ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΠΕΫΖΙΑΝΗ ΠΡΟΣΕΓΓΙΣΗ ΓΙΑ ΤΙΣ ΠΡΟΒΛΕΨΕΙΣ ΤΟΥ ΚΑΤΑ ΗΛΙΚΙΑ ΣΥΝΤΕΛΕΣΤΗ ΘΗΝΗΣΙΜΟΤΗΤΑΣ

Χαράλαμπος Ν. Φάρος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών στο πλαίσιο του

Προπτυχιακού Προγράμματος Σπουδών

Αθήνα

Σεπτέμβριος 2022

ΕΥΧΑΡΙΣΤΙΕΣ

Με την παρούσα εργασία θα ήθελα να ευχαριστήσω τον καθηγητή κ. Νικόλαο Δεμίρη για την εξαιρετική συνεργασία και καθοδήγηση τόσο στο επιστημονικό κομμάτι όσο και στην συγγραφή της παρούσας διπλωματικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω και τον κ. Αλεξόπουλο, ο οποίος με βοήθησε στην κατανόηση της επιστημονικής εργασίας του και ήταν αρωγός σε αυτή τη συνεχή προσπάθεια.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου που με υποστήριξε όλα αυτά τα χρόνια, σε όλες τις καλές και κακές στιγμές αυτού του δύσκολου αλλά ευχάριστου ταξιδιού.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Γεννήθηκα στο Μαρούσι στις 8 Μαρτίου το 2000. Μεγάλωσα στη Νίκαια όπου και αποφοίτησα από το 3ο ΓΕΛ Νίκαιας. Ως φοιτητής του τμήματος Στατιστικής διακρίθηκα για τις ακαδημαϊκές μου επιδόσεις στα έτη 2018-2019 και 2020-2021. Από εργασιακές εμπειρίες, διετέλεσα την τρίμηνη πρακτική μου άσκηση στον ΙΟΒΕ, ενώ πλέον δουλεύω στη Grifon Capital (εταιρία που υφίσταται για λογαριασμό του hedge fund Fortress).

ΠΕΡΙΛΗΨΗ

Χαράλαμπος Φάρος

ΜΠΕΫΖΙΑΝΗ ΠΡΟΣΕΓΓΙΣΗ ΓΙΑ ΤΙΣ ΠΡΟΒΛΕΨΕΙΣ ΤΟΥ ΚΑΤΑ ΗΛΙΚΙΑ ΣΥΝΤΕΛΕΣΤΗ ΘΝΗΣΙΜΟΤΗΤΑΣ

Σεπτέμβριος 2022

Στην παρούσα εργασία εξετάζεται η μοντελοποίηση των πιθανοτήτων θανάτου χρησιμοποιώντας μπεϋζιανές μεθόδους. Αρχικά, δίνεται το μαθηματικό υπόβαθρο για την κατανόηση της μπεϋζιανής στατιστικής καθώς και των μεθόδων Markov Chain Monte Carlo (MCMC) που χρησιμοποιούνται σε αυτή. Στη συνέχεια, παρουσιάζεται αφενός το μοντέλο Heligman – Pollard το οποίο χρησιμοποιείται ευρέως στη μοντελοποίηση των πιθανοτήτων θανάτου και αφετέρου η γενικότερη μεθοδολογία μοντελοποίησης του Gaussian Markov Random Field (GMRF).

Με βάση την επιστημονική δημοσίευση των: Αλεξόπουλος, Δελλαπόρτας, Forster (2018) όπου τροποποίησαν τα παραπάνω μοντέλα/μεθόδους κατά Μπεϋζ για να προβλέψουν τις πιθανότητες θανάτου, γίνεται μια πλήρης περιγραφή των δύο μεθόδων καθώς και εφαρμογή αυτών στα δεδομένα της Ελλάδας και της Νορβηγίας.

Τέλος, βάσει των αποτελεσμάτων που προκύπτουν, αξιολογείται και σχολιάζεται κατά περίπτωση η εφαρμογή της ακολουθούμενης μεθοδολογίας.

ABSTRACT

Charalampos Faros

BAEYSIAN APPROACH FOR PROJECTING AGE-SPECIFIC MORTALITY RATES

September 2022

In this thesis, the death probability modeling is considered using Bayesian methods. At first, the mathematical background for understanding Bayesian statistics and Markov Chain Monte Carlo (MCMC) methods, are presented. Then, Helligman – Pollard model, which is used widely for modelling age-specific mortality rates, is given along with the general modelling method of Gaussian Markov Random Field (GMRF).

This work is based on the scientific publication of: Alexopoulos, Dellaportas, Forster (2018), who modified above mentioned models/ methods according to a Bayesian approach, to predict the mortality rates. A full description of both methods of their approach, and the application on Greek and Norwegian data, is given.

At last, the implementation of these methods is evaluated and commented based of the applications' results

Περιεχόμενα

Πίνακας Πινάκων

Πίνακας Εικόνων

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Στις μέρες μας οι προβλέψεις των πιθανοτήτων θανάτου ενδιαφέρουν αναλογιστές, κράτη, ασφαλιστικές εταιρίες, ασφαλιστικούς οργανισμούς, δημογράφους, στατιστικούς κ.α. Ενδιαφέρουν δε, οι έγκυρες προβλέψεις με κάποιο μέτρο αβεβαιότητας, δηλαδή οι πιθανοτικές προβλέψεις. Ο κύριος οργανισμούς που κάνει πληθυσμιακές προβλέψεις και οι οποίες χρησιμοποιούνται ευρέως είναι τα Ηνωμένα Έθνη. Μέχρι πρόσφατα, οι προβλέψεις του ΟΗΕ γίνονταν με ντετερμινιστικό τρόπο, συνεπώς δεν υπήρχε ουσιαστικό μέτρο αβεβαιότητας. Για την αντιμετώπιση αυτού, αναπτύχθηκαν μπεϋζιανές μέθοδοι οι οποίες πλέον καθιερώθηκαν [20].

Εστιάζοντας στις πιθανότητες θανάτου, έχουν προταθεί από τη βιβλιογραφία διάφορα υποδείγματα που μοντελοποιούν το μοτίβο των πιθανοτήτων θανάτου ανά ηλικία. Το μοντέλο των Heligman – Pollard, Lee – Carter και διάφορα άλλα γενικευμένα γραμμικά μοντέλα [3] είναι τα πιο διαδεδομένα.

Οι μέθοδοι που θα αναπτυχθούν παρακάτω δεν αποτελούν τις καθιερωμένες μεθόδους που χρησιμοποιεί ο ΟΗΕ, αντίθετα αποτελούν επιστημονικό έργο του [3]. Οι δύο μέθοδοι αυτοί μπορούν συνοπτικά να περιγραφούν ως εξής:

Η πρώτη μέθοδος αποτελεί εξέλιξη του μοντέλου Heligman – Pollard [10]. Συγκεκριμένα μοντελοποιήθηκαν οι 8 παράμετροι του μοντέλου αυτού σε ένα ιεραρχικό Μπεϋζιανό υπόδειγμα (με κάποιες επιπλέον παραμέτρους). Έτσι χρησιμοποιώντας τα πληθυσμιακά διαθέσιμα δεδομένα γίνεται πρώτα εκτίμηση και έπειτα οι μελλοντικές προβλέψεις των παραμέτρων. Οι προβλέψεις αυτές εφαρμόζονται στο μοντέλο Η – Ρ δίνοντας έτσι το μοτίβο θνησιμότητας.[3]

Η δεύτερη μέθοδος είναι πιο «άμεση» καθώς υποθέτει ότι οι (μετασχηματισμένες) πιθανότητες θανάτου είναι μεταβλητές και ακολουθούν μία («ακατάλληλη») πολυμεταβλητή κανονική κατανομή επιτρέποντας τις μεταξύ τους εξαρτήσεις. Επίσης εισάγονται λίγες επιπλέον μεταβλητές και όλες μαζί οργανώνονται σε ένα ιεραρχικό Μπεϋζιανό μοντέλο. Η πρόβλεψη γίνεται απευθείας στις μετασχηματισμένες πιθανότητες θανάτου.[3]

Η παρούσα εργασία έχει ως κύριο στόχο να περιγράψει αναλυτικά τις δύο παραπάνω μεθόδους, να τις εφαρμόσει και να τις αξιολογήσει για τα παρακάτω δεδομένα. Τα δεδομένα που χρησιμοποιήθηκαν είναι οι θάνατοι και ο πληθυσμός ανά ηλικία για τις χώρες: Ελλάδα, Νορβηγία και για τις χρονιές 1983-1992. Τα δεδομένα αυτά πάρθηκαν από το Human Mortality Database και αφορούν και τα δύο φύλα μαζί. Για την εφαρμογή και την αξιολόγηση των παραπάνω τεχνικών χρησιμοποιήθηκε η γλώσσα προγραμματισμού « R». Η αξιολόγηση των μεθόδων επικεντρώθηκε την ποιότητα της σύγκλισης του MCMC χρησιμοποιώντας γραφήματα καθώς και το Effective Sample Size.

ΚΕΦΑΛΑΙΟ 2

ΤΑ ΜΑΘΗΜΑΤΙΚΑ ΤΗΣ ΜΠΕΫΖΙΑΝΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Η βασική ιδέα

Η μπεϋζιανή στατιστική διαφέρει με την κλασσική στατιστική σε 2 σημεία:

1. Οι παράμετροι δεν είναι απλά άγνωστοι αριθμοί που χρήζουν εκτίμησης, αλλά τυχαίες μεταβλητές με δικιά τους κατανομή.
2. Δεν υπάρχει «αντικειμενικότητα», αλλά κάθε πιθανότητα/κατανομή είναι μοναδική για κάθε άτομο καθώς προκύπτει από συνδυασμό των πεποιθήσεων και των δεδομένων

Επί του πρακτέου, έστω μια παράμετρος θ η οποία είναι άγνωστη και πρέπει να εκτιμηθεί. Τις πεποιθήσεις (πριν παρουσιαστούν τα δεδομένα) για τις τιμές που παίρνει η θ τις αντιπροσωπεύει μια κατανομή, έστω $\pi(\theta)$, αυτή η κατανομή ονομάζεται και a-priori κατανομή ή -πιο γρήγορα- prior. Έστω τώρα ότι παρουσιάζονται τα δεδομένα x, τα οποία ακολουθούν την κατανομή $f(x|\theta)$.

Υπενθύμιση

Θεώρημα ολικής πιθανότητας

$$\mathrm{P(A)} = \sum \mathrm{P(A|B_i)P(B_i)} \ \ \mu\varepsilon\ B_i\ \delta\iota\alpha\mu\acute{\varepsilon}\rho\iota\sigma\eta\ \tau\sigma\upsilon\ \Omega \qquad (2.1)$$

και στην συνεχή περίπτωση:

$$P(A) = \int P(A|X = x)P(X = x)dx = \int P(A|X = x)f(x)dx \qquad (2.2)$$

Χρησιμοποιώντας τώρα τον τύπο του Bayes θα υπολογιστεί η κατανομή -posterior- του θ ως εξής:

$$p(\theta|x) = \frac{f(x,\theta)}{f(x)} = \frac{f(x|\theta)\pi(\theta)}{f(x)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} (\propto f(x|\theta)\pi(\theta)) \qquad (2.3)$$

Αξίζει να σημειωθεί, ότι η κεντρική ιδέα παραμένει ίδια είτε το θ είναι μια παράμετρος είτε είναι διάνυσμα παραμέτρων $\theta = (\theta_1, \theta_2, \dots, \theta_n)^T$.

Κατανομή πρόβλεψης

Υπενθύμιση

Ανεξαρτησία: Τα ενδεχόμενα A, B θα λέμε ότι είναι ανεξάρτητα αν

$$P(A, B) = P(A)P(B) \tag{2.4}$$

Υπό συνθήκη ανεξαρτησία: Τα ενδεχόμενα A, B θα λέμε ότι είναι υπό συνθήκη ανεξάρτητα δοθέντος C αν

$$P(A, B|C) = P(A|C)P(B|C) \text{ ή ισοδύναμα}$$

$$P(A|B, C) = P(A|C) \tag{2.5}$$

Βάσει της posterior κατανομής, δηλαδή τις πεποιθήσεις και τα δεδομένα, δύναται πλέον να υπολογιστεί μια κατανομή για τις μελλοντικές τιμές των δεδομένων. Έστω y η επόμενη παρατήρηση που θα συλλεχθεί, και έστω x τα δεδομένα έως τώρα. Η κατανομή $y|x$ δίνεται παρακάτω:

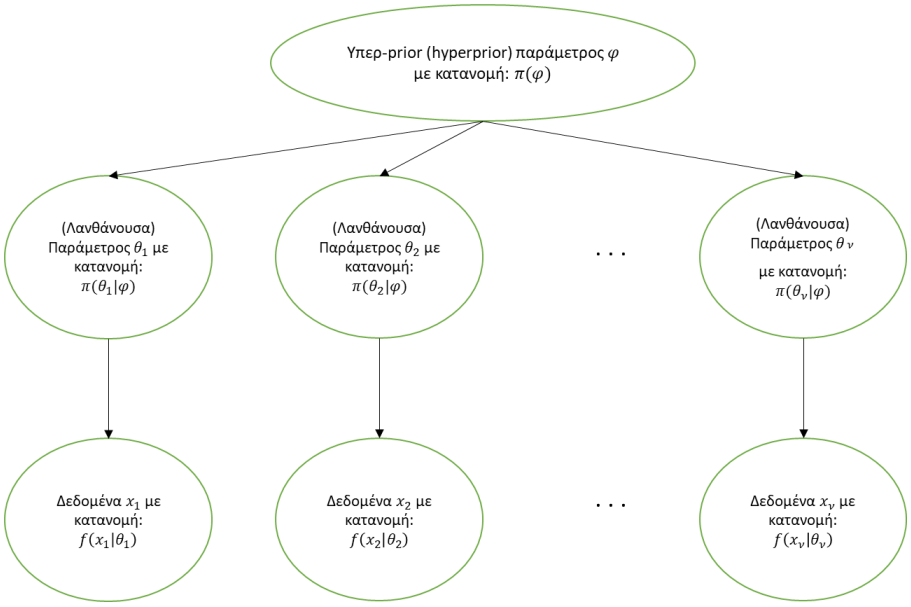
$$p(y|x) = \int p(y, \theta|x)\, d\theta = \int \frac{p(y, \theta, x)}{p(x)} d\theta = \int \frac{p(y, \theta, x)}{p(\theta, x)} \frac{p(\theta, x)}{p(x)} d\theta = \int p(y|\theta, x)p(\theta|x)\, d\theta \tag{2.6}$$

Και αν θωρηθεί ότι τα y, x είναι υπό συνθήκη ανεξάρτητα δοθέντος του θ, τότε η παραπάνω εξίσωση γίνεται:

$$p(y|x) = \int p(y|\theta)p(\theta|x)\, d\theta \tag{2.7}$$

Μπεϋζιανά ιεραρχικά μοντέλα

Τα μπεϋζιανά ιεραρχικά μοντέλα βασίζονται σε μια δομή, η οποία έχει ως τελικό σκοπό να δημιουργήσει πιο «ανθεκτικά» μοντέλα. Το βασικό μοντέλο της ιεραρχικής δομής φαίνεται στην Εικόνα 1.:



Εικόνα 1 Αναπαράσταση της ιεραρχικής δομής ενός Μπεϋζιανού ιεραρχικού μοντέλου

Παράδειγμα

Έστω, ένα μηχάνημα που φτιάχνει πλακέτες ν ειδών. Τα $x_i (i = 1, \dots, \nu)$ είναι ο αριθμός των χαλασμένων transistor στην πλακέτα του είδους i και ακολουθούν την κατανομή $f(x_i|\theta_i)$. Οι παράμετροι θ_i, που σχετίζονται με την κατανομή του αριθμού των χαλασμένων transistor στην πλακέτα του είδους i, προφανώς διαφέρουν μεταξύ του από το γεγονός ότι η πλακέτα του κάθε είδους έχει διαφορετικό σχήμα/πολυπλοκότητα κλπ. Όμως, επειδή οι πλακέτες όλων των ειδών φτιάχνονται από το ίδιο μηχάνημα υπάρχει κάποια σύνδεση μεταξύ των θ_i. Η σύνδεση αυτή αναπαρίσταται από το γεγονός ότι οι παράμετροι θ_i προκύπτουν από μια άλλη «υπερπαράμετρο» φ, δηλαδή $\pi(\theta_i|\varphi)$.

Ένα από τα πιο βασικά προτερήματα αυτής της μεθόδου είναι ότι αν δεν έχουμε πολύ πληροφορία για κάποιο i, δηλαδή ελλιπή x_i, τότε η εκτίμηση του θ_i «δανείζεται δύναμη» από τα υπόλοιπα δεδομένα, εξαιτίας της παραμέτρου φ που όλα τα θ_i υπάγονται.

Μία συνοπτική περιγραφή της Εικόνας 1, είναι ότι η παράμετρος θ_i έχει prior την $\pi(\theta_i|\varphi)$ που με τη σειρά της η παράμετρος φ έχει prior την $\pi(\varphi)$.

Σε αυτό το σημείο να σημειωθεί ότι τα μπεϋζιανά ιεραρχικά μοντέλα βασίζονται στην ιδιότητα της δεσμευμένης ανεξαρτησίας.

Συνεπώς για την εύρεση της posterior κατανομής των θ, φ:

$$p(\theta, \varphi|x) \propto f(x|\theta, \varphi) f(\theta, \varphi) = f(x|\theta) f(\theta, \varphi) = f(x|\theta) f(\theta|\varphi) \pi(\varphi) \qquad (2.8)$$

ΚΕΦΑΛΑΙΟ 3

MARKOV CHAIN MONTE CARLO (MCMC)

Θεωρία

Η Μπεϋζιανή στατιστική αντιμετωπίζει πολύ συχνά το εξής πρόβλημα: Για να βρεθεί η ακριβής posterior κατανομή πρέπει να μπορεί να υπολογιστεί ακριβώς η τιμή του ολοκληρώματος στον παρανομαστή (σταθερά κανονικοποίησης), όπως αυτή φαίνεται στην εξίσωση (1). Αυτό όμως στην πράξη είναι αδύνατον, καθώς το ολοκλήρωμα δύναται να μην μπορεί να υπολογιστεί με κλειστό τύπο, ενώ για ένα διάνυσμα παραμέτρων θα έπρεπε να υπολογιστεί το αντίστοιχο πολλαπλό ολοκλήρωμα. Ακόμα και η αριθμητική προσέγγιση ενός τέτοιου πολλαπλού ολοκληρώματος θα απαιτούσε τεράστια υπολογιστική δύναμη. Για τη λύση αυτού του προβλήματος, αναπτύχθηκαν αλγόριθμοι ώστε να προσομοιώνονται τυχαία δείγματα από την posterior κατανομή, αποφεύγοντας έτσι πλήρως τον υπολογισμό του παρανομαστή/σταθεράς κανονικοποίησης [6]-[7].

Η βασική ιδέα του MCMC είναι η δημιουργία μιας Μαρκοβιανής αλυσίδας $\theta_1, \theta_2, \dots, \theta_i, \theta_{i+1}, \dots, \theta_n$, η οποία να συγκλίνει στην κατανομή-στόχο, δηλαδή για την περίπτωσή μας στην $p(\theta|x)$

<u>Θεωρία MCMC</u>

Μια ακολουθία $X_1, X_2, \dots$ τυχαίων μεταβλητών θα ονομάζεται Μαρκοβιανή Αλυσίδα αν $P(X_n|X_1, X_2, \dots, X_{n-1}) = P(X_n|X_{n-1})$, δηλαδή η δεσμευμένη κατανομή της X_n δοθέντος των $X_1, \dots, X_{n-1}$ εξαρτάται μόνο από την X_{n-1}. Ουσιαστικά είναι μία στοχαστική ανέλιξη που έχει τη Μαρκοβιανή ιδιότητα. Το σύνολο στο οποίο παίρνει τιμές η X_t ονομάζεται χώρος καταστάσεων και συνήθως τον συμβολίζουμε με Ω. Το $t \in T$ αφορά τον χρόνο και μπορεί να είναι είτε συνεχής είτε διακριτός, στην περίπτωση του MCMC είναι διακριτός.

Για την μαθηματική ακρίβεια, μαρκοβιανή αλυσίδα ονομάζεται μόνο η στοχαστική ανέλιξη που έχει τη μαρκοβιανή ιδιότητα και **διακριτό** χώρο καταστάσεων, είτε σε συνεχή χρόνο είτε σε διακριτό. Στην περίπτωση του MCMC που κατά κύριο λόγο ο χώρος καταστάσεων είναι συνεχής, δεν μπορεί να ονομασθεί μαρκοβιανή αλυσίδα αλλά απλά στοχαστική ανέλιξη. Παρόλα αυτά, για χάρη του ονόματος «MCMC», όταν θα γίνεται αναφορά σε μαρκοβιανή αλυσίδα, στην πραγματικότητα θα εννοείται στοχαστική ανέλιξη σε διακριτό χρόνο. [8]

Επιπλέον, για τις MCMC μεθόδους που θα αναλυθούν, θα ισχύει ότι η πιθανότητα μεταπήδησης, και πιο συγκεκριμένα, η κατανομή της πιθανότητας μεταπήδησης (λόγω συνεχούς χώρου καταστάσεων), θα είναι ανεξάρτητη του χρόνου. Θα εννοείται δηλαδή, ότι η μαρκοβιανή αλυσίδα θα είναι ομοιογενής ή ότι θα έχει στάσιμες πιθανότητες μεταπήδησης.

Ωστόσο, το κλειδί της λειτουργείας του MCMC, βασίζεται στο θεώρημα [8], κατά το οποίο , μια ανάγωγη και εργοδική[1] μαρκοβιανή αλυσίδα συγκλίνει σε μία στάσιμη συνάρτηση πιθανότητας Πρακτικά, αν μπορεί να κατασκευαστεί μια μαρκοβιανή αλυσίδα που θα μπορεί να πηγαίνει από την κατάσταση i στη j $\forall i, j \in \Omega$ και αντίθετα, (ανάγωγη), θα μπορεί να πηγαίνει από την i στην i (έμμονη) σε κάποιο πεπερασμένο -μέσο- χρόνο μ_i (θετική) και τέλος θα μπορεί να πηγαίνει από την i στην i για κάθε t (απεριοδική) τότε, η μαρκοβιανή αυτή αλυσίδα θα συγκλίνει σε μία συνάρτηση πιθανότητας π.[8]

Οι βασικοί αλγόριθμοι που θα αναλυθούν είναι ο Metropolis – Hasting και ο δειγματολήπτης Gibbs καθώς και ο συνδυασμός τους (Metropolis-within-Gibbs) [6]-[7].

Δειγματολήπτης Gibbs

Έστω ότι υπάρχει ενδιαφέρον για την λήψη ενός δείγματος από την κατανομή $\pi(\theta)$ που $\theta = (\theta_1, \theta_2, \dots, \theta_d)$. **Αν** υπάρχει γνώση των d πλήρων δεσμευμένων κατανομών, δηλαδή των: $\pi\left(\theta_j \middle| \theta_1, \theta_2, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d\right) για\, j = 1, \dots, d$, **και** είναι εύκολο να παρθούν δείγματα από αυτές τις κατανομές, **τότε** είναι δυνατή η λήψη δείγματος από την $\pi(\theta)$ με τον εξής αλγόριθμο:

Βήμα 0: Αρχικοποίησε το διάνυσμα θ με κάποιες τιμές δηλαδή: $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)}\right)$

Βήμα 1: Πάρε τυχαία δείγματα βάσει των πλήρων δεσμευμένων κατανομών ως εξής:

$$\theta_1{}^{(1)} \sim \pi\left(\theta_1 \middle| \theta_2^{(0)}, \theta_3^{(0)}, \theta_4{}^{(0)}, \dots, \theta_d^{(0)}\right)$$

$$\theta_2{}^{(1)} \sim \pi\left(\theta_2 \middle| \theta_1^{(1)}, \theta_3^{(0)}, \theta_4{}^{(0)}, \dots, \theta_d^{(0)}\right)$$

$$\theta_3{}^{(1)} \sim \pi\left(\theta_3 \middle| \theta_1^{(1)}, \theta_2^{(1)}, \theta_4{}^{(0)}, \dots, \theta_d^{(0)}\right)$$

[1] Εργοδική χαρακτηρίζεται μια διαδικασία που είναι έμμονη, θετική και απεριοδική.

$$\vdots$$

$$\theta_d{}^{(1)} \sim \pi\left(\theta_d \middle| \theta_1^{(1)}, \theta_2^{(1)}, \theta_3{}^{(1)}, \dots, \theta_{d-1}^{(1)}\right)$$

$$\vdots$$

Βήμα $(\boldsymbol{\nu+1})$: Πάρε τυχαία δείγματα βάσει των πλήρων δεσμευμένων κατανομών ως εξής:

$$\theta_1{}^{(\nu+1)} \sim \pi\left(\theta_1 \middle| \theta_2^{(\nu)}, \theta_3^{(\nu)}, \theta_4{}^{(\nu)}, \dots, \theta_d^{(\nu)}\right)$$

$$\theta_2{}^{(\nu+1)} \sim \pi\left(\theta_2 \middle| \theta_1^{(\nu+1)}, \theta_3^{(\nu)}, \theta_4{}^{(\nu)}, \dots, \theta_d^{(\nu)}\right)$$

$$\theta_3{}^{(\nu+1)} \sim \pi\left(\theta_3 \middle| \theta_1^{(\nu+1)}, \theta_2^{(\nu+1)}, \theta_4{}^{(\nu)}, \dots, \theta_d^{(\nu)}\right)$$

$$\vdots$$

$$\theta_d{}^{(\nu+1)} \sim \pi\left(\theta_d \middle| \theta_1^{(\nu+1)}, \theta_2^{(\nu+1)}, \theta_3{}^{(\nu+1)}, \dots, \theta_{d-1}^{(\nu+1)}\right)$$

Όσο το ν αυξάνεται, η μαρκοβιανή αλυσίδα συγκλίνει στην $\pi(\theta)$.

Για παράδειγμα, αν έχει κατασκευαστεί μια αλυσίδα από 500.000 θ, επιλέγεται να απορριφθούν –«καούν»- οι πρώτες 100.000 τιμές επειδή πιστεύεται ότι δεν έχει επιτευχθεί η σύγκλιση, ενώ για τις υπόλοιπες 400.000 τιμές επιλέγονται ανά κάποιο βήμα, έστω 100, ώστε να απαλειφθεί η μεταξύ τους συσχέτιση. Συνεπώς, αναμένεται η δημιουργία ενός τυχαίου δείγματος 4.000 μονάδων από την κατανομή $\pi(\theta)$.

Metropolis – Hasting

Ο αλγόριθμος Metropolis – Hasting, είναι πιο γενικός αλγόριθμός (ο δειγματολήπτης Gibbs αποτελεί ειδική περίπτωση), και για επιτευχθεί η σύγκλιση της μαρκοβιανής αλυσίδας στην κατανομή $\pi(\theta)$, κατά γενικό κανόνα, ακολουθούνται τα εξής βήματα:

Βήμα 0: Αρχικοποίησε το διάνυσμα θ με κάποιες τιμές δηλαδή: $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)}\right)$

Βήμα 1: Πάρε μια προτεινόμενη τιμή φ από την συνάρτηση πρότασης q

$$\varphi \sim q\left(\theta^{(0)}, \cdot\right)$$

Βήμα 2: Υπολόγισε την πιθανότητα αποδοχής

$$\alpha\left(\theta^{(0)},\varphi\right)=min\left\{1,\frac{\pi(\varphi)q\left(\varphi,\theta^{(0)}\right)}{\pi(\theta^{(0)})q(\theta^{(0)},\varphi)}\right\}$$

Βήμα 3: Αν δεχτείς την τιμή τότε

$$\theta^{(1)}\leftarrow\varphi$$

Αλλιώς

$$\theta^{(1)}\leftarrow\theta^{(0)}$$

⋮

Βήμα $\boldsymbol{\nu}$: Πάρε μια προτεινόμενη τιμή φ από την συνάρτηση πρότασης q

$$\varphi\sim q\left(\theta^{(\nu-1)},\cdot\right)$$

Βήμα $(\boldsymbol{\nu+1})$: Υπολόγισε την πιθανότητα αποδοχής

$$\alpha\left(\theta^{(\nu-1)},\varphi\right)=min\left\{1,\frac{\pi(\varphi)q\left(\varphi,\theta^{(\nu-1)}\right)}{\pi(\theta^{(\nu-1)})q(\theta^{(\nu-1)},\varphi)}\right\}$$

Βήμα $(\boldsymbol{\nu+2})$: Αν δεχτείς την τιμή τότε

$$\theta^{(\nu)}\leftarrow\varphi$$

Αλλιώς

$$\theta^{(\nu)}\leftarrow\theta^{(\nu-1)}$$

Η επιλογή της συνάρτησης πρότασης q είναι ένα κρίσιμο και περίπλοκο ζήτημα. Η συνάρτηση αυτή θα επηρεάσει πολύ τον ρυθμό και την ποιότητα σύγκλισης. Για χάρη όμως της απλότητας της παρούσας εργασίας, δεν θα γίνει περεταίρω ανάλυση. Όπως και με τον δειγματολήπτη Gibbs, έτσι και εδώ, η μετέπειτα ενέργειες που ακολουθούνται είναι οι ίδιες:

«κάψιμο»→ επιλογή θ ανά k βήματα.

Metropolis-within-Gibbs

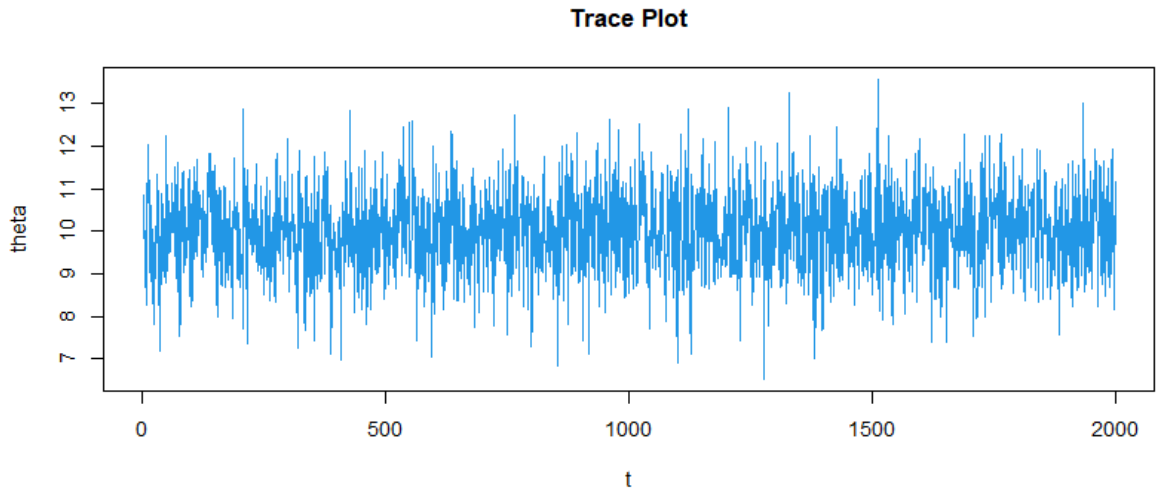
Αυτή η μέθοδος είναι μια υβριδική μέθοδος και απαντάει στο ερώτημα, τι μπορούμε να κάνουμε αν δεν ξέρουμε όλες τις πλήρως δεσμευμένες κατανομές. Η λύση είναι σχετικά αναμενόμενη και αφορά την εφαρμογή του δειγματολήπτη Gibbs ώστε να βρεθούν όσα θ_i επιτρέπουν οι πλήρως δεσμευμένες κατανομές, και για τα υπόλοιπα θ_i γίνεται εφαρμογή του αλγόριθμου Metropolis – Hasting.

Διαγνωστικά κριτήρια σύγκλισης

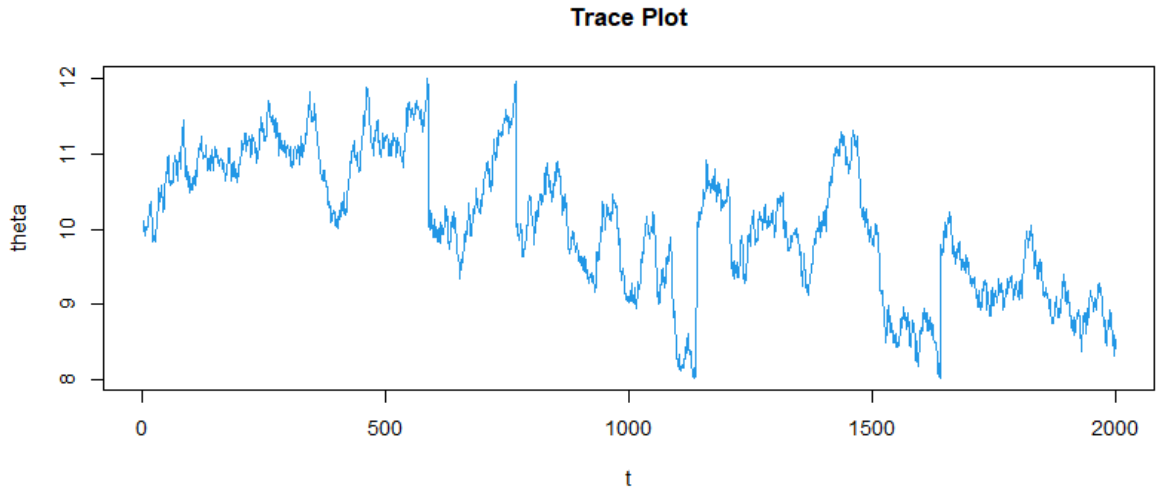
Συνήθως, για να ελεγχθεί η ποιότητα της σύγκλισης υπολογίζονται και σχεδιάζονται δύο γραφήματα και ένα μέτρο, τα οποία παρουσιάζονται στη συνέχεια. Για την καλύτερη επεξήγηση αυτών, θεωρείται ένα σχήμα MCMC που υλοποιήθηκε με σκοπό να προσεγγιστεί μια posterior κατανομή $p(\theta|x)$. Επίσης, θεωρείται ότι έγιναν 1,000,000 επαναλήψεις με απόρριψη («κάψιμο») τα πρώτα 500,000 και επιλογή ανά 250 τιμές. Συνεπώς έχει δημιουργηθεί ένα «δείγμα» από 2,000 θ_t.

Trace plot

Το πρώτο γράφημα που παρουσιάζεται είναι το trace plot και δεν είναι τίποτε άλλο από ένα σύστημα κάθετων αξόνων που στον οριζόντιο άξονα είναι ο χρόνος t και στον κάθετο άξονα είναι η τιμή του θ. Σε ένα τέτοιο σχήμα, η ιδανική του μορφή θα ήταν μία ταλάντωση του θ γύρω από μία οριζόντια ευθεία με κάποιο σταθερό θόρυβο (βλ. Εικόνα 2). Από την άλλη πλευρά, ένα εμφανές πρόβλημα θα ήταν αν είτε η αλυσίδα να μην είχε σταθεροποιηθεί, είτε η αλυσίδα να «κολλάει» σε διαφορετικές ευθείες και να μην μπορεί να επιστρέψει στην επιθυμητή της ευθεία (βλ. Εικόνα 3).



Εικόνα 2 Αναπαράσταση επιθυμητού Trace Plot αλυσίδας MCMC



Εικόνα 3 Αναπαράσταση Trace Plot προβληματικής σύγκλισης αλυσίδας MCMC

Διάγραμμα Αυτοσυσχετίσεων

Υπενθύμιση

Γραμμικός Συντελεστής Συσχέτισης του Pearson

Ο γραμμικός συντελεστής συσχέτισης του Pearson, ή εν συντομία ρ, δείχνει τη γραμμική σχέση που έχουν 2 μεταβλητές και παίρνει τιμές στο [-1,1]. Όταν $|\rho| = 1$ τότε οι δύο αυτές μεταβλητές έχουν τέλεια γραμμική σχέση (τα σημεία βρίσκονται σε ευθεία γραμμή), ενώ όταν $|\rho| = 0$ τότε οι δύο αυτές μεταβλητές δεν συσχετίζονται καθόλου γραμμικά. Έστω X, Y δύο τυχαίες μεταβλητές τότε:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Και δειγματικά:

$$\rho_{X,Y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

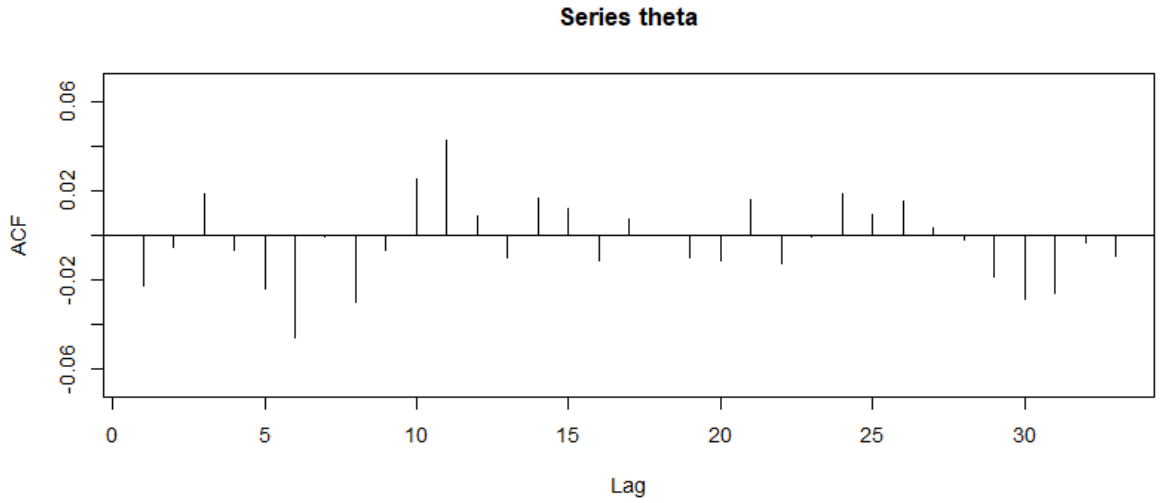
Το δεύτερο γράφημα που παρουσιάζεται είναι το διάγραμμα αυτοσυσχετίσεων και αναπαρίσταται συνήθως ως ένα σχήμα με 2 άξονες όπου στον οριζόντιο άξονα είναι οι χρονικές υστερήσεις, και στον κάθετο άξονα είναι η αυτοσυσχέτιση.

Η αυτοσυσχέτιση δεν είναι τίποτε άλλο παρά ο συντελεστής του Pearson της ίδιας μεταβλητής για διάφορες χρονικές υστερήσεις. Για παράδειγμα αν πρέπει να υπολογιστεί η αυτοσυσχέτιση μίας μεταβλητής, έστω $X_i, i = 1, \dots, 100$, για χρονική υστέρηση 1 τότε πρακτικά δημιουργούνται 2 καινούριες μεταβλητές ως εξής:

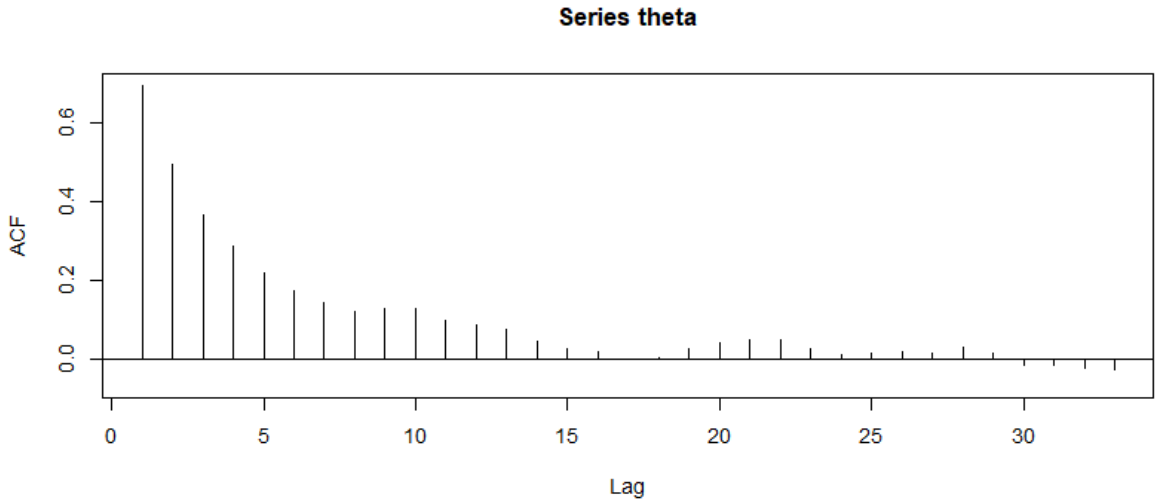
$$\begin{cases} Y_1 = X_i, i = 1, \dots, 99 \\ Y_2 = X_i, i = 2, \dots, 100 \end{cases}$$

Σε αυτές υπολογίζεται το ρ_{Y_1,Y_2} που είναι πρακτικά η αυτοσυσχέτιση σε χρονική υστέρηση 1

Στο γράφημα αυτοσυσχέτισης μίας αλυσίδας που δεν παρουσιάζει πρόβλημα αυτοσυσχέτισης, αναμένεται για κάθε χρονική υστέρηση, η αυτοσυσχέτιση να είναι πολύ κοντά στο μηδέν (βλ. Εικόνα 4). Αντίθετα σε ένα γράφημα αυτοσυσχέτισης μίας αλυσίδας που παρουσιάζει πρόβλημα αυτοσυσχέτισης, αναμένεται να υπάρχουν αυτοσυσχετίσεις αρκετά μακριά από το 0 και κοντά στα -1 ή 1 (βλ. Εικόνα 5).



Εικόνα 4 Διάγραμμα αυτοσυσχετίσεων για αλυσίδα χωρίς πρόβλημα αυτοσυσχέτισης



Εικόνα 5 Διάγραμμα αυτοσυσχετίσεων για αλυσίδα με πρόβλημα αυτοσυσχέτισης

Effective Sample Size

Αυτό το μέτρο «διορθώνει» το πλήθος του δείγματος που προκύπτει μετά το MCMC βάσει των συσχετίσεων. Δηλαδή, αν για παράδειγμα υπάρχουν παρατηρήσεις συσχετισμένες, τότε επειδή προσφέρουν παρόμοια πληροφορία, θα ισοδυναμούσαν με ανεξάρτητο δείγμα μικρότερου μεγέθους. Αυτό το μέγεθος του ισοδύναμου ανεξάρτητου δείγματος είναι το effective sample size [6]-[7]. Δηλαδή, το ESS δείχνει το μέγεθος της ακρίβειας της εκτίμησης κάποιας ποσότητας, π.χ. σε ένα δείγμα από MCMC μεγέθους 2,000 με ESS=100, δείχνει ότι

η ακρίβεια της εκτίμησης ισοδυναμεί με την ακρίβεια ενός ανεξάρτητου δείγματος μεγέθους 100.

Το ESS υπολογίζεται με τον εξής τύπο:

$$ESS = \frac{N}{1 + 2\sum_{t=1}^{\infty} \rho_t} \tag{3.1}$$

όπου N = το πλήθος του συσχετισμένου δείγματος και ρ_t = η αυτοσυσχέτιση σε lag=t

Στην πράξη, επειδή δεν μπορεί να υπολογιστεί το άπειρο άθροισμα του παρανομαστή, από τη βιβλιογραφία έχουν προταθεί διάφοροι τύποι που εκτιμούν τον παρανομαστή του (3.1). Στην παρούσα εργασία, για τον υπολογισμό του ESS χρησιμοποιήθηκε στην R η εντολή effectiveSize από τη βιβλιοθήκη coda, που για την εκτίμηση του αθροίσματος αυτού βασίζεται στην σε ένα AR(p) υπόδειγμα.

ΚΕΦΑΛΑΙΟ 4

ΤΟ ΥΠΟΔΕΙΓΜΑ HELIGMAN – POLLARD

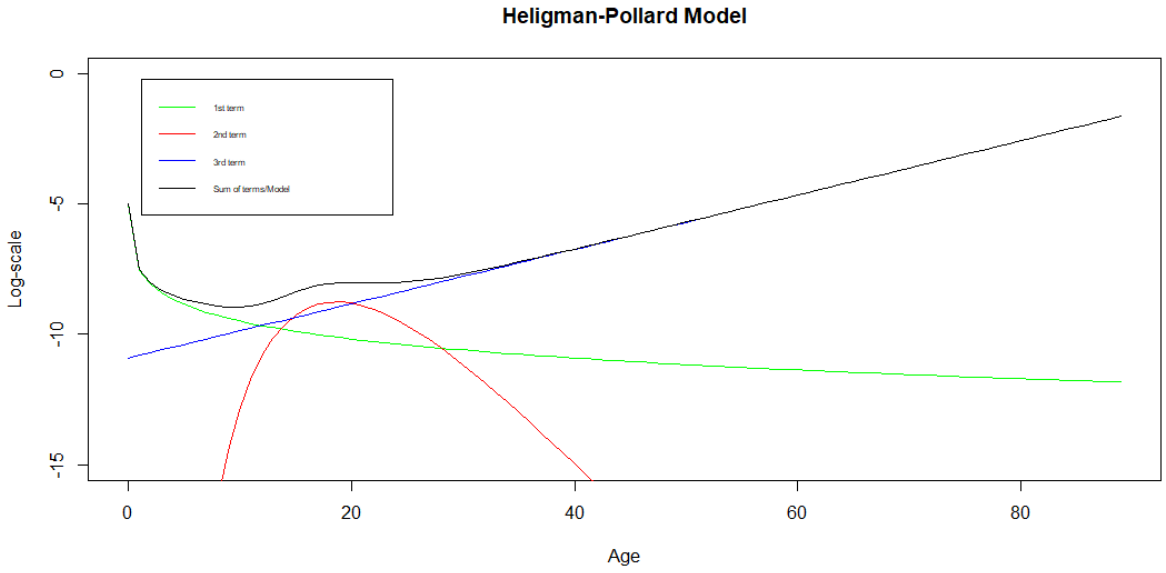
Το υπόδειγμα Heligman – Pollard, είναι μη γραμμικό και μοντελοποιεί τα odds των πιθανοτήτων θανάτου. Συγκεκριμένα, χωρίζεται σε 3 όρους [10]:

- ο $1^{ος}$ όρος αφορά την καμπύλη της παιδικής θνησιμότητας
- ο $2^{ος}$ όρος αφορά την απότομη αύξηση της νεανικής θνησιμότητας (accident hump)
- ο $3^{ος}$ όρος αφορά την θνησιμότητα των ενηλίκων.

Ο τύπος αναλυτικά φαίνεται παρακάτω (που z είναι η ηλικία):

$$\frac{\mathrm{p_z}}{1-\mathrm{p_z}} = \mathrm{A}^{(\mathrm{z}+\mathrm{B})^{\mathrm{C}}} + \mathrm{De}^{-\mathrm{E}\left(\log\left(\frac{\mathrm{z}}{\mathrm{F}}\right)\right)^2} + \mathrm{G}H^{z} \qquad (4.1)$$

Η επίδραση του κάθε όρου φαίνεται πολύ καθαρά στην εικόνα 6 :



Εικόνα 6 Το μοντέλο H-P και τα γραφήματα των τριών όρων που το αποτελούν

Και η επεξήγηση των παραμέτρων [2] της εξισωσης 4.1 παρουσιάζεται συνοπτικά στον παρακάτω πίνακα:

Πίνακας 1 Επεξήγηση μεταβλητών του μοντέλου H-P

Παράμετρος	**Πεδίο Ορισμού**	**Επεξήγηση**
A	(0,1)	Βρεφικός συντελεστής θνησιμότητας
B	(0,1)	Συντελεστής θνησιμότητας $1^{ος}$ έτους
C	(0,1)	Σχετίζεται με το ρυθμό πτώσης του συντελεστή θνησιμότητας
D	(0,1)	Αντιπροσωπεύει την σοβαρότητα του accident hump
E	(0,∞)	Συνδέεται με την έκταση του accident hump (όσο μεγαλύτερο, τόσο πιο στενή η συγκέντρωση)
F	(15,110)	Αφορά τη θέση του accident hump
G	(0,1)	Αφορά τη θέση/ύψος του της ενήλικης θνησιμότητας
H	(0,∞)	Σχετίζεται με το ρυθμό αύξησης τους συντελεστή θνησιμότητας των ενηλίκων

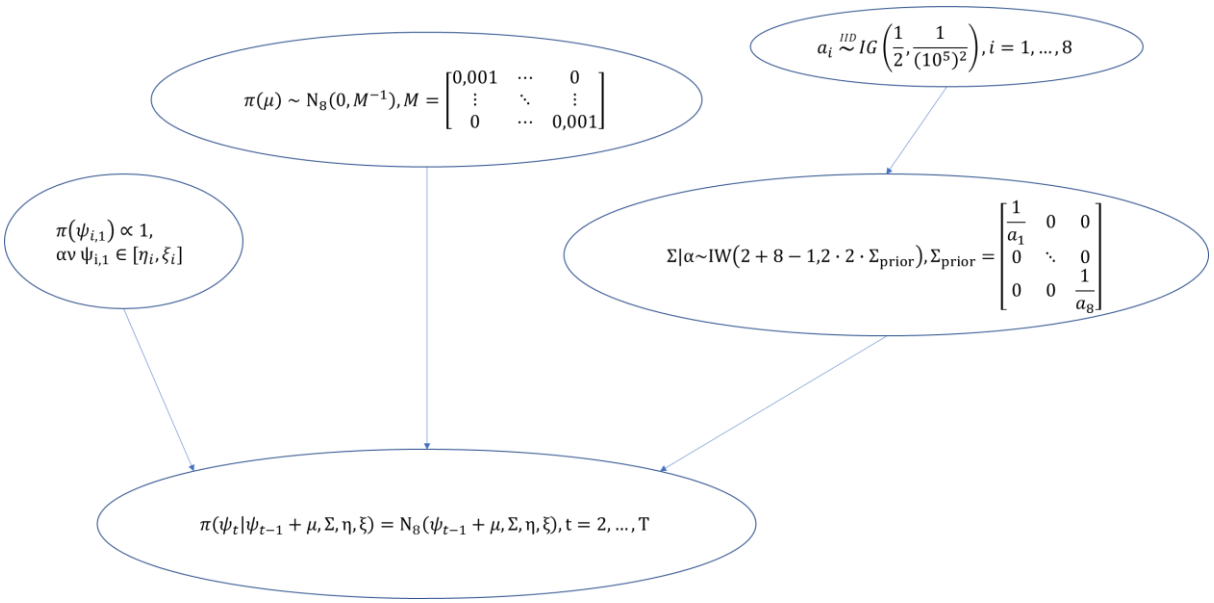
Το συγκεκριμένο μοντέλο παρόλο που χρησιμοποιείται ευρέως, πάσχει από υπερπαραμετροποίηση. Αυτό έχει ως αποτέλεσμα την έντονη αριθμητική αστάθεια των εκτιμήσεων των παραμέτρων. Για να λυθεί το πρόβλημα αυτό, οι Heligman – Pollard πρότειναν τη χρησιμοποίηση σταθμισμένων ελαχίστων τετραγώνων με βάρη $w_z = \frac{1}{m_z^2}$ που m_z είναι οι κατά ηλικία συντελεστές θνησιμότητας των δεδομένων. Επιστημονικά ενδιαφέρον, αποτελεί το γεγονός ότι δεν έχει βρεθεί άλλος τρόπος να καθοριστούν τα βάρη και να επιτευχθεί σύγκλιση όπως έχουν δείξει οι [12] και [14]. Η μόνη διαφορετική προσέγγιση είναι των [13]που διατήρησαν 2 παραμέτρους σταθερές.

ΚΕΦΑΛΑΙΟ 5

ΤΟ ΔΥΝΑΜΙΚΟ ΥΠΟΔΕΙΓΜΑ HELIGMAN – POLLARD

Σε αυτό το σημείο θα αναλυθεί η θεωρία για το Δυναμικό Υπόδειγμα Heligman – Pollard όπως αυτή προτάθηκε από το [3].

Οι παράμετροι του Heligman – Pollard μοντέλου μετασχηματίζονται ώστε να είναι πιο κοντά στην κανονικότητα, [11]. Η μέθοδος αυτή βασίζεται και υποθέτει ένα ιεραρχικό μπεϋζιανό μοντέλο, όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 7 Η δομή που χρησιμοποιεί το Dynamic H-P

Το $\psi_{i,t}$ είναι η παράμετρος i από το διάνυσμα των 8 παραμέτρων του μοντέλου στον χρόνο t.

Τα δεδομένα που είναι οι θάνατοι (d) και ο πληθυσμός σε ρίσκο (n) για κάθε ηλικιακή ομάδα και κάθε προηγούμενη χρονιά, ακολουθούν διωνυμική κατανομή με κάποια πιθανότητα επιτυχίας που μεταβάλλεται, δηλαδή: $d|p \sim Binom(n, p)$

Οι υπερπαράμετροι συγκεντρώνονται στο διάνυσμα $\theta = (\Sigma, \mu, \alpha)$ και η posterior κατανομή γράφεται ως εξής:

$$\pi(\theta, \psi|d, \eta, \xi) \propto f(d|\psi) \prod_{t=2}^{T} \varphi_8(\psi_t; \psi_{t-1} + \mu, \Sigma, \eta, \xi)\, \pi(\theta) \tag{5.1}$$

όπου:

$$f(d|\psi) = \prod_{t=1}^{T}\prod_{z=0}^{89}\binom{n_{z,t}}{d_{z,t}}K(z,\psi_t)^{d_{z,t}}\left(1+K(z,\psi_t)\right)^{-n_{z,t}} \quad (5.2)$$

και

$$\pi(\theta) \propto \varphi_8(\mu)\pi(\Sigma|\alpha)\pi(\alpha) \quad (5.3)$$

Η «δειγματοληψία» από την posterior κατανομή γίνεται με την μέθοδο του MCMC Metropolis -within- Gibbs, διότι είναι δυνατόν να παρθούν δείγματα από τις πλήρες δεσμευμένες κατανομές των Σ,μ και α. Όμως, για την πλήρη δεσμευμένη κατανομή των ψ_t θα γίνει εφαρμογή του Metropolis - Hasting αλγορίθμου. Ο αλγόριθμος παρατίθεται αναλυτικά ως εξής [3]:

Βήμα 0: Κάνουμε μη-γραμμικά σταθμισμένα ελάχιστα τετράγωνα για T προηγούμενες χρονιές, με στάθμιση $w_{z,t} = \frac{1}{q_{z,t}^2}$ στο μοντέλο των Heligman - Pollard, όπως δηλαδή αναφέρθηκε στο Κεφάλαιο 4, και παίρνουμε τα m_t,V_t που είναι οι εκτιμήσεις των παραμέτρων και ο πίνακας διακύμανσης – συνδιακύμανσης των παραμέτρων αντίστοιχα.

Βήμα 1: Αρχικοποίησε:

$\psi_t^{(0)} \leftarrow m_t, για\ t = 1, \dots, T$

$\alpha_j^{(0)} \sim \frac{1}{G\left(\frac{1}{2}, 10^{10}\right)}, για\ j = 1, \dots, 8$

$\Sigma^{(0)} \sim IW(9, I_8)$

$\mu^{(0)} \leftarrow [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^T$

Βήμα 2: Για κάθε iteration i επανέλαβε:

$$\alpha_j^{(i)} \sim \frac{1}{G\left(\frac{1}{2}, 10^{-10} + 2\left(\Sigma_{jj}^{(i-1)}\right)^{-1}\right)}, για\ j = 1, \dots, 8$$

$$\mu^{(i)} \sim N_8\left(\left(M + (T-1)\left(\Sigma^{(i-1)}\right)^{-1}\right)^{-1}(\psi_T - \psi_1), \left(M + (T-1)\left(\Sigma^{(i-1)}\right)^{-1}\right)^{-1}\right)$$

$$\Sigma^{(i)} \sim IW(T+8, 4\Sigma_{prior} + \sum_{t=2}^{T}(\psi_t - \psi_{t-1} - \mu)(\psi_t - \psi_{t-1} - \mu)')$$

Βήμα α: Για κάθε προηγούμενη χρονιά $t = 1, \dots, T$ κάνε πρόταση

$$\psi_{prop} \sim N_8\left(\psi_t^{(i-1)}, c_t V_t\right)$$

που c_t είναι μια προκαθορισμένη σταθερά που έχει σκοπό την εξασφάλιση καλύτερης σύγκλισης

Βήμα β: Αν $t = 1$ τότε:

$$\pi(\cdot) \sim N_8\left(\psi^{(i-1)}{}_{t+1} - \mu, \Sigma; \eta, \xi\right)$$

Αν $t = T$ τοτε:

$$\pi(\cdot) \sim N_8\left(\psi^{(i-1)}{}_{t-1} + \mu, \Sigma; \eta, \xi\right)$$

Αν $1 < t < T$ τότε:

$$\pi(\cdot) \sim N_8\left(\frac{1}{2}\left(\psi^{(i-1)}{}_{t-1} + \psi^{(i-1)}{}_{t+1}\right), \frac{1}{2}\Sigma; \eta, \xi\right)$$

Βήμα γ: Δέξου το $prop$ με πιθανότητα:

$$\alpha = \frac{f\left(d|\psi_{prop}\right)\pi\left(\psi_{prop}\right)}{f\left(d|\psi_t{}^{(i-1)}\right)\pi\left(\psi_t{}^{(i-1)}\right)}$$

Για παράδειγμα, εφαρμόζεται ο αλγόριθμος για 2.700.000 επαναλήψεις, ενώ απορρίπτονται -«καίγονται»- οι 900.000 πρώτοι παράμετροι και στη συνέχεια επιλέγονται οι υπόλοιπες ανά κάθε 900 «παρατηρήσεις. Έτσι προκύπτει ένα ανεξάρτητο δείγμα από τις επιθυμητές κατανομές, με την προϋπόθεση πάντα ότι το MCMC συγκλίνει.

Για τα βρεθεί η κατανομή πρόβλεψης των παραμέτρων ψ_t, δηλαδή η κατανομή $p(\psi_{T+1}|d)$, θα πρέπει να υπολογιστεί το παρακάτω ολοκλήρωμα:

$$p(\psi_{T+1}|d) = \int p(\psi_{T+1}|\psi_T, d)\pi(\psi, \theta|d)\, d\psi d\theta \qquad (5.4)$$

Προφανώς το ολοκλήρωμα αυτό δεν υπολογίζεται αναλυτικά, αλλά προσεγγίζεσαι από το δείγμα μεγέθους M της posteriorως εξής:

Για κάθε i από 1 μέχρι M πάρε:

$$\psi_{T+1}{}^{(i)} \sim N_8\left(\psi_T^{(i)} + \mu^{(i)}, \Sigma^{(i)}, \eta, \xi\right) \tag{5.5}$$

Αντιστοίχως, αν χρειάζεται να παρθεί δείγμα από την $T+2$ θα πρέπει πρώτα να υπολογιστεί το δείγμα από την $T+1$ και για κάθε i από 1 μέχρι M με την χρήση της 5.6:

$$\psi_{T+2}{}^{(i)} \sim N_8\left(\psi_{T+1}^{(i)} + \mu^{(i)}, \Sigma^{(i)}, \eta, \xi\right) \tag{5.6}$$

Με τον ίδιο τρόπο υπολογίζονται τα δείγμα για κάθε $T+k$.

ΚΕΦΑΛΑΙΟ 6

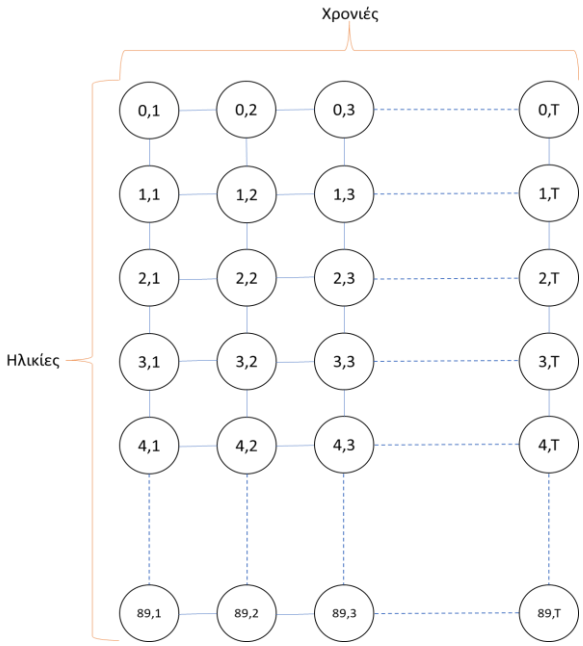
GAUSSIAN MARKOV RANDOM FIELD (GMRF)

Σε αυτό το κεφάλαιο θα αναλυθεί η θεωρία του GMRF και θα συσχετιστεί με την εφαρμογή της στη μπευζιανή στατιστική για τις προβλέψεις των πιθανοτήτων θανάτου [3].

Η βασική ιδέα είναι η εξής: Τα δεδομένα ανήκουν σε ένα διάνυσμα, έστω $x = (x_1, x_2, \dots, x_p)^T$, που ακολουθεί πολυμεταβλητή κανονική κατανομή με πίνακα ακρίβειας $Q = \Sigma^{-1}$, που εκτός από τα διαγώνια στοιχεία, περιέχει και κάποια επιπλέον μη μηδενικά στοιχεία πράγμα που υποδηλώνουν υπό συνθήκη εξάρτηση μεταξύ των αντίστοιχων μεταβλητών. Για να γίνει πρόβλεψη, απλά επεκτείνεται το διάνυσμα, έστω $x = (x_1, x_2, \dots, x_p, x_{p+1}, \dots, x_{p+k})^T$, καθώς και ο πίνακας Q με την ίδια πάντα δομή και έπειτα γίνεται η δέσμευση της μορφής: $x_{p+1}, \dots, x_{p+k} | \mathrm{x}_1, \dots, \mathrm{x}_\mathrm{p}$. Η κατανομή της δέσμευσης είναι πάλι κανονική, και μάλιστα γνωστής μορφής οπότε γίνεται εύκολα διαχειρίσιμη [5].

Γράφοι

Γράφος $G(V, E)$ ονομάζεται ένα σύνολο από κορυφές/κόμβους V και ακμές/συνδέσεις E που μοιάζει πρακτικά σαν ένα δίκτυο. Οι ακμές μπορεί να έχουν κατευθύνσεις ή και βάρη. Οι ακμές επίσης μπορούν να επιστρέφουν στον ίδιο κόμβο, ενώ δύναται να υπάρχουν κόμβοι που να είναι πλήρως ασύνδετοι. Για το πρόβλημα που πραγματεύεται η παρούσα εργασία, ο γράφος που χρησιμοποιείται παρουσιάζεται στην εικόνα 8. **Ο κάθε κόμβος είναι η πιθανότητα θανάτου** και οι **συνδέσεις** είναι πρακτικά οι «συσχετίσεις» που φανερώνουν τις **υπό συνθήκη εξαρτήσεις**. [5].



Εικόνα 8 Ο γράφος που θα μας απασχολήσει για το GMRF

Θεωρία των GMRF

Ένα τυχαίο διάνυσμα $x \in \mathbb{R}^n$ λέγεται GMRF σύμφωνα με τον γράφο $G = (V, E)$ αν και μόνο αν

$$x \sim N_n(\mu, Q^{-1}) \tag{6.1}$$

Και

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E \ \forall i \neq j \tag{6.2}$$

Τότε ξέρουμε ότι

$$E(x_i | x_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_j Q_{ij}(x_j - \mu_j) \tag{6.3}$$

$$Prec(x_i | x_{-i}) = Q_{ii} \tag{6.4}$$

Και από θέμα συνόλων

$$\mathrm{x} = \begin{pmatrix} \mathrm{x}_A \\ x_B \end{pmatrix}, \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, Q = \begin{pmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{pmatrix}$$

$$\mu_{A|B} = \mu_A - Q_{AA}^{-1} Q_{AB}(x_B - \mu_B)$$

$$Q_{A|B} = Q_{AA}$$

Και επειδή $x \sim N_n(\mu, Q^{-1})$, τότε $x_A|x_B \sim N\left(\mu_{A|B}, Q_{AA}^{-1}\right)$ [5].

Επίσης, αξίζει να σημειωθεί το αποτέλεσμα ενός θεωρήματος [5] που δίνει τη δυνατότητα να ορισθεί ανάποδα το GMRF ως εξής:

Δοθέντος n κανονικών πλήρων δεσμευμένων κατανομών με μέση τιμή και ακρίβεια

$$\boldsymbol{E}(\boldsymbol{x_i}|\boldsymbol{x_{-i}}) = \boldsymbol{\mu_i} - \sum_{\boldsymbol{j}} \boldsymbol{\beta_{ij}}\left(\boldsymbol{x_j} - \boldsymbol{\mu_j}\right) \tag{6.5}$$

$$Perc(x_i|x_{-i}) = \kappa_i > 0 \tag{6.6}$$

αντίστοιχα, τότε το x είναι GMRF σύμφωνα με τον γράφο $G = (V, E)$ με μέση τιμή μ, και πίνακα ακρίβειας Q:

$$\boldsymbol{Q} = \boldsymbol{Q}_{ij} = \begin{cases} \boldsymbol{\kappa_i \beta_{ij}}, \boldsymbol{i \neq j} \\ \boldsymbol{\kappa_i}, \boldsymbol{i = j} \end{cases}$$

δοθέντος ότι $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}, i \neq j$ και $Q > 0$

Τέλος ένα χρήσιμο -για την κατανόηση- αποτέλεσμα που αφορά GMRF είναι το εξής [5]:

Έστω x ένα GMRF σύμφωνα με τον γράφο $G = (V, E)$,τότε οι τοπικές, ολικές και κατά ζεύγη μαρκοβιανές ιδιότητες είναι ισοδύναμες. Αυτός είναι ουσιαστικά ο λόγος που υπάρχει ο όρος «Markov» στο GMRF.

Intrinsic GMRF

Το Intrinsic GMRF ή εν συντομία IGMRF είναι μια ειδική περίπτωση του GMRF με την διαφορά ότι ο πίνακας Q είναι πλέον ένας **μη αντιστρέψιμος πίνακας**. Το IGMRF έχεις πολλές χρησιμότητες, μία από αυτές είναι ότι συνηθίζεται να είναι μια improper prior. Όσον αφορά την -ακατάλληλη- κατανομή της, αν το x είναι IGMRF τότε είναι της μορφής

$$\pi(x) = (2\pi)^{\frac{-(n-k)}{2}} (|Q|^*)^{\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T Q (x-\mu)} \tag{6.7}$$

που $|Q|^*$ η γενικευμένη ορίζουσα, δηλαδή το γινόμενο των μη μηδενικών ιδιοτιμών. Πρακτικά είναι μία **σχεδόν** πολυμεταβλητή κανονική κατανομή.

IGMRF τάξης k θα λέμε ένα IGMRF που ο πίνακας Q είναι τάξης $n - k$.

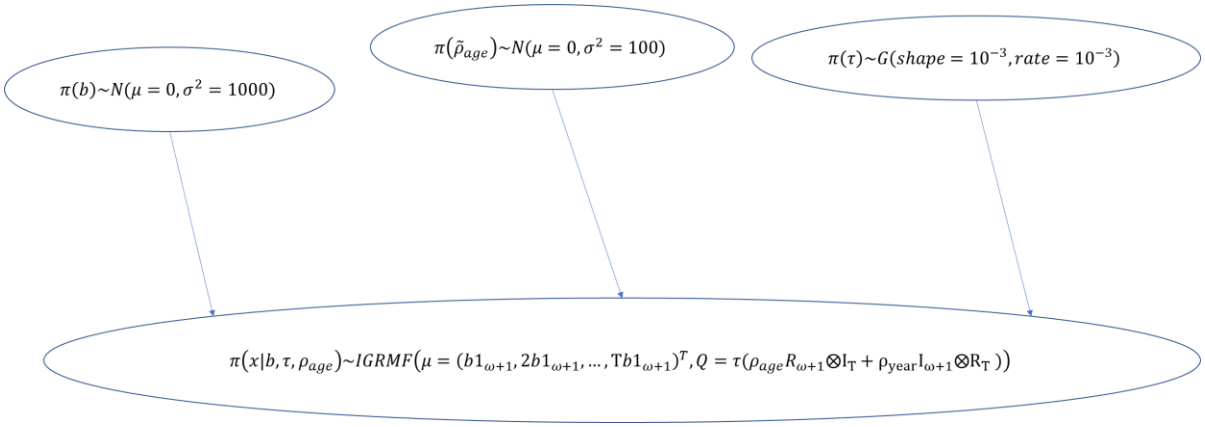
Ένας τρόπος που μπορεί να μας οδηγήσει από ένα GMRF σε ένα IGMRF είναι να δεσμεύσουμε το GMRF με γραμμικούς περιορισμούς, κάτι που δεν θα γίνει στην παρούσα εργασία.

ΚΕΦΑΛΑΙΟ 7

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ IGMRF

Σε αυτό το κεφάλαιο θα αναλυθεί η θεωρία για τη μοντελοποίηση βάσει ενός IGMRF όπως αυτή προτάθηκε από τον [3].

Αρχικά, η κάθε πιθανότητα θανάτου, για κάθε ηλικία και κάθε χρονιά, μετασχηματίζεται με ένα logit μετασχηματισμό, δηλαδή $x_{z,t} = log\left(\frac{p_{z,t}}{1-p_{z,t}}\right)$. Η μέθοδος που θα περιγραφεί βασίζεται σε ένα ιεραρχικό μπεϋζιανό μοντέλο, όπως φαίνεται στο σχήμα παρακάτω [3]:



Εικόνα 9 Η ιεραρχική δομή της μεθόδου με το IGMRF υπόδειγμα

Η επεξήγηση των βασικών συμβόλων του μπεϋζιανού μοντέλου (Εικ. 9) είναι η ακόλουθη:

- ρ_{age} και ρ_{year} είναι τα μέτρα που «πιάνουν» τις συσχετίσεις των πιθανοτήτων θανάτου μεταξύ των ηλικιών και των χρονιών αντίστοιχα. Για αυτές τις ποσότητες γίνεται η υπόθεση ότι:

  $$\rho_{age} + \rho_{year} = 2$$

  Συνεπώς αρκεί να υπάρχει γνώση μίας από τις δύο παραμέτρους, έστω της ρ_{age}. Αυτή η παράμετρος πλέον μετασχηματίζεται ως εξής:

  $$\tilde{\rho}_{age} = log(\rho_{age}) - log(2 - \rho_{age})$$

- Ο πίνακας R_t είναι τριδιαγώνιος τετραγωνικός $(t \times t)$ της μορφής:

$$R = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

- Ο πίνακας Q που προκύπτει είναι τετραγωνικός ($90T \times 90T$) και είναι της μορφής:

$$Q = \tau \cdot \begin{bmatrix} \rho_{age} + \rho_{year} & -\rho_{year} & 0 & \cdots & -\rho_{age} & 0 & 0 & \cdots & 0 \\ -\rho_{year} & \rho_{age} + 2\rho_{year} & -\rho_{year} & \cdots & 0 & -\rho_{age} & 0 & \cdots & 0 \\ 0 & -\rho_{year} & \rho_{age} + 2\rho_{year} & \cdots & 0 & 0 & -\rho_{age} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 & 0 & \ddots & \vdots \\ -\rho_{age} & 0 & 0 & 0 & \ddots & 0 & 0 & 0 & -\rho_{age} \\ 0 & -\rho_{age} & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & -\rho_{age} & 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & -\rho_{age} & 0 & 0 & \cdots & \rho_{age} + \rho_{year} \end{bmatrix}$$

Προφανώς τα δεδομένα, που είναι οι θάνατοι (d) και ο πληθυσμός σε ρίσκο (n) για κάθε ηλικιακή ομάδα και κάθε προηγούμενη χρονιά, ακολουθούν διωνυμική κατανομή με κάποια πιθανότητα επιτυχίας που μεταβάλλεται, δηλαδή: $d|p \sim Binom(n, p)$.

Οι υπερπαραμέτροι συγκεντρώνονται στο διάνυσμα $\theta = (\mathrm{b}, \tilde{\rho}_{age}, \mathrm{t})$ και η posterior κατανομή γράφεται ως εξής:

$$\pi(\theta, x|d) \propto f(d|x)\pi(x|\theta)\pi(\theta) \tag{7.1}$$

με:

$$f(d|p) = \prod_{t=1}^{T} \prod_{z=0}^{89} \binom{n_{z,t}}{d_{z,t}} {p_{z,t}}^{d_{z,t}} \left(1 - p_{z,t}\right)^{n_{z,t} - d_{z,t}} \tag{7.2}$$

και

$$\pi(x|\theta) \propto IGMRF \tag{7.3}$$

και

$$\pi(\theta) \propto \pi(b)\pi\left(\tilde{\rho}_{age}\right)\pi(\tau) \tag{7.4}$$

Ο τρόπος με τον οποίο θα γίνει η «δειγματοληψία» από την posterior είναι με την μέθοδο του MCMC Metropolis -within- Gibbs, διότι είναι δυνατόν να παρθούν δείγματα από τις πλήρες δεσμευμένες κατανομές των b και τ. Όμως, για την πλήρη δεσμευμένη κατανομή των $\tilde{\rho}_{age}$ και $x_{i,t}$ θα γίνει εφαρμογή του Metropolis - Hasting αλγορίθμου. Σε αυτό το σημείο

πρέπει να τονιστεί ότι για την πλήρη δεσμευμένη κατανομή του $x_{i,t}$, δεν χρησιμοποιείται ο Metropolis – Hasting αλγόριθμος, αλλά μία παραλλαγή αυτού. Στην πραγματικότητα εισάγεται μία βοηθητική μεταβλητή η οποία χρησιμοποιεί την κατευθυνόμενη παράγωγο της λογαριθμικής πιθανοφάνειας των δεδομένων [19] ώστε να επιτευχθεί καλύτερη σύγκλιση.

Ο αλγόριθμος παρατίθεται αναλυτικά ως εξής [3]:

Βήμα 0: Θέσε:

$p = \frac{deaths}{number\ of\ people}$ για κάθε ηλικία $x = 1, \dots, \omega$ και κάθε χρονιά $t = 1, \dots, T$

Δηλαδή p είναι ένα διάνυσμα μήκους $(\omega + 1) \times T$

$L = [1 \quad 2 \quad \dots \quad T \quad 1 \quad 2 \quad \dots \quad T \quad \dots \quad 1 \quad 2 \quad \dots \quad T]^T$

Δηλαδή L είναι ένα διάνυσμα μήκους $(\omega + 1) \times T$

Βήμα 1: Αρχικοποίησε:

$$z^{(0)} = \log\left(\frac{p}{1-p}\right)$$

$$b^{(0)} = 0$$

$$\tilde{\rho}_{age}^{(0)} = \log\left(\frac{0.5}{2-0.5}\right), \delta\eta\lambda\alpha\delta\acute{\eta}\ \rho_{age} = 0.5$$

$$Q^{(0)} = \rho_{age} R_{\omega+1} \otimes \mathrm{I_T} + (2 - \rho_{age}) \mathrm{I}_{\omega+1} \otimes \mathrm{R_T}$$

Βήμα 2: Για κάθε iteration i επανέλαβε:

$$\tau^{(i)} \sim G\left(\frac{T(\omega+1) + 0.002}{2}, \frac{0.002 + \left(z^{(i-1)} - b^{(i-1)}L\right)^T Q^{(i-1)} \left(z^{(i-1)} - b^{(i-1)}L\right)}{2}\right)$$

$$b^{(i)} \sim N\left(\left(\tau^{(i)} L^T Q^{(i-1)} L + 0.001\right)^{-1} L^T \left(\tau^{(i)} Q^{(i-1)}\right) z^{(i-1)}, \left(\tau^{(i)} L^T Q^{(i-1)} L + 0.001\right)^{-1}\right)$$

Κάνε πρόταση για το $\tilde{\rho}_{age}$:

$\tilde{\rho}_{age_{Prop}} = \tilde{\rho}_{age}^{(i-1)} + N(0,0.5^2)$

Άρα: $\rho_{age_{Prop}} = 2\frac{e^{\tilde{\rho}_{age_{Prop}}}}{1+e^{\tilde{\rho}_{age_{Prop}}}}$

Άρα: $Q_{Prop} = \rho_{age_{Prop}} R_{\omega+1} \otimes \mathrm{I}_{\mathrm{T}} + (2 - \rho_{age_{Prop}}) \mathrm{I}_{\omega+1} \otimes \mathrm{R}_{\mathrm{T}}$

Δέξου το $\tilde{\rho}_{age_{Prop}}$ άρα και το Q_{Prop} με πιθανότητα

$$\alpha = \frac{N\left(\tilde{\rho}_{age_{Prop}}; 0,100\right)\sqrt{logdet\left(\rho_{age_{Prop}}\right)}\sqrt{\tau^{(i)}(x-\mu)^T Q^{(i-1)}(x-\mu)}}{N\left(\tilde{\rho}_{age}^{(i-1)}; 0,100\right)\sqrt{logdet\left(\rho_{age}^{(i-1)}\right)}\sqrt{\tau^{(i)}(x-\mu)^T Q_{Prop}(x-\mu)}}$$

Υπολόγισε την κατευθυνόμενη παράγωγο λογαριθμικής πιθανοφάνειας:

$$derivative_z^{(i)} = \frac{deaths}{1+e^{z^{(i-1)}}} - \frac{number\ of\ people - deaths}{1+e^{-z^{(i-1)}}}$$

Υπολόγισε τη βοηθητική μεταβλητή

$$u_x^{(i)} \sim N_{(\omega+1)T}\left(z^{(i-1)} + \frac{\delta}{2} derivative_z^{(i)}, \frac{\delta}{2} I_{(\omega+1)T}\right)$$

Κάνε πρόταση για το z:

$z_{prop} \sim N_{(\omega+1)T}(M^{-1}b, M^{-1})$

Όπου:

$$M = \frac{2}{\delta} I_{(\omega+1)T} + \tau^{(i)} Q^{(i)}$$

$$b = \frac{2}{\delta} u_x + \tau^{(i)} Q^{(i)} Lb^{(i)}$$

Δέξου το z_{prop} με πιθανότητα:

$$\alpha = \frac{f\left(d|z_{prop}, \theta\right)}{f(d|z,\theta)} e^{g(u_x, z_{prop}) - g(u_x, z)}$$

Όπου:

$$g(u_x, z) = \left(u_x - z - \frac{\delta}{4} derivative_z\right)^T (derivative_z)$$

Για παράδειγμα, αν ο παραπάνω αλγόριθμος εφαρμοστεί για 220.000 επαναλήψεις, τότε απορρίπτονται -«καίγονται»- οι 20.000 πρώτοι παράμετροι και έπειτα επιλέγονται οι

υπόλοιπες ανά κάθε 200 «παρατηρήσεις». Έτσι προκύπτει ένα ανεξάρτητο δείγμα από τις επιθυμητές κατανομές, με την προϋπόθεση πάντα ότι το MCMC συγκλίνει.

Για τα βρεθεί η κατανομή πρόβλεψης των παραμέτρων, δηλαδή την κατανομή $p(x^*|d)$, με $x^* = (x'_{t+1}, x'_{t+2}, \dots, x'_{t+k})'$ θα πρέπει να υπολογιστεί το παρακάτω ολοκλήρωμα:

$$p(x^*|d) = \int p(x^*|x,d)\pi(x,\theta|d)\, dxd\theta \tag{7.5}$$

Προφανώς το ολοκλήρωμα αυτό δεν υπολογίζεται αλλά προσεγγίζεσαι από το δείγμα μεγέθους M της posterior ως εξής:

$L^* = [t+1 \quad t+2 \quad \dots \quad t+k \quad t+1 \quad t+2 \quad \dots t+k \quad \dots \quad t+1 \quad t+2 \quad \dots \quad t+k]^T$

Δηλαδή L^* είναι ένα $(\omega+1)\times k$ διάνυσμα.

Για κάθε i από 1 μέχρι M πάρε:

$$\tilde{Q}^{(i)} = \begin{bmatrix} Q^{(i)} & Q^{**} \\ (Q^{**})^T & Q^* \end{bmatrix}$$

Δηλαδή ο πίνακας εξακολουθεί να έχει την ίδια δομή με τον Q

$\mu^{(i)} = b^{(i)}L$

$\mu^{*(i)} = b^{(i)}L^*$

$$x^{*(i)} \sim N_{(\omega+1)\times k}\left(\mu^{*(i)} - (Q^*)^{-1}Q^{**}\left(x^{(i)} - \mu^{(i)}\right), (Q^*)^{-1}\right)$$

ΚΕΦΑΛΑΙΟ 8

ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ: DYNAMIC H – P

Στο παρόν κεφάλαιο θα γίνει εφαρμογή του Dynamic Heligman – Pollard [3] για τα δεδομένα της Ελλάδας. Τα δεδομένα αυτά, αφορούν θανάτους και πληθυσμούς ανά ηλικία και πάρθηκαν από το Human Mortality Database.

Αρχικά σημειώνεται ότι η εφαρμογή της Dynamic H – P αναμένεται να είναι προβληματική εξαιτίας της υπερπαραμετροποίησης του μοντέλου H – P. Παρά τους πολλούς περιορισμούς από τις prior κατανομές, και το γεγονός ότι λειτούργησε για τα δεδομένα Ηνωμένου Βασιλείου και Ουαλίας, δεν υπάρχει διαβεβαίωση ότι θα λειτουργήσει για κάθε χώρα σε κάθε χρονιά.

Για να ελεγχθεί λοιπόν η «συμπεριφορά» της μεθόδου, αποφασίστηκε να χρησιμοποιηθούν τα συνολικά δεδομένα (άνδρες + γυναίκες) στην περίοδο 1983 με 1992. Έγιναν 2.700.000 επαναλήψεις, με απόρριψη -«κάψιμο»- των 900.000 πρώτων τιμών και επιλογή των υπόλοιπων ανά 900 «παρατηρήσεις» (thining). Η σταθερά για την εξασφάλιση καλύτερης σύγκλισης ρυθμίστηκε στο $0{,}008$ έτσι επιτεύχθηκε πιθανότητα αποδοχής για τα για τα ψ_t:

Πίνακας 2 Πιθανότητες αποδοχής των ψ, για κάθε χρονιά

$$\begin{bmatrix} 0.27 & 0.29 & 0.21 & 0.27 & 0.23 & 0.28 & 0.13 & 0.17 & 0.17 & 0.26 \end{bmatrix}$$

Στο τελικό δείγμα μεγέθους 2,000 υπολογίζονται τα Effective Sample Size για τις παραμέτρους, όπως φαίνεται στους πίνακες παρακάτω.

Πίνακας 3 Effective Sample Size για κάθε στοιχείο του της παραμέτρου (πίνακα) Σ

$$\begin{bmatrix} 511 & 1197 & 1395 & 1142 & 1169 & 1197 & 2000 & 1754 \\ 1197 & 743 & 794 & 2000 & 1570 & 1516 & 2000 & 2000 \\ 1395 & 794 & 1357 & 2000 & 1966 & 1244 & 2000 & 2000 \\ 1142 & 2000 & 2000 & 1654 & 2000 & 1764 & 2000 & 1795 \\ 1169 & 1570 & 1966 & 2000 & 1004 & 1100 & 1462 & 2000 \\ 1197 & 1516 & 1244 & 1764 & 1100 & 265 & 1762 & 2000 \\ 2000 & 2000 & 2000 & 2000 & 1462 & 1762 & 2000 & 1724 \\ 1754 & 2000 & 2000 & 1795 & 2000 & 2000 & 1724 & 1842 \end{bmatrix}$$

Πίνακας 4 Effective Sample Size για κάθε στοιχείο της παραμέτρου μ

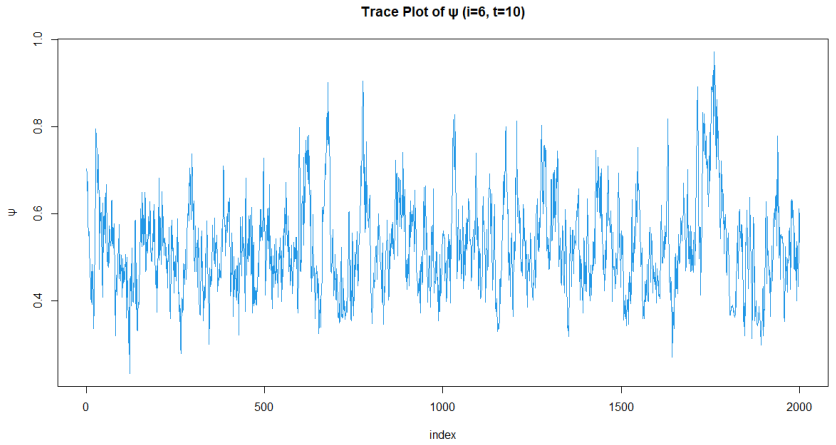
$$[1475 \quad 487 \quad 481 \quad 2000 \quad 1489 \quad 762 \quad 1791 \quad 1625]$$

Πίνακας 5 Effective Sample Size για τις παραμέτρους ψi,t, για κάθε χρονιά t (γραμμή) και κάθε στοιχείο i (στήλη)

$$\begin{bmatrix} 400 & 231 & 228 & 882 & 439 & 567 & 1446 & 1629 \\ 288 & 199 & 205 & 699 & 395 & 477 & 1317 & 1462 \\ 338 & 184 & 176 & 868 & 426 & 571 & 1116 & 1220 \\ 301 & 175 & 168 & 770 & 500 & 590 & 660 & 559 \\ 368 & 162 & 155 & 938 & 465 & 584 & 937 & 973 \\ 287 & 203 & 193 & 780 & 529 & 394 & 946 & 1012 \\ 283 & 190 & 181 & 876 & 452 & 342 & 716 & 760 \\ 257 & 207 & 222 & 855 & 283 & 281 & 408 & 412 \\ 232 & 186 & 219 & 826 & 245 & 158 & 350 & 375 \\ 318 & 195 & 191 & 966 & 226 & 127 & 302 & 339 \end{bmatrix}$$

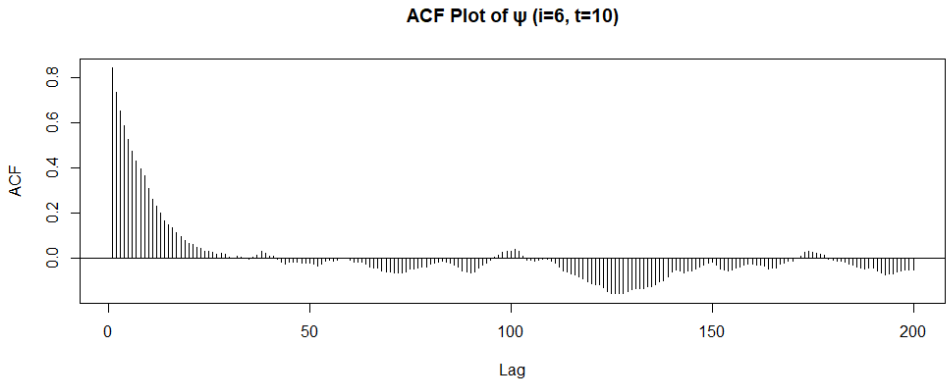
Εύκολα παρατηρείται ότι το Effective Sample Size των στοιχείων είναι αρκετά ικανοποιητικό σε σχέση με το μέγεθος του δείγματος.

Παραθέτουμε το trace plot της παραμέτρου με τη χειρότερο ESS, δηλαδή τη παραμέτρου $\psi_{i=6,t=10}$.



Εικόνα 10 Trace Plot της παραμέτρου ψ (i=6, t=10)

Παρατηρώντας το trace plot φαίνεται η αλυσίδα να έχει σταθεροποιηθεί, ενώ δεν φαίνεται να «κολλάει» σε διαφορετικά ύψη. Πράγματα που υποδηλώνουν καλή σύγκλιση. Στην εικόνα 11 φαίνεται το διάγραμμα αυτοσυσχέτισης της συγκεκριμένης μεταβλητής και παρατηρείται ότι παρά το thining υπάρχει πρόβλημα αυτοσυσχετίσεων μέχρι περίπου την χρονική υστέρηση 25.



Εικόνα 11 Διάγραμμα αυτοσυσχετίσεων της παραμέτρου ψ (i=6, t=10)

Η πρόβλεψη των παραμέτρων του δείγματος, όπως αναφέρθηκε προηγουμένως γίνεται ως εξής:

$$\psi_{T+1}{}^{(i)} \sim N_8\left(\psi_T^{(i)} + \mu^{(i)}, \Sigma^{(i)}, \eta, \xi\right) \tag{8.1}$$

Στην εφαρμογή του αλγόριθμου για τα ελληνικά δεδομένα, η μέση τιμή της κατανομής $\psi_T^{(i)} + \mu^{(i)}$ βγαίνει γρήγορα έξω από τα όρια που έχουν καθοριστεί για περικόψουν την κανονική κατανομή. Συνεπώς, δεν μπορεί να επιλεγεί μια τιμή από ένα διάστημα με κάποια μέση τιμή εκτός αυτού. Για παράδειγμα δεν μπορεί να επιλεγεί μια τιμή περικομμένης κανονικής κατανομής στο [-3,3] με μέση τιμή 5.

Τέλος, δεν κρίνεται σκόπιμο να γίνει προσπάθεια για μία πιθανή διόρθωση της συγκεκριμένης μεθόδου, αλλάζοντας για παράδειγμα τις prior κατανομές, συνεπώς θα αρκεστούμε στην παρουσίαση του παραπάνω προβλήματος που προέκυψε.

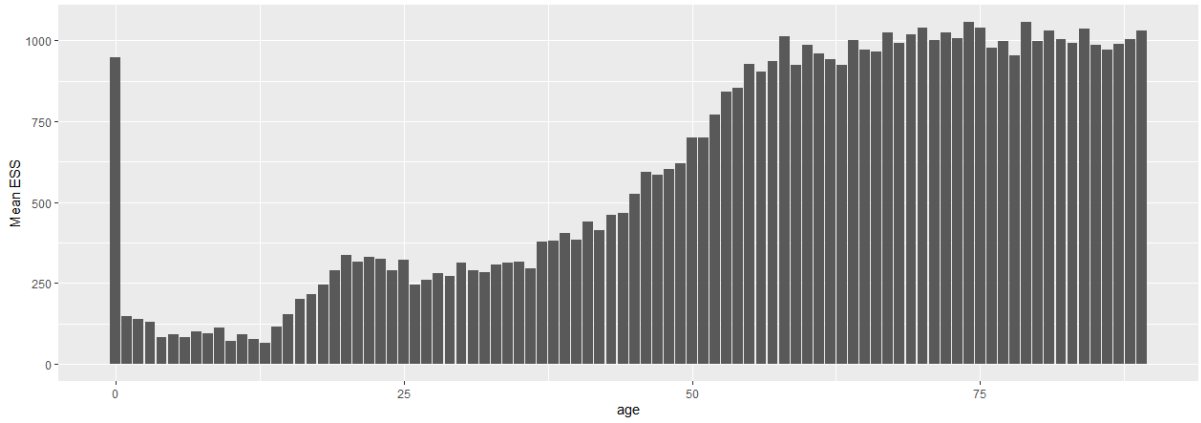
ΚΕΦΑΛΑΙΟ 9

ΕΦΑΡΜΟΓΗ IGMRF

Στο παρόν κεφάλαιο θα γίνει εφαρμογή της μεθόδου που χρησιμοποιεί ένα IGMRF για τα δεδομένα της Ελλάδας. Τα δεδομένα αυτά, αφορούν θανάτους και πληθυσμούς ανά ηλικία και πάρθηκαν από το Human Mortality Database.

Για να ελεγχθεί λοιπόν η «συμπεριφορά» της μεθόδου, αποφασίστηκε να χρησιμοποιηθούν τα συνολικά δεδομένα (άνδρες + γυναίκες) στην περίοδο 1983 με 1992. Έγιναν 220.000 επαναλήψεις, με απόρριψη -«κάψιμο»- των 20.000 πρώτων τιμών και επιλογή των υπόλοιπων ανά 200 «παρατηρήσεις» (thining). Η σταθερά για την εξασφάλιση καλύτερης σύγκλισης ρυθμίστηκε στο $0{,}000055$ έτσι επιτεύχθηκε πιθανότητα αποδοχής 55% για τα για τα z και 34% για το $\tilde{\rho}_{age}$.

Στο τελικό δείγμα μεγέθους 1.000, παρουσιάζονται τα μέσα ESS για τα z στο γράφημα που ακολουθεί:



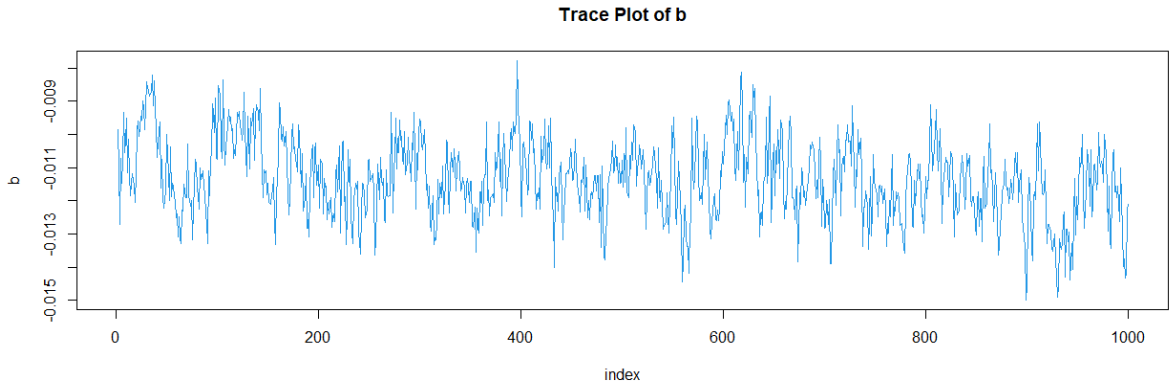
Εικόνα 12 Μέσο ESS των z ανα ηλικία (Ελλάδα)

Από την Εικόνα 12 παρατηρείται ότι το ESS στην παιδική ηλικία μειώνεται κατακόρυφα και σε κάποιες περιπτώσεις πλησιάζει το 50, το ελάχιστο ESS (38.8) παρατηρείται στην ηλικία 4 της χρονιάς 1988. Συνεπώς, φαίνεται να υπάρχει κακή σύγκλιση για τις μικρές ηλικίες. Το αποτέλεσμα αυτό, όπως θα δούμε παρακάτω, θα επηρεάσει αρνητικά τις προβλέψεις των z σε αυτές τις ηλικίες.

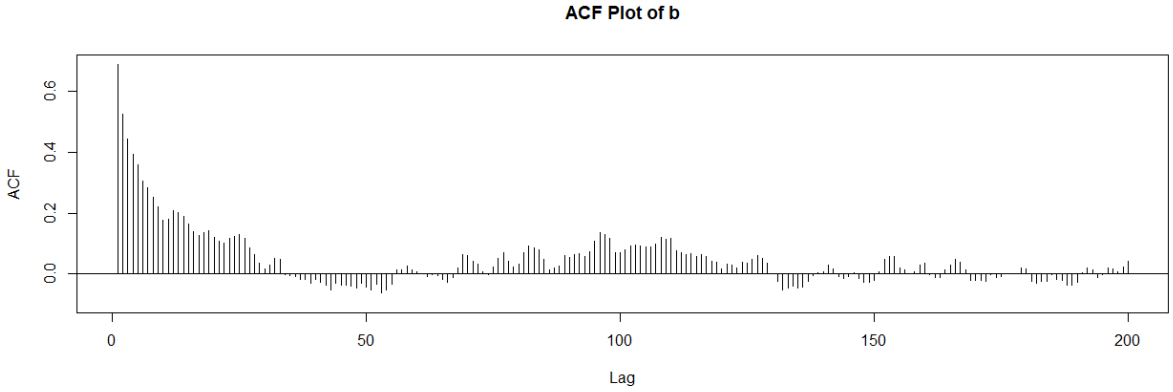
Επίσης υπολογίζεται και το ESS για τα $\tilde{\rho}_{age}, \tau, b$ όπως φαίνονται παρακάτω:

Πίνακας 6 ESS των $\tilde{\rho}_{age}, \tau, b$ αντίστοιχα (Ελλάδα)

[679 513 96]



Εικόνα 14 Trace Plot της μεταβλητής b

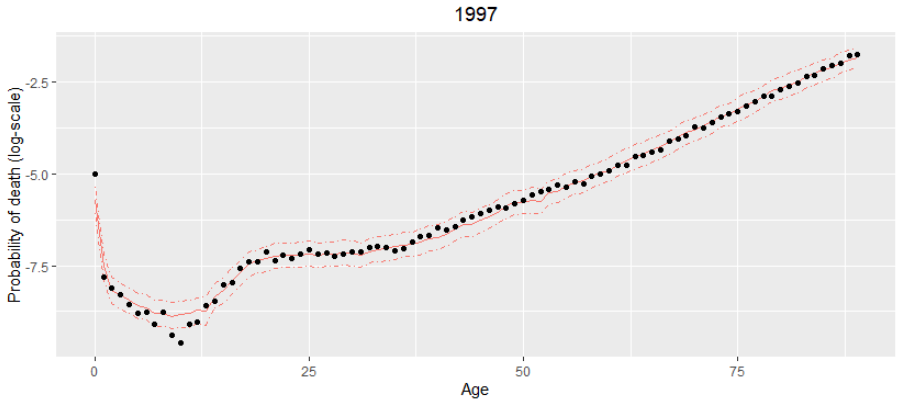


Εικόνα 13 ACF Plot της μεταβλητής b

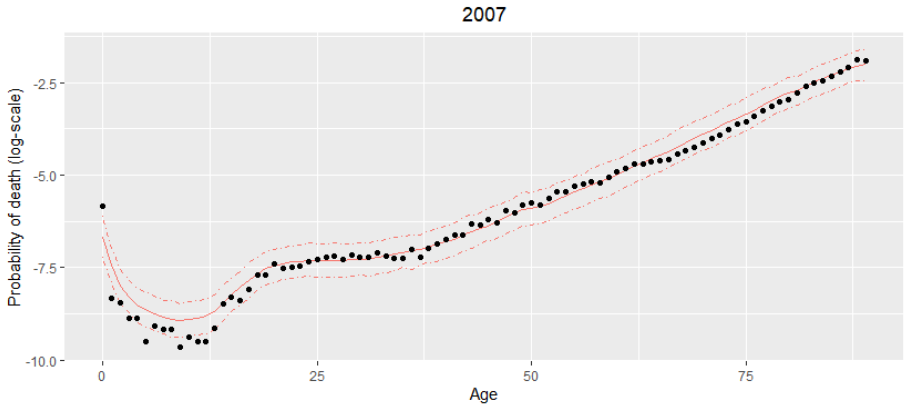
Επίσης και η μεταβλητή b έχει αρκετά χαμηλό ESS αφού το μέγεθος του ισοδύναμου ανεξάρτητου δείγματος είναι μικρότερο του 100 (δηλαδή 96)

Από το διάγραμμα αυτοσυσχετίσεων (εικ. 13)φαίνεται να υπάρχουν σχετικά υψηλές αυτοσυσχετίσεις σε αρκετές χρονικές υστερήσεις πράγμα που αιτιολογεί το χαμηλό ESS.

Περνώντας σε αυτό το σημείο στις προβλέψεις, αξίζει να παρατεθούν τα γραφήματα (εικ. 15-18) για τα επόμενα 5, 10, 15, 25 χρόνια. Στα γραφήματα αυτά αποτυπώνονται οι μέσες προβλεπόμενες πιθανότητες θανάτου σε λογαριθμική κλίμακα, μαζί με το 95% διάστημα αξιοπιστίας, καθώς και οι πραγματικές πιθανότητες θανάτου (εφόσον είναι σε παλαιότερες χρονολογίες που υπάρχουν δεδομένα).

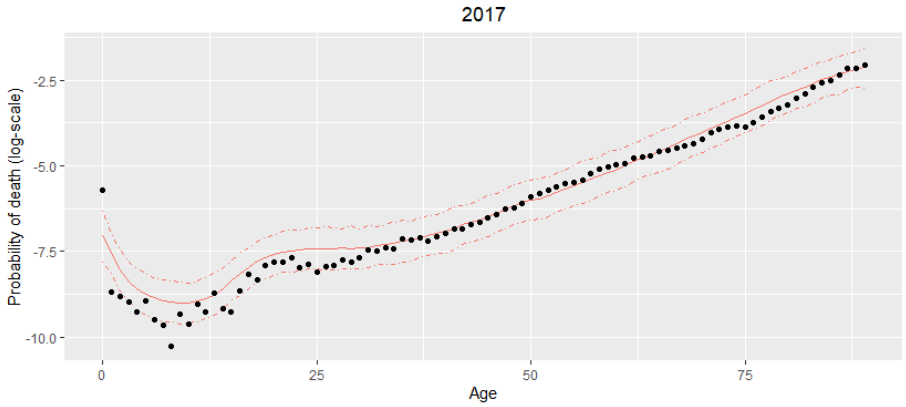


Εικόνα 15 Ελλάδα. Πρόβλεψη σε 5 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα



Εικόνα 16 Ελλάδα. Πρόβλεψη σε 10 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα

Εικόνα 17 Ελλάδα. Πρόβλεψη σε 15 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα

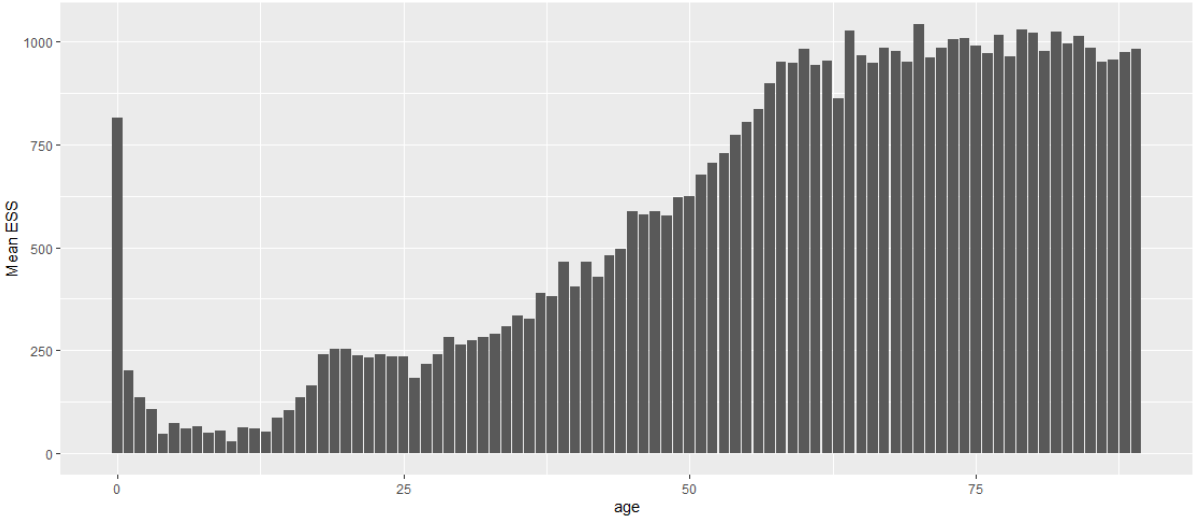


Εικόνα 18 Ελλάδα. Πρόβλεψη σε 25 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα

Από τα παραπάνω γραφήματα παρατηρείται ότι τα 95% διαστήματα αξιοπιστίας παραμένουν αρκετά «στενά» ακόμα και για τις προβλέψεις 25 έτη μετά. Παρόλα αυτά στις μικρές ηλικίες, στις οποίες το ESS ήταν εξαιρετικά χαμηλό, φαίνεται οι εκτιμήσεις να μεροληπτούν θετικά, δηλαδή οι προβλεπόμενες πιθανότητες θανάτου είναι μεγαλύτερες από τις πραγματικές. Επίσης, στις μικρές ηλικίες, σε αρκετές περιπτώσεις, το διάστημα αξιοπιστίας των προβλέψεων δεν περιέχει τις πραγματικές τιμές των πιθανοτήτων θανάτου. Στις ηλικίες άνω των 25 ετών φαίνεται η μοντελοποίηση να λειτουργεί πολύ ικανοποιητικά.

Με την ίδια λογική, για να ελεγχθεί η μέθοδος, αποφασίστηκε να χρησιμοποιηθούν τα συνολικά δεδομένα (άνδρες + γυναίκες) στην περίοδο 1983 με 1992 της Νορβηγίας. Έγιναν 220.000 επαναλήψεις, με «κάψιμο» τις 20.000 πρώτες τιμές και επιλογή ανά 200 «παρατηρήσεις» (thining). Η σταθερά για την εξασφάλιση καλύτερης σύγκλισης ρυθμίστηκε στο 0.000115 έτσι επιτεύχθηκε πιθανότητα αποδοχής 55% για τα για τα z και 34% για το $\tilde{\rho}_{age}$.

Στο τελικό δείγμα μεγέθους 1.000, παρουσιάζονται τα μέσα ESS για τα z στο γράφημα που ακολουθεί:



Εικόνα 19 Μέσο ESS των z ανά ηλικία (Νορβηγία)

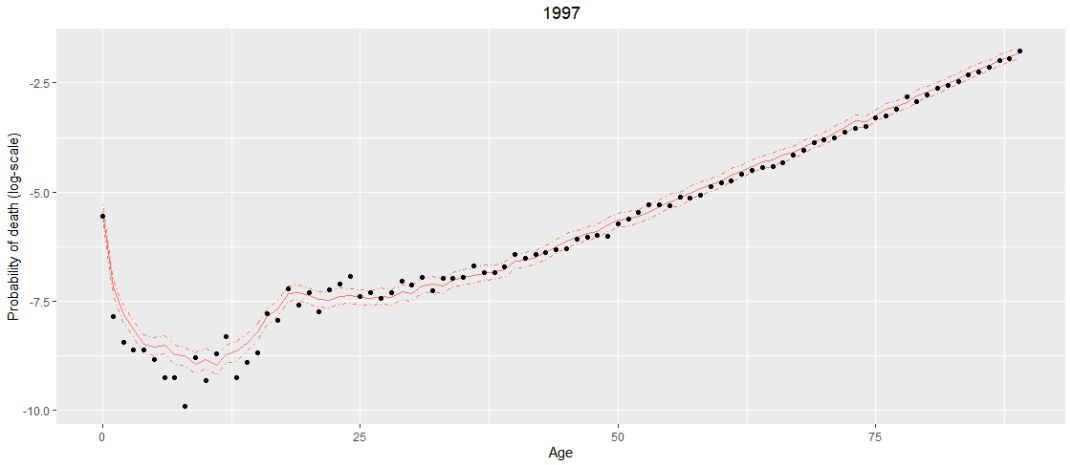
Από το γράφημα παρατηρούμε ότι το ESS στην παιδική ηλικία μειώνεται κατακόρυφα και σε κάποιες περιπτώσεις πέφτει κάτω από το 50, το ελάχιστο ESS (28.3) παρατηρείται στην ηλικία 10 της χρονιάς 1988. Συνεπώς, φαίνεται να υπάρχει κακή σύγκλιση για τις μικρές ηλικίες.

Επίσης υπολογίζεται και το ESS για τα $\tilde{\rho}_{age}, \tau, b$ όπως φαίνονται παρακάτω

Πίνακας 7 ESS των $\tilde{\rho}_{age}$, τ, b αντίστοιχα (Νορβηγία)

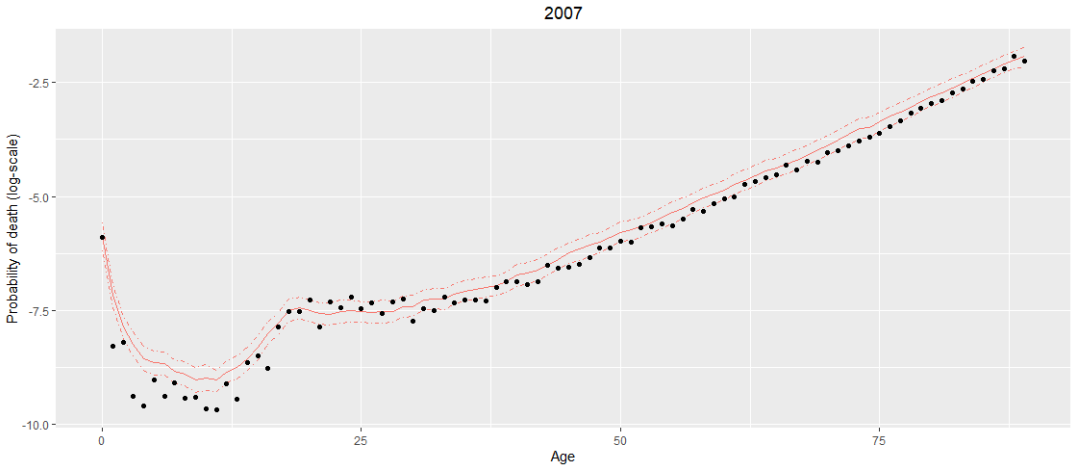
$$[355 \quad 246 \quad 300]$$

Περνώντας σε αυτό το σημείο στις προβλέψεις, αξίζει να παρατεθούν τα γραφήματα (εικ. 20-23) για τα επόμενα 5, 10, 15, 25 χρόνια. Στα γραφήματα αυτά αποτυπώνονται οι μέσες προβλεπόμενες πιθανότητες θανάτου σε λογαριθμική κλίμακα, μαζί με το 95% διάστημα αξιοπιστίας, καθώς και οι πραγματικές πιθανότητες θανάτου (εφόσον είναι σε παλαιότερες χρονολογίες που υπάρχουν δεδομένα).

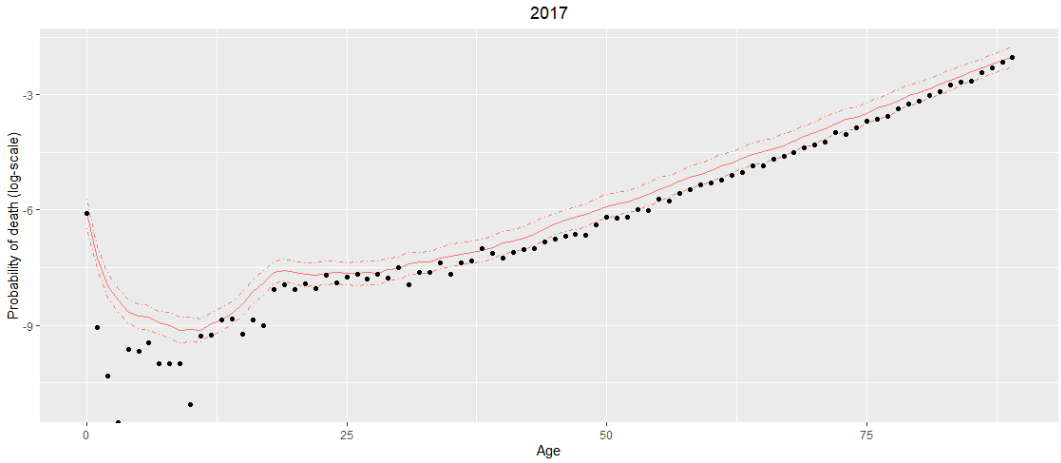


Εικόνα 20 Νορβηγία. Πρόβλεψη σε 5 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα

Εικόνα 21 Νορβηγία. Πρόβλεψη σε 10 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα



Εικόνα 22 Νορβηγία. Πρόβλεψη σε 15 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα



Εικόνα 23 Νορβηγία. Πρόβλεψη σε 25 έτη μετά των πιθανοτήτων θανάτου σε λογαριθμική κλίμακα

Από τα παραπάνω γραφήματα παρατηρείται ότι τα 95% διαστήματα αξιοπιστίας παραμένουν αρκετά «στενά» (πιο πολύ από την περίπτωση των ελληνικών δεδομένων) ακόμα και για τις προβλέψεις 25 έτη μετά. Παρόλα αυτά στις μικρές ηλικίες, στις οποίες το ESS ήταν εξαιρετικά χαμηλό, φαίνεται η μοντελοποίηση να μην λειτουργεί αρκετά καλά. Επίσης, στις μικρές ηλικίες, κατά κύριο λόγο, το διάστημα αξιοπιστίας των προβλέψεων δεν περιέχει τις πραγματικές τιμές των πιθανοτήτων θανάτου. Στις ηλικίες άνω των 25 ετών φαίνεται η μοντελοποίηση να λειτουργεί σχετικά καλά καθώς το 95% διαστήματα αξιοπιστίας περιέχει τις πραγματικές πιθανότητες θανάτου.

ΚΕΦΑΛΑΙΟ 10

ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο σκοπός της παρούσης διπλωματικής εργασίας ήταν να περιγράψει δύο μπεϋζιανές μεθόδους που αναπτύχθηκαν ώστε να μοντελοποιηθεί οι πιθανότητες θανάτου ανά ηλικία. Η πρώτη μέθοδος βασίστηκε στο υπόδειγμα Heligman – Pollard, και συγκεκριμένα, έγινε προσπάθεια να μοντελοποιηθούν οι παράμετροι του μοντέλου αυτού. Στη δεύτερη μέθοδο, εφαρμόστηκε ένα GMRF πετυχαίνοντας έτσι μία πιο άμεση μοντελοποίηση των πιθανοτήτων θανάτου.

Για να ελεγχθεί η ποιότητα των μεθόδων χρησιμοποιήθηκαν πρώτα τα δεδομένα της Ελλάδας για την περίοδο 1983-1992. Η πρώτη μέθοδος δεν κατάφερε να λειτουργήσει στο κομμάτι των προβλέψεων, η υπερπαραμετροποίηση του μοντέλου Heligman – Pollard και συνεπώς η πολύπλοκη δομή της μοντελοποίησης συνέβαλε στο να καταστήσει αδύνατη την πρόβλεψη των παραμέτρων. Για χάρη της παρουσίασης των μεθόδων δεν έγινε κάποια προσπάθεια διόρθωσης του συγκεκριμένου προβλήματος.

Η δεύτερη μέθοδος λειτούργησε τόσο για τα δεδομένα της Ελλάδας όσο και για της Νορβηγίας (ίδιες χρονολογίες). Παρόλα αυτά, η εφαρμογή της μεθόδου παρουσίασε κάποια θέματα: πολύ χαμηλά ESS σε κάποιες ηλικίες και κάποιες παραμέτρους και συνεπώς υψηλές αυτοσυσχετίσεις σε αυτές τις αλυσίδες (του MCMC). Επίσης και στις δύο χώρες για μικρές ηλικίες (<25 ετών) οι προβλέψεις φάνηκαν να μεροληπτούν θετικά καθώς υπερεκτιμούσαν τις πιθανότητες θανάτου σε σημείο που πολλές φορές οι πραγματικές πιθανότητες θανάτου δεν περιλαμβάνονταν στα 95% διαστήματα αξιοπιστίας. Παρόλα αυτά, για τα δεδομένα της Νορβηγίας επιτεύχθηκαν εμφανώς στενότερα διαστήματα αξιοπιστίας, ενώ και για τις δύο χώρες για ηλικίες άνω των 25 ετών φάνηκε να λειτουργούν αρκετά καλά.

Τέλος, κάποιες ιδέες για περεταίρω μελέτη θα ήταν: όσον αφορά την πρώτη μέθοδο θα μπορούσαν να αλλάξουν σε κάποιο -μικρό μάλλον- βαθμό οι prior κατανομές των παραμέτρων ώστε να γίνει η μέθοδος πιο «ανθεκτική» σε διαφορετικά δεδομένα. Όσον αφορά την δεύτερη μέθοδο θα μπορούσε να αλλάξει λίγο η δομή του πίνακα ακρίβειας του IGMRF ώστε να βελτιωθούν τα ESS των μικρών ηλικιών, ενώ ακόμα θα ήταν δυνατό να μοντελοποιηθούν οι πιθανότητες θανάτου ανδρών και γυναικών μαζί, διπλασιάζοντας ουσιαστικά σε μήκος το διάνυσμα των πιθανοτήτων θανάτου και μεγαλώνοντας και

αλλάζοντας αντίστοιχα το πίνακα ακρίβειας, επιτρέποντας έτσι τις εξαρτήσεις των πιθανοτήτων θανάτου μεταξύ των δύο φύλων.

Αναφορές

[1] **Gelman, A. et al. (2021).** *Bayesian Data Analysis Third edition (with errors fixed as of 15 February 2021)*, Columbia

[2] **Dellaportas, P. and Smith, A. F. M. and Stavropoulos, F. (2001).** Bayesian Analysis of Mortality Data, *Wiley for the Royal Statistical Society*, Vol. 164, No2, pp. 274-291

[3] **Alexopoulos, A. and Dellaportas, P. and Forster, J. J. (2018).** Bayesian forecasting of mortality rates by using latent Gaussian models, *Journal of the Royal Statistical Society*, Vol. 184, Part 2, pp. 689-711

[4] **Geweke, J. and Amisano, G. (2010).** Comparing and evaluating Bayesian predictive distributions of asset returns, *International Journal of Forecasting*, 26, 216-230

[5] **Havard, R. amd Held, L. (2005).** *Gaussian Markov Random Fields Theory and Applications*, Chapman & Hall/CRC

[6] **Brooks, S. and Jones, A. and Jones, G. L. and Meng, Xiao-Li (2011).** *Handbook of Markov Chain Monte Carlo*, Chapman & Hall/CRC

[7] **Gamerman, D. and Lopes, H. F. (2011).** *Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC

[8] **Χρυσαφίνου, Ο. (2016).** *ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΣΤΟΧΑΣΤΙΚΕΣ ΑΝΕΛΙΞΕΙΣ*, σοφία, Θεσσαλονίκη

[9] **Congdon, P. (1993).** Statistical graduation in local demographic analysis and projection, *J. R. Statist. Soc.* A, 156, 237–270

[10] **Heligman, L. and Pollard, J. H. (1980).** The age pattern of mortality, *J. Inst. Act.*, 107, 49-80

[11] **Hills, S. E. and Smith, A. F. M. (1992).** Parameterization issues in Bayesian inference, *In Bayesian Statistics 4* (eds J. M. Bernardo, J. 0. Berger, A. P. Dawid and A. F. M. Smith), pp. 227-246. Oxford: Oxford University Press

[12] **Kostaki, A. (1992a).** Methodology and applications of the Heligman-Pollard formula, *PhD Thesis*, Department of Statistics, University of Lund, Lund.

[13] **Rogers, A. (1986).** Parametrized multistate population dynamics and projections, *J. Am. Statist. Ass.*, 81, 48-61

[14] **Hartmann, M. (1983).** Past and recent attempts to model mortality at all ages, *J. Off. Statist.*, 3, 19-36

[15] **Κούτρας, Μ.Β. (2016).** *Εισαγωγή στη Θεωρία Πιθανοτήτων και Εφαρμογές*, Εκδόσεις Σταμούλη Α.Ε.

[16] **Dooren, P.V. (2009).** Graph Theory and Applications, *University Notes*, Université catholique de Louvain, Louvain-la-Neuve, Belgium

[17] **Molitierno, J.J. (2012).** *Applications of Combinatorial Matrix Theory to Laplacian Matrices of Graphs*, Chapman and Hall/CRC, New York

[18] **Kastner, G. and Frühwirth-Schnatter, S. (2014).** *Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models*, Computational Statistics & Data Analysis, Vol. 76, Pages 408-423

[19] **Titsias, M. K. and Papaspiliopoulos, O. (2018).** *Auxiliary gradient-based sampling algorithms*, J. R. Statist. Soc. B, 80, 749–767

[20] **Department of Economic and Social Affairs Population Division. (2019).** *Methodology of the United Nations population estimates and projections*