Παρουσιάζεται στη συνέχεια ο ορισμός της μεθόδου Γκαουσιανής τυχαιοποίησης για το  $\mu_1$  υποθέτοντας επιπλέον ότι  $M_1 = \mathbb{R}^{d_{\mu_1}}$ . Φυσικά, η τελευταία υπόθεση μπορεί να χαλαρώσει αν υπάρχει κατάλληλη αναπαραμέτρηση, έστω  $\mu_1 \mapsto g(\mu_1)$  με  $g(M_1) = \mathbb{R}^{d_{\mu_1}}$ . Επίσης, σε όσα ακολουθούν, υποθέτουμε ότι το  $M_2$  είναι ανοικτό σύνολο.

**Ορισμός 6.1** (Γκαουσιανή Τυχαιοποίηση). Έστω  $m$  ένα στατιστικό μοντέλο που ικανοποιεί την υπόθεση 6.1 και  $\mu = (\mu_1, \mu_2)$ , με  $\mu_1 \in \mathbb{R}^{d_{\mu_1}}$ . Το στατιστικό μοντέλο  $\tilde{m}(\mu_1)$  καλείται Γκαουσιανή τυχαιοποίηση του  $m$  ως προς το  $\mu_1$  αν:

Το  $\tilde{m}(\mu_1)$  είναι ένα μοντέλο μη πλήρων δεδομένων (incomplete data model) που αποτελείται από:

- (i) ένα μη παρατηρήσιμο τυχαίο διάνυσμα  $\Theta$ , όπου

$$\Theta \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma), \tag{6.2.1}$$

με  $\Sigma = \text{diag}\{\sigma_i^2\}_{1 \leq i \leq d_{\mu_1}}$ , όπου  $\sigma_i^2 > 0$  και

- (ii) ένα παρατηρήσιμο διάνυσμα  $Y$ , όπου η δεσμευμένη κατανομή του  $Y|\Theta = \theta$  εξαρτάται μόνο από την παράμετρο  $\mu_2$  και ικανοποιεί την:

$$p(y|\theta; \mu_2) = L(\theta, \mu_2). \tag{6.2.2}$$

Από τον παραπάνω ορισμό παρατηρούμε ότι το  $\tilde{m}(\mu_1)$  περιλαμβάνει τις παραμέτρους του αρχικού μοντέλου που είναι επιθυμητό να εκτιμηθούν καθώς και το διάνυσμα παραμέτρων  $(\sigma_i^2)_{1 \leq i \leq d_{\mu_1}} \in (\mathbb{R}_{>0})^{d_{\mu_1}}$ . Επομένως, η παραμετροποίηση που αντιστοιχεί στο μοντέλο  $\tilde{m}(\mu_1)$  θα είναι:

$$\phi = (\mu_1, \mu_2, \sigma^2) \in \Phi = M \times (\mathbb{R}_{>0})^{d_{\mu_1}} \subset \mathbb{R}^{d_\phi}. \tag{6.2.3}$$

και  $M = M_1 \times M_2$

Η περίπτωση όπου  $d_\mu = d_{\mu_1}$  θα αναφέρεται ως ολική τυχαιοποίηση. Η απαίτηση (ii) του παραπάνω ορισμού τροποποιείται ως εξής: Η δεσμευμένη κατανομή του  $Y|\Theta = \theta$  δεν εξαρτάται μόνο από την παράμετρο  $\mu$  και ικανοποιεί την:

$$p(y|\theta) = L(\theta). \quad (6.2.4)$$

Οι λόγοι που γίνεται αυτή η διάκριση θα γίνει καταστούν σαφείς στη συνέχεια.

### Παρατήρηση.

Η μέθοδος της Γκαουσιανής τυχαιοποίησης δεν θα πρέπει να συγχέεται με τη Μπεϋζιανή προσέγγιση. Οι παράμετροι των διασπορών που εμπλέκονται σε αυτήν, είναι άγνωστες σταθερές, σε αντίθεση με την επιλογή (κανονικών) εκ των προτέρων κατανομών (με γνωστές διασπορές) για τις παραμέτρους του μοντέλου, που γίνεται στο πλαίσιο της Μπεϋζιανής προσέγγισης.

Η συνάρτηση πιθανοφάνειας του επεκτεταμένου μοντέλου  $\tilde{m}(\mu_1)$  για  $\phi \in \Phi$  είναι:

$$\begin{aligned} \tilde{L}(\phi) &= \int p(\theta, y; \phi) d\theta \\ &= \int p(\theta; \mu_1, \sigma^2) p(y|\theta; \mu_2) d\nu(\theta) \\ &= E_{\mu_1, \sigma^2} [p(y|\Theta; \mu_2)], \end{aligned} \quad (6.2.5)$$

όπου  $\Theta = (\Theta_i)_{i=1}^{d_{\mu_1}}$  με  $\Theta_i \sim \mathcal{N}(\mu_{1i}, \sigma_i^2)$  και  $p(\theta; \mu_1, \sigma^2)$  η από κοινού σ.π.π. των  $\Theta_i$ , για  $1 \leq i \leq d_{\mu_1}$ .

Στην επόμενη Πρόταση παρουσιάζονται κάποιες βασικές ιδιότητες που μοιράζονται οι συναρτήσεις πιθανοφάνειας των δύο μοντέλων.

### Πρόταση 6.1

- (i) Η  $\tilde{L}$  είναι φραγμένη στο  $\Phi$ .
- (ii) Η  $\tilde{L}$  είναι συνεχής στο  $\Phi$ .
- (iii) Όταν η  $p(\theta, y; \phi)$  έχει συνεχή και φραγμένη παράγωγο (ως προς  $\phi$ ), τότε η  $\tilde{L}$  είναι διαφορίσιμη στο  $\Phi$ .

### Απόδειξη

Για το (i). Από τον ορισμό (1-(ii)), την υπόθεση (1-(ii)) και την (6.2.5) έχουμε ότι για οποιοδήποτε  $\phi \in \Phi$ :

$$\tilde{L}(\phi) = E_{\mu_1, \sigma^2} [p(y|\Theta; \mu_2)] = E_{\mu_1, \sigma^2} [L(\Theta; \mu_2)] \leq L(\mu^*)$$

Για το (ii). Έστω  $\phi_n = (\mu_{1,n}, \mu_{2,n}, \sigma_n^2) \in \Phi$  η οποία συγκλίνει στο  $\phi_0 = (\mu_{1,0}, \mu_{2,0}, \sigma_0^2) \in \Phi$ . Το  $\Phi$  είναι ανοικτό σύνολο, ως καρτεσιανό γινόμενο ανοικτών συνόλων. Θα δείξουμε πως  $\tilde{L}(\phi_n) \rightarrow \tilde{L}(\phi_0)$ .

## Ορισμός και Βασικές Ιδιότητες

Η πυκνότητα  $p(\theta; \mu_1, \sigma_2)$  είναι συνεχής ως προς τα  $\mu_1$  και  $\sigma^2$ . Έπεται πως  $p(\theta; \mu_{1,n}, \sigma_n^2) \rightarrow p(\theta; \mu_{1,0}, \sigma_0^2)$ . Το  $\Phi$  είναι ανοικτό (ως καρτεσιανό γινόμενο ανοικτών), οπότε από το Λήμμα του Scheffe, είναι  $\Theta_{\mu_{1,n}, \sigma_n^2} \xrightarrow{D} \Theta_{\mu_{1,0}, \sigma_0^2} \sim \mathcal{N}_{d_{\mu_1}}(\mu_{1,0}, \sigma_0^2)$ . Άρα:

$$(\Theta_{\mu_{1,n}, \sigma_n^2}; \mu_{2,n}) \xrightarrow{D} (\Theta_{\mu_{1,0}, \sigma_0^2}; \mu_{2,0})$$

Επειδή η  $L$  είναι συνεχής και φραγμένη, από τον χαρακτηρισμό της ασθενούς σύγκλισης, έπεται πως:

$$\tilde{L}(\phi_n) = E[L(\Theta_{\mu_{1,n}, \sigma_n^2}, \mu_{2,n})] \rightarrow E[L(\Theta_{\mu_{1,0}, \sigma_0^2}, \mu_{2,0})] = \tilde{L}(\phi_0).$$

καθώς  $n \rightarrow \infty$

Για το (iii) Έστω  $y \in \mathbb{R}$  και  $f_y : \mathbb{R} \times \Phi \rightarrow \mathbb{R}$  με  $f_y(\theta, \phi) = p(\theta, y; \phi)$ . Υποθέτουμε πως  $|\frac{\partial}{\partial \phi} p(\theta, y; \phi)| \leq g(\theta, y)$  με  $\int |g(\theta, y)| d\nu < \infty$ . Τότε, από το θεώρημα κυριαρχημένης σύγκλισης:

$$\nabla_{\phi} \tilde{L}(\phi) = \nabla_{\phi} \int f_y(\theta, \phi) d\nu(\theta) = \int \nabla_{(\mu_1, \sigma^2)} p(\theta; \mu_1, \sigma^2) \nabla_{\mu_2} p(y|\theta; \mu_2) d\nu(\theta),$$

καθώς

$$\lim_{h \rightarrow 0} \frac{p(\theta, y; \phi_i + h, \phi^{(-i)}) - p(\theta, y; \phi)}{h} = \frac{\partial}{\partial \phi_i} p(\theta, y; \phi),$$

για κάποιο  $t \in (0, 1)$  και για κάθε  $i$  (από το Θεώρημα μέσης τιμής). □

### Πλήρη δεδομένα από την εκθετική οικογένεια.

Πολύ σημαντική είναι η περίπτωση που τα πλήρη δεδομένα,  $X = (Y, \Theta)$ , βρίσκονται στην εκθετική οικογένεια. Αυτό μπορεί να συμβεί αν για παράδειγμα η  $Y$  έχει πυκνότητα που ανήκει στην εκθετική οικογένεια ή αν έχουμε ολική τυχαιοποίηση. Στην πρώτη περίπτωση, ο ισχυρισμός είναι αληθής καθώς το γινόμενο πυκνοτήτων εκθετικής οικογένειας είναι εξίσου μέλος της, ενώ στη δεύτερη έχουμε ότι: για  $\phi = (\mu, \sigma^2) \in M \times \mathbb{R}_{>0}^{d_{\mu}} = \Phi$ , είναι:

$$p(x; \phi) = p(y, \theta; \phi) = p(y|\theta)p(\theta, \phi) = cp(\theta, \phi),$$

καθώς το  $p(y|\theta)$  δεν έχει παράμετρο.

Στην περίπτωση αυτή επομένως, η πυκνότητα για τα πλήρη δεδομένα έχει τη μορφή:

$$p(y, \theta; \phi) = h(y, \theta) \exp\{s(\phi) \cdot t(y, \theta) - \alpha(\phi)\},$$

όπου με  $\cdot$  συμβολίζεται το εσωτερικό γινόμενο στον  $\mathbb{R}^n$ . Η κανονική μορφή θα είναι:

$$p(y, \theta; \phi) = h(y, \theta) \exp\{\eta \cdot t(y, \theta) - b(\eta)\},$$

όπου:

$$e^{b(\eta)} = \int h(y, \theta) \exp\{\eta \cdot t(y, \theta)\} \lambda(d\theta, dy).$$

## Ορισμός και Βασικές Ιδιότητες

Συγκεκριμένα, η  $b(\eta)$  είναι ο λογάριθμος του μετασχηματισμού Laplace του μέτρου  $\nu(d\theta, dy) = h(y, \theta)\lambda(d\theta, dy)$ , όπου το  $\lambda$  είναι το μέτρο Lebesgue διάστασης ίσης με τη διάσταση των πλήρη δεδομένων. Σύμφωνα με τη σχετική θεωρία στην ενότητα 5.5, η  $b(\eta)$  είναι αναλυτική και οι παράγωγοι περνάνε μέσα στο ολοκλήρωμα. Από το τελευταίο γεγονός παίρνουμε και ότι:

$$\nabla b(\eta) = E_\eta[t(y, \theta)].$$

Ωστόσο, κάτω από τις υποθέσεις του μοντέλου, αν  $\phi = (\phi_1, \phi_2)$ , όπου  $\phi_1 = (\mu_1, \sigma^2)$  και  $\phi_2 = \mu_2$ , τότε θα είναι:

$$p(y, \theta; \phi) = \exp\{\eta_1 \cdot t_1(\theta) - b_1(\eta_1)\} \exp\{\eta_2 \cdot t_2(y, \theta) - b_2(\eta_2)\},$$

όπου:

$$\eta_1 = s_1(\phi_1) = (\Sigma^{-1}\mu_1, \Sigma^{-1})^T.$$

$$t_1(\theta) = (\sum_{i=1}^{d_{\mu_1}} \theta_i, -\frac{1}{2} \sum_{i=1}^{d_{\mu_1}} \theta_i \theta_i^T)^T.$$

$$b_1(\eta_1) = \frac{d_{\mu_1}}{2} \log 2\pi + \frac{d_{\mu_1}}{2} \log |\Sigma| + \frac{d_{\mu_1}}{2} \mu_1^T \Sigma^{-1} \mu_1.$$

Σύμφωνα με τους συγγραφείς, προκειμένου να δικαιολογηθεί η ύπαρξη της Ε.Μ.Π. για το μοντέλο  $\tilde{m}(\mu_1)$ , θα χρειαστεί να επεκταθεί η  $\tilde{L}$  ώστε να παίρνει τιμές που να περιέχουν μηδενικές διασπορές (δηλαδή, θα επιτραπεί στις συνιστώσες του  $\sigma^2$  (κάποιες, ή όλες) να μπορούν να πάρουν την τιμή 0) και θα δοθούν σωστές ερμηνείες των συνοριακών αυτών τιμών. Αν η  $j$ -οστή (για παράδειγμα) συνιστώσα του διανύσματος  $\Theta$  έχει παράμετρο διασποράς ίση με 0, τότε η τ.μ.  $\Theta_j$  είναι εκφυλισμένη. Συγκριμένα, δίνει όλη της τη μάζα στο σημείο  $\mu_{1,j}$  (αντιστοιχεί στη  $j$ -οστή συνιστώσα του  $\mu_1$ ). Ισοδύναμα, θα μπορούσαμε να θεωρήσουμε ότι στην τυχαιοποίηση δεν περιλαμβάνεται το  $\mu_{1,j}$ , και άρα σύμφωνα με το (i) του ορισμού, δεν έχουμε τυχαιοποίηση για το  $\mu_1$  αλλά για το  $\mu_1 \setminus \{\mu_{1,j}\}$ . Έτσι, μπορεί να δημιουργηθεί και ένα ενιαίο πλαίσιο στο οποίο το  $m$ , με  $\mu = (\mu_1, \mu_2)$ , να αποτελεί υπομοντέλο του  $\tilde{m}(\mu_1)$ . Ορίζεται λοιπόν ο χώρος που συμπεριλαμβάνει μηδενικές διασπορές:

$$\tilde{\Phi} = M \times (\mathbb{R}_{\geq 0})^{d_{\mu_1}} \tag{6.2.6}$$

Για τη συνέχεια, θα χρειαστούμε τον ακόλουθο συμβολισμό, με τον οποίο θα μπορούμε να επεκτείνουμε την  $\tilde{L}$  στο  $\tilde{\Phi}$

Έστω τα σύνολα  $I \subset \{1, \dots, d_{\mu_1}\}$  και  $J = \{1, \dots, d_{\mu_1}\} \setminus I$ . Με  $\mu_{1,I}, \sigma_I^2$  συμβολίζονται τα υποδιανύσματα των  $\mu, \sigma^2$  με συνιστώσες που αντιστοιχούν στους δείκτες από το  $I$ .

Έστω επίσης τα σύνολα  $I' = \{i \in \{1, \dots, d_{\mu_1}\} : \sigma_i^2 > 0\}$  και  $J' = \{1, \dots, d_{\mu_1}\} \setminus I'$ . Με  $\mu_{1,I'}$  συμβολίζεται το υποδιάνυσμα του  $\mu_1$  του οποίου οι δείκτες αντιστοιχούν στους δείκτες των μη-μηδενικών διασπορών. Επομένως, στο  $\mu_\theta$  αντικατοπτρίζεται το γεγονός ότι δεν έχουμε καθόλου τυχαιοποίηση. Με το συμβολισμό αυτό, θα είναι  $\tilde{m}(\mu_\theta) = m$ . Οπότε το  $\tilde{m}(\mu_1)$  περιλαμβάνει το  $m$  ως ειδική περίπτωση (όταν  $\mu_1 = \mu_\theta$ ).

## Ορισμός και Βασικές Ιδιότητες

Προς την κατεύθυνση αυτή, θα χρειαστεί επίσης το επόμενο Λήμμα.

**Λήμμα 6.1.** Έστω  $x \in \mathbb{R}^{d_x}$  και τα σύνολα  $I$  και  $J$  όπως παραπάνω. Θεωρούμε την οικογένεια των διανυσμάτων:  $\{X_{\sigma^2} : X_{\sigma^2} \sim \mathcal{N}_{d_x}(x, \Sigma), \text{ όπου } \Sigma = \text{diag}\{\sigma_i^2\}_{1 \leq i \leq d_{\mu_1}}\}$  αντιστρέψιμος πίνακας συνδιακύμανσης. Τότε:

$$X_{\sigma^2} \xrightarrow{D} X_{\sigma_I^2} \sim \mathcal{N}_{d_x}(x, \Sigma_I), \text{ καθώς } \|\sigma_J^2\| \rightarrow 0,$$

όπου  $\Sigma_I$  ένας μη αντιστρέψιμος πίνακας συνδιακύμανσης που προκύπτει από τον  $\Sigma$  τοποθετώντας  $\sigma_J^2 = 0_J$ . Στην ειδική περίπτωση που  $I = \emptyset$  τότε:

$$X_{\sigma^2} \xrightarrow{D} x, \text{ καθώς } \|\sigma_J^2\| \rightarrow 0.$$

### Απόδειξη

Αν  $\phi_{\sigma^2}$ ,  $\phi_{\sigma_I^2}$  και  $\phi_0$  οι χαρακτηριστικές συναρτήσεις των  $X_{\sigma^2}$ ,  $X_{\sigma_I^2}$  και του σταθερού διανύσματος  $x$  αντίστοιχα, τότε αρκεί να δείξουμε πως καθώς  $\sigma_J^2 \rightarrow 0_J$  ισχύει  $\phi_{\sigma^2} \rightarrow \phi_{\sigma_I^2}$  κατά σημείο. Έστω  $t = (t_1, \dots, t_{d_x})^T \in \mathbb{R}^{d_x}$ . Τότε:

$$\phi_{\sigma^2}(t) = e^{it^T x - \frac{1}{2}t^T \Sigma t} \rightarrow e^{it^T x - \frac{1}{2}t^T \Sigma_I t} = \phi_{\sigma_I^2}(t),$$

καθώς  $\|\sigma_J^2\| \rightarrow 0$ .

Όταν  $I = \emptyset$ , τότε ο  $\Sigma_I$  είναι ο μηδενικός πίνακας και άρα από την παραπάνω, έπεται πως  $\phi_{\sigma^2}(t) \rightarrow e^{it^T x} = \phi_0(t)$  □

Θα επεκτείνουμε τη συνάρτηση πιθανοφάνειας,  $\tilde{L}$ , του επεκταμένου μοντέλου από το  $\Phi$  στο  $\tilde{\Phi}$ . Προς τούτο, αν (για παράδειγμα)  $J'_{\mu_1} \neq \emptyset$ , δηλαδή  $\sigma_J^2 = 0$ , για κάποιο  $j$  Τότε:

$$\tilde{L}(\phi) = \tilde{L}(\mu_1, \mu_2, \sigma^2) = \tilde{L}(\mu_{1,J'}, \mu_{1,J'}, \mu_2, \sigma_{I'}^2, \sigma_{J'}^2) = \tilde{L}(\mu_{1,J'}, \mu'_2, \sigma_{I'}^2, 0_{J'}),$$

όπου  $\mu'_2 = (\mu_{1,J'}, \mu_2)$  και η τυχαιοποίηση γίνεται στο  $\mu_{1,J'}$

Επίσης, η  $\{\tilde{m}(\mu_I)\}_{I_{\mu_1}}$  θα είναι η κλάση των  $2^{d_{\mu_1}} - 1$  Γκαουσιανών τυχαιοποιήσεων του  $m$  ως προς τα υποδιανύσματα του  $\mu_1$  μαζί με το αρχικό μοντέλο  $m$  που αντιστοιχεί στην επιλογή  $I_{\mu_1} = \emptyset$ . Η επόμενη Πρόταση συνδέει τα μοντέλα της παραπάνω κλάσης.

**Πρόταση 6.2** Έστω  $I \subset \{1, \dots, d_{\mu_1}\}$  και  $J = \{1, \dots, d_{\mu_1}\} \setminus I$ . Τότε,

$$\tilde{L}(\phi) = \tilde{L}(\mu, \sigma_I^2, \sigma_J^2) \rightarrow \tilde{L}(\mu, \sigma_I^2), \text{ καθώς } \|\sigma_J^2\| \rightarrow 0, \quad (6.2.7)$$

όπου η  $\tilde{L}(\mu, \sigma_I^2)$  είναι η συνάρτηση πιθανοφάνειας που αντιστοιχεί στο μοντέλο  $\tilde{m}(\mu_I)$ . Συγκεκριμένα, αν  $I = \emptyset$ , τότε,

$$\tilde{L}(\phi) = \tilde{L}(\mu, \sigma^2) \rightarrow L(\mu), \text{ καθώς } \|\sigma^2\| \rightarrow 0, \quad (6.2.8)$$

όπου η  $L$  είναι η συνάρτηση πιθανοφάνειας του αρχικού μοντέλου  $m$ .

### Απόδειξη



## Ορισμός και Βασικές Ιδιότητες

Είναι  $\tilde{L}(\phi) = E[L(\Theta_{\mu_1, \sigma^2}, \mu_2)]$ . Από το παραπάνω λήμμα έχουμε εξασφαλίσει πως:

$$\Theta_{\sigma^2} \xrightarrow{D} \Theta_{\sigma_I^2} \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma_I), \quad \text{καθώς } \|\sigma_J^2\| \rightarrow 0,$$

Για οποιοδήποτε  $\mu_2 \in M_2$ , η συνάρτηση  $g$  με  $g(\mu_1) = L(\mu)$  είναι συνεχής στο  $M_1$ . Πράγματι, επιλέγοντας μία  $(\mu_1^n) \in M_1$  με  $\mu_1^n \rightarrow \mu_1$  καθώς  $n \rightarrow \infty$ , τότε  $\mu^n \equiv (\mu_1^n, \mu_2) \rightarrow (\mu_1, \mu_2) = \mu$  και άρα  $g(\mu_1^n) = L(\mu^n) \rightarrow L(\mu) = g(\mu_1)$  καθώς η  $L$  είναι συνεχής στο  $M$ . Η  $g$  είναι επίσης φραγμένη στο  $M_1$ .

Από το παραπάνω, και τον χαρακτηρισμό της ασθενούς σύγκλισης θα είναι:

$$\begin{aligned} E[L(\Theta_{\mu_1, \sigma^2}, \mu_2)] &= E[g(\Theta_{\mu_1, \sigma^2})] \rightarrow E[g(\Theta_{\mu_1, \sigma_I^2})] \\ &= E[L(\Theta_{\mu_1, \sigma_I^2}, \mu_2)], \end{aligned}$$

όπου με  $\Theta_{\mu_1, \sigma_I^2}$  συμβολίζεται το υποδιάνυσμα του  $\Theta_{\mu_1, \sigma^2}$  με δείκτες από το  $I$ , δηλαδή για δείκτες με  $\sigma_i^2 > 0$ , Επομένως ο τελευταίος όρος συμπίπτει με ανάλογη αναπαράσταση της  $\tilde{L}$  για το μοντέλο  $\tilde{m}(\mu_I)$ . Όταν  $I = \emptyset$ , τότε συμπίπτει με την  $L(\mu)$ .  $\square$

Επειδή η  $\tilde{L}$  είναι συνεχής στο  $\Phi$  (Πρόταση 6.1), αν θέσουμε:

$$\tilde{L}(\mu, \sigma_I^2, 0_J) = \tilde{L}(\mu, \sigma_I^2), \quad (6.2.9)$$

$$\tilde{L}(\mu, 0_{d_{\mu_1}}) = L(\mu), \quad (6.2.10)$$

τότε έπεται άμεσα από την Πρόταση 6.2 πως η επέκταση της  $\tilde{L}$  στο  $\tilde{\Phi} = M \times (\mathbb{R}_{\geq 0})^{d_{\mu_1}}$  που ικανοποιεί τις (6.2.9, 6.2.10), είναι συνεχής.

Ωστόσο, η  $\tilde{L}$  δεν είναι διαφορίσιμη στο  $\tilde{\Phi}$ . Πράγματι, αν  $\sigma_J^2 = 0$ , τότε δεν υπάρχει η πυκνότητα  $p(\phi; \mu_1, \sigma^2)$ .

**Πρόταση 6.3** Η  $\tilde{L} : \tilde{\Phi} \rightarrow \mathbb{R}$  είναι άνω φραγμένη από το  $L(\hat{\mu})$ .

### Απόδειξη

Έστω  $I = \{i \in \{1, \dots, d_{\mu_1}\} : \sigma_i^2 > 0\}$  και  $J = \{1, \dots, d_{\mu_1}\} \setminus I$ . Θα δείξουμε πως  $\tilde{L}(\mu, \sigma^2) \leq L(\hat{\mu})$ .

Αν  $I = \emptyset$  δεν έχουμε τίποτα να δείξουμε καθώς τότε  $\tilde{m}(\mu_\emptyset) = m$  και από την (6.2.10) είναι  $\tilde{L}(\mu, 0_{d_{\mu_1}}) = L(\mu)$ . Η περίπτωση  $I = \{1, \dots, d_{\mu_1}\}$  (δηλαδή όλες οι διασπορές είναι γνήσια θετικές) έχει λυθεί στη Πρόταση 6.1. Υποθέτουμε λοιπόν ότι  $I \subsetneq \{1, \dots, d_{\mu_1}\}$ . Από την (6.2.9) είναι  $\tilde{L}(\mu, \sigma_I^2, 0_J) = \tilde{L}(\mu, \sigma_I^2)$  και η τελευταία αντιστοιχεί στο μοντέλο  $\tilde{m}(\mu_I)$  για το οποίο  $\Theta_I \sim \mathcal{N}_{d_{\mu_I}}(\mu_I, \Sigma_I)$ .

Οπότε,

$$\begin{aligned} \tilde{L}(\mu, \sigma_I^2) &= \int p(y|\theta; \mu_{1,J}, \mu_2) p(\theta; \mu_{1,I}, \sigma_I^2) d\theta \\ &= E_{\mu_{1,I}, \sigma_I^2}[L(\Theta_I, (\mu_2)')] \leq E_{\mu_{1,I}, \sigma_I^2}[L(\mu_{\hat{1}, I}, (\hat{\mu}_2)')] \\ &= L(\hat{\mu}) \quad (\mathbb{P}_{\mu_{1,I}, \sigma_I^2} - \text{σ.π.}), \end{aligned}$$

όπου  $(\mu_2)' = (\mu_{1,J}, \mu_2)$ . Η ανισότητα δικαιολογείται διότι  $\Theta_I \in M_I$  και από την υπόθεση 1, για οποιοδήποτε  $\theta_I \in M_I$  ισχύει  $L(\theta_I, \mu_{1,J}, \mu_2) \leq L(\hat{\mu}_{1,I}, \hat{\mu}_{1,J}, \hat{\mu}_2) \equiv L(\hat{\mu})$ .  $\square$

**Θεώρημα 6.1** Έστω  $\{\tilde{m}(\mu_I)\}_I$  μία κλάση μοντέλων Γκαουσιανής τυχαιοποίησης με αρχικό μοντέλο το  $m$ , όπου  $I \subset \{1, \dots, d_{\mu_1}\}$  (όπου για  $I = \emptyset$ ,  $\tilde{m}(\mu)_\emptyset = m$ ). Κάτω από την υπόθεση 1, αν  $\hat{\mu}$ ,  $\eta$  (μοναδική) Ε.Μ.Π. για το μοντέλο  $m$ , τότε αυτή καθορίζει τη μοναδική Ε.Μ.Π.  $\hat{\phi}$ , για το μοντέλο  $\tilde{m}(\mu_1)$ . Συγκεκριμένα έχουμε ότι:

$$\hat{\phi} = (\hat{\mu}, 0_{d_{\mu_1}}), \quad (6.2.11)$$

$$\tilde{L}(\hat{\phi}) = L(\hat{\mu}). \quad (6.2.12)$$

Αντίστροφα, αν η  $L$  είναι συνεχής στο  $M$  (υπόθεση 1-(i)) και υπάρχει μοναδική Ε.Μ.Π.  $\hat{\phi}$  για το μοντέλο  $\tilde{m}(\mu)_1$ , τότε αυτή καθορίζει τη μοναδική Ε.Μ.Π.  $\hat{\mu}$  για το μοντέλο  $m$  και ικανοποιεί τις παραπάνω εξισώσεις 6.2.11, 6.2.12.

### Απόδειξη

Για το ευθύ: Υποθέτουμε ότι υπάρχει μοναδική Ε.Μ.Π.,  $\hat{\mu}$  στο  $M$ . Από την (6.2.10) είναι  $\tilde{L}(\hat{\mu}, 0_{d_{\mu_1}}) = L(\hat{\mu})$ . Άρα, η Ε.Μ.Π. στο  $\tilde{\Phi}$  υπάρχει, και αν αυτή είναι μοναδική θα πρέπει να ικανοποιεί τις (6.2.11, 6.2.12). Επειδή  $(\hat{\mu}, 0_{d_{\mu_1}}) \in M \times \{0\}^{d_{\mu_1}}$ , από την υπόθεση 1 και την (6.2.10), είναι μοναδική στο σύνολο αυτό. Ας υποθέσουμε πως υπάρχει κάποιο  $\phi = (\mu, \sigma^2) \in M \times (\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}})$  τέτοιο ώστε  $\tilde{L}(\phi) = L(\hat{\mu})$ . Θα δείξουμε πως αυτό δεν μπορεί να γίνει. Θεωρούμε τα σύνολα  $I = \{i \in \{1, \dots, d_{\mu_1}\} : \sigma_i^2 > 0\}$  και  $J = \{1, \dots, d_{\mu_1}\} \setminus I$  όπου το πρώτο είναι μη κενό αφού  $(\mu, \sigma^2) \in M \times (\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}})$ . Επειδή  $\tilde{L}(\mu, \sigma^2) = E_{\mu_{1,I}, \sigma_I^2}[L(\Theta_I, \mu'_2)]$ , και λόγω της  $\tilde{L}(\phi) = L(\hat{\mu})$ , θα είναι:

$$L(\hat{\mu}) = E_{\mu_{1,I}, \sigma_I^2}[L(\Theta_I, (\mu_2)')] ]$$

όπου  $\mu'_2 = (\mu_{1,J}, \mu_2)$ . Από την Πρόταση 6.3 η  $\tilde{L}$  είναι άνω φραγμένη στο  $\tilde{\Phi}$  από το  $L(\hat{\mu})$  και άρα η τ.μ.  $L(\Theta_I, (\mu_2)')$ , θα είναι άνω φραγμένη από την ίδια τιμή. Δηλαδή η  $L(\hat{\mu}) - L(\Theta_I, (\mu_2)')$  θα είναι μη αρνητική. Ωστόσο, από την παραπάνω προκύπτει πως:

$$E_{\mu_{1,I}, \sigma_I^2}[L(\hat{\mu}) - L(\Theta_I, (\mu_2)')] = 0$$

και παίρνουμε επομένως λόγω της μη αρνητικότητας ότι  $L(\hat{\mu}) = L(\Theta_I, (\mu_2)')$  σ.π. ως προς το  $\mathbb{P}_{\mu_{1,I}, \sigma_I^2}$ . Δηλαδή:

$$1 = \mathbb{P}_{\mu_{1,I}, \sigma_I^2}(L(\hat{\mu}) = L(\Theta_I, (\mu_2)')) = \mathbb{1}_{\{(\mu_2)' = (\hat{\mu}_2)'\}} \mathbb{P}_{\mu_{1,I}, \sigma_I^2}(\Theta_I = \hat{\mu}_{1,I}) = 0$$

Η δεύτερη ισότητα στην παραπάνω έπεται από την υπόθεση 1 και η τελευταία από το γεγονός πως η  $\Theta_I$  είναι απόλυτα συνεχής. Καταλήγουμε λοιπόν πως για οποιοδήποτε  $\phi \in M \times (\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}})$  ισχύει ότι  $\tilde{L}(\phi) < L(\hat{\mu})$ . Οπότε η Ε.Μ.Π.  $\hat{\phi}$  είναι μοναδική.

Για το αντίστροφο, υποθέτουμε πως υπάρχει μοναδική Ε.Μ.Π.,  $\hat{\phi}$  στο  $\tilde{\Phi}$ . Στην περίπτωση αυτή, για κάθε  $\phi \in \tilde{\Phi}$  ισχύει ότι  $\tilde{L}(\phi) \leq \tilde{L}(\hat{\phi})$ . Άρα θα ισχύει και  $\tilde{L}(\phi) \leq \tilde{L}(\hat{\phi})$  για κάθε  $\phi \in M \times \{0\}^{d_{\mu_1}}$  αφού αυτό αποτελεί υποσύνολο του  $\tilde{\Phi}$ . Επομένως, θα είναι:

$$\tilde{L}(\mu, 0_{d_{\mu_1}}) \leq \tilde{L}(\hat{\phi})$$

## Εφαρμογή του αλγορίθμου EM

για κάθε  $\mu \in M$ . Αν  $\hat{\phi} \in M \times \{0\}^{d_{\mu_1}}$ , τότε οι εξισώσεις (6.2.11 και 6.2.12) ικανοποιούνται για κάποιο  $\mu^* \in M$ . Δηλαδή για αυτό το  $\mu^*$  θα είναι:  $\hat{\phi} = (\mu^*, 0_{d_{\mu_1}})$  και  $\tilde{L}(\hat{\phi}) = L(\mu^*)$ . Επειδή το άνω φράγμα της  $L$  στο  $M$  πιάνεται στο  $\hat{\mu}$  η Ε.Μ.Π. υπάρχει. Είναι επίσης μοναδική διότι αν υπήρχαν 2 διαφορετικές Ε.Μ.Π.,  $\mu^{(1)}, \mu^{(2)}$  στο  $M$ , τότε τα  $(\mu^{(1)}, 0_{d_{\mu_1}})$  και  $(\mu^{(2)}, 0_{d_{\mu_1}})$  θα ήταν Ε.Μ.Π. στο  $\tilde{\Phi}$  το οποίο παραβιάζει την υπόθεση μοναδικότητας του  $\hat{\phi}$ . Θα δείξουμε τώρα πως το  $\hat{\phi} \in M \times \{0\}^{d_{\mu_1}}$  είναι αναγκαία συνθήκη. Ας υποθέσουμε πως  $\hat{\phi} \in M \times (\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}})$  και θεωρούμε τα σύνολα  $I = \{i \in \{1, \dots, d_{\mu_1}\} : (\sigma_i^2)^* > 0\}$  και  $J = \{1, \dots, d_{\mu_1}\} \setminus I$ . Θα είναι:

$$\tilde{L}(\hat{\phi}) = E_{(\mu_{1,I})^*, (\sigma_I^2)^*} [L(\Theta_I, (\mu'_2)^*)]$$

Δεν θα μπορούσε να είναι  $L(\mu_{1,I}, (\mu'_2)^*) < \tilde{L}(\hat{\phi})$  για όλα τα  $\mu_{1,I} \in \mathbb{R}^{d_{\mu_1}}$  καθώς αυτό έρχεται σε αντίθεση με την παραπάνω ισότητα. Επομένως υπάρχει κάποιο  $(\mu_I)^{**}$  τέτοιο ώστε  $L((\mu_I)^{**}, (\mu'_2)^*) = \tilde{L}(\hat{\phi})$ , ή ισοδύναμα, κάποιο  $\phi^* = ((\mu_I)^{**}, (\mu'_2)^*, 0_{d_{\mu_1}}) \neq \hat{\phi}$  το οποίο θα είναι Ε.Μ.Π. Το τελευταίο έρχεται σε αντίθεση με την υπόθεση μοναδικότητας. Επομένως καταλήγουμε πως  $\hat{\phi} \in M \times \{0\}^{d_{\mu_1}}$ .  $\square$

### Παρατηρήσεις.

1. Κάτω από την υπόθεση 1, η (μοναδική) Ε.Μ.Π. για το μοντέλο  $\tilde{m}(\mu)_1$  αποτελεί στοιχείο στο σύνορο του παραμετρικού χώρου  $\tilde{\Phi} = M \times (\mathbb{R}_{\geq 0})^{d_{\mu_1}}$ .

2. Το παραπάνω Θεώρημα επιτρέπει τη μεταφορά του προβλήματος μεγιστοποίησης που αντιστοιχεί στο αρχικό μοντέλο  $m$ , σε ένα ισοδύναμο πρόβλημα μεγιστοποίησης που αντιστοιχεί στο μοντέλο  $\tilde{m}(\mu_1)$ . Ισοδύναμο με την έννοια πως αν το ένα έχει (μοναδική) λύση, τότε έχει και το άλλο και οι λύσεις ταυτίζονται.

3. Το πρόβλημα μεγιστοποίησης που αντιστοιχεί στο  $\tilde{m}(\mu_1)$ , μπορεί να επιλυθεί με κατασκευή κατάλληλης έκδοσης του EM καθώς το  $\tilde{m}(\mu_1)$  είναι μοντέλο μη πλήρων δεδομένων.

4. Η υπόθεση μοναδικότητας της Ε.Μ.Π. στο αρχικό μοντέλο είναι πολύ ισχυρή και γενικά δύσκολα επαληθεύσιμη. Ωστόσο, αν η υπόθεση χαλαρωθεί, τότε, επαναλαμβάνοντας τα βήματα της απόδειξης, μπορούμε να συμπεράνουμε ότι το αρχικό μοντέλο έχει πεπερασμένο πλήθος Ε.Μ.Π. αν το επεκταμένο μοντέλο έχει πεπερασμένο πλήθος Ε.Μ.Π. και ισχύει η (6.2.11).

## 6.3 Εφαρμογή του αλγορίθμου EM

**Πρόταση 6.4** Έστω  $\phi' = (\mu_1, \mu_2, \sigma^2)' \in M \times (\mathbb{R}_{>0})^{d_{\mu_1}}$  η τρέχουσα ανανέωση των παραμέτρων για το μοντέλο  $\tilde{m}(\mu_1)$ . Τότε, οι εξισώσεις για την ανανέωση των παραμέτρων  $(\mu_1, \sigma^2)$  είναι:

$$\hat{\mu}_1 = E_{\phi'}[\Theta|Y = y] \tag{6.3.1}$$

$$\hat{\sigma}^2 = (\hat{\sigma}_i^2)_{1 \leq i \leq d_{\mu_1}} = (Var_{\phi'}[\Theta_i|Y = y])_{1 \leq i \leq d_{\mu_1}}. \tag{6.3.2}$$

## Εφαρμογή του αλγορίθμου EM

και το  $\mu_2$  μπορεί να ανανεωθεί ανεξάρτητα, μεγιστοποιώντας την:

$$Q_2(\mu_2; \phi') = E_{\phi'}[\log p(y|\Theta; \mu_2)|Y = y]. \quad (6.3.3)$$

### Απόδειξη

$$\begin{aligned} Q(\phi; \phi') &= E_{\phi'}[\log p(\Theta, y; \mu_1, \mu_2, \sigma^2)|Y = y] \\ &= E_{\phi'}[\log p(\Theta, \mu_1, \sigma^2)|Y = y] + E_{\phi'}[\log p(y|\Theta; \mu_2)|Y = y] \\ &= Q_1(\mu_1, \sigma^2; \phi') + Q_2(\mu_2; \phi'). \end{aligned}$$

Ο τελευταίος όρος στο δεξί μέλος της παραπάνω εξίσωσης εξαρτάται μόνο από το  $\mu_2$  και συμπίπτει με την εξίσωση (6.3.3) ενώ ο πρώτος όρος δεν εξαρτάται από το  $\mu_2$ . Επομένως ισχύει ο τελευταίος ισχυρισμός στην Πρόταση. Επειδή οι συνιστώσες του διανύσματος  $\Theta$  είναι ανεξάρτητες κανονικές, θα είναι:

$$Q_1(\mu_1, \sigma^2; \phi') \stackrel{max}{\sim} -\frac{1}{2} \sum_{i=1}^{d_{\mu_1}} \log \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{d_{\mu_1}} \frac{E_{\phi'}[(\Theta_i - \mu_{1,i})^2|Y = y]}{\sigma_i^2}.$$

και επομένως ο πρώτος ισχυρισμός έπεται άμεσα. □

**Παρατήρηση.** Στην περίπτωση της ολικής τυχαιοποίησης είναι φανερό πως το βήμα M είναι αναλυτικό. Ωστόσο, το βήμα-E θα είναι σχεδόν πάντα μη αναλυτικό και για τον υπολογισμό του, ενδεχομένως να χρειάζεται να εφαρμοστούν αλγοριθμικές τεχνικές. Τέλος, μία ακόμα σημαντική παρατήρηση είναι πως η  $Q_1$  όπως φαίνεται από την παραπάνω έκφραση, είναι συνεχής συνάρτηση ως προς τα 2 ορίσματα. Η ιδιότητα αυτή είναι βασική για τη θεωρία σύγκλισης που ακολουθεί.

### 6.3.1 Εφαρμογή στο μοντέλο LNAS

Στην παράγραφο αυτή θα εφαρμόσουμε τη μέθοδο της Γκαουσιανής τυχαιοποίησης στο μοντέλο LNAS που περιγράφηκε στο κεφάλαιο 2. Έτσι θα έχουμε τη δυνατότητα να παρατηρήσουμε τις σχετικές δυσκολίες ως προς την εφαρμογή καθώς και τα προβλήματα σύγκλισης που προκύπτουν.

Αρχικά θα θεωρήσουμε πως οι παράμετροι για τα σφάλματα  $\{\sigma_{\gamma\gamma}^2, \sigma_g^2, \sigma_r^2\}$  είναι γνωστές σταθερές. Επομένως εστιάζουμε στο πρόβλημα της εκτίμησης του διανύσματος που περιέχει τις παραμέτρους που συνδέονται με τη δομή του φυτού:

$$\mu = (\mu_\alpha, \lambda, \mu_\gamma, \sigma_\gamma^2, \gamma_0, \gamma_f). \quad (6.3.4)$$

Προκειμένου να εφαρμόσουμε τη μέθοδο Γκαουσιανής τυχαιοποίησης θα χρειαστεί να κάνουμε αναπαραμέτρηση ώστε να διασφαλίσουμε πως οι παράμετροι θα μπορούν να έχουν πραγματικές τιμές. Θέτουμε:

$$\eta = (\log \mu_\alpha, \lambda, \mu_\gamma, \log \sigma_\gamma^2, \log(\gamma_0^{-1} - 1), \log(\gamma_f^{-1} - 1)), \quad (6.3.5)$$

διότι  $\mu_\alpha, \sigma_\gamma^2 > 0$  και  $\gamma_0, \gamma_f \in (0, 1)$ . Με την αναπαραμέτρηση αυτή μπορούμε να εφαρμόσουμε τη μέθοδο Γκαουσιανής τυχαιοποίησης στο διάνυσμα  $\eta$ . Συγκεκριμένα,

## Ζητήματα σύγκλισης

έστω  $\tilde{m}(\eta)$  η Γκαουσιανή τυχαιοποίηση του  $m$  ως προς το  $\eta$ , όπου με  $m$  συμβολίζουμε το μοντέλο LNAS. Τότε, σύμφωνα με τον ορισμό 6.1 έχουμε ότι το  $\tilde{m}(\eta)$  είναι ένα μοντέλο κρυμμένων δεδομένων το οποίο αποτελείται από:

(i) Ένα κρυμμένο Γκαουσιανό διάνυσμα  $\Theta$  όπου:

$$\Theta \sim \mathcal{N}_6(\eta, \Sigma), \quad (6.3.6)$$

όπου  $\Sigma = \text{diag}\{\sigma_i^2\}_{1 \leq i \leq 6}$ , με  $\sigma_i^2 > 0$

(ii) Ένα παρατηρούμενο διάνυσμα  $Y$ , όπου δοθέντος  $\Theta = \theta$ , είναι μοντέλο χώρου καταστάσεων που ικανοποιεί τις εξισώσεις στην (2.4.12) όταν η παράμετρος  $\mu$  αντικατασταθεί από το  $\theta$ .

Οπότε η παράμετρος για το επεκταμένο μοντέλο είναι:  $\phi = (\eta, \sigma^2) \in \mathbb{R}^6 \times (\mathbb{R}_+^*)^6$ . Ξεκινώντας με μία αρχική τιμή, έστω  $\phi^{(0)}$ , ο αλγόριθμος EM που δίνεται από την Πρόταση 6.3 παράγει μία ακολουθία  $\phi^{(n)}$  η οποία είναι επιθυμητό να συγλίνει σε κάποιο στάσιμο σημείο της αρχικής συνάρτησης πιθανοφάνειας. Οι δεσμευμένες ροπές που εμπλέκονται στις εξισώσεις αυτές, δεν έχουν αναλυτική έκφραση για μη-γραμμικά μοντέλα τέτοιου τύπου. Το πρόβλημα αντιμετωπίζεται στο [12] με την εφαρμογή στοχαστικής έκδοσης του αλγορίθμου EM.

## 6.4 Ζητήματα σύγκλισης

Στο κεφάλαιο 5 παρουσιάστηκαν συνθήκες κάτω από τις οποίες ο Wu (1983) έδειξε τη σύγκλιση τόσο μιας ακολουθίας EM, όσο και της αντίστοιχης ακολουθίας τιμών πιθανοφάνειας  $(L(\phi^{(k)}))_{k \geq 0}$ . Τα θεωρήματα αυτά δεν μπορούν να εφαρμοστούν στην περίπτωση του EM που κατασκευάστηκε για το μοντέλο  $\tilde{m}(\mu_1)$ . Αυτό διότι το επιθυμητό σημείο σύγκλισης της ακολουθίας EM βρίσκεται στο σύνορο του παραμετρικού χώρου  $\Phi = M \times (\mathbb{R}_{>0})^{d_{\mu_1}}$  και επομένως παραβιάζεται η συνθήκη R1. στην υπόθεση 5.3. Επιπλέον, η  $\tilde{L}$  δεν είναι διαφορίσιμη σε αυτό. Οπότε το  $\hat{\phi} = (\hat{\mu}, 0_{d_{\mu_1}})$  δεν είναι στάσιμο σημείο της  $\tilde{L}$  και το Θεώρημα 5.2 δεν εφαρμόζεται.

Στη παρούσα εργασία θα ασχοληθούμε με τη περίπτωση της ολικής τυχαιοποίησης ( $d_{\mu_1} = d_{\mu}$ ). Στην περίπτωση αυτήν έχουμε από την υπόθεση 6.1 ότι υπάρχει μοναδική Ε.Μ.Π. για το μοντέλο  $m$ , έστω  $\mu^*$ . Το Θεώρημα 6.1 μας λέει ότι η Ε.Μ.Π για το μοντέλο  $\tilde{m}(\mu)$  θα είναι η  $\phi^* = (\mu^*, 0_{d_{\mu}}) \in \Phi = \mathbb{R}^{d_{\mu}} \times \mathbb{R}_{>0}^{d_{\mu}}$ .

Έστω επίσης  $\mathcal{M} : \Phi \mapsto \Phi$  απεικόνιση τέτοια ώστε για κάθε  $\phi \in \Phi$  ισχύει:

$$\mathcal{M}(\phi) = \underset{\psi \in \Phi}{\text{argmax}} Q(\psi, \phi)$$

Επομένως, ο αλγόριθμος EM παράγει μία ακολουθία  $(\phi^{(k)})_{k \geq 0} = (\mu^{(k)}, (\sigma^2)^{(k)})_{k \geq 0}$ . Στη περίπτωση αυτή είναι  $(\phi^{(k)}) = ((\mu, \sigma^2)^{(k)}) \in \Phi = M \times (\mathbb{R}_{>0})^{d_{\mu}}$ , όπου  $M = M_1 \times M_2$  με  $M_1 = \mathbb{R}^{d_{\mu}}$  και  $M_2$  ανοικτό υποσύνολο του  $\mathbb{R}^{d_{\mu_2}}$ .

## Ζητήματα σύγκλισης

Στη περίπτωση της μερικής τυχαιοποίησης είναι  $(\phi^{(k)}) = ((\mu_1, \mu_2, \sigma^2)^{(k)})$  και όπως δείχθηκε στην Πρόταση 6.4, η ακολουθία μπορεί να ανανεωθεί από 2 διαδικασίες. Συγκεκριμένα, η  $((\mu_1, \sigma^2)^{(k)})$  παράγεται από την  $Q_1$ , ενώ η ακολουθία  $\mu_2^{(k)}$  παράγεται από την  $Q_2$  και είναι ανεξάρτητη της πρώτης. Επομένως, στην περίπτωση αυτή προκύπτει ένα γενικότερο πρόβλημα σύγκλισης καθώς αυτές οι δύο ακολουθίες ανανεώνονται διαδοχικά.

Θα αναζητήσουμε λοιπόν συνθήκες για την περίπτωση της ολικής τυχαιοποίησης, με τι οποίες μπορούμε να εξασφαλίσουμε ότι:

1.  $\tilde{L}(\phi^{(k)}) \rightarrow \tilde{L}(\hat{\phi}) = \tilde{L}(\hat{\mu}, 0_{d_\mu}) = L(\hat{\mu})$ , και
2.  $\phi^{(k)} \rightarrow \hat{\phi} = (\hat{\mu}, 0_{d_\mu})$ , καθώς  $k \rightarrow \infty$ .

Σε όσα ακολουθούν θα ισχύει η παρακάτω υπόθεση:

**Υπόθεση 6.3 (Συνθήκες κανονικότητας)** .

- A1. Το σύνολο  $\Phi_{\phi_0} = \{\phi \in \Phi : \tilde{L}(\phi) \geq \tilde{L}(\phi_0)\}$  είναι συμπαγές για οποιαδήποτε επιλογή του  $\phi_0$  τέτοια ώστε  $\tilde{L}(\phi_0) > -\infty$ .
- A2. Η συνάρτηση  $H(\cdot; \phi)$  είναι διαφορίσιμη για κάθε  $\phi \in \Phi$ .

Από την Πρόταση 6.1, η  $\tilde{L}$  είναι συνεχής και διαφορίσιμη στο  $\Phi$ .

Θα εφαρμόσουμε το Θεώρημα Ολικής Σύγκλισης για το παρακάτω σύνολο λύσεων:

$$S = \{\bar{\phi} \in \tilde{\Phi} : \frac{d\tilde{L}}{d\lambda}(\bar{\phi} + \lambda(\phi - \bar{\phi}))|_{\lambda=0} \leq 0, \quad \phi \in \Phi\},$$

όπου το παραπάνω όριο παίρνεται από δεξιά.

Συγκεκριμένα το Θεώρημα Ολικής Σύγκλισης εφαρμόζεται για την  $\tilde{L}$  με σύνολο λύσεων το  $S$  δίνοντας το επόμενο αποτέλεσμα ως ειδική περίπτωση:

**Θεώρημα 6.2** Έστω  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία ΕΜ που παράγεται από την  $\phi^{(k+1)} = \mathcal{M}(\phi^{(k)})$ . Αν:

1. Όλα τα οριακά σημεία της  $\phi^{(k)}$  βρίσκονται σε συμπαγές υποσύνολο του  $\Phi$ .
2. Για την  $\tilde{L}$  ισχύει:
  - (i) Αν  $\phi^{(k)} \notin S$ , τότε  $\tilde{L}(\phi^{(k)}) > \tilde{L}(\phi^{(k+1)}) \forall \phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ .
  - (ii) Αν  $\phi^{(k)} \in S$ , τότε  $\tilde{L}(\phi^{(k)}) \geq \tilde{L}(\phi^{(k+1)}) \forall \phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$ .
3. Η απεικόνιση  $\mathcal{M}$  είναι κλειστή στο  $\Phi \setminus S$

## Ζητήματα σύγκλισης

Τότε, όλα τα οριακά σημεία οποιασδήποτε υπακολουθίας  $(\phi^{(k)})_{k \geq 0}$  είναι στο  $S$ . Ακόμα,  $\lim_{x \rightarrow \infty} \tilde{L}(\phi^{(k)}) = \tilde{L}(\phi^*)$  για όλα τα οριακά σημεία  $\phi^* \in S$ .

Υποθέτουμε πως η  $\phi^{(k)}$  βρίσκεται στο σύνολο  $\Phi_{\phi_0} = \{\phi \in \Phi : \tilde{L}(\phi) \geq \tilde{L}(\phi_0)\}$  που είναι συμπαγές από την υπόθεση 6.3. Επίσης, από την ιδιότητα της μονοτονίας του αλγορίθμου EM, η συνθήκη 2 (ii) ικανοποιείται. Ακόμα, μια ικανή συνθήκη για την κλειστότητα όπως έχουμε δει στο κεφάλαιο 5, είναι η συνέχεια της  $Q$  ως προς τα δύο ορίσματα που στη περίπτωση της ολικής τυχαιοποίησης ικανοποιείται. Τέλος, λόγω της κυρτότητας του  $\Phi$ , η συνθήκη 2 (i) ικανοποιείται επίσης. Συγκεκριμένα:

**Πρόταση 6.5** Έστω  $(\phi^{(k)})_{k \geq 0} = (\phi^{(k)})_{k \geq 0}$  μία ακολουθία που παράγεται από την  $\phi^{(k+1)} = \mathcal{M}(\phi^{(k)})$ . Τότε:

$$\tilde{L}(\phi^{(k+1)}) > \tilde{L}(\phi^{(k)}) \text{ για κάθε } \phi^{(k)} \in \Phi \setminus S.$$

### Απόδειξη

Έστω  $\phi^{(k)} \notin S$ . Η  $H(\cdot; \phi^{(k)})$  είναι διαφορίσιμη στο  $\phi^{(k)}$  από την υπόθεση (6.3) και επειδή αυτή ελαχιστοποιείται στο  $\phi^{(k)}$ , έπεται ότι για κάθε  $\phi \in \Phi$ :

$$\frac{dH}{d\lambda}(\phi^{(k)} + \lambda(\phi - \phi^{(k)}); \phi^{(k)})|_{\lambda=0} = 0$$

Επομένως, για  $\phi \in \Phi$  ισχύει:

$$\frac{dQ}{d\lambda}(\phi^{(k)} + \lambda(\phi - \phi^{(k)}); \phi^{(k)})|_{\lambda=0} = \frac{d\tilde{L}}{d\lambda}(\phi^{(k)} + \lambda(\phi - \phi^{(k)}); \phi^{(k)})|_{\lambda=0} > 0$$

Το σύνολο  $\Phi$  είναι κυρτό και επομένως υπάρχει κάποιο  $\nu \in (0, 1)$  και κάποιο κυρτό σύνολο  $K \subset \Phi$ , και κάποιο  $\phi \in \Phi$  ώστε:

$$Q(\phi^{(k)} + \nu(\phi - \phi^{(k)}); \phi^{(k)}) > Q(\phi^{(k)}; \phi^{(k)})$$

Από τον ορισμό του  $\phi^{(k+1)}$  και την κυρτότητα του  $K$  έπεται πως

$$Q(\phi^{(k+1)}; \phi^{(k)}) \geq Q(\phi^{(k)} + \nu(\phi - \phi^{(k)}); \phi^{(k)}).$$

Τελικά,  $Q(\phi^{(k+1)}; \phi^{(k)}) > Q(\phi^{(k)}; \phi^{(k)})$  και  $\tilde{L}(\phi^{(k+1)}) > \tilde{L}(\phi^{(k)})$  □

Από τα παραπάνω έχουμε το ακόλουθο αποτέλεσμα:

**Πόρισμα 6.1** Έστω  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία που παράγεται από την  $\phi^{(k+1)} = \mathcal{M}(\phi^{(k)})$ . Τότε κάθε οριακό σημείο της  $(\phi^{(k)})_{k \geq 0}$  βρίσκεται στο  $S$  και η  $\tilde{L}(\phi^{(k)})_{k \geq 0}$  συγκλίνει στο  $\tilde{L}^*$ , όπου  $\tilde{L}^* = \tilde{L}(\phi^*)$  για κάποιο  $\phi^* \in S$ .

Θα εξετάσουμε ορισμένες ιδιότητες για το σύνολο λύσεων που έχει επιλεγθεί. Συγκεκριμένα, με την Πρόταση 6.6 ενώ με το Λήμμα 6.2 εξασφαλίζουμε πως το επιθυμητό σημείο σύγκλισης βρίσκεται στο  $S$ .

**Πρόταση 6.6** Το όριο  $\frac{d\tilde{L}}{d\lambda}(\bar{\phi} + \lambda(\phi - \bar{\phi}))|_{\lambda=0}$ , υπάρχει για κάθε  $\bar{\phi} \in \tilde{\Phi}$  και για κάθε  $\phi \in \Phi$ .

**Απόδειξη**

Έστω  $\phi = (\mu, \sigma^2) \in \Phi$ . Διακρίνουμε τις περιπτώσεις:

Έστω  $\bar{\phi} = (\bar{\mu}, \bar{\sigma}^2)$  τέτοιο ώστε  $\bar{\sigma}_i^2 > 0$  για κάθε  $i \in \{1, \dots, d_\mu\}$ . Από την Πρόταση (6.1) η  $\tilde{L}$  είναι διαφορίσιμη στο  $\Phi$  και άρα είναι διαφορίσιμη σε κάθε σημείο  $\bar{\phi} + \lambda(\phi - \bar{\phi})$  για οποιοδήποτε  $\lambda \in [0, 1]$ . Ειδικότερα, στην περίπτωση αυτή είναι:

$$\frac{d\tilde{L}}{d\lambda}(\bar{\phi} + \lambda(\phi - \bar{\phi}))|_{\lambda=0} = \nabla_{\phi} \tilde{L}(\phi)|_{\phi=\bar{\phi}} \cdot (\phi - \bar{\phi}) \in \mathbb{R}$$

Έστω τώρα ότι το  $\bar{\phi} = (\bar{\mu}, \bar{\sigma}^2)$  έχει κάποιες διασπορές ίσες με μηδέν. Θεωρούμε το σύνολο:

$I = \{i \in \{1, \dots, d_\mu\} : \bar{\sigma}_i^2 > 0\}$  και με  $J$ , συμβολίζεται το συμπλήρωμά τους ως προς το  $\{1, \dots, d_\mu\}$ .

Τότε για  $\lambda \in (0, 1]$ :

$$\bar{\phi} + \lambda(\phi - \bar{\phi}) = (\bar{\mu}, \bar{\sigma}_I^2, 0_J) + \lambda(\mu - \bar{\mu}, \sigma_I^2 - \bar{\sigma}_I^2, \sigma_J^2) = (\bar{\mu} + \lambda(\mu - \bar{\mu}), \bar{\sigma}_I^2 + \lambda(\sigma_I^2 - \bar{\sigma}_I^2), \lambda\sigma_J^2) \in \Phi$$

$$\lim_{\lambda \rightarrow 0^+} \frac{\tilde{L}(\bar{\phi} + \lambda(\phi - \bar{\phi})) - \tilde{L}(\bar{\phi})}{\lambda} = \lim_{\lambda \rightarrow 0^+} \frac{\tilde{L}(\bar{\mu} + \lambda(\mu - \bar{\mu}), \bar{\sigma}_I^2 + \lambda(\sigma_I^2 - \bar{\sigma}_I^2), \lambda\sigma_J^2) - \tilde{L}(\bar{\mu}, \bar{\sigma}_I^2, 0_J)}{\lambda}$$

**Λήμμα 6.2** Ισχύουν τα παρακάτω:

(i) Η Ε.Μ.Π.  $\hat{\phi} = (\hat{\mu}, 0) \in S$ .

(ii) Αν  $\bar{\phi} \in \Phi$  δεν είναι στάσιμο σημείο της  $\tilde{L}$ , τότε υπάρχουν  $\phi_1, \phi_2 \in \Phi$  ώστε:

$$\frac{d\tilde{L}}{d\lambda}(\bar{\phi} + \lambda(\phi_1 - \bar{\phi}))|_{\lambda=0} < 0 \quad \text{και} \quad \frac{d\tilde{L}}{d\lambda}(\bar{\phi} + \lambda(\phi_2 - \bar{\phi}))|_{\lambda=0} > 0$$

**Απόδειξη**

Για το (i). Έστω  $\phi = (\mu, \sigma^2) \in \tilde{\Phi}$  και  $\lambda \in (0, 1)$ . Από την Πρόταση 6.3, η  $\tilde{L}$  είναι άνω φραγμένη από το  $\tilde{L}(\hat{\phi})$ . Τότε:

$$\tilde{L}(\hat{\phi} + \lambda(\phi - \hat{\phi})) - \tilde{L}(\hat{\phi}) \leq 0$$

Επομένως:

$$0 \geq \lim_{\lambda \rightarrow 0^+} \frac{\tilde{L}(\hat{\phi} + \lambda(\phi - \hat{\phi})) - \tilde{L}(\hat{\phi})}{\lambda} = \frac{d\tilde{L}}{d\lambda}(\hat{\phi} + \lambda(\phi - \hat{\phi}))|_{\lambda=0}$$

Για το (ii) Αρκεί να δείξουμε ότι  $\nabla \tilde{L}(\bar{\phi})^t(\phi_1 - \bar{\phi}) \nabla \tilde{L}(\bar{\phi})^t(\phi_2 - \bar{\phi}) < 0$ , όπου  $\nabla \tilde{L}(\bar{\phi})^t$  ο ανάστροφος του  $\nabla_{\phi} \tilde{L}(\phi)|_{\phi=\bar{\phi}}$



## Ζητήματα σύγκλισης

Έστω  $\delta > 0$  και επιλέγουμε  $\phi_1 = -\delta \nabla \tilde{L}(\bar{\phi}) + \bar{\phi}$  και  $\phi_2 = \delta \nabla \tilde{L}(\bar{\phi}) + \bar{\phi}$ . Τότε έπεται πως:

$$\nabla \tilde{L}(\bar{\phi})^t(\phi_1 - \bar{\phi}) = -\delta \|\nabla \tilde{L}(\bar{\phi})\|^2 < 0 \quad \text{και} \quad \nabla \tilde{L}(\bar{\phi})^t(\phi_2 - \bar{\phi}) = \delta \|\nabla \tilde{L}(\bar{\phi})\|^2 > 0$$

αφού  $\nabla \tilde{L}(\bar{\phi}) \neq 0$ . □

Το σύνολο  $S$  περιέχει, σύμφωνα με το Λήμμα, όλα τα στάσιμα σημεία της  $\tilde{L}$  μαζί με κάποια σημεία που η κλίση (gradient) της  $\tilde{L}$  δεν είναι απαραίτητα 0. Τα σημεία αυτά είναι συνοριακά σημεία (κάποιο  $\sigma_i^2 = 0$ ) για τα οποία η πιθανοφάνεια αυξάνεται μόνο σε σημεία έξω από το  $\Phi$ .

Προκειμένου ο ΕΜ να συγκλίνει σε σημεία τοπικού μεγίστου, θα χρειαστούν αυστηρότερες συνθήκες, όπως και στη γενική περίπτωση (Θεώρημα 5.3). Αν  $\mathcal{M}$  είναι το σύνολο τοπικών μεγίστων της  $\tilde{L}$  στο  $\Phi$ , τότε η συνθήκη C3, για  $\phi \in S \setminus \mathcal{M}$  είναι μια ικανή συνθήκη:

**Θεώρημα 6.3** Έστω  $(\phi^{(k)})_{k \geq 0}$  μία ακολουθία που παράγεται από την  $\phi^{(k+1)} = \mathcal{M}(\phi^{(k)})$  και έστω  $\mathcal{M}$  το σύνολο τοπικών μεγίστων της  $\tilde{L}$  στο  $\Phi$ . Υποθέτουμε ότι ισχύει η C3:

$$\sup_{\phi' \in \Phi} Q(\phi', \phi) \geq Q(\phi, \phi) \quad \forall \phi \in S \setminus \mathcal{M}.$$

Τότε κάθε οριακό σημείο της  $(\phi^{(k)})_{k \geq 0}$  είναι τοπικό μέγιστο της  $\tilde{L}$  και η  $\tilde{L}(\phi^{(k)})_{k \geq 0}$  συγκλίνει στο  $\tilde{L}^*$ , όπου  $\tilde{L}^* = \tilde{L}((\phi)^*)$  για κάποιο τοπικό μέγιστο  $(\phi)^* \in S$ .

### Απόδειξη

Αρκεί να δείξουμε ότι για  $\phi^{(k)} \in \Phi \setminus \mathcal{M}$  ισχύει  $\tilde{L}(\phi^{(k+1)}) > \tilde{L}(\phi^{(k)})$ . Από την Πρόταση 6.5 έχουμε πως  $\tilde{L}(\phi^{(k+1)}) > \tilde{L}(\phi^{(k)})$  για κάθε  $\phi^{(k)} \in \Phi \setminus S$ . Από την Πρόταση 5.1, το πόρισμα 5.1 και την C3, για  $\phi^{(k)} \in S \setminus \mathcal{M}$  θα είναι:

$$\begin{aligned} \tilde{L}(\phi^{(k+1)}) &= Q(\phi^{(k+1)}; \phi^{(k)}) - H(\phi^{(k+1)}; \phi^{(k)}) = \sup_{\phi' \in \Phi} Q(\phi'; \phi^{(k)}) - H(\phi^{(k+1)}; \phi^{(k)}) \\ &> Q(\phi^{(k)}; \phi^{(k)}) - H(\phi^{(k)}; \phi^{(k)}) = \tilde{L}(\phi^{(k)}). \end{aligned}$$

□

Μπορούμε να έχουμε και εντελώς ανάλογα αποτελέσματα με εκείνα του Wu για τη σύγκλιση της ακολουθίας  $(\phi^{(k)})_{k \geq 0}$ , με ελάχιστες τροποποιήσεις στις αποδείξεις. Συγκεκριμένα, αν θέσουμε  $S(a) = \{\phi \in S : \tilde{L}(\phi) = a\}$  για  $a \in \mathbb{R}$ , τότε τα παρακάτω είναι ανάλογα των θεωρημάτων 5.4 και 5.5:

**Θεώρημα 6.4** Έστω  $\phi^{(k)}$  μία ακολουθία που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$  που ικανοποιεί τις συνθήκες του Θεωρήματος 6.2. Υποθέτουμε ότι

$$\mathcal{S}(\tilde{L}^*) = \{\phi^*\},$$

όπου  $\tilde{L}^*$  το όριο της  $(\tilde{L}(\phi^{(k)}))_{k \geq 0}$  στο Θεώρημα 6.2. Τότε  $\phi^{(k)} \rightarrow \phi^*$  καθώς  $k \rightarrow \infty$ .

## Ζητήματα σύγκλισης

Θέτουμε  $R(\psi; \phi) = \sup\{\frac{d}{d\lambda}Q((\psi + \lambda(\phi' - \psi); \phi)|_{\lambda=0} : \phi' \in \Phi\}$  για κάθε  $\psi, \phi \in \Phi$ . Επίσης θέτουμε  $\mathcal{L}(a) = \{\phi \in \Phi : \tilde{L}(\phi) = a\}$ .

**Θεώρημα 6.5** Έστω  $\phi^{(k)}$  μία ακολουθία που παράγεται από την  $\phi^{(k+1)} \in \mathcal{M}(\phi^{(k)})$  για την οποία ισχύει ότι  $R(\phi^{(k+1)}; \phi^{(k)}) \leq 0$ . Υποθέτουμε επίσης ότι η  $R$  είναι συνεχής ως προς και τα δύο ορίσματα. Τότε η  $(\phi^{(k)})_{k \geq 0}$  συγκλίνει σε ένα σημείο  $\phi^* \in S$  με  $\tilde{L}(\phi^*) = \tilde{L}^*$ , το όριο της  $(\tilde{L}(\phi^{(k)}))_{k \geq 0}$  στο Θεώρημα 6.2. Μάλιστα, αν το  $S(\tilde{L}^*)$  είναι διακριτό, τότε  $\phi^{(k)} \rightarrow \phi^*$  καθώς  $k \rightarrow \infty$ .

### Απόδειξη

Από τις αποδείξεις των θεωρημάτων 5.4 και 5.5 έχουμε πως  $\phi^{(k)} \rightarrow \phi^* \in \mathcal{L}(\tilde{L}^*)$ . Θα δείξουμε ότι  $\phi^* \in S$ . Επειδή η  $R$  είναι συνεχής και  $R(\phi^{(k+1)}; \phi^{(k)}) \leq 0$  όπως επίσης και  $R(\phi^*; \phi^*) \leq 0$ , θα είναι  $\frac{d}{d\lambda}H(\phi' + \lambda(\phi - \phi'); \phi')|_{\lambda=0} = 0$  για κάθε  $\phi \in \Phi$ . Οπότε:

$$\sup_{\phi \in \Phi} \frac{d}{d\lambda} \tilde{L}(\phi' + \lambda(\phi - \phi'))|_{\lambda=0} = R(\phi^*; \phi^*) \leq 0,$$

και το συμπέρασμα έπεται. □

Τέλος, έχουμε το ακόλουθο αποτέλεσμα:

**Πόρισμα 6.2** Υποθέτουμε πως το  $\hat{\phi}$  είναι μοναδικό μέγιστο της  $\tilde{L}$  στο  $\Phi$  και είναι επιπλέον και μοναδικό στοιχείο του  $S$ . Αν η  $R$  είναι συνεχής ως προς και τα 2 ορίσματα, τότε, για οποιαδήποτε ακολουθία ΕΜ,  $(\phi^{(k)})_{k \geq 0}$ , ισχύει  $\phi^{(k)} \rightarrow \hat{\phi}$ .

# Κεφάλαιο 7

## Αλγοριθμικές τεχνικές για το μοντέλο GreenLab

### 7.1 Εισαγωγή

Ο κύριος στόχος του κεφαλαίου αυτού είναι η παρουσίαση ενός αλγορίθμου Μπεϋζιανής λογικής που σχεδιάστηκε στη MatLab στα πλαίσια της παρούσας εργασίας και αφορά τη συγκεκριμένη έκδοση του μοντέλου GreenLab που περιγράφηκε στο κεφάλαιο 2. Επίσης, θα αναφερθούν και άλλες αλγοριθμικές τεχνικές (ενότητα 7.3) που έχουν σχεδιαστεί σε πρόσφατες εργασίες (cite και αφορούν το μοντέλο αυτό, και θα γίνει συνολική παρουσίαση των αποτελεσμάτων για την εκτίμηση των παραμέτρων. Επίσης θα γίνει και μία σύντομη περιγραφή της προσαρμογής της μεθόδου Gaussian Randomization που περιγράφηκε στο κεφάλαιο 6 στο μοντέλο LNAS.

Στην ενότητα 7.2 περιγράφεται αναλυτικά ο αλγόριθμος MCMC Bayes για το μοντέλο GreenLab και γίνεται μία περιγραφή των χαρακτηριστικών του. Τα αποτελέσματα για την εκτίμηση των παραμέτρων, τόσο του MCMC Bayes παρουσιάζονται στην ενότητα 7.4. Η εργασία ολοκληρώνεται με την ενότητα 7.5 στην οποία εφαρμόζεται στο μοντέλο LNAS η μέθοδος Gaussian Randomization για την εκτίμηση των παραμέτρων του.

### 7.2 Αλγόριθμος MCMC-Bayes

Στην ενότητα αυτή παρουσιάζεται ένας αλγόριθμος Μπεϋζιανής προσέγγισης για την εκτίμηση κάποιων από τις παραμέτρους του μοντέλου GreenLab στην περίπτωση του φυτού ζαχαρότευτλου. Όπως αναφέρθηκε στην εισαγωγή του κεφαλαίου 2, οι τύποι των οργάνων που συμπεριλαμβάνονται στο σύνολο  $\mathcal{O}$  είναι η ρίζα, τα ελάσματα και ο μίσχος. Δηλαδή  $\mathcal{O} = \{b, p, r\}$ . Ο αλγόριθμος που σχεδιάστηκε στα πλαίσια αυτής της εργασίας αφορά ορισμένες από τις παραμέτρους του παραπάνω μοντέλου. Συγκεκριμένα:

1. Από τις παραμέτρους που υπεισέρχονται στο μηχανισμό διανομής της βιομάζας μέσω των συναρτήσεων απορρόφησης,  $p_{al}^o = (a_o, b_o, p_o)$  για  $o \in \mathcal{O}$  θα εκτιμηθούν οι παράμετροι  $a_b, a_p, p_p$ . Οι υπόλοιπες παράμετροι του διανύσματος αυτού θεωρούνται γνωστές και οι τιμές τους παρουσιάζονται στο συγκεντρωτικό πίνακα παραμέτρων (πίνακας 7.2).
2. Παράμετρος προς εκτίμηση θεωρείται και η ανθεκτικότητα του φύλλου στη χρήση της ακτινοβολίας ( $\mu$ ), που εμφανίζεται στην αναδρομική συνάρτηση  $F_n$ , στη σχέση (2.2.5), που δίνει τις θεωρητικές βιομάζες  $(q_n)_{n=0}^N$ . Οι υπόλοιπες ποσότητες  $(s^{pr}, k_b)$  που εμφανίζονται στη σχέση αυτή θεωρούνται γνωστές.
3. Τέλος, το διάνυσμα των παραμέτρων θα περιλαμβάνει και τις τυπικές αποκλίσεις των σφαλμάτων στην παρατηρούμενη διαδικασία  $\mathbf{Y}_N$  που εκφράζει τα δεδομένα

(μάζες διαθέσιμων οργάνων στον κύκλο ανάπτυξης  $GC(N)$ ), καθώς και τη συσχέτιση τους. Οι υπόλοιπες παράμετροι θα θεωρούνται γνωστές ποσότητες. Το διάνυσμα των πραγματικών δεδομένων δίνεται στον πίνακα 7.1 στην ενότητα 7.4.

Από την υπόθεση 1 (ii) του κεφαλαίου 2, ο χρόνος που απαιτείται μέχρι τα ελάσματα και ο μίσχος να φτάσουν στο μέγιστο στάδιο της ανάπτυξης τους (από τη στιγμή τη δημιουργίας τους) είναι ίσος με τη φωτοσυνθετικά ενεργή περίοδο η οποία υποθέτουμε πως είναι ίση με 10 (κύκλους ανάπτυξης). Δηλαδή  $t_b = t_p = T = 10$ . Για τη ρίζα υποθέτουμε ότι  $t_r = 60$ . Επίσης, είναι  $N_n^o = 1$  για κάθε  $n \geq 0$  και  $o \in \{b, p\}$  για το ζαχαρότευτλο. Δηλαδή σε κάθε κύκλο ανάπτυξης δημιουργείται ακριβώς ένα νέο όργανο κάθε τύπου εκτός της ρίζας.

Σύμφωνα λοιπόν με τα παραπάνω το παραμετρικό διάνυσμα είναι το:

$$\theta = (\mu, a_b, a_p, p_p, \sigma_b, \sigma_p, \rho). \quad (7.2.1)$$

Η Πρόταση 2.1 στο κεφάλαιο 2 καθορίζει το πλαίσιο στο οποίο μπορούμε να κάνουμε στατιστική ανάλυση. Σύμφωνα λοιπόν με αυτή, οι δύο βασικές στοχαστικές διαδικασίες  $\mathbf{Q}$  και  $\mathbf{Y}$ , που επίσης ορίστηκαν στο κεφάλαιο 2, έχουν κανονικές κατανομές με την πρώτη να είναι κρυμμένη. Αναλυτικότερα, χρησιμοποιώντας την αναπαράσταση ως KMM:

Η κρυμμένη διαδικασία  $\mathbf{Q}$ , με τιμές στον  $\mathbb{R}_+^*$ , εξελίσσεται ως μία χρονικά μη-ομογενής  $T$ -οστής τάξης Μαρκοβιανή αλυσίδα η οποία ξεκινά από το  $q_0$  (αρχική βιομάζα σπόρου) και οι πιθανότητες μετάβασης καθορίζονται από την:

$$\mathbb{P}_\theta(Q_{n+1} \in \cdot \mid Q_{(n-T+1)+:n}) \stackrel{D}{\approx} \mathcal{N}(F_n(Q_{(n-T+1)+:n}; p_1), \sigma_Q^2 F_n^2(Q_{(n-T+1)+:n}; p_1)), \quad (7.2.2)$$

όπου  $p_1 = (\mu, a_b, a_p, p_p)$ . Επίσης, η παρατηρούμενη διαδικασία  $\mathbf{Y}$ , με τιμές στον  $(\mathbb{R}_+)^d$ , δοθείσης της  $\mathbf{Q}$ , σχηματίζει μία ακολουθία δεσμευμένα ανεξάρτητων τυχαίων διανυσμάτων όπου κάθε  $Y_n$  δοθείσης της  $\mathbf{Q}$  εξαρτάται μόνο από το διάνυσμα  $Q_{n:(n+T-1)}$  με δεσμευμένη κατανομή:

$$\mathbb{P}_\theta(Y_n \in \cdot \mid Q_{n:(n+T-1)}) \stackrel{D}{\approx} \mathcal{N}_2(G_n(Q_{n:(n+T-1)}; p_2), \Sigma), n \geq 0, \quad (7.2.3)$$

όπου  $p_2 = (a_b, a_p, p_p)$  και ο πίνακας συνδιακύμανσης, λαμβάνοντας υπόψη την επιλογή του  $\theta$ , είναι :

$$\Sigma = \begin{bmatrix} \sigma_b^2 & \rho \sigma_b \sigma_p \\ \rho \sigma_b \sigma_p & \sigma_p^2 \end{bmatrix} \quad (7.2.4)$$

Οι συναρτήσεις  $F_n, G_n$ , έχουν οριστεί στις (2.2.5) και (2.3.1) και όπως έχει αναφερθεί, η κανονικότητα ισχύει προσεγγιστικά (με μικρές διασπορές) καθώς οι τ.μ. είναι θετικές.

Ισοδύναμα, μπορούμε να χρησιμοποιήσουμε και την έκφραση ως μοντέλο χώρου καταστάσεων:

$$Q_{n+1} = F_n(Q_{(n-T+1)+:n}; p_1)(1 + W_n), \quad (7.2.5)$$

$$Y_n = G_n(Q_{n:(n+T-1)}; p_2) + V_n, \quad (7.2.6)$$

όπου  $W_n \sim \mathcal{N}(0, \sigma_Q^2)$  και  $V_n \sim \mathcal{N}_d(0, \Sigma)$ .

### 7.2.1 Η εκ των υστέρων κατανομή

Υποθέτουμε ότι η εκ των προτέρων κατανομή του  $\theta$  έχει πυκνότητα, και τη συμβολίζουμε με:

$$\pi^{(0)}(\theta) = \pi^{(0)}(\mu^{-1}, a_b, a_p, p_p, \Sigma) \quad (7.2.7)$$

Με  $(q_{0:N}, y_{0:N})$  συμβολίζουμε το διάνυσμα που αντιπροσωπεύει τις πραγματοποιήσεις από τη διαδικασία  $(\mathbf{Q}_N, \mathbf{Y}_N) = (Q_{0:N}, Y_{0:N})$ . Επιπλέον, υποθέτουμε γνωστή αρχική κατανομή, δηλαδή  $\mathbb{P}(Q_0 = q_0) = 1$ . Για να υπολογίσουμε την εκ των υστέρων κατανομή ( $\pi(\theta|q_{0:N}, y_{0:N})$ ) θα χρειαστεί πρώτα να κάνουμε ορισμένους υπολογισμούς:

Θα συμβολίσουμε με  $p_\theta(q_{1:N})$  την από κοινού πυκνότητα της  $\mathbf{Q}_N$  και με  $p_\theta(y_{0:N}|q_{0:N})$  την από κοινού πυκνότητα της  $\mathbf{Y}_N$  δοθείσης της  $\mathbf{Q}_N$ . Από την Πρόταση 2.1:

$$\begin{aligned} p_\theta(q_{1:N}) &= p_\theta(q_1) p_\theta(q_2|q_{(1-T)+:1}) \dots p_\theta(q_{N-1}|q_{(N-T-1)+:(N-2)}) p_\theta(q_N|q_{(N-T)+:(N-1)}) \\ &= \prod_{i=0}^{N-1} f_i(\cdot; \theta), \end{aligned}$$

όπου  $f_i(\cdot; \theta)$  η πυκνότητα της  $\mathcal{N}_+(F_i(q_{(i-T+1)+:i}; \theta), \sigma_Q^2 F_i^2(q_{(i-T+1)+:i}; \theta))$ .

Οπότε καταλήγουμε στην παρακάτω έκφραση για την πυκνότητα της κρυμμένης διαδικασίας:

$$p_\theta(q_{1:N}) = \prod_{i=0}^{N-1} (2\pi\sigma_Q^2 F_i^2(q_{(i-T+1)+:i}; \theta))^{-\frac{1}{2}} \exp \left\{ -\frac{(q_i - F_i(q_{(i-T+1)+:i}; \theta))^2}{2\sigma_Q^2 F_i^2(q_{(i-T+1)+:i}; \theta)} \right\} c_i^{-1} \mathbb{1}_{\{q_{i+1} > 0\}}, \quad (7.2.8)$$

όπου  $c_i = 1 - \Phi\left(\frac{-F_i(q_{(i-T+1)+:i}; \theta)}{\sigma_Q F_i(q_{(i-T+1)+:i}; \theta)}\right) = 1 - \Phi\left(-\frac{1}{\sigma_Q}\right)$ , και  $\Phi$  η συνάρτηση κατανομής της τυπικής κανονικής. Η ποσότητα αυτή είναι επομένως σταθερός αριθμός λόγω της υπόθεσης πως το  $\sigma_Q$  είναι γνωστό.

Με ανάλογο τρόπο μπορούμε να δείξουμε ότι:

$$p_\theta(y_{0:N}|q_{0:N}) = \prod_{i=0}^{N-T+1} (2\pi|\Sigma|)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y_i - G_i(q_{i:(i+T-1)}; \theta))^t \Sigma^{-1} (y_i - G_i(q_{i:(i+T-1)}; \theta)) \right\} (c'_i)^{-1} \mathbb{1}_{\{y_i > 0\}}, \quad (7.2.9)$$

όπου το ενδεχόμενο  $\{Y_i > 0\}$  είναι το  $\{Y_{1i} > 0, Y_{2i} > 0\}$  και  $c'_i = \mathbb{P}(Y_i > 0)$ .

Επίσης, στην (7.2.8) τα  $q_{0:N}$  εξαρτώνται από το  $\theta$  μόνο μέσω του  $p_1$  ενώ στην (7.2.9), μόνο μέσω του  $p_2$ . Τέλος, η παραπάνω σχέση, όπως αναφέρθηκε στην Πρόταση 2.1 ισχύει μόνο κατά προσέγγιση και επομένως το  $c'_i$  μπορεί να παραληφθεί.

Επομένως, έχουμε το ακόλουθο αποτέλεσμα:

**Πρόταση 7.1** Η εκ των υστέρων κατανομή του μοντέλου που ορίζεται από τις υποθέσεις 1-4 και από την Πρόταση 2.1, έχει την ακόλουθη μορφή:

$$\begin{aligned} \pi(\theta|q_{0:N}, y_{0:N}) &\propto \\ &\prod_{i=0}^{N-1} (\sigma_Q F_i(q_{(i-T+1)+:i}; \theta))^{-1} \exp \left\{ -\frac{(q_i - F_i(q_{(i-T+1)+:i}; \theta))^2}{2\sigma_Q^2 F_i^2(q_{(i-T+1)+:i}; \theta)} \right\} \mathbb{1}_{\{q_{i+1}>0\}} \\ &\prod_{i=0}^{N-T+1} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(y_i - G_i(q_{i:(i+T-1)}; \theta))^t \Sigma^{-1} (y_i - G_i(q_{i:(i+T-1)}; \theta)) \right\} (c'_i)^{-1} \mathbb{1}_{\{y_i>0\}} \\ &\pi^{(0)}(\mu^{-1}, a_b, a_p, p_p, \Sigma). \end{aligned} \quad (7.2.10)$$

### Απόδειξη

Είναι:

$$\begin{aligned} \pi(\theta|q_{0:N}, y_{0:N}) &= \frac{L(q_{0:N}, y_{0:N}|\theta) \pi^{(0)}(\theta)}{\int L(q_{0:N}, y_{0:N}|s) \pi^{(0)}(s) ds} \propto L(q_{0:N}, y_{0:N}|\theta) \pi^{(0)}(\theta) \\ &\propto L(q_{0:N}|\theta) L(y_{0:N}|q_{0:N}, \theta) \pi^{(0)}(\theta). \end{aligned}$$

Στην τελευταία έκφραση, οι συναρτήσεις πιθανοφάνειας αντιστοιχούν στις πυκνότητες  $p_\theta(y_{0:N}|q_{0:N})$  και  $p_{\theta'}(q_{1:N})$  καθώς για δεδομένο  $\theta$  εκφράζουν την ίδια ποσότητα. Κάνοντας τις αντικαταστάσεις από τις (7.2.8) και (7.2.9), το ζητούμενο έπεται άμεσα.  $\square$

Είναι φανερό πως η εκ των υστέρων κατανομή έχει μία πολύπλοκη έκφραση και επομένως δεν είναι εύκολη (ενδεχομένως ανέφικτη) η απευθείας συμπερασματολογία. Για το λόγο αυτό θα χρησιμοποιηθούν τεχνικές MCMC, προκειμένου να μπορέσουμε να προσομοιώσουμε δείγμα από αυτήν.

### 7.2.2 Σχεδιασμός του MCMC-Bayes

Ο στόχος του αλγορίθμου MCMC-Bayes που θα περιγραφτεί σε αυτή την παράγραφο είναι η προσομοίωση δείγματος από την εκ των υστέρων κατανομή (Πρόταση 7.1). Για αυτό το σκοπό θα χρησιμοποιηθεί ο Υβριδικός δειγματολήπτης Gibbs (Hybrid Gibbs sampler) όπου όπως έγινε λόγος στο κεφάλαιο 4, είναι η επέκταση του Gibbs sampler, στην οποία, για κάποια(-ες) από τις πλήρως δεσμευμένες κατανομές από τις οποίες γίνεται προσομοίωση, χρησιμοποιείται ο αλγόριθμος Metropolis-Hastings. Ο Υβριδικός Gibbs θα παράγει μία ακολουθία  $(\mathbf{X}^{(n)})_{n \geq 0}$  η οποία είναι επιθυμητό να συγκλίνει κατά κατανομή στην εκ των υστέρων κατανομή  $\pi(\theta|q_{0:N}, y_{0:N})$ .

Αρχικά θα περιγράψουμε πως πραγματοποιείται μία μετάβαση από την  $\mathbf{X}^{(n)} \stackrel{def}{=} ((\mathbf{Q}_N, \mathbf{Y}_N)|\Theta = \theta)^{(n)}$  στην  $\mathbf{X}^{(n+1)}$ .

Μπορούμε για ευκολία, να θεωρούμε τα εξής 2 στάδια:

1. Στο πρώτο στάδιο γίνεται προσομοίωση του  $((\mathbf{Q}_N^{(n)}|\mathbf{y}_N, \Theta = \theta')$  από τη συνάρτηση  $L(q_{0:N}|y_{0:N}, \theta')$ , θεωρώντας τις συντεταγμένες του διανύσματος  $(Y_{0:N}, \Theta)$

γνωστές, όπου  $\theta'$  η τρέχουσα τιμή του παραμετρικού διανύσματος. Με άλλα λόγια, στο πρώτο στάδιο κάνουμε προσομοίωση των κρυμμένων καταστάσεων.

2. Στο δεύτερο στάδιο, θεωρώντας τις κρυμμένες καταστάσεις γνωστές, κάνουμε προσομοίωση του διανύσματος  $\Theta^{(n)}$  για το οποίο μπορούμε να θεωρήσουμε περαιτέρω υποστάδια στα οποία πραγματοποιείται προσομοίωση από κάθε συνιστώσα.

Στη συνέχεια, θα συζητηθούν οι τεχνικές λεπτομέρειες όπως για παράδειγμα ο τρόπος μετάβασης (κυκλικός ή τυχαίος), οι συγκεκριμένες κατανομές που θα επιλέξουμε ως κατανομές Πρόταση για τα βήματα που εφαρμόζεται ο αλγόριθμος MH κ.α. Επίσης θα συζητηθούν θέματα σύγκλισης της κατανομής και του εργοδικού μέσου.

### 7.2.3 Περιγραφή των μεταβάσεων

#### 1ο Στάδιο.

Στο στάδιο αυτό θα περιγράψουμε πως θα γίνει η προσομοίωση από την  $L(q_{0:N}, y_{0:N} | \theta')$ , όπου  $\theta'$  η τρέχουσα τιμή του παραμετρικού διανύσματος (δηλαδή το  $\theta'$  θα είναι γνωστό και σταθερό στο στάδιο αυτό). Είναι επιθυμητό, η προσομοίωση αυτή να γίνει με τον πιο αποδοτικό τρόπο.

Για να χρησιμοποιήσουμε τον Gibbs sampler θα πρέπει αρχικά να προσδιοριστούν οι πλήρεις δεσμευμένες κατανομές. Χρησιμοποιώντας το συμβολισμό του κεφαλαίου 4, θα είναι:  $\pi_i(\cdot | q^{(-i)}, y_{0:N}, \theta')$  η πλήρης δεσμευμένη κατανομή του  $Q_i$  δοθέντος όλων των υπολοίπων, για  $i = 1, \dots, N$  (καθώς το  $q_0$  είναι σταθερά) και όπου  $x^{(-i)}$  το διάνυσμα  $x$  χωρίς την  $i$ -συνιστώσα. Τότε:

$$\begin{aligned} \pi_i(s | q^{(-i)}, y_{0:N}, \theta') &= \frac{L(q_{0:(i-1)}, s, q_{(i+1):N}, y_{0:N} | \theta')}{\int L(q_{0:(i-1)}, s, q_{(i+1):N}, y_{0:N} | \theta') ds} \\ &\propto L(q_{0:(i-1)}, s, q_{(i+1):N}, y_{0:N} | \theta') \\ &\propto L(q_{0:N} | \theta') L(y_{0:N} | q_{0:(i-1)}, s, q_{(i+1):N}, \theta') \end{aligned} \quad (7.2.11)$$

Στην τελευταία έκφραση, αν κρατήσουμε μόνο τους όρους που εξαρτώνται από το  $s$  θα καταλήξουμε στην:

**Πρόταση 7.2** Η πλήρης δεσμευμένη πυκνότητα στο βήμα  $i = 1, \dots, N$  του Gibbs sampler είναι:

$$\begin{aligned} \pi_i(s | q^{(-i)}, y_{0:N}, \theta') &\propto \prod_{k=i+1}^{(i+T) \wedge N} \left( 1 - \exp \left\{ -k_B \frac{s_k^{act}(s)}{s^{pr}} \right\} \right) \\ &\times \exp \left\{ -\frac{1}{2\sigma_Q^2} \left( \left( \frac{s}{F_{i-1}(q_{(i-T)^+:(i-1)})} - 1 \right)^2 + \sum_{k=i+1}^{(i+T) \wedge N} \left( \frac{q_k}{F_{k-1}(s)} - 1 \right)^2 \right) \right\} \\ &\times \exp \left\{ -\frac{1}{2} \sum_{k=(i-T+1)^+}^{(N-T+1) \wedge i} (y_k - G_k(s))^t \Sigma^{-1} (y_k - G_k(s)) \right\}, \end{aligned} \quad (7.2.12)$$

όπου οι συναρτήσεις  $F$ ,  $s^{act}$  έχουν οριστεί στις (2.2.4) και (2.2.5) και εδώ εκφράζονται ως συναρτήσεις μονάχα του  $s$ . Επίσης, η  $F_{i-1}(q_{(i-T)+:(i-1)})$  δεν εξαρτάται από το  $s$ .

Είναι φανερό πως δεν είναι εύκολο να προσομοιώσουμε απευθείας από την (7.2.12). Για το λόγο αυτό, μπορούμε να χρησιμοποιήσουμε τον MH για τον οποίο έγινε λόγος στην ενότητα 4.1.

Έστω  $\pi_i$ ,  $i = 1, \dots, N$ , οι μη κανονικοποιημένες πυκνότητες στην Πρόταση 7.2, και  $f_i(z_i | q_{1:i-1}, q_{i+1:N})$ ,  $i = 1, \dots, N$  οι πυκνότητες που αντιστοιχούν στην κατανομή Πρότασης. Τότε ο υβριδικός δειγματολήπτης Gibbs περιγράφεται ως εξής:

Στο βήμα  $n$ , δοθέντος του  $q_{1:N}^{(n-1)}$ :  
 Για  $k = 1$  μέχρι  $N$   
 1. Προσομοίωση της:  $z_k \sim f_k(\cdot | q_{1:k-1}^{(n)}, q_{k+1:N}^{(n-1)})$   
 2. Θέτουμε  $q_k^{(n)} = z_k$  με πιθανότητα  $\alpha(q_k^{(n-1)}, z_k)$ ,  
 διαφορετικά θέτουμε  $q_k^{(n)} = q_k^{(n-1)}$

Η πιθανότητα αποδοχής για το βήμα  $n$  είναι:

$$\alpha(q_i^{(n-1)}, z_i) = \min \left\{ 1, \frac{\pi_i(z_i | q_{1:i-1}, q_{i+1:N}, r)}{\pi_i(q_i^{(n-1)} | q_{1:i-1}, q_{i+1:N}, r)} \times \frac{f_i(q_i^{(n-1)} | q_{1:i-1}, z_i, q_{i+1:N}^{(n-1)})}{f_i(z_i | q_{1:i-1}, q_i^{(n-1)}, q_{i+1:N}^{(n-1)})} \right\}, \quad (7.2.13)$$

όπου  $r = (y_{0:N}, \theta')$ .

Για την κατανομή-πρόταση, υπάρχουν πολλές επιλογές συναρτήσεων με τις οποίες να ικανοποιούνται οι συνθήκες (4.2.1) και (4.2.2) που διασφαλίζουν τη σύγκλιση στη κατανομή-στόχο. Ωστόσο, η σύγκλιση είναι πιο γρήγορη όταν η κατανομή-πρόταση είναι κατά προσέγγιση κοντά στην κατανομή-στόχο. Έτσι, ως κατανομή-πρόταση για τις κρυμμένες καταστάσεις  $((Q_n)_{n=1}^N)$  έχει επιλεγθεί η:

$$\mathbb{P}_\theta(Q_{n+1} \in \cdot | Q_{(n-T+1)+:n}) \stackrel{D}{\approx} \mathcal{N}(F_n(Q_{(n-T+1)+:n}; p'_1), \sigma_Q^2 F_n^2(Q_{(n-T+1)+:n}; p'_1)), \quad (7.2.14)$$

δοθέντος της τρέχουσας τιμής για τις παραμέτρους  $p'_1 \subset \theta'$ .

Η παραπάνω κατανομή είναι εύκολο να προσομοιωθεί και δεν έχει μεγάλο υπολογιστικό κόστος, γεγονός που προστίθεται στα υπέρ για τη συγκεκριμένη επιλογή οικογένειας.

## 2ο Στάδιο.

Για να προχωρήσουμε στην προσομοίωση από την εκ των υστέρων κατανομή για δεδομένο διάλυμα κρυμμένων καταστάσεων, θα πρέπει πρώτα να προσδιοριστεί η εκ των προτέρων κατανομή  $(\pi^{(0)}(\theta))$ .



**Υπόθεση 7.1** Υποθέτουμε τις ακόλουθες κατανομές για το  $\theta = (\mu^{-1}, a_b, a_p, p_p, \Sigma)$ :

$$\begin{aligned}\mu^{-1} &\sim \mathcal{N}(m_{\mu^{-1}}, \sigma_{\mu^{-1}}^2) \mathbb{1}_{\{\mu^{-1} > 0\}} \\ a_b &\sim \mathcal{N}(m_{a_b}, \sigma_{a_b}^2) \mathbb{1}_{\{a_b > 1\}}, \\ a_p &\sim \mathcal{N}(m_{a_p}, \sigma_{a_p}^2) \mathbb{1}_{\{a_p > 1\}}, \\ p_p &\sim \mathcal{N}(m_{p_p}, \sigma_{p_p}^2) \mathbb{1}_{\{p_p > 0\}}, \\ \Sigma &\sim \mathcal{IW}_2(\Psi, \nu)\end{aligned}\tag{7.2.15}$$

Υποθέτουμε ότι έχει γίνει η προσομοίωση των κρυμμένων καταστάσεων στο 1ο στάδιο και αυτές θα τις συμβολίζουμε με  $q'_{1:N}$  (θα είναι  $q'_i > 0$  για κάθε  $i$ ). Συνεχίζοντας τον δειγματολήπτη Gibbs στο στάδιο αυτό, θα συμβολίσουμε με  $\pi_j(\cdot | \theta^{(-j)}, q_0, q'_{1:N}, y_{0:N})$  τις αναλυτικές εκφράσεις για τις πλήρεις δεσμευμένες πυκνότητες (διατηρώντας το συμβολισμό του κεφαλαίου 4), όπου με  $\theta^{(-j)}$  συμβολίζεται το διάνυσμα των παραμέτρων χωρίς την  $i$ -συνιστώσα. Ωστόσο, όποτε είναι βολικό, θα συμβολίζουμε τις συναρτήσεις αυτές και με  $p(j | \dots)$ . Έστω  $r = (q_0, q'_{1:N}, y_{0:N})$ . Θα είναι:

$$\begin{aligned}\pi_j(s | \theta^{(-j)}, r) &= \frac{L((r | s, \theta^{(-j)}) \pi^{(0)}(s, \theta^{(-j)})}{\int L((r | s, \theta^{(-j)}) \pi^{(0)}(s, \theta^{(-j)}) ds} \\ &= \frac{L((r | s, \theta^{(-j)}) \pi^{(0)}(s))}{\int L((r | s, \theta^{(-j)}) \pi^{(0)}(s) ds} \\ &\propto L((r | s, \theta^{(-j)}) \pi^{(0)}(s)) \\ &= L(q_0, q'_{1:N} | s, \theta^{(-j)}) L(y_{0:N} | q_0, q'_{1:N}, s, \theta^{(-j)}) \pi^{(0)}(s)\end{aligned}\tag{7.2.16}$$

Όπως προαναφέρθηκε, μπορούμε να χωρίσουμε το στάδιο αυτό σε περαιτέρω υποστάδια (μέρη (i)- (v)), τα οποία θα περιγράψουμε στη συνέχεια.

## 2ο Στάδιο, μέρος (i): Για το $\mu^{-1}$ :

Στο μέρος αυτό θα κάνουμε προσομοίωση (για  $j = \mu^{-1}$ ) από την  $\pi_j(\cdot | \theta^{(-j)}, q_0, q'_{1:N}, y_{0:N})$  Από την (7.2.16) είναι:

$$\begin{aligned}\pi_j(\cdot | \theta^{(-j)}, q_0, q'_{1:N}, y_{0:N}) &\propto L(q_0, q'_{1:N} | \mu^{-1}, \theta^{(-j)}) L(y_{0:N} | q_0, q'_{1:N}, \mu^{-1}, \theta^{(-j)}) \pi^{(0)}(\mu^{-1}) \\ &\propto L(q_0, q'_{1:N} | \mu^{-1}, \theta^{(-j)}) \pi^{(0)}(\mu^{-1}).\end{aligned}$$

Διότι η δεύτερη συνάρτηση στο δεξί μέλος δεν εξαρτάται από το  $\mu^{-1}$ , παρά μόνο από το  $p_2 = (a_b, a_p, p_p)$  το οποίο συμπεριλαμβάνεται στο  $\theta^{(-j)}$ . Επίσης, έχουμε ότι:

$$\begin{aligned}L(q_0, q'_{1:N} | \mu, \theta^{(-j)}) &= \prod_{i=0}^{N-1} (2\pi\sigma_Q^2 (\mu K_i(q_{(i-T+1)+:i}; \theta^{(-j)})))^{-\frac{1}{2}} \exp \left\{ -\frac{(q_i - \mu K_i(q_{(i-T+1)+:i}; \theta^{(-j)}))^2}{2\sigma_Q^2 (\mu K_i(q_{(i-T+1)+:i}; \theta^{(-j)}))^2} \right\} \\ &\propto \mu^{-N} \exp \left\{ -\frac{1}{2\sigma_Q^2} \sum_{i=0}^{N-1} \left( \frac{q'_i}{\mu K_i(q'_{(i-T+1)+:i}; \theta^{(-j)})} - 1 \right)^2 \right\},\end{aligned}$$

όπου  $\mu K_i(q_{(i-T+1)+:i}; \theta^{(-j)}) = F_i(q_{(i-T+1)+:i}; \mu^{-1}, \theta^{(-j)})$ .

Επομένως, κάνοντας αντικατάσταση της εκ των προτέρων κατανομής  $\pi^{(0)}(\mu^{-1})$  θα έχουμε δείξει ότι:

**Πρόταση 7.3** Η πλήρης δεσμευμένη κατανομή για το  $\mu^{-1}$  δίνεται από την παρακάτω:

$$p(\mu^{-1}|\dots) \propto \mu^{-N} \exp \left\{ -\frac{\mu^{-1}}{2\sigma_Q^2} \sum_{i=0}^{N-1} \left( \frac{q'_i}{K_i(q'_{(i-T+1)+:i}; \theta^{(-j)})} - 1 \right)^2 \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma_{\mu^{-1}}^2} (\mu^{-1} - m_{\mu^{-1}})^2 \right\}. \quad (7.2.17)$$

### Παρατήρηση

Εδώ, όπως και στα επόμενα βήματα ((ii)-(iv)), η εκ των προτέρων κατανομή είναι περικομμένη (truncated). Πάραυτα:

$$f_{\mu^{-1}|\mu^{-1}>0}(\cdot) = \frac{f_{\mu^{-1}}(\cdot)}{\mathbb{P}_{(m_{\mu^{-1}}, \sigma_{\mu^{-1}}^2)}(\mu^{-1} > 0)} = \frac{f_{\mu^{-1}}(\cdot)}{\Phi\left(\frac{m_{\mu^{-1}}}{\sigma_{\mu^{-1}}}\right)} \\ \propto f_{\mu^{-1}}(\cdot) \propto \exp \left\{ -\frac{1}{2\sigma_{\mu^{-1}}^2} (\mu^{-1} - m_{\mu^{-1}})^2 \right\}$$

Δεν είναι εύκολο να κάνουμε απευθείας προσομοίωση από την (7.2.17) και έτσι επιλέγουμε να χρησιμοποιήσουμε τον αλγόριθμο ΜΗ τυχαίου περιπάτου από οικογένεια μετατοπισμένων κανονικών κατανομών. Έτσι:

- Στο βήμα  $n$ , δοθέντος του  $(1/\mu)^{(n-1)}$
1. Προσομοίωση της:  $z \sim \mathcal{N}_+((1/\mu)^{(n-1)}, \sigma_{MH(\mu^{-1})}^2)$
  2. Θέτουμε  $(1/\mu)^{(n)} = z$  με πιθανότητα  $\alpha((1/\mu)^{(n-1)}, z)$ , διαφορετικά θέτουμε  $(1/\mu)^{(n)} = (1/\mu)^{(n-1)}$

Η πιθανότητα αποδοχής για το βήμα  $n$  είναι:

$$\alpha\left(\frac{1}{\mu}^{(n-1)}, z\right) = \min \left\{ 1, \frac{p(z|\dots)}{p\left(\frac{1}{\mu}^{(n-1)}|\dots\right)} \times \frac{g\left(\frac{1}{\mu}^{(n-1)}|z\right) \Phi^{-1}\left(\frac{z}{\sigma_{MH\left(\frac{1}{\mu}^{(n-1)}}\right)}\right)}{g\left(z|\frac{1}{\mu}^{(n-1)}\right) \Phi^{-1}\left(\frac{\frac{1}{\mu}^{(n-1)}}{\sigma_{MH\left(\frac{1}{\mu}^{(n-1)}}\right)}\right)} \right\} \\ = \min \left\{ 1, \frac{p(z|\dots)}{p\left(\frac{1}{\mu}^{(n-1)}|\dots\right)} \times \frac{\Phi^{-1}\left(\frac{z}{\sigma_{MH\left(\frac{1}{\mu}^{(n-1)}}\right)}\right)}{\Phi^{-1}\left(\frac{\frac{1}{\mu}^{(n-1)}}{\sigma_{MH\left(\frac{1}{\mu}^{(n-1)}}\right)}\right)} \right\}, \quad (7.2.18)$$

όπου  $g(\cdot|x)$  η πυκνότητα της  $\mathcal{N}(x, \sigma_{MH(\mu^{-1})}^2)$  και επομένως η δεύτερη ισότητα δικαιολογείται λόγω της συμμετρίας της  $g$ . Οι νέοι όροι που εμφανίζονται στην παραπάνω αφορούν τη μετατόπιση.

**2ο Στάδιο, μέρη (ii)-(iv):** Για τις  $a_b, a_p, p_p$

Οι παράμετροι  $a_b, a_p, p_p$  εμφανίζονται στις συναρτήσεις απορρόφησης (;;) και οι τελευταίες αποτελούν βασικό μέρος τόσο για την παραγωγή της βιομάζας (συνάρτηση  $F_i(q_{(i-T+1)+:i}; \theta)$  όσο και για τη συνάρτηση που εκφράζει τις θεωρητικές τελικές μάζες ( $G_n$  στην (2.3.1)).

Συγκεκριμένα, οι  $a_b, a_p$  επηρεάζουν κατά ανάλογο τρόπο τη διανομή βιομάζας και έτσι η προσομοίωση από τις αντίστοιχες πλήρεις δεσμευμένες κατανομές:

$$p(a_b | \dots) \propto L(q_0, q'_{1:N} | a_b, \theta^{(-j)}) L(y_{0:N} | q_0, q'_{1:N}, a_b, \theta^{(-j)}) \pi^{(0)}(a_b), \quad (7.2.19)$$

και

$$p(a_p | \dots) \propto L(q_0, q'_{1:N} | a_p, \theta^{(-j)}) L(y_{0:N} | q_0, q'_{1:N}, a_p, \theta^{(-j)}) \pi^{(0)}(a_p), \quad (7.2.20)$$

γίνεται κατά όμοιο τρόπο.

Για τη  $p_p$ , η πλήρης δεσμευμένη κατανομή είναι:

$$p(p_p | \dots) \propto L(q_0, q'_{1:N} | p_p, \theta^{(-j)}) L(y_{0:N} | q_0, q'_{1:N}, p_p, \theta^{(-j)}) \pi^{(0)}(p_p). \quad (7.2.21)$$

Για την προσομοίωση από τις παραπάνω εφαρμόζουμε τον αλγόριθμο MH τυχαίου περιπάτου από οικογένειες μετατοπισμένων κανονικών κατανομών.

Για  $\gamma \in \{a_b, a_p, p_p\}$  ο αλγόριθμος MH είναι:

- Στο βήμα  $n$ , δοθέντος του  $\gamma^{(n-1)}$
1. Προσομοίωση της:  $z \sim \mathcal{N}(\gamma^{(n-1)}, \sigma_{MH(\gamma)}^2) \mathbb{1}_{\{\gamma > r_\gamma\}}$
  2. Θέτουμε  $\gamma^{(n)} = z$  με πιθανότητα  $\alpha(\gamma^{(n-1)}, z)$ ,  
διαφορετικά θέτουμε  $\gamma^{(n)} = \gamma^{(n-1)}$

Όπου το  $r_\gamma$  είναι 0 όταν  $\gamma = p_p$  και 1 διαφορετικά.

Για τις πιθανότητες αποδοχής, επειδή η κατανομή-πρόταση είναι από οικογένεια περικομμένων κανονικών, θα υπάρχει, όπως είχαμε και στην περίπτωση του  $\mu^{-1}$ , ο όρος που αντιστοιχεί στη μετατόπιση, συναρτήσεως της σ.κ. της τυποποιημένης. Συγκεκριμένα, για  $\gamma \in \{a_b, a_p, p_p\}$  θα είναι:

$$\alpha(x, z) = \min \left\{ 1, \frac{p(z | \dots)}{p(x | \dots)} \times \frac{\Phi^{-1}\left(\frac{y - r_\gamma}{\sigma_{MH(p_p)}}\right)}{\Phi^{-1}\left(\frac{x - r_\gamma}{\sigma_{MH(p_p)}}\right)} \right\} \quad (7.2.22)$$

**2ο Στάδιο, μέρος (v): Προσομοίωση του πίνακα  $\Sigma$**

Στο μέρος αυτό, προσομοιώνουμε τις παραμέτρους των σφαλμάτων στη διαδικασία των παρατηρήσεων  $\mathbf{Y}_N$ , δηλαδή τις τυπικές αποκλίσεις  $\sigma_b, \sigma_p$  για τις μετρήσεις των ελασμάτων και του μίσχου καθώς και τη συσχέτιση τους:  $\rho$ . Στην περίπτωση αυτή, η προσομοίωση μπορεί να γίνει απευθείας. Πράγματι:

Οι μόνες παράμετροι που μεταβάλλονται στο μέρος αυτό είναι οι  $\sigma_b, \sigma_p, \rho$ . Οπότε αν  $q'_{1:N}$  και  $(\theta^{(-j)})' = ((\mu^{-1})', a'_p, a'_b, p'_p)$  οι τρέχουσες τιμές των παραμέτρων για αυτό το μέρος, η εκ των υστέρων κατανομή γίνεται:

$$\begin{aligned} \pi(\Sigma, (\theta^{(-j)})' | q_0, q'_{1:N}, y_{0:N}) &\propto L(q_0, q'_{1:N} | p'_1) L(y_{0:N} | q_0, q'_{1:N}, a'_p, a'_b, p'_p, \Sigma) \pi^{(0)}(\Sigma) \\ &\propto L(y_{0:N} | q_0, q'_{1:N}, a'_p, a'_b, p'_p, \Sigma) \pi^{(0)}(\Sigma). \end{aligned} \quad (7.2.23)$$

Η συνάρτηση πιθανοφάνειας στην τελευταία έκφραση αντιστοιχεί στην πυκνότητα της  $\mathbf{Y}_N$  δοθείσης της  $\mathbf{Q}$  που από την Πρόταση 2.1 είναι πυκνότητα πολυδιάστατης κανονικής με μέσο το διάνυσμα  $G$  και πίνακα συνδιακύμανσης τον  $\Sigma$ . Αφού ως εκ των προτέρων κατανομή έχει την αντίστροφη Wishart κατανομή, η εκ των υστέρων κατανομή θα έχει επίσης την αντίστροφη Wishart. Αυτό είναι το αποτέλεσμα της επόμενης Πρότασης για το οποίο η απόδειξη παραλείπεται.

**Πρόταση 7.4** Έστω  $\Psi$  θετικά ορισμένος πίνακας διάστασης  $p \times p$  και  $v > p - 1$ . Έστω επίσης  $\Sigma$  ένας επίσης θετικά ορισμένος πίνακας διάστασης  $p \times p$ . Αν η εκ των προτέρων κατανομή του  $\Sigma$ , έστω  $\pi^{(0)}(\Sigma)$  έχει την  $\mathcal{IW}_p(\Psi, v)$  κατανομή, και αν  $\mathbf{X}_N \sim \mathcal{N}_p(0, \Sigma)$ , τότε για την εκ των υστέρων κατανομή ισχύει:

$$\Sigma | \mathbf{X}_N \sim \mathcal{IW}_p(\Psi + A, v + N), \quad (7.2.24)$$

όπου  $A = \mathbf{X}_N \mathbf{X}_N^t$ .

Με άμεση εφαρμογή της παραπάνω για την πλήρη δεσμευμένη κατανομή του  $\Sigma$ , παίρνουμε ότι:

**Πρόταση 7.5** Υπό τις υποθέσεις 1-4 του κεφαλαίου 2 και της Πρότασης 2.1 καθώς και των υποθέσεων 7.1, η πλήρης δεσμευμένη κατανομή για τον πίνακα συνδιακυμάνσεων στη διαδικασία παρατηρήσεων έχει την κατανομή αντίστροφη Wishart:

$$p(\Sigma | \dots) \sim \mathcal{IW}_2(\Psi + A, v + N - T + 1), \quad (7.2.25)$$

όπου  $A = Y^* Y^{*t}$  και  $y_i^* = y_i - G_i(q'_{i:(i+T-1)}, a'_p, a'_b, p'_p)$ .

Περιγράψαμε επομένως τον τρόπο με τον οποίο γίνονται οι μεταβάσεις στην αλυσίδα  $(\mathbf{X}^{(n)})_{n \geq 0}$  που παράγεται από τον Υβριδικό Gibbs.

**7.2.4 Ανάλυση της αλυσίδας**

Στην υποενότητα αυτή θα συζητηθούν ορισμένες τεχνικές λεπτομέρειες που αφορούν την αποδοτικότητα του αλγορίθμου. Από τα προηγούμενα είδαμε τον τρόπο με τον

οποίο μπορούμε να προσομοιώσουμε δείγμα από τις πλήρεις δεσμευμένες κατανομές. Ωστόσο, δεν καθορίστηκε η σειρά με την οποία επιλέγονται αυτές (δηλαδή η σειρά με την οποία ανανεώνονται οι συνιστώσες της αλυσίδας  $(\mathbf{X}^{(n)})_{n \geq 0}$ ). Σχετικά με αυτό έχει επιλεγεί ο συστηματικός (κυκλικός τρόπος). Όπως έχουμε δει στο κεφάλαιο 4, με τον τρόπο αυτό η αλυσίδα που παράγεται, δεν είναι αντιστρέψιμη.

Οι διασπορές στους ΜΗ των (i)-(iv):  $\sigma_{MH(\mu^{-1})}^2, \sigma_{MH(a_b)}^2, \sigma_{MH(a_p)}^2$ , και  $\sigma_{MH(p_p)}^2$  έχουν επιλεγεί ώστε οι πιθανότητες αποδοχής που αντιστοιχούν στα μέρη αυτά, να είναι κατά προσέγγιση βέλτιστες, δηλαδή κοντά στο 0.44 (κεφάλαιο 4). Ένας τρόπος που αυτό μπορεί να γίνει είναι να ρυθμίζουμε κατάλληλα τη διασπορά ανά τακτά χρονικά διαστήματα (με βάση το τρέχων ποσοστό αποδοχής). Συγκεκριμένα, για  $\gamma \in \{\mu^{-1}, a_b, a_p, p_p\}$ , η διασπορά  $(\sigma_{MH(j)}^2)^{(n)}$ , αφού αρχικοποιηθεί ( $n = 0$ ), στη συνέχεια, όταν  $n = 0 \pmod{k}$ , ρυθμίζεται ως εξής:

$$(\sigma_{MH(j)}^2)^{(n+1)} = (\sigma_{MH(j)}^2)^{(n)}(1 + (\alpha_n - 0.44)), \quad (7.2.26)$$

όπου  $k$  κάποιος θετικός ακέραιος ανάλογος του πλήθους επαναλήψεων και  $\alpha_n$  το ποσοστό αποδοχής μέχρι την  $n$ -οστή επανάληψη.

Το μπερν  $in$  ( $m_{burn\ in}$ ) (πλήθος επαναλήψεων στο οποίο θεωρούμε πως έχει επιτευχθεί η σύγκλιση, δηλαδή:  $\mathbf{X}^{(n)} \sim p(j | \dots)$  για  $n > m_{burn\ in}$ ) θεωρούμε πως είναι 10%.

#### Δυνατές επεκτάσεις του αλγορίθμου

Στα πλαίσια της παρούσας εργασίας έχουμε υποθέσει ότι οι παράμετροι για εκτίμηση είναι αυτές που συμπεριλαμβάνονται στο διάλυμα  $\theta$ . Ένας μελλοντικός στόχος είναι να επεκταθεί ο αλγόριθμος ώστε να περιλαμβάνει περισσότερες παραμέτρους. Για παράδειγμα, στα άμεσα σχέδια είναι να συμπεριληφθεί στο  $\theta$  η παράμετρος των σφαλμάτων ( $\sigma_Q$ ) των θεωρητικών βιομαζών ( $Q_n$ ). Αυτό μάλιστα μπορεί να γίνει χωρίς ιδιαίτερο υπολογιστικό κόστος υποθέτοντας μία φυσική επιλογή εκ των προτέρων κατανομής, όπως για παράδειγμα η αντίστροφη Γάμμα (Inverse Gamma). Επίσης στα άμεσα πλάνα είναι και η εφαρμογή του αλγορίθμου για μοντέλα με πολλαπλασιαστικά σφάλματα κλπ.

### 7.3 Άλλες αλγοριθμικές τεχνικές για το μοντέλο GreenLab

Στην ενότητα αυτή θα αναφερθούν ορισμένοι τρόποι με τους οποίους είναι εφικτή η εκτίμηση τόσο των κρυμμένων καταστάσεων όσο και των παραμέτρων για το συγκεκριμένο φυτό στο οποίο εφαρμόστηκε ο αλγόριθμος MCMC-Bayes. Το πρώτο ζήτημα που θα συζητηθεί είναι η εκτίμηση της δεσμευμένης κατανομής της κρυμμένης ακολουθίας καταστάσεων  $Q_{0:N}$  δοθέντος της ακολουθίας των παρατηρήσεων  $Y_{0:N}$  με τη μέθοδο της μέγιστης πιθανοφάνειας. Συγκεκριμένα, η εκτίμηση του  $p_\phi(q_{0:N} | y_{0:N})$  είναι εφικτή σε περιπτώσεις KMM με πεπερασμένο χώρο καταστάσεων (Baum et al. 1970) ή για KMM που μπορούν να αναπαρασταθούν ως γραμμικά κανονικά μοντέλα καταστάσεων (Carré, Moulines, and Rydén (2005) κεφάλαιο 5). Ωστόσο, το συγκεκριμένο μοντέλο GreenLab που περιγράφηκε στην εργασία αυτή έχει συνεχή χώρο καταστάσεων και η κρυμμένη ακολουθία καταστάσεων είναι μη-γραμμική. Για το λόγο

αυτό, αλγόριθμοι τύπου EM είναι μία φυσική επιλογή καθώς, όπως συζητήθηκε στο κεφάλαιο 5, η προσαρμογή του EM σε προβλήματα με μη πλήρη δεδομένα είναι πολύ καλή. Στο Trevezas & Cournéde (2013), προτείνεται η προσαρμογή του αλγορίθμου EM και πιο συγκεκριμένα, μιας στοχαστικής έκδοσης του (ECM), όπου για την πραγματοποίηση του (Monte Carlo) E-Βήματος, χρησιμοποιούν ακολουθιακό δειγματολήπτη σπουδαιότητας με επαναδειγματοληψία (Sequential Importance Sampling with Resampling (SISR)). Για περισσότερες λεπτομέρειες για των τεχνικών αυτών παραπέμπουμε στο Cappé, Moulines, and Rydén (2005). Πιο αναλυτικά, για το μοντέλο αυτό έχει πραγματοποιηθεί σύγκριση αλγοριθμικών τεχνικών της παραπάνω κατεύθυνσης η οποία παρουσιάζεται αναλυτικά στο Trevezas, Malefaki, Cournéde (2014). Μάλιστα, στη δημοσίευση αυτή, χρησιμοποιούνται και MCMC τεχνικές για την εκτίμηση την διαδικασίας των κρυμμένων καταστάσεων με τη λογική του σταδίου 1 που θεωρήθηκε εδώ.

Στην παρούσα εργασία αντιπαραβάλλονται τα αποτελέσματα από τον MCMC-Bayes με εκείνα από τις τρεις εκδόσεις του SISR αλλά και με τον MCMC-ECM.

Συγκεκριμένα, οι εκδόσεις του (SISR) αφορούν:

- (i) Τον SISR που παρουσιάστηκε στο [41], όπου η επαναδειγματοληψία είναι πολυωνυμική (Gordon et al., 1993).
- (ii) Την τροποποίηση του τελευταίου, στον οποίο το βήμα της επαναδειγματοληψίας γίνεται με συνδυασμό των καταλοίπων και στρωματοποιημένης επαναδειγματοληψίας (stratified resampling).
- (iii) Τον SIS με μερικό έλεγχο απόρριψης (πχ, Liu et al., 1998, 2001)

## 7.4 Αριθμητικά Αποτελέσματα

### 7.4.1 Ο MCMC-Bayes σε πραγματικά δεδομένα

Προκειμένου να ελεγχθεί η αποδοτικότητα του αλγορίθμου MCMC- Bayes θα χρησιμοποιηθούν πραγματικά δεδομένα που αφορούν τις μετρήσεις των βιομαζών από 42 ελάσματα και μίσχους για τα οποία υποθέτουμε ότι έχουν περίοδο ανάπτυξης ίση με 10 κύκλους ανάπτυξης ( $T = 10$ ). Με την υπόθεση αυτή, όλες οι μετρήσεις αντιστοιχούν σε φύλλα που είχαν ολοκληρώσει την ανάπτυξη τους όταν κόπηκε το φυτό. Τα συγκεκριμένα δεδομένα παρουσιάστηκαν στο [41] στο οποίο προτάθηκε η εφαρμογή του KMM που θεωρήσαμε στην παρούσα εργασία, ως η βέλτιστη επιλογή μεταξύ άλλων μοντέλων, αυτά παρουσιάζονται στον Πίνακα 7.1.

ελάσμ.	0,021	0,069	0,084	0,138	0,246	0,414	0,604	0,85	0,892	0,99	1,398	1,627	1,568	1,774
μισχ	0,01	0,014	0,023	0,045	0,079	0,29	0,475	0,529	0,537	0,649	0,857	0,988	1,059	1,216
ελάσμ.	1,728	1,625	1,349	1,297	1,212	1,184	1,097	1,028	0,943	0,856	0,744	0,615	0,555	0,476
μισχ	1,317	1,263	1,154	1,204	1,134	1,106	1,056	0,964	0,904	0,889	0,797	0,687	0,655	0,532
ελάσμ.	0,422	0,361	0,326	0,277	0,238	0,191	0,179	0,15	0,124	0,117	0,079	0,089	0,106	0,095
μισχ	0,52	0,471	0,392	0,365	0,296	0,241	0,242	0,186	0,167	0,126	0,091	0,094	0,094	0,083

Πίνακας 7.1: Πραγματικά δεδομένα από το φυτό ζαχαρότευτλο που αντιστοιχούν σε μετρήσεις της βιομάζας από 42 ελάσματα (blades) και μίσχους(petioles)

Τα αποτελέσματα που θα παρουσιαστούν αφορούν 2 σετ επιλογών για τις εκ των προτέρων κατανομές. Συγκεκριμένα:

#### Bayes 1:

$$\mu^{-1} \sim U(50, 300), a_b \sim U(1.5, 4), a_p \sim U(1.5, 4)$$

$$p_p \sim U(0.6, 0.9), \Sigma \sim \mathcal{IW}_2(0.01I_2, 4). \quad (7.4.1)$$

#### Bayes 2:

$$\mu^{-1} \sim \mathcal{N}_+(100, 15^2)\mathbb{1}_{\{\mu^{-1}>0\}}, a_b \sim \mathcal{N}(3, 0.5^2)\mathbb{1}_{\{a_b>1\}}, a_p \sim \mathcal{N}(3, 0.5^2)\mathbb{1}_{\{a_p>1\}},$$

$$p_p \sim \mathcal{N}(0.85, 0.03^2)\mathbb{1}_{\{p_p>0\}}, \Sigma \sim \mathcal{IW}(0.01I_2, 4) \quad (7.4.2)$$

Οι υπόλοιπες παράμετροι, παρουσιάζονται στον πίνακα (7.2) και αντιστοιχούν στο μοντέλο για το οποίο έχει δείχθει πως είναι το καταλληλότερο για το συγκεκριμένο φυτό ανάμεσα στα μοντέλα που θεωρήθηκαν στο Trevezas & Cournéde (2013). Το πλήθος επαναλήψεων για τις παραπάνω επιλογές είναι 200.000 με περίοδο μπερν in 20.000 (10%). Επίσης, και για τις 2 αυτές επιλογές, οι αρχικοποιημένες διασπορές για τους αλγορίθμους MH (i)-(iv) του 2ου σταδίου είναι:

$$\begin{aligned} \sigma^2_{MH(\mu^{-1})} &= 3, & \sigma^2_{MH(a_b)} &= .06 \\ \sigma^2_{MH(a_p)} &= .08, & \sigma^2_{MH(p_p)} &= .035 \end{aligned}$$

## Αριθμητικά Αποτελέσματα

		Παράμετροι							
$\mu^{-1}$	εκτ	$\rho$	εκτ	$a_r$	3.1	$e_b$	.0083	$b_b$	2
$a_b$	εκτ	$\sigma_p$	εκτ	$p_b$	1	$k_B$	.7	$b_p$	2
$a_p$	εκτ	$\sigma_b$	εκτ	$p_r$	329.48	$s^{pr}$	500	$b_r$	2
$p_p$	εκτ	$\sigma_Q$	.1	$t_r$	60	$T$	10		

Πίνακας 7.2: Τιμές των παραμέτρων για το φυτό ζαχαρότευτλο. (με εκτ συμβολίζονται οι παράμετροι που εκτιμούνται στο μοντέλο)

Αρχικά, στον πίνακα (7.3) παρουσιάζονται ορισμένα περιγραφικά στατιστικά στοιχεία (μέσος, τυπική απόκλιση, διάμεσος, μέγιστη και ελάχιστη παρατήρηση και 95% διάστημα αξιοπιστίας credible interval (CI)) για ένα δείγμα από τους Bayes 1 και Bayes 2.

Bayes 1						
Παράμετρος	Μέσος	T. απόκλιση	Διάμεσος	min, max	Διάστημα Αξιοπιστίας	
$\mu^{-1}$	97.02	4.88	97.14	(80.844, 115.78)	( 87.266, 106.280)	
$a_b$	2.922	.273	2.89	(2.212, 3.881)	(2.458, 3.518)	
$a_p$	1.921	.239	1.89	(1.500, 3.013)	(1.546, 2.474)	
$p_p$	.8146	.015	.08147	(.745, .883)	(.785, .844)	
$\sigma_b$	.074	.012	.083	(.039, .155)	(.054, .101)	
$\sigma_p$	.062	.009	.060	(.034, .141)	(.047, .082)	
$\rho$	.081	.018	.084	(-.686, .753)	(-.287, .437)	

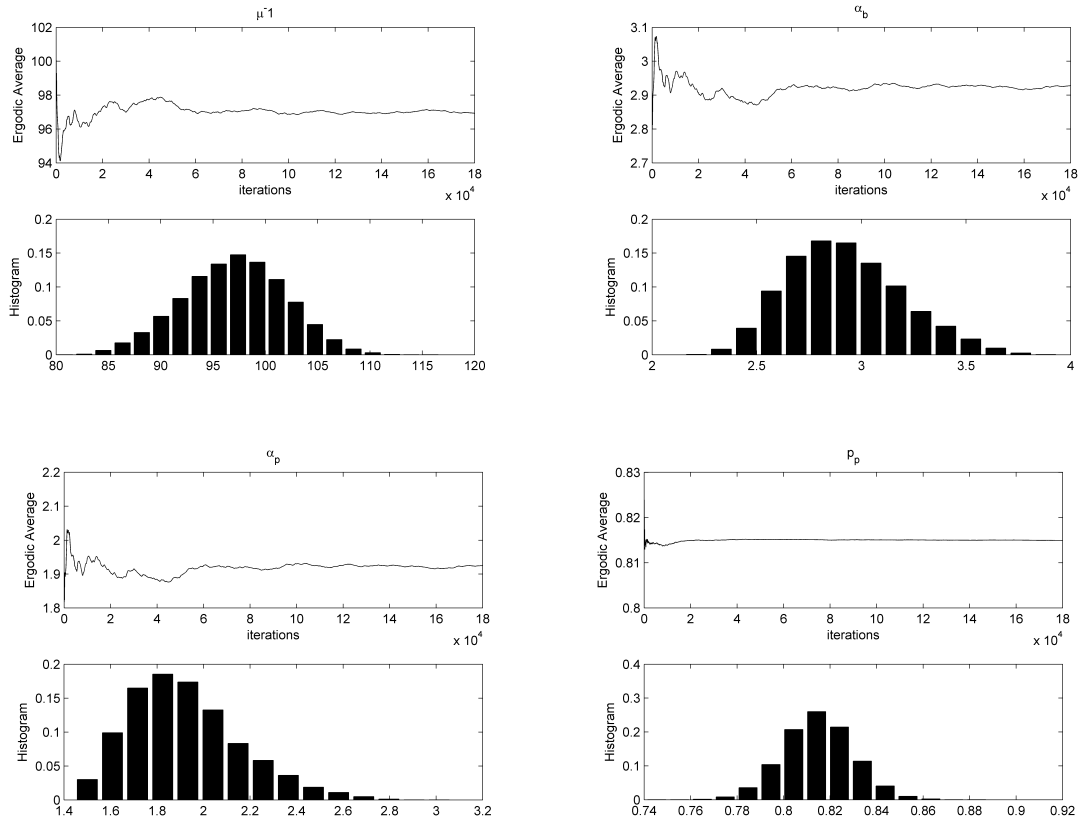
Bayes 2						
Παράμετρος	Μέσος	T. απόκλιση	Διάμεσος	min, max	Διάστημα Αξιοπιστίας	
$\mu^{-1}$	94.506	3.999	94.362	(79.7, 111.6)	(86.877, 102.711)	
$a_b$	3.096	.231	3.094	(2.33, 3.97)	(2.639, 3.566)	
$a_p$	2.095	.222	2.081	(1.42, 3.17)	(1.695, 2.570)	
$p_p$	0.823	.014	.0822	(.76, .89)	(.795, .851)	
$\sigma_b$	.0747	.012	.074	(.039, .16)	(.054, .103)	
$\sigma_p$	.0635	.010	.062	(.036, .13)	(.047, .085)	
$\rho$	.0230	.019	.030	(-.67, .77)	(-.339, .402)	

Πίνακας 7.3: Περιγραφικά στατιστικά για ένα δείγμα από τους Bayes 1 και Bayes 2



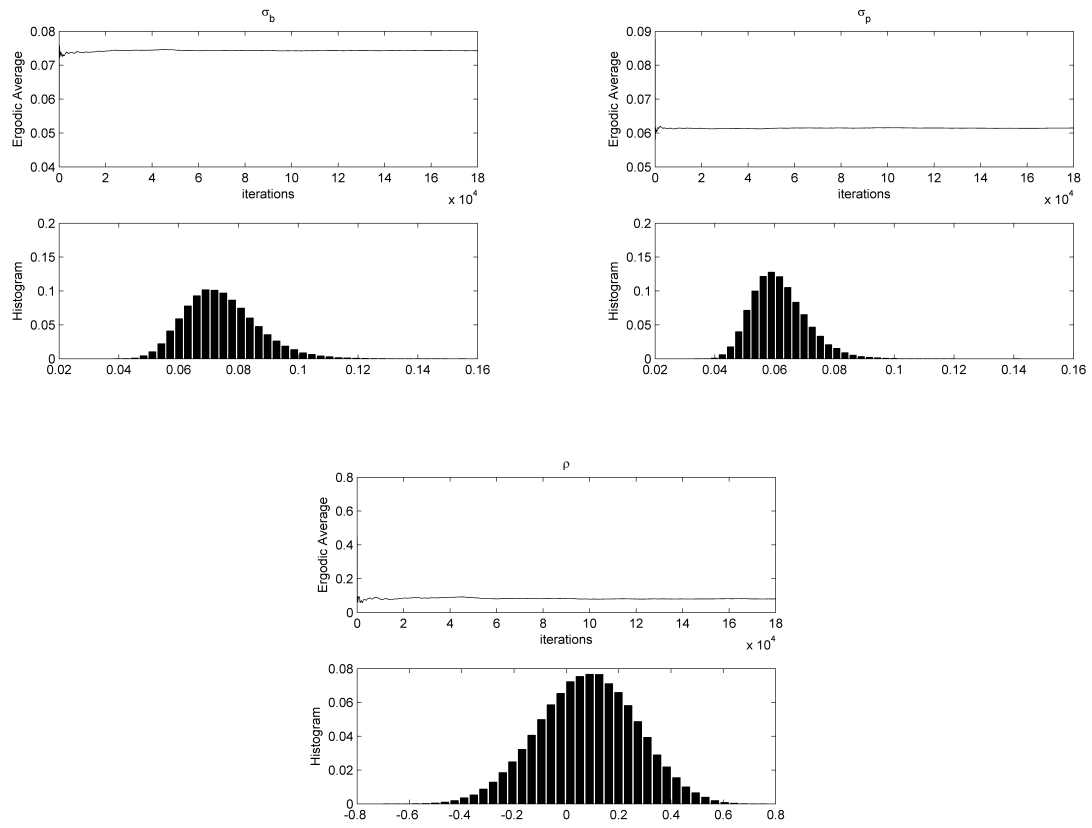
## Αριθμητικά Αποτελέσματα

Στα σχήματα (7.1) και (7.2), έχει υπολογιστεί το ιστόγραμμα για τις παραμέτρους του μοντέλου Bayes 1. Τα γραφήματα αντιστοιχούν επομένως στις περιθώριες συναρτήσεις της εκ των υστέρων κατανομής. Επίσης, στα υπογραφήματα παρουσιάζεται η σύγκλιση του εργοδικού μέσου που αντιστοιχεί στην κάθε παράμετρο.



Σχήμα 7.1: Εργοδικός μέσος και ιστόγραμμα για τις λειτουργικές παραμέτρους του μοντέλου Bayes 1

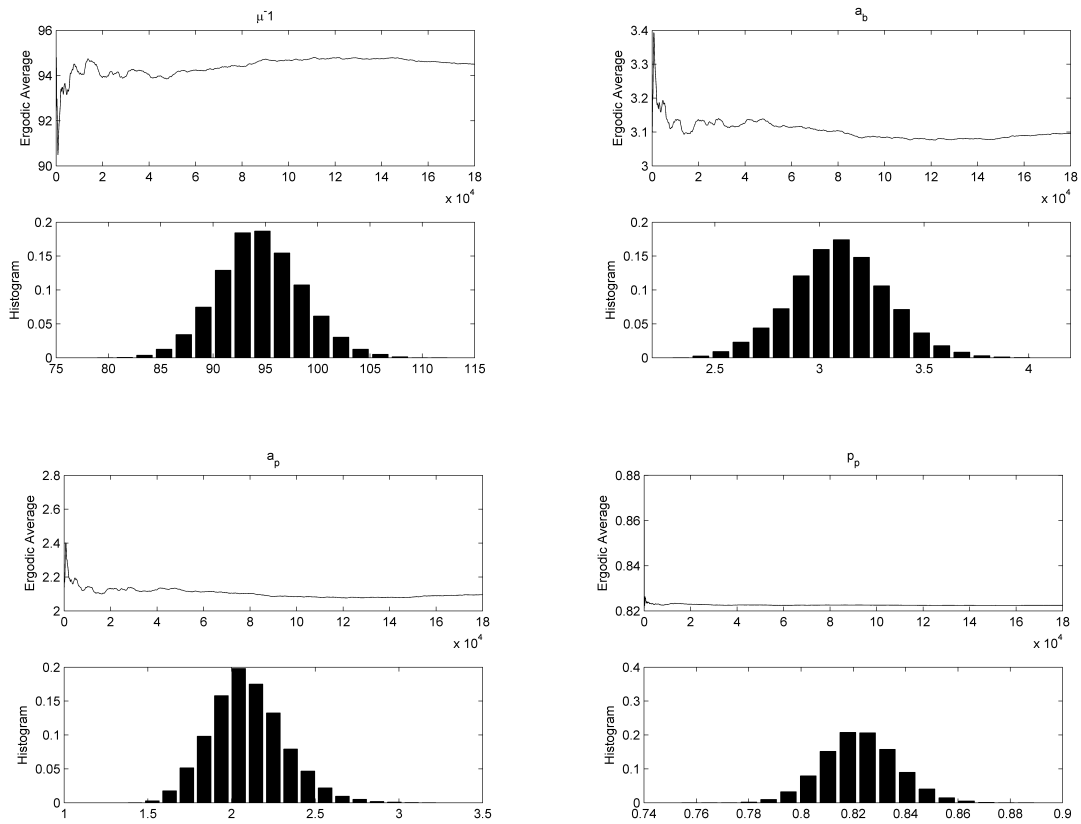
## Αριθμητικά Αποτελέσματα



Σχήμα 7.2: Εργοδικός μέσος και ιστόγραμμα για τις παραμέτρους του πίνακα συνδιακύμανσης Bayes 1

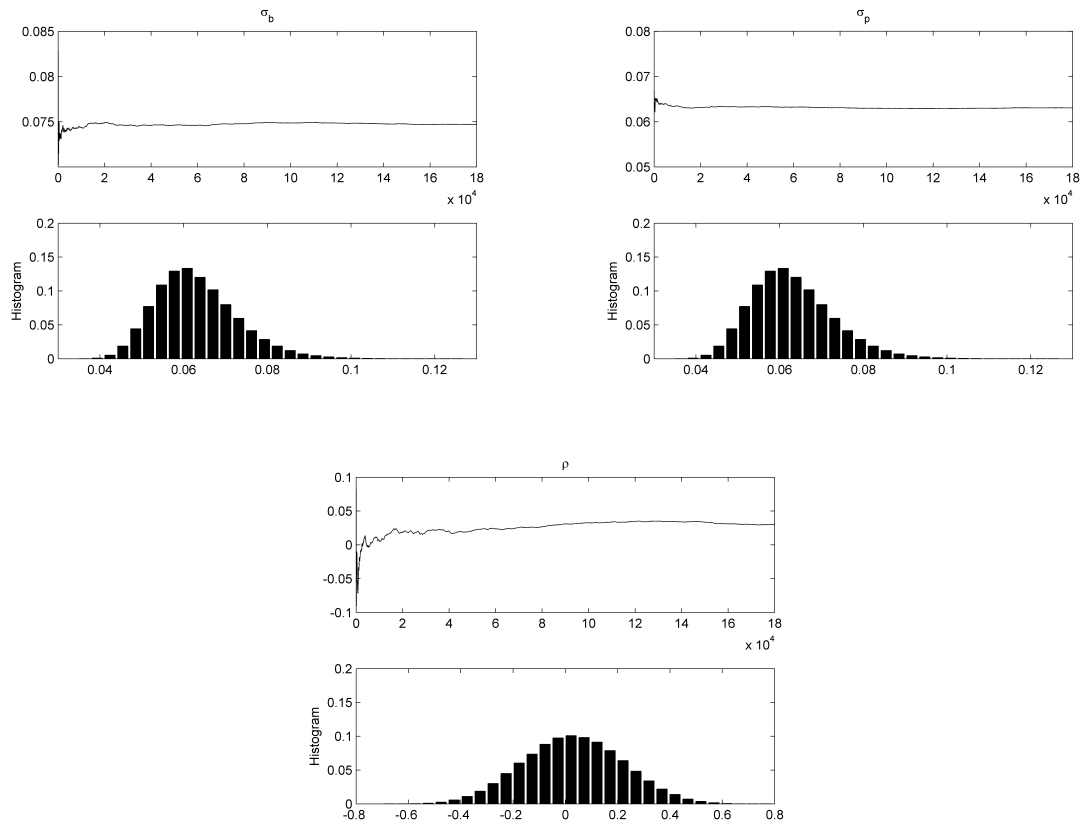
## Αριθμητικά Αποτελέσματα

Τα αντίστοιχα ιστογράμματα για την περίπτωση του Bayes 2 παρουσιάζονται στα σχήματα (7.3) και (7.4) που ακολουθούν:



Σχήμα 7.3: Εργοδικός μέσος και ιστογράμματα για τις λειτουργικές παραμέτρους του μοντέλου Bayes 2

## Αριθμητικά Αποτελέσματα



Σχήμα 7.4: Εργοδικός μέσος και ιστόγραμμα για τις παραμέτρους του πίνακα συνδιακύμανσης Bayes 2

Στους πίνακες (7.4) και (7.5) παρουσιάζονται οι μέσες τιμές και οι τυπικές αποκλίσεις από 50 ανεξάρτητα δείγματα από τους MCMC-Bayes 1-2 αλλά και από τις άλλες αλγοριθμικές τεχνικές που όπως αναφέρθηκε, έχουν εφαρμοστεί για το συγκεκριμένο μοντέλο. Με τον τρόπο αυτό θα έχουμε κάποια ένδειξη για το τυπικό σφάλμα των εκτιμητριών.

Παράμετρος	Μέσος					
	SISmR	SISrsR	SISprc	MCMC-ECM	Bayes1	Bayes2
$\mu^{-1}$	98.2	98.1	98.2	97.9	96.9	94.2
$a_b$	2.83	2.83	2.82	2.83	2.93	3.11
$a_p$	1.8	1.82	1.81	1.83	1.93	2.1
$p_p$	.815	.8147	.8147	.8147	.8147	.8224
$\sigma_p$	.075	.076	.076	.076	.075	.074
$\sigma_p$	.058	.058	.058	.058	.062	.063
$\rho$	.13	.12	.12	.12	.08	.02

Πίνακας 7.4: Μέσες τιμές για τις εκτιμήσεις έχουν υπολογιστεί από 50 ανεξάρτητες επαναλήψεις και για τους έξι αλγορίθμους.

## Αριθμητικά Αποτελέσματα

Παράμετρος	Τυπική απόκλιση					
	SISmR	SISrsR	SISprc	MCMC-ECM	Bayes1	Bayes2
$\mu^{-1}$	.545	.42	.50	.23	.23	.21
$a_b$	.029	.023	.027	.013	.014	.014
$a_p$	.021	.017	.019	.009	.011	.013
$p_p$	.0001	.0001	.0001	.00002	.0001	.0001
$\sigma_b$	.0002	.0002	.0002	.0001	.0001	.0001
$\sigma_p$	.0001	.0001	.0001	.00005	.00008	.0001
$\rho$	.0030	.0028	.0026	.0012	.003	.004

Πίνακας 7.5: Τυπικές αποκλίσεις για τις εκτιμήσεις έχουν υπολογιστεί από 50 ανεξάρτητες επαναλήψεις και για τους έξι αλγορίθμους.

7.4.2 Ο MCMC-Bayes σε προσομοιωμένα δεδομένα

Προκειμένου να ελεγχθεί η αποδοτικότητα του αλγορίθμου MCMC-Bayes πραγματοποιήθηκαν όμοια τεστ με αυτά των οποίων τα αποτελέσματα παρουσιάστηκαν στην προηγούμενη ενότητα, χρησιμοποιώντας προσομοιωμένα δεδομένα. Συγκεκριμένα, σε κάθε επανάληψη του αλγορίθμου, προσομοιώνονται 42 δεδομένα, με βάση τις τιμές του παρακάτω παραμετρικού διανύσματος, και τα οποία αντικαθιστούν τα πραγματικά δεδομένα.

$$\theta_{simul} = (\mu^{-1}, a_b, a_p, p_p, \sigma_b, \sigma_p, \rho) = (100, 3, 3, 0.81654, 0.05, 0.05, 0.8). \quad (7.4.3)$$

Πραγματοποιήθηκαν 300 ανεξάρτητες επαναλήψεις με κάθε έναν από τους Bayes 1 και Bayes 2 όπως αυτοί παρουσιάζονται στις (7.4.1) και (7.4.2) και στους πίνακες (7.6) και (7.7) έχει υπολογιστεί η μεροληψία (BIAS) και το μέσο τετραγωνικό σφάλμα (MSE).

	$\mu^{-1}$		$a_b$	
	Bayes 1	Bayes 2	Bayes 1	Bayes 2
BIAS	.832	.329	-.048	-.022
MSE	14,106	5.178	.0381	.009

	$a_p$		$p_p$	
	Bayes 1	Bayes 2	Bayes 1	Bayes 2
BIAS	.0718	-.071	-.001	.002
MSE	.0704	.009	1,90E-05	1,88E-05

Πίνακας 7.6: Μεροληψία και μ.τ.σ των λειτουργικών παραμέτρων του μοντέλου βασισμένες σε 300 προσομοιωμένα δείγματα δεδομένων.

	$\sigma_b$		$\sigma_p$		$\rho$	
	Bayes 1	Bayes 2	Bayes 1	Bayes 2	Bayes 1	Bayes 2
BIAS	.006	.05	.004	.047	-.101	-.05
MSE	.001	.001	.001	.001	.06	.007

Πίνακας 7.7: Μεροληψία και ΜΤΣ των παραμέτρων που αντιστοιχούν στον πίνακα συνδιακύμανσης του μοντέλου βασισμένες σε 300 προσομοιωμένα δείγματα δεδομένων.

## Αριθμητικά Αποτελέσματα

Τέλος, στον πίνακα (7.8) έχουν υπολογιστεί ορισμένα περιγραφικά στατιστικά στοιχεία για τους Bayes 1 και Bayes 2.

Bayes 1						
Παράμετρος	Μέσος	T. απόκλιση	Διάμεσος	min, max	Διάστημα Αξιοπιστίας	
$\mu^{-1}$	100.689	2.907	98.39	(85.844, 113.11)	(91.627, 110.561)	
$a_b$	3.023	.140	3.127	(2.712, 3.692)	(2.778, 3.231)	
$a_p$	2.959	.159	3.089	(2.610, 3.713)	(2.798, 3.491)	
$p_p$	.819	.004	.0819	(.798, .843)	(.808, .824)	
$\sigma_b$	.050	.006	.051	(.031, .081)	(.039, .059)	
$\sigma_p$	.059	.007	.058	(.039, .091)	(.048, .065)	
$\rho$	.721	.069	.774	(.382, .943)	(.607, .876)	

Bayes 2						
Παράμετρος	Μέσος	T. απόκλιση	Διάμεσος	min, max	Διάστημα Αξιοπιστίας	
$\mu^{-1}$	102.710	4.193	102.670	(87.346, 118.491)	(94.606, 110.965)	
$a_b$	2.930	.212	2.934	(2.132, 3.617)	(2.513, 3.331)	
$a_p$	2.888	.218	2.887	(2.096, 3.731)	(2.473, 3.313)	
$p_p$	.8166	.0052	.8165	(.794, .841)	(.806, .821)	
$\sigma_b$	.047	.005	.047	(.029, .084)	(.038, .060)	
$\sigma_p$	.055	.006	.054	(.034, .100)	(.044, .069)	
$\rho$	.708	.0795	.717	(.241, .938)	(.528, 0.839)	

Πίνακας 7.8: Περιγραφικά στατιστικά για ένα δείγμα από τους Bayes 1 και Bayes 2

### 7.4.3 Συμπεράσματα και μελλοντικοί στόχοι

Ο αλγόριθμος MCMC-Bayes που κατασκευάστηκε (ενότητες 7.1-7.2) και του οποίου τα αποτελέσματα παρουσιάστηκαν (ενότητα 7.4) αποτελεί την πρώτη αμιγώς Μπεϋζιανή προσέγγιση για την εκτίμηση των παραμέτρων της συγκεκριμένης κλάσης μοντέλων ανάπτυξης φυτών. Η κλάση αυτή χαρακτηρίζεται από ντετερμινιστική διαδικασία οργανογένεσης και στοχαστική διαδικασία παραγωγής και διανομής της βιομάζας. Τα πρώτα αποτελέσματα είναι αρκετά ελπιδοφόρα, καθώς, σε μία πρώτη αντιπαραβολή με άλλες αλγοριθμικές τεχνικές που έχουν αποδειχθεί σημαντικά αποδοτικές για το συγκεκριμένο μοντέλο [41], [42], δεν διαπιστώθηκε ότι κάποια εκτιμήτρια μπορεί να ξεχωρίσει ως βέλτιστη. Η Μπεϋζιανή προσέγγιση στην οποία στηρίχθηκε ο συγκεκριμένος αλγόριθμος, μας προσφέρει ιδιαίτερη ευελιξία στη διαχείριση πολύπλοκων δομών, όπως αυτές που διέπουν την ανάπτυξη των φυτών, το οποίο είναι αναμφίβολα ένα σημαντικό πλεονέκτημα. Ακόμη, η μεθοδολογική πλευρά στην οποία στηρίζεται έχει σημαντικά πλεονεκτήματα, καθώς μπορεί με σχετική ευκολία να τροποποιηθεί ώστε να συμπεριλάβει γενικότερα μοντέλα. Για παράδειγμα, μοντέλα ανάπτυξης φυτών με στοχαστική διαδικασία οργανογένεσης (π.χ. δέντρα), δηλαδή μοντέλα στα οποία ο συνολικός αριθμός των οργάνων κάθε κλάσης σε κάθε κύκλο ανάπτυξης, είναι τυχαίος.

# Κεφάλαιο 8

## Παράρτημα

### 8.0.1 A.1 Πιθανοθεωρητικά

#### 1. Στοχαστικές Διαδικασίες

Στην παρούσα διπλωματική ασχοληθήκαμε με ακολουθίες Markov και γενικότερα είδαμε πως συνδέονται με τις στάσιμες ακολουθίες. Αυτά είναι 2 είδη στοχαστικών διαδικασιών. Άλλα είδη είναι για παράδειγμα οι ανανεωτικές ακολουθίες, οι ανταλλάξιμες ακολουθίες, τα Martingales κ.α. Αυτό που έχει ιδιαίτερο μαθηματικό ενδιαφέρον είναι το πως αυτές οι σ.δ. συνδέονται μεταξύ τους. Για παράδειγμα, κάθε ακολουθία από α.ι.τ.μ είναι Markov, ανταλλάξιμη και στάσιμη, ενώ αν επιπλέον οι μοναδικές τιμές που παίρνουν είναι 0 και 1, τότε είναι και ανανεωτικές ακολουθίες. Οι τυχαίοι περίπατοι είναι πάντα Markov και όταν τα βήματα έχουν μέσο 0, είναι και Martingales.

Το σπουδαιότερο ωστόσο δεν είναι οι μαθηματικές ιδιότητες αλλά τα πραγματικά προβλήματα που μπορούν να μοντελοποιηθούν και στη συνέχεια να μελετηθούν με τη βοήθεια των στοχαστικών διαδικασιών. Τι ακριβώς σημαίνει όμως αυτό. Σε γενικές γραμμές, οι στοχαστικές διαδικασίες είναι ένας κλάδος της θεωρίας πιθανοτήτων στο οποίο μελετώνται πιθανοθεωρητικά μοντέλα που εξελίσσονται στο χρόνο. Αυτό μπορούμε να το δούμε καλύτερα και ως τον κλάδο των μαθηματικών που ξεκινά με τα αξιώματα των πιθανοτήτων και περιέχει πλούσια και ενδιαφέροντα μαθηματικά αποτελέσματα που έπονται από αυτά τα αξιώματα. Για να εφαρμόσουμε τα αποτελέσματα αυτά σε πραγματικά σενάρια, απαιτείται αρχικά η κατασκευή ενός πιθανοθεωρητικού μοντέλου για το συγκεκριμένο πρόβλημα. Στη συνέχεια, μέσω του μοντέλου διατυπώνονται με ακρίβεια τα μαθηματικά αποτελέσματα και στη συνέχεια μεταφράζονται στον 'πραγματικό κόσμο'. Το μοντέλο έχει κατασκευαστεί ώστε να προσεγγίζει το πραγματικό πρόβλημα. Έτσι, τα ακριβή αποτελέσματα που παίρνουμε από αυτό, θα προσεγγίζουν τα πραγματικά. Ο τζόγος, τα ασφαλιστικά ζητήματα, οι οικονομικές προβλέψεις, τα προβλήματα παραγωγής και ζήτησης, τα συστήματα εξυπηρέτησης, τα δίκτυα επικοινωνίας, και άλλα πολλά, αποτελούν πραγματικά σενάρια τα οποία μπορούν μελετηθούν με την εφαρμογή των στοχαστικών διαδικασιών. Στη παρούσα εργασία είδαμε έναν τρόπο με τον οποίο μπορεί να μοντελοποιηθεί η εξέλιξη ενός ζωντανού οργανισμού (φυτά) με τη χρήση στοχαστικών διαδικασιών ενώ η μελέτη δεν βασίστηκε μόνο σε πιθανοθεωρητικά αποτελέσματα αλλά και σε εργαλεία της στατιστικής, τονίζοντας για άλλη μία φορά το πόσο στενά συνδεδεμένοι είναι μεταξύ τους οι δύο αυτοί κλάδοι.

Στο μέρος αυτό θα αναφέρουμε συνοπτικά, ορισμένα βασικά στοιχεία θεωρίας των λεγόμενων στοχαστικών διαδικασιών (σ.δ.).

**Ορισμός** (Στοχαστική διαδικασία). Έστω  $(\Omega, \mathcal{F}, \mathbb{P})$  χώρος πιθανότητας και  $(\mathcal{X}, \mathcal{X})$  μετρήσιμος χώρος. Μία στοχαστική διαδικασία (σ.δ.) είναι μία οικογένεια τυχαίων



μεταβλητών  $\{X_t : t \in T\}$  ορισμένη στον παραπάνω χώρο πιθανότητας με τιμές σε κάποιον γενικό χώρο  $(\mathcal{X}, \mathcal{X})$ .

1. Στην ουσία, μία σ.δ. είναι μία συνάρτηση:  $X : T \times \Omega \rightarrow \mathcal{X}$ , όπου για δοσμένα  $\omega, t$ , το  $Q(\omega, t)$  αναφέρεται ως η κατάσταση στην οποία βρίσκεται η σ.δ.
2. Για κάθε  $t$ , η συνάρτηση  $X_t = X(\cdot, t) : \Omega \rightarrow \mathcal{X}$  είναι τυχαία μεταβλητή ορισμένη στον  $(\Omega, \mathcal{F}, \mathbb{P})$  ενώ, για κάθε  $\omega$ , η συνάρτηση  $X(\omega, \cdot) : T \rightarrow \mathcal{X}$  περιγράφει μία συγκεκριμένη εξέλιξη της σ.δ. μέσα στο χρόνο και ονομάζεται πραγματοποίηση της διαδικασίας και μπορεί σε ορισμένες περιπτώσεις να παρασταθεί γραφικά.
3. Ο χώρος  $\mathcal{X}$  αναφέρεται ως χώρος καταστάσεων (χ.κ.) ενώ με τον όρο: γενικός χώρος καταστάσεων, αναφερόμαστε σε υπεραριθμήσιμο μετρικό χώρο ή μετρήσιμο χώρο (πιο γενικές περιπτώσεις). Στην περίπτωση αριθμήσιμου χ.κ. η  $X$  είναι αυτόματα μετρήσιμη ως προς την  $\mathcal{P}(\mathcal{X})$ .
4. Μία διήθηση (Filtration) στον  $(\Omega, \mathcal{F})$  είναι μία αύξουσα ακολουθία  $(\mathcal{F}_t)_{t \in T}$  από υπο-σ-άλγεβρες της  $\mathcal{F}$ .
5. Ένας χώρος με φίλτρο συμβολίζεται με  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T})$  ενώ ο  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, \mathbb{P})$  λέγεται χώρος πιθανότητας με φίλτρο.
6. Διακρίνοντας το χρόνο σε συνεχή ή διακριτό, έχουμε τις στοχαστικές διαδικασίες συνεχούς και διακριτού χρόνου.

Σε όσα ακολουθούν υποθέτουμε πως  $T = \mathbb{N}_0$ . Έτσι, μία στοχαστική διαδικασία είναι πρακτικά μία ακολουθία τ.μ.  $(X_n)_{n \geq 0}$  ορισμένη στον ίδιο χώρο πιθανότητας.

Συμβολίζουμε με  $\mathcal{F}^0$  τη φυσική διήθηση, δηλαδή  $\mathcal{F}^0 = \sigma(X_0 X_1 \dots X_n) = \sigma(\{X_k\} : 0 \leq k \leq n) = \sigma(\cup_{k=0}^n \sigma(X_k))$ , η σ άλγεβρα που κάνει όλες τις  $X_k$  μέχρι τη στιγμή  $n$ , μετρήσιμες.

Μία συνάρτηση  $\Phi$  στον  $\Omega$  είναι  $\mathcal{F}^0$ -μετρήσιμη, αν  $\Phi = \phi(X_0 X_1 \dots X_n)$  για μία  $\mathcal{X}^{\otimes(n+1)}$ -μετρήσιμη συνάρτηση  $\phi$ . Συγκεκριμένα, ένα ενδεχόμενο βρίσκεται στην  $\mathcal{F}^0$  αν γνωρίζουμε ότι έχει πραγματοποιηθεί ή όχι, μόλις οι τιμές των  $(X_0, X_1, \dots, X_n)$  γίνουν γνωστές. Ωστόσο, σε αρκετές περιπτώσεις, έχουμε και περισσότερες πληροφορίες μέχρι τη χρονική στιγμή  $n$  εκτός των τιμών των  $(X_0, X_1, \dots, X_n)$ , όπως πχ οι τιμές άλλων διαδικασιών.

Μία προσαρμοσμένη σ.δ. στον  $(\mathcal{X}, \mathcal{X})$  συμβολίζεται με  $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbb{P})$  και για κάθε  $n$ , η  $X_n$  είναι τ.μ. με τιμές στον  $(\mathcal{X}, \mathcal{X})$  και  $(\mathcal{F}_n)$ -μετρήσιμη. (ειδικά η τελευταία ιδιότητα εκφράζεται λέγοντας ότι η  $(X_n)_{n \geq 0}$  είναι προσαρμοσμένη στη διήθηση  $(\mathcal{F}_n)_{n \geq 0}$ ).

### Πεπερασμένης-διάστασης κατανομές.

Καταρχάς, οι κατανομές αυτές είναι μέτρα στον  $(\Omega, \mathcal{F})$ . Επειδή γενικά μία στοχαστική διαδικασία είναι ένα απειροδιάστατο αντικείμενο, χρειαζόμαστε ένα μέτρο το

οποίο να παρέχει ουσιαστικές πληροφορίες για την διαδικασία. Οι πεπερασμένης διάστασης κατανομές είναι ένα μέτρο που κάνει ακριβώς αυτό. Φυσικά, υπάρχει σημαντική διαφορά όταν αυτές ορίζονται για σ.δ. συνεχούς χρόνου. Για σ.δ. διακριτού χρόνου ωστόσο, αποδεικνύεται πως μας παρέχουν όλη την απαραίτητη πληροφορία για τη διαδικασία. Αυτό είναι το αποτέλεσμα από το Θεώρημα επέκτασης του Kolmogorov. Με άλλα λόγια, το Θεώρημα αυτό (κάτω από ορισμένες συνθήκες 'συνέπειας') μας επιτρέπει να ορίσουμε μία διαδικασία  $(X_n)_{n \geq 0}$  (που είναι απειροδιάστατο αντικείμενο) χρησιμοποιώντας μόνο 'πεπερασμένα' αντικείμενα όπως οι κατανομές αυτές. Επομένως, η εύρεση της κατανομής πεπερασμένης διάστασης για μία σ.δ., αποτελεί ένα κεντρικό ζήτημα.

Έστω  $\mathbf{X}$  μία διαδικασία με τιμές στον  $(X, \mathcal{X})$ . Συμβολίζουμε με  $\mu_n$  την κατανομή του τυχαίου διανύσματος  $(X_0, X_1, \dots, X_n)$  που παίρνει τιμές στον  $(\mathcal{X}^{\otimes(n+1)}, \mathcal{X}^{\otimes(n+1)})$ . Τα μέτρα πιθανότητας  $(\mu_n)_{n \geq 0}$  λέγονται πεπερασμένης διάστασης κατανομές της διαδικασίας  $\mathbf{X}$ .

Στην περίπτωση διακριτού χρόνου, οι κατανομές αυτές ικανοποιούν απευθείας την παρακάτω πολύ βασική ιδιότητα:

$$\mu_{n-1}(A_0 \times A_1 \times \dots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \dots \times \mathcal{X}). \quad (8.0.1)$$

Αυτό είναι άμεσο διότι:

$$\mu_n(A_0 \times A_1 \times \dots \times A_n) = \mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n)$$

Για το ζήτημα που αναφέρθηκε, του ορισμού μιας σ.δ. από τις κατανομές αυτές, θα χρειαστούμε τον κανονικό χώρο μιας σ.δ. Έτσι θα δώσουμε μία καταφατική απάντηση στο ερώτημα αν υπάρχει σ.δ. διακριτού χρόνου.

**Ορισμός** (Κανονικός χώρος)

$$\Omega = \mathcal{X}^{\mathbb{N}}, \omega = (\omega)_{n \geq 0}, X_n(\omega) = \omega_n, \quad (8.0.2)$$

$$\mathcal{F}_n = \sigma(X_k, k \leq n), \mathcal{F} = \sigma(X_k, k \geq 0), \quad (8.0.3)$$

Για τη κατασκευή λοιπόν μιας σ.δ., ορίζουμε αρχικά ένα μέτρο πιθανότητας  $\mathbb{P}_n$  στον  $(\Omega, \mathcal{F}_n)$ , θέτοντας, για  $A \in \mathcal{X}^{\otimes(n+1)}$ ,

$$\mathbb{P}_n(A \times X \dots \times X \times \dots) = \mu_n(A)$$

και στη συνέχεια ορίζουμε μία συνολοσυνάρτηση στον  $\cup_{n \geq 0} \mathcal{F}_n$ :

$$\mathbb{P}(A) = \mathbb{P}_n(A) \text{ αν } A \in \mathcal{F}_n.$$

Αρκεί να επεκτείνουμε τη  $\mathbb{P}$  στο  $\mathcal{F} = \sigma(\cup_{n=0}^{\infty} \mathcal{F}_n)$ . Την απάντηση δίνει τελικά το παρακάτω Θεώρημα:

**Θεώρημα** (Kolmogorov) Έστω  $(\mu_n)_{n \geq 0}$  οικογένεια μέτρων πιθανότητας όπου κάθε  $\mu_n$  είναι μέτρο πιθανότητας στον χώρο γινόμενο  $(X^{\otimes(n+1)}, \mathcal{X}^{\otimes(n+1)})$  και ικανοποιεί

την 8.0.1. Τότε στον κανονικό χώρο που ορίζεται στην ;; υπάρχει μοναδικό μέτρο πιθανότητας  $\mathbb{P}$  τέτοιο ώστε η στοχαστική διαδικασία  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbb{P})$  έχει την  $(\mu_n)_{n \geq 0}$  ως κατανομή πεπερασμένης διάστασης.

Επιπλέον, το μέτρο αυτό είναι και μοναδικό. Αυτό έπεται από το Θεώρημα π-λ καθώς η κλάση  $\cup_{n \geq 0} \mathcal{F}_n$  είναι κλειστή στις πεπερασμένες τομές.

Διαισθητικά, η ιδέα πίσω από τον κανονικό χώρο και τον ορισμό του  $\mathbb{P}_\mu$ , είναι η εξής: Ένα τρόπος να κατασκευαστούν διαφορετικές στοχαστικές διαδικασίες είναι να ξεκινήσουμε από τον  $(\Omega, \mathcal{F})$  και να ορίσουμε σε αυτόν διάφορες συναρτήσεις,  $X, Y, Z$  κ.τ.λ., με τιμές προς τον  $(\mathcal{X}, \mathcal{X})$  για παράδειγμα, στον οποίο θα ορίζονται και τα αντίστοιχα επαγόμενα μέτρα τους. Ένα εναλλακτικός τρόπος είναι να ξεκινήσουμε από τον  $(\Omega, \mathcal{F})$  και να ορίσουμε ακριβώς μία συνάρτηση  $X : \Omega \mapsto \mathcal{X}$ , την ταυτοτική,  $(X(\omega) = \omega)$ . Μπορούμε στη συνέχεια να της δώσουμε πολλές διαφορετικές ιδιότητες ορίζοντας διαφορετικά μέτρα στον  $(\Omega, \mathcal{F})$ , πχ τα  $\mathbb{P}_1, \mathbb{P}_x, \mathbb{P}_\mu$ .

## 8.0.2 A.2 Πλαίσιο στατιστικής

### 1. Βασικές έννοιες στατιστικής.

Στα προβλήματα στατιστικής γενικά, ασχολούμαστε κυρίως με τη συλλογή των δεδομένων, αλλά και με την ανάλυση και την ερμηνεία τους. Με το πρώτο ασχολείται ο κλάδος δειγματοληψία και περιγραφική στατιστική και δεν θα μπούμε σε περισσότερη ανάλυση. Θα θεωρούμε πάντα πως έχουμε δεδομένο δείγμα και ασχολούμαστε με τις πληροφορίες μπορούμε να αποκομίσουμε από αυτό. Η απάντηση δεν εξαρτάται μόνο από τα δεδομένα (δηλαδή από το τι έχει παρατηρηθεί), αλλά και από τη συνολική γνώση που έχουμε για την κατάσταση. Το τελευταίο διατυπώνεται με τις υποθέσεις τις οποίες κάνουμε για τη συγκεκριμένη κατάσταση. Υπάρχουν δύο κύριες προσεγγίσεις: Η κλασική και η Μπεϋζιανή:

1. Σχετικά με την πρώτη, θεωρούμε τυπικά μία παραμετρική οικογένεια μέτρων,  $(\mathbb{P}_\phi)$  όπου η παράμετρος  $\phi$  παίρνει τιμές σε κάποιος χώρο (συνήθως Ευκλείδειο)  $\Phi$ . Ο πληθυσμός του οποίου τα χαρακτηριστικά μελετάμε υποθέτουμε πως αντιπροσωπεύεται από το μοντέλο  $\mathbb{P}$ . Το ζητούμενο είναι η εύρεση ενός κατάλληλου  $\phi^*$  τέτοιο ώστε  $\mathbb{P}_{\phi^*} = \mathbb{P}$ , δηλαδή ο πληθυσμός έχει πλήρως καθοριστεί. Αυτό γίνεται παρατηρώντας μία ακολουθία δεδομένων  $y_1, \dots, y_n$  η οποία υποθέτουμε ότι αποτελεί πραγματοποίηση του τ. διανύσματος  $Y_1, \dots, Y_n$ . Κάθε συνιστώσα του διανύσματος αυτού, αποτελεί μία ανεξάρτητη τ.μ. που έχει κατανομή  $\mathbb{P}$ . Επομένως, όταν το  $\phi^*$  καθοριστεί, θα είμαστε σε θέση να αναπαράγουμε την ακολουθία  $y_1, \dots, y_n$ . Ο στόχος λοιπόν της ανάλυσης είναι είτε να αναζητήσουμε τρόπους εύρεσης της τιμής του  $\phi^*$  (σημειακή εκτίμηση), είτε να αναζητήσουμε κάποιο σύνολο για το οποίο να μπορούμε να ισχυριστούμε πως περιέχει ή δεν περιέχει το  $\phi^*$  (διαστήματα αξιοπιστίας, έλεγχοι υποθέσεων). Συχνά, το πρόβλημα εύρεσης της τιμής του  $\phi^*$ , προσεγγίζεται συνήθως με την μέθοδο μεγιστοποίησης της συνάρτησης πιθανοφάνειας.
2. Σχετικά με τη Μπεϋζιανή προσέγγιση, έχουμε την επιπλέον υπόθεση ότι το ίδιο το  $\phi$  είναι μία άγνωστη (τυχαία) ποσότητα, η οποία αντιμετωπίζεται ως τ.μ. (μη

παρατηρήσιμη) με γνωστή κατανομή. Αυτή η εκ των προτέρων κατανομή (που καθορίζεται σύμφωνα με το πρόβλημα) τροποποιείται υπό το φώς των δεδομένων και καθορίζει την εκ των υστέρων κατανομή (τη δεσμευμένη κατανομή του *phi* δοθέντος των δεδομένων), η οποία εκφράζει το τι μπορεί να ειπωθεί για το  $\phi$  υπό το πρίσμα των υποθέσεων που έγιναν για τα δεδομένα.

## 2. Ορολογία:

1. Με τον όρο πληθυσμό εννοούμε τον χώρο πιθανότητας  $(\Omega, \mathcal{F}, \mathbb{P})$ . Ωστόσο, επειδή μας ενδιαφέρουν τα χαρακτηριστικά του πληθυσμού και αυτά περιγράφονται από το  $\mathbb{P}$ , με τον όρο πληθυσμό καλούμε συχνά το μέτρο  $\mathbb{P}$ .
2. Ένα τυχαίο δείγμα μεγέθους  $n$  είναι ένα τυχαίο διάνυσμα (διάστασης  $n$ ), ορισμένο στο παραπάνω χώρο με τιμές στον  $\mathbb{R}^n$ ,  $\mathcal{B}(\mathbb{R}^n), \lambda$ ). Το σετ δεδομένων (ή απλά, τα δεδομένα) αποτελούν μία πραγματοποίηση του παραπάνω διανύσματος. Συχνά, για λόγους απλότητας, και το τυχαίο διάνυσμα και τα δεδομένα καλούνται δείγμα από το  $\mathbb{P}$  ενώ το μέγεθος θεωρείται πως είναι  $n$ .
3. Λέμε πως ένα πληθυσμός είναι γνωστός αν το  $\mathbb{P}(A)$  έχει γνωστή τιμή για οποιοδήποτε σύνολο  $A \in \mathcal{F}$ .
4. Ως στατιστικό μοντέλο καλείται ένα σύνολο υποθέσεων για τον πληθυσμό  $\mathbb{P}$
5. Ένα σύνολο μέτρων πιθανότητας  $\mathbb{P}_\theta$  στον  $(\Omega, \mathcal{F})$  με δείκτες  $\theta \in \Theta$  (παράμετροι), λέγεται παραμετρική οικογένεια αν  $\Theta \subset \mathbb{R}^d$  για κάποιο (σταθερό) θετικό ακέραιο  $d$  και το  $\mathbb{P}_\theta$  είναι γνωστό μέτρο πιθανότητας όταν το  $\theta$  είναι γνωστό. Το σύνολο  $\Theta$  λέγεται παραμετρικός χώρος και  $d$  είναι η διάσταση του.
6. Ο πληθυσμός  $\mathbb{P}$  ανήκει σε μία παραμετρική οικογένεια  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ .
7. Η παραμετρική οικογένεια  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$  είναι ταυτοποιήσιμη αν για οποιαδήποτε  $\theta_i, \theta_j \in \Theta$  με  $\theta_i \neq \theta_j$  έχουμε ότι  $\mathbb{P}_{\theta_i} \neq \mathbb{P}_{\theta_j}$ . (Στις περισσότερες περιπτώσεις, μία ταυτοποιήσιμη παραμετρική οικογένεια μπορεί να αποκτηθεί μέσω αναπαραμέτρησης.
8. Μία οικογένεια πληθυσμών  $\mathcal{P}$  κυριαρχείται από το ( $\sigma$ -πεπερασμένο) μέτρο  $\nu$  αν  $\mathbb{P} \ll \nu$  για οποιοδήποτε  $\mathbb{P} \in \mathcal{P}$ . Η  $\mathcal{P}$  μπορεί να ταυτοποιηθεί από την οικογένεια πυκνοτήτων  $\{\frac{d\mathbb{P}}{d\nu} : \mathbb{P} \in \mathcal{P}\}$  ή την  $\{\frac{d\mathbb{P}_\theta}{d\nu} : \theta \in \Theta\}$ .

## 3. Μη πλήρη δεδομένα (Incomplete Data)

Έστω πως η (μονοδιάστατη) τ.μ.  $X$  αντιπροσωπεύει τα δεδομένα και έχει πυκνότητα  $f(x; \phi)$ . Υποθέτουμε πως επιθυμούμε να εκτιμήσουμε την παράμετρο  $\phi$ . Αν το διάνυσμα  $x$  μπορεί να παρατηρηθεί μέσω του δείγματος, τότε μία εκτίμηση για το  $\phi$  μπορεί να γίνει με τη μέθοδο πιθανοφάνειας. Αν όμως το  $x$  δεν μπορεί να παρατηρηθεί πλήρως, αλλά αντί αυτού μπορούν να παρατηρηθούν τα  $y = h(x, \delta)$  και το  $\delta$ . Το  $y = h(x, \delta)$  είναι μία μη-πλήρης έκδοση του  $x$  που ικανοποιεί την  $h(x, \delta = 1) = x$  και η  $\delta$  είναι μία δείκτρια με τιμές 0 και 1. Η εκτίμησης του  $\phi$  από τις παρατηρήσεις των  $(y, \delta)$  αποτελεί το κεντρικό πρόβλημα της ανάλυσης με μη πλήρη δεδομένα.

Προκειμένου να αντιμετωπιστεί το πρόβλημα, η περιθώρια πυκνότητα της  $(y, \delta)$  θα πρέπει να εκφράζεται ως συνάρτηση της αρχικής πυκνότητας  $f(x; \phi)$ . Εκτίμηση με τη μέθοδο πιθανοφάνειας είναι εφικτή κάτω από ορισμένες συνθήκες ταυτοποιησιμότητας και επίσης μπορούν να αναπτυχθούν στατιστικές θεωρίες για την εύρεση της εκτιμήτριας μέγιστης πιθανοφάνειας από το παρατηρούμενο δείγμα. Στην περίπτωση που το  $X$  είναι διάνυσμα, οι υπολογισμοί δυσκολεύουν σημαντικά.

Η συνάρτηση πιθανοφάνειας υπό την ύπαρξη κρυμμένων δεδομένων.

Η περιθώρια πυκνότητα των παρατηρούμενων δεδομένων λέγεται και παρατηρούμενη πιθανοφάνεια. Έστω  $Y = (Y_1, \dots, Y_d)$  ένα τυχαίο διάνυσμα που αντιπροσωπεύει τα δεδομένα, το οποίο έχει πυκνότητα  $f(y; \phi)$  και ζητάμε να κάνουμε εκτίμηση του  $\phi$ . Υποθέτουμε τώρα πως για τη  $i$ -οστή συνιστώσα, παρατηρούμε μόνο ένα μέρος του  $Y_i = (Y_{i1}, \dots, Y_{id})$ . Έστω  $\delta_{ij}$  η δείκτρια:

$$\delta_{ij} = \begin{cases} 1, & \text{αν το } Y_{ij} \text{ παρατηρείται} \\ 0, & \text{διαφορετικά} \end{cases}$$

Για να εκτιμήσουμε το  $\phi$ , θα χρειαστούμε κάποιο μηχανισμό:  $Pr(\delta|y)$ . Συχνά υποθέτουμε πως  $Pr(\delta|y) = \mathbb{P}(\delta|y; \psi)$ , όπου  $\psi$  η παράμετρος του μοντέλου απάντησης και  $\delta = (\delta_1, \dots, \delta_d)$ .

Έστω  $(y_{i,obs}, y_{i,miss})$ , το παρατηρούμενο και το χαμένο μέρος των  $y_i$ , αντίστοιχα. Για κάθε  $i$  παρατηρούμε το  $(y_{i,obs}, \delta_i)$  αντί για το  $y_i$ . Από τα παραπάνω, η περιθώρια πυκνότητα του  $(y_{i,obs}, \delta_i)$  είναι:

$$g(y_{i,obs}, \delta_i; \phi, \psi) = \int f(y_i, \phi) \mathbb{P}(\delta_i|y_i; \psi) d\mu(y_{i,miss}).$$

Υποθέτοντας ανεξαρτησία και ισονομία, η από κοινού της παραπάνω περιθώριας γράφεται:

$$g(y_{obs}, \delta; \phi, \psi) = \prod_{i=1}^n g(y_{i,obs}, \delta_i; \phi, \psi),$$

όπου,  $y_{obs} = (y_{1,obs}, \dots, y_{n,obs})$ . Η τελευταία λέγεται παρατηρούμενη πιθανοφάνεια.

Έστω

$$\mathcal{R}(y_{obs}, \delta) = \{y : y_{obs}(y, \delta_i) = y_{i,obs}, i = 1, \dots, n\},$$

το σύνολο όλων των δυνατών τιμών που μπορεί να πάρει το  $y$  έχοντας κοινή την τιμή της  $y_{obs}$ , δοθέντος  $\delta$ . Η συνάρτηση  $y_{obs}(y, \delta_i)$  επιστρέφει την τιμή του  $y_{ij}$  όταν  $\delta_{ij} = 1$ .

**Ορισμός.** (Παρατηρούμενη πιθανοφάνεια). Έστω  $f(y; \phi)$  η από κοινού πυκνότητα των πλήρη δεδομένων  $y = (y_1, \dots, y_n)$  και έστω  $\mathbb{P}(\delta|y; \psi)$  η δεσμευμένη πιθανότητα του  $\delta = (\delta_1, \dots, \delta_n)$  δοθέντος του  $y$ . Η παρατηρούμενη πιθανοφάνεια της  $\phi, \psi$  βασισμένη στα παρατηρούμενα δεδομένα  $y_{obs}, \delta$  δίνεται από την:

$$L_{obs}(\phi, \psi) = \int_{\mathcal{R}(y_{obs}, \delta)} f(y; \phi) \mathbb{P}(\delta|y; \psi) d\mu(y). \quad (8.0.4)$$

Κάτω από την υπόθεση ανεξαρτησίας και ισονομίας η παραπάνω γράφεται:

$$L_{obs}(\phi, \psi) = \prod_{i=1}^n \left[ \int f(y; \phi) \mathbb{P}(\delta_i | y_i; \psi) d\mu(y_{i,miss}) \right]. \quad (8.0.5)$$

θεωρώντας πως αν  $y_i = y_{i,obs}$  και το  $y_{i,miss}$  είναι κενό, there is nothing to integrate out. Όταν έχουμε μονοδιάστατο δείγμα, η παραπάνω γίνεται:

$$L_{obs}(\phi, \psi) = \prod_{\{i:\delta_i=1\}} [f(y_i; \phi) p(y_i; \psi)] \times \prod_{\{i:\delta_i=0\}} \left[ \int f(y; \phi) [1 - p(y; \psi)] d\mu(y) \right]. \quad (8.0.6)$$

όπου  $p(y; \psi) = \mathbb{P}(\delta = 1 | y; \psi)$ . Τέλος, η παράμετρος  $\psi$  μπορεί να θεωρηθεί nuisance με την έννοια ότι δεν ενδιαφερόμαστε άμεσα για την εκτίμηση της, ωστόσο χρειαζόμαστε την εκτίμηση της για την εκτίμηση του  $\phi$ .

### 8.0.3 A.3 Επιλεγμένες αποδείξεις θεωρημάτων και προτάσεων

**Θεώρημα 3.1** (i) Έστω  $\mathbf{X}$  μία Μ.Α. με αρχική κατανομή  $\mu$  και πυρήνα  $Q$ . Τότε για κάθε  $B \in \mathcal{X}^{n+1}$ , της μορφής  $B = B_0 \times \dots \times B_n$  ισχύει:

$$\mathbb{P}(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \dots \int_{B_n} Q(x_{n-1}, dx_n), \quad (8.0.7)$$

(ii) Αντίστροφα, αν μία σ.δ.  $\mathbf{X} = (\Omega, \mathcal{F}, (X_n)_{n \geq 0}, \mathbb{P})$  με τιμές στον  $(\mathcal{X}, \mathcal{X})$ , ικανοποιεί την 3.1.3 για ένα μέτρο πιθανότητας  $\mu$  και ένα πυρήνα  $Q$ , τότε η  $\mathbf{X}$  είναι Μ.Α με αρχική κατανομή το  $\mu$  και πυρήνα το  $Q$  (ως προς τη φυσική διήθηση  $(\mathcal{F}_n^0)_{n \geq 0}$ )

#### Απόδειξη

Το (i) είναι άμεσο. Θα δείξουμε το (ii).

Έστω  $n \geq 0$  και  $A \in \mathcal{X}$ . Θα δείξουμε ότι για κάθε  $C \in \mathcal{F}_n$  ισχύει:

$$\int_C \mathbb{1}_{\{X_{n+1} \in A\}} d\mathbb{P} = \int_C Q(X_n, A) d\mathbb{P} \quad (8.0.8)$$

Τότε έπεται πως  $E[\mathbb{1}_{\{X_{n+1} \in A\}} | \mathcal{F}_n] = Q(X_n, A)$  σ.π. και επομένως η  $\mathbf{X}$  είναι μ.α.

Έστω  $B = B_0 \times \dots \times B_n$  και  $C = \{X_0 \in B_0, \dots, X_n \in B_n\}$ . Τότε χρησιμοποιώντας τις ιδιότητες του ολοκληρώματος και την υπόθεση θα είναι:

$$\begin{aligned} \int_C \mathbb{1}_{\{X_{n+1} \in A\}} d\mathbb{P} &= \mathbb{P}(X_0 \in B_0, \dots, X_n \in B_n, X_{n+1} \in A) \\ &= \int_{B_0} \mu(dx_0) \int_{B_1} Q(x_0, dx_1) \dots \int_{B_n} Q(x_{n-1}, dx_n) Q(x_n, A). \end{aligned}$$

Η τελευταία έκφραση είναι ίση με:

$$\int_C Q(X_n, A) d\mathbb{P}.$$

Πράγματι, ο ισχυρισμός αυτός ισχύει αν αντικατασταθεί η συνάρτηση  $Q(x_n, A)$  από μία φραγμένη και μετρήσιμη  $f(x_n)$ . Αν η  $f$  είναι δείκτρια τότε προφανώς η ισότητα ισχύει. Από τη γραμμικότητα το ολοκληρώματος, η ισότητα ισχύει και για απλές μετρήσιμες. Τέλος, από το Θεώρημα φραγμένης σύγκλισης, ο ισχυρισμός επαληθεύεται. Οπότε για την επιλογή του  $C$  έχουμε πως:

$$\int_C \mathbb{1}_{\{X_{n+1} \in A\}} d\mathbb{P} = \int_C Q(X_n, A) d\mathbb{P}.$$

Τώρα, εύκολα δείχνεται πως η οικογένεια των συνόλων για τα οποία ισχύει η παραπάνω ισότητα είναι ένα  $\lambda$ -σύστημα. Από την άλλη, η οικογένεια των συνόλων για τα οποία αποδείχθηκε είναι ένα  $\pi$ -σύστημα. Οπότε από το Θεώρημα  $\pi - \lambda$  το παραπάνω ισχύει για όλα τα  $C \in \mathcal{F}_n$ .  $\square$

**Θεώρημα 3.6** Έστω μία  $\phi$ -αδιαχώριστη Μ.Α. για την οποία υπάρχει atom, έστω  $a$ . Η Μ.Α. λέγεται θετική αν  $(E_a[\tau_a])^{-1} < \infty$ . Τότε η στάσιμη κατανομή που επάγει ικανοποιεί την:

$$\pi(a) = (E_a[\tau_a])^{-1} \quad (8.0.9)$$

### Απόδειξη

Λήμμα: Αν μία Μ.Α. είναι θετική, τότε είναι επαναληπτική.

Αν  $E_a[\tau_a] < \infty$  τότε  $\mathbb{P}_a(\tau_a < \infty) = 1$ , και άρα η αλυσίδα είναι επαναληπτική από το Λήμμα. Έστω το μέτρο:

$$\pi(A) = \sum_{n \geq 1} \mathbb{P}_a(X_n \in A, \tau_a \geq n)$$

Επειδή  $\pi(a) = \mathbb{P}_a(\tau_a < \infty) = 1$ , και:

$$\begin{aligned} \int Q(x, A)\pi(dx) &= \pi(a)Q(a, A) + \int_a^c \sum_{n \geq 1} Q(x_n, A)\mathbb{P}_a(\tau_a \geq n, dx_n) \\ &= Q(a, A) + \sum_{n \geq 2} \mathbb{P}_a(X_n \in A, \tau_a \geq n) = \pi(A) \end{aligned}$$

το  $\pi$  είναι στάσιμο. Επίσης, είναι και πεπερασμένο καθώς

$$\begin{aligned} \pi(X) &= \sum_{n \geq 1} \mathbb{P}_a(\tau_a \geq n) = \sum_{k \geq 1} k\mathbb{P}_a(\tau_a = k) \\ &= E_a[\tau_a] < \infty \end{aligned}$$

Κανονικοποιώντας ξανά το  $\pi$  σε  $\pi/\pi(X)$  παίρνουμε  $\pi(a) = (E_a[\tau_a])^{-1}$ .

**Πρόταση 3.2** Έστω  $(X_n, Y_n)_{n \geq 0}$  μία δισδιάστατη ορισμένη στον  $(\mathcal{X} \times \mathcal{Y})$  με πυρήνα τον  $T$  που δίνεται στην (3.3.1). Τότε, για οποιοδήποτε ακέραιο  $m$  και οποιοδήποτε διατεταγμένο σύνολο δεικτών  $\{k_1 < \dots < k_m\}$  και όλες τις μετρήσιμες και φραγμένες συναρτήσεις  $f_1, \dots, f_m$  ισχύει:

$$E_\nu \left[ \prod_{i=1}^m f_i(Y_{k_i}) | X_{k_1}, \dots, X_{k_m} \right] = \prod_{i=1}^m E_\nu [f_i(Y_{k_i}) | X_{k_i}]. \quad (8.0.10)$$

Για οποιαδήποτε μετρήσιμη και φραγμένη συνάρτηση  $h$  είναι:

$$\begin{aligned} E_\nu \left[ \prod_{i=1}^m f_i(Y_{k_i}) h(X_{k_1}, \dots, X_{k_m}) \right] &= \int \dots \int \nu(dx_0) G(x_0, dy_0) \left[ \prod_{i=1}^{k_m} Q(x_{i-1}, d_{x_i}) G(x_i, dy_i) \right] \\ &\times \left[ \prod_{i=1}^m f_i(y_{k_i}) \right] h(x_{k_1}, \dots, X_{k_m}) = \int \dots \int \nu(dx_0) \prod_{i=1}^{k_m} Q(x_{i-1}, d_{x_i}) h(x_{k_1}, \dots, X_{k_m}) \\ &\int \dots \int \left[ \prod_{i \notin \{k_1, \dots, k_m\}} G(x_i, dy_i) \right] \left[ \prod_{i \in \{k_1, \dots, k_m\}} \int f_i(y_i) G(x_i, dy_i) \right] \end{aligned}$$

Επειδή για κάθε  $i$  είναι  $\int G(x_i, dy_i) = 1$  παίρνουμε ότι:

$$E_\nu \left[ \prod_{i=1}^m f_i(Y_{k_i}) h(X_{k_1}, \dots, X_{k_m}) \right] = E_\nu [h(x_{k_1}, \dots, x_{k_m}) \prod_{i \in \{k_1, \dots, k_m\}} \int f_i(y_i) G(x_i, dy_i)]$$



## Bibliography

- [1] Robert B Ash. *Real Analysis and Probability: Probability and Mathematical Statistics: a Series of Monographs and Textbooks*. Academic Press, 2014.
- [2] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.
- [3] Yves F Atchadé, Gareth O Roberts, and Jeffrey S Rosenthal. Towards optimal scaling of metropolis-coupled markov chain monte carlo. *Statistics and Computing*, 21(4):555–568, 2011.
- [4] Krishna B Athreya and Peter Jagers. *Classical and modern branching processes*, volume 84. Springer Science & Business Media, 2012.
- [5] Charlotte Baey, Samis Trevezas, and Paul-Henry Cournede. A non linear mixed effects model of plant growth and estimation via stochastic variants of the em algorithm. *Communications in Statistics-Theory and Methods*, 45(6):1643–1669, 2016.
- [6] Daniel Barthélémy and Yves Caraglio. Plant architecture: a dynamic, multi-level and comprehensive approach to plant form, structure and ontogeny. *Annals of botany*, 99(3):375–407, 2007.
- [7] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.
- [8] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.
- [9] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [10] Zdzislaw Brzezniak and Tomasz Zastawniak. *Basic stochastic processes: a course through exercises*. Springer Science & Business Media, 1999.
- [11] Olivier Cappé, Eric Moulines, and Tobias Rydén. Inference in hidden markov models. In *Proceedings of EUSFLAT Conference*, pages 14–16, 2009.
- [12] Yuting Chen, Samis Trevezas, and Paul-Henry Cournede. A regularized particle filter em algorithm based on gaussian randomization with an application to plant growth modeling. *Methodology and Computing in Applied Probability*, 17(4):847–870, 2015.

## Bibliography

- [13] P-H Cournede, Veronique Letort, Amélie Mathieu, Meng Zhen Kang, Sébastien Lemaire, Samis Trevezas, François Houllier, and Philippe De Reffye. Some parameter estimation issues in functional-structural plant modelling. *Mathematical Modelling of Natural Phenomena*, 6(2):133–159, 2011.
- [14] Paul-Henry Cournède, Meng-Zhen Kang, Amélie Mathieu, Jean-François Barczi, Hong-Pin Yan, Bao-Gang Hu, and Philippe De Reffye. Structural factorization of plants to compute their functional and architectural growth. *Simulation*, 82(7):427–438, 2006.
- [15] Philippe De Reffye and BG Hu. Relevant qualitative and quantitative choices for building an efficient dynamic plant growth model: Greenlab case. In *International Symposium on Plant Growth Modeling, Simulation, Visualization and their Applications-PMA'03*, pages 87–107. Springer and Tsinghua University Press, 2003.
- [16] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.
- [17] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.
- [18] William Feller. *An introduction to probability theory and its applications: volume I*, volume 3. John Wiley & Sons London-New York-Sydney-Toronto, 1968.
- [19] Bert E Fristedt and Lawrence F Gray. *A modern approach to probability theory*. Springer Science & Business Media, 2013.
- [20] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.
- [21] YAN Guo, Yuntao Ma, Zhigang Zhan, Baoguo Li, Michael Dingkuhn, Delphine Luquet, and Philippe De Reffye. Parameter optimization and field validation of the functional–structural model greenlab for maize. *Annals of botany*, 97(2):217–230, 2006.
- [22] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [23] JL Hibey. Markov chains and stochastic stability-sp meyn and rl tweedie. *IEEE TRANSACTIONS ON AUTOMATIC CONTROL AC*, 40(5):979–979, 1995.
- [24] WS Kendall. E. nummelin general irreducible markov chains and non-negative operators (cambridge tracts in mathematics 83, cambridge university press, 1984), 156 pp.£ 20. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 29(02):283–284, 1986.
- [25] Jae Kwang Kim and Jun Shao. *Statistical methods for handling incomplete data*. CRC Press, 2013.

## Bibliography

- [26] Vidyadhar G Kulkarni. *Modeling and analysis of stochastic systems*. CRC Press, 1996.
- [27] Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- [28] Xiao-Li Meng and Donald B Rubin. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278, 1993.
- [29] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [30] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [31] Per Mykland, Luke Tierney, and Bin Yu. Regeneration in markov chain samplers. *Journal of the American Statistical Association*, 90(429):233–241, 1995.
- [32] Dan Nettleton. Convergence properties of the em algorithm in constrained parameter spaces. *Canadian Journal of Statistics*, 27(3):639–648, 1999.
- [33] James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.
- [34] Fernando Pérez-Cruz. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE international symposium on information theory*, pages 1666–1670. IEEE, 2008.
- [35] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [36] Christian P Robert, Tobias RydÉn, and David M Titterington. Convergence controls for mcmc algorithms, with applications to hidden markov chains. *Journal of Statistical Computation and Simulation*, 64(4):327–355, 1999.
- [37] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statist. Sci.*, 16(4):351–367, 11 2001.
- [38] Jeffrey S Rosenthal et al. Optimal proposal distributions and adaptive mcmc. *Handbook of Markov Chain Monte Carlo*, pages 93–112, 2011.
- [39] Sheldon M Ross et al. *Stochastic processes*, volume 2. John Wiley & Sons New York, 1996.
- [40] Rolf Sundberg. Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, pages 49–58, 1974.
- [41] Samis Trevezas and Paul-Henry Cournede. A sequential monte carlo approach for mle in a plant growth model. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2):250–270, 2013.

## Bibliography

- [42] Samis Trevezas, Sonia Malefaki, and P-H Cournède. Parameter estimation via stochastic variants of the ecm algorithm with applications to plant growth modeling. *Computational Statistics & Data Analysis*, 78:82–99, 2014.
- [43] Florin Vaida. Parameter convergence for em and mm algorithms. *Statistica Sinica*, pages 831–840, 2005.
- [44] CF Jeff Wu. On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103, 1983.
- [45] Willard I Zangwill. Convergence conditions for nonlinear programming algorithms. *Management Science*, 16(1):1–13, 1969.
- [46] Willard I Zangwill. *Nonlinear programming: a unified approach*. 1969.