ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Δ.Μ.Δ.Ε.: ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ
ΚΑΤΕΥΘΥΝΣΗ: ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΠΙΘΑΝΟΤΗΤΕΣ

"Μπεϋζιανή επιλογή μοντέλων και μεταβλητών στα γενικευμένα γραμμικά μοντέλα και εφαρμογή του αλγορίθμου MC^3 "

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΙΩΑΝΝΗ ΔΕΔΑΚΗ

Επιβλέπων: ΔΗΜΗΤΡΗΣ ΦΟΥΣΚΑΚΗΣ
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2011

ΕΥΧΑΡΙΣΤΙΕΣ

Θεωρώ υποχρέωσή μου να ευχαριστήσω θερμά τον επιβλέποντά μου, Επίκουρο Καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου, **κ. Δημήτρη Φουσκάκη** για την υπομονή που επέδειξε και την καθοριστική βοήθειά του που τόσο πρόθυμα μου πρόσφερε καθ'όλη τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας.

Επίσης θα ήθελα να ευχαριστήσω τους Καθηγητές του Εθνικού Μετσόβιου Πολυτεχνείου κυρίους **Γιώργο Κοκολάκη** και **Βασίλη Παπανικολάου**, μέλη της τριμελούς εξεταστικής επιτροπής.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ

Πολλές φορές για να ερμηνεύσουμε τη συμπεριφορά μιας μεταβλητής Y (μεταβλητή απόκρισης) έχουμε στη διάθεσή μας ένα μεγάλο πλήθος επεξηγηματικών μεταβλητών $X_1, X_2, \ldots$ χωρίς αυτές να είναι όλες στατιστικά σημαντικές. Εύκολα μπορούμε να καταλάβουμε ότι η θεώρηση παραπάνω μεταβλητών από αυτές που πραγματικά επεξηγούν το φαινόμενο σε μελλοντικές μελέτες ή ο καθορισμός και ο έλεγχος μη σημαντικών ποσοτήτων για την πρόβλεψη της τιμής της μεταβλητής απόκρισης, μπορεί να είναι χρονοβόρος αλλά και οικονομικά ασύμφορος. Γεννιέται λοιπόν το ερώτημα για το αν μπορούμε να εντοπίσουμε το μικρότερο υποσύνολο επεξηγηματικών μεταβλητών, από το αρχικό μας σύνολο, αναπτύσσοντας έτσι ένα μοντέλο το οποίο όμως θα εξακολουθεί να επεξηγεί τη συμπεριφορά της μεταβλητής απόκρισης μας σε ικανοποιητικό βαθμό.

Κατά την επιλογή ενός στατιστικού μοντέλου το οποίο περιγράφει ένα υπό εξέταση φαινόμενο έχουμε να αντιμετωπίσουμε δύο αντικρουόμενα κριτήρια. Το πρώτο κριτήριο αφορά την όσο το δυνατόν καλύτερη πρόβλεψη της μεταβλητής απόκρισης μέσω του μοντέλου ενώ το δεύτερο κριτήριο αφορά την «οικονομία» του μοντέλου. Πιο συγκεκριμένα, όσες περισσότερες επεξηγηματικές μεταβλητές συμπεριλάβουμε στο μοντέλο, τόσο καλύτερη πρόβλεψη της μεταβλητής απόκρισης θα έχουμε. Ταυτόχρονα όμως, ένα μεγάλο πλήθος επεξηγηματικών μεταβλητών κάνει το μοντέλο μη «οικονομικό», με την έννοια ότι καθίσταται πολλές φορές οικονομικά και χρονικά δαπανηρή η συλλογή δεδομένων από ένα μεγάλο πλήθος μεταβλητών. Στα πλαίσια της κλασικής Στατιστικής θεωρίας έχουν αναπτυχθεί κατά καιρούς διάφορες μέθοδοι επιλογής μοντέλου που βασίζονται κατά κύριο λόγο στον συντελεστή προσδιορισμού R^2, στο κριτήριο C_p-Mallows καθώς και στον στατιστικό έλεγχο F-test.

Στην παρούσα διπλωματική εργασία, η ανάπτυξη του προβλήματος επιλογής μοντέλου και μεταβλητών εξετάζεται από τη σκοπιά της Μπεϋζιανής Στατιστικής. Συγκεκριμένα, εξετάζεται η γενική θεωρία επιλογής μοντέλων και μεταβλητών με αναφορά στα δημοφιλή γενικευμένα γραμμικά μοντέλα καθώς επίσης και ο τρόπος με τον οποίο προσεγγίζουμε το παραπάνω πρόβλημα από τη σκοπιά της Μπεϋζιανής θεωρίας. Ο όρος «Μπεϋζιανή» έχει αναφορά στον Thomas Bayes (1702-1761), ο οποίος απέδειξε μια ειδική περίπτωση αυτού που καλείται τώρα το Θεώρημα του Bayes. Ωστόσο, ήταν ο Pierre Simon Laplace (1749-1827) ο οποίος παρουσίασε μια γενική μορφή του Θεωρήματος και το χρησιμοποίησε για την προσέγγιση των προβλημάτων στην ουράνια μηχανική, στην επεξεργασία ιατρικών στοιχείων και στη νομολογία. Στη δεκαετία του 1980 υπήρξε μια δραματική αύξηση στον τομέα της έρευνας και των εφαρμογών των Μπεϋζιανών μεθόδων, γεγονός που ως επί το πλείστον οφείλεται στην ανακάλυψη Markov Chain Monte Carlo τεχνικών, οι οποίες ήραν πολλά από τα υπολογιστικά προβλήματα που παρουσιάζονταν μέχρι τότε κατά την εφαρμογή των μεθόδων αυτών.

Η Στατιστική κατά Bayes βασίζεται σε μία απλή ιδέα: η μόνη ικανοποιητική περιγραφή της αβεβαιότητας μας επιτυγχάνεται μέσω της πιθανότητας. Η Μπεϋζιανή προσέγγιση μας δίνει, μέσω του υπολογισμού πιθανοτήτων, ένα ισχυρό εργαλείο να καταλάβουμε, να χειριστούμε και να ελέγξουμε την αβεβαιότητα. Ο βασικός κανόνας στη Μπεϋζιανή συμπερασματολογία είναι ότι όλες οι άγνωστες ποσότητες θεωρούνται τυχαίες μεταβλητές και πρέπει να περιγράφονται δια μέσου πιθανοτήτων.

Η Στατιστική συμπερασματολογία χρησιμοποιείται για την εξαγωγή συμπερασμάτων από τα δεδομένα που έχει στη διάθεση του ο ερευνητής για τον πληθυσμό. Βασικό εργαλείο όλων των Μπεϋζιανών μεθόδων είναι οι εκ των προτέρων (prior) κατανομές. Οι κατανομές αυτές εκφράζουν τις εκ των προτέρων γνώσεις και πεποιθήσεις του ερευνητή για τις άγνωστες παραμέτρους του μοντέλου και μέσω της Μπεϋζιανής μεθοδολογίας οδηγούν σε εκ των υστέρων (posterior) κατανομές. Στις εκ των υστέρων κατανομές εμπεριέχεται όλη η στατιστική συμπερασματολογία των αγνώστων αυτών παραμέτρων όπως αυτή έχει προκύψει από την Μπεϋζιανή ανάλυση. Η ιδέα της εκ των προτέρων κατανομής αποτελεί και την «καρδιά» της θεωρίας κατά Bayes και μπορεί να αποτελέσει το μεγαλύτερο πλεονέκτημα ή το σοβαρότερο μειονέκτημα έναντι της κλασικής Στατιστικής.

Η παρούσα διπλωματική διαρθρώνεται σε τέσσερα κεφάλαια ως εξής: Στο *πρώτο κεφάλαιο* γίνεται μια εισαγωγή στα γενικευμένα γραμμικά μοντέλα και αναφέρονται κάποιες βασικές έννοιες και ιδιότητες των μοντέλων που ανήκουν σε αυτήν την κατηγορία. Στο *δεύτερο κεφάλαιο* αναπτύσσονται οι βασικές αρχές της Μπεϋζιανής Στατιστικής θεωρίας όπου μεταξύ άλλων δίνεται ο ορισμός της εκ των προτέρων κατανομής, της εκ των υστέρων κατανομής και του Θεωρήματος Bayes. Το *τρίτο κεφάλαιο* περιγράφει τον τρόπο με τον οποίο η Μπεϋζιανή θεωρία αντιμετωπίζει το πρόβλημα της επιλογής μοντέλων και μεταβλητών στα γενικευμένα γραμμικά μοντέλα και αναφέρονται όλες οι βασικές έννοιες που συνδέονται με το πρόβλημα αυτό. Τέλος, στο *τέταρτο κεφάλαιο* αναλύεται κατά κύριο λόγο ο αλγόριθμος MC^3 (Markov Chain Monte Carlo Model Composition) που αποτελεί μία από τις πολλές Μπεϋζιανές υπολογιστικές μεθόδους για την επιλογή μοντέλων καθώς και τρεις εφαρμογές του αλγορίθμου αυτού σε πραγματικά δεδομένα.

ΚΕΦΑΛΑΙΟ 1

ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

1.1 Εισαγωγή

Τα γενικευμένα γραμμικά μοντέλα (Generalized Linear Models) (GLM) είναι στατιστικά μοντέλα που αποτελούν τη φυσική γενίκευση των κλασικών γραμμικών μοντέλων. Στο ***κλασικό γραμμικό μοντέλο*** ασχολούμαστε με το πρόβλημα της πρόβλεψης μιας μεταβλητής απόκρισης Y όταν γνωρίζουμε τις τιμές των επεξηγηματικών μεταβλητών $X_1, X_2, ..., X_p$ τις οποίες συμβολίζουμε για συντομία με το διάνυσμα $\mathbf{X}$. Για δεδομένο $\mathbf{X} = \mathbf{x}$, όπου $\mathbf{x} = [x_1, x_2, ..., x_p]$, θεωρούμε ότι η τυχαία μεταβλητή $Y \mid \mathbf{X} = \mathbf{x} \sim N(\mu, \sigma^2)$ με

$$\mu = E\left[Y \mid \mathbf{X} = \mathbf{x}\right] = \beta_0 + \beta_1 \cdot x_1 + \ \dots \ + \beta_p \cdot x_p. \quad (1.1.1)$$

Οι ποσότητες $\beta_0, \beta_1, ..., \beta_p$ ονομάζονται ***συντελεστές*** του γραμμικού μοντέλου τους οποίους καλούμαστε να εκτιμήσουμε.

Το παραπάνω μοντέλο μπορεί να γενικευτεί ως προς τις υποθέσεις και πιο συγκεκριμένα, η επέκταση αφορά περισσότερο τις εξής δύο περιπτώσεις

- Οι τυχαία μεταβλητή $Y \mid \mathbf{X} = \mathbf{x}$ μπορεί να ακολουθεί κατανομή διαφορετική της Κανονικής, (π.χ Poisson, Διωνυμική).
- Η σχέση μεταξύ της δεσμευμένης μέσης τιμής $E\left[Y \mid \mathbf{X} = \mathbf{x}\right]$ και των τιμών $x_1, x_2, ..., x_p$ των μεταβλητών $X_1, X_2, ..., X_p$ δεν είναι ανάγκη να προσδιορίζεται από την απλή γραμμική μορφή της σχέσης (1.1.1).

Η πρώτη γενίκευση βασίζεται στο γεγονός ότι μπορούμε να θεωρήσουμε μια ευρύτερη κλάση κατανομών, γνωστή ως Εκθετική Οικογένεια Κατανομών. Η δεύτερη γενίκευση αναφέρεται στη σχέση (1.1.1) η οποία μπορεί να αντικατασταθεί από τη σχέση

$$g(\mu) = \beta_0 + \beta_1 \cdot x_1 + \ \dots \ + \beta_p \cdot x_p,$$

όπου g είναι μια μονότονη και διαφορίσιμη συνάρτηση που συνδέει τη μέση τιμή της τυχαίας μεταβλητής $Y \mid \mathbf{X} = \mathbf{x}$ με τη γραμμική συνιστώσα του δευτέρου μέλους.

1.2 Εκθετική Οικογένεια Κατανομών

Η Εκθετική Οικογένεια Κατανομών περιλαμβάνει ένα πλήθος από γνωστές κατανομές και ανάλογα με τον αριθμό των παραμέτρων που χαρακτηρίζουν μια κατανομή χωρίζεται σε μονοπαραμετρική και πολυπαραμετρική. Στην παρούσα ενότητα θα αναφερθούμε στη μονοπαραμετρική και τη διπαραμετρική Εκθετική Οικογένεια Κατανομών μιας και αυτές καλύπτουν τις περισσότερες περιπτώσεις κατανομών που συναντάμε στις εφαρμογές.

Μια τυχαία μεταβλητή Y ανήκει στη ***διπαραμετρική Εκθετική Οικογένεια Κατανομών*** αν η κατανομή της μπορεί να εκφραστεί στη μορφή

$$f(y \mid \theta,\phi)=\exp\left\{\frac{y\cdot\theta-b(\theta)}{a(\phi)}-c(y,\phi)\right\},$$

όπου a,b και c είναι γνωστές συναρτήσεις. Η θ ονομάζεται κανονική παράμετρος και είναι παράμετρος κεντρικής τάσης (location parameter) ενώ η ϕ ονομάζεται παράμετρος διασποράς (dispersion parameter). Ορίζουμε επίσης το στήριγμα της κατανομής ως εξής

$$S=\left\{y\in\mathbb{R} : f(y,\theta,\phi)>0\right\}.$$

<u>Παράδειγμα</u>

<u>Κανονική κατανομή</u>: Η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής Y που ακολουθεί την Κανονική κατανομή $N(\mu,\sigma^2)$ μπορεί να γραφτεί ως εξής

$$\begin{aligned} f(y \mid \mu,\sigma^2)&=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}\\ &=\exp\left\{\left(\frac{y\cdot\mu-\mu^2/2}{\sigma^2}\right)-\frac{y^2/\sigma^2+\log(2\pi\sigma^2)}{2}\right\}. \end{aligned}$$

Στη συγκεκριμένη περίπτωση έχουμε

$$\theta=\mu\ ,\ \phi=\sigma^2,\ b(\theta)=\theta^2/2\ ,\ a(\phi)=\phi \text{ και } c(y,\phi)=\frac{y^2/\sigma^2+\log(2\pi\sigma^2)}{2}.$$

Θεωρούμε τώρα μια τυχαία μεταβλητή Y της οποίας η κατανομή εξαρτάται από μία μόνο παράμετρο θ. Η κατανομή της ανήκει στην ***μονοπαραμετρική Εκθετική Οικογένεια Κατανομών*** αν μπορεί να γραφτεί στη μορφή

$$f(y \mid \theta)=s(y)t(\theta)\exp\left\{a(y)b(\theta)\right\},$$

όπου a,b,s και t είναι γνωστές συναρτήσεις. Την παραπάνω σχέση τη συναντάμε πολλές φορές και στην ισοδύναμη μορφή

$$f(y \mid \theta) = \exp\left\{a(y)b(\theta) + c(\theta) + d(y)\right\},$$

όπου $s(y) = \exp\left\{d(y)\right\}$ και $t(\theta) = \exp\left\{c(\theta)\right\}$. Όπως και στην περίπτωση της διπαραμετρικής Εκθετικής Οικογένειας Κατανομών ορίζεται και εδώ το στήριγμα

$$S = \left\{y \in \mathbb{R} : f(y, \theta) > 0\right\}.$$

Αν $a(y) = y$, η κατανομή λέγεται ότι είναι σε ***κανονική μορφή*** και η συνάρτηση $b(\theta)$ καλείται ***φυσική παράμετρος*** της κατανομής. Αν υπάρχουν και άλλες παράμετροι στην κατανομή της παραμέτρου θ που μας ενδιαφέρει, αυτές θεωρούνται ως 'οχληρές' (nuisance) παράμετροι. Αποδεικνύεται ότι αν η κατανομή μιας τυχαίας μεταβλητής Y είναι μέλος της Εκθετικής Οικογένειας Κατανομών, ισχύουν οι παρακάτω ιδιότητες

$$E\left[a(Y)\right] = -\frac{c'(\theta)}{b'(\theta)}$$

και

$$Var\left[a(Y)\right] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{\left[b'(\theta)\right]^3},$$

από τις οποίες, αν $a(y) = y$ δηλαδή η κατανομή είναι σε κανονική μορφή, μπορούμε να υπολογίσουμε τη μέση τιμή και τη διασπορά της τυχαίας μεταβλητής Y.
Η μονοπαραμετρική Εκθετική Οικογένεια Κατανομών περιλαμβάνει μια σειρά από κατανομές όπως η Εκθετική, η Poisson, η Γάμμα με μία παράμετρο, η Διωνυμική και η Κανονική με γνωστή διακύμανση.

<u>Παραδείγματα</u>

1) <u>Κατανομή Poisson</u>: Μία διακριτή τυχαία κατανομή Y, με τιμές στο $\{0,1,2,...\}$, ακολουθεί την κατανομή $Poisson(\theta)$ αν έχει συνάρτηση μάζας πιθανότητας την

$$f(y \mid \theta) = \frac{\theta^y \cdot e^{-\theta}}{y!}, \ \theta > 0, \ y = 0,1,2,... \ .$$

Αν γράψουμε την παραπάνω σχέση στη μορφή

$$f(y \mid \theta) = \exp\left\{y \log(\theta) - \theta - \log(y!)\right\},$$

παρατηρούμε ότι η κατανομή ανήκει στην Εκθετική Οικογένεια Κατανομών με $a(y) = y$ (είναι δηλαδή και σε κανονική μορφή), $b(\theta) = \log(\theta)$ (φυσική παράμετρος), $c(\theta) = -\theta$ και $d(y) = -\log(y!)$.

2) <u>Κανονική κατανομή με γνωστή διακύμανση</u>: Ως γνωστόν, η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής που ακολουθεί την Κανονική κατανομή $N(\mu,\sigma^2)$ δίνεται από τον τύπο

$$f(y\,|\,\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\},\ y\in\mathbf{R},$$

όπου $\mu\in\mathbf{R}$ είναι η παράμετρος ενδιαφέροντος και η $\sigma>0$ θεωρείται ως οχληρή παράμετρος. Η Κανονική κατανομή μπορεί να γραφτεί στην ισοδύναμη μορφή

$$f(y\,|\,\mu,\sigma^2)=\exp\left\{y\cdot\frac{\mu}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\frac{1}{2}\log(2\pi\sigma^2)-\frac{y^2}{2\sigma^2}\right\},$$

απ'όπου καταλαβαίνουμε ότι ανήκει στην Εκθετική Οικογένεια Κατανομών με $a(y)=y$ (κανονική μορφή), $b(\mu)=\dfrac{\mu}{\sigma^2}$ (φυσική παράμετρος), $c(\mu)=-\dfrac{\mu^2}{2\sigma^2}-\dfrac{1}{2}\log(2\pi\sigma^2)$ και $d(y)=-\dfrac{y^2}{2\sigma^2}$.

3) <u>Διωνυμική κατανομή</u>: Έστω ότι η τυχαία μεταβλητή Y αφορά των αριθμό των επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli με πιθανότητα επιτυχίας π. Τότε η Y είναι διακριτή και ακολουθεί τη Διωνυμική κατανομή $Binomial(n,\pi)$ με συνάρτηση μάζας πιθανότητας

$$f(y\,|\,\pi)=\binom{n}{y}\pi^y(1-\pi)^{n-y},\ \ y=0,1,...,n,$$

με π να είναι η παράμετρος που μας ενδιαφέρει. Γράφοντας την κατανομή της Y στη μορφή

$$f(y\,|\,\pi)=\exp\left\{y\log\left(\frac{\pi}{1-\pi}\right)+n\log(1-\pi)+\log\binom{n}{y}\right\},$$

συμπεραίνουμε ότι ανήκει στην Εκθετική Οικογένεια Κατανομών με $a(y)=y$ (είναι σε κανονική μορφή), $b(\pi)=\log\left(\dfrac{\pi}{1-\pi}\right)$ (φυσική παράμετρος), $c(\pi)=n\log(1-\pi)$ και $d(y)=\log\dbinom{n}{y}$.

1.3 Γενικευμένα γραμμικά μοντέλα – Βασικοί ορισμοί

Οι Nelder και Wedderburn (1972) προσπάθησαν να ενοποιήσουν αρκετές στατιστικές μεθόδους χρησιμοποιώντας την ιδέα του γενικευμένου γραμμικού μοντέλου. Το μοντέλο αυτό ορίζεται σε σχέση με ένα σύνολο από ανεξάρτητες τυχαίες μεταβλητές $Y_1, Y_2, ..., Y_n$ όπου κάθε μία από αυτές ακολουθεί κατανομή που ανήκει στην Εκθετική Οικογένεια Κατανομών και έχει τις εξής ιδιότητες:

- Η κατανομή κάθε μιας από τις Y_i, $i = 1, ..., n$, είναι Κανονικής μορφής και εξαρτάται από μία μόνο παράμετρο θ_i, κατά συνέπεια

$$f(y_i \mid \theta_i) = \exp\left\{ y_i\, b_i(\theta_i) + c_i(\theta_i) + d_i(y_i) \right\}.$$

- Οι κατανομές όλων των Y_i, $i = 1, ..., n$, είναι της ίδιας μορφής (π.χ όλες Κανονικές ή όλες Διωνυμικές) και έτσι οι δείκτες i στις συναρτήσεις b, c και d δεν χρειάζονται.

Η από κοινού κατανομή των $Y_1, Y_2, ..., Y_n$ ή διαφορετικά η κατανομή του τυχαίου διανύσματος $\mathbf{Y} = \left[Y_1, Y_2, ..., Y_n \right]^T$, η οποία καλείται ***μεταβλητή απόκρισης*** (response variable), (αναφέρεται και ως ***εξαρτημένη μεταβλητή***), είναι η εξής

$$f(y_1, ..., y_n \mid \theta_1, ..., \theta_n) = \prod_{i=1}^{n} f(y_i \mid \theta_i) = \exp\left\{ \sum_{i=1}^{n} y_i\, b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i) \right\}.$$

Ας δούμε λοιπόν αναλυτικά στο σημείο αυτό την περιγραφή του προβλήματος. Έστω ότι έχουμε ένα τυχαίο διάνυσμα $\mathbf{Y} = \left[Y_1, Y_2, ..., Y_n \right]^T$ (μεταβλητή απόκρισης) και $(p+1)$ μεταβλητές $X_0, X_1, ..., X_p$ οι οποίες ονομάζονται ***επεξηγηματικές μεταβλητές*** (explanatory variables), (αναφέρονται και ως ***ανεξάρτητες μεταβλητές***). Θεωρούμε την X_0 σαν σταθερή και συγκεκριμένα $X_0 = 1$. Έστω επίσης ότι έχουμε συλλέξει n παρατηρήσεις (δείγμα) για κάθε μία από τις μεταβλητές

Α/Α ΠΑΡΑΤΗΡΗΣΕΩΝ	ΤΙΜΕΣ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ			
	Y	X_1	$\cdots$	X_p
1	y_1	$x_1^{(1)}$	$\cdots$	$x_p^{(1)}$
2	y_2	$x_1^{(2)}$	$\cdots$	$x_p^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	$x_1^{(n)}$	$\cdots$	$x_p^{(n)}$

όπου y_i είναι η i-παρατήρηση της τ.μ $\mathbf{Y}$, (ή αλλιώς η τιμή της τ.μ Y_i) και $x_j^{(i)}$ είναι η τιμή της μεταβλητής X_j στην i-παρατήρηση της, για $i = 1, ..., n$ και $j = 1, ..., p$. Συμβολίζουμε με $E\left[Y_i \mid X_j = x_j^{(i)},\ j = 0, ..., p \right] = \mu_i$ τη μέση τιμή της Y_i δοθέντων των παρατηρήσεων $x_j^{(i)}$, $j = 0, ..., p$. <u>Για ευκολία συμβολισμού</u> θα γράφουμε $E\left[Y_i \right] = \mu_i$. Ορίζουμε ακόμα την ποσότητα

$$\eta_i := \beta_0 \cdot 1 + \beta_1 \cdot x_1^{(i)} + \ \dots \ + \beta_p \cdot x_p^{(i)},\ i = 1, ..., n,$$

η οποία καλείται ***γραμμικός προσδιορισμός*** (linear predictor) του μοντέλου. Στα γενικευμένα γραμμικά μοντέλα, θεωρούμε μια μονότονη και διαφορίσιμη συνάρτηση g η οποία εκφράζει τον τρόπο σύνδεσης, συνήθως της μέσης τιμής μ_i με τον γραμμικό προσδιορισμό η_i και ονομάζεται ***συνδετική συνάρτηση*** (link function). Οι παρακάτω ισότητες δείχνουν ακριβώς αυτόν τον τρόπο σύνδεσης

$$\begin{array}{l} g(\mu_1) = \beta_0 \cdot 1 + \beta_1 \cdot x_1^{(1)} + \ldots + \beta_p \cdot x_p^{(1)} \\ g(\mu_2) = \beta_0 \cdot 1 + \beta_1 \cdot x_1^{(2)} + \ldots + \beta_p \cdot x_p^{(2)} \\ \quad\vdots \qquad\quad \vdots \qquad \vdots \qquad\qquad\quad \vdots \\ g(\mu_n) = \beta_0 \cdot 1 + \beta_1 \cdot x_1^{(n)} + \ldots + \beta_p \cdot x_p^{(n)}. \end{array}$$

Τις παραπάνω σχέσεις μπορούμε να τις εκφράσουμε και σε μορφή πινάκων ως εξής

$$\begin{bmatrix} g(\mu_1) \\ g(\mu_2) \\ \vdots \\ g(\mu_n) \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & \ldots & x_p^{(1)} \\ 1 & x_1^{(2)} & \ldots & x_p^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(n)} & \ldots & x_p^{(n)} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Θέτοντας $\boldsymbol{\eta} = [\eta_1, \eta_2, ..., \eta_n]^T$, $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_p]^T$ και ορίζοντας ως ***πίνακα σχεδιασμού*** (design matrix) τον $n \times (p+1)$ πίνακα

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & \ldots & x_p^{(1)} \\ 1 & x_1^{(2)} & \ldots & x_p^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(n)} & \ldots & x_p^{(n)} \end{bmatrix},$$

γράφουμε συνοπτικά το αποτέλεσμα στη μορφή $\boldsymbol{\eta} = \mathbf{X} \cdot \boldsymbol{\beta}$ ή στη μορφή

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \cdot \boldsymbol{\beta} = \sum_{j=0}^{p} x_j^{(i)} \beta_j,$$

για $i = 1, ..., n$, όπου $\mathbf{x}_i^T$ είναι η i-γραμμή του πίνακα σχεδιασμού $\mathbf{X}$.

<u>Παρατηρήσεις</u>:

1) Σε όλα τα παραπάνω έχουμε κάνει την υπόθεση ότι οι επεξηγηματικές μεταβλητές X_j δεν υπόκεινται σε σφάλματα, υπό την έννοια ότι η όλη ανάλυση γίνεται δοθέντων των παρατηρήσεων $x_j^{(i)}$, $j = 1, ..., p$, $i = 1, ..., n.$

2) Ο συντελεστής β_0 εκφράζει την αναμενόμενη τιμή της τυχαίας μεταβλητής Y όταν όλες οι μεταβλητές X_j, $j = 1, ..., p$ είναι μηδέν, ενώ ο συντελεστής β_j, για

$j = 1,...,p,$ εκφράζει την αναμενόμενη μεταβολή της Y όταν η X_j αυξηθεί κατά μία μονάδα αλλά οι υπόλοιπες επεξηγηματικές μεταβλητές παραμείνουν σταθερές.

1.4 Το Κανονικό μοντέλο

Η πιο δημοφιλής περίπτωση των γενικευμένων γραμμικών μοντέλων είναι το λεγόμενο Κανονικό γραμμικό μοντέλο

$$E\left[Y_i \mid \mathbf{x}_i\right] = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad Y_i \mid \mathbf{x}_i \sim N(\mu_i, \sigma^2),$$

με τις Y_i, $i = 1,...,n$ ανεξάρτητες μεταξύ τους. Στη συγκεκριμένη περίπτωση η συνάρτηση σύνδεσης είναι η ταυτοτική $g(\mu_i) = \mu_i$. Το μοντέλο αυτό συνήθως γράφεται και στη μορφή

$$\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{e},$$

όπου $\mathbf{e} = \left[e_1,...,e_n\right]^T$, με τα e_i, $i = 1,...,n$, ανεξάρτητα και να ακολουθούν την κατανομή $N(0,\sigma^2)$. Για το Κανονικό μοντέλο, ο ***εκτιμητής μέγιστης πιθανοφάνειας*** των συντελεστών $\boldsymbol{\beta} = \left[\beta_0, \beta_1,...,\beta_p\right]^T$ είναι ο

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\mathbf{T}\mathbf{X}\right)^{-1} \mathbf{X}^\mathbf{T}\mathbf{y},$$

όπου $\mathbf{y} = \left[y_1,...,y_n\right]^T$, το διάνυσμα των παρατηρήσεων της τ.μ Y.

1.5 Μοντέλο Poisson

Αν $Y_i \sim Poisson(\theta_i)$, $i = 1,...,n$, τότε ως γνωστόν $E\left[Y_i\right] = \mu_i = \theta_i > 0$. Η συνδετική συνάρτηση που χρησιμοποιούμε στην περίπτωση αυτή είναι η $g(\mu_i) = \log(\mu_i)$, δηλαδή έχουμε

$$\log(\mu_i) = \eta_i = \beta_0 + \beta_1 \cdot x_1^{(i)} + \ \dots \ + \beta_p \cdot x_p^{(i)} = \mathbf{x}_i^T \cdot \boldsymbol{\beta},$$

από την οποία, λύνοντας ως προς $\mu_i = \theta_i$ προκύπτει

$$\mu_i = \exp\left\{\mathbf{x}_i^T \cdot \boldsymbol{\beta}\right\} > 0.$$

1.6 Διωνυμικό μοντέλο

Στην περίπτωση όπου οι μεταβλητές $Y_i \sim Binomial(\nu_i, \pi_i)$, η μέση τιμή κάθε μιας είναι $E[Y_i] = \mu_i = \nu_i \cdot \pi_i$. Η παράμετρος που μας ενδιαφέρει είναι η πιθανότητα π_i και η σύνδεσή της με τον γραμμικό προσδιορισμό η_i γίνεται μέσω της συνάρτησης $\mathrm{g}(\pi_i) = \mathrm{logit}(\pi_i)$ και πιο συγκεκριμένα

$$\mathrm{logit}(\pi_i) := \log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i = \mathbf{x}_i^T \cdot \boldsymbol{\beta}.$$

Όταν $\nu_i = 1$, προκύπτει η κατανομή $Bernoulli(\pi_i)$ για την οποία χρησιμοποιούμε συνήθως την ίδια συνάρτηση σύνδεσης. Λύνοντας την παραπάνω σχέση ως προς π_i παίρνουμε

$$\pi_i = \frac{\exp\left\{\mathbf{x}_i^T \cdot \boldsymbol{\beta}\right\}}{1+\exp\left\{\mathbf{x}_i^T \cdot \boldsymbol{\beta}\right\}} \in (0,1).$$

Είναι ενδιαφέρον να παρατηρήσουμε στο σημείο αυτό την αναγκαιότητα χρήσης της συνδετικής συνάρτησης g. Αν για παράδειγμα στο Διωνυμικό μοντέλο θέταμε $g(\pi_i) = \pi_i = \mathbf{x}_i^T \cdot \boldsymbol{\beta}$, τότε θα υπήρχε πρόβλημα διότι το γραμμικό μέρος $\eta_i = \mathbf{x}_i^T \cdot \boldsymbol{\beta}$ μπορεί να πάρει οποιαδήποτε τιμή στο $\mathbf{R}$ ενώ η πιθανότητα $\pi_i \in (0,1)$. Για το λόγο αυτό, χρειαζόμαστε μια κατάλληλη συνδετική συνάρτηση g που να απεικονίζει το διάστημα $(0,1)$ σε όλο το διάστημα $\mathbf{R}$, ώστε η αντίστροφή της (που υπάρχει αφού η g είναι μονότονη) να απεικονίζει το $\mathbf{R}$ στο $(0,1)$. Υπάρχουν και άλλες συναρτήσεις που χρησιμοποιούνται στα Διωνυμικά (και στα Bernoulli) μοντέλα αντί της logit, με πιο γνωστές την probit (ή αντίστροφή Κανονική) $g(\pi_i) = \Phi^{-1}(\pi_i)$, όπου Φ η τυποποιημένη Κανονική κατανομή και την συμπληρωματική $\log-\log$ συνάρτηση με τύπο $g(\pi_i) = \log\left[-\log(1-\pi_i)\right]$.

ΚΕΦΑΛΑΙΟ 2

ΜΠΕΫΖΙΑΝΉ ΣΤΑΤΙΣΤΙΚΉ ΘΕΩΡΊΑ ΚΑΙ MCMC

2.1 Εισαγωγή

Παρότι οι αρχές της Μπεϋζιανής θεωρίας ανάγονται στον 18ο αιώνα, η Στατιστική κατά Bayes δεν είχε ευρεία διάδοση για αρκετά χρόνια, καθώς στις περισσότερες περιπτώσεις οι ερευνητές ήταν αναγκασμένοι να αντιμετωπίσουν δυσεπίλυτα ολοκληρώματα υψηλής τάξης. Όμως η ανάπτυξη των υπολογιστών και η εμφάνιση μεθόδων προσομοίωσης έκανε την εφαρμογή της Μπεϋζιανής θεωρίας πρακτικά δυνατή. Ειδικότερα, οι πρόσφατες μέθοδοι προσομοίωσης Monte Carlo με την χρήση Μαρκοβιανών αλυσίδων (MCMC) συνέβαλλαν αποφασιστικά στην διεκπεραίωση της Μπεϋζιανής ανάλυσης.

Η Μπεϋζιανή συμπερασματολογία είναι μία μεθοδολογική προσέγγιση η οποία συνθέτει τη θεωρητική ή εμπειρική αντίληψη μιας τυχαίας δοκιμασίας, με τα παρατηρηθέντα δεδομένα. Δηλαδή, είναι η διαδικασία κατά την οποία βελτιώνεται η εκ των προτέρων γνώση (a-priori) που έχουμε για μία παράμετρο, μέσω της πιθανοφάνειας, στην εκ των υστέρων γνώση (posterior) που θα έχουμε για αυτή. Με άλλα λόγια, η Μπεϋζιανή Στατιστική βασίζεται στην υποκειμενική πιθανότητα, τις πεποιθήσεις, τις γνώσεις που έχουμε πριν αναλύσουμε τα δεδομένα κάτι το οποίο εκφράζεται με την εκ των προτέρων πληροφορία. Μέσω του Θεωρήματος του Bayes η πληροφορία αναβαθμίζεται στην εκ των υστέρων πληροφορία όπου συνδυάζεται η εκ των προτέρων και η πληροφορία από τα δεδομένα.

Το πλαίσιο στο οποίο κινείται η συμπερασματολογία κατά Bayes είναι παρόμοιο με αυτό της κλασικής Στατιστικής: υπάρχει η παράμετρος $\boldsymbol{\theta}$ του πληθυσμού την οποία θέλουμε να εκτιμήσουμε, καθώς και η πιθανότητα $f(\mathbf{y}\,|\,\boldsymbol{\theta})$ η οποία καθορίζει την πιθανότητα παρατήρησης διαφορετικών $\mathbf{y}$ κάτω από διαφορετικές τιμές της παραμέτρου $\boldsymbol{\theta}$. Όμως η <u>θεμελιώδης διαφορά</u> είναι ότι στη Μπεϋζιανή Στατιστική η παράμετρος $\boldsymbol{\theta}$ δεν θεωρείται σταθερός αριθμός αλλά χρησιμοποιείται σαν τυχαία ποσότητα (τυχαία μεταβλητή). Αν και η διαφορά αυτή μπορεί να φανεί όχι και τόσο ουσιαστική, οδηγεί σε μία τελείως διαφορετική προσέγγιση, ως προς την ερμηνεία, από αυτήν την κλασικής Στατιστικής. Χρησιμοποιώντας την παράμετρο $\boldsymbol{\theta}$ σαν τυχαία, όλη η ανάπτυξη της Μπεϋζιανής συμπερασματολογίας πηγάζει και εξαρτάται μόνο από την θεωρία πιθανοτήτων. Αυτό έχει πολλά πλεονεκτήματα και σημαίνει πως όλα τα συμπεράσματα μπορούν να παρουσιαστούν με την μορφή πιθανοτήτων για την παράμετρο $\boldsymbol{\theta}$.

Η Μπεϋζιανή προσέγγιση θέτει εκ των προτέρων πληροφορία στις άγνωστες παραμέτρους, μέσω μιας εκ των προτέρων κατανομής, και μας παρέχει τις εκ των υστέρων κατανομές των παραμέτρων. Τα συμπεράσματα βασίζονται στους μέσους (ή σε άλλες ποσότητες) των εκ των υστέρων κατανομών. Η Μπεϋζιανή συμπερασματολογία δηλαδή, τυποποιεί την ενσωμάτωση των εκ των προτέρων (a-priori) πληροφοριών, κάτι που στην κλασική Στατιστική συνήθως γίνεται όχι φανερά.

Κλείνοντας αυτήν την εισαγωγική παράγραφο και συνοψίζοντας, μπορούμε να πούμε ότι η Μπεϋζιανή Στατιστική θέτει με σαφήνεια την ιδέα ότι όλες οι πιθανότητες είναι υποκειμενικές και εξαρτώνται από τις πεποιθήσεις του κάθε ατόμου και τις γνώσεις που μπορεί να έχει καθένας για ένα δεδομένο πρόβλημα.

2.2 Εκ των προτέρων κατανομή - Εκ των υστέρων κατανομή - Θεώρημα Bayes

Σημαντικό κομμάτι της Μπεϋζιανής Στατιστικής και ιδιαίτερου ενδιαφέροντος είναι η επιλογή της εκ των προτέρων κατανομής κάθε φορά. Κάθε πρόβλημα είναι μοναδικό και έχει το δικό του περιεχόμενο. Από αυτό ακριβώς το περιεχόμενο εξάγονται οι a-priori πληροφορίες και είναι η διατύπωση και η εκμετάλλευση της προηγούμενης γνώσης που διαχωρίζουν την Μπεϋζιανή θεωρία από αυτήν της κλασικής Στατιστικής. Όταν προσπαθούμε να εκτιμήσουμε την παράμετρο $\boldsymbol{\theta}$ θα πρέπει να έχουμε κάποια γνώση ή κάποια πεποίθηση σχετικά με την τιμή της $\boldsymbol{\theta}$ προτού λάβουμε υπόψη μας τα δεδομένα. Αυτό που επιτυγχάνεται με το Θεώρημα του Bayes, είναι η βελτίωση της εκ των προτέρων γνώσης που έχουμε για την παράμετρο, σύμφωνα με την καινούρια γνώση που μας δίνει το παρατηρηθέν δείγμα εκφρασμένη μέσω της πιθανοφάνειας. Ο συνδυασμός της εκ των προτέρων και της δειγματικής πληροφορίας, μας δίνει την εκ των υστέρων πληροφορία που εκφράζεται μέσω της εκ των υστέρων κατανομής. Για να μπορέσει όμως να επιτευχθεί αυτό θα πρέπει να καθορίσουμε την εκ των προτέρων κατανομή $f(\boldsymbol{\theta})$ (prior distribution), η οποία αντιπροσωπεύει όπως είπαμε τις πεποιθήσεις μας για την κατανομή του $\boldsymbol{\theta}$ προτού αποκτήσουμε οποιαδήποτε πληροφορία από τα δεδομένα μας.

Η επιλογή της εκ των προτέρων κατανομής πρέπει να γίνεται με πολύ προσοχή διότι καθιστά την ανάλυση υποκειμενική. Διαφορετική εκ των προτέρων επιλογή οδηγεί σε διαφορετικά αποτελέσματα. Ωστόσο η επιρροή της εκ των προτέρων κατανομής γίνεται ολοένα και μικρότερη καθώς προστίθενται νέα δεδομένα. Συνήθως διαλέγουμε κατανομές που απλοποιούν τους υπολογισμούς οι οποίες όμως «περιορίζουν» την επιλογή της $f(\boldsymbol{\theta})$ σε κάποια από τις γνωστές οικογένειες κατανομών. Σε πολλές περιπτώσεις η διαθέσιμη εκ των προτέρων πληροφορία είναι περιορισμένη ή ακόμα και ανύπαρκτη. Στην περίπτωση αυτή επιθυμούμε η πληροφορία που προέρχεται από τα δεδομένα να κυριαρχήσει στον τελικό υπολογισμό της εκ των υστέρων κατανομής και είναι αρκετά συνηθισμένο να χρησιμοποιούμε μια a-priori η οποία αντανακλά την άγνοια μας για την παράμετρο. Γενικά οι κατανομές διακρίνονται σε δύο μεγάλες κατηγορίες, τις πληροφοριακές (informative) και τις μη-πληροφοριακές (non-informative).

Υπάρχουν τέσσερα βασικά βήματα για την Μπεϋζιανή προσέγγιση ενός πιθανοθεωρητικού προβλήματος :

1. Καθορισμός του μοντέλου πιθανοφάνειας $f(\mathbf{y} \mid \boldsymbol{\theta})$.
2. Καθορισμός της εκ των προτέρων κατανομής $f(\boldsymbol{\theta})$.
3. Υπολογισμός της εκ των υστέρων κατανομής $f(\boldsymbol{\theta} \mid \mathbf{y})$.
4. Εξαγωγή συμπερασμάτων από την εκ των υστέρων κατανομή.

Πιο συγκεκριμένα, έστω $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ είναι ένα τυχαίο δείγμα από γνωστή κατανομή με σ.π.π. (ή σ.μ.π. στη διακριτή περίπτωση) f και άγνωστη παράμετρο $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m) \in \boldsymbol{\Theta} \subset \mathbf{R}^m$. Η συνάρτηση πιθανοφάνειας (likelihood) είναι η

$f(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta})$, η οποία όπως έχει αναφερθεί καθορίζει την πιθανότητα παρατήρησης διαφορετικών y_i κάτω από διαφορετικές τιμές της παραμέτρου $\boldsymbol{\theta}$. Υποθέτοντας ακόμα ότι $f(\boldsymbol{\theta})$ είναι η εκ των προτέρων κατανομή για την παράμετρο $\boldsymbol{\theta}$, η συμπερασματολογία μας τότε δεν προκύπτει μόνο από την μελέτη των δεδομένων, δηλαδή την πιθανοφάνεια $f(\mathbf{y} \mid \boldsymbol{\theta})$, αλλά τελικά από την συνάρτηση πιθανότητας της κατανομής της παραμέτρου δοθέντος των δεδομένων $f(\boldsymbol{\theta} \mid \mathbf{y})$, η οποία καλείται εκ των υστέρων κατανομή (posterior distribution). Σε πολλές περιπτώσεις αυτό οδηγεί σε περισσότερο φυσικά συμπεράσματα σε σχέση με την κλασική Στατιστική. Ο υπολογισμός της γίνεται με χρήση του ***Θεωρήματος Bayes***:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{f(\boldsymbol{\theta}) \cdot f(\mathbf{y} \mid \boldsymbol{\theta})}{\int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} f(\boldsymbol{\theta}) \cdot f(\mathbf{y} \mid \boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Σημειώνουμε ότι όταν η παράμετρος είναι διακριτή, τότε ο παρονομαστής αντικαθιστάται από το άθροισμα $\sum_j f(\theta_j) f(\mathbf{y} \mid \theta_j)$. Θα πρέπει να προσέξουμε ιδιαίτερα το γεγονός ότι από την στιγμή που ολοκληρώσαμε ως προς $\boldsymbol{\theta}$, ο παρανομαστής στο Θεώρημα του Bayes είναι συνάρτηση μόνο ως προς $\mathbf{y}$. Συνεπώς, για δεδομένες παρατηρήσεις $\mathbf{y}$, ο παρανομαστής είναι σταθερά και ονομάζεται σταθερά κανονικοποίησης. Με βάση αυτά ένας εναλλακτικός τρόπος παρουσίασης του Θεωρήματος του Bayes είναι ο εξής:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) \propto f(\boldsymbol{\theta}) \cdot f(\mathbf{y} \mid \boldsymbol{\theta}),$$

δηλαδή η εκ των υστέρων κατανομή είναι ανάλογη της a-priori κατανομής πολλαπλασιαζόμενης με την συνάρτηση πιθανοφάνειας.

2.3 Εκθετική Οικογένεια Κατανομών - Συζυγείς εκ των προτέρων κατανομές

Η εφαρμογή του Θεωρήματος του Bayes στην πράξη είναι αρκετά δύσκολη όσον αφορά τους μαθηματικούς υπολογισμούς και η δυσκολία αυτή οφείλεται κυρίως στο ολοκλήρωμα (σταθερά κανονικοποίησης) το οποίο υπάρχει στον παρανομαστή. Πολύ 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 διακύμανσή 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 η οποία θα είναι το αποτέλεσμα των πεποιθήσεων μας και ταυτόχρονα θα κάνει τους μαθηματικούς υπολογισμούς σχετικά εύκολους. Αυτό είναι πολύ σημαντικό καθώς κάποιες, ακόμα και απλές, επιλογές εκ των προτέρων κατανομών, εμφανίζουν μεγάλες υπολογιστικές δυσκολίες. Για το λόγο αυτό για αρκετό διάστημα η Μπεϋζιανή θεωρία περιορίστηκε σε κατανομές που διευκόλυναν τον υπολογισμό των εκ των υστέρων κατανομών. Τέτοιες κατανομές είναι οι ***συζυγείς εκ των προτέρων κατανομές***.

Ως συζυγείς ορίζονται οι εκ των προτέρων κατανομές οι οποίες εφαρμόζοντας το Θεώρημα του Bayes καταλήγουν σε εκ των υστέρων κατανομές που ανήκουν στην ίδια οικογένεια κατανομών με την εκ των προτέρων κατανομή. Χρησιμοποιούμε τις κατανομές αυτές όταν είναι συμβατές με τις πεποιθήσεις μας, όταν περιγράφουν την προηγούμενη γνώση που έχουμε για την παράμετρο. Σε πολλές περιπτώσεις ο πλούτος της οικογένειας των συζυγών κατανομών είναι αρκετά μεγάλος ώστε να μπορέσει να προσδιοριστεί η συζυγής εκ των προτέρων κατανομή η οποία είναι αρκετά κοντά στις εκ των προτέρων πεποιθήσεις μας. Αυτό όμως το οποίο θα πρέπει να τονίσουμε είναι πως οι συζυγείς εκ των προτέρων κατανομές δεν θα πρέπει να χρησιμοποιούνται απλά επειδή κάνουν τους μαθηματικούς υπολογισμούς ευκολότερους και αν δεν υπάρχει a-priori κατανομή μέσα στην οικογένεια αυτή η οποία αντανακλά τι πραγματικά πιστεύουμε, τότε θα πρέπει να αποφύγουμε την εν λόγω προσέγγιση .

Η μόνη περίπτωση στην οποία οι συζυγείς κατανομές προκύπτουν εύκολα, είναι για τα υποδείγματα που ανήκουν στην Εκθετική Οικογένεια Κατανομών. Δηλαδή, όταν η συνάρτηση πυκνότητας πιθανότητας των παρατηρήσεων του δείγματος εκφράζεται με τη μορφή:

$$f(y \mid \boldsymbol{\theta}) = h(y) g(\boldsymbol{\theta}) \exp\left\{\phi^T(\boldsymbol{\theta}) \cdot u(y)\right\},$$

όπου h, u είναι συναρτήσεις των παρατηρήσεων y και g, ϕ συναρτήσεις της παραμέτρου $\boldsymbol{\theta}$, ενώ για τις συναρτήσεις h , u , g και ϕ ισχύει:

$$\int f(y \mid \boldsymbol{\theta}) dy = g(\boldsymbol{\theta}) \int h(y) \exp\left\{\phi^T(\boldsymbol{\theta}) u(y)\right\} dy = 1 .$$

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα $\mathbf{y} = \left[y_1, y_2, ..., y_n\right]^T$ από γνωστή κατανομή f η οποία ανήκει στην Εκθετική Οικογένεια Κατανομών, δηλαδή μπορεί να γραφτεί στη μορφή

$$f(y \mid \boldsymbol{\theta}) = h(y) g(\boldsymbol{\theta}) \exp\left\{\phi^T(\boldsymbol{\theta}) \cdot u(y)\right\}.$$

Η συνάρτηση πιθανοφάνειας θα είναι

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}) &= \left\{\prod_{i=1}^{n} h(y_i)\right\} \left\{g(\boldsymbol{\theta})\right\} \exp\left\{\phi^T(\boldsymbol{\theta}) \sum_{i=1}^{n} u(y_i)\right\} \\ &\propto \left\{g(\boldsymbol{\theta})\right\} \exp\left\{\phi^T(\boldsymbol{\theta}) \sum_{i=1}^{n} u(y_i)\right\}. \end{aligned}$$

Χρησιμοποιώντας μια εκ των προτέρων κατανομή της μορφής

$$f(\boldsymbol{\theta}) \propto g(\boldsymbol{\theta}) \exp\left\{\phi^T(\boldsymbol{\theta}) \cdot b\right\},$$

η εκ των υστέρων κατανομή θα υπολογιστεί κατά τα γνωστά από το Θεώρημα του Bayes

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) = \{g(\boldsymbol{\theta})\}^{n+d} \exp\left\{\phi^T(\boldsymbol{\theta})\left[b+\sum_{i=1}^{n} u(y_i)\right]\right\}$$
$$= \{g(\boldsymbol{\theta})\}^{p} \exp\{\phi^T(\boldsymbol{\theta})\cdot q\},$$

όπου $p = n + d$ και $q = b + \sum_{i=1}^{n} u(y_i)$.

Το συμπέρασμα είναι ότι η εκ των υστέρων ανήκει στην ίδια οικογένεια κατανομών με την εκ των προτέρων κατανομή αλλά με προσαρμοσμένες παραμέτρους. Πρέπει να τονιστεί στο σημείο αυτό για να ισχύουν όλα τα προηγούμενα, θα πρέπει οι χώροι Θ και Ψ, όπου παίρνουν τιμές οι μεταβλητές $\boldsymbol{\theta}$ και $\mathbf{y}$ αντίστοιχα, να είναι ανεξάρτητοι (ασύνδετοι), δηλαδή οι μεταβολές στις τιμές του $\boldsymbol{\theta}$ να μην επηρεάζουν τις τιμές του $\mathbf{y}$.

Συνοψίζοντας την ιδέα των συζυγών εκ των προτέρων κατανομών αναφέρουμε το παρακάτω

Σκοπός: Δοθέντος της κατανομής από την οποία προέρχονται τα δεδομένα, να βρούμε κατάλληλη εκ των προτέρων κατανομή ώστε η εκ των υστέρων να ανήκει στην ίδια οικογένεια κατανομών με την εκ των προτέρων αλλά με προσαρμοσμένες παραμέτρους και έτσι να αποφύγουμε τον υπολογισμό δύσκολων ολοκληρωμάτων.

<u>Παράδειγμα</u>: Έστω ότι $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ είναι ένα τυχαίο δείγμα από κατανομή Poisson με άγνωστη παράμετρο λ, δηλαδή $f(y_i|\theta) = e^{-\theta}\cdot\frac{\theta^{y_i}}{y_i!}$, όπου έχουμε θέσει $\lambda = \theta$. Η κατανομή αυτή μπορεί να γραφτεί στην ακόλουθη μορφή

$$f(y_i|\theta) = \frac{1}{y_i!}\exp\{-\theta\}\exp\{\ln\theta\cdot y_i\},$$

από την οποία καταλαβαίνουμε ότι ανήκει στην Εκθετική Οικογένεια Κατανομών με

$$h(y) = \frac{1}{y!}$$
$$g(\theta) = \exp\{-\theta\}$$
$$\phi(\theta) = \ln\theta$$
$$u(y) = y.$$

Η <u>πιθανοφάνεια</u> υπολογίζεται ως ακολούθως

$$f(\mathbf{y}|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n}\left\{\frac{1}{y_i!}\right\}\{\exp(-\theta)\}^n \exp\left\{\ln(\theta)\sum_{i=1}^{n} y_i\right\}$$
$$\propto \{\exp(-\theta)\}^n \exp\left\{\ln(\theta)\sum_{i=1}^{n} y_i\right\},$$

την οποία αν γράψουμε στη μορφή $\theta^{(\sum_{i=1}^{n} y_i+1)-1}\exp(-n\theta)$ βλέπουμε ότι είναι η κατανομή $Gamma\left(\sum_{i=1}^{n} y_i+1, n\right)$.

Ορίζουμε ως εκ των προτέρων την εξής συνάρτηση

$$f(\theta) \propto \exp\{-\theta\}^{d} \exp\{\ln\theta \cdot b\} = \theta^{b} e^{-d\theta} = \theta^{(b+1)-1} e^{-d\theta},$$

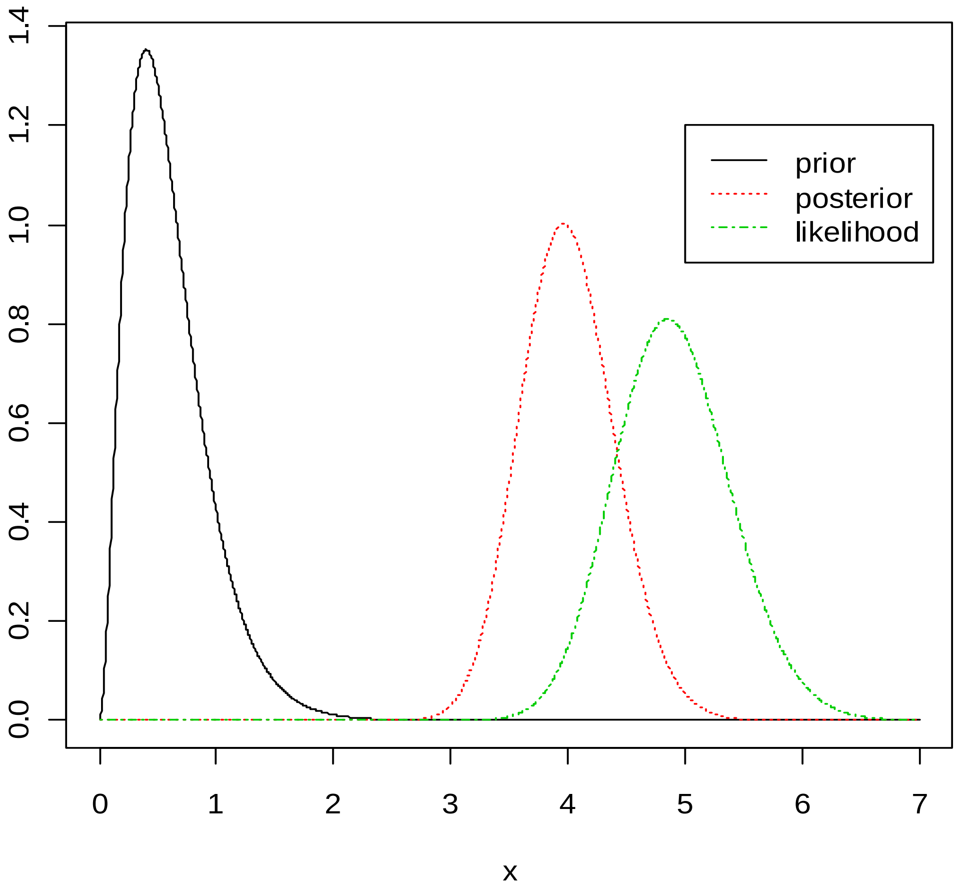
που όπως παρατηρούμε αποτελεί τον πυρήνα της κατανομής Gamma(b+1,d). Από την θεωρία λοιπόν προέκυψε ότι για το συγκεκριμένο πρόβλημα πρέπει να χρησιμοποιήσουμε για συζυγή prior μια κατανομή Gamma και όπως αναμένεται, η εκ των υστέρων κατανομή (posterior) θα είναι και αυτή Gamma με προσαρμοσμένες βέβαια παραμέτρους. Πράγματι, από το Θεώρημα Bayes έχουμε

$$\begin{aligned} f(\theta \mid \mathbf{y}) \propto f(\theta) f(\mathbf{y} \mid \theta) &= \{\exp(-\theta)\}^{n+d} \exp\left\{\ln(\theta)\left[\sum_{i=1}^{n} y_i + b\right]\right\} \\ &= \{\exp(-\theta)\}^{n+d} \cdot \theta^{\sum_{i=1}^{n} y_i + b} = \theta^{\left(\sum_{i=1}^{n} y_i + b + 1\right)-1} \cdot e^{-(n+d)\theta} \\ &= Gamma\left(\sum_{i=1}^{n} y_i + b + 1, n+d\right). \end{aligned}$$

<u>Αριθμητική Εφαρμογή</u>

Έχουμε ένα τυχαίο δείγμα από κατανομή Poisson με άγνωστη παράμετρο $\lambda = \theta$ $\mathbf{y} = [3,5,2,5,4,7,4,5,6,5,4,6,6,4,5,6,7,3,4,6]^{\mathrm{T}}$. Ας υποθέσουμε ότι οι εκ των προτέρων πεποιθήσεις μας για την παράμετρο θ αντιπροσωπεύονται ικανοποιητικά από την κατανομή $\text{Gamma}(3,5)$. Τότε, σύμφωνα με τα προηγούμενα, η εκ των υστέρων κατανομή της $\theta \mid \mathbf{y}$ θα είναι η $\text{Gamma}(100,25)$. Ο παρακάτω κώδικας στην R μας δίνει τα γραφήματα της εκ των προτέρων (prior) και της εκ των υστέρων (posterior) κατανομής.

```
y<-c(3,5,2,5,4,7,4,5,6,5,4,6,6,4,5,6,7,3,4,6)
n<-length(y)
s<-sum(y)
x<-seq(0,7,length=2000)
b<-2
d<-5
plot(x,dgamma(x,b+1,d),ylab="",col=1,lty=1,type="l")
lines(x,dgamma(x,s+b+1,n+d),col=2,lty=3)
lines(x,dgamma(x,s+1,n),col=3,lty=4)
legend(5,1.2,legend=c("prior","posterior","likelihood"),col=c(1,2,3),
lty=c(1,3,4))
```



2.4 Συνεχής αναθεώρηση

Είδαμε ότι το Θεώρημα του Bayes προσφέρει τον μηχανισμό με βάση τον οποίο οι εκ των προτέρων πληροφορίες μας αναθεωρούνται από τα δεδομένα και δίνουν την εκ των υστέρων πληροφορία. Όμως η εκ των προτέρων και η εκ των υστέρων κατανομή είναι έννοιες σχετικές, καθώς η κατανομή που είναι σήμερα εκ των υστέρων, αύριο μπορεί να είναι εκ των προτέρων. Αυτό γιατί το Θεώρημα του Bayes βρίσκει εφαρμογή και όταν έχουμε διαδοχική συλλογή δεδομένων. Έτσι προκύπτει η εξής ερώτηση: εάν πάρουμε μια σειρά από δεδομένα και αναθεωρήσουμε τις εκ των προτέρων πεποιθήσεις μας την στιγμή που λαμβάνουμε κάθε ένα από τα δεδομένα μας, θα πάρουμε διαφορετικό αποτέλεσμα από ότι θα παίρναμε αν περιμέναμε να φτάσουν στα χέρια μας όλα τα δεδομένα μαζί και μετά να αναθεωρήσουμε την εκ των προτέρων πληροφόρηση μας;

Ας υποθέσουμε ότι έχουμε δύο ανεξάρτητα δείγματα $\mathbf{y}_1$ και $\mathbf{y}_2$ από την ίδια σ.π.π. f. Θεωρούμε αρχικά μόνο το πρώτο δείγμα $\mathbf{y}_1$ και υπολογίζουμε την εκ των υστέρων κατανομή του $\boldsymbol{\theta} \mid \mathbf{y}_1$ από το Θεώρημα Bayes

$$f(\boldsymbol{\theta} \mid \mathbf{y}_1) \propto f(\boldsymbol{\theta}) f(\mathbf{y}_1 \mid \boldsymbol{\theta}).$$

Ύστερα λαμβάνουμε και το δεύτερο δείγμα $\mathbf{y}_2$ και υπολογίζουμε την εκ των υστέρων κατανομή του $\boldsymbol{\theta}|\mathbf{y}_2$ χρησιμοποιώντας όμως ως εκ των προτέρων την ήδη υπολογισμένη κατανομή $f(\boldsymbol{\theta}|\mathbf{y}_1)$

$$f(\boldsymbol{\theta}|\mathbf{y}_2) \propto f(\boldsymbol{\theta}|\mathbf{y}_1)f(\mathbf{y}_2|\boldsymbol{\theta}) \propto f(\boldsymbol{\theta})f(\mathbf{y}_1|\boldsymbol{\theta})f(\mathbf{y}_2|\boldsymbol{\theta}) = f(\boldsymbol{\theta})f(\mathbf{y}_1,\mathbf{y}_2|\boldsymbol{\theta}).$$

Θεωρούμε τώρα ότι έχουμε ταυτόχρονα και τα δύο δείγματα $\mathbf{y}_1$ και $\mathbf{y}_2$, (οπότε είναι σαν να έχουμε ένα ενιαίο δείγμα). Από το Θεώρημα Bayes υπολογίζουμε την εκ των υστέρων κατανομή του $\boldsymbol{\theta}|\mathbf{y}_1,\mathbf{y}_2$ και έχουμε

$$f(\boldsymbol{\theta}|\mathbf{y}_1,\mathbf{y}_2) \propto f(\boldsymbol{\theta})f(\mathbf{y}_1,\mathbf{y}_2|\boldsymbol{\theta}).$$

<u>Συμπέρασμα</u>: Η εκ των υστέρων κατανομή του $\boldsymbol{\theta}|\mathbf{y}_1,\mathbf{y}_2$ είναι η ίδια με την εκ των υστέρων κατανομή που λαμβάνουμε όταν θεωρήσουμε αρχικά το πρώτο δείγμα $\mathbf{y}_1$, βρούμε την εκ των υστέρων του $\boldsymbol{\theta}|\mathbf{y}_1$ και χρησιμοποιώντας αυτήν ως prior για τον υπολογισμό της εκ των υστέρων του $\boldsymbol{\theta}|\mathbf{y}_2$. Δηλαδή είτε διαδοχικά είτε ενιαία λάβουμε τα δύο δείγματα, καταλήγουμε στο ίδιο αποτέλεσμα.

Η παραπάνω διαδικασία, που αποτελεί βασικό πλεονέκτημα της Μπεϋζιανής Στατιστικής, γενικεύεται για οποιονδήποτε αριθμό ανεξάρτητων δειγμάτων και είναι ιδιαίτερα εύχρηστη για την βελτίωση της εκ των υστέρων πληροφορίας όταν τα δεδομένα συλλέγονται διαδοχικά με το πέρας του χρόνου.

2.5 Προβλεπτική κατανομή

Ως τώρα έχουμε επικεντρώσει την προσοχή μας στην εκτίμηση των παραμέτρων. Έχουμε δηλαδή, καθορίσει ένα μοντέλο πιθανότητας με σκοπό να περιγράψει την τυχαία διαδικασία με την οποία παρατηρούνται τα δεδομένα μας, και έχουμε ακόμα δείξει πώς η Μπεϋζιανή θεωρία συνδυάζει την πληροφορία την οποία παρέχει το δείγμα και την εκ των προτέρων πληροφόρηση η οποία χρησιμοποιείται στην εκτίμηση των παραμέτρων με την μορφή της εκ των υστέρων κατανομής. Συνήθως, ο λόγος για τον οποίο «προτείνονται» τα στατιστικά μοντέλα, είναι η διεξαγωγή προβλέψεων σχετικά με τις μελλοντικές τιμές της διαδικασίας. Την διαδικασία αυτή μπορούμε να την χειριστούμε πολύ καλύτερα στην Μπεϋζιανή Στατιστική απ' ότι στην κλασική θεωρία.

Ας υποθέσουμε λοιπόν ότι έχουμε ένα τυχαίο δείγμα (δεδομένα) $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ μίας τυχαίας μεταβλητής Y με συνάρτηση πιθανοφάνειας $f(\mathbf{y}|\boldsymbol{\theta})$ και θέλουμε να εξάγουμε συμπεράσματα σχετικά με την κατανομή της μελλοντικής τιμής y_{n+1} μέσα από την διαδικασία αυτή. Με μία εκ των προτέρων κατανομή $f(\boldsymbol{\theta})$, το Θεώρημα του Bayes οδηγεί σε μία εκ των υστέρων κατανομή $f(\boldsymbol{\theta}|\mathbf{y})$. Η συνάρτηση πρόβλεψης για το y_{n+1} δεδομένου του διανύσματος y θα είναι:

$$
\begin{aligned}
f(y_{n+1} \mid \mathbf{y}) &= \int_{\Theta} f(y_{n+1}, \boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \\
&= \int_{\Theta} f(y_{n+1} \mid \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \\
&= \int_{\Theta} f(y_{n+1} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta},
\end{aligned}
$$

με την τελευταία ισότητα να ισχύει επειδή όταν ξέρω το $\boldsymbol{\theta}$, που είναι η παράμετρος της κατανομής, τότε οι παρατηρήσεις $\mathbf{y} = \left[y_1, y_2, ..., y_n \right]^T$, δεν μου χρειάζονται αφού ξέρω όλη την κατανομή μέσω του $\boldsymbol{\theta}$ και άρα θα ισχύει $f(y_{n+1} \mid \boldsymbol{\theta}, \mathbf{y}) = f(y_{n+1} \mid \boldsymbol{\theta})$. Επομένως η συνάρτηση πρόβλεψης είναι ένα ολοκλήρωμα της πιθανοφάνειας (για μία και μόνο παρατήρηση) πολλαπλασιαζόμενο με την εκ των υστέρων κατανομή.

<u>Παράδειγμα</u>: Έστω τυχαίο δείγμα $\mathbf{y} = \left[y_1, y_2, ..., y_n \right]^T$ με $y_i = 0$ ή $y_i = 1$, δηλαδή

$$
y_i \sim bernoulli(\theta) \ , \ i = 1,...,n \ , \quad 0 < \theta < 1
$$

$$
f(y_i \mid \theta) = \begin{Bmatrix} \theta & , & y_i = 1 \\ 1-\theta & , & y_i = 0 \end{Bmatrix} = \theta^{y_i} (1-\theta)^{1-y_i}.
$$

Η συνάρτηση πιθανοφάνειας θα είναι

$$
f(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta) = \theta^{\sum_{i=1}^{n} y_i} \cdot (1-\theta)^{n - \sum_{i=1}^{n} y_i}.
$$

Θεωρούμε ότι η prior για το θ είναι η Βήτα(a_0, b_0) και άρα η σ.π.π. του θ θα είναι η

$$
f(\theta) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0) \cdot \Gamma(b_0)} \cdot \theta^{a_0 - 1} (1-\theta)^{b_0 - 1} \propto \theta^{a_0 - 1} (1-\theta)^{b_0 - 1}.
$$

Από το Θεώρημα Bayes υπολογίζουμε την εκ των υστέρων κατανομή του $\theta \mid \mathbf{y}$

$$
\begin{aligned}
f(\theta \mid \mathbf{y}) \propto f(\theta) f(\mathbf{y} \mid \theta) &\propto \theta^{a_0 - 1} (1-\theta)^{b_0 - 1} \times \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n - \sum_{i=1}^{n} y_i} \\
&= \theta^{(a_0 + \sum_{i=1}^{n} y_i) - 1} (1-\theta)^{(n - \sum_{i=1}^{n} y_i + b_0) - 1} \\
&= \text{Βήτα}\left(a, b\right),
\end{aligned}
$$

$$
\text{όπου} \quad a := a_0 + \sum_{i=1}^{n} y_i \quad \text{και} \quad b = n - \sum_{i=1}^{n} y_i + b_0.
$$

$$
f(y_{n+1} = 1 \mid \mathbf{y}) = \int_0^1 f(y_{n+1} = 1 \mid \theta) \cdot f(\theta \mid \mathbf{y}) d\theta = \int_0^1 \theta \cdot f(\theta \mid \mathbf{y}) d\theta = E\left[\theta \mid \mathbf{y}\right].
$$

Άρα η πιθανότητα η $n+1$- παρατήρηση να είναι '1' ισούται με τη μέση τιμή της εκ των υστέρων κατανομής του $\theta \mid y$, δηλαδή είναι ίση με

$$E[\theta \mid \mathbf{y}] = \frac{a}{a+b} = \frac{a_0 + \sum_{i=1}^{n} y_i}{a_0 + b_0 + n}.$$

2.6 Markov Chain Monte Carlo (MCMC)

Έστω ότι $f(\theta \mid y)$ είναι η σ.π.π. της τυχαίας μεταβλητής θ και ενδιαφερόμαστε γενικά να υπολογίσουμε τη μέση τιμή $E[h(\theta)]$, όπου h μια οποιαδήποτε συνάρτηση του θ. Τότε ως γνωστόν

$$E[h(\theta)] = \int_{\Theta} h(\theta) f(\theta \mid y)\, d\theta.$$

Την παραπάνω μέση τιμή, που σε πολλές περιπτώσεις δεν υπολογίζεται με αναλυτικό τρόπο, μπορούμε να την εκτιμήσουμε με την ***Monte Carlo εκτιμήτρια***:

$$\hat{E}[h(\theta)] = \frac{1}{n} \cdot \sum_{i=1}^{n} h(\theta_i^*),$$

όπου θ_i^*, $i = 1,...,n$, είναι τυχαίες προσομοιωμένες τιμές από την κατανομή $f(\theta \mid y)$. Συνήθως ενδιαφερόμαστε για την εκτίμηση της μέσης τιμής $E[\theta]$ και της διασποράς $V[\theta]$ της τ.μ θ. Στην πρώτη περίπτωση θέτουμε $h(\theta) = \theta$ οπότε παίρνουμε την εκτιμήτρια

$$\hat{E}[\theta] = \frac{1}{n} \cdot \sum_{i=1}^{n} \theta_i^* := \overline{\theta}^*,$$

ενώ στη δεύτερη περίπτωση θέτουμε $h(\theta) = (\theta - \overline{\theta}^*)^2$ και προκύπτει η εκτίμηση

$$\hat{V}[\theta] = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\theta_i^* - \overline{\theta}^*)^2.$$

Αποδεικνύεται ότι οι παραπάνω Monte Carlo εκτιμήτριες είναι συνεπείς και αμερόληπτές και για μεγάλο n συγκλίνουν στην ως προς εκτίμηση ποσότητα με αρκετά μεγάλη πιθανότητα, υπό την προϋπόθεση το δείγμα των προσομοιωμένων τιμών θ_i^*, $i = 1,...,n$ να είναι τυχαίο.

Από τα προηγούμενα προκύπτει το ερώτημα για τον τρόπο με τον οποίο μπορούμε να προσομοιώσουμε τυχαίες τιμές από μια κατανομή. Υπάρχουν διάφορες τεχνικές που εξυπηρετούν το σκοπό αυτό όπως η μέθοδος Απόρριψης που παράγει ένα IID δείγμα προσομοιωμένων τιμών. Στη συνέχεια, θα αναφερθούμε πολύ συνοπτικά στη γενική ιδέα των αλγορίθμων MCMC καθώς επίσης και σε δύο τέτοιους αλγορίθμους, τον Metropolis-Hastings και τον Gibbs Sampling όπου το προσομοιωμένο δείγμα συνιστά πλέον μια Μαρκοβιανή Αλυσίδα. Τα βασικά στοιχεία της θεωρίας των Στοχαστικών Ανελίξεων θεωρούνται γνωστά και δεν αναπτύσσονται στην παρούσα παράγραφο.

Έστω $\boldsymbol{\theta} \in \Theta$ η παράμετρος της κατανομής f και $\mathbf{y}$ τα δεδομένα. Οι Metropolis et al (1953) πρότειναν την ιδέα της προσομοίωσης τιμών από την εκ των υστέρων κατανομή $f(\boldsymbol{\theta} \mid \mathbf{y})$ στοχεύοντας στη δημιουργία μιας Μαρκοβιανής Αλυσίδας με τις εξής ιδιότητες:

1. Χώρο καταστάσεων Θ.
2. Στάσιμη κατανομή $f(\boldsymbol{\theta} \mid \mathbf{y})$.

Αν η Μαρκοβιανή Αλυσίδα είναι εργοδική (γνήσια επαναληπτική και απεριοδική) και η $f(\boldsymbol{\theta} \mid \mathbf{y})$ είναι η στάσιμη κατανομή της αλυσίδας, τότε μπορούμε να μάθουμε πληροφορίες όπως π.χ ο εκ των υστέρων μέσος απλά περιμένοντας να επιτευχθεί η στασιμότητα (να έχει συγκλίνει η αλυσίδα). Εν συνεχεία απλά καταγράφουμε τις τιμές μετά από αυτό το χρονικό διάστημα.

.

<u>Αλγόριθμος Metropolis - Hastings</u>

- Για $t = 0$, $\theta^{(0)}$, (αρχική τιμή).
- Για $t = 1,\ldots,n$, επανέλαβε :
- Προσομοίωσε $\theta^{*} \sim g(\theta \mid \theta^{(t)}, y)$. (***κατανομή εισήγησης***)
- Υπολόγισε $\alpha_{MH} = \min\left\{1,\ \dfrac{f(\theta^{*} \mid y) / g(\theta^{*} \mid \theta^{(t)}, y)}{f(\theta^{(t)} \mid y) / g(\theta^{(t)} \mid \theta^{*}, y)}\right\}$.(***πιθανότητα αποδοχής***)
- Θέσε $\theta^{(t+1)} = \theta^{*}$ με πιθανότητα α_{MH} και $\theta^{(t+1)} = \theta^{(t)}$ με πιθανότητα $1 - \alpha_{MH}$.

<u>Αλγόριθμος Gibbs Sampling</u>

Έστω $\boldsymbol{\theta} = \left[\theta_1, \theta_2, \ldots, \theta_p\right]^T$ το διάνυσμα των παραμέτρων

- Αρχικές τιμές $\boldsymbol{\theta}^{(0)} = \left[\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right]^T$, για $t = 0$.
- Σε κάθε επανάληψη ο αλγόριθμος κάνει την εξής διαδικασία

$$
\begin{array}{l}
\theta_1^{(t+1)} \sim f(\theta_1 \mid \theta_2^{(t)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)}, y) \\
\theta_2^{(t+1)} \sim f(\theta_2 \mid \theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)}, y) \\
\theta_3^{(t+1)} \sim f(\theta_3 \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \ldots, \theta_p^{(t)}, y) \\
\vdots \quad\ \vdots \quad \vdots \quad \vdots \qquad \vdots \qquad\quad \vdots \\
\theta_p^{(t+1)} \sim f(\theta_p \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \ldots, \theta_{p-1}^{(t+1)}, y),
\end{array}
$$

δηλαδή ως κατανομές εισήγησης χρησιμοποιεί τις πλήρους δέσμευσης εκ των υστέρων κατανομές των αντίστοιχων παραμέτρων (***full conditional***). Ένα άλλο χαρακτηριστικό του αλγορίθμου Gibbs είναι ότι η πιθανότητα αποδοχής είναι ένα, που σημαίνει πως κάθε τιμή που προσομοιώνεται γίνεται αποδεκτή (σε αντίθεση με τον αλγόριθμο Metropolis- Hastings).

ΚΕΦΑΛΑΙΟ 3

ΜΠΕΫΖΙΑΝΗ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ ΣΤΑ ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

3.1 Εισαγωγή

Η επιλογή μοντέλου είναι ένα σημαντικό στάδιο της Στατιστικής συμπερασματολογίας. Πιο συγκεκριμένα, ένα από τα ζητήματα που καλούμαστε να αντιμετωπίσουμε στα γενικευμένα γραμμικά μοντέλα, είναι και αυτό της επιλογής επεξηγηματικών μεταβλητών. Το πρόβλημα προκύπτει όταν ο αριθμός των επεξηγηματικών μεταβλητών, που έχουμε προς μελέτη, είναι αρκετά μεγάλος και ενδιαφερόμαστε να επιλέξουμε ένα υποσύνολο από αυτές. Αλλά και όταν ο αριθμός των επεξηγηματικών μεταβλητών δεν είναι αρκετά μεγάλος αντιμετωπίζουμε ανάλογο πρόβλημα εάν το ενδιαφέρον μας συγκεντρώνεται στην επιλογή των πιο "σημαντικών" μεταβλητών. Ένα μοντέλο με μικρότερο αριθμό επεξηγηματικών μεταβλητών, και συνεπώς και παραμέτρων, είναι πιο επιθυμητό, διότι είναι πιο οικονομικό.

Για να επιλέξουμε μεταξύ διαφορετικών μοντέλων μπορούμε να εφαρμόσουμε τόσο "υποκειμενικά", όσο και "αντικειμενικά" κριτήρια. Στα υποκειμενικά κριτήρια συγκαταλέγονται α) οικονομία μοντέλου (φειδωλότητα), β) πραγματικότητα γ) καλή προσαρμογή και δ) συμφωνία μεθόδων επιλογής. Ας δούμε καθ' ένα από τα κριτήρια αυτά χωριστά.

Οικονομία μοντέλου: Σύμφωνα με το κριτήριο αυτό επιλέγουμε εκείνο το μοντέλο που έχει τον μικρότερο αριθμό ανεξάρτητων μεταβλητών.

Πραγματικότητα: Ας υποθέσουμε ότι από προηγούμενες μελέτες ή από εμπειρία είναι γνωστό ότι ένας ή περισσότεροι παράγοντες θα πρέπει να συμπεριλαμβάνονται (ή να μην συμπεριλαμβάνονται) στο μοντέλο. Στην περίπτωση αυτή επιλέγουμε εκείνο το μοντέλο το οποίο περιέχει τους περισσότερους (ή τους λιγότερους) από τους παράγοντες αυτούς.

Καλή προσαρμογή: Πόσο καλά προσεγγίζει το θεωρητικό μοντέλο τα δεδομένα.

Συμφωνία μεθόδων: Σύμφωνα με το κριτήριο αυτό επιλέγουμε εκείνο το μοντέλο στο οποίο συμφωνούν οι περισσότερες από τις μεθόδους επιλογής. Σε μερικές περιπτώσεις τέτοιο μοντέλο μπορεί να μην υπάρχει.

Σε αυτό το κεφάλαιο εξετάζουμε την επιλογή μοντέλου και μεταβλητών από την Μπεϋζιανή σκοπιά. Αυτό σημαίνει ότι το διάνυσμα των παραμέτρων $\boldsymbol{\theta}$ του μοντέλου θα είναι τυχαία μεταβλητή. Επειδή το υπό μελέτη πρόβλημα αφορά τα γενικευμένα γραμμικά μοντέλα

$$g(\mu_i) = \eta_i = \sum_{j=0}^{p} x_j^{(i)} \beta_j \quad , \quad i = 1, \ldots, n$$

ένα μέρος του διανύσματος $\boldsymbol{\theta}$ θα αποτελούν οι συντελεστές β_j, οι οποίοι κατά συνέπεια θεωρούνται και αυτοί τυχαίες μεταβλητές. Για παράδειγμα, στο Κανονικό γραμμικό μοντέλο όπου $\mathbf{Y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, το διάνυσμα των παραμέτρων είναι το $\boldsymbol{\theta}^T = \left[\boldsymbol{\beta}^T, \sigma^2\right]$. Όταν οι επεξηγηματικές μεταβλητές ενσωματώνονται στον φορμαλισμό του μοντέλου m, το ενδιαφέρον συνήθως εστιάζεται στον υπολογισμό της posterior κατανομής $f(\boldsymbol{\beta}|\mathbf{y}, m)$ παρά της $f(\boldsymbol{\theta}|\mathbf{y}, m)$.

Σε προβλήματα επιλογής μεταβλητών, μπορούμε να αντικαταστήσουμε τον δείκτη m για ένα μοντέλο με τον δείκτη-διάνυσμα $\boldsymbol{\gamma} = \left[\gamma_0, \gamma_1, ..., \gamma_p\right]^T \in \{0,1\}^p$, όπου ο δείκτης $\boldsymbol{\gamma}$ αντιπροσωπεύει ποιες από τις p μεταβλητές $X_0, X_1, ..., X_p$, περιλαμβάνονται στο μοντέλο. Για τις συνιστώσες του διανύσματος $\boldsymbol{\gamma}$ ισχύει

$$\gamma_j = \begin{cases} 1\,, & \text{αν η } X_j \text{ περιλαμβάνεται στο μοντέλο } \boldsymbol{\gamma} \\ 0\,, & \text{διαφορετικά} \end{cases}.$$

Το λανθάνων διάνυσμα των δυαδικών δεικτών $\boldsymbol{\gamma}$, το εισήγαγαν οι George και McCulloch (1993) στην πρώτη προσπάθειά τους να χρησιμοποιήσουν MCMC αλγόριθμους σε προβλήματα επιλογής μοντέλου.

3.2 Εκ των υστέρων λόγος πιθανοτήτων και ο παράγοντας Bayes

Η κατά Bayes σύγκριση μοντέλων βασίζεται στον υπολογισμό των περιθώριων πιθανοφανειών των παρατηρούμενων δεδομένων των υπο σύγκριση μοντέλων. Μέσω αυτών των δεσμευμένων πιθανοτήτων μπορούν να υπολογιστούν οι παράγοντες Bayes (Bayes Factors) και οι εκ των υστέρων λόγοι σχετικών πιθανοτήτων (Posterior Odds) μεταξύ των συγκρινόμενων μοντέλων.

Ας θεωρήσουμε δύο ανταγωνιστικά μοντέλα m_0 και m_1 και ας υποθέσουμε ότι έχουμε παρατηρήσει κάποια δεδομένα $\mathbf{y}$ τα οποία έχουν παραχθεί από ένα εκ των δύο αυτών μοντέλων. Κάθε ένα από τα μοντέλα $m \in \{m_0, m_1\}$ καθορίζει την κατανομή της τυχαίας μεταβλητής Y, $f(\mathbf{y}|\boldsymbol{\theta}_{(m)}, m)$, όπου με $\boldsymbol{\theta}_{(m)}$ συμβολίζουμε το άγνωστο διάνυσμα παραμέτρων που αντιστοιχεί στο μοντέλο m. Αν $f(m)$ είναι η εκ των προτέρων πιθανότητα του μοντέλου m τότε, χρησιμοποιώντας το Θεώρημα Bayes, η εκ των υστέρων πιθανότητα του αντίστοιχου μοντέλου θα είναι

$$f(m|\mathbf{y}) = \frac{f(\mathbf{y}|m) \cdot f(m)}{f(\mathbf{y}|m_0) \cdot f(m_0) + f(\mathbf{y}|m_1) \cdot f(m_1)} \qquad (3.1.1)$$

όπου $m \in \{m_0, m_1\}$ και $f(m_0) + f(m_1) = 1$. Έχουμε λοιπόν συνοπτικά τις παρακάτω έννοιες

$f(m)$: ***Εκ των προτέρων πιθανότητα του μοντέλου*** m (Prior model probability).

$f(m|\mathbf{y})$: ***Εκ των υστέρων πιθανότητα του μοντέλου*** m (Posterior model probability).

$f(\mathbf{y} \mid m)$ **:** ***Περιθωριακή πιθανοφάνεια των δεδομένων στο μοντέλο*** m (marginal likelihood of model m).

Η περιθωριακή πιθανοφάνεια των δεδομένων στο μοντέλο m δίνεται από τη σχέση

$$f(\mathbf{y} \mid m) = \int_{\Theta} f(\mathbf{y} \mid \boldsymbol{\theta}_{(m)}, m) \cdot f(\boldsymbol{\theta}_{(m)} \mid m)\, d\boldsymbol{\theta}_{(m)},$$

όπου $f(\boldsymbol{\theta}_{(m)} \mid m)$ είναι η εκ των προτέρων κατανομή των παραμέτρων στο μοντέλο m.

Μια πολύ χρήσιμη ποσότητα για τη σύγκριση των δύο μοντέλων αποτελεί ο λόγος της εκ των υστέρων πιθανότητας του μοντέλου m_0 προς την εκ των υστέρων πιθανότητα του μοντέλου m_1 που ονομάζεται εκ των υστέρων λόγος πιθανοτήτων του μοντέλου m_0 έναντι του μοντέλου m_1 και συμβολίζεται με PO_{01} (Posterior Odds). Από τη σχέση (3.1.1) έχουμε

$$PO_{01} := \frac{f(m_0 \mid \mathbf{y})}{f(m_1 \mid \mathbf{y})} = \frac{\dfrac{f(\mathbf{y} \mid m_0) \cdot f(m_0)}{f(\mathbf{y} \mid m_0) \cdot f(m_0) + f(\mathbf{y} \mid m_1) \cdot f(m_1)}}{\dfrac{f(\mathbf{y} \mid m_1) \cdot f(m_1)}{f(\mathbf{y} \mid m_0) \cdot f(m_0) + f(\mathbf{y} \mid m_1) \cdot f(m_1)}} \Rightarrow$$

$$PO_{01} = \frac{f(\mathbf{y} \mid m_0) \cdot f(m_0)}{f(\mathbf{y} \mid m_1) \cdot f(m_1)} \Rightarrow$$

$$PO_{01} = \frac{f(\mathbf{y} \mid m_0)}{f(\mathbf{y} \mid m_1)} \times \frac{f(m_0)}{f(m_1)}.$$

Σύμφωνα με την τελευταία σχέση έχουμε τις ακόλουθες έννοιες

$\dfrac{f(m_0)}{f(m_1)}$ **:** Εκ των προτέρων λόγος πιθανοτήτων του μοντέλου m_0 έναντι του μοντέλου m_1 (Prior model odds).

$\dfrac{f(\mathbf{y} \mid m_0)}{f(\mathbf{y} \mid m_1)} := B_{01}$ **:** ***Παράγοντας Bayes*** του μοντέλου m_0 έναντι του μοντέλου m_1 (Bayes Factor).

$\dfrac{f(m_0 \mid \mathbf{y})}{f(m_1 \mid \mathbf{y})} := PO_{01}$ **:** ***Εκ των υστέρων λόγος πιθανοτήτων του μοντέλου*** m_0 ***έναντι του μοντέλου*** m_1 (Posterior model odds).

Αποδείξαμε λοιπόν ότι τα παραπάνω τρία μεγέθη συνδέονται με την εξής σχέση

Posterior model odds = Bayes Factor × Prior model odds.

Η σχέση (3.1.1) μπορεί να γενικευτεί και στην περίπτωση που έχουμε περισσότερα από δύο ανταγωνιστικά μοντέλα. Θεωρώντας ένα σύνολο μοντέλων $M=\{m_1,m_2,...,m_{|M|}\}$, η εκ των υστέρων πιθανότητα του μοντέλου $m \in M$ υπολογίζεται ως εξής

$$
\begin{aligned}
f(m\,|\,\mathbf{y}) &= \frac{f(\mathbf{y}\,|\,m)\cdot f(m)}{\sum_{k=1}^{|M|} f(\mathbf{y}\,|\,m_k)\cdot f(m_k)} = \\
&\frac{1}{\dfrac{\sum_{k=1}^{|M|} f(\mathbf{y}\,|\,m_k)\cdot f(m_k)}{f(\mathbf{y}\,|\,m)\cdot f(m)}} = \\
&\frac{1}{\sum_{k=1}^{|M|}\left(\dfrac{f(\mathbf{y}\,|\,m_k)}{f(\mathbf{y}\,|\,m)}\times\dfrac{f(m_k)}{f(m)}\right)} = \\
&\left(\sum_{k=1}^{|M|} PO_{m_k m}\right)^{-1},
\end{aligned}
$$

όπου $|M|$ είναι το πλήθος των μοντέλων που απαρτίζουν το σύνολο M.

Ο παράγοντας Bayes αποτελεί ένα μέτρο για την βαρύτητα της πληροφορίας, η οποία περιλαμβάνεται στα δεδομένα, υπέρ ενός μοντέλου έναντι ενός άλλου μοντέλου. Ο παράγοντας αυτός μπορεί να χρησιμοποιηθεί ως σχετικό μέτρο για τη σύγκριση δύο μοντέλων. Ερμηνείες του παράγοντα Bayes, σύμφωνα με τους Kass και Raftery, δίνονται στους Πίνακες 1 και 2.

Πίνακας 1: Ερμηνεία του παράγοντα Bayes σύμφωνα με τους Kass και Raftery (log_{10})

$\log_{10}(B_{10})$	B_{10}	ΕΝΔΕΙΞΗ ΕΝΑΝΤΙ ΤΟΥ ΜΟΝΤΕΛΟΥ m_0
0.0 έως 0.5	1.0 έως 3.2	Όχι άξια αναφοράς
0.5 έως 1.0	3.2 έως 10	Ουσιαστική
1.0 έως 2.0	10 έως 100	Ισχυρή
>2	>100	Αποφασιστική

Πίνακας 2: Ερμηνεία του παράγοντα Bayes σύμφωνα με τους Kass και Raftery (φυσικός λογάριθμος)

$\ln(B_{10})$	B_{10}	ΕΝΔΕΙΞΗ ΕΝΑΝΤΙ ΤΟΥ ΜΟΝΤΕΛΟΥ m_0
0 έως 1	1 έως 3	Όχι άξια αναφοράς
1 έως 3	3 έως 20	Θετική
3 έως 5	20 έως 150	Ισχυρή
>5	>150	Πολύ ισχυρή

Παρατήρηση: Εάν τα δύο μοντέλα έχουν την ίδια εκ των προτέρων πιθανότητα, δηλαδή αν $f(m_0) = f(m_1)$, τότε

$$PO_{01} = \frac{f(\mathbf{y} \mid m_0)}{f(\mathbf{y} \mid m_1)} \times \frac{f(m_0)}{f(m_1)} = \frac{f(\mathbf{y} \mid m_0)}{f(\mathbf{y} \mid m_1)} = B_{01},$$

που σημαίνει ότι στην ειδική αυτή περίπτωση ισχύει Posterior model odds = Bayes Factor.

Παρ'όλο που ο παράγοντας Bayes αποτελεί ένα σημαντικό εργαλείο για την επιλογή του καταλληλότερου μοντέλου, θα διαπιστώσουμε στην επόμενη παράγραφο ότι επηρεάζεται πολύ από την μεταβλητότητα της εκ των προτέρων κατανομής των παραμέτρων.

3.3 Το παράδοξο του Lindley

Η επιλογή των εκ των προτέρων κατανομών είναι καίριας σημασίας για τις εκ των υστέρων πιθανότητες των μοντέλων και πρέπει να γίνεται με πολύ μεγάλη προσοχή. Οι εκ των υστέρων πιθανότητες είναι πολύ ευαίσθητες στο βαθμό των εκ των προτέρων διασπορών έχοντας μια τάση να ευνοούν τα μοντέλα με πιο απλή δομή καθώς οι εκ των προτέρων διασπορές αυξάνουν. Πιο συγκεκριμένα, ο Lindley (1957), παρατήρησε ότι όταν το μέγεθος του δείγματος n αυξάνεται, επίσης αυξάνεται και ο λόγος των εκ των υστέρων συμπληρωματικών πιθανοτήτων, γεγονός που οδηγεί σε παράδοξο, αφού για μεγάλα δείγματα αυξάνεται και η εκ των υστέρων πιθανότητα του μοντέλου της αρχικής υπόθεσης. Ο Bartlett (1957) παρατήρησε ότι το παράδοξο του Lindley σχετίζεται και με τη διασπορά της εκ των προτέρων κατανομής, συγκεκριμένα όσο μεγαλύτερη είναι η διασπορά της, τόσο αυξάνεται και ο παράγοντας Bayes. Με άλλα λόγια επιλέγοντας εκ των προτέρων κατανομές με μεγάλη διασπορά καταλήγουμε σε ένδειξη υπέρ του μοντέλου της αρχικής υπόθεσης. Επιλέγουμε μεγάλες διασπορές στις εκ των προτέρων κατανομές, όταν θέλουμε να δείξουμε τη μη επαρκή γνώση για μια παράμετρο.

Έστω ότι έχουμε ένα τυχαίο δείγμα από κανονική κατανομή $y_i \sim N(\theta, \sigma^2)$, $i = 1, \ldots, n$, με γνωστή διασπορά σ^2 και θέλουμε να ελέγξουμε την υπόθεση $H_0 : \theta = \theta_0$ (μοντέλο m_0) έναντι της υπόθεσης $H_1 : \theta \neq \theta_0$ (μοντέλο m_1). Υποθέτουμε ότι η εκ των προτέρων πεποίθησή μας για την άγνωστη παράμετρο θ μπορεί να αντιπροσωπευτεί ικανοποιητικά από μια κανονική κατανομή, $\theta \sim N(\theta_1, \sigma_1^2)$ όπου τα θ_1 και σ_1^2 είναι γνωστά. Οι περιθώριες πιθανοφάνειες των δύο μοντέλων είναι οι εξής:

$$\text{Μοντέλο } m_0 \textbf{:}\ f(\mathbf{y} \mid m_0) = \prod_{i=1}^{n} N(y_i \mid \theta_0, \sigma^2).$$

$$\text{Μοντέλο } m_1 \textbf{:}\ f(\mathbf{y} \mid m_1) = \int_{\Theta} \prod_{i=1}^{n} N(y_i \mid \theta, \sigma^2) \times N(\theta \mid \theta_1, \sigma_1^2) d\theta .$$

Υπολογίζουμε τώρα τον παράγοντα Bayes (Bayes Factor) και παίρνουμε το ακόλουθο αποτέλεσμα

$$B_{01} = \frac{f(\mathbf{y}\,|\,m_0)}{f(\mathbf{y}\,|\,m_1)} = \frac{N(\bar{\mathbf{y}}\,|\,\theta_0,\, \sigma^2/n)}{\int_\Theta N(\bar{\mathbf{y}}\,|\,\theta,\, \sigma^2/n) \times N(\theta\,|\,\theta_1, \sigma_1^2)\, d\theta}$$

$$= \left(\frac{\sigma_1^{-2} + n/\sigma^2}{\sigma_1^{-2}} \right)^{\frac{1}{2}} \cdot \frac{\exp\left\{ \frac{1}{2} \left(\sigma_1^2 + \sigma^2/n \right)^{-1} \cdot (\bar{\mathbf{y}} - \theta_1)^2 \right\}}{\exp\left\{ \frac{1}{2} \cdot \frac{n}{\sigma^2} \cdot (\bar{\mathbf{y}} - \theta_0)^2 \right\}},$$

που δείχνει ότι για δεδομένη τιμή του $\bar{\mathbf{y}}$, (δηλαδή για δεδομένες παρατηρήσεις y_i , $i = 1,...,n$), έχουμε ότι $B_{01} \to +\infty$, όταν $\sigma_1^2 \to +\infty$ που σημαίνει πως όταν η prior κατανομή της παραμέτρου θ έχει πολύ μεγάλη διασπορά σ_1^2 (πολύ μεγαλύτερη από αυτήν του δείγματος σ^2), τότε η Μπεϋζιανή ανάλυση οδηγεί στην υποστήριξη του απλούστερου μοντέλου m_0.

3.4 Τρόποι υπολογισμού της περιθώριας πιθανοφάνειας (marginal likelihood)

Όπως θα δούμε στο επόμενο κεφάλαιο, στα Κανονικά γραμμικά μοντέλα και με χρήση συζυγών εκ των προτέρων κατανομών, μπορούμε να έχουμε αναλυτικά αποτελέσματα για τις αντίστοιχες εκ των υστέρων κατανομές. Το γεγονός αυτό όμως αποτελεί μια ειδική περίπτωση καθώς σε πολλές περιπτώσεις, τα ολοκληρώματα που πρέπει να επιλυθούν για τον υπολογισμό των εκ των υστέρων πιθανοτήτων των μοντέλων και του παράγοντα Bayes πολλές φορές παρουσιάζουν μεγάλη δυσκολία. Πιθανόν να μην είναι δυνατή η χρήση συζυγών εκ των προτέρων κατανομών, οπότε στις περιπτώσεις αυτές χρησιμοποιούνται ασυμπτωτικές προσεγγίσεις, όπως η μέθοδος Laplace, ή χρησιμοποιούνται τεχνικές βασισμένες σε μεθόδους Markov Chain Monte Carlo (MCMC). Στην παράγραφο αυτή αναφέρονται ορισμένες μέθοδοι εκτίμησης της περιθώριας πιθανοφάνειας $f(\mathbf{y}\,|\,m)$ που όπως έχουμε δει υπεισέρχεται στον υπολογισμό του παράγοντα Bayes.

<u>Προσέγγιση Laplace</u>

Η πιο δημοφιλής προσέγγιση της περιθώριας πιθανοφάνειας είναι η προσέγγιση Laplace που χρησιμοποιήθηκε από τους Tierney και Kadame (1986), Tierney et al (1989) και Erkanli (1984)

$$f(\mathbf{y}\,|\,m) \approx (2\pi)^{d(m)/2} \left| I_{\tilde{\boldsymbol{\theta}}_{(m)}} \right|^{\frac{1}{2}} f(\mathbf{y}\,|\,\tilde{\boldsymbol{\theta}}_{(m)}, m) \cdot f(\tilde{\boldsymbol{\theta}}_{(m)}\,|\,m), \quad (3.4.1)$$

όπου $d(m)$ είναι η διάσταση του μοντέλου m , $I_{\tilde{\boldsymbol{\theta}}_{(m)}}$ είναι ο αρνητικός αντίστροφος του Εσσιανού πίνακα (Hessian matrix) των δευτέρων παραγώγων της log-posterior κατανομής $f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y}, m)$ υπολογισμένος στην posterior κορυφή $\tilde{\boldsymbol{\theta}}_{(m)}$ της κατανομής αυτής. Η $f(\mathbf{y}\,|\,\tilde{\boldsymbol{\theta}}_{(m)}, m)$ και η $f(\tilde{\boldsymbol{\theta}}_{(m)}\,|\,m)$ είναι η πιθανοφάνεια του μοντέλου και η prior κατανομή της διανυσματικής παραμέτρου $\boldsymbol{\theta}_{(m)}$ αντίστοιχα υπολογισμένες και αυτές στην εκ των υστέρων κορυφή $\tilde{\boldsymbol{\theta}}_{(m)}$.

Laplace - Metropolis εκτιμητής

Για την αποφυγή αναλυτικών υπολογισμών των ποσοτήτων $\tilde{\theta}_{(m)}$ και $I_{\tilde{\theta}_{(m)}}$ που υπεισέρχονται στη σχέση (3.4.1), μπορούμε να εκτιμήσουμε αυτές τις ποσότητες με χρήση ενός MCMC προσομοιωμένου δείγματος από τον μέσο $\bar{\theta}_{(m)}$ και τον πίνακα διασποράς-συνδιασποράς $\hat{\Sigma}$ των τιμών του δείγματος. Το προσομοιωμένο δείγμα το λαμβάνουμε από την εκ των υστέρων κατανομή των παραμέτρων $f(\boldsymbol{\theta}_{(m)} \mid \mathbf{y}, m)$. Η διαδικασία αυτή, που συνδυάζει τα ασυμπτωτικά αποτελέσματα της προσέγγισης Laplace με ένα MCMC δείγμα, αναφέρεται ως Laplace-Metropolis εκτιμητής (Lewis και Raftery, 1997) και δουλεύει αποδοτικά όταν η κατανομή $f(\mathbf{y} \mid m)$ είναι συμμετρική και κατά συνέπεια δεν αποτελεί πρόβλημα η αντικατάσταση της κορυφής από τον μέσο. Έχουμε λοιπόν την ακόλουθη εκτίμηση

$$\hat{f}_{LM}(\mathbf{y} \mid m) \approx (2\pi)^{d(m)/2} \left|\hat{\boldsymbol{\Sigma}}\right|^{\frac{1}{2}} f(\mathbf{y} \mid \bar{\boldsymbol{\theta}}_{(m)}, m) \cdot f(\bar{\boldsymbol{\theta}}_{(m)} \mid m),$$

όπου

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{t=1}^{N} \left(\boldsymbol{\theta}_{(m)}^{(t)} - \bar{\boldsymbol{\theta}}\right) \cdot \left(\boldsymbol{\theta}_{(m)}^{(t)} - \bar{\boldsymbol{\theta}}\right)^{T},$$

είναι ο πίνακας διασποράς-συνδιασποράς του δείγματος $\left\{\theta_{(m)}^{(t)},\ t = 1, 2, \ldots, N\right\}$.

Άμεση Monte Carlo εκτίμηση

Μέθοδοι Monte Carlo μπορούν να χρησιμοποιηθούν για την εκτίμηση της περιθωριακής πιθανοφάνειας $f(\mathbf{y} \mid m)$. Η πιο απλή διαδικασία είναι η παραγωγή ενός προσομοιωμένου δείγματος $\left\{\theta_{(m)}^{(t)},\ t = 1, 2, \ldots, N\right\}$ από την <u>εκ των προτέρων</u> κατανομή $f(\theta_{(m)} \mid m)$ και στη συνέχεια η εκτίμηση της ζητούμενης πιθανοφάνειας

$$f(\mathbf{y} \mid m) = \int_{\Theta} f(\mathbf{y} \mid \boldsymbol{\theta}, m) f(\boldsymbol{\theta} \mid m) d\boldsymbol{\theta},$$

από την ποσότητα

$$\hat{f}(\mathbf{y} \mid m) = \frac{1}{N} \cdot \sum_{t=1}^{N} f(\mathbf{y} \mid \theta_{(m)}^{(t)}, m),$$

που είναι η στάθμιση της πιθανοφάνειας για κάθε τιμή του δείγματος. Ο παραπάνω εκτιμητής είναι αρκετά ασταθής ειδικά όταν η εκ των προτέρων κατανομή της παραμέτρου $f(\boldsymbol{\theta}_{(m)} \mid m)$ διαφέρει πολύ από την εκ των υστέρων $f(\boldsymbol{\theta}_{(m)} \mid \mathbf{y}, m)$ (για παράδειγμα όταν ως prior έχουμε επιλέξει μια επίπεδη κατανομή) ή όταν η πιθανοφάνεια έχει πολύ μικρότερη διασπορά από την prior. Επιπλέον, η διασπορά του συγκεκριμένου εκτιμητή μπορεί να είναι μεγάλη με συνέπεια η σύγκλιση στην πραγματική τιμή να γίνεται με πολύ αργό ρυθμό.

Εκτιμητής Αρμονικού Μέσου

Σε αντίθεση με τη μέθοδο της άμεσης Monte Carlo εκτίμησης της κατανομής $f(\mathbf{y}\,|\,m)$, στη μέθοδο του Αρμονικού μέσου (Newton και Raftery, 1994) η προσομοίωση των τιμών $\left\{\theta_{(m)}^{(t)},\ t=1,2,...,N\right\}$ γίνεται από την εκ των υστέρων κατανομή $f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m)$. Έχουμε το ακόλουθο αποτέλεσμα

$$\begin{aligned}
\frac{1}{f(\mathbf{y}\,|\,m)} &= \frac{1}{f(\mathbf{y}\,|\,m)}\cdot\int_{\Theta}\frac{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)}\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)\,d\boldsymbol{\theta}_{(m)}\\
&=\int_{\Theta}\frac{1}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)}\cdot\frac{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}{f(\mathbf{y}\,|\,m)}\,d\boldsymbol{\theta}_{(m)}\\
&\overset{\theta.Bayes}{=}\int_{\Theta}\frac{1}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)}\cdot f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m)\,d\boldsymbol{\theta}_{(m)}\\
&=E_{\boldsymbol{\theta}_{(m)}|\mathbf{y},m}\left[\frac{1}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)}\right].
\end{aligned}$$

Βασιζόμενοι στην παραπάνω σχέση, μπορούμε να εκτιμήσουμε την περιθώρια πιθανοφάνεια $f(\mathbf{y}\,|\,m)$ από την εξής ποσότητα

$$\hat{f}_{HM}(\mathbf{y}\,|\,m)=\left[\frac{1}{N}\cdot\sum_{t=1}^{N}\left\{f(\mathbf{y}\,|\,\theta_{(m)}^{(t)},m)\right\}^{-1}\right]^{-1}.$$

Πρέπει να τονίσουμε ότι ο παραπάνω εκτιμητής είναι αρκετά ασταθής, ειδικά με τιμές που δίνουν μικρή πιθανοφάνεια. Ο εκτιμητής του Αρμονικού μέσου μπορεί να επεκταθεί στον ***γενικευμένο εκτιμητή Αρμονικού μέσου*** θεωρώντας μια γενική συνάρτηση πυκνότητας των παραμέτρων $g(\boldsymbol{\theta}_{(m)})$ και παρατηρώντας ότι ισχύει το ακόλουθο

$$\begin{aligned}
\frac{1}{f(\mathbf{y}\,|\,m)} &= \int_{\Theta}\frac{1}{f(\mathbf{y}\,|\,m)}\cdot g(\boldsymbol{\theta}_{(m)})\,d\boldsymbol{\theta}_{(m)}\\
&=\int_{\Theta}\frac{1}{f(\mathbf{y}\,|\,m)}\cdot g(\boldsymbol{\theta}_{(m)})\,d\boldsymbol{\theta}_{(m)}\\
&\overset{\theta.Bayes}{=}\int_{\Theta}\frac{f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m)}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}\cdot g(\boldsymbol{\theta}_{(m)})\,d\boldsymbol{\theta}_{(m)}\\
&=\int_{\Theta}\frac{g(\boldsymbol{\theta}_{(m)})}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}\cdot f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m)\,d\boldsymbol{\theta}_{(m)}\\
&=E_{\boldsymbol{\theta}_{(m)}|\mathbf{y},m}\left[\frac{g(\boldsymbol{\theta}_{(m)})}{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}\right].
\end{aligned}$$

Ως εκ τούτου, μπορούμε να προσομοιώσουμε τιμές $\left\{\theta_{(m)}^{(t)}\ ,\ t=1,2,...,N\right\}$ από την εκ των υστέρων κατανομή $f(\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m)$ και να εκτιμήσουμε την $f(\mathbf{y}\,|\,m)$ από τον εκτιμητή

$$\hat{f}_{GHM}(\mathbf{y}\,|\,m)=\left[\frac{1}{N}\cdot\sum_{t=1}^{N}\frac{g(\theta_{(m)}^{(t)})}{f(\mathbf{y}\,|\,\theta_{(m)}^{(t)},m)\cdot f(\theta_{(m)}^{(t)}\,|\,m)}\right]^{-1}.$$

Όσον αφορά την επιλογή της συνάρτησης πυκνότητας g, αυτή συνήθως είναι μια πολυμεταβλητή Κανονική ή μια Student κατανομή με μέση τιμή και διασπορά ίδια με την εκ των υστέρων μέση τιμή και διασπορά της $\boldsymbol{\theta}_{(m)}\,|\,\mathbf{y},m$ εκτιμώμενη από ένα MCMC δείγμα. Να σημειώσουμε τέλος ότι στην ειδική περίπτωση όπου $g(\boldsymbol{\theta}_{(m)})=f(\boldsymbol{\theta}_{(m)}\,|\,m)$, αναγόμαστε στην περίπτωση του απλού Αρμονικού μέσου.

Δειγματολήπτης Σπουδαιότητας

Μια περισσότερο ακριβής Monte Carlo εκτίμηση γίνεται με την τεχνική του δειγματολήπτη σπουδαιότητας (*Importance sampling)*. Στη μέθοδο αυτή προσομοιώνουμε τιμές $\left\{\theta_{(m)}^{(t)}\ ,\ t=1,2,...,N\right\}$ από μια αυθαίρετη κατανομή $g(\theta_{(m)})$. Ισχύει το παρακάτω αποτέλεσμα

$$\begin{aligned}
f(\mathbf{y}\,|\,m) &= \int_{\Theta} f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)\,d\boldsymbol{\theta}_{(m)} = \\
&\int_{\Theta}\frac{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}{g(\boldsymbol{\theta}_{(m)})}\cdot g(\boldsymbol{\theta}_{(m)})\,d\boldsymbol{\theta}_{(m)} = \\
&E_g\left[\frac{f(\mathbf{y}\,|\,\boldsymbol{\theta}_{(m)},m)\cdot f(\boldsymbol{\theta}_{(m)}\,|\,m)}{g(\boldsymbol{\theta}_{(m)})}\right].
\end{aligned}$$

Ως εκ τούτου, μπορούμε να εκτιμήσουμε την κατανομή $f(\mathbf{y}\,|\,m)$ από την ποσότητα

$$\hat{f}_{IS}(\mathbf{y}\,|\,m)=\frac{1}{N}\cdot\sum_{t=1}^{N}\frac{f(\mathbf{y}\,|\,\theta_{(m)}^{(t)},m)\cdot f(\theta_{(m)}^{(t)}\,|\,m)}{g(\theta_{(m)}^{(t)})}.$$

Όταν η κατανομή g είναι γνωστή μέχρι μια σταθερά, δηλαδή $g(\theta_{(m)})=C\cdot g^{*}(\theta_{(m)})$, τότε έχουμε

$$\hat{f}_{IS}(\mathbf{y}\,|\,m)=\frac{\sum_{t=1}^{N}\left\{ f(\mathbf{y}\,|\,\theta_{(m)}^{(t)},m)\cdot f(\theta_{(m)}^{(t)}\,|\,m) \big/ g^{*}(\theta_{(m)}^{(t)})\right\}}{\sum_{t=1}^{N}\left\{ f(\theta_{(m)}^{(t)}\,|\,m) \big/ g^{*}(\theta_{(m)}^{(t)})\right\}} \quad , \ \theta_{(m)}^{(t)}\sim g(\theta_{(m)}), t=1,2,...,N$$

αφού ισχύει το εξής

$$C=\int_{\Theta} C\cdot f(\theta_{(m)}\mid m)\,d\theta_{(m)} \overset{C=g/g^*}{=} \int_{\Theta} \frac{f(\boldsymbol{\theta}_{(m)}\mid m)}{g^*(\boldsymbol{\theta}_{(m)})}\cdot g(\boldsymbol{\theta}_{(m)})\,d\boldsymbol{\theta}_{(m)} = E_g\left[\frac{f(\boldsymbol{\theta}_{(m)}\mid m)}{g^*(\boldsymbol{\theta}_{(m)})}\right].$$

Ειδικές περιπτώσεις:

(i) Αν $g(\boldsymbol{\theta}_{(m)})=f(\boldsymbol{\theta}_{(m)}\mid m)$, τότε έχουμε την περίπτωση της Άμεσης Monte Carlo εκτίμησης.

(ii) Αν $g(\boldsymbol{\theta}_{(m)})=f(\boldsymbol{\theta}_{(m)}\mid \mathbf{y},m)$, τότε έχουμε την περίπτωση του Αρμονικού μέσου.

(iii) Αν $g(\boldsymbol{\theta}_{(m)})=\omega\cdot f(\boldsymbol{\theta}_{(m)}\mid m)+(1-\omega)\cdot f(\boldsymbol{\theta}_{(m)}\mid \mathbf{y},m)$, $\omega\in(0,1)$, τότε έχουμε ένα συνδυασμό των περιπτώσεων (i) και (ii) μέσω του παραπάνω κυρτού συνδυασμού της prior και της posterior κατανομής του $\boldsymbol{\theta}_{(m)}$.

3.5 Παραλλαγές του παράγοντα Bayes

Η ανάγκη για χρήση σε ορισμένες περιπτώσεις μη πληροφοριακών εκ των προτέρων κατανομών οδήγησε στην κατασκευή παραλλαγών του παράγοντα Bayes όπως ο εκ των υστέρων (Posterior BF), ο κλασματικός (Fractional BF) και ο ενδογενής παράγοντας Bayes (Intrinsic BF).

- Ο ***εκ των υστέρων παράγοντας Bayes*** (Posterior Bayes Factor) (Aitkin, 1991) είναι μια φυσική παραλλαγή του παράγοντα Bayes που βασίζεται στον λόγο των εκ των υστέρων προβλεπτικών κατανομών των παρατηρούμενων δεδομένων

$$\frac{f(\mathbf{y}\mid \mathbf{y},m_0)}{f(\mathbf{y}\mid \mathbf{y},m_0)}:=PBF_{01}=\frac{\int_{\Theta_{(m_0)}} f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_0)},m_0)\cdot f(\boldsymbol{\theta}_{(m_0)}\mid m_0,\mathbf{y})\,d\boldsymbol{\theta}_{(m_0)}}{\int_{\Theta_{(m_1)}} f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_1)},m_1)\cdot f(\boldsymbol{\theta}_{(m_1)}\mid m_1,\mathbf{y})\,d\boldsymbol{\theta}_{(m_1)}}=$$

$$\frac{E_{\boldsymbol{\theta}_{(m_0)}\mid \mathbf{y},m_0}\left[f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_0)},m_0)\right]}{E_{\boldsymbol{\theta}_{(m_1)}\mid \mathbf{y},m_1}\left[f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_1)},m_1)\right]}.$$

Το κύριο μειονέκτημα του είναι το γεγονός ότι η πληροφορία των δεδομένων χρησιμοποιείται δύο φορές και επομένως παραβιάζει την αρχή της πιθανοφάνειας.

- Ο ***κλασματικός παράγοντας Bayes*** (Fractional Bayes Factor) (O'Hagan, 1995):

$$PBF_{01}=\frac{\int_{\Theta_{(m_0)}} f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_0)},m_0)^{1-b}\cdot f_b(\boldsymbol{\theta}_{(m_0)}\mid m_0,\mathbf{y})\,d\boldsymbol{\theta}_{(m_0)}}{\int_{\Theta_{(m_1)}} f(\mathbf{y}\mid \boldsymbol{\theta}_{(m_1)},m_1)^{1-b}\cdot f_b(\boldsymbol{\theta}_{(m_1)}\mid m_1,\mathbf{y})\,d\boldsymbol{\theta}_{(m_1)}}$$

όπου

$$f_b(\boldsymbol{\theta}_{(m)}\mid m,\mathbf{y})=\int f(\mathbf{y}\mid \boldsymbol{\theta}_{(m)},m)^b\cdot f_b(\boldsymbol{\theta}_{(m)}\mid m,\mathbf{y})\,d\boldsymbol{\theta}_{(m)},$$

και $b<1$ (κλασματικός παράγοντας). Αυτή η παραλλαγή του παράγοντα Bayes βασίζεται στην έννοια της χρήσης ενός μέρους της πιθανοφάνειας για εκτίμηση και το υπόλοιπο για την επιλογή μοντέλου. Αν και ο κλασματικός παράγοντας Bayes είναι ένα χρήσιμο στατιστικό εργαλείο για την επιλογή μοντέλου, δεν αποτελεί έναν καθαρά Μπεϋζιανό τρόπο προσέγγισης του προβλήματος.

- Ο ***ενδογενής παράγοντας Bayes*** (Intrinsic Bayes Factor) (Berger και Pericchi, 1996), βασίστηκε στην αρχική ιδέα των (Spiegelhalter και Smith, 1982) που είχαν εισάγει την έννοια του *μερικού παράγοντα Bayes* (partial Bayes factor) στον οποίο χρησιμοποιούμε ένα μικρό μέρος των δεδομένων για εκτίμηση και τα υπόλοιπα δεδομένα για την επιλογή μοντέλου. Πιο συγκεκριμένα, τα δεδομένα διαιρούνται σε δύο μέρη, υπολογίζονται οι εκ των υστέρων κατανομές για το πρώτο μέρος και αυτές χρησιμοποιούνται ως εκ των προτέρων πληροφορίες για το δεύτερο μέρος. Η διαίρεση γίνεται με τέτοιο τρόπο ώστε το πρώτο μέρος να είναι το ελάχιστο δυνατό και να καταλήγει σε κατάλληλη εκ των υστέρων κατανομή. Δοθέντος ενός μικρού μέρους του δείγματος $y(l)$, ο μερικός παράγοντας Bayes ορίζεται ως εξής

$$B_{01}\left(\mathbf{y}(l)\right)=\frac{\int_{\Theta_{(m_0)}} f(\mathbf{y}(\backslash l)\,|\,\boldsymbol{\theta}_{(m_0)},m_0)\cdot f(\boldsymbol{\theta}_{(m_0)}\,|\,m_0,\mathbf{y}(l))\,d\boldsymbol{\theta}_{(m_0)}}{\int_{\Theta_{(m_1)}} f(\mathbf{y}(\backslash l)\,|\,\boldsymbol{\theta}_{(m_1)},m_1)\cdot f(\boldsymbol{\theta}_{(m_1)}\,|\,m_1,\mathbf{y}(l))\,d\boldsymbol{\theta}_{(m_1)}},$$

όπου $\mathbf{y}(\backslash l)$ είναι τα υπόλοιπα δεδομένα που θα χρησιμοποιηθούν για την επιλογή του μοντέλου. Ο ενδογενής παράγοντας Bayes εκτιμάται από τη διάμεσο ή τον μέσο των διαφορετικών μερικών παραγόντων Bayes $B_{01}\left(\mathbf{y}(l)\right)$ που μπορούν να προκύψουν από ένα δείγμα.

3.6 Μπεϋζιανή στάθμιση μοντέλων

Κατά την προσπάθεια να βγάλουμε συμπεράσματα για μια ποσότητα που μας ενδιαφέρει και είναι καλά καθορισμένη για κάθε μοντέλο, μπορούμε να εκφράσουμε την αβεβαιότητά μας (ως προς το ποιο μοντέλο είναι καταλληλότερο) χρησιμοποιώντας τις εκ των υστέρων πιθανότητες των μοντέλων σαν βάρη (Kass και Raftery, 1995). Η τεχνική αυτή, γνωστή ως Μπευζιανή στάθμιση μοντέλων (Bayesian Model Averaging, BMA) παράγει ουσιαστικά καλύτερες προβλέψεις από τις μεθόδους που βασίζονται σε μεμονωμένα μοντέλα.

Αν Δ είναι η παράμετρος που μας ενδιαφέρει (π.χ η μελλοντική παρατήρηση), τότε η εκ των υστέρων κατανομή της δεδομένου των παρατηρήσεων θα είναι

$$f(\Delta\,|\,\mathbf{y})=\sum_{m\in\mathrm{M}} f(m\,|\,\mathbf{y})\cdot f(\Delta\,|\,m,\mathbf{y}),$$

όπου $f(m\,|\,y)$ είναι η posterior πιθανότητα του μοντέλου m και $f(\Delta\,|\,m,y)$ είναι η posterior κατανομή της ποσότητας Δ στο m-μοντέλο. Η τεχνική αυτή παρουσιάζει ορισμένες δυσκολίες κατά την εφαρμογή της, όπως ότι αριθμός των συγκρινόμενων μοντέλων κάποιες φορές είναι πολύ μεγάλος.
Η προβλεπτική ικανότητα των οποιονδήποτε μοντέλων μπορεί να μετρηθεί από την ποσότητα

$$LS = -E\left\{\log\left[\sum_{m\in \mathrm{M}} f(m\mid \mathbf{y})\cdot f(\Delta\mid m,\mathbf{y})\right]\right\}$$, (logarithmic scoring rule)

για την Μπευζιανή στάθμιση μοντέλων και από την ποσότητα

$$LS_m = -E\left\{\log\left[f(\Delta\mid m,\mathbf{y})\right]\right\},$$

για το μοντέλο m ξεχωριστά. Χαμηλότερες τιμές του LS δείχνουν καλύτερη προβλεπτική δυνατότητα. Η μέθοδος Bayesian Model Averaging προσφέρει πάντα καλύτερες προβλέψεις για μια ποσότητα που μας ενδιαφέρει και αυτό διότι ισχύει $LS \le LS_m$, $\forall m \in \mathrm{M}$.

3.7 Κριτήρια πληροφορίας

Μία εναλλακτική και πολλές φορές ευκολότερη λύση για να συγκρίνουμε διαφορετικά μοντέλα είναι η χρησιμοποίηση κριτηρίων πληροφορίας. Τα πιο δημοφιλή κριτήρια είναι το Akaike's Information Criterion (AIC) (Akaike, 1974), το Bayes Information Criterion (BIC) (Schwarz, 1978), που είναι γνωστό και ως Schwarz Criterion και το πιο πρόσφατο Deviance Information Criterion (DIC) (Spiegelhalter et al, 2002). Όλα αυτά τα κριτήρια βασίζονται στη συνάρτηση Deviance η οποία για το μοντέλο m ορίζεται ως εξής

$$D(\boldsymbol{\theta}_{(m)}, m) = -2\log f(\mathbf{y}\mid \boldsymbol{\theta}_{(m)}, m).$$

Γενικά, τα περισσότερα από τα κριτήρια ελαχιστοποιούν την ποσότητα

$$IC_{(m)} = -2\log f(\mathbf{y}\mid \hat{\boldsymbol{\theta}}_{(m)}, m) + d(m)\cdot F,$$

όπου $\hat{\boldsymbol{\theta}}_{(m)}$ και $d(m)$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας και η διάσταση της παραμέτρου $\boldsymbol{\theta}_{(m)}$ στο μοντέλο m αντίστοιχα. Είναι προφανές ότι οι τιμές του διανύσματος $\hat{\boldsymbol{\theta}}_{(m)}$ που μεγιστοποιούν την πιθανοφάνεια, ελαχιστοποιούν ταυτόχρονα τη συνάρτηση Deviance. Ο όρος F εκφράζει μία συνάρτηση ποινής (penalty function) η οποία επιβάλλεται στην ποσότητα -2log-likelihood($D(\hat{\theta}_{(m)}, m)$) για κάθε μία επιπλέον παράμετρο που προστίθεται στο μοντέλο. Διαφορετικές συναρτήσεις ποινής δίνουν και διαφορετικά κριτήρια.

Για $F = 2$ έχουμε το AIC κριτήριο

$$AIC_{(m)} = D(\hat{\boldsymbol{\theta}}_{(m)}, m) + 2\cdot d(m),$$

ενώ για $F = \log(n)$ έχουμε το κριτήριο BIC

$$BIC_{(m)} = D(\hat{\boldsymbol{\theta}}_{(m)}, m) + d(m)\cdot \log(n),$$

όπου n είναι ο αριθμός των παρατηρήσεων του δείγματος. Αν θέλουμε να συγκρίνουμε δύο μοντέλα m_0 και m_1, επιλέγουμε αυτό που έχει τη μικρότερη τιμή του εκάστοτε κριτηρίου που χρησιμοποιούμε ($IC_{(m)}$) και επομένως ορίζουμε ως IC_{01} τη διαφορά τους

$$IC_{(m_0)} - IC_{(m_1)} := IC_{01} = -2\log\left(\frac{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_0)}, m_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_1)}, m_1)}\right) + \left[d(m_0) - d(m_1)\right] \cdot F. \quad (3.7.1)$$

Αν $IC_{01} < 0$ επιλέγουμε το μοντέλο m_0 ενώ αν $IC_{01} > 0$ επιλέγουμε το μοντέλο m_1. Η παραπάνω σχέση μπορεί να γενικευτεί ορίζοντας

$$IC_{01} = -2\log\left(\frac{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_0)}, m_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_1)}, m_1)}\right) + \psi,$$

όπου ψ είναι μια συνάρτηση ποινής που εξαρτάται από τη διαφορά των διαστάσεων των μοντέλων $d(m_0) - d(m_1)$, το μέγεθος n του δείγματος και από τους πίνακες σχεδιασμού $\mathbf{X}_{(m_0)}$, $\mathbf{X}_{(m_1)}$ των δύο μοντέλων.

Το κριτήριο BIC μπορεί να χρησιμοποιηθεί για τον προσεγγιστικό υπολογισμό του παράγοντα Bayes (BF) και της εκ των υστέρων πιθανότητας ενός μοντέλου m, $f(m \mid y)$. Με χρήση του κριτηρίου Schwarz (Schwarz, 1978)

$$S_{01} = \log\left(\frac{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_0)}, m_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_1)}, m_1)}\right) - \frac{1}{2}\left[d(m_0) - d(m_1)\right] \cdot \log(n),$$

αποδεικνύεται ότι

$$\frac{S_{01} - \log(B_{01})}{\log(B_{01})} \to 0, \ \text{όταν } n \to \infty,$$

οπότε

$$\log(B_{01}) \to S_{01} \Rightarrow -2\log(B_{01}) \to -2 \cdot S_{01}. \quad (3.7.2)$$

Επίσης έχουμε το ακόλουθο

$$\begin{aligned}
\Delta BIC_{01} &:= BIC_{(m_0)} - BIC_{(m_1)} \\
&= -2\log\left(\frac{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_0)}, m_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_1)}, m_1)}\right) + \left[d(m_0) - d(m_1)\right] \cdot \log(n) \\
&= -2 \cdot \left\{\log\left(\frac{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_0)}, m_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_{(m_1)}, m_1)}\right) - \frac{1}{2} \cdot \left[d(m_0) - d(m_1)\right] \cdot \log(n)\right\} \\
&= -2 \cdot S_{01},
\end{aligned}$$

δηλαδή

$$\Delta BIC_{01} = -2 \cdot S_{01}. \qquad (3.7.3)$$

Από τις σχέσεις (3.7.2) και (3.7.3) συμπεραίνουμε ότι

$$-2\log(B_{01}) \to \Delta BIC_{01}, \quad \text{όταν } n \to \infty$$

και συνεπώς παίρνουμε την προσέγγιση του παράγοντά Bayes από τη σχέση

$$B_{01} \approx \exp\left\{-\frac{1}{2} \cdot \Delta BIC_{01}\right\}.$$

Όπως αναφέραμε προηγουμένως, το κριτήριο BIC μπορεί να χρησιμοποιηθεί και για τον υπολογισμό της εκ των υστέρων πιθανότητας ενός μοντέλου *m*. Θεωρώντας ένα σύνολο μοντέλων $M = \{m_1, m_2, ..., m_{|M|}\}$, η εκ των υστέρων πιθανότητα του μοντέλου $m \in M$ υπολογίζεται ως γνωστόν από τη σχέση

$$f(m \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid m) \cdot f(m)}{\sum_{k=1}^{|M|} f(\mathbf{y} \mid m_k) \cdot f(m_k)} = \left(\sum_{k=1}^{|M|} PO_{m_k m} \right)^{-1}.$$

Ας υποθέσουμε ότι όλα τα μοντέλα έχουν την ίδια εκ των προτέρων πιθανότητα. Τότε όμως θα είναι $PO_{m_k m} = B_{m_k m}$, για κάθε $k = 1, 2, ..., |M|$ και συνεπώς

$$\begin{aligned} f(m \mid \mathbf{y}) &= \left(\sum_{k=1}^{|M|} B_{m_k m} \right)^{-1} = \left(\sum_{k=1}^{|M|} \exp\left\{ -\frac{1}{2} \cdot \Delta BIC_{m_k m} \right\} \right)^{-1} \\ &= \frac{\exp\left(-\frac{1}{2} BIC_m \right)}{\sum_{k=1}^{|M|} \exp\left(-\frac{1}{2} BIC_{m_k} \right)}. \end{aligned}$$

Μέχρι τώρα στην παρούσα ενότητα έχουμε χρησιμοποιήσει τους εκτιμητές μέγιστης πιθανοφάνειας $\hat{\boldsymbol{\theta}}_{(m)}$ στα κριτήρια πληροφορίας AIC και BIC. Σύμφωνα με τον Brooks (2002) μπορούμε να ορίσουμε τις Μπεϋζιανές εκδοχές των κριτηρίων αυτών

$$AIC_{(m)}(\overline{D}) = \overline{D}(\boldsymbol{\theta}_{(m)}, m) + 2 \cdot d(m), \quad AIC_{(m)}(\overline{\boldsymbol{\theta}}_{(m)}) = D(\overline{\boldsymbol{\theta}}_{(m)}, m) + 2 \cdot d(m)$$

και

$$BIC_{(m)}(\overline{D}) = \overline{D}(\boldsymbol{\theta}_{(m)}, m) + d(m) \cdot \log(n), \quad BIC_{(m)}(\overline{\boldsymbol{\theta}}_{(m)}) = D(\overline{\boldsymbol{\theta}}_{(m)}, m) + d(m) \cdot \log(n),$$

όπου $\overline{D}(\boldsymbol{\theta}_{(m)}, m)$ είναι η εκ των υστέρων μέση τιμή της Deviance (η Deviance είναι τυχαία μεταβλητή αν θεωρήσουμε το διάνυσμα των παραμέτρων $\boldsymbol{\theta}_{(m)}$ ως τυχαία μεταβλητή). Αυτή η μέση τιμή αποτελεί ένα μέτρο για το πόσο καλά το μοντέλο παρεμβάλει τα δεδομένα. Ο όρος $D(\overline{\boldsymbol{\theta}}_{(m)}, m)$ εκφράζει την τιμή της Deviance υπολογισμένη στον εκ των υστέρων μέσο $\overline{\boldsymbol{\theta}}_{(m)}$ των παραμέτρων.

Αναφέρουμε στη συνέχεια ένα ακόμα κριτήριο γνωστό ως Deviance Information Criterion (DIC) το οποίο αποτελεί μια γενίκευση του AIC και ορίζεται ως εξής

$$DIC_{(m)} = 2 \cdot \overline{D}(\boldsymbol{\theta}_{(m)}, m) - D(\overline{\boldsymbol{\theta}}_{(m)}, m).$$

Θέτοντας $p_{(m)} := \overline{D}(\theta_{(m)}, m) - D(\overline{\theta}_{(m)}, m)$ τον αριθμό των δραστικών (effective) παραμέτρων στο μοντέλο m τότε το κριτήριο DIC μπορεί να εκφραστεί με τις σχέσεις

$$\begin{aligned} DIC_{(m)} &= D(\overline{\boldsymbol{\theta}}_{(m)}, m) + 2 \cdot p_{(m)} \\ &= \overline{D}(\boldsymbol{\theta}_{(m)}, m) + p_{(m)}. \end{aligned}$$

<u>Παρατήρηση</u>: Για Κανονικά μοντέλα (Normal models), ισχύει $p_{(m)} \approx d(m)$ και κατά συνέπεια από τη σχέση $DIC_{(m)} = D(\overline{\boldsymbol{\theta}}_{(m)}, m) + 2 \cdot p_{(m)}$ έχουμε ότι $DIC_{(m)} \approx AIC_{(m)}$.

Τέλος, αξίζει να αναφερθεί ότι στις περισσότερες περιπτώσεις υπάρχει ισοδυναμία ανάμεσα στις διάφορες παραλλαγές του παράγοντα Bayes και στα κριτήρια πληροφορίας. Για παράδειγμα, ο εκ των υστέρων παράγοντας Bayes (PBF) είναι προσεγγιστικά ίσος με το κριτήριο του τύπου (3.7.1) με συνάρτηση ποινής $F = \log(2)$.

3.8 Εκ των προτέρων κατανομή στον χώρο των μοντέλων

Ο υπολογισμός, με χρήση του Θεωρήματος Bayes, της εκ των υστέρων πιθανότητας $f(m \mid \mathbf{y})$ οποιουδήποτε μοντέλου, απαιτεί όπως είναι γνωστό τον καθορισμό μιας εκ των προτέρων πιθανότητας για κάθε ένα μοντέλο $m \in M$. Η πιο συνηθισμένη και προφανής επιλογή, κυρίως όταν δεν έχουμε καμία εκ των προτέρων πληροφορία, είναι η χρήση μιας Ομοιόμορφης κατανομής, η οποία δίνει το ίδιο βάρος (την ίδια αρχική πιθανότητα) σε όλα τα μοντέλα του χώρου M και ως εκ τούτου χαρακτηρίζεται ως μή-πληροφοριακή (non-informative). Επομένως ορίζουμε

$$f(m) = \frac{1}{|M|}, \quad \forall m \in M, \quad (3.8.1)$$

όπου $|M|$ το πλήθος όλων των μοντέλων του χώρου M. Κατά το πρόβλημά της επιλογής μεταβλητών σε ένα γενικευμένο γραμμικό μοντέλο, είναι βολικό να ορίσουμε μια κατανομή *Bernoulli* για κάθε ένα όρο γ_j, δηλαδή

$$\gamma_j \sim Bernoulli(\pi_j), \quad j = 1, 2, \ldots p,$$

με π_j να είναι η εκ των προτέρων πιθανότητα να περιλαμβάνεται η μεταβλητή X_j στο μοντέλο. Θεωρώντας τα γ_j ανεξάρτητα, η παραπάνω prior δίνει την συνολική prior του μοντέλου $\boldsymbol{\gamma}$

$$f(\boldsymbol{\gamma}) = \prod_{j=1}^{p} \pi_j^{\gamma_j} (1-\pi_j)^{1-\gamma_j},$$

όπου $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p) \in \{0,1\}^p$. Πολλές φορές θεωρούμε την ίδια εκ των προτέρων πιθανότητα για όλες τις μεταβλητές, δηλαδή $\pi_j = \pi$ για κάθε $j = 1, 2, ...p$. Τότε έχουμε

$$f(\boldsymbol{\gamma}) = \pi^{d(\boldsymbol{\gamma})} (1-\pi)^{p-d(\boldsymbol{\gamma})},$$

με $d(\boldsymbol{\gamma}) = \sum_{j=1}^{p} \gamma_j \ (= d(m))$ τη διάσταση του μοντέλου γ. Για $\pi = 0.5$, δηλαδή όταν $\gamma_j \sim Bernoulli(0.5)$, $j = 1, 2, ...p$, παίρνουμε το ισοδύναμο της prior (3.8.1) (non-informative prior). Από την παραπάνω σχέση προκύπτει επίσης ότι

$$f(\boldsymbol{\gamma}) = (1-\pi)^p \cdot \left(\frac{\pi}{1-\pi} \right)^{d(\boldsymbol{\gamma})} \propto \left(\frac{\pi}{1-\pi} \right)^{d(\boldsymbol{\gamma})},$$

απ'όπου καταλαβαίνουμε ότι η prior πιθανότητα ενός μοντέλου εξαρτάται στη γενική περίπτωση από τη διάσταση που έχει το συγκεκριμένο μοντέλο. Στην ειδική περίπτωση όμως που θεωρήσουμε μη πληροφοριακή prior ($\pi = 0.5$), έχουμε το εξής

$$f(\boldsymbol{\gamma}) \propto (1)^{d(\boldsymbol{\gamma})} = 1,$$

πράγμα που δείχνει ότι δεν υπάρχει εξάρτηση από τη διάσταση του μοντέλου γ, κάτι που διαπιστώνουμε και από την μη-πληροφοριακή prior της σχέσης (3.8.1), αφού και αυτή μας δείχνει ότι η εκ των προτέρων πιθανότητα κάθε μοντέλου m δεν εξαρτάται από τη διάστασή του.

3.9 Εκ των προτέρων κατανομή για τους συντελεστές του μοντέλου

3.9.1 Εισαγωγικά

Ο καθορισμός της εκ των προτέρων κατανομής για τους συντελεστές β_j ενός γενικευμένου γραμμικού μοντέλου όταν καμία εκ των προτέρων πληροφορία δεν είναι διαθέσιμη, παρουσιάζει μεγάλο ενδιαφέρον κυρίως λόγω του παραδόξου των Lindley και Bartlett. Οι εκ των υστέρων πιθανότητες των μοντέλων είναι πολύ ευαίσθητες στο βαθμό των εκ των προτέρων διασπορών, έχοντας μια τάση να ευνοούν τα μοντέλα με πιο απλή δομή (με λιγότερες μεταβλητές) καθώς οι εκ των προτέρων διασπορές αυξάνουν (Bartllet 1957, Lindley 1957). Επομένως, η επιλογή των εκ των προτέρων κατανομών είναι καίριας σημασίας για την εκ των υστέρων υποστήριξη των μοντέλων. Στην δημοσίευσή του ο Lindley (1957), δίνει έμφαση στην επίδραση του μεγέθους του δείγματος στις εκ των υστέρων πιθανότητες των μοντέλων και στα αντιφατικά αποτελέσματα μεταξύ των Μπεϋζιανών και κλασικών ελέγχων σημαντικότητας. Ο Bartllet (1957), τονίζει αντιστοίχως την επίδραση της διασποράς της εκ των προτέρων κατανομής των παραμέτρων, στις εκ των υστέρων πιθανότητες των μοντέλων.

Συνοψίζοντας όλα τα παραπάνω, έστω m_0 και m_1 δύο μοντέλα με $d(m_0) < d(m_1)$, όπου $d(m_0)$, $d(m_1)$ οι διαστάσεις των μοντέλων m_0 και m_1 αντίστοιχα.

- Αν το μέγεθος του δείγματος $n \to \infty$ τότε $B_{10} \to 0$,
  (ο παράγοντας Bayes υποστηρίζει απλούστερα μοντέλα (Lindley, 1957)).

- Αν η εκ των προτέρων διασπορά των συντελεστών $\to \infty$ τότε $B_{10} \to 0$,
  (Bartllet, 1957).

Η πιο συνηθισμένη εκ των προτέρων κατανομή για να εκφράσουμε τις εκ των προτέρων πεποιθήσεις μας για τις παραμέτρους του μοντέλου $m \in M$ είναι η πολυμεταβλητή Κανονική κατανομή, δηλαδή

$$f(\boldsymbol{\beta}_{(m)} \mid m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, \boldsymbol{\Sigma}_{(m)}), \quad (3.9.1.1)$$

όπου $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}$ και $\boldsymbol{\Sigma}_{(m)}$ είναι ο εκ των προτέρων μέσος και ο εκ των προτέρων πίνακας συνδυακύμανσης αντίστοιχα των $\boldsymbol{\beta}_{(m)}$. Λόγω του παραδόξου Lindley και Bartlett, όταν καμία εκ των προτέρων πληροφορία δεν είναι διαθέσιμη, πρέπει να επιλέξουμε μια prior κατανομή που παρέχει μικρή πληροφορία για τις παραμέτρους $\boldsymbol{\beta}_{(m)}$ η οποία όμως δεν είναι τελείως επίπεδη. Ο εκ των προτέρων πίνακας συνδυακύμανσης μπορεί να γραφτεί εναλλακτικά ως $\boldsymbol{\Sigma}_{(m)} = c^2 \mathbf{V}_{(m)}$, όπου ο συντελεστής c^2 ελέγχει την διασπορά της prior κατανομής και ο πίνακας $\mathbf{V}_{(m)}$ καθορίζει τον εκ των προτέρων συσχετισμό των συντελεστών $\boldsymbol{\beta}_{(m)}$. Συνηθισμένη επιλογή για τον prior μέσο, όταν καμία εκ των προτέρων πληροφορία δεν είναι διαθέσιμη, είναι το μηδενικό διάνυσμα, $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \mathbf{0}$.

3.9.2 Ανεξάρτητες εκ των προτέρων κατανομές για τους συντελεστές

Σε προβλήματα επιλογής μοντέλου και μεταβλητών μία από τις μεθόδους που μπορούμε να χρησιμοποιούμε είναι να θέτουμε ***ανεξάρτητες*** εκ των προτέρων κατανομές στις παραμέτρους β_j, $j = 1, ..., p$. Σ'αυτή την περίπτωση η prior δίνεται από τη μορφή

$$\beta_j \sim N(\mu_{\beta_j}, \sigma_j^2), \; j = 1, ..., p,$$

όπου μ_{β_j} και σ_j^2 είναι ο εκ των προτέρων μέσος και εκ των προτέρων διασπορά αντίστοιχα για τον όρο j, ανεξάρτητα από το μοντέλο *m*. Για μη πληροφοριακές περιπτώσεις θέτουμε $\mu_{\beta_j} = 0$.

3.9.3 Εκ των προτέρων κατανομές εξαρτημένες από το μοντέλο

Στα Κανονικά γραμμικά μοντέλα η εκ των προτέρων κατανομή (3.9.1.1) είναι αυτή που συνήθως χρησιμοποιείται. Επιπρόσθετα χρειάζεται να ορίσουμε μια prior για την διασπορά σ^2. Συνήθως η *Gamma* κατανομή χρησιμοποιείται για την παράμετρο ακρίβειας $\tau = \sigma^{-2}$ (ή η Αντίστροφη *Gamma* για το σ^2). Επομένως έχουμε

$$\tau \sim Gamma(\alpha_\tau, b_\tau). \quad (3.9.3.1)$$

Μια ακατάλληλη (improper) εκ των προτέρων στην τ δεν επηρεάζει σε μεγάλο βαθμό τις εκ των υστέρων πιθανότητες (posterior odds) και ως εκ τούτου μπορούμε να θέσουμε $\alpha_\tau = b_\tau = 0$ χωρίς κανένα πρόβλημα. Στα Κανονικά γραμμικά μοντέλα είναι βολικό να χρησιμοποιήσουμε εκ των προτέρων κατανομές στους συντελεστές του μοντέλου δεδομένου της παραμέτρου σ^2. Έτσι αντί της (3.9.1.1) μπορούμε να χρησιμοποιήσουμε την prior

$$f(\boldsymbol{\beta}_{(m)} \mid \sigma^2, m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, \sigma^2 \boldsymbol{\Sigma}_{(m)}). \quad (3.9.3.2)$$

Η από κοινού εκ των προτέρων κατανομή $f(\boldsymbol{\beta}_{(m)}, \sigma^2 \mid m)$ δίνεται τότε από το γινόμενο των κατανομών (3.9.3.1) και (3.9.3.2) η οποία καλείται Κανονική-Αντίστροφη Γάμμα κατανομή (Normal-Inverse Gamma) και είναι ***συζυγής*** αφού και η εκ των υστέρων $f(\boldsymbol{\beta}_{(m)}, \sigma^2 \mid m, \mathbf{y})$ είναι επίσης Normal-Inverse Gamma (βλέπε π.χ. Bernardo και Smith (1994) ή O'Hagan (1994)).

Οι Smith και Kohn (1996) και οι George και Foster (1997) υιοθέτησαν την εκ των προτέρων κατανομή

$$f(\boldsymbol{\beta}_{(m)} \mid \sigma^2, m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, c^2 \mathbf{V}_{(m)} \sigma^2), \; \tau \sim \text{Gamma}(\alpha_\tau, b_\tau), \quad (3.9.3.3)$$

με

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \mathbf{0} \text{ και } \mathbf{V}_{(m)} = (\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1}, \; (3.9.3.4)$$

καταλήγοντας σε μία απλή μορφή για τις εκ των υστέρων πιθανότητες με πολύ καλές ιδιότητες και ερμηνείες. Για τον καθορισμό της παραμέτρου c^2, οι Smith και Kohn (1997) πρότειναν τιμές μεταξύ 10 και 1000 ενώ οι τιμές $c^2 = 100$ και $c^2 = n$ προτάθηκαν ως πολύ καλές πρακτικές λύσεις.

Οι Ibrahim και Laud (1994) και οι Laud και Ibrahim (1995) χρησιμοποίησαν παρόμοιες εκ των προτέρων κατανομές στη δική τους εφαρμογή επιλογής μοντέλου. Πρότειναν πίνακα συνδυακύμανσης όπως αυτόν της (3.9.3.4) και μέσο ίσο με τους εκτιμητές μέγιστης πιθανοφάνειας, δηλαδή $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \hat{\boldsymbol{\beta}}_{(m)}$.

Για τα υπόλοιπα γενικευμένα γραμμικά μοντέλα (πλην του Κανονικού), μπορούμε να χρησιμοποιήσουμε μια περισσότερο γενική μορφή για τις εκ των προτέρων κατανομές που δίνονται από τη σχέση

$$f(\boldsymbol{\beta}_{(m)} \mid m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, c^2 \mathbf{V}_{(m)}). \quad (3.9.3.5)$$

Όπως αναφέραμε νωρίτερα, η συνήθης επιλογή όταν καμία πληροφορία δεν είναι διαθέσιμη, είναι $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \mathbf{0}$. Για τον πίνακα συνδυακύμανσης μπορούμε επίσης να θεωρήσουμε

$$\mathbf{V}_{(m)} = (\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1}. \qquad (3.9.3.6)$$

Μια εναλλακτική επιλογή είναι να θέσουμε

$$\mathbf{V}_{(m)} = \mathbf{I}_{\hat{\boldsymbol{\beta}}_{(m)}} = -\left[\frac{\partial^2 l(\boldsymbol{\beta}_{(m)})}{\partial \beta_v \partial \beta_j} \right]_{\boldsymbol{\beta}_{(m)} = \hat{\boldsymbol{\beta}}_{(m)}}^{-1} = \left(\mathbf{X}_{(m)}^T \hat{\mathbf{H}}_{(m)} \mathbf{X}_{(m)} \right)^{-1}, \quad (3.9.3.7)$$

όπου $\mathbf{I}_{\hat{\boldsymbol{\beta}}_{(m)}}$ ο πίνακας πληροφορίας Fisher, $l(\boldsymbol{\beta}_{(m)})$ η συνάρτηση πιθανοφάνειας του μοντέλου m, $\hat{\boldsymbol{\beta}}_{(m)}$ ο εκτιμητής μέγιστης πιθανοφάνειας των συντελεστών $\boldsymbol{\beta}_{(m)}$ και $\mathbf{H}_{(m)} = diag(h_i)$, όπου $h_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{\alpha_i(\phi) b''(\theta)}$.

ΜΟΝΤΕΛΟ	ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ	ΒΑΡΗ h_i
Κανονικό	Ταυτοτική	σ^{-2}
Poisson	Log	λ_i
Διωνυμικό	Logit	$N_i p_i (1 - p_i)$

Αν και οι εκ των προτέρων κατανομές που χρησιμοποιούν πίνακα συνδυακύμανσης όπως αυτός της σχέσης (3.9.3.7) είναι εξαρτημένες, μπορούμε να τις βλέπουμε ως μη πληροφοριακές μιας και δεν επηρεάζουν την εκ των υστέρων κατανομή $f(\boldsymbol{\beta}_{(m)} \mid m, \mathbf{y})$ για μεγάλες τιμές του c^2. Να σημειωθεί ότι η prior των Smith και Kohn (1996) είναι μια ειδική περίπτωση της παραπάνω γενίκευσης μιας και $h_i = \sigma^{-2}$ για τα Κανονικά μοντέλα. Ειδική περίπτωση της prior (3.9.3.3) με $\mathbf{V}_{(m)}$ της μορφής (3.9.3.7) είναι η ***μοναδιαίας πληροφορίας εκ των προτέρων κατανομή*** (unit information prior) για $c^2 = n$. Αυτή η prior έχει ακρίβεια προσεγγιστικά ίση με την ακρίβεια που παρέχεται από ένα δεδομένο. Ο πίνακας πληροφορίας Fisher μετρά το ποσό της πληροφορίας που παρέχονται από τα δεδομένα και επομένως η ακρίβεια από ένα δεδομένο είναι ασυμπτωτικά ίση με $n^{-1} \mathbf{I}_{\boldsymbol{\beta}_{(m)}}^{-1}$.

3.9.4 Εκ των προτέρων κατανομές στους συντελεστές προερχόμενες από το μοντέλο με τυποποιημένες μεταβλητές

Ο Raftery (1996) ανέπτυξε μια εκ των προτέρων κατανομή για γενικευμένα γραμμικά μοντέλα με μονοδιάστατους όρους χρησιμοποιώντας τυποποιημένες μεταβλητές οι οποίες σημειώνονται με $\mathbf{Y}^s$ και X_j^s. Αρχικά θεώρησε την περίπτωση όπου τα h_i είναι σταθερά για όλες τις παρατηρήσεις και ταυτοτική συνδετική συνάρτηση $g(\mu)=\mu$. Αν θεωρήσουμε το διάνυσμα των συντελεστών του πλήρους μοντέλου $\boldsymbol{\beta}^T=(\beta_0,\beta_1,...,\beta_p)$, όπου β_0 είναι ο συντελεστής του σταθερού όρου, τότε το νέο μετασχηματισμένο μοντέλο δίνεται από τη σχέση

$$\mu_i^s = E\left[Y_i^s\right] = \beta_0^s + \sum_{j=1}^{p} x_{ij}^s \beta_j^s \ , \ i=1,...,n.$$

Ο Raftery χρησιμοποίησε ανεξάρτητες Κανονικές εκ των προτέρων στα β_0^s και β_j^s, δηλαδή

$$\beta_0^s \sim N(\mu_{\beta_0^s}, c_0^2), \quad \beta_j^s \sim N(0,c^2), \ j=i=1,...,p,$$

όπου c_0^2, c^2 είναι οι prior παράμετροι που πρέπει να καθοριστούν. Από τα παραπάνω έχουμε

$$\beta_0 = \overline{\mathbf{y}} + s_{\mathbf{y}}\beta_0^s - \sum_{j=1}^{p} \frac{s_{\mathbf{y}}}{s_j}\overline{\mathbf{x}}_j \beta_j^s$$

$$\beta_j = \frac{s_{\mathbf{y}}}{s_j}\beta_j^s, \ j=1,...,p,$$

όπου $\overline{\mathbf{y}}$ και $s_{\mathbf{y}}^2$ είναι ο δειγματικός μέσος και η διασπορά της μεταβλητής απόκρισης $\mathbf{Y}$, $\overline{\mathbf{x}}_j$ και s_j^2 είναι ο δειγματικός μέσος και η διασπορά των παρατηρήσεων που αφορούν την επεξηγηματική μεταβλητή X_j. Όλα αυτά οδηγούν στην πολυμεταβλητή Κανονική εκ των προτέρων κατανομή για τους συντελεστές $\boldsymbol{\beta}$ που δίνεται από τη σχέση (3.9.3.3) με εκ των προτέρων μέσο

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}^T = (\mu_{\beta_0^s} + \overline{\mathbf{y}}, 0, ..., 0)$$

και εκ των προτέρων πίνακα συνδυασποράς

$$\boldsymbol{\Sigma}_{(m)} = c^2 s_{\mathbf{y}}^2 \begin{bmatrix} c^{-2}c_0^2 - \sum_j s_j^{-2}\overline{\mathbf{x}}_j^2 & -s_2^{-2}\overline{\mathbf{x}}_2 & -s_3^{-2}\overline{\mathbf{x}}_3 & \cdots & -s_{d(m)}^{-2}\overline{\mathbf{x}}_{d(m)} \\ -s_2^{-2}\overline{\mathbf{x}}_2 & s_2^{-2} & 0 & \cdots & 0 \\ -s_3^{-2}\overline{\mathbf{x}}_3 & 0 & s_3^{-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -s_{d(m)}^{-2}\overline{\mathbf{x}}_{d(m)} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

όπου $d(m)$ η διάσταση του μοντέλου m (ο αριθμός των επεξηγηματικών μεταβλητών που λαμβάνουν μέρος στο μοντέλο m). Να σημειωθεί ότι ο παραπάνω πίνακας είναι συμμετρικός.

Για γενικευμένα γραμμικά μοντέλα με διαφορετική της ταυτοτικής συνδετική συνάρτηση, ο Raftery πρότεινε παρόμοια διαδικασία με την παραπάνω αλλά με κατάλληλα κάθε φορά βάρη h_i όπως αυτά δίνονται π.χ στο Poisson και Διωνυμικό μοντέλο στον πίνακα (3.14). Χρησιμοποίησε επίσης δύο κριτήρια για να ορίσει κατάλληλες τιμές για την παράμετρο c^2. Ανέφερε την τιμή $c=1.65$ που ικανοποιεί και τα δύο κριτήρια και πρότεινε διάφορες τιμές από 1 ως 5 ($1 \le c \le 5$).

Μια παρόμοια prior ορίστηκε από τους Raftery et al (1997) για Κανονικά γραμμικά μοντέλα. Η εκ των προτέρων για το σταθερό συντελεστή προτάθηκε να είναι $\beta_0 \sim N(\hat{\beta}_0, s_y^2\sigma^2)$ ενώ για τους υπόλοιπους συντελεστές

$$\beta_j \sim N(0, c^2 s_j^{-2}\sigma^2),\ j=1,...,p.$$

Οι Raftery et al (1997) ακολουθώντας συγκεκριμένα κριτήρια πρότειναν τις τιμές $c=2.85$, $a_\tau=1.27$, $b_\tau=0.3612$ (όπου a_τ, b_τ είναι οι παράμετροι της εκ των προτέρων Γάμμα κατανομής που υποθέτουμε ότι ακολουθεί η παράμετρος ακρίβειας $\tau=\sigma^{-2}$).

3.9.5 Καθορισμός εκ των προτέρων κατανομής στο πλήρες μοντέλο

Μια άλλη στρατηγική για την επιλογή εκ των προτέρων κατανομών είναι να χρησιμοποιήσουμε prior για τους συντελεστές $\mathbf{\beta}$ του πλήρους μοντέλου (με όλες τις επεξηγηματικές μεταβλητές) και τις περιθώριες κατανομές για τα μοντέλα χαμηλότερης διάστασης. Αυτή η εφαρμογή χρησιμοποιήθηκε από τους Kuo και Mallick (1998) αλλά κατέληξαν στο ότι μια εκ των προτέρων στο πλήρες μοντέλο μπορεί να αποτελέσει ακατάλληλη ή ανεπιθύμητη prior για μοντέλα με μικρότερη διάσταση. Στην περίπτωση πού ο εκ των προτέρων πίνακας συνδυακύμανσης είναι διαγώνιος, η κατανομή του $\mathbf{\beta}$ (δηλαδή η από κοινού κατανομή των συντελεστών β_j) ανάγεται σε ανεξάρτητες εκ των προτέρων κατανομές για κάθε συντελεστή β_j, $j=0,...,p$. Έτσι λοιπόν αν υποθέσουμε ότι $\mathbf{\beta}_{(m)} \mid m \sim N(\mathbf{0}, c^2\mathbf{V}_{(m)})$ με $\mathbf{V}_{(m)} = diag(v_j^2)$ τότε έχουμε ισοδύναμα ότι $\beta_j \sim N(0, c^2 v_j^2)$, $j=0,...,p$. Άλλες επιλογές για τον πίνακα συνδυακύμανσης είναι είτε $\mathbf{V}_{(m)} = (\mathbf{X}_{(m)}^T\mathbf{X}_{(m)})^{-1}$ είτε $\left(\mathbf{X}_{(m)}^T\hat{\mathbf{H}}_{(m)}\mathbf{X}_{(m)}\right)^{-1}$, όπου $\mathbf{X}_{(m)}$ και $\hat{\mathbf{H}}_{(m)}$ είναι ο πίνακας σχεδιασμού και ο σταθμισμένος πίνακας του πλήρους μοντέλου αντίστοιχα. Για Κανονικά μοντέλα η παραπάνω prior μπορεί επίσης να οριστεί δεδομένου του σ^2 παρόμοια με την prior των Smith και Kohn (1996).

3.9.6 Ενδογενής και «φανταστικών» δεδομένων εκ των προτέρων κατανομές

Μια άλλη πιο περίπλοκη εφαρμογή είναι να θεωρήσουμε διαμορφωτικά (training) δεδομένα ώστε να καθορίσουμε μια εκ των προτέρων κατανομή για τους συντελεστές

β_j. Τα διαμορφωτικά δεδομένα μπορεί να είναι είτε ένα υποσύνολο του πραγματικού δείγματος $\mathbf{y}$ είτε ένα «φανταστικό» δείγμα που χρησιμοποιείται για να εκφράσει τις εκ των προτέρων πεποιθήσεις μας. Οι Spiegelhalter και Smith (1982) είχαν την ιδέα για τα φανταστικά δεδομένα ενώ οι Berger και Pericchi (1996) εισήγαγαν την έννοια του ***ενδογενή παράγοντα Bayes*** που βασίζεται στην ιδέα του να χρησιμοποιήσουμε ένα ελάχιστο μέρος από τα δεδομένα για εκτίμηση και τα υπόλοιπα για την επιλογή μοντέλου. Κατέληξαν στο ότι ο ενδογενής παράγοντας Bayes αντιστοιχεί στον πραγματικό παράγοντα Bayes αν χρησιμοποιηθεί η κατάλληλη εκ των προτέρων κατανομή. Αυτές οι priors καλούνται ***ενδογενής εκ των προτέρων κατανομές***

$$f(\boldsymbol{\theta}_{(m)}) = \int f(\boldsymbol{\theta}_{(m)} \mid m, \mathbf{y}^*) f(\mathbf{y}^* \mid m_0) d\mathbf{y}^*,$$

όπου m_0 ένα μοντέλο αναφοράς, π.χ το σταθερό μοντέλο (Null).

Παρουσιάζουμε στη συνέχεια μια οικογένεια εκ των προτέρων κατανομών που σχετίζεται με την Zellner's g-prior, Zellner (1986). Περιγράφουμε μια εκδοχή που βασίζεται σε φανταστικά δεδομένα καθώς και την power prior των Ibrahim και Chen (2000) και Chen, Ibrahim και Shao (2000). Οι εκ των προτέρων αυτές κατανομές παρουσιάζονται για Κανονικά, Poisson και Διωνυμικά μοντέλα αλλά κάτι αντίστοιχο μπορεί να οριστεί και για τα υπόλοιπα μοντέλα χρησιμοποιώντας παρόμοια επιχειρήματα.

Η ***Zellner's g-prior*** μπορεί απλά να οριστεί αν θέσουμε $\mathbf{V}_{(m)} = (\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1}$ στην συζυγή εκ των προτέρων κατανομή $f(\boldsymbol{\beta}_{(m)} \mid \sigma^2, m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, c^2 \mathbf{V}_{(m)} \sigma^2)$, για το Κανονικό μοντέλο. Η παραπάνω prior μπορεί να ερμηνευτεί χρησιμοποιώντας την ***power prior*** των Ibrahim και Chen (2000) και Chen et al (2000). Ας θεωρήσουμε φανταστικά δεδομένα $\mathbf{y}^*$ τα οποία έχουν αποκτηθεί από τον ίδιο πίνακα σχεδιασμού $\mathbf{X}_{(m)}$. Τότε θέτουμε δική μας εκ των προτέρων κατανομή ανάλογη με την πιθανοφάνεια πού αποκτούμε από τα φανταστικά δεδομένα, υψωμένη στη δύναμη $1/c^2$, δηλαδή

$$f(\boldsymbol{\beta}_{(m)} \mid \mathbf{y}^*, m) \propto \left[f(\mathbf{y}^* \mid \boldsymbol{\beta}_{(m)}, m) \right]^{1/c^2}.$$

Η παραπάνω εκ των προτέρων κατανομή χρησιμοποιεί n/c^2 δεδομένα. Στο Κανονικό μοντέλο η κατανομή αυτή καταλήγει στην ακόλουθη μορφή

$$f(\boldsymbol{\beta}_{(m)} \mid \mathbf{y}^*, m) \sim N\left((\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1} \mathbf{X}_{(m)} \mathbf{y}^*, c^2 (\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1} \sigma^2 \right), \quad (3.9.6.1)$$

όπου $(\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1} \mathbf{X}_{(m)} \mathbf{y}^*$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας των συντελεστών $\boldsymbol{\beta}_{(m)}$ αν θεωρήσουμε ως παρατηρήσεις τα φανταστικά δεδομένα $\mathbf{y}^*$. Καταλήγουμε λοιπόν στο συμπέρασμα ότι η (3.9.6.1) αποτελεί την Zellner g-prior για φανταστικά δεδομένα ίσα με $\mathbf{y}^* = \mathbf{X}_{(m)} \cdot \boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}$. Εάν καμία εκ των προτέρων πληροφορία δεν είναι διαθέσιμη, είναι φυσικό να «κεντράρουμε» τις εκ των προτέρων πεποιθήσεις μας για το διάνυσμα των συντελεστών $\boldsymbol{\beta}_{(m)}$ γύρω από το μηδέν και ως εκ τούτου θέτουμε $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \mathbf{0}$. Επιπλέον, μία επιλογή για το c^2 είναι να θέσουμε $c^2 = n$ το οποίο αντιστοιχεί στο να προσθέσουμε εκ των προτέρων πληροφορία ίση την πληροφορία που προσφέρει ένα δεδομένο. Με βάση λοιπόν αυτά παίρνουμε την ακόλουθη prior

$$f(\boldsymbol{\beta}_{(m)} \mid \mathbf{y}^*, m) \sim N\left(\mathbf{0}, n(\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1} \sigma^2\right).$$

Με την παραπάνω θεώρηση, εκ των προτέρων υποστηρίζουμε το απλούστερο μοντέλο αλλά με ελάχιστο τρόπο μιας και η prior που θέτουμε μετράει μόνο για ένα δεδομένο στην τελική ανάλυση που θα κάνουμε. Αυτή η προσέγγιση είναι επίσης λογική στα πλαίσια της αρχής της «οικονομίας» του μοντέλου. Οι εκ των υστέρων πιθανότητες των μοντέλων και οι διάφορες παραλλαγές του παράγοντα Bayes επιβάλουν ποινή στην πιθανοφάνεια του μοντέλου για τις παρεκκλίσεις από τα πραγματικά δεδομένα και την εκ των προτέρων κατανομή. Εφόσον η παραπάνω prior κατανομή μπορεί να προκύψει χρησιμοποιώντας ένα σύνολο από ελάχιστα σταθμισμένα φανταστικά δεδομένα, τα οποία πλήρως υποστηρίζουν το σταθερό μοντέλο, παρέχει μια λογική εκ των προτέρων υποστήριξη σε περισσότερο «οικονομικά» μοντέλα.

Παρόμοια επιχειρήματα χρησιμοποιήθηκαν από τους Fouskakis et al (2008) για να υιοθετήσουν την εκ των προτέρων των Ntzoufras et al (2003)

$$f(\boldsymbol{\beta}_{(m)} \mid m) \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}}, n[\mathbf{I}_{\boldsymbol{\beta}_{(m)}}]^{-1}), \quad (3.9.6.2)$$

για την λογιστική παλινδρόμηση, όπου $\mathbf{I}_{\boldsymbol{\beta}_{(m)}}$ είναι ο πίνακας πληροφορίας που δίνεται από τη σχέση

$$\mathbf{I}_{\boldsymbol{\beta}_{(m)}} = \mathbf{X}_{(m)}^T \hat{\mathbf{H}}_{(m)} \mathbf{X}_{(m)},$$

με τον $\hat{\mathbf{H}}_{(m)}$ να είναι ένας διαγώνιος πίνακας ο οποίος στη Διωνυμική περίπτωση παίρνει τη μορφή

$$\hat{\mathbf{H}}_{(m)} = diag\left(N_i \pi_i (1-\pi_i)\right),$$

όπου N_i ο αριθμός των επαναλήψεων και π_i η πιθανότητα επιτυχίας.

Αυτή είναι η ***εκ των προτέρων κατανομή μοναδιαίας πληροφορίας*** (unit information prior), που εισήχθη από τους Kass και Wasserman (1995). Εδώ χρησιμοποιούμε αυτή την prior σαν βάση, αλλά καθορίζουμε τα π_i στον πίνακα πληροφορίας σύμφωνα με τις δικές μας εκ των προτέρων πληροφορίες. Με αυτόν τον τρόπο αποφεύγουμε, έστω και ελάχιστα, την επαναχρησιμοποίηση των δεδομένων στη εκ των προτέρων κατανομή. Παρόμοια με την περίπτωση του Κανονικού μοντέλου, όταν μικρή η μηδαμινή εκ των προτέρων πληροφορία είναι διαθέσιμη, μια λογική επιλογή για την prior μέση τιμή του διανύσματος $\boldsymbol{\beta}_{(m)}$ είναι να θέσουμε $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(m)}} = \mathbf{0}$. Αυτό αντιστοιχεί στην υπόθεση ότι μια λογική εκ των προτέρων εκτίμηση (όταν καμία πληροφορία δεν είναι διαθέσιμη), για όλες τις πιθανότητες του Διωνυμικού μοντέλου παλινδρόμησης, είναι $\pi_i = 1/2$. Με αυτή την επιλογή και για δυαδικά δεδομένα ($N_i = 1$ για όλα τα i), η εξίσωση (3.9.6.2) απλουστεύεται παίρνοντας την ακόλουθη μορφή

$$f(\boldsymbol{\beta}_{(m)} \mid m) \sim N\left(\mathbf{0}, 4n\,(\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1}\right), \quad (3.9.6.3)$$

όπου n είναι ο συνολικός αριθμός των δοκιμών Bernoulli. Αυτή η εκ των προτέρων κατανομή μπορεί επίσης να «αιτιολογηθεί» χρησιμοποιώντας την power prior προσέγγιση των Chen et al (2000). Μετά τον καθορισμό του πίνακα σχεδιασμού $\mathbf{X}_{(m)}$ για κάθε μοντέλο m, θεωρούμε ένα σύνολο από φανταστικά δεδομένα $y_i^* = N_i$ και $N_i^* = 2N_i$ για $i = 1,...,n$, που δίνουν πιθανότητες $1/2$ για όλες τις Διωνυμικές παρατηρήσεις i και επομένως υποστηρίζεται το απλούστερο μοντέλο. Θεωρούμε μια εκ των προτέρων κατανομή η οποία παράγεται χρησιμοποιώντας την πιθανοφάνεια αυτών των φανταστικών δεδομένων

$$f(\boldsymbol{\beta}_{(m)} \mid m, \mathbf{y}^*) \propto \left(\prod_{i=1}^{n} \pi_i^{N_i} \cdot (1-\pi_i)^{N_i} \right)^{1/2N}, \quad (3.9.6.4)$$

όπου $N = \sum_{i=1}^{n} N_i$. Σημειωτέον, όταν τα δυαδικά δεδομένα έχουν θεωρηθεί, τότε $N_i = 1$ για όλα τα $i = 1,...,n$, καταλήγοντας στην μορφή των Fouskakis et al (2008). Χρησιμοποιώντας την παραπάνω prior, η εκ των υστέρων κατανομή δίνεται από την ακόλουθη σχέση

$$f(\boldsymbol{\beta}_{(m)} \mid m, \mathbf{y}) \propto \prod_{i=1}^{n} \pi_i^{y_i + \frac{1}{2N}} \cdot (1-\pi_i)^{N_i\left(1+\frac{1}{N}\right) - \left(y_i + \frac{N_i}{2N}\right)}.$$

Συνεπώς, αυτό είναι ισοδύναμο με το να αποκτήσουμε πληροφορία από $\sum_{i=1}^{n}\left(N_i + \frac{N_i}{n} \right) = (N+1)$ δοκιμές Bernoulli, αντί για N όταν χρησιμοποιούμε επίπεδη (μη πληροφοριακή) εκ των προτέρων κατανομή. Έτσι λοιπόν, η εκ των προτέρων (3.9.6.4) εισάγει στην εκ των υστέρων κατανομή επιπρόσθετη πληροφορία ισοδύναμη με ένα δεδομένο. Με εφαρμογή της προσέγγισης Laplace στην (3.9.6.4), προκύπτει το ακόλουθο προσεγγιστικό αποτέλεσμα

$$f(\boldsymbol{\beta}_{(m)} \mid m, \mathbf{y}^*) \sim N\left(\hat{\boldsymbol{\beta}}^*_{(m)}, 2N[\mathbf{I}_{\hat{\boldsymbol{\beta}}_{(m)}}]^{-1} \right),$$

όπου $\hat{\boldsymbol{\beta}}^*_{(m)}$ είναι οι εκτιμητές μέγιστης πιθανοφάνειας με δεδομένα y_i^* και $\mathbf{I}_{\hat{\boldsymbol{\beta}}^*_{(m)}}$ είναι ο παρατηρούμενος πίνακας πληροφορίας που δίνεται από τη σχέση

$$\mathbf{I}_{\hat{\boldsymbol{\beta}}^*_{(m)}} = \mathbf{X}^T_{(m)} diag\left(2N_i \hat{\pi}_i^* (1-\hat{\pi}_i^*) \right) \mathbf{X}_{(m)},$$

όπου $\hat{\pi}_i^* = \left(1 + \exp(-\mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}^*_{(m)} \right)^{-1}$ είναι η κατάλληλη πιθανότητα επιτυχίας για όλα τα i υπό το m μοντέλο όταν παρατηρούμε τα δεδομένα $\mathbf{y}^*$. Με βάση αυτά τα φανταστικά δεδομένα, $\hat{\boldsymbol{\beta}}_{(m)} = \mathbf{0}$ και $\hat{\pi}_i = 1/2$ για όλα τα i και έτσι προκύπτει ότι $\mathbf{I}_{\hat{\boldsymbol{\beta}}_{(m)}} = \frac{1}{2}\left(\mathbf{X}^T_{(m)} \mathbf{X}_{(m)} \right)$. Συνεπώς καταλήγουμε στην εκ των προτέρων κατανομή της σχέσης (3.9.6.3).

Παρόμοια επιχειρήματα μπορούν να χρησιμοποιηθούν για την περίπτωση που έχουμε Poisson μοντέλο. Αν όλα τα φανταστικά δεδομένα y_i^* είναι ίσα με ένα, τότε η αντίστοιχη εκ των προτέρων η οποία λαμβάνει υπ'όψιν ένα επιπρόσθετο δεδομένο, δίνεται από τη σχέση

$$f(\boldsymbol{\beta}_{(m)} \mid m) \sim N\left(\mathbf{0}, n(\mathbf{X}_{(m)}^T \mathbf{X}_{(m)})^{-1}\right). \quad (3.9.6.5)$$

Η παραπάνω εκ των προτέρων κατανομή μπορεί να είναι εξαιρετικά πληροφοριακή για τη σταθερή παράμετρο β_0 και για το λόγο αυτό μπορούμε να θέσουμε μία ξεχωριστή prior για τον συντελεστή β_0 και να χρησιμοποιήσουμε την prior (3.9.6.5) για τους υπόλοιπους συντελεστές $\boldsymbol{\beta}_{\backslash 0(m)} := (\beta_1, ..., \beta_p)$, δηλαδή να έχουμε

$$f(\boldsymbol{\beta}_{\backslash 0(m)} \mid m) \sim N\left(\mathbf{0}, n(\mathbf{X}_{\backslash 0(m)}^T \mathbf{X}_{\backslash 0(m)})^{-1}\right),$$

όπου $\mathbf{X}_{\backslash 0(m)}$ είναι ο πίνακας σχεδιασμού μετά την αφαίρεση της πρώτης στήλης με τα '1' η οποία αντιστοιχεί στον σταθερό συντελεστή β_0.

ΚΕΦΑΛΑΙΟ 4

MCMC ΤΕΧΝΙΚΕΣ ΣΤΗΝ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

4.1 Εισαγωγή

Οι δυσκολίες στην Μπεϋζιανή επιλογή μοντέλων και μεταβλητών σχετίζονται με τον υπολογισμό των ολοκληρωμάτων που εμφανίζονται στις εκ των υστέρων πιθανότητες των μοντέλων. Αυτά τα ολοκληρώματα μπορούν να υπολογιστούν αναλυτικά μόνο σε κάποιες εξειδικευμένες περιπτώσεις. Έτσι, υπάρχει η ανάγκη για χρήση MCMC μεθόδων για την προσομοίωση δείγματος από την από κοινού εκ των υστέρων κατανομή $f(\boldsymbol{\gamma}, \boldsymbol{\beta}_{(\gamma)} \mid \mathbf{y})$. Ακόμα όμως και στην περίπτωση όπου μπορούν να υπάρξουν αναλυτικές εκφράσεις για ορισμένα μεγέθη, όπως η περιθώρια πιθανοφάνεια $f(\boldsymbol{\gamma} \mid \mathbf{y})$ και ο παράγοντας Bayes (BF), η χρήση MCMC μεθόδων για την επιλογή μοντέλου κρίνεται επιτακτική μιας και σε πολλά προβλήματα ο χώρος M των υποψήφιων μοντέλων είναι πολύ μεγάλος.

Οι MCMC μέθοδοι έχουν αναδειχτεί σε ένα πολύτιμο εργαλείο για την εκ των υστέρων διερεύνηση σε προβλήματα Μπεϋζιανης επιλογής μοντέλων και μεταβλητών. Τέτοιες μέθοδοι χρησιμοποιούνται για την δημιουργία μιας Μαρκοβιανής Αλυσίδας $\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \ldots$ με στάσιμη κατανομή την εκ των υστέρων κατανομή $f(\boldsymbol{\gamma} \mid \mathbf{y})$. Σε εφαρμογές όπου ο αναλυτικός υπολογισμός της από κοινού εκ των υστέρων κατανομής $f(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2 \mid \mathbf{y})$, δεν είναι διαθέσιμος, γίνεται η προσομοίωση μιας Μαρκοβιανής Αλυσίδας της μορφής

$$\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)}, \sigma^{2(1)}, \quad \boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)} \sigma^{2(2)}, \ldots ,$$

που συγκλίνει στην από κοινού εκ των υστέρων κατανομή $f(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{y})$. Πιο αναλυτικά, με τις μεθόδους MCMC προσομοιώνονται παρατηρήσεις από την από κοινού εκ των υστέρων κατανομή, οι οποίες συνιστούν μια αλυσίδα Markov. Η ιδιότητες των τιμών της Μαρκοβιανής αλυσίδας δίνει τη δυνατότητα στην επόμενη τιμή κάθε παρατήρησης να εξαρτάται από την παρούσα τιμή, όχι όμως από την προηγούμενη. Το πλεονέκτημα αυτής της μεθόδου, είναι ότι όταν ο αλγόριθμος της προσομοίωσης επαναλαμβάνεται πολλές φορές, η προσέγγιση της εκ των υστέρων κατανομής βελτιώνεται σε κάθε βήμα. Έτσι, δίνεται η ικανότητα στους Μπεϋζιανούς να εκτιμούν με ακρίβεια τις εκ των υστέρων κατανομές. Οι πιο δημοφιλείς μέθοδοι MCMC, είναι ο αλγόριθμος Metropolis-Hastings, και ο Gibbs Sampler. Περισσότερες λεπτομέρειες από τους George και McCulloch (1993, 1996), Bernado και Smith (1994), Green (1995), Gelman et al (1998), Ntzoufras (1999), Lopes (2002).

Αν ένα δείγμα $(\boldsymbol{\gamma}^{(t')}, \boldsymbol{\beta}^{(t')},\ t' = 1, \ldots, t)$ μπορεί να παραχθεί από αυτή την κατανομή, τότε οι εκ των υστέρων πιθανότητες των μοντέλων μπορούν να εκτιμηθούν από τη σχέση

$$\hat{f}(\boldsymbol{\gamma}) = \frac{1}{t}\sum_{t'=1}^{t} I(\boldsymbol{\gamma}^{(t')} = \boldsymbol{\gamma}), \quad \boldsymbol{\gamma} \in M,$$

όπου $I(.)$ είναι η δείκτρια συνάρτηση. Επίσης η εκτίμηση για τους συντελεστές $\boldsymbol{\beta}_{(\gamma)}$ του μοντέλου $\boldsymbol{\gamma}$ μπορεί να γίνει με παραγωγή δείγματος από την περιθώρια εκ των υστέρων κατανομή $f(\boldsymbol{\beta}_{(\gamma)} \mid \boldsymbol{\gamma}, \mathbf{y})$.

4.2 Το Κανονικό γραμμικό μοντέλο

Υποθέτουμε ότι Y είναι η μεταβλητή απόκρισης και $X_0, X_1, ..., X_p$ οι επεξηγηματικές μεταβλητές και επίσης έχουμε συλλέξει n παρατηρήσεις για κάθε μια από τις προηγούμενες μεταβλητές. Το πρόβλημα της επιλογής μεταβλητών προκύπτει όταν θέλουμε να μοντελοποιήσουμε τη σχέση μεταξύ της Y και ενός υποσυνόλου των $X_0, X_1, ..., X_p$ αλλά υπάρχει αβεβαιότητα για το ποιο υποσύνολο θα χρησιμοποιήσουμε. Μια τέτοια κατάσταση είναι ιδιαίτερα ενδιαφέρουσα κυρίως όταν το p είναι μεγάλο (άρα και τα υποψήφια πιθανά μοντέλα) και το σύνολο $X_0, X_1, ..., X_p$ περιέχει αρκετές μη σημαντικές μεταβλητές.

Στο πρόβλημα επιλογής μεταβλητών κάθε θεωρούμενο μοντέλο αντιστοιχεί σε ένα ευδιάκριτο υποσύνολο των $X_0, X_1, ..., X_p$. Αυτό το πρόβλημα είναι περισσότερο οικείο στο πλαίσιο της πολλαπλής παλινδρόμησης όπου η προσοχή εστιάζεται στα Κανονικά γραμμικά μοντέλα. Πολλές από τις θεμελιώδεις εξελίξεις στην επιλογή μεταβλητών έχουν εμφανιστεί στο πλαίσιο των Κανονικών γραμμικών μοντέλων, επειδή ο αναλυτικός προσδιορισμός ορισμένων μεγεθών όπως η περιθώρια πιθανοφάνεια $f(\mathbf{y} \mid m)$ με χρήση συζυγών κατανομών, συνεπάγεται τη μείωση του υπολογιστικού κόστους και επειδή παρέχει μια απλή πρώτη προσέγγιση για περισσότερο πολύπλοκα μοντέλα.

Ας υποθέσουμε τώρα ότι το Κανονικό γραμμικό μοντέλο χρησιμοποιείται για να συσχετίσει την εξάρτηση της Y από τις $X_0, X_1, ..., X_p$ μεταβλητές, δηλαδή

$$\mathbf{Y} \mid \mathbf{X} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right), \ (4.2.1)$$

όπου $\mathbf{X} = \left[\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_p\right]$, $\boldsymbol{\beta}$ είναι ένα $(p+1) \times 1$ διάνυσμα με τους άγνωστους συντελεστές της παλινδρόμησης και σ^2 είναι η άγνωστη παράμετρος διασποράς. Είναι πολύ βολικό να αναπαραστήσουμε κάθε ένα από τα 2^p μοντέλα m με το διάνυσμα

$$\boldsymbol{\gamma} = \left[\gamma_0, \gamma_1, ..., \gamma_p\right]^{\mathrm{T}},$$

όπου $\gamma_j = 0$ αν η μεταβλητή X_J δεν βρίσκεται στο μοντέλο και $\gamma_j = 1$ στην αντίθετη περίπτωση. Ορίζουμε επίσης $q_\gamma = \boldsymbol{\gamma}^{\mathrm{T}} \cdot \mathbf{1} = \sum_{j=0}^{p} \gamma_j$ που δηλώνει το πλήθος των επεξηγηματικών μεταβλητών του μοντέλου $\boldsymbol{\gamma}$.

Για τον καθορισμό της εκ των προτέρων κατανομής στο χώρο των μοντέλων M, στις περισσότερες εφαρμογές επιλογής μεταβλητών χρησιμοποιούνται ανεξάρτητες εκ των προτέρων κατανομές της μορφής

$$f(\boldsymbol{\gamma}) = \prod_{j=0}^{p} \omega_j^{\gamma_j} \left(1-\omega_j\right)^{1-\gamma_j}, \qquad (4.2.2)$$

οι οποίες ουσιαστικά μειώνουν το υπολογιστικό κόστος και συχνά οδηγούν σε λογικά αποτελέσματα, Clyde, Desimone και Parmigiani (1996), George και McCulloch (1993, 1997), Raftery, Madigan και Hoeting (1997) και Smith και Kohn (1996). Κάτω από αυτή την εκ των προτέρων κατανομή, κάθε X_J εισάγεται στο μοντέλο ανεξάρτητα από τις άλλες επεξηγηματικές μεταβλητές με πιθανότητα

$$f(\gamma_j = 1) = 1 - f(\gamma_j = 0) = \omega_j .$$

Μια χρήσιμη απλούστευση της (4.2.2) είναι να θέσουμε $\omega_j = \omega$, για κάθε $j = 0,1,...,p$ και ως εκ τούτου να έχουμε

$$f(\boldsymbol{\gamma}) = \omega^{q_\gamma} (1-\omega)^{p-q_\gamma}, \quad (4.2.3)$$

όπου σ'αυτή την περίπτωση η υπερπαράμετρος ω είναι η εκ των προτέρων πιθανότητα να βρίσκεται η μεταβλητή X_J , $j = 0,1,...,p$ στο μοντέλο $\boldsymbol{\gamma}$. Ειδικότερα, θέτοντας $\omega = 1/2$, οδηγούμαστε στη πιο δημοφιλή ίσως εκ των προτέρων κατανομή (Ομοιόμορφη) στο χώρο των μοντέλων

$$f(\boldsymbol{\gamma}) = \frac{1}{2^p},$$

η οποία χρησιμοποιείται συνήθως για να εκφράσει την εκ των προτέρων άγνοιά μας. Ωστόσο, αυτή η εκ των προτέρων κατανομή δίνει περισσότερη βαρύτητα σε μοντέλα μεγέθους $q_\gamma = p/2$ επειδή υπάρχουν περισσότερα τέτοια στο σύνολο M. Αν θέλουμε να αυξήσουμε τη βαρύτητα πιο φειδωλών μοντέλων, μπορούμε να θέσουμε μια πολύ μικρή τιμή στο ω. Εναλλακτικά, μπορούμε να ορίσουμε μια εκ των προτέρων κατανομή και στο ω, για παράδειγμα $\omega \sim B(\alpha,\beta)$ και σε συνδυασμό με την (4.2.3) οδηγούμαστε στο αποτέλεσμα :

$$f(\boldsymbol{\gamma}) = \frac{B\left(\alpha + q_\gamma, \beta + p - q_\gamma\right)}{B(\alpha,\beta)}, \qquad (4.2.4)$$

όπου είναι $B(\alpha,\beta)$ η Βήτα κατανομή. Πιο γενικά, μπορούμε να ορίσουμε μια εκ των προτέρων στη διάσταση του μοντέλου q_γ έστω $h(q_\gamma)$ και τότε

$$f(\boldsymbol{\gamma}) = \binom{p}{q_\gamma}^{-1} h\left(q_\gamma\right). \quad (4.2.5)$$

Η (4.2.4) αποτελεί τότε μια ειδική περίπτωση της (4.2.5).

Εκτός από τον καθορισμό εκ των προτέρων κατανομών στο χώρο των μοντέλων, πρέπει να ορίσουμε εκ των προτέρων κατανομές και στους συντελεστές του Κανονικού γραμμικού μοντέλου καθώς και στην παράμετρο διασποράς σ^2 που αμφότερα θεωρούνται άγνωστα. Ο στόχος είναι να απορρίψουμε τις μεταβλητές X_j για τις οποίες $\beta_J = 0$ ή β_j πολύ κοντά στο 0 στη σχέση (4.1.1) . Στην πραγματικότητα, το πρόβλημα πλέον είναι να επιλέξουμε ένα υποσύνολο των επεξηγηματικών μεταβλητών και να έχουμε:

$$f(\mathbf{y} \mid \boldsymbol{\beta}_{(\gamma)}, \sigma^2, \boldsymbol{\gamma}) = N_n(\mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}, \sigma^2\mathbf{I}), \quad (4.2.6)$$

όπου $\mathbf{X}_{(\gamma)}$ είναι ένας $n \times q_\gamma$ πίνακας του οποίου οι στήλες αντιστοιχούν στο υποσύνολο των $X_0, X_1, ..., X_p$ που συνθέτουν το μοντέλο $\boldsymbol{\gamma}$. Επίσης $\boldsymbol{\beta}_{(\gamma)}$ είναι ένα $q_\gamma \times 1$ διάνυσμα με τους αντίστοιχους άγνωστους συντελεστές του μοντέλου και σ^2 η άγνωστη παράμετρος διασποράς. Εδώ το ζεύγος $(\boldsymbol{\beta}_{(\gamma)}, \sigma^2)$ παίζει το ρόλο της παραμέτρου $\boldsymbol{\theta}_{(m)} \equiv \boldsymbol{\theta}_{(\gamma)}$ όπως συμβολιζόταν στα προηγούμενα κεφάλαια. Ο πιο δημοφιλής και κοινός εκ των προτέρων καθορισμός για τις άγνωστες παραμέτρους $\boldsymbol{\beta}_{(\gamma)}, \sigma^2$ ενός μοντέλου $\boldsymbol{\gamma}$ διάστασης q_γ ιδιαίτερα σε μεγάλα προβλήματα, είναι η συζυγής Normal-Inverse-Gamma(NIG) κατανομή, η οποία συντίθεται από μια $q_\gamma -$διάστατη Κανονική κατανομή για το $\boldsymbol{\beta}_{(\gamma)}$

$$f(\boldsymbol{\beta}_{(\gamma)} \mid \sigma^2, \boldsymbol{\gamma}) = N_{q_\gamma}(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \sigma^2\boldsymbol{\Sigma}_{(\gamma)}), \quad (4.2.7)$$

συνδυαζόμενη με μια Inverse-Gamma (αντίστροφη Γάμμα) κατανομή για το σ^2

$$f(\sigma^2 \mid \boldsymbol{\gamma}) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(-\frac{b}{\sigma^2}\right) := IG(a,b),\ a,b>0. \quad (4.2.8)$$

Ένα πολύ σημαντικό χαρακτηριστικό του συνδυασμού των δύο εκ των προτέρων κατανομών (4.2.7) και (4.2.8) είναι το γεγονός ότι οδηγούν σε αναλυτικές εκφράσεις των εκ των υστέρων κατανομών $f(\boldsymbol{\beta}_{(\gamma)} \mid \sigma^2, \boldsymbol{\gamma}, \mathbf{y})$, $f(\sigma^2 \mid \boldsymbol{\gamma}, \mathbf{y})$ καθώς και της περιθώριας πιθανοφάνειας $f(\mathbf{y} \mid \boldsymbol{\gamma})$. Περιγράφουμε στη συνέχεια τη διαδικασία με την οποία φτάνουμε στον υπολογισμό αυτών των κατανομών:

Η από κοινού εκ των προτέρων κατανομή για το ζεύγος $(\boldsymbol{\beta}_{(\gamma)}, \sigma^2)$ γράφεται ως εξής:

$$f(\boldsymbol{\beta}_{(\gamma)}, \sigma^2 \mid \boldsymbol{\gamma}) = f(\boldsymbol{\beta}_{(\gamma)} \mid \sigma^2, \boldsymbol{\gamma}) f(\sigma^2 \mid \boldsymbol{\gamma}) = N_{q_\gamma}(\mu_{\boldsymbol{\beta}_{(\gamma)}}, \sigma^2\boldsymbol{\Sigma}_{(\gamma)}) \times IG(a,b) = NIG(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \boldsymbol{\Sigma}_{(\gamma)}, a, b)$$

$$= \frac{b^a}{(2\pi)^{q_\gamma/2}\left|\boldsymbol{\Sigma}_{(\gamma)}\right|^{1/2}\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+q_\gamma/2+1} \times \exp\left[-\frac{1}{\sigma^2}\left\{b + \frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \boldsymbol{\Sigma}_{(\gamma)}^{-1}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)\right\}\right]$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{a+q_\gamma/2+1} \times \exp\left[-\frac{1}{\sigma^2}\left\{b + \frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \boldsymbol{\Sigma}_{(\gamma)}^{-1}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)\right\}\right].$$

Ολοκληρώνοντας την παραπάνω *NIG* εκ των προτέρων κατανομή ως προς σ^2 παίρνουμε το ακόλουθο ενδιαφέρον αποτέλεσμα:

$$
\begin{aligned}
&\int NIG(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \boldsymbol{\Sigma}_{(\gamma)}, a, b) d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{q_\gamma/2} \left|\boldsymbol{\Sigma}_{(\gamma)}\right|^{1/2} \Gamma(a)} \int \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left[-\frac{1}{\sigma^2}\left\{b + \frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \boldsymbol{\Sigma}_{(\gamma)}^{-1} \left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)\right\}\right] d\sigma^2 \\
&= \frac{b^a \Gamma\left(a + \frac{q_\gamma}{2}\right)}{(2\pi)^{q_\gamma/2} \left|\boldsymbol{\Sigma}_{(\gamma)}\right|^{1/2} \Gamma(a)} \left[b + \frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \boldsymbol{\Sigma}_{(\gamma)}^{-1} \left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)\right]^{-\left(a + \frac{q_\gamma}{2}\right)} \\
&= \frac{\Gamma\left(a + \frac{q_\gamma}{2}\right)}{\pi^{q_\gamma/2} \left|(2a)\frac{b}{a}\boldsymbol{\Sigma}_{(\gamma)}\right|^{1/2} \Gamma(a)} \left[1 + \frac{\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T (\frac{b}{a} \cdot \boldsymbol{\Sigma}_{(\gamma)})^{-1} \left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)}{2a}\right]^{-\left(\frac{2a+q_\gamma}{2}\right)}.
\end{aligned}
$$

Αυτή είναι η πολυμεταβλητή κατανομή Student:

$$
MVSt_\nu\left(\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \mathbf{Z}\right) = \frac{\Gamma\left(\frac{\nu + q_\gamma}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi^{q_\gamma/2} \left|\nu \boldsymbol{\Sigma}_{(\gamma)}\right|^{1/2}} \left[1 + \frac{\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \mathbf{Z}^{-1} \left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)}{\nu}\right]^{-\frac{\nu + q_\gamma}{2}},
$$

με $\nu = 2a$ και $\mathbf{Z} = \left(\frac{b}{a}\right)\boldsymbol{\Sigma}_{(\gamma)}$.

Η πιθανοφάνεια του μοντέλου δίνεται από την παρακάτω σχέση:

$$
f(\mathbf{y} \mid \boldsymbol{\beta}_{(\gamma)}, \sigma^2, \boldsymbol{\gamma}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}\right)^T \left(\mathbf{y} - \mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}\right)\right].
$$

Με χρήση του Θεωρήματος Bayes μπορούμε να υπολογίσουμε την από κοινού εκ των υστέρων κατανομή του $(\boldsymbol{\beta}_{(\gamma)}, \sigma^2)$ καθώς ισχύει

$$
f(\boldsymbol{\beta}_{(\gamma)}, \sigma^2 \mid \mathbf{y}, \boldsymbol{\gamma}) \propto f(\boldsymbol{\beta}_{(\gamma)}, \sigma^2 \mid \boldsymbol{\gamma}) \cdot f(\mathbf{y} \mid \boldsymbol{\beta}_{(\gamma)}, \sigma^2, \boldsymbol{\gamma}).
$$

Επιπλέον, θεωρούμε την ακόλουθη ταυτότητα από τη γραμμική άλγεβρα:

$$
\mathbf{u}^T \mathbf{A} \mathbf{u} - 2\boldsymbol{\alpha}^T \mathbf{u} = (\mathbf{u} - \mathbf{A}^{-1}\boldsymbol{\alpha})^T \mathbf{A} (\mathbf{u} - \mathbf{A}^{-1}\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T \mathbf{A}^{-1} \boldsymbol{\alpha},
$$

όπου $\mathbf{A}$ είναι ένας συμμετρικός θετικά ορισμένος (ως εκ τούτου αντιστρέψιμος) πίνακας. Με εφαρμογή αυτής της ταυτότητας προκύπτει η παρακάτω σχέση:

$$\frac{1}{\sigma^2}\left[b+\frac{1}{2}\left\{\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)^T\boldsymbol{\Sigma}_{(\gamma)}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}\right)+\left(\mathbf{y}-\mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}\right)^T\left(\mathbf{y}-\mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}\right)\right\}\right]$$

$$=\frac{1}{\sigma^2}\left[b^*+\frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)^T\boldsymbol{\Sigma}^{*-1}_{(\gamma)}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)\right],$$

με την οποία μπορούμε να γράψουμε την από κοινού εκ των υστέρων κατανομή του $(\boldsymbol{\beta}_{(\gamma)},\sigma^2)$ ως ακολούθως:

$$f(\boldsymbol{\beta}_{(\gamma)},\sigma^2\mid\mathbf{y},\boldsymbol{\gamma})\propto\left(\frac{1}{\sigma^2}\right)^{a+(n+q_\gamma)/2+1}\times\exp\left\{-\frac{1}{\sigma^2}\left[b^*+\frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)^T\boldsymbol{\Sigma}^{*-1}_{(\gamma)}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)\right]\right\},$$

όπου

$$\begin{aligned}
\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}} &=\left(\boldsymbol{\Sigma}^{-1}_{(\gamma)}+\mathbf{X}^T_{(\gamma)}\mathbf{X}_{(\gamma)}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}+\mathbf{X}^T_{(\gamma)}\mathbf{y}\right),\\
\boldsymbol{\Sigma}^*_{(\gamma)} &=\left(\boldsymbol{\Sigma}^{-1}_{(\gamma)}+\mathbf{X}^T_{(\gamma)}\mathbf{X}_{(\gamma)}\right)^{-1},\\
a^* &= a+n/2,\\
b^* &= b+\frac{1}{2}\left[\boldsymbol{\mu}^T_{\boldsymbol{\beta}_{(\gamma)}}\boldsymbol{\Sigma}^{-1}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}+\mathbf{y}^T\mathbf{y}-\boldsymbol{\mu}^{*T}_{\boldsymbol{\beta}_{(\gamma)}}\boldsymbol{\Sigma}^{*-1}_{(\gamma)}\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right].
\end{aligned}$$

Αυτή η εκ των υστέρων κατανομή μπορούμε να καταλάβουμε εύκολα ότι είναι μια $NIG(\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}},\boldsymbol{\Sigma}^*_{(\gamma)},a^*,b^*)$ αποδεικνύοντας ότι πρόκειται για συζυγή οικογένεια κατανομών για το Κανονικό γραμμικό μοντέλο. Είναι σχεδόν άμεσο ότι η περιθώρια εκ των υστέρων κατανομή του σ^2 είναι μία $IG(a^*,b^*)$ ενώ η περιθώρια εκ των υστέρων κατανομή του $\boldsymbol{\beta}_{(\gamma)}$ προκύπτει ολοκληρώνοντας την από κοινού NIG εκ των υστέρων κατανομή του $(\boldsymbol{\beta}_{(\gamma)},\sigma^2)$ ως ακολούθως:

$$\begin{aligned}
f(\boldsymbol{\beta}_{(\gamma)}\mid\mathbf{y},\boldsymbol{\gamma}) &=\int f(\boldsymbol{\beta}_{(\gamma)},\sigma^2\mid\mathbf{y},\boldsymbol{\gamma})d\sigma^2=\int NIG(\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}},\boldsymbol{\Sigma}^*_{(\gamma)},a^*,b^*)d\sigma^2\\
&\propto\int\left(\frac{1}{\sigma^2}\right)^{a^*+1}\times\exp\left\{-\frac{1}{\sigma^2}\left[b^*+\frac{1}{2}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)^T\boldsymbol{\Sigma}^{*-1}_{(\gamma)}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)\right]\right\}d\sigma^2\\
&\propto\left[1+\frac{\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)^T\boldsymbol{\Sigma}^{*-1}_{(\gamma)}\left(\boldsymbol{\beta}_{(\gamma)}-\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)}{2b^*}\right]^{-(a^*+q_\gamma/2)}.
\end{aligned}$$

Η παραπάνω κατανομή είναι μια πολυμεταβλητή Student κατανομή:

$$MVSt_{\nu^*}\left(\boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}, \mathbf{Z}^*\right) = \frac{\Gamma\left(\frac{\nu^* + q_\gamma}{2}\right)}{\Gamma\left(\frac{\nu^*}{2}\right)\pi^{q_\gamma/2}\left|\nu^*\mathbf{Z}^*\right|^{1/2}}\left[1 + \frac{\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)^T \mathbf{Z}^{*-1}\left(\boldsymbol{\beta}_{(\gamma)} - \boldsymbol{\mu}^*_{\boldsymbol{\beta}_{(\gamma)}}\right)}{\nu^*}\right]^{-\frac{\nu^*+q_\gamma}{2}},$$

με $\nu^* = 2a^*$ και $\mathbf{Z}^* = \left(\frac{b^*}{a^*}\right)\boldsymbol{\Sigma}^*_{(\gamma)}$.

Όπως είχαμε αναφέρει και προηγουμένως, η χρήση των συζυγών εκ των προτέρων κατανομών (4.2.7) και (4.2.8) οδηγεί και στον αναλυτικό υπολογισμό της περιθώριας πιθανοφάνειας $f(\mathbf{y}\,|\,\boldsymbol{\gamma})$. Πράγματι, αρκεί να παρατηρήσουμε ότι ισχύουν τα εξής:

$$\mathbf{Y} = \mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)} + \boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_1 \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right),$$
$$\boldsymbol{\beta}_{(\gamma)} = \boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}} + \boldsymbol{\varepsilon}_2,\ \boldsymbol{\varepsilon}_2 \sim N\left(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}_{(\gamma)}\right),$$

όπου τα $\boldsymbol{\varepsilon}_1$ και $\boldsymbol{\varepsilon}_2$ είναι ανεξάρτητα μεταξύ τους. Από αυτό προκύπτει ότι

$$\mathbf{Y} = \mathbf{X}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}} + \mathbf{X}_{(\gamma)}\boldsymbol{\varepsilon}_2 + \boldsymbol{\varepsilon}_1,$$

το οποίο οδηγεί με τη σειρά του στην δεσμευμένη κατανομή

$$f(\mathbf{y}\,|\,\sigma^2, \boldsymbol{\gamma}) = N_n\left(\mathbf{X}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \sigma^2(\mathbf{I}_n + \mathbf{X}_{(\gamma)}\boldsymbol{\Sigma}_{(\gamma)}\mathbf{X}^T_{(\gamma)}\right).$$

Η περιθώρια πιθανοφάνεια $f(\mathbf{y}\,|\,\boldsymbol{\gamma})$ δίνεται τελικά από το ολοκλήρωμα

$$\begin{aligned} f(\mathbf{y}\,|\,\boldsymbol{\gamma}) &= \int f(\mathbf{y}\,|\,\sigma^2, \boldsymbol{\gamma}) f(\sigma^2\,|\,\boldsymbol{\gamma})d\sigma^2 \\ &= \int N_n\left(\mathbf{X}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \sigma^2(\mathbf{I}_n + \mathbf{X}_{(\gamma)}\boldsymbol{\Sigma}_{(\gamma)}\mathbf{X}^T_{(\gamma)}\right) \times IG(a,b)d\sigma^2 \\ &= \int NIG\left(\mathbf{X}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, (\mathbf{I}_n + \mathbf{X}_{(\gamma)}\boldsymbol{\Sigma}_{(\gamma)}\mathbf{X}^T_{(\gamma)}), a, b\right)d\sigma^2 \\ &= MVSt_{2a}\left(\mathbf{X}_{(\gamma)}\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}, \frac{b}{a}(\mathbf{I}_n + \mathbf{X}_{(\gamma)}\boldsymbol{\Sigma}_{(\gamma)}\mathbf{X}^T_{(\gamma)})\right). \end{aligned}$$

Η χρήση αυτών των αναλυτικών εκφράσεων διευκολύνει την εκ των υστέρων αξιολόγηση και επιταχύνει ουσιαστικά την διερεύνηση του χώρου των μοντέλων μέσω των MCMC αλγορίθμων.

Επιστρέφοντας τώρα στην (4.2.7), η πιο συνηθισμένη εκ των προτέρων επιλογή για το μέσο $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}}$, είναι η $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}} = \mathbf{0}$. Για την επιλογή του πίνακα συνδιασποράς $\boldsymbol{\Sigma}_{(\gamma)}$, ο καθορισμός γίνεται ουσιαστικά θέτοντας $\boldsymbol{\Sigma}_{(\gamma)} = c^2\mathbf{V}_{(\gamma)}$, όπου c σταθερά και ο $\mathbf{V}_{(\gamma)}$ έχει συνήθως τη μορφή $\mathbf{V}_{(\gamma)} = \left(\mathbf{X}^{\mathbf{T}}_{(\gamma)}\mathbf{X}_{(\gamma)}\right)^{-1}$ ή $\mathbf{V}_{(\gamma)} = \mathbf{I}_{\mathbf{q}_\gamma}$ ($q_\gamma \times q_\gamma$ μοναδιαίος πίνακας). Σημειωτέον ότι κάτω από αυτή τη θεώρηση του πίνακα $\mathbf{V}_{(\gamma)}$, η υπό συνθήκη εκ των προτέρων κατανομή (4.2.7) παρέχει μια σταθερή (συνεπή) περιγραφή της αβεβαιότητας με τη λογική ότι είναι η υπό συνθήκη κατανομή για τα μη μηδενικά στοιχεία του διανύσματος $\boldsymbol{\beta}_{(\gamma)}$ δεδομένου του $\boldsymbol{\gamma}$ όταν $\boldsymbol{\beta}_{(\gamma)} \sim N_p\left(\mathbf{0}, c^2\sigma^2\left(\mathbf{X}_{(\gamma)}^{\mathrm{T}}\mathbf{X}_{(\gamma)}\right)^{-1}\right)$ ή $\boldsymbol{\beta}_{(\gamma)} \sim N_p\left(\mathbf{0}, c^2\sigma^2\mathbf{I}\right)$ αντίστοιχα. Όπως έχει αναφερθεί και στο προηγούμενο κεφάλαιο, η επιλογή $\mathbf{V}_{(\gamma)} = \left(\mathbf{X}_{(\gamma)}^{\mathrm{T}}\mathbf{X}_{(\gamma)}\right)^{-1}$ αποτελεί την Zellner's g prior (1986) ενώ θέτοντας $\mathbf{V}_{(\gamma)} = \mathbf{I}_{\mathbf{q}_\gamma}$ άμεσα προκύπτει ότι οι συντελεστές $\beta_j, j = 1,...,p$ είναι ανεξάρτητοι.

Έχοντας σταθεροποιήσει τον πίνακα $\mathbf{V}_{(\gamma)}$, ο επόμενος στόχος είναι η επιλογή του c^2 που θα πρέπει να λαμβάνει μεγάλη τιμή ώστε η εκ των προτέρων κατανομή του $\boldsymbol{\beta}_{(\gamma)}$ να είναι σχεδόν επίπεδη και έτσι να μειώνεται η εκ των προτέρων επίδραση στο εκ των υστέρων αποτέλεσμα. Ταυτόχρονα όμως, θα πρέπει να αποφύγουμε υπερβολικά μεγάλες τιμές για το c^2 επειδή η εκ των προτέρων τότε θα προσδώσει αυξημένη βαρύτητα στο μηδενικό μοντέλο καθώς $c^2 \to \infty$ ενεργοποιώντας έτσι το παράδοξο των Bartllet-Lindley, (Bartllet 1957). Διάφορες τιμές έχουν προταθεί στη βιβλιογραφία για την επιλογή του συντελεστή c^2, όπου ενδεικτικά αναφέρουμε τις τιμές $c^2 = 2.85^2$ από τους Raftery, Madigan και Hoeting (1997), $c^2 = 100$ από τους Smith και Kohn (1996), $c^2 = \max\left\{p^2, n\right\}$ από τους Fernandez, Ley και Steel (2001).

4.3 Ο αλγόριθμος MC^3 (Markov Chain Monte Carlo Model Composition)

Ο αλγόριθμος MC^3 εισήχθη από τους Madigan και York (1995) στο γραφικό προσδιορισμό επιλογής μοντέλου. Παραλλαγές του MC^3 είχαν χρησιμοποιηθεί σε Κανονικά γραμμικά μοντέλα από τον Hoeting et al (1995, 1996), τον Raftery et al (1997). Ο MC^3 είναι ένας απλός Metropolis αλγόριθμος ο οποίος βοηθάει να εξερευνηθεί ο χώρος των μοντέλων όταν ο αριθμός των υποψηφίων μοντέλων είναι πολύ μεγάλος. Ορίζουμε ως γειτονιά του μοντέλου $\boldsymbol{\gamma}$ το σύνολο $nb(\boldsymbol{\gamma})$ που περιέχει όλα τα μοντέλα που διαφέρουν από το $\boldsymbol{\gamma}$ κατά έναν όρο ή μεταβλητή. Επίσης επιλέγουμε τη συνάρτηση μεταπήδησης $j(\boldsymbol{\gamma},\boldsymbol{\gamma}')$ για όλα τα $\boldsymbol{\gamma},\boldsymbol{\gamma}' \in M$ που δείχνει την πιθανότητα του προτεινόμενου μοντέλου $\boldsymbol{\gamma}'$ όταν βρισκόμαστε στο τρέχον μοντέλο $\boldsymbol{\gamma}$. Σημειώνουμε ότι $j(\boldsymbol{\gamma},\boldsymbol{\gamma}') = |nb(\boldsymbol{\gamma})|^{-1}$, για όλα τα $\boldsymbol{\gamma}' \in nb(\boldsymbol{\gamma})$ και $j(\boldsymbol{\gamma},\boldsymbol{\gamma}') = 0$, για όλα τα $\boldsymbol{\gamma}' \notin nb(\boldsymbol{\gamma})$, όπου $|nb(\boldsymbol{\gamma})|$ είναι ο αριθμός των μοντέλων που ανήκουν στο $nb(\boldsymbol{\gamma})$. Αν η αλυσίδα βρίσκεται στην κατάσταση $\boldsymbol{\gamma}$, μετά προτείνουμε το μοντέλο $\boldsymbol{\gamma}'$ με πιθανότητα $j(\boldsymbol{\gamma},\boldsymbol{\gamma}')$ και αποδεχόμαστε το προτεινόμενο μοντέλο με πιθανότητα

$$a = \min\left(1, \frac{f(\boldsymbol{\gamma}' \mid \mathbf{y})|nb(\boldsymbol{\gamma})|}{f(\boldsymbol{\gamma} \mid \mathbf{y})|nb(\boldsymbol{\gamma}')|}\right).$$

Η παραπάνω διαδικασία συνθέτει τον ορισμό του MC^3 αλγορίθμου για τον γραφικό προσδιορισμό μοντέλου από τους Madigan και Raftery (1994). Στην περίπτωση όπου $|nb(\boldsymbol{\gamma})| = |nb(\boldsymbol{\gamma}')|$, η προηγούμενη πιθανότητα αποδοχής απλοποιείται στην έκφραση $a = \min(1, PO\boldsymbol{\gamma}',\boldsymbol{\gamma})$ όπως δόθηκε από τους Kass και Raftery (1995), Madigan et al

(1995) και Raftery et al (1997). Μπορούμε εύκολα να γενικεύσουμε τον αλγόριθμο MC^3 χρησιμοποιώντας οποιαδήποτε συνάρτηση $j(\gamma,\gamma')$ και επομένως να αποδεχτούμε το προτεινόμενο μοντέλο $\boldsymbol{\gamma}'$ με πιθανότητα

$$a=\min\left(1,\frac{f(\boldsymbol{\gamma}'\mid\mathbf{y})\,j(\boldsymbol{\gamma}',\boldsymbol{\gamma})}{f(\boldsymbol{\gamma}\mid\mathbf{y})\,j(\boldsymbol{\gamma}',\boldsymbol{\gamma})}\right). \tag{4.3.1}$$

Στην πράξη, προτείνουμε ένα νέο μοντέλο με την ακόλουθη διαδικασία:
Έστω ότι ο αλγόριθμος βρίσκεται στο τρέχων μοντέλο $\boldsymbol{\gamma}$. Για $j=1,2,...,p$ προτείνουμε το νέο μοντέλο $\boldsymbol{\gamma}'$ με στοιχεία

$$\gamma'_j=1-\gamma_j$$
$$\text{και}$$
$$\gamma'_k=\gamma_k \text{ για } k\neq j,$$

με πιθανότητα 1.
Αποδεχόμαστε το προτεινόμενο μοντέλο $\boldsymbol{\gamma}'$ με πιθανότητα

$$a=\min\left\{1,\frac{f(\boldsymbol{\gamma}'\mid\mathbf{y})}{f(\boldsymbol{\gamma}\mid\mathbf{y})}\right\}=\min\left\{1,PO_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}\right\}.$$

Σε περιπτώσεις όπου οι εκ των υστέρων λόγοι πιθανοτήτων (ή ο παράγοντας Bayes) δεν μπορούν να υπολογιστούν αναλυτικά, μπορούμε να χρησιμοποιήσουμε προσεγγιστικές τεχνικές και συγκεκριμένα την προσέγγιση BIC ή Laplace. Πιο αναλυτικά όταν βρισκόμαστε στο μοντέλο $\boldsymbol{\gamma}$ και προτείνουμε το μοντέλο $\boldsymbol{\gamma}'$, η πιθανότητα αποδοχής δίνεται από τη σχέση

$$a=\min\left(1,\frac{\left|\mathbf{X}^T_{(\gamma')}\hat{\mathbf{H}}_{(\gamma')}\mathbf{X}_{(\gamma')}\right|^{\frac{1}{2}}f(\mathbf{y}\mid\hat{\boldsymbol{\beta}}_{(\gamma')},\boldsymbol{\gamma}')f(\hat{\boldsymbol{\beta}}_{(\gamma')}\mid\boldsymbol{\gamma}')j(\boldsymbol{\gamma}',\boldsymbol{\gamma})}{\left|\mathbf{X}^T_{(\gamma)}\hat{\mathbf{H}}_{(\gamma)}\mathbf{X}_{(\gamma)}\right|^{\frac{1}{2}}f(\mathbf{y}\mid\hat{\boldsymbol{\beta}}_{(\gamma)},\boldsymbol{\gamma})f(\hat{\boldsymbol{\beta}}_{(\gamma)}\mid\boldsymbol{\gamma})j(\boldsymbol{\gamma},\boldsymbol{\gamma}')}\left(2\pi\right)^{\frac{[d(\gamma')-d(\gamma)]}{2}}\right),$$

όταν υιοθετούμε την προσέγγιση Laplace και από τη σχέση

$$a=\min\left(1,\frac{f(\mathbf{y}\mid\hat{\boldsymbol{\beta}}_{(\gamma')},\boldsymbol{\gamma}')j(\boldsymbol{\gamma}',\boldsymbol{\gamma})}{f(\mathbf{y}\mid\hat{\boldsymbol{\beta}}_{(\gamma)},\boldsymbol{\gamma})j(\boldsymbol{\gamma},\boldsymbol{\gamma}')}\,n^{-\frac{[d(\gamma')-d(\gamma)]}{2}}\right),$$

αν χρησιμοποιήσουμε την προσέγγιση BIC.

4.4 Αριθμητικές εφαρμογές

Στην ενότητα αυτή παρουσιάζονται τρεις εφαρμογές του αλγορίθμου MC^3 εκ των οποίων οι δύο αφορούν Κανονικά γραμμικά μοντέλα ενώ η τρίτη αναφέρεται σε ένα πρόβλημα λογιστικής παλινδρόμησης.

4.4.1 Κανονικό γραμμικό μοντέλο: Προσομοιωμένα δεδομένα

Στην παρούσα εφαρμογή θεωρούμε ένα Κανονικό γραμμικό μοντέλο με $p = 15$ επεξηγηματικές μεταβλητές $X_1, X_2, ..., X_{15}$ και μέγεθος δείγματος $n = 50$. Τα δεδομένα είναι τεχνητά και έχουν κατασκευαστεί με τέτοιο τρόπο ώστε οι μεταβολές των τιμών της μεταβλητής απόκρισης Y να εξαρτώνται πολύ ισχυρά από τις τιμές των μεταβλητών X_4 και X_5, λιγότερο από τη μεταβλητή X_{12} και τέλος να εχουν πολύ ασθενή εξάρτηση από τις τιμές των υπόλοιπων μεταβλητών. Επίσης θεωρούμε την παρουσία του σταθερού όρου X_0. Το μοντέλο που χρησιμοποιούμε για να εκφράσουμε την μεταβλητή απόκρισης Y συναρτήσει των επεξηγηματικών μεταβλητών $X_1, X_2, ..., X_{15}$ είναι το ακόλουθο

$$f(\mathbf{y} \mid \boldsymbol{\beta}_{(\gamma)}, \sigma^2, \boldsymbol{\gamma}) = N_n(\mathbf{X}_{(\gamma)}\boldsymbol{\beta}_{(\gamma)}, \sigma^2\mathbf{I}),$$

ενώ χρησιμοποιούμε την συζυγή Normal-Inverse-Gamma κατανομή των σχέσεων (4.2.7) και (4.2.8) ως εκ των προτέρων κατανομή για τις παραμέτρους $\boldsymbol{\beta}$ και σ^2 με $\boldsymbol{\mu}_{\boldsymbol{\beta}_{(\gamma)}} = \mathbf{0}$, $a = b = 1.5$, $\boldsymbol{\Sigma}_{(\gamma)} := c^2\left(\mathbf{X}_{(\gamma)}^T\mathbf{X}_{(\gamma)}\right)^{-1}$ και $c^2 = n$. Στον χώρο των μοντέλων θεωρούμε την Ομοιόμορφη κατανομή έτσι ώστε $f(\boldsymbol{\gamma}) = 1/2^p$, για κάθε μοντέλο $\boldsymbol{\gamma}$. Σύμφωνα με όσα δείξαμε στην προηγούμενη παράγραφο, η χρήση της συγκεκριμένης εκ των προτέρων κατανομής οδηγεί στον αναλυτικό προσδιορισμό της περιθώριας πιθανοφάνειας $f(\mathbf{y} \mid \boldsymbol{\gamma})$ η οποία συμμετέχει στον υπολογισμό της πιθανότητας αποδοχής $a = \min\left\{1, \dfrac{f(\boldsymbol{\gamma}' \mid \mathbf{y})}{f(\boldsymbol{\gamma} \mid \mathbf{y})}\right\}$ αφού ισχύει $f(\boldsymbol{\gamma} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\gamma}) \cdot f(\boldsymbol{\gamma})$. Στο παρόν πρόβλημα η διάσταση του χώρου των μοντέλων (2^{15} =32768) είναι τέτοια που μας επιτρέπει να υπολογίσουμε τις εκ των υστέρων πιθανότητες όλων των μοντέλων με 'πλήρη απαρίθμηση' του χώρου και έτσι έχουμε την δυνατότητα να συγκρίνουμε τα αποτελέσματα με αυτά που παίρνουμε ύστερα από την υλοποίηση του αλγορίθμου MC^3 μετά από 10.000 επαναλήψεις:

Πίνακας 4.4.1.1: Εκ των υστέρων πιθανότητες των 10 καλύτερων μοντέλων ύστερα από υλοποίηση του αλγορίθμου MC^3.

	Εκ των υστέρων πιθανότητα του μοντέλου	
Μοντέλο	**MC^3**	**Πλήρης απαρίθμηση**
X4+X5+X12	0.09	0.08
X4+X5	0.06	0.05
X4+X5+X12+X15	0.04	0.03
X4+X5+X11+X12	0.02	0.02
X4+X5+X6+X12	0.02	0.01
X4+X5+X15	0.02	0.02
X4+X5+X11	0.02	0.01
X4+X5+X10+X12	0.02	0.01
X2+X4+X5+X12	0.01	0.01
X4+X5+X11+X12+X15	0.01	0.01

Πίνακας 4.4.1.2: Εκ των υστέρων περιθώριες πιθανότητες αποδοχής των επεξηγηματικών μεταβλητών του μοντέλου.

	$f(\gamma_j = 1 \mid \mathbf{y})$	
Μεταβλητές	**MC^3**	**Πλήρης απαρίθμηση**
X1	0.15	0.12
X2	0.14	0.13
X3	0.11	0.13
X4	0.99	0.98
X5	1.00	1.00
X6	0.15	0.16
X7	0.10	0.14
X8	0.11	0.13
X9	0.15	0.17
X10	0.15	0.15
X11	0.23	0.20
X12	0.60	0.62
X13	0.14	0.14
X14	0.11	0.13
X15	0.27	0.28

Σύμφωνα με τους παραπάνω Πίνακες, παρατηρούμε ότι επιβεβαιώνεται η ισχυρή εξάρτηση της μεταβλητής απόκρισης Y από τις επεξηγηματικές μεταβλητές X_4 και X_5 κατά κύριο λόγο και κατά δεύτερο λόγο από την μεταβλητή X_{12}. Επίσης έχουμε μία σχεδόν ταύτιση των αποτελεσμάτων που παίρνουμε ύστερα από την υλοποίηση του MC^3 αλγορίθμου σε σχέση με αυτά της πλήρους απαρίθμησης.

4.4.2 Κανονικό γραμμικό μοντέλο: Μελέτη όζοντος

Η εφαρμογή πού αναλύεται σε αυτή την παράγραφο αφορά τη συγκέντρωση όζοντος σε ορεινές περιοχές της Καλιφόρνιας (CA). Τα δεδομένα περιγράφονται από τον ακόλουθο Πίνακα:

Μεταβλητές		**Περιγραφή**
Y	Ozone(depended)	Μέγιστο ημερήσιο επίπεδο όζοντος
X1	Month	Μήνας : 1 = Γενάρης,..., 12 = Δεκέμβριος
X2	Mday	Ημέρα του μήνα
X3	Wday	Ημέρα της εβδομάδας : 1 = Δευτέρα,...,7 = Κυριακή
X4	Vh	500 millibar ύψος πίεσης (m) που μετριέται στο Vandenberg AFB
X5	Wind	Ταχύτητα ανέμου (mph) στο διεθνές αεροδρόμιο του Λος Άντζελες (LAX)
X6	Hum	Υγρασία (%) στο LAX
X7	temp1	Θερμοκρασία (°F) που μετράται στο Sandburg, CA
X8	temp2	Θερμοκρασία (°F) που μετράται στο El Monte, CA
X9	Ibh	Αναστροφή ύψους βάσης (πόδια) στο LAX
X10	Ibt	Κλίση πίεσης (mm Hg) από το LAX για το Daggett, CA
X11	Dpg	Αναστροφή θερμοκρασιακής βάσης (βαθμούς F) στο LAX
X12	Vis	Ορατότητα (μίλια) που μετράται στο LAX

Στο παρόν πρόβλημα ακολουθήθηκε ακριβώς η ίδια διαδικασία με αυτήν της προηγούμενης παραγράφου και προέκυψαν τα ακόλουθα απότελέσματα:

Πίνακας 4.4.2.1: Εκ των υστέρων πιθανότητες των 10 καλύτερων μοντέλων ύστερα από υλοποίηση του αλγορίθμου MC^3.

	Εκ των υστέρων πιθανότητα του μοντέλου	
Μοντέλο	**MC^3**	**Πλήρης απαρίθμηση**
X1+X6+X8	0.26	0.27
X1+X4+X6+X8	0.10	0.10
X1+X6+X7+X8	0.08	0.09
X1+X4+X6+X7+X8	0.04	0.04
X1+X6+X8+X9	0.04	0.04
X1+X6+X8+X11	0.03	0.02
X1+X5+X6+X8	0.03	0.02
X1+X6+X8+X10	0.03	0.03
X1+X6+X8+X12	0.03	0.03
X1+X2+X6+X8	0.02	0.02

Πίνακας 4.4.2.2: Εκ των υστέρων περιθώριες πιθανότητες αποδοχής των επεξηγηματικών μεταβλητών του μοντέλου.

	$f(\gamma_j = 1 \mid \mathbf{y})$	
Μεταβλητές	**MC^3**	**Πλήρης απαρίθμηση**
month	0.99	1.00
mday	0.08	0.08
wday	0.07	0.07
vh	0.28	0.29
wind	0.09	0.07
hum	1.00	1.00
temp1	0.27	0.28
temp2	1.00	1.00
ibh	0.17	0.15
ibt	0.10	0.10
dpg	0.11	0.18
vis	0.11	0.88

Βλέπουμε από τα αποτελέσματα ότι οι παράγοντες που επηρεάζουν περισσότερο το επίπεδο του όζοντος είναι ο Μήνας (X1), η Υγρασία (%) στο LAX (X6) καθώς και η Θερμοκρασία (°F) που μετράται στο El Monte (X8). Οι υπόλοιπες μεταβλητές επεξηγούν σε πολύ μικρότερο βαθμό το επίπεδο του όζοντος στην συγκεκριμένη περιοχή. Αξίζει να παρατηρήσουμε επίσης ότι και σε αυτήν την εφαρμογή τα αποτελέσματα που προκύπτουν από τον αλγόριθμο MC^3 είναι πάρα πολύ κοντά με αυτά της μεθόδου της πλήρους απαρίθμησης.

4.4.3 Λογιστικό μοντέλο: Μέτρηση ποιότητας της υγειονομικής περίθαλψης

Ένα σημαντικό θέμα στην πολιτική για την υγεία είναι η αξιολόγηση της ποιότητας της υγειονομικής περίθαλψης που προσφέρεται στους νοσοκομειακούς ασθενείς. Η ποιότητα της φροντίδας συνήθως θεωρείται ότι εξαρτάται από τρία συστατικά [π.χ. Donabedian και Bashshur, (2002)]: (i) την διαδικασία, η οποία περιλαμβάνει όλες τις ενέργειες που οι παροχείς υγειονομικής περίθαλψης κάνουν για λογαριασμό των ασθενών, (ii) τα αποτελέσματα, τα οποία είναι ό,τι συμβαίνει στους ασθενείς ως αποτέλεσμα της φροντίδας που λαμβάνουν, και (iii) την κατάσταση της υγείας των ασθενών κατά την εισαγωγή, επειδή η καταλληλότητα των αποτελεσμάτων δεν μπορεί να κριθεί μη λαμβάνοντας υπόψη την κατάσταση της υγείας των ασθενών κατα την εισαγωγή τους σε ένα νοσοκομείο.

Το πρόβλημα με το οποίο ασχολούμαστε σε αυτήν την παράγραφο έχει να κάνει με την αξιολόγηση της ποιότητας της ιατρικής περίθαλψης η οποία παρέχεται σε νοσοκομεία των Η.Π.Α. σχετικά με την αντιμετώπιση της πνευμονίας. Πιο συγκεκριμένα, τα δεδομένα που αναλύονται στην παρούσα εφαρμογή αποτελούν ένα αντιπροσωπευτικό δείγμα $n = 2532$ ηλικιωμένων Αμερικανών ασθενών που νοσηλεύτηκαν κατά την περίοδο 1980-86 με πνευμονία σε διάφορα νοσοκομεία των Η.Π.Α. και μετράται το ενδεχόμενο επιβίωσης ή θανάτου τους ύστερα από 30 μέρες παραμονής στο νοσοκομείο. Το ενδεχόμενο αυτό θεωρούμε ότι εξαρτάται από 83 συνολικά παράγοντες (δείκτες ασθενείας) οι οποίοι είναι για παράδειγμα η συστολική αρτηριακή πίεση την πρώτη ημέρα της εισαγωγής, η παρουσία ή απουσία δύσπνοιας και το επίπεδο του αζώτου στην ουρία του αίματος (ένα μέτρο της λειτουργίας των νεφρών). Το στατιστικό μοντέλο που χρησιμοποιούμε για να μετρήσουμε την επίδραση των $p = 83$ συνολικά παραγόντων (επεξηγηματικές μεταβλητές) πάνω στην μεταβλητή απόκρισης Y που δηλώνει την επιβίωση(0) ή τον θάνατο(1) ενός ασθενούς είναι το λογιστικό μοντέλο παλινδρόμισης:

$$
\begin{aligned}
&f(Y_i \mid \boldsymbol{\gamma}) = Bernoulli\left[p_i(\boldsymbol{\gamma})\right],\ i = 1,\ldots,n,\\
&\eta_i(\boldsymbol{\gamma}) = \log\left[\frac{p_i(\boldsymbol{\gamma})}{1 - p_i(\boldsymbol{\gamma})}\right] = \sum_{j=0}^{p} \beta_j \gamma_j X_{ij},\ i = 1,\ldots,n,\\
&\eta(\boldsymbol{\gamma}) = \mathbf{X}_{(\gamma)} \boldsymbol{\beta}_{(\gamma)},
\end{aligned}
$$

όπου X_{ij}, $j = 0,1,\ldots,p$, $i = 1,\ldots,n$ είναι ο j-δείκτης ασθενείας για τον i-ασθενή και $p_i(\boldsymbol{\gamma})$ η πιθανότητα, για το μοντέλο $\boldsymbol{\gamma}$, να πεθάνει ο ασθενής i ύστερα από 30 ημέρες νοσηλείας, $\boldsymbol{\gamma} = (\gamma_0,\ldots,\gamma_p)^T$, $\boldsymbol{\beta} = (\beta_0,\ldots,\beta_p)^T$, $\boldsymbol{\beta}_{(\gamma)} = (\beta_j : \gamma_j = 1,\ j = 0,\ldots p)$ και $\mathbf{X}_{(\gamma)}$ είναι ο υποπίνακας του πίνακα σχεδιασμού $\mathbf{X}$ με στήλες που αντιστοιχούν στις μεταβλητές οι οποίες καθορίζονται από το μοντέλο $\boldsymbol{\gamma}$. Θεωρούμε επίσης ότι $X_{i0} = 1,\ i = 1,\ldots,n$ και $\gamma_0 = 1$, δηλαδή ότι ο σταθερός όρος X_0 περιέχεται σε όλα τα μοντέλα. Αξίζει να σημειωθεί στο σημείο αυτό ότι η εξέταση όλων των δυνατών μοντέλων, πλήθους $2^{83} = 9.7 \times 10^{24}$, με στόχο την επιλογή του καλύτερου υποσυνόλου μεταβλητών, είναι ανέφικτη καθώς ο χώρος των μοντέλων είναι εξαιρετικά μεγάλος. Σε ένα τέτοιο πρόβλημα λοιπόν κρίνεται επιτακτική η ανάγκη χρήσης μεθόδων όπως ο αλγόριθμος MC^3 πού 'εξερευνούν' τον χώρο των μοντέλων δίνοντας μια πολύ καλή ένδειξη για το ποιο πρέπει να είναι το βέλτιστο μοντέλο.

Ένα ακόμη σημείο το οποίο πρέπει να τονίσουμε είναι ότι στο παρόν πρόβλημα η περιθώρια πιθανοφάνεια $f(\mathbf{y}\,|\,\boldsymbol{\gamma})$ δεν μπορεί να υπολογιστεί σε 'κλειστή μορφή' και ως εκ τούτου είμαστε αναγκασμένοι να καταφύγουμε σε προσεγγιστικές μεθόδους για τον υπολογισμό της. Χρησιμοποιούμε λοιπόν την BIC προσέγγιση για τον υπολογισμό της πιθανότητας αποδοχής a (βλέπε παράγραφο 4.3) και μετά από 10.000 επαναλήψεις του αλγορίθμου προκύπτουν τα ακόλουθα αποτελέσματα

Πίνακας 4.4.3.1: Εκ των υστέρων πιθανότητες των 10 καλύτερων μοντέλων ύστερα από υλοποίησητου αλγορίθμου MC^3.

Μοντέλο	Εκ των υστέρων πιθανότητα του μοντέλου
X1+X2+X3+X4+X5+X8+X12+X15+X27+X37+X70+X73	0.013
X1+X2+X3+X4+X5+X8+X12+X14+X15+X37+X70+X73	0.008
X1+X2+X3+X4+X5+X8+X12+X15+X27+X37+X46+X70+X73	0.008
X1+X2+X3+X4+X5+X8+X12+X15+X26+X27+X37+X56+X70+X73	0.007
X1+X2+X3+X4+X5+X8+X12+X15+X37+X70+X73	0.007
X1+X2+X3+X4+X5+X12+X15+X37+X46+X62+X64+X70+X73	0.006
X1+X2+X3+X4+X5+X8+X12+X15+X16+X37+X42+X46+X54+X70+X73	0.006
X1+X2+X3+X4+X5+X8+X12+X15+X16+X23+X37+X46+X58+X70+X73	0.006
X1+X2+X3+X4+X5+X8+X12+X15+X27+X37+X46+X48+X64+X70+X73	0.006
X1+X2+X3+X4+X5+X8+X12+X26+X37+X46+X47+X70+X73	0.006

Όπως μπορούμε να καταλάβουμε, η εκτίμηση των εκ των υστέρων πιθανοτήτων των μοντέλων σε τόσο μεγάλους χώρους είναι προβληματική και έτσι σε αυτές τις περιπτώσεις εστιάζουμε περισσότερο στην εκτίμηση των εκ των υστέρων περιθώριων πιθανοτήτων αποδοχής $f(\gamma_j = 1\,|\,\mathbf{y})$.

Πίνακας 4.4.3.2 Εκ των υστέρων περιθώριες πιθανότητες αποδοχής των επεξηγηματικών μεταβλητών του μοντέλου.

Μεταβλητές	$f(\gamma_j = 1\,|\,\mathbf{y})$	Μεταβλητές	$f(\gamma_j = 1\,|\,\mathbf{y})$	Μεταβλητές	$f(\gamma_j = 1\,|\,\mathbf{y})$
X1	1.00	**X29**	0.00	**X57**	0.03
X2	0.96	**X30**	0.00	**X58**	0.04
X3	1.00	**X31**	0.03	**X59**	0.01
X4	1.00	**X32**	0.07	**X60**	0.07
X5	0.95	**X33**	0.02	**X61**	0.04
X6	0.08	**X34**	0.03	**X62**	0.13
X7	0.07	**X35**	0.05	**X63**	0.00
X8	0.85	**X36**	0.03	**X64**	0.05
X9	0.04	**X37**	0.94	**X65**	0.00
X10	0.00	**X38**	0.01	**X66**	0.00
X11	0.09	**X39**	0.01	**X67**	0.03
X12	0.97	**X40**	0.02	**X68**	0.14

X13	0.05	**X41**	0.00	**X69**	0.02
X14	0.11	**X42**	0.02	**X70**	0.97
X15	0.68	**X43**	0.05	**X71**	0.03
X16	0.15	**X44**	0.02	**X72**	0.06
X17	0.01	**X45**	0.04	**X73**	0.97
X18	0.01	**X46**	0.58	**X74**	0.02
X19	0.01	**X47**	0.08	**X75**	0.02
X20	0.04	**X48**	0.01	**X76**	0.01
X21	0.05	**X49**	0.13	**X77**	0.01
X22	0.00	**X50**	0.03	**X78**	0.02
X23	0.04	**X51**	0.01	**X79**	0.02
X24	0.06	**X52**	0.06	**X80**	0.05
X25	0.03	**X53**	0.04	**X81**	0.04
X26	0.12	**X54**	0.07	**X82**	0.01
X27	0.51	**X55**	0.10	**X83**	0.07
X28	0.04	**X56**	0.03		

Πίνακας 4.4.3.3: Οι μεταβλητές με εκ των υστέρων περιθώρια πιθανότητα αποδοχής $f(\gamma_j = 1 | \mathbf{y})$>0.5

Μεταβλητές	**Όνομα**	$f(\gamma_j = 1 \| \mathbf{y})$
X1	Systolic blood pressure	1.00
X2	Age	0.96
X3	Blood urea nitrogen	1.00
X4	APACHE II coma score	1.00
X5	Shortness of breath day 1	0.95
X8	Septic complications	0.85
X12	Initial temperature	0.97
X15	Cardiomegaly score	0.68
X27	Hematologic history score	0.51
X37	APACHE respiratory rate score	0.94
X46	Admission SBP	0.58
X70	APACHE pH score	0.97
X73	Morbid+ comorbid score	0.97

Στον Πίνακα 4.4.3.3 φαίνεται ποιοί απο τους δείκτες ασθενείας είναι οι σημαντικότεροι και καθορίζουν ουσιαστικά το αν ένας ασθενής θα πεθάνει ή θα επιβιώσει ύστερα από 30 ημέρες νοσηλείας στο νοσοκομείο. Επιπλέον, όπως μπορούμε να παρατηρήσουμε στον Πίνακα 4.4.3.2, όλες οι υπόλοιπες μεταβλητές έχουν περιθώρια εκ των υστέρων πιθανότητα μικρότερη από 0.15.

ΕΠΙΛΟΓΟΣ

Στόχος της παρούσας διπλωματικής εργασίας ήταν να εισάγει τον ενδιαφερόμενο αναγνώστη στην περιοχή της επιλογής μοντέλων και μεταβλητών στα δημοφιλή γενικευμένα γραμμικά μοντέλα από τη σκοπιά της Μπεϋζιανής Στατιστικής μεθοδολογίας. Στην παρούσα διπλωματική παρουσιάστηκαν όλες εκείνες οι βασικές έννοιες και αναπτύχθηκαν αρκετές από τις διαδικασίες με τις οποίες χειρίζεται η Μπεϋζιανή Στατιστική το συγκεκριμένο θέμα. Ιδιαίτερη έμφαση δόθηκε στον καθορισμό εκ των προτέρων κατανομών για τους συντελεστές και τις παραμέτρους του μοντέλου καθώς επίσης και στις διάφορες τεχνικές υπολογισμού της περιθώριας πιθανοφάνειας σε περιπτώσεις όπου αυτή δεν μπορεί να υπολογιστεί σε αναλυτική μορφή. Εκτός από τις απαραίτητες θεωρητικές έννοιες που αφορούν το πρόβλημα επιλογής μοντέλων και μεταβλητών και οι οποίες αναφέρθηκαν κατά κύριο λόγο στα τρία πρώτα κεφάλαια, στο τέταρτο και τελευταίο κεφάλαιο της εργασίας περιγράφτηκε αναλυτικά ο αλγόριθμος MC^3 καθώς και τρεις εφαρμογές του εν λόγω αλγορίθμου σε αριθμητικά δεδομένα. Ειδικά στην περίπτωση της τρίτης εφαρμογής (λογιστικό μοντέλο) φάνηκε πόσο επιτακτική είναι η ανάγκη χρήσης MCMC μεθόδων, όπως ο αλγόριθμος MC^3, σε προβλήματα όπου ο αριθμός των υποψηφίων μοντέλων είναι πολύ μεγάλος και αυτό γιατί η πλήρης διερεύνηση τόσο μεγάλων χώρων είναι πρακτικά ανέφικτη.

Ελπίζω η παρούσα διπλωματική εργασία να έχει θετική απήχηση σε όποιον αναγνώστη ενδιαφέρεται να μελετήσει ή και να εφαρμόσει ακόμα τη Μπεϋζιανή μεθοδολογία σε προβλήματα επιλογής μοντέλων και μεταβλητών στα γενικευμένα γραμμικά μοντέλα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Φουσκάκης, Δ.(2007). *Μπεϋζιανή Στατιστική και MCMC*. Σημειώσεις, Ε.Μ.Π., Αθήνα.
- Δελλαπόρτας, Π. και Τσιαμυρτζής, Π.(2004). *Στατιστική κατά Bayes*. Σημειώσεις, Ο.Π.Α., Αθήνα.

Ξενόγλωσση

- Aitkin, M. (1991). Posterior Bayes Factors. *Journal of the Royal Statistical Society*, B, **53**, 111-142.
- Berger, J. and Pericchi, L. (2001). *Objective Bayesian Methods for Model Selection: Introduction and Comparison, Model Selection*. Hayward Institute of Mathematical Statistics, Hayward, U.S.A.
- Carlin, B. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society* B, **57**, 473-484.
- Dellaportas, P. and Forster, J. (1999). Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Log-Linear Models. *Biometrika*, **86**, 615-633.
- Dellaportas, P., Forster, J. and Ntzoufras, I. (2002). On Bayesian Model and Variable Selection Using MCMC. *Statistics and Computing*, **12**, 27-36.
- Dobson, A. (2002). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, Boca Raton, U.S.A.
- Donabedian, A. and Bashshur, R. (2002). *An Introduction to Quality Assurance in Health Care*. Oxford University Press, Oxford, U.K.
- Draper, N. and Smith, H. (1966). *Applied Regression Analysis*. John Wiley & Sons, New York, U.S.A.
- Fouskakis, D., Ntzoufras, I. and Draper, D. (2009). Bayesian Variable Selection Using Cost-Adjusted BIC with Application to Cost-Effective Measurement of Quality of Health Care. *The Annals of Applied Statistics*, **3**, 663-690.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, U.S.A.
- George, E. and Foster, D. (2000). Calibration and Empirical Bayes Variable Selection. *Biometrika*, **87**, 731-747.
- Godsill, S. (1998). *On the Relationship between MCMC Model Uncertainty Methods*. Technical Report, Signal Processing Group, Engineering Department, Cambridge University, Cambridge, U.K.

- Hansen, M. and Yu, B. (2001). Model Selection and the Principle of Minimum Description Length. *Journal of the American Statistical Association*, **96**, 746-774.
- Healy, M. (1988). *Generalized Linear Models: An Introduction*. Clarendon Press, Gloucestershire, U.K.
- Hocking, R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, **32**, 1-49.
- Ibrahim, J. and Laud, P. (1994). A Predictive Approach to the Analysis of Designed Experiments. *Journal of the Royal Statistical Society* B, **89**, 309-319.
- Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.
- Koch, K. (2007). *Introduction to Bayesian Statistics*. Springer, Berlin, Germany.
- Lindley, D. (1957). A Statistical Paradox, *Biometrika*, **44**, 187-192.
- Madigan, D. and York J. (1995). Bayesian Graphical Models for Discrete Data. *International Statistical Review*, **63**, 215-232.
- Nelder, J. and Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society* A, **135**, 370-384.
- Ntzoufras, I. (1999). *Aspects of Bayesian Model and Variable Selection Using MCMC*, Ph.D. Thesis, Department of Statistics, Athens University of Economics and Business, Athens, Greece.
- O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society* B, **57**, 99-138.
- Raftery, A. (1995). *Bayesian Model Selection in Social Research, Sociological Methodology*. Blackwell, Oxford, U.K.
- Raftery, A. (1996). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika*, **83**, 251-266.
- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, **22**, 1701-1728.
- Venables, W. and Smith, D. (2007). *An Introduction to R*: A Programming Environment for Data Analysis and Graphics, User Manual for Version 2.5.1 by the R Development Core Team.
- Zellner, A. (1986). *On Assessing Prior Distributions and Bayesian Regression Analysis Using G-Prior distributions, Bayesian Inference and Decision Techniques*: Essays in Honor of Bruno de Finetti, Amsterdam, Holland.