



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ ΣΤΙΣ ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ
ΕΠΙΣΤΗΜΕΣ**

**Κυρτοί Συνδυασμοί Γενικευμένων Γραμμικών
Μοντέλων και Βελτιστοποίηση**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΦΑΝΑΡΑ ΜΗΝΑ

Τριμελής Επιτροπή: Γ. Κοκολάκης (Επιβλέπων Καθηγητής)

Χ. Καρόνη

Δ. Φουσκάκης

Αθήνα, Οκτώβριος 2012

Ευχαριστίες

Η πραγματοποίηση της παρούσας εργασίας θα ήταν αδύνατη χωρίς την βοήθεια των παρακάτω ατόμων, τα οποία θέλω να ευχαριστήσω.

Κατ' αρχάς, ευχαριστώ θερμά τον κ. Γεώργιο Κοκολάκη για την καθοδήγησή του και τις πολύτιμες συμβουλές και διορθώσεις που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της εργασίας, αφιερώνοντας πολύ από τον περιορισμένο χρόνο του.

Επιπλέον, ευχαριστώ την κα Χρυσήδα Καρώνη για την καθοριστική βοήθειά της στην αναζήτηση δεδομένων και την κα Ευστρατία Χαριτίδου, η οποία με βοήθησε με μεγάλη προθυμία στις πολλές περιπτώσεις που της το ζήτησα.

Τέλος, ευχαριστώ ιδιαιτέρως την οικογένειά μου και όλα τα κοντινά μου πρόσωπα, που με στηρίζουν γενικώς σε ό,τι κάνω.

Φαναράς Μηνάς

Πρόλογος

Κατά την ανάλυση δεδομένων στην Κλασική Στατιστική, συνηθίζεται να επιλέγουμε ένα συγκεκριμένο μοντέλο μίας κλάσης και ακολούθως να δρούμε σαν αυτό το μοντέλο να παρήγαγε όντως τα δεδομένα. Με την στρατηγική αυτή αγνοούμε την αβεβαιότητα που υπάρχει σχετικά με την επιλογή μοντέλου [model uncertainty], έναν παράγοντα που αποτελεί μεγάλο ποσοστό της συνολικής αβεβαιότητας που έχουμε για το εκάστοτε πρόβλημα. Κατά συνέπεια, η ανάλυση με αυτόν τον τρόπο εμπεριέχει αρκετό ρίσκο.

Στην Μπεϋζιανή Στατιστική, η παραπάνω αδυναμία μπορεί να αντιμετωπιστεί χρησιμοποιώντας κυρτούς συνδυασμούς των μοντέλων που θεωρούμε «υποψήφια» να περιγράψουν το πρόβλημα [Bayesian model averaging], αντί να επιλέξουμε μόνο ένα από αυτά για την περιγραφή του προβλήματος. Η στρατηγική αυτή θα μας απασχολήσει στο παρόν κείμενο. Ιδιαίτερη σημασία θα δοθεί στην προβλεπτική ικανότητα της μεθόδου και στην εφαρμογή της σε γενικευμένα γραμμικά μοντέλα.

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή.....	6
1.1 Το Μπεϋζιανό Μοντέλο Στατιστικής Μεθοδολογίας.....	6
1.2 Μπεϋζιανή Εκτίμηση.....	9
1.3 Μέθοδοι MCMC.....	10
Κεφάλαιο 2: Κυρτοί Συνδυασμοί Μοντέλων.....	15
2.1 Μπεϋζιανή Επιλογή Μοντέλου.....	15
2.2 Η Μέθοδος των Κυρτών Συνδυασμών Μοντέλων.....	17
2.3 Επιλογή των Πρότερων Κατανομών.....	20
2.4 Υπολογισμός των Ολοκληρωμάτων.....	21
2.5 Υπολογισμός του Αθροίσματος.....	23
2.6 Προβλεπτική Ικανότητα των Κυρτών Συνδυασμών Μοντέλων.....	27
2.7 Επιλογή της Κλάσης για την Εφαρμογή της Μεθόδου.....	30
Κεφάλαιο 3: Κυρτοί Συνδυασμοί Γενικευμένων Γραμμικών Μοντέλων.....	32
3.1 Εισαγωγή.....	32
3.2 Η Μέθοδος των Κυρτών Συνδυασμών στα Γενικευμένα Γραμμικά Μοντέλα.....	35
3.3 Επιλογή των Πρότερων Πιθανοτήτων των Μοντέλων.....	37
3.4 Επιλογή των Πρότερων Κατανομών των Παραμέτρων.....	39
3.5 Υπολογισμός των Ύστερων Πιθανοτήτων των Μοντέλων.....	42
3.6 Εφαρμογή της Μεθόδου.....	44
3.7 Η Συνάρτηση bic.glm της R.....	49
Κεφάλαιο 4: Εμφάνιση Στεφανιαίου Επεισοδίου σε Άντρες.....	51
4.1 Περιγραφή του Προβλήματος.....	51
4.2 Αποτελέσματα.....	52
4.3 Εξαγωγή Συμπερασμάτων από p-τιμές και από Ύστερες Πιθανότητες.....	58
4.4 Σύγκριση της Προβλεπτικής Ικανότητας.....	61
4.5 Ανάλυση χωρίς την Μεταβλητή της Ηλικίας.....	66
Παράρτημα.....	71
Βιβλιογραφία.....	78

Κεφάλαιο 1

Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε συνοπτικά τις βασικές έννοιες που θα χρησιμοποιηθούν στα επόμενα κεφάλαια, δίνοντας έτσι το πλαίσιο στο οποίο θα κινηθούμε.

1.1 Το Μπεϋζιανό Μοντέλο Στατιστικής Μεθοδολογίας

Σε αντίθεση με το κλασικό μοντέλο για την εκτίμηση μίας παραμέτρου θ , στο μπεϋζιανό μοντέλο εκτός από την συνάρτηση πιθανοφάνειας χρησιμοποιούμε και μία συνάρτηση βάρους $p(\theta)$, η οποία εκφράζει τη γνώση μας για την παράμετρο πριν από την λήψη του τυχαίου δείγματος. Η συνάρτηση αυτή καλείται **πρότερη [prior]** κατανομή της παραμέτρου θ (η οποία βέβαια μπορεί να είναι και διανυσματική).

Έτσι, το **εκ των προτέρων** [a priori] μπεϋζιανό στατιστικό μοντέλο είναι το :

$$\{X, \mathcal{X}, p(x|\theta), p(\theta), \theta \in \Theta\}, \quad (1.1)$$

όπου X μία τυχαία μεταβλητή που συμβολίζει το χαρακτηριστικό των μονάδων του πληθυσμού που μελετάμε, \mathcal{X} το πεδίο τιμών της X , $p(x|\theta)$ η σ.π.π. της τ.μ. X και Θ ο χώρος στον οποίο λαμβάνει τιμές η άγνωστη παράμετρος θ , με $\Theta \subseteq \mathbb{R}^n$ στη γενική περίπτωση που $\theta = (\theta_1, \theta_2, \dots, \theta_n)$.

Μετά τη δειγματοληψία, μπορούμε να υπολογίσουμε την λεγόμενη **ύστερη [posterior]** κατανομή της παραμέτρου θ με βάση το *Θεώρημα Bayes*. Έτσι, για δεδομένο δείγμα \mathbf{x} , έχουμε:

$$p(\theta | \mathbf{x}) = \frac{p(\theta)p(\mathbf{x} | \theta)}{p(\mathbf{x})} = K \times p(\theta) \prod_{i=1}^n p(x_i | \theta), \quad (1.2)$$

όπου η σταθερά κανονικοποίησης K προκύπτει από τη συνθήκη:

$$K^{-1} = \int p(\boldsymbol{\theta}) \prod_{i=1}^n p(x_i | \boldsymbol{\theta}) d\boldsymbol{\theta} = p(\mathbf{x}) . \quad (1.3)$$

Ισοδύναμα, έχουμε την παρακάτω έκφραση για την ύστερη κατανομή της παραμέτρου $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta} | \mathbf{x}) \propto p(\boldsymbol{\theta}) L(\boldsymbol{\theta} : \mathbf{x}), \quad (1.4)$$

με το δεξιό μέλος να αποτελεί τον *πυρήνα* της κατανομής και με $L(\boldsymbol{\theta} : \mathbf{x})$ να συμβολίζει τη συνάρτηση πιθανοφάνειας της παραμέτρου $\boldsymbol{\theta}$.

Κατά τον τρόπο αυτό, μετά τη λήψη του τυχαίου δείγματος \mathbf{x} προκύπτει το **εκ των υστέρων** μπεϋζιανό στατιστικό μοντέλο:

$$\{X, \mathcal{X}^n, p(x | \boldsymbol{\theta}), p(\boldsymbol{\theta} | \mathbf{x}), \boldsymbol{\theta} \in \Theta\}, \quad (1.5)$$

πάνω στο οποίο στηρίζονται όλα τα συμπεράσματά μας σχετικά με την παράμετρο $\boldsymbol{\theta}$.

Το παραπάνω μοντέλο μπορεί να χρησιμοποιηθεί ως το εκ των προτέρων μοντέλο σε μία καινούρια δειγματοληψία κ.ο.κ. Δηλαδή, η τρέχουσα ύστερη κατανομή μπορεί να χρησιμοποιείται ως πρότερη κατανομή όταν έχουμε ένα νέο τυχαίο δείγμα στη διάθεσή μας, με την διαδικασία αυτή να επαναλαμβάνεται συνεχώς. Κάθε φορά, η ύστερη κατανομή εκπροσωπεί συνοπτικά όλη την πληροφορία που έχουμε στη διάθεσή μας και σε αυτήν βασίζονται τα συμπεράσματά μας, αλλά και οι αποφάσεις που παίρνουμε για ζητήματα σχετικά με την τιμή της παραμέτρου $\boldsymbol{\theta}$.

Η κατάλληλη επιλογή της πρότερης κατανομής είναι ένα βασικό ζητούμενο στη διαδικασία που περιγράψαμε και θα μας απασχολήσει στις ειδικές περιπτώσεις στις οποίες θα επικεντρωθούμε στα επόμενα κεφάλαια. Προς το παρόν, θα αναφέρουμε μία οικογένεια κατανομών όπου το πρόβλημα της επιλογής prior κατανομής απλοποιείται. Αυτή είναι η **Εκθετική Οικογένεια Κατανομών**.

Λέμε ότι η συνάρτηση πιθανότητας $p(\mathbf{x} | \boldsymbol{\theta})$ ανήκει στην Εκθετική Οικογένεια Κατανομών όταν είναι της μορφής:

$$\begin{aligned} p(\mathbf{x} | \boldsymbol{\theta}) &= f(\mathbf{x}) g(\boldsymbol{\theta}) \exp\{\mathbf{c}(\boldsymbol{\theta})^T \mathbf{d}(\mathbf{x})\} \\ &= f(\mathbf{x}) g(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^k c_j(\boldsymbol{\theta}) d_j(\mathbf{x}) \right\}, \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n \end{aligned} \quad (1.6)$$

όπου οι συναρτήσεις $f(\mathbf{x}), d(\mathbf{x})$ είναι συναρτήσεις του \mathbf{x} μόνο, ενώ οι $g(\boldsymbol{\theta}), c(\boldsymbol{\theta})$ είναι συναρτήσεις του $\boldsymbol{\theta}$ μόνο. Επιπλέον, το *στήριγμα* (support) της σ.π. πρέπει να μην εξαρτάται από την παράμετρο $\boldsymbol{\theta}$.

Τότε η **συζυγής** πρότερη κατανομή για την παράμετρο $\boldsymbol{\theta}$ είναι της μορφής:

$$p(\boldsymbol{\theta}) = \{g(\boldsymbol{\theta})\}^\eta \exp\{\mathbf{v}^T \mathbf{c}(\boldsymbol{\theta})\}, \boldsymbol{\theta} \in \Theta, \quad (1.7)$$

με *υπερπαραμέτρους* η και $\mathbf{v} = (v_1, \dots, v_k)^T$.

Η παραπάνω prior κατανομή καλείται συζυγής, επειδή σε συνδυασμό με την συνάρτηση πιθανοφάνειας $p(\mathbf{x}|\boldsymbol{\theta})$ – και με εφαρμογή του Θεωρήματος Bayes – δίνει posterior σ.π. της ίδιας μορφής με αυτήν.

Συγκεκριμένα, η πιθανοφάνεια στην περίπτωση μας είναι της μορφής:

$$p(\boldsymbol{\theta}) \propto \{g(\boldsymbol{\theta})\}^\eta \exp\{\mathbf{t}(\mathbf{x})^T \mathbf{c}(\boldsymbol{\theta})\}, \quad (1.8)$$

όπου $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_n$ τυχαίο δείγμα και $\mathbf{t}(\mathbf{x}) = (t_1(\mathbf{x}), \dots, t_k(\mathbf{x}))$ η διανυσματική

συνάρτηση με $t_j(\mathbf{x}) = \sum_{i=1}^n d_j(\mathbf{x}_i)$, $j = 1, \dots, k$.

Με εφαρμογή του Θεωρήματος Bayes προκύπτει η εξής posterior κατανομή για την παράμετρο $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta} | \mathbf{x}) \propto g(\boldsymbol{\theta})^{\eta+n} \exp\{\mathbf{c}(\boldsymbol{\theta})^T (\mathbf{v} + \mathbf{t}(\mathbf{x}))\}, \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^n, \quad (1.9)$$

δηλαδή της ίδιας μορφής με την prior, με επικαιροποιημένες παραμέτρους $\eta+n$, $\mathbf{v} + \mathbf{t}(\mathbf{x})$.

Πολλές γνωστές κατανομές, όπως η κατανομή Bernoulli, η Διωνυμική, η Γεωμετρική, η Εκθετική, καθώς και η Κανονική και η Γάμμα (όταν η μία από τις δύο παραμέτρους είναι γνωστή) ανήκουν στην Εκθετική Οικογένεια Κατανομών.

Ωστόσο, τις περισσότερες φορές καλούμαστε να χρησιμοποιήσουμε πιο πολύπλοκες κατανομές και κατ' επέκταση να υπολογίσουμε πολύπλοκες (και πολυδιάστατες) ύστερες κατανομές. Επίσης, ενδέχεται ακόμα και στην περίπτωση που χρησιμοποιούμε κάποιο μοντέλο από την Εκθετική Οικογένεια Κατανομών να μην θέλουμε να χρησιμοποιήσουμε συζυγή prior κατανομή. Οι μέθοδοι MCMC που θα αναφέρουμε συνοπτικά στο τέλος του κεφαλαίου είναι ιδιαίτερα αποτελεσματικές σε τέτοιες περιπτώσεις.

1.2 Μπεϋζιανή Εκτίμηση

Τίθεται τώρα λογικά το ερώτημα σχετικά με το ποιά είναι η εκτίμηση που χρησιμοποιούμε για την παράμετρό μας. Η εκτίμηση μίας παραμέτρου θ στο πλαίσιο της Μπεϋζιανής Στατιστικής βασίζεται στην ύστερη κατανομή της $p(\theta|x)$ και στην ελαχιστοποίηση της αναμενόμενης ζημιάς (ή στην μεγιστοποίηση της αναμενόμενης ωφελιμότητας) που συνεπάγεται η όποια απόκλιση μεταξύ πραγματικής και εκτιμώμενης τιμής της παραμέτρου. Η διαδικασία αυτή της εκτίμησης εντάσσεται στο ευρύτερο πλαίσιο της **Μπεϋζιανής Θεωρίας Αποφάσεων**.

Στο πλαίσιο της Θεωρίας Αποφάσεων, η παράμετρος θ εκφράζει την κατάσταση στην οποία βρισκόμαστε και η απόφαση σχετικά με την τιμή της εκφράζει τη **δράση** ή **ενέργεια** που επιλέγουμε να κάνουμε, την οποία συμβολίζουμε με α . Το σύνολο όλων των δυνατών δράσεων που μπορούμε να επιλέξουμε συμβολίζεται με \mathcal{A} . Η ζημιά που μπορεί να επέλθει από την επιλογή μίας συγκεκριμένης δράσης α ενώ η αληθής κατάσταση είναι θ καθορίζεται από μία **συνάρτηση ζημιάς** $L(\alpha, \theta)$, η οποία είναι φραγμένη από κάτω. Επειδή η αληθής κατάσταση θ δεν μας είναι γνωστή, άμεσα συνδεδεμένη με την επιλογή μίας συγκεκριμένης δράσης α είναι η **κατά Bayes αναμενόμενη ζημιά**, που δίνεται από τον τύπο:

$$R(\alpha) = E[L(\alpha, \theta) | x] = \int_{\Theta} L(\alpha, \theta) p(\theta | x) d\theta. \quad (1.10)$$

Έτσι, η βέλτιστη δράση είναι αυτή που συνεπάγεται τη χαμηλότερη κατά Bayes αναμενόμενη ζημιά, δηλαδή η δράση α^* είναι βέλτιστη αν και μόνο αν:

$$R(\alpha^*) = \inf \{R(\alpha) : \alpha \in \mathcal{A}\}. \quad (1.11)$$

Η δράση α καλείται τότε **δράση κατά Bayes**.

Στην περίπτωση που στη θέση του χώρου δράσεων \mathcal{A} έχουμε τον παραμετρικό χώρο Θ , η δράση είναι η επιλογή μίας τιμής θ για την παράμετρο θ . Συνεπώς, στο πλαίσιο της Μπεϋζιανής Στατιστικής, η κατά Bayes δράση δίνει τη βέλτιστη - με βάση την συνάρτηση ζημιάς $L(\theta, \theta)$ που εφαρμόζουμε - εκτίμηση του

θ . Δηλαδή, λαμβάνοντας υπόψη την (1,10), η κατά Bayes εκτίμηση του θ είναι η ποσότητα θ που ελαχιστοποιεί την κατά Bayes αναμενόμενη ζημιά:

$$R(\theta) = E[L(\theta, \theta) | \mathbf{x}] = \int_{\Theta} L(\theta, \theta) p(\theta | \mathbf{x}) d\theta. \quad (1.12)$$

Όταν χρησιμοποιούμε τετραγωνική συνάρτηση κόστους, όταν δηλαδή $L(\theta, \theta) = \|\theta - \theta\|^2$, τότε η βέλτιστη εκτίμηση του θ είναι η μέση τιμή του με βάση την ύστερη κατανομή $p(\theta | \mathbf{x})$, δηλαδή η δεσμευμένη μέση τιμή $E(\theta | \mathbf{x})$. Έχουμε δηλαδή:

$$E(\theta | \mathbf{x}) = \arg \min_{\theta \in \Theta} \int_{\Theta} \|\theta - \theta\|^2 p(\theta | \mathbf{x}) d\theta, \quad (1.13)$$

η οποία είναι η εκτιμήτρια ελάχιστου μέσου τετραγωνικού σφάλματος με βάση την ύστερη κατανομή της παραμέτρου. Την ίδια εκτίμηση (δηλαδή την ύστερη μέση τιμή) παίρνουμε και για οποιαδήποτε συνάρτηση ζημιάς από την **κλάση αποκλίσεων Bregman** [Bregman Divergences], μία ιδιαίτερα πλούσια κλάση «αποστάσεων», λεπτομέρειες για την οποία ο αναγνώστης μπορεί να αναζητήσει στη βιβλιογραφία (Kokolakis et al., 2006).

Πριν προχωρήσουμε σε μία σύντομη παρουσίαση των μεθόδων MCMC, να αναφέρουμε ότι περισσότερες λεπτομέρειες καθώς και αναλυτικότεροι υπολογισμοί για όσα παραθέσαμε ως τώρα μπορούν να αναζητηθούν στη βιβλιογραφία (Κοκολάκης και Σπηλιώτης, 2010).

1.3 Μέθοδοι MCMC

Οι μέθοδοι MCMC (Markov Chain Monte Carlo) έχουν βοηθήσει πολύ την πρακτική εφαρμογή της Μπεϋζιανής μεθοδολογίας, αντιμετωπίζοντας διεξοδικά το πρόβλημα του υπολογισμού της posterior κατανομής (ή καλύτερα των χαρακτηριστικών της που μας ενδιαφέρουν). Η βασική ιδέα τους είναι απλή: αντί να υπολογίσουμε αναλυτικά τις ποσότητες που μας ενδιαφέρουν σε πολύπλοκες posterior κατανομές, προσομοιώνουμε ένα δείγμα τιμών από μία κατάλληλη Μαρκοβιανή Αλυσίδα που βρίσκεται σε κατάσταση ισορροπίας. Έτσι μπορούμε να

υπολογίσουμε τα χαρακτηριστικά που θέλουμε (μέση τιμή, διασπορά κ.ά.) μέσω των αντίστοιχων τιμών του δείγματος που κατασκευάσαμε.

Πιο συγκεκριμένα, έστω ότι έχουμε μία συνάρτηση πιθανότητας $\pi(\cdot)$ από την οποία θέλουμε να προσομοιώσουμε ένα τυχαίο δείγμα. Τότε επιχειρούμε να κατασκευάσουμε μία Μαρκοβιανή Αλυσίδα με πυρήνα μετάβασης [transition kernel] $P(x, A)$ (πιθανότητα μετακίνησης από το σημείο x στο σύνολο A), με την ιδιότητα η συνάρτηση $P(x, dy)$ να συγκλίνει στην κατανομή $\pi(\cdot)$. Δηλαδή, η κατανομή-στόχος πρέπει να αποτελεί την κατανομή ισορροπίας της Μαρκοβιανής Αλυσίδας που κατασκευάζουμε. Έτσι, αν η αλυσίδα έχει αρχικοποιηθεί με μία τιμή x_0 και έχουν περάσει τα πρώτα N_0 στάδια της αλυσίδας [burn-in period], τότε για αρκετά μεγάλο N_0 οι τιμές από το στάδιο αυτό και μετά μπορούν να θεωρηθούν ως δείγμα από την κατανομή $\pi(\cdot)$. Το πρόβλημά μας λοιπόν είναι να βρούμε κατάλληλη συνάρτηση $P(x, dy)$.

Θα δούμε πώς επιτυγχάνεται αυτό με τη βοήθεια δύο αλγορίθμων που χρησιμοποιούνται ευρύτατα τα τελευταία χρόνια. Αυτοί είναι ο αλγόριθμος Metropolis-Hastings και ο αλγόριθμος Gibbs Sampling. Μάλιστα, οι αλγόριθμοι αυτοί μπορούν να εφαρμοστούν και αν δεν γνωρίζουμε την σταθερά κανονικοποίησης, δηλαδή αν γνωρίζουμε μόνο τον πυρήνα της ύστερης κατανομής. Να σημειώσουμε ότι ο δεύτερος αλγόριθμος μπορεί να προκύψει ως ειδική περίπτωση του πρώτου (Chib and Greenberg, 1995). Ωστόσο, λόγω της εκτεταμένης χρήσης αυτής της ειδικής περίπτωσης θα τους αναφέρουμε ξεχωριστά, όπως συνηθίζεται.

A) Αλγόριθμος Metropolis-Hastings

Έστω πως θέλουμε να προσομοιώσουμε τιμές από μία κατανομή $\pi(x)$. Η βασική ιδέα του αλγορίθμου είναι να χρησιμοποιήσουμε μία *κατανομή εισήγησης* $q(x, \cdot)$ από την οποία μπορούμε εύκολα να προσομοιώσουμε και κάθε φορά να δεχόμαστε ή να απορρίπτουμε την τιμή y που προσομοιώσαμε σύμφωνα με μία συγκεκριμένη *πιθανότητα αποδοχής* a_{MH} . Αν απορρίψουμε την τιμή y παραμένουμε στην τωρινή κατάσταση της αλυσίδας, έστω x .

Για να οδηγηθούμε στην επιθυμητή στάσιμη κατανομή αποδεικνύεται ότι πρέπει να ορίσουμε ως πιθανότητα αποδοχής την:

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}. \quad (1.14)$$

Η συγκεκριμένη επιλογή βασίζεται στην έννοια της *αντιστρεψιμότητας* και ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στην βιβλιογραφία (Chib and Greenberg, 1995) για να δει πώς πηγάζει.

Σύμφωνα με τα παραπάνω, προκύπτει ο ακόλουθος αλγόριθμος για την κατασκευή μίας Μαρκοβιανής Αλυσίδας με στάσιμη κατανομή $p(\theta | \mathbf{x})$, όπου έχουμε πάρει (αυθαίρετα) ως αρχική τιμή την θ_0 :

- Επανάλαβε για $j=1, 2, \dots, N$.
- Προσομοίωσε y από την κατανομή $q(x_j, \cdot)$ και u από την ομοιόμορφη κατανομή $U(0, 1)$.
- Αν $u \leq \alpha(x_j, y)$ τότε θέσε $x_{j+1} = y$, αλλιώς θέσε $x_{j+1} = x_j$.
- Επίστρεψε τις τιμές $\{x_1, x_2, \dots, x_N\}$.

Να επισημάνουμε ότι οι προσομοιωμένες τιμές μπορούν να θεωρηθούν ως δείγμα από την κατανομή $\pi(x)$ μόνο αφού η αλυσίδα έχει φτάσει στη στάσιμη κατανομή, ενώ η σύγκλιση σε αυτήν επιτυγχάνεται με τις συνθήκες της *απεριοδικότητας* και της *μη υποβιβασιμότητας*. Ωστόσο, η ταχύτητα της σύγκλισης δεν καθορίζεται από αυτές τις συνθήκες και κατά συνέπεια το πλήθος των τιμών N_0 που αφήνουμε να «τρέξει» η αλυσίδα μέχρι να φτάσει στη στάσιμη κατανομή επιλέγεται συνήθως εμπειρικά.

Δύο μεγάλα πλεονεκτήματα του αλγορίθμου είναι αφενός ότι δεν χρειάζεται να γνωρίζουμε την σταθερά κανονικοποίησης της ύστερης κατανομής που μας ενδιαφέρει και αφετέρου ότι η επιλογή της κατανομής εισήγησης δεν επηρεάζει την σύγκλιση της αλυσίδας, δηλαδή η αλυσίδα θα συγκλίνει στη στάσιμη κατανομή που θέλουμε (κάτω από τις συνθήκες που αναφέραμε) ανεξαρτήτως της κατανομής εισήγησης που θα επιλέξουμε. Βέβαια, η επιλογή αυτή επηρεάζει την ταχύτητα της

σύγκλισης και καλό είναι όταν μπορούμε να επιλέγουμε κατανομές εισήγησης που μοιάζουν με υπερκαλυπτόμενες εκδοχές της ύστερης κατανομής που μας ενδιαφέρει.

Επίσης, για λόγους απλότητας, συχνά επιλέγονται συμμετρικές κατανομές εισήγησης, δηλαδή κατανομές q τέτοιες ώστε $q(x, y) = q(y, x)$. Χρησιμοποιώντας τέτοιες κατανομές, η πιθανότητα αποδοχής (1.14) απλοποιείται ως:

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}. \quad (1.15)$$

Αυτό σημαίνει ότι αν $\pi(y) \geq \pi(x)$ η αλυσίδα πηγαίνει στο y , ενώ αν $\pi(y) < \pi(x)$ η μετακίνηση στο y γίνεται με πιθανότητα $\pi(y) / \pi(x)$. Άλλες συνηθισμένες επιλογές της κατανομής εισήγησης αναφέρονται στη βιβλιογραφία (Chib and Greenberg, 1995).

Πολλές παραλλαγές και ειδικές περιπτώσεις του παραπάνω αλγορίθμου χρησιμοποιούνται και, όπως ήδη αναφέραμε, μία τέτοια ειδική περίπτωση που χρησιμοποιείται ευρύτατα είναι ο αλγόριθμος Gibbs Sampling που θα παρουσιάσουμε τώρα.

B) Αλγόριθμος Gibbs Sampling

Ο αλγόριθμος Gibbs Sampling είναι ιδιαίτερα αποτελεσματικός στα πολυδιάστατα προβλήματα και υλοποιείται ως ακολούθως.

Έστω ένα k -διάστατο διάνυσμα $\mathbf{x} = (x_1, \dots, x_k) \in X$, $k \geq 2$, με κατανομή $\pi(\mathbf{x})$, όπου τα x_i δεν είναι απαραίτητα μονοδιάστατα. Υποθέτουμε ακόμα πως μπορούμε να προσομοιώσουμε από τις δεσμευμένες κατανομές $\pi_i(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$, $i = 1, \dots, k$ [full conditionals]. Τότε ο αλγόριθμος που ακολουθεί κατασκευάζει δείγματα από την κατανομή $\pi(\mathbf{x})$ μέσα από διαδοχικές προσομοιώσεις από τις δεσμευμένες κατανομές που περιγράψαμε:

Διάλεξε μία τυχαία αρχική τιμή $\mathbf{x}^0 = (x_1^0, \dots, x_k^0) \in X$ και θέσε $j=1$.

1. Προσομοίωσε $x_1^j \sim \pi_1(x_1 | x_2^{j-1}, \dots, x_k^{j-1})$,

προσομοίωσε $x_2^j \sim \pi_2(x_2 | x_1^j, x_3^{j-1}, \dots, x_k^{j-1})$,

...

προσομοίωσε $x_k^j \sim \pi_k(x_k | x_1^j, \dots, x_{k-1}^j)$.

2. Θέσε $j=j+1$ και πήγαινε στο 1.

Όπως και με τον αλγόριθμο Metropolis-Hastings, ένα βασικό ζήτημα είναι η ταχύτητα σύγκλισης της αλυσίδας, πόσο μεγάλη δηλαδή πρέπει να είναι η τιμή του N_0 που θα αφήσουμε την αλυσίδα να «τρέξει» μέχρι να μπορούμε να θεωρούμε ότι έχουμε τυχαίο δείγμα από την κατανομή που μας ενδιαφέρει. Πάλι η επιλογή γίνεται τις περισσότερες φορές με βάση την εμπειρία μας, ωστόσο ο αναγνώστης μπορεί να ανατρέξει στη βιβλιογραφία για αναλυτικότερες πληροφορίες (Casella and George, 1992).

Τέλος, να σημειώσουμε ότι οι δύο αλγόριθμοι που παραθέσαμε μπορούν να συνδυαστούν και να προκύψουν άλλες μορφές που χρησιμοποιούνται συχνά (Chib and Greenberg, 1995).

Κεφάλαιο 2

Κυρτοί Συνδυασμοί Μοντέλων

Η μέθοδος των κυρτών συνδυασμών μοντέλων [Bayesian Model Averaging] που θα αναπτύξουμε στο κεφάλαιο αυτό χρησιμοποιείται ως μία λύση στο πρόβλημα της αβεβαιότητας σχετικά με τη χρήση ενός συγκεκριμένου μοντέλου για την περιγραφή ενός προβλήματος. Θα ξεκινήσουμε με μία εισαγωγή σχετικά με την επιλογή μοντέλου στην Μπεϋζιανή Στατιστική και ακολούθως θα περιγράψουμε τη μέθοδο που προαναφέραμε, εξηγώντας παράλληλα τα πλεονεκτήματά της, τις δυσκολίες της και μεθόδους που έχουν προταθεί για την αντιμετώπισή τους.

2.1 Μπεϋζιανή Επιλογή Μοντέλου

Σε κάθε στατιστικό πρόβλημα, υπάρχει ένας μεγάλος αριθμός μοντέλων τα οποία ενδέχεται να το περιγράφουν ικανοποιητικά. Πολλές φορές, ωστόσο, το ενδιαφέρον μας εστιάζεται στην επιλογή ενός μόνο από αυτά τα μοντέλα, είτε γιατί αυτό δείχνει να περιγράφει πολύ καλά τα δεδομένα μας είτε για απλοποίηση στους υπολογισμούς μας. Αυτή είναι η συνήθης πρακτική στην στατιστική, αν και - όπως θα αναλύσουμε στη συνέχεια - εμπεριέχει αρκετό ρίσκο. Προς το παρόν, θα δούμε πώς συγκρίνουμε και επιλέγουμε μοντέλα στα πλαίσια της Μπεϋζιανής στατιστικής.

Όπως συμβαίνει και με τις παραμέτρους που είδαμε στην εισαγωγή, έτσι και τα πιθανά μοντέλα είναι άγνωστες ποσότητες που πρέπει να εκτιμηθούν. Η Μπεϋζιανή επιλογή μοντέλου, λοιπόν, χρησιμοποιεί συγκρίσεις μεταξύ των πιθανών μοντέλων, οι οποίες μας υποδεικνύουν ποιο είναι το βέλτιστο μοντέλο, υπό την έννοια ποιο μοντέλο είναι αυτό που ελαχιστοποιεί ενδεχόμενη ζημιά από την επιλογή του (ή ποιο μοντέλο μεγιστοποιεί μία συνάρτηση χρησιμότητας). Έχουμε λοιπόν να κάνουμε με την εύρεση του *βέλτιστου* για την περίπτωση μας μοντέλου και όχι του «σωστού» μοντέλου (αν κάτι τέτοιο μπορούμε να πούμε ότι υπάρχει). Όπως και στην περίπτωση των παραμέτρων, έτσι και εδώ τα συμπεράσματά μας βασίζονται σε συγκεκριμένες posterior πιθανότητες, τις οποίες όταν δεν μπορούμε να υπολογίσουμε

αναλυτικά (που είναι και το πιο σύνηθες) χρησιμοποιούμε μεθόδους MCMC για να εκτιμήσουμε τα χαρακτηριστικά τους.

Πιο συγκεκριμένα, έστω $\mathcal{M} = \{M_1, \dots, M_k\}$ ένα σύνολο από k πιθανά μοντέλα που θεωρούμε για τα δεδομένα μας, που τα συμβολίζουμε με \mathbf{Y} . Έστω ακόμα ότι για το μοντέλο M_j η μεταβλητή Y έχει συνάρτηση πυκνότητας πιθανότητας $p(Y | \theta_j, M_j)$, όπου θ_j το διάνυσμα των άγνωστων παραμέτρων του μοντέλου M_j . Κατά την Μπεϋζιανή προσέγγιση πρέπει αρχικά να καθορίσουμε μία prior κατανομή $p(\theta_j | M_j)$ στις παραμέτρους του κάθε μοντέλου και μία prior πιθανότητα $p(M_j)$ σε κάθε μοντέλο. Αυτή η διαδικασία ισοδυναμεί με το ακόλουθο *ιεραρχικό* μοντέλο τριών σταδίων το οποίο θεωρούμε ότι παράγει τα δεδομένα \mathbf{Y} :

1. Το μοντέλο M_j παράγεται από την κατανομή $p(M_1), \dots, p(M_k)$.
2. Η διανυσματική παράμετρος θ_j παράγεται από την $p(\theta_j | M_j)$.
3. Τα δεδομένα \mathbf{Y} παράγονται από την $p(Y | \theta_j, M_j)$.

Το πρόβλημα της επιλογής μοντέλου συνίσταται τότε στο να βρούμε ποιο από τα μοντέλα της κλάσης \mathcal{M} παρήγαγε τα δεδομένα μας, δηλαδή ποιο μοντέλο παρήγαγε η κατανομή $p(M_1), \dots, p(M_k)$ στο πρώτο βήμα της παραπάνω διαδικασίας.

Η posterior πιθανότητα να είναι το μοντέλο M_j αυτό που παρήγαγε τα δεδομένα \mathbf{Y} παίζει καθοριστικό ρόλο στα συμπεράσματά μας και δίνεται από το Θεώρημα Bayes ως:

$$p(M_j | \mathbf{Y}) = \frac{p(\mathbf{Y} | M_j)p(M_j)}{\sum_{i=1}^k p(\mathbf{Y} | M_i)p(M_i)}, \quad (2.1)$$

όπου:

$$p(\mathbf{Y} | M_i) = \int p(\mathbf{Y} | \theta_i, M_i)p(\theta_i | M_i)d\theta_i, \quad i = 1, \dots, k, \quad (2.2)$$

είναι η *περιθώρια πιθανοφάνεια* του μοντέλου M_j . Να σημειώσουμε εδώ ότι όλες οι πιθανότητες που θεωρήσαμε είναι δεσμευμένες στο \mathcal{M} , δηλαδή το σύνολο όλων των πιθανών μοντέλων που έχουμε θεωρήσει.

Η σύγκριση μεταξύ δύο μοντέλων, έστω M_i και M_j , περιγράφεται τότε από τον λόγο των posterior πιθανοτήτων:

$$\frac{p(M_i | \mathbf{Y})}{p(M_j | \mathbf{Y})} = \frac{p(\mathbf{Y} | M_i) p(M_i)}{p(\mathbf{Y} | M_j) p(M_j)}. \quad (2.3)$$

Ο πρώτος παράγοντας στο γινόμενο του δεύτερου μέλους είναι γνωστός ως **παράγοντας Bayes**, ενώ ο δεύτερος είναι ο λόγος των prior πιθανοτήτων των δύο μοντέλων.

Η posterior κατανομή των μοντέλων $\{p(M_1 | \mathbf{Y}), \dots, p(M_k | \mathbf{Y})\}$ συνοψίζει όλη την πληροφορία που έχουμε στη διάθεσή μας μετά την παρατήρηση των δεδομένων \mathbf{Y} και μας παρέχει μία πλήρη αναπαράσταση της αβεβαιότητάς μας σχετικά με την επιλογή του μοντέλου. Έτσι, οι όποιες αποφάσεις μας λαμβάνονται με γνώμονα αυτήν την κατανομή. Χρησιμοποιώντας την πιθανότητα $p(M_j | \mathbf{Y})$ ως μία μονάδα της καταλληλότητας του μοντέλου M_j , μία απλή στρατηγική θα είναι να διαλέξουμε το μοντέλο για το οποίο αυτή η πιθανότητα μεγιστοποιείται.

Ωστόσο, η συγκεκριμένη στρατηγική πολλές φορές δεν είναι η καλύτερη, όπως θα δούμε. Εμείς θα ασχοληθούμε εκτενώς με μία άλλη στρατηγική, γνωστή στην ξενόγλωσση βιβλιογραφία ως «**Bayesian Model Averaging**». Εναλλακτικά, θα χρησιμοποιούμε και τον όρο «**κυρτοί συνδυασμοί μοντέλων**» για να την περιγράψουμε, μιας και όπως θα δούμε για αυτό ακριβώς πρόκειται.

2.2 Η Μέθοδος των Κυρτών Συνδυασμών Μοντέλων

Όπως είπαμε και προηγουμένως, η συνήθης πρακτική στην στατιστική περιλαμβάνει τον έλεγχο κάποιων μοντέλων ως προς το πόσο καλά περιγράφουν τα δεδομένα μας και, αν κάτι τέτοιο συμβαίνει ικανοποιητικά για κάποιο από αυτά, τη χρήση του για την εκτίμηση των μεγεθών που μας ενδιαφέρουν, αλλά και την πρόβλεψή τους. Έτσι, όλα τα συμπεράσματά μας βασίζονται αποκλειστικά και μόνο στην υιοθέτηση ενός μοντέλου, σαν να πρόκειται όντως για το μοντέλο από το οποίο προέρχονται τα δεδομένα μας. Όμως, πολλές φορές ενδέχεται να υπάρχουν και εναλλακτικά μοντέλα τα οποία περιγράφουν εξίσου καλά τα δεδομένα μας, αλλά

οδηγούν σε εντελώς διαφορετικές εκτιμήσεις και προβλέψεις. Ο αναγνώστης μπορεί να δει τέτοια παραδείγματα στη βιβλιογραφία (Draper, 1995). Κατά συνέπεια, η εξαγωγή συμπερασμάτων που βασίζεται σε ένα μόνο μοντέλο εμπεριέχει αρκετό κίνδυνο.

Η μέθοδος που θα περιγράψουμε τώρα και που θα μας απασχολήσει σε όλο το υπόλοιπο κείμενο χρησιμοποιεί κυρτούς συνδυασμούς μοντέλων, καθένα από τα οποία έχει ως συντελεστή βάρους την posterior πιθανότητά του. Κατά τον τρόπο αυτό χειριζόμαστε καλύτερα την αβεβαιότητα που υπάρχει κατά την επιλογή ενός μοντέλου, η οποία αποτελεί σημαντικό μέρος της συνολικής αβεβαιότητάς μας για το εκάστοτε πρόβλημα. Μία πολύ θετική συνέπεια αυτής της διαδικασίας είναι ότι μας δίνει κατά μέσο όρο καλύτερη προβλεπτική ικανότητα. Θα αναφερθούμε σε αυτό εκτενέστερα στη συνέχεια.

Προς το παρόν, ας θεωρήσουμε εκ νέου το σύνολο $\mathcal{M} = \{M_1, \dots, M_k\}$ που περιλαμβάνει k πιθανά μοντέλα για τα δεδομένα μας \mathbf{Y} και ας συμβολίσουμε με Δ την ποσότητα για την οποία μας ενδιαφέρει να εξάγουμε συμπεράσματα. Για παράδειγμα, η ποσότητα Δ μπορεί να είναι κάποια παράμετρος ή μία μελλοντική παρατήρηση Y_f . Τότε η μέθοδος του Bayesian Model Averaging συνίσταται στο να θεωρήσουμε την εξής ύστερη κατανομή για την Δ :

$$p(\Delta | \mathbf{Y}) = \sum_{j=1}^k p(\Delta | M_j, \mathbf{Y}) p(M_j | \mathbf{Y}), \quad (2.4)$$

όπου η ύστερη κατανομή $p(\Delta | M_j, \mathbf{Y})$ της Δ δεδομένου ενός συγκεκριμένου μοντέλου M_j , δίνεται από τη σχέση:

$$p(\Delta | M_j, \mathbf{Y}) = \int p(\Delta | \boldsymbol{\theta}_j, M_j, \mathbf{Y}) p(\boldsymbol{\theta}_j | M_j, \mathbf{Y}) d\boldsymbol{\theta}_j, \quad (2.5)$$

ενώ η ύστερη πιθανότητα του μοντέλου M_j , $p(M_j | \mathbf{Y})$, δίνεται από την (2.1).

Έχουμε λοιπόν να κάνουμε με έναν μέσο όρο των ύστερων κατανομών που προκύπτουν από κάθε μοντέλο, όπου σαν συντελεστή βάρους βάζουμε κάθε φορά την αντίστοιχη ύστερη πιθανότητα μοντέλου. Δεδομένου ότι αυτοί οι συντελεστές έχουν ως άθροισμα τη μονάδα, γίνεται τώρα εμφανής η χρήση της ονομασίας «κυρτοί

συνδυασμοί μοντέλων» που προαναφέραμε. Να σημειώσουμε επίσης ότι στην ακραία περίπτωση που όλη η ύστερη πιθανότητα συγκεντρώνεται σε ένα μοντέλο, δεν υπάρχει αβεβαιότητα για την επιλογή μοντέλου και ο μέσος όρος που περιγράφεται από τη σχέση (2.4) θα μας δώσει το ίδιο αποτέλεσμα με την επιλογή ενός μοντέλου που περιγράψαμε στην προηγούμενη παράγραφο. Έτσι, η επιλογή αυτή μπορεί να προκύψει σαν υποπερίπτωση της μεθοδολογίας που αναπτύσσουμε τώρα.

Η posterior μέση τιμή που προκύπτει για το Δ είναι σύμφωνα με τα παραπάνω η:

$$\begin{aligned} E[\Delta | \mathbf{Y}] &= E_{\mathcal{M}}[E(\Delta | \mathbf{Y}, \mathcal{M})] \\ &= \sum_{j=1}^k E[\Delta | M_j, \mathbf{Y}] p(M_j | \mathbf{Y}), \end{aligned} \quad (2.6)$$

ενώ η posterior διασπορά προκύπτει ως

$$\begin{aligned} \text{Var}[\Delta | \mathbf{Y}] &= E_{\mathcal{M}}[\text{Var}(\Delta | \mathbf{Y}, \mathcal{M})] + \text{Var}_{\mathcal{M}}[E(\Delta | \mathbf{Y}, \mathcal{M})] \\ &= \sum_{j=1}^k \{ \text{Var}[\Delta | M_j, \mathbf{Y}] + (E[\Delta | M_j, \mathbf{Y}])^2 \} p(M_j | \mathbf{Y}) - E[\Delta | \mathbf{Y}]^2. \end{aligned} \quad (2.7)$$

Οι παραπάνω τύποι είναι γενικοί και μπορούν να εφαρμοστούν σε οποιαδήποτε κατηγορία μοντέλων. Στο επόμενο κεφάλαιο θα εστιάσουμε στα γενικευμένα γραμμικά μοντέλα. Μέχρι τότε θα συνεχίσουμε με την παρουσίαση των βασικών ζητημάτων σχετικά με τη μέθοδο, ξεκινώντας από τις δυσκολίες που καλούμαστε να αντιμετωπίσουμε για να τη θέσουμε σε εφαρμογή.

Οι δυσκολίες αυτές αφορούν τα εξής: Πρώτον, πρέπει να καθοριστούν οι prior πιθανότητες για κάθε μοντέλο, $p(M_j)$, $j=1, \dots, k$, και οι prior κατανομές για τις παραμέτρους κάθε μοντέλου, $p(\theta_j | M_j)$, $j=1, \dots, k$. Δεύτερον, τόσο τα ολοκληρώματα της μορφής (2.2) όσο και της μορφής (2.5) που απαιτούνται για τον υπολογισμό του αθροίσματος (2.4) είναι πολύ δύσκολο να υπολογιστούν, με αποτέλεσμα να καταφεύγουμε σε προσεγγιστικές λύσεις. Τέλος, το πλήθος k των μοντέλων που θεωρούμε είναι συχνά πολύ μεγάλο, με αποτέλεσμα ο αναλυτικός υπολογισμός του αθροίσματος (2.4) να καθίσταται αδύνατος.

Για όλα τα παραπάνω προβλήματα θα παρουσιάσουμε διάφορες λύσεις που έχουν προταθεί για την αντιμετώπισή τους.

2.3 Επιλογή των Πρότερων Κατανομών

Η αποτελεσματικότητα της μεθόδου που περιγράψαμε εξαρτάται προφανώς και από τον καθορισμό αφενός των πρότερων πιθανοτήτων των μοντέλων $p(M_1), \dots, p(M_k)$ και αφετέρου των πρότερων κατανομών των παραμέτρων $p(\theta_j | M_j)$, $j = 1, \dots, k$. Όλα τα αποτελέσματα που έχουμε δώσει στις προηγούμενες παραγράφους θεωρούν δεδομένο τον «ορθό» καθορισμό αυτών των κατανομών. Θεωρητικά, αυτές οι prior κατανομές πρέπει να αναπαριστούν την εκ των προτέρων αβεβαιότητά μας για όλες τις εμπλεκόμενες άγνωστες ποσότητες, οπότε η posterior που δίνεται στην (2.4) ερμηνεύεται ως η επικαιροποίηση αυτής της αβεβαιότητας μετά την παρατήρηση των δεδομένων \mathbf{Y} .

Στην πράξη, ωστόσο, ο καθορισμός αυτών των πρότερων κατανομών υπόκειται σε πολλούς περιορισμούς. Τα πιθανά μοντέλα που θεωρούμε για τα προβλήματά μας είναι συχνά πολυάριθμα, ενώ και οι παράμετροι των μοντέλων είναι συνήθως πολυδιάστατες και πολύπλοκες. Έτσι γίνεται ουσιαστικά ανέφικτο να περιγράψουμε την αβεβαιότητά μας επακριβώς με μαθηματικό τρόπο και ως εκ τούτου καταφεύγουμε σε απλοποιήσεις.

Η πιο απλή και συνηθισμένη πρακτική προσέγγιση στο πρόβλημα του καθορισμού των πρότερων πιθανοτήτων των μοντέλων είναι η κατασκευή *μη πληροφοριακών* πρότερων κατανομών. Έτσι, στην περίπτωση της κλάσης $\mathcal{M} = \{M_1, \dots, M_k\}$ μία πολύ απλή και δημοφιλής επιλογή είναι να θέσουμε την ομογενή prior κατανομή:

$$p(M_j) = \frac{1}{k}, \quad j = 1, \dots, k, \quad (2.8)$$

η οποία είναι μη πληροφοριακή, υπό την έννοια ότι δίνει την ίδια a priori πιθανότητα σε καθένα από τα k το πλήθος θεωρούμενα μοντέλα. Με την συγκεκριμένη prior, η posterior κάθε μοντέλου $p(M_j | \mathbf{Y})$ είναι ανάλογη της πιθανοφάνειας $p(\mathbf{Y} | M_j)$.

Η επιλογή μίας prior όπως η παραπάνω ενδείκνυται περισσότερο για περιπτώσεις όπου η εκ των προτέρων γνώση μας για την καταλληλότητα των μοντέλων είναι περιορισμένη. Ωστόσο, αν και βολική για τους υπολογισμούς μας, η χρήση ομοιόμορφων prior κατανομών παρουσιάζει και αρκετά μειονεκτήματα και συχνά προτείνονται επιλογές πληροφοριακών prior κατανομών. Αυτές μπορούν να περιλαμβάνουν, για παράδειγμα, τη χρήση κάποιων φανταστικών δεδομένων για το πρόβλημά μας, τα οποία πρέπει να έχουν δοθεί από κάποιον ειδικό (Hoeting et al., 1999). Επίσης, οι επιλογές των prior πιθανοτήτων των μοντέλων ποικίλλουν ανάλογα με την κλάση των μοντέλων (μοντέλα γραμμικής παλινδρόμησης, μοντέλα αναλογικής διακινδύνευσης κτλ.) που θεωρούμε κάθε φορά. Στο επόμενο κεφάλαιο θα εξετάσουμε πιο διεξοδικά τον καθορισμό των prior κατανομών στην περίπτωση των γενικευμένων γραμμικών μοντέλων.

Αφού έχουμε καθορίσει την prior κατανομή στο χώρο των πιθανών μοντέλων, καλούμαστε να καθορίσουμε και τις prior κατανομές των παραμέτρων $p(\theta_j | M_j)$. Και εδώ προσπαθούμε να κάνουμε επιλογές που να διευκολύνουν τους υπολογισμούς μας, και κυρίως τον υπολογισμό των ολοκληρωμάτων $p(\mathbf{Y} | M_j)$, $j = 1, \dots, k$, από τη σχέση (2.2). Στην περίπτωση μοντέλων της Εκθετικής Οικογένειας Κατανομών τα πράγματα απλοποιούνται με τη χρήση συζυγών πρότερων κατανομών. Σε κάθε περίπτωση, είναι καλό να χρησιμοποιούμε κατανομές $p(\theta_j | M_j)$ με μικρό αριθμό υπερπαραμέτρων. Και εδώ, το ενδιαφέρον μας θα εστιαστεί στην περίπτωση των γενικευμένων γραμμικών μοντέλων.

2.4 Υπολογισμός των Ολοκληρωμάτων

Αφού καθορίσουμε τις prior κατανομές που χρειάζεται, καλούμαστε να υπολογίσουμε το άθροισμα (2.4). Σε αυτό εμπλέκονται τα ολοκληρώματα (2.2) και (2.5), τα οποία σπάνια είναι εφικτό να υπολογιστούν αναλυτικά, με αποτέλεσμα να καταφεύγουμε σε προσεγγιστικές μεθόδους.

Κατ' αρχάς, τα ολοκληρώματα (2.2) μπορούν όπως είπαμε να υπολογιστούν αναλυτικά στην περίπτωση που έχουμε μοντέλο της Εκθετικής Οικογένειας Κατανομών και χρησιμοποιούμε συζυγή prior κατανομή, αλλά στην συνηθέστερη

περίπτωση που αυτό δεν είναι δυνατό υποχρεωνόμαστε να χρησιμοποιήσουμε προσεγγιστικές τεχνικές για τον υπολογισμό τους.

Μία τέτοια τεχνική που χρησιμοποιείται συχνά είναι η **μέθοδος ολοκλήρωσης του Laplace**, η οποία σε συγκεκριμένες περιπτώσεις δίνει την πολύ απλή **προσέγγιση BIC**. Η μέθοδος του Laplace για ολοκληρώματα βασίζεται στην ανάλυση κατά Taylor μίας πραγματικής συνάρτησης $f(\mathbf{u})$ ενός διανύσματος \mathbf{u} διάστασης d (Tierney and Kadane, 1986) και δίνει την εξής προσέγγιση για τον υπολογισμό του ολοκληρώματος (2.2):

$$p(\mathbf{Y} | M_j) \approx (2\pi)^{d_j/2} |A_j|^{-1/2} p(\mathbf{Y} | \tilde{\boldsymbol{\theta}}_j, M_j) p(\tilde{\boldsymbol{\theta}}_j | M_j), \quad (2.9)$$

όπου $\boldsymbol{\theta}_j$ η τιμή που μεγιστοποιεί την συνάρτηση $f(\boldsymbol{\theta}_j) = \log \{p(\mathbf{Y} | \boldsymbol{\theta}_j, M_j) p(\boldsymbol{\theta}_j | M_j)\}$, d_j το πλήθος των παραμέτρων του μοντέλου (δηλαδή η διάσταση του διανύσματος $\boldsymbol{\theta}_j$), A_j μείον τον αντίστροφο του Εσσιανού πίνακα της f υπολογισμένου στο $\boldsymbol{\theta}_j$ και n το μέγεθος του δείγματος.

Για μεγάλα δείγματα, έχουμε $\boldsymbol{\theta}_j \approx \tilde{\boldsymbol{\theta}}_j$, όπου $\tilde{\boldsymbol{\theta}}_j$ η εκτιμήτρια μέγιστης πιθανοφάνειας της παραμέτρου $\boldsymbol{\theta}_j$, και $\log |A_j| = -d_j \log n + O(1)$, οπότε τελικά καταλήγουμε στην προσέγγιση:

$$\log p(\mathbf{Y} | M_j) = \log p(\mathbf{Y} | \boldsymbol{\theta}_j, M_j) - d_j \log n + O(1). \quad (2.10)$$

Το παραπάνω αποτέλεσμα έχει υπολογιστεί με διαφορετικό τρόπο και από τον Schwarz, ο οποίος είχε βρει την προσέγγιση:

$$\log p(\mathbf{Y} | M_j) \approx \log p(\mathbf{Y} | \boldsymbol{\theta}_j, M_j) - (d_j / 2) \log n, \quad (2.10^*)$$

μία προσέγγιση που προτείνει - στα πλαίσια του προβλήματος της επιλογής μοντέλου - να επιλέγουμε το μοντέλο M για το οποίο η ποσότητα στα δεξιά της τελευταίας εξίσωσης μεγιστοποιείται. Το κριτήριο αυτό είναι γνωστό ως BIC (*Bayesian Information Criterion*). Θα επανέλθουμε στο θέμα του υπολογισμού αυτών των ολοκληρωμάτων και στο επόμενο κεφάλαιο, όπου θα το εξετάσουμε στην περίπτωση των γενικευμένων γραμμικών μοντέλων.

Όσον αφορά τον υπολογισμό των ολοκληρωμάτων (2.5), τώρα, μία απλή προσέγγιση που χρησιμοποιείται συχνά είναι η προσέγγιση που χρησιμοποιεί την εκτιμήτρια μέγιστης πιθανοφάνειας θ_j (*MLE approximation*), δηλαδή:

$$p(\Delta | M_j, \mathbf{Y}) \approx p(\Delta | M_j, \theta_j, \mathbf{Y}) , \quad (2.11)$$

η οποία είναι αρκετά αποτελεσματική σε πολλές περιπτώσεις (Draper, 1995; Raftery et al., 1995).

2.5 Υπολογισμός του Αθροίσματος

Όπως αναφέραμε, η άλλη δυσκολία που καλούμαστε να αντιμετωπίσουμε είναι το ενδεχομένως μεγάλο μέγεθος της κλάσης $\mathcal{M} = \{M_1, \dots, M_k\}$ των πιθανών μοντέλων. Για να αντιμετωπίσουμε αυτό το πρόβλημα, καταλήγουμε στο να υπολογίζουμε το άθροισμα (2.4) ή το (2.6) (εάν μας ενδιαφέρει η ύστερη μέση τιμή της Δ) σε ένα υποσύνολο του \mathcal{M} που υποστηρίζεται περισσότερο από τα δεδομένα.

Υπάρχουν δύο μεγάλες κατηγορίες τεχνικών για την εύρεση κατάλληλων υποσυνόλων. Η πρώτη από αυτές αφορά περιπτώσεις που το πλήθος των θεωρούμενων μοντέλων δεν είναι ιδιαίτερα μεγάλο. Τότε μπορούμε να βρούμε ένα κατάλληλο υποσύνολο με ντετερμινιστικό τρόπο. Η δεύτερη κατηγορία αφορά περιπτώσεις που το πλήθος k των μοντέλων είναι μεγαλύτερο, οπότε οι τεχνικές της προηγούμενης κατηγορίας κοστίζουν πολλούς υπολογισμούς, ενώ ενδέχεται να μην είναι και αποδοτικές. Τότε οδηγούμαστε στη χρήση MCMC τεχνικών, αξιοποιώντας τους αλγορίθμους Metropolis-Hastings και Gibbs Sampling για να βρούμε ένα κατάλληλο υποσύνολο στο οποίο θα αθροίσουμε.

Αναλύοντας την πρώτη κατηγορία, θέλουμε να δούμε ποιούς τρόπους και ποιά κριτήρια θα χρησιμοποιήσουμε για να καταλήξουμε σε κάποιο κατάλληλο υποσύνολο του χώρου των μοντέλων. Διάφορες τεχνικές έχουν προταθεί προς αυτήν την κατεύθυνση. Μία από αυτές είναι η μέθοδος του παράθυρου του Occam [**Occam's window**] των Madigan και Raftery (1994).

Η μέθοδος αυτή περιλαμβάνει δύο βασικές αρχές. Σύμφωνα με την πρώτη από αυτές, αν ένα μοντέλο εξηγεί τα δεδομένα πολύ χειρότερα από το βέλτιστο βάσει

ύστερης πιθανότητας μοντέλο, τότε πρέπει να απορριφθεί και να μην περιλαμβάνεται στα μοντέλα που λαμβάνουμε υπ' όψιν. Συγκεκριμένα, μοντέλα που δεν ανήκουν στην κλάση:

$$\mathcal{A}' = \left\{ M_j : \frac{\max_i \{p(M_i | \mathbf{Y})\}}{p(M_j | \mathbf{Y})} \leq C \right\}, \quad (2.12)$$

όπου C σταθερά που επιλέγεται κάθε φορά από τον ερευνητή, πρέπει να αποκλείονται από το άθροισμα (2.4).

Η δεύτερη αρχή που θέτουν οι Raftery και Madigan λέει ότι από το άθροισμα πρέπει να αποκλείονται επίσης πολύπλοκα μοντέλα που υποστηρίζονται από τα δεδομένα λιγότερο από τα απλούστερα μοντέλα των οποίων αποτελούν υποπερίπτωση. Συγκεκριμένα, ένα δεύτερο «φιλτράρισμα» των μοντέλων της αρχικής κλάσης \mathcal{M} προκύπτει αποκλείοντας από τη σχέση (2.4) μοντέλα που δεν ανήκουν στην κλάση:

$$\mathcal{B} = \left\{ M_j : \exists M_i \in \mathcal{A}' \text{ με } M_i \subset M_j \text{ και } \frac{p(M_i | \mathbf{Y})}{p(M_j | \mathbf{Y})} > 1 \right\}. \quad (2.13)$$

Τελικά, στη σχέση (2.4) αθροίζουμε στα μοντέλα που ανήκουν στην κλάση $\mathcal{A} = \mathcal{A}' \setminus \mathcal{B}$, όπου οι πιθανότητες πλέον είναι δεσμευμένες στο σύνολο των μοντέλων \mathcal{A} .

Οι δύο παραπάνω συνθήκες ελαττώνουν κατά πολύ τον αριθμό των μοντέλων στη σχέση (2.4). Αυτό που απαιτείται λοιπόν είναι μία στρατηγική αναζήτησης που να βρίσκει τα μοντέλα της κλάσης \mathcal{A} . Οι Madigan και Raftery (1994) πρότειναν μία τέτοια στρατηγική, η οποία βασίζεται σε δύο βασικές ιδέες.

Πρώτον, όταν ο αλγόριθμος συγκρίνει δύο μοντέλα που το ένα αποτελεί υποσύνολο του άλλου και απορρίπτει το πιο απλό μοντέλο, τότε και όλα τα υποσύνολα του πιο απλού μοντέλου θα πρέπει να απορρίπτονται άμεσα.

Η δεύτερη ιδέα αφορά την ερμηνεία του λόγου των ύστερων πιθανοτήτων μοντέλων $p(M_0 | \mathbf{Y}) / p(M_1 | \mathbf{Y})$, όπου το μοντέλο M_0 είναι «μικρότερο» από το M_1 . Συγκεκριμένα, αν υπάρχουν ενδείξεις υπέρ του M_0 τότε το μοντέλο M_1 απορρίπτεται, αλλά για να απορρίψουμε το μοντέλο M_0 θα πρέπει να έχουμε

ισχυρότερες ενδείξεις για το (μεγαλύτερο) μοντέλο M_1 . Αν ο λόγος των δύο πιθανοτήτων είναι τέτοιος ώστε να μην μπορούμε να αποφανθούμε, τότε κανένα από τα δύο μοντέλα δεν απορρίπτεται. Η περιοχή αυτή που οδηγεί σε ανικανότητα να αποφασίσουμε καλείται *παράθυρο του Occam*.

Διάφορες τιμές μπορούν να προταθούν ως κριτήρια για τη λήψη της απόφασής μας. Οι Madigan και Raftery (1994) πρότειναν να δεχόμαστε το M_0 αν $p(M_0 | \mathbf{Y}) / p(M_1 | \mathbf{Y}) > 1 = O_R$, δηλαδή αν απλώς η posterior πιθανότητα του M_0 είναι μεγαλύτερη από αυτήν του M_1 , ενώ να δεχόμαστε το M_1 αν $p(M_0 | \mathbf{Y}) / p(M_1 | \mathbf{Y}) < 1/20 = O_L$, δίνοντας έτσι μία πολύ αυστηρότερη συνθήκη για να δεχτούμε το μεγαλύτερο μοντέλο. Το παράθυρο του Occam σε αυτή την περίπτωση ταυτίζεται με το διάστημα $[1/20, 1]$. Μία άλλη επιλογή πρότειναν οι Raftery et al. (1995), υποστηρίζοντας ότι παίρνοντας $O_R = 20$ και $O_L = 1/20$ έχουμε σε πολλές περιπτώσεις καλύτερη προβλεπτική ικανότητα. Η επιλογή αυτή ουσιαστικά ακυρώνει τη δεύτερη από τις βασικές ιδέες που αναφέραμε.

Οι παραπάνω αρχές συνοψίζουν τη στρατηγική μας και μπορούν να «ρίξουν» τον αριθμό των μοντέλων που υπεισέρχονται στο άθροισμα (2.4) σε πολύ λιγότερα μοντέλα από αυτά που θεωρήσαμε αρχικά. Μάλιστα, οι Madigan και Raftery (1994) παρουσιάζουν έναν συγκεκριμένο αλγόριθμο για την στρατηγική αυτή.

Μία άλλη στρατηγική για την αναζήτηση των καλύτερων μοντέλων περιλαμβάνει τη χρήση του αλγόριθμου «leaps and bounds» για την επιλογή των μοντέλων που θα εισαχθούν στο άθροισμα (2.4). Θα αναφερθούμε σε αυτό και στο επόμενο κεφάλαιο.

Η δεύτερη κατηγορία τεχνικών για τον υπολογισμό του μέσου όρου (2.4) περιλαμβάνει, όπως είπαμε, τη χρήση αλγορίθμων MCMC. Μία τέτοια τεχνική είναι η \mathbf{MC}^3 (Markov Chain Monte Carlo Model Composition), η οποία προτάθηκε από τους Madigan και York (1995). Η τεχνική αυτή κάνει χρήση του αλγορίθμου Metropolis-Hastings, που αναφέραμε στο πρώτο κεφάλαιο. Στην περίπτωσή μας, η κατανομή από την οποία θέλουμε να προσομοιώσουμε είναι αυτή που προκύπτει από τις ύστερες πιθανότητες $p(M_j | \mathbf{Y})$ των μοντέλων που θεωρούμε στο άθροισμα (2.4).

Συγκεκριμένα, ας υποθέσουμε εκ νέου ότι $\mathcal{M} = \{M_1, \dots, M_k\}$ είναι ο χώρος των μοντέλων που θεωρούμε για το πρόβλημά μας. Τότε μπορούμε να κατασκευάσουμε μία μαρκοβιανή αλυσίδα $\{M_{(t)} : t = 1, 2, \dots\}$, με χώρο καταστάσεων \mathcal{M} και κατανομή ισορροπίας $p(M_j | \mathbf{Y})$, $j = 1, \dots, k$, και να προσομοιώσουμε αυτή την αλυσίδα για να πάρουμε παρατηρήσεις $M_{(1)}, \dots, M_{(N)}$. Για να θεωρούνται οι παρατηρήσεις αυτές δείγμα από την επιθυμητή κατανομή, θα πρέπει βέβαια να έχουμε αφήσει αρκετές τιμές της αλυσίδας ώστε να επέλθει η στασιμότητα.

Οι προσομοιωμένες παρατηρήσεις $M_{(t)}$ μπορούν τότε να χρησιμοποιηθούν για την εκτίμηση μίας άγνωστης ποσότητας Δ με τον ακόλουθο τρόπο. Έχοντας την έκφραση (2.6) για τον posterior μέσο της Δ σε ένα πρόβλημα που ο χώρος \mathcal{M} περιλαμβάνει πολλά μοντέλα, σκοπός μας είναι να προσεγγίσουμε αυτό το άθροισμα εκτελώντας το σε ένα υποσύνολο S του \mathcal{M} όπου αυτό γίνεται πιο εύκολα εφικτό. Δηλαδή, θέλουμε να υπολογίσουμε ένα άθροισμα της μορφής:

$$E[\Delta | \mathbf{Y}] = \sum_{M_j \in S} E[\Delta | M_j, \mathbf{Y}] p(M_j | \mathbf{Y}, S), \quad (2.14)$$

όπου S κατάλληλο υποσύνολο του \mathcal{M} και $p(M_j | \mathbf{Y}, S)$ μία κατανομή πιθανότητας στο S .

Μία επιλογή για το άθροισμα (2.14) είναι να θεωρήσουμε την μαρκοβιανή αλυσίδα που κατασκευάσαμε ως χώρο S και να υπολογίσουμε το άθροισμα:

$$E_f[\Delta | \mathbf{Y}] = \sum_{M_j \in S} E[\Delta | M_j, \mathbf{Y}] p_f(M_j | \mathbf{Y}, S), \quad (2.15)$$

όπου $p_f(M_j | \mathbf{Y}, S)$ είναι η συχνότητα εμφάνισης του M_j στην αλυσίδα S . Το άθροισμα (2.15) αποτελεί τότε μία εκτιμήτρια του ύστερου μέσου $E[\Delta | \mathbf{Y}]$.

Ακόμα καλύτερα, αν είναι εφικτό, μπορούμε να υπολογίσουμε την εξής εκτίμηση του μέσου:

$$E_g[\Delta | \mathbf{Y}] = \sum_{M_j \in S} E[\Delta | M_j, \mathbf{Y}] p_g(M_j | \mathbf{Y}, S), \quad (2.16)$$

όπου $p_s(M_j | \mathbf{Y}, S) = p(M_j | \mathbf{Y}) / p(S | \mathbf{Y})$ η κανονικοποιημένη τιμή του $p(M_j | \mathbf{Y})$ για κάθε $M_j \in S$ (Chipman et al., 2001).

Έτσι το ενδιαφέρον μας στρέφεται στην κατασκευή της μαρκοβιανής αλυσίδας $\{M_{(t)} : t=1, 2, \dots\}$. Αυτή μπορεί να πραγματοποιηθεί με εφαρμογή του αλγορίθμου Metropolis-Hastings. Για να γίνει αυτό, πρέπει αρχικά για κάθε μοντέλο $M \in \mathcal{M}$ να ορίσουμε μία γειτονιά $nbd(M)$ και έναν πίνακα πιθανοτήτων μετάβασης q θέτοντας $q(M \rightarrow M') = 0 \quad \forall M' \notin nbd(M)$ και $q(M \rightarrow M') \neq 0$ αν $M' \in nbd(M)$. Αν η αλυσίδα βρίσκεται στο μοντέλο M , τότε προσομοιώνουμε ένα μοντέλο M' μέσω της $q(M \rightarrow M')$ και το δεχόμαστε με πιθανότητα:

$$\min \left\{ 1, \frac{q(M | M') p(M' | \mathbf{Y})}{q(M' | M) p(M | \mathbf{Y})} \right\}, \quad (2.17)$$

αλλιώς το απορρίπτουμε και η αλυσίδα παραμένει στο ίδιο μοντέλο, M . Χρησιμοποιώντας συμμετρικούς πίνακες q , η παραπάνω πιθανότητα απλοποιείται ως:

$$\min \left\{ 1, \frac{p(M' | \mathbf{Y})}{p(M | \mathbf{Y})} \right\}, \quad (2.18)$$

Όπως συμβαίνει γενικά με τις MCMC τεχνικές, έτσι και με την MC³ το βασικό πλεονέκτημα είναι η ευελιξία της, ενώ το βασικό μειονέκτημα είναι τα θέματα που αφορούν τη σύγκλιση των μαρκοβιανών αλυσίδων. Θα δούμε στο επόμενο κεφάλαιο πώς χρησιμοποιείται αυτή η τεχνική στην περίπτωση των γενικευμένων γραμμικών μοντέλων. Θα αναφερθούμε επιπλέον και σε μία άλλη τεχνική MCMC που χρησιμοποιείται σε αυτή την περίπτωση και που κάνει χρήση του αλγορίθμου Gibbs Sampling, τη μέθοδο SSVS.

2.6 Προβλεπτική Ικανότητα των Κυρτών Συνδυασμών Μοντέλων

Έχουμε ήδη κάνει λόγο για την καλύτερη προβλεπτική ικανότητα που έχει η χρήση κυρτών συνδυασμών μοντέλων σε σύγκριση με τη χρήση ενός μόνο μοντέλου για την περιγραφή των δεδομένων μας. Αυτό αποτελεί το σημαντικότερο ίσως πλεονέκτημα της μεθόδου, μιας και ένας από τους βασικούς σκοπούς της στατιστικής

ανάλυσης είναι να κάνει προβλέψεις, δίνοντας μεγαλύτερες πιθανότητες σε γεγονότα που είναι πιο πιθανόν να συμβούν. Ως εκ τούτου, ένας τρόπος για να κρίνουμε την επιτυχία μίας στρατηγικής είναι να εξετάσουμε και αν είναι δυνατόν να μετρήσουμε το πόσο καλά προβλέπει τις μελλοντικές εμφανίσεις γεγονότων. Αυτό θα κάνουμε και στο τελευταίο κεφάλαιο όπου θα παρουσιάσουμε ένα συγκεκριμένο παράδειγμα με εφαρμογή του Bayesian Model Averaging.

Ένας τρόπος για να εξετάσουμε αυτή την προβλεπτική ικανότητα είναι ο ακόλουθος. Έχοντας κάποια δεδομένα D στη διάθεσή μας, τα χωρίζουμε με τυχαίο τρόπο σε δύο υποσύνολα. Στο πρώτο υποσύνολο εφαρμόζουμε τις τεχνικές επιλογής μοντέλου που θέλουμε να συγκρίνουμε. Το υποσύνολο αυτό το συμβολίζουμε με D_T (**training data**). Η προβλεπτική ικανότητα μπορεί να μετρηθεί τότε στο δεύτερο υποσύνολο, το οποίο συμβολίζουμε με D_p (**prediction data**).

Ένα μέτρο της προβλεπτικής ικανότητας μάς παρέχει τότε ο ακόλουθος κανόνας του Good (**logarithmic scoring rule of Good**). Σύμφωνα με αυτόν, η παρακάτω τιμή (**predictive log score**) μετρά την προβλεπτική ικανότητα ενός (μεμονωμένου) μοντέλου M χρησιμοποιώντας το άθροισμα των λογαρίθμων των παρατηρημένων τιμών της προβλεπτικής κατανομής για κάθε παρατήρηση του συνόλου D_p :

$$-\sum_{d \in D_p} \log p(d | M, D_T). \quad (2.19)$$

Αντίστοιχα, μετράει την προβλεπτική ικανότητα του κυρτού συνδυασμού μοντέλων μέσω της ποσότητας:

$$-\sum_{d \in D_p} \log \left\{ \sum_{M \in \mathcal{M}} p(d | M, D_T) p(M | D_T) \right\}, \quad (2.20)$$

όπου \mathcal{M} είναι το σύνολο των μοντέλων που θεωρούμε.

Όσο μικρότερη είναι η τιμή (2.19) ή (2.20) για ένα μοντέλο ή έναν κυρτό συνδυασμό μοντέλων αντίστοιχα, τόσο καλύτερη είναι η προβλεπτική ικανότητα. Δουλεύοντας στα πλαίσια της Μπεϋζιανής Θεωρίας Αποφάσεων, μπορούμε να εξηγήσουμε την καλύτερη πρόβλεψη των κυρτών συνδυασμών μοντέλων μέσω των συναρτήσεων ζημίας που μπορούμε να χρησιμοποιήσουμε, ανάλογα με το ρόλο της

ποσότητας Δ . Συγκεκριμένα, συμβολίζοντας με α τη δράση μας σχετικά με την επιλογή μοντέλου και με $L(\alpha, \Delta)$ την συνάρτηση ζημίας που χρησιμοποιούμε, ο σκοπός μας είναι να ελαχιστοποιήσουμε την αναμενόμενη ζημία:

$$\int L(\alpha, \Delta) p(\Delta | \mathbf{Y}) d\Delta, \quad (2.21)$$

όπου η ύστερη κατανομή $p(\Delta | \mathbf{Y})$ δίνεται από την (2.4).

Αν τώρα υποθέσουμε ότι η δράση α αφορά την επιλογή της $p(\Delta | M_j, \mathbf{Y})$ ως προβλεπτικής κατανομής για μία μελλοντική παρατήρηση Δ , τότε μία συνηθισμένη επιλογή είναι να θέσουμε ως συνάρτηση χρησιμότητας την συνάρτηση $u(\alpha, \Delta) = \log p(\Delta | M_j, \mathbf{Y})$ (*logarithmic score function*), ή ισοδύναμα να θέσουμε ως συνάρτηση ζημίας το αντίθετο αυτής της συνάρτησης, δηλαδή :

$$L(\alpha, \Delta) = -\log p(\Delta | M_j, \mathbf{Y}). \quad (2.22)$$

Τότε η βέλτιστη επιλογή είναι η δράση α που ελαχιστοποιεί την αναμενόμενη ζημία:

$$R(\alpha) = E[L(\alpha, \Delta) | \mathbf{Y}]$$

και λόγω της σχέσης:

$$-E \left[\log \left\{ \sum_{j=1}^k p(\Delta | M_j, \mathbf{Y}) p(M_j | \mathbf{Y}) \right\} \right] \leq -E \left[\log \{ p(\Delta | M_j, \mathbf{Y}) \} \right], \quad j = 1, \dots, k \quad (2.23)$$

(Madigan and Raftery, 1994) καταλήγουμε ότι είναι προτιμότερο να επιλέξουμε τον μέσο όρο $\sum_{j=1}^k p(\Delta | M_j, \mathbf{Y}) p(M_j | \mathbf{Y})$ από το να επιλέξουμε οποιοδήποτε μεμονωμένο μοντέλο M .

Έτσι έχουμε μία θεωρητική εξήγηση γιατί κατά μέσο όρο περιμένουμε με την εκτίμηση βάσει κυρτών συνδυασμών μοντέλων να έχουμε καλύτερη πρόβλεψη από ότι με την εκτίμηση βάσει ενός μόνο μοντέλου. Η θεωρητική αυτή εξήγηση επαληθεύεται στην πράξη σε πολλά παραδείγματα (Hoeting et al., 1999; Madigan and Raftery, 1994; Raftery et al., 1995). Να σημειώσουμε εδώ ότι λόγω της δυσκολίας του ακριβούς υπολογισμού του log predictive score σε πολλές περιπτώσεις, αντί

αυτού υπολογίζεται μία προσέγγισή του (*partial predictive score*), όπως μπορεί να δει κανείς στη βιβλιογραφία (Hoeting et al., 1999; Raftery et al., 1995).

Βεβαίως, ο υπολογισμός του log predictive score δεν είναι ο μοναδικός τρόπος για να εξετάσουμε την προβλεπτική ικανότητα. Ο χωρισμός του συνόλου των δεδομένων D σε δύο υποσύνολα γενικώς μας επιτρέπει να εκτιμήσουμε τις ποσότητες που θέλουμε στο υποσύνολο D_T και μέσω των εκτιμήσεων που παίρνουμε να προβλέψουμε τις τιμές που μας ενδιαφέρουν στο υποσύνολο D_P . Αυτό θα κάνουμε και στο παράδειγμα του τελευταίου κεφαλαίου.

2.7 Επιλογή της Κλάσης για την Εφαρμογή της Μεθόδου

Είδαμε μέχρι τώρα πως ένας τρόπος να αντιμετωπίσουμε την αβεβαιότητά μας σχετικά με την επιλογή ενός μοντέλου είναι να χρησιμοποιήσουμε ένα άθροισμα της μορφής (2.4) ή (2.6) θεωρώντας όλα τα πιθανά μοντέλα, αντί να βασίσουμε τις αναλύσεις και τις προβλέψεις μας σε ένα μόνο μοντέλο. Θεωρήσαμε ωστόσο βέβαιο το ενδεχόμενο ότι το βέλτιστο μοντέλο βρίσκεται ανάμεσα σε αυτά τα μοντέλα, αφού θεωρήσαμε το άθροισμα των πιθανοτήτων τους ίσο με τη μονάδα.

Προχωρώντας τον προβληματισμό μας ένα βήμα παραπέρα, κατανοούμε ότι αβεβαιότητα υπάρχει και ως προς το ποιά είναι τα μοντέλα που πρέπει να θεωρήσουμε, ποιά είναι δηλαδή η κλάση $\mathcal{M} = \{M_1, \dots, M_k\}$ που θεωρούμε αρχικά. Η επιλογή αυτή βασίζεται συνήθως στην ήδη υπάρχουσα γνώση που έχουμε για το πρόβλημα που μελετούμε. Έτσι, συχνά θεωρούμε μοντέλα που ξέρουμε πως περιγράφουν καλά αντίστοιχα προβλήματα σε παλαιότερες έρευνες και εν συνεχεία ελέγχουμε αν είναι καλή η προσαρμογή ή όχι. Για παράδειγμα, στην περίπτωση που έχουμε να κάνουμε με ένα πρόβλημα ανάλυσης επιβίωσης, μία συνηθισμένη επιλογή είναι να συμπεριλάβουμε στην κλάση \mathcal{M} μοντέλα αναλογικής διακινδύνευσης.

Στο επόμενο κεφάλαιο θα ασχοληθούμε εκτενώς με την εφαρμογή του Bayesian Model Averaging σε προβλήματα που μπορούν να περιγραφούν ικανοποιητικά από γενικευμένα γραμμικά μοντέλα. Η πιθανότητα ότι τέτοια μοντέλα είναι κατάλληλα θα θεωρείται ίση με τη μονάδα και μέλημά μας θα είναι – με τις

τεχνικές που είδαμε αλλά και με αυτές που θα αναπτύξουμε - να υπολογίσουμε όσο το δυνατόν καλύτερα τις ποσότητες που μας ενδιαφέρουν.

Κεφάλαιο 3

Κυρτοί Συνδυασμοί Γενικευμένων Γραμμικών Μοντέλων

Στο κεφάλαιο αυτό θα εστιάσουμε στην εφαρμογή του Bayesian Model Averaging στα γενικευμένα γραμμικά μοντέλα, βλέποντας πώς εξειδικεύονται οι μέθοδοι που είδαμε στο προηγούμενο κεφάλαιο στην συγκεκριμένη κλάση μοντέλων.

3.1 Εισαγωγή

Ως εισαγωγή, θα παραθέσουμε μία συνοπτική περιγραφή των γενικευμένων γραμμικών μοντέλων, για να καταστεί παράλληλα σαφές ποιές είναι οι ποσότητες που καλούμαστε να εκτιμήσουμε.

Έστω λοιπόν Y μία μεταβλητή της οποίας η κατανομή γνωρίζουμε ότι εξαρτάται από κάποιες μεταβλητές X_j , $j=1, \dots, p$. Καλούμε τη μεταβλητή Y **εξαρτημένη**, ενώ τις μεταβλητές X_j τις καλούμε **επεξηγηματικές** ή **ανεξάρτητες**.

Υπάρχουν πολλοί τρόποι με τους οποίους οι επεξηγηματικές μεταβλητές μπορεί να επηρεάζουν την εξαρτημένη. Ο πιο απλός ίσως από αυτούς είναι ο μέσος της εξαρτημένης μεταβλητής να επηρεάζεται από έναν γραμμικό συνδυασμό των επεξηγηματικών μεταβλητών. Αν επιπλέον υποθέσουμε ότι η Y ακολουθεί κανονική κατανομή, οδηγούμαστε στο πολύ δημοφιλές στις εφαρμογές μοντέλο της **γραμμικής παλινδρόμησης**.

Σε ένα μοντέλο γραμμικής παλινδρόμησης λοιπόν, υποθέτουμε ότι για τις παρατηρήσεις y_i ισχύει :

$$y_i \sim N(x_{i1}\beta_1 + \dots + x_{ip}\beta_p, \sigma^2), \quad i=1, \dots, n, \quad (3.1)$$

όπου x_{i1}, \dots, x_{ip} είναι οι τιμές των p το πλήθος επεξηγηματικών μεταβλητών για την i -παρατήρηση και β_1, \dots, β_p είναι οι συντελεστές παλινδρόμησης που συνδέονται με αυτές τις μεταβλητές (τους οποίους καλούμαστε να εκτιμήσουμε). Ισοδύναμα

μπορούμε να χρησιμοποιήσουμε τη διατύπωση $y_i \sim N(\mu_i, \sigma^2)$ με $\mu_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p$, $i = 1, \dots, n$, αντί της (3.1).

Στην περίπτωση που υποθέτουμε ανεξάρτητες παρατηρήσεις με την ίδια διασπορά, μπορούμε να γράψουμε την εξίσωση (3.1) σε μορφή πινάκων ως εξής:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \quad (3.2)$$

όπου ο \mathbf{X} είναι ο $n \times p$ πίνακας με γραμμές τις παρατηρήσεις x_{i1}, \dots, x_{ip} , $i = 1, \dots, n$, $\boldsymbol{\beta}$ είναι το p -διάστατο διάνυσμα με στοιχεία τους συντελεστές β_1, \dots, β_p και \mathbf{I}_n είναι ο μοναδιαίος $n \times n$ πίνακας.

Μία πιο γενική υπόθεση συνίσταται στο να θεωρήσουμε $\text{Var}(\mathbf{y}) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, ενώ γενικεύοντας ακόμα περισσότερο μπορούμε να υποθέσουμε ότι $\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$, με $\boldsymbol{\Sigma}$ έναν οποιοδήποτε συμμετρικό και θετικά ορισμένο πίνακα. Στα υπόλοιπα θα θεωρούμε το μοντέλο (3.2) που είδαμε προηγουμένως.

Τα μοντέλα γραμμικής παλινδρόμησης μπορούν να επεκταθούν ως εξής: Οι ανεξάρτητες παρατηρήσεις y_i μπορούν να ακολουθούν κάποια κατανομή f της εκθετικής οικογένειας κατανομών με παράμετρο $\boldsymbol{\theta}$, ενώ η παράμετρος αυτή (ή κάποια συνιστώσα της παραμέτρου) μπορεί να συνδέεται μέσω μίας συνάρτησης g με τον γραμμικό συνδυασμό $\eta_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p$. Έχουμε δηλαδή το ακόλουθο σχήμα:

$$f(y_i | \boldsymbol{\theta}_i) = \alpha(y_i) g(\boldsymbol{\theta}_i) \exp\{c(\boldsymbol{\theta}_i)^T d(x_i)\}, \quad (3.3)$$

$$g(\boldsymbol{\theta}_i) = \eta_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p, \quad i = 1, \dots, n. \quad (3.4)$$

Κατά τον τρόπο αυτό προκύπτουν τα λεγόμενα **γενικευμένα γραμμικά μοντέλα**. Η συνάρτηση g καλείται **συνδετική συνάρτηση** [link function] και επιλέγεται συνήθως με τέτοιο τρόπο ώστε να ικανοποιούνται τυχόν περιορισμοί στα θ_i . Επίσης, το η_i καλείται **γραμμικός προσδιορισμός** του μοντέλου [linear predictor] και είναι αυτό που εισάγει στο μοντέλο μας την συνολική επίδραση των επεξηγηματικών μεταβλητών.

Η επιλογή μίας συνδετικής συνάρτησης λοιπόν, πέρα από το ότι μπορεί να προκύψει εξαιτίας της καλύτερης περιγραφής των δεδομένων μας, σε κάποιες περιπτώσεις είναι επιβεβλημένη. Για παράδειγμα, στην περίπτωση που τα y_i αφορούν απαριθμήσεις γεγονότων, η πιο συνηθισμένη επιλογή είναι η χρήση της κατανομής Poisson για την περιγραφή τους, δηλαδή $Y_i \sim \text{Poisson}(\lambda_i)$. Τότε όμως δεν μπορούμε να θέσουμε $\lambda_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p$, αφού για τα λ_i απαιτούμε $\lambda_i > 0$ ενώ ο περιορισμός αυτός δεν υπάρχει στο δεξί μέλος της τελευταίας εξίσωσης. Έτσι, πρέπει να χρησιμοποιήσουμε μία συνάρτηση g που να απεικονίζει το σύνολο \mathbb{R}_+ στο \mathbb{R} . Μία τέτοια συνάρτηση που χρησιμοποιείται συχνά είναι η $g(\lambda) = \log(\lambda)$. Ανάλογα, αν τα y_i αφορούν δίτιμες μεταβλητές και χρησιμοποιήσουμε κατανομή Bernoulli με παράμετρο p_i για την περιγραφή τους, τότε θα χρειαζόμαστε μία συνδετική συνάρτηση g με πεδίο ορισμού το διάστημα $[0,1]$ και πεδίο τιμών το \mathbb{R} . Μία συνηθισμένη επιλογή είναι τότε η $g(p_i) = \text{logit}(p_i) = \log[p_i / (1 - p_i)]$. Και στις δύο περιπτώσεις, βέβαια, υπάρχουν και άλλες συναρτήσεις που πληρούν τις προϋποθέσεις μας και η εύρεση της καταλληλότερης τέτοιας συνάρτησης αποτελεί έναν από τους στόχους μας.

Να σημειώσουμε εδώ πως το μοντέλο γραμμικής παλινδρόμησης που περιγράφεται από τις σχέσεις (3.1), (3.2) αποτελεί ειδική περίπτωση του γενικευμένου γραμμικού μοντέλου και προκύπτει θέτοντας την κανονική κατανομή στη σχέση (3.3), δηλαδή $f(y_i | \mu) = N(y_i | \mu, \sigma^2)$, και την ταυτοτική συνάρτηση στη σχέση (3.4), δηλαδή $g(\mu_i) = \mu_i$.

Στα πλαίσια της Μπεϋζιανής εκτίμησης των παραμέτρων β_i , σε πρώτο στάδιο πρέπει να καθορίσουμε τις prior κατανομές τους. Θα δούμε αυτό και τα άλλα ζητήματα που προκύπτουν, αφού πρώτα περιγράψουμε πώς εξειδικεύεται η επιλογή μοντέλου και η μέθοδος του Bayesian Model Averaging στα γενικευμένα γραμμικά μοντέλα.

3.2 Η Μέθοδος των Κυρτών Συνδυασμών στα Γενικευμένα Γραμμικά Μοντέλα

Στην ειδική περίπτωση των μοντέλων γραμμικής παλινδρόμησης, έχοντας μία εξαρτημένη μεταβλητή Y και ένα σύνολο p το πλήθος επεξηγηματικών μεταβλητών $\{X_1, \dots, X_p\}$, το πρόβλημα επιλογής μοντέλου συνίσταται στο να επιλέξουμε το υποσύνολο των επεξηγηματικών μεταβλητών που θα χρησιμοποιήσουμε στην ανάλυσή μας. Έχουμε λοιπόν να κάνουμε με ένα **πρόβλημα επιλογής μεταβλητών**.

Το πρόβλημα αυτό έχει ιδιαίτερη σημασία όταν ο αριθμός p είναι μεγάλος και υποψιαζόμαστε ότι πολλές από τις επεξηγηματικές μεταβλητές έχουν μικρή ή και μηδαμινή επίδραση στην Y . Για $y \sim N(X\beta, \sigma^2 I_n)$, λοιπόν, μας ενδιαφέρει να δούμε αν κάποιοι από τους συντελεστές β_i είναι αρκετά κοντά στο μηδέν, ώστε να μπορούμε να τους αγνοήσουμε. Γενικότερα, το πρόβλημα επιλογής μεταβλητών αναφέρεται σε κάθε πρόβλημα επιλογής μοντέλου όπου κάθε ένα από τα θεωρούμενα μοντέλα αντιστοιχεί σε ένα συγκεκριμένο σύνολο επεξηγηματικών μεταβλητών. Ως εκ τούτου, ειδική περίπτωσή του αποτελεί και κάθε περίπτωση γενικευμένων γραμμικών μοντέλων όπου η συνδετική συνάρτηση έχει καθοριστεί.

Θεωρούμε τώρα ένα πρόβλημα επιλογής μεταβλητών με p επεξηγηματικές μεταβλητές. Τότε το πλήθος των πιθανών μοντέλων είναι 2^p . Χρησιμοποιώντας τον συμβολισμό του προηγούμενου κεφαλαίου, η κλάση \mathcal{M} των υποψήφιων μοντέλων είναι η $\mathcal{M} = \{M_1, \dots, M_k\}$, όπου $k = 2^p$. Είναι ιδιαίτερα βολικό να αντιστοιχίσουμε σε καθένα από τα k αυτά μοντέλα M_j ένα διάνυσμα $\gamma_j = (\gamma_j^1, \dots, \gamma_j^p)$, όπου $\gamma_j^i = 0$ ή 1 ανάλογα με το αν η μεταβλητή X_i περιλαμβάνεται στο μοντέλο ή όχι, ή ισοδύναμα αν η τιμή του συντελεστή β_i είναι 0 ή όχι. Το πλήθος των μεταβλητών που περιλαμβάνει το μοντέλο M_j είναι τότε $q_{\gamma_j} = \sum_{i=1}^p \gamma_j^i$.

Εστιάζοντας στην ειδική περίπτωση των μοντέλων γραμμικής παλινδρόμησης, τα συμπεράσματά μας βασίζονται στο ακόλουθο ιεραρχικό μοντέλο:

$$\begin{aligned}
\mathbf{Y} | \boldsymbol{\beta}, \sigma^2, M_j &\sim N(X\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \\
\boldsymbol{\beta} | \sigma^2, M_j &\sim p(\boldsymbol{\beta} | M_j, \sigma^2) \\
\sigma^2 | M_j &\sim p(\sigma^2 | M_j) \\
M_j &\sim p(M_j),
\end{aligned} \tag{3.5}$$

όπου οι συναρτήσεις $p(\boldsymbol{\beta} | M_j, \sigma^2)$, $p(\sigma^2 | M_j)$, $p(M_j)$ αφορούν prior κατανομές παραμέτρων και μοντέλων (στην τελευταία περίπτωση) που πρέπει να καθοριστούν. Θα δούμε συνήθεις επιλογές για αυτό στην επόμενη ενότητα.

Στην ευρύτερη περίπτωση των γενικευμένων γραμμικών μοντέλων, πέρα από την επιλογή των ανεξάρτητων μεταβλητών καλούμαστε να επιλέξουμε και τη συνδετική συνάρτηση. Κάθε διαφορετικός συνδυασμός επιλογών οδηγεί σε διαφορετικό μοντέλο, με συνέπεια το πλήθος των πιθανών μοντέλων να είναι ακόμη μεγαλύτερο (πολλαπλάσιο του 2^p).

Οι συνήθεις τεχνικές στην επιλογή μεταβλητών, όπως η προς τα εμπρός επιλογή [forward selection], η διαδοχική αφαίρεση ή απαλοιφή [backward elimination], η κατά βήματα εμπρός πίσω επιλογή [stepwise selection] και τα κριτήρια C_p και R^2 , οδηγούν συχνά στην επιλογή διαφορετικού μοντέλου. Γίνεται έτσι εμφανές το πρόβλημα της αβεβαιότητας κατά την επιλογή μοντέλου που περιγράψαμε στο προηγούμενο κεφάλαιο και το οποίο θα επιχειρήσουμε να αντιμετωπίσουμε με τις τεχνικές των κυρτών συνδυασμών μοντέλων, όπως αυτές εξειδικεύονται στην περίπτωση των γραμμικών μοντέλων.

Χάριν ευκολίας θα ξαναπαραθέσουμε εδώ τον βασικό τύπο:

$$p(\Delta | \mathbf{Y}) = \sum_{j=1}^k p(\Delta | M_j, \mathbf{Y}) p(M_j | \mathbf{Y}), \tag{3.6}$$

που μας δίνει την ύστερη κατανομή μίας άγνωστης ποσότητας Δ που μας ενδιαφέρει, ενώ η αντίστοιχη ύστερη μέση τιμή είναι η:

$$E(\Delta | \mathbf{Y}) = \sum_{j=1}^k E(\Delta | M_j, \mathbf{Y}) p(M_j | \mathbf{Y}). \tag{3.7}$$

Θα εξετάσουμε τώρα τα διάφορα ζητήματα που προκύπτουν κατά τον υπολογισμό αυτών των ποσοτήτων, ξεκινώντας από τον καθορισμό των πρότερων πιθανοτήτων των μοντέλων και των πρότερων κατανομών των παραμέτρων κάθε μοντέλου.

3.3 Επιλογή των Πρότερων Πιθανοτήτων των Μοντέλων

Θα ξεκινήσουμε αυτή την ενότητα παραδίδοντας κάποιες συνηθισμένες επιλογές όσον αφορά τις πρότερες πιθανότητες των μοντέλων. Στην περίπτωση που έχουμε να κάνουμε με ένα πρόβλημα επιλογής μεταβλητών (π.χ. μοντέλα γραμμικής παλινδρόμησης ή γενικευμένα γραμμικά μοντέλα με συγκεκριμένη συνδετική συνάρτηση), μία πολύ διαδεδομένη επιλογή είναι να υποθέσουμε ότι τα διανύσματα γ_j που ορίσαμε προηγουμένως είναι ανεξάρτητα a priori και να καθορίσουμε την εξής prior κατανομή για αυτά:

$$p(\gamma_j) = \prod_{i=1}^p \omega_i^{\gamma_j^i} (1 - \omega_i)^{1 - \gamma_j^i}, \quad (3.8)$$

όπου η υπερπαραμέτρος ω_i αντιστοιχεί στην prior πιθανότητα ότι η μεταβλητή X_i περιλαμβάνεται στο μοντέλο, ανεξαρτήτως της παρουσίας ή απουσίας των άλλων μεταβλητών, δηλαδή $\omega_i = \Pr[\gamma_j^i = 1] = 1 - \Pr[\gamma_j^i = 0]$. Έτσι, οι πιθανότητες των μοντέλων είναι εύκολο να καθοριστούν, αρκεί να έχουν καθοριστεί οι πιθανότητες ω_i .

Η επιλογή των πιθανοτήτων αυτών πρέπει να βασίζεται στην πληροφορία που έχουμε πριν τη δειγματοληψία. Για παράδειγμα, αν είμαστε πεπεισμένοι ότι μία μεταβλητή X_i πρέπει να περιλαμβάνεται σε όλα τα μοντέλα, είναι λογικό να θέσουμε $\omega_i = 1$. Μία ενδιαφέρουσα απλοποίηση προκύπτει αν θέσουμε $\omega_i \equiv \omega \in (0,1)$, οπότε παίρνουμε:

$$p(\gamma_j) = \prod_{i=1}^p \omega^{q_{\gamma_j^i}} (1 - \omega)^{p - q_{\gamma_j^i}}. \quad (3.9)$$

Η ιδιαίτερα δημοφιλής επιλογή της ομοιόμορφης prior κατανομής προκύπτει τότε θέτοντας $\omega = 1/2$, οπότε παίρνουμε:

$$p(\gamma_j) = \frac{1}{2^p} = \frac{1}{k} \quad \forall j \in \{1, \dots, k\}, \quad (3.10)$$

όπου 2^p το πλήθος των πιθανών μοντέλων που θεωρούμε.

Μία εναλλακτική επιλογή είναι να καθορίσουμε μία prior κατανομή για το ω . Για παράδειγμα, θέτοντας $\omega \sim \text{Beta}(\alpha, \beta)$ προκύπτει:

$$p(\gamma_j) = \frac{B(\alpha + q_{\gamma_j}, \beta + p - q_{\gamma_j})}{B(\alpha, \beta)}, \quad (3.11)$$

όπου $B(\bullet)$ η συνάρτηση Βήτα.

Ακόμη γενικότερα, θα μπορούσαμε να τοποθετήσουμε μία prior πιθανότητα $h(q_{\gamma_j})$ για το μέγεθος κάθε μοντέλου και να θέσουμε:

$$p(\gamma_j) = h(q_{\gamma_j}) \binom{p}{q_{\gamma_j}}^{-1}, \quad (3.12)$$

της οποίας η (3.11) αποτελεί ειδική περίπτωση.

Πρέπει να σημειώσουμε ότι αν και η επιλογή μίας ομοιόμορφης prior κατανομής για τα μοντέλα απλοποιεί πολύ τη διαδικασία, εντούτοις αντιστοιχεί σε μία διωνυμική $(q_{\gamma_j}, 0.5)$ prior κατανομή για το μέγεθος του μοντέλου και μία εκ των προτέρων πεποίθηση ότι οι μισές από τις επεξηγηματικές μεταβλητές θα περιληφθούν στο μοντέλο. Αυτό ενδέχεται να μην είναι ιδιαίτερα ρεαλιστικό, ειδικά αν το q είναι πολύ μεγάλο. Ένας τρόπος να κάνουμε λιγότερο πιθανή την επιλογή των μοντέλων με μεγάλο q_{γ_j} είναι να θέσουμε $\omega_i < 1/2$ για κάθε i στην (3.8) ή να θέσουμε μικρές πιθανότητες για τα μοντέλα με μεγάλο μέγεθος στην (3.12).

Επιπλέον, να αναφέρουμε ότι η υπόθεση για ανεξαρτησία των prior κατανομών δεν ενδείκνυται σε περιπτώσεις που έχουμε ενδείξεις ότι υπάρχουν συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών. Ο αναγνώστης μπορεί να εξετάσει την βιβλιογραφία για μία προσέγγιση σε τέτοιες περιπτώσεις (Chipman et al., 2001).

Τέλος, να τονίσουμε ότι στην περίπτωση που θέλουμε να χρησιμοποιήσουμε μία ομοιόμορφη prior κατανομή για γενικευμένα γραμμικά μοντέλα, τότε το πλήθος των πιθανών μοντέλων που εισάγεται στον παρονομαστή στη σχέση (3.10) μπορεί να είναι μεγαλύτερο από 2^p , αν θεωρήσουμε παραπάνω από μία πιθανές συνδετικές συναρτήσεις.

3.4 Επιλογή των Πρότερων Κατανομών των Παραμέτρων

Έχοντας καθορίσει τις prior πιθανότητες για τα πιθανά μοντέλα μας, καλούμαστε να καθορίσουμε και τις prior κατανομές για τις παραμέτρους των μοντέλων. Θα εξετάσουμε αρχικά την περίπτωση των μοντέλων γραμμικής παλινδρόμησης, όπου ο καθορισμός είναι πιο απλός.

Θεωρούμε λοιπόν ένα μοντέλο της μορφής $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, το οποίο για να ολοκληρωθεί χρειάζεται μία prior κατανομή για τη διανυσματική παράμετρο $(\boldsymbol{\beta}, \sigma^2)$. Σε μοντέλα αυτής της μορφής μπορούμε να χρησιμοποιήσουμε μία συζυγή prior κατανομή με την εξής επιλογή:

$$\boldsymbol{\beta} | \sigma^2 \sim N(\mathbf{b}_0, \sigma^2 \mathbf{B}_0), \quad (3.13)$$

$$\sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right), \quad (3.14)$$

όπου IG η κατανομή *Inverse Gamma*. Η (3.14) γράφεται ισοδύναμα ως $\frac{n_0 S_0}{\sigma^2} \sim \chi_{n_0}^2$.

Εναλλακτικά, μπορούμε αντί της παραμέτρου σ^2 να χρησιμοποιήσουμε την ακρίβεια $\phi = 1/\sigma^2$, οπότε προκύπτει:

$$\boldsymbol{\beta} | \phi \sim N\left(\mathbf{b}_0, \frac{\mathbf{B}_0}{\phi}\right), \quad (3.13^*)$$

$$\phi \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right). \quad (3.14^*)$$

Τότε για την από κοινού prior κατανομή των $\boldsymbol{\beta}, \phi$ έχουμε:

$$\begin{aligned}
p(\boldsymbol{\beta}, \phi) &\propto p(\boldsymbol{\beta} | \phi) p(\phi) \\
&\propto \phi^{d/2} \exp\left\{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_0)^T B_0^{-1}(\boldsymbol{\beta} - \mathbf{b}_0)\right\} \phi^{n_0/2-1} \exp\left\{-\frac{\phi}{2}n_0 S_0\right\} \\
&\propto \phi^{(n_0+d)/2-1} \exp\left\{-\frac{\phi}{2}[n_0 S_0 + (\boldsymbol{\beta} - \mathbf{b}_0)^T B_0^{-1}(\boldsymbol{\beta} - \mathbf{b}_0)]\right\}. \tag{3.15}
\end{aligned}$$

Επιπλέον, για ένα συγκεκριμένο δείγμα \mathbf{y} , η πιθανοφάνεια είναι:

$$\begin{aligned}
L(\mathbf{y}; \boldsymbol{\beta}, \phi) &\propto \phi^{-n/2} \exp\left\{-\frac{\phi}{2}(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})\right\} \\
&= \phi^{-n/2} \exp\left\{-\frac{\phi}{2}[Q(\boldsymbol{\beta}) + S_e]\right\}, \tag{3.16}
\end{aligned}$$

όπου $Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\beta})^T X^T X (\boldsymbol{\beta} - \boldsymbol{\beta})$, $\boldsymbol{\beta} = (X^T X)^{-1} X^T \mathbf{y}$ και $S_e = (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) = \sum_{i=1}^n e_i^2$, με $e_i = \mathbf{y}_i - \mu_i$ και $\mu_i = x_{i1}\beta_1 + \dots + x_{id}\beta_d$, $i = 1, \dots, n$, ενώ με $\boldsymbol{\beta}$ συμβολίζουμε την τιμή που μεγιστοποιεί την συνάρτηση $L(\mathbf{y}; \boldsymbol{\beta}, \phi)$, βλέποντάς την ως συνάρτηση του $\boldsymbol{\beta}$.

Συνδυάζοντας την prior κατανομή (3.15) με την πιθανοφάνεια (3.16) και εφαρμόζοντας το Θεώρημα του Bayes, έχουμε την εξής έκφραση για την από κοινού posterior κατανομή των $\boldsymbol{\beta}, \phi$:

$$\begin{aligned}
p(\boldsymbol{\beta}, \phi | \mathbf{y}) &\propto p(\boldsymbol{\beta}, \phi) L(\mathbf{y}; \boldsymbol{\beta}, \phi) \\
&\propto \phi^{(n_0+n+d)/2-1} \exp\left\{-\frac{\phi}{2}[n_0 S_0 + S_e + (\boldsymbol{\beta} - \mathbf{b}_0)^T B_0^{-1}(\boldsymbol{\beta} - \mathbf{b}_0) + Q(\boldsymbol{\beta})]\right\} \\
&\propto \phi^{d/2} \exp\left\{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_1)^T B_1^{-1}(\boldsymbol{\beta} - \mathbf{b}_1)\right\} \times \phi^{n_1/2-1} \exp\left\{-\frac{\phi}{2}n_1 S_1\right\}, \tag{3.17}
\end{aligned}$$

όπου $n_1 = n_0 + n$, $n_1 S_1 = n_0 S_0 + (\mathbf{y} - X\mathbf{b}_1)^T \mathbf{y} + (\mathbf{b}_0 - \mathbf{b}_1)^T B_0^{-1} \mathbf{b}_0$, $\mathbf{b}_1 = B_1(B_0^{-1} \mathbf{b}_0 + X^T \mathbf{y})$ και $B_1^{-1} = B_0^{-1} + X^T X$.

Δηλαδή, η ύστερη από κοινού κατανομή των $\boldsymbol{\beta}, \phi$ προκύπτει με αναπαραμετροποίηση της πρότερης κατανομής τους και συγκεκριμένα είναι η

κατανομή $\text{NormalGamma}(\mathbf{b}_1, B_1, n_1, S_1)$. Στην περίπτωση που θέλουμε το αποτέλεσμα συναρτήσει της διασποράς σ^2 , έχουμε:

$$\boldsymbol{\beta} | \sigma^2, \mathbf{y} \sim N(\mathbf{b}_1, \sigma^2 B_1) \quad \text{και} \quad (3.18)$$

$$\sigma^2 | \mathbf{y} \sim \text{IG}\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right). \quad (3.19)$$

Για να είναι βέβαια πλήρης ο καθορισμός των prior κατανομών στις (3.13) και (3.14), θα πρέπει να δώσουμε τιμές στις υπερπαραμέτρους \mathbf{b}_0 , B_0 , n_0 και S_0 . Συνήθως επιλέγουμε να δίνουμε τιμές που οδηγούν σε μη πληροφοριακές prior κατανομές.

Αναφορικά με τις υπερπαραμέτρους n_0 και S_0 , μία παρατήρηση που μας βοηθάει στην επιλογή είναι ότι η (3.14) ισοδυναμεί με την πληροφορία που θα είχαμε για την παράμετρο σ^2 , αν είχαμε στη διάθεσή μας ένα δείγμα n_0 ανεξάρτητων παρατηρήσεων που ακολουθούν $N(0, S_0)$ κατανομή. Χρησιμοποιώντας τα δεδομένα μας, μία επιλογή είναι να θέσουμε το n_0 μικρό (π.χ. μέχρι 3) και το S_0 ίσο με τη δειγματική διασπορά s^2 του Y , η οποία μπορεί να θεωρηθεί ως ένα άνω φράγμα για το σ^2 .

Κοιτώντας τώρα τις υπερπαραμέτρους \mathbf{b}_0 και B_0 , μία συνηθισμένη επιλογή για τον μέσο \mathbf{b}_0 είναι να τον θέσουμε ίσο με $\mathbf{0}$, ώστε να μην ευνοούμε ούτε τις θετικές ούτε τις αρνητικές τιμές. Για τον πίνακα B_0 , μία επιλογή που απλοποιεί την κατάσταση είναι να θέσουμε $B_0 = gV$, όπου g είναι ένας αριθμός και V είναι ένας πίνακας προκαθορισμένης μορφής. Επιλέγοντας $V = (X^T X)^{-1}$ οδηγούμαστε στην δημοφιλή επιλογή $B_0 = g \cdot (X^T X)^{-1}$, η οποία είναι γνωστή ως *g-prior του Zellner*. Τότε το πρόβλημά μας είναι μόνο ο καθορισμός της τιμής g , η οποία πρέπει να είναι αρκετά μεγάλη για να μην υπάρχει πληροφορία στην prior κατανομή, αλλά όχι ιδιαίτερα υψηλή, γιατί κάτι τέτοιο θα ευνοεί τελικά το μοντέλο χωρίς καμία επεξηγηματική μεταβλητή (Chipman et al., 2001). Συγκεκριμένες τιμές για το g , καθώς και άλλες επιλογές για τις υπερπαραμέτρους υπάρχουν στη βιβλιογραφία (Chipman et al., 2001; Fernandez et al., 2001; Raftery et al., 1997).

Συνοψίζοντας, στην περίπτωση των μοντέλων γραμμικής παλινδρόμησης έχουμε τη δυνατότητα να υπολογίσουμε αναλυτικά την ύστερη κατανομή των παραμέτρων, χρησιμοποιώντας συζυγείς πρότερες κατανομές. Δυστυχώς, δεν μπορεί να γίνει το ίδιο στην περίπτωση των γενικευμένων γραμμικών μοντέλων, με συνέπεια να είναι αναγκαστική η χρήση προσεγγίσεων και τεχνικών MCMC. Χρησιμοποιώντας τέτοιες τεχνικές, μία συχνή επιλογή είναι να θεωρούμε μη πληροφοριακές πρότερες κατανομές για την διανυσματική παράμετρο β , όπως για παράδειγμα κανονική κατανομή με μέση τιμή το $\mathbf{0}$ και αρκετά μεγάλη διασπορά.

3.5 Υπολογισμός των Ύστερων Πιθανοτήτων των Μοντέλων

Έχοντας καθορίσει τις prior κατανομές για τα πιθανά μοντέλα και τις παραμέτρους τους, καλούμαστε να υπολογίσουμε τις posterior πιθανότητες των μοντέλων.

Στην περίπτωση των μοντέλων γραμμικών παλινδρόμησης, είναι δυνατόν να υπολογίσουμε αναλυτικά αυτές τις πιθανότητες, εφόσον έχουμε χρησιμοποιήσει τις συζυγείς prior κατανομές που παραθέσαμε προηγουμένως. Συγκεκριμένα, για κάθε μοντέλο M_j μπορούμε να υπολογίσουμε την περιθώρια κατανομή των δεδομένων $p(\mathbf{Y} | M_j)$ μέσω της σχέσης:

$$\begin{aligned} p(\mathbf{Y} | M_j) &= \iint p(\mathbf{Y} | \beta, \sigma^2, M_j) p(\beta, \sigma^2 | M_j) d\beta d\sigma^2 \\ &= \iint p(\mathbf{Y} | \beta, \sigma^2, M_j) p(\beta | \sigma^2, M_j) p(\sigma^2 | M_j) d\beta d\sigma^2, \quad j=1, \dots, k, \end{aligned} \quad (3.20)$$

η οποία αποτελεί εφαρμογή της σχέσης (2.2) στην περίπτωσή μας. Ακολούθως, μπορούμε να υπολογίσουμε την ύστερη πιθανότητα κάθε μοντέλου από τη σχέση (2.1):

$$p(M_j | \mathbf{Y}) = \frac{p(\mathbf{Y} | M_j) p(M_j)}{\sum_{i=1}^k p(\mathbf{Y} | M_i) p(M_i)}, \quad j=1, \dots, k.$$

Ωστόσο, ο υπολογισμός αυτός δεν είναι εφικτός στην ευρύτερη περίπτωση των γενικευμένων γραμμικών μοντέλων, οπότε και αναγκαζόμαστε να καταφεύγουμε

σε προσεγγιστικές λύσεις. Μία τέτοια προσεγγιστική λύση που χρησιμοποιείται συχνά, λόγω της εύκολης υλοποίησής της, είναι αυτή που προκύπτει με τη χρήση του κριτηρίου BIC που είδαμε στο προηγούμενο κεφάλαιο. Έχοντας λοιπόν ένα μοντέλο M_j και θεωρώντας την τιμή:

$$\text{BIC}(M_j) = -2\log(\text{maximizedlikelihood} | M_j) + q_j \log n, \quad (3.21)$$

όπου n είναι το μέγεθος του δείγματος και q_j το πλήθος των μεταβλητών του μοντέλου, έχουμε ότι:

$$-0.5\text{BIC}(M_j) = \log(\text{maximizedlikelihood} | M_j) - (q_j / 2) \log n \approx \log p(\mathbf{Y} | M_j),$$

και συνεπώς μπορούμε να πάρουμε προσεγγιστικές ύστερες πιθανότητες για τα μοντέλα M_j , μέσω της σχέσης:

$$p(M_j | \mathbf{Y}) \approx \frac{p(M_j) e^{-0.5\text{BIC}(M_j)}}{\sum_{i=1}^k p(M_i) e^{-0.5\text{BIC}(M_i)}}, \quad (3.22)$$

όπου $p(M_j)$ η prior πιθανότητα του μοντέλου, η οποία ταυτίζεται με την πιθανότητα $p(\gamma_j)$ που είδαμε στις σχέσεις (3.8) έως (3.12). Η posterior πιθανότητα $p(M_j | \mathbf{Y})$ θα συμβολίζεται κατά αντιστοιχία στην περίπτωση ενός προβλήματος επιλογής μεταβλητών και ως $p(\gamma_j | \mathbf{Y})$.

Η προσέγγιση (3.22) απλοποιεί πολύ τους υπολογισμούς μας, ενώ έχει υπάρξει ιδιαίτερα αποτελεσματική σε πολλές εφαρμογές. Επίσης, χρησιμοποιείται συχνά και στην περίπτωση των μοντέλων γραμμικής παλινδρόμησης, επειδή με τη χρήση της αποφεύγουμε να καθορίσουμε τις τιμές των υπερπαραμέτρων. Για τους λόγους αυτούς, στα περισσότερα στατιστικά πακέτα υπάρχουν εντολές που υπολογίζουν κατευθείαν την τιμή BIC, όπως θα δούμε και στη συνέχεια.

3.6 Εφαρμογή της Μεθόδου

Συνοψίζοντας, οι posterior πιθανότητες των μοντέλων μπορούν να υπολογιστούν είτε αναλυτικά (στην περίπτωση των μοντέλων γραμμικής

παλινδρόμησης και με χρήση συζυγών πρότερων κατανομών), είτε προσεγγιστικά στην γενικότερη περίπτωση των γενικευμένων γραμμικών μοντέλων. Ακολουθώντας, η ποσότητα $p(\Delta | M_j, \mathbf{Y})$ του αθροίσματος (3.6) μπορεί να υπολογιστεί, ενδεχομένως προσεγγιστικά χρησιμοποιώντας την εκτίμηση μέγιστης πιθανοφάνειας για την παράμετρο του κάθε μοντέλου. Έτσι, στην περίπτωση που το πλήθος των θεωρούμενων μοντέλων δεν είναι ιδιαίτερα μεγάλο, είμαστε σε θέση να υπολογίσουμε το άθροισμα (3.6) και την εκτίμηση (3.7) για την άγνωστη ποσότητα που μας ενδιαφέρει.

Πολλές φορές όμως το πλήθος των πιθανών μοντέλων που θεωρούμε είναι ιδιαίτερα μεγάλο, με αποτέλεσμα να μην καθίσταται πρακτικό να υπολογίσουμε για κάθε μοντέλο χωριστά τις ποσότητες που εμπεριέχονται στο άθροισμα (3.6). Για παράδειγμα, αν θεωρήσουμε 20 πιθανές επεξηγηματικές μεταβλητές σε ένα πρόβλημα, προκύπτουν περισσότερα από 1.000.000 πιθανά μοντέλα, ενώ αν θεωρήσουμε και περισσότερες από μία πιθανές συνδετικές συναρτήσεις ο αριθμός αυτός πολλαπλασιάζεται.

Στο προηγούμενο κεφάλαιο, είδαμε την τεχνική του «παράθυρου του Occam» η οποία μπορεί να ελαττώσει αισθητά το πλήθος των μοντέλων μας. Η τεχνική αυτή μπορεί να εφαρμοστεί εναλλακτικά χρησιμοποιώντας τις προσεγγίσεις BIC για κάθε μοντέλο στις σχέσεις (2.12), (2.13), όταν δεν μπορούμε να υπολογίσουμε αναλυτικά τις posterior πιθανότητες των μοντέλων. Οι αλγόριθμοι «branch and bounds» και «leaps and bounds» μπορούν να μας δώσουν έναν ακόμη τρόπο για να αναζητήσουμε τα καλύτερα μοντέλα χωρίς να εξετάσουμε όλες τις πιθανές επιλογές, γλιτώνοντας έτσι πολλούς υπολογισμούς. Πληροφορίες για το πώς εφαρμόζονται οι αλγόριθμοι αυτοί σε προβλήματα επιλογής μεταβλητών μπορούν να αναζητηθούν στη βιβλιογραφία (Furnival and Wilson, 1974; Hand, 1981). Εφαρμόζοντας τις τεχνικές αυτές, όπως είδαμε, αθροίζουμε τελικά σε ένα υποσύνολο του αρχικού χώρου των μοντέλων \mathcal{M} και έτσι οι υπολογισμοί μας απλοποιούνται κατά πολύ.

Οι παραπάνω στρατηγικές αναζήτησης είναι ντετερμινιστικές και χρησιμοποιούνται σε διάφορα προγράμματα, όπως στις συναρτήσεις `bicreg` και `bic.glm` της R που θα χρησιμοποιήσουμε στην συνέχεια. Για πολύ μεγάλο αριθμό μοντέλων όμως (>30 μεταβλητών), οι στρατηγικές αυτές δεν είναι συμφέρουσες από υπολογιστικής πλευράς, ενώ παράλληλα δεν μπορούν να διερευνήσουν το

μεγαλύτερο τμήμα του χώρου των μοντέλων. Κατά συνέπεια σε τέτοιες περιπτώσεις είναι απαραίτητο να καταφεύγουμε στη χρήση τεχνικών MCMC.

Στην περίπτωση που έχουμε να κάνουμε με ένα πρόβλημα επιλογής μεταβλητών, σκοπός μας είναι να προσομοιώσουμε μία ακολουθία:

$$\mathcal{Y}_{(1)}, \mathcal{Y}_{(2)}, \dots, \quad (3.23)$$

(όπου κάθε διάνυσμα $\mathcal{Y}_{(i)}$ αντιστοιχεί σε μία συγκεκριμένη επιλογή μεταβλητών) που να συγκλίνει κατά κατανομή στην $p(\mathcal{Y} | \mathbf{Y})$. Έτσι είναι δυνατόν, ακόμα και στην περίπτωση που το δείγμα μας είναι μικρότερο από το πλήθος των πιθανών μοντέλων, να εντοπίζουμε τα μοντέλα που εμφανίζονται συχνότερα στην ακολουθία. Οι προσομοιωμένες τιμές μπορούν να χρησιμοποιηθούν για την εκτίμηση μίας άγνωστης ποσότητας Δ στα πλαίσια του Bayesian Model Averaging με τους τρόπους που είδαμε στο προηγούμενο κεφάλαιο.

Σε αυτό είδαμε και μία γενική περιγραφή της μεθόδου **MC³**. Θα δούμε τώρα πιο αναλυτικά πώς εφαρμόζεται η μέθοδος αυτή σε προβλήματα επιλογής μεταβλητών. Ο αλγόριθμος Metropolis-Hastings μπορεί να γραφεί για το πρόβλημά μας ως εξής:

1. Προσομοίωσε ένα υποψήφιο \mathcal{Y}^* από τον πίνακα πιθανοτήτων μετάβασης $q(\mathcal{Y}^* | \mathcal{Y}_{(j)})$.
2. Θέσε $\mathcal{Y}_{(j+1)} = \mathcal{Y}^*$ με πιθανότητα $\alpha(\mathcal{Y}^* | \mathcal{Y}_{(j)}) = \min \left\{ \frac{q(\mathcal{Y}_{(j)} | \mathcal{Y}^*) p(\mathcal{Y}^* | \mathbf{Y})}{q(\mathcal{Y}^* | \mathcal{Y}_{(j)}) p(\mathcal{Y}_{(j)} | \mathbf{Y})}, 1 \right\}$,
αλλιώς θέσε $\mathcal{Y}_{(j+1)} = \mathcal{Y}_{(j)}$.

Όπως ήδη αναφέραμε, συχνή είναι η χρήση συμμετρικών q για τα οποία $q(\mathcal{Y}_{(i)} | \mathcal{Y}_{(j)}) = q(\mathcal{Y}_{(j)} | \mathcal{Y}_{(i)})$, οπότε η πιθανότητα αποδοχής γίνεται

$$\alpha(\mathcal{Y}^* | \mathcal{Y}_{(j)}) = \min \left\{ \frac{p(\mathcal{Y}^* | \mathbf{Y})}{p(\mathcal{Y}_{(j)} | \mathbf{Y})}, 1 \right\}.$$

Τίθεται τώρα το ερώτημα πώς θα ορίσουμε τη γειτονιά κάθε διανύσματος \mathcal{Y}_j ή ισοδύναμα κάθε μοντέλου M_j . Μία συχνή επιλογή είναι να ορίζουμε ως γειτονιά

ενός μοντέλου M_j το σύνολο όλων των μοντέλων που έχουν μία μεταβλητή λιγότερη ή μία μεταβλητή περισσότερη, ή ισοδύναμα ως γειτονιά κάθε διανύσματος γ_j όλα τα διανύσματα που έχουν διαφορετική τιμή σε μία μόνο θέση.

Όπως έχουμε πει, ο πίνακας q ορίζεται τότε ώστε να δίνει μηδενική πιθανότητα μετάβασης για κάθε μοντέλο που δεν ανήκει στη γειτονιά του γ_j και μη μηδενική πιθανότητα για κάθε μοντέλο που ανήκει στη συγκεκριμένη γειτονιά. Μία απλή επιλογή για έναν τέτοιο συμμετρικό πίνακα q είναι να ορίσουμε:

$$q(\gamma_{(i)} | \gamma_{(j)}) = \frac{1}{p} \text{ αν } \sum_{l=1}^p |\gamma_{(i)}^l - \gamma_{(j)}^l| = 1 \text{ και } 0 \text{ διαφορετικά.} \quad (3.24)$$

Με την παραπάνω επιλογή, ο αλγόριθμός μας μπορεί να διαμορφωθεί ως εξής:

1. Προσομοίωσε ένα υποψήφιο γ^* αλλάζοντας κατά τυχαίο τρόπο ένα στοιχείο του $\gamma_{(j)}$ με πιθανότητα $1/p$.
2. Θέσε $\gamma_{(j+1)} = \gamma^*$ με πιθανότητα $\alpha(\gamma^* | \gamma_{(j)}) = \min \left\{ \frac{p(\gamma^* | \mathbf{Y})}{p(\gamma_{(j)} | \mathbf{Y})}, 1 \right\}$, αλλιώς θέσε $\gamma_{(j+1)} = \gamma_{(j)}$.

Ο πυρήνας μετάβασης που ορίζεται από την (3.24) είναι ειδική περίπτωση της ακόλουθης κλάσης συμμετρικών πινάκων:

$$q(\gamma_{(i)} | \gamma_{(j)}) = q_d \text{ αν } \sum_{l=1}^p |\gamma_{(i)}^l - \gamma_{(j)}^l| = d, \quad (3.25)$$

οι οποίοι οδηγούν σε αλγορίθμους της μορφής:

1. Προσομοίωσε ένα υποψήφιο γ^* αλλάζοντας κατά τυχαίο τρόπο d στοιχεία του $\gamma_{(j)}$ με πιθανότητα q_d .
2. Θέσε $\gamma_{(j+1)} = \gamma^*$ με πιθανότητα $\alpha(\gamma^* | \gamma_{(j)}) = \min \left\{ \frac{p(\gamma^* | \mathbf{Y})}{p(\gamma_{(j)} | \mathbf{Y})}, 1 \right\}$, αλλιώς θέσε $\gamma_{(j+1)} = \gamma_{(j)}$.

Η πιθανότητα q_d παραπάνω είναι η πιθανότητα ότι το γ^* θα έχει d καινούργια στοιχεία. Έτσι, βάζοντας κάποια μικρή πιθανότητα q_d για μεγάλα d , ο αλγόριθμος έχει τη δυνατότητα να κάνει ενίοτε και πιο μεγάλα βήματα σε άλλες τιμές γ .

Τέλος, ενδέχεται να θέλουμε να χρησιμοποιήσουμε και μη συμμετρικούς πίνακες που να ευνοούν την επιλογή πιο απλών (με λιγότερες επεξηγηματικές μεταβλητές) μοντέλων. Μία επιλογή σε αυτήν την κατεύθυνση είναι η:

$$q(\gamma_{(i)} | \gamma_{(j)}) = q_d \text{ αν } \sum_{l=1}^p \gamma_{(i)}^l - \gamma_{(j)}^l = d, \quad (3.26)$$

όπου q_d είναι η πιθανότητα να μεταβούμε σε ένα μοντέλο με d περισσότερες ή λιγότερες μεταβλητές (το d μπορεί να είναι και αρνητικός ακέραιος). Χρησιμοποιώντας κατάλληλες πιθανότητες q_d , μπορούμε να κατασκευάσουμε έναν αλγόριθμο που να διερευνά πιο συχνά την ύστερη κατανομή των μοντέλων στην περιοχή των μικρότερων μοντέλων.

Εκτός όμως από τον αλγόριθμο Metropolis-Hastings, μπορούμε να χρησιμοποιήσουμε και τον αλγόριθμο Gibbs Sampling για να κατασκευάσουμε ένα δείγμα από την ύστερη κατανομή $p(\gamma | \mathbf{Y})$. Οι George και McCulloch (1993) πρότειναν έναν τέτοιο αλγόριθμο στα πλαίσια της τεχνικής **SSVS** (Stochastic Search Variable Selection) για το πρόβλημα της επιλογής μεταβλητών.

Εξειδικεύοντας τον αλγόριθμο που είδαμε στο πρώτο κεφάλαιο για την περίπτωση μας, έχουμε τον αλγόριθμο:

Διάλεξε μία τυχαία αρχική τιμή $\gamma_{(0)} = (\gamma_{(0)}^1, \dots, \gamma_{(0)}^k)$ και θέσε $j = 1$.

1. Προσομοίωσε $\gamma_{(j)}^1 \sim \pi_1(\gamma^1 | \gamma_{(j-1)}^2, \dots, \gamma_{(j-1)}^k, \mathbf{Y})$,

προσομοίωσε $\gamma_{(j)}^2 \sim \pi_2(\gamma^2 | \gamma_{(j)}^1, \gamma_{(j-1)}^3, \dots, \gamma_{(j-1)}^k, \mathbf{Y})$,

...

προσομοιώσε $\gamma_{(j)}^k \sim \pi_k(\gamma^k | \gamma_{(j)}^1, \dots, \gamma_{(j)}^{k-1}, \mathbf{Y})$.

2. Θέσε $j = j+1$ και πήγαινε στο 1.

Υπενθυμίζεται ότι βασική προϋπόθεση για την εφαρμογή του αλγορίθμου είναι να μπορούμε να προσομοιώσουμε από τις δεσμευμένες κατανομές $\pi_i(\gamma^i | \gamma^1, \dots, \gamma^{i-1}, \gamma^{i+1}, \dots, \gamma^k, \mathbf{Y})$, $i = 1, \dots, k$ (full conditionals).

Να σημειώσουμε ότι οι Chipman et al. (2001) χρησιμοποιούν την παραπάνω τεχνική για να προσομοιώσουν από την ύστερη κατανομή $p(\gamma | \mathbf{Y})$ μοντέλων γραμμικής παλινδρόμησης και σε περιπτώσεις που δεν έχουν χρησιμοποιηθεί συζυγείς πρότερες κατανομές για τις παραμέτρους και συνεπώς η $p(\gamma | \mathbf{Y})$ δεν υπάρχει σε κλειστή μορφή. Τέτοιες επιλογές μπορεί να υιοθετούνται για να δίνονται πολύ μικρές πιθανότητες σε μοντέλα που περιέχουν μία μεταβλητή X_i για την οποία ο αντίστοιχος συντελεστής β_i είναι πολύ μικρός. Ο αλγόριθμος Gibbs Sampling κατασκευάζει τότε μία μαρκοβιανή αλυσίδα:

$$\boldsymbol{\beta}_{(1)}, \sigma_{(1)}, \boldsymbol{\gamma}_{(1)}, \boldsymbol{\beta}_{(2)}, \sigma_{(2)}, \boldsymbol{\gamma}_{(2)}, \dots, \quad (3.27)$$

αντί της (3.23), προσομοιώνοντας διαδοχικά από τις κατανομές:

$$p(\boldsymbol{\beta} | \sigma^2, \boldsymbol{\gamma}, \mathbf{Y}),$$

$$p(\sigma^2 | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{Y}),$$

$$p(\gamma^i | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}^{(i)}, \mathbf{Y}), \quad i = 1, \dots, k,$$

με $\boldsymbol{\gamma}^{(i)} = (\gamma^1, \gamma^2, \dots, \gamma^{i-1}, \gamma^{i+1}, \dots, \gamma^k)$, όπου σε κάθε στάδιο οι κατανομές είναι δεσμευμένες στις πιο πρόσφατα προσομοιωμένες τιμές των παραμέτρων.

3.7 Η Συνάρτηση `bic.glm` της R

Κλείνοντας αυτό το κεφάλαιο, θα περιγράψουμε συνοπτικά τη λειτουργία της συνάρτησης `bic.glm` της R, η οποία μπορεί να χρησιμοποιηθεί για την εφαρμογή του Bayesian Model Averaging στα γενικευμένα γραμμικά μοντέλα. Το παράδειγμα που θα παρουσιάσουμε στο επόμενο κεφάλαιο θα πραγματοποιηθεί με τη βοήθεια αυτής της συνάρτησης.

Η συνάρτηση `bic.glm` βρίσκεται στο πακέτο BMA, το οποίο περιλαμβάνει και άλλες συναρτήσεις για την εφαρμογή του Bayesian Model Averaging, όπως την `bic.surv` που μπορεί να χρησιμοποιηθεί για μοντέλα αναλογικής διακινδύνευσης και την `bicreg` που χρησιμοποιείται για μοντέλα γραμμικής παλινδρόμησης. Οι συναρτήσεις αυτές χρησιμοποιούνται συχνά στη βιβλιογραφία (π.χ. Hoeting et al., 1999). Αναλυτικές πληροφορίες για το πώς χρησιμοποιούνται οι συναρτήσεις του πακέτου BMA μπορεί να αναζητήσει κανείς στην ηλεκτρονική διεύθυνση <http://cran.r-project.org/web/packages/BMA/BMA.pdf>.

Η συνάρτηση `bic.glm`, που μας ενδιαφέρει εδώ, χρησιμοποιεί προσεγγίσεις BIC για τις ύστερες πιθανότητες των μοντέλων. Επιπλέον, αναζητά τα καλύτερα μοντέλα μέσω του αλγορίθμου «leaps and bounds», ενώ περιορίζει και άλλο τον αριθμό των υποψήφιων μοντέλων μέσω του παραθύρου του Occam, ανάλογα με τις επιλογές που καθορίζει ο χρήστης.

Υπάρχουν πολλά πεδία που μπορεί να επιλέξει και να εισάγει ο χρήστης όταν καλεί την συγκεκριμένη συνάρτηση. Τα κυριότερα από αυτά αφορούν τον πίνακα των δεδομένων των επεξηγηματικών μεταβλητών, το διάνυσμα των δεδομένων της εξαρτημένης μεταβλητής και τον συνδυασμό κατανομής και συνδυαστικής συνάρτησης που θα χρησιμοποιήσουμε. Οι επιλογές αυτές είναι οι μόνες απαραίτητες για να καλέσουμε την συνάρτηση. Ωστόσο, μπορούμε να κάνουμε πιο συγκεκριμένη την επιλογή μας, καθορίζοντας και άλλα πεδία, τα οποία αφορούν το καθορισμό των πρότερων πιθανοτήτων των μοντέλων (αν δεν επιθυμούμε την ομοιόμορφη κατανομή), την χρήση ή όχι του παραθύρου του Occam για περαιτέρω μείωση των μοντέλων και, αν ναι, τον καθορισμό της τιμής O_R , τον τρόπο που χειριζόμαστε τις κατηγορικές μεταβλητές κ.ά.

Ιδιαίτερα σημαντικά είναι τα συμπεράσματα που μπορούμε να πάρουμε μετά την δημιουργία ενός αντικειμένου της `bic.glm`, καλώντας τις αντίστοιχες εντολές. Καταρχάς, οι εντολές `summary`, `plot` και `imageplot` μας δίνουν αντίστοιχα μία σύνοψη των αποτελεσμάτων, γραφήματα για τις ύστερες κατανομές των συντελεστών και μία εικόνα των μοντέλων που επιλέχθηκαν στο άθροισμα. Άλλες εντολές μας δίνουν τις ύστερες πιθανότητες των επιλεχθέντων μοντέλων, τους μέσους και τις διασπορές για κάθε συντελεστή, τις τιμές BIC για κάθε μοντέλο κ.ά.

Θα δούμε πώς μπορούμε να αξιοποιήσουμε την παραπάνω συνάρτηση στο επόμενο κεφάλαιο, όπου θα εφαρμόσουμε τη μέθοδο του Bayesian Model Averaging σε συγκεκριμένα δεδομένα και θα συγκρίνουμε τα αποτελέσματα που προκύπτουν με την εφαρμογή του με αυτά που προκύπτουν χωρίς αυτήν.

Κεφάλαιο 4

Εμφάνιση Στεφανιαίου Επεισοδίου σε Άντρες

Στο κεφάλαιο αυτό θα εφαρμόσουμε τη μέθοδο των κυρτών συνδυασμών μοντέλων σε συγκεκριμένα δεδομένα. Θα δοθεί έμφαση στη σύγκριση των αποτελεσμάτων με εφαρμογή της μεθόδου με αυτά που παίρνουμε χρησιμοποιώντας ένα μόνο μοντέλο. Η ανάλυση θα γίνει με τη βοήθεια της R και κυρίως της συνάρτησης `bic.glm`, που περιγράψαμε στο τέλος του προηγούμενου κεφαλαίου.

4.1 Περιγραφή του Προβλήματος

Τα δεδομένα που θα χρησιμοποιηθούν στη συνέχεια (Chapman data) παρουσιάστηκαν στο *Introduction to Statistical Analysis* των Dixon και Massey (1983) και παρατίθενται σε παράρτημα μετά το τέλος του κεφαλαίου. Τα δεδομένα αποτελούνται από παρατηρήσεις σε 200 άντρες, στους οποίους έχει μετρηθεί η ηλικία (σε έτη), η συστολική και η διαστολική πίεση (σε χιλιοστά στήλης υδραργύρου), η χοληστερόλη (σε mg/dl), το ύψος (σε ίντσες) και το βάρος (σε pounds). Οι παραπάνω μεταβλητές αποτελούν τις επεξηγηματικές μεταβλητές. Η εξαρτημένη μεταβλητή είναι δίτιμη και αφορά την εμφάνιση ή όχι στεφανιαίου επεισοδίου μέσα σε δέκα χρόνια από την εξέταση (1 για εμφάνιση).

Στην ανάλυση που θα κάνουμε αντί των δύο επεξηγηματικών μεταβλητών του ύψους και του βάρους, θα χρησιμοποιήσουμε ως μεταβλητή τον δείκτη μάζας σώματος $BMI = (703.07 * \text{Βάρος}) / (\text{Υψος}^2)$, που αποτελεί έναν τρόπο μέτρησης της παχυσαρκίας κάθε ατόμου. Να σημειώσουμε ότι για άτομα με φυσιολογικό βάρος ο δείκτης BMI παίρνει τιμές μεταξύ του 20 και του 25. Να αναφέρουμε ακόμα ότι την μεταβλητή αυτή είχαν χρησιμοποιήσει και οι Chen et al. (2008) σε ένα άρθρο που είχε σκοπό να δείξει τη σχέση μεταξύ των επιλογών μεταβλητών με διαφορετικά κριτήρια (AIC, BIC κ.ά.). Εμάς ο κύριος σκοπός μας είναι να συγκρίνουμε τα αποτελέσματα που παίρνουμε με την εφαρμογή του Bayesian Model Averaging με

αυτά που παίρνουμε με την ανάλυση με βάση μόνο ένα μοντέλο, βλέποντας παράλληλα ποιες είναι οι πιο σημαντικές μεταβλητές για τα δεδομένα μας.

4.2 Αποτελέσματα

Βλέποντας πως έχουμε μία δίτιμη εξαρτημένη μεταβλητή και πέντε ποσοτικές επεξηγηματικές μεταβλητές, είναι εύλογο να οδηγηθούμε στη χρήση λογιστικής παλινδρόμησης, θεωρώντας ως συνδετική συνάρτηση την $\text{logit}\{\text{Pr}[Y = 1]\} = \text{log}\{\text{Pr}[Y = 1] / (1 - \text{Pr}[Y = 1])\}$. Η ανάλυση με κυρτούς συνδυασμούς μοντέλων θα γίνει μέσω της συνάρτησης `bic.glm`, με την οποία μπορούμε να δούμε και ποιά είναι τα μοντέλα με τις μεγαλύτερες ύστερες πιθανότητες. Στην περίπτωση μας έχουμε $2^5 = 32$ πιθανά μοντέλα.

Μπορούμε αρχικά να εφαρμόσουμε τη μέθοδο χρησιμοποιώντας λογιστική παλινδρόμηση μέσω της εντολής:

```
glm.out<-bic.glm(X,Y,glm.family=binomial(link="logit")),
```

όπου X είναι ο πίνακας με τις τιμές των επεξηγηματικών μεταβλητών και Y το διάνυσμα με τις τιμές της εξαρτημένης μεταβλητής.

Καθώς δεν έχουμε επιλέξει συγκεκριμένη *prior* πιθανότητα για την εισαγωγή κάθε μεταβλητής, θα τεθεί πιθανότητα ίση με $1/2$ για κάθε μεταβλητή (που είναι η *default* επιλογή), η οποία οδηγεί – βάσει της υπόθεσης της ανεξαρτησίας μεταξύ των μεταβλητών - στην ομοιόμορφη πρότερη κατανομή στο χώρο των 32 πιθανών μοντέλων.

Οι ύστερες πιθανότητες που προκύπτουν τότε για κάθε μεταβλητή, δηλαδή η πιθανότητα ο αντίστοιχος συντελεστής να είναι μη μηδενικός, προκύπτουν από το άθροισμα των ύστερων πιθανοτήτων των μοντέλων στα οποία αυτή περιλαμβάνεται και δίνονται στον ακόλουθο πίνακα:

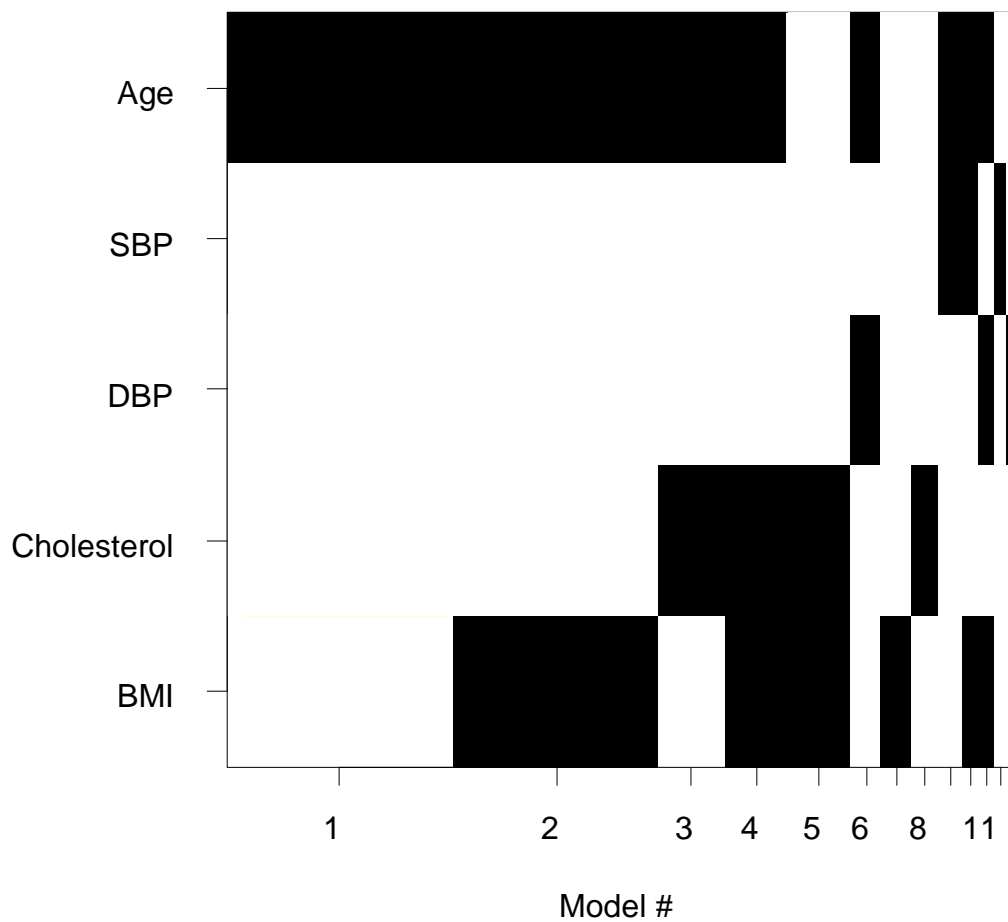
Μεταβλητή	P($\beta \neq 0$ D)
Ηλικία	81.9%
SBP	6.8%
DBP	7.3%
Χοληστερόλη	27.7%
BMI	49.4%

Πίνακας 1: Ύστερες πιθανότητες μη μηδενικών συντελεστών για κάθε μεταβλητή.

Από τα παραπάνω προκύπτει ότι μόνο ο παράγοντας της ηλικίας είναι σχεδόν βέβαιο ότι επηρεάζει την εξαρτημένη μεταβλητή μας, ενώ μεγάλη πιθανότητα (σχεδόν 50%) συγκεντρώνει και ο δείκτης BMI. Τρίτη πιο σημαντική μεταβλητή φαίνεται να είναι η χοληστερόλη, ενώ οι άλλες δύο (συστολική και διαστολική πίεση) συγκεντρώνουν πολύ μικρές πιθανότητες.

Η εντολή `imageplot.bma(glm.out)` μας παρέχει μία εικόνα των επικρατέστερων μοντέλων, δείχνοντας ποιές μεταβλητές περιλαμβάνει το καθένα από αυτά:

Models selected by BMA



Γράφημα 1: Τα μοντέλα με τις μεγαλύτερες ύστερες πιθανότητες.

Γίνεται φανερό ότι η ηλικία υπεισέρχεται σε όλα τα επικρατέστερα μοντέλα, σε αντίθεση με τις μεταβλητές SBP και DBP που χρησιμοποιούνται σε λίγα μοντέλα και μάλιστα όχι στα επικρατέστερα. Η εισαγωγή της χοληστερόλης και του BMI σε πολλά μοντέλα είναι επίσης εμφανής.

Τα συμπεράσματα συμφωνούν με την ανάλυση μέσω backward stepwise selection, καθώς το μοντέλο M_1 που περιλαμβάνει τις τρεις επικρατέστερες μεταβλητές (ηλικία, BMI, χοληστερόλη) είναι αυτό που επιλέγεται μέσω του κριτηρίου AIC, όπως αναφέρεται και στο άρθρο των Chen et al. που προαναφέραμε (2008). Αντίθετα, το κριτήριο BIC (που ευνοεί τα μικρότερα μοντέλα) θεωρεί ως

επικρατέστερο μοντέλο το μοντέλο M_2 που περιλαμβάνει μόνο την ηλικία, ενώ το μοντέλο M_1 έρχεται τέταρτο. Οι ύστερες πιθανότητες που μάς βγάζει η συνάρτηση `bic.glm` πηγάζουν βάσει του κριτηρίου BIC και ως εκ τούτου το μοντέλο με τη μεγαλύτερη πιθανότητα είναι το M_2 με πιθανότητα 28.7%, ενώ το μοντέλο M_1 συγκεντρώνει μόλις το 8.1%. Σε κάθε περίπτωση, είναι εμφανές ότι υπάρχει μεγάλη αβεβαιότητα γύρω από την επιλογή ενός συγκεκριμένου μοντέλου και αυτό θα πρέπει να ληφθεί υπόψη στα συμπεράσματά μας.

Είναι ενδιαφέρον να δούμε τώρα ποιές είναι οι p-τιμές που μας δίνει μία ανάλυση με λογιστική παλινδρόμηση (χωρίς τη μέθοδο των κυρτών συνδυασμών μοντέλων) για τα δεδομένα μας. Προς το σκοπό αυτό θα χρησιμοποιήσουμε τη συνάρτηση `glm` της R και συγκεκριμένα τις εντολές:

```
out<-glm(Y~Age+SBP+DBP+Cholesterol+BMI,family=binomial(link=logit))
```

```
summary(out)
```

Τα αποτελέσματα συγκεντρώνονται στον παρακάτω πίνακα:

Μεταβλητή	p-τιμή
Ηλικία	0.0499
SBP	0.7609
DBP	0.8501
Χοληστερόλη	0.0865
BMI	0.0355

Πίνακας 2: Οι p-τιμές για κάθε μεταβλητή.

Βλέπουμε ότι οι δύο μεταβλητές που φαίνονται σημαντικές είναι η ηλικία και το BMI, με p-τιμές < 0.05. Δηλαδή, υπάρχουν ισχυρές ενδείξεις ότι οι συντελεστές τους είναι μη μηδενικοί. Η τιμή για τη χοληστερόλη ακολουθεί με $p=0.0865$, τιμή που δείχνει ότι και αυτός ο παράγοντας δεν πρέπει να αγνοηθεί, καθώς υπάρχουν αρκετές ενδείξεις κατά της υπόθεσης για μηδενικό συντελεστή. Αναφορικά με τις μεταβλητές SBP και DBP υπάρχει συμφωνία με τα όσα είδαμε πριν, καθώς οδηγούμαστε στο συμπέρασμα ότι έχουν μικρή επιρροή στην εξαρτημένη μεταβλητή μας.

Θα παραθέσουμε τώρα τις εκτιμημένες τιμές και τις τυπικές αποκλίσεις για τους συντελεστές κάθε μεταβλητής, όπως αυτοί εκτιμώνται από τις συναρτήσεις `glm` (χωρίς BMA) και `bic.glm` (με BMA). Δίνουμε αρχικά στον παρακάτω πίνακα τις τιμές που υπολογίζονται χωρίς εφαρμογή του BMA:

Μεταβλητή	Εκτιμημένος συντελεστής	Τυπική απόκλιση
Σταθερός όρος	-9.527542	2.216748
Ηλικία	0.045352	0.023129
SBP	0.006127	0.020137
DBP	-0.007248	0.038355
Χοληστερόλη	0.006198	0.003616
BMI	0.138956	0.066105

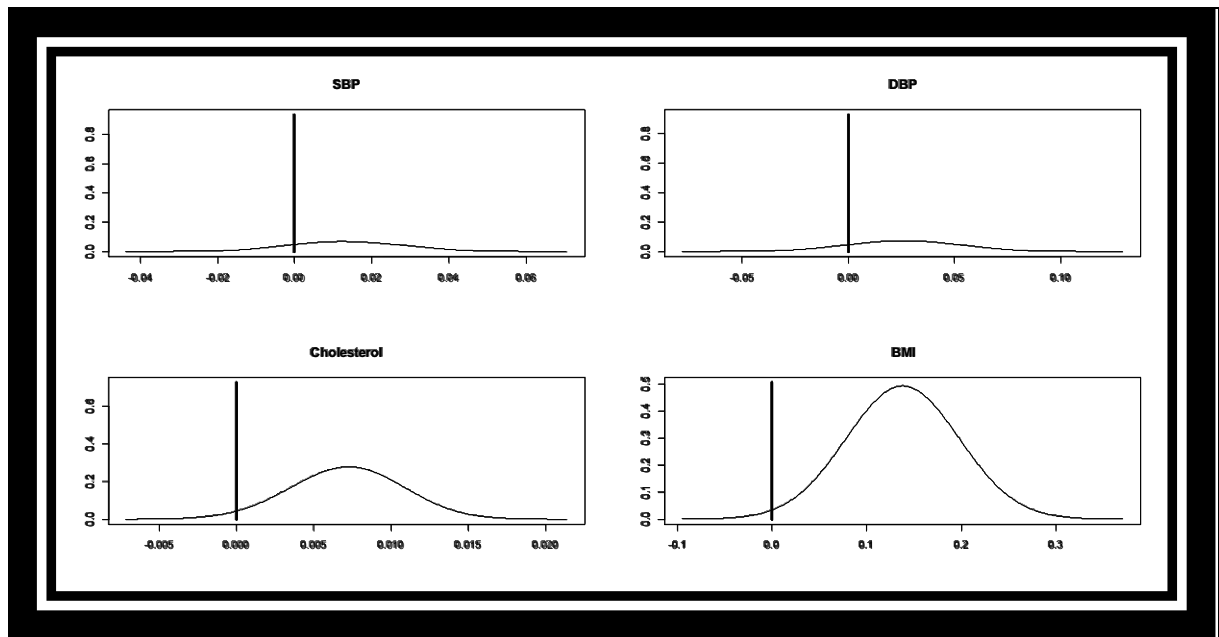
Πίνακας 3: Εκτιμημένες τιμές και αποκλίσεις συντελεστών χωρίς BMA.

Ακολουθούν οι τιμές που υπολογίζονται με εφαρμογή του BMA:

Μεταβλητή	Εκτιμημένος συντελεστής	Τυπική απόκλιση
Σταθερός όρος	-6.6631950	2.233265
Ηλικία	0.0474846	0.029098
SBP	0.0009066	0.005085
DBP	0.0018778	0.009833
Χοληστερόλη	0.0019696	0.003723
BMI	0.0682742	0.080892

Πίνακας 4: Εκτιμημένες τιμές και αποκλίσεις συντελεστών με BMA.

Για την περίπτωση της ανάλυσης με BMA, η R μας δίνει τα γραφήματα των ύστερων κατανομών του συντελεστή κάθε μεταβλητής μέσω της εντολής `plot.bic.glm(glm.out)`. Θα δούμε τα γραφήματα αυτά συγκεντρωμένα για τους συντελεστές των μεταβλητών πλην της ηλικίας, η οποία από τα παραπάνω είναι βέβαιο πως θα περιληφθεί στην ανάλυσή μας:



Γράφημα 2: Ύστερες κατανομές των συντελεστών κάθε μεταβλητής.

Διαπιστώνουμε έτσι και πάλι ότι η επίδραση των μεταβλητών SBP και DBP είναι μικρή και οι αντίστοιχοι συντελεστές είναι πιο κοντά στο μηδέν από τους υπόλοιπους. Να σημειώσουμε εδώ ότι στα παραπάνω γραφήματα οι κάθετες γραμμές στο σημείο 0 των οριζόντιων αξόνων αντιστοιχούν στην πιθανότητα ότι οι αντίστοιχοι συντελεστές είναι μηδενικοί (πάνω από 90% στην περίπτωση των SBP, DBP), ενώ οι κατανομές έχουν κανονικοποιηθεί ώστε να παίρνουν για μέγιστη τιμή την πιθανότητα $P(\beta \neq 0 | D)$.

4.3 Εξαγωγή Συμπερασμάτων από p-τιμές και από Ύστερες Πιθανότητες

Ένα σημείο στο οποίο πρέπει να σταθούμε είναι η σύγκριση μεταξύ των συμπερασμάτων που μπορεί να εξάγει κάποιος από τις p-τιμές και από τις ύστερες πιθανότητες να είναι ο συντελεστής κάθε μεταβλητής μη μηδενικός. Όπως έχουν τονίσει και οι Hoeting et al. (1999), οι ύστερες πιθανότητες κάθε μεταβλητής προσφέρουν μία διάκριση που οι p-τιμές δεν μπορούν να προσφέρουν. Συγκεκριμένα,

τα κριτήρια για τις p-τιμές αφορούν στο αν αυτές είναι μικρότερες από 0.01, 0.05 και 0.1 οδηγώντας αντίστοιχα στα συμπεράσματα ότι υπάρχουν πολύ ισχυρές, ισχυρές και αρκετές ενδείξεις κατά της μηδενικής υπόθεσης. Σε περίπτωση που είναι μεγαλύτερες από αυτές τις τιμές, και ειδικότερα από την τιμή 0.1, σπάνια απορρίπτουμε την μηδενική υπόθεση. Αντίθετα, οι ύστερες πιθανότητες $P(\beta_i \neq 0 | D)$ μπορούν να μας οδηγήσουν σε ανικανότητα να απορρίψουμε τη μηδενική υπόθεση για δύο λόγους: είτε επειδή τα δεδομένα δείχνουν να υποστηρίζουν τη μηδενική υπόθεση είτε επειδή δεν υπάρχουν αρκετά δεδομένα για να εξάγουμε ασφαλή συμπεράσματα. Για παράδειγμα, στην περίπτωση της χοληστερόλης (πιθανότητα 27.7%) το συμπέρασμα που βγάζουμε είναι ότι για να οδηγηθούμε σε πιο ασφαλή συμπεράσματα θα πρέπει να έχουμε περισσότερα δεδομένα στη διάθεσή μας.

Η παραπάνω διάκριση μπορεί να γίνει πιο σαφής αν συγκρίνουμε τα συμπεράσματα που εξάγουμε με την ανάλυση όλων των δεδομένων με αυτά που θα βγάγαμε αν είχαμε τις μισές (τις 100 πρώτες) παρατηρήσεις στη διάθεσή μας. Συγκεκριμένα, η ανάλυση με τη μέθοδο BMA δίνει τις ακόλουθες ύστερες πιθανότητες $P(\beta_i \neq 0 | D_T)$, όπου με D_T συμβολίζουμε το πρώτο μισό των δεδομένων μας:

Μεταβλητή	$P(\beta_i \neq 0 D_T)$
Ηλικία	37.9%
SBP	26.1%
DBP	22.3%
Χοληστερόλη	34.3%
BMI	39.8%

Πίνακας 5: Ύστερες πιθανότητες μη μηδενικών συντελεστών για τα μισά δεδομένα.

Τα παραπάνω αποτελέσματα δίνουν πάλι τις μεγαλύτερες πιθανότητες στις μεταβλητές της ηλικίας, της χοληστερόλης και του δείκτη BMI και μικρότερες πιθανότητες (αλλά με όχι τόσο μεγάλη διαφορά) στις υπόλοιπες δύο μεταβλητές. Σε κάθε περίπτωση, το συμπέρασμα που βγάζουμε είναι ότι χρειαζόμαστε επιπλέον δεδομένα, καθώς προς το παρόν δεν μπορούμε να έχουμε ασφαλή συμπεράσματα για το ποιές μεταβλητές ασκούν όντως επίδραση στην εξαρτημένη μεταβλητή μας.

Ας δούμε τώρα τί συμπεράσματα προκύπτουν από την αντίστοιχη ανάλυση (των μισών δεδομένων) με βάση τη συνάρτηση glm. Οι p-τιμές που προκύπτουν για κάθε μεταβλητή δίνονται στον ακόλουθο πίνακα:

Μεταβλητή	p-τιμή
Ηλικία	0.142397
SBP	0.561714
DBP	0.877044
Χοληστερόλη	0.093563
BMI	0.095913

Πίνακας 6: Οι p-τιμές για κάθε μεταβλητή με βάση τα μισά δεδομένα.

Με βάση τα παραπάνω, οι μόνες μεταβλητές για τις οποίες έχουμε αρκετές ενδείξεις κατά της μηδενικής υπόθεσης είναι η χοληστερόλη και το BMI, για τις οποίες $p < 0.1$. Αντίθετα, για τη μεταβλητή της ηλικίας έχουμε $p > 0.1$, με συνέπεια να μην μπορούμε να απορρίψουμε τη μηδενική υπόθεση για την (όπως αποδεικνύεται από την συνολική ανάλυση) επιδραστικότερη μεταβλητή μας.

Ως εκ τούτου, βλέπουμε ότι είναι ασφαλέστερο να κρίνουμε την αβεβαιότητά μας με βάση τις ύστερες πιθανότητες από ότι με βάση τις p-τιμές. Αυτό σχετίζεται με την ασυμμετρία με την οποία χειριζόμαστε την μηδενική και την εναλλακτική

υπόθεση στην Κλασσική Στατιστική, καθώς τείνουμε να μην απορρίπτουμε την μηδενική υπόθεση κάθε φορά που δεν έχουμε αρκετές ενδείξεις εναντίον της, όπως έχει τονιστεί και στη βιβλιογραφία (Κοκολάκης και Σπηλιώτης, 2010).

4.4 Σύγκριση της Προβλεπτικής Ικανότητας

Ο χωρισμός του συνόλου των δεδομένων σε δύο υποσύνολα είναι σημαντικός και για να μετρήσουμε την προβλεπτική ικανότητα του BMA, συγκρίνοντάς την με αυτήν ενός μεμονωμένου μοντέλου. Σύμφωνα με τα όσα είπαμε στο δεύτερο κεφάλαιο, μπορούμε να χωρίσουμε τα δεδομένα μας σε δύο ισοπληθικά υποσύνολα. Στο πρώτο από αυτά, το σύνολο D_T , θα κάνουμε ανάλυση και με τη μέθοδο BMA και χωρίς αυτήν και στο δεύτερο σύνολο, το D_p , θα εξετάσουμε αν όντως η χρήση του BMA δίνει καλύτερη πρόβλεψη.

Είδαμε στην προηγούμενη παράγραφο τις ύστερες πιθανότητες και τις p-τιμές μη μηδενικού συντελεστή για κάθε μεταβλητή που προκύπτουν με βάση το σύνολο D_T . Τώρα θα παραθέσουμε τις εκτιμήσεις των συντελεστών κάθε μεταβλητής που προκύπτουν χωρίς την εφαρμογή του BMA και με αυτήν αντίστοιχα.

Μεταβλητή	Εκτιμημένος συντελεστής	Τυπική απόκλιση
Σταθερός όρος	-12.073090	3.347859
Ηλικία	0.046482	0.031687
SBP	0.017375	0.029941
DBP	-0.008347	0.053949
Χοληστερόλη	0.008715	0.005197
BMI	0.152344	0.091498

Πίνακας 7: Εκτιμημένες τιμές και αποκλίσεις συντελεστών χωρίς BMA με βάση τα μισά δεδομένα.

Μεταβλητή	Εκτιμημένος συντελεστής	Τυπική απόκλιση
Σταθερός όρος	-7.143847	3.271011
Ηλικία	0.020361	0.031420
SBP	0.007606	0.015842
DBP	0.010637	0.024985
Χοληστερόλη	0.003130	0.005234
BMI	0.063327	0.093579

Πίνακας 8: Εκτιμημένες τιμές και αποκλίσεις συντελεστών με BMA με βάση τα μισά δεδομένα.

Θα χρησιμοποιήσουμε τώρα τις παραπάνω εκτιμήσεις για να προβλέψουμε τις τιμές της εξαρτημένης μεταβλητής στο δεύτερο υποσύνολο των δεδομένων, D_p . Έχοντας στη διάθεσή μας τα υπόλοιπα δεδομένα με τις τιμές των ανεξάρτητων μεταβλητών, μπορούμε να πάρουμε μία πρόβλεψη για τις τιμές της εξαρτημένης μεταβλητής στις αντίστοιχες θέσεις μέσω της αντίστροφης συνάρτησης της $\text{logit}(p)$, δηλαδή της $\text{logit}^{-1}(a) = \frac{\exp(a)}{1 + \exp(-a)}$. Στην περίπτωση μας, στη θέση του a θα αντικαταστήσουμε τον γραμμικό συνδυασμό $a_0 + a_1 \text{Age}_i + a_2 \text{SBP}_i + a_3 \text{DBP}_i + a_4 \text{Cholesterol}_i + a_5 \text{BMI}_i$, $i = 101, \dots, 200$, όπου a_0, \dots, a_5 οι αντίστοιχοι εκτιμημένοι συντελεστές.

Εκτελώντας αυτήν την διαδικασία με τις εκτιμήσεις που προέκυψαν χωρίς την μέθοδο του BMA και με αυτήν, παίρνουμε δύο διανύσματα με προβλέψεις για την εξαρτημένη μεταβλητή. Καλώντας «Prediction1» το πρώτο διάνυσμα και «Prediction2» το δεύτερο, οι τιμές που παίρνουν τα δύο αυτά διανύσματα στην i -θέση παρέχονται αντίστοιχα από τις εντολές:

```
Prediction1[i]<-exp(-12.073090 + 0.046482*Age2[i] + 0.017375 *SBP2[i] -
0.008347*DBP2[i] + 0.008715*Cholesterol2[i] + 0.152344 *BMI2[i]) / (1 + exp(-
12.073090 + 0.046482*Age2[i] + 0.017375 *SBP2[i] - 0.008347*DBP2[i] +
0.008715*Cholesterol2[i] + 0.152344 *BMI2[i]))
```

και

```
Prediction1[i]<-exp( -7.14387 +0.020361*Age2[i] + 0.007606*SBP2[i] +
0.010637*DBP2[i] + 0.003130* Cholesterol2[i] + 0.063327*BMI2[i]) / (1 + exp( -
7.14387 + 0.020361*Age2[i] + 0.007606*SBP2[i] + 0.010637*DBP2[i] + 0.003130*
Cholesterol2[i] + 0.063327*BMI2[i])).
```

Θα εξετάσουμε τώρα πόσο καλά αυτά τα διανύσματα μπορούν να προβλέψουν τις περιπτώσεις που η εξαρτημένη μεταβλητή παίρνει την τιμή 1, δηλαδή τις περιπτώσεις που έχουμε εμφάνιση στεφανιαίου επεισοδίου.

Ψάχνοντας αρχικά τις θέσεις στις οποίες τα δύο διανύσματα παίρνουν τις μεγαλύτερες τιμές τους, δηλαδή αυτές που είναι πιο κοντά στη μονάδα, συμπεραίνουμε ότι οι θέσεις αυτές σχεδόν ταυτίζονται. Συγκεκριμένα, οι εννέα από τις δέκα αυτές θέσεις είναι κοινές για τα δύο διανύσματα: οι θέσεις 8, 11, 12, 13, 14, 16, 21, 23 και 93. Η διαφορά είναι ότι η δέκατη θέση για το διάνυσμα «Prediction1» είναι η θέση 49, ενώ για το «Prediction2» η θέση 57. Το ερώτημα που προκύπτει, ασφαλώς, είναι σε ποιές από αυτές τις θέσεις έχουμε όντως εμφάνιση στεφανιαίου επεισοδίου. Η απάντηση είναι στις θέσεις 11 και 23, δηλαδή στις θέσεις αυτές τα δύο διανύσματα παίρνουν δικαιολογημένα σχετικά υψηλές τιμές.

Αυτό που έχει ενδιαφέρον στο σημείο αυτό, είναι να συγκρίνουμε τις τιμές των δύο διανυσμάτων στις υπόλοιπες θέσεις που η εξαρτημένη μεταβλητή ισούται με τη μονάδα. Αυτές είναι εννέα το πλήθος και συγκεκριμένα οι: 24, 26, 29, 51, 59, 84, 88, 90 και 92. Οι τιμές των διανυσμάτων στις θέσεις αυτές συγκεντρώνονται στον επόμενο πίνακα:

Θέση	Πρόβλεψη χωρίς BMA	Πρόβλεψη με BMA
24	0.08761016	0.1147227
26	0.1290799	0.1393436
29	0.1236877	0.1585387
51	0.0509151	0.08363773
59	0.03764762	0.07436111
84	0.162247	0.1762899
88	0.119935	0.1340938
90	0.2834756	0.2069562
92	0.07236727	0.1385301

Πίνακας 9: Προβλέψεις της εξαρτημένης μεταβλητής με και χωρίς BMA.

Παρατηρούμε ότι στις οκτώ από τις εννέα αυτές θέσεις, η τιμή που προκύπτει με την εφαρμογή του Bayesian Model Averaging είναι πιο κοντά στη μονάδα από την αντίστοιχη τιμή χωρίς την εφαρμογή της μεθόδου. Καταλήγουμε έτσι στο συμπέρασμα ότι η εφαρμογή του BMA οδήγησε σε καλύτερες προβλέψεις στις περιπτώσεις που είχαμε εμφάνιση στεφανιαίου επεισοδίου.

4.5 Ανάλυση χωρίς την Μεταβλητή της Ηλικίας

Στην ανάλυση που προηγήθηκε, εστίασαμε στην σύγκριση των αποτελεσμάτων με και χωρίς την χρήση κυρτών συνδυασμών μοντέλων χρησιμοποιώντας πέντε μεταβλητές. Είδαμε ότι η μεταβλητή της ηλικίας είχε τη μεγαλύτερη πιθανότητα να έχει μη μηδενικό συντελεστή, με μεγάλη διαφορά από τις υπόλοιπες μεταβλητές. Μάλιστα, το κριτήριο BIC έδινε ως επικρατέστερο μοντέλο αυτό που περιλαμβάνει μόνο την μεταβλητή της ηλικίας.

Αυτό εξηγείται από το γεγονός ότι η ηλικία σχετίζεται με όλες τις άλλες επεξηγηματικές μεταβλητές. Κάνοντας μία απλή ανάλυση για κάθε μία από τις υπόλοιπες μεταβλητές για να δούμε τη σχέση τους με την ηλικία, έχουμε σε κάθε περίπτωση πολύ ισχυρές ενδείξεις ότι η ηλικία τις επηρεάζει, με τις αντίστοιχες p-τιμές να είναι όλες πολύ μικρότερες από 0.01. Ως εκ τούτου, θα ήταν καλό να κάνουμε ανάλυση του προβλήματος και χωρίς την μεταβλητή της ηλικίας.

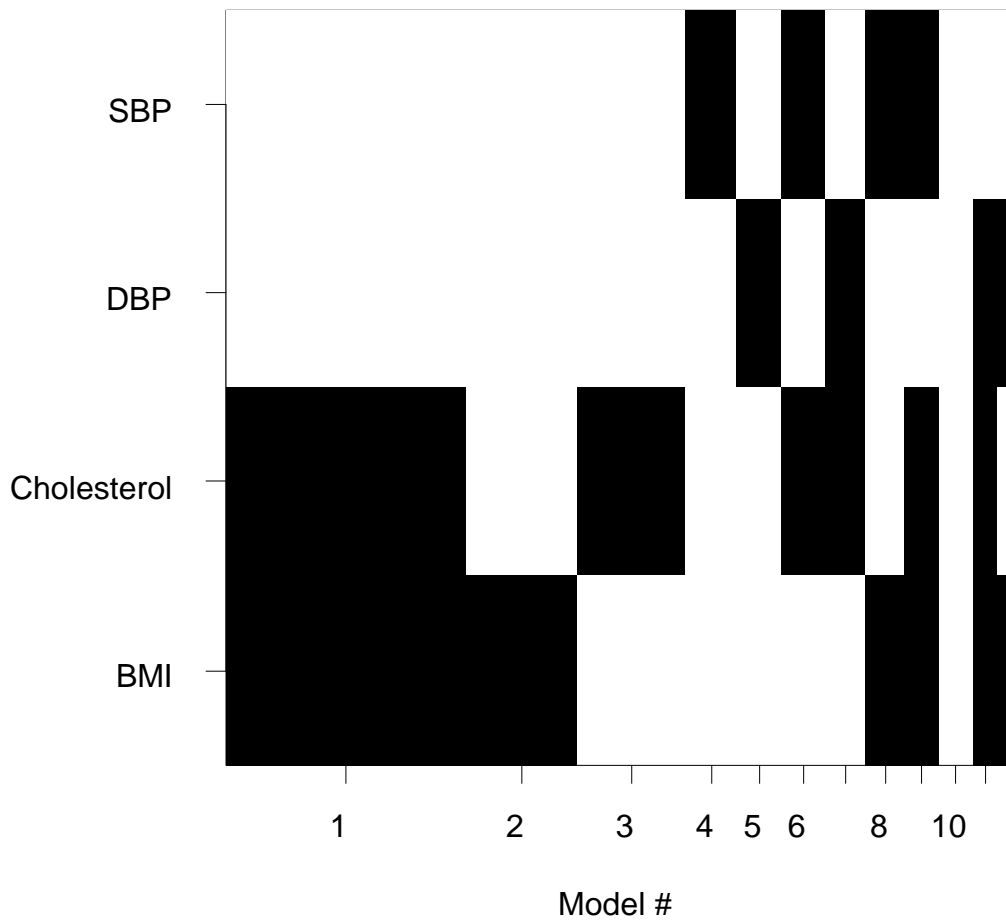
Με ανάλογο τρόπο με πριν, θα χρησιμοποιήσουμε την συνάρτηση `bic.glm` για να εφαρμόσουμε τη μέθοδο του Bayesian Model Averaging, θεωρώντας τώρα ως επεξηγηματικές μεταβλητές τις SBP, DBP, την χοληστερόλη και τον δείκτη BMI. Με τέσσερις μεταβλητές, το πλήθος των πιθανών μοντέλων γίνεται $2^4 = 16$. Και εδώ θα θεωρήσουμε ομοιόμορφη πρότερη κατανομή στο χώρο των μοντέλων. Οι ύστερες πιθανότητες για μη μηδενικό συντελεστή που προκύπτουν τότε για κάθε μεταβλητή, δίνονται στον ακόλουθο πίνακα:

Μεταβλητή	$P(\beta \neq 0 D)$
SBP	21.0%
DBP	16.6%
Χοληστερόλη	62.1%
BMI	59.3%

Πίνακας 10: Ύστερες πιθανότητες μη μηδενικών συντελεστών για κάθε μεταβλητή (πλην ηλικίας).

Παρατηρούμε ότι η χοληστερόλη και ο δείκτης BMI συγκεντρώνουν τώρα τα μεγαλύτερα ποσοστά, ενώ οι δύο πιέσεις συγκεντρώνουν μεγαλύτερα ποσοστά από πριν αλλά και πάλι πολύ μικρότερα σε σχέση με τις υπόλοιπες μεταβλητές. Τα τρία καλύτερα μοντέλα που μας δίνει η ανάλυση με κυρτούς συνδυασμούς μοντέλων είναι κατά σειρά αυτό που περιλαμβάνει την χοληστερόλη και το BMI, αυτό που περιλαμβάνει μόνο το BMI και αυτό που περιλαμβάνει μόνο την χοληστερόλη. Μπορούμε να έχουμε μία οπτική αναπαράσταση των καλύτερων μοντέλων μέσω της εντολής `imageplot.bma`, η οποία μας δίνει το ακόλουθο γράφημα:

Models selected by BMA



Γράφημα 3: Τα μοντέλα με τις μεγαλύτερες ύστερες πιθανότητες

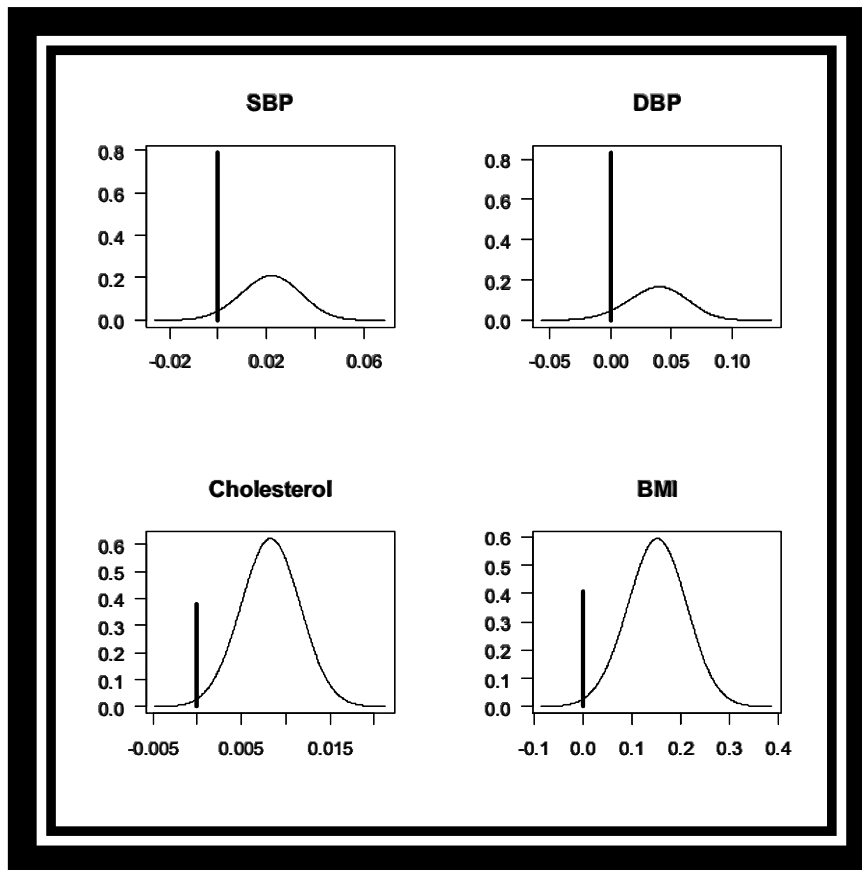
Βλέπουμε λοιπόν ότι οι μεταβλητές της χοληστερόλης και του BMI περιλαμβάνονται στα περισσότερα και στα σημαντικότερα μοντέλα. Μάλιστα, οι δύο αυτές μεταβλητές είναι οι μόνες που δείχνουν να είναι στατιστικά σημαντικές μετά από μία ανάλυση με λογιστική παλινδρόμηση με χρήση της συνάρτησης $\text{glm}(Y \sim \text{SBP} + \text{DBP} + \text{Cholesterol} + \text{BMI}, \text{family} = \text{binomial}(\text{link} = \text{logit}))$. Οι p-τιμές που προκύπτουν για κάθε μεταβλητή δίνονται συγκεντρωμένες στον παρακάτω πίνακα:

Μεταβλητή	p-τιμή
SBP	0.3494
DBP	0.8267
Χοληστερόλη	0.0228
BMI	0.0319

Πίνακας 11: Οι p-τιμές για κάθε μεταβλητή.

Βλέπουμε ότι οι p-τιμές για τις δύο τελευταίες μεταβλητές είναι μικρότερες από 0.05, σε αντίθεση με τις υψηλές τιμές των δύο πρώτων μεταβλητών. Μάλιστα, η p-τιμή για την διαστολική πίεση παραμένει στα ίδια επίπεδα με πριν, παρά την απομάκρυνση της μεταβλητής της ηλικίας από την ανάλυσή μας.

Οι γραφικές παραστάσεις των ύστερων κατανομών του συντελεστή κάθε μεταβλητής που μας δίνει η εντολή `plot.bic.glm` για την ανάλυση με κυρτούς συνδυασμούς μοντέλων δίνονται συγκεντρωμένες παρακάτω:



Γράφημα 4: Ύστερες κατανομές των συντελεστών κάθε μεταβλητής.

Έχουμε έτσι μία εικόνα του ότι οι συντελεστές των δύο πιέσεων είναι πιο κοντά στο μηδέν. Όλα τα παραπάνω οδηγούν στο συμπέρασμα πως οι μεταβλητές της χοληστερόλης και του δείκτη BMI είναι αυτές που σχετίζονται περισσότερο με την εξαρτημένη μεταβλητή.

Παράρτημα

Στον ακόλουθο πίνακα παρουσιάζονται τα δεδομένα του Charman, τα οποία χρησιμοποιήσαμε στο τέταρτο κεφάλαιο. Με Y συμβολίζουμε την εξαρτημένη μεταβλητή που αφορά την εμφάνιση ή όχι στεφανιαίου επεισοδίου (1 για εμφάνιση).

	Ηλικία	Συστολική Πίεση	Διαστολική Πίεση	Χοληστερόλη	Ύψος	Βάρος	Y
1	44	124	80	254	70	190	0
2	35	110	70	240	73	216	0
3	41	114	80	279	68	178	0
4	31	100	80	284	68	149	0
5	61	190	110	315	68	182	1
6	61	130	88	250	70	185	0
7	44	130	94	298	68	161	0
8	58	110	74	384	67	175	0
9	52	120	80	310	66	144	0
10	52	120	80	337	67	130	0
11	52	130	80	367	69	162	0
12	40	120	90	273	68	175	0
13	49	130	75	273	66	155	0
14	34	120	80	314	74	156	0
15	37	115	70	243	65	151	0
16	63	140	90	341	74	168	0
17	28	138	80	245	70	185	0
18	40	115	82	302	69	225	0
19	51	148	110	302	69	247	1
20	33	120	70	386	66	146	0
21	37	110	70	312	71	170	1
22	33	132	90	302	69	161	0
23	41	112	80	394	69	167	0
24	38	114	70	358	69	198	0
25	52	100	78	336	70	162	0

26	31	114	80	251	71	150	0
27	44	110	80	322	68	196	1
28	31	108	70	281	67	130	0
29	40	110	74	336	68	166	1
30	36	110	80	314	73	178	0
31	42	136	82	383	69	187	0
32	28	124	82	360	67	148	0
33	40	120	85	369	71	180	0
34	40	150	100	333	70	172	0
35	35	100	70	253	68	141	0
36	32	120	80	268	68	176	0
37	31	110	80	257	71	154	0
38	52	130	90	474	69	145	0
39	45	110	80	391	69	159	1
40	39	106	80	248	67	181	0
41	40	130	90	520	68	169	1
42	48	110	70	285	66	160	1
43	29	110	70	352	66	149	0
44	56	141	100	428	65	171	1
45	53	90	55	334	68	166	0
46	47	90	60	278	69	121	0
47	30	114	76	264	73	178	0
48	64	140	90	243	71	171	1
49	31	130	88	348	72	181	0
50	35	120	88	290	70	162	0
51	65	130	90	370	65	153	1
52	43	122	82	363	69	164	0
53	53	120	80	343	71	159	0
54	58	138	82	305	67	152	1
55	67	168	105	365	68	190	1
56	53	120	80	307	70	200	0
57	42	134	90	243	67	147	0

58	43	115	75	266	68	125	0
59	52	110	75	341	69	163	0
60	68	110	80	268	62	138	0
61	64	105	68	261	66	108	0
62	46	138	90	378	67	142	0
63	41	120	90	279	70	212	0
64	58	130	90	416	68	188	0
65	50	160	110	261	66	145	0
66	45	100	70	332	67	144	0
67	59	156	90	337	67	158	0
68	56	120	92	365	65	154	0
69	59	126	96	292	67	148	0
70	47	110	80	304	67	155	0
71	43	95	70	341	69	154	0
72	37	120	74	317	74	184	0
73	27	100	60	296	67	140	0
74	44	110	80	390	66	167	0
75	41	120	80	274	69	138	0
76	33	120	80	355	68	169	0
77	29	115	80	225	70	186	0
78	24	120	80	218	69	131	0
79	36	108	66	298	67	160	0
80	23	110	78	178	66	142	0
81	47	120	80	341	70	218	1
82	26	110	75	274	70	147	0
83	45	130	90	285	65	161	0
84	41	164	110	259	66	245	0
85	55	125	88	266	67	167	0
86	34	110	80	214	67	139	1
87	51	110	75	267	66	150	0
88	58	120	80	256	67	175	0
89	51	112	88	273	74	123	0

90	35	120	75	348	72	174	0
91	34	118	78	322	69	192	0
92	26	120	70	267	70	140	0
93	25	110	75	270	74	195	0
94	44	100	70	280	65	144	0
95	57	130	85	320	69	193	0
96	67	110	80	320	64	134	0
97	59	160	90	330	63	144	0
98	62	130	88	274	69	179	0
99	40	140	90	269	63	111	0
100	52	120	90	269	68	164	0
101	28	130	80	135	67	168	0
102	34	115	80	403	69	175	0
103	43	122	78	294	68	173	0
104	38	125	80	312	71	158	0
105	45	110	80	311	69	154	0
106	26	120	84	222	72	214	0
107	35	112	90	303	67	176	0
108	51	120	90	269	70	262	0
109	55	120	82	311	71	181	0
110	45	130	90	286	73	143	0
111	69	160	90	370	67	185	1
112	58	140	90	403	66	140	0
113	64	190	100	244	66	187	0
114	70	190	112	353	66	163	0
115	27	112	78	252	68	164	0
116	53	155	104	453	66	170	0
117	28	115	65	260	66	150	0
118	29	110	80	269	68	141	0
119	23	120	82	235	65	135	0
120	40	120	78	264	71	135	0
121	53	160	90	420	67	141	0

122	25	110	75	235	69	148	0
123	63	130	90	420	69	160	1
124	48	110	78	277	71	180	1
125	36	120	80	319	70	157	0
126	28	120	86	386	70	189	1
127	57	110	68	353	71	166	0
128	39	110	80	344	66	175	0
129	52	130	90	210	65	172	1
130	51	140	90	286	67	134	0
131	37	120	90	260	67	188	0
132	28	110	75	252	67	149	0
133	44	120	90	336	72	175	0
134	35	100	70	216	66	126	0
135	41	100	65	208	69	165	0
136	29	120	80	352	68	160	0
137	46	125	90	346	63	155	0
138	55	148	90	259	71	140	0
139	32	100	70	290	70	181	0
140	40	125	90	239	67	178	0
141	61	154	80	333	66	141	0
142	29	100	60	173	69	143	0
143	52	110	80	253	70	139	0
144	25	120	80	156	67	136	0
145	27	110	70	156	67	150	0
146	27	130	90	208	69	185	0
147	53	130	80	218	73	185	0
148	42	120	80	172	68	161	0
149	64	115	75	357	67	180	0
150	27	130	90	178	74	198	0
151	55	110	70	283	70	128	1
152	33	120	80	275	67	177	0
153	58	110	80	187	71	224	0

154	51	110	80	282	71	160	0
155	37	134	80	282	71	181	0
156	47	120	70	254	65	136	0
157	49	120	90	273	71	245	0
158	46	150	104	328	71	187	0
159	40	110	70	244	70	161	1
160	26	110	80	277	74	190	0
161	28	110	70	195	73	180	0
162	23	108	68	206	71	165	0
163	52	125	90	327	65	147	0
164	42	110	75	246	67	146	0
165	27	115	88	203	70	182	0
166	29	120	80	185	72	187	0
167	43	122	78	224	66	128	0
168	34	110	70	246	68	140	0
169	40	120	80	227	67	163	0
170	28	115	70	229	70	144	0
171	30	115	70	214	71	150	0
172	34	108	75	206	71	137	0
173	26	120	80	173	67	141	0
174	34	135	90	248	70	141	0
175	35	100	70	222	73	190	0
176	34	110	60	230	73	167	0
177	45	100	70	219	69	159	0
178	47	120	80	239	67	157	0
179	54	134	90	258	66	170	0
180	30	110	80	190	69	132	0
181	29	110	80	252	68	155	0
182	47	110	80	253	72	178	0
183	37	120	88	172	71	168	0
184	43	138	94	320	65	159	1
185	31	124	90	166	67	160	0

186	48	140	85	266	71	165	0
187	34	110	80	176	71	194	0
188	42	130	80	271	70	191	1
189	49	122	94	295	73	198	0
190	50	130	85	271	68	212	1
191	42	120	80	259	64	147	0
192	50	140	95	178	68	173	1
193	60	160	100	317	72	206	0
194	27	124	80	192	70	190	0
195	29	120	85	187	68	181	0
196	29	110	70	238	72	143	0
197	49	112	78	283	64	149	0
198	49	100	70	264	70	166	0
199	50	128	92	264	70	176	0
200	31	105	68	193	67	141	0

Βιβλιογραφία

- Bernardo, J.M. and Smith, A.F.M (1994). *Bayesian Theory*. New York: Wiley.
- Casella, G. and George, E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician* **46**, 167-174.
- Chen, M.H., Huang, L., Ibrahim, J.G. and Kim, S. (2008). Bayesian Variable Selection and Computation for Generalized Linear Models with Conjugate Priors. *Bayesian Analysis* **3**, 585-614.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* **40**, 327-335.
- Chipman, H., George, E. and McCulloch, R. (2001). The Practical Implementation of Bayesian Model Selection. In *IMS Lecture Notes - Monograph Series* **38**, 65-134.
- Dixon, W. J. and Massey, F.J. (1983). *Introduction to Statistical Analysis*. New York: McGraw-Hill, the fourth edition.
- Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *J.R. Statist. Soc. Ser. B* **57**, 45-97.
- Fernández, C., Ley, E. and Steel, M.F.J. (2001). Benchmark Priors for Bayesian Model Averaging. *J. Econometrics* **100**, 381-427.
- Furnival, G.M. and Wilson, R.W. (1974). Regression by Leaps and Bounds. *Technometrics* **16**, 499-511.
- Gamerman, D. and Lopez, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC Texts in Statistical Science, second edition.
- George, E. and McCulloch, R. (1993). Variable Selection via Gibbs Sampling. *J. Amer. Statist. Assoc.* **88**, 881-889.
- Hand, D. J. (1981). Branch and Bound in Statistical Data Analysis. *The Statistician* **30**, 1-13.
- Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial (with discussion). *Statist. Sc.* **14**, 382-414.
- Hoeting, J.A., Raftery, A.E. and Madigan, D. (1996). A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression. *J. Comput. Statist.* **22**, 251-271.
- Kokolakis, G., Nanopoulos, Ph. and Fouskakis, D. (2006). Bregman Divergences in (mxk)-Partitioning Problem. *Comput. Statistics and Data Analysis* **51**, 668-678.

- Madigan, D. and Raftery, A.E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *J.Amer. Statist. Assoc.* **89**, 1535-1546.
- Madigan, D. and York, J. (1995). Bayesian Graphical Models for Discrete Data. *Internat. Statist. Rev.* **63**, 215-232.
- Press, J.S. (2003). *Subjective and Objective Bayesian Statistics. Principles Models and Applications*. Wiley.
- Raftery, A.E., Madigan, D. and Hoeting, J. (1997). Bayesian Model Averaging for Linear Regression Models. *J. Amer. Statist. Assoc.* **92**, 179-191.
- Raftery, A.E., Madigan, D. and Volinsky, C.T. (1995). Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance (with discussion). *Bayesian Statistics 5* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.), 323-349. Oxford University Press.
- Smith, A.F.M. and Roberts, G.O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion). *J. Roy. Statist. Soc. Ser. B*, **55**, 3-23.
- Tierney, L. and Kadane, J.D. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *J. Amer. Statist. Assoc.* **81**, 82-86.
- Κοκολάκης, Γ. και Σπηλιώτης, Ι. (2010). *Θεωρία Πιθανοτήτων και Στατιστική με Εφαρμογές*. Εκδόσεις Συμμεών.

Διαδικτυακοί Τόποι

1. <http://cran.r-project.org/web/packages/BMA/BMA.pdf>
2. <http://www.math.ntua.gr/~fouskakis/>