ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΙΑΤΡΙΚΗ ΣΧΟΛΗ, ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Μεταπτυχιακό Πρόγραμμα Σπουδών

Τμήμα Βιοστατιστικής

Ακαδημαικό Έτος 2015-2016

ΜΠΕΥΖΙΑΝΗ ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΜΗ ΤΟΠΙΚΕΣ ΕΚ-ΤΩΝ-ΠΡΟΤΕΡΩΝ ΚΑΤΑΝΟΜΕΣ

Φοιτητής: ΑΛΕΞΙΟΣ ΠΟΛΥΜΕΡΟΠΟΥΛΟΣ

Επιβλέπων: Δρ. ΙΩΑΝΝΗΣ ΝΤΖΟΥΦΡΑΣ

27 Απριλίου 2015

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΒΙΟΣΤΑΤΙΣΤΙΚΗ

που απονέμει η Ιατρική Σχολή και το Τμήμα Μαθηματικών του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών και το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την ..... από την εξεταστική επιτροπή:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΒΑΘΜΙΔΑ	ΥΠΟΓΡΑΦΗ
Ι. ΝΤΖΟΥΦΡΑΣ	ΑΝ. ΚΑΘΗΓΗΤΗΣ	.....
Λ. ΜΕΛΙΓΚΟΤΣΙΔΟΥ	ΕΠ. ΚΑΘΗΓΗΤΡΙΑ	.....
Β. ΒΑΣΔΕΚΗΣ	ΑΝ. ΚΑΘΗΓΗΤΗΣ	.....

Abstract

The most important purpose of statistics inference is the model selection. The same goal turns out to be from the Bayesian perspective which we will be further analyzed in this thesis. Although, the Bayesian statistics made the first steps in the early of 18th century, there was no succesful development because of the calculation of the multiple integrals was inefficient, thus the only exact calculation was regarding the univariate integrals. As year passed, the important improvement of the technology and the use of computers made this calculation reality, such as in handsome domains of science like physics, maths etc. The use of MCMC made more efficient and fast calculations and the inference turns out to be simplier, thus there was a significant increase in the applications of Bayesian method regarding the different fields of science.
In this thesis, we will focus on Bayesian statistics for linear models and we deal with problems where much uncertainty lies from the selection of variables and from the unknown parameters of the model. We make a description of Bayesian model selection and we will further discuss about the different choices of prior distributions. The use of advanced MCMC in problems regarding the model selection such as the estimation of aposteriori quantities of interest is called variable selection. Furthermore, the traditional priors that are used in the Bayesian model selection, are sensitive to larger values of the sample and larger values of the prior variance, also these densities assign larger weight to parameter spaces consistent with the null hypothesis, so we make an introduction and further analysis on an alternative Bayeasian approach based on non local priors regarding the variable selection and it is this most striking feature of this thesis. In addition, these non local priors densities give more mass probability to regions that are more probable under the alternative hypothesis than the null and take more in consideration the model uncertainty in such way that increase the shrinkage of the non important covariates towards zero provided by the data. The most important properties of these densities is that they introduce more sparsity models and provide more accurate predictions. Moreover, in this thesis we analyze further apllications of medical data and simulated in the enviroment of R and WinBUGS regarding the Bayesian variable selection and the model averaging.

Περίληψη

Ο σημαντικότερος σκοπός της συμπερασματολογίας στην στατιστική είναι η επιλογή μοντέλου. Ο ίδιος ο σκοπός προκύπτει και απο την Μπεϋζιανή πλευρά ο οποίος θα αναλυθεί, σε αυτήν την διπλωματική. Παρόλου που η Μπεϋζιανη συμπερασματολογία έκανε τα πρώτα της βήματα στις αρχές του 18ου αιώνα, δεν υπήρξε σημαντική ανάπτυξη καθώς οι υπολογισμοί των πολλαπλών ολοκληρωμάτων δεν ήτανε εφικτοί και οι υπολογισμοί περιορίζονταν αποκλειστικά σε μονοδιάστατα ολοκληρώματα. Καθώς τα χρόνια πέρασαν, η πρόοδος της τεχνολογίας και των ηλεκτρονικών υπολογιστών κατάφερε να κάνει εφικτό τον υπολογισμό πολλαπλών ολοκληρωμάτων σε πολλά πεδία επιστημών, όπως της φυσικής, των μαθηματικών κ.τ.λ.π. Η χρήση των μεθόδων MCMC κατάφερε να απλοποίησει όλους αυτούς τους πολύπλοκους υπολογισμούς και η συμπερασματολογία εγίνε απλούστερη και έτσι αναπτύχθηκαν πολυάριθμες εφαρμογές των Μπεϋζιανών μεθόδων σε διάφορα επιστημονικά πεδία.

Σε αυτήν την διπλωματική θα ασχοληθούμε αποκλειστικά και μόνο με Μπεϋζιανή συμπερασματολογία για τα απλά γραμικά μοντέλα και θα αντιμετωπίσουμε προβλήματα όταν υπάρχει πλήρης αβεβαιότητα για το ποιες μεταβλητές πρέπει να συμπεριληφθούν στο μοντέλο και αβεβαιότητα που πηγάζει απο τις άγνωστες παραμέτρους του μοντέλου. Θα γίνει περιγραφή της Μπεϋζιανής επιλογής μοντέλων και θα γίνει αναφορά για τις εναλλακτικές επιλογές των εκ-των-προτέρων κατανομών που θα χρησιμοποιηθούν σε τέτοια προβλήματα. Η χρήση προχωρημένων μεθόδων MCMC σε τέτοια προβλήματα εκτίμησης των εκ-των-υστέρων παραμέτρων καλείται επιλογή μεταβλητών. Οι παραδοσιακές εκ-των-προτέρων κατανομές που χρησιμοποιούνται στην Μπεϋζιανή επιλογή μεταβλητών είναι ευαίσθητες στον καθορισμό των εκ-των-προτέρων υπερπαραμέτρων και στην αύξηση του μέγεθος του δείγματος n, και επίσης δίνουν μεγαλύτερο βάρος σε περιοχές που σχετίζονται με την μηδενική υπόθεση, γιαυτό προτείνονται οι μη τοπικές εκ-των-προτέρων κατανομές που αποτελούν ενα καινοτόμο προσέγγισης στην επιλογή μεταβλητών και είναι το κύριο κομμάτι αυτής της διπλωματικής. Επιπλέον, οι μη τοπικές εκ-των-προτέρων κατανομές δίνουν μεγαλύτερη μάζα πιθανότητα σε περιοχές που είναι πιο πιθανές για την εναλλακτική υπόθεση, έτσι λαμβάνουν υπόψην μεγάλο μέρος της αβεβαιότητας του μοντέλου καθώς συρρικνώνουν στο μηδέν τις μη σημαντικές μεταβλητές. Τα βασικά χαρακτηριστικά είναι ότι υποστηρίζουν

απλούστερα μοντέλα και παρέχουν καλύτερες προβλέψεις. Στο πλαίσιο αυτης της διπλωματικής, εφαρμογές των προγραμμάτων R και WinBUGS παρουσιάζονται για την επιλογή μεταβλητών και την Μπεϋζιανη στάθμιση μοντέλων. Οι εκάστοτε μεθοδολογίες θα εφαρμοστούν σε ιατρικά δεδομένα και προσομοιωμένα δεδομένα.

Ευχαριστίες

Αρχικά, θα ήθελα να εκφράσω την πλήρη ευγνωμοσύνη και τον σεβασμό για τον Επιβλέπων μου Καθηγητή Ιωάννη Ντζούφρα για την συνεργασία και για όλο τον διαθέσιμο χρόνο που μου παρείχε, για την επίλυση όλων των αποριών, για όλα αυτά που μου έμαθε, κυρίως για την υπομονή του και πραγματικά για όλη την βοήθεια του. Χωρίς την συνεισφορά του συγκεκριμένου Καθηγητή σε αυτήν την διπλωματική, το όλο αποτέλεσμα δεν θα ήτανε ικανοποιητικό και πλήρες.

Επίσης, θα ήθελα να ευχαριστήσω την Καθηγήτρια Λουκία Μελιγκοτσίδου που μου δημιούργησε ενδιαφέρον για την Μπεϋζιανη Συμπερασματολογία και για την δυνατότητα απόκτησης εμπειρίας μέσω της διδασκαλίας στα μαθήματα Μπεϋζιανη Συμπερασματολογία 1 και Μπεϋζιανή Συμπερασματολογία 2.

Επιπλέον, θα ήθελα να ευχαριστήσω τον Καθηγητή Βασίλη Βασδέκη για την εξαιρετική μεταδοτικότητα του μέσω των εργαστηριών και της διδασκαλίας στα γραμμικά μοντέλα στο τελευταίο προπτυχιακό μου έτος 2011-2012, ακόμα τον ευχαριστώ και για την συγγραφή της συστατικής μου επιστολής για την επίτευξη αυτού του μεταπτυχιακού την περίοδο 2012.

Ακόμα, θα ήθελα να ευχαριστήσω όλα τους διδάσκοντες του ΠΜΣ βιοστατιστικής για την εξαιρετική διδασκαλία και την μεθοδική δουλειά που μας παρείχαν μέσω των διαλέξεων και των εργασιών.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και όλους του κοντινούς φίλου μου για όλη την υποστήριξη και ενθάρρυνση για την επίτευξη αυτής της διπλωματικής εργασίας.

Περιγραφή Κεφαλαιών

Στο κεφάλαιο 1 γίνεται μία εισαγωγή στην ιστορία της Μπεϋζιανής συμπερασματολογίας. Στην συνέχεια γίνεται αναφορά στα κυριώς χαρακτηριστικά της Μπεϋζιανής συμπερασματολογίας και προσπαθούμε να αποτυπόσουμε τα βασικά σημεία των εκ-των-προτέρων και των εκ-των-υστέρων κατανομών που αποτελούν σημαντικό ρόλο στην Μπεϋζιανή στατιστική. Επιπλέον, περιγράφεται η Μπεϋζιανή συμπερασματολογία για τα απλά γραμμικά μοντέλα και γίνεται ανάλυση δεδομένων για ιατρικά δεδομένα που σχετίζονται με ελλιποβαρή μωρά με την χρήση μαρκοβιανών αλυσίδων MCMC, καθώς περιλαμβάνονται γραφήματα. Ωστόσο, θα αναφερθούμε στον έλεγχο υποθέσεων ο οποίος ισοδυναμεί με την Μπεϋζιανή επιλογή μοντέλων που πραγματοποιείται με την βοήθεια της περιθώριας πιθανοφάνειας και γίνεται υπολογισμός της περιθώριας πιθανοφάνειας για το απλο γραμμικό μοντέλο για τα δεδομένα που σχετίζονται με τα ελλιποβαρή μωρά. Τέλος, γίνεται αναφορά στο περιφήμο παράδοξο Barttlet και Lindley και παρουσιάζονται γραφήματα για τα δεδομένα των ελλιποβαρών μωρών.

Στο κεφάλαιο 2 γίνεται μια σύντομη περιγραφή των κλασσικών και Μπεϋζιανών μεθόδων επιλογής. Γινέται αναφορά στους κλασσικούς τρόπους επιλογής μοντέλων οι οποίοι περιλαμβάνουν τις βηματικές ή κλιμακωτές διαδικασίες (forward, backward και Stepwise) και τα κριτήρια πληροφορίας όπως AIC και BIC, καθώς αναφέρονται τα πλεονεκτήματα και τα μειονεκτήματα. Απο την άλλη περιγράφονται οι Μπεϋζιανες μέθοδοι επιλογής μοντέλων και αναφέρονται τα πλεονεκτήματα και τα μειονεκτήματα, ενω παράλληλα περιγράφονται διάφορες εναλλακτικές εκ-των-προτέρων κατανομές που χρησιμοποιούνται στην επιλογή μοντέλων, όταν πρακτικά ο χώρος των μοντέλων είναι μεγάλος. Ακόμα, περιγράφονται οι αλγόριθμοι Μπεϋζιανής επιλογής μεταβλητών ο (Gibbs variable selection) , η στοχαστική διερεύνηση επιλογής μεταβλητών (Stochastic search variable) και ο δειγματολήπτης Gibbs. Επιπλέον περιγράφονται πιο γένικοι αλγόριθμοι επιλογής μεταβλητών όπως Άλλοι πιο γενικοί αλγόριθμοι επιλογής μοντέλων είναι η μέθοδος των (Carlin and Chib), η μέθοδος MCMC αναστρέψιμου άλματος MCMC (Reversible jump MCMC και η κατά Μετρόπολις εκδοχή της μεθόδου Carlin και Chib (Metropolized Carlin and Chib) στην περίπτωση που ό χώρος των μοντέλων είναι πολύ μεγάλος και η πλήρη απαρίθμηση του καθίσταται

αδύνατη. Επιπροσθέτως, περιγράφεται η εκ-των-υστέρων ανάλυση μέσω της Μπεϋζιανής στάθμισης μοντέλων και η επιλογή μοντέλων μέσω του πακέτου BAS της R και μέσω του Winbugs. Τέλος, πραγματοποιείται ανάλυση δεδομένων με την χρήση BAS της R και μέσω του Winbugs για δεδομένα της παχυσαρκίας και παρουσιάζονται τα αποτελέσματα.

Στο κεφάλαιο 3 γίνεται μια μεγάλη εισαγωγή στις προύπαρχουσες παραδοσιακές εκ-των-προτέρων κατανομές συμπεριλαμβανομένου και αυτές που περιγράφηκαν στο κεφάλαιο 1 και 2 καθώς γίνεται αναφόρα στα μειονεκτήματα τους στην αύξηση του μεγέθους του δείγματος και στον καθορισμό των εκ-των προτέρων υπερπαραμέτρων. Επιπλέον, οι παραδοσιακές εκ-των-προτέρων κατανομές που περιγράφηκαν στα προηγούμενα κεφάλαια περιγράφονται ώς εκ-των-προτέρων τοπικές εναλλακτικές κατανομές. Στην συνέχεια, γίνεται αναφορά στις εκ-των-προτέρων μη τοπικές εναλλακτικές κατανομές που είναι και το βασικό σημείο αυτής της διπλωματικής και παρουσιάζονται οι βασικές ιδιότητες και γίνεται ο διαχωρισμός απο τις αντίστοιχες εκ-των-προτέρων τοπικές εναλλακτικές. Ακόμα, περιγράφονται οι 2 βασικές κατηγορίες εκ-των-προτέρων μη τοπικών κατανομών, δηλαδή οι κατανομές γινομένου ροπών r-τάξης και οι κατανομές γινομένου αντίστροων ροπών r-τάξης και περιγράφονται αντίστοιχα οι μη τοπικές μονομεταβλητές και πολυμεταβλητές κατανομές Επιπροσθέτως, γίνεται αναφορά στον σωστό καθορισμό των υπερπαραμέτρων που διέπουν τις μη τοπικές κατανομές. Τέλος, γίνεται αναφορά στις εφαρμογές των μη τοπικών εκ-των-προτέρων κατανομών στο απλό γραμμικό μοντέλο και περιγράφεται η Μπεϋζιανή επιλογή μεταβλητών για τις μή-τοπικές εκ-των-προτέρων κατανομές. Στο πλάισιο αυτού του κεφαλαίου, πραγματοποιείται ανάλυση δεδομένων στα δεδομένα που σχετίζονται με την παχυσαρκία και με δεδομένα που σχετίζονται με τον καρκίνο του προστάτη μέσω του πακέτου mombf της R.

Στο κεφάλαιο 4 γίνεται μια ανασκόπηση στις μη τοπικές εκ-των-προτέρων κατανομές και στις ιδίοτητες τους και τονίζεται ότι ο εκ-των-προτέρων καθορισμός των παραμέτρων τ και r πρέπει να γίνεται προσεκτικά. Στο συγκεκριμένο κεφάλαιο, χρησιμοποιούμε προσομοιωμένα δείγματα μέσω μιας προσομοιωμένης μελέτης και προσπαθούμε να εξάγουμε συμπεράσματα απο τα αποτελέσματα μέσω πινάκων και γραφημάτων για τις διάφορες τιμές των παραμέτρων τ και r, καθώς πραγματοποιείται επιπλέον ανάλυση ευαισθησίας για την άγνωστη παράμετρο τ με επιπλέον πίνακες και γραφήματα. Τέλος, γίνεται σύγκριση των αποτελεσμάτων των προσομοιωμένων δεδομένων για τις εκ-των-προτέρων μη τοπικές κατανομές και εκ-των-προτέρων τοπικές κατανομές με διαγράμματα.

Περιεχόμενα

Κατάλογος διαγραμμάτων

Κατάλογος πινάκων

Κεφάλαιο 1

Εισαγωγή στην Μπεϋζιανή στατιστική

1.1 Εισαγωγή

Η θεωρία πιθανότητων ξεκίνησε απο δύο Γάλλους μαθηματικούς, τους Pierre de Fermat (1601-1665) και Blaise Pascal (1623-1662), με αφορμή τις συνθήκες πονταρίσματος σέ ενα δίασημο παιχνίδι με ζάρια εκείνης της εποχής. Μια αλληλογραφία μεταξύ των δύο μαθηματικών περιλάμβανε τα αρχικά θεμέλια της θεωρίας πιθανοτήτων. Ο Γερμανός επιστήμονας Christian Huygens (1629-1695), όντας δάσκαλος του Leibniz (1646-1716), έμαθε για την αλληλογραφία αυτή και δημοσίευσε το πρώτο βιβλίο με αναφορά στην θεωρία πιθανοτήτων το 1657. Ο Thomas Bayes (1702-1761) διατύπωσε το πασίγνωστο θέωρημα του Bayes, το οποίο δημοσιεύτηκε μετά τον θάνατο του το 1763. Το θέωρημα του Bayes είναι η καρδιά της Μπεϋζιανής στατιστικής. Έχοντας άγνοια για τον Bayes, ο Pierre-Simon Laplace (1749-1827) ανέπτυξε το θέωρημα του Bayes και το δημοσίευσε το 1774, 11 χρόνια μετά τον Bayes, ένα απο τα βασικότερα επιτεύγματα του Laplace (Laplace, 1774, p.366-367). Το 1812, ο Laplace εισήγαγε καινούργιες ιδέες και μαθηματικές τεχνικές στο βιβλίο Theorie analytique des probabilites. Πριν απο τον Laplace η θεωρία πιθανοτήτων είχε ως σκοπό να αναπτύξει μια μαθηματική ανάλυση για τα τυχερά παιχνίδια. Ο Laplace εφάρμοσε την θεωρία πιθανοτήτων σε πολλά επιστημονικά και πρακτικά προβλήματα. Το 1814, ο Laplace δημοσίευσε το Essai philosophique sur le probabilites, το οποίο εισήγαγε ένα επαγωγικό συλλογισμό που βασίζεται στις πιθανότητες. Μέσα απο αυτό η Μπεϋζιανή προσέγγιση της πιθανότητας αναπτύχθηκε περισσότερο απο τον Laplace απ΄ ότι τον Bayes, καθώς κάποιοι Μπεϋζιανοί αναφέρουν την Μπεϋζιανή συμπερασματολογία ως Λαπλασιανή συμπερασματολογία. Στην ίδια δημοσίευση ο Laplace ανέπτυξε την προσέγγιση Laplace συνοψίζοντας την σε μία απόδειξη και την χρησιμοποιήσε για να εκτιμήσει εκ-των υστέρων ροπές. Η Μπεϋζιανή ή Λαπλασιανή συμπερασματολογία χρησιμοποιήθηκε ευρέως το 1800, μέχρι που καταπολεμήθηκε

απο τους Ronald A. Fisher (1890-1962) και Jersy Neyman (1894-1981).

Ακολούθησαν διαμάχες ανάμεσα στους υποστηρικτές της υποκειμενικότητας και της αντικειμενικότητας της πιθανότητας. Στις αρχές του 1920, ο John Maynard Keynes (1883-1946) πρότεινε την ιδέα ότι η πιθανότητα πρέπει να ερμηνευθεί ως ενα στοιχείο υποκειμενικότητας που έξαρτάται απο τις συνθήκες στις οποίες στοιχηματίζουμε. Η υποκειμενικότητα της πιθανότητας αναπτύχθηκε απο τους Frank Plumpton Ramsey (1903-1930), Bruno de Fineti (1906-1985), Leonard Jimmie Savage (1917-1971) και απο άλλους.

Η πρώιμη προσέγγιση του Laplace θεωρήθηκε αντικειμενική, και αναπτύχθηκε επιπλεόν απο τον Harold Jeffreys (1891-1989). Ο Harold Jeffreys δημοσίευσε το Theory of probability το 1939 και αναγράφεται ώς η αρχή της μπεύζιανης συμπερασματολογίας. Ο Richard T. Cox (1898-1991) απέδειξε το 1946 ότι οι κανόνες της Μπεϋζιανής αναφοράς έχουνε μια αξιωματική θεμελίωση, σε αντίθεση με την κλασσική προσέγγιση, και ότι μπορεί να προκύψει απο ένα μικρό σύνολο προσωπικών πεποιθήσεων. Ο ίδιος απέδειξε ότι η Μπεϋζιανη συμπερασματολογία είναι συνεπής. Το 1950, ο Leonard Jimmie Savage (1917-1971) έκανε ευρέως πιο γνωστή την υποκειμενική πιθανότητα. Εν έτει 1906 ο Andrej Markov (1856-1922) εισήγαγε τις μαρκοβιανές αλυσίδες. Σε αυτόν βασίστηκαν οι Stanislaw Ulam (1909-1984) και John Von Neumann (1903-1957) που ανέπτυξαν την μεθοδολογία (Monte Carlo) με αναφορά στην παραγωγή τυχαίων αριθμών για την επίλυση αριθμητικών προβλήματων. Η πρώτη δημοσίευση έγινε στο περιοδικό του American statistical association το 1949 απο τον Nicholas Metropolis (1915-1999). Επίσης ο Metropolis και οι υπόλοιποι συνεργάτες του εισήγαγαν αυτό που ονόμαζουμε μέχρι τώρα Markov chain Monte Carlo (MCMC) αλγόριθμο στο περιοδικό Chemical Physics το 1953. Οι αλγόριθμοι (MCMC) απέκτησαν αργότερα πολυ μεγάλη φήμη για την δειγματοληψία απο κατανομές πιθανοτήτων και έπαιξαν σημαντικό ρόλο στην εξέλιξη της Μπεϋζιανης στατιστικής. Ο αλγόριθμος του Metropolis γενικεύτηκε στον Metropolis-Hastings αλγόριθμο απο τον W. Keith Hastings (Biometrika 1970). Το 1971 ο Valentin Fedorovich Turchin επινόησε τον δειγματολήπτη Gibbs. Ανεξαρτήτως του Turchin, τα αδέρφια Stuart και Donald German εισήγαγαν τον δειγματολήπτη Gibbs το 1984 ως ειδική περίπτωση του αλγόριθμου (Metropolis-Hastings). Το αποκορύφωμα για την Μπεϋζιανη συμπερασματολογία ήταν η ανάπτυξη λογισμικού του προγράμματος BUGS το 1989 με το οποίο οι αλγόριθμοι (MCMC) απέκτησαν αρκετές εφαρμογές σε πολλα επιστημονικά πεδία.

1.2 Εκ-των-προτέρων πεποίθηση

Η κλασσική προσέγγιση υποθέτει μια κατανομή πιθανότητας $f(x|\theta)$ για τα δεδομένα και όλες οι άγνωστες ποσότητες θεωρούνται ως σταθερές. Η αβεβαιότητα στην κλασσική στατιστική πηγάζει απο τα επαναλαμβανόμενα δείγματα έτσι ώστε η πραγματοποιήση των διάφορων διαδικασιών που βασίζεται στην επαναλαμβανόμενη δειγματοληψία απεικονίζει μία άπειρη επανάληψη του ίδιου προβλήματος για προκαθορισμένες τιμές των άγνωστων παραμέτρων.

Το πλαίσιο στο οποίο κινείται η συμπερασματολογία κατά Bayes διαφοροποιείται με αυτό της κλασικής στατιστικής: υπάρχει η παράμετρος θ του πληθυσμού που θέλουμε να εκτιμήσουμε, καθώς και η πιθανότητα $f(x|\theta)$ η οποία εκφράζει την πιθανότητα παρατήρησης των δεδομένων, κάτω απο διαφορετικές τιμές της θ. Το θ χρησιμοποιείται σαν τυχαία μεταβλητή και αυτή η διαφορά οδηγεί σε μια εντελώς διαφορετική ερμηνεία απο αυτήν της κλασικής στατιστικής. Όλη η συμπερασματολογία βασίζεται στην εκ-των-υστέρων κατανομή $f(\theta|x)$ και όχι στην $f(x|\theta)$ δηλαδή στην πιθανότητα της κατανομής της παραμέτρου θ δεδομένης της x (δεδομένα). Σε πολλές περιπτώσεις αυτό οδηγεί σε περισσότερο φυσικά συμπεράσματα σε σχέση με την κλασική στατιστική, για να μπορέσει όμως να επιτευχθεί αυτό πρέπει να καθορίσουμε την εκ-των-προτέρων κατανομή $f(\theta)$ η οποία αντιπροσωπεύει τις πεποιθήσεις μας για την κατανομή του θ προτού αποκτήσουμε οποιαδήποτε πληροφορία για τα δεδομένα μας.

Η ιδέα της εκ-των-προτέρων κατανομής της παραμέτρου θ αποτελεί και την καρδιά της θεωρίας κατά Bayes, και βασιζόμενοι στο αν μιλάμε σε έναν συνήγορο ή σε έναν αντιμαχόμενο της συγκεκριμένης μεθοδολογίας, η εκ-των-προτέρων κατανομή μπορεί να αποτελέσει το μεγαλύτερο πλεονέκτημα ή το σοβαρότερο μειονέκτημα έναντι της κλασικής στατιστικής.

1.3 Χαρακτηριστικά Μπεϋζιανής Συμπερασματολογίας

Η Μπεϋζιανη συμπερασματολογία αποτελεί ενα απο τα σπουδαιότερα εργαλεία για την επίλυση πολλών προβλημάτων σε ευρεία επιστημονικά πεδία γιατί σχετίζεται με πραγματικά προβλήματα και έχει ως συνέπεια έναν λογικό και ευαίσθητο τρόπο αντιμετώπισης. Τα βασικά χαρακτηριστικά που τη θέτουν ως συναρπαστικό και ξεχωριστό εργαλείο είναι:

- Εκ-των-προτέρων Πληροφορία: Κάθε πρόβλημα είναι μοναδικό και έχει το δικό του περιεχόμενο. Απο αυτο ακριβώς το περιεχόμενο εξάγονται εκ-των-προτέρων πληροφορίες και είναι η διατύπωση και η εκμετάλλευση της προηγούμενης γνώσης που διαχωρίζουν την Μπεϋζιανή θεωρία απο αυτήν της κλασικής στατιστικής.

- Υποκειμενική Πιθανότητα: Η κλασική στατιστική εξαρτάται απο μια μακροχρόνια συχνότητα καθορισμού των πιθανοτήτων. Αν και αυτό είναι επιθυμητό, οδηγεί σε δυσνόητα συμπεράσματα. Αντίθετα, η Μπεϋζιανη στατιστική θέτει με σαφήνεια την ιδέα ότι όλες οι πιθανότητες είναι υποκειμενικές και εξαρτώνται απο τις πεποιθήσεις του κάθε ατόμου και τις γνώσεις που μπορεί να έχει καθένας από εμας για μια δεδομένη κατάσταση. Όλη η συμπερασματολογία βασίζεται στην εκ-των-υστέρων κατανομή $f(\theta|x)$, η μορφή της οποίας εξαρτάται απο τον τρόπο καθορισμού της εκ-των -προτέρων κατανομής $f(\theta)$
- Συνέπεια: Χρησιμοποιώντας την παράμετρο θ σαν τυχαία μεταβλητή, όλη ανάπτυξη της Μπεϋζιανης συμπερασματολογίας γίνεται με τον λογισμό πιθανοτήτων. Αυτο έχει ως αποτέλεσμα ότι ολα τα συμπεράσματα μπορούνε να παρουσιαστούν με την μορφή πιθανοτήτων για την παράμετρο θ, που προκύπτουν άμεσα απο την εκ-των-υστέρων κατανομή.

Μη προσκόλληση σε συνταγές: Επειδή η κλασική στατιστική δεν είναι σε θέση να χρησιμοποιήσει όρους πιθανοτήτων για την παράμετρο θ, έχουνε αναπτυχθεί κριτήρια με σκοπό να καθορίσουνε πότε ένας συγκεκριμένος εκτιμητής θεωρείται ως καλός.

Για παράδειγμα ο συνήθης τρόπος εκτίμησης μιας παραμέτρου θ είναι ο εξής:

√ Επιλέγουμε μια στατιστική συνάρτηση δεδομένων $t = t(x)$ η οποία συνδέεται με την άγνωστη παράμετρο θ.
√ Βρίσκουμε την δειγματική κατανομή του t, $p(t \mid x)$.
√ Μετράμε την πιθανότητα κάθε τιμής του θ, ελέγχοντας τη δεδομένη τιμή του t σε σχέση με την αναμενόμενη συμπεριφορά του, δεδομένης της θ. Για μία συγκεκριμένη τιμή της $\theta = \theta_0$, αν η τιμή t βρίσκεται μέσα σε μια περιοχή που ανήκει η περισσότερη πυκνότητα πιθανότητας της $p(t \mid \theta_0)$ λέμε ότι η θ_0 είναι συμβατή με τα δεδομένα. Αλλιώς ή κάτι σπάνιο έχει συμβεί, ή η θ_0 δεν είναι η αληθινή τιμή της θ. Η Μπεϋζιανη στατιστική έχει ξεπεράσει το πρόβλήμα αυτό δημιουργώντας μια σειρά απο κριτήρια τα οποία κρίνουν και συγκρίνουν τους εκτιμητές βασιζόμενα στην εκ-των-υστέρων κατανομή.

1.4 Εκ-των-υστέρων συμπερασματολογία

Έαν θέλαμε να επικεντρωθούμε σε δύο βασικά σημεία της Μπεϋζιανής προσέγγισης: (α) θα έπρεπε να ορίσουμε εξ΄ αρχής την εκ-των προτέρων κατανομή $f(\theta)$ και (β) να υπολογίσουμε την εκ-των-υστέρων κατανομή $f(\theta|x)$ με βάση την εξίσωση $f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}$. Ας θεωρήσουμε τυχαίο δείγμα $x = (x_1, \ldots, x_n)$, με $f(x_i|\theta)$ ως η συνάρτηση κατανομής που περιγράφει τις τυχαίες μεταβλητές $x_1, \ldots, x_n$.

Έτσι η συνάρτηση πιθανοφάνειας δίνεται απο τη σχέση $f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$ η οποία ορίζει την πιθανότητα να παρατηρήσω τα δεδομένα για συγκεκριμένες τιμές τις παραμέτρου θ.

Το θεώρημα του Bayes κάνει αναθεώρηση των εκ-των-προτέρων πεποιθήσεων μας, λαμβάνοντας υπόψην τα δεδομένα x μεταβαίνοντας στην εκ-των-υστέρων κατανομή $f(\theta|x)$ μέσω της οποίας γίνεται όλη η συμπερασματολογία.

Για την περίπτωση που η άγνωστη παράμετρος θ είναι συνεχής είδαμε και παραπάνω ότι η ακόλουθη εξίσωση αναπαριστά την εκ-των-υστέρων κατανομή $f(\theta|x)$:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} \tag{1.1}$$

όπου $f(x) = \int f(x|\theta)f(\theta)d\theta$ είναι η περιθώρια πιθανοφάνεια των δεδομένων x, $f(\theta)$ η εκ-των-προτέρων κατανομή για την παράμετρο θ και $f(x|\theta)$ η πιθανοφάνεια των δεδομένων δοθέντος της παραμέτρου θ.

Σε περίπτωση που η παράμετρος θ είναι διακριτή η εξίσωση αλλάζει σε

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\sum f(x|\theta)f(\theta)d\theta} \tag{1.2}$$

όπου $f(x) = \sum f(x|\theta)f(\theta)d\theta$ είναι η περιθώρια πιθανοφάνεια των δεδομένων.

Σύμφωνα με το θεώρημα του Bayes, η εκ-των- υστέρων κατανομή της παραμέτρου θ δοθέντος των δεδομένων x μπορεί να γραφτεί ως $f(\theta|x) \propto f(x|\theta)f(\theta)$ (1.2). Απο αυτό μπορούμε να εξάγουμε το συμπέρασμα ότι η εκ-των-υστέρων κατανομή $f(\theta|x)$ περιέχει πληροφορία από τις προσωπικές πεποιθήσεις της εκ-των-προτέρων κατανομής της παράμετρου ϑ και των δεδομένων μέσω της συνάρτησης πιθανοφάνειας $f(x|\theta)$. Μπορούμε να ελέγξουμε την εκ-των-προτέρων πεποίθηση για την παράμετρο θ ορίζοντας αναλόγως την διακύμανση, δηλαδή σε περίπτωση που έχουμε πλήρη γνώση εκ-των-προτέρων για την παράμετρο θ ορίζουμε μικρή εκ-των-προτέρων διακύμανση, αντίθετα οταν έχουμε άγνοια για την παράμετρο θ ορίζουμε μεγάλη εκ-των -προτέρων διακύμανση.

Είναι σημαντικό να τονιστεί ότι στην πραγματικότητα η εκ-των-προτέρων πληροφορία δεν είναι διαθέσιμη. Υπάρχουν ποικίλοι τρόποι προκειμένου να εκφράσουμε την εκ-των -προτέρων άγνοια για την παράμετρο θ. Από την άλλη υπάρχουν περιπτώσεις όπου χρησιμοποιώντας μια ακατάλληλη εκ-των-προτέρων κατανομή για την παράμετρο θ θα έχει ως αποτελέσμα δυσκολία στους υπολογισμούς των ολοκληρωμάτων καθώς θα έχει αντίκτυπο σε ακατάλληλη εκ-των-υστέρων κατανομή που συνεπάγεται ακατόρθωτη συμπερασματολογία.

1.5 Επιλογή εκ-των-προτέρων κατανομής

Στην Μπεϋζιανη συμπερασματολογία με τον όρο εκ-των-προτέρων κατανομή για την παράμετρο θ εννοούμε την κατανομή που εκφράζει τις προσωπικές πεποιθήσεις ενός ατόμου πριν λάβει υπόψην του τα δεδομένα. Το βασικό πρόβλημα που θα μας απασχολήσει μέχρι το τέλος αυτής της διπλωματικής είναι η κατάλληλη επιλογή της εκ-των -προτέρων κατανομής της παραμέτρου θ. Κάθε ενας πρέπει να είναι πολύ προσεκτικός και να έχει σωστή μεθοδολογία και αιτιολογία που επέλεξε κάποια εκ-των-προτέρων κατανομή.

Όπως σχολιάστηκε προηγουμένως, μία απο τις κύριες διαφορές μεταξύ της κλασσικής και της Μπεϋζιανης προσέγγισης είναι ότι στην Μπεϋζιανή προσέγγιση οι παράμετροι θεωρούνται ως τυχαίες μεταβλητές και προκειμένου να γίνει η συμπερασματολογία, πρέπει να ορισθεί μια εκ-των-προτέρων κατανομή για τις παραμέτρους αυτές. Οι παράμετροι της εκ-των-προτέρων κατανομής αποκαλούνται υπερ-παράμετροι, έτσι ώστε να είναι πιο εύκολο να τις ξεχωρίσουμε απο τις υπόλοιπες και μπορεί να δοθεί ακόμα και για αυτές μια υπερ-εκ-των-προτέρων κατανομή. Μια μεγάλη οικογένεια εκ-των-προτέρων κατανομών είναι οι συζυγείς εκ-των-προτέρων κατανομές εξαιτίας των ιδιαίτερων χαρακτηριστικών που διευκολύνουνε κυρίως τους υπολογιστικούς χειρισμούς πριν απο την ανάπτυξη των ηλεκτρονικών υπολογιστών. Αυτές οι μεγάλες οικογένειες κατανομών σχολίαζονται ακολούθως:

1. υποκειμενικές - εκ-των-προτέρων κατανομές, το αποτέλεσμα της γνώμης ενός ατόμου που γνωρίζει πολύ καλα το πρόβλημα,
2. πληροφοριακές - εκ-των-προτέρων κατανομές , οι οποίες αντιπροσωπεύουν την ποσότητα συγκεκριμένης πληροφορίας για το υπο εξέταση πρόβλημα,
3. συζυγείς κατανομές, μια ευρέως γνωστή και πολύ χρήσιμη κατανομή μέλος της εκθετικής οικογένειας (Morris 1983),
4. μη-πληροφοριακές - εκ-των-προτέρων κατανομές, οι οποίες ξεκινούν απο το γεγονός ότι δεν υπάρχει εκ-των προτέρων πληροφορία. Η εκ-των-προτέρων κατανομές του Jeffreys αποτελούν ειδική περίπτωση μη-πληροφοριακών κατανομών.

Συνοψίζοντας, είδαμε ότι η επιλογή της εκ-των-προτέρων κατανομής αποτελεί ενα απο τα θέματα που πρέπει να προκαθοριστούν. Παρόλα ταύτα, θα πρέπει να λάβουμε υπόψην τα παρακάτω σημεία:

- Θα δούμε αργότερα ότι στην περίπτωση που η εκ-των-προτέρων δεν είναι εντελώς παράλογη η επιρροή της γίνεται ολοένα μικρότερη καθώς προστίθενται νέα δεδομένα.
- Από την στιγμή που η εκ-των-προτέρων κατανομή αντιπροσωπεύει τις πεποιθήσεις μας για την παράμετρο ϑ προτού μελετηθούνε τα δεδομένα μας, είναι φυσικό ότι η μεταγενέστερη ανάλυση είναι μοναδική για εμάς. Δηλαδή η εκ-των-προτέρων κατανομή που θέτει κάποιος άλλος,θα οδηγήσει σε διαφορετική μεταγενέστερη εκ-των-υστέρων συμπεραματολογία. Με αυτή την έννοια η ανάλυση είναι καθαρά υποκειμενική.
- Συχνά έχουμε μια γενική ιδέα για το ποια θα πρέπει να είναι η εκ-των-προτέρων κατανομή (πιθανότατα να μπορούμε να πούμε ποιος είναι ο μέσος και η διακύμανση της), χωρίς όμως να μπορούμε να είμαστε πιο συγκεκριμένοι για την μορφή της. Σε αυτές τις περιπτώσεις μπορούμε να χρησιμοποιήσουμε μια βολική μορφή της εκ-των-προτέρων κατανομή η οποία θα είναι το αποτέλεσμα των πεποιθήσεων μας και ταυτόχρονα θα απλοποιήσει τους μαθηματικούς υπολογισμούς σχετικά εύκολους.
- Πολλές φορές ο τελικός χρήστης ή ο ειδικός δεν εχει εκ-των-προτέρων πληροφορία ή είναι απλά αδιάφορος ως προς τις τιμές της παραμέτρου θ. Σε αυτές τις περιπτώσεις είναι αρκετά συνηθισμένο να χρησιμοποιούμε μια εκ-των-προτέρων η οποία αντανακλά την άγνοια μας για την παράμετρο ή απλα εκ-των-προτέρων κατανομή με μεγάλη διακύμανση.

1.5.1 Συζυγείς εκ-των-προτέρων κατανομές

Από τα βασικότερα προβλήματα της Μπεϋζιανης προσέγγισης είναι ότι προκύπτουν υπολογιστικές δυσκολίες όταν η σταθερά κανονικοποιήσης στον παρανομαστή της εκ-των-υστέρων κατανομής, $f(\theta|x)$, (Εξίσωση: 1.1) πρέπει να υπολογιστεί.

Ακόμα και οι πιο απλές επιλογές εκ-των-προτέρων κατανομών $f(\theta)$ μπορούν να οδηγήσουν σε υπολογιστικές δυσκολίες. Για αυτό το λόγο, κάθε επιστήμονας πρέπει να ήταν ολοένα και πιο προσεκτικός προτού ορίσει την εκ-των-προτέρων κατανομή. Μια συζυγής εκ-των-προτέρων κατανομή , $f(\theta)$, έχει ακριβώς την ίδια συναρτησιακή μορφή με την εκ-των-υστέρων κατανομή $f(\theta|x)$. Αυτό έχει ώς αποτέλεσμα ότι οι συζυγείς εκ-των-προτέρων κατανομές ανήκουν στην ίδια οικογένεια κατανομών με τις εκ-των-υστέρων κατανομές χωρίς την χρήση απαιτητικών υπολογιστικών πράξεων. Όλες οι κατανομές που ανήκουν στην εκθετική οικογένεια κατανομών έχουν συζυγείς εκ-των-προτέρων κατανομές (Gelman κ.α. 2013). Οι πιο γνωστές συζυγείς κατανομές παρουσιάζονται στον Πίνακα 1.1.

Πίνακας 1.1: Οι πιο γνωστές συζυγείς εκ-των-προτέρων κατανομές και οι ιδιότητες τους.

Κατανομή	Πιθανοφάνεια	Εκ-των- προτέρων κατανομή	Εκ-των-υστέρων παράμετροι
Poisson	$Y_i \sim Poisson(\lambda)$	$\lambda \sim Gama(a,b)$	$\hat{\alpha} = n\bar{y} + \alpha$ $\hat{b} = n + b$
Binomial	$Y_i \sim Binomial(p)$	$p \sim Beta(a,b)$	$\hat{\alpha} = \sum_n^{i=1} y_i + \alpha$ $\hat{b} = \sum_n^{i=1} N_i + b$
Normal(γνωστό σ^2)	$Yi \sim N(\lambda)$	$\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2)$	$\hat{\mu} = w\bar{y} + (1-w)\mu_0$ $\hat{\sigma^2} = \frac{w\sigma^2}{n}$ $w = \frac{\sigma^2}{\sigma_0^2 + \frac{\sigma^2}{n}}$

Η χρήση των συζυγών εκ-των-προτέρων κατανομών διευκολύνει τους υπολογισμούς της έυρεσης για της εκ-των-υστέρων κατανομής $f(\theta|x)$. Ενδεχομένως,η χρήση ορισμένης εκ-των-προτέρων κατανομής προτιμάται να επιλέγεται καθε αυτού λόγω σημαντικού πλεονεκτήματος στον υπολογισμό της εκ-των-υστέρων κατανομής σε σχέση με άλλες εκ-των- προτέρων κατανομές. Με άλλα λόγια, μπορούμε να επιλέξουμε προσεκτικά μια τέτοια εκ-των-προτέρων κατανομή μέλος της εκθετικής οικογένειας που είναι συζυγής για το συγκεκριμένο μοντέλο πιθανοφάνειας $f(x|\theta)$ καταλήγοντας σε μια εκ-των-υστέρων κατανομή που ανήκει στην ίδια οικογένεια κατανομών με την εκ-των-προτέρων κατανομή $f(\theta)$ αποφεύγοντας την χρήση πολύπλοκων υπολογισμών.

Επιπλέον, οι μείξεις των συζυγών εκ-των-προτέρων κατανομών εφαρμόζονται σε προβλήματα όπου είναι ακατάλληλη η εφαρμόγης μίας μόνο εκ-των προτέρων κατανομής, διατηρώντας εξίσου γρήγορο και εφικτό τον υπολογισμό της εκ-των-υστέρων κατανομής. Μια μείξη απο συζυγείς κανονικές κατανομές θα χρησιμοποιηθεί αργότερα οταν θα μιλήσουμε για την επιλογή μεταβλητών (για μία πιο ιδιαίτερη έκδοση του Πίνακα 1.1, βλέπε Ντζουφρας 2011, κεφ. 2). Μια σύνοψη των συζυγών εκ-των-προτέρων κατανομών βρίσκεται στον Fink (1997).

1.5.2 Μη-πληροφοριακές εκ-των-προτέρων κατανομές

Στην Μπεϋζιανή ανάλυση, πολλές φορές η εκ-των-προτέρων πληροφορία μπορεί να μην υπάρχει για την υπο μελέτη παράμετρο ή ο ερευνητής προκειμένου να αποφύγει λάθος υποκειμενική συμπερασματολογία, τα δεδομένα θα τον οδηγήσουνε στον κατάλληλο ορισμό της εκ-των-προτέρων πεποίθησης.

Σε αυτές τις περιπτώσεις η εκ-των-προτέρων κατανομή $f(\theta)$ πρέπει να περιέχει καθόλου ή λιγότερη πληροφορία συγκεκριμένα για την παράμετρο θ με την έννοια ότι καμία τιμή δεν πρέπει να

υπερισχύει απο τις υπόλοιπες, δηλαδή να έιναι ισοπίθανες. Τέτοιες εκ-των προτέρων κατανομές καλούνται ασαφής ή μη πληροφοριακές.

Σε ορισμένες περιπτώσεις μπορούμε να δημιουργήσουμε μια μή-πληροφοριακή εκ-των-προτέρων κατανομή μετατρέπωντας την κατανομή οσο το δυνατόν πιο επίπεδη. Αυτο μπορεί να επιτευχθεί πολύ εύκολα ορίζοντας μεγάλες τιμές (στην διακύμανση της κανονικής κατανομής), το οποίο θα συνεισφέρει στο άπλωμα της κατανομής και προσεγγιστικά να είναι επίπεδη για τις τιμές της παραμέτρου θ που ενδιαφερόμαστε.

Στην περίπτωση που έχουμε φραγμένο συνεχές παραμέτρικό χώρο $\theta \in \Theta = [a, b]$ μπορούμε να χρησιμοποιήσουμε ως μη-πληροφοριακή εκ-των-προτέρων κατανομή την ομοιόμορφη (Bolstad 2007)

$$f(\theta) = \frac{1}{b-a}, \quad a < \theta < b$$

Αντίθετα, εάν η παράμετρος δεν βρίσκεται υπο φραγμένο διάστημα η εκ-των-προτέρων κατανομή έχει την ακόλουθη μορφή, $f(\theta) = c, c > 0$.

Ο ερευνητής θα πρέπει να είναι ιδιαίτερα προσεκτικός όταν χρησιμοποιήσει τέτοιες εκ-των-προτέρων κατανομές. Έαν η εκ-των-προτέρων κατανομή ολοκληρώνει στο άπειρο ορίζεται ως ακατάλληλη εκ του αποτελέσματος ($\int f(\theta)d\theta = \infty$) και καθιστά αδύνατη την συμπερασματολογία.

Ακόμα και στις περιπτώσεις που αναφέραμε παραπάνω, υπάρχει τρόπος να αντιμετωπίσουμε την ακατάλληλη συμπερασματολογία κάνοντας την πιθανοφάνεια $f(x|\theta)$ με αναφορά ως προς το θ να ολοκληρώνει σε κλειστό σύνολο.

Η χρήση ομοιόμορφης εκ-των-προτέρων κατανομής ενδείκνυται για να αντιπροσωπεύσουμε την άγνοια για την παράμετρο θ και αποτελεί εύκολη λύση, όμως το κυριότερο μειονέκτημα αυτής της επιλογής είναι ότι η ομοιόμορφη κατανομή είναι μεταβλητή σε ¨1-1' μετασχηματισμούς, το οποίο έχει ως συνέπεια ότι αν η $f(\theta)$ περιέχει λίγη ή καθόλου πληροφορία για το θ, τότε η $f(\phi)$ θα είναι πληροφοριακή για το ϕ όταν $\phi = g(\theta)$. Η χρήση της μη-πληροφοριακής εκ-των -προτέρων κατανομής του Jeffreys δίνει άμεσα λύση στο συγκεκριμένο πρόβλημα καθώς είναι αμετάβλητη σε ¨1-1' μετασχηματισμούς (Jeffreys 1946).

1.5.3 **Jeffreys** εκ-των-προτέρων κατανομές

Γενικά όταν χρησιμοποιούμε επίπεδες εκ-των-προτέρων κατανομές, δεν μπορούμε να καταλήξουμε σε κλειστή μορφή των εκ-των-υστέρων κατανομών και έτσι η χρήση των τεχνικών (MCMC) είναι αναπόφευκτη.

Η μη-πληροφοριακή εκ-των-προτέρων κατανομή του Jeffreys όπως προαναφέρθηκε στην προηγούμενη παράγραφο είναι αμετάβλητη σε ¨1-1' μετασχηματισμό και έχει την μορφή $f(\theta) \propto |I(\theta)|^{\frac{1}{2}}$ όπου $I(\theta)$ είναι ο αναμενόμενος πίνακας πληροφορίας του Fisher ο οποίο ορίζεται ως εξής:

$$I(\theta) = -E(\frac{d^2 log f(x|\theta)}{d\theta^2}) = E(\frac{d log f(x|\theta)^2}{d\theta})$$

Ωστόσο ο υπολογισμός του $I(\theta)$ είναι ιδιαίτερα πολύπλοκος σε προβλήματα μεγάλων διαστάσεων. Πρέπει να τονιστεί ιδιαίτερα οτι η εκ-των-προτέρων κατανομή του Jeffreys δεν είναι απαραίτητα επίπεδη. Ωστόσο η εκ-των-προτέρων κατανομή του Jeffreys κάνει χρήση της μορφής της πιθανοφάνειας και όχι των δεδομένων καθε αυτών. Παρόλα αυτά μπορεί να είναι υποκειμενική, όπως προαναφέρθηκε προηγουμένως, για πολλές παραμετροποιήσεις, καθώς, ευνοεί τιμές δίνοντας σε αυτές περισσότερο βάρος έναντι άλλων σε ορισμένες περιπτώσεις.

1.6 Μπεϋζιανή γραμμική παλινδρόμηση

Έστω οτι έχουμε τυχαίο δείγμα $y_1, \ldots, y_n$ απο μια μεταβλητή απόκρισης Y που κατανέμεται ώς:

$$Y_i \mid \boldsymbol{\beta}, \sigma^2 \sim N(x_i^T \boldsymbol{\beta}, \sigma^2 I_n)$$

όπου

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

το διάνυσμα των παραμέτρων διάστασης $p \times 1$

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

ο πίνακας σχεδιασμού διάστασης $n \times p$

$$\boldsymbol{I_n} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

είναι ο μοναδιαίος πίνακας διάστασης $n \times n$

Το γραμμικό μοντέλο που θέλουμε να εφαρμόσουμε είναι αυτό που θα μας περιγράφει την σχέση της μεταβλητής μας απόκρισης Y σε σχέση με τις ανεξάρτητες μεταβλητές $x_1, x_2, \dots, x_p$ και θα έχει την μορφή $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + x_{pi}\beta_p$. Το ενδιαφέρον επικεντρώνεται στην εκτίμηση και την ερμηνεία της επίδρασης των ανεξάρτητων μεταβλητών στην μεταβλητή απόκρισης Y. Η παράμετρος σ^2 είναι η διακύμανση και x_i^T είναι η i-στήλη του πίνακα σχεδιασμού X. Για να εκτιμήσουμε τους συντελεστές παλινδρόμησης $\beta_0, \beta_1, \dots, \beta_p$ χρησιμοποιούμε μεθόδους βελτιστοποιήσης. Ο εκτιμητής είναι ο $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y}$ και ονομάζεται εκτιμητής ελαχίστων τετραγώνων ή μέγιστης πιθανοφάνειας.

Για να καταλήξουμε στην συμπερασματολογία πρέπει να ορίσουμε καταλλήλες εκ-των-προτέρων κατανομές για το διάνυσμα $\boldsymbol{\beta}$ των παραμέτρων του μοντέλου και για την διακύμανση σ^2. Στις επόμενες υποπαραγράφους αναλύονται δυο διαφορετικές Μπεϋζιανές προσεγγίσεις, η συζυγής ανάλυση και η δεσμευμένη συζυγής ανάλυση, και συγκρίνονται με την παραδοσιακή μέθοδο της μέγιστης πιθανοφάνειας.

1.6.1 Συζυγής ανάλυση

Στην Μπεϋζιανή συμπερασματολογία ενα απο τα πιο χρήσιμα εργαλεία που απλοποιεί τους μαθηματικούς υπολογισμούς και κάνει την μεταγενέστηρη ανάλυση άμεση είναι η χρήση των συζυγών εκ-των-προτέρων κατανομών, $f(\theta)$, για ένα συγκεκριμένο μοντέλο πιθανοφάνειας, $f(x|\theta)$, που έχει ως αποτέλεσμα η εκ-των-υστέρων κατανομή να ανήκει στην ίδια οικογένεια με την εκ-των-προτέρων κατανομή. Έστω ότι έχουμε πάλι τυχαίες παρατηρήσεις, $y_1, \dots, y_n$, απο μια μεταβλητή απόκρισης Y απο κανονική κατανομή και το $\boldsymbol{\theta}$ ορίζεται ως το διάνυσμα των παραμέτρων

$\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)$. Η πιθανοφάνεια γράφεται ως εξής:

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = (\sigma^2 2\pi)^{\frac{-n}{2}} e^{\frac{-\sum_{i=1}^{n}(y_i - x_i^T \boldsymbol{\beta})^2}{2\sigma^2}} \tag{1.3}$$

Για το απλό γραμμικό μοντέλο ας υποθέσουμε εκ-των -προτέρων κατανομή για το διάνυσμα $\boldsymbol{\beta}$ των παραμέτρων του μοντέλου $\boldsymbol{\beta}|\sigma^2 \sim N_p(\boldsymbol{\mu_\beta}, c^2\sigma^2\boldsymbol{I_p})$ όπου c θετική σταθερά. Τότε η εκ-των -υστέρων κατανομή για το διάνυσμα $\boldsymbol{\beta}$ των παραμέτρων του μοντέλου παραμένει στην οικογένεια των κανονικών κατανομών, δηλαδή $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y} \sim N_p(\widehat{\boldsymbol{\mu}}, \sigma^2\widehat{\boldsymbol{C}})$, όπου $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{C}^{-1}}(\boldsymbol{x}^T\boldsymbol{y} + c^{-2}\boldsymbol{I_p}^{-1}\boldsymbol{\mu_\beta})$ και $\widehat{\boldsymbol{C}} = (\boldsymbol{x^T x} + c^{-2}\boldsymbol{I_p}^{-1})^{-1}$. Για τις εκ-των-προτέρων υπερπαραμέτρους του διανύμαστος $\boldsymbol{\beta}$ έχουμε:

$$\boldsymbol{\mu_\beta} = \begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \\ \vdots \\ \mu_{\beta_p} \end{pmatrix}$$

το δίανυσμα της εκ-των-προτέρων μέσης τιμής του διανύσματος $\boldsymbol{\beta}$ διάστασης $p \times 1$

$$\boldsymbol{I_n} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

είναι ο μοναδιαίος πίνακας διάστασης $p \times p$.

Για τις εκ-των-υστέρων υπερπαραμέτρους του διανύμαστος $\boldsymbol{\beta}$ έχουμε :

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} \widehat{\mu_0} \\ \widehat{\mu_1} \\ \vdots \\ \widehat{\mu_p} \end{pmatrix}$$

το δίανυσμα της εκ-των-υστέρων μέσης τιμής του διανύσματος β διάστασης $p \times 1$

$$\widehat{\boldsymbol{C}} = \begin{pmatrix} \widehat{C_{11}} & \widehat{C_{12}} & \ldots & \widehat{C_{1p}} \\ \widehat{C_{21}} & \widehat{C_{22}} & \ldots & \widehat{C_{2p}} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C_{n1}} & \widehat{C_{n2}} & \ldots & \widehat{C_{np}} \end{pmatrix}$$

ο εκ-των-υστέρων πίνακας διακύμανσης του $\boldsymbol{\beta}$ διάστασης $p \times p$

Για την διακύμανση σ^2 του απλού γραμμικού μοντέλου υποθέτουμε εκ-των-προτέρων κατανομή $\sigma^2 \sim IGamma(\alpha_0, \lambda_0)$ όπου $\alpha_0, \lambda_0 > 0$. Τότε η εκ-των-υστέρων κατανομή για το διάνυσμα σ^2 των παραμέτρων του μοντέλου παραμένει στην οικογένεια των γάμμα κατανομών, δηλαδή $\sigma^2|\boldsymbol{y} \sim IGamma\hat{\alpha}, \hat{\lambda}$, όπου $\hat{\alpha}, \hat{\lambda} > 0$ και $\hat{\alpha} = \frac{n}{2} + \alpha_0$, $\hat{\lambda} = \frac{SS}{2} + \lambda_0$. Για τις εκ-των-υστέρων υπερπαραμέτρους της διακύμανσης σ^2 έχουμε: $SS = \boldsymbol{y^T y} - \widehat{\boldsymbol{\mu}^T \boldsymbol{C}^{-1}} \widehat{\boldsymbol{\mu}} + c^{-2} \boldsymbol{\mu_\beta}^T \boldsymbol{I_p}^{-1} \boldsymbol{\mu_\beta}$. Επιπλέον αν θέλαμε να βρούμε την περιθώρια εκ-των- υστέρων κατανομή του διανύσματος $\boldsymbol{\beta}$ θα ολοκληρώναμε την απο κοινού εκ-των-υστέρων κατανομή των παραμέτρων $\boldsymbol{\beta}$ και σ^2 ως προς σ^2 ως εξής:

$$f(\boldsymbol{\beta}|\boldsymbol{y}) \propto \left\{ 1 + \frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{C}}^{-1} (\boldsymbol{\beta} - \widehat{\boldsymbol{\mu}})}{\hat{\lambda}} \right\}^{-(\hat{\alpha} + \frac{p}{2})}. \tag{1.4}$$

Απο την παραπάνω εξίσωση 1.4 συνεπάγεται ότι η περιθώρια εκ-των-υστέρων κατανομή του διανυσματος $\boldsymbol{\beta}$ είναι πολυμεταβλητή $t - \textit{student}$ κατανομή με παραμέτρους $\hat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$, $\widehat{\nu}$ με την εξής μορφή:

$$f(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{\Gamma\left(\frac{\hat{\nu}+p}{2}\right)}{\Gamma\left(\frac{\hat{\nu}}{2}\right) \pi^{\frac{p}{2}} \left|\hat{\nu}\widehat{\boldsymbol{\Sigma}}\right|^{\frac{1}{2}}} \left\{ 1 + \frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta} - \widehat{\boldsymbol{\mu}})}{\hat{\nu}} \right\}^{-(\hat{\alpha} + \frac{p}{2})}, \tag{1.5}$$

όπου $\hat{\nu} = 2\hat{\alpha}$, $\hat{\boldsymbol{\Sigma}} = \left(\frac{\hat{\lambda}}{\hat{\alpha}}\right) \hat{\boldsymbol{C}}$. Συνήθως, επικεντρωνόμαστε στις περιθώριες εκ-των-υστέρων κατανομές $f(\beta_j|y)$, $j = 1, \ldots, p$ οι οποίες είναι μονομεταβλητές $t - \textit{student}$ κατανομές (Nadarajah και Dey 2005) τέτοιες ώστε:

$$\beta_j|\boldsymbol{y} \sim MSt_1(\hat{\mu_j}, \frac{SS + 2\lambda_0}{n + 2\alpha_0} \widehat{\Sigma_{jj}}, n + 2\alpha_0)$$

. Επιπλέον, στην περίπτωση που θα θέλαμε να υπολογίσουμε τα περιγραφικά μέτρα των εκ-των-υστέρων παραμέτρων β_j, σ^2 όπως η μέση τιμή, η τυπική απόκλιση και τα 95 % διαστήματα εμπιστοσύνης που υπολογίζονται απο τα εκ-των-υστέρων 2.5 % και 97.5 % ποσοστημόρια, θα εφαρμόζαμε την προσέγγιση του πίνακα 1.2

Πίνακας 1.2: Περιγραφικά μέτρα για τις εκ-των-υστέρων παραμέτρους β_j, σ^2 στην συζυγή εκ-των-προτέρων ανάλυση

Παράμετρος μοντέλου	Εκ-των-υστέρων μέση τιμή	Εκ-των-υστέρων διακύμανση	q-οστό ποσοστημόριο
β_j	$\widehat{\mu_j}$	$\frac{\hat{\lambda}}{\hat{\alpha}-1} \widehat{\Sigma_{jj}}$	$\widehat{\mu_j} + t_{2\hat{\alpha},q} \sqrt{\frac{\hat{\lambda}}{\hat{\alpha}} \widehat{\Sigma_{jj}}}$
σ^2	$\frac{\hat{\lambda}}{\hat{\alpha}-1}$	$\frac{\hat{\lambda}}{\hat{\alpha}-1}^2 \frac{1}{\hat{\alpha}-1}$	$\frac{1}{\Gamma_{\hat{\alpha},\hat{\lambda},1-q}}$

Απο τα παραπάνω είδαμε ότι στην περίπτωση της γραμμικής παλινδρόμησης οι εκ-των-υστέρων κατανομές των παραμέτρων $\boldsymbol{\beta}$ και σ^2 ανήκουν στην ίδια οικογένεια κατανομών με τις εκ-των-προτέρων κατανομές των παραμέτρων $\boldsymbol{\beta}$ και σ^2 διευκολύνοντας έτσι την συμπερασματολογία χωρίς πολύπλοκους μετασχηματισμούς επιτρέποντας την απευθείας προσομοιώση απο τις κατανομές. Σε πολλά προβλήματα όπου η παράμετρος θ είναι η μοναδική παράμετρος όπως στο παράδειγμα της γραμμικής παλινδρόμησης που εξετάσαμε παραπάνω τότε είναι εύκολο να βρούμε μία συζυγή εκ-των-προτέρων κατανομή. Παρόλα αυτά όσο αυξάνεται η διάσταση της παραμέτρου θ τόσο πιο πολύ αυξάνει η πολυπλοκότητα της συζυγούς εκ-των-προτέρων κατανομής και καθιστά αδύνατη την εύρεση της. Απο την άλλη όμως πολλα πολυπαραμετρικά Μπεϋζιανά προβλήματα επιλύονται με την δεσμευμένη συζυγής ανάλυση.

1.6.2 Δεσμευμένη συζυγής ανάλυση

Η Μπεϋζιανή στατιστική σημείωσε μεγάλη πρόοδο με την ανάπτυξη των ηλεκτρονικών υπολογιστών καθώς η ευρεία χρήση τους με την μέθοδο (MCMC) κατάφερε να λύσει πολλα πρακτικά προβλήματα σε διάφορους επιστημονικούς κλάδους που προηγουμένως παρέμεναν άλυτα. Παρόλο που αυτή η μέθοδος χρησιμοποιείται για την προσομοίωση απο την απο κοινού κατανομή, η κυρίως εφαρμογή της μέσω του δειγματολήπτη Gibbs συνδέεται άμεσα με την δεσμευμένη συζυγής ανάλυση.

Έστω ότι έχουμε πάλι τυχαίες παρατηρήσεις $y_1,\ldots,y_n$ απο μια μεταβλητή απόκρισης Y απο κανονική κατανομή και το $\boldsymbol{\theta}$ ορίζεται ως το διάνυσμα των παραμέτρων $\boldsymbol{\theta} = (\beta_0,\beta_1,\ldots,\beta_p,\omega)$ όπου ω είναι η ακρίβεια της εκτίμησης και ορίζεται ως $\omega = \frac{1}{\sigma^2}$ και η πιθανοφάνεια γράφεται σύμφωνα με την εξίσωση (1.2) όπου $\sigma^2 = \frac{1}{\omega}$. Για το απλό γραμμικό μοντέλο ας υποθέσουμε εκ-των προτέρων ανεξαρτησία των παραμέτρων $\boldsymbol{\beta}$ και ω όπου για την παράμετρο $\boldsymbol{\beta}$ υποθέτουμε εκ-των =προτέρων κατανομή $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu_\beta}, c^{-2}\boldsymbol{I_p})$ ενώ για την $\omega \sim Gamma(\alpha_0,\lambda_0)$. Η απο κοινού κατανομή εκ-των υστέρων κατανομή για το $\boldsymbol{\theta}$ γράφεται :

$$f(\boldsymbol{\theta}) = f(\boldsymbol{\beta})f(\omega) \tag{1.6}$$

Τότε η απο κοινού εκ-των-υστέρων κατανομή για το διάνυσμα $\boldsymbol{\theta} = (\beta_0,\beta_1,\ldots,\beta_p,\omega)$ των παραμέτρων του μοντέλου γράφεται, σύμφωνα με την εξίσωση (1.2), ως εξής:

$$f(\boldsymbol{\beta},\omega|y_1,\ldots,y_n) \propto e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})^T \boldsymbol{I_p}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})}\omega^{\alpha_0-1}e^{-\lambda_0\omega}\omega^{\frac{n}{2}}e^{-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-x_i^T\boldsymbol{\beta})^2} \tag{1.7}$$

Παρόλο που η μορφή της εκ-των-προτέρων απο κοινού κατανομής για την παράμετρο $\boldsymbol{\theta}$ είναι γνωστή κατανομή, η μορφή της εκ-των-υστέρων απο κοινού κατανομής είναι μία ιδιαίτερη πολύπλοκη συνάρτηση πολλών διαστάσεων που καθιστά αδύνατη την συμπερασματολογία. Όμως μπορούμε να παρατηρήσουμε ότι οι δεσμευμένες συζυγείς εκ-των-υστέρων κατανομές ανήκουν στην ίδια οικογένεια με τις εκ-των-προτέρων κατανομές των παραμέτρων:

$$f(\boldsymbol{\beta}|\omega, \boldsymbol{y}) \propto e^{-\frac{1}{2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{C}^{-1}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\mu}})} \tag{1.8}$$

$$f(\omega|\boldsymbol{\beta}, \boldsymbol{y}) \propto \omega^{\hat{\alpha}-1} e^{-\omega\hat{\lambda}} \tag{1.9}$$

Απο τις εξισώσεις 1.8 και 1.9 προκύπτουν αντίστοιχα οι δεσμευμένες εκ-των-υστέρων κατανομές των παραμέτρων $\boldsymbol{\beta}$ και σ^2 που είναι $\boldsymbol{\beta}|\omega, \boldsymbol{y} \sim N_p(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{C}})$, όπου $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{C}^{-1}}(\omega \boldsymbol{x}^T \boldsymbol{y} + c^{-2}\boldsymbol{I_p}^{-1}\boldsymbol{\mu_\beta})$, $\widehat{\boldsymbol{C}} = (c^{-2}\boldsymbol{I}_p^{-1} + \omega \boldsymbol{x}^T \boldsymbol{x})^{-1}$ και $\omega|\boldsymbol{\beta}, \boldsymbol{y} \sim Gamma(\hat{\alpha}, \hat{\lambda})$, όπου $\hat{\alpha} = \alpha_0 + \frac{n}{2}$, $\hat{\lambda} = \lambda_0 + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta})$.

Για τις εκ-των-υστέρων υπερπαραμέτρους του διανύσματος $\boldsymbol{\beta}$ έχουμε :

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} \widehat{\mu_0} \\ \widehat{\mu_1} \\ \vdots \\ \widehat{\mu_p} \end{pmatrix}$$

το δίανυσμα της εκ-των-υστέρων μέσης τιμής του διανύσματος $\boldsymbol{\beta}$ διάστασης $p \times 1$

$$\widehat{\boldsymbol{C}} = \begin{pmatrix} \widehat{C_{11}} & \widehat{C_{12}} & \dots & \widehat{C_{1p}} \\ \widehat{C_{21}} & \widehat{C_{22}} & \dots & \widehat{C_{2p}} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C_{n1}} & \widehat{C_{n2}} & \dots & \widehat{C_{np}} \end{pmatrix}$$

ο εκ-των-υστέρων πίνακας διακύμανσης του $\boldsymbol{\beta}$ διάστασης $p \times p$

Σε Μπεϋζιανά πολυπαραμετρικά προβλήματα όπως η γραμμική παλινδρόμηση που εξετάσαμε παραπάνω, είναι σχεδόν πρακτικά αδύνατο να μπορέσουμε να υπολογίσουμε περιγραφικά μέτρα όπως η εκ-των-υστέρων μέση τιμή ή διακύμανση ή ακόμα και εκ-των-υστέρων πιθανότητες. Έτσι είναι απαραίτητο να εκτιμήσουμε τις ποσότητες που μας ενδιαφέρουνε χρησιμοποιώντας την μέθοδο (MCMC). Η προσομοίωση απο μια πολυδιάστατη κατανομή είναι αδύνατη τις περισσότερες φορές και δεν μπορεί να γίνει απευθείας. Αντίθετα, η μέθοδος (MCMC) χρησιμοποιείται για να προσομοιώσει μια Μαρκοβιανή αλυσίδα, όπου η στάσιμη κατανομή συγκλίνει στην υπο-εξέταση

εκ-των-υστέρων κατανομή που θέλουμε να προσομοιώσουμε. Η ιδιότητα της δεσμευμένης συζυγούς ανάλυσης είναι καθοριστικής σημασίας για την κατασκευή μιας απο τις βασικότερες μορφές (MCMC), τον δειγματολήπτη Gibbs. Στην Μπεϋζιανή ανάλυση η υπο-εξέταση κατανομή είναι η εκ-των-υστέρων απο κοινού κατανομή των παραμέτρων. Στην περίπτωση της γραμμικής παλινδρόμησης ο δειγματολήπτης Gibbs θα προσομοιώσει δείγμα επαναληπτικά και ακολουθιακά απο τις δεσμευμένες εκ-των-υστέρων κατανομές του διανύσματος $\boldsymbol{\beta}$ και της ακρίβειας ω. Ο αλγόριθμος του Gibbs δηλαδή για την γραμμική παλινδρόμηση έχει την εξής μορφή:

Βήμα 1

► Γεννάω μια αρχική τιμή $\omega^{(0)} \sim Gamma(\alpha_0, \lambda_0)$

► Γεννάω μια αρχική τιμή $\boldsymbol{\beta}^{(0)} = \begin{pmatrix} \beta_0^{(0)} \\ \beta_1^{(0)} \\ \vdots \\ \beta_p^{(0)} \end{pmatrix} \sim N_p(\boldsymbol{\mu_\beta}^{(0)}, c^{-2(0)}\boldsymbol{I_p}^{(0)})$

Βήμα 2

► Γεννάω μια αρχική τιμή $\omega^{(1)} \sim Gamma(\hat{\alpha^{(0)}}, \hat{\lambda^{(0)}})$

► Γεννάω μια αρχική τιμή $\boldsymbol{\beta}^{(1)} = \begin{pmatrix} \beta_0^{(1)} \\ \beta_1^{(1)} \\ \vdots \\ \beta_p^{(1)} \end{pmatrix} \sim N_p(\widehat{\boldsymbol{\mu}^{(1)}}, \widehat{\boldsymbol{C}^{(1)}})$

όπου $\widehat{\boldsymbol{\mu}^{(1)}} = \widehat{\boldsymbol{C}^{-1(1)}}(\omega^{(1)}\boldsymbol{x^T y} + c^{-2(0)})\boldsymbol{I_p}^{(0)}\boldsymbol{\mu_\beta}$ και $\widehat{\boldsymbol{C}^{(1)}} = (c^{-2(0)}\boldsymbol{I_p}^{(0)} + \omega^{(1)}\boldsymbol{x^T x})^{-1}$.

Επαναλαμβάνουμε την διαδικασία μέχρι να επιτευχθεί σύγκλιση στην στάσιμη κατανομή που είναι η απο κοινού κατανομή $(\boldsymbol{\beta}, \omega)$ απο την οποία θέλουμε να προσομοιώσουμε δείγμα μέσω των δεσμευμένων συζυγών εκ-των-υστέρων κατανομών. Τέλος μπορούμε να υπολογίσουμε τα επιθυμήτα εκ-των-υστέρων περιγραφικά μέτρα όπως μέση τιμή, τυπική απόκλιση κ.α και να εξάγουμε γενικά συμπεράσματα.

Παράδειγμα 1 Μπεϋζιανη γραμμική παλινδρόμηση

Εδώ θα αναλύσουμε τα δεδομένα που προέρχονται απο το μεταπτυχιακό της βιοστατιστικής και θα χρησιμοποιήσουμε ως επεξηγηματική μεταβλητή το βάρος (weight) σε χιλιόγραμμα. Ώς ανεξάρτητες μεταβλητές έχουμε το ύψος (height) σε εκατοστόμετρα και την ηλικία (age) σε έτη. Τα δεδομένα είναι αποθηκευμένα για ένα τυχαίο δείγμα 12 παιδιών που δεν έχουν σωστή διατροφή. Το ενδιαφέρον επικεντρώνεται στο να περιγράψουμε την σχέση του μέσου βάρους ώς συνάρτηση του ύψους και της ηλικίας των παιδιών που δεν έχουν σωστή διατροφή.

Το μοντέλο που θα χρησιμοποιήσουμε είναι το μοντέλο παλινδρόμησης $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, όπου μ_i το μέσο βάρος του i-παιδιού, x_{1i} το ύψος (height) του i-παιδιού και x_{2i} η ηλικία (age) του i-παιδιού. Η Μπεϋζιανή συμπερασματολογία των δεδομένων έγινε σύμφωνα με την συζυγή εκ-των-προτέρων και δεσμευμένη συζυγή εκ-των -προτέρων ανάλυση. Για την συζυγή εκ-των-προτέρων ανάλυση, υποθέτοντας εκ-των-προτέρων κατανομές για τις παραμέτρους $\beta_0, \beta_1, \beta_2, \sigma^2$ αντίστοιχα $\beta_0 \sim N(0, 10000)$, $\beta_1 \sim N(0, 10000)$ και $\beta_2 \sim N(0, 10000)$ και $\sigma^2 \sim IGamma(0.1, 0.1)$ χρησιμοποιήθηκε απευθείας προσομοίωση των εν γνώση εκ-των-υστέρων κατανομών που παραθέσαμε στην προηγούμενη παράγραφο για τις παραμέτρους β και σ^2 αντίστοιχα απο τις κατανομές $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y} \sim N_3(\widehat{\boldsymbol{\mu}}, \sigma^2\widehat{\boldsymbol{C}})$ και $\sigma^2|\boldsymbol{\beta}, \boldsymbol{y} \sim IGamma(\widehat{\alpha}, \widehat{\lambda})$.

Αντίθετα, σύμφωνα με την δεσμευμένη συζυγή εκ-των -προτέρων ανάλυση, υποθέτοντας εκ-των-προτέρων κατανομές για τις παραμέτρους $\beta_0, \beta_1, \beta_2, \omega$ αντίστοιχα $\beta_0 \sim N(0, 10000)$, $\beta_1 \sim N(0, 10000)$ και $\beta_2 \sim N(0, 10000)$ και $\sigma^2 \sim Gamma(0.1, 0.1)$ έχοντας υπολογίσει τις δεσμευμένες εκ-των-υστέρων κατανομές των παραμέτρων $\boldsymbol{\beta}$ και ω αντίστοιχα απο τις κατανομές $\boldsymbol{\beta}|\omega, \boldsymbol{y} \sim N_3(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{C}})$ και $\omega|\boldsymbol{\beta}, \boldsymbol{y} \sim Gamma(\hat{\alpha}, \hat{\lambda})$ και ακολουθώντας την επαναληπτική διαδικάσια προσομοίωσης (MCMC) σύμφωνα με τον δειγματολήπτη Gibbs εξάγουμε δέιγμα απο την απο κοινού κατανομή των εκ-των-υστέρων παραμέτρων $\boldsymbol{\beta}$, ω 5000 επαναλήψεων.

Πρέπει να αναφερθεί ότι οι ποσότητες $\widehat{\boldsymbol{C}}$, $\widehat{\boldsymbol{\mu}}$, $\hat{\alpha}$ και $\hat{\boldsymbol{\lambda}}$ είναι διαφορετικές στις 2 ξεχωριστές εκ-των-προτέρων αναλύσεις. Ακόμα για τη συζυγή ανάλυση χρησιμοποιήσαμε την παράμετρο σ^2 ενώ για την δεσμευμένη συζυγή ανάλυση χρησιμοποιήθηκε η ακρίβεια ω. Τα αποτελέσματα των 2 αναλύσεων συνοψίζονται στους Πίνακες 1.3 και 1.4. Ακόμα έγινε έλεγχος σύγκλισης του αλγορίθμου Gibbs των υπο-μελέτη εκ-των-υστέρων παραμέτρων β_0, β_1, β_2, ω και ιστογράμματα που παρατίθενται στο Σχήμα 1.1.

Πίνακας 1.3: Εκ-των-υστέρων μέσες τιμές και τυπικές αποκλίσεις των παραμέτρων β_0, β_1, β_2, ω σε σχέση με τους εκτιμητές μέγιστης πιθανοφάνειας για το παράδειγμα 1

	Συζυγής εκ-των-προτέρων ανάλυση		Δεσμευμένη συζυγής εκ-των προτέρων κατανομή		Μέγιστη πιθανοφάνεια	
Παράμετροι	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
β_0	6.51	10.25	6.57	11.17	6.55	10.94
β_1	0.72	0.24	0.73	0.273	0.72	0.26
β_2	2.05	0.88	2.05	1.036	2.05	0.93
ω	0.045	0.018	0.046	0.021	0.046	-

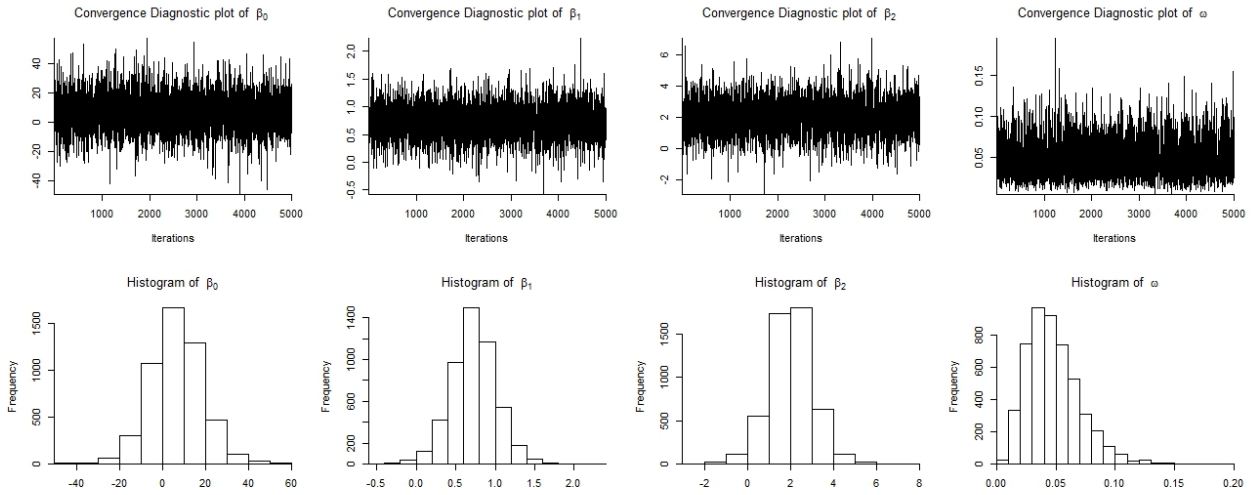
Απο τον παραπάνω Πίνακα 1.3 συμπεραίνουμε ότι τα εκ-των-υστέρων περιληπτικά μέτρα των παραμέτρων β_0, β_1, β_2 είναι πολύ κοντά στα αντίστοιχα αποτελέσματα της μέγιστης πιθανοφάνειας. Το εκτιμώμενο μοντέλο για το βάρος των παιδιών είναι $\mu_i = 6.55 + 0.72x_{1i} + 2.05x_{2i}$ για κάθε i-παιδί.

Πίνακας 1.4: Ποσοστημόρια των εκ-των-υστέρων παραμέτρων β_0, β_1, β_2

	Συζυγής εκ-των-προτέρων ανάλυση		Δεσμευμένη συζυγής εκ-των-προτέρων ανάλυση 5000 επαναλήψεων		Μέγιστη πιθανοφάνεια	
Παράμετροι	2.5%	97.5%	2.5%	97.5%	95% Διάστημα εμπιστοσύνης	
β_0	13.82	26.97	-18.04	30.45	-18.2	31.31
β_1	0.23	1.20	0.13	1.31	0.13	1.31
β_2	0.29	3.80	-0.033	4.15	-0.07	4.17

Απο τον Πίνακα 1.4 βλέπουμε ότι η συζυγής εκ-των-προτέρων ανάλυση και η δεσμευμένη συζυγής εκ-των-προτέρων ανάλυση δίνουνε ακριβώς τα ίδια αποτελέσματα με την μέθοδο εκτίμησης της μέγιστης πιθανοφάνειας. Παρατηρούμε επίσης ότι η ηλικία (age) με την κλασσική μέθοδο δεν είναι στατιστικά σημαντική λόγω ότι το 95% διάστημα εμπιστοσύνης περιέχει το μηδέν, πιθανόν λόγω ότι το δειγμα είναι μικρό ενώ σύμφωνα με τις δύο μεθόδους Μπεϋζιανής ανάλυσης το ύψος και η ηλικία έχουν σημαντικό προγνωστικό ρόλο στην πρόβλεψη του βάρους των παιδιών. Με άλλα λόγια για κάθε μια επιπλέον αύξηση ενός εκατοστόμετρου στο ύψος των παιδιών το αναμενόμενο εκ-των-υστέρων βάρος αυξάνεται κατά 0.72 χιλιόγραμμα και για κάθε

μια επιπλέον αύξηση ενός έτους της ηλικιάς των παιδιών το αναμενόμενο εκ-των-υστέρων βάρος αυξάνεται κατά 2.05 χιλιόγραμμα.



Διάγραμμα 1.1: Διαγράμματα ελέγχου σύγκλισης του αλγόριθμου Gibbs καί ιστογράμματα των εκ-των-υστέρων κατανομών των παραμέτρων $\beta_0, \beta_1, \beta_2, \omega$

1.6.3 Σύγκριση μοντέλων

Όταν διαθέτουμε έναν μεγάλο αριθμό απο ανεξάρτητες μεταβλητές είναι πολυ δύσκολο να βρούμε το βέλτιστο μοντέλο που θα περιγράφει σωστά τα υπο-μελέτη δεδομένα. Ερωτήσεις του τύπου προκύπτουν, για το ποιες κύριες επιδράσεις των μεταβλητών θα έπρεπε να συμπεριλάβουμε και ποιές αλληλεπιδράσεις θα έπρεπε να λάβουμε υπόψην προκειμένου να προχωρίσουμε στην ανάλυση των δεδομένων.

Η επιλογή μοντέλου προσπαθεί να διευκολύνει το πρόβλημα που ειπώθηκε παραπάνω για τον ερευνητή. Προκειμένου να ολοκληρώσουμε αυτην την διαδικασία που αποκαλούμε επιλογή μοντέλου, θα έπρεπε να καθοριστεί ένα κριτήριο σύγκρισης των 2 μοντέλων και μία εκάστοτε στρατηγική σύγκρισης των δύο μοντέλων. Έτσι σύμφωνα με αυτόν τον τρόπο, η επιλογή μοντέλου προσπαθεί να βρεί το σύνολο των ανεξάρτητων μεταβλητών που μπορούν να ερμηνεύσουν με τον καλύτερο τρόπο την μεταβλητότητα της εξαρτημένης μεταβλητής. Καθε μία απο τις μή σημαντικές ανεξάρτητες μεταβλητές δεν θα πρέπει να ληφθούν υπόψην γιατί εισάγουν μεροληψία στις υπο-εκτίμηση ποσότητες. Η κλασσική προσέγγιση έχει αναπτύξει στατιστικούς ελέγχους σημαντικότητας χρησιμοποιώντας F και t ελέγχους αντίστοιχα, ενώ έχουν αναπτυχθεί εναλλακτικές μέθοδοι χρησιμοποιώντας κριτήρια μοντέλου.

1.7 Μπεϋζιανή σύγκριση μοντέλων

Η βασική ιδέα της Μπεϋζιανης επιλογής μοντέλου βασίζεται στον Jeffreys (1961). Το σημαντικότερο είναι ότι τα μοντέλα δεν χρείαζεται να είναι φωλιασμένα μεταξύ τους και ότι δεν περιορίζεται ο αριθμός των συνεχόμενων υπόθέσεων, απο εδώ και στο εξής οι υποθέσεις θα έχουν την μορφή μοντέλων $m_i, i = 0, \ldots, M$ (Carlin και Louis 1997).

Η σύγκριση μεταξύ μοντέλων, περιλαμβάνει σύγκριση μεταξύ των περιθώριων πιθανοφάνειων. Η περιθώρια πιθανοφάνεια είναι η πιθανότητα των δεδομένων x δοθέντος κάποιο μοντέλο m_i και υπολογίζεται ολοκληρώνοντας την απο κοινού αβεβαιότητα μέσω των εκ-των-προτέρων παραμέτρων του μοντέλου m_i. Η σύγκριση δύο ανταγωνιστικών μοντέλων βασίζεται στον λόγο των περιθωρίων πιθανοφάνειων ή στον παράγοντα του Bayes (Bayes factor) του μοντέλου m_1 έναντι m_0. Το σημαντικότερο είναι ότι η σταθερά κανονικοποιήσης απλοποιείται στον αριθμητή και τον παρανομαστή στον τύπο του παράγοντα του Bayes.

Ας υποθέσουμε ότι έχουμε σύγκριση δύο μοντέλων m_0 και m_1 αντίστοιχα, και θ_0 και θ_1 είναι οι παράμετροι που αντιστοιχούν σε κάθε ένα απο τα μοντέλα m_0, m_1. Κάθε μοντέλο έχει μια εκ-των-προτέρων πιθανότητα $f(m_i)$, $i = 0,1$ και $f(m_0) + f(m_1) = 1$. Εφαρμόζοντας το θεώρημα του Bayes σύμφωνα με την εξίσωση 1.1 η εκ-των-υστέρων πιθανότητα καθε ένος απο τα μοντέλα δίνεται από:

$$f(m_i|x) = \frac{f(m_i)f(x \mid m_i)}{\sum_{j=0}^{1} f(m_j)f(x|m_j)}, \; i = 0,1 \tag{1.10}$$

Ο εκ-των-υστέρων λόγος συμπληρωματικών πιθανοτήτων (posterior odds) PO_{10} του μοντέλου m_1 έναντι του μοντέλου m_0 δίνεται απο:

$$PO_{10} = \frac{f(m_1|x)}{f(m_0|x)} = \frac{f(x|m_1)f(m_1)}{f(x|m_0)f(m_0)} \tag{1.11}$$

Το πηλίκο BF_{10} ονομάζεται παράγοντας του Bayes του μοντέλου m_1 σε σχέση με το μοντέλο m_0 και το πηλίκο $\frac{f(m_1)}{f(m_0)}$ ονομάζεται εκ-των-προτέρων λόγος συμπληρωματικών πιθανοτήτων. Με άλλα λόγια η εξίσωση 1.11 μπορεί να γραφτεί ως:

$$(PosteriorOdds = BayesFactor \times PriorOdds)$$

$f(x|m_i)$ είναι η περιθώρια πιθανοφάνεια των δεδομένων x για ένα συγκεκριμένο μοντέλο m_i και υπολογίζεται ώς εξής:

$$f(x|m_i) = \int_{\theta_i} f(x|\theta_i, m_i) f(\theta_i|m_i) d\theta_i \tag{1.12}$$

Η σύγκριση μεταξύ 2 μοντέλων που αναφέρθηκε προηγουμένως μπορεί να γενικευτεί και για περισσότερες συγκρίσεις. Ας υποθέσουμε ότι έχουμε υποψήφια μοντέλα για σύγκριση $m_0, m_1, \ldots, m_M$. Κάθε μοντέλο $m_1, \ldots, m_M$ συγκρίνεται με το μηδενικό μοντέλο m_0 και προκύπτουν έτσι οι παράγοντες Bayes BF_{10}, $BF_{20}, \ldots$, BF_{M0}. Οι εκ-των-υστέρων πιθανότητες κάθε μοντέλου m_i, για $i = 0, \ldots, M$ δίνονται ως εξής:

$$f(m_i|x) = \frac{w_i BF_{i0}}{\sum_{j=1}^{M} w_j BF_{j0}},\ i = 0, \ldots, M \tag{1.13}$$

Ο όρος w_i είναι ο εκ-των-προτέρων λόγος συμπληρωματικών πιθανοτήτων για το Μοντέλο m_i σε σχέση με το Μοντέλο m_0, με $BF_{00} = w_0 = 1$

Όταν θέλουμε να κάνουμε συμπερασματολογία για μια ποσότητα που μας ενδιαφέρει που έχει καθοριστεί για κάθε μοντέλο μπορούμε να λάβουμε υπόψην την αβεβαιότητα του μοντέλου χρησιμοποιώντας τις εκ-των-υστέρων πιθανότητες των μοντέλων ως βάρη Kass και Raftery (1995). Σύμφωνα με τους συγγραφείς, αυτή η τεχνική ονομάζεται Μπεϋζιανή στάθμιση μοντέλων και οδηγεί σε σημαντικά καλύτερες προβλέψεις σε σχέση με τις μεθόδους που επιλέγουν ατομικά μοντέλα, για περισσότερα στην Μπεϋζιανή στάθμιση μοντέλων βλέπε Hoeting κ.α (1999) και Raftery κ.α (1997). Προσεγγιστικές τιμές για την ερμηνεία του παράγοντα του Bayes παρέχονται απο τους Kass και Raftery (1995) στους πίνακες 1.5 και 1.6. Για τιμές του $BF_{01} < 1$ έχουμε ενδείξεις υπέρ του μοντέλου m_M. Επιπλεόν, παίρνοντας το διπλάσιο του φυσικού λογάριθμου του παράγοντα του Bayes επιστρέφουμε στην ίδια κλίμακα με την (deviance) και το τέστ πηλίκου πιθανοφανειών (Likelihood ratio test) (βλέπε Condon 2007, σελίδα 27). Παρόλα αυτά, ο ειδκός πρέπει με μεγάλη προσοχή να τα εφαρμόσει στην πράξη ακόμα και αν αυτές θεωρούνται οι καταλληλότερες ερμηνείες για τον BF. Οι λογάριθμοι της περιθώριας πιθανότητας των δεδομένων μπορούν να χρησιμοποηθούν και ώς μέτρο της προβλεπτικής ικανότητας.

Πίνακας 1.5: Ερμηνεία του παράγοντα του Bayes και του δεκαδικού λογάριθμου του παράγοντα του Bayes.

$log_{10}BF_{10}$	BF_{10}	Κατά της H_0
0-0.5	1-3.2	Ασήμαντη
0.5-1	3.2-10	Σημαντική
1-2	10-100	Ισχυρή
> 2	> 100	Καθοριστική

Πίνακας 1.6: Ερμηνεία του παράγοντα του Bayes και του διπλάσιου φυσικού λογάριθμου του παράγοντα του Bayes.

$2log_{10}BF_{10}$	BF_{10}	Κατά της H_0
0-2	1-3	Ασήμαντη
2-6	3-20	Σημαντική
6-10	20-150	Ισχυρή
> 10	> 150	Καθοριστική

1.7.1 Περιθώρια Πιθανοφάνεια

Η περιθώρια πιθανοφάνεια στην Μπεϋζιανή σύγκριση μοντέλων παίζει σημαντικό ρόλο στον υπολογισμό του παράγοντα του Bayes, στις εκ-των-υστέρων πιθανότητες και στα posterior odds και υπολογίζεται ως:

$$f(x|m_i) = \int f(x|\theta_{m_i}, m_i) f(\theta_{m_i}|m_i) d\theta_{m_i} \tag{1.14}$$

καθώς το m_i είναι μοντέλο υπο εκτίμηση, $f(\theta_{m_i}|m_i)$ είναι η πυκνότητα της παραμέτρου θ_{m_i} για το μοντέλο m_i. Απο εδώ και στο εξής, σε αυτήν την ενότητα, για λόγους ευκολίας θα αποφύγουμε τον όρο m_i που παρουσιάστηκε στις προηγούμενες εξισώσεις, εκτός απο τις περιπτώσεις που είναι απαραίτητο.

Στην πραγματικότητα είναι πολύ δύσκολο να υπολογίσουμε την περιθώρια πιθανοφάνεια του μοντέλου σε προβλήματα μεγάλων διαστάσεων που περιλαμβάνουνε πολλαπλά ολοκληρώματα. Η κατανομή πρόβλεψης είναι η μέση τιμή της πιθανοφάνειας με αναφορά στην εκ-των-προτέρων κατανομή:

$$f(x) = E_p\left[f(x|\theta)\right] \tag{1.15}$$

Στην Μπεϋζιανή προσέγγιση, οι περιθώριες πιθανοφάνειες μπορούν να δώσουν μια φυσιολογική ποινή σε πιο πολύπλοκα μοντέλα. Σε σύγκριση με την πιθανοφάνεια, η οποία αυξάνεται συνεχώς με την επιπλέον εισαγωγή παραμέτρων, η περιθώρια πιθανοφάνεια, αρχίζει να μειώνεται όσο πιο πολύπλοκο είναι το μοντέλο την οποία την καθιστά αμετάβλητη σε προβλήματα υπερπροσαρμοστικότητας. Τις περισσότερες φορές δεν είναι εφικτή η απευθείας ολοκλήρωση για τον υπολογισμό της περιθώριας πιθανοφάνειας, έτσι πρέπει να γίνει εκτίμηση της περιθώριας πιθανοφάνειας χρησιμοποιώντας άλλες προσεγγίσεις όπως είναι οι ασυμπτωτικές μέθοδοι όπως η μέθοδος του Laplace ή μέθοδοι προσομοίωσης (MCMC)

Υπολογισμός περιθώριας πιθανοφάνειας στην Μπεϋζιανή γραμμική παλινδρόμηση

Επιστρέφοντας στο παράδειγμα 1 κανονικής παλινδρόμησης με τα δεδομένα που είναι αποθηκευμένα για ένα τυχαίο δείγμα 12 παιδιών που δεν έχουνε σωστή διατροφή, το ενδιαφέρον επικεντρώνεται στο να περιγράψουμε την σχέση του μέσου βάρους ώς συνάρτηση του ύψους και της ηλικίας των παιδιών που δεν έχουνε σωστή διατροφή. Το γραμμικό μοντέλο που εκτιμήσαμε μέσω της μπεϋζιανής ανάλυσης σε προηγούμενη ενότητα είναι το , $\mu_i = 6.55 + 0.72x_{1i} + 2.05x_{2i}$, όπου μ_i το μέσο βάρος του i-παιδιού, x_{1i} όπου το μέσο βάρος του i-παιδιού, το ύψος (height) του i-παιδιού και x_{2i} η ηλικία (age) του i-παιδιού. Ο υπολογισμός της περιθώριας πιθανοφάνειας του γραμμικού μοντέλου έγινε σύμφωνα με την συζυγής εκ-των-προτέρων ανάλυση γιατί μόνο σε αυτήν την περίπτωση μπορεί να γίνει ο απευθείας υπολογισμός των ολοκληρωμάτων των υπο-εξέταση παραμέτρων $\beta_0, \beta_1, \beta_2, \sigma^2$. Ως μοντέλο πιθανοφάνειας με βάση την εξίσωση 1.3 χρησιμοποιώντας ως εκ-των -προτέρων συζυγείς κατανομές των παραμέτρων $\boldsymbol{\beta}$ και σ^2 αντίστοιχα έχουμε $\beta_0 \sim N(0, 10000)$, $\beta_1 \sim N(0, 10000)$, $\beta_2 \sim N(0, 10000)$ και $\sigma^2 \sim IGamma(0.1, 0.1)$. Τότε, σύμφωνα με την εξίσωση 1.1 οι εκ-των-υστέρων κατανομές των παραμέτρων β και ειναι αντίστοιχα $\boldsymbol{\beta}|\sigma^2, \boldsymbol{y} \sim N_3(\widehat{\boldsymbol{\mu}}, \sigma^2\widehat{\boldsymbol{C}})$ και $\sigma^2|y \sim IGamma\hat{\alpha}, \hat{\lambda}$. Έτσι για τον υπολογισμό της περιθώριας πιθανοφάνειας του γραμμικού μοντέλου σύμφωνα με την εξίσωση 1.14 έχουμε:

$$f(x|m_1) = \int_{\beta}\int_{\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 d\beta = \int_{\beta}\int_{\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}|\sigma^2) f(\sigma^2) d\sigma^2 d\beta$$

όπου m_1 το γραμμικό μοντέλο $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ Μετά απο πράξεις καταλήγουμε στην ακόλουθη μορφή περιθώρειας πιθανοφάνειας:

$$f(x|m_1) = c^{-p}\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left|\widehat{\boldsymbol{C}}\right|^{0.5}) \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\hat{\alpha})}{\hat{\lambda}^{\hat{\alpha}}} \tag{1.16}$$

όπου p η διάσταση του διανύσματος $\boldsymbol{\beta}$ και $\widehat{C}$, $\widehat{\boldsymbol{\mu}}$, $\hat{\alpha}$, $\hat{\lambda}$ ορίστηκαν σε προηγούμενη ενότητα. Στο παράδειγμα με τα παιδιά που δεν τρέφονται σωστά έχουμε την λογαριθμησμένη περιθώρια

πιθανοφάνεια του μοντέλου m_1, $\ln f(x|m_1) = -42.55$. Έστω ότι θέλουμε να ελέγξουμε τα μοντέλα m_0 και m_1 αντίστοιχα, δηλαδή την συνεισφορά του ύψους των παιδιών στην πρόβλεψη του αναμενόμενου βάρους δοθείσως της ηλικίας:

$$m_0 : \mu_i = \beta_0 + \beta_2 x_{2i}$$

$$m_1 : \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

Ισοδύναμα, αυτο μπορεί να γραφτεί με την μορφή ελέγχου υποθέσεων ως εξής:

$$H_0 : \beta_1 = 0$$

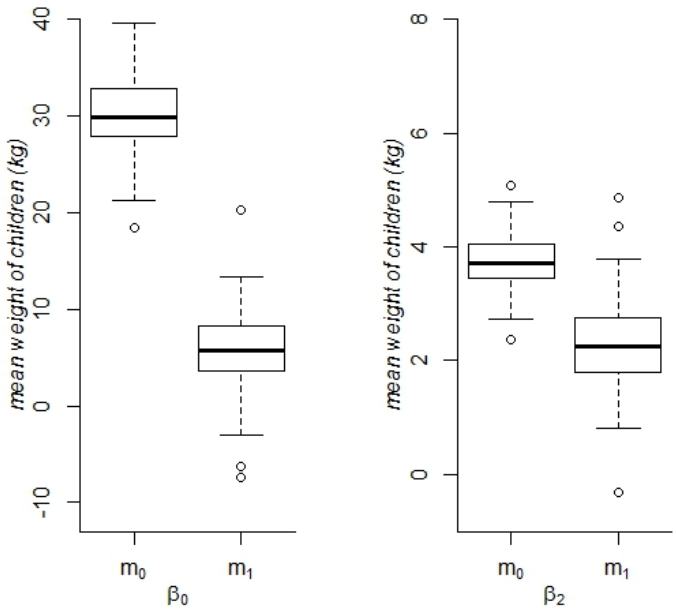
$$H_1 : \beta_2 \neq 0$$

Υποθέτοντας εκ-των-προτέρων κατανομές για τις παραμέτρους $\beta_0, \beta_2, \sigma^2$ του μοντέλου m_0 : $\mu_i = \beta_0 + \beta_2 x_{2i}$, $\beta_0 \sim N(0, 10000)$, $\beta_2 \sim N(0, 10000)$, $\sigma^2 \sim IGamma(0.1, 0.1)$ και αντίστοιχα για το μοντέλο $m_1 : \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, $\beta_0 \sim N(0, 10000)$, $\beta_1 \sim N(0, 10000)$, $\beta_2 \sim N(0, 10000)$, $\sigma^2 \sim IGamma(0.1, 0.1)$. Σύμφωνα με τις εξισώσεις 1.10 και 1.14 υπολογίστηκαν οι λογαριθμηκές περιθώριες πιθανοφάνειες και οι εκ-των-υστέρων πιθανότητες των μοντέλων και αντίστοιχα, τα αποτελέσματα παρέχονται στον πίνακα 1.7

Πίνακας 1.7: Περιθώριες λογαριθμησμένες πιθανοφάνειες και εκ των-υστέρων πιθανότητες των μοντέλων m_0, m_1 για την συζυγή ανάλυση

Μοντέλο	Μεταβλητές	Περιθώρια λογαριθμηκή πιθανοφάνεια	Εκ-των-υστέρων πιθανότητα μοντέλου
m_0	ηλικία	-38.82	0.976
m_1	ύψος +ηλικία	-42.55	0.023

Άρα το μειωμένο μοντέλο m_0 με ανεξάρτητη μεταβλητή μόνο την ηλικία είναι καλύτερο για την πρόβλεψη του βάρους σε σχέση με το πλήρες μοντέλο m_1 που περιέχει την ηλικία και το ύψος. Ακόμα για την σύγκριση των συντελεστών β_0, β_2 για να έχουμε μια καλύτερη γενική εικόνα και να είναι πιο άμεση η σύγκριση στο Διάγραμμα 1.2 φαίνονται τα διαγράμματα πλαισίου απολήξεων των εκ-των-υστέρων παραμέτρων β_0, β_2 αντίστοιχα για τα μοντέλα m_0 και m_1.



Διάγραμμα 1.2: Διαγράμματα σύγκρισης πλαισίου-απολήξεων των παραμέτρων β_0, β_2 κάτω απο τα μοντέλα m_0, m_1

1.7.2 Το παράδοξο του **Bartlett-Lindley**

Το (1957) ο Lindley παρατήρησε μια περίεργη συμπεριφορά του παράγοντα του Bayes και το ονόμασε παράξοδο. Καθώς το μέγεθος του δείγματος n τείνει στο ∞, τότε ο εκ-των-υστέρων λόγος συμπληρωματικών πιθανοτήτων (posterior odds) πάει στο ∞ για οποιοδήποτε επίπεδο σημαντικότητας υποστηρίζοντας το απλούστερο μοντέλο. Στην κλασσική στατιστική, οι έλεγχοι σημαντικότητας, καθώς το μέγεθος του δείγματος n αυξάνει, τείνουν να απορρίπτουν το πιο απλό μοντέλο. Έτσι, ανάλογα με την μέθοδο που επιλέγεται για να γίνει η συμπερασματολογία, ο ειδικός θα καταλήξει "σωστά" υποστηρίζοντας διαφορετικές υποθέσεις.

Ο Bartlett αργότερα παρατήρησε ότι για μεγάλες εκ-των-προτέρων διακυμάνσεις ο εκ-των-υστέρων συμπληρωματικός λόγος πιθανοτήτων (posterior odds) υποστηρίζει το πιο απλό μοντέλο. Δοθέντος αυτού του ευρήματος, ο ερευνητής θα πρεπει προσεκτικά να διαλέξει εκ-των-προτέρων κατανομές με μεγάλη διακύμανση και πρέπει να αποφεύγεται η χρήση των ακατάλληλων εκ-των -προτέρων κατανομών. Πιο συγκεκριμένα, αυτό το παράδοξο οδήγησε σε πολλές προτεινόμενες εκ-των-προτέρων κατανομές για την επιλογή μοντέλου, ορισμένες απο τις οποιές πρόκειται να αναλύσουμε σε επόμενη ενότητα.

Παράδειγμα 1. Το παράδοξο του **Bartlett-Lindley** στην γραμμική παλινδρόμηση

Επιστρέφοντας στο παράδειγμα 1 κανονικής παλινδρόμησης υποθέτοντας τις ίδιες εκ-των-προτέρων κατανομές για τις παραμέτρους $\boldsymbol{\beta}$ και σ^2 σύμφωνα με την συζυγή ανάλυση για το μοντέλο πιθανοφάνειας με βάση την εξίσωση 1.3 είδαμε ότι οι εκ-των-υστέρων κατανομές είναι αντίστοιχα $N_3(\widehat{\boldsymbol{\mu}},\sigma^2\widehat{\boldsymbol{C}})$ και $\sigma^2|\boldsymbol{y} \sim IGamma(\hat{\alpha},\hat{\lambda})$. Δηλαδή για οποιοδήποτε μοντέλο υποθέτουμε για τις παραμέτρους $\boldsymbol{\beta},\sigma^2$ εκ-των-προτέρων κατανομές αντίστοιχα $\boldsymbol{\beta}|\sigma^2 \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{m_i}}}, c^2\sigma^2\boldsymbol{I}_{\boldsymbol{p}_{\boldsymbol{m_i}}})$ καί $\sigma^2 \sim IGamma(\alpha_0,\lambda_0)$ όπου $\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{m_i}}}$ είναι το διάνυσμα της εκ-των-προτέρων μέσης τιμής του διανύσματος $\boldsymbol{\beta}$ κάτω απο οποιοδήποτε μοντέλο m_i διάστασης $p\times 1$, $\boldsymbol{I}_{\boldsymbol{p}_{\boldsymbol{m_i}}}$ είναι ο μοναδιαίος πίνακας διάστασης $p\times p$ του μοντέλου m_i και c θετική σταθερά. Δοθέντων των δεδομένων και σύμφωνα με την εξίσωση (1.2) οι εκ-των -υστέρων κατανομές των παραμέτρων $\boldsymbol{\beta},\sigma^2$ για οποιοδήποτε γραμμικό μοντέλο m_i είναι είναι $\boldsymbol{\beta}|\sigma^2,\boldsymbol{y} \sim N_p(\widehat{\boldsymbol{\mu}_{\boldsymbol{m_i}}},\sigma^2\widehat{\boldsymbol{C}_{\boldsymbol{m_i}}})$, όπου $\widehat{\boldsymbol{\mu}_{\boldsymbol{m_i}}} = \widehat{\boldsymbol{C}_{\boldsymbol{m_i}}}^{-1}(\boldsymbol{x}_{\boldsymbol{m_i}}^{\boldsymbol{T}}\boldsymbol{y} + c^{-2}\boldsymbol{I}_{\boldsymbol{p}_{\boldsymbol{m_i}}}{}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{m_i}}}), \widehat{\boldsymbol{C}_{\boldsymbol{m_i}}} = (\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x} + c^{-2}\boldsymbol{I}_{\boldsymbol{p}}{}^{-1})^{-1}$ όπου $\widehat{\boldsymbol{\mu}_{\boldsymbol{m_i}}}$ είναι το διάνυσμα της εκ-των-υστέρων μέσης τιμής του διανύσματος $\boldsymbol{\beta}$ του μοντέλου m_i δίαστασης $p\times 1$, $\boldsymbol{x}_{\boldsymbol{m_i}}$ είναι ο πίνακας σχεδιασμού του μοντέλου m_i , $\widehat{\boldsymbol{C}_{\boldsymbol{m_i}}}$ είναι ο εκ-των-υστέρων πίνακας διακύμανσης του μοντέλου m_i και $\sigma^2|\boldsymbol{y} \sim IGamma(\hat{\alpha},\hat{\lambda_{m_i}})$ όπου $\hat{\alpha},\hat{\lambda} > 0$, $\hat{\alpha} = \frac{n}{2} + \alpha_0$, $p\times 1$, $\hat{\lambda}_{m_i} = \frac{SS_{m_i}}{2} + \lambda_0$, $SS = \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \widehat{\boldsymbol{\mu}_{\boldsymbol{m_i}}^{\boldsymbol{T}}}\widehat{\boldsymbol{C}_{\boldsymbol{m_i}}^{-1}}\widehat{\boldsymbol{\mu}_{\boldsymbol{m_1}}} + c^{-2}\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{m_i}}}{}^{T}\boldsymbol{I}_{\boldsymbol{p}}{}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}_{\boldsymbol{m_i}}}$. Σύμφωνα με την προηγούμενη ενότητα ο υπολογισμός της περιθώριας πιθανοφάνειας για το πλήρες μοντέλο m_1 έγινε σύμφωνα με την εξίσωση 1.15. Στο παράδειγμα 1 θέλαμε να ελέγξουμε την επίδραση του ύψους στο μέσο βάρος των παιδιών δοθείσως της ηλικίας. Υποθέτοντας εκ-των-προτέρων κατανομές για τις παραμέτρους β_0,β_2,σ^2 του μοντέλου $m_0 : \mu_i = \beta_0 + \beta_2 x_{2i}$, $\beta_0 \sim N(0,10)$, $\beta_2 \sim N(0,10)$, $\sigma^2 \sim IGamma(\alpha_0,\lambda_0)$ και αντίστοιχα για το μοντέλο $m_1 : \mu_i = \beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}$, $\beta_0 \sim N(0,10)$, $\beta_1 \sim N(0,d^2)$, $\beta_2 \sim N(0,10)$, $\sigma^2 \sim IGamma(\alpha_0,\lambda_0)$, χρησιμοποιώντας την εξίσωση 1.15 θα υπολογίσουμε τον εκ-των-υστέρων συμπληρωματικό λόγο πιθανοτήτων (odds) του μοντέλου $m_0 : \mu_i = \beta_0 + \beta_2 x_{2i}$ σε σχέση με το μοντέλο $m_1 : \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ και μη έχοντας εκ-των-προτέρων πληροφορία θεωρούμε ισοπίθανες τις $f(m_1) = f(m_0)$ ως εξής:

$$P_{10} = c^{p_{m_0}-p_{m_1}}\frac{\left|\widehat{\boldsymbol{C}_{\boldsymbol{m_1}}}\right|^{0.5}}{\left|\widehat{\boldsymbol{C}_{\boldsymbol{m_0}}}\right|^{0.5}}\frac{(0.5SS_{m_1}+\lambda_0)^{-\hat{\alpha}}}{(0.5SS_{m_0}+\lambda_0)^{-\hat{\alpha}}} \tag{1.17}$$

όπου p_{m_i} ο αριθμός των παραμέτρων του μοντέλου m_i. Στο συγκεκριμένο παράδειγμα, με την διατροφή των παιδιών έχουμε για τον εκ-των-υστέρων συμπληρωματικό λόγο πιθανοτήτων odds:

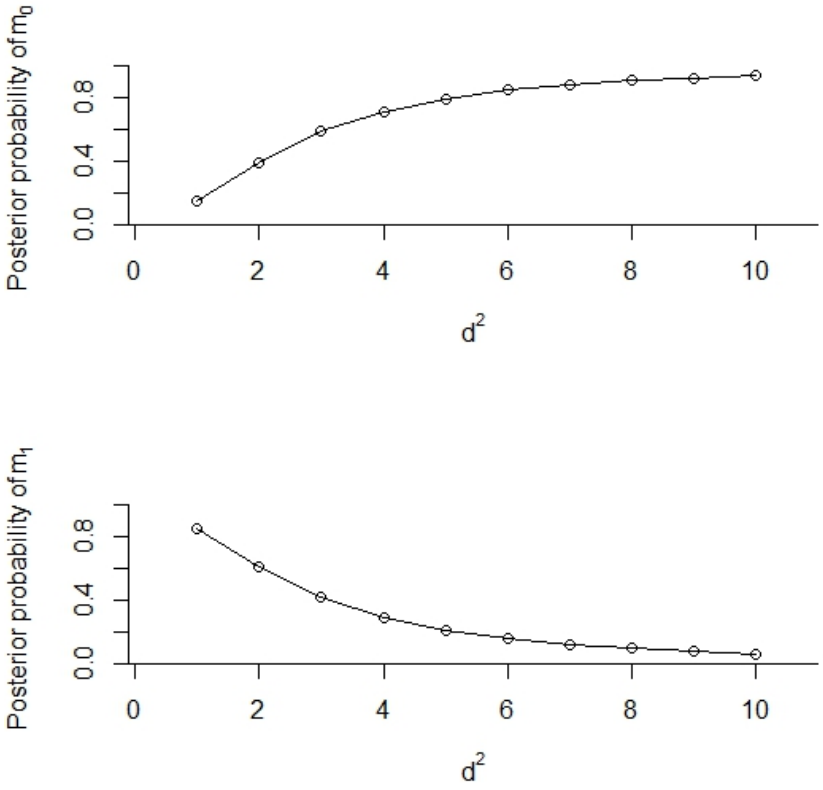
$$P_{01} = d^{p_{m_0}}\frac{\left|\widehat{\boldsymbol{C}_{\boldsymbol{m_1}}}\right|^{0.5}}{\left|\widehat{\boldsymbol{C}_{\boldsymbol{m_0}}}\right|^{0.5}}\frac{(0.5SS_{m_1}+\lambda_0)^{-\hat{\alpha}}}{(0.5SS_{m_0}+\lambda_0)^{-\hat{\alpha}}} \tag{1.18}$$

Απο την εξίσωση 1.18 μπορούμε να δούμε ότι ο εκ-των-υστέρων λόγος συμπληρωματικών πιθανοτήτων odds του μοντέλου m_1 σε σχέση με το m_0 εξαρτάται αποκλειστικά απο την εκ-των-προτέρων υπερπαράμετρο διακύμανσης της παραμέτρου β_1 που μας δείχνει την επίδραση του ύψους. Προκειμένου να δούμε αν όντως η επίδραση β_1 της ανεξάρτητης μεταβλητής είναι σημαντική για το ύψος των παιδιών κάναμε την ανάλυση δεδομένων για διάφορες τιμές της σταθεράς d^2 στον πίνακα 1.8 και είδαμε αν τελικά επηρεάζει τον εκ-των-υστέρων συμπληρωματικό λόγο πιθανοτήτων odds.

Πίνακας 1.8: Οι λογαριθμηκές περιθώριες πιθανοφάνειες, οι εκ-των-υστέρων πιθανότητες των μοντέλων και ο παράγοντας Bayes των μοντέλων m_0, m_1

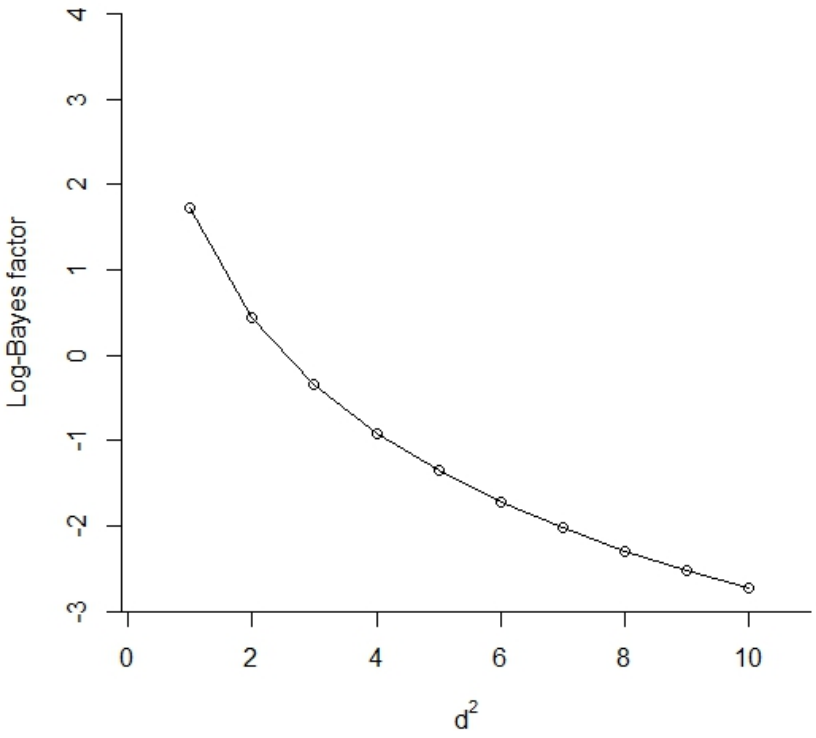
Τιμές σταθεράς	Λογαριθμική περιθώρια πιθανοφάνεια		Εκ-των-υστέρων πιθανότητα μοντέλων		Λογαριθμισμένος παράγοντας **Bayes**
d^2	m_0	m_1	m_0	m_1	$log_e BF_{10}$
1	-35.92	-34.19	0.1512043	0.8487957	1.72
2	-37.31	-36.86	0.3906697	0.6093303	0.44
3	-38.12	-38.46	0.5856641	0.4143359	-0.34
4	-38.69	-39.61	0.7138679	0.2861321	-0.91
5	-39.14	-40.49	0.795303	0.204697	-1.35
6	-39.50	-41.22	0.8481312	0.1518688	-1.72
7	-39.81	-41.84	0.8836259	0.1163741	-2.02
8	-40.08	-42.37	0.9083435	0.0916564	-2.29
9	-40.31	-42.84	0.9261262	0.0738737	-2.52
10	-42.85	-47.34	0.9889158	0.0110842	-4.49

Πρέπει να επισημανθεί ότι εφόσον δεν υπήρχε διαθέσιμη εκ-των-προτέρων πληροφορία για τα μοντέλα m_0 και m_1, τα δύο μοντέλα θεωρούνται εκ-των-προτέρων ισοπίθανα με εκ-των-προτέρων πιθανότητες $f(m_0) = f(m_1) = 0.5$ και ο εκ-των-υστέρων λόγος συμπληρωματικών πιθανοτήτων (odds) είναι ίσος με το παραγόντα Bayes για το μοντέλο m_0 σε σχέση με το μοντέλο m_1. Σύμφωνα με τον Πίνακα 1.8 παρατηρούμε ότι ο παράγοντας Bayes για τις τιμές της εκ-των -προτέρων διακύμανσης d^2 απο 1-2 υπερισχυεί για το πλήρες μοντέλο m_1 έναντι του m_0, αντίθετα για εύρος τιμών 3-10 υπερισχυεί με μεγάλη διαφορά για το μοντέλο m_0 έναντι του m_1, βασισμένοι στους πίνακες 1.5 και 1.6 των Kass και Raftery της προηγούμενης ενότητας άρα για μεγάλες τιμές ενεργοποιείται το παράδοξο. Ακόμα στο σχήμα 1.3 έγιναν τα διαγράμματα για τις εκ-των-υστέρων πιθανότητες των μοντέλων m_0 και m_1 ως συναρτήση συγκεκριμένων τιμών της εκ-των -προτέρων διακύμανσης d^2.



Διάγραμμα 1.3: Διαγράμματα των εκ-των-υστέρων πιθανοτήτων των μοντέλων m_0 και m_1 για τιμές της εκ-των -προτέρων διακύμανσης d^2 απο 1-10

Απο τα παραπάνω συμπεραίνουμε ότι αρχικά για χαμηλές τιμές της σταθεράς d^2 επιλέγεται ως καλύτερο μοντέλο το m_1 και καθώς αυξάνεται η τιμή της εκ-των-προτέρων διακύμανσης d^2 αυξάνει κατα πολύ την επιλογή του μειωμένου μοντέλου m_0 έναντι του m_1. Επιπλέον πληροφορίες μπορούμε να έχουμε απο το γράφημα του παράγοντα Bayes σε συνάρτηση με την εκ-των -προτέρων διακύμανση d^2 στο σχήμα 1.4.



Διάγραμμα 1.4: Διάγραμμα log-Bayes factor του μοντέλου m_0 σε σχέση με το μοντέλο m_1 για τιμές της εκ-των-προτέρων διακύμανσης d^2 απο 1-10

Απο το παραπάνω διάγραμμα μπορούμε να συμπεράνουμε ότι καθώς αυξάνει η σταθερά d^2 που καθορίζει την εκ-των-προτέρων διακύμανση ο παράγοντας Bayes του μοντέλου m_1 έναντι m_0 μειώνεται υποστηρίζοντας το μειωμένο μοντέλο m_0 έναντι του πλήρες μοντέλου m_1 .

1.8 Συμπεράσματα

Σε αυτό το κεφάλαιο προσπαθήσαμε να συνοψίσουμε τα βασικότερα σημεία της Μπεϋζιανής ανάλυσης. Είδαμε ότι όλη η Μπεϋζιανή θεωρία βασίζεται στον καθορισμό τριών κατανομών: την εκ-των-προτέρων κατανομή, την πιθανοφάνεια και την εκ-των-υστέρων κατανομή. Όλες οι επακόλουθες εφαρμογές της Μπεϋζιανής θεωρίας είναι άμεσες συνέπειες των σχέσεων μεταξύ αυτών των κατανομών. Επιπλέον, αναφερθήκαμε στις εφαρμογες της Μπεϋζιανής γραμμικής παλινδρόμησης και εφαρμόστηκαν δύο ξεχωριστές εκ-των-προτέρων αναλύσεις: η συζυγής ανάλυση και η δεσμευμένη ανάλυση με την βοήθεια των μεθόδων MCMC. Ακόμα, αναφερθήκαμε στην Μπεϋζιανή σύγκριση μοντέλων που επιτυγχάνεται μέσω των εκ-των-υστέρων πιθανοτήτων του μοντέλου, του παράγοντα Bayes και της περιθώριας πιθανοφάνειας. Τέλος, αναφερθήκαμε στο περίφημο παράδοξο του Barttlet και Lindley που επηρεάζει την εκ-των-υστέρων ανάλυση των παραμέτρων των μοντέλων και το επιβεβαιώσαμε για το απλό γραμμικό μοντέλο. Στο επόμενο κεφάλαιο θα αναφερθούμε στην Μπεϋζιανή επιλογή μεταβλητών, που υπάγεται στην κατηγορία της Μπεϋζιανής επιλογής μοντέλων και είναι το σημαντικότερο αντικείμενο αυτής της διπλωματικής εργασίας με μια πληθώρα εφαρμογών.

Κεφάλαιο 2

Μπεϋζιανή επιλογή μεταβλητών

2.1 Το πρόβλημα της επιλογής μεταβλητών

Ένα απο τα πιο σημαντικά προβλήματα στην στατιστική μοντελοποιήση είναι η επιλογή μεταβλητών που εξηγούν ή προβλέπουν ένα φαινόμενο που μετριέται μέσω της μεταβλητής απόκρισης Y. Προκειμένου να εφαρμοστεί αυτή η διαδικασία, αρχικά θα πρέπει να ορίσουμε ένα κριτήριο αξιολόγησης της αποτελεσματικότητας του μοντέλου έτσι ώστε να επιλέξουμε μεταξύ δυο ή περισσότερων μοντέλων.

Κλασσικές μέθοδοι επιλογής μεταβλητών έχουν αναπτυχθεί χρησιμοποιώντας ελέγχους σημαντικότητας όπως το t-test και το F-test καθώς και εναλλακτικές μέθοδοι χρησιμοποιώντας κυρίως κριτήρια πληροφορίας (Information Criteria) για την επιλογή μοντέλων.

Οι κλιμακωτές ή βηματικές διαδικασίες επιλογής μοντέλου (forward, backward και Stepwise) είναι ευρέως χρησιμοποιούμενες καθώς είναι εύκολα εφαρμόσιμες ακόμα και σε δεδομένα με σχετικά μεγάλο αριθμό μεταβλητών (Efroymson 1960). Στους ελέγχους σημαντικότητας προβλήματα πηγάζουν όμως με την χρήση αυτών των μεθόδων καθώς σε μεγάλα σέτ δεδομένων τα p-values τέινουν να είναι πολυ μικρά και ο συνεχόμενος υπολογισμός των πολλαπλών ελέγχων σημαντικότητας αυξάνει το σφάλμα τύπου I, με αποτέλεσμα να περιλαμβάνονται μη σημαντικές μεταβλητές. Χρησιμοποιώντας αυτές τις μεθόδους επιλέγουμε ενα μοντέλο ως "το πραγματικό" αγνοώντας την αβεβαιότητα. Έτσι εισάγεται μεροληψία στην συμπερασματολογία των εκτιμήσεων γίνεται και υπο-εκτίμηση της αβεβαιότητας σχετικά με την επιλογή.

Άλλο κριτήριο προσαρμοστικότητας που βασίζεται στην ιδέα της ελαχιστοποιήσης του μέσου τετραγωνικού αθροίσματος των καταλοίπων MSE είναι το στατιστικό C_p (Mallows 1973) που ορίζεται ως εξής:

$$C_p = \frac{SSE}{\hat{\sigma}^2_{full}} + 2p - n$$

όπου SSE είναι το άθροισμα των τετραγώνων των καταλοίπων και $\hat{\sigma}^2_{full}$ η εκτίμηση της διακύμανσης του πλήρους μοντέλου. Ο Mallows παρακινημένος απο τις προβλετικές ικανότητες του κριτηρίου, πρότεινε τα γραφήματα του κριτηρίου για την επιλογή μοντέλου (George 2000). Μια άλλη ποσότητα προσαρμογής μοντέλου είναι ο συντελεστής προσδιορισμού R^2 ο οποίος δείχνει την προβλεπτική ικανότητα των ανεξάρτητων μεταβλητών. Ορίζεται ως εξής:

$$R^2 = 1 - \frac{SSE}{SST}$$

όπου SST είναι το συνολικό άθροισμα τετραγώνων της μεταβλητής απόκρισης y. Η κατανομή του συντελεστή προσδιορισμού περιγράφηκε απο τον Zirphile (1975) στην περίπτωση που η y ακολουθεί κανονική κατανομή ανεξάρτητη απο τις ανεξάρτητες μεταβλητές.

Το βασικό μειονέκτημα του R^2 είναι ότι με την οποιαδήποτε εισαγωγή ανεξάρτητης μεταβλητής αυξάνεται η τιμή του ανεξαρτήτως αν έχει πραγματική ή ψευδή συνεισφόρα στην συνολική μεταβλητότητα της μεταβλητής απόκρισης y. Για τον λόγο αυτόν μια ποσότητα που διακρίνει αν είναι σημαντική η συνεισφορά μιας ανεξάρτητης μεταβλητής στο μοντέλο είναι ο σταθμισμένος συντελεστής προσδιορισμού R^2_{adj}.

Εναλλακτικά μπορούμε να χρησιμοποιούμε τα κριτήρια επιλογής μοντέλου που βασίζονται στην ιδέα μεγιστοποιήσης της πιθανοφάνειας. Η οικογένεια αυτων των κριτήρίων περιγράφεται στην γενική της μορφή ως

$$IC = -2 \log f(\theta|y) + d(p,n)$$

όπου η προηγούμενη εξίσωση περιέχει τον φυσικό λογάριθμο της πιθανοφάνειας και μια συνάρτηση $d(p,n)$ ποινής, που καθορίζεται απο τον αριθμό των παραμέτρων p του μοντέλου και τον αριθμό n των παρατηρήσεων. Δηλαδή δεν είναι τίποτα άλλο απο μια ποινικοποιημένη πιθανοφάνεια της οποίας η ποινή καθορίζεται κυρίως απο την διάσταση του μοντέλου. Τα κριτήρια πληροφορίας χρησιμοποιούνται για την σύγκριση μοντέλων διαφορετικού μεγέθους και όχι απαραίτητα φωλιασμένων όπως γενικά στους ελέγχους σημαντικότητας.

Τα πιο σημαντικά απο αυτά τα κριτήρια είναι το κριτήριο πληροφορίας του Akaike AIC (Akaike 1971) και το κριτήριο πληροφορίας του Schwarz BIC (Schwarz 1978). Τα κριτήρια αυτά βασίζονται στην ένοια της απόκλισης (Deviance). Η απόκλιση για ένα μοντέλο m ορίζεται ως

$D(\theta_m) = -2\log f(y|\theta_m)$. Το κριτήριο πληροφορίας του Akaike ορίζεται ως εξής:

$$AIC_m = D(\hat{\theta}_m) + 2p_m$$

όπου $D(\widehat{\theta_m})$ είναι η ελάχιστη τιμή της απόκλισης για το μοντέλο m και p_m είναι ο αριθμός των παραμέτρων του μοντέλου. Το κριτήριο πληροφορίας του Schwarz ορίζεται ως εξής:

$$BIC_m = D(\hat{\theta}_m) + p_m \log n$$

Αντίθετα η αντιμετώπιση της αβεβαιότητας με κατανομές πιθανοτήτων και η απότομη εξέλιξη των ηλεκτρονικών υπολογιστών αποτέλεσε απο τα σπουδαιότερα εφόδια στην Μπεϋζιανή επιλογή μεταβλητών. Οι Μπεϋζιανές μέθοδοι επιλογής μεταβλητών επιτρέπουν έναν σημαντικό τρόπο εισαγωγής της εκ-των-προτέρων πληροφορίας για τις άγνωστες παραμέτρους. Οι O'Hara και Sillanpaa (2009) περιέγραψαν το πρόβλημα επιλογής μεταβλητών ως εξής "σε οποιοδήποτε σέτ δεδομένων, είναι σχεδόν απίθανο ότι οι πραγματικοί συντελεστές παλινδρόμησης είναι είτε μηδέν είτε πολυ μεγάλοι· τα μεγέθη των επιδράσεων των ανεξάρτητων μεταβλητών πλησίαζουν το μηδέν. Έτσι, το πρόβλημα δεν είναι να βρούμε τους συντελεστές που είναι μηδέν, αλλα να βρούμε αυτούς που είναι αρκετά μικροί για να είναι μη σημαντικοί, παγιδεύοντας τους προς το μηδέν". Έτσι μέσω της Μπεϋζιανης προσέγγισης λαμβάνεται υπόψην η αβεβαιότητα του μοντέλου που πηγάζει απο τις εκτιμώμενες παραμέτρους σταθμίζοντας ως προς όλα τα πιθανά μοντέλα μέσω των εκ-των υστέρων πιθανοτήτων των μοντέλων. Επιπλέον, οι Μπεϋζιανές μέθοδοι επιλογής μεταβλητών καθιστούν εφικτή την εξερεύνηση του χώρου όλων των πιθανών μοντέλων μέσω των μεθόδων MCMC καθώς επίσης και την αυτόματη επιλογή βέλτιστου μοντέλου. Το κυριότερο όμως πλεονέκτημα είναι η σύγκριση μη φωλιασμένων μοντέλων.

Αντίθετα, το κυριότερο μειονέκτημα των Μπεϋζιανών μεθόδων, όπως είδαμε στην προηγούμενη ενότητα, είναι ότι οι εκ-των υστέρων πιθανότητες των μοντέλων και ο παράγοντας του Bayes είναι ευαίσθητα στην αύξηση του δείγματος και της διακύμανσης των παραμέτρων Bartlett και Lindley (1957) υποστηρίζοντας όλο και πιο ισχυρά το απλούστερο μοντέλο. Ακόμα ο παράγοντας του Bayes δεν μπορεί να υπολογιστεί αναλυτικά παρα μόνο σε λίγες περιπτώσεις.

Επιπλέον, ο χώρος των μοντέλων σε πραγματικά προβλήματα είναι πολυ μεγάλος κάνοντας την εύρεση των καλύτερων μοντέλων και την εκτίμηση των σχετικών ποσοτήτων ανέφικτη.

Η χρήση μεθόδων MCMC για την εύρεση του χώρου των μοντέλων χωρίς την πλήρη απαρίθμηση και υπολογισμό όλων των περιθωρίων πιθανοφάνειων αποφεύγει (έστω και μερικώς) πολλά απο αυτά τα προβλήματα. Οι πιο γνωστοί αλγόριθμοι Μπεϋζιανής επιλογής μεταβλητών είναι η επιλογή μεταβλητών με το δειγματολήπτη Gibbs (Gibbs variable selection) των Dellaportas κ.α (2002), η στοχαστική διερεύνηση επιλογής μεταβλητών (Stochastic search variable) των George

και McCulloch (1993) και ο δειγματολήπτης Gibbs των Kuo και Mallick (1998).

Άλλοι πιο γενικοί αλγόριθμοι επιλογής μοντέλων είναι η μέθοδος των Carlin και Chib (1995), η μέθοδος MCMC αναστρέψιμου άλματος (Reversible jump MCMC, Green, 1995) και η κατά Μετρόπολις εκδοχή της μεθόδου Carlin και Chib (Metropolized Carlin and Chib) όπως περιφράφηκε απο τον (Dellaportas et al. 2002).

2.2 Βασικές ιδέες επιλογής μεταβλητών

Είναι ευρέως αποδεκτό οτι η Μπεϋζιανή ανάλυση ξεκινάει με το προσδιορισμό εκ-των-προτέρων κατανομών για τις άγνωστες παράμετρους του υπο-εξέταση μοντέλου.

Όπως είδαμε παραπάνω, η αβεβαιότητα πηγάζει όχι μόνο απο τις ανεξάρτητες μεταβλητές που θα πρέπει να εισαχθούν στο μοντέλο αλλά και απο ολόκληρο το μοντέλο συμπεριλαμβάνοντας και την αβεβαίοτητα των παραμέτρων.

Για τις μεθόδους επιλογής μεταβλητών, σύμφωνα με τους George και McCulloch (1993, 1997), υποθέτουμε λανθάνων διάνυσμα $\boldsymbol{\gamma}$ τέτοιο ώστε $\gamma \in (0,1)^p$ όπου το $\boldsymbol{\gamma}$ περιέχει δείκτες εισαγωγής για όλες τις πιθανές ανεξάρτητες μεταβλητές p. Ας υποθέσουμε ότι εξετάζαμε το γραμμικό μοντέλο της μορφής:

$$\boldsymbol{\mu} = \sum_{j=0}^{p} \gamma_j X_j \beta_j \tag{2.1}$$

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I_n})$$

όπου X_j είναι η j-στήλη του πίνακα σχεδιασμού και β_j το j-στοιχείο του διανύσματος $\boldsymbol{\beta}$ των παραμέτρων του πλήρες μοντέλου, της j-μεταβλητής που θα εισαχθεί στον γραμμικό συνδιασμό 2.1, β_0 είναι η σταθερά και X_0 ισούται με την πρώτη στήλη του πίνακα σχεδιασμού όπου συμπληρώνεται με μονάδες, $\boldsymbol{\mu}$ το διάνυσμα των μέσων τιμών και $\boldsymbol{I_n}$ ο μοναδιαίος πίνακας.

Σύμφωνα με τον Ntzoufras (2011), ορίζουμε τις ποσότητες του $\boldsymbol{\gamma}$ τέτοιες ώστε $(\boldsymbol{\beta_\gamma}, \boldsymbol{\beta_{-\gamma}})$ ξεχωρίζοντας έτσι τις μεταβλητές που εισάγονται με $(\gamma = 1)$ και αυτές που δεν περιλαμβάνονται με $(\gamma = 0)$. Ακόμα καθε πιθανό μοντέλο απο το σύνολο των μοντέλων που είναι 2^p αντιπροσωπεύεται απο μια τιμή του διανύσματος $\boldsymbol{\gamma}$. Αφού επιλέξουμε ενα μοντέλο $\boldsymbol{\gamma}$, έχουμε το ακόλουθο απλό γραμμικό μοντέλο:

$$\boldsymbol{\mu_\gamma} = \boldsymbol{X_\gamma \beta_\gamma} \tag{2.2}$$

$$\boldsymbol{y} \sim N(\boldsymbol{X_\gamma \beta_\gamma}, \sigma^2 \boldsymbol{I_\gamma})$$

όπου $\boldsymbol{\beta_\gamma}$ περιέχει τις μη μηδενικές επιδράσεις του διανύσματος $\boldsymbol{\beta}$ και $\boldsymbol{X_\gamma}$ ο πίνακας σχεδιασμού που περιέχει τις στήλες που αντιστοιχούν στις μη μηδενικές επιδράσεις του διανύσματος $\boldsymbol{\beta}$ και $\boldsymbol{\mu_\gamma}$ το διάνυσμα των μέσων τιμών των επιδράσεων που έχουν $\gamma = 1$.

Ο γραμμικός συνδιασμός 2.1 χρησιμοποιείται στις μέθοδους τόσο των Kuo και Mallick (1998) όσο και του Gibbs variable selection (Ntzoufras κ.α 2002). Εναλλακτικά της χρήσης των δεικτών εισαγωγής των μεταβλητών $\boldsymbol{\gamma}$ μέσα απο τον γραμμικό συνδιασμό 2.1, μπορούμε να χρησιμοποιήσουμε την προσέγγιση των George και McCullogh (1993), δηλαδή στην στοχαστική διερεύνηση επιλογής μεταβλητών, αντικαθιστώντας το $\gamma_j \beta_j$ με το β_j. Σύμφωνα με την προσέγγιση αυτή το διάνυσμα των παραμέτρων περιέχει τις επιδράσεις των ανεξάρτητων μεταβλητών και η παράμετρος $\boldsymbol{\gamma}$ λαμβάνεται υπόψην μόνο στην εκ-των-προτέρων απο κοινού κατανομή που υπολογίζεται. Στην περίπτωση που μια μεταβλητή δεν είναι σημαντική τότε δεν αφαιρείται απο το μοντέλο αλλά συρικνώνεται προς το μηδέν μέσω πληροφοριακών εκ-των-προτέρων κατανομών.

2.3 Βασικές οικογένειες εκ-των-προτέρων κατανομών για την επιλογή μεταβλητών

Μία άλλη μεγάλη οικογένεια εκ-των-προτέρων κατανομών που χρησιμοποιείται ευρέως για την επιλογή μεταβλητών είναι οι (Dirac Spikes-Slab) εκ-των-προτέρων κατανομές που περιλαμβάνουν την ανεξάρτητη κατανομή εκ-των-προτέρων κατανομή, την g εκ-των-προτέρων κατανομή του (Zellner 1980), η οποία έχει πολύ ενδιαφέρουσες ιδιότητες και διευκολύνει τους υπολογισμούς στην επιλογή μεταβλητών και η δυναμική (power) εκ-των-προτέρων κατανομή. Επιπλέον, οι γενικέυσεις των g εκ-των-προτέρων κατανομών περιλαμβάνουν την υπερ g εκ-των -προτέρων κατανομή (Cui και George 2007) και την Zellner-Siow εκ-των-προτέρων κατανομή των (Zellner και Siow 1980).

2.3.1 Συζυγείς κατανομές

Είδαμε στην προηγούμενη ενότητα ότι οι συζυγείς κατανομές για ενα συγκεκριμένο μοντέλο πιθανοφάνειας δίνουν την δυνατότητα στην εκ-των-υστέρων κατανομή να έχει ακριβώς την ιδια αλγεβρική μορφή με την εκ-των-προτέρων κατανομή. Έστω πάλι, το γραμμικό μοντέλο σύμφωνα με την εξίσωση 2.2 όπου οι άγνωστες παράμετροι είναι το διάνυσμα $\boldsymbol{\beta}$, η διακύμανση σ^2 και η παράμετρος εισαγωγής $\boldsymbol{\gamma}$, τότε για την απο κοινού κατανομή των παραμέτρων έχουμε:

$$f(\boldsymbol{\gamma},\boldsymbol{\beta},\sigma^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\gamma},\boldsymbol{\beta},\sigma^2)f(\boldsymbol{\gamma},\boldsymbol{\beta},\sigma^2). \tag{2.3}$$

Προκειμένου να ορίσουμε την μορφή της απο κοινού εκ-των προτέρων κατανομής $f(\boldsymbol{\gamma},\boldsymbol{\beta},\sigma^2)$ υποθέτουμε την μορφή:

$$f(\boldsymbol{\gamma},\boldsymbol{\beta},\sigma^2) = f(\sigma^2,\boldsymbol{\beta}|\boldsymbol{\gamma})\prod_{j=0}^{m} f(\gamma_j), \tag{2.4}$$

όπου γ_j είναι δίτιμη μεταβλητή και μια απευθείας επιλογή εκ-των-προτέρων κατανομής είναι

$$f(\gamma_j = 1) = \pi, \quad j = 1,\ldots,m, \tag{2.5}$$

όπου π είναι μια προεπιλεγμένη πιθανότητα εισαγωγής μεταξύ μηδέν και ένα. Για τις επιδράσεις του διανύσματος $\boldsymbol{\beta}$ που είναι διαφορετικές απο το μηδέν, $\boldsymbol{\beta_\gamma}$ και για την παράμετρο σ^2 χρησιμοποιούμε συζυγείς εκ-των-προτέρων κατανομές αντίστοιχα $\boldsymbol{\beta_\gamma}|\sigma^2$, $\gamma \sim N(\boldsymbol{\mu_{\beta\gamma}}, c^2\boldsymbol{I_\gamma}\sigma^2)$ και $\sigma^2 \sim Igamma(\alpha_0,\lambda_0)$.

2.3.2 Δεσμευμένες συζυγείς κατανομές

Στο προηγούμενο κεφάλαιο είδαμε ότι οι δεσμευμένες συζυγείς κατανομές έχουν ιδιαίτερο ρόλο στην προσομοίωση της απο κοινού εκ-των-υστέρων κατανομής των παραμέτρων μέσω του δειγματολήπτη Gibbs.

Πιο συγκεκριμένα, είδαμε ότι οι δεσμευμένες συζυγείς εκ-των-υστέρων κατανομές των παραμέτρων ανήκουν στην ίδια οικογένεια με τις αντίστοιχες εκ-των-προτέρων κατανομές των παραμέτρων. Στην περίπτωση του γραμμικού μοντέλου σύμφωνα με τις προηγούμενες ενότητες μπορούμε να προσομοιώσουμε δείγμα απο την σχέση 2.4 καθορίζοντας την απο κοινού εκ-των-προτέρων κατανομή των παραμέτρων $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ και σ^2 με βάση την σχέση 2.5.

Άρα σύμφωνα με την δεσμευμένη συζυγή ανάλυση έχουμε για τις επιδράσεις του διανύσματος $\boldsymbol{\beta}$ που είναι διαφορετικές απο το μηδέν, $\boldsymbol{\beta_\gamma}$, για την παράμετρο σ^2 και $\boldsymbol{\gamma}$ εκ-των προτέρων κατανομές αντίστοιχα $\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma} \sim N(\boldsymbol{\mu_{\beta\gamma}}, c^2\boldsymbol{I_\gamma})$ και $\sigma^2 \sim Igamma(\alpha_0, \lambda_0)$ και $f(\gamma_j = 1) = \pi, \quad j = 0, \ldots, p.$

2.3.3 **Dirac Spikes-Slabs** εκ-των-προτέρων κατανομές

Εισάγοντας την δείκτρια μεταβλητή εισαγωγής, γ_j, στο απλο γραμμικό μοντέλο παλινδρόμησης, έχουμε ότι η εκ-των προτέρων κατανομή για το διάνυσμα $\boldsymbol{\beta}$ των συντελεστών παλινδρόμησης είναι ενα παράδειγμα μιας Spike-Slab εκ-των-προτέρων κατανομής.

Μια Spike-Slab εκ-των-προτέρων κατανομή είναι μείξη δύο κατανομών. Μιάς ακίδας (Spike) και μιάς επίπεδης (Slab) κατανομής αντίστοιχα, όπου η κατανομή ακίδας συγκεντρώνεται πολύ κοντά στο μηδέν ή ακριβώς στο μηδέν με μάζα συγκεντρωμένη γύρω απο το μηδέν και η επίπεδη κατανομή εκφράζει την πραγματική μη μηδενική εκ-των -προτέρων κατανομή. Οι Mitchell και Beauchamp (1988) εισήγαγαν αυτήν την εκ-των-προτέρων κατανομή για να διευκολύνει την επιλογή μεταβλητών βάζοντας περιορισμό στις επιδράσεις των ανεξάρτητων μεταβλητών να είναι μηδέν ή όχι. Πιο συγκεκριμένα η εκ-των-προτέρων κατανομή μπορεί να γραφτεί ως εξής:

$$f(\beta_j|\gamma_j) = \gamma_j f_{slab}(\beta_j) + (1 - \gamma_j) f_{spike}(\beta_j)$$

Θεωρούμε δηλαδή μια δείκτρια μάζας στο μηδεν (Spike) και κανονικές κατανομές ως επίπεδες (Slab):

$$f_{slab}(\boldsymbol{\beta}) = N(\boldsymbol{\mu_{\beta\gamma}}, c^{-2}\boldsymbol{I_{p\gamma}}\sigma^2)$$

$$f_{spike}(\beta_j) = I_0(\beta_j)$$

Αναλόγως με την τιμή του γ_j, η επίδραση β_j θα ανήκει είτε στην Spike είτε στην Slab κατανομή. Οι Spike-Slab εκ-των -προτέρων κατανομές μπορούν να κατατάξουν τους συντελεστές παλινδρόμησης σε μηδενικούς και μή-μηδενικούς αναλύοντας τις εκ-των-υστέρων πιθανότητες εισαγωγής.

Εναλλακτικά η κατανομή Dirac Spike θα αντικατασταθεί απο μια συνεχή κατανομή με μέση τιμή μηδέν και με πολυ μικρή διακύμανση βλέπε στοχαστική διερεύνηση επολογής μεταβλητών (George και McCullogh 1993).

Θα θεωρήσουμε ειδικές περιπτώσεις την εκ-των-προτέρων παραμέτρων $\boldsymbol{\mu_{\beta\gamma}}$ και $c^{-2}\boldsymbol{I_{p\gamma}}$ της κατανομής $N(\boldsymbol{\mu_{\beta\gamma}}, c^{-2}\boldsymbol{I_{p\gamma}}\sigma^2)$ Slab:

- Ανεξάρτητη εκ-των-προτέρων κατανομή: $\boldsymbol{\mu_{\beta\gamma}}$=0, $c^{-2}\boldsymbol{I_{p\gamma}}$, ανάλογο της ridge παλινδρόμησης
- g εκ-των-προτέρων κατανομή: $\boldsymbol{\mu_{\beta\gamma}}$=0, $g(\boldsymbol{X_\gamma^T X_\gamma})^{-1}$ Liang κ.α (2008)
- Δυναμική εκ-των-προτέρων κατανομή: $\boldsymbol{\mu_{\beta\gamma}} = (\boldsymbol{X_\gamma^T X_\gamma})^{-1}\boldsymbol{X_\gamma^T y_c}$, $\frac{1}{b}(\boldsymbol{X_\gamma^T X_\gamma})^{-1}$ Ibrahim και Chen (2000)

όπου c, g, $\frac{1}{b}$ είναι σταθερές και $\boldsymbol{y_c}$ είναι η κεντροποιημένη μεταβλητή απόκρισης. Όλες οι παραπάνω εκ-των-προτέρων κατανομές είναι συζυγείς κατανομές και επιτρέπουν τον απευθείας υπολογισμό των εκ-των-υστέρων πιθανοτήτων των μοντέλων.

2.3.4 Ιεραρχική εκ-των-προτέρων κατανομή της πιθανότητας εισαγωγής

Για να υπάρχει μεγαλύτερη ευελιξία σε σχέση με μία προκαθορισμένη εκ-των-προτέρων εισαγωγή πιθανότητα $p(\gamma_j = 1) = \pi$, μπορούμε να χρησιμοποιήσουμε ιεραρχική εκ-των -προτέρων κατανομή για την πιθανότητα εισαγωγής μια βήτα κατανομή ως εξής:

$$\pi \sim Beta(c_0, d_0)$$

Έτσι η εκ-των-προτέρων κατανομή για το διάνυσμα $\boldsymbol{\gamma}$ έχει την ακόλουθη βήτα κατανομή:

$$f(\boldsymbol{\gamma}) \sim B(p_\gamma + c_0, p - p_\gamma + d_0)$$

όπου p_γ είναι ο αριθμός των μη-μηδενικών στοιχειών στο μοντέλό $\boldsymbol{\gamma}$ και p ο αριθμός των ανεξάρτητων μεταβλητών. Εάν c_0 και d_0 ισούται με ένα, η εκ-των-προτέρων κατανομή είναι μη-πληροφοριακή, παρόλα αυτά μπορεί να χρησιμοποιηθεί μια εκ-των-προτέρων πληροφοριακή κατανομή. Οι Ley και Steel (2007) έδειξαν οτι η ιεραρχική εκ-των-προτέρων κατανομή είναι προτιμότερη προσέγγιση απο την εκ-των-προτέρων κατανομή με προκαθορισμένες πιθανότητες εισαγωγής.

2.3.5 Ανεξάρτητη εκ-των-προτέρων κατανομή

Στην πιο απλή περίπτωση μιας μη-πληροφοριακής εκ-των -προτέρων κατανομής για τους συντελεστές παλινδρόμησης έχουμε ανεξαρτησία και την ίδια κατανομή ως εξής:

$$f(\boldsymbol{\beta_\gamma}|\sigma^2) = N(0, c^{-2}\boldsymbol{I_\gamma}\sigma^2),$$

όπου c είναι μια σταθερά. Έτσι έχουμε:

$$f(\beta_j) = f(\beta_j|\gamma_j = 1)f(\gamma_j = 1) + f(\beta_j|\gamma_j = 0)f(\gamma_j = 0).$$

Η εκ-των-προτέρων κατανομή για το διάνυσμα $\boldsymbol{\beta}$ των παραμέτρων είναι μείξη μιάς επίπεδης κανονικής κατανομής για $\gamma_j = 1$ και μιάς Spike δείκτριας μάζας συγκεντρωμένη γύρω απο το μηδέν για $\gamma_j = 0$.

2.3.6 g εκ-των-προτέρων κατανομή του **Zellner**

Πρωτού αναφερθούμε στις τεχνικές επιλογής μεταβλητών, θα παρουσιάσουμε την g εκ-των-προτέρων κατανομή του Zellner (1986). Όπως στην περίπτωση της ανεξάρτητης εκ-των-προτέρων κατανομής η g εκ-των -προτέρων κατανομή υποθέτει ότι οι επιδράσεις είναι κεντροποιημένες στο μηδέν. Το 1986 ο Zellner εισήγαγε την g εκ-των- προτέρων κατανομή για την περίπτωση της κανονικής παλινδρόμησης με την μορφή:

$$\boldsymbol{\beta_\gamma}|\sigma^2, \boldsymbol{\gamma} \sim N(0, g\sigma^2(\boldsymbol{X_\gamma^T X_\gamma})^{-1}),$$

όπου $\boldsymbol{X_\gamma}$ είναι ο πίνακας σχεδιασμού που περιέχει τις επιδράσεις του διανύσματος $\boldsymbol{\beta}$ που είναι διαφορετικές απο το μηδέν και η παράμετρος σ^2 ακολουθεί την εκ-των-προτέρων κατανομή του Jeffreys. Χρησιμοποιώντας την (εξίσωση 2.4) προσομοιώνουμε απο την απο κοινού κατανομή των παραμέτρων $\boldsymbol{\beta}$, σ^2 και $\boldsymbol{\gamma}$.

Η συγκεκριμένη κατανομή έγινε ευρέως γνωστή λόγω του εφικτού υπολογισμού της περιθώριας πιθανοφάνειας (Liang κ.α 2008).

Η περιθώρια πιθανοφάνεια μπορεί να εκφραστεί ως συνάρτηση του συντελεστή προσδιορισμού σύμφωνα με την εξής μορφή:

$$f(\boldsymbol{y}|\boldsymbol{\gamma}) = \frac{(1+g)^{\frac{p_\gamma}{2}}\Gamma(\frac{n-1}{2})}{\sqrt{n}\pi^{\frac{(n-1)}{2}}S_\gamma^{\frac{(n-1)}{2}}}$$

όπου $S_\gamma = \frac{y_c^2}{1+g}(1+g(1-R_\gamma))$.

Πολυ μεγάλες τιμές της υπερπαραμέτρου οδήγησαν στην διαδικασία επιλογής μικρότερων μοντέλων έναντι πολυπλοκότερων, αυτη η συμπεριφορά όπως προαναφέρθηκε είναι το παράδοξο των Lindley και Bartlett (1957).

Η υπερπαράμετρος g καθώς $g > 0$ επηρεάζει κατα πολύ τον τρόπο με τον οποίο αλλάζει η εκ-των-προτέρων κατανομή.

Έτσι, ο καθορισμός της υπερπαραμέτρου g είναι ιδιαίτερα σημαντικός και πολλές προσεγγίσεις για τον καθορισμό αυτόν παρουσιάζονται τα τελευταία χρόνια (βλέπε για παράδειγμα τους George και Foster (2000), Cui και George (2008).

Άρα ανάλογα με τις τιμές της υπερπαραμέτρου g εχουμε τις εξής ειδικές κατηγορίες των g εκ-των-προτέρων κατανομών για το απλό γραμμικό μοντέλο:

- Εκ-των-προτέρων κατανομή μοναδιαίας πληροφορίας (Unit information): $g = n$, όπου n ο αριθμός των παρατηρήσεων που αντιστοιχεί στην εκ-των-προτέρων κατανομή (Unit information) των Kass και Wasserman (1995).
- Εκ-των-προτέρων κατανομή (Benchmark): $g = max(n, p^2)$, όπου p ο αριθμός των ανεξάρτητων μεταβλητών που αντιστοιχεί στην εκ-των-προτέρων κατανομή ορόσημο (Benchmark) των Fernandez κ.α (2001).
- Εκ-των-προτέρων-κατανομή πληθωριστικού κινδύνου (Risk inflation criterion): $g = p^2$ των (Foster και George 1994).
- Εμπειρικές Μπεϋζιανές μέθοδοι, (Liang κ.α 2008).
- Μείξη των g εκ-των-προτέρων κατανομών (Liang κ.α 2008) και (Ley και Steel 2012) όπου η g υπερ-παράμετρος θεωρείται τυχαία μεταβλητή κάτω απο αυτές τις προσεγγίσεις.

2.3.7 Δυναμική εκ-των-προτέρων κατανομή

Η βασική ιδέα της δυναμικής power εκ-των-προτέρων κατανομής που εισήχθει για πρώτη φορά απο τους Ibrahim και Chen (2000) είναι η χρήση ενός πηλίκου , b, της πιθανοφάνειας των κεντροποιημένων δεδομένων , $\boldsymbol{y_c}$, για την κατασκευή κατάλληλης εκ-των-προτέρων κατανομής σε σχέση με την ακατάλληλη εκ-των-προτέρων κατανομή $f(\boldsymbol{\beta_\gamma}) \propto 1$. Έτσι η εκ-των- προτέρων κατανομή για την παράμετρο $\boldsymbol{\beta}$ των επιδράσεων που είναι μη μηδενικές γράφεται ως εξής:

$$f(\boldsymbol{\beta_\gamma}|\sigma^2) = N((\boldsymbol{X_\gamma^T X_\gamma})^{-1})\boldsymbol{X_\gamma^T y_c}, \frac{1}{b}(\boldsymbol{X_\gamma^T X_\gamma})^{-1}\sigma^2).$$

Η δυναμική εκ-των-προτέρων κατανομή είναι κεντραρισμένη στην τιμή του εκτιμητή ελαχίστων τετραγώνων, με τον πίνακα διακύμανσης-συνδιακύμανσης να πολλαπλασιάζεται με τον παράγοντα $\frac{1}{b}$.

Για τιμές $0 < b < 1$, συνήθως $b \ll 1$, η εκ-των-προτέρων κατανομή ειναι πιο ευέλικτη απο την κατανομή δειγματοληψίας. Η περιθώρια πιθανοφάνεια μπορεί να γραφτεί με την εξής μορφή:

$$f(\boldsymbol{y}|\boldsymbol{\gamma}) = \frac{b^{\frac{p_\gamma}{2}}\Gamma(\hat{\alpha})\lambda_0^{\alpha_0}}{2\pi^{\frac{(n-1)(1-b)}{2}}}\Gamma(\alpha_0)(\hat{\lambda})^{(\hat{\alpha})},$$

όπου $\hat{\alpha} = \alpha_0 + \frac{(n-1)(1-b)}{2}$ και $\hat{\lambda} = (\lambda_0 + \frac{(1-b)}{2}\boldsymbol{y_c^T y_c} - ((\boldsymbol{X_\gamma^T X_\gamma^T})^{-1}(\boldsymbol{X_\gamma})^T\boldsymbol{y_c})^T((\boldsymbol{X_\gamma^T X_\gamma}^T)^{-1})^{-1}(\boldsymbol{X_\gamma^T X_\gamma})^{-1}(\boldsymbol{X_\gamma})^{\boldsymbol{T}}\boldsymbol{y_c})$.

Η δυναμική εκ-των-προτέρων κατανομή συνδιάζεται με το μέρος της πιθανοφάνειας που απομένει και έτσι οι εκ-των-υστέρων κατανομές των παραμέτρων β και σ^2 γίνονται (Fruhwirth-Schnatter και Tuchler 2008):

$$f(\boldsymbol{\beta_\gamma}|\sigma^2, g, \gamma) = N((\boldsymbol{X_\gamma^T X_\gamma})^{-1}(\boldsymbol{X_\gamma})^T\boldsymbol{y_c}, (X_\gamma^T X_\gamma)^{-1}\sigma^2),$$

$$f(\sigma^2|\boldsymbol{y}, \boldsymbol{\gamma}) = IG(\hat{\alpha}, \hat{\lambda}).$$

2.3.8 Υπερ g εκ-των-προτέρων κατανομή

Οι Liang κ.α (2008) εισήγαγαν την υπερ g εκ-των-προτέρων κατανομή καθιστώντας την υπερπαράμετρο g ως τυχαία μεταβλητή. Η προσέγγιση αυτή καταλήγει σε Μπεϋζιανή πιθανοφάνεια που μπορεί να υπολογιστεί και αναλυτικά σε κλειστή μορφή. Οι Sabanes Bove και Held (2011) σε μια δημοσίευση τους, πρότειναν μια επέκταση της κλασσικής εκ-των-προτέρων κατανομής g για εφαρμογές στα γενικευμένα γραμμικά μοντέλα, σύμφωνα με την οποία η υπερ g εκ-των-προτέρων κατανομή μπορεί να είναι μια συνεχής κατανομή $f(g)$. Καθώς το μέγεθος του δείγματος τείνει στο άπειρο η εκ-των-προτέρων κατανομή $\boldsymbol{\beta_\gamma}$ συγκλίνει στην κανονική κατανομή με μορφή:

$$\boldsymbol{\beta_\gamma}|\sigma^2, g, \boldsymbol{\gamma} \sim N(0, gc\sigma^2(\boldsymbol{X_\gamma W X_\gamma})^{-1}),$$

όπου c σταθερά όπου παίρνει διάφορες τιμές, καθώς $\boldsymbol{W}$ είναι πίνακας διαγώνιος με τα βάρη.

Η μορφή της συνεχούς κατανομής $f(g)$ για την υπερπαράμετρο g παίρνει την μορφή:

$$f(g) = \frac{\alpha - 2}{2}(1+g)^{\frac{-\alpha}{2}}, \quad g > 0.$$

Στην λογιστική παλινδρόμηση, για παράδειγμα, κάτω απο την τιμή $c = 4$ έχουμε τη συνάρτηση σύνδεσμο (logit), για $c = \pi/2$ την συνάρτηση σύνδεσμο (probit) και για $c = e-1$ την συνάρτηση σύνδεσμο (cloglog). Για περιπτώσεις όχι λογιστικής παλινδρόμησης έχουμε $c = 1$. Η οικογένεια αυτών των εκ-των-προτέρων κατανομών μελετήθηκαν επίσης απο τους Cui και George (2008) για το πρόβλημα της επιλογής μεταβλητών στην περίπτωση γνωστής διακύμανσης. Για τιμές $\alpha \leq 2$ η υπερ g εκ-των-προτέρων κατανομή είναι μη ολοκληρώσιμη στο μηδέν.

Για τιμές $1 < \alpha \leq 2$ η περιθώρια πιθανοφάνεια υπολογίζεται σε κλειστή μορφή καθώς η εκ-των-υστέρων κατανομή είναι κατάλληλη. Θα ασχοληθούμε κυρίως για τιμές $\alpha > 2$ της υπερ g εκ-των-προτέρων κατανομής. Επιπλεόν, μπορεί να χρησιμοποιηθεί μια εκ-των-προτέρων κατανομή για τον παράγοντα $\frac{g}{g+1}$ τέτοια ώστε $\frac{g}{g+1} \sim \boldsymbol{Beta}(1, \frac{\alpha}{2} - 1)$. Η εκ-των- υστέρων κατανομή της υπερ παραμέτρου g παίρνει την μορφή:

$$f(g|\boldsymbol{y}) = \frac{p_\gamma + \alpha - 2}{{}_2F_1(\frac{n-2}{2}, 1; \frac{p_\gamma+\alpha}{2}; R_\gamma)^2} \frac{(1+g)^{(n-1-p_\gamma-\alpha)}}{2}(1 + (1-R_\gamma^2)g)^{\frac{-(n-1)}{2}},$$

όπου ${}_2F_1(a, b; c; z)$ είναι η γκαουσιανή υπεργεωμετρική συνάρτηση Abramowitz και Milton (1970).

2.3.9 **Zellner-Siow** εκ-των-προτέρων κατανομή

Στον έλεγχο υποθέσεων για την μέση τιμή ο Jeffreys (1961) απέρριψε τις κανονικές εκ-των-προτέρων κατανομές για τους λόγους που προαναφέραμε στο προηγούμενο κεφάλαιο και βρήκε ότι η Cauchy εκ-των-προτέρων κατανομή είναι η πιο κατάλληλη κατανομή προκειμένου να μην επηρρεάζεται απο τα παράδοξα. Οι (Zellner και Siow 1980) εισήγαγαν πολυμεταβλητές Cauchy εκ-των-προτέρων κατανομές στους συντελέστες παλινδρόμησης βασιζόμενοι στην δουλειά του Jeffreys στο πρόβλημα μίας μέσης τιμής. Έτσι η εκ-των-προτέρων κατανομή Cauchy για την παράμετρο $\boldsymbol{\beta}$ παίρνει την ακόλουθη μορφή:

$$f(\boldsymbol{\beta_\gamma}|\sigma^2) \propto \frac{\Gamma(\frac{p_\gamma}{2})}{\pi^{p_\gamma/2}} \left|\frac{\sigma^2 \boldsymbol{X_\gamma^T X_\gamma}}{n}\right|^{0.5} \left(1 + \frac{\sigma^2 \boldsymbol{\beta_\gamma^T X_\gamma^T X_\gamma \beta_\gamma}}{n}\right)^{\frac{p_\gamma}{2}},$$

ενώ για την εκ-των-προτέρων κατανομή της παραμέτρου σ^2 υποθέτουμε κατανομή $f(\sigma^2) \propto \sigma^2$. Μια απο τις κυριότερες αιτίες που δεν έγινε γνωστή αυτή η οικογένεια κατανομών όπως η g εκ-των- προτέρων κατανομή είναι ότι η περιθώρια πιθανοφάνεια δεν μπορεί να υπολογιστεί σε κλειστή μορφή.

Ακόμα είναι γνωστό οτι η κατανομή Cauchy μπορεί να αναπαρασταθεί ώς μείξη κανονικών γι΄ αυτό τον λόγο χρησιμοποιήθηκε η μείξη των g εκ-των-προτέρων κατανομών με inverse gamma εκ-των-προτέρων κατανομές για την υπερ παράμετρο g, συγκεκριμένα έχουμε:

$$f(\boldsymbol{\beta_\gamma}|\sigma^2) \propto \int N(\boldsymbol{\beta_\gamma}|0, g\sigma^2(\boldsymbol{X_\gamma^T X_\gamma})^{-1}) f(g) dg,$$

όπου

$$f(g) = \frac{(\frac{n}{2})^{0.5}}{\Gamma(0.5)} g^{-3/2} e^{-n/2g}.$$

2.4 Αλγόριθμοι δεικτών επιλογής μεταβλητών

2.4.1 Στοχαστική διερεύνηση επιλογής μεταβλητών

Η στοχαστική διερεύνηση επιλογής μεταβλητών εισήχθει πρώτη φορά απο τους George και McCullogh (1993, 1997). Σύμφωνα με αυτήν την προσέγγιση ο γραμμικός συνδιασμός έχει την εξής μορφή:

$$\eta = \sum_{j=0}^{p} X_j \beta_j.$$

Οι George και McCullogh εισήγαγαν την δίτιμη μεταβλητή γ που δείχνει αν η i-ανεξάρτητη μεταβλητή θα συμπεριληφθεί ή δεν θα συμπεριληφθεί στο μοντέλο $(\gamma_i = 1)$ ή $(\gamma_i = 0)$.

Αν ονομάσουμε m το σύνολο όλων των μοντέλων, αυτό το σύνολο μπορεί να αναπαρασταθεί μέσω του διανύσματος γ για όλες τις πιθανές ανεξάρτητες μεταβλητές και να μας δείχνει ποιες συμπεριλήφθηκαν και ποιες όχι. Κάθε μια ανεξάρτητη μεταβλητή μοντελοποιείται ως μείξη δύο κανονικών κατανομών. Η μια συγκεντρώνεται με πυκνότητα γύρω απο το μηδέν και η άλλη με πυκνότητα μεγάλης διακύμανσης.

Αυτό μπορεί να γίνει χρησιμοποιώντας μια εκ-των-προτέρων κατανομή (O'Hara και Sillanpaa 2009) για κάθε μία παράμετρο β_j. Οι George και McCullogh (1993) πρότειναν μια μίξη δύο κανονικών κατανομών για τις παραμέτρους β_j. Σύμφωνα με αυτόν τον τρόπο οι δίτιμες μεταβλητές συμπεριλαμβάνονται στο μοντέλο και παίρνουν την εξής μορφή:

$$\beta_j|\gamma_j \sim (1-\gamma_j)N(0,\tau_j^2) + \gamma_j N(0, g_j^2\tau_j^2). \tag{2.6}$$

Η εκ-των-προτέρων κατανομή βάση της εξίσωσης 2.6 δεν θέτει τις ανεξάρτητες μεταβλητές ίσες με μηδέν αλλα σε μια περιοχή κοντά στο μηδέν.

Καθορίζοντας μια πολυ μικρή τιμή για το $\tau_j > 0$ επιτυγχάνεται η εκτίμηση της επίδρασης των ανεξάρτητων μεταβλητών β_j, έαν $\gamma_j = 0$, τότε κοντά στο μηδέν. Καθώς $g_j > 1$, τότε αν $\gamma_j = 1$ προκύπτει β_j διάφορο του μηδενός. Έτσι όταν $\gamma_j = 0$, οι εκτιμήσεις των επιδράσεων β_j δεν είναι υποψήφιες για εισαγωγή, ενώ όταν $\gamma_j = 1$ οι εκτιμήσεις των επιδράσεων β_j είναι υποψήφιες για εισαγωγή. Το σημαντικότερο είναι ότι οι διαστάσεις του μοντέλου δεν αλλάζουν. Οι George και McCullogh (1993) περιγράφουν ακριβώς την επιλογή g_j^2 και τ_j^2.

Οι δεσμευμένες εκ-των-υστέρων κατανομές των β_j χρησιμοποιώντας την εξίσωση (5.7) παίρνουνε την μορφή:

$$f(\beta_j|\boldsymbol{y}, \gamma, \beta_{-j}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\beta_j|\gamma_j), \tag{2.7}$$

$$για\ \gamma_j = 1 : f(\beta_j|\boldsymbol{y}, \boldsymbol{\gamma}, \beta_{-j}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\beta_j|\gamma_j), \tag{2.8}$$

$$\gamma\iota\alpha\ \gamma_j = 0 : f(\beta_j|\boldsymbol{y}, \gamma, \beta_{-j}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}). \tag{2.9}$$

Η κατασκευή του μοντέλου είναι τέτοια ώστε να υπάρχει ανεξαρτησία μεταξύ $\boldsymbol{\beta}$ και $\boldsymbol{\gamma}$. Έτσι, οι δεσμευμένες εκ-των- υστέρων κατανομές των δεικτρειών γ_j έχουν την εξής μορφή:

$$\gamma_j|\boldsymbol{\beta}, \gamma_{-j}, \boldsymbol{y} \propto Bernoulli(\frac{O_j}{1+O_j}),$$

όπου

$$O_j = \frac{f(\gamma_j = 1|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})}{f(\gamma_j = 0|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})} = \frac{f(\boldsymbol{\beta}|\gamma_j = 1, \gamma_{-j})}{f(\boldsymbol{\beta}|\gamma_j = 0, \gamma_{-j})} \frac{f(\gamma_j = 1|\gamma_{-j})}{f(\gamma_j = 0|\gamma_{-j})}. \tag{2.10}$$

2.4.2 Ο δειγματολήπτης **Gibbs** των **Kuo** και **Mallick**

Η πιο άμεση προσέγγιση για την εφαρμογή της επιλογής μεταβλητής είναι η χρήση εκ-των προτέρων κατανομών (Spike-Slab) θέτωντας ως επίπεδη κατανομή (Slab) $\beta_j|(\gamma_j = 1)$ τον αριθμό των παραμέτρων που εισάγονται στο μοντέλο και ως κατανομή ακίδα (Spike) $\beta_j|(\gamma_j = 0)$ των αριθμό των παραμέτρων που είναι μηδέν. Η εκ-των-προτέρων απο κοινού κατανομή των παραμέτων μπορεί να γραφτεί με βάση την εξίσωση (2.2).

Ο δειγματολήπτης Gibbs των Kuo και Mallick (1998) θεωρεί απαραίτητη προϋπόθεση ότι οι δείκτριες εισαγωγής γ_j είναι ανεξάρτητες των επιδράσεων β_j, δηλαδή $f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = f(\boldsymbol{\gamma})f(\boldsymbol{\beta})$. Διασπώντας το διάνυσμα $\boldsymbol{\beta}$ σε $(\boldsymbol{\beta_\gamma}, \boldsymbol{\beta_{-\gamma}})$ έχουμε $f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = f(\gamma)f(\beta_\gamma|\beta_{-\gamma})$.

Ως συνέπεια της ανεξαρτησίας έχουμε ότι οι δεσμευμένες εκ-των-υστέρων κατανομές για τα β_j και γ_j είναι:

$$\gamma\iota\alpha\ \gamma_j = 1 : f(\beta_j|\boldsymbol{y}, \gamma, \beta_{-j}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\beta_j|\beta_{-j}), \tag{2.11}$$

$$\gamma\iota\alpha\ \gamma_j = 0 : f(\beta_j|\boldsymbol{y}, \boldsymbol{\gamma}, \beta_{-j}) \propto f(\beta_j|\beta_{-j}). \tag{2.12}$$

Παρατηρούμε ότι οταν η επίδραση β_j εκτιμάται με μηδέν, δηλαδή $\gamma_j = 0$, οι προτεινόμενες τιμές βασίζονται αποκλειστικά στην ποσότητα $f(\beta_j|\beta_{-j})$.

Η δεσμευμένη εκ-των-υστέρων κατανομή των δεικτριών γ_j δίνεται απο:

$$\boldsymbol{\gamma_j}|\boldsymbol{\beta}, \gamma_{-j}, \boldsymbol{y} \propto Bernoulli(\frac{O_j}{1+O_j}),$$

όπου

$$O_j = \frac{f(\gamma_j = 1|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})}{f(\gamma_j = 0|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})} = \frac{f(\boldsymbol{\beta}|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta})}{f(\boldsymbol{\beta}|\gamma_j = 0, \gamma_{-j}, \boldsymbol{\beta})} \frac{f(\gamma_j = 1|\gamma_{-j})}{f(\gamma_j = 0|\gamma_{-j})}. \tag{2.13}$$

Στην εξίσωση 2.13 μπορούμε να δούμε ότι η εκ-των-προτέρων κατανομή των επιδράσεων $\boldsymbol{\beta_j}$ δεν συμπεριλαμβάνεται στον υπολογισμό της εκ-των- υστέρων πιθανότητας εξαιτίας της υπόθεσης της ανεξαρτησίας και εξαρτάται μόνο απο την πιθανοφάνεια. Η μέθοδος των Kuo και Mallick είναι εύκολο να εφαρμοστεί αλλά πρέπει να προσέχει κανείς ότι ο αλγόριθμος θεωρεί μια εκ-των-προτέρων κατανομή $f(\boldsymbol{\beta})$ για όλα τα μοντέλα. Μια τέτοια υπόθεση είναι περιοριστκή. Παρόλο που ο αλγόριθμος ειναί πολύ απλός για να εφαρμοστεί, δεν υπάρχει τρόπος βελτίωσης της ακρίβειας για τον ερευνητή.

2.4.3 Η επιλογή μεταβλητών με το δειγματολήπτη **Gibbs**

Οι Dellaportas κ.α (2002) πρότειναν τον δειγματολήπτη Gibbs επιλογής μεταβλητών, παρόμοια με τον δειγματολήπτη Gibbs των Kuo και Mallick, που η κύρια διαφορά είναι ότι δεν υπάρχει ανεξαρτησία μεταξύ των επιδράσεων β_j και των δεικτριών εισαγωγής γ_j και χρησιμοποιείται το ιεραρχικό σχήμα $f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = f(\boldsymbol{\gamma}) f(\boldsymbol{\beta}|\boldsymbol{\gamma})$. Θεωρώντας ότι το $\boldsymbol{\beta}$ το διαμερίζουμε στις ποσότητες $(\boldsymbol{\beta_\gamma}, \boldsymbol{\beta_{-\gamma}})$ έχουμε $f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = f(\boldsymbol{\gamma}) f(\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma}) f(\boldsymbol{\beta_{-\gamma}}|\boldsymbol{\beta_\gamma}, \boldsymbol{\gamma})$.

Για τις δεσμευμένες εκ-των-υστέρων κατανομές για τα β_j και γ_j έχουμε ως εξής:

$$\gamma\iota\alpha\ \gamma_j = 1 : f(\beta_j|\boldsymbol{y}, \boldsymbol{\gamma}, \beta_{-j}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma}) f(\boldsymbol{\beta_{-\gamma}}|\boldsymbol{\beta_\gamma}, \boldsymbol{\gamma}), \tag{2.14}$$

$$\gamma\iota\alpha\ \gamma_j = 0 : f(\beta_j|\boldsymbol{y}, \boldsymbol{\gamma}, \beta_{-j}) \propto f(\boldsymbol{\beta_{-\gamma}}|\boldsymbol{\beta_\gamma}, \boldsymbol{\gamma}). \tag{2.15}$$

Η δεσμευμένη εκ-των-υστέρων κατανομή των δεικτριών γ_j δίνεται απο:

$$\gamma_j|\boldsymbol{\beta}, \gamma_{-j}, \boldsymbol{y} \propto Bernoulli(\frac{O_j}{1+O_j}),$$

όπου

$$O_j = \frac{f(\gamma_j = 1|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})}{f(\gamma_j = 0|\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{-j})} = \frac{f(\boldsymbol{\beta}|\gamma_j = 1, \gamma_{-j}, \boldsymbol{\beta})}{f(\boldsymbol{\beta}|\gamma_j = 0, \gamma_{-j}, \boldsymbol{\beta})} \frac{f(\boldsymbol{\beta}|\gamma_j = 1, \gamma_{-j})}{f(\boldsymbol{\beta}|\gamma_j = 0, \gamma_{-j})} \frac{f(\gamma_j = 1|\gamma_{-j})}{f(\gamma_j = 0|\gamma_{-j})}. \tag{2.16}$$

Πρέπει να αναφέρουμε ότι οι ψευδό-εκ-των-προτέρων κατανομές $f(\boldsymbol{\beta_{-\gamma}}|\boldsymbol{\beta_\gamma}, \boldsymbol{\gamma})$ έχουν σημαντικό ρόλο για την βελτίωση της ακρίβειας του αλγόριθμου αλλά όχι στην ίδια την εκ-των-υστέρων κατανομή. Για την βέλτιστη σύγκλιση του αλγορίθμου πρέπει να εκτιμηθούν τις περισσότερες φορές μέσω μιας πιλοτικής μελέτης προκειμένου να χρησιμοποίησουμε την προτεινόμενη κατανομή για την ψευδό-εκ-των-προτέρων κατανομή (Dellaportas κ.α 2002).

2.4.4 Γενικοί αλγόριθμοι επιλογής μεταβλητών

Ας υποθέσουμε ότι έχουμε εναν συγκεκριμένο αριθμό $\boldsymbol{\gamma}$ απο ανταγωνιστικά μοντέλα και τα δεδομένα προκύπτουν απο ενα $\boldsymbol{\gamma} \in (0,1)^p$. Θα επεκτείνουμε την σύγκριση μοντέλων απο δύο ανταγωνιστικά μοντέλα σε περισσότερα απο δύο ανταγωνιστικά μοντέλα. Αν $f(\boldsymbol{\gamma})$ είναι η εκ-των-προτέρων κατανομή του μοντέλου $\boldsymbol{\gamma}$, τότε η εκ-των-υστέρων κατανομή $f(\boldsymbol{\gamma}|\boldsymbol{y}) = \frac{f(\boldsymbol{\gamma})f(\boldsymbol{y}|\boldsymbol{\gamma})}{\sum f(\boldsymbol{\gamma})f(\boldsymbol{y}|\boldsymbol{\gamma})}$, όπου $f(\boldsymbol{y}|\boldsymbol{\gamma}) = \int f(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\beta_\gamma}) f(\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma}) d\boldsymbol{\beta_\gamma}$ είναι η περιθώρια πιθανοφάνεια του $\boldsymbol{\gamma}$ και $f(\boldsymbol{\beta_\gamma})$ είναι η δεσμευμένη κατανομή των $\boldsymbol{\beta_m}$, $\boldsymbol{\gamma}$ με $\boldsymbol{\beta_\gamma} \in B_\gamma$, όπου B_γ είναι όλες οι πιθανές τιμές των συντελεστών παλινδρόμησης των μοντέλων.

Στους περισσότερους αλγόριθμους εύρεσης βέλτιστου μοντέλου χρείαζεται να διερευνηθεί ικανοποιητικά ο χώρος των μοντέλων και των παραμέτρων μέσω των μεθόδων MCMC.

Οι εκ-των-υστέρων κατανομές των μοντέλων μπορούν να υπολογιστούν μέσω της $f(\boldsymbol{\gamma}|\boldsymbol{y})$, καθώς η εκ-των-υστέρων εκτίμηση των συντελεστών παλινδρόμησης κάθε μοντέλου $f(\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma}, \boldsymbol{y})$ είναι εφικτή για τα μοντέλα γ για τα οποία ο αλγόριθμος (Carlin 2001).

Τα τελευταία χρόνια πολλές τεχνικές έχουν προταθεί για να παράγουν τιμές απο την εκ-των-υστέρων κατανομή $f(\boldsymbol{\gamma}, \boldsymbol{\beta_\gamma})$ και έπειτα να εκτιμήσουμε την εκ-των-υστέρων πιθανότητα των μοντέλων και την δεσμευμένη εκ-των-υστέρων κατανομή $f(\boldsymbol{\beta_\gamma}|\boldsymbol{\gamma}, \boldsymbol{y})$.

Ο πιο εύκολος και απευθείας τρόπος να παράγουμε δείγμα απο την απο κοινού εκ-των-υστέρων κατανομή $(\boldsymbol{\beta_\gamma}, \boldsymbol{\gamma})$ είναι ο δειγματολήπτης ανεξαρτησίας. Μια προτεινόμενη τιμή $(\boldsymbol{\gamma_{can}}, \boldsymbol{\beta_{\gamma_{can}}})$ παράγεται απο μία ανεξάρτητη γεννήτρια πιθανοτήτων και η προτεινόμενη τιμή είναι αποδεκτή με πιθανότητα αποδοχής τέτοια ώστε:

$$\alpha = min(1, \frac{f(\boldsymbol{y}|\boldsymbol{\gamma_{can}}, \boldsymbol{\beta_{\gamma_{can}}}) f(\boldsymbol{\beta_{\gamma_{can}}}) f(\boldsymbol{\gamma}^{can}) q(\boldsymbol{\gamma}, \boldsymbol{\beta_\gamma}|\boldsymbol{\gamma_{can}}, \boldsymbol{\beta_{\gamma_{can}}})}{f(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\beta_\gamma}) f(\boldsymbol{\beta_\gamma}) f(\boldsymbol{\gamma}) q(\boldsymbol{\gamma_{can}}, \boldsymbol{\beta_{\gamma_{can}}}|\boldsymbol{\gamma}, \boldsymbol{\beta_\gamma})}).$$

Θα πρέπει να σημειωθεί ότι ο δειγματολήπτης είναι ακριβής όταν η ανεξάρτητη γεννήτρια πιθανοτήτων q προσεγγίζει ικανοποιητικά την κατανομή στόχο. Στην πράξη ο ερευνητής θα πρέπει να εκτιμήσει τις ποσότητες $f(\boldsymbol{\gamma}|\boldsymbol{y})$ και $f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma},\boldsymbol{y})$ για κάθε μοντέλο $\boldsymbol{\gamma}$. Όταν ο αριθμός των πιθανών μοντέλων είναι μεγάλος οι υπολογισμοί γίνονται πολύπλοκοι και ένας διαφορετικός τρόπος εκ-των-υστέρων εκτίμησης θα πρέπει να χρησιμοποιηθεί.

Οι Dellaportas, Forster και Ntzoufras (2002) πρότειναν την κατά Μετρόπολις εκδοχή των Carlin και Chib, ο οποίος περιλαμβάνει ένα επιπλέον βήμα το οποίο βασίζεται σε μια προτεινόμενη κατανομή για την μεταπήδηση απο ένα μοντέλο $\boldsymbol{\gamma}$ σε $\boldsymbol{\acute{\gamma}}$. Έτσι, εισάγοντας τις μεθόδους MCMC στα βήματα επιλογής μοντέλου, η προηγούμενη μέθοδος δειγματοληπτεί μόνο απο την ψευδό-εκ-των-προτέρων κατανομή του μοντέλου $\boldsymbol{\acute{\gamma}}$.

Η μέθοδος αναστρέψιμου άλματος MCMC (Reversible jump MCMC) του Green (1995) είναι ένας αλγόριθμος επιλογής μεταβλητών που μας βοηθά να πάρουμε δείγμα απο τον χώρο των μοντέλων. Το σημαντικότερο είναι ότι έχει την ικανότητα να μετακινείται απο ένα μοντέλο σε ένα άλλο με διαφορετικό αριθμό παραμέτρων ενώ παράλληλα διατηρεί την ικανότητα των Μαρκοβιανών αλυσίδων για την σύγκλιση του αλγορίθμου. Η βασική ιδέα είναι οτι χρησιμοποιείται μια υποψήφια ανεξάρτητη γεννήτρια πιθανοτήτων για να μεταπηδήσουμε απο ένα μοντέλο $\boldsymbol{\gamma}$ σε ενα άλλο μοντέλο $\boldsymbol{\acute{\gamma}}$, δηλαδή χρησιμοποιούνται επιπλέον ψευδοπαράμετροι για να γίνει ταύτιση των διαστάσεων των δύο καταστάσεων της αλυσίδας, δηλαδή της τρέχουσας και της προτεινόμενης.

2.5 Εκ-των-υστέρων ανάλυση αποτελεσμάτων **MCMC**

Ύστερα απο την παραγωγή δειγμάτων απο τις εκ-των υστέρων κατανομές των παραμέτρων του μοντέλου, οι παράμετροι εκτιμώνται από τους εκ-των-υστέρων δειγματικούς μέσους ως εξής:

$$\widehat{\boldsymbol{\beta}_{\boldsymbol{\gamma}}} = \frac{1}{K}\sum_{k=1}^{K} I(\boldsymbol{\beta}_{\boldsymbol{\gamma}}{}^{(k)} = \boldsymbol{\beta}_{\boldsymbol{\gamma}})$$

$$\hat{\sigma}^2_{\gamma} = \frac{1}{K}\sum_{k=1}^{K} I(\sigma_{\gamma}{}^{2(k)} = \sigma_{\gamma})$$

$$\widehat{f}(\gamma_j = 1|\boldsymbol{y}) = \frac{1}{K}\sum_{k=1}^{K} I(\gamma_j)^{(k)} = \boldsymbol{\gamma}, \quad j = 1,\ldots,p$$

Πρέπει να τονιστεί ότι οι εκτιμήσεις των μ, β και σ^2 είναι μέσες τιμές διάφορων απλών γραμμικών μοντέλων και επιλέγονται με δεσμευμένο τρόπο για διάφορες τιμές της δείκτριας μεταβλητής $\boldsymbol{\gamma}$, όπου K είναι ο αριθμός των συνολικών επαναλήψεων και k ο τρέχων αριθμός επαναλήψεων των μεθόδων MCMC. Αυτό είναι γνωστό ως Μπεϋζιανή στάθμιση μοντέλων (Bayesian model averaging).

Η εκ-των-υστέρων πιθανότητα εισαγωγής της ανεξάρτητης μεταβλητής x_j μπορεί να εκτιμηθεί απο την δειγματική μέση τιμή γ_j ή απο την μέση τιμή $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$. Για να γίνει η επιλογή του τελικού μοντέλου πρέπει να χρησιμοποιήσουμε μια απο τις παρακάτω επιλογές:

- Επιλογή του μοντέλου διάμεσης πιθανότητας: στο μοντέλο συμπεριλάμβανονται μόνο οι ανεξάρτητες μεταβλητές με εκ-των-υστέρων πιθανότητα εισαγωγής μεγαλύτερη απο 50%.
- Επιλογή εκ-των υστέρων επικρατέστερου μοντέλου: με την έννοια μεγαλύτερη πιθανότητα εννοούμε την δείκτρια μεταβλητή $\boldsymbol{\gamma}$ που εχεί συμβεί τις περισσότερες φορές στις επαναλήψεις MCMC.

Οι Barbieri and Berger (2004) έδειξαν ότι το μοντέλο εκ-των-υστέρων δίαμεσης πιθανότητας υπερισχυεί έναντι του αντίστοιχου μοντέλου μεγίστης εκ-των-υστέρω πιθανότητας σε όρους προβλεπτικής ικανότητας.

2.6 Επιλογή μεταβλητών με την χρήση του πακέτου *BAS* της *R*

Οι Clyde κ.α (2011) εισήγαγαν ενα διαφορετικό τρόπο δειγματοληψίας απο τον χώρο των μοντέλων. Δημιούργησαν εναν καινούριο τρόπο ανανέωσης των αρχικών πιθανοτήτων κάθε ανεξάρτητης μεταβλητής χρησιμοποιώντας τις εκτιμώμενες περιθώριες πιθανότητες εισαγωγής. Εισάγωντας τα δέντρα αποφάσεων, δειγματοληπτούν μοντέλα απο τον χώρο των μοντέλων χρησιμοποιώντας τεχνικές με υψηλές πιθανότητες εισαγωγής.
Συγκεκριμένα, οι Clyde κ.α (2011) πρότειναν μια καινούργια τεχνική δειγματοληψίας χωρίς επανατοποθέτηση, με την χρήση των δέντρων αποφάσεων. Οι ίδιοι υποστήριξαν ότι το πακέτο *BAS* σε σύγκριση με την απλή δειγματοληψία χωρίς επανατοποθέτηση καταφέρνει να μην δειγματοληπτεί μόνο το 20% του χώρου των μοντέλων, ενώ η απλή δειγματοληψία χωρίς επανατοποθέτηση αφήνει ενα πολυ μεγάλο μέρος του χώρου των μοντέλων (συγκεκριμένα 95%) ως μη δειγματοληπτούμενο. Το οποίο είναι εύκολο να κατανοηθεί λόγω της εντελώς τυχαίας επιλογής των μοντέλων απο τον χώρο των μοντέλων. Επίσης το πακέτο BAS χρησιμοποιεί το μοντέλο εκ-των-υστέρων διάμεσης πιθανότητας σε σχέση με τις άλλες μεθόδους που χρησιμοποιούν το μοντέλο εκ-των-υστέρων υψίστης πιθανότητας (Barbieri και Berger 2004).

Όταν ο αριθμός των ανεξάρτητων μεταβλητών που εισάγονται στο πακέτο BAS ξεπερνάει τις 30, το πακέτο δεν κάνει πλήρη απαρίθμηση των μοντέλων, αλλά χρησιμοποιεί τον αλγόριθμο που περιγράφηκε παραπάνω.

Ο υπολογισμός των αρχικών πιθανοτήτων δεν είναι προφανής. Για αυτόν τον λόγο οι Clyde κ.α (2011), πρότειναν τρείς διαφορετικούς τρόπους υπολογισμού των αρχικών πιθανοτήτων, οι οποίοι ενσωματώνονται στο πακέτο BAS, συγκεκριμένα ομοιόμορφες πιθανότητες, μια στάθμιση με βάση p-value και μια στάθμιση MCMC.

Το πακέτο BAS δίνει την δυνατότητα για πολλές επιλογές στις υπερ παραμέτρους και στις εκ-των-προτέρων κατανομές των μοντέλων. Οι ακόλουθες επιλογές εκ-των-προτέρων κατανομών είναι διαθέσιμες για τις παραμέτρους των μοντέλων: το κριτήριο πληροφορίας του Akaike, η g εκ-των-προτέρων κατανομή, η Zellner's Siow εκ-των- προτέρων κατανομή, η υπερ g εκ-των-προτέρων κατανομή, η υπερ g εκ-των-προτέρων κατανομή με την προσέγγιση του Laplace, οι εμπειρικές τοπικές Μπεϋζιανές μέθοδοι και οι εμπερικές σφαιρικές Μπεϋζιανές μέθοδοι.

Επίσης, για τις εκ-των-προτέρων κατανομές των μοντέλων μπορούμε να χρησιμοποιήσουμε την ομοιόμορφη κατανομή, την διωνυμική και την Βήτα-διωνυμική κατανομή. Τέλος, παρέχει γραφήματα για τον ελέγχο καλής προσαρμογής των δεδομένων από τα μοντέλα και γράφημα των εκ-των-υστέρων πιθανοτήτων εισαγωγής.

2.7 Επιλογή μεταβλητών με την χρήση πακέτου του **WinBugs**

Οι τρεις τεχνικές επιλογής μεταβλητών που χρησιμοποιούν την δείκτρια μεταβλητή $\boldsymbol{\gamma}$ είναι η στοχαστική διερεύνηση μεταβλητών (George και McCulloch 1993), ο δειγματολήπτης Gibbs των Kuo και Mallick (1998) και ο δειγματολήπτης Gibbs επιλογής μεταβλητών (Dellaportas κ.α 2000), οι οποίες μπορούν να εφαρμοστούν χωρίς προβλήματα και στα γενικευμένα γραμμικά μοντέλα για log-γραμμικά μοντέλα (Ntzoufras κ.α 2000) και για μοντέλα πολυμεταβλητής παλινδρόμησης (Brown κ.α 1998). Η βασική ιδέα αυτών αλγορίθμων είναι ότι εντοπίζονται οι πιο πιθανές ανεξάρτητες μεταβλητές με την χρήση των εκ-των-υστέρων πιθανοτήτων. Το βέλτιστο υποσύνολο των ανεξάρτητων μεταβλητών είναι αυτό που εμφανίζεται συχνότερα κατα την πραγματοποιήση των μεθόδων MCMC. Με άλλα λόγια, το βέλτιστο μοντέλο μπορεί να βρεθεί μέσω της ανάλυσης της εκ-των- υστέρων κατανομής της παραμέτρου $\boldsymbol{\gamma}$.

Θα γίνει εφαρμογή των τριών μεθόδων MCMC που περιγράφηκαν παραπάνω προκειμένου να δούμε περαιτέρω αποτελέσματα.

Οι διαφορές μεταξύ των διαφορετικών μεθόδων πηγάζουν απο το πως εισάγεται και χρησιμο-

ποιείται η παράμετρος $\boldsymbol{\gamma}$ στο μοντέλο. Στην στοχαστική διερεύνηση επιλογής μεταβλητών η παράμετρος $\boldsymbol{\gamma}$ λαμβάνει μέρος μόνο στην εκ-των- προτέρων κατανομή των συντελεστών της παλινδρόμησης, στην μέθοδο των Kuo και Mallick, η παράμετρος συμμετέχει μόνο μέσω του γραμμικού συνδιασμού, ενώ ο δειγματολήπτης Gibbs επιλογής μεταβλητών συνδιάζει τα χαρακτηριστικά και των δύο προηγούμενων μεθόδων. Θα παρουσιάσουμε αυτούς τους 3 αλγόριθμους και τις διαφορές τους για ενα μοντέλο απλής παλιδρόμησης για δεδομένα της παχυσαρκίας. Δεν θα λάβουμε υπόψην τυχόν αλληλεπιδράσεις των ανεξάρτητων μεταβλητών καθώς με τον όρο πλήρες μοντέλο θα αναφερόμαστε στο μοντέλο με όλες τις κύριες επιδράσεις.

2.7.1 Παράδειγμα επιλογής μεταβλητών με *R* και **WinBugs**

Σε αυτό το παράδειγμα, θα αναλύσουμε τα δεδομένα που προέρχονται απο την διπλωματική για την διευρεύνηση της παχυσαρκίας στην Έλλαδα (Ιωάννης Κυριάκος 2010).
Ο σκοπός αυτής της διπλωματικής ήτανε η εκτίμηση του επιπολασμού της παχυσαρκίας στους φοιτητές που σπουδάζουν στην ευρύτατη περιοχή της Αττικής και να εντοπιστούν οι πιθανοί παράγοντες που προκαλούν το φαινόμενο αυτό σε ενα νεανικό κομμάτι του πληθυσμού.
Στήν μελέτη αυτή συμπεριλαμβάνονται 608 φοιτητές απο πανεπιστημιακούς χώρους της Αττικής. Επειδή ο αριθμός των μεταβλητών ήτανε πολύ μεγάλος, επιλέξαμε έναν συγκεκριμένο αριθμό μεταβλητών προκειμένου να περιγράψουμε με λίγες μεταβλητές την Μπεϋζιανή επιλογή μεταβλητών.
Τα δεδομένα έχουν $n = 597$ παρατηρήσεις (εξαιρέσαμε 10 φοιτητές των οποίων είχαμε ελλιποϋσες τιμές στο δείγμα). Η μεταβλητή απόκρισης Y είναι ο δείκτης μάζας σώματος, οι ανεξάρτητες μεταβλητές είναι το φύλο (C2) ώς δίτιμη μεταβλητή με κατηγορίες άνδρας, γυναίκα, η σωματική κατάσταση του/της φοιτητή/φοιτήτριας 12-18 (C_11_) ώς τρίτιμη μεταβλητή με κατηγορίες λεπτός/ή, κανονικός/ή, παχύς/ιά, η σωματική κατάσταση μητέρας (C_13_) ως τρίτιμη μεταβλητή με κατηγορίες λεπτή, κανονική, παχιά, η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών (A_49_) ώς τρίτιμη μεταβλητή με κατηγορίες σπάνια/λίγες φορές, αρκετές φορές, πολλές φορές/σχεδόν καθημερινά.
Επιπλεόν για την κατάλληλη μοντέλοποιήση των δεδομένων δημιουργήσαμε ψευδομεταβλητές για τις κατηγορικές μεταβλητές. Συγκρεκριμένα, επειδή η μεταβλητή φύλο (C2) είναι δίτιμη εισήχθη απευθείας στο μοντέλο και η επίδραση της καθορίζει τον εκ-των-υστέρων αναμενόμενο δείκτη μάζας σώματος στους άνδρες σε σχέση με τις γυναίκες (επίπεδο αναφοράς), για την μεταβλητή σωματική κατάσταση του/της φοιτητή/φοιτήτριας 12-18 (C_11_) δημιουργήσαμε τις ψευδομεταβλητές (C_11_2) και (C_11_3) των οποίων οι επιδράσεις είναι αντίστοιχα ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της κανονικής σωματικής κατάστασης του/της φοιτητή/φοιτήτριας 12-18 σε σχέση με την λεπτή σωματική κατάσταση του/της φοιτητή/φοιτήτριας

12-18(κατηγορία αναφοράς) και ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της παχιάς σωματικής κατάστασης του/της φοιτητή/φοιτήτριας 12-18 σε σχέση με την λεπτή σωματική κατάσταση του/της φοιτητή/φοιτήτριας 12-18 (κατηγορία αναφοράς), για την μεταβλητή σωματική κατάσταση μητέρας (C_13_) δημιουργήσαμε τις ψευδομεταβλητές (C_13_2) και (C_13_3) των οποίων οι επιδράσεις είναι αντίστοιχα ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της κανονικής σωματικής κατάστασης της μητέρας του/της φοιτητή/φοιτήτριας σε σχέση με την λεπτή σωματική κατάσταση της μητέρας του/της φοιτητή/φοιτήτριας (κατηγορία αναφοράς) και ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της παχιάς σωματικής κατάστασης της μητέρας του/της φοιτητή/φοιτήτριας σε σχέση με την λεπτή σωματική κατάσταση της μητέρας του/της φοιτητή/φοιτήτριας (κατηγορία αναφοράς), για την μεταβλητή κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών (A_49_) δημιουργήσαμε τις ψευδομεταβλητές (A_49_2) και (A_49_3) των οποίων οι επιδράσεις είναι αντίστοιχα ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της κατανάλωσης συσκευασμένων έτοιμων μαγειρευτών φαγητών αρκετές φορές του/της φοιτητή/φοιτήτριας σε σχέση με την κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών σπάνια/λίγες φορές του/της φοιτητή/φοιτήτριας (κατηγορία αναφοράς) και ο εκ-των-υστέρων αναμενομένος δείκτης μάζας σώματος της κατανάλωσης συσκευασμένων έτοιμων μαγειρευτών φαγητών πολλές φορές/σχεδόν καθημερινά του/της φοιτητή/φοιτήτριας σε σχέση με την κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών σπάνια/λίγες φορές του/της φοιτητή/φοιτήτριας (κατηγορία αναφοράς).

Προκειμένου να αξιολογήσουμε αν όντως οι κατηγορικές μεταβλητές C2, C_11_ , C_13_ A_49_ είναι σημαντικές θα πρέπει οι εκ-των-υστέρων πιθανότητες εισαγωγής των επιμέρους ψευδομεταβλητών κάθε κατηγορικής μεταβλητής να είναι σημαντική. Δηλαδή, το ενδιαφέρον θα επικεντρωθεί στην συνολική επίδραση κάθε κατηγορικής μεταβλητής που προκύπτει απο τις επιμέρους επιδράσεις των ψευδομεταβλητών κάθε κατηγορικής μεταβλητής. Η εφαρμογή έγινε μέσω του πακέτου BAS προκειμένου να γίνει κατάλληλα η επιλογή μεταβλητών για τα δεδομένα παχυσαρκίας.

Τα συγκεκριμένα δεδομένα αποτελούνται απο 7 ψευδομεταβλητές με 597 παρατηρήσεις, άρα έχουμε $2^7 = 128$ πιθανά μοντέλα. Επιπλέον μέσω του πακέτου BAS έγινε πλήρη απαρίθμηση του χώρου των μοντέλων και η οποία οδήγησε στον υπολογισμό των εκ-των-υστέρων ποσοτήτων που μας ενδιαφέρουν. Παρουσιάζουμε αποτελέσματα με βάση το κριτήριο πληροφορίας του Akaike (AIC), Μπεϋζιανό κριτήριο πληροφορίας (BIC), την g εκ-των-προτέρων κατανομή (G-p), την Zellner-Siow εκ-των-προτέρων κατανομή (ZS-p), την υπέρ-g-εκ-των-προτέρων κατανομή (HyperG-p), την υπερ-g-εκ-των-προτέρων κατανομή με προσέγγιση του Laplace (HyperGl-p), τις εμπειρικές τοπικές Μπεϋζιανές μεθόδους (EB-L) και τις τις εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους (EB-G).

Ακόμα, χρησιμοποιήσαμε υπερ εκ-των-προτέρων κατανομή $beta - binomial \sim (1,1)$ για τις εκ-

των-προτέρων πιθανότητες εισαγωγής γ_j. Όλες οι λεπτομέρειες συνοψίζονται στον Πίνακα 2.1 και χρησιμοποιούνται ως αναφορά στα διαγράμματα και τους πίνακες αυτου του παραδείγματος.

Για όλες τις μεθόδους επιλογής μεταβλητών του Πίνακα 2.2 χρησιμοποιήσαμε Μπεϋζιανή στάθμιση μοντέλών για το πλήρες μοντέλο. Όλες οι μέθοδοι συγκρίνονται χρησιμοποιώντας τις εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ κάθε ψευδομεταβλητής, τις εκ-των-υστέρων πιθανότητες μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y_i}$ για τα εκ-των-υστέρων μοντέλα $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$.
Απο τα αποτελέσματα του Πίνακα 2.2 και του Διαγράμματος 3.5 παρατηρούμε για όλες τις μεθόδους επιλογής μεταβλητών στο μοντέλο περιλαμβάνονται το φύλο (C2), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_2, κανονική), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_3, παχιά) καθώς έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y}) > 0.95$. Η σωματική κατάσταση μητέρας (C_13_2, κανονική) και η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_3, πολλές φορές/σχεδόν καθημερινά) είναι σχετικά σημαντικές καθώς έχουν εύρος μέσης εκ-των-υστέρων πιθανότητας εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ 0.50-0.84 και 0.37-0.93 αντίστοιχα.
Αντίθετα, η σωματική κατάσταση μητέρας (C_13_3, παχιά) και η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_2, αρκετές φορές) έχουν πολυ χαμηλές εκ-των-υστέρων πιθανότητες εισαγωγής σε όλες τις μεθόδους με εξαίρεση σε κάποιες περιπτώσεις (οι ψευδομεταβλητές A_49_2 και C_13_3 έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.77 και 0.79 αντίστοιχα στην μέθοδο AIC.
Οι μέθοδοι HyperG-P, HyperGl-P, EB-L, EB-G έχουν παρόμοια συμπεριφορά με αυτήν της μεθόδου ZS-P, όμως οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι συστηματικά υψηλότερες.
Συγκεκριμένα, για τις σημαντικές ψευδομεταβλητές (C2), (C_11_2), (C_11_3), οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι > 0.99, καθώς οι υπόλοιπες δύο ψευδομεταβλητές (C_13_2), (A_49_3) είναι σημαντικές με εκ-των-υστέρων πιθανότητες εισαγωγής 0.64 και 0.66 αντίστοιχα. Επιπλέον, οι μη σημαντικές ψευδομεταβλητές είχαν εύρος μέσης εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ απο 0.38 (για την A_49_2) εώς 0.56 (για την C_13_3).

Οι μέθοδοι BIC, G-P παρουσιάζουν παρόμοια αποτελέσματα όσον αναφορά τις μέσες εκ-των-υστέρων πιθανότητες εισαγωγής. Για τις σημαντικές ψευδομεταβλητές, οι μέσες εκ-των-υστέρων πιθανότητες εισαγωγής είναι παρόμοιες με αυτές των μεθόδων HyperG-P, HyperGl-P, EB-L, EB-G, ενώ για τις μη σημαντικές ψευδομεταβλητές μειώνονται με κατεύθυνση προς το 0.5 με εύρος μέσης εκ-των-υστέρων πιθανότητα εισαγωγής απο 0.14 (για την A_49_2) εώς 0.43 (για την C_13_3).

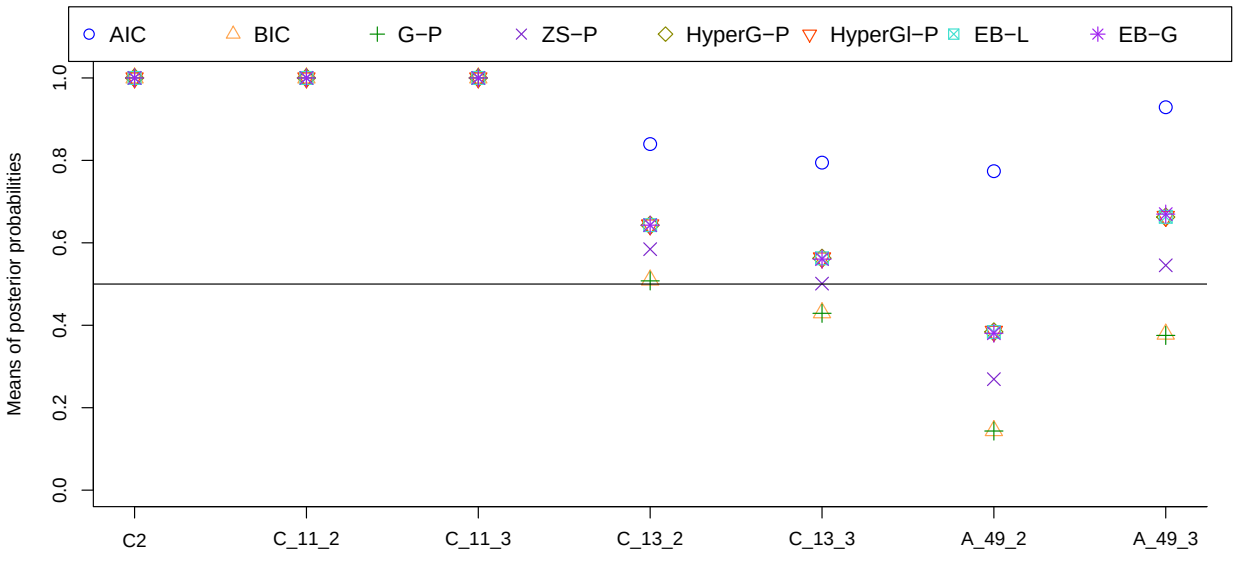
Η μέθοδος AIC έχει μια εντελώς διαφορετική συμπεριφορά απο τις υπόλοιπες μεθόδους επιλογής μεταβλητών. Παρόλο που η συγκεκριμένη μέθοδος δίνει υψηλές πιθανότητες εισαγωγής στις σημαντικές ψευδομεταβλητές (C2), (C_11_2), (C_11_3), (C_13_2), (A_49_3), υποστηρίζει και την εισαγωγή των ψευδομεταβλητών (C_13_3), (A_49_2) με μέσες εκ-των-υστέρων πιθανότητες εισαγωγής 0.79, 0.77 αντίστοιχα. Επιπλέον, η μέθοδος AIC επιλέγει πολύπλοκα μοντέλα και οι πιθανότητες εισαγωγής των μη σημαντικών ψευδομεταβλητών αυξάνονται σε πολυ μεγάλο βαθμό.

Πίνακας 2.1: Συντομογραφίες και λεπτομέρειες για τις μεθόδους

	Συντομογραφία	Μέθοδος
1	AIC	Κριτήριο πληροφορίας Akaike
2	BIC	Μπεϋζιανό κριτήριο πληροφορίας
3	G-P	g εκ-των προτέρων κατανομή για $g = 597$
4	ZS-P	Zellner-Siow εκ-των προτέρων κατανομή
5	HyperG-P	Υπέρ g εκ-των-προτέρων κατανομή για $g = 3$
6	HyperGl-P	Υπερ g εκ-των-προτέρων κατανομή με μέθοδο του Laplace για $g = 3$
7	EB-L	Εμπειρικές τοπικές Μπεϋζιανές μεθόδους
8	EB-G	Εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους

Πίνακας 2.2: Αποτελέσματα επιλογής μεταβλητών μέσω του πακέτου BAS στα οποία αναγράφονται για κάθε μία απο τις ανεξάρτητες μεταβλητές οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 2.1

	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$						
	Ανεξάρτητες μεταβλητές						
Μέθοδος επιλογής μεταβλητών	C2	C_11_2	C_11_3	C_13_2	C_13_3	A_49_2	A_49_3
AIC	1	1	1	0.84	0.79	0.77	0.93
BIC	1	1	1	0.51	0.43	0.14	0.37
G-P	1	1	1	0.50	0.43	0.14	0.37
ZS-P	1	1	1	0.58	0.50	0.27	0.54
HyperG-P	1	1	1	0.64	0.56	0.38	0.66
HyperGl-P	1	1	1	0.64	0.56	0.38	0.66
EB-L	1	1	1	0.64	0.56	0.38	0.66
EB-G	1	1	1	0.64	0.56	0.38	0.66



Διάγραμμα 2.1: Διάγραμμα των μέσων εκ-των-υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 2.1

Απο τα αποτελέσματα του Πίνακα 2.3 παρατηρούμε ότι οι ψευδομεταβλητές (C2), (C_11_2), (C_11_3), (C_13_2), (A_49_3) περιλαμβάνονται σχεδόν σε όλα τα εκ-των-υστέρων μοντέλα $\boldsymbol{\gamma}$ που επιλέγονται απο όλες τις μεθόδους επιλογής μεταβλητών του Πίνακα 2.2. Ακόμα, το μέσο τετραγωνικό σφάλμα RMSE των προβλέπομενων τιμών $\widehat{y_i}$ όλων των μεθόδων επιλογής μεταβλητών είναι παρόμοιας κλίμακας.

Σε όρους RMSE των προβλέπομενων τιμών $\widehat{y_i}$ η χειρότερη μέθοδος είναι η AIC καθώς επιλέγει πολύπλοκα μοντέλα συμπεριλαμβάνοντας μη σημαντικές ψευδομεταβλητές και μικραίνει ψευδώς η αβεβαιότητα (μεγάλη εκ-των-υστέρων πιθανότητα μοντέλου $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ διότι συμπεριλαμβάνει όλες τις ανεξάρτητες μεταβλητές), ενώ καλύτερες στα ίδια RMSE είναι ZS-P, HyperG-P, HyperGl-P, EB-L, EB-G παρόλου που αυξάνεται η αβεβαιότητα (μικρές εκ-των-υστέρων πιθανότητες μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$). Για τις μεθόδους BIC, G-P σε όρους RMSE των προβλέπομενων τιμών $\widehat{y_i}$ είναι ελαφρώς χειρότερες σε σχέση με τις ZS-P, HyperG-P, HyperGl-P, EB-L, EB-G και επιλέγουν απλούστερα μοντέλα μικραίνοντας ψευδώς την αβεβαιότητα του μοντέλου (η τιμή $g = 597$ είναι πολύ μεγάλη και οδηγεί σε αποδοχή απλούστερων μοντέλων μηδενίζοντας τυχόν σημαντικές ψευδομεταβλητές).

Συνοψίζοντας, για όλες τις μεθόδους επιλογής μεταβλητών του Πίνακα 2.1 συμπεραίνουμε ότι οι μεταβλητές φύλο (C2), σωματική κατάσταση του/της φοιτητή/φοιτήτριας 12-18 (C_11_) πρέπει να συμπεριληφθούν στο μοντέλο, καθώς οι ψευδομεταβλητές (C2, άνδρας), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_2, κανονική), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_3, παχιά) περιλαμβάνονται σε όλα τα εκ-των-υστέρων μοντέλα και έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y}) > 0.99$. Ισοδύναμα η συνολική επίδραση των κατηγορικών μεταβλητών C2, C_11_ είναι σημαντική για τον αναμενόμενο δείκτη μάζας σώματος. Αντίθετα, οι μεταβλητές, σωματική κατάσταση μητέρας (C_13_), κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών (A_49_) δεν μπορούν να συμπεριληφθούν στο μοντέλο καθώς μόνο οι ψευδομεταβλητές σωματική κατάσταση μητέρας (C_13_2, κανονική), η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_3, πολλές φορές/σχεδόν καθημερινά) είναι σχετικά σημαντικές καθώς περιλαμβάνονται σε όλα σχεδόν τα εκ-των-υστέρων μοντέλα, ενώ οι ψευδομεταβλητές σωματική κατάσταση μητέρας (C_13_3, παχιά) και η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_2, αρκετές φορές) έχουν χαμηλές εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ και δεν περιλαμβάνονται σε όλα τα εκ-των-υστέρων μοντέλα. Άρα, η συνολική επίδραση των κατηγορικών μεταβλητών A_49_, C_13_ δεν είναι σημαντική για τον αναμενόμενο δείκτη μάζας σώματος.

Πίνακας 2.3: Αποτελέσματα επιλογής μεταβλητών μέσω του πακέτου BAS στα οποία αναγράφονται για κάθε μοντέλο οι εκ-των-υστέρων πιθανότητες $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y}_i$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 2.1

Μέθοδος	$\boldsymbol{\gamma}$	$\widehat{f}(\boldsymbol{\gamma}\|\boldsymbol{y})$	$RMSE(\widehat{y}_i)$
AIC	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2+$		
	$+C\_13\_3 + A\_49\_2 + A\_49\_3$	0.51	4.95
BIC	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2$	0.23	4.92
G-P	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2$	0.23	4.91
ZS-P	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.16	4.91
HyperG-P	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.15	4.89
HyperGl-P	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.15	4.89
EB-L	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.15	4.89
EB-G	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.16	4.89

Σύμφωνα με τους Ntzoufras (2002, 1999, 2011) και Dellaportas κ.α (2002) , ο κώδικας στο Παράρτημα (9.4.1) τροποποιήθηκε καταλλήλως για τον δειγματολήπτη των Kuo και Mallick, την στοχαστική διερεύνηση επιλογής μεταβλητών και τον Gibbs επιλογής μεταβλητών για το απλό γραμμικό μοντέλο.

Χρησιμοποιώντας την επιπλέον ανάλυση στο Winbugs, η ακρίβεια της κάθε παραμέτρου του μοντέλου υπολογίστηκε απο μια πιλοτική μελέτη για το πλήρες μοντέλο, $\beta_j \sim N(0, nS^2_{\beta j})$.
Οι δείκτριες μεταβλητές $\boldsymbol{\gamma}$ κατανέμονται ως $\gamma \sim Bin(0.5)$ για όλες τις ανεξάρτητες μεταβλητές εκτός απο την σταθερά η οποία έτσι και αλλιώς συμπεριλαμβάνεται στο μοντέλο.
Στον Πίνακα 2.4 έχει γίνει η ανάλυση για τις μέσες εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ και για το σφάλμα MC για κάθε μία απο τις ανεξάρτητες μεταβλητές με την χρήση των μεθόδων MCMC οι οποιες περιλαμβάνουν τον δειγματολήπτη των Kuo και Mallick, την στοχαστική διερεύνηση επιλογής μεταβλητών και τον Gibbs επιλογής μεταβλητών 5000 επαναλήψεων με βάση την εκ-των- προτέρων εμπειρική ανεξάρτητη κατανομή των παραμέτρων του μοντέλου.
Συγκεκριμένα, όλες οι μέθοδοι MCMC συγκρίνονται χρησιμοποιώντας τις εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ κάθε ανεξάρτητης μεταβλητής, τις εκ-των-υστέρων πιθανότητες μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y_i}$ για τα εκ-των-υστέρων μοντέλα $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$.
Απο τα αποτελέσματα του Πίνακα 2.4 παρατηρούμε για όλες τις μεθόδους MCMC επιλογής μεταβλητών στο μοντέλο περιλαμβάνονται το φύλο (C2), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_2, κανονική), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_3, παχιά) καθώς έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y}) > 0.95$ (με μοναδική εξαίρεση την μεταβλητή C_11_2 στην στοχαστική διερεύνηση επιλογής μεταβλητών με πιθανότητα εισαγωγής 0.53).
Αντίθετα, η σωματική κατάσταση μητέρας (C_13_2, κανονική), η σωματική κατάσταση μητέρας (C_13_3, παχιά),η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας(A_49_2, αρκετές φορές), η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας(A_49_3, πολλές φορές/σχεδόν καθημερινά) έχουν πολυ χαμηλές εκ-των-υστέρων πιθανότητες εισαγωγής σε όλες τις μεθόδους με εξαίρεση σε κάποιες περιπτώσεις (η μεταβλητή C_13_3 έχει μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.46 στην μέθοδο του Δειγματολήπτη των Kuo και Mallick και η μεταβλητή C_13_2 έχει μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.39 στην μέθοδο του δειγματολήπτη Gibbs).
Οι μέθοδοι Δειγματολήπτης των Kuo και Mallick, Δειγματολήπτης Gibbs επιλογής μεταβλητών έχουν παρόμοια συμπεριφορά με αυτήν της στοχαστικής διερεύνησης επιλογής μεταβλητών, όμως οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι συστηματικά υψηλότερες. Συγκεκριμένα, για τις σημαντικές μεταβλητές (C2), (C_11_2), (C_11_3), οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι > 0.95. Επιπλέον, οι μη σημαντικές μεταβλητές είχαν εύρος μέσης εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ απο 0.06 (για την A_49_2) εώς 0.46 (για την C_13_3). Ακόμα, οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ των μη σημαντικών μεταβλητών (C_13_2, C_13_3, A_49_2, A_49_3) αυξάνονται με κατεύθυνση προς το 0.5.

Πίνακας 2.4: Αποτελέσματα επιλογής μεταβλητών μέσω των μεθόδων MCMC στα οποία αναγράφονται για κάθε μία απο τις ανεξάρτητες μεταβλητές οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ και τα σφάλματα MCMC χρησιμοποιώντας εμπειρική ανεξάρτητη εκ-των-προτέρων κατανομή για τις παραμέτρους του μοντέλου.

	Μέθοδος MCMC					
	Ανεξάρτητη εκ-των-προτέρων κατανομή για τις επιδράσεις					
	Στοχαστική διερεύνηση επιλογής μεταβλητών		Δειγματολήπτης Kuo-Mallick		Δειγματολήπτης Gibbs	
Ανεξάρτητες μεταβλητές	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$	Σφάλμα MC	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$	Σφάλμα MC	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$	Σφάλμα MC
C2	0.99	< .0001	0.99	< .0001	1	< .0001
C_11_2	0.53	0.012	0.99	0.001	1	< .0001
C_11_3	0.99	< .0001	0.99	< .0001	1	< .0001
C_13_2	0.11	0.004	0.33	0.034	0.39	0.026
C_13_3	0.12	0.005	0.46	0.040	0.38	0.028
A_49_2	0.11	0.004	0.07	0.005	0.06	< .0001
A_49_3	0.13	0.005	0.25	0.016	0.24	0.011

Η στοχαστική διερεύνηση επιλογής μεταβλητών διαφέρει απο τις υπόλοιπες μεθόδους MCMC. Παρόλο που η συγκεκριμένη μέθοδος δίνει υψηλές πιθανότητες εισαγωγής στις σημαντικές μεταβλητές (C2), (C_11_2), (C_11_3), (C_13_2), (A_49_3), οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ των μη σημαντικών μεταβλητών (C_13_2, C_13_3, A_49_2, A_49_3) αντιστοιχούν σε πολύ χαμηλότερα επίπεδα σε σχέση με τις άλλες μεθόδους MCMC υποστηρίζοντας απλούστερα μοντέλα.

Απο τα αποτελέσματα του Πίνακα 2.5 παρατηρούμε ότι οι μεταβλητές (C2), (C_11_2), (C_11_3) περιλαμβάνονται σχεδόν σε όλα τα εκ-των-υστέρων μοντέλα $\boldsymbol{\gamma}$ που επιλέγονται απο όλες τις μεθόδους MCMC επιλογής μεταβλητών. Ακόμα, το μέσο τετραγωνικό σφάλμα RMSE των προβλέπομενων τιμών $\widehat{y_i}$ όλων των μεθόδων MCMC επιλογής μεταβλητών κυμαίνεται στις ίδιες τιμές.

Σε όρους RMSE των προβλέπομενων τιμών $\widehat{y_i}$ ο δειγματολήπτης των Kuo και Mallick επιλογής μεταβλητών, ο δειγματολήπτης Gibbs επιλογής μεταβλητών είναι χειρότερες σε σχέση με την στοχαστική διευρεύνηση επιλογής μεταβλητών, καθώς υποστηρίζουν πολυπλοκότερα μοντέλα (επιπλέον εισαγωγή της μεταβλητής C_13_3) μικραίνοντας την αβεβαιότητα του μοντέλου (χαμηλή εκ-των-υστέρων πιθανότητα εισαγωγής).

Πίνακας 2.5: Αποτελέσματα επιλογής μοντέλων μέσω των μεθόδων MCMC στα οποία αναγράφονται για κάθε μοντέλο οι εκ-των-υστέρων πιθανότητες εισαγωγής για ανεξάρτητη εκ-των-προτέρων κατανομή για τις παραμέτρους του μοντέλου.

Μέθοδος MCMC	$\boldsymbol{\gamma}$	$\widehat{f}(\boldsymbol{\gamma}\|\boldsymbol{y})$	$RMSE(\widehat{y}_i)$
Στοχαστική διευρεύνηση επιλογής μεταβλητών Δειγματολήπτης	$C2 + C\_11\_2 + C\_11\_3$	0.32	8.71
Kuo-Mallick επιλογής μεταβλητών Δειγματολήπτης	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_3$	0.25	8.78
Gibbs επιλογής μεταβλητών	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_3$	0.25	8.96

Συνοψίζοντας, για όλες τις μεθόδους επιλογής μεταβλητών MCMC συμπεραίνουμε ότι οι μεταβλητές φύλο (C2), σωματική κατάσταση του/της φοιτητή/φοιτήτριας 12-18 (C_11_) πρέπει να συμπεριληφθούν στο μοντέλο, καθώς οι ψευδομεταβλητές (C2, άνδρας), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_2, κανονική), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_3, παχιά) περιλαμβάνονται σε όλα τα εκ-των-υστέρων μοντέλα και έχουν υψηλές μέσες εκ-των-υστέρων πιθανότητες εισαγωγής. Ισοδύναμα η συνολική επίδραση των κατηγορικών μεταβλητών C2, C_11_ είναι σημαντική για τον αναμενόμενο δείκτη μάζας σώματος.
Αντίθετα, οι μεταβλητές, σωματική κατάσταση μητέρας (C_13_), κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών (A_49_) δεν μπορούν να συμπεριληφθούν στο μοντέλο καθώς οι ψευδομεταβλητές σωματική κατάσταση μητέρας (C_13_2, κανονική), η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_3, πολλές φορές/σχεδόν καθημερινά), η σωματική κατάσταση μητέρας (C_13_3, παχιά), η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_2, αρκετές φορές) έχουν χαμηλές εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ και δεν περιλαμβάνονται στα εκ-των-υστέρων μοντέλα. Άρα, η συνολική επίδραση των κατηγορικών μεταβλητών A_49_, C_13_ δεν είναι σημαντική για τον αναμενόμενο δείκτη μάζας σώματος.

2.8 Συμπεράσματα

Στο κεφάλαιο αυτό προσπαθήσαμε να παρουσιάσουμε εν συντομία τα σημαντικότερα χαρακτηριστικά της Μπεϋζιανής επιλογής μεταβλητών. Απο Μπεϋζιανής πλευράς, η επιλογή μεταβλητών φαίνεται πιο λογική εκ φύσεως και ανταποκρίνεται στην πραγματικότητα. Επιπλέον, η Μπεϋζιανή στάθμιση μοντέλων παρέχει την δυνάτοτητα σωστής και ακριβής εκτίμησης των επιδράσεων, καθώς λαμβάνει υπόψην την αβεβαιότητα του μοντέλου και προτιμάται σε σχέση με τις υπόλοιπες κλασσικές μεθόδους.

Όσον αφορά τους αλγόριθμους MCMC επιλογής μεταβλητών, είδαμε ότι για την στοχαστική διερεύνηση επιλογής μεταβλητών χρησιμοποιείται ως εκ-των-προτέρων κατανομή για το διάνυσμα των επιδράσεων $\boldsymbol{\beta}$ η μείξη (Spike-Slab) κατανομών που έχει σκοπό να συρρικνώσει στο μηδέν τις μη σημαντικές μεταβλητές και στην μονάδα τις σημαντικές. Ακόμα, η διάσταση των παραμέτρων του μοντέλου παραμένει σταθερή καθώς δεν εξαρτάται απο το διάνυσμα $\boldsymbol{\gamma}$ και χρησιμοποιείται η ίδια πιθανοφάνεια.

Για τον δειγματολήπτη Gibbs των Kuo και Mallick η διάσταση των μοντέλων παραμένει σταθερή και η πιθανοφάνεια εξαρτάται απο το διάνυσμα $\boldsymbol{\gamma}$, καθώς καθορίζεται μία απο κοινού εκ-των-προτέρων κατανομή για το πλήρες μοντέλο.

Για τον Gibbs επιλογής μεταβλητών η πιθανοφάνεια είναι ίδια για όλα τα μοντέλα. Ωστόσο, η εκ-των-προτέρων κατανομή εξαρτάται αποκλειστικά και μόνο απο την κατασκευή του μοντέλου, δηλαδή την πραγματική εκ-των-προτέρων κατανομή και την ψευδό-εκ-των-προτέρων κατανομή, καθώς χρησιμοποιεί χαρακτηριστικά των δύο προηγούμενων αλγόριθμων.

Το σημαντικότερο μειονέκτημα αυτών των μεθόδων είναι ότι οι αλγόριθμοι είναι μη ακριβείς όταν υπάρχει πολυσυγραμμικότητα μεταξύ ανεξάρτητων μεταβλητών.

Επιπλέον, είδαμε πως οι μείξεις g εκ-των-προτέρων κατανομών λύνουν τα προβλήματα που προκύπτουν απο τις προκαθορισμένες g εκ-των-προτέρων κατανομές (συγκεκριμένος καθορισμός της παραμέτρου g), και γίνονται ιδιαίτερα προσιτές για τον τρόπο υπολογισμό των εκ-των-υστέρων ποσοτήτων. Οι εμπειρικές Μπεϋζιανες Μέθοδοι προσεγγίζονται όταν ο παραμετρικός χώρος των μοντέλων είναι πολυ μεγάλος, ενώ οι μέιξεις g εκ-των-προτέρων κατανομών, όπως οι Zellner-Siow εκ-των-προτέρων κατανομές και οι υπερ g εκ-των-προτέρων κατανομές δίνουν ικανοποιητικά αποτελέσματα σε όρους προσαρμοστηκότητας και συρρίκνωσης καθώς είναι εύρωστες στον μη καθορισμό της g παραμέτρου και παρέχουν γρήγορους υπολογισμούς της περιθώριας πιθανοφάνειας.

Ιδιαίτερη προσοχή πρέπει να δοθεί στον καθορισμό των εκ-των-προτέρων κατανομών των μοντέλων. Η χρήση της $\boldsymbol{beta - binomial}$ υπερ εκ-των-προτέρων κατανομής αυξάνει την ικανότητα επιλογής απλούστερων μοντέλων.

Στην περίπτωση που ο χώρος των μοντέλων είναι πολυ μεγάλος οι Μπεϋζιανοι αλγόριθμοι επιλογής μεταβλητών αδυνατούν και χρησιμοποιούνται γενικότεροι αλγόριθμοι επιλογής μεταβλητών όπως (Carlin και Chib), η μέθοδος MCMC αναστρέψιμου άλματος MCMC (Reversible jump MCMC και η κατά Μετρόπολις εκδοχή της μεθόδου Carlin και Chib (Metropolized Carlin and Chib).

Κεφάλαιο 3

Επιλογή μεταβλητών με την χρήση μη τοπικών εκ-των- προτέρων κατανομών

3.1 Εισαγωγή

Η ανάπτυξη των αλγορίθμων MCMC στις αρχές του 1990 συνέβαλε στην ανάπτυξη των Μπεϋζιανών μεθόδων σε προβλήματα εκτιμήσεων και ελέγχου υποθέσεων σε πολλούς επιστημονικούς κλάδους. Πολλές απο τις εφαρμογές αυτές βασίζονταν στην χρήση αντικειμενικών Μπεϋζιανών μοντέλων ή στην χρήση ασαφών και μη-πληροφοριακών εκ-των-προτέρων κατανομών για τις άγνωστες παραμέτρους των μοντέλων. Η αντικειμενική Μπεϋζιανή μεθοδολογία χρησιμοποιείται τα τελευταία χρόνια για την εκτίμηση των παραμέτρων και για συμπερασματολογία, υπο το καθεστώς άγνοιας.

Όπως αναφέραμε παραπάνω η επιλογή μοντέλου μεταφράζεται άμεσα σε έλεγχο υποθέσεων που αποσκοπεί στην εύρεση του βέλτιστου μοντέλου που περιγράφει τα δεδομένα για κάποια τιμή της άγνωστης παραμέτρου θ. Έστω ότι θέλουμε να ελέγξουμε την ακόλουθη υπόθεση, όπου $\Theta_0 \in \Theta_1$:

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1$$

Ο έλεγχος είναι ισοδύναμος, αφού καθορίσουμε τις εκ-των- προτέρων κατανομές $f_0(\theta)$ και $f_1(\theta)$, με:

$$H_0 : \theta \sim f_0(\theta), \theta \in \Theta_0$$

$$H_1 : \theta \sim f_1(\theta), \theta \in \Theta_1$$

Οι περισσότεροι Μπεϋζιανοί έλεγχοι υποθέσεων ορίζονται με εναλλακτικές εκ-των-προτέρων πυκνότητες πιθανότητας $f_1(\theta)$ που είναι θετικές στον παραμετρικό χώρο Θ_0. Τέτοιες εκ-των-προτέρων κατανομές ονομάζονται τοπικές εναλλακτικές εκ-των-προτέρων κατανομές. Η εναλλακτική υπόθεση H_1 που ορίζεται σύμφωνα με την τοπική εκ-των-προτέρων κατανομή $f_1(\theta)$ ονομάζεται τοπική εναλλακτική υπόθεση. Μια εναλλακτική υπόθεση πρέπει να αντικατοπτρίζει μια εντελώς διαφορετική θεωρία απο την αντίστοιχη μηδενική υπόθεση. Αυτό πρέπει να επιτυγχάνεται με τον καθορισμό μιας παραμέτρου που δείχνει την σημαντική απόκλιση της εναλλακτικής υπόθεσης απο την μηδενική υπόθεση.

Μια σημαντική δυσκολία σε πολλές εφαρμογές της Μπεϋζιανης συμπερασματολογίας σε ελέγχους υποθέσεων είναι ο καθορισμός κατάλληλων εκ-των-προτέρων κατανομών στις άγνωστες παραμέτρους των μοντέλων προκειμένου να γίνει η συμπερασματολογία. Έτσι ακατάλληλες εκ-των-προτέρων κατανομές δεν μπορούνε να χρησιμοποιηθούν στον υπολογισμό του παράγοντα Bayes για τον έλεγχο υποθέσεων και ακόμα οι συνέπειες των ασαφών εκ-των-προτέρων κατανομών δεν περιορίζονται με την αύξηση του δείγματος (Lindley και Barttlet, 1957).

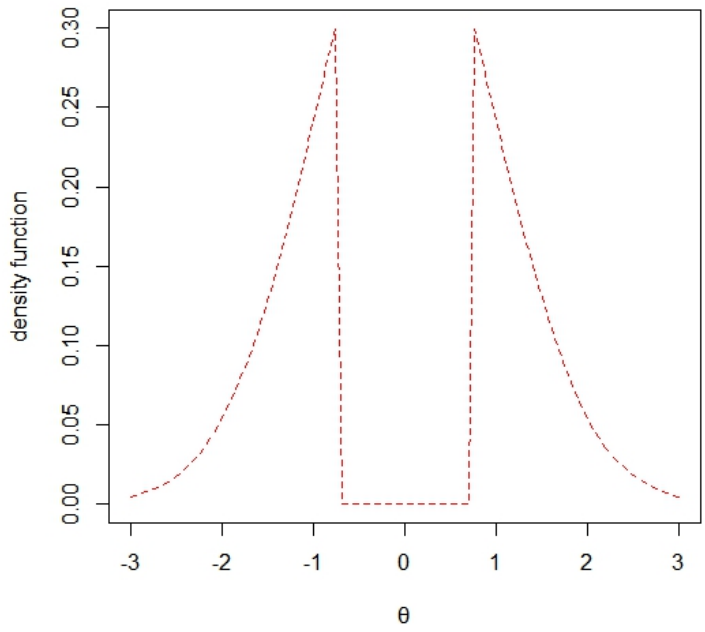
Πολλές προσεγγίσεις προτάθηκαν για την λύση αυτού του προβλήματος και συγκεκριμένα ο Jeffreys στις αρχές του 1939 πρότεινε τον καθορισμό εκ-των-προτέρων κατανομών υπό την εναλλακτική υπόθεση κεντραρισμένες στην μηδενική υπόθεση. Οι περισσότερες δημοσιεύσεις για την διαδικασία των Μπεϋζιανών ελέγχων βασίζονται στην ίδεα των εναλλακτικών εκ-των-προτερών κατανομών (Kass και Raftery 1995, Lahiri 2001, Walker 2004). Πιο πρόσφατα, ο κλασματικός παράγοντας Bayes (O'Hagan 1995, 1997, Conigliani και O'Hagan 2000, De Santis και Spezzaferi 2011) και ο ενδογενής παράγοντας Bayes (Berger και Pericchi 1996, 1998, Berger και Mortera 1999, Perez και Berger 2002) χρησιμοποιήθηκαν ως εναλλακτικές προσεγγίσεις. Όταν τα δεδομένα προέρχονται απο ένα μοντέλο που σχετίζεται με την μηδενική υπόθεση, τότε ο δυναμικός παράγοντας Bayes και ο ενδογενής παράγοντας Bayes παράγουν εναλλακτικές εκ-των-προτέρων κατανομές που είναι θετικές σε τιμές των παραμέτρων που σχετίζονται με την μηδενική υπόθεση. Σε πολλές περιπτώσεις, περιορισμένα σχόλια μπορούν να γίνουν σχετικά με τις εκ-των-προτέρων κατανομές στις οποίες βασίζονται οι ενδογενείς παράγοντες Bayes, γιατί οι εναλλακτικές ενδογενείς εκ-των-προτέρων κατανομές ορίζονται σύμφωνα με μια μηδενική υπόθεση και έτσι συγκεντρώνουν την μάζα πιθανότητας στις τιμές των παραμέτρων που σχετίζονται

με την μηδενική υπόθεση.

Επιπλέον οι παράγοντες Bayes που υπολογίζονται μέσω των εναλλακτικών εκ-των-προτέρων κατανομών εμφανίζουν παθογένεια σε μεγάλες τιμές του δείγματος.
Όσο το μέγεθος του δείγματος αυξάνει, ο παράγοντας Bayes συσσωρεύει ταχύτατα ενδείξεις και αποδειχτικά στοιχεία σε περίπτωση που η εναλλακτική υπόθεση είναι σωστή. Για παράδειγμα για τον έλεγχο μιας παραμέτρου ο παράγοντας Bayes υπέρ της μηδενικής υπόθεσης αυξάνει με ρυθμό $O_p\left(n^{\frac{1}{2}}\right)$ όταν τα δεδομένα προέρχονται απο την μηδενική υπόθεση, ενώ οταν τα δεδομένα προέρχονται απο την εναλλακτική υπόθεση ο παράγοντας Bayes υπερ της εναλλακτικής υπόθεσης αυξάνει με εκθετικό ρυθμό σε σχέση με το μέγεθος του δείγματος n (Bahadur και Bickel 1967, Walker και Hjort 2001, Walker 2004). Συνεπώς οι ρυθμοί υπο την εναλλακτική και υπο την μηδενική ειναι πολύ ασσύμετροι, παρόλο που αυτή η κατάσταση δεν αντιπροσωπεύεται μέ όρους πιθανότητας στα αποτελέσματα των Μπεϋζιανών ελέγχων υποθέσεων (Vlachos και Gelfand 2003).

Μέθοδοι της κλασσικής στατιστικής αναφέρουν αυτήν την ασσυμετρία στην διεξαγωγή των ελέγχων υποθέσεων. Επειδή δεν ορίζεται η πιθανότητα αποδοχής της μηδενικής υπόθεσης ο ρυθμός συσσώρευσης των ενδείξεων υπερ της μηδενικής είναι λιγότερο προβληματικός. Σε αντίθεση, ένας Μπεϋζιανος έλεγχος υποθέσεων με την χρήση των τοπικών εναλλακτικών εκ-των-προτέρων κατανομών στοχεύει στον υπολογισμό της εκ-των-υστέρων πιθανότητας ότι η μηδενική υπόθεση είναι σωστή.

Οι Verdinelli και Wasserman (1996), Rousseau (2007) αναφέρθηκαν σε αυτό το πρόβλημα προτείνοντας μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές της μορφής $f(\theta) = 0$ για όλα τα θ που ανήκουν σε κοντινή γειτονία του παραμετρικού χώρου Θ_0, όμως αυτή η προσέγγιση υστερεί στον καθορισμό του ρυθμού με τον οποίο η $f(\theta)$ πλησιάζει το μηδέν σε κοντινή περιοχή του παραμετρικού χώρου Θ_0 και δεν παρέχει κάποιο μηχανισμό απόρριψης της μηδενικής υπόθεσης για τιμές της παραμέτρου θ σε γειτονικές περιοχές του παραμετρικού χώρου Θ_0. Μια τέτοια αναπαράσταση της εκ-των-προτέρων κατανομής $f(\theta)$ σύμφωνα με την οποία δεν καθορίζεται ο ρυθμός με τον οποίο πλησιάζει το μηδέν παρουσιάζεται στο Διάγραμμα 3.1.



Διάγραμμα 3.1: Απεικόνιση των μη-τοπικών εκ-των-προτέρων κατανομών των Verdinelli και Wasserman (1996) και Rouseau (2007)

Οι μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές που προτάθηκαν απο τους (Valen Johnson και David Rossel 2010) εξισορροπούν το ρυθμό σύγκλισης των παραγόντων Bayes υπέρ της μηδενικής και εναλλακτικής υπόθεσης. Συγκεκριμένα προσφέρουν σημαντική λύση ανάμεσα στην χρήση των ασαφών εκ-των-προτέρων κατανομών που τείνουν στην αποδοχή του πιο απλού μοντέλου (Jeffreys 1998, Lindley και Barttlet 1957) και των τοπικών εναλλακτικών εκ-των-προτέρων κατανομών που περιορίζουνε τον ρυθμό σύγκλισης υπερ της μηδενικής υπόθεσης. Αυτές οι συναρτήσεις πυκνότητες πιθανότητας βασίζονται στον καθορισμό μιας παραμέτρου που καθορίζει την κλίμακα των αποκλίσεων μεταξύ μηδενικής και εναλλακτικής υπόθεσης. Λογικές επιλογές της παραμέτρου αυξάνουν το λογάριθμο του παράγοντα Bayes ταυτόχρονα υπερ της μηδενικής και της εναλλακτικής υπόθεσης.

Το συγκεκριμένο κεφάλαιο επικεντρώνεται στην παρουσίαση και την περιγραφή των ιδιοτήτων των μη-τοπικών εκ-των-προτέρων κατανομών οι οποίες περιλαμβάνουνε τις εκ-των-προτέρων κατανομές ροπών και τις κατανομές αντίστροφων ροπών (Johnson και Rossel 2010). Ιδιαίτερη έμφαση θα δοθεί στην περιγραφή της επιλογής μεταβλητών όπως έγινε στο Κεφάλαιο 2 για την εύρεση του βέλτιστου μοντέλου.

3.2 Τοπικές εναλλακτικές εκ-των-προτέρων κατανομές

Οι παράγοντες Bayes που ορίζονται με βάση τις τοπικές εναλλακτικές εκ-των-προτέρων-κατανομές παρουσίαζουν σημαντικές ασυμπτωτικές ιδιότητες. Ιδιαίτερη προσοχή δίνεται στην ασυμπτωτική συμπεριφορά της εκ-των- υστέρων κατανομής (Walker 1969). Έστω πάλι ο έλεγχος υποθέσεων που αναφέραμε προηγουμένως:

$$H_0 : \theta \sim f_0(\theta), \theta \in \Theta_0$$

$$H_1 : \theta \sim f_1(\theta), \theta \in \Theta_1$$

Έστω ότι έχουμε τυχαίο δείγμα $x_1, \ldots, x_n$ απο την πιθανοφάνεια $f(x|\theta)$ και ορίζουμε τις εκ-των-προτέρων κατανομές $f_0(\theta)$, $f_1(\theta)$ στους αντίστοιχους παραμετρικους χώρους Θ_0, Θ_1 αντίστοιχα για την μηδενική και εναλλακτική υπόθεση. Αν υποθέσουμε ότι $f_j(\theta)$ είναι συνεχής πυκνότητα, τότε η ποσότητα:

$$f_j(x) = \int_\Theta f(x|\theta) f_j(\theta) d\theta$$

ορίζεται ως η περιθώρια πιθανοφάνεια των δεδομένων κάτω απο την εκ-των-προτέρων κατανομή $f_j(\theta)$ και την οποία προαναφέραμε στο Κεφάλαιο 1 (βλέπε Ενότητα 1.7.1). Ακόμα ο παράγοντας του Bayes συναρτήσει του δείγματος n (βλέπε ενότητα 1.7.1) ορίζεται ως:

$$BF_n(1|0) = \frac{f_1(x)}{f_0(x)}$$

Η κατανομή $f_1(\theta)$ που χρησιμοποιείται υπο την εναλλακτική υπόθεση ονομάζεται τοπική εναλλακτική εκ-των-προτέρων κατανομή αν $f_1(\theta) > \epsilon$, για κάθε $\theta \in \Theta_0$. Οι τοπικές εναλλακτικές εκ-των -προτέρων κατανομές δίνουν μεγαλύτερη μάζα πιθανότητας στούς παραμετρικούς χώρους που ορίζονται υπο-την μηδενική υπόθεση, ενώ δίνουν μικρότερη μάζα πιθανότητας στούς παραμετρικούς χώρους που ορίζονται υπο-την εναλλακτική υπόθεση. Έτσι, εάν τα δεδομένα δεν δείξουν ισχυρή ένδειξη κατά της μηδενικής υπόθεσης το οποίο συμβαίνει σπάνια σε συγκεκριμένα δείγματα, θα υπάρχει πάντα μεροληψία υπερ της μηδενικής υπόθεσης. (Consonni, Forster και La Rocca 2013).

Επιπλέον, παρόλο που η προσέγγιση του Jeffreys (1939) που αφορά το καθορισμό εκ-των-προτέρων κατανομών υπο-την εναλλακτική υπόθεση κεντραρισμένες στην μηδενική υπόθεση είναι εύλογη, στην πραγματικότητα όμως υστέρει για τον λογο ότι επαναπροσδιορίζοντας την μάζα πιθανότητας σε παραμετρικούς χώρους που σχετίζονται με την μηδενική υπόθεση για συγκεκριμένα μεγέθους δείγματα, μειώνει την ένδειξη υπέρ της εναλλακτικής υπόθεσης οταν οι τιμές των παραμέτρων που γεννούν τα δεδομένα είναι πολύ μακριά απο την μηδενική υπόθεση Consonni, Forster και La Rocca (2013).

Οι τοπικές εναλλακτικές εκ-των-προτέρων κατανομές χρησιμοποιούνται ευρύτατα σε ελέγχους υποθέσεων όπου τα μοντέλα υπο σύγκριση είναι φωλιασμένα, δηλαδή προκύπτει το ένα ως ειδική περίπτωση του άλλου. Τέλος, για τα συμπεράσματα που προκύπτουνε μέσω των ιδιοτήτων του (Walker 1969) για τις ασυμπτωτικές ιδιότητες των παραγόντων Bayes έχουμε:

1. Για μία πραγματική μηδενική υπόθεση, ο παράγοντας Bayes υπερ της εναλλακτικής υπόθεσης μειώνεται με ρυθμό $O_p\left(n^{\frac{1}{2}}\right)$
2. Για μια πραγματική εναλλακτική υπόθεση, ο παράγοντας Bayes υπερ της μηδενικής υπόθεσης μειώνεται εκθετικά

Αυτοί οι εντελώς διαφορετικοί ρυθμοί σύγκλισης οδηγούν στο συμπέρασμα ότι είναι πολυ πιο πιθανό ενα πείραμα να παρέχει στοιχεία υπερ της αληθινής εναλλακτικής υπόθεσης απο ότι της αληθινής μηδενικής υπόθεσης, γεγονός που επιβεβαιώνει ότι οι τοπικές εναλλακτικές εκ-των-προτέρων κατανομές υπόκεινται σε μεροληψία Consonni, Forster και La Rocca (2013). Τετοιοι έλεγχοι υποθέσεων αντικατοπτρίζονται στον έλεγχο υπόθεσης για μία μόνο παράμετρο, καθώς και σε προβλήματα επιλογής μεταβλητών.

3.3 Μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές

Οι μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές ξεπερνούν τα περιοριστικά προβλήματα των τοπικών εναλλακτικών εκ-των-προτέρων κατανομών που αναφέραμε στην εισαγωγή και στην Ενότητα 3.2 και βετιώνουν τον ρυθμό σύγκλισης υπερ της αληθινής μηδενικής υπόθεσης για μια δοθείσα εναλλακτική υπόθεση. Οι μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές $f_1(\theta)$ ορίζονται έτσι ώστε $f_1(\theta) < \epsilon$, για κάθε $\theta \in \Theta_0$: $\inf|\theta - \theta_0| < \zeta$. Δηλαδή, η βασική ιδέα είναι ο καθορισμός εκ-των-προτέρων κατανομών που ελαχιστοποιούν την μάζα πιθανότητας (πρακτικά την θέτουν ίση με μηδέν) σε παραμετρικούς χώρους που σχετίζονται με την μηδενική υπόθεση και σε γειτονικές περιοχές γύρω απο την μηδενική υπόθεση δίνουν μεγαλύτερη μάζα πιθανότητας. Θα ξεκινήσουμε απο τον ακόλουθο σημειακό έλεγχο υπόθεσης:

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

Στην υπο-Ενότητα 3.3.1 θα αναφερθούμε στην οικογένεια εκ-των-προτέρων κατανομών ροπών οι οποίες παρόλο που δεν παρέχουνε εκθετικό ρυθμό σύγκλισης των παραγόντων Bayes υπέρ της αληθινής μηδενικής υπόθεσης, προσφέρουν σημαντική βελτίωση στον ρυθμό σύγκλισης υπό

την αληθινή μηδενική υπόθεση. Επιπλέον οι παράγοντες Bayes που καθορίζονται απο αυτήν την οικογένεια κατανομών είναι διαθέσιμοι σε κλειστή μορφή για αρκετα κοινά στατιστικά μοντέλα. Στην υπο-Ενότητα 3.3.2 περιγράφονται οι εκ-των-προτέρων κατανομές αντίστροφων ροπών οι οποίες παρέχουν εκθετική σύγκλιση στον ρυθμό υπερ της αληθινής μηδενικής και της αληθινής εναλλακτικής.

3.3.1 Μη τοπικές εκ-των-προτέρων κατανομές ροπών μιας παραμέτρου

Οι εκ-των-προτέρων κατανομές ροπών μπορούν να προκύψουν και ως το γινόμενο ακόμα και δυνάμεων της υπο-εξέταση παραμέτρου και αυθαίρετων πυκνοτήτων πιθανότητας, οι οποίες ονομάζονται εκ-των-προτέρων κατανομές γινομένου ροπών. Ας υποθέσουμε ότι η $f_b(\theta)$ ορίζει μια εκ-των-προτέρων κατανομή ώς βάση με $2r$ πεπερασμένες ακέραιες ροπές. Έτσι η εκ-των-προτέρων κατανομή ροπή r-τάξης ορίζεται ως εξής:

$$f_M(\theta) = \frac{(\theta - \theta_0)^{2r}}{\tau_r} f_b(\theta), \tag{3.1}$$

όπου η σταθερά κανονικοποιήσης ορίζεται ως:

$$\tau_r = \int_\Theta (\theta - \theta_0)^{2r} f_b(\theta) d\theta. \tag{3.2}$$

Για ευκολία κατανόησης, ορίζουμε τ_1 την σταθερά κανονικοποίησης για της πρώτης τάξής ροπής εκ-των-προτέρων κατανομής. Ακόμα μονόπλευροι έλεγχοι υποθέσεων μπορούν να πραγματοποιηθούν λαμβάνοντας υπόψην ότι $f_M(\theta) = 0$ είτε για $\theta < 0$ είτε $\theta > 0$, ενώ ισχύει ότι$f_M(\theta_0) = 0$. Οι ρυθμοί σύγκλισης των παραγόντων Bayes υπέρ της αληθινής μηδενικής υπόθεσης όταν η εναλλακτική υπόθεση ορίζεται με εκ-των-προτέρων κατανομή ροπών μπορεί να επιτευχθεί μέσω προσεγγίσεων Laplace (Tierney κ.α. 1989, de Bruijn 1981).

Η επιλογή της εκ-των-προτέρων κατανομής βάσης $f_b(\theta)$ μπορεί να χρησιμοποιηθεί για να αντιπροσωπεύσουμε τις πεποιθήσεις μας μέσω των ουρών της κατανομής υπο την εναλλακτική υπόθεση. Η συγκεκριμένη επιλογή καθορίζει την ασυμπτωτική συμπεριφορά των παραγόντων Bayes συναρτήσει του δείγματος όταν η εναλλακτική υπόθεση είναι σωστή.

3.3.2 Μη τοπικές εκ-των-προτέρων κατανομές αντίστροφων ροπών μιας παραμέτρου

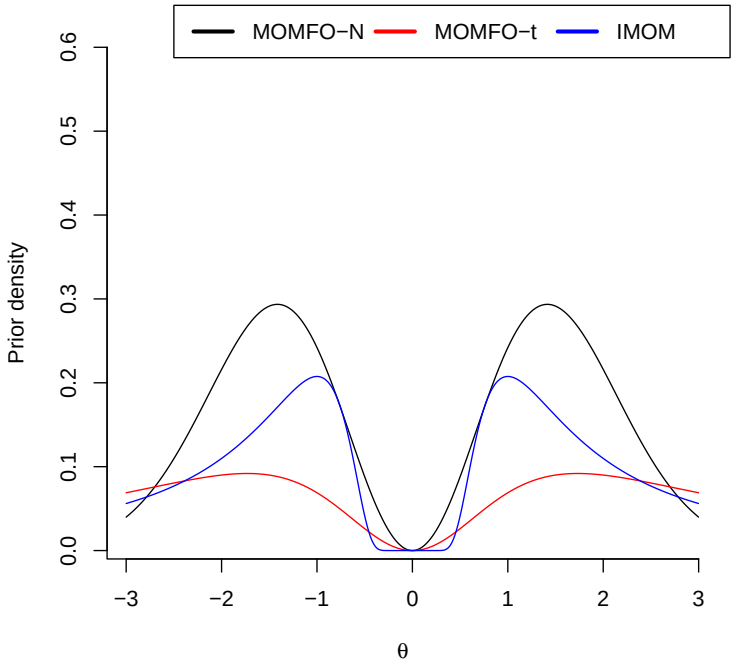
Οι εκ-των-προτέρων κατανομές αντίστροφων ροπών μπορούν να προκύψουν και ως το γινόμενο ακόμα και δυνάμεων της υπο-εξέταση παραμέτρου και αυθαίρετων πυκνοτήτων πιθανότητας, οι οποίες ονομάζονται εκ-των- προτέρων κατανομές γινομένου αντίστροφων ροπών. Οι εκ-των-προτέρων κατανομές αντίστροφων ροπών r-τάξης ορίζονται σύμφωνα με την μορφή:

$$f_I(\theta) = \frac{r\tau^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2r}\right)}[(\theta-\theta_0)^2]^{-\frac{\nu+1}{2}} \exp\left[-\left\{\frac{(\theta-\theta_0)^2}{\tau}\right\}^{-r}\right] \tag{3.3}$$

για $r, \nu, \tau > 0$

Η συναρτησιακή μορφή τους σχετίζεται με τις αντίστροφες γάμμα πυκνότητες πιθανότητας που σημαίνει ότι η συμπεριφορά τους κοντά στο θ_0 είναι παρόμοια με την συμπεριφορά μιάς αντίστροφης γάμμα κατανομής κοντά στο μηδέν.
Προσεγγίσεις Laplace (Tierney κ.α. 1989, de Bruijn 1981) μπορεί να χρησιμοποιηθούν για τον υπολογισμό των περιθώριων πιθανοφανειών υπό την εναλλακτική υπόθεση όταν η μηδενική υπόθεση είναι σωστή και η εκ-των-υστέρων επικρατούσα τιμή $\hat{\theta}^*$ ικανοποιεί την σχέση $|\theta^* - \theta_0| = O_p\left(n^{-\frac{1}{2r+2}}\right)$. Για μεγάλες τιμές του r, ο παράγοντας Bayes με βάση τις εκ-των-προτέρων κατανομές αντίστροφες ροπές γινομένου συγκλίνει εκθετικά στο δείγμα υπερ της μηδενικής και της εναλλακτικής υπόθεσης. Στην πράξη, τιμές του r απο το ένα εώς το δύο παρέχουν κατάλληλο ρυθμό σύγκλισης υπερ της μηδενικής υπόθεσης. Εάν χρησιμοποιήσουμε ως βάση την κατανομή $t-student$ με ν βαθμούς ελευθερίας, τότε η επιλογή της τιμής $\nu = 1$ παράγει κατανομές με ουρές σαν της κατανομής Cauchy βελτιώνοντας την ισχύ υπέρ της εναλλακτικής υπόθεσης ακόμα και σε μικρά δείγματα.

Οι διαφορές στους ρυθμούς σύγκλισης των παραγόντων Bayes μεταξύ των ροπών και αντίστροφων ροπών απεικονίζονται στο Διάγραμμα 3.2. Στο Διάγραμμα 3.2 αναπαριστώνται οι εκ-των-προτέρων κατανομές ροπών πρώτης τάξης χρησιμοποιώντας ως βάση κανονική κατανομή για $\tau = 1$ (MOMFO-N), οι εκ-των-προτέρων κατανομές ροπών πρώτης τάξης χρησιμοποιώντας ως βάση $t-Student$ κατανομή για $\tau = 1$, $\nu = 3$ (MOMFO-t) και οι εκ-των προτέρων κατανομές αντίστροφων ροπών πρώτης τάξης για $\tau = 1$ (IMOM).



Διάγραμμα 3.2: Απεικόνιση των εκ-των-προτέρων κατανομών ροπών πρώτης τάξης χρησιμοποιώντας ως βάση κανονική κατανομή για $\tau = 1$ (MOMFO-N), των εκ-των-προτέρων κατανομών ροπών πρώτης τάξης χρησιμοποιώντας ως βάση $t-Student$ κατανομή για $\tau = 1$, $\nu = 3$ (MOMFO-t) και των εκ-των προτέρων κατανομών αντίστροφων ροπών πρώτης τάξης για $\tau = 1$ (IMOM)

Όπως φαίνεται το Διάγραμμα 3.2 οι εκ-των-προτέρων κατανομες αντίστροφων ροπών (IMOM) πλησίαζουν πολύ πιο γρήγορα το μηδέν σε σχέση με τις εκ-των-προτέρων κατανομές ροπών πρώτης τάξης χρησιμοποιώντας ως βάση κανονική κατανομή (MOMFO-N) και τις εκ-των-προτέρων κατανομές ροπών πρώτης τάξης χρησιμοποιώντας ως βάση $t-Student$ κατανομή, έτσι είναι λογικό που παρέχουν εκθετικό ρυθμό συσσώρευσης για την αληθινή μηδενική υπόθεση ακόμα και σε μικρά δείγματα.

3.3.3 Πολυμεταβλητές μη τοπικές εκ-των-προτέρων κατανομές

Οι εκ-των-προτέρων κατανομές ροπών και αντίστροφων ροπών μπορούν να γενικευτούν για πολυμεταβλητές κατανομές ορίζοντας την τετραγωνική μορφή:

$$Q(\boldsymbol{\theta}) = \frac{(\boldsymbol{\theta} - \boldsymbol{\theta_0})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta_0})}{n\tau\sigma^2} \tag{3.4}$$

όπου $\boldsymbol{\theta}$ είναι διάνυσμα διάστασης $e\times 1$, $\boldsymbol{\Sigma}$ είναι θετικά ορισμένος πίνακας και τ είναι μονοδιάστατο. Για την διευκόλυνση της εξαγωγής των αποτελεσμάτων για τα γραμμικά μοντέλα, χρησιμοποιήθηκε εναλλακτική παραμετροποιήση. Συγκεκριμένα, ο παράγοντας $\frac{1}{n}$ συμπεριλαμβάνεται στον πίνακα $\boldsymbol{\Sigma}^{-1}$ και παρέχει μια τυποποιημένη κλίμακα όταν χρησιμοποιείται ο πίνακας πληροφορίας, ενώ ο παράγοντας $\frac{1}{\sigma^2}$ μας επιτρέπει να λάβουμε υπόψην την παρατηρούμενη διακύμανση στην περίπτωση του απλού γραμμικού μοντέλου και στα γενικευμένα γραμμικά μοντέλα αντιπροσωπεύει παράμετρο υπερδιασποράς. Στην μονομεταβλητή περίπτωση, η παράμετρος τ εκφράζει την διασπορά γύρω απο το $\boldsymbol{\theta_0}$. Έτσι η πολυμεταβλητή κατανομή αντίστροφης ροπής για το διάνυσμα $\boldsymbol{\theta}$ ορίζεται ως εξής:

$$f_I(\boldsymbol{\theta}) = c_I Q(\boldsymbol{\theta})^{-\frac{\nu+e}{2}} \exp\left\{-Q(\boldsymbol{\theta})\right\}, \tag{3.5}$$

όπου

$$c_I = \left|\frac{\boldsymbol{\Sigma}^{-1}}{n\tau\sigma^2}\right| \frac{r}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{e}{2}\right)}{\pi^{\frac{e}{2}}}.$$

Καθώς $Q(\boldsymbol{\theta})$ αυξάνει, η επιρροή του εκθετικού όρου στην εξίσωση 3.5 εξαφανίζεται και οι ουρές της συνάρτησης πυκνότητας πιθανότητας $f_I(\boldsymbol{\theta})$ μοιάζουν με αυτές της πολυμεταβλητής $t-$ ***Student*** κατανομής με ν βαθμούς ελευθερίας.

Η επέκταση για τις κατανομές γινομένου ροπών γίνεται με παρόμοιο τρόπο. Έστω $f_b(\boldsymbol{\theta})$ μια πολυμεταβλητή κατανομή βάση για το δίανυσμα $\boldsymbol{\theta}$, η αναμενόμενη τιμή $E_b[Q(\boldsymbol{\theta})^r]$ είναι πεπερασμένη και $f_b(\boldsymbol{\theta_0})$ (Johnson και Rossel 2010). Έτσι μπορούμε να ορίσουμε την πολυμεταβλητή εκ-των-προτέρων ροπή r-τάξης ως εξής:

$$f_M(\boldsymbol{\theta}) = \frac{Q(\boldsymbol{\theta})^r}{E_b[Q(\boldsymbol{\theta})^r]} f_b(\boldsymbol{\theta}). \tag{3.6}$$

Για τιμές $r = 0$ επιστρέφουμε στις τοπικές εκ-των-προτέρων κατανομές για την εναλλακτική υπόθεση. Έαν για βάση χρησιμοποιηθεί η κανονική κατανομή $N_e(\boldsymbol{\theta_0}, n\tau\sigma^2\boldsymbol{\Sigma})$ τότε $E_b[Q(\boldsymbol{\theta})^r] = \prod_{i=0}^{r-1}(e+2i)$, αντιστοιχεί στην ροπή r-τάξης μιας χ^2 κατανομής με e βαθμούς ελευθερίας, ενώ αν για βάση χρησιμοποιηθεί μια πολυμεταβλητή $t-$ ***student*** κατανομή $t_\nu(\boldsymbol{\theta_0}, n\tau\sigma^2\boldsymbol{\Sigma})$ με $\nu > 2$ βαθμούς ελευθερίας και $r = 1$, τότε $E_b[Q(\boldsymbol{\theta})]^r = \frac{\nu e}{\nu-2}$.

3.3.4 Εκ-των-προτέρων καθορισμός των υπερπαραμέτρων

Η εκ-των-προτέρων πολυμεταβλητή κατανομή ροπών με βάση την πολυμεταβλητή κανονική κατανομή περιέχει τις υπερπαράμετρους: r, τ, σ , ενώ με βάση την πολυμεταβλητή $t-Student$ κατανομή περιέχεται μια ακόμα επιπλέον παράμετρος ν. Και οι δύο κατηγορίες μη-τοπικών εκτων-προτέρων κατανομές απαιτούν τον καθορισμό του πίνακα $\mathbf{\Sigma}$.

Σε πολλές εφαρμογές, είναι χρήσιμο να αντικαταστήσουμε τον όρο $\frac{\mathbf{\Sigma}^{-1}}{\sigma^2}$ με τον πίνακα πληροφορίας του Fisher. Πολλαπλασιάζοντας αυτόν τον πίνακα με το μέγεθος του δείγματος n διευκολύνει στην στάθμιση των συνολικών επιδράσεων των παραμέτρων. Σύμφωνα με αυτήν την παραμετροποιήση, η παράμετρος σ^2 αντιπροσωπεύει την παρατηρούμενη διακύμανση είτε παίρνει την τιμή ένα και εισάγεται στον πίνακα $\mathbf{\Sigma}$.

Η επιλογή $r = 1$ είναι μια βολική τιμή εξίσου για το γινομένο ροπών και το γινόμενο αντίστροφων ροπών. Συγκεκριμένα, για τις κατανομές ροπών, για $r = 1$ και χρησιμοποιώντας μια κανονική κατανομή ως βάση καταλήγουμε σε απλές μορφές των παραγόντων Bayes για τα απλά γραμμικά μοντέλα καθώς και προσεγγίσεις των παραγόντων Bayes σε μεγάλα δείγματα. Για τις κατανομές αντίστροφων ροπών, για $r = 1$ έχουμε σύγκλιση των παραγόντων Bayes απο κοινού για την μηδενική και εναλλακτική υπόθεση. Για την πολυμεταβλητή $t-Student$ κατανομή προτείνονται οι τιμές $r = 1$ και $\nu = 3$ της οποίας οι ουρές μοιάζουνε με την κατανομή Cauchy. Για τις κατανομές αντίστροφων ροπών, προτείνεται η τιμή $\nu = 1$ για την οποία οι ούρες μοιάζουν με μια πολυμεταβλητή κατανομή Cauchy.

Τέλος, προτείνονται δύο μέθοδοι για τον καθορισμό της παραμέτρου τ όταν δεν υπάρχει διαθέσιμη εκ-των-προτέρων πληροφορία (Johnson και Rossel 2010). Στην πρώτη μέθοδο, ορίζουμε την παράμετρο τ έτσι ώστε η εκ-των-προτέρων επικρατούσα τιμή να είναι πιο πιθανή υπο την εναλλακτική υπόθεση. Έτσι η επικρατούσα τιμή ορίζεται ως εξής:

$$w = \left(\frac{(\boldsymbol{\theta} - \boldsymbol{\theta_0})}{\sigma}\right)^T \frac{\mathbf{\Sigma}^{-1}}{n} \left(\frac{(\boldsymbol{\theta} - \boldsymbol{\theta_0})}{\sigma}\right) \tag{3.7}$$

Για την επικρατούσα τιμή για την κανονική κατανομή ροπών θα πρεπε να ισχυεί $w = 2k\tau$, ενώ για την επικρατούσα τιμή υπό την $t-Student$ κατανομή ροπών θα πρέπει $w = \tau\frac{2\nu}{\nu-2+e}$. Ορίζοντας έτσι τις αναμενόμενες διαφορές της $\boldsymbol{\theta}$ απο της $\boldsymbol{\theta_0}$, αυτές οι ποσότητες χρησιμοποιούνται για τον καθορισμό της παράμετρου τ η οποία τοποθετεί την επικρατούσα τιμή σε μια κανονικοποιημένη απόσταση απο $\boldsymbol{\theta_0}$. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη στα άρχικα στάδια των κλινικών δοκιμών και μπορεί να πραγματοποιηθεί καθορίζοντας τις εκ-των-προτέρων επικρατούσες τιμές

στους ρυθμούς απόκρισης που χρησιμοποιούνται για τον υπολογισμό του δείγματος.

Μια εναλλακτική προσέγγιση είναι η παράμετρος τ να οριστεί έτσι ώστε οι σταθμισμένες επιδράσεις να είναι μεγαλύτερες απο ένα συγκεκριμένο όριο. Για παράδειγμα, οι σταθμισμένες επιδράσεις μικρότερες του 0.2 δεν θεωρούνται συχνά σημαντικές στις κοινωνικές επιστήμες (Cohen 1992), δηλαδή η παράμετρος τ ορίζεται έτσι ώστε η εκ-των-προτέρων πιθανότητα που σχετίζεται με το γεγονός ότι οι σταθμισμένες επιδράσεις ειναι μικρότερες απο 0.2. Στην περίπτωση όπου $r = 1$, η πιθανότητα που αντιστοιχεί σε ένα διάστημα $(-a, a)$ με βάση την εκ-των-προτέρων κανονική κατανομή ροπών κεντραρισμένη στο μηδέν, με τ και $\nu = 1$ είναι $2\left\{\Phi\left(\frac{\alpha}{\sqrt{\tau}}\right) - \frac{\alpha}{\sqrt{2\pi\tau}}\exp\left(-\frac{\alpha^2}{2\tau}\right) - \frac{1}{2}\right\}$, όπου Φ είναι η αθροιστική συνάρτηση της τυπικής κανονικής κατανομής. Ακόμα η πιθανότητα που αντιστοιχεί σε κατανομή αντίστροφων ροπών είναι $1 - G\left\{\left(\frac{\alpha}{\sqrt{\tau}}\right)^{-2r}; \frac{\nu}{2r}, 1\right\}$, όπου G είναι η αθροιστική συνάρτηση της γάμμα κατανομής.

3.4 Γραμμική παλινδρόμηση με την χρήση μη τοπικών εκ-των -προτέρων κατανομών

Σε αυτήν την ενότητα θα περιγράψουμε μια νέα κατηγορία Μπεϋζιανής επιλογής μεταβλητών χρησιμοποιώντας μη-τοπικές εκ-των-προτέρων κατανομές (Johnson και Rossel 2012) για τις παράμετρους του απλού γραμμικού μοντέλου υιοθετώντας και εδώ τις βασικές αρχές της επιλογής μεταβλητών George και McCulloch (1993, 1997), Ntzoufras (2011) βλέπε Ενότητα 2.2. Οι μη τοπικές εκ-των-προτέρων κατανομές είναι συναρτήσεις πυκνότητας πιθανότητας που είναι μηδέν όταν οποιαδήποτε παράμετρος του μοντέλου είναι ίση με την τιμή της μηδενικής υπόθεσης, στην οποία ο συντελεστής β_j ισούται με μηδέν στην διαδικασία επιλογής μεταβλητών. Αντίθετα, οι τοπικές εναλλακτικές εκ-των προτέρων κατανομές είναι πάντα θετικές υπο τις τιμές της μηδενικής υπόθεσης. Οι περισσότερες Μπεϋζιανές μέθοδοι επιλογής μεταβλητών γίνονται με τοπικές εναλλακτικές εκ-των- προτέρων κατανομές.

Αποδεικνύεται ότι οι Μπεϋζιανές μέθοδοι επιλογής μεταβλητών που βασίζονται σε μη τοπικές εκ-των-προτέρων κατανομές δίνουν εκ-των υστέρων πιθανότητα για το πραγματικό μοντέλο που τείνει στο ένα όσο το μέγεθος του δείγματος αυξάνεται, όταν ο αριθμός των ανεξάρτητων μεταβλητών p φράσεται απο το μέγεθος του δείγματος, n, και ισχύουν συγκεκριμένες ιδιότητες για τον πίνακα σχεδιασμού. Απο την άλλη, οι Μπεϋζιανές μέθοδοι επιλογής μεταβλητών που βασίζονται σε τοπικές εναλλακτικές εκ-των-προτέρων κατανομές δίνουν εκ-των- υστέρων πιθανότητα μηδέν στο πραγματικό μοντέλο όσο το μέγεθος του δείγματος, n, αυξάνεται (Johnson

και Rossel 2010). Μεταξύ αυτων των Μπεϋζιανών μεθόδων που έχουνε αύτην την αδυναμία είναι οι ενδογενείς εκ-των-προτέρων κατανομές (Berger και Pericchi 1996), οι g εκ-των-προτέρων κατανομές (Liang κ.α 2008) και οι δυναμικές εκ-των-προτέρων κατανομές (O'Hagan 1995 και Ibrahim και Chen 2002).

Σε μεγάλα δείγματα, αποδεικνύεται ότι οι μέθοδοι επιλογής μεταβλητών που βασίζονται σε μη-τοπικές εκ-των-προτέρων κατανομές είναι πιο ευέλικτες στην εύρεση του σωστού μοντέλου και έχουν λιγότερα προβλεπτικά σφάλματα σε σχέση με τις παραδοσιακές Μπεϋζιανές μεθόδους (Johnson και Rossel 2012). Στην πραγματικότητα, είναι πολύ χρήσιμο όχι μόνο να βρεθεί το πιο πιθανό μοντέλο για ένα σετ δεδομένων, αλλα να βρεθεί και η πιθανότητα το παρατηρούμενο μοντέλο να είναι το σωστό. Ένα σημαντικό πλεονέκτημα των Μπεϋζιανών μεθόδων επιλογής μεταβλητών που βασίζονται στις μη τοπικές εκ-των-προτέρων κατανομές είναι ότι παρέχουνε εκτίμηση των εκ-των-υστέρων πιθανότητων για το κάθε μοντέλο ότι είναι το σωστό. Σε προσομοιωμένες μελέτες οι (Johnson και Rossel 2012) έδειξαν ότι αυτές οι εκ-των-υστέρων πιθανότητες συγκλίνουνε στις εμπειρικές πιθανότητες των σωστών μοντέλων.

Αντίθετα, οι πιο γνωστοί κλασσικοί αλγόριθμοι βρίσκουν το μοντέλο που μεγιστοποιεί μια ποινικοποιημένη πιθανοφάνεια, ενώ οι περισσότεροι Μπεϋζιανοί αλγόριθμοι παρέχουν εκ-των-υστέρων πιθανότητες των μοντέλων οι οποιές δεν μπορούνε να ερμηνευτούνε λογικώς ώς εκ-των-υστέρων πιθανότητες με όλη την έννοια καθώς δεν παρέχουν άλλα μέτρα εκτίμησης. Για παράδειγμα, οι συνηθισμένες Μπεϋζιανές μέθοδοι δίνουν πολύ μικρές εκ-των υστέρων πιθανότητες μοντέλων σε προβλήματα πολλών διαστάσεων, ακόμα και όταν η μέγιστη εκ-των-υστέρων πιθανότητα του μοντέλου δίνει σχετικά μεγάλη πιθανότητα στο πραγματικό μοντέλο. Γιάυτόν τον λόγο οι δημοσιεύσεις που περιγράφουν τις Μπεϋζιανές επιλογές μεταβλητών δεν αναφέρουνε την εκ-των-υστέρων πιθανότητα του πιο πιθανού μοντέλου και αναφέρουν τις περιθώριες πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών στο μοντέλο δειγματοληπτώντας απο την απο κοινού εκ-των-υστέρων κατανομή των παραμέτρων.

3.4.1 Μη τοπικές εκ-των-προτέρων κατανομές στην γραμμική παλινδρόμηση

Μεγάλο ενδιαφέρον επικεντρώνεται στον τρόπο με τον οποίο καθορίζονται οι μη τοπικές εκ-των-προτέρων κατανομές για τις παραμέτρους του απλού γραμμικού μοντέλου. Παρόλο που η μεθοδολογία μπορεί να επεκταθεί και για τα γενικευμένα γραμμικά μοντέλα, θα περιορίσουμε την προσοχή μας στην μελέτη του απλού γραμμικού μοντέλου. Επίσης, κάνουμε την παραδοχή ότι το πραγματικό μοντέλο είναι στοιχείο του χώρου των μοντέλων. Επιπλέον, είναι ιδιαίτερα σημαντικό να αναφερθεί ότι οι μη τοπικές κατανομές γινομένου ροπών και γινομένου αντίστροφων ροπών

προκύπτουν απο το γινόμενο επιμέρους ανεξάρτητων ροπών και αντίστροφων ροπών (Johnson και Rossel 2010). Έστω οτι έχουμε τυχαίο δείγμα $y_1, \ldots, y_n$ απο μια μεταβλητή απόκρισης $\boldsymbol{Y}$, $\boldsymbol{X_\gamma}$ ο πίνακας σχεδιασμού διάστασης $n \times p_\gamma$ και $\boldsymbol{\beta_\gamma}$ το διάνυσμα των παραμέτρων του μοντέλου, $\boldsymbol{\gamma}$ είναι η παράμετρος εισαγωγής των ανεξάρτητων μεταβλητών, θα επικεντρωθούμε σε γραμμικά μοντέλα της μορφής:

$$\boldsymbol{Y} \mid \boldsymbol{\beta_\gamma}, \sigma_\gamma^2 \sim N(\boldsymbol{X_\gamma \beta_\gamma}, \sigma_\gamma^2 \boldsymbol{I_\gamma}) \tag{3.8}$$

Οι εκ-των-προτέρων κατανομές γινομένου ροπών r-τάξης για το απλό γραμμικό μοντέλο της μορφής 3.8 για το διάνυσμα $\boldsymbol{\beta_\gamma}$ των παραμέτρων του μοντέλου με κατανομή ως βάση την κανονική κατανομή $N_p(0, \tau_\gamma \sigma_\gamma^2 \boldsymbol{A_{p\gamma}}^{-1})$ ορίζονται ως εξής:

$$f(\boldsymbol{\beta_\gamma} | \tau_\gamma, \sigma_\gamma^2, r) = d_{p_\gamma} (2\pi)^{-\frac{p_\gamma}{2}} (\tau_\gamma \sigma_\gamma^2)^{-r p_\gamma - \frac{p_\gamma}{2}} |\boldsymbol{A_{p_\gamma}}|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2\tau_\gamma \sigma_\gamma^2} \boldsymbol{\beta_\gamma}^T \boldsymbol{A_{p_\gamma}} \boldsymbol{\beta_\gamma} \right\} \prod_{i=1}^{p_\gamma} \beta_{\gamma_i}^{2r} \tag{3.9}$$

όπου $\tau > 0$, $\boldsymbol{A_{p_\gamma}}$ είναι πίνακας ακρίβειας διάστασης $p_\gamma \times p_\gamma$, $r = 1, 2, \ldots..$ είναι η παράμετρος τάξης ροπής και $d_{p_\gamma} = \left\{ \int (2\pi)^{\frac{-p_\gamma}{2}} |\boldsymbol{A_{p_\gamma}}|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\beta_\gamma}^T \boldsymbol{A_{p_\gamma}} \boldsymbol{\beta_\gamma} \right\} \prod_{i=1}^{p_\gamma} \beta_{\gamma_i}^{2r} d\beta_\gamma \right\}^{-1}$. Η σταθερά κανονικοποιήσης d_{p_γ} είναι ανεξάρτητη των παραμέτρων σ_γ^2 και τ_γ, η μορφή της έχει άμεση σχέση με την εξίσωση 3.1. Οι (Consonni και La Rocca 2010) πρότειναν μια παρόμοια οικογένεια εκ-των-προτέρων κατανομών για εφαρμογές σε γραφικά μοντέλα επιλογής μεταβλητών.

Απο την άλλη, οι εκ-των-προτέρων κατανομές γινομένου αντίστροφων ροπών για το απλό γραμμικό μοντέλο ορίζονται ως εξής:

$$f(\boldsymbol{\beta_\gamma} | \tau_\gamma, \sigma_\gamma^2, r) = \frac{(\tau_\gamma \sigma_\gamma^2)^{r \frac{p_\gamma}{2}}}{\Gamma(\frac{r}{2})_\gamma^p} \prod_{i=1}^{p_\gamma} |\beta_{\gamma_i}|^{-(r+1)} \exp\left(-\frac{\tau_\gamma \sigma_\gamma^2}{\beta_{\gamma i}^2} \right) \tag{3.10}$$

Όταν ορίζουμε την τιμή της παραμέτρου τ_γ είναι πολυ σημαντικό να ξέρουμε σε τι μονάδες βρίσκονται οι ανεξάρτητες μεταβλητές. Για απλότητα, υποθέτουμε ότι οι στήλες του πίνακα σχεδιασμού έχουν τυποποιηθεί καί έτσι μια τίμη της παραμέτρου τ_γ είναι κατάλληλή για οποιοδήποτε στοιχείο του διανύσματος $\boldsymbol{\beta_\gamma}$. Έαν δεν ισχύει αυτή η υπόθεση, τότε ξεχωριστές υπερπαράμετροι τ_{γ_i} πρέπει να οριστούν για καθε μία απο τις επιδράσεις β_{γ_i}. Ακόμα θα ασχοληθούμε με τον υπολογισμό της περιθώριας πιθανοφάνειας που γίνεται σύμφωνα με την εξίσωση 1.14 για τις εκ-των-προτέρων κατανομές γινόμενου ροπών της 3.1 ως εξής:

$$f(\boldsymbol{\gamma} | \boldsymbol{y}) = d_k (2\pi)^{-\frac{n}{2}} 2^{\hat{\nu}_\gamma} 2\tau^{-r p_\gamma - \frac{p_\gamma}{2}} \left[\frac{|A_\gamma|}{|\widehat{C}_\gamma|} \right]^{\frac{1}{2}} \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} (\hat{\nu}_\gamma \hat{s}_\gamma^2)^{-\frac{\hat{\nu}_\gamma}{2}} \Gamma\left(\frac{\hat{\nu}_\gamma}{2} \right) E_\gamma \left(\prod_{i=1}^{p_\gamma} \beta_{\gamma_i}^{2r} \right) \tag{3.11}$$

$$\widehat{\mu}_\gamma = \widehat{C}_\gamma^{-1} X_\gamma^T y$$

$$\widehat{C}_\gamma = X_\gamma^T X_\gamma + \frac{1}{\tau} A_\gamma$$

$$R_\gamma = y^T (I_n - X_\gamma \widehat{C_\gamma^{-1}} X_\gamma^T) y$$

όπου $\hat{\nu}_\gamma = n + 2rp_\gamma + 2\alpha_0$, $\hat{s}_\gamma^2 = \frac{2\lambda_0 + R_\gamma}{\hat{\nu}_\gamma}$, $E_\gamma(.)$ είναι η μέση τιμή πολυμεταβλητής $t - Student$ κατανομής με παραμέτρους $\widehat{\mu}_\gamma$, $\hat{s}_\gamma^2 (\hat{C}_\gamma)^{-1}$, $\hat{\nu}_\gamma$. Απο την άλλη, δεν υπάρχουν σε κλειστή μορφή οι περιθώριες πιθανοφάνειες που βασίζονται στις εκ-των-προτέρων κατανομές γινομένου αντίστροφων ροπών της εξίσωσης 3.10. Οι συναρτήσεις πυκνότητας πιθανότητας που περιγράφονται στις εξισώσεις 3.9 και 3.10 είναι μη τοπικές εκ-των-προτέρων κατανομές στο μηδέν γιατί είναι μηδέν οταν οποιοδήποτε στοιχείο του διανύσματος $\boldsymbol{\beta_\gamma}$ είναι μηδέν. Οι Μπεϋζιανές μέθοδοι επιλογής μεταβλητών με την χρήση των μη τοπικών εκ-των-προτέρων κατανομών έχουνε την δυνατότητα να εκμηδενίσουν τυχόν μοντέλα που περιέχουν μη σημαντικές μεταβλητές. Αντίθετα, οι Μπεϋζιανές μέθοδοι με την χρήση των τοπικών εκ-των-προτέρων κατανομών δίνουνε θετικές τιμές πυκνότητας πιθανότητας στο διάνυσμα των παραμέτρων $\boldsymbol{\beta_\gamma}$ του μοντέλου το οποίο μπορεί να περιέχει στοιχεία που είναι μηδέν. Οι μη τοπικές εκ-των-προτέρων κατανομές των εξισώσεων 3.9 και 3.10 διαφέρουν κατα πολύ απο τις πολυμεταβλητές μη τοπικές εκ-των-προτέρων κατανομές που παρουσίασαμε στην Ενότητα 3.3.3. Καθώς οι πολυμεταβλητές μη τοπικές κατανομές της ενότητας 3.3.3 είναι μηδέν όταν όλα τα στοιχεία του διανύσματος θ είναι μηδέν. Έτσι αυτές οι συναρτήσεις πυκνότητας πιθανότητας εισάγουν μικρή ή μηδενική ποινή σε μοντέλα με πολλές παράμετρους που έχουνε εκτιμήσεις κοντά στο μηδέν, ακόμα και όταν μία ή περισσότερες παραμέτροι που εισήχθησαν στο μοντέλο δεν είναι μηδέν. Απο την άλλη, οι μη τοπικές εκ-των-προτέρων κατανομές του γραμμικού μοντέλου είναι μηδέν όταν κάποιο στοιχείο του διανύσματος $\boldsymbol{\beta_\gamma}$ είναι μηδέν, έτσι αυτό εισάγει μεγαλύτερη ποινή.

3.5 Επιλογή μεταβλητών με μη τοπικές εκ-των-προτέρων κατανομές

Η εύρεση του βέλτιστου μοντέλου μεγίστης εκ-των-υστέρων πιθανότητας είναι συνήθως απαιτητική και επίπονη για δύο λόγους. Αρχικά, όπως αναφέραμε και στο Κεφάλαιο 2, ο χώρος των μοντέλων εχει πλήθος 2^p, και έτσι καθιστάται πρακτικά αδύνατος ο υπολογισμός των περιθώριων πιθανοφάνειων για κάθε μοντέλο. Επιπλέον, ο υπολογισμός της περιθώριας πιθανοφάνειας του κάθε μοντέλου απαιτεί τον υπολογισμό πολλαπλών ολοκληρωμάτων.

Θα αντιμετωπίσουμε το πρόβλημα της μεγάλης διάστασης υιοθετώντας έναν απο τους γνωστούς αλγόριθμους εύρεσης μοντέλων που προτάθηκαν για την κλασσική επιλογή μεταβλητών

όπως η παλινδρόμηση ελάχιστης γωνίας (Efron κ.α. 2004) και η τοπική τετραγωνική προσέγγιση (Fan και Li 2001). Ωστόσο, προκειμένου να βρούμε το πιο πιθανό μοντέλο, ενδιαφερόμαστε για την εύρεση της πιθανότητας αυτού του μοντέλου και στον υπολογισμό των πιθανοτήτων και άλλων πιθανών μοντέλων. Για τον λόγο αυτό, χρησιμοποιούνται οι μέθοδοι MCMC οι οποίες μας επιτρέπουν να πάρουμε δείγμα απο την απο κοινού εκ-των-υστέρων κατανομή, δηλαδή να πάρουμε δείγμα απο τον χώρο των μοντέλων.

Οι υπολογιστικές δυσκολίες που σχετίζονται με τον υπολογισμό της περιθώριας πιθανοφάνειας ποικίλουν ανάλογα με την επιλογή των μη τοπικών εκ-των-προτέρων κατανομών για το διάνυσμα $\boldsymbol{\beta}$ των παραμέτρων του μοντέλου. Στην περίπτωση των κατανομών γινομένου ροπών, η περιθώρια πιθανοφάνεια της εξίσωσης 3.12 είναι διαθέσιμη και στην βιβλιογραφία (Kan 2008). Ωστόσο, η υπολογιστική δυσκολία αυτών των ποσοτήτων αυξάνει με μεγάλο ρυθμό όσο αυξάνουνε και οι διαστάσεις του μοντέλου. Επιπλέον, έαν ο πίνακας ακρίβειας A_γ δεν είναι ο ταυτοτικός πίνακας, τότε η σταθερά κανονικοποιήσης d_{p_γ} είναι εξίσου δύσκολο να υπολογισθεί.

Για να αντιμετωπίσουμε αυτά τα προβλήματα, θέτουμε $A_\gamma = I_\gamma$ ακόμα και όταν δεν υπάρχει υποκειμενική πληροφορία για την εκ-των-προτέρων συσχέτιση των παραμέτρων του μοντέλου $\boldsymbol{\gamma}$. Θα εφαρμόσουμε την προσέγγιση του Laplace (Tierney και Kadane 1986) για να εκτιμήσουμε την περιθώρια πιθανοφάνεια των επιμέρους μοντέλων. Για την περίπτωση των εκ-των-προτέρων ροπών γινομένου με πίνακα ακρίβειας $A_\gamma = I_\gamma$, με σταθερά κανονικοποιήσης $d_{p_\gamma} = (2r-1)^{-p_\gamma}$ και η εκτίμηση της περιθώριας πιθανοφάνειας για το μοντέλο $\boldsymbol{\gamma}$ γίνεται ως εξής:

$$\hat{f}(\boldsymbol{\gamma}|\boldsymbol{y}) = \frac{\Gamma\left(\frac{\hat{\nu}_\gamma}{2}\right)\lambda_0^{\alpha_0}2^{\frac{\hat{\nu}_\gamma}{2}}\left(2\lambda_0 + \boldsymbol{y^T y} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}^{\boldsymbol{T}}\widehat{\boldsymbol{C}}_{\boldsymbol{\gamma}}\widehat{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}\right)^{-\frac{\hat{\nu}_\gamma}{2}}}{\Gamma(\alpha_0)\left[(2r-1)\right]^{-p_\gamma}(2\pi)^{\frac{n}{2}}\tau^{\frac{p_\gamma}{2}+rp_\gamma}} \frac{\prod_{i=1}^{p_\gamma}(\hat{\beta}_i)^{2r}\exp\left\{-\frac{\hat{\nu}_\gamma-2}{2\hat{\nu}_\gamma}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\gamma}})^{\boldsymbol{T}}\frac{\widehat{\boldsymbol{C}}_{\boldsymbol{\gamma}}}{\hat{s}_\gamma^2}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\gamma}})\right\}}{\left|\widehat{\boldsymbol{C}}_{\boldsymbol{\gamma}} + 2r\frac{\hat{\nu}_\gamma\hat{s}_\gamma^2}{\hat{\nu}_\gamma-2}\widehat{\boldsymbol{D}}_{(}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}})\right|} \tag{3.12}$$

όπου $\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}})$ είναι ο διαγώνιος πίνακας με στοιχεία στην κύρια διαγώνιο $\frac{1}{\hat{\beta}_i^2}$,
$\hat{\beta}_\gamma = argmax_{\boldsymbol{\beta}_{\boldsymbol{\gamma}}}\left\{N_{p_\gamma}\left(\boldsymbol{\beta}_{\boldsymbol{\gamma}}; \hat{\boldsymbol{\mu}}_{\boldsymbol{\gamma}}, \frac{\hat{\nu}_\gamma}{\hat{\nu}_\gamma-2}\hat{s}_\gamma^2\widehat{\boldsymbol{C}}_{\boldsymbol{\gamma}}^{-}\boldsymbol{1}\right)\prod_{i=1}^{p_\gamma}\beta_i^{2r}\right\}$, δηλαδή η επικρατούσα τιμή της συγκεκριμένης παράστασης. Αντίθετα για την περίπτωση των κατανομών γινομένου αντίστροφων ροπών έχουμε ότι η εκτίμηση της περιθώριας πιθανοφάνειας δίνεται ως εξής για την περίπτωση $r = 1$:

$$\hat{f}(\boldsymbol{\gamma}|\boldsymbol{y}) = \frac{\lambda_0^{\alpha_0}(2\tau)^{p_\gamma}}{(2\pi)^{\frac{n}{2}}\Gamma(\alpha_0)} \frac{\exp f(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}, \hat{\eta}_\gamma)}{\left|\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}, \hat{\eta}_\gamma)\right|^{\frac{1}{2}}} \tag{3.13}$$

όπου $(\widehat{\beta}_\gamma, \hat{\eta}_\gamma) = argmax_{(\beta_\gamma,\eta_\gamma)}f(\beta_\gamma,\eta), \eta = \log\sigma^2$,
$f(\beta_\gamma,\eta_\gamma) = -\frac{2\lambda_0+(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}})^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}})}{2\exp\eta_\gamma} - \frac{\eta(n-rp_\gamma+2\alpha_0)}{2} - (r+1)\left[\sum_{i=1}^{p_\gamma}\frac{\tau\exp\eta_\gamma}{\beta_{\gamma_i}^2} + \log|\beta_{\gamma_i}|\right]$,

$\widehat{V}(\boldsymbol{\beta_\gamma}, \eta_\gamma)$ είναι πίνακας τάξης $p_\gamma \times p_\gamma$ με στοιχεία ως εξής:
$\widehat{V}_{11} = -\exp -\eta_\gamma(\boldsymbol{X_\gamma^T X_\gamma}) + (r+1)\sum_{i=1}^{p_\gamma} \frac{1}{\beta_{\gamma_i}^2} - \sum_{i=1}^{p_\gamma} \frac{6\tau \exp \eta}{\beta_{\gamma_i}^4}$,
$\widehat{V}_{12} = \exp -\eta_\gamma(-\boldsymbol{X_\gamma^T y} + \boldsymbol{\beta_\gamma^T}(\boldsymbol{X_\gamma^T X_\gamma}) + \sum_{i=1}^{p_\gamma} \frac{2\tau \exp \eta_\gamma}{{\beta_{\gamma_i}}^3}$,
$\widehat{V}_{22} = -\frac{1}{2}\exp -\eta_\gamma(\boldsymbol{y - X_\gamma \beta_\gamma})^T(\boldsymbol{y - X_\gamma \beta_\gamma}) - \sum_{i=1}^{p_\gamma} \frac{\tau \exp \eta_\gamma}{\beta_{\gamma_i}^2} - \lambda_0 \exp -\eta_\gamma$

3.5.1 Αλγόριθμος επιλογής μεταβλητών με μη-τοπικές κατανομές

Σύμφωνα με τις περιθώριες πιθανοφάνειες των εξισώσεων 3.12 και 3.13, προτείνεται τον ακόλουθο αλγόριθμο MCMC που εξερευνά των χώρο των μοντέλων ως εξής (Johnson και Rossel 2012) :

Βήμα 1

► Επιλέγω ένα συγκεκριμένο μοντέλο $\boldsymbol{\gamma_{curr}}$

Βήμα 2

Για $i = 1, \ldots, p$
► Καθορίζω ενα υποψήφιο μοντέλο $\boldsymbol{\gamma_{cand}}$ εισάγοντας ή μή εισάγοντας τις επιδράσεις β_{γ_i} απο το συγκεκριμένο μοντέλο $\boldsymbol{\gamma_{curr}}$

► Υπολογίζω την εξής ποσότητα:
$r = \frac{f(\boldsymbol{\gamma_{cand}}|\boldsymbol{y})f(\boldsymbol{\gamma_{cand}})}{f(\boldsymbol{\gamma_{cand}}|\boldsymbol{y})f(\boldsymbol{\gamma_{cand}})+f(\boldsymbol{\gamma_{curr}}|\boldsymbol{y})f(\boldsymbol{\gamma_{curr}})}$
► Παράγω $u \sim U(0,1)$, εάν $r > u$, τότε έχουμε για το μοντέλο $\boldsymbol{\gamma_{curr}} = \boldsymbol{\gamma_{cand}}$

Βήμα 3
► Επαναλαμβάνουμε το Βήμα 2 μέχρι να γίνει σύγκλιση στην στάσιμη κατανομή που είναι η απο κοινού κατανομή που προσομοιώσαμε δείγμα

Η συχνότητα των δειγματοληπτούμενων μοντέλων χρησιμοποιείται για την εύρεση του μέγιστου εκ-των-υστέρων μοντέλου. Για την επιλογή ως αρχικού μοντέλου προτείνεται συνήθως το μηδενικό μοντέλο $\boldsymbol{\gamma}$, δηλαδή αυτό που περιέχει μόνο την σταθερά, και στην συνέχεια μεταπηδάμε στο υποψήφιο μοντέλο $\boldsymbol{\gamma_{cand}}$

3.5.2 Παράδειγμα 1 επιλογής μεταβλητών με μη-τοπικές εκ-των προτέρων κατανομές

Επιστρέφοντας στο παράδειγμα του κεφάλαιου 2 για τα δεδομένα της παχυσαρκίας, εφαρμόσαμε την Μπεϋζιανή στάθμιση μοντέλων χρησιμοποιώντας τις μεθόδους επιλογής μεταβλητών των τοπικών εκ-των-προτέρων κατανομών μέσω του πακέτου BAS.
Για τα ίδια δεδομένα θα χρησιμοποιήσουμε τις μεθόδους επιλογής μεταβλητών με μη τοπικές εκ-των-προτέρων κατανομές μέσω του πακέτου **mombf**. Ακόμα, παρουσιάζουμε αποτελέσματα με βάση το γινόμενο ροπών πρώτης τάξης MOMFO, το γινόμενο ροπών δευτέρης τάξης MOMSO, το γινόμενο αντίστροφων ροπών πρώτης τάξης IMOMFO .
Η ανάλυση των δεδομένων έγινε χρησιμοποιώντας ως υπερ εκ-των προτέρων κατανομή για την παράμετρο $\boldsymbol{\gamma}$ την $beta - binomial(1,1)$, ενώ για την παράμετρο σ^2 την $IG(0.01, 0.01)$.
Ο καθορισμός της εκ-των-προτέρων παραμέτρου τ έγινε με βάση τις προσομοιωμένες μελέτες των (Johnson και Rossell 2012), για την μέθοδο MOMFO $\tau = 0.348$, για την μέθοδο MOMSO $\tau = 0.072$ και για την μέθοδο IMOMFO $\tau = 0.113$.
Όλες οι λεπτομέρειες συνοψίζονται στον Πίνακα 3.1 και χρησιμοποιούνται ως αναφορά στα διαγράμματα και τους πίνακες αυτου του παραδείγματος. Για τις μεθόδους επιλογής μεταβλητών του Πίνακα 3.1 χρησιμοποίησαμε το πακέτο **mombf** για 5500 επαναλήψεις MCMC χωρίς να λάβουμε υπόψην της πρώτες 500 επαναλήψεις ως περίοδο ζεστάματος.
Για όλες τις μεθόδους επιλογής μεταβλητών με μη τοπικές κατανομές εφαρμόσαμε Μπεϋζιανή στάθμιση μοντέλών για το πλήρες μοντέλο. Θα αξιολογήσουμε και θα συγκρίνουμε όλες τις μεθόδους με μη τοπικές κατανομές υπολογίζοντας τις εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ κάθε ψευδομεταβλητής, τις εκ-των-υστέρων πιθανότητες μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y_i}$ για τα εκ-των-υστέρων μοντέλα $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$.
Απο τα αποτελέσματα του Πίνακα 3.2 και του Διαγράμματος 3.3 βλέπουμε για τις μεθόδους με μη τοπικές κατανομές στο μοντέλο περιλαμβάνονται το φύλο (C2), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_2, κανονική), η σωματική κατάσταση φοιτητή/φοιτήτριας 12-18 (C_11_3, παχιά) καθώς έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y}) > 0.95$. Η σωματική κατάσταση μητέρας (C_13_2, κανονική), (C_13_3, παχιά) και η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας(A_49_3, πολλές φορές/σχεδόν καθημερινά) είναι σχετικά σημαντικές καθώς έχουν εύρος μέσης εκ-των-υστέρων πιθανότητας εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ 0.03-0.99, 0.41-0.85 και 0.52-0.91 αντίστοιχα.
Ωστόσο, η κατανάλωση συσκευασμένων έτοιμων μαγειρευτών φαγητών του/της φοιτητή/φοιτήτριας (A_49_2, αρκετές φορές) έχει πολυ χαμηλές εκ-των-υστέρων πιθανότητες εισαγωγής σχεδόν σε όλες τις μεθόδους με εξαίρεση την εξής περίπτωση (έχει πιθανότητα εισαγωγής 0.66 στην MOM-

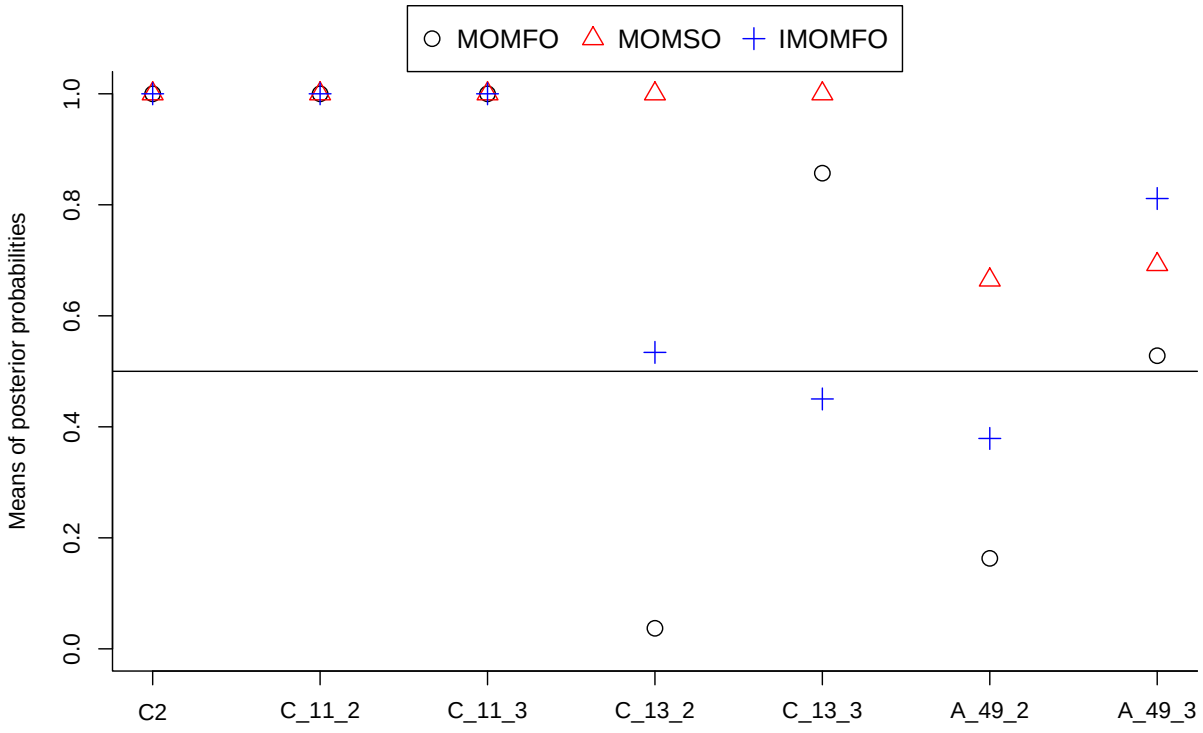
SO).

Πίνακας 3.1: Συντομογραφίες και λεπτομέρειες για τις μεθόδους

	Συντομογραφία	Μέθοδος
1	MOMFO	Γινόμενο ροπών πρώτης τάξης για $r = 1$, $\tau = 0.348$
2	MOMSO	Γινόμενο ροπών δευτέρης τάξης για $r = 2$, $\tau = 0.072$
3	IMOMFO	Γινόμενο αντίστροφων ροπών πρώτης τάξης για $\tau = 0.113$

Πίνακας 3.2: Αποτελέσματα επιλογής μεταβλητών μέσω του πακέτου **mombf** στα οποία αναγράφονται για κάθε μία απο τις ανεξάρτητες μεταβλητές οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.1

	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$						
	Ανεξάρτητες μεταβλητές						
Μέθοδος επιλογής μεταβλητών	$C2$	C_11_2	C_11_3	C_13_2	C_13_3	A_49_2	A_49_3
MOMFO	1	1	1	0.03	0.85	0.16	0.52
MOMSO	1	1	1	0.99	1	0.66	0.69
IMOMFO	1	0.99	1	0.56	0.41	0.38	0.81



Διάγραμμα 3.3: Διάγραμμα των μέσων εκ-των-υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.1

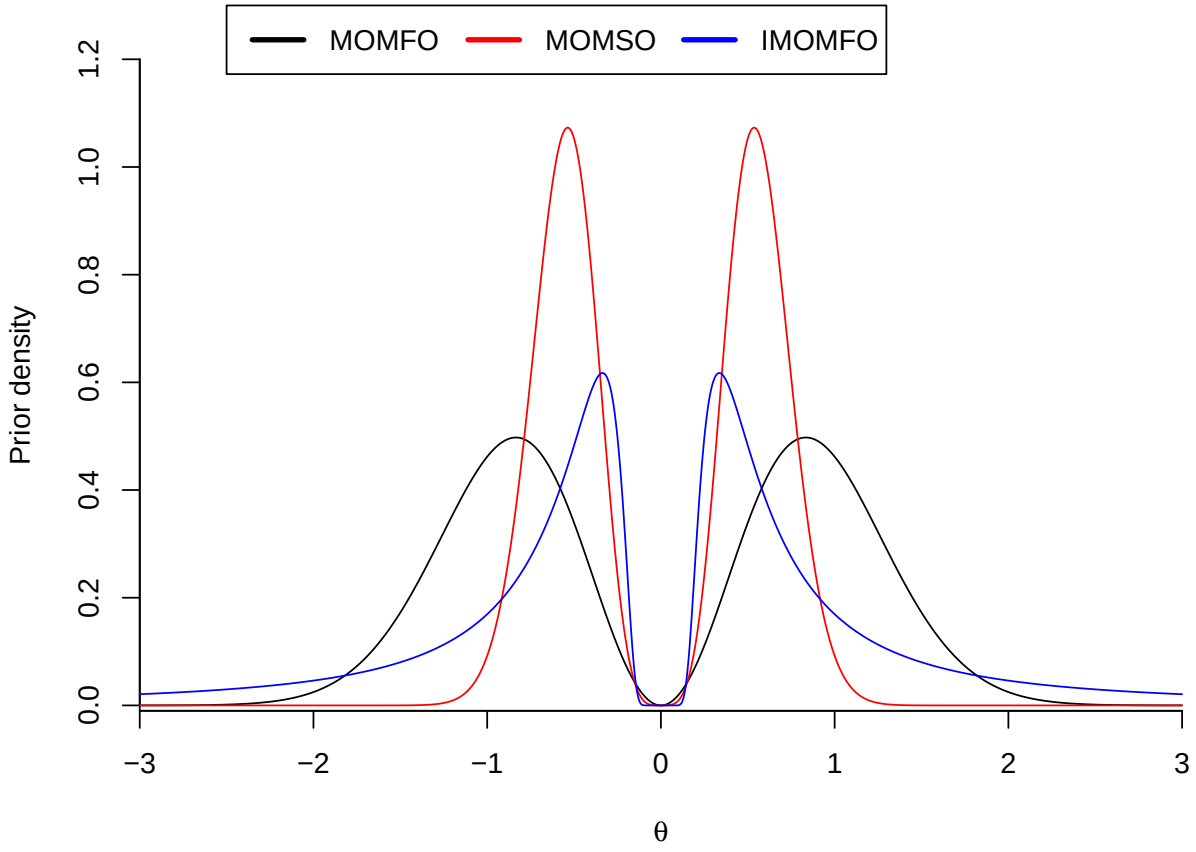
Η μέθοδος IMOMFO παρουσιάζει σχεδόν παρόμοια συμπεριφορά με την MOMFO, με την μόνη διαφορά ότι αποδίδει σχεδόν σε όλες τις ψευδομεταβλητές υψηλότερες μέσες εκ-των-υστέρων πιθανότητες εισαγωγής.

Συγκεκριμένα, για τις σημαντικές ψευδομεταβλητές (C2), (C_11_2), (C_11_3), οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι > 0.99 , καθώς η ψευδομεταβλητή (A_49_3) είναι σχετικά σημαντική με μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.52, 0.69 αντίστοιχα για τις μεθόδους MOMFO, IMOMFO, ενώ η ψευδομεταβλητή (A_49_2) είναι μη σημαντική με μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.52, 0.81 για τις ίδιες μεθόδους επιλογής μεταβλητών.

Επιπλέον, οι ψευδομεταβλητές (C_13_2), (C_13_3) επιλέγονται ως σημαντικές με μέσες εκ-των-υστέρων πιθανότητες εισαγωγής 0.56, 0.85 αντίστοιχα για τις μεθόδους επιλογής μεταβλητών IMOMFO, MOMFO, ενώ οι ίδιες μέθοδοι δεν υποστηρίζουν την επιλογή των ψευδομεταβλητών (C_13_3), (C_13_2) με μέση εκ-των-υστέρων πιθανότητα εισαγωγής 0.41, 0.03.

Πίνακας 3.3: Αποτελέσματα επιλογής μεταβλητών μέσω του πακέτου BAS στα οποία αναγράφονται για κάθε μοντέλο οι εκ-των-υστέρων πιθανότητες $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y}_i$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.1

Μέθοδος	$\boldsymbol{\gamma}$	$\widehat{f}(\boldsymbol{\gamma}\|\boldsymbol{y})$	$RMSE(\widehat{y}_i)$
MOMFO	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_3 + A\_49\_3$	0.35	2.94
MOMSO	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + C\_13\_3 + A\_49\_2 + A\_49\_3$	0.52	3.04
IMOMFO	$C2 + C\_11\_2 + C\_11\_3 + C\_13\_2 + A\_49\_3$	0.27	2.92



Διάγραμμα 3.4: Διάγραμμα των εκ-των-προτέρων κατανομών γινομένου ροπών πρώτης τάξης για $\tau = 0.348$, των εκ-των-προτέρων κατανομών γινομένου ροπών δεύτερης τάξης για $\tau = 0.072$ και των εκ-των-προτέρων κατανομών γινομένου αντίστροφων ροπών πρώτης τάξης για $\tau = 0.072$

Ωστόσο, η μέθοδος MOMSO υποστηρίζει πολυπλόκοτερα μοντέλα σε σχέση με τις υπόλοιπες μεθόδους . Η συγκεκριμένη μέθοδος εκτός απο το ότι δίνει υψηλές πιθανότητες εισαγωγής > 0.99 (C2), (C_11_2), (C_11_3), επιπλέον επιλέγει ως σημαντικές τις ψευδομεταβλητές (C_13_2), (C_13_3) με μέσες εκ-των-υστέρων πιθανότητες εισαγωγής 0.99, 1 αντίστοιχα. Επιπλέον, υποστηρίζει την εισαγωγή των ψευδομεταβλητών (A_49_2), (A_49_3) με μέσες εκ-των-υστέρων πιθανότητες εισαγωγής 0.66, 0.69 αντίστοιχα.

Απο τα αποτελέσματα του Πίνακα 3.3 παρατηρούμε ότι οι ψευδομεταβλητές (C2), (C_11_2), (C_11_3), (C_13_2), (A_49_3) περιλαμβάνονται σχεδόν σε όλα τα εκ-των-υστέρων μοντέλα $\boldsymbol{\gamma}$ που επιλέγονται απο όλες τις μεθόδους επιλογής μεταβλητών του Πίνακα 3.1.

Επιπλέον, το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y_i}$ για τις μεθόδους του Πίνακα 3.1 είναι παρόμοιας κλίμακας.

Σε όρους RMSE καλύτερη μέθοδος είναι η IMOMFO (σχεδόν παρόμοιο με αυτό της MOMFO), ενώ η MOMSO είναι χειρότερη. Μέσω του Διαγράμματος 3.4, παρατηρούμε ότι τα συγκεκριμένα αποτελέσματα αποδίδονται στο γεγονός ότι οι εκ-των-προτέρων κατανομές γινομένου αντίστροφων ροπών έχουν τις παχύτερες ουρές (ισοδύναμα και τις μικρότερες μεροληψίες), ενώ οι εκ-των-προτέρων κατανομές γινομένου ροπών δεύτερης τάξης έχουν λεπτότερες ουρές σε μεγάλες τιμές των επιδράσεων $\boldsymbol{\beta}_j$ (δίνει μικρό βάρος σε τιμές των επιδράσεων > 1.2).

3.5.3 Παράδειγμα 2 επιλογής μεταβλητών με μη-τοπικές εκ-των-προτέρων κατανομές

Σε αυτό το παράδειγμα θα αναλύσουμε τα δεδομένα για τον καρκίνο του προστάτη (Stamey κ.α 1989) και θα συγκρίνουμε τα αποτελέσματα των μη τοπικών εκ-των-προτέρων κατανομών σε σχέση με αυτά των τοπικών εκ-των-προτέρων κατανομών που χρησιμοποιούνται μέσω του πακέτου BAS. Τα δεδομένα έχουν $n = 97$ παρατηρήσεις και $p = 8$ μεταβλητές.
Η μεταβλητή απόκρισης Y είναι το επίπεδο ειδικού αντιγόνου στον προστάτη, οι ανεξάρτητες μεταβλητές είναι ο λογάριθμος καρκινικού χώρου (X_1), ο λογάριθμος βάρους του προστάτη (X_2), η ηλικία του ασθενή (X_3), ο λογάριθμος καλοηθούς υπερπλασίας του προστάτη (X_4), εισβολή σπερματοδόχου κύστης (X_5), ο λογάριθμος εισόδου της κάψουλας (X_6), το σκόρ Gleason (X_7), τα ποσοστά του σκόρ Gleason 4 και 5 (X_8).
Επιπλέον, παρουσιάζουμε αποτελέσματα με βάση το γινόμενο ροπών πρώτης τάξης MOMFO, το γινόμενο ροπών δευτέρης τάξης MOMSO, το γινόμενο αντίστροφων ροπών πρώτης τάξης IMOMFO σε σχέση με το κριτήριο πληροφορίας του Akaike (AIC), Μπεϋζιανό κριτήριο πληροφορίας (BIC), την g εκ-των-προτέρων κατανομή (G-p), την Zellner-Siow εκ-των-προτέρων κατανομή (ZS-p), την υπέρ-g-εκ-των-προτέρων κατανομή (HyperG-p), την υπερ-g-εκ-των-προτέρων κατανομή με προσέγγιση του Laplace (HyperGl-p), τις εμπειρικές τοπικές Μπεϋζιανές μεθόδους (EB-L) και τις τις εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους (EB-G).
Η ανάλυση των δεδομένων πραγματοποιήθηκε χρησιμοποιώντας υπερ εκ-των-προτέρων κατανομή $\boldsymbol{beta - binomial}(1,1)$ για την παράμετρο $\boldsymbol{\gamma}$, ενώ για την παράμετρο σ^2 χρησιμοποιήσαμε $\boldsymbol{IG}(0.01, 0.01)$.
Όλες οι λεπτομέρειες συνοψίζονται στον Πίνακα 3.5 και χρησιμοποιούνται ως αναφορά στα διαγράμματα και τους πίνακες αυτου του παραδείγματος.
Για τις μεθόδους επιλογής μεταβλητών με μη τοπικές κατανομές χρησιμοποίησαμε το πακέτο **mombf** για 5500 επαναλήψεις MCMC χωρίς να λάβουμε υπόψην της πρώτες 500 επαναλήψεις ως περίοδο ζεστάματος, ενω για τις μεθόδους επιλογής μεταβλητών με τοπικές κατανομές χρησιμοποιήσαμε το πακέτο BAS με πλήρη απαρίθμηση του χώρου των μοντέλων.
Για όλες τις μεθόδους επιλογής μεταβλητών με μη τοπικές κατανομές και τοπικές κατανομές χρησιμοποιήσαμε Μπεϋζιανή στάθμιση μοντέλών για το πλήρες μοντέλο. Όλες οι μέθοδοι συγκρίνονται χρησιμοποιώντας τις εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y})$ κάθε ανεξάρτητης μεταβλητής, τις εκ-των-υστέρων πιθανότητες μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y_i}$ για τα εκ-των-υστέρων μοντέλα $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$.
Απο τα αποτελέσματα του Πίνακα 3.6 και του Διαγράμματος 3.5 βλέπουμε για τις μεθόδους μη τοπικών και τοπικών κατανομών οτι οι μεταβλητές που περιλαμβάνονται στο μοντέλο είναι ο λογάριθμος καρκινικού χώρου (X_1),ο λογάριθμος βάρους του προστάτη (X_2), εισβολή σπερματο-

δόχου κύστης (X_5) καθώς έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j|\boldsymbol{y}) > 0.93$. Οι εκ-των-υστέρων πιθανότητες εισαγωγής αυτών των ανεξάρτητων μεταβλητών είναι υψηλότερες για τις μεθόδους μη τοπικών κατανομών, ενώ είναι ελαφρώς χαμηλότερές για τις μεθόδους με τοπικές κατανομές με μοναδική εξαίρεση την μεταβλητή X_1 που παραμένει ίδια για όλες τις μεθόδους επιλογής μεταβλητών.
Επιπλέον, η μέθοδος AIC επιλέγει πολύπλοκα μοντέλα καθώς οι εκ-των-υστέρων πιθανότητες εισαγωγής όλων των ανεξάρτητων μεταβλητών είναι υψηλότερες σε σχέση με τις υπόλοιπες μεθόδους.

Πίνακας 3.4: Πραγματικές τιμές του συντελεστή συσχέτισης Pearson και τιμές μερικής συσχέτισης των συντελεστών (σε απόλυτες τιμές)

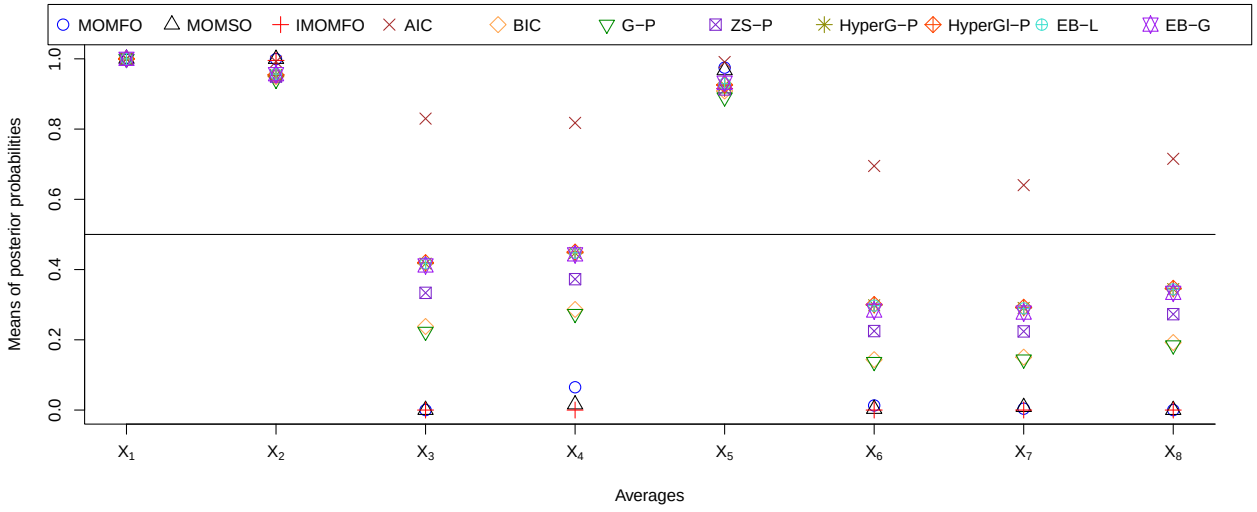
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$\text{corr}(y,\, X_j)$	0.73	0.43	0.17	0.18	0.56	0.55	0.37	0.42
$\text{corr}(y,\, X_j, X_{-j})$	0.56	0.31	0.20	0.17	0.32	0.12	0.03	0.10

Πίνακας 3.5: Συντομογραφίες και λεπτομέρειες για τις μεθόδους

	Συντομογραφία	Μέθοδος
1	MOMFO	Γινόμενο ροπών πρώτης τάξης για $r = 1$, $\tau = 0.348$
2	MOMSO	Γινόμενο ροπών δευτέρης τάξης για $r = 2$, $\tau = 0.072$
3	IMOMFO	Γινόμενο αντίστροφων ροπών πρώτης τάξης για $\tau = 0.113$
4	AIC	Κριτήριο πληροφορίας Akaike
5	BIC	Μπεϋζιανό κριτήριο πληροφορίας
6	G-P	g εκ-των προτέρων κατανομή για $g = 97$
7	ZS-P	Zellner-Siow εκ-των προτέρων κατανομή
8	HyperG-P	Υπέρ g εκ-των-προτέρων κατανομή για $g = 3$
9	HyperGl-P	Υπερ g εκ-των-προτέρων κατανομή με μέθοδο του Laplace για $g = 3$
10	EB-L	Εμπειρικές τοπικές Μπεϋζιανές μεθόδους
11	EB-G	Εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους

Πίνακας 3.6: Αποτελέσματα επιλογής μεταβλητών μέσω του πακέτου **mombf** και του πακέτου BAS στα οποία αναγράφονται για κάθε μία απο τις ανεξάρτητες μεταβλητές οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.5

	$\widehat{f}(\gamma_j = 1\|\boldsymbol{y})$							
	Ανεξάρτητες μεταβλητές							
Μέθοδος επιλογής μεταβλητών	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
MOMFO	1	1	$< .01$	0.01	0.96	$< .01$	$< .01$	$< .01$
MOMSO	1	1	$< .01$	0.06	0.97	0.01	$< .01$	$< .01$
IMOMFO	1	1	$< .01$	$< .01$	0.91	$< .01$	$< .01$	$< .01$
AIC	1	1	0.83	0.81	0.99	0.70	0.64	0.71
BIC	1	0.95	0.24	0.28	0.91	0.14	0.15	0.19
G-P	1	0.94	0.22	0.27	0.89	0.13	0.14	0.18
ZS-P	1	0.95	0.33	0.37	0.91	0.22	0.22	0.27
HyperG-P	1	0.95	0.41	0.44	0.92	0.30	0.29	0.34
HyperGl-P	1	0.95	0.42	0.44	0.92	0.30	0.29	0.34
EB-L	1	0.95	0.41	0.44	0.93	0.30	0.29	0.34
EB-G	1	0.95	0.41	0.44	0.93	0.28	0.27	0.33



Διάγραμμα 3.5: Διάγραμμα των μέσων εκ-των-υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.5

Πίνακας 3.7: Αποτελέσματα επιλογής μεταβλητών μέσω των πακέτων **mombf** και **BAS** στα οποία αναγράφονται για κάθε μοντέλο οι εκ-των-υστέρων πιθανότητες $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ και το μέσο τετραγωνικό σφάλμα RMSE των προβλεπόμενων τιμών $\widehat{y}_i$ για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 3.5

Μέθοδος	$\boldsymbol{\gamma}$	$\widehat{f}(\boldsymbol{\gamma}\|\boldsymbol{y})$	$RMSE(\widehat{y}_i)$
MOMFO	$X_1 + X_2 + X_5$	0.57	0.693
MOMSO	$X_1 + X_2 + X_5$	0.69	0.699
IMOMFO	$X_1 + X_2 + X_5$	0.71	0.696
AIC	$X_1 + X_2 + X_5$	0.35	2.592
BIC	$X_1 + X_2 + X_5$	0.36	3.048
G-P	$X_1 + X_2 + X_5$	0.36	3.056
Zellner-Siow	$X_1 + X_2 + X_5$	0.27	2.916
HyperG-P	$X_1 + X_2 + X_5$	0.21	2.816
HyperGl-P	$X_1 + X_2 + X_5$	0.21	2.816
EB-L	$X_1 + X_2 + X_5$	0.21	2.839
EB-G	$X_1 + X_2 + X_5$	0.2	2.845

Επιπροσθέτως, για τις μη σημαντικές μεταβλητές με χρήση τοπικών εκ-των-προτέρων κατανομών, οι εκ-των-υστέρων πιθανότητες εισαγωγής είναι κοντα στο 0.5, το οποίο αυξάνει την αβεβαιότητα του μοντέλου και ωθεί την κάθε μέθοδο να επιλέξει πολύπλοκα μοντέλα με μη σημαντικές μεταβλητές.

Αντίθετα για τις μέθοδους MOMFO MOMSO IMOMFO των μη τοπικών εκ-των-προτέρων κατανομών οι εκ-των-υστέρων πιθανότητες εισαγωγής των μη σημαντικών μεταβλητών συρρικνώνονται στο μηδέν και έτσι μικραίνεται η αβεβαιότητα του μοντέλου υποστηρίζοντας απλούστερα μοντέλα.

Απο τον Πίνακα 3.7 βλέπουμε ότι οι μεταβλητές λογάριθμος καρκινικού χώρου (X_1), λογάριθμος βάρους του προστάτη (X_2), εισβολή σπερματοδόχου κύστης (X_5) είναι σημαντικές καθώς περιλαμβάνονται σε όλα τα μοντέλα που επιλέγονται απο όλες τις μεθόδους επιλογής μεταβλητών.

Το σημαντικότερο είναι ότι για τις μεθόδους MOMFO, MOMSO, IMOMFO με μη τοπικές κατανομές παρατηρούμε ότι τα μέσα τετραγωνικά σφάλματα RMSE των προβλεπόμενων τιμών $\widehat{y}_i$ είναι πολυ μικρότερα και οι εκ-των υστέρων πιθανότητες των μοντέλων $\widehat{f}(\boldsymbol{\gamma}|\boldsymbol{y})$ είναι υψηλότερες σε σχέση με αυτά των μεθόδων AIC, BIC, G-P, Zellner-Siow, HyperG-P, HyperGl-P, EB-L, EB-G. Το γεγονός αυτο επιβεβαιώνει ότι οι μέθοδοι επιλογής μεταβλητών με μη τοπικές κατανομές έχουνε καλύτερες προβλεπτικές δυνατότητες αφου επιλέγουν απλούστερα μοντέλα συμπεριλαμβάνοντας μόνο τις πραγματικές μη μηδενικές μεταβλητές και συρρικνώνουν τις μη σημαντικές μεταβλητές στο μηδέν το οποίο συμφωνεί με τα αποτελέσματα του Πίνακα 3.6 και του Διαγράμ-

ματος 3.5.
Ο Πίνακας 3.4 παρουσιάζει τις πραγματικές συσχετίσεις των δεδομένων για τον καρκίνο του προστάτη. Οι μεταβλητές λογάριθμος καρκινικού χώρου (X_1), λογάριθμος βάρους του προστάτη (X_2), εισβολή σπερματοδόχου κύστης (X_5) που έχουν τις υψηλότερες εκ-των-υστέρων πιθανότητες εισαγωγής, έχουν και τις υψηλότερες μερικές συσχετίσεις με το ειδικό επίπεδο αντιγόνου στον προστάτη (Y) δοθέντος της παρουσίας των υπόλοιπων.
Όλα τα αποτελέσματα που παρουσιάστηκαν σε αυτό το παράδειγμα δείχνουν ότι οι μέθοδοι επιλογής μεταβλητών με μη τοπικές κατανομές ανταποκρίνονται με επιτυχία στην προηγούμενη κατάσταση και οδηγούν στα ίδια συμπεράσματα με τα αποτελέσματα επιλογής μεταβλητών (Fouskakis και Ntzoufras 2012).

3.6 Συμπεράσματα

Στο κεφάλαιο αυτό, το ενδιαφέρον επικεντρώνεται στον καθορισμό κατάλληλων εκ-των-προτέρων κατανομών που ορίζονται από την εναλλακτική υπόθεση. Τέτοιες κατανομές ονομάζονται εναλλακτικές εκ-των-προτέρων κατανομές. Οι εναλλακτικές εκ-των-προτέρων κατανομές που ορίζονται έτσι ώστε να δίνουν μεγαλύτερο βάρος σε περιοχές που σχετίζονται με την μηδενική υπόθεση ονομάζονται εναλλακτικές τοπικές εκ-των-προτέρων κατανομές (γιατί σχετίζονται με τους παραμετρικούς χώρους που ορίζονται απο την μηδενική υπόθεση).
Αυτές οι κατανομές δεν είναι άλλες απο τις κατανομές που παρουσιάσαμε στα δυο προηγούμενα κεφάλαια και είναι ευαίσθητες σε μεγάλες τιμές της εκ-των-προτέρων διακύμανσης και σε μεγάλες τιμές του μεγέθους του δείγματος, καθώς η ασυμπωτική συμπεριφορά των παραγόντων Bayes που προκύπτουν απο τις κατανομές αυτές εισάγουν μεροληψία κάνοντας του ρυθμούς συσσώρευσης εντελώς ασσύμετρους.
Έτσι, αναζητάμε εκ-των-προτέρων εναλλακτικές κατανομές οι οποίες να ορίζονται έτσι ώστε η εναλλακτική υπόθεση να είναι μια εντελώς διαφορετική θεωρία απο την αντίστοιχη της μηδενικής υπόθεσης. Τέτοιες κατανομές προϋποθέτουν τον καθορισμό μιας παραμέτρου τ η οποία εκφράζει κατα πόσο απέχει η μηδενική απο την εναλλακτική υπόθεση. Αυτές οι κατανομές ονομάζονται μη τοπικές εκ-των-προτέρων κατανομές γιατί μηδενίζουν την μάζα πιθανότητας στην περιοχή της μηδενικής υπόθεσης και δίνουν μεγαλύτερη πιθανότητα σε κοντινές περιοχές γύρω απο την μηδενική υπόθεση ενισχύοντας έτσι τιμές που ορίζονται απο την εναλλακτική υπόθεση.

Τέλος, ισορροπούν τον ρυθμό συσσώρευσης των παραγόντων Bayes χωρίς να εκτείθενται σε μεροληψίες. Όσον αφορά το πρόβλημα της γραμμικής παλινδρόμησης στο οποίο θέλουμε να βρούμε τις πιθανές ανεξάρτητες μεταβλητές που επηρεάζουν μια μεταβλητή απόκρισης Y, οι μέθοδοι επιλογής μεταβλητών που βασίζονται στις μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές συρρικνώνουν με επιτυχία τις μηδενικές επιδράσεις στο μηδέν και τις πραγματικές επιδράσεις στο ένα.

Κεφάλαιο 4

Μελέτη προσομοίωσης

Στο κεφάλαιο 4 είδαμε οτι οι μη τοπικές εκ-των-προτέρων κατανομές εξισσοροπούν τον ρυθμό σύγκλισης της μηδενικής και εναλλακτικής υπόθεσης ως συνάρτηση του μέγεθους του δειγματος, n, όταν η αντίστοιχη υπόθεση είναι αληθής.
Επιπλέον, δεν είναι ευαίσθητες στις υπερ-παραμέτρους που καθορίζουν την εκ-των-προτέρων διακύμανση και στην αύξηση του μεγέθους του δείγματος (Johnson και Rossel, 2010).
Το βασικότερο, οι μη τοπικές εκ-των προτέρων κατανομές χαρακτηρίζονται απο την παράμετρο r η οποία εκφράζει την τάξη της ροπής και την παράμετρο τ η οποία εκφράζει την εκ-των-προτέρων απόκλιση απο το μήδεν.
Έτσι, ο καθορισμός των συγκεκριμένων παραμέτρων πρέπει να γίνεται ιδιαίτερα προσεκτικά προκειμένου να είναι σωστή η εκ-των-υστέρων συμπερασματολογία.
Συγκεκριμένα, θα επικεντρωθούμε στην ανάλυση προσομοιωμένων δειγμάτων προκειμένου να γίνει κατάλληλη ανάλυση ευαισθησίας της παραμέτρου r, της παραμέτρου τ και ως προς το μέγεθος του δείγματος. Είναι ιδιαίτερα σημαντικό να αναφερθεί ότι για τις κατανομές γινομένου αντίστροφων ροπών η ανάλυση των αποτελεσμάτων θα περιοριστεί μόνο στις κατανομές γινομένου αντίστροφων ροπών πρώτης τάξης (δηλαδή για $r = 1$).
Παράλληλα θα γίνει επιπλέον ανάλυση λαμβάνοντας υπόψην την παράμετρο τ ως άγνωστη παράμετρο στην επιλογή μεταβλητών για τις κατανομές γινομένου ροπών, καθώς οι υπολογισμοί για τις κατανομές γινομένου αντίστροφων ροπών είναι ιδιαίτερα πολύπλοκοι. Η ενότητα χωρίζεται σε δύο κύρια μέρη. Στο πρώτο μέρος παρουσιάζουμε αποτελέσματα της προσομοίωσης για συγκεκριμένες τιμές των παραμέτρων r και τ των μη τοπικών εκ-των-προτέρων κατανομών. Στο δεύτερο παρουσιάζουμε την ανάλυση λαμβάνοντας υπόψην την παράμετρο τ ως άγνωστη και χρησιμοποιώντας ως υπερ εκ-των-προτέρων κατανομή $IG \sim (h,u)$.

4.1 Σχέδιο προσομοίωσης

Στην μελέτη προσομοίωσης του κεφάλαιου αυτού, χρησιμοποιούμε το πλάνο προσομοίωσης των **Nott και Kohn** (2005), με 15 επεξηγηματικές μεταβλητές 50 παρατηρήσεων. Συγκεκριμένα, οι πρώτες 10 μεταβλητές ακολουθούν ανεξάρτητες τυποποιημένες κανονικές κατανομές και οι υπόλοιπες 5 μεταβλητές δημιουργούνται ως εξής:

$$(X_{11},\ldots,X_{15}) = (X_1,\ldots,X_5) \times (0.3, 0.5, 0.7, 0.9, 11)^T \times (1,1,1,1,1) + E,$$

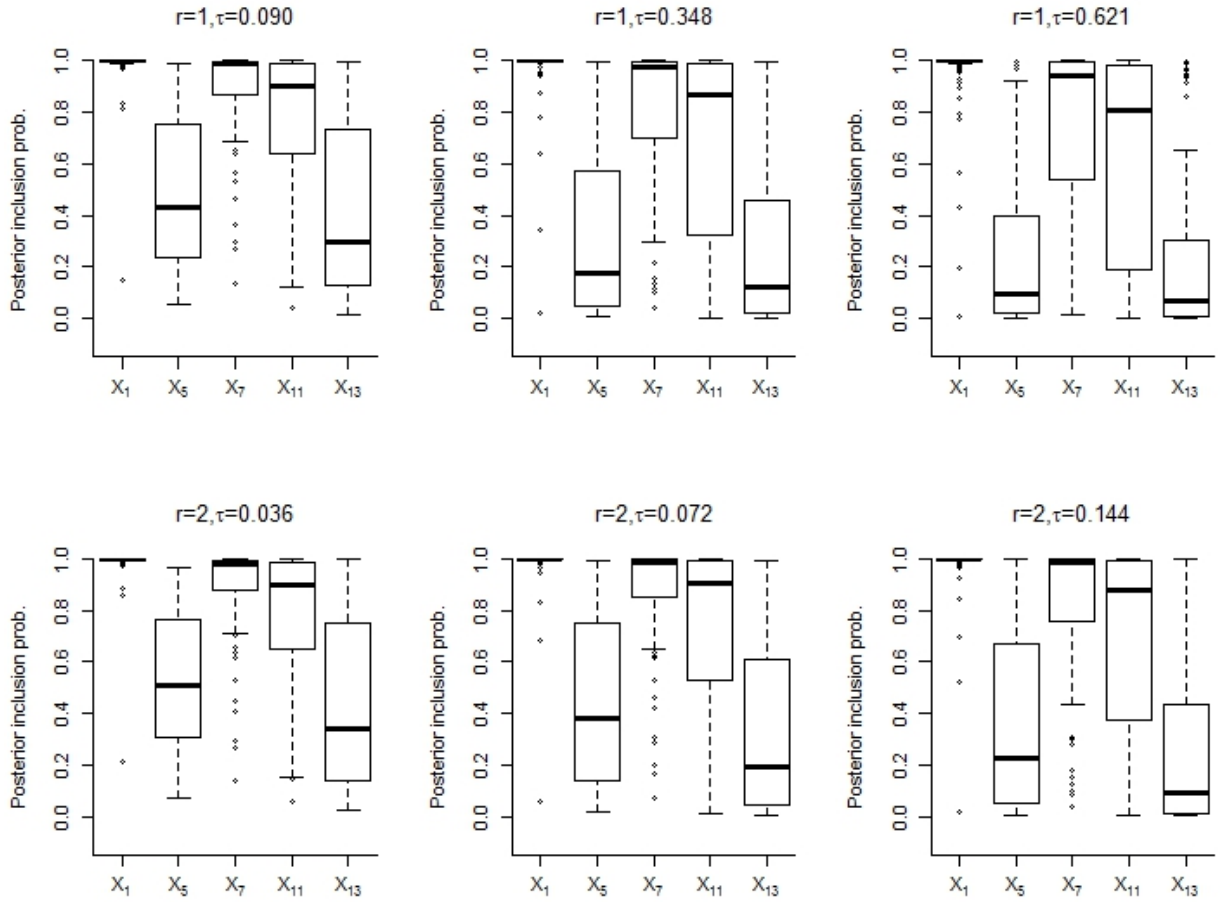
όπου E αποτελείται απο 5 ανεξάρτητες μεταβλητές $N(0,1)$. Σύμφωνα με αυτόν τον σχεδιασμό, οι 5 τελευταίες μεταβλητές είναι υψηλά συσχετισμένες, ενώ είναι μέτριως συσχετισμένες με τις πρώτες 5 μεταβλητές. Η μεταβλητή απόκρισης δημιουργείται ως εξής:

$$Y = 4 + 2X_1 - X_5 + 1.5X_7 + X_{11} + 0.5X_{13} + \epsilon,$$

όπου $\epsilon \sim N(0, 2.5^2)$. Αυτα τα προσομοιωμένα δεδομένα περιλαμβάνουν μεταβλητές οι οποίες είναι συσχετισμένες η μία με την άλλη και έτσι η επιλογή μεταβλητών δεν είναι εύκολη καθώς κάποιες πραγματικές επιδράσεις αλλοιώνονται λόγω της πολυσυγγραμικότητας των επιδράσεων των υπόλοιπων μεταβλητών. Το πηλίκο του σήματος προς τον θόρυβο (**Signal-to-noise ratio**) είναι ίσο με 2.15, όπου **Signal-to-noise ratio** είναι το πηλίκο της διακύμανσης του γραμμικού συνδιασμού προς την διακύμανση των τυχαίων σφαλμάτων. Οι πραγματικές συσχετίσεις ανά ανεξάρτητη μεταβλητή είναι $corr(X_1, X_j) = 0.15$, $corr(X_2, X_j) = 0.26$, $corr(X_3, X_j) = 0.36$, $corr(X_4, X_j) = 0.46$, $corr(X_l, X_j) = 0.74$ για $l, j = 11,\ldots,15$ και $l \neq j$.

4.2 Αποτελέσματα της μελέτης προσομοίωσης για διάφορες τιμές r και τ

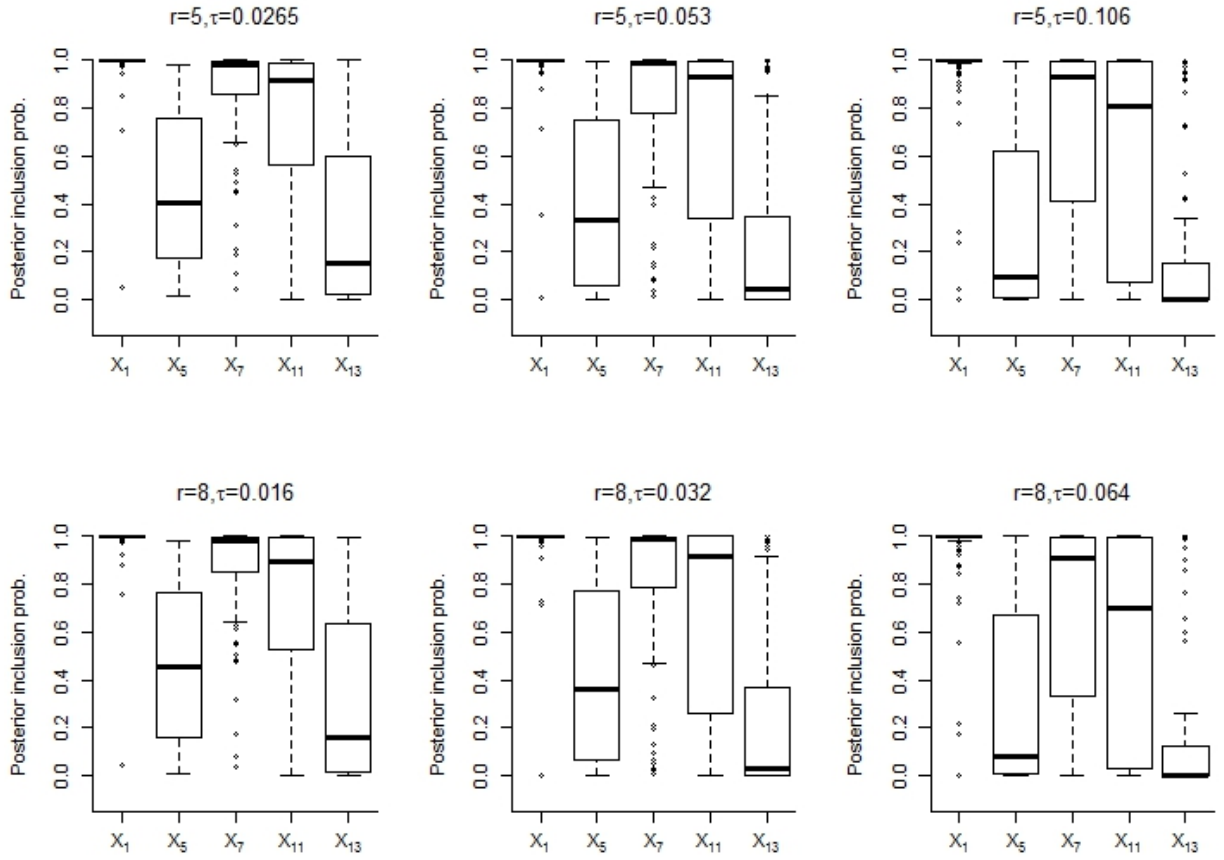
Θα παρουσιάσουμε τα αποτελέσματα χρησιμοποιώντας ενα εύρος τιμών για τις εκ-των προτέρων παράμετρους r και τ. Συγκεκριμένα, κάθε μέθοδος MCMC πραγματοποιήθηκε για 10000 επαναλήψεις χωρίς να λάβουμε υπόψην τις πρώτες 1000 επαναλήψεις ως περίοδο ζεστάματος. Όλα τα αποτελέσματα παρουσιάζονται για 100 σέτ προσομοιωμένων δεδομένων που δημιουργήθηκαν σύμφωνα με το σχέδιο δειγματοληψίας που περιγράφηκε στην Ενότητα 4.1. Τα διαγράμματα 4.1 και 4.2 αναπαριστούν τα διαγράμματα πλαισίου-απολήξεων (για τα 100 σέτ δεδομένων) των εκ-των-υστέρων πιθανοτήτων για τις μεταβλητές $X_1, X_5, X_7, X_{11}, X_{13}$ (δηλαδή των μεταβλητών με τις μη μηδενικές επιδράσεις) για τις τιμές της υπερ-παραμέτρου τ που λάβαμε υπόψην μας.



Διάγραμμα 4.1: Διαγράμματα πλαισίου απολήξεων για τις εκ-των-υστέρων πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών με μη μηδενικές πραγματικές επιδράσεις των 100 σέτ δεδομένων για τις κατανομές γινομένου ροπών πρώτης και δεύτερης τάξης

Συγκεκριμένα, απο τα διαγράμματα 4.1, 4.2, 4.3 παρατηρούμε ότι οι ανεξάρτητες μεταβλητές X_1, X_7, X_{11} είναι σημαντικές γιατί έχουν εκ-των-υστέρων πιθανότητα εισαγωγής > 0.5 για όλες τις τιμές της παραμέτρου τ για όλες τις μη τοπικές εκ-των-προτέρων κατανομές, καθώς οι μεταβλητές X_5, X_{13} παραμένουν μη-σημαντικές λογω ότι η εκ-των-υστέρων πιθανότητα εισαγωγής είναι < 0.5. Επιπλέον οι πιθανότητες εισαγωγής των δύο αυτών μεταβλητών μικραίνουν καθώς οι τιμές της παραμέτρου τ αυξάνουν. Ακόμα, παρατηρείται η μεγάλη μεταβλητότητα των εκ-των-υστέρων πιθανοτήτων εισαγωγής που παρατηρείται μεταξύ των προσομοιωμένων δεδομένων. Οι εκ-των-υστέρων πιθανότητες εισαγωγής για τις υπόλοιπες ανεξάρτητες μεταβλητές $X_2, X_3, X_4, X_6, X_8, X_9, X_{10}, X_{12}, X_{14}, X_{15}$ είναι σχετικά χαμηλές με εύρος απο 0.16 εώς 0.23 για $\tau = 0.090, r = 1$, με εύρος απο 0.07 εώς 0.13 για $\tau = 0.348, r = 1$, με εύρος απο 0.03 εώς 0.10 για $\tau = 0.621, r = 1$, με εύρος απο 0.17 εώς 0.24 για $\tau = 0.036, r = 2$, με εύρος απο 0.09 εώς

0.16 για $\tau = 0.072, r = 2$, με εύρος απο 0.04 εώς 0.13 για $\tau = 0.144, r = 2$, με εύρος απο 0.07 εώς 0.15 για $\tau = 0.0265, r = 5$, με εύρος απο 0.02 εώς 0.12 για $\tau = 0.053, r = 5$, με εύρος απο 0.01 εώς 0.09 για $\tau = 0.106, r = 5$, με εύρος απο 0.06 εώς 0.14 για $\tau = 0.016, r = 8$, με εύρος απο 0.01 εώς 0.11 για $\tau = 0.032, r = 8$, με εύρος απο 0.01 εώς 0.08 για $\tau = 0.064, r = 8$, με εύρος απο 0.18 εώς 0.25 για $\tau = 0.05$, με εύρος απο 0.10 εώς 0.18 για $\tau = 0.113$, με εύρος απο 0.02 εώς 0.10 για $\tau = 0.445$

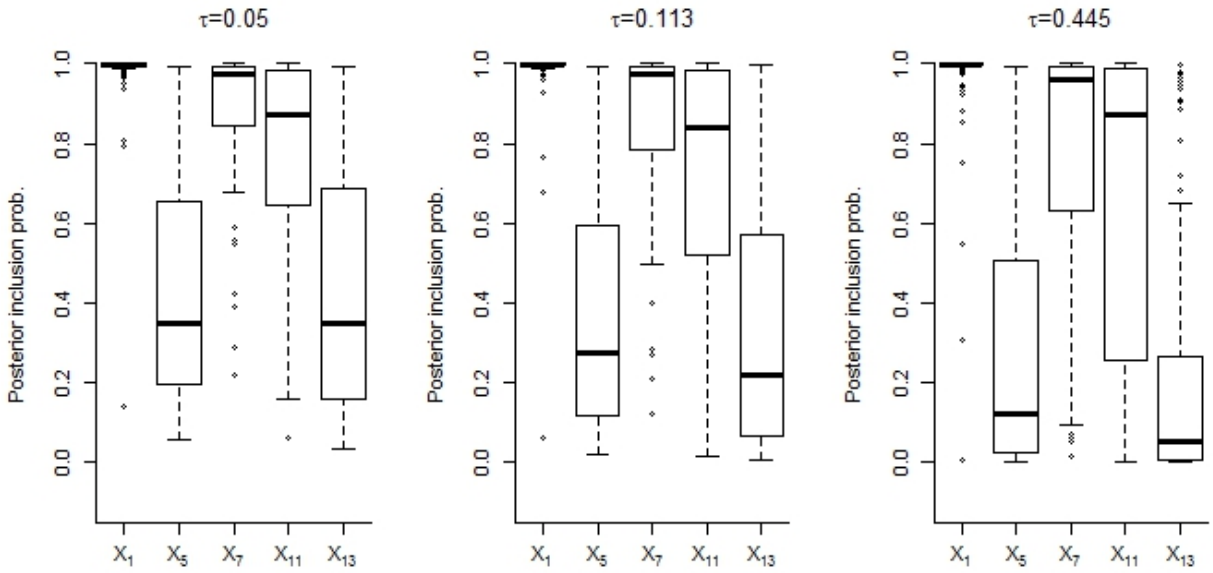


Διάγραμμα 4.2: Διαγράμματα πλαισίου απολήξεων για τις εκ-των-υστέρων πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών με μη μηδενικές πραγματικές επιδράσεις των 100 σέτ δεδομένων για τις κατανομές γινομένου ροπών πέμπτης και όγδοης τάξης

Πίνακας 4.1: Πραγματικές τιμές του συντελεστή συσχέτισης Pearson και τιμές μερικής συσχέτισης των συντελεστών (σε απόλυτες τιμές) για την Ενότητα 4.1 (Lykou και Ntzoufras 2013)

	X_1	X_2	X_3	X_4	X_5	X_6, X_8, X_9, X_{10}	X_7	X_{11}	X_{12}, X_{14}, X_{15}	X_{13}
$corr(y, X_j)$	0.56	0.17	0.24	0.31	0.15	0	0.34	0.56	0.44	0.5
$corr(y, X_j, X_{-j})$	0.55	0	0	0	0.15	0	0.52	0.37	0	0.2

Ο Πίνακας 4.1 παρουσιάζει τις πραγματικές συσχετίσεις των προσομοιωμένων δεδομένων της Ενότητας 4.1 (Lykou και Ntzoufras 2013). Οι μεταβλητές X_1, X_7, X_{11} που έχουν τις υψηλότερες εκ-των υστέρων πιθανότητες, έχουν και τις υψηλότερες μερικές συσχετίσεις με την μεταβλητή απόκρισης Y δοθέντος της παρουσίας των υπόλοιπων. Καθώς οι μεταβλητές X_5, X_{13} χρησιμοποιήθηκαν για την δημιουργία της μεταβλητής απόκρισης Y, οι αντίστοιχες μερικές συσχετίσεις είναι πολύ χαμηλές εξαιτίας των υψηλών συσχετίσεων των μεταβλητών X_5, X_{11}, X_{13}.



Διάγραμμα 4.3: Διαγράμματα πλαισίου απολήξεων για τις εκ-των-υστέρων πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών με μή-μηδενικές πραγματικές επιδράσεις των 100 σέτ δεδομένων για τις κατανομές γινομένου αντίστροφων ροπών πρώτης τάξης

Οι μέθοδοι επιλογής μεταβλητών με βάση τις μη τοπικές κατανομές τείνουν να επιλέξουν μία απο τις τρείς υψηλώς συσχετισμένες μεταβλητές, και γιαυτό οι εκ-των-υστέρων πιθανότητες εισαγωγής των μεταβλητών X_5, X_{13} είναι χαμηλές.

Όλα τα αποτελέσματα που παρουσιάστηκαν σε αυτην την μικρή ενότητα δείχνουν ότι οι μέθοδοι επιλογής μεταβλητών με μη τοπικές κατανομές ανταποκρίνονται με επιτυχία στην προηγούμενη κατάσταση που περιγράψαμε. Οι πραγματικές επιδράσεις των μεταβλητών X_5, X_{13} δεν μπορούν να εντοπιστούν λόγω της πολυσυγραμικότητας.

Συγκεκριμένα, η μεταβλητή X_5 έχει χαμηλή συσχέτιση Pearson και μερική συσχέτιση με την μεταβλητή απόκρισης Y, καθώς η μεταβλητή X_{13} σε όλες σχεδόν τις περιπτώσεις δεν είναι σημαντική παρόλου που έχει υψηλη συσχέτιση Pearson ίση με 0.5 και μερική συσχέτιση ίση με 0.2.

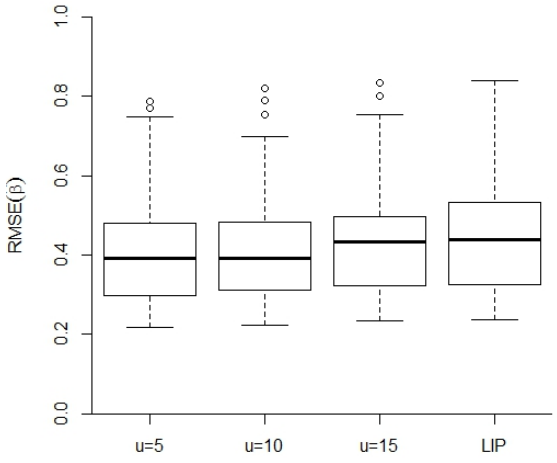
4.3 Ανάλυση ευαισθησίας για την αντίστροφη γάμμα υπερ εκ-των προτέρων κατανομή

Στην Ενότητα αυτή θα παρουσιάσουμε την ανάλυση ευαισθησίας για διάφορες τιμές των υπερ-παραμέτρων της αντίστροφης γάμμα κατανομής για την παράμετρο τ, δηλαδή $\tau \sim IG(h,u)$. Η συγκεκριμένη ανάλυση θα περιοριστεί μόνο στις εκ-των-προτέρων ροπές γινομένου, καθώς μόνο σε αυτην την περίπτωση είναι εφικτοί οι υπολογισμοί των εκ-των υστέρων εκτιμήσεων.
Ώς τιμές αναφοράς θα θεωρήσουμε αυθαίρετα τις επιλογές των τιμών $h = 3$ και $u = 10$ προκειμένου να ξεκινήσουμε την ανάλυση ευαισθησίας.
Στην πρώτη ανάλυση χρησιμοποιούμε τιμές $u = 5$, $u = 10$, $u = 15$ κρατώντας την μέση τιμή ίση με 2.5, 5, 7.5 και την διακύμανση να παίρνει τιμές 6.25, 25, 56.25 αντίστοιχα.
Στην δεύτερη ανάλυση χρησιμοποιούμε τιμές $h = 2$, $h = 3$, $h = 5$, $h = 7$ κρατώντας την διακύμανση ίση με 25, 2, 0.55, 0.22 και την μέση τιμή να παίρνει τιμές 5, 2.5, 1.66, 1.25, αντίστοιχα.
Παρουσιάζουμε επίσης τα αποτελέσματα για την εκ-των-προτέρων κατανομή $IG \sim (3,18)$, η οποία θεωρείται ως χαμηλής πληροφορίας κατανομή, επειδή η διακύμανση της ειναι υψηλή (ίση με 81). Τα Διαγράμματα 4.4 και 4.5 παρουσιάζουν τα αποτελέσματα απο αυτές τις επιμέρους αναλύσεις ευαισθησίας. Πιο συγκεκριμένα, το κάθε Διάγραμμα αναπαριστά τα διαγράμματα πλαισίου-απολήξεων 4.4α΄, 4.5α΄ της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος των εκτιμήσεων των επιδράσεων $\hat{\beta}_j$ σε σχέση με τις πραγματικές επιδράσεις, τα Διαγράμματα πλαισίου-απολήξεων 4.4β΄, 4.5β΄ τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (RMEs) των προβλεπόμενων τιμών $\hat{y}_i$ σε σχέση με τις πραγματικές τιμές, τα Διαγράμματα 4.4γ΄, 4.5γ΄ των μέσων εκ-των-υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|\boldsymbol{y})$, όπου $j = 1,\ldots,15$ καθώς και διαγράμματα πλαισίου απολήξεων 4.4δ΄, 4.5δ΄ της εκ-των-υστέρων μέσης τιμής της παραμέτρου τ των δημιουργημένων δεδομένων.
Οι ποσότητες RMSEs υπολογίζονται σύμφωνα με τις ακόλουθες εξισώσεις:

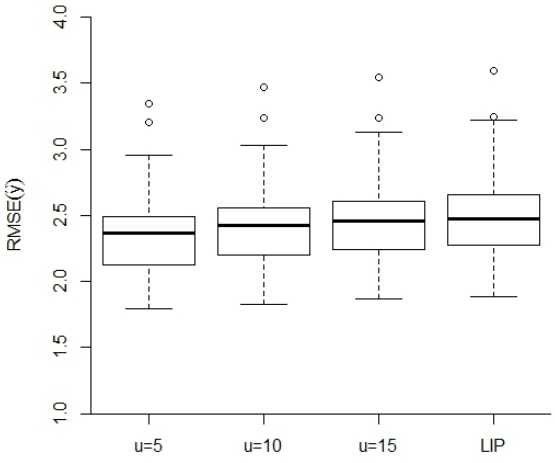
$$RMSE(\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4.1}$$

$$RMSE(\hat{\beta}) = \sqrt{\frac{1}{p}\sum_{j=1}^{n}(\beta_j - \hat{\beta}_j)^2}, \tag{4.2}$$

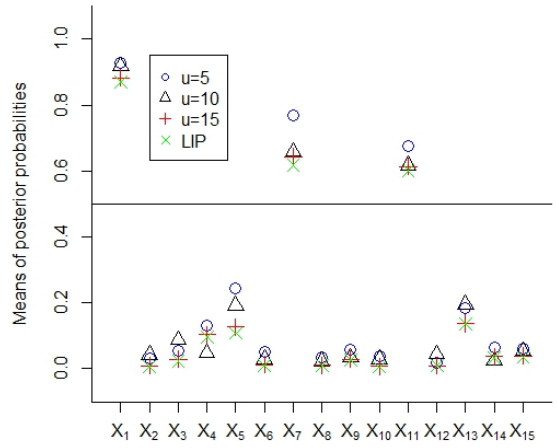
όπου για τις προβλεπόμενες τιμές $\hat{y}_i$ και τις εκτιμώμενες επιδράσεις $\hat{\beta}_j$ έχουμε αντίστοιχα, ότι $\hat{\beta}_j$ ισούνται με τις εκ-των-υστέρων επικρατούσες τιμές, καθώς ότι $\hat{y}_i$ ισούνται με τις $\hat{y}_i = \sum_{j=1}^{p} X_{ij}\hat{\beta}_j$.



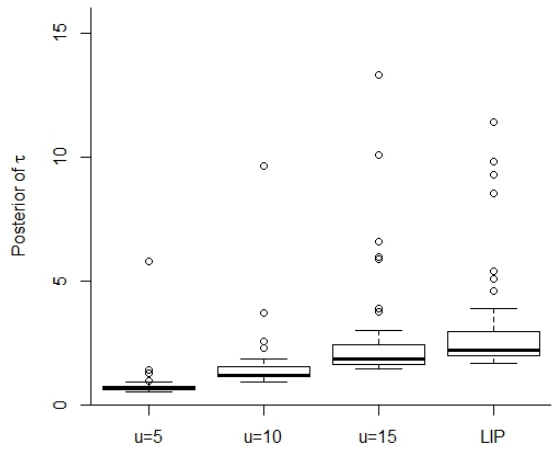
(α′) RMEs των εκτιμώμενων επιδράσεων β_j



(β′) RMEs των προβλεπόμενων τιμών y_i



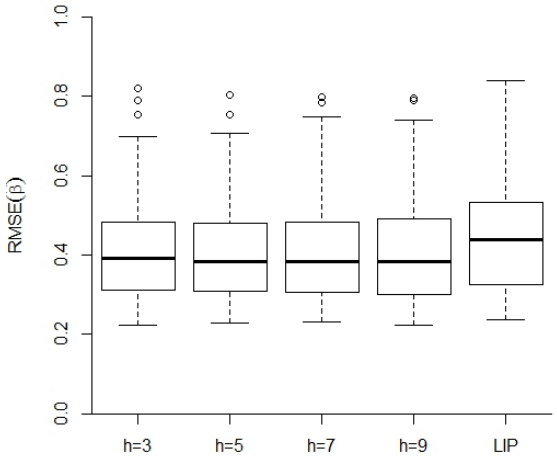
(γ′) Εκ-των-υστέρων μέσες τιμές πιθανότητας $\widehat{f(\gamma_j = 1|y)}$ για 50 δεδομένα προσομοίωσης



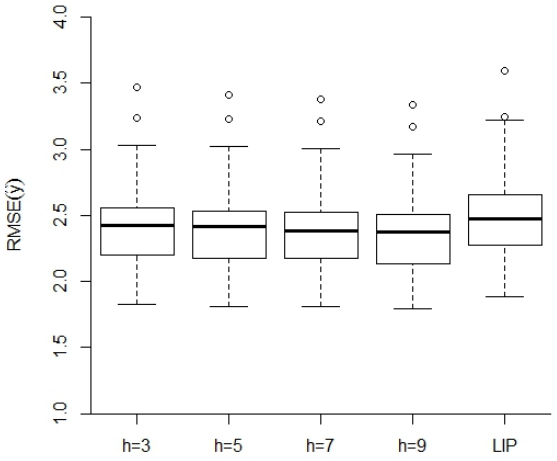
(δ′) Εκ-των-υστέρων μέση τιμή τ για 50 δεδομένα προσομοίωσης

Οι ποσότητες $\boldsymbol{RMSEs}$ υπολογίστηκαν σύμφωνα με τις εξισώσεις 4.1 και 4.2, οι εκτιμώμενες επιδράσεις $\hat{\beta}_j$ είναι ίσες με τις εκ-των-υστέρων διάμεσους της ποσότητας $\beta_j = \gamma_j \beta_j, \hat{y}_i = \sum_{j=1}^{p} X_{ij} \hat{\beta}_j$ και $\boldsymbol{LIP}$: $\boldsymbol{Gamma} \sim (3,18)$ εκ-των-προτέρων κατανομή χαμηλής πληροφορίας

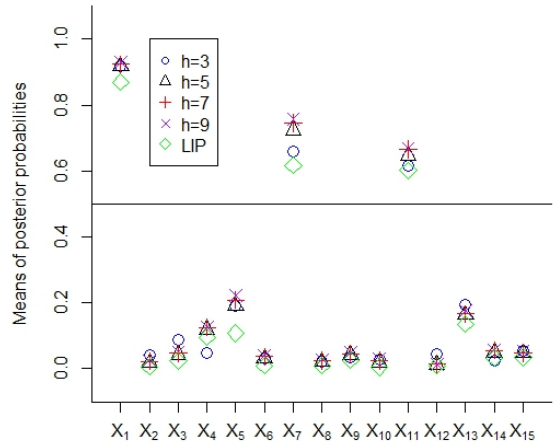
Διάγραμμα 4.4: Ανάλυση ευαισθησίας της μελέτης προσομοίωσης , για τρία ζεύγη εκ-των-προτέρων παραμέτρων για μέση τιμή ίση με 2.5, 5, 7.5 και την διακύμανση να παίρνει τιμές 6.25, 25, 56.25 αντίστοιχα για ($h = 3$ για όλες τις αναλύσεις, $u \in \{5,10,15\}$), για 50 προσομοιωμένα δεδομένα



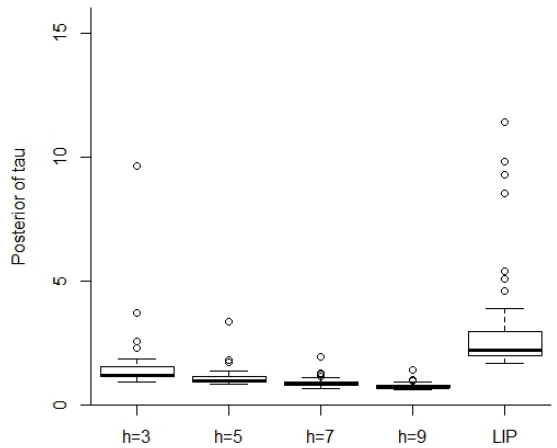
(α′) RMEs των εκτιμώμενων επιδράσεων β_j



(β′) RMEs των προβλεπόμενων τιμών y_i



(γ′) Εκ-των-υστέρων μέσες τιμές πιθανότητας $\widehat{f(\gamma_j = 1|y)}$ για 50 δεδομένα προσομοίωσης



(δ′) Εκ-των-υστέρων μέση τιμή τ για 50 δεδομένα προσομοίωσης

Οι ποσότητες $\boldsymbol{RMSEs}$ υπολογίστηκαν σύμφωνα με τις εξισώσεις 4.1 και 4.2, οι εκτιμώμενες επιδράσεις $\hat{\beta}_j$ είναι ίσες με τις εκ-των-υστέρων διάμεσους της ποσότητας $\beta_j = \gamma_j \beta_j, \hat{y}_i = \sum_{j=1}^{p} X_{ij} \hat{\beta}_j$ και $\boldsymbol{LIP}$: $\boldsymbol{Gamma} \sim (3,18)$ εκ-των-προτέρων κατανομή χαμηλής πληροφορίας

Διάγραμμα 4.5: Δεύτερη ανάλυση ευαισθησίας της μελέτης προσομοίωσης, για 4 ζεύγη εκ-των-προτέρων παραμέτρων για διακύμανση ίση με 25, 2, 0.55, 0.22 και την μέση τιμή να παίρνει τιμές 5, 2.5, 1.66, 1.25 αντίστοιχα για ($u = 10$ για όλες τις αναλύσεις, $h \in \{3,5,7,9\}$), για 50 προσομοιωμένα δεδομένα

Από τα αποτελέσματα της πρώτης ανάλυσης (Διάγραμμα 4.4), παρατηρούμε ότι η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος RMSEs των $\hat{y}$ και η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος RMEs των $\hat{\beta}_j$ είναι σχεδόν όμοια κατανεμημένες για τις τιμές $u = 5$, $u = 10$, $u = 15$ με μικρές διαφορές.

Οι εκ-των-υστέρων πιθανότητες εισαγωγής παρουσιάζουν ελάχιστες διαφορές μεταξύ τους για τις διάφορες τιμές του u, καθώς επίσης παρατηρούνται κάποιες διαφορές στην εκ-των- υστέρων μέση τιμή της παραμέτρου τ καθώς ενδείκνυται μια τάση αύξησης καθώς αυξάνουν οι τιμές της παραμέτρου u οι οποίες δεν φαίνεται να έχουν σημαντική επίδραση στις εκ-των-υστέρων πιθανότητες εισαγωγής και στις ποσότητες RMSEs.

Ακόμα για την τιμή $u = 5$, φαίνεται ότι οι εκ-των-υστέρων πιθανότητες των μη σημαντικών μεταβλητών και των σημαντικών μεταβλητών τείνουν προς υψηλότερες τιμές.

Για την δεύτερη ανάλυση ευαισθησίας (Διάγραμμα 4.5) η εκ-των-προτέρων μέση τιμή αλλάζει για δοθείσες τιμές της υπερπαραμέτρου $h = 3$, $h = 5$, $h = 7$, $h = 9$. Τα αποτελέσματα απο το Διάγραμμα 4.5 είναι παρόμοια με αυτά της πρώτης ανάλυσης ευαισθησίας, δηλαδή οι ποσότητες RMSEs και οι εκ-των-υστέρων πιθανότητες είναι εύρωστες για τις τιμές της υπερπαραμέτρου $h \in \{3,5,7,9\}$), καθώς ότι κατανέμονται με παρόμοιο τρόπο, ενώ παρατηρούμε για την ποσοτητα RMSEs των προβλεπόμενων τιμών μια ελαφρώς τάση μείωσης όσο το h αυξάνεται.

Ελάχιστες διαφορές παρατηρούνται ξανά στην κατανομή της εκ-των-υστέρων μέσης τιμής της παραμέτρου τ, αφού παρατηρείται μια ελάφρώς μειωμένη τάση καθώς αυξάνουν οι τιμές της παραμέτρου h. Είναι ιδιαίτερα σημαντικό να αναφερθεί ότι απο τα Διαγράμματα 4.4 και 4.5 ότι οι εκ-των-υστέρων επικρατούσες τιμές των επιδράσεων β_j^* είναι εύρωστες για τις τιμές των υπερπαραμέτρων h και u που δοκιμάστηκαν.

Τέλος, θα κλείσουμε αυτήν την ενότητα με μία σύντομη σύγκριση των αποτελεσμάτων με αυτά που προέκυψαν απο την $IG \sim (3,18)$ υπερ εκ-των-προτέρων κατανομή. Με βάση αυτήν την εκ-των-προτέρων κατανομή για την πρώτη ανάλυση μέσω του Διαγράμματος 4.4, τα RMSEs είναι παρόμοιας κλίμακας με τα υπόλοιπα, οι εκ-των υστέρων πιθανότητες εισαγωγής των σημαντικών μεταβλητών των τιμών αναφοράς έναντι των τιμών της χαμηλής εκ-των προτέρων πληροφορίας μικραίνουν με ποσοστιαία μείωση 1.6%-8.4% (οι μέσες εκ-των-υστέρων πιθανότητες εισαγωγής έχουν εύρος 0.19-0.91 έναντι 0.13-0.86 , ενώ των μη σημαντικών μεταβλητών μικραίνουν με κατεύθυνση προς το μηδέν με ποσοστιαία μείωση 0.9%-3.6% (οι μέσες εκ-των-υστέρων πιθανότητες εισαγωγής έχουν εύρος 0.02-0.08 έναντι 0.005-0.05.

Επιπλέον, η εκ-των υστέρων μέση τιμή της παραμέτρου τ για την χαμηλή πληροφορίας εκ-των-προτέρων κατανομή φαίνεται να διαφέρει σημαντικά απο τις υπόλοιπες μέσες εκ-των-υστέρων μέσες τιμές της παραμέτρου τ.

4.4 Σύγκριση με τις μεθόδους που εφαρμόζονται μέσω του πακέτου **BAS**

Σε αυτήν την ενότητα, θα συγκρίνουμε τα αποτελέσματα των μη τοπικών-εκ-των-προτέρων κατανομών σε σχέση με τις τοπικές εκ-των-προτέρων κατανομές που χρησιμοποιούνται μέσω του πακέτου BAS και παρουσιάστηκαν στο κεφάλαιο 2. Όλες οι μέθοδοι παρουσιάζονται για το σχήμα προσομοιωμένων δεδομένων της Ενότητας 4.1.

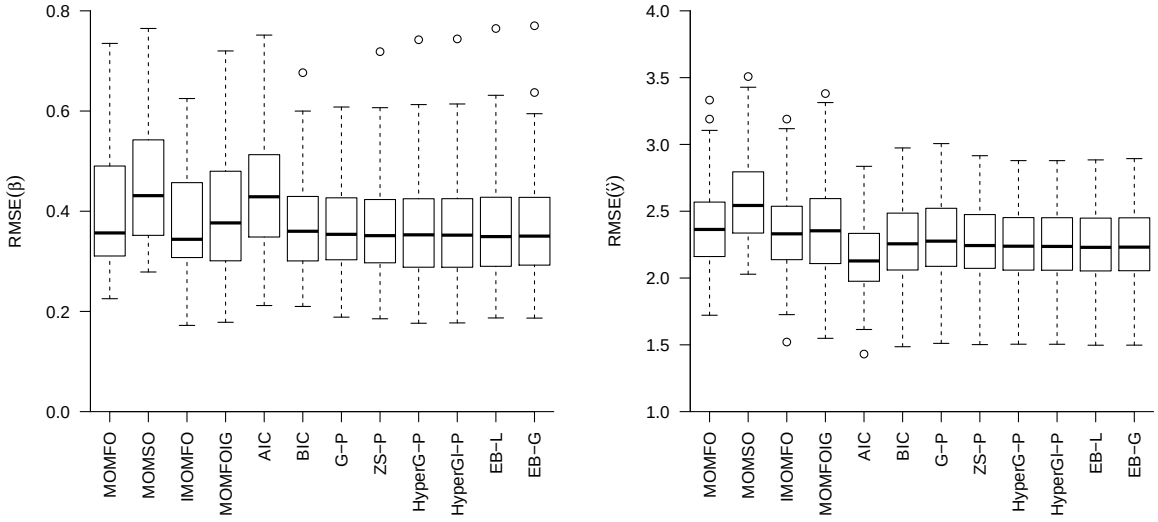
Σε όλα τα Διαγράμματα, παρουσιάζουμε αποτελέσματα με βάση το γινόμενο ροπών πρώτης τάξης MOMFO, το γινόμενο ροπών δευτέρης τάξης MOMSO, το γινόμενο αντίστροφων ροπών πρώτης τάξης IMOMFO και για το γινόμενο ροπών πρώτης τάξης με υπερ εκ-των-προτέρων κατανομή $IG \sim (3, 10)$ για την παράμετρο τ MOMFOIG σε σχέση με το κριτήριο πληροφορίας του Akaike (AIC), Μπεϋζιανό κριτήριο πληροφορίας (BIC), την g εκ-των-προτέρων κατανομή (G-p), την Zellner-Siow εκ-των-προτέρων κατανομή (ZS-p), την υπερ g εκ-των-προτέρων κατανομή (HyperG-p), την υπερ g εκ-των-προτέρων κατανομή με προσέγγιση του Laplace (HyperGl-p), τις εμπειρικές τοπικές Μπεϋζιανές μεθόδους (EB-L) και τις τις εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους (EB-G).

Για κάθε σύγκριση, χρησιμοποιήσαμε 100 σέτ προσομοιωμένα δεδομένα. Όλες οι μέθοδοι συγκρίνονται χρησιμοποιώντας τις ποσότητες RMSEs των προβλεπόμενων τιμών $\hat{y}$ και των εκτιμώμενων επιδράσεων $\hat{\beta}_j$ όπως περιγράφηκε στην ενότητα 4.1.3. Για όλες τις μεθόδους μη τοπικών κατανομών χρησιμοποιήσαμε τις εκ-των-υστέρων επικρατούσες τιμές για τις εκτιμώμενες επιδράσεις $\hat{\beta}_j$, ενω για τις μεθόδους τοπικών κατανομών χρησιμοποιήσαμε τις εκ-των-υστέρων διαμέσους. Όλες οι προβλεπόμενες τιμές υπολογίστηκαν όπως στην Ενότητα 4.3. Όλες οι λεπτομέρειες συνοψίζονται στον Πίνακα 4.2 και χρησιμοποιουνται ως αναφορά σε όλες τα διαγράμματα και τους πίνακες αυτής της ενότητας.

Πίνακας 4.2: Συντομογραφίες και λεπτομέρειες για τις μεθόδους

	Συντομογραφία	Μέθοδος
1	MOMFO	Γινόμενο ροπών πρώτης τάξης για $r = 1$, $\tau = 0.348$
2	MOMSO	Γινόμενο ροπών δευτέρης τάξης για $r = 2$, $\tau = 0.072$
3	IMOMFO	Γινόμενο αντίστροφων ροπών πρώτης τάξης για $\tau = 0.113$
4	MOMFOIG	Γινόμενο ροπών πρώτης τάξης με εκ-των-προτέρων κατανομή $IG \sim (3, 10)$ για $r = 1$
5	AIC	Κριτήριο πληροφορίας Akaike
6	BIC	Μπεϋζιανό κριτήριο πληροφορίας
7	G-P	g εκ-των προτέρων κατανομή για $g = 50$
8	ZS-P	Zellner-Siow εκ-των προτέρων κατανομή
9	HyperG-P	Υπερ g εκ-των-προτέρων κατανομή για $g = 3$
10	HyperGl-P	Υπερ g εκ-των-προτέρων κατανομή με μέθοδο του Laplace για $g = 3$
11	EB-L	Εμπειρικές τοπικές Μπεϋζιανές μεθόδους
12	EB-G	Εμπειρικές σφαιρικές Μπεϋζιανές μεθόδους

Η κατανομή των RMSEs των $\hat{y}_i$ και των $\hat{\beta}_j$ για 100 προσομοιωμένα σετ δεδομένων με βάση το σχήμα προσομοιώσης της Ενότητας 4.1 παρουσιάζονται στο Διάγραμμα 4.6 για όλες τις μεθόδους του Πίνακα 4.2. Τα πρώτα 4 διαγράμματα πλαισίου-απολήξεων αφορούν την επιλογή μεταβλητών με βάση τις μη τοπικές εκ-των-προτέρων κατανομές και τα υπόλοιπα την επιλογή μεταβλητών με βάση τις τοπικές κατανομές.



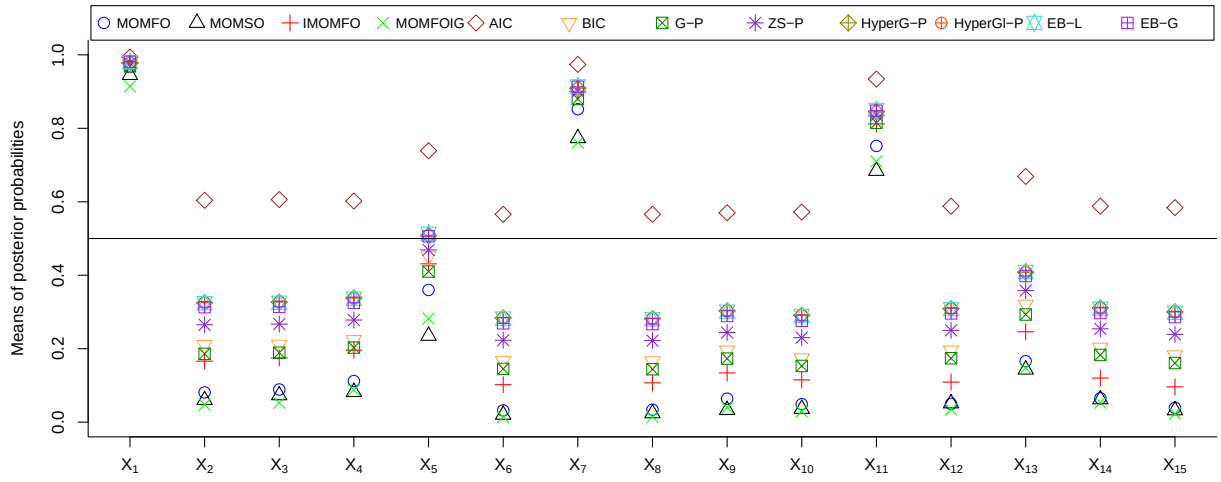
(α′) RMEs των εκτιμώμενων επιδράσεων β_j (β′) RMEs των προβλεπόμενων τιμών y_i

Όλες οι συντομογραφίες και λεπτομέρειες των μεθόδων συνοψίζονται στον πίνακα 4.2, οι ποσότητες $\boldsymbol{RMSEs}$ υπολογίστηκαν σύμφωνα με τις εξισώσεις 4.1 και 4.2

Διάγραμμα 4.6: Διαγράμματα πλαισίου-απολήξεων των ποσοτήτων $\boldsymbol{RMSEs}$ των προβλεπόμενων τιμών $\hat{y}_i$ και των εκτιμώμενων επιδράσεων $\hat{\beta}_j$ για 100 προσομοιωμένα σετ δεδομένων για τις διαφορετικές μεθόδους επιλογής μεταβλητών του Πίνακα 4.2

Οι κατανομές των ποσοτήτων RMSEs φαίνονται να είναι παρόμοιας κλίμακας με πολύ μικρές διαφορές μεταξύ τους λαμβάνοντας υπόψην την συνολική μεταβλητότητα.
Παρόλα αυτά, η μέθοδος AIC φαίνεται να είναι η χειρότερη απο τις μεθόδους τοπικών κατανομών σε όρους RMSEs των $\hat{\beta}_j$, ενώ παρατηρούμε για τις μεθόδους μη τοπικών κατανομών ότι η χειρότερη μέθοδος είναι η MOMSO για τα ίδια RMSEs.
Επιπλέον, αυτο συμβαίνει γιατί η μέθοδος AIC τείνει να επιλέξει πολυπλοκότερα μοντέλα (ενδεχομένως να συμπεριλάβει επιδράσεις μεταβλητών ως μη μηδενικές που στην πραγματικότητα είναι μηδενικές), ενώ οι εκ-των-προτέρων κατανομές γινομένου ροπών δεύτερης τάξης έχουν ουρές λεπτότερες δίνοντας μικρότερη μάζα πιθανότητας σε τιμές των επιδράσεων $\beta_j > 1.2$, έτσι δεν μπορούν να εντοπίσουν τις πραγματικές επιδράσεις επηρεασμένες απο την χαμηλή πληροφο-

ρία του μικρού μεγέθους δείγματος n και απο την τιμή της παραμέτρου $\tau = 0.072$.
Ακόμα, οι εκ-των-προτέρων κατανομές γινομένου ροπών πρώτης τάξης χρησιμοποιώντας για την άγνωστη παράμετρο τ υπερ εκ-των-προτέρων κατανομή $IG \sim (3, 18)$ επηρεάζεται αποκλειστικά απο τον καθορισμό των υπερ παραμέτρων. Έτσι, υποεκτιμά τις πραγματικές επιδράσεις β_j ενεργοποιώντας το παράδοξο Bartlett και Lindley (μηδενίζονται τυχον σημαντικές επιδράσεις ανεξάρτητων μεταβλητών), γιάυτό είναι αναμενόμενο να προκύπτουν υψηλότερα RMSEs των $\hat{\beta}_j$ σε σχέση με τις μεθόδους MOMFO, IMOMFO.
Επιπροσθέτως, οι μέθοδοι (BIC), (G-p), (Zellner-Siow), (ZS-p), (HyperG-p), (HyperGl-p), (EB-L), (EB-G) έχουν παρόμοια RMSEs των $\hat{\beta}_j$
Επιπλέον, παρατηρούμε για τις μεθόδους μη τοπικών κατανομών ότι η πρώτη καλύτερη είναι η IMOMFO και η δεύτερη καλύτερη η MOMFO για τα ίδια RMSEs. Αυτό συμβαίνει γιατί οι εκ-των-προτέρων κατανομές γινομένου αντίστροφων ροπών πρώτης τάξης έχουν παχύτερες ουρές και δίνουν μικρότερη μάζα πιθανότητας σε τιμές των επιδράσεων $\beta_j > 2.5$, έτσι εκτείθονται σε μικρότερες μεροληψίες σε σχέση με τις υπόλοιπες μεθόδους μη τοπικών εκ-των-προτέρων κατανομών και μπορούν να εντοπίσουν τις πραγματικές επιδράσεις των ανεξάρτητων μεταβλητών ακόμα και σε μικρό μέγεθος δείγματος $n = 50$.
Επιπλέον, παρόλο που οι εκ-των-προτέρων κατανομές γινομένου ροπών πρώτης τάξης συγκεντρώνουν μικρότερη μάζα πιθανότητας γύρω απο τις επικρατούσες τιμές β_j σε σχέση με τις υψηλότερες τάξεις γινομένου ροπών και δίνουν μικρότερη μάζα πιθανότητας σε τιμές των επιδράσεων $\beta_j > 2$, τείνουν να εντοπίσουν τις πραγματικές επιδράσεις β_j καθώς έχουν καλύτερη συμπεριφορά στο μέγεθος του δείγματος $n = 50$ (ακόμα και σε μικρά μεγέθη δείγματος εντοπίζουν τις πραγματικές επιδράσεις). Ακόμα, σε όρους RMSEs των $\hat{y}_i$ η μέθοδος AIC είναι η καλύτερη, ενώ η μέθοδος MOMFOIG είναι η χειρότερη.



Όλες οι συντομογραφίες και λεπτομέρειες των μεθόδων συνοψίζονται στον πίνακα 4.2, οι ποσότητες $\boldsymbol{RMSEs}$ υπολογίστηκαν σύμφωνα με τις εξισώσεις 4.1 και 4.2

Διάγραμμα 4.7: Διάγραμμα των μέσων εκ-των υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ για 100 προσομοιωμένα δεδομένα για τις διαφορετικές μεθόδους επιλογής μεταβλητών για τις μεθόδους του Πίνακα 4.2

Παρόλο που η μέθοδος AIC δίνει καλύτερα αποτελέσματα σε όρους προβλεψιμότητας RMSEs των $\hat{y}_i$ λόγω του ορισμού της, το ενδιαφέρον επικεντρώνεται στις μεθόδους IMOMFO, MOMFO, MOMFOIG, ενώ χειρότερη είναι και πάλι η MOMSO. Ακόμα, οι μέθοδοι (BIC), (G-p), (Zellner-Siow), (ZS-p), (HyperG-p), (HyperGl-p), (EB-L), (EB-G) δίνουν και πάλι παρόμοια στα ίδια RMSEs. Το Διάγραμμα 4.7 απεικονίζει τις μέσες τιμές των εκ-των-υστέρων πιθανοτήτων εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ κάθε ανεξάρτητης μεταβλητής.
Συγκεκριμένα, παρατηρούμε ότι οι μεταβλητές X_1, X_7, X_{11} είναι σημαντικές για όλες τις μεθόδους επιλογής μεταβλητών (με εξαίρεση την μέθοδο MOMFOIG της οποιας αντιστοιχούν σε μικρότερες τιμές). Οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ αυτών των ανεξάρτητων μεταβλητών είναι υψηλότερες για την μέθοδο AIC, και χαμηλότερες για την μέθοδο MOMFOIG. Οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ για την μέθοδο AIC είναι συστηματικά πολυ υψηλότερες σε σχέση με τις υπόλοιπες μεθόδους, με μοναδική εξαίρεση την X_1 που είναι η σημαντικότερη για όλες τις μεθόδους. Η σημαντική διαφορά αυτής της μεθόδου απο τις υπόλοιπες μεθόδους είναι οτι φαίνεται να αυξάνεται η σημαντικότητα της κάθε μεταβλητής (οι μη σημαντικές μεταβλητές έχουν μέση εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j = 1|y) > 0.5$) και συγκεκριμένα μεγιστοποιείται η μέση εκ-των-υστέρων πιθανότητα εισαγωγής για τις τρείς μεταβλητές X_1, X_7 και X_{11} που είναι σημαντικές απ΄όλες τις μεθόδους.

Ακόμα, για τις μη σημαντικές μεταβλητές με χρήση τοπικών εκ-των-προτέρων κατανομών, όλες οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ είναι κοντα στο 0.5, το οποίο αυξάνει την αβεβαιότητα του μοντέλου και ωθεί την κάθε μεθόδο να συμπεριλάβει μη σημαντικά πολύπλοκα μοντέλα (δηλαδή επιλέγει ως σημαντικές μεταβλητές που έχουν στην πραγματικότητα μηδενικές επιδράσεις).
Αντίθετα, για τις μεθόδους MOMFO, MOMSO, MOMFOIG των μη τοπικών εκ-των-προτέρων κατανομών, όλες οι εκ-των-υστέρων πιθανότητες εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ είναι κοντα στο μηδέν, το οποίο μειώνει την αβεβαιότητα του μοντέλου και ωθεί την κάθε μεθόδο να συμπεριλάβει απλούστερα μοντέλα (δηλαδή όλες οι μη σημαντικές επιδράσεις συρρικνώνονται στο μηδέν). Συνοψίζοντας, οι μη τοπικές μέθοδοι IMOMFO, MOMFO, με εξάιρεση τις MOMFO MOMFOIG παρουσιάζουν ικανοποιητικά RMSEs σε σχέση με τις υπόλοιπες μεθόδους. Όλες οι μέθοδοι των μη τοπικών κατανομών επιλέγουν απλούστερα μοντέλα θέτοντας την εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ για τις μη σημαντικές μεταβλητές κοντά στο μηδέν, ενώ για τις σημαντικές μεταβλητές υψηλές παρομοιώς για όλες τις μεθόδους, ενώ οι μέθοδοι των τοπικών κατανομών επιλέγουν πολύπλοκα μοντέλα θέτοντας την εκ-των-υστέρων πιθανότητα εισαγωγής $\widehat{f}(\gamma_j = 1|y)$ για τις μη σημαντικές μεταβλητές κοντά στο 0.5, ενώ για τις σημαντικές μεταβλητές υψηλές παρομοιώς για όλες τις μεθόδους.

4.5 Συμπεράσματα

Σε αυτό το κεφάλαιο προσπαθήσαμε να δούμε την συμπεριφορά των μη-τοπικών εκ-των-προτέρων κατανομών σε σχέση με την παράμετρο r και την παράμετρο τ.
Συγκεκριμένα, παρατηρήσαμε ότι για διαφορετικές τιμές της παραμέτρου r καθώς η τ αυξάνει οι μη σημαντικές μεταβλητές συρρικνώνονται πιο γρήγορα στο μηδέν. Ακόμα, καθώς $r > 2$ δεν βλέπουμε καμιά οιδοποιώς διαφορά. Αυτό συμβαίνει γιατί είδαμε ότι οι εκ-των-προτέρων κατανομές είναι μη πληροφοριακές λόγω της μάζας πιθανότητας που προκύπτει απο τις λεπτές ουρές, άρα καθώς αυξάνεται η παράμετρος r οι κατανομές γίνονται περισσότερο μη πληροφοριακές.
Επιπλέον, είδαμε ότι ο καθορισμός της παραμέτρου τ είναι ιδιαίτερα καθοριστικός αφου συρρικνώνει τις μη σημαντικές μεταβλητές προς το μηδέν όσο αυξάνεται επηρεάζοντας ελάχιστα τις σημαντικές.
Επιπλέον, η χρήση υπερ εκ-των-προτέρων $IG(h,u)$ κατανομής για την άγνωστη παράμετρο τ είναι απαραίτητη γιατί στην πραγματικότητα δεν θα γνωρίζουμε ποτε την πραγματική τιμή της παραμέτρου τ που διαχωρίζει ενα μηδενικό μοντέλο απο ενα εναλλακτικό μοντέλο, οπότε είναι αντικειμενική η χρήση της υπερ εκ-των-προτέρων κατανομής.
Τα αποτελέσματα βασισμένα στην ανάλυση ευαισθησίας για καθορισμένες τιμές των παραμέτρων

h, u παρουσιάζουν ελάχιστες διαφορές σε όρους RMSEs.
Μόνο τα αποτελέσματα για την χαμηλή πληροφορίας $IG(3, 18)$ εκ-των-προτέρων κατανομή φαίνεται να διαφέρουν σημαντικά απο τα υπόλοιπα το οποίο είναι και λογικό.
Όσον αφορά την σύγκριση των μη τοπικών κατανομών σε σχέση με τις τοπικές κατανομές, τα αποτελέσματα διαφέρουν ελάχιστα σε όρους RMSEs, σημειώνεται ότι οι μέθοδοι μη τοπικών κατανομών MOMFO, IMOMFO έχουν τα καλύτερα RMSEs.
Επιπροσθέτως, οι μη τοπικές εκ-των-προτέρων κατανομές δίνουν μηδενική εκ-των-υστέρων πιθανότητα εισαγωγής στις μη σημαντικές μεταβλητές, ενώ οι τοπικές δίνουν μεγαλύτερη εκ-των-υστέρων πιθανότητα εισαγωγής απο αυτήν που αντιστοιχεί στην πραγματικότητα. Έτσι, οι μη τοπικές μέθοδοι υποστηρίζουν απλούστερα μοντέλα σε σχέση με τις τοπικές μεθόδους που υποστηρίζουν πολυπλοκότερα μοντέλα.

Κεφάλαιο 5

Συζήτηση-περαιτέρω διερεύνηση

Σε αυτήν την διπλωματική εργασία αναπτύχθηκε η Μπεϋζιανή προσέγγιση στην επιλογή μοντέλου. Στην Μπεϋζιανή συμπερασματολογία η σύγκριση μοντέλων επιτυγχάνεται μέσω του λόγου συμπληρωματικών πιθανοτήτων της εκ-των-υστέρων και της εκ-των-προτέρων πληροφορίας που καλείται παράγοντας Bayes. Οι παράγοντες Bayes υπολογίζονται έτσι ώστε να συγκρίνουν την αντίστοιχη μηδενική υπόθεση με την αντίστοιχη εναλλακτική υπόθεση και η ερμηνεία των αποτελεσμάτων να συμβαδίζει με αυτήν των Kass και Raftery. Όπως είδαμε, η Μπεϋζιανή στατιστική ξεπερνάει τα μειονεκτήματα της Κλασσικής στατιστικής και έτσι είναι μια αξιόπιστη μεθοδολογία.
Παρόλα αυτά, η Μπεϋζιανή μεθοδολογία υστερεί στα υπολογιστικά προβλήματα και γιάυτόν τον λόγο δεν είναι ευρέως διαδεδομένη. Το σημαντικότερο πρόβλημα είναι η πολυπλοκότητα στον υπολογισμό των εκ-των-υστέρων ποσοτήτων και της αντίστοιχης περιθώριας πιθανοφάνειας.
Ιδιαίτερη σημασία δίνεται σε μοντέλα πιθανοφάνειας που ανήκουν στην εκθετική οικογένεια (συζυγείς κατανομές) τα οποία παρέχουν σημαντικές ιδιότητες και δίνουν παράγοντες Bayes και περιθώρειες πιθανοφάνειες σε κλειστή μορφή.
Επιπλέον, οι εκ-των-προτέρων κατανομές των υπο-εξέταση μοντέλων είναι ευαίσθητες σε μεγάλες τιμές των εκ-των-προτέρων υπερ παραμέτρων που καθορίζουν την διακύμανση και σε μεγάλες τιμές του δείγματος υποστηρίζοντας το απλούστερο μοντέλο Bartlett και Lindley (1957). Ακόμα, όσον αναφορά την επιλογή μεταβλητών στην περίπτωση που ο παραμετρικός χώρος των μοντέλων είναι μεγάλος, εφαρμόζουμε τις βασικές ιδέες επιλογής μεταβλητών που προτείνονται μέσω των George και McCulloch (1993, 1997) μέσω δεικτριών μεταβλητών που καθορίζουν ανάλογα την εισαγωγή ανεξάρτητης μεταβλητής ή όχι. Συγκεκριμένα, οι μέθοδοι επιλογής μεταβλητών με χρήση MCMC όπως η στοχαστική διερεύνηση, ο δείγματολήπτης των Kuo και Mallick κα ο δειγματολήπτης Gibbs βασίζονται σε διαφορετικές εκ-των-προτέρων κατανομές οι οποίες τελικά καταλήγουν σε παρόμοια συμπεράσματα με την στοχαστική διερεύνηση να διαφέρει.

Οι αλγόριθμοι εφαρμόστηκαν στο παράδειγμα του Κεφαλαίου 2 για τα δεδομένα της παχυσαρκίας μέσω του πακέτου WinBugs και είδαμε ότι οι μέθοδοι των Kuo και Mallick και Gibbs επιλογής μεταβλητών καταλήγουν στα ίδια συμπεράσματα όσον αναφορά, τις εκ-των-υστέρων πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών και το πιθανότερο εκ-των-υστέρων μοντέλο, ενώ η στοχαστική η διερεύνηση διέφερε.

Επιπλέον, η πλήρης απαρίθμηση του χώρου των μοντέλων με την χρήση του πακέτου BAS της R δίνει την δυνατότητα επιλογής μεταβλητών για διάφορες εκ-των-προτέρων κατανομές. Συγκεκριμένα, η εφαρμογή των g εκ-των-προτέρων κατανομών για προκαθορισμένες τιμές της υπερ παραμέτρου g σε σχέση με τις μείξεις g εκ-των-προτέρων κατανομών όπου η g χρησιμοποιείται ως τυχαία μεταβλητή υστερούν σε σημαντικό βαθμό καθώς οι προκαθορισμένες g εκ-των-προτέρων κατανομές υποστηρίζουν απλούστερα μοντέλα για μεγάλες τιμές της υπερ παραμέτρου g. Τα αποτελέσματα για τα δεδομένα της παχυσαρκίας έδειξαν ότι οι υπερ g εκ-των-προτέρων κατανομές, Zellner-Siow και οι εμπειρικές Μπεϋζιανές μέθοδοι (τοπικές και σφαιρικές) είναι παρόμοιες στην περίπτωση που ο μηχανισμός που γεννάει τα δεδομένα είναι το μηδενικό μοντέλο.

Παρόλα αυτά, αναζητάμε εκ-των-προτέρων κατανομές που να ορίζονται υπο-την εναλλακτική υπόθεση έτσι ώστε να αντανακλούν μια εντελώς ξεχωριστή θεωρία απο την αντίστοιχη μηδενική υπόθεση. Επιπλέον, παραδοσιακές εκ-των-προτέρων εναλλακτικές κατανομές όπως οι συζυγείς, οι δυναμικές και οι ενδογενείς ορίζονται σε παραμετρικούς χώρους που συσχετίζονται με την μηδενική. Έτσι, η μάζα πιθανότητας γύρω απο τιμές της εναλλακτικής υπόθεσης σπαταλάται και δίνεται μεγαλύτερο βάρος σε τιμές που σχετίζονται με την μηδενική υπόθεση. Το σημαντικότερο πρόβλημα είναι ότι εισάγεται μεροληψία η οποία σχετίζεται με το παράδοξο Bartlett και Lindley με αποτέλεσμα, ο ρυθμός συσσώρευσης για μια αληθινή εναλλακτική υπόθεση να αυξάνεται με γρηγορότερο ρυθμό σε σχέση με αυτόν της αληθινής μηδενικής υπόθεσης.

Για αυτόν τον λόγο ορίζονται οι μη τοπικές εναλλακτικές εκ-των-προτέρων κατανομές οι οποίες καταφέρνουν να μηδενίσουν την περιοχή που σχετίζεται με την μηδενική υπόθεση και να δόσουν μεγαλύτερη μάζα πιθανότητας σε μια γειτονική περιοχή της μηδενικής. Τέτοιες κατανομές είναι οι κατανομές ροπών και οι κατανομές αντίστροφων ροπών. Τα χαρακτηριστικά που διέπουν αυτές τις κατανομές είναι η τάξη r και η παράμετρος τ. Οι Κατανομές ροπών και αντίστροφων ροπών προκύπτουν και από το γινομένο επιμέρους πυκνοτήτων πιθανότητας.

Η εφαρμογή αυτών των κατανομών σίγουρα δίνει μια ευέλικτη προσέγγιση καθώς στην πραγματικότητα όσον αναφορά την επιλογή μεταβλητών θα εντοπίζει τις πραγματικές επιδράσεις και θα συρρικνώνει τις μηδενικές επιδράσεις στο μηδέν.

Σε όρους προβλεψιμότητας, σίγουρα αναζητάμε ενα απλούστερο μοντέλο το οποίο να είναι εύκολο στην ερμηνεία των παραμέτρων και να κάνει την πρόβλεψη βέλτιστη.

Η εφαρμογή των μεθόδων όσον αναφορά την προσομοιωμένη μελέτη για διάφορες τιμές της παραμέτρου r και τ έδειξε ότι όσο αυξάνεται η παράμετρος τ οι επιδράσεις των μη σημαντικών

μεταβλητών συρρικνώνονται προς το μηδέν ενώ των σημαντικών παραμένουν στα ίδια επίπεδα. Η αύξηση της παραμέτρου r δεν φάινεται να έχει καμια επιρροή στις εκ-των-υστέρων πιθανότητες όσον αφορά στις κατανομές γινομένου ροπών και κατανομές γινομένου αντίστοφων ροπών.
Στην ανάλυση ευαισθησίας, χρησιμοποιήσαμε υπερ εκ-των-προτέρων κατανομή $IG(h,u)$ για την παράμετρο τ και κάναμε δύο αναλύσεις επιπλέον για διάφορες τιμές της h και u. Οι τιμές αναφοράς ηταν οι $h = 3$ και $u = 10$.
Τα αποτελέσματα σε όρους RMSEs ήταν παρόμοια με εξαίρεση αυτά της χαμηλής εκ-των-προτέρων κατανομής των οποίων ο καθορισμός των υπερ παραμέτρων φάινεται να επηρεάζει σημαντικά. Οι εκ-των-υστέρων πιθανότητες εισαγωγής των ανεξάρτητων μεταβλητών τείνουν να μειώνονται όσο η υπερ παράμετρος u αυξάνει.
Συγκεκριμένα, οι εκ-των-υστέρων πιθανότητες των μη σημαντικών μεταβλητών συρρικνώνονται στο μηδέν, ενώ των σημαντικών μεταβλητών μειώνονται ελαφρώς.
Τέλος, η σύγκριση μη τοπικών κατανομών και τοπικών κατανομών παρέχει παρόμοια αποτελέσματα σε όρους RMSEs. Οι κατανομές γινομένου ροπών πρώτης τάξης και οι κατανομές γινομένου αντίστροφων ροπών πρώτης τάξης δίνουν τα καλύτερα RMSEs γιατι οι πρώτες έχουν πολύ καλή συμπεριφορά στο μέγεθος του δείγματος και η δεύτερη έχει παχιές ουρές και άρα εντοπίζει τις πραγματικές επιδράσεις συρρικνώνοντας στο μηδέν αυτές που είναι μηδέν. Το σημαντικότερο ,όμως, είναι ότι η αβεβαιότητα του μοντέλου αυξάνει ως προς το ποιές ανεξάρτητες μεταβλητές θα πρέπει να συμπεριλάβουμε στο μοντέλο με βάση την επιλογή μεταβλητών που βασίζεται στις τοπικές κατανομές, ενώ η αβεβαιότητα του μοντέλου για τις μη τοπικές κατανομές μικραίνει καθώς οι πραγματικές μεταβλητές συρρικνώνονται στο μηδέν.
Περαιτέρω διερεύνηση χρήζουν οι ιδιότητες της υπερ εκ-των-προτέρων κατανομής $IG(h,u)$ με τον τρόπο που επηρεάζει την συνολική προβλεψιμότητα. Επιπλέον, περαιτέρω διερεύνηση χρήζει η μέθοδος επιλογής μεταβλητών που βασίζονται στον συνδιασμό εκ-των-προτέρων τοπικών και μη τοπικών κατανομών, δηλαδή υβριδικών κατανομών Consonni, Forster και La Rocca (2013).

Βιβλιογραφία

[1] Abramowitz, Milton and Stegun, Irene A and others. Handbook of mathematical functions. *Applied Mathematics Series*, 55:62, 1966.

[2] Altomare, D. and Consonni, G. and La Rocca, L. Objective Bayesian Search of Gaussian Directed Acyclic Graphical Models for Ordered Variables with Non-Local Priors. *Biometrics*, 69(2):478–487, 2013.

[3] Bahadur, R. R. and Bickel, P. J. Asymptotic optimality of Bayes' test statistics. *TechnicalReport. University of Chicago, Chicago. Unpublished*, 1967.

[4] Barbieri, M. M. and Berger, J. O. Optimal predictive model selection. *Annals of statistics*, 32(math. ST/0406464):870–897, 2004.

[5] Bartlett, M. S. Comment on 'A statistical paradox' by DV Lindley. *Biometrika*, 44(1-2):533–534, 1957.

[6] Berger, J. O. and Mortera, J. Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, 94(446):542–554, 1999.

[7] Berger, J. O. and Pericchi, L. R. The intrinsic Bayes factor for linear models. *Bayesian statistics*, 5:25–44, 1996.

[8] Berger, J. O. and Pericchi, L. R. Accurate and Stable Bayesian Model Selection: The Median Intrinsic Bayes Factor. 1998.

[9] Bolstad, W. M. *Introduction to Bayesian statistics*. John Wiley & Sons, 2007.

[10] Bové, D. S. and Held, L. and others. Hyper-g priors for generalized linear models. *Bayesian Analysis*, 6(3):387–410, 2011.

[11] Bruijn, N. G. Asymptotic methods in analysis. *Bibliotheca mathematica* , (4), 1958.

[12] Carlin, B. P. and Chib, S. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(3):473–484, 1995.

[13] Carlin, B. P. and Louis, T. A. Bayes and empirical Bayes methods for data analysis. *Statistics and Computing*, 7(2):153–154, 1997.

[14] Clyde, M. A. and Ghosh, J. and Littman, M. L. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, 20(1), 2011.

[15] Cohen, J. A power primer. *Psychological bulletin*, 112(1):155, 1992.

[16] Congdon, P. Bayesian Statistical Modelling (Wiley Series in Probability and Statistics-Applied Probability and Statistics Section). 2001.

[17] Conigliani, C. and O'Hagan, A. Sensitivity of the fractional Bayes factor to prior distributions. *Canadian Journal of Statistics*, 28(2):343–352, 2000.

[18] Consonni, G. and Forster, J. J. and La Rocca, L. and others. The whetstone and the alum block: Balanced objective Bayesian comparison of nested models for discrete data. *Statistical Science*, 28(3):398–423, 2013.

[19] Consonni, G. and La Rocca, L. Tests based on intrinsic priors for the equality of two correlated proportions. *Journal of the American Statistical Association*, 103(483):1260–1269, 2008.

[20] Cui, W. and George, E. I. Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138(4):888–900, 2008.

[21] De Santis, F. and Spezzaferri, F. Consistent fractional Bayes factor for nested normal linear models. *Journal of statistical planning and inference*, 97(2):305–321, 2001.

[22] Dellaportas, P. and Forster, J. J. and Ntzoufras, I. On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1):27–36, 2002.

[23] Efron, B. and Hastie, T. and Johnstone, I. and Tibshirani, R. and others. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[24] Efroymson, M. A. Multiple regression analysis. *Mathematical methods for digital computers*, 1:191–203, 1960.

[25] Fan, J. and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[26] Fernandez, C. and Ley, E. and Steel, Mark F. J. Model uncertainty in cross-country growth regressions. *Journal of applied Econometrics*, 16(5):563–576, 2001.

[27] Fink, D. A compendium of conjugate priors. *Environmental Statistics Group*, 1997.

[28] Foster, D. P. and George, E. I. The risk inflation criterion for multiple regression. *Annals of statistics*, 22(4):1947–1975, 1994.

[29] Frühwirth-Schnatter, S. and Tüchler, R. Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Statistics and Computing*, 18(1):1–13, 2008.

[30] Gelman, A. and Carlin, J. B. and Stern, H. S. and Dunson, D. B. and Vehtari, A. and Rubin, D. B. *Bayesian data analysis*. CRC press, 2013.

[31] George, E. and Foster, D. P. Calibration and empirical Bayes variable selection. *Biometrika*, 87(4):731–747, 2000.

[32] George, E. I. and McCulloch, R. E. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

[33] George, E. I. and McCulloch, R. E. Approaches for Bayesian variable selection. *Statistica sinica*, 7(2):339–373, 1997.

[34] Green, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[35] Ibrahim, J. G. and Chen, M. H. and others. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, 2000.

[36] Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

[37] Jeffreys, H. *The theory of probability*. OUP Oxford, 1961.

[38] Johnson, V. E. and Rossell, D. On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):143–170, 2010.

[39] Johnson, V. E. and Rossell, D. Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660, 2012.

[40] Kan, R. moments of sum to moments of product. *Journal of Multivariate Analysis*, 99(3):542–554, 2008.

[41] Kass, R. E. and Raftery, A. E. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[42] Kass, R. E. and Wasserman, L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the american statistical association*, 90(431):928–934, 1995.

[43] Kuo, L. and Mallick, B. Variable Selection for Regression Models. *Sankhyā: The Indian Journal of Statistics, Series B*, 1998.

[44] Lahiri, P. Model Selection, Institute of Mathematical Statistics. *Lecture Notes-Monograph Series*, 38, 2001.

[45] Ley, E. and Steel, M. F. J. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. 2007.

[46] Ley, E. and Steel, M. F. J. Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics*, 171(2):251–266, 2012.

[47] Liang, F. and Paulo, R. and Molina, G. and Clyde, M. A. and Berger, J. O. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 2008.

[48] Lindley, D. V. A STATISTICAL PARADOX. *Biometrika*, 44:187, 1957.

[49] Mallows, C. L. Some comments on C p. *Technometrics*, 15(4):661–675, 1973.

[50] Mitchell, T. J. and Beauchamp, J. J. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

[51] Morris, C. N. and others. Natural Exponential Families with Quadratic Variance Functions: Statistical Theory. *The Annals of Statistics*, 11(2):515–529, 1983.

[52] Nadarajah, S. and Dey, D. K. Multitude of multivariate t-distributions. *Statistics*, 39(2):149–181, 2005.

[53] Ntzoufras, I. Gibbs variable selection using BUGS. *Journal of Statistical Software*, 7(7):1–19, 2002.

[54] Ntzoufras, I. Bayesian Modeling Using WinBUGS. *AMC*, 10:12, 2011.

[55] O'Hagan, A. Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):99–138, 1995.

[56] O'Hagan, A. Properties of intrinsic and fractional Bayes factors. *Test*, 6(1):101–118, 1997.

[57] O'Hara, R. B. and Sillanpää, M. J. and others. A review of Bayesian variable selection methods: what, how and which. *Bayesian analysis*, 4(1):85–117, 2009.

[58] Pérez, J. M. and Berger, J. O. Expected-posterior prior distributions for model selection. *Biometrika*, 89(3):491–512, 2002.

[59] Rossell, D. and Telesca, D. Non-Local Priors for High-Dimensional Estimation. *arXiv preprint arXiv:1402.5107*, 2014.

[60] Rousseau, J. Approximating interval hypothesis: p-values and Bayes factors. 2007.

[61] Schwarz, G. and others. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[62] Tierney, L. and Kass, R. E. and Kadane, J. B. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84(407):710–716, 1989.

[63] Verdinelli, I. and Wasserman, L. Bayes factors, nuisance parameters, and imprecise tests. *Bayesian Statistics*, 5:765–771, 1996.

[64] Vlachos, P. K. and Gelfand, A. E. On the calibration of Bayesian model choice criteria. *Journal of statistical planning and inference*, 111(1):223–234, 2003.

[65] Walker, A. On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1969.

[66] Walker, S. and Hjort, N. L. On Bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821, 2001.

[67] Walker, S. G. and others. Modern Bayesian Asymptotics. *Statistical Science*, 19(1):111–117, 2004.

[68] Zellner, A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243, 1986.

[69] Zellner, A. and Siow, A. Posterior odds ratios for selected regression hypotheses. *Trabajos de estadística y de investigación operativa*, 31(1):585–603, 1980.

[70] Zirphile, J. and Avancées, Etudes. The EDITOR, TECHNOMETRICS. *Technometrics*, 17(1):145–145, 1975.