

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ
ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΓΙΑ ΤΗΝ ΑΠΟΚΤΗΣΗ ΤΟΥ Δ.Μ.Δ.Ε. ΣΤΙΣ
«ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ»
ΣΤΗΝ ΚΑΤΕΥΘΥΝΣΗ
«ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΠΙΘΑΝΟΤΗΤΕΣ»
ΜΕ ΤΙΤΛΟ

*«Μπεϋζιανή Επιλογή μοντέλου με χρήση
Ενδογενών εκ των προτέρων κατανομών»*

ΠΕΤΡΑΚΗΣ ΝΙΚΟΛΑΟΣ

ΑΘΗΝΑ 2016

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω αρχικά τον καθηγητή μου **κ. Δημήτρη Φουσκάκη**, Αναπληρωτή Καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου και κύριου επιβλέποντά μου. Θα ήθελα να τον ευχαριστήσω για την εμπιστοσύνη που μου έδειξε, την καθοδήγηση του και την βοήθειά του κατά την διάρκεια των μεταπτυχιακών μου σπουδών και της εκπόνησης της παρούσας διπλωματικής εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τον καθηγητή μου **κ. Γεώργιο Ηλιόπουλο**, Καθηγητή του Πανεπιστημίου Πειραιώς. Θα ήθελα τον ευχαριστήσω αρχικά για την βοήθεια που μου προσέφερε σχετικά με την επίλυση μίας εφαρμογής αυτής της εργασίας, όπως επίσης και για την καθοδήγηση του κατά την διάρκεια των προπτυχιακών μου σπουδών, καθώς ήταν ο πρώτος που μου δίδαξε Μπεϋζιανή στατιστική και με προετοίμασε κατάλληλα για να μπορέσω να προχωρήσω στις μεταπτυχιακές σπουδές μου.

Κλείνοντας θα ήθελα να ευχαριστήσω και την φίλη μου **Εύα Σπυρίδη** για την υποστήριξη της καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Πετράκης Νίκος

Απρίλιος 2016

Περιεχόμενα

Περιεχόμενα..... 3

Περίληψη..... 5

Κεφάλαιο 1. Μπεϋζιανή Στατιστική

1.1 Το θεώρημα Μπεϋες.....7

1.2 Εκ των προτέρων κατανομές.....11

1.2.1 Πληροφοριακές εκ των προτέρων κατανομές.....11

1.2.1.1 Συζυγείς οικογένειες εκ των προτέρων κατανομών.....11

1.2.1.2 Εκ των προτέρων κατανομές βασισμένες σε δυνάμεις πιθανοφάνειας.....13

1.2.2 Μη-πληροφοριακές κατανομές.....14

1.2.2.1 Εκ των προτέρων κατανομή του Jeffreys.....15

1.2.2.2 Εκ των προτέρων κατανομή αναφοράς.....20

1.2.3 Ιεραρχικές εκ των προτέρων κατανομές.....20

1.3 Επίλογος 1^ο Κεφαλαίου.....23

Κεφάλαιο 2. Μπεϋζιανή επιλογή μοντέλου

2.1 Ο έλεγχος υποθέσεων στην κλασσική στατιστική.....24

2.2 Η σύγκριση μοντέλων στην Μπεϋζιανή στατιστική.....26

2.3 Μπεϋζιανή Στάθμιση Μοντέλων.....30

2.4 Η επιλογή μεταβλητών στην κλασσική στατιστική.....32

2.5 Η επιλογή μεταβλητών στην Μπεϋζιανή στατιστική.....34

2.6 Ασυμπτωτικές Προσεγγίσεις του Παράγοντα Μπεϋες.....35

2.6.1 Μέθοδος του Laplace.....36

2.6.2 Κριτήριο Schwarz.....37

2.6.3 Δειγματοληψία Σπουδαιότητας και Monte Carlo ολοκλήρωση.....38

2.6.4 Προσομοίωση από την εκ των υστέρων κατανομή.....39

2.7 Επίλογος 2^ο Κεφαλαίου.....40

Κεφάλαιο 3. Παραλλαγές του Παράγοντα Μπεϋες

3.1 Το παράδοξο των Lindley-Bartlett.....41

3.2 Χρήση μη-κατάλληλων πληροφοριακών κατανομών.....43

3.3 Ο εκ των υστέρων Παράγοντας Μπεϋες.....44

3.4 Ο Κλασματικός Παράγοντας Μπεϋες.....45

3.5 Ο Ενδογενής Παράγοντας Μπεϋες.....48

3.5.1 Αριθμητικός και Γεωμετρικός Ενδογενής Παράγοντας Μπεϋες.....50

3.5.2 Αναμενόμενος Αριθμητικός και Γεωμετρικός Ενδογενής Παράγοντας Μπεϋες.....51

3.5.3 Πλεονεκτήματα και Μειονεκτήματα του Ενδογενή Παράγοντα Μπεϋες.....	52
3.6 Επίλογος 3 ^{ου} Κεφαλαίου.....	53

Κεφάλαιο 4. Ενδογενείς εκ των προτέρων κατανομές

4.1 Ενδογενείς εκ των προτέρων κατανομές.....	54
4.1.1 Ενδογενείς εξισώσεις.....	55
4.2 Σύγκριση εμφωλευμένων μοντέλων.....	59
4.3 Σύγκριση μη-εμφωλευμένων μοντέλων.....	62
4.4 Αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές.....	65
4.4.1 Επιλογή της προβλεπτικής κατανομές.....	66
4.4.1.1 Επιλογή της μη-πληροφοριακής περιθώριας κατανομής του μοντέλου βάσης.....	66
4.4.1.2 Χρήση εμπειρικής κατανομής για προβλεπτική κατανομή.....	68
4.5 Πλεονεκτήματα EPPs.....	69
4.6 Ειδική περίπτωση ισοδυναμίας EPP και Ενδογενών εκ των προτέρων κατανομών.....	69
4.7 Επίλογος 4 ^{ου} κεφαλαίου.....	70

Κεφάλαιο 5. Εφαρμογές

5.1 Υπολογισμός Ενδογενών Παραγόντων Μπεϋες σε πραγματικά δεδομένα.....	71
5.1.1 Εφαρμογή Δεδομένων του Proschan (1963).....	75
5.2 Έλεγχος παραμέτρου Εκθετικής κατανομής.....	76
5.2.1 Παράδειγμα με προσομοιωμένα δεδομένα.....	83
5.3 Έλεγχος μέσης τιμής Κανονικής κατανομής με γνωστή διακύμανση.....	89
5.3.1 Παράδειγμα με προσομοιωμένα δεδομένα.....	93
5.4 Έλεγχος μέσης τιμής Κανονικής κατανομής με άγνωστη διακύμανση.....	99
5.4.1. Υπολογισμός Ενδογενών εκ των προτέρων κατανομών.....	99
5.4.2. Υπολογισμός Αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών.....	109
5.4.3. Παράδειγμα με προσομοιωμένα δεδομένα.....	114
Επίλογος.....	119
Παράρτημα 1.....	120
Βιβλιογραφία.....	124

Περίληψη

Η Μπεϋζιανή Στατιστική αποτελεί μία διαφορετική προσέγγιση της στατιστικής επιστήμης, η οποία γνωρίζει σημαντική ανάπτυξη τις τελευταίες δύο δεκαετίες, την οποία την οφείλει στην εμφάνιση των αλγορίθμων MCMC. Ένα από τα κυρίαρχα προβλήματα της Στατιστικής είναι η επιλογή ενός μοντέλου το οποίο θα είναι σε θέση να προσαρμοστεί στα δεδομένα τα οποία διαθέτουμε. Η παρούσα διπλωματική εργασία επικεντρώνεται στη μελέτη και παρουσίαση των μεθόδων επίλυσης του προβλήματος επιλογής μοντέλου μέσα από την μπεϋζιανή οπτική, εστιάζοντας στη χρήση ενδογενών και αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών. Η διάρθρωση της θεματολογίας που αναπτύσσεται χωρίζεται σε πέντε κεφάλαια και ένα παράρτημα.

Το πρώτο κεφάλαιο αποτελεί μία εισαγωγή στην Μπεϋζιανή Στατιστική και στη σημασία της εκ των προτέρων κατανομής. Επίσης μελετώνται και ορισμένες απλές εφαρμογές ώστε να γίνει κατανοητή η μετάβαση από την εκ των προτέρων στην εκ των υστέρων κατανομή.

Στο δεύτερο κεφάλαιο παρουσιάζονται οι βασικές έννοιες και διαδικασίες της επιλογής μοντέλου καθώς και επεξηγηματικών μεταβλητών σε προβλήματα παλινδρόμησης, κριτήρια αξιολόγησης και επιλογής μοντέλων καθώς επίσης και η σύγκριση της μπεϋζιανής και της κλασσικής σχολής σχετικά με το εν λόγω πρόβλημα.

Στο τρίτο κεφάλαιο παρουσιάζονται οι αδυναμίες των παραγόντων Μπεϋες στην περίπτωση όπου χρησιμοποιούμε εκ των προτέρων κατανομές οι οποίες δεν είναι κατάλληλες είτε έχουν πολύ μεγάλη μεταβλητότητα και παρουσιάζονται εναλλακτικές μορφές παραγόντων Μπεϋες. Συγκεκριμένα μελετάται εκτενέστερα ο ενδογενής παράγοντας Μπεϋες προκειμένου στο επόμενο κεφάλαιο να προχωρήσουμε στον κύριο πυρήνα της παρούσας διπλωματικής εργασίας.

Το τέταρτο κεφάλαιο όπως αναφέραμε αποτελεί τον κύριο πυρήνα της παρούσας διπλωματικής εργασίας, καθώς μελετάται και παρουσιάζεται αναλυτικά ο καθορισμός των ενδογενών εκ των προτέρων κατανομών, όπως επίσης και των αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών, καθώς επίσης προτερήματα και μειονεκτήματα αυτών.

Στο πέμπτο και τελευταίο κεφάλαιο μελετώνται τέσσερις εφαρμογές της παραπάνω μεθοδολογίας για πραγματικά και προσομοιωμένα δεδομένα. Μελετάμε προβλήματα σύγκρισης εμφωλευμένων και μη εμφωλευμένων μοντέλων και στις περιπτώσεις στις οποίες οι υπολογισμοί δεν γίνονται αναλυτικά παρουσιάζονται εναλλακτικοί τρόποι υπολογισμών. Στο τέλος της παρούσας διπλωματικής εργασίας παρουσιάζονται οι αλγόριθμοι στην στατιστική γλώσσα προγραμματισμού R, που χρησιμοποιήθηκαν για τη δημιουργία διαγραμμάτων και την διενέργεια των εφαρμογών του πέμπτου κεφαλαίου.

Κεφάλαιο 1

Μπεϋζιανή Στατιστική

1.1 Το Θεώρημα Μπεϋες

Η *Μπεϋζιανή Στατιστική (Bayesian Statistics)* αποτελεί μία ιδιαίτερη προσέγγιση της Στατιστικής Επιστήμης. Πατέρας της Μπεϋζιανής Στατιστικής είναι ο Thomas Bayes. Ο Thomas Bayes γεννημένος το 1702 στο Λονδίνο, ήταν κληρικός και στατιστικός ο οποίος χρησιμοποίησε επαγωγικά την θεωρία πιθανοτήτων για να δημιουργήσει έναν μαθηματικό τύπο ώστε να εξάγει συμπεράσματα μέσω πιθανοτήτων. Αυτός ο μαθηματικός τύπος καλείται **Θεώρημα Μπεϋες (Bayes theorem)**.

Έστω χώρος πιθανότητας (Ω, Σ, P) και A, B δύο ενδεχόμενα του δειγματοχώρου Ω με θετικές πιθανότητες. Τότε από την Θεωρία Πιθανοτήτων, η δεσμευμένη πιθανότητα του ενδεχομένου A δοθέντος του ενδεχομένου B θα είναι

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Κάνοντας χρήση του Θεωρήματος Ολικής Πιθανότητας παίρνουμε ότι

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

όπου A^c είναι το συμπληρωματικό ενδεχόμενο του A . Με αυτόν τον τρόπο καταλήγουμε στην απλούστερη μορφή του Θεωρήματος Μπεϋες η οποία είναι

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

Η ουσία του Θεωρήματος Μπεϋες βρίσκεται στην προσθήκη πληροφορίας που εκχωρείται από την πραγματοποίηση του ενδεχομένου B , στην πραγματοποίηση του ενδεχομένου A . Επομένως η πιθανότητα $P(A)$ που είχαμε αρχικά, αναβαθμίζεται στην πιθανότητα $P(A|B)$. Αν αντιστοιχίσουμε τα παραπάνω ενδεχόμενα A και B , με τις τυχαίες μεταβλητές Y και X αντίστοιχα, και τις πιθανότητες με τις αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας (π.χ. στην συνεχή περίπτωση) παίρνουμε

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_y f_{X|Y}(x|y)f_Y(y)dy}.$$

Η βασική διαφορά της Μπεϋζιανής από την Κλασσική Στατιστική βρίσκεται στην αντιμετώπιση της άγνωστης παραμέτρου θ . Στην Κλασσική Στατιστική η παράμετρος θ είναι μία σταθερή αλλά άγνωστη προς εκτίμηση ποσότητα. Στην Μπεϋζιανή Στατιστική η παράμετρος θ αντιμετωπίζεται ως μία τυχαία μεταβλητή η οποία έχει την

δική της κατανομή. Αυτή η βασική φιλοσοφία είναι και ο κορμός της Μπεϋζιανής Στατιστικής, δηλαδή την οποιαδήποτε πληροφορία ή άγνοια ή αβεβαιότητα έχουμε γύρω από την παράμετρο θ , την εκφράζουμε μέσω κάποιας κατανομής πιθανότητας και προκύπτουν συμπεράσματα τα οποία βασίζονται στη θεωρία πιθανοτήτων

Έστω δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ με κατανομή που εξαρτάται από κάποια παράμετρο θ και $f(\mathbf{x}|\theta)$ η *συνάρτηση πιθανοφάνειας (likelihood function)* του θ . Η κατανομή της τυχαίας μεταβλητής θ καλείται εκ των *προτέρων κατανομή (prior distribution)* και αντιπροσωπεύει την γνώση μας πριν την συλλογή των δεδομένων \mathbf{x} ή την προσωπική μας πεποίθηση σχετικά με την κατανομή της παραμέτρου θ και συμβολίζεται με $\pi(\theta)$. Στην Μπεϋζιανή Στατιστική η εξαγωγή συμπερασμάτων βασίζεται στη *εκ των υστέρων κατανομή (posterior distribution)* της παραμέτρου θ και συμβολίζεται $\pi(\theta|\mathbf{x})$. Ο υπολογισμός της εκ των υστέρων κατανομής βασίζεται στο Θεώρημα Μπεϋες και δίνεται από τον τύπο

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta}.$$

Αν η τυχαία μεταβλητή θ είναι διακριτή, τότε ο παρονομαστής θα αντικατασταθεί από το άθροισμα

$$\sum_{\theta} f(\mathbf{x}|\theta)\pi(\theta).$$

Η ποσότητα στον παρονομαστή είναι η περιθωριακή κατανομή των δεδομένων \mathbf{x} όπου προκύπτει είτε αθροίζοντας είτε ολοκληρώνοντας ως προς την παράμετρο θ και (στην συνεχή περίπτωση π.χ.) δίνεται από την σχέση

$$m(\mathbf{x}) = \int_{\theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta.$$

Η ποσότητα $m(\mathbf{x})$ καλείται επίσης και σταθερά κανονικοποίησης, καθώς δεν φέρει κάποια πληροφορία σχετικά με την παράμετρο θ αφού είναι η περιθωριακή κατανομή των δεδομένων $\mathbf{x} = (x_1, x_2, \dots, x_n)$, έχοντας ως αποτέλεσμα την αντιμετώπισή της ως σταθερό αριθμό. Παρόλα αυτά όπως θα δούμε και στα επόμενα κεφάλαια αποτελεί σημαντικό μέτρο για την σύγκριση στατιστικών μοντέλων.

Επομένως αφού η περιθωρια κατανομή των δεδομένων \mathbf{x} δεν παρέχει κάποια πληροφορία για το θ στην εκ των υστέρων κατανομή, το Θεώρημα Μπεϋες συναντάται και στην μορφή

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta).$$

Δηλαδή η εκ των υστέρων κατανομή είναι ανάλογη της πιθανοφάνειας του θ και της εκ των προτέρων κατανομής για το θ , καθώς μόνο αυτές οι δύο συναρτήσεις περιέχουν πληροφορία για την εκ των υστέρων κατανομή του θ .

Επομένως η εκ των προτέρων κατανομή, σε συνδυασμό με την πληροφορία που παρέχουν τα δεδομένα, αναβαθμίζεται στην εκ των υστέρων κατανομή για το θ .

Παράδειγμα 1.1

Έστω παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενες από τυχαίο δείγμα μεγέθους n όπου ακολουθεί την εκθετική κατανομή με άγνωστη παράμετρο $\theta > 0$ και συνάρτηση πυκνότητας πιθανότητας

$$f(x_i|\theta) = \theta e^{-\theta x_i} I_{[0,\infty)}(x_i).$$

Η εκ των προτέρων κατανομή για την παράμετρο θ θα είναι έστω μία Γάμμα με γνωστές παραμέτρους $\alpha, \beta > 0$ και συνάρτηση πυκνότητας πιθανότητας

$$\pi(\theta) = \frac{\theta^{\alpha-1} \beta^\alpha e^{-\theta\beta}}{\Gamma(\alpha)} I_{(0,\infty)}(\theta).$$

Η συνάρτηση πιθανοφάνειας του θ υπολογίζεται ως

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} I_{[0,\infty)}(x_i) = \theta^n e^{-\theta n\bar{x}} I_{[0,\infty)}^n(\mathbf{x}).$$

Εφαρμόζοντας το Θεώρημα Μπεϋες η εκ των υστέρων κατανομή για την παράμετρο θ υπολογίζεται

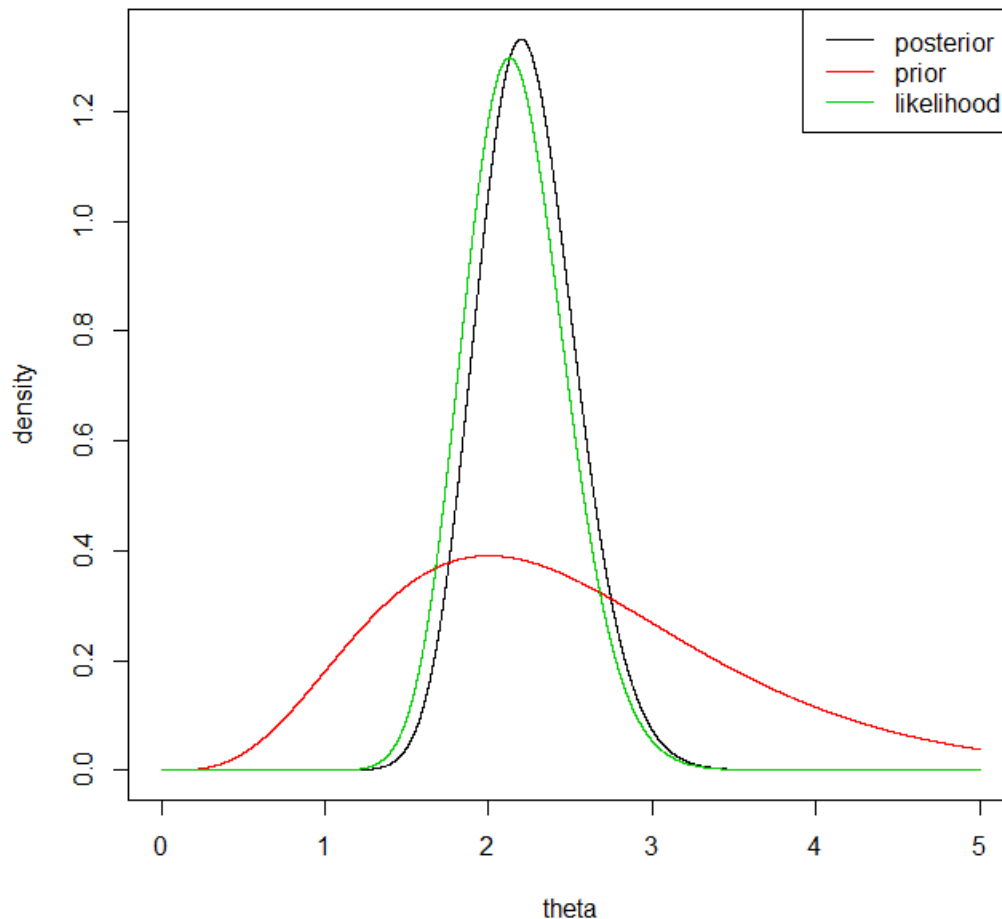
$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\ &= \theta^n e^{-\theta n\bar{x}} I_{[0,\infty)}^n(\mathbf{x}) \frac{\theta^{\alpha-1} \beta^\alpha e^{-\theta\beta}}{\Gamma(\alpha)} I_{(0,\infty)}(\theta) \\ &\propto \theta^n e^{-\theta n\bar{x}} \theta^{\alpha-1} e^{-\theta\beta} I_{(0,\infty)}(\theta) \\ &= \theta^{n+\alpha-1} e^{-\theta(n\bar{x}+\beta)} I_{(0,\infty)}(\theta). \end{aligned}$$

Παρατηρούμε ότι ο πυρήνας αυτής της κατανομής προέρχεται από μία κατανομή γάμμα με παραμέτρους $\alpha' = n + \alpha$ και $\beta' = n\bar{x} + \beta$, επομένως για να γίνει συνάρτηση πυκνότητας πιθανότητας αρκεί να πολλαπλασιάσουμε με την σταθερά κανονικοποίησης, η οποία προκύπτει από τον υπολογισμό της περιθωριακής κατανομής των δεδομένων. Επομένως η εκ των υστέρων πυκνότητα πιθανότητας της παραμέτρου θ δίνεται από

$$\pi(\theta|\mathbf{x}) = \frac{\theta^{n+\alpha-1}(n\bar{x} + \beta)^{n+\alpha} e^{-\theta(n\bar{x}+\beta)}}{\Gamma(n + \alpha)} I_{(0,\infty)}(\theta).$$

Εφαρμογή

Προσομοιώνουμε 50 παρατηρήσεις x_i από την εκθετική κατανομή με παράμετρο $\theta = 6$ και επιλέγουμε ως εκ των προτέρων κατανομή για την παράμετρο θ μία κατανομή γάμμα με παραμέτρους $\alpha = 5$ και $\beta = 2$. Στο **Διάγραμμα 1.1** παρουσιάζονται η εκ των προτέρων κατανομή (κόκκινο χρώμα) και η συνάρτηση πιθανοφάνειας (πράσινο χρώμα) καθώς και πώς διαμορφώνεται βάσει αυτών η εκ των υστέρων κατανομή (μαύρο χρώμα).



Διάγραμμα 1.1 : Διάγραμμα απεικόνισης σχηματισμού της εκ των υστέρων κατανομής προσομοιωμένων τιμών x_i

Ο κώδικας που χρησιμοποιήθηκε για την παραγωγή του παραπάνω διαγράμματος μπορεί να βρεθεί στο **Παράρτημα 1**.

1.2 Εκ των προτέρων κατανομές

Όπως αναφέραμε η εκ των υστέρων κατανομή υπολογίζεται χρησιμοποιώντας την συνάρτηση πιθανοφάνειας και την εκ των προτέρων κατανομή της παραμέτρου θ . Στο παραπάνω παράδειγμα διαλέξαμε ως εκ των προτέρων κατανομή για το θ μία γάμμα με γνωστές παραμέτρους α και β , και καταλήξαμε πάλι σε μία κατανομή γάμμα με κάποιες καινούριες παραμέτρους. Αν διαλέγαμε κάποια άλλη εκ των προτέρων κατανομή για την παράμετρο θ , για παράδειγμα μία Weibull, θα καταλήγαμε σε κάποια άλλη εκ των υστέρων κατανομή. Όπως αναφέραμε η επιλογή αυτή βασίζεται είτε στη γνώση που έχουμε για την παράμετρο θ είτε σε κάποια προσωπική μας πεποίθηση. Η εκ των προτέρων κατανομή μπορεί να έχει συναρτησιακή μορφή είτε να είναι κάποια σταθερή πιθανότητα.

Έστω ότι θέλουμε να εκτιμήσουμε το ποσοστό θ επιτυχημένων εγχειρήσεων ανοιχτής καρδιάς που πραγματοποιούνται στο Ωνάσειο. Οι επιλογές μας για την εκ των προτέρων κατανομή της παραμέτρου θ είναι πάρα πολλές. Αρχικά μπορούμε να πάρουμε το ποσοστό των επιτυχημένων εγχειρήσεων του τελευταίου εξαμήνου ως εκ των προτέρων πληροφορία, μπορούμε να πούμε ότι η εκ των προτέρων κατανομή της παραμέτρου θ θα είναι μία κατανομή βήτα με συγκεκριμένες παραμέτρους ή μπορούμε επίσης να βασιστούμε στην προσωπική μας πεποίθηση και να ισχυριστούμε ότι η εκ των προτέρων πιθανότητα επιτυχίας θα παίρνει μία συγκεκριμένη τιμή.

Όπως βλέπουμε η επιλογή της εκ των προτέρων κατανομής καθορίζει σημαντικά την πορεία της ανάλυσης στην Μπεϋζιανή Στατιστική και παίζει πολύ σημαντικό ρόλο. Σκοπός μας είναι να βρούμε την κατάλληλη εκ των προτέρων κατανομή και να την ενσωματώσουμε στην στατιστική ανάλυση ώστε να πάρουμε όσο το δυνατόν περισσότερο αξιόπιστα αποτελέσματα.

Γενικά οι εκ των προτέρων κατανομές μπορούν να χωριστούν σε αρκετές κατηγορίες, όμως μία βασική κατηγοριοποίηση τους είναι σε *πληροφοριακές (informative priors)* και *μη πληροφοριακές (non informative priors)* εκ των προτέρων κατανομές.

1.2.1 Πληροφοριακές εκ των προτέρων κατανομές

Αυτή η κατηγορία εκ των προτέρων κατανομών αφορά περιπτώσεις όπου υπάρχει διαθέσιμη πληροφορία σχετικά με την άγνωστη κατανομή ή πιθανότητα της παραμέτρου θ και μπορεί να οριστεί με ακρίβεια. Στην συνέχεια θα παρουσιαστούν κάποιες βασικές κατηγορίες εκ των προτέρων κατανομών στις οποίες εκχωρείται πληροφορία σχετικά με την κατανομή της παραμέτρου θ σύμφωνα με συγκεκριμένες μεθοδολογίες.

1.2.1.1 Συζυγείς οικογένειες εκ των προτέρων κατανομών

Πριν την εμφάνιση των αλγορίθμων MCMC η μπεϋζιανή συμπερασματολογία θεωρείτο πρακτικά δύσχρηστη λόγω των υπολογισμών των εκ των υστέρων

κατανομών. Πρακτικά αυτό το πρόβλημα αντιμετωπίζεται με την χρήση *συζυγών εκ των προτέρων κατανομών (conjugate priors)*.

Αν F οικογένεια κατανομών από την οποία προέρχεται ένα τυχαίο δείγμα x και P είναι μία οικογένεια κατανομών για την παράμετρο θ , τότε λέμε ότι η οικογένεια P είναι συζυγής της F όταν

$$\pi(\theta|x) \in P \text{ για κάθε } f(\cdot|\theta) \in F \text{ και } \pi(\cdot) \in P.$$

Η συγκεκριμένη κατηγορία εκ των προτέρων κατανομών ήταν αρκετά διαδεδομένη τα παλαιότερα χρόνια, καθώς η εξεύρεση της εκ των υστέρων κατανομής καθώς και διάφοροι άλλοι υπολογισμοί ήταν αρκετά απλουστευμένοι. Επίσης όπως αναφέρεται στο βιβλίο των Carlin & Stern (2004) βασικό πλεονέκτημα των συζυγών εκ των προτέρων κατανομών, είναι η εναλλακτική ερμηνεία τους ως ένα επιπλέον δείγμα παρατηρήσεων. Το Παράδειγμα 1.1 που αναπτύχθηκε παραπάνω αποτελεί εφαρμογή της συγκεκριμένης κατηγορίας εκ των προτέρων, καθώς ορίσαμε για εκ των προτέρων κατανομή μία γάμμα και καταλήξαμε σε εκ των υστέρων κατανομή επίσης γάμμα.

Άμεση εφαρμογή των συζυγών εκ των προτέρων κατανομών βρίσκουμε σε κατανομές οι οποίες ανήκουν στην εκθετική οικογένεια κατανομών. Όπως αναφέρεται στο σύγγραμμα των Carlin & Stern (2004), μία κλάση κατανομών F καλείται κλάση εκθετικής οικογένειας κατανομών αν όλες οι κατανομές πιθανότητας που ανήκουν σε αυτήν είναι της μορφής

$$f(x_i|\theta) = p(x_i)g(\theta)e^{\varphi(\theta)^T u(x_i)}.$$

Οι όροι $\varphi(\theta)$ και $u(x_i)$ αποτελούν διανύσματα ίδιας διάστασης με το διάνυσμα των παραμέτρων θ , όπου ο όρος $\varphi(\theta)$ καλείται *‘φυσική παράμετρος’* της οικογένειας F . Για ένα τυχαίο δείγμα παρατηρήσεων $x = (x_1, x_2, \dots, x_n)$ το οποίο ορίζεται από μία κατανομή $f \in F$, η συνάρτηση πιθανοφάνειας του τυχαίου δείγματος θα δίνεται από

$$f(x|\theta) = \left[\prod_{i=1}^n p(x_i) \right] g(\theta)^n \exp \left[\varphi(\theta)^T \sum_{i=1}^n u(x_i) \right].$$

Για δοσμένο διάνυσμα δεδομένων x μεγέθους n , η συνάρτηση πιθανοφάνειας εκφράζεται συναρτήσει της παραμέτρου θ ως

$$f(x|\theta) \propto g(\theta)^n \exp[\varphi(\theta)^T t(x)]$$

όπου η ποσότητα $t(x) = \sum_{i=1}^n u(x_i)$ είναι η επαρκής και πλήρης στατιστική συνάρτηση για την παράμετρο θ . Εάν η εκ των προτέρων κατανομή είναι της μορφής

$$\pi(\theta) \propto g(\theta)^\eta \exp[\varphi(\theta)^T \nu]$$

τότε η εκ των υστέρων κατανομή θα δίνεται από

$$\pi(\boldsymbol{\theta}|\mathbf{x}) \propto g(\boldsymbol{\theta})^{\eta+n} \exp[\boldsymbol{\varphi}(\boldsymbol{\theta})^T (\boldsymbol{v} + t(\mathbf{x}))]$$

όπου υποδεικνύεται ότι η συγκεκριμένη επιλογή εκ των προτέρων κατανομής είναι μία συζυγής εκ των προτέρων κατανομή.

1.2.1.2 Εκ των προτέρων κατανομές βασισμένες σε δυνάμεις πιθανοφάνειας

Όπως αναφέραμε οι εκ των προτέρων κατανομές αντιπροσωπεύουν είτε κάποια πεποίθηση μας σχετικά με την παράμετρο $\boldsymbol{\theta}$, είτε κάποια εκ των προτέρων γνώση που έχουμε για αυτή. Οι Chen & Ibrahim (2000) εισήγαγαν μία γενικευμένη κλάση εκ των προτέρων κατανομών για μοντέλα παλινδρόμησης, η οποία ονομάζεται **δυναμική εκ των προτέρων κατανομή (power-priors)** και χρησιμοποιείται σε περιπτώσεις όπου η επιλογή της εκ των προτέρων κατανομής βασίζεται σε προηγούμενη γνώση. Για να προχωρήσουμε στον ορισμό των δυναμικών εκ των προτέρων κατανομών, θα καλούμε την προηγούμενη μας γνώση ως **ιστορικά δεδομένα**. Η ιδέα στην οποία βασίζεται η μεθοδολογία, είναι η ύπαρξη μίας παραμέτρου α_0 με $0 \leq \alpha_0 \leq 1$, η οποία είναι η δύναμη στην οποία θα υψώνεται η συνάρτηση πιθανοφάνειας, προκειμένου να καθορισθεί η συμβολή των ιστορικών δεδομένων.

Έστω $\mathbf{D} = (n, \mathbf{Y}, \mathbf{X})$ τα δεδομένα της παρούσας έρευνας, όπου n το μέγεθος δείγματος, \mathbf{Y} ο $n \times 1$ πίνακας απόκρισης και \mathbf{X} ο $n \times p$ πίνακας των εξηγηματικών μεταβλητών και ορίζουμε την συνάρτηση πιθανοφάνειας της παρούσας έρευνας ως $L(\boldsymbol{\theta}|\mathbf{D})$ όπου $\boldsymbol{\theta}$ το διάνυσμα των παραμέτρων. Υποθέτουμε ότι διαθέτουμε ιστορικά δεδομένα μιας προηγούμενης έρευνας τα οποία συμβολίζονται $\mathbf{D}_0 = (n_0, \mathbf{Y}_0, \mathbf{X}_0)$ και $\pi_0(\boldsymbol{\theta}|\cdot)$ η εκ των προτέρων κατανομή της παραμέτρου $\boldsymbol{\theta}$, πριν την παρατήρηση των ιστορικών δεδομένων \mathbf{D}_0 . Η $\pi_0(\boldsymbol{\theta}|\cdot)$ καλείται η **αρχική εκ των προτέρων κατανομή (initial prior)** για την παράμετρο $\boldsymbol{\theta}$.

Ορίζουμε ως **δυναμική εκ των προτέρων κατανομή** της παραμέτρου $\boldsymbol{\theta}$ της παρούσας έρευνας την

$$\pi(\boldsymbol{\theta}|\mathbf{D}_0, \alpha_0) \propto [L(\boldsymbol{\theta}|\mathbf{D}_0)]^{\alpha_0} \pi_0(\boldsymbol{\theta}|c_0)$$

όπου c_0 παράμετρος από την οποία εξαρτάται η εκ των προτέρων κατανομή του αρχικού μοντέλου και α_0 μία παράμετρος κλίμακας, η οποία καθορίζει το βάρος των ιστορικών δεδομένων στην δυναμική εκ των προτέρων κατανομή.

Μία από τις βασικές ιδιότητες της παραμέτρου α_0 είναι ο καθορισμός του βάρους της ουράς της δυναμικής εκ των προτέρων κατανομής, καθώς όσο μικρότερες τιμές λαμβάνει τόσο πιο μεγάλη θα είναι και η ουρά της. Θέτοντας $\alpha_0 = 0$ είναι προφανές ότι τα ιστορικά δεδομένα δεν έχουν καμία απολύτως επίδραση στον καθορισμό της δυναμικής εκ των προτέρων κατανομής, η οποία συμπίπτει με

$$\pi(\boldsymbol{\theta}|\mathbf{D}_0, \alpha_0) \equiv \pi_0(\boldsymbol{\theta}|c_0).$$

Συνεπώς η παράμετρος α_0 δίνει στον ερευνητή την δυνατότητα να καθορίσει με ακρίβεια την επίδραση των ιστορικών δεδομένων στην παρούσα έρευνα. Το χαρακτηριστικό αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις όπου η παρούσα έρευνα εμφανίζει *ανομοιογένεια (heterogeneity)* συγκριτικά με προηγούμενες έρευνες ή σε περιπτώσεις όπου εμφανίζονται σημαντικές διαφορές στα μεγέθη των δειγμάτων από έρευνα σε έρευνα.

Οι Chen & Ibrahim (2000) εισήγαγαν την δυναμική εκ των προτέρων κατανομή ως μία ιεραρχική εκ των προτέρων κατανομή καθορίζοντας την κατανομή της παραμέτρου α_0 , λαμβάνοντας την μορφή

$$\pi(\boldsymbol{\theta}, \alpha_0 | \mathbf{D}_0) \propto [L(\boldsymbol{\theta} | \mathbf{D}_0)]^{\alpha_0} \pi_0(\boldsymbol{\theta} | c_0) \pi(\alpha_0 | \gamma_0)$$

όπου γ_0 ένα διάνυσμα υπερπαραμέτρων.

Στην εργασία τους χρησιμοποιούν ως εκ των προτέρων κατανομή για το α_0 μία κατανομή βήτα καθώς η ερμηνεία της ταιριάζει καλύτερα στον ρόλο τον οποίο διαδραματίζει η παράμετρος.

1.2.2 Μη πληροφοριακές κατανομές

Σε περιπτώσεις όπου δεν έχουμε κάποια εκ των προτέρων πληροφορία για την κατανομή της παραμέτρου $\boldsymbol{\theta}$, είτε θέλουμε κρατήσουμε αντικειμενική στάση απέναντι της, καταφεύγουμε σε εκ των προτέρων κατανομές οι οποίες έχουν πολύ μεγάλη ή και άπειρη διασπορά. Αυτές οι κατανομές καλούνται *μη πληροφοριακές εκ των προτέρων κατανομές (non informative priors, diffuse priors etc.)* για την παράμετρο $\boldsymbol{\theta}$. Δίνοντας μικρή εκ των προτέρων πληροφορία, η εκ των υστέρων πληροφορία διαμορφώνεται κυρίως από τις πληροφορίες που παρέχουν τα δεδομένα. Όπως και στις πληροφοριακές εκ των προτέρων κατανομές, οι μη πληροφοριακές εκ των προτέρων κατανομές μπορούν να χωριστούν σε δύο βασικές κατηγορίες. Οι εκ των προτέρων κατανομές οι οποίες ανήκουν στην πρώτη κατηγορία καλούνται *κατάλληλες (proper)* όπου βασικό τους χαρακτηριστικό είναι ότι πρόκειται για συναρτήσεις κατανομών όπου η συνάρτηση πυκνότητας τους ολοκληρώνει στην μονάδα. Παράδειγμα κατάλληλης μη-πληροφοριακής εκ των προτέρων κατανομής είναι η Κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση $\sigma = 1000$. Η δεύτερη κατηγορία αφορά εκ των προτέρων κατανομές οι οποίες καλούνται *μη-κατάλληλες (improper)* καθώς πρόκειται για συναρτήσεις οι οποίες δεν αποτελούν συναρτήσεις πυκνότητας πιθανότητας διότι δεν ολοκληρώνουν στην μονάδα. Παράδειγμα τέτοιας μη πληροφοριακής εκ των προτέρων κατανομής μπορεί να θεωρηθεί η

$$\pi(\theta) \propto \frac{1}{\theta} I_{(0,\infty)}(\theta).$$

Η συγκεκριμένη εκ των προτέρων κατανομή δεν ολοκληρώνει στην μονάδα, κάτι το οποίο επιτυγχάνεται πολλαπλασιάζοντας με μία άγνωστη σταθερά κανονικοποίησης c ,

η οποία θα μας εξασφαλίσει την κανονικοποίηση της. Συνεπώς η εκ των προτέρων κατανομή

$$\pi(\theta) = \frac{c}{\theta} I_{(0,\infty)}(\theta)$$

θα αποτελεί μία κατάλληλη εκ των προτέρων κανονική κατανομή. Όμως η σταθερά κανονικοποίησης c είναι αυθαίρετη, συνεπώς δεν μπορεί να αξιοποιηθεί στην μπεϋζιανή συμπερασματολογία και σε περιπτώσεις όπου χρειάζεται να επιλέξουμε κάποιο στατιστικό μοντέλο όπως θα δούμε στην συνέχεια, μπορεί να μας δημιουργήσει σημαντικά προβλήματα.

Παράδειγμα 1.2

Έστω οι παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενες από τυχαίο δείγμα μεγέθους n όπου ακολουθεί την κατανομή Poisson με παράμετρο $\theta > 0$ και εκ των προτέρων κατανομή για την παράμετρο θ την $\pi(\theta) \propto \frac{1}{\theta} I_{(0,\infty)}(\theta)$.

Η πιθανοφάνεια των δεδομένων \mathbf{x} υπολογίζεται

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} I_{Z_+}(x_i) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} I_{Z_+^n}(\mathbf{x})$$

Επομένως η εκ των υστέρων κατανομή για το θ υπολογίζεται

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\ &= \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} I_{Z_+^n}(\mathbf{x}) \frac{1}{\theta} I_{(0,\infty)}(\theta) \\ &\propto e^{-n\theta} \theta^{n\bar{x}-1} I_{(0,\infty)}(\theta). \end{aligned}$$

Ο παραπάνω πυρήνας της εκ των υστέρων κατανομής είναι από μία κατανομή γάμμα με παραμέτρους $a = n\bar{x}$ και $b = n$. Είναι εμφανές ότι η εκ των υστέρων κατανομή καθορίζεται αποκλειστικά από πληροφορίες που παρέχουν τα δεδομένα.

Στην συνέχεια θα εξετάσουμε κάποιες βασικές μη πληροφοριακές εκ των προτέρων κατανομές.

1.2.2.1 Εκ των προτέρων κατανομή του Jeffreys

Το 1961 ο Harold Jeffreys εισήγαγε μία μη πληροφοριακή εκ των προτέρων κατανομή η οποία διακρίνεται για την ευκολία ως προς τον υπολογισμό της. Η συγκεκριμένη αυτή κατανομή ονομάστηκε η *εκ των προτέρων κατανομή του Jeffreys (Jeffreys Prior)* και ορίζεται στην μονοδιάστατη περίπτωση ως

$$\pi(\theta) \propto [I(\theta)]^{\frac{1}{2}}$$

όπου $I(\theta)$ είναι η αναμενόμενη πληροφορία Fisher του αντίστοιχου μοντέλου που εξετάζουμε, η οποία υπολογίζεται ως

$$I(\theta) = -E_{x|\theta} \left[\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right] = E_{x|\theta} \left\{ \left[\frac{\partial}{\partial \theta} \log f(x|\theta) \right]^2 \right\}.$$

Στην πολυδιάστατη περίπτωση η εκ των προτέρων κατανομή του Jeffreys ορίζεται ως

$$\pi(\boldsymbol{\theta}) \propto [\det(I(\boldsymbol{\theta}))]^{\frac{1}{2}}$$

όπου $I(\boldsymbol{\theta})$ είναι ο πίνακας πληροφορίας Fisher όπου το ij -στοιχείο του δίνεται ως

$$I_{ij}(\boldsymbol{\theta}) = -E_{x|\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta) \right]$$

Βασικό χαρακτηριστικό της συγκεκριμένης εκ των προτέρων κατανομής είναι ότι παραμένει *αναλλοίωτη ως προς ένα προς ένα μετασχηματισμούς* του θ .

Απόδειξη (για την μονοδιάστατη περίπτωση)

Έστω $k = g(\theta)$ ένας ένα προς ένα μετασχηματισμός του θ τότε

$$\begin{aligned} I(\theta) &= E_{\theta} \left\{ \left[\frac{\partial}{\partial \theta} \log f(X|\theta) \right]^2 \right\} \\ &= E_{g(\theta)} \left\{ \left[\frac{\partial}{\partial g(\theta)} \log f(X|g(\theta)) \right]^2 \right\} [g'(\theta)]^2 \\ &= E_k \left\{ \left[\frac{\partial}{\partial k} \log f(X|k) \right]^2 \right\} \left(\frac{dk}{d\theta} \right)^2 \\ &= I(k) \left(\frac{dk}{d\theta} \right)^2 \end{aligned}$$

Επομένως η πληροφορία Fisher για το k θα είναι

$$I(k) = I(\theta) \left(\frac{d\theta}{dk} \right)^2 = I(g^{-1}(k)) \left(\frac{d\theta}{dk} \right)^2$$

άρα η εκ των προτέρων κατανομή του Jeffreys για το k θα είναι

$$\pi_j(k) = \pi_j(g^{-1}(k)) \left(\frac{d\theta}{dk} \right)^2 \propto [I(\theta)]^{\frac{1}{2}} \left| \frac{d\theta}{dk} \right| = [I(k)]^{\frac{1}{2}}.$$

Παράδειγμα 1.3

Έστω παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενες από τυχαίο δείγμα μεγέθους n όπου ακολουθεί κανονική κατανομή $N(\mu, \sigma^2)$ όπου $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ με συνάρτηση πυκνότητας πιθανότητας

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} I_{\mathbb{R}}(x).$$

Η εκ των προτέρων κατανομή του Jeffreys δίνεται από

$$\pi(\theta) \propto [I(\theta)]^{\frac{1}{2}}$$

Επομένως για τον υπολογισμό της πληροφορίας Fisher έχουμε

$$I(\theta) = E_{\mathbf{x}|\theta} \left\{ \left[\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) \right]^2 \right\}.$$

Η συνάρτηση πιθανοφάνειας υπολογίζεται ως

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\} I_{\mathbb{R}}(x_i) \right\} \\ &= (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} I_{\mathbb{R}^n}(\mathbf{x}). \end{aligned}$$

Στην συνέχεια υπολογίζεται ο λογάριθμος της συνάρτησης πιθανοφάνειας ως

$$\begin{aligned} \log f(\mathbf{x}|\theta) &= \log \left\{ (\sigma\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \right\} \\ &= -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (K1.1) \end{aligned}$$

Μετά παραγωγίζουμε την (K1.1) ως προς μ και προκύπτει ότι

$$\frac{\partial}{\partial \mu} \log f(\mathbf{x}|\theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu)(-1) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu).$$

Υστερα παραγωγίζουμε την σχέση (K1.1) ως προς σ^2 και προκύπτει ότι

$$\frac{\partial}{\partial \sigma^2} \log(f(\mathbf{x}|\theta)) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

Για να προχωρήσουμε στον υπολογισμό του πίνακα πληροφορίας Fisher, θέτουμε αρχικά

$$g(\theta) = \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta)$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{bmatrix} \\ = A.$$

Συνεπώς έχουμε ότι

$$AA' = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) & -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{bmatrix} \\ = \begin{bmatrix} \frac{1}{\sigma^4} \left(\sum_{i=1}^n (x_i - \mu) \right)^2 & -\frac{n}{2\sigma^4} \sum_{i=1}^n (x_i - \mu) + \frac{1}{2\sigma^6} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \left(\sum_{i=1}^n (x_i - \mu) \right) \\ -\frac{n}{2\sigma^4} \sum_{i=1}^n (x_i - \mu) + \frac{1}{2\sigma^6} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \left(\sum_{i=1}^n (x_i - \mu) \right) & \left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right)^2 \end{bmatrix}$$

Στην συνέχεια για να υπολογίσουμε τον πίνακα πληροφορίας Fisher θα χρειαστεί να υπολογίσουμε την μέση τιμή του πίνακα AA' .

Η μέση τιμή του $(AA)'_{11}$ θα είναι

$$E \left[\frac{1}{\sigma^4} \left(\sum_{i=1}^n (X_i - \mu) \right)^2 \right] = \frac{1}{\sigma^2} E \left[\left(\sum_{i=1}^n \frac{(X_i - \mu)}{\sigma} \right)^2 \right]$$

όπου η ποσότητα μέσα στην παρένθεση ακολουθεί X_n^2 κατανομή με n βαθμούς ελευθερίας, όπου η μέση της τιμή είναι οι βαθμοί ελευθερίας της, συνεπώς η μέση τιμή του $(AA)'_{11}$ είναι

$$\frac{n}{\sigma^2}.$$

Η μέση τιμή του $(AA)'_{12}$ θα είναι

$$E \left[-\frac{n}{2\sigma^4} \sum_{i=1}^n (x_i - \mu) + \frac{1}{2\sigma^6} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) \left(\sum_{i=1}^n (x_i - \mu) \right) \right] \\ = -\frac{n}{2\sigma^4} (nE[\bar{X}] - n\mu) + \frac{1}{2\sigma^6} \left(\sum_{i=1}^n E(x_i - \mu)^2 \right) \left(\sum_{i=1}^n E(x_i - \mu) \right) \\ = -\frac{n}{2\sigma^4} (n\mu - n\mu) + \frac{n}{2\sigma^6} \sigma^2 0 = 0.$$

Αντίστοιχα η μέση τιμή του $(AA)'_{21}$ θα είναι 0 καθώς είναι το ίδιο με το $(AA)'_{12}$.

Η μέση τιμή του $(AA)'_{22}$ θα είναι

$$\begin{aligned}
 E \left[\left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right)^2 \right] &= \\
 &= \text{Var} \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right] + E^2 \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right] \\
 &= \text{Var} \left[\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} \right] + \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n E(x_i - \mu)^2 \right]^2 \\
 &= \text{Var} \left[\frac{1}{2\sigma^2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] + \left[-\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4} \sigma^2 \right]^2 = \text{Var} \left[\frac{1}{2\sigma^2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right]. \quad (K1.2)
 \end{aligned}$$

Όταν για την τυχαία μεταβλητή X ισχύει ότι $X \sim N(\mu, \sigma^2)$, τότε η τυποποιημένη εκδοχή της, $Z = \frac{X - \mu}{\sigma}$ ακολουθεί την τυπική κανονική κατανομή, συνεπώς η τυχαία μεταβλητή $Y = \sum_{i=1}^n Z_i^2$ θα ακολουθεί κατανομή χ_n^2 και η διακύμανσή της αντίστοιχα θα είναι $\text{Var}(Y) = 2n$.

Επομένως η σχέση (K1.2) γίνεται

$$\begin{aligned}
 E \left[\left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right)^2 \right] &= \frac{1}{4\sigma^4} \text{Var} \left[\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] \\
 &= \frac{1}{4\sigma^4} 2n \\
 &= \frac{n}{2\sigma^4}.
 \end{aligned}$$

Συνεπώς ο πίνακας πληροφορίας Fisher διαμορφώνεται ως

$$I(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

Άρα η εκ των προτέρων κατανομή του Jeffreys θα είναι

$$\pi_j(\mu, \sigma) \propto |I(\theta)|^{\frac{1}{2}} = \left(\frac{n}{\sigma^2} \frac{n}{2\sigma^4} \right)^{\frac{1}{2}} \propto \frac{1}{\sigma^3}.$$

1.2.2.2 Εκ των προτέρων κατανομή αναφοράς

Μία άλλη διαδεδομένη κατηγορία μη πληροφοριακών κατανομών είναι η κατηγορία των εκ των προτέρων κατανομές αναφοράς, την οποία την εισήγαγε αρχικά ο Bernardo (1979) και στην συνέχεια την εξέλιξε μαζί με τον Berger (1992) σε μία κοινή τους εργασία. Αξίζει να σημειωθεί ότι μία εναλλακτική ονομασία αυτής της κατανομής είναι η μέθοδος *Berger-Bernardo* (*Berger-Bernardo Method*).

Έστω ένα τυχαίο δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Ορίζουμε την *απόκλιση Kullback-Leiber* (*Kullback-Leiber distance*) μεταξύ της εκ των προτέρων και της εκ των υστέρων κατανομής ως

$$K_n(\pi(\theta|\mathbf{x}), \pi(\theta)) = \int_{\theta \in \Theta} f(\mathbf{x}|\theta)\pi(\theta) \log\left(\frac{\pi(\theta|\mathbf{x})}{\pi(\theta)}\right) d\theta .$$

Η *εκ των προτέρων κατανομή αναφοράς* (*reference prior*) ορίζεται ως η εκ των προτέρων κατανομή η οποία μεγιστοποιεί την απόκλιση Kullback-Leiber μεταξύ της εκ των προτέρων και της εκ των υστέρων κατανομής.

Η λογική αυτής της μεθόδου βασίζεται στο μέγεθος της πληροφορίας που παρέχουν τα δεδομένα, καθώς όσο μεγαλύτερη πληροφορία προσδίδουν τα δεδομένα, τόσο λιγότερη πληροφορία παρέχει η εκ των προτέρων κατανομή. Όταν η παράμετρος θ είναι μονοδιάστατη η κατανομή αυτή συμπίπτει με την εκ των προτέρων κατανομή του Jeffreys, όμως όταν μελετάμε προβλήματα σε μεγαλύτερες διαστάσεις η εκ των προτέρων κατανομή αναφοράς προσαρμόζεται καλύτερα συγκριτικά με αυτή του Jeffreys.

1.2.3 Ιεραρχικές εκ των προτέρων κατανομές

Οι κατανομές γενικά στην Στατιστική ορίζονται από ορισμένες παραμέτρους οι οποίες αντιπροσωπεύουν κάποια ποιοτικά χαρακτηριστικά της κατανομής, όπως μέση τιμή, διακύμανση, ρυθμό κλπ. Αυτά τα χαρακτηριστικά μπορούν να οριστούν είτε ως σταθεροί γνωστοί όροι είτε να είναι τυχαίες μεταβλητές ακολουθώντας κάποια κατανομή. Στην περίπτωση όπου οι παράμετροι της εκ των προτέρων κατανομής είναι τυχαίες μεταβλητές θα καλούνται *υπερπαραμέτροι* (*hyper-parameters*) και θα έχουν συμβολή στον υπολογισμό της εκ των υστέρων κατανομής.

Έστω δεδομένα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ με συνάρτηση πυκνότητας πιθανότητας $f(\mathbf{x}|\boldsymbol{\theta})$ και $\boldsymbol{\theta}$ τυχαία μεταβλητή με εκ των προτέρων κατανομή $\pi(\boldsymbol{\theta}, \eta)$ όπου $\eta \sim \pi(\eta)$. Τότε η εκ των προτέρων κατανομή της τυχαία μεταβλητής $\boldsymbol{\theta}$ θα δίνεται από

$$\pi(\boldsymbol{\theta}, \eta) = \pi(\boldsymbol{\theta}|\eta)\pi(\eta).$$

Τέτοιου είδους εκ των προτέρων κατανομές θα καλούνται **ιεραρχικές εκ των προτέρων κατανομές (hyperpriors)**, όπου κάθε υπερπαράμετρος αποτελεί και ένα **επίπεδο (level)** της εκ των υστέρων κατανομής.

Τότε η εκ των υστέρων κατανομή της τυχαίας μεταβλητής θ θα υπολογίζεται από

$$\pi(\theta, \eta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta, \eta) \pi(\theta, \eta)}{\iint f(\mathbf{x} | \theta, \eta) \pi(\theta, \eta) d\theta d\eta}.$$

Συνεπώς μία εκ των προτέρων κατανομή μπορεί να οριστεί από πολλές υπερπαραμέτρους και να αποτελείται από πολλά επίπεδα. Αυτού του είδους κατασκευής εκ των προτέρων κατανομής καλείται **ιεραρχική μοντελοποίηση (hierarchical modelling)**. Οι ιεραρχικές εκ των προτέρων κατανομές χρησιμοποιούνται προκειμένου να περιγράψουμε ένα πρόβλημα του οποίου οι παράμετροι εμφανίζουν κάποια εξάρτηση ή συσχέτιση, αντιπροσωπεύοντας την εξάρτηση αυτή μέσω μίας από κοινού κατανομής των παραμέτρων και των υπερπαραμέτρων τους

Παράδειγμα 1.4

Έστω παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενες από τυχαίο δείγμα μεγέθους n όπου ακολουθεί την κατανομή Poisson με παράμετρο $\lambda > 0$ όπου η συνάρτηση πιθανοφάνειας του τυχαίου δείγματος θα δίνεται από τον τύπο

$$f(\mathbf{x} | \lambda) = e^{-\lambda n} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} I_{Z^+}(\mathbf{x}).$$

Διαλέγουμε ως εκ των προτέρων κατανομή της παραμέτρου λ μία κατανομή γάμμα με παραμέτρους a και b , όπου a και b υπερπαράμετροι οι οποίες ακολουθούν κατανομές γάμμα με παραμέτρους γ, δ και τ, ρ αντίστοιχα, όπου $\gamma, \delta, \tau, \rho$ γνωστές θετικές σταθερές.

Οι εκ των προτέρων κατανομές θα δίνονται από τους τύπους

$$\pi(\lambda | a, b) = \frac{\lambda^{\alpha-1} b^\alpha e^{-b\lambda}}{\Gamma(\alpha)} I_{(0, \infty)}(\lambda)$$

$$\pi(\alpha) = \frac{a^{\gamma-1} \delta^\gamma e^{-\delta a}}{\Gamma(\gamma)} I_{(0, \infty)}(\alpha)$$

$$\pi(b) = \frac{b^{\tau-1} \rho^\tau e^{-\rho b}}{\Gamma(\tau)} I_{(0, \infty)}(b).$$

Η από κοινού εκ των υστέρων κατανομή των τυχαίων μεταβλητών λ, a, b θα δίνεται από

$$\begin{aligned} \pi(\lambda, a, b | \mathbf{x}) &\propto f(\mathbf{x} | \lambda, a, b) \pi(\lambda | a, b) \pi(a) \pi(b) \\ &= e^{-\lambda n} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} I_{z^+}(\mathbf{x}) \frac{\lambda^{\alpha-1} b^\alpha e^{-b\lambda}}{\Gamma(\alpha)} I_{(0,\infty)}(\lambda) \frac{a^{\gamma-1} \delta^\gamma e^{-\delta a}}{\Gamma(\gamma)} I_{(0,\infty)}(a) \\ &\quad \times \frac{b^{\tau-1} \rho^\tau e^{-\rho b}}{\Gamma(\tau)} I_{(0,\infty)}(b) \\ &\propto e^{-\lambda n} \lambda^{n\bar{x}} \frac{\lambda^{\alpha-1} b^\alpha e^{-b\lambda}}{\Gamma(\alpha)} a^{\gamma-1} e^{-\delta a} b^{\tau-1} e^{-\rho b} I_{(0,\infty)^3}(\lambda, a, b). \end{aligned}$$

Στη συνέχεια για να καθοριστούν οι εκ των υστέρων κατανομές της τυχαίας μεταβλητής λ αλλά και των υπερπαραμέτρων a και b προχωράμε στον υπολογισμό των **κατανομών πλήρους δέσμευσης (full conditional distributions)**. Οι κατανομές πλήρους δέσμευσης υπολογίζονται δεσμεύοντας την μεταβλητή ενδιαφέροντος ως προς τις υπόλοιπες τυχαίες μεταβλητές του μοντέλου, απαλείφοντας τους όρους οι οποίοι δεν προσδίδουν κάποια πληροφορία σε αυτήν.

Επομένως η κατανομή πλήρους δέσμευσης για την παράμετρο λ θα είναι

$$\begin{aligned} \pi(\lambda | a, b) &\propto e^{-\lambda n} \lambda^{n\bar{x}} \lambda^{\alpha-1} e^{-b\lambda} I_{(0,\infty)}(\lambda) \\ &= e^{-\lambda(b+n)} \lambda^{\alpha+n\bar{x}-1} I_{(0,\infty)}(\lambda). \end{aligned}$$

Παρατηρούμε ότι η κατανομή πλήρους δέσμευσης για την παράμετρο λ είναι μία κατανομή γάμμα με παραμέτρους $\alpha' = \alpha + n\bar{x}$ και $b' = b + n$.

Αντίστοιχα η κατανομή πλήρους δέσμευσης για την υπερπαραμέτρο b θα είναι

$$\pi(b | \lambda, a) \propto b^{\alpha+\tau-1} e^{-(\lambda+\rho)b} I_{(0,\infty)}(b),$$

όπου είναι μία κατανομή γάμμα με παραμέτρους $\alpha' = \alpha + \tau$ και $b' = \lambda + \rho$

Αντίστοιχα η κατανομή πλήρους δέσμευσης για την υπερπαραμέτρο a θα είναι

$$\pi(a | \lambda, b) \propto \frac{\lambda^{\alpha-1} b^\alpha a^{\gamma-1} e^{-\delta a}}{\Gamma(\alpha)} I_{(0,\infty)}(a).$$

Παρατηρούμε ότι η κατανομή πλήρους δέσμευσης της υπερπαραμέτρου a δεν είναι κάποια "γνωστή" κατανομή η οποία ορίζεται από μία συνάρτηση κλειστού τύπου όπως στις προηγούμενες περιπτώσεις, επομένως δεν μπορούμε να προχωρήσουμε σε ανάλυση βασικών ποιοτικών χαρακτηριστικών της.

Παρατηρούμε ότι σε περιπτώσεις όπου η εκ των υστέρων κατανομή δεν δίνεται σε κλειστό τύπο, αδυνατούμε να εξάγουμε συμπεράσματα. Αυτή ήταν και η βασική αδυναμία της Μπεϋζιανής Στατιστικής μέχρι τα τέλη της δεκαετίας του 80, διότι σε πολλές πρακτικές εφαρμογές είναι αναγκαία η ιεραρχική μοντελοποίηση η οποία όμως οδηγεί σε εκ των υστέρων κατανομές οι οποίες δεν δίνονται σε κλειστό τύπο. Την επίλυση αυτού του δύσκολου προβλήματος έδωσαν οι MCMC μεθοδολογίες οι οποίες βασίζονται σε προσομοίωση από μαρκοβιανές αλυσίδες, δίνοντάς μας την δυνατότητα να παράγουμε τυχαία δείγματα από τέτοιου είδους μη αναγνωρίσιμες εκ των υστέρων κατανομές.

1.3 Επίλογος 1^ο Κεφαλαίου

Το πρώτο βήμα μίας μπεϋζιανής ανάλυσης, όπως παρουσιάστηκε σε αυτό το κεφάλαιο, είναι η κατασκευή ενός **στατιστικού μοντέλου** το οποίο βασίζεται στις εκ των προτέρων και τις εκ των υστέρων κατανομές, έχοντας ως σκοπό την μελέτη βασικών ποιοτικών και ποσοτικών χαρακτηριστικών τα οποία ανταποκρίνονται στις απαιτήσεις του προβλήματος. Μετά την ολοκλήρωση της κατασκευής του μοντέλου αυτού προκύπτουν ορισμένες φυσιολογικές απορίες όπως "Είναι το βέλτιστο δυνατό μοντέλο;", "Υπάρχει κάποιο καλύτερο μοντέλο;", "Ποιο από τα παραπάνω μοντέλα προσαρμόζεται καλύτερα στη φύση του προβλήματος;" κλπ. Απαντήσεις σε τέτοιου είδους ερωτήματα δίνει η **μπεϋζιανή επιλογή μοντέλου** και οι **έλεγχοι υποθέσεων** όπως θα παρουσιαστούν στο **Κεφάλαιο 2**.

Κεφάλαιο 2

Μπεϋζιανή επιλογή μοντέλου

Κυρίαρχο μέλημα της στατιστικής επιστήμης αποτελεί η επιλογή ενός **στατιστικού μοντέλου** (*statistical model*) το οποίο θα είναι αρκετά ικανό να περιγράψει ένα πρόβλημα καθώς και διάφορα βασικά χαρακτηριστικά αυτού. Για να καταλήξουμε σε ένα μοντέλο το οποίο θα έχει την καλύτερη δυνατή προσαρμογή στα δεδομένα που διαθέτουμε, καλούμαστε να επιλέξουμε μέσα από μία κλάση υποψηφίων μοντέλων. Αυτή η διαδικασία αξιολόγησης και σύγκρισης μοντέλων αποσκοπώντας στην ανάδειξη του επικρατέστερου μοντέλου ανάμεσα από τις διαθέσιμες επιλογές, αναφέρεται συχνά στην βιβλιογραφία ως το **πρόβλημα επιλογής μοντέλου** (*model selection problem*). Η Μπεϋζιανή σχολή, όπως και η Κλασσική σχολή αντίστοιχα, έχουν μελετήσει εκτενώς το πρόβλημα επιλογής μοντέλου, οδηγώντας στη δημιουργία μεθοδολογιών οι οποίες λαμβάνουν υπόψιν διάφορα ποιοτικά και ποσοτικά χαρακτηριστικά των υποψηφίων μοντέλων, όπως την πολυπλοκότητα ενός μοντέλου, την δομή, την προβλεπτική ικανότητα κλπ., βοηθώντας στην εξερεύνηση των διαθέσιμων επιλογών.

Για να προχωρήσουμε στην μελέτη του προβλήματος επιλογής μοντέλου, χρειάζεται αρχικά να μελετήσουμε τη έννοια του **ελέγχου υποθέσεων** (*hypothesis testing*), καθώς αποτέλεσε το πρώτο βήμα σύγκρισης δύο διαφορετικών μοντέλων και στην συνέχεια λειτούργησε ως βάση για την δημιουργία μεθοδολογιών που θα εξετάσουμε στην συνέχεια. Αρχικά θα εξετάσουμε την κλασσική προσέγγιση αναδεικνύοντας τις πρακτικές δυσκολίες ενός κλασσικού ελέγχου υποθέσεων και στην συνέχεια θα παρουσιάσουμε πως δομείται ένας έλεγχος υποθέσεων στην Μπεϋζιανή Στατιστική, καθώς και τα πλεονεκτήματα χρήσης του συγκριτικά με τον κλασσικό έλεγχο υποθέσεων.

2.1 Ο έλεγχος υποθέσεων στην Κλασσική Στατιστική

Στην Κλασσική Στατιστική ο έλεγχος υποθέσεων πραγματοποιείται μέσω των μεθόδων που εισήγαγαν οι Fisher, Neyman και Pearson. Η δομή ενός κλασσικού ελέγχου υποθέσεων όπως παρουσιάζεται στο βιβλίο των Κούτρα & Δαμιανού (2003) είναι η εξής:

- 1) Έστω ένα τυχαίο δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ από μία οικογένεια κατανομών

$$\mathbf{F} = \{F(\mathbf{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

με υποοικογένειες κατανομών

$$\mathbf{F} = \{F(\mathbf{x}, \boldsymbol{\theta}_0) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\} \text{ και } \mathbf{F} = \{F(\mathbf{x}, \boldsymbol{\theta}_1) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1\}$$

όπου

$$\theta_0 \cap \theta_1 = \emptyset$$

- 2) Ορίζουμε ως *μηδενική υπόθεση (null hypothesis)* την $H_0 : \theta \in \theta_0$. Ο ερευνητής ορίζει σε αυτήν την υπόθεση αυτό που θέλει να αμφισβητήσει ή να αποκλείσει.
- 3) Ορίζουμε ως *εναλλακτική υπόθεση (alternative hypothesis)* την $H_1 : \theta \in \theta_1$.
- 4) Ορίζουμε μία στατιστική συνάρτηση $T(X)$ για τον έλεγχο της H_0 εναντίον της H_1 .
- 5) Υπολογίζουμε την περιοχή απόρριψης K της H_0 και την τιμή p του ελέγχου.

Η *τιμή p (p-value)* ορίζεται ως η πιθανότητα κάτω από την H_0 να πάρουμε μεγαλύτερες τιμές της στατιστικής συνάρτησης $T(X)$ από την τιμή $T(X_{obs})$ που παρατηρήσαμε και πρακτικά υπολογίζει το μέγιστο *επίπεδο σημαντικότητας (significance level)* για το οποίο δεν θα απορρίπταμε τη μηδενική υπόθεση.

Ακολουθώντας αυτή τη διαδικασία οι έλεγχοι υποθέσεων της κλασσικής στατιστικής εδραιώθηκαν στις στατιστικές αναλύσεις οι οποίες διενεργούνται στα περισσότερα επιστημονικά πεδία. Παρ' όλα αυτά η κλασσική προσέγγιση παρουσιάζει αρκετές αδυναμίες ως προς την ερμηνεία και την πρακτική εφαρμογή των ελέγχων υποθέσεων.

Αρχικά η κλασσική προσέγγιση έχει άμεση εφαρμογή μόνο όταν οι δύο υποθέσεις που εξετάζουμε είναι εμφωλευμένες η μία μέσα στην άλλη και η μηδενική υπόθεση αποτελεί απλούστερη εκδοχή της εναλλακτικής υπόθεσης. Όμως σε πολλές πρακτικές εφαρμογές καλούμαστε να συγκρίνουμε περισσότερα από δύο υποψήφια μοντέλα τα οποία δεν περιγράφονται από την ίδια κατανομή, όπως για παράδειγμα σύγκριση εκθετικής και λογαριθμοκανονικής κατανομής ως προς την προσαρμογή τους σε δεδομένα απώλειας ασφαλιστικών αποζημιώσεων.

Ένα άλλο πρόβλημα που εμφανίζει η κλασσική προσέγγιση είναι πως επικεντρώνεται στην συγκέντρωση στοιχείων εναντίον της μηδενικής υπόθεσης και όχι στην άμεση παραδοχή κάποιου μοντέλου ως επικρατέστερο. Για μικρή τιμή p , ο έλεγχος υποδεικνύει ότι το εναλλακτικό μοντέλο το οποίο είναι πολυπλοκότερο συγκριτικά με το μοντέλο που εξετάζεται στην μηδενική υπόθεση, μπορεί να ερμηνεύσει καλύτερα το υπάρχον πρόβλημα. Όμως για μεγάλη τιμή p , δεν ισχύει ότι το υπό εξέταση μοντέλο της μηδενικής υπόθεσης είναι επικρατέστερο, αλλά ότι δεν υπάρχουν αρκετές ενδείξεις ώστε να απορριφθεί.

Η τιμή p επίσης, δεν μπορεί να ερμηνευτεί ως μία πιθανότητα υπέρ κάποιας υπόθεσης, παρά μόνο ως πιθανότητα εμφάνισης δεδομένων τα οποία προσδίδουν πιο ακραίες τιμές στην στατιστική συνάρτηση $T(X)$ η οποία χρησιμοποιήθηκε για τον έλεγχο της μηδενικής υπόθεσης. Συχνά επικρατεί η λανθασμένη ερμηνεία της τιμής p ως η πιθανότητα να ισχύει η μηδενική υπόθεση, ερμηνεία η οποία δεν συμβαδίζει με την πραγματική σημασία της τιμής p .

Επίσης σημαντικό ρόλο στην διενέργεια των κλασικών ελέγχων υποθέσεων διαδραματίζει το μέγεθος του δείγματος, καθώς όσο αυτό αυξάνεται αντίστοιχα μειώνεται η τιμή p οδηγώντας στην απόρριψη της μηδενικής υπόθεσης. Ο 21^{ος} αιώνας χαρακτηρίζεται ως η εποχή των *μεγάλων σετ δεδομένων (big data)*, συνεπώς δημιουργείται η ανάγκη εναλλακτικών τρόπων αντιμετώπισης της σύγκρισης μοντέλων κάτω από μεγάλο όγκο διαθέσιμων παρατηρήσεων.

2.2 Η σύγκριση μοντέλων στην Μπεϋζιανή Στατιστική

Η Μπεϋζιανή Στατιστική προσεγγίζει τους ελέγχους υποθέσεων και την επιλογή μοντέλων με έναν πιο πρακτικό και άμεσο τρόπο χωρίς να παρουσιάζονται δυσκολίες στην ερμηνεία των αποτελεσμάτων, όπως γίνεται μέσω της κλασικής μεθόδου, καθώς βασίζεται σε μοντέλα τα οποία κάνουν αποκλειστικά χρήση πιθανοτήτων. Ο έλεγχος υποθέσεων στην Μπεϋζιανή Στατιστική είναι στην ουσία η σύγκριση δύο ή και παραπάνω υποψηφίων μοντέλων, με βασικό πλεονέκτημα την ταυτόχρονη και πολλαπλή σύγκριση μοντέλων χωρίς να είναι υποχρεωτικό το ένα μοντέλο να είναι εμφωλευμένο ενός άλλου. Επίσης σημαντικό προτέρημα της Μπεϋζιανής Στατιστικής αποτελεί η χρήση συγκεκριμένων ποσοτικών δεικτών οι οποίοι βασίζονται σε λόγους πιθανοτήτων και υποδεικνύουν ξεκάθαρα ποιο μοντέλο είναι το επικρατέστερο. Η βασική δομή και σύγκριση υποψηφίων μοντέλων στην Μπεϋζιανή Στατιστική θα παρουσιαστεί όπως αυτή περιγράφεται στην εργασία των Kass & Raftery (1995).

Για να προχωρήσουμε στην περιγραφή της διαδικασίας κατασκευής αλλά και σύγκρισης μοντέλων στην Μπεϋζιανή Στατιστική ορίζουμε αρχικά ένα αριθμήσιμο σύνολο μοντέλων προς αξιολόγηση $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$. Η κατασκευή ενός μοντέλου στην Μπεϋζιανή Στατιστική μπορεί να θεωρηθεί ως ένα ιεραρχικό μοντέλο τριών επιπέδων.

Στο πρώτο στάδιο κατασκευής ενός μοντέλου καθορίζεται η εκ των προτέρων πιθανότητα των μοντέλων M_i , $\pi(M_i)$, η οποία εκπροσωπεί την εκ των προτέρων πεποίθησή μας ότι το μοντέλο M_i ότι παρήγαγε τα δεδομένα, όπου ισχύει ότι

$$\sum_{i=1}^k \pi(M_i) = 1.$$

Μία συνηθισμένη πρακτική είναι η επιλογή μίας ομοιόμορφης πιθανότητας για όλα τα μοντέλα, έτσι ώστε να εξασφαλίζεται η αντικειμενικότητα σε περιπτώσεις όπου δεν υπάρχει κάποια εκ των προτέρων πληροφορία. Τότε η ομοιόμορφη εκ των προτέρων πιθανότητα του μοντέλου M_i θα ορίζεται ως

$$\pi(M_i) = \frac{1}{k}.$$

Στην συνέχεια θεωρούμε τις παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενες από τυχαίο δείγμα μεγέθους n , με βάση τις οποίες υπολογίζουμε την συνάρτηση πυκνότητας πιθανότητας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ κάτω από το μοντέλο M_i , με αντίστοιχη εκ των προτέρων πυκνότητα πιθανότητας για την παράμετρο $\boldsymbol{\theta}_i$, $\pi(\boldsymbol{\theta}_i|M_i)$. Επομένως εφαρμόζοντας το θεώρημα Μπεϋες η εκ των υστέρων πιθανότητα του μοντέλου M_i θα δίνεται ως

$$\pi(M_i|\mathbf{x}) = \frac{f(\mathbf{x}|M_i)\pi(M_i)}{\sum_{M_l \in \mathcal{M}} f(\mathbf{x}|M_l)\pi(M_l)}.$$

όπου η ποσότητα $f(\mathbf{x}|M_i)$ καλείται και συνάρτηση *περιθωριακής πιθανοφάνειας (marginal likelihood)* του μοντέλου M_i , η οποία προκύπτει ολοκληρώνοντας την συνάρτηση πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ ως προς το διάνυσμα των παραμέτρων $\boldsymbol{\theta}_i$, δηλαδή

$$m_i(\mathbf{x}) = \int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i \quad (K2.1).$$

Καταφεύγουμε στην χρήση της περιθωριακής πυκνότητας $m_i(\mathbf{x})$ αντί της συνάρτησης πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$, διότι η συνάρτηση πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ εξαρτάται από την παράμετρο $\boldsymbol{\theta}_i$.

Για να προχωρήσουμε στην σύγκριση των δύο υποψηφίων μοντέλων M_i και M_j του συνόλου \mathcal{M} θα προχωρήσουμε στον υπολογισμό του λόγου συμπληρωματικών πιθανοτήτων. Με τον όρο *λόγος συμπληρωματικών πιθανοτήτων (odds)* αναφερόμαστε στον λόγο της πιθανότητας πραγματοποίησης ενός ενδεχομένου A ως προς την πιθανότητα πραγματοποίησης του συμπληρωματικού του, δηλαδή

$$odds(A) = \frac{P(A)}{1 - P(A)}.$$

Επομένως ο *εκ των προτέρων λόγος συμπληρωματικών πιθανοτήτων (prior odds)* του μοντέλου M_i ως προς του μοντέλου M_j δίνεται από

$$O_{ij} = \frac{\pi(M_i)}{\pi(M_j)}.$$

Αντίστοιχα ο *εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων (posterior odds)* του μοντέλου M_i ως προς του μοντέλου M_j δίνεται από

$$PO_{ij} = \frac{\pi(M_i|\mathbf{x})}{\pi(M_j|\mathbf{x})}.$$

Συνεπώς έχουμε ότι

$$PO_{ij} = \frac{f(\mathbf{x}|M_i) \pi(M_i)}{f(\mathbf{x}|M_j) \pi(M_j)}.$$

Η ποσότητα

$$BF_{ij} = \frac{f(\mathbf{x}|M_i)}{f(\mathbf{x}|M_j)}$$

καλείται **παράγοντας Μπεϋες (Bayes factor)** υπέρ του μοντέλου M_i και προκύπτει από το πηλίκο του λόγου των εκ των υστέρων συμπληρωματικών πιθανοτήτων ως προς τον λόγο των εκ των προτέρων συμπληρωματικών πιθανοτήτων του μοντέλου M_i , δηλαδή

$$PO_{ij} = BF_{ij} O_{ij}.$$

Επίσης ένας εναλλακτικός τρόπος υπολογισμού της εκ των υστέρων πιθανότητας του μοντέλου M_i προκύπτει ως

$$\begin{aligned} \pi(M_i|\mathbf{x}) &= \frac{f(\mathbf{x}|M_i)\pi(M_i)}{\sum_{M_l \in \mathcal{M}} f(\mathbf{x}|M_l)\pi(M_l)} \\ &= \frac{1}{\sum_{M_l \in \mathcal{M}} \frac{f(\mathbf{x}|M_l)\pi(M_l)}{f(\mathbf{x}|M_i)\pi(M_i)}} \\ &= \frac{1}{\sum_{M_l \in \mathcal{M}} PO_{l,i}} \\ &= \left(\sum_{M_l \in \mathcal{M}} PO_{l,i} \right)^{-1}. \end{aligned}$$

Δηλαδή η εκ των υστέρων πιθανότητα του μοντέλου M_i μπορεί να εκφραστεί μέσω του αντιστρόφου του αθροίσματος των εκ των υστέρων λόγων συμπληρωματικών πιθανοτήτων όλων των μοντέλων του συνόλου \mathcal{M} ως προς το μοντέλο M_i .

Αξίζει να τονίσουμε ότι στην περίπτωση όπου οι εκ των προτέρων πιθανότητες των δύο μοντέλων είναι ίσες, δηλαδή $\pi(M_i) = \pi(M_j)$, τότε ο παράγοντας Μπεϋες ισούται με τον εκ των υστέρων λόγο συμπληρωματικών πιθανοτήτων, δηλαδή

$$PO_{ij} = BF_{ij}.$$

Επίσης μπορούμε εύκολα να παρατηρήσουμε ότι ο παράγοντας Μπεϋες του μοντέλου M_j εναντίον του μοντέλου M_i είναι

$$BF_{ji} = \frac{f(x|M_j)}{f(x|M_i)} = \frac{1}{\frac{f(x|M_i)}{f(x|M_j)}} = \frac{1}{BF_{ij}}.$$

Είναι εμφανές ότι ο παράγοντας Μπεϋες δεν εξαρτάται από τις εκ των προτέρων πιθανότητες $\pi(M_i)$ ή $\pi(M_j)$ που ορίσαμε, συνεπώς αποτελεί ένα αντικειμενικό κριτήριο επιλογής για την σύγκριση μοντέλων. Επίσης εκφράζεται ως ο λόγος των περιθωρίων πιθανοφανειών των εκάστοτε προς σύγκριση μοντέλων, όμοια με τον γενικευμένο λόγο πιθανοφανειών που χρησιμοποιείται στην Κλασσική Στατιστική.

Μετά την μαθηματική θεμελίωση του παράγοντα Μπεϋες, ο Jeffreys (1961) και οι Kass&Raftery (1995) εισήγαγαν αριθμητικές κλίμακες αξιολόγησης του Παράγοντα Μπεϋες, όπως περιγράφονται από τον **Πίνακα 2.1** και τον **Πίνακα 2.2**, προκειμένου να υπάρχει ένας γενικός κανόνας αποφάσεων στους ελέγχους υποθέσεων.

Πίνακας 2.1 Τιμές παράγοντα Bayes σύμφωνα με τον Jeffreys

Κλίμακα του Jeffreys (1961)		
$\log_{10} BF_{ij}(x)$	$BF_{ij}(x)$	Στοιχεία εναντίον του M_j
0 έως 0.5	1 έως 3.2	Όχι αρκετά
0.5 έως 1	3.2 έως 10	Ουσιαστικά
1 έως 2	10 έως 100	Ισχυρά
Τιμές μεγαλύτερες του 2	Τιμές μεγαλύτερες του 100	Καθοριστικά

Πίνακας 2.2 Τιμές παράγοντα Bayes σύμφωνα με τους Kass&Raftery

Κλίμακα των Kass και Raftery (1995)		
$2\log BF_{ij}(x)$	$BF_{ij}(x)$	Στοιχεία εναντίον του M_j
0 έως 2	1 έως 3	Όχι αρκετά
2 έως 6	3 έως 20	Ουσιαστικά
6 έως 10	20 έως 150	Ισχυρά
Τιμές μεγαλύτερες του 10	Τιμές μεγαλύτερες του 150	Καθοριστικά

Ολοκληρώνοντας την περιγραφή της δομής και διενέργειας της σύγκρισης μοντέλων στην Μπεϋζιανή Στατιστική, θα προχωρήσουμε στην μπεϋζιανή στάθμιση μοντέλων.

2.3 Μπεϋζιανή στάθμιση μοντέλων

Η ιδέα γύρω από την επιλογή μοντέλου όπως αναφέραμε είναι η εύρεση ενός βέλτιστου μοντέλου, το οποίο θα είναι σε θέση να περιγράψει με την μεγαλύτερη δυνατή ακρίβεια το πρόβλημα το οποίο μελετάμε. Στην πραγματικότητα όμως υπάρχουν δύο ζητήματα τα οποία θέτουν υπό αμφισβήτηση τον όρο "βέλτιστο" γύρω από την διαδικασία επιλογής μοντέλου. Αρχικά η συνηθισμένη πρακτική που ακολουθούν οι ερευνητές είναι η επιλογή ενός μοντέλου M_i μέσα από μία κλάση μοντέλων \mathcal{M} , θεωρώντας ότι το μοντέλο M_i είναι το βέλτιστο και αυτό παρήγαγε τα δεδομένα. Αποτέλεσμα αυτής της τακτικής είναι η παράβλεψη της αβεβαιότητας που δημιουργείται γύρω από την επιλογή ενός μοντέλου, οδηγώντας σε αποφάσεις οι οποίες κρύβουν κινδύνους. Επίσης είναι σύνηθες να υπάρχουν δύο ή και περισσότερα υποψήφια μοντέλα τα οποία μπορούν να έχουν εξίσου καλή προσαρμογή στα δεδομένα τα οποία διαθέτουμε είτε να μη υπάρχει κάποια ξεκάθαρη ένδειξη υπέρ κάποιου μοντέλου. Η **Μπεϋζιανή στάθμιση μοντέλων (Bayesian model averaging)** λύνει τα παραπάνω προβλήματα συνδυάζοντας τα διαθέσιμα μοντέλα μέσω των εκ των υστέρων πιθανοτήτων τους.

Η πρώτη αναφορά στην στατιστική βιβλιογραφία σχετικά με την μίξη μοντέλων έγινε από τον Barnard (1963) σε μία μελέτη όπου συνδυάζε μοντέλα για την πρόβλεψη του συνολικού πλήθους μιλίων που θα διανύσει αεροπορικά ένας επιβάτης. Στην συνέχεια ο Roberts (1965) πρότεινε την χρήση μίας κατανομής η οποία συνδυάζει τα αποτελέσματα δύο διαφορετικών μοντέλων και προκύπτει ως το αποτέλεσμα της στάθμισης των δύο εκ των υστέρων κατανομών των μοντέλων. Ο Leamer (1978) χρησιμοποίησε ως βάση την δουλειά του Roberts (1965) και δημιούργησε ένα βασικό παράδειγμα της μπεϋζιανής στάθμισης μοντέλων στο οποίο απέδειξε πως η συγκεκριμένη μεθοδολογία λαμβάνει υπόψιν στην επιλογή του μοντέλου την αβεβαιότητα γύρω από αυτό. Στην συνέχεια αρκετοί ερευνητές όπως οι Draper (1995), Kass&Raftery (1995) επέκτειναν την ιδέα της μπεϋζιανής στάθμισης μοντέλων στην Στατιστική εδραιώνοντας την ως μία βασική μεθοδολογία για την επίλυση του προβλήματος επιλογής μοντέλου.

Έστω παρατηρήσεις $\mathbf{x} = (x_1, x_2, \dots, x_n)$, ένα πεπερασμένο σύνολο μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$ και Δ η παράμετρος ενδιαφέροντος (συνάρτηση ζημίας, μελλοντική παρατήρηση κλπ.). Η εκ των υστέρων κατανομή της παραμέτρου Δ θα δίνεται ως

$$\pi(\Delta|\mathbf{x}) = \sum_{i=1}^k \pi(\Delta|M_i, \mathbf{x})\pi(M_i|\mathbf{x}) \quad (K2.2)$$

όπου $\pi(\Delta|M_i, \mathbf{x})$ είναι η εκ των υστέρων κατανομή της παραμέτρου Δ για το κάθε μοντέλο M_i και $\pi(M_i|\mathbf{x})$ η εκ των υστέρων πιθανότητα που δίνεται από

$$\pi(M_i|\mathbf{x}) = \frac{f(\mathbf{x}|M_i)\pi(M_i)}{\sum_{M_l \in \mathcal{M}} f(\mathbf{x}|M_l)\pi(M_l)}.$$

Η μπεϋζιανή στάθμιση μοντέλων παρουσιάζει δύο βασικές δυσκολίες ως προς την εφαρμογή της.

1) Μία πρώτη δυσκολία αφορά στο πλήθος των εξεταζόμενων μοντέλων, καθώς όσο περισσότερα μοντέλα εξετάζονται τόσο περισσότεροι θα είναι και οι όροι του αθροίσματος της σχέσης (K2.2), κάνοντας τον υπολογισμό της ποσότητας $\pi(\mathbf{A}|\mathbf{x})$ αρκετά χρονοβόρο.

Για την επίλυση αυτού του προβλήματος χρησιμοποιείται το *ξυράφι του Occam* (*Occam's razor*), έχοντας ως στόχο τον αποκλεισμό των μοντέλων με την μικρότερη συμβολή στην ποσότητα $\pi(\mathbf{A}|\mathbf{x})$. Η απόρριψη των λιγότερο σημαντικών μοντέλων μπορεί να γίνει σε δύο στάδια.

Στο πρώτο στάδιο δημιουργούμε ένα σύνολο $A' \subset \mathcal{M}$ όπου

$$A' = \left\{ M_k : \frac{\max_l \{\pi(M_l|\mathbf{x})\}}{\pi(M_k|\mathbf{x})} \leq C \right\}.$$

Η διαδικασία επιλογής των στοιχείων του συνόλου A' επιτυγχάνεται μέσω της σύγκρισης της εκ των υστέρων πιθανότητας του μοντέλου M_l , το οποίο έχει την μεγαλύτερη συμβολή στην $\pi(\mathbf{A}|\mathbf{x})$, με τις εκ των υστέρων πιθανότητες των υπόλοιπων μοντέλων από την κλάση \mathcal{M} . Επομένως με αυτόν τον τρόπο τα μοντέλα τα οποία δημιουργούν μεγαλύτερους λόγους εκ των υστέρων πιθανοτήτων από την τιμή C δεν θα εμπεριέχονται στο σύνολο A' . Η τιμή C καθορίζεται υποκειμενικά από τον ερευνητή και οι Kass & Raftery (1995) πρότειναν να λαμβάνει την τιμή $C = 20$.

Στην συνέχεια αφού έχουν αφαιρεθεί τα μοντέλα με την μικρότερη συμβολή στην εκ των υστέρων κατανομή της ποσότητας \mathbf{A} , προχωράμε στο δεύτερο στάδιο καθορισμού του συνόλου των μοντέλων που πρόκειται να αξιοποιηθούν.

Στο δεύτερο στάδιο δημιουργούμε ένα σύνολο B όπου $B \subset A'$ με

$$B = \left\{ M_k : \exists M_l \in A', M_l \subset M_k, \frac{\pi(M_l|\mathbf{x})}{\pi(M_k|\mathbf{x})} > 1 \right\}$$

όπου γίνεται έλεγχος για ύπαρξη σύνθετων μοντέλων τα οποία έχουν μικρότερες εκ των υστέρων πιθανότητες από μοντέλα όπου απορρίφθηκαν κατά την κατασκευή του συνόλου A' , ευνοώντας έτσι απλούστερα μοντέλα.

Ολοκληρώνοντας την διαδικασία ταξινόμησης των μοντέλων ορίζουμε το σύνολο

$$A = A' \setminus B$$

και πραγματοποιούμε την Μπεϋζιανή στάθμιση μοντέλων στο σύνολο A υπολογίζοντας την εκ των υστέρων κατανομή

$$\pi(\Delta|\mathbf{x}) = \sum_{M_k \in A} f(\Delta|M_k, \mathbf{x})\pi(M_k|\mathbf{x}) / \sum_{M_k \in A} \pi(M_k|\mathbf{x})$$

έχοντας μειώσει σημαντικά τον αριθμό των υποψήφια μοντέλων.

2) Η επόμενη δυσκολία η οποία συναντάται στην μπεϋζιανή στάθμιση μοντέλων είναι ο υπολογισμός των ολοκληρωμάτων που πραγματοποιούνται μέσα στους αλγορίθμους, όμως μπορούν να ξεπεραστούν μέσω μεθοδολογιών *MCMC*.

Σημαντικό ρόλο επίσης έχει ο καθορισμός των εκ των προτέρων πιθανοτήτων των μοντέλων $\pi(M_i)$, που όπως αναφέρουν και οι Hoeting & Madigan & Raftery (1999) δεν έχει λάβει την ανάλογη προσοχή από τους ερευνητές.

2.4 Η επιλογή μεταβλητών στην Κλασσική Στατιστική

Ένα συνηθισμένο πρόβλημα που αντιμετωπίζει η Κλασσική Στατιστική είναι η **επιλογή των επεξηγηματικών μεταβλητών** που θα χρησιμοποιηθούν σε ένα τελικό μοντέλο παλινδρόμησης (*variable selection problem*). Όταν μελετάμε ένα πρόβλημα επιλογής μεταβλητών σκοπός μας είναι η επιλογή ενός μοντέλου με την καλύτερη προβλεπτική ικανότητα και παράλληλα να είναι όσο το λιγότερο πολύπλοκο, καθώς δεν είναι απαραίτητο να χρησιμοποιηθούν όλες οι δυνατές επεξηγηματικές μεταβλητές.

Για την αξιολόγηση των διαθέσιμων μοντέλων καταφεύγουμε σε διάφορα κριτήρια τα οποία ποινικοποιούν την πιθανοφάνεια και επιλέγουμε το μοντέλο το οποίο ελαχιστοποιεί την τιμή του κριτηρίου αυτού. Τα κριτήρια αυτά χωρίζονται σε δύο μέρη όπου το ένα μέρος ποινικοποιεί τα μοντέλα με μεγάλο πλήθος επεξηγηματικών μεταβλητών και το δεύτερο μέρος ευνοεί τα μοντέλα με την καλύτερη προβλεπτική ικανότητα.

Ένα κλασσικό κριτήριο είναι το **κριτήριο του Akaike (Akaike Information Criterion)** όπου η τιμή του κριτηρίου *AIC* για ένα μοντέλο m δίνεται από

$$AIC(m) = -2\log L(m) + 2d_m$$

όπου $L(m)$ είναι η μεγιστοποιημένη τιμή της πιθανοφάνειας υπό το μοντέλο m ως προς τις παραμέτρους του και d_m είναι ο αριθμός των άγνωστων παραμέτρων του μοντέλου m . Με την χρήση του κριτηρίου *AIC* στοχεύουμε στην εύρεση του μοντέλου το οποίο προσεγγίζει όσο το δυνατόν καλύτερα το "πραγματικό" μοντέλο το οποίο παρήγαγε τα δεδομένα.

Επίσης ένα ακόμα αρκετά διαδεδομένο κριτήριο είναι το ***BIC (Bayesian Information Criterion)***, το οποίο χρησιμοποιείται στην Μπεϋζιανή Στατιστική για την προσέγγιση του παράγοντα Μπεϋες και λειτουργεί παρόμοια με το κριτήριο *AIC*. Η τιμή του

κριτηρίου BIC για ένα μοντέλο m όταν το μέγεθος δείγματος n είναι μεγαλύτερο του 7.39 δίνεται από

$$BIC(m) = -2\log L(m) + d_m \log(n)$$

όπου $L(m)$ είναι η μεγιστοποιημένη τιμή της πιθανοφάνειας του μοντέλου m ως προς τις παραμέτρους του και d_m είναι ο αριθμός των άγνωστων παραμέτρων του μοντέλου m και n είναι το μέγεθος του δείγματος.

Εάν δεχθούμε ότι εξετάζουμε ένα μοντέλο παλινδρόμησης το οποίο αποτελείται από 12 επεξηγηματικές μεταβλητές, τότε θα χρειαστεί να μελετήσουμε και να συγκρίνουμε συνολικά $2^{12} = 4096$ μοντέλα. Για να μπορέσουμε να συγκρίνουμε όλα τα υποψήφια μοντέλα προκειμένου να καταλήξουμε στο βέλτιστο δυνατό θα χρειαστεί ένας υπερβολικά μεγάλος όγκος χρονοβόρων υπολογισμών. Για να αποφευχθεί αυτή η δυσκολία εισήχθησαν υβριδικές τμηματικές μεθοδολογίες οι οποίες εξερευνούν "έξυπνα" μέσα στον χώρο των υποψηφίων μοντέλων και χωρίζονται σε τρεις κατηγορίες:

- 1) Μέθοδος προς τα εμπρός επιλογής (*forward selection*).
- 2) Μέθοδος προς τα πίσω απαλοιφής (*backward elimination*).
- 3) Κατά τμήματα εμπρός και πίσω επιλογής (*stepwise regression*).

Μέθοδος προς τα εμπρός επιλογής

Στην περίπτωση της διαδικασίας επιλογής μοντέλου χρησιμοποιώντας την μέθοδο της προς τα εμπρός επιλογής εκτελούμε έναν τμηματικό αλγόριθμο πρόσθεσης επεξηγηματικών μεταβλητών. Χρησιμοποιούμε ως μοντέλο εκκίνησης το απλούστερο μοντέλο το οποίο αποτελείται μόνο από την σταθερά και καλείται το **μηδενικό μοντέλο** (*null model*). Στην συνέχεια πραγματοποιώντας διαδοχικούς ελέγχους προσθέτουμε κάθε φορά την περισσότερο "σημαντική" επεξηγηματική μεταβλητή βάσει κάποιου επιλεγμένου κριτηρίου (π.χ. AIC). Το βέλτιστο δυνατό μοντέλο θα είναι το μοντέλο στο οποίο δεν θα είναι εφικτό να προστεθεί κάποια εναπομένουσα επεξηγηματική μεταβλητή.

Μέθοδος προς τα πίσω απαλοιφής

Στην περίπτωση της διαδικασίας επιλογής μοντέλου χρησιμοποιώντας την μέθοδο της προς τα πίσω απαλοιφής εκτελούμε έναν τμηματικό αλγόριθμο αφαίρεσης επεξηγηματικών μεταβλητών. Χρησιμοποιούμε ως μοντέλο εκκίνησης το **πλήρες μοντέλο** (*full model*) το οποίο περιέχει όλες τις διαθέσιμες επεξηγηματικές μεταβλητές. Στην συνέχεια πραγματοποιώντας διαδοχικούς ελέγχους απαλείφουμε την επεξηγηματική μεταβλητή η οποία κρίνεται ως λιγότερο "σημαντική" βάσει κάποιου επιλεγμένου κριτηρίου (π.χ. AIC). Το βέλτιστο δυνατό μοντέλο θα είναι το μοντέλο από το οποίο δεν θα είναι εφικτό να αφαιρεθεί κάποια εναπομένουσα επεξηγηματική μεταβλητή.

Κατά τμήματα εμπρός και πίσω επιλογής

Στην περίπτωση διαδικασίας επιλογής μοντέλου χρησιμοποιώντας κατά τμήματα εμπρός και πίσω επιλογής πραγματοποιείται ένας συνδυασμός των δύο παραπάνω μεθόδων όπου επιλέγουμε ως μοντέλο εκκίνησης το μηδενικό μοντέλο και σε κάθε βήμα ελέγχουμε για την εισαγωγή αλλά και την αφαίρεση κάποιας επεξηγηματικής μεταβλητής βασιζόμενοι σε κάποιο επιλεγμένο κριτήριο (π.χ. *AIC*). Το βέλτιστο δυνατό μοντέλο θα είναι το μοντέλο στο οποίο δεν θα μπορεί να προστεθεί κάποια από τις εναπομένουσες επεξηγηματικές μεταβλητές αλλά ούτε και να αφαιρεθεί κάποια από τις εισηγμένες επεξηγηματικές μεταβλητές.

Οι τμηματικές διαδικασίες εξεύρεσης ενός βέλτιστου μοντέλου καθιερώθηκαν στην Κλασσική Στατιστική τα προηγούμενα χρόνια και είναι διαθέσιμες σχεδόν σε όλα τα στατιστικά πακέτα, όμως παρουσιάζουν ορισμένες αδυναμίες. Αρχικά όπως αναφέρει και ο Graybill (1976), όταν εξερευνούμε ένα συγκεκριμένο σύνολο μοντέλων κάνοντας χρήση και των τριών μεθοδολογιών που αναφέραμε είναι δυνατό να καταλήξουμε σε διαφορετικά αποτελέσματα. Επίσης, με τον τρόπο με τον οποίο λειτουργεί στην βάση του ο αλγόριθμος, δεν είναι σε θέση να εξερευνήσει όλα τα διαθέσιμα μοντέλα, παραλείποντας έτσι έναν σημαντικό αριθμό μοντέλων. Επίσης σημαντικό πρόβλημα δημιουργείται από το γεγονός ότι οι μεθοδολογίες αυτές επηρεάζονται σημαντικά από την παρουσία πολυσυγγραμικότητας, δηλαδή ύπαρξη υψηλής συσχέτισης μεταξύ των επεξηγηματικών μεταβλητών.

2.5 Η επιλογή μεταβλητών στην Μπεϋζιανή Στατιστική

Η Μπεϋζιανή Στατιστική όπως αναφέρει και ο George (2000) αντιμετωπίζει το πρόβλημα της επιλογής μεταβλητών ως μία ειδική περίπτωση του προβλήματος επιλογής μοντέλου όπως αναφερθήκαμε στην **Ενότητα 2.2**. Εάν υποθέσουμε ότι διαθέτουμε μία μεταβλητή απόκρισης Y και $\mathbf{X} = \{X_1, X_2, \dots, X_p\}$ ένα σύνολο επεξηγηματικών μεταβλητών, τότε το σύνολο \mathcal{M} όλων των διαθέσιμων μοντέλων θα αποτελείται συνολικά από 2^p μοντέλα, δηλαδή $\mathcal{M} = \{M_1, M_2, \dots, M_{2^p}\}$ όπου το M_γ μοντέλο ανταποκρίνεται στο γ υποσύνολο του συνόλου \mathbf{X} των επεξηγηματικών μεταβλητών. Για το M_γ μοντέλο ορίζουμε την εκ των προτέρων πιθανότητα του $\pi(M_\gamma)$ και τις εκ των προτέρων κατανομές των συντελεστών του μοντέλου $\pi(\boldsymbol{\beta}_\gamma | M_\gamma, \gamma)$. Για να προχωρήσουμε στην σύγκριση των υποψηφίων μοντέλων του συνόλου \mathcal{M} καταφεύγουμε μέσω του θεωρήματος Μπεϋες στον υπολογισμό της εκ των υστέρων πιθανότητας $\pi(M_\gamma | Y)$ του αντίστοιχου μοντέλου M_γ και το μοντέλο με την μεγαλύτερη εκ των υστέρων πιθανότητα θεωρείται το καλύτερο δυνατό για την περιγραφή του προβλήματος που μελετάμε.

Παρόλο που η Μπεϋζιανή Στατιστική προσεγγίζει πιο άμεσα το πρόβλημα επιλογής μεταβλητών εμφανίζονται ορισμένες δυσκολίες σχετικά με την επιλογή της εκ των προτέρων κατανομής των συντελεστών του μοντέλου M_γ , $\pi(\boldsymbol{\beta}_\gamma | M_\gamma, \gamma)$, καθώς και με

τον υπολογισμό της εκ των υστέρων κατανομής του μοντέλου M_γ , όταν το σύνολο των επεξηγηματικών μεταβλητών είναι αρκετά μεγάλο. Αντίστοιχα, όπως αναφέρουν και οι Garthwaite & Dickey (1996), όταν το σύνολο των επεξηγηματικών μεταβλητών είναι μικρό προκύπτουν οι ανάλογες δυσκολίες.

Παλαιότερα αρκετές μέθοδοι (π.χ. *BIC Ενότητα 2.6.1.2*) εστιάζονταν σε ημιαυτόματες μεθόδους οι οποίες προσπαθούσαν να εξαλείψουν αυτήν την εξάρτηση από τις εκ των προτέρων κατανομές των συντελεστών του μοντέλου. Μία άλλη αντίστοιχη ιδέα είχε προταθεί από τον Lindley (1968) όπου υποδείκνυε την χρήση ομοιόμορφων κατανομών και για την εξαγωγή συμπερασμάτων τα οποία θα βασίζονταν στην χρήση μίας συνάρτησης ζημίας.

Με την δημιουργία των αλγορίθμων MCMC η μπεϋζιανή επιλογή μεταβλητών προχώρησε στην εξερεύνηση μοντέλων με πλήρως καθορισμένες τις εκ των προτέρων κατανομές των συντελεστών του M_γ μοντέλου $\pi(\beta_\gamma | M_\gamma, \gamma)$. Για τον υπολογισμό των εκ των υστέρων πιθανοτήτων των διαθέσιμων μοντέλων του συνόλου \mathcal{M} χρησιμοποιούνται πλέον στοχαστικοί αλγόριθμοι εξερεύνησης μοντέλων υψηλής εκ των υστέρων πιθανότητας, χωρίς να δημιουργούνται τα κλασσικά προβλήματα υπολογισμών εκ των υστέρων πιθανοτήτων. Η ιδέα των αλγορίθμων αυτών βασίζεται στην προσομοίωση μίας αλυσίδας Markov η οποία συγκλίνει στην εκ των υστέρων κατανομή $\pi(\beta_\gamma | Y, X)$ των συντελεστών του M_γ μοντέλου και θα οδηγεί τα μοντέλα με υψηλές εκ των υστέρων πιθανότητες να εμφανίζονται συχνότερα και νωρίτερα κατά την εξερεύνηση στον χώρο των διαθέσιμων μοντέλων.

2.6 Ασυμπτωτικές Προσεγγίσεις του παράγοντα Μπεϋες

Το ολοκλήρωμα το οποίο παρουσιάστηκε στην σχέση (K2.1) είναι εφικτό να υπολογιστεί αναλυτικά μόνο σε πολύ συγκεκριμένες περιπτώσεις όπου ο παραμετρικός χώρος Θ είναι μονοδιάστατος, κάτι το οποίο συμβαίνει πολύ σπάνια. Επιπλέον όπως αναφέρεται σε μία εργασία του DeGroot (1970) οι αναλυτικοί υπολογισμοί του ολοκληρώματος (K2.1) είναι εφικτοί μέσω της χρήσης συζυγών εκ των προτέρων κατανομών. Σε περιπτώσεις όπου ο παραμετρικός χώρος είναι αρκετά μεγάλης διάστασης οι υπολογισμοί είναι αρκετά περίπλοκοι και συνήθως καταφεύγουμε σε αριθμητικές μεθόδους υπολογισμού ολοκληρωμάτων. Επίσης σημαντικό ρόλο διαδραματίζει το μέγεθος του δείγματος το οποίο εξετάζεται, καθώς όσο μεγαλύτερο είναι το δείγμα τόσο δυσκολότερη είναι η εξεύρεση της μέγιστης τιμής του ολοκληρώματος (K2.1) λόγω της δυσκολίας εξερεύνησης χώρων μεγάλων διαστάσεων. Τέτοιων ειδών προβλήματα επιλύονται μέσω διαφόρων αριθμητικών μεθόδων ολοκλήρωσης και ασυμπτωτικών προσεγγίσεων.

2.6.1 Μέθοδος του Laplace

Οι Tierney & Kadane (1986) χρησιμοποίησαν μία ενδιαφέρουσα προσέγγιση του παράγοντα Μπεϋες μέσω της *προσέγγισης Laplace (Laplace approximation)* κάνοντας χρήση του τύπου

$$m_i(\mathbf{x}) \approx (2\pi)^{\frac{d_i}{2}} |\tilde{\Sigma}_i|^{-\frac{1}{2}} f(\mathbf{x}|\tilde{\theta}_i, M_i) \pi(\tilde{\theta}_i|M_i),$$

όπου d_i η διάσταση του μοντέλου M_i , $\tilde{\Sigma} = \left(-D^2\tilde{l}(\tilde{\theta}_i)\right)^{-1}$ ο πίνακα συνδιασπορών όπου $D^2\tilde{l}(\tilde{\theta}_i)$ ο *Εσσιανός πίνακας (Hessian matrix)* του λογαρίθμου της εκ των υστέρων κατανομής, υπολογισμένος στην κορυφή της $\tilde{\theta}_i$. Η συγκεκριμένη προσέγγιση βρίσκει εφαρμογή όταν η εκ των υστέρων κατανομή παρουσιάζει μόνο μία κορυφή ή υπάρχει μία κορυφή η οποία υπερέχει ισχυρότερα των υπολοίπων.

Συχνά οι υπολογισμοί που απαιτούνται για τον καθορισμό του εσσιανού πίνακα αλλά και της εκ των υστέρων κορυφής, είναι αρκετά πολύπλοκοι όπως και χρονοβόροι. Προκειμένου να περιορίσουμε το υπολογιστικό κόστος που προκύπτει κατά την εφαρμογή της μεθόδου, καταφεύγουμε στον εκτιμητή Laplace-Metropolis ο οποίος προτάθηκε από τους Raftery (1996b) και Lewis & Raftery (1997). Πραγματοποιώντας την προσέγγιση αυτή καταφεύγουμε στην χρήση αλγορίθμων MCMC, οι οποίοι οδηγούν στην εκτίμηση της εκ των υστέρων κορυφής και του εσσιανού πίνακα, από τον αριθμητικό μέσο και τον πίνακα διασποράς-συνδιασποράς αντίστοιχα των προσομοιωμένων τιμών $\theta_i^{(t)}$. Συνεπώς ο εκτιμητής Laplace-Metropolis θα δίνεται από

$$f(\mathbf{x}|M_i) \approx f_1(\mathbf{x}|M_i) = (2\pi)^{d_i/2} |\mathbf{S}_i|^{1/2} f(\mathbf{x}|\bar{\theta}_i, M_i) f(\bar{\theta}_i|M_i),$$

όπου

$$\bar{\theta}_i = \frac{1}{T} \sum_{t=1}^T \theta_i^{(t)} \quad \text{και} \quad \mathbf{S}_i = \frac{1}{T-1} \sum_{t=1}^T (\theta_i^{(t)} - \bar{\theta}_i)(\theta_i^{(t)} - \bar{\theta}_i)^T.$$

Είναι σημαντικό να επισημάνουμε ότι για να λειτουργήσει αποτελεσματικά η συγκεκριμένη μέθοδος, είναι απαραίτητο η εκ των υστέρων κατανομή της παραμέτρου θ_i να είναι συμμετρική, συνεπώς η αντικατάσταση της εκ των υστέρων κορυφής από τον αριθμητικό μέσο των προσομοιωμένων παρατηρήσεων της εκ των υστέρων κατανομής, δεν θα επηρεάσει τα αποτελέσματα.

Οι Kass & Vaidyanathan (1992) χρησιμοποίησαν μία παραλλαγή της παραπάνω προσέγγισης, αντικαθιστώντας τον Εσσιανό πίνακα $\tilde{\Sigma}$ με τον πίνακα $\hat{\Sigma}$, όπου ο $\hat{\Sigma}^{-1}$ είναι ο αντίστροφος Εσσιανός πίνακας της λογαριθμικής πιθανοφάνειας υπολογισμένος στον εκτιμητή μέγιστης πιθανοφάνειας $\hat{\theta}_i$, καταλήγοντας στον τύπο

$$\hat{m}_i(\mathbf{x}) = (2\pi)^{\frac{d_i}{2}} |\hat{\Sigma}|^{-\frac{1}{2}} f(\mathbf{x}|\hat{\theta}_i, M_i) \pi(\hat{\theta}_i|M_i). \quad (K2.3)$$

Για να καταστεί εφικτή η προσέγγιση των Kass & Vaidyanathan (1992) είναι απαραίτητη η ικανοποίηση των προϋποθέσεων που θέσανε οι Kass & Tierney & Kadane (1990). Τα αποτελέσματα της προσέγγισης των Kass & Vaidyanathan (1992) παρουσιάζουν μικρότερη ακρίβεια σχετικά με την αρχική προσέγγιση που παρουσίασαν οι Tierney & Kadane (1986), όμως οι απαραίτητοι υπολογισμοί για την διενέργεια της δεύτερης προσέγγισης είναι απλούστεροι της αρχικής καθώς πραγματοποιούνται από την πλειοψηφία των διαθέσιμων στατιστικών πακέτων.

2.6.2 Κριτήριο Schwarz

Μία διαφορετική προσέγγιση του παράγοντα Μπεϋες για την σύγκριση δύο μοντέλων M_i και M_j πραγματοποιείται μέσω της τιμής S του *κριτηρίου Schwarz* (*Schwarz criterion*) η οποία περιγράφεται από

$$S_{ij} = \log f(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, M_i) - \log f(\mathbf{x}|\hat{\boldsymbol{\theta}}_j, M_j) - \frac{1}{2}(d_i - d_j)\log(n) \quad (K2.4)$$

όπου $\hat{\boldsymbol{\theta}}_i$ ο εκτιμητής μέγιστης πιθανοφάνειας των παραμέτρων του μοντέλου M_i , διάστασης d_i και n το μέγεθος δείγματος. Όπως αναφέρουν οι Kass & Raftery (1995), ικανοποιεί την συνθήκη

$$\frac{S_{ij} - \log BF_{ij}}{\log BF_{ij}} \rightarrow 0 \text{ για } n \rightarrow \infty \quad (K2.5)$$

συνεπώς η τιμή S μπορεί να προσεγγίσει τον λογάριθμο του παράγοντα Μπεϋες του μοντέλου M_i εναντίον του μοντέλου M_j χωρίς να είναι αναγκαίος ο καθορισμός των εκ των προτέρων κατανομών των παραμέτρων $\pi(\boldsymbol{\theta}_i|M_i)$ και $\pi(\boldsymbol{\theta}_j|M_j)$. Επομένως μετά τον καθορισμό της τιμής S , βασιζόμαστε στην λογαριθμική κλίμακα του Παράγοντα Μπεϋες όπως περιγράφεται στον **Πίνακα 2.1** και τον **Πίνακα 2.2** για την επιλογή ενός εκ των δύο μοντέλων.

Μέσω της τιμής S υπολογίζεται επίσης το *κριτήριο BIC* (*Bayesian Information Criterion*) όπου για το μοντέλο M_i ορίζεται ως

$$BIC(i) \equiv D(\hat{\boldsymbol{\theta}}_i, i) + d_i \log n,$$

με $D(\hat{\boldsymbol{\theta}}_i, i) = -2 \log f(\mathbf{x}|\hat{\boldsymbol{\theta}}_i, i)$ το κλασσικό μέτρο deviance για το μοντέλο M_i υπολογισμένο στον εκτιμητή μέγιστης πιθανοφάνειας $\hat{\boldsymbol{\theta}}_i$ των παραμέτρων $\boldsymbol{\theta}_i$ διάστασης d_i του μοντέλου M_i . Ο Schwarz χρησιμοποιώντας την σχέση (K2.5), απέδειξε ότι για δύο ανταγωνιστικά μοντέλα M_i και M_j ,

$$-2 \log BF_{ij} \approx BIC(i) - BIC(j).$$

Άρα το κριτήριο BIC χρησιμοποιείται για την προσέγγιση της ποσότητας $-2\log BF_{ij}$ και με την βοήθεια του **Πίνακα 2.1** και του **Πίνακα 2.2** μπορούμε να προχωρήσουμε στην σύγκριση των μοντέλων M_i και M_j αντίστοιχα.

Σε περιπτώσεις αρκετά μεγάλων δειγμάτων όπως θα δούμε και στην συνέχεια, ο παράγοντας Μπεϋες εμφανίζει ορισμένες αδυναμίες, όμως το κριτήριο Schwarz σε συνδυασμό με την χρήση της λογαριθμικής κλίμακας του παράγοντα Μπεϋες είναι ικανό να μας δώσει σαφείς ενδείξεις υπέρ κάποιου μοντέλου. Ένα επίσης σημαντικό πλεονέκτημα του κριτηρίου Schwarz είναι εμφανές παρατηρώντας την σχέση (Κ2.4), καθώς κατά την διαδικασία υπολογισμού της τιμής S δεν διαδραματίζουν κάποιο ρόλο οι εκ των προτέρων κατανομές των παραμέτρων και οι εκ των προτέρων πιθανότητες των μοντέλων. Συνεπώς το κριτήριο Schwarz βρίσκει εφαρμογή σε περιπτώσεις όπου δεν μπορεί να καθοριστούν οι εκ των προτέρων κατανομές.

2.6.3 Δειγματοληψία Σπουδαιότητας και Monte Carlo ολοκλήρωση

Ένας άλλος τρόπος προσέγγισης του ολοκληρώματος (Κ2.1) επιτυγχάνεται μέσω της ολοκλήρωσης Monte Carlo σε συνδυασμό με την χρήση **δειγματοληψίας σπουδαιότητας (importance sampling)**. Για το μοντέλο M_i θεωρούμε ένα τυχαίο δείγμα μεγέθους m $\{\theta_i^{(k)} : k = 1, \dots, m\}$ από την εκ των προτέρων κατανομή $\pi(\theta_i | M_i)$.

Ένας τρόπος εκτίμησης της περιθώριας κατανομής του μοντέλου M_i πραγματοποιείται μέσω του υπολογισμού της **σταθμισμένης πιθανοφάνειάς (weighted likelihood)** του, δηλαδή

$$\hat{m}_i(x) = \frac{1}{m} \sum_{i=1}^m f(x | \theta_i^{(k)}).$$

Όταν η εκ των προτέρων κατανομή έχει ανόμοιο σχήμα με αυτό της εκ των υστέρων κατανομής, τότε η προσομοίωση κρίνεται αναποτελεσματική δημιουργώντας έτσι αστάθεια στην εκτίμηση της σταθμισμένης πιθανοφάνειας του μοντέλου M_i . Οι McCulloch & Rossi (1991) μελέτησαν αρκετές εφαρμογές της συγκεκριμένης μεθοδολογίας και διαπίστωσαν ότι η διασπορά της εκτιμώμενης πιθανοφάνειας ήταν αρκετά μεγάλη και συνέκλινε ασυμπτωτικά με αρκετά αργό ρυθμό στην Κανονική κατανομή.

Ένας τρόπος βελτίωσης της ακρίβειας της ολοκλήρωσης Monte Carlo επιτυγχάνεται μέσω της δειγματοληψίας σπουδαιότητας, όπου προσομοιώνουμε τιμές $\{\theta_i^{(k)} : k = 1, \dots, m\}$ από μία κατανομή $\pi^*(\theta_i)$ η οποία ονομάζεται **συνάρτηση δειγματοληψίας σπουδαιότητας (importance sampling function)**, η οποία αναλύεται εκτενέστερα στην εργασία του Geweke (1989). Μετά την προσομοίωση των τιμών $\theta_i^{(k)}$ προχωράμε στην εκτίμηση της σταθμισμένης περιθώριας πιθανοφάνεια μέσω της ποσότητας

$$\hat{m}_i(\mathbf{x}) = \frac{\sum_{k=1}^m w_k f(\mathbf{x} | \boldsymbol{\theta}_i^{(k)}, M_i)}{\sum_{k=1}^m w_k} \quad (K2.6)$$

όπου $w_k = \pi(\boldsymbol{\theta}_i^{(k)} | M_i) / \pi^*(\boldsymbol{\theta}_i^{(k)})$ (K2.7). Οι McCulloch & Rossi (1991) αναφέρουν χαρακτηριστικά ότι η συγκεκριμένη μεθοδολογία εκτίμησης της περιθώριας πιθανοφάνειας βρίσκει άμεση εφαρμογή όταν ο παραμετρικός χώρος $\boldsymbol{\theta}$ έχει το πολύ 9 διαστάσεις.

2.6.4 Προσομοίωση από την εκ των υστέρων κατανομή

Τα τελευταία χρόνια έχουν αναπτυχθεί αρκετές μεθοδολογίες σχετικά με την προσομοίωση τιμών από την εκ των υστέρων κατανομή. Σε απλά παραδείγματα δημιουργούμε απευθείας τιμές από την εκ των υστέρων κατανομή ή καταφεύγουμε σε μεθόδους προσομοίωσης τιμών κάνοντας χρήση της μεθόδου Απόρριψης. Σε πιο δύσκολες περιπτώσεις καταφεύγουμε σε MCMC αλγόριθμους όπως ο *Metropolis-Hastings* και ο *δειγματολήπτης Gibbs*. Σκοπός των μεθοδολογιών αυτών είναι η δημιουργία ενός προσομοιωμένου τυχαίου δείγματος $\{(\boldsymbol{\theta} | \mathbf{x})_i^{(k)} : k = 1, \dots, t\}$ από την εκ των υστέρων κατανομή της παραμέτρου $\boldsymbol{\theta}_i$ με συνάρτηση πυκνότητας πιθανότητας

$$\pi(\boldsymbol{\theta}_i | \mathbf{x}) = \frac{f(\mathbf{x} | \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x} | \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i}.$$

Αντικαθιστώντας στην σχέση (K2.7) την συνάρτηση $\pi^*(\boldsymbol{\theta}_k^{(i)})$ με την $\pi(\boldsymbol{\theta}_k^{(i)} | \mathbf{x})$ η σχέση (K2.6) διαμορφώνεται ως

$$\hat{m}_i(\mathbf{x}) = \left\{ \frac{1}{t} \sum_{k=1}^t [f(\mathbf{x} | \boldsymbol{\theta}_i^{(k)}, M_i)]^{-1} \right\}^{-1}$$

όπου όπως αναφέρουν οι Newton & Raftery (1994) είναι ο αρμονικός μέσος των τιμών της περιθώριας πιθανοφάνειας. Όσο το μέγεθος t των προσομοιωμένων τιμών τείνει στο άπειρο, η εκτιμώμενη συνάρτηση περιθώριας πιθανοφάνειας $\hat{m}_i(\mathbf{x})$ θα συγκλίνει στην πραγματική περιθώρια πιθανοφάνεια $m_i(\mathbf{x} | M_i)$, χωρίς να είναι απαραίτητο να συγκλίνει ασυμπτωτικά στην κανονική κατανομή. Οι Carlin & Chib (1993) αναφέρουν χαρακτηριστικά ότι παρά την ευκολία του υπολογισμού της εκτιμώμενης συνάρτησης περιθώριας πιθανοφάνειας $\hat{m}_i(\mathbf{x} | M_i)$, παρουσιάζονται ορισμένες αστάθειες στον υπολογισμό των τελικών παραγόντων Μπεϋες οι οποίοι μπορούν να παρακαμφθούν μερικώς κάνοντας χρήση της λογαριθμικής κλίμακας του παράγοντα Μπεϋες.

2.7 Επίλογος 2^{ου} Κεφαλαίου

Σε αυτό το κεφάλαιο έγινε μία αρχική παρουσίαση του προβλήματος επιλογής ενός μοντέλου και πως διαμορφώνεται στην Μπεϋζιανή Στατιστική. Για τον υπολογισμό των παραγόντων Μπεϋες όπως παρουσιάστηκε, είναι αναγκαίος ο υπολογισμός της περιθώριας πιθανοφάνειας των δεδομένων κάτω από το κάθε μοντέλο, όπου για τον καθορισμό της όπως παρουσιάστηκε στην σχέση (Κ2.1), χρειάζεται προσεκτικός χειρισμός της εκ των προτέρων κατανομής των παραμέτρων. Στο **Κεφάλαιο 3** θα μελετηθεί αναλυτικά ο ρόλος της εκ των προτέρων κατανομής των παραμέτρων καθώς και πως πραγματοποιείται η επιλογή ενός μοντέλου σε περιπτώσεις μη κατάλληλων εκ των προτέρων κατανομών.

Κεφάλαιο 3

Παραλλαγές του παράγοντα Μπεϋες

Όπως παρουσιάστηκε στο *Κεφάλαιο 2*, για να προχωρήσουμε στην σύγκριση μοντέλων καταφεύγουμε στον υπολογισμό παραγόντων Μπεϋες και λόγων εκ των υστέρων συμπληρωματικών πιθανοτήτων, όπου συνήθως συμπίπτουν λόγω της χρήσης ομοιόμορφης εκ των προτέρων πιθανότητας στον χώρο \mathcal{M} των μοντέλων.

Συνεπώς είναι απαραίτητος ο υπολογισμός της περιθώριας πιθανοφάνειας του εκάστοτε μοντέλου $M_i \in \mathcal{M}$ όπως αυτή παρουσιάστηκε στον τύπο (K2.1) της *Ενότητας 2.2*. Για την διενέργεια των κατάλληλων υπολογισμών της περιθώριας πιθανοφάνειας είναι αναγκαίος ο καθορισμός των εκ των προτέρων κατανομών των παραμέτρων $\pi(\theta_i | M_i)$. Επομένως το πρώτο κλασσικό ζήτημα που συναντάμε στην μπεϋζιανή επιλογή μοντέλου είναι η σωστή επιλογή της εκ των προτέρων κατανομής των παραμέτρων και στην συνέχεια εξετάζουμε την επίδραση της στον παράγοντα Μπεϋες.

Συνεπώς συναντάμε μία βασική αδυναμία της Μπεϋζιανής επιλογής μοντέλου η οποία αφορά στην επίδραση των εκ των προτέρων κατανομών των παραμέτρων. Στην συνέχεια θα παρουσιάσουμε ένα παράδοξο που συνέβη κατά την διενέργεια ενός απλού ελέγχου υποθέσεων, όπου εξετάστηκε η ίδια υπόθεση από Μπεϋζιανή και Κλασσική σχολή και έδωσαν αντίθετα αποτελέσματα.

3.1 Το παράδοξο των Lindley-Bartlett

Το 1957 ο Lindley σε μία εργασία του παρουσίασε ένα παράδειγμα ελέγχου υποθέσεων στο οποίο οι μέθοδοι Κλασσικής και Μπεϋζιανής Στατιστικής έδιναν αντίθετα αποτελέσματα. Έστω τυχαίο δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενο από κανονική κατανομή $N(\theta, \sigma^2)$ με άγνωστη μέση τιμή θ και γνωστή διασπορά σ^2 . Οι υποθέσεις οι οποίες θέλουμε να εξετάσουμε είναι οι

$$H_0: \theta = \theta_0 \text{ με } \theta_0 \text{ γνωστό}$$

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

Επιλέγουμε ως εκ των προτέρων πιθανότητα της μηδενικής υπόθεσης $\pi(H_0) = c$ με $c \in [0,1]$ και αντίστοιχα ως εκ των προτέρων πιθανότητα της εναλλακτικής υπόθεσης θα είναι $\pi(H_1) = 1 - c$.

Παρατηρούμε ότι κάτω από την μηδενική υπόθεση το μοντέλο το οποίο εξετάζουμε δεν έχει παραμέτρους, ενώ κάτω από την εναλλακτική υπόθεση έχουμε την παράμετρο θ , όπου η εκ των προτέρων κατανομή της θ ως υποθέσουμε ότι είναι κανονική κατανομή

με μέση τιμή θ_0 και διασπορά σ_θ^2 . Ο εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων υπέρ της H_0 θα δίνεται από

$$PO_{01} = \frac{\pi(H_0)}{\pi(H_1)} \sqrt{1 + n \frac{\sigma_\theta^2}{\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \theta_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n(\bar{x} - \theta_0)^2}{1 + \frac{n\sigma_\theta^2}{\sigma^2}} \right] \right\}.$$

Ο Lindley εφαρμόζοντας τον έλεγχο t της Κλασσικής Στατιστικής αποφάσισε την απόρριψη της μηδενικής υπόθεσης $H_0: \theta = \theta_0$ σε επίπεδο σημαντικότητας α , όταν η τιμή της στατιστικής συνάρτησης \bar{X} λάμβανε τιμές εκτός του διαστήματος $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$. Στην συνέχεια μελέτησε δείγματα παρατηρήσεων για τα οποία η τιμή της στατιστικής συνάρτησης \bar{X} κρινόταν οριακά στατιστικά σημαντική στο επίπεδο σημαντικότητας g , δηλαδή η στατιστική συνάρτηση \bar{X} θα λαμβάνει τιμές στο διάστημα $\bar{x} \pm z_{g/2}\sigma/\sqrt{n}$, όπου $z_{g/2}$ είναι το αντίστοιχο σημείο της τυπικής κανονικής κατανομής στο οποίο ο Lindley παρήγαγε τα αντίστοιχα τυχαία δείγματα.

Επομένως αφού βρισκόμαστε στα όρια της στατιστικής σημαντικότητας δεν μπορούμε να ισχυριστούμε εάν θα απορρίπταμε την μηδενική υπόθεση ή όχι με βεβαιότητα. Όμως στην Μπεϋζιανή Στατιστική μπορούμε να εξαγάγουμε ένα ασφαλές συμπέρασμα μέσω του εκ των υστέρων λόγου συμπληρωματικών πιθανοτήτων. Όμως ο Lindley παρατήρησε ότι όσο το μέγεθος του δείγματος αυξάνεται, ο λόγος των εκ των υστέρων συμπληρωματικών πιθανοτήτων θα παίρνει όλο και μεγαλύτερες τιμές, συνεπώς θα ισχυροποιείται περισσότερο η μηδενική υπόθεση.

Στην συνέχεια ο Bartlett την ίδια χρονιά δουλεύοντας στην ίδια εργασία κατέληξε στο εξής παρόμοιο συμπέρασμα. Ακολουθώντας την ίδια διαδικασία έφτασε στον ίδιο λόγο εκ των υστέρων συμπληρωματικών πιθανοτήτων, όπου παρατήρησε ότι για μεγάλες τιμές της σ_θ^2 ισχυροποιείται η μηδενική υπόθεσή καθώς ο παραπάνω λόγος συμπληρωματικών πιθανοτήτων θα λάμβανε όλο και μεγαλύτερες τιμές. Δίνοντας λοιπόν μεγάλες τιμές στην διακύμανση σ_θ^2 , η πυκνότητα της εκ των προτέρων πιθανότητας της τυχαίας μεταβλητής θ γίνεται όλο και περισσότερο διακεχυμένη.

Συνοψίζοντας τα αποτελέσματα των Lindley και Bartlett έχουμε

- Για $n \rightarrow \infty$, $PO_{01} \rightarrow \infty$, Lindley.
- Για $\sigma_\theta^2 \rightarrow \infty$, $PO_{01} \rightarrow \infty$, Bartlett.

Επομένως σε περιπτώσεις που η Κλασσική Στατιστική δυσκολεύεται να δώσει αποτελέσματα ή απορρίπτει το απλούστερο μοντέλο, η Μπεϋζιανή Στατιστική υποστηρίζει το απλούστερο μοντέλο.

Μία λύση του αρχικού παραδόξου του Lindley, είναι η επιλογή ως διακύμανση της εκ των προτέρων κατανομής της σ_θ^2/n . Όμως η επιλογή της σ_θ^2 ξεχωριστά είναι ιδιαίτερα δύσκολη σε περιπτώσεις όπου εξετάζουμε μη πληροφοριακές κατανομές.

3.2 Χρήση μη κατάλληλων πληροφοριακών κατανομών

Κατά την διενέργεια μίας Μπεϋζιανής επιλογής μοντέλου σπάνια έχουμε επαρκή πληροφορία για τον καθορισμό της εκ των προτέρων κατανομής των παραμέτρων του εκάστοτε μοντέλου, γεγονός το οποίο μας οδηγεί στην χρήση μη πληροφοριακών κατανομών. Χρησιμοποιώντας τον διαχωρισμό των μη-πληροφοριακών εκ των προτέρων κατανομών της **Ενότητας 1.2.2** σε κατάλληλες και μη κατάλληλες κατανομές, καλούμαστε να εξετάσουμε πως συμπεριφέρονται οι παράγοντες Μπεϋες.

Ως μη κατάλληλες κατανομές θεωρούμε τις εκ των προτέρων κατανομές οι οποίες δίνονται στην μορφή

$$\pi^N(\boldsymbol{\theta}_i|M_i) = \frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i} \quad (K3.1)$$

όπου ο όρος $\pi^{N^*}(\boldsymbol{\theta}_i|M_i)$ αποτελεί τον πυρήνα της κατανομής καθώς περιέχει την διαθέσιμη πληροφορία της παραμέτρου $\boldsymbol{\theta}_i$ και ο όρος c_i εκπροσωπεί την αυθαίρετη σταθερά κανονικοποίησης.

Θεωρούμε δύο ανταγωνιστικά μοντέλα M_i και M_j από έναν χώρο μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με ίσες εκ των προτέρων πιθανότητες $\pi(M_i) = \pi(M_j) = \frac{1}{k}$ και ένα τυχαίο δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ το οποίο έχει συνάρτηση πυκνότητας πιθανότητας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ κάτω από το μοντέλο M_i . Επιλέγουμε ως εκ των προτέρων κατανομές των παραμέτρων $\boldsymbol{\theta}_i$ και $\boldsymbol{\theta}_j$ αντίστοιχα τις μη κατάλληλες εκ των προτέρων κατανομές $\pi^N(\boldsymbol{\theta}_i|M_i)$ και $\pi^N(\boldsymbol{\theta}_j|M_j)$ όπως ορίζονται από την σχέση (K3.1). Τότε ο παράγοντας Μπεϋες υπέρ του μοντέλου M_i θα δίνεται ως

$$\begin{aligned} BF_{ij} &= \frac{m_i(\mathbf{x})}{m_j(\mathbf{x})} \\ &= \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \pi^N(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \pi^N(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j} \\ &= \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} d\boldsymbol{\theta}_j} \\ &= \frac{c_j \int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \pi^{N^*}(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{c_i \int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \pi^{N^*}(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j}. \quad (K3.2) \end{aligned}$$

Παρατηρούμε ότι ο παράγοντας Μπεϋες υπέρ του μοντέλου M_i δεν μπορεί να υπολογιστεί καθώς εξαρτάται από τον λόγο των δύο αυθαίρετων σταθερών c_j/c_i , επομένως δεν είμαστε σε θέση να εξαγάγουμε συμπεράσματα.

Βλέπουμε ότι ενώ αναγκαζόμαστε να χρησιμοποιούμε αρκετά συχνά μη κατάλληλες εκ των προτέρων κατανομές, δεν είναι εύκολο να προχωρήσουμε σε ελέγχους υποθέσεων αλλά και σύγκριση μοντέλων διότι καταλήγουμε σε αυθαίρετους παράγοντες Μπεϋες και κατά συνέπεια σε αυθαίρετους εκ των υστέρων λόγους συμπληρωματικών πιθανοτήτων.

Λύση αυτού του προβλήματος έρχεται να δώσει μία συγκεκριμένη κατηγορία παραγόντων Μπεϋες οι οποίοι καλούνται **μερικοί παράγοντες Μπεϋες** (*partial Bayes factors*). Βασικό χαρακτηριστικό αυτών αποτελεί η χρήση ενός μέρους των δεδομένων για την μετατροπή των μη κατάλληλων εκ των προτέρων κατανομών σε κατάλληλες εκ των προτέρων κατανομές προκειμένου να γίνει εφικτός ο υπολογισμός των παραγόντων Μπεϋες. Το κομμάτι των δεδομένων που θα χρησιμοποιείται για τον καθορισμό των μη-κατάλληλων εκ των προτέρων κατανομών θα καλείται **δείγμα εκπαίδευσης** (*training sample*).

3.3 Ο εκ των υστέρων Παράγοντας Μπεϋες

Ο Aitkin (1991) εισήγαγε την έννοια του **εκ των υστέρων παράγοντα Μπεϋες** (*posterior Bayes factor*) προκειμένου να καταστεί εφικτός ο υπολογισμός παραγόντων Μπεϋες όπως περιγράφονται από την σχέση (Κ3.2). Παρατηρώντας τη μορφή της περιθώριας κατανομής ενός τυχαίου δείγματος όπως περιγράφεται από την σχέση (Κ2.1), ο Aitkin (1991) την περιέγραψε ως την εκ των προτέρων αναμενόμενη τιμή της συνάρτησης πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$. Η ιδέα του εκ των υστέρων παράγοντα Μπεϋες βασίζεται στην χρήση της εκ των υστέρων αναμενόμενης τιμής της συνάρτησης πιθανοφάνειας του τυχαίου δείγματος όπου εξετάζεται.

Κάνοντας χρήση μη κατάλληλων εκ των προτέρων κατανομών της μορφής (Κ3.1) είναι εφικτό να υπολογίσουμε κατάλληλες εκ των υστέρων κατανομές, διότι

$$\begin{aligned} \pi_i^{N^*}(\boldsymbol{\theta}_i | \mathbf{x}, M_i) &= \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i} \\ &= \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i}}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i}d\boldsymbol{\theta}_i} \\ &= \frac{c_i}{c_i} \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i} \end{aligned}$$

$$= \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i}.$$

Επομένως παρατηρούμε ότι η εκ των υστέρων κατανομή δεν εξαρτάται από κάποια άγνωστη σταθερά κανονικοποίησης, συνεπώς η εκ των υστέρων μέση τιμή της συνάρτησης πιθανοφάνειας για το μοντέλο M_i θα είναι

$$\begin{aligned}\bar{L}_i^A &= \int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi_i^{N^*}(\boldsymbol{\theta}_i|\mathbf{x}, M_i)d\boldsymbol{\theta}_i \\ &= \int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i} d\boldsymbol{\theta}_i \\ &= \int_{\boldsymbol{\theta}_i} \frac{f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)^2\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i} d\boldsymbol{\theta}_i.\end{aligned}$$

Ως εκ των υστέρων παράγοντας Μπεϋες υπέρ του μοντέλου M_i ορίζεται ο λόγος της εκ των υστέρων μέσης τιμής της συνάρτησης πιθανοφάνειας των δεδομένων κάτω από το μοντέλο M_i προς τον αντίστοιχο κάτω από το μοντέλο M_j , δηλαδή

$$PBF_{ij} = \frac{\bar{L}_i^A}{\bar{L}_j^A}$$

και η ερμηνεία του είναι παρόμοια με του κλασσικού παράγοντα Μπεϋες. Ως μεθοδολογία έχει το πλεονέκτημα ότι μας επιτρέπει την χρήση μη κατάλληλων εκ των προτέρων κατανομών, όμως δεν χρησιμοποιείται γενικά ως εναλλακτική μέθοδος υπολογισμού παραγόντων Μπεϋες. Βασικό μειονέκτημα του εκ των υστέρων Παράγοντα Μπεϋες είναι η χρήση της πληροφορίας που παρέχουν τα δεδομένα δύο φορές έχοντας ως αποτέλεσμα την παραβίαση της αρχής της πιθανοφάνειας. Επίσης όπως αναφέρει και ο Aitkin (1991) ο εκ των υστέρων παράγοντας Μπεϋες δεν προκύπτει άμεσα μέσω του θεωρήματος Μπεϋες συνεπώς δεν μπορεί να θεωρηθεί ως μία αμιγώς Μπεϋζιανή μέθοδος.

3.4 Ο κλασματικός παράγοντας Μπεϋες

Ο O'Hagan (1995) εισήγαγε τον *κλασματικό παράγοντα Μπεϋες (fractional Bayes factor)* προκειμένου να καταστεί εφικτός ο υπολογισμός παραγόντων Μπεϋες όπως περιγράφονται από την σχέση (Κ3.2). Η ιδέα πίσω από την μεθοδολογία αυτή είναι η χρήση ενός μέρους των δεδομένων ως δείγμα εκπαίδευσης ώστε μπορέσουμε να απαλείψουμε τις αυθαίρετες σταθερές κανονικοποίησης και τα υπόλοιπα δεδομένα να χρησιμοποιηθούν για συμπερασματολογία.

Για να προχωρήσουμε στον ορισμό του κλασματικού παράγοντα Μπεϋες ορίζουμε πρώτα την τιμή b με $0 \leq b \leq 1$, η οποία υποδεικνύει το ποσοστό της πληροφορίας που

παρέχουν τα δεδομένα η οποία θα χρησιμοποιηθεί για την απαλοιφή των αυθαίρετων σταθερών κανονικοποίησης.

Συνεπώς ο κλασματικός παράγοντας Μπεϋες υπέρ του μοντέλου M_i ορίζεται ως

$$\text{FBF}_{ij} = \frac{q_i(b, \mathbf{x})}{q_j(b, \mathbf{x})}$$

όπου

$$q_i(b, \mathbf{x}) = \frac{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \pi^N(\theta_i|M_i) d\theta_i}{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i)^b \pi^N(\theta_i|M_i) d\theta_i}.$$

Οι ποσότητες q_i όπως αποδεικνύεται δεν εξαρτώνται από τις αυθαίρετες σταθερές κανονικοποίησης και κατά συνέπεια ούτε και ο κλασματικός παράγοντας Μπεϋες, επομένως είμαστε σε θέση να προχωρήσουμε σε χρήση μη κατάλληλων εκ των προτέρων κατανομών της μορφής (Κ3.1). Επομένως

$$\begin{aligned} q_i(b, \mathbf{X}) &= \frac{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \pi^N(\theta_i|M_i) d\theta_i}{\int_{\theta_i} (f(\mathbf{x}|\theta_i, M_i))^b \pi^N(\theta_i|M_i) d\theta_i} \\ &= \frac{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \frac{\pi^{N^*}(\theta_i|M_i)}{c_i} d\theta_i}{\int_{\theta_i} (f(\mathbf{x}|\theta_i, M_i))^b \frac{\pi^{N^*}(\theta_i|M_i)}{c_i} d\theta_i} \\ &= \frac{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \pi^{N^*}(\theta_i|M_i) d\theta_i}{\int_{\theta_i} (f(\mathbf{x}|\theta_i, M_i))^b \pi^{N^*}(\theta_i|M_i) d\theta_i}. \end{aligned}$$

Οπότε η τελική μορφή του κλασματικού παράγοντα Μπεϋες θα δίνεται από

$$\text{FBF}_{ij} = \frac{\frac{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \pi^{N^*}(\theta_i|M_i) d\theta_i}{\int_{\theta_i} (f(\mathbf{x}|\theta_i, M_i))^b \pi^{N^*}(\theta_i|M_i) d\theta_i}}{\frac{\int_{\theta_j} f(\mathbf{x}|\theta_j, M_j) \pi^{N^*}(\theta_j|M_j) d\theta_j}{\int_{\theta_j} (f(\mathbf{x}|\theta_j, M_j))^b \pi^{N^*}(\theta_j|M_j) d\theta_j}}.$$

Ο Κλασματικός Παράγοντας Μπεϋες του Ο'Hagan (1995) είναι ικανός να εξαλείψει την εξάρτηση που έχουν οι παράγοντες Μπεϋες από τις αυθαίρετες σταθερές κανονικοποίησης, όμως εμφανίζει την ίδια αδυναμία με αυτήν του εκ των υστέρων παράγοντα Μπεϋες του Aitkin (1991), καθώς χρησιμοποιεί την πληροφορία που παρέχουν τα δεδομένα δύο φορές.

Εάν εφαρμόσουμε την μεθοδολογία του εκ των υστέρων παράγοντα Μπεϋες, χρησιμοποιώντας όμως ένα ποσοστό b της πιθανοφάνειας έχουμε ότι

$$\pi_i^{N^*}(\boldsymbol{\theta}_i | \mathbf{x}, M_i, b) = \frac{(f(\mathbf{x} | \boldsymbol{\theta}_i, M_i))^b \pi^{N^*}(\boldsymbol{\theta}_i | M_i)}{\int_{\boldsymbol{\theta}_i} (f(\mathbf{x} | \boldsymbol{\theta}_i, M_i))^b \pi^{N^*}(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i}$$

με απλοποιημένες τις σταθερές κανονικοποίησης c_i . Επομένως ο εκ των υστέρων Παράγοντας Μπεϋες του μοντέλου M_j εναντίον του μοντέλου M_i σε αυτήν την περίπτωση θα δίνεται από

$$\begin{aligned} PBF_{ij} &= \frac{m_i(\mathbf{x}|b)}{m_j(\mathbf{x}|b)} \\ &= \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x} | \boldsymbol{\theta}_i, M_i)^{1-b} \pi_i^{N^*}(\boldsymbol{\theta}_i | \mathbf{x}, M_i, b) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x} | \boldsymbol{\theta}_j, M_j)^{1-b} \pi_j^{N^*}(\boldsymbol{\theta}_j | \mathbf{x}, M_j, b) d\boldsymbol{\theta}_j} \\ &= \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x} | \boldsymbol{\theta}_i, M_i)^{1-b} \frac{(f(\mathbf{x} | \boldsymbol{\theta}_i, M_i))^b \pi^{N^*}(\boldsymbol{\theta}_i | M_i)}{\int_{\boldsymbol{\theta}_i} (f(\mathbf{x} | \boldsymbol{\theta}_i, M_i))^b \pi^{N^*}(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x} | \boldsymbol{\theta}_j, M_j)^{1-b} \frac{(f(\mathbf{x} | \boldsymbol{\theta}_j, M_j))^b \pi^{N^*}(\boldsymbol{\theta}_j | M_j)}{\int_{\boldsymbol{\theta}_j} (f(\mathbf{x} | \boldsymbol{\theta}_j, M_j))^b \pi^{N^*}(\boldsymbol{\theta}_j | M_j) d\boldsymbol{\theta}_j} d\boldsymbol{\theta}_j} \\ &= \frac{\frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x} | \boldsymbol{\theta}_i, M_i) \pi^{N^*}(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_i} (f(\mathbf{x} | \boldsymbol{\theta}_i, M_i))^b \pi^{N^*}(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i}}{\frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x} | \boldsymbol{\theta}_j, M_j) \pi^{N^*}(\boldsymbol{\theta}_j | M_j) d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_j} (f(\mathbf{x} | \boldsymbol{\theta}_j, M_j))^b \pi^{N^*}(\boldsymbol{\theta}_j | M_j) d\boldsymbol{\theta}_j}} = FBF_{ij}. \end{aligned}$$

Συνεπώς βλέπουμε ότι ο Κλασματικός Παράγοντας Μπεϋες συμπίπτει με τον εκ των υστέρων παράγοντα Μπεϋες όταν χρησιμοποιείται ένα ποσοστό των δεδομένων, συνεπώς παραβιάζεται και εδώ η αρχή της πιθανοφάνειας οπότε δεν μπορεί να αποτελέσει μία αξιόπιστη εναλλακτική παραλλαγή του παράγοντα Μπεϋες.

Ο κλασματικός παράγοντας Μπεϋες εξαρτάται αποκλειστικά από την τιμή b και γι' αυτό ο O'Hagan (1995) πρότεινε τρεις διαφορετικές επιλογές της βασιζόμενος στην ανθεκτικότητα του, προκειμένου να θεσπίσει έναν γενικό κανόνα καθορισμού του κλασματικού παράγοντα Μπεϋες.

- 1) $b = \frac{m_0}{n}$ όταν η ανθεκτικότητα του κλασματικού παράγοντα Μπεϋες δεν μας ενδιαφέρει ιδιαίτερα, όπου m_0 είναι το μέγεθος του μικρότερου δυνατού δείγματος εκπαίδευσης.
- 2) $b = n^{-1} \max\{m_0, n^{\frac{1}{2}}\}$ όταν η ανθεκτικότητα του κλασματικού παράγοντα Μπεϋες είναι ιδιαίτερα σημαντική.

- 3) $b = n^{-1} \max\{m_0, \log n\}$ ως μία άμεση λύση σε περιπτώσεις όπου επιλέγουμε να κρατήσουμε ουδέτερη στάση.

3.5 Ο Ενδογενής Παράγοντας Μπεϋες

Το 1996 οι Berger & Perrichi δημιούργησαν μία ευρηματική διαδικασία κατασκευής κατάλληλων εκ των προτέρων κατανομών η οποία ανταποκρίνεται στην χρήση αρχικών εκ των προτέρων κατανομών της μορφής (Κ3.1), η οποία θα παρουσιαστεί αναλυτικά στο **Κεφάλαιο 4**. Για τον καθορισμό αυτής της κατηγορίας εκ των προτέρων κατανομών χρησιμοποιείται ο *ενδογενής παράγοντας Μπεϋες (intrinsic Bayes factor)*.

Έστω τυχαίο δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και το σύνολο των υποψήφιων μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$, όπου κάτω από το μοντέλο M_i ορίζεται η συνάρτηση πυκνότητας πιθανότητας του τυχαίου δείγματος $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$, επιλέγοντας την μη κατάλληλη εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_i όπως αυτή ορίζεται από την σχέση (Κ3.1). Τότε ο παράγοντας Μπεϋες δύο ανταγωνιστικών μοντέλων M_i και M_j προκύπτει από την σχέση (Κ3.2).

Η φιλοσοφία του ενδογενή παράγοντα Μπεϋες βασίζεται στην χρήση ενός μέρους των δεδομένων \mathbf{x} ως δείγμα εκπαίδευσης το οποίο θα χρησιμοποιηθεί για την μετατροπή των μη κατάλληλων εκ των προτέρων κατανομών $\pi^N(\boldsymbol{\theta}_i|M_i)$ σε κατάλληλες εκ των υστέρων κατανομές και στην συνέχεια αξιοποιώντας τα υπόλοιπα δεδομένα να καταστεί εφικτή η σύγκριση μοντέλων. Ορίζουμε ως δείγμα εκπαίδευσης το διάνυσμα των παρατηρήσεων $\mathbf{x}(l)$ και τα υπόλοιπα δεδομένα θα περιγράφονται από το διάνυσμα $\mathbf{x}(-l)$. Ένα δείγμα εκπαίδευσης $\mathbf{x}(l)$ θα θεωρείται κατάλληλο όταν για όλα τα υποψήφια μοντέλα M_i οι περιθώριες κατανομές ικανοποιούν την σχέση $0 < m_i^N(\mathbf{x}(l)) < \infty$. Επίσης ένα δείγμα εκπαίδευσης θα καλείται ελάχιστο δείγμα εκπαίδευσης όταν δεν θα υπάρχει κάποιο υποδιάνυσμα του $\mathbf{x}(l)$ το οποίο θα ικανοποιεί την παραπάνω σχέση.

Μία κλασική κατασκευή ενός ελαχίστου δείγματος εκπαίδευσης αποτελεί η επιλογή πλήθους παρατηρήσεων ίσο με το συνολικό πλήθος των παραμέτρων στο σύνολο των μοντέλων \mathcal{M} , όπως επίσης μπορεί να είναι μεγέθους $\max\{k_i\}$ όπου k_i είναι η διάσταση της παραμέτρου $\boldsymbol{\theta}_i$ του εκάστοτε μοντέλου M_i . Τα ελάχιστα δείγματα εκπαίδευσης είναι εφικτό να κατασκευαστούν ακόμα και σε περιπτώσεις όπου δεν μελετάμε τυχαία δείγματα ανεξάρτητων παρατηρήσεων. Όταν εμφανίζεται εξάρτηση στο δείγμα των παρατηρήσεων, όπως γίνεται στις χρονοσειρές, τότε η δομή της μεταφέρεται και στην κατασκευή του δείγματος εκπαίδευσης.

Επομένως για το μοντέλο M_i θα χρησιμοποιήσουμε ως εκ των προτέρων κατανομή την

$$\pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}(l)) = \frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)}{m_i^N(\mathbf{x}(l))}$$

$$\begin{aligned}
&= \frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i} \\
&= \frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i}}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i}d\boldsymbol{\theta}_i} \\
&= \frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i}.
\end{aligned}$$

Είναι προφανές ότι η $\pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}(l))$ είναι μία κατάλληλη κατανομή, συνεπώς είναι εφικτό να χρησιμοποιηθεί ως εκ των προτέρων κατανομή για τον υπολογισμό περιθώριων πιθανοφαινιών. Συνεπώς δοθέντος ενός δείγματος εκπαίδευσης $\mathbf{x}(l)$ και δοθέντος του ότι η εκ των υστέρων κατανομή $\pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}(l))$ είναι κατάλληλη, ορίζουμε τον ενδογενή παράγοντα Μπεϋες υπέρ του μοντέλου M_j ως

$$\begin{aligned}
IBF_{ji} &= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(-l)|\boldsymbol{\theta}_j, \mathbf{x}(l), M_j)\pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}(l))d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(-l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)\pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}(l))d\boldsymbol{\theta}_i} \\
&= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(-l)|\boldsymbol{\theta}_j, \mathbf{x}(l), M_j)\frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_j, \mathbf{x}(l), M_j)\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(l)|\boldsymbol{\theta}_j, M_j)\pi^{N^*}(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j}d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(-l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)\frac{f(\mathbf{x}(l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i}d\boldsymbol{\theta}_i} \\
&= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(-l)|\boldsymbol{\theta}_j, \mathbf{x}(l), M_j)f(\mathbf{x}(l)|\boldsymbol{\theta}_j, \mathbf{x}(l), M_j)\pi^{N^*}(\boldsymbol{\theta}_j|M_j)\frac{1}{m_j^{N^*}(\mathbf{x}(l))}d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(-l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)f(\mathbf{x}(l)|\boldsymbol{\theta}_i, \mathbf{x}(l), M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)\frac{1}{m_i^{N^*}(\mathbf{x}(l))}d\boldsymbol{\theta}_i} \\
&= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)\pi^{N^*}(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j \frac{1}{m_i^{N^*}(\mathbf{x}(l))}}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^{N^*}(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i \frac{1}{m_j^{N^*}(\mathbf{x}(l))}} \\
&= \frac{m_j^{N^*}(\mathbf{x}) \frac{1}{m_i^{N^*}(\mathbf{x}(l))}}{m_j^{N^*}(\mathbf{x}) \frac{1}{m_j^{N^*}(\mathbf{x}(l))}} = BF_{ji}^{N^*} BF_{ij}^{N^*}(\mathbf{x}(l)). \quad (K3.4)
\end{aligned}$$

Ο δείκτης N υποδηλώνει την χρήση μίας μη κατάλληλης εκ των προτέρων κατανομής και ο δείκτης N^* υποδηλώνει την χρήση του πυρήνα της μη κατάλληλης εκ των προτέρων κατανομής απαλλαγμένο από τις αυθαίρετες σταθερές.

3.5.1 Αριθμητικός και γεωμετρικός ενδογενής παράγοντας Μπεϋες

Είναι εμφανές ότι ο ενδογενής παράγοντας Μπεϋες όπως τον ορίσαμε στην σχέση (K3.4) εξαρτάται άμεσα από την επιλογή του ελαχίστου δείγματος εκπαίδευσης. Όταν μελετάμε ένα τυχαίο δείγμα $\mathbf{x} = (x_1, x_2, \dots, x_n)$ υπάρχουν πολλοί διαθέσιμοι συνδυασμοί ελαχίστων δειγμάτων εκπαίδευσης. Οι Berger & Perrichi (1996) προκειμένου να απαλείψουν αυτήν την εξάρτηση από το ελάχιστο δείγμα εκπαίδευσης υπολόγισαν τον αριθμητικό και γεωμετρικό μέσο αντίστοιχα του Ενδογενούς παράγοντα Μπεϋες.

Για ένα σετ δεδομένων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ορίζουμε το σύνολο όλων των διαθέσιμων επιλογών ελαχίστων δειγμάτων εκπαίδευσης ως

$$\mathcal{X}_L = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(L)\}.$$

Δοθέντος του συνόλου \mathcal{X}_L ορίζουμε τον **αριθμητικό ενδογενή παράγοντα Μπεϋες** (*arithmetic intrinsic Bayes factor*) υπέρ του μοντέλου M_j ως

$$\begin{aligned} \text{AIBF}_{ji} &= \frac{1}{L} \sum_{l=1}^L \text{IBF}_{ji}(\mathbf{x}(l)) \\ &= \frac{1}{L} \sum_{l=1}^L \text{BF}_{ji}^{N^*} \text{BF}_{ij}^{N^*}(\mathbf{x}(l)) \\ &= \text{BF}_{ji}^{N^*} \frac{1}{L} \sum_{l=1}^L \text{BF}_{ij}^{N^*}(\mathbf{x}(l)). \quad (\text{K3.5}) \end{aligned}$$

Δοθέντος του συνόλου \mathcal{X}_L ορίζουμε επίσης τον **γεωμετρικό ενδογενή παράγοντα Μπεϋες** (*geometric intrinsic Bayes factor*) υπέρ του μοντέλου M_j ως

$$\begin{aligned} \text{GIBF}_{ji} &= \left(\prod_{l=1}^L \text{IBF}_{ji}(\mathbf{x}(l)) \right)^{1/L} \\ &= \left(\prod_{l=1}^L \text{BF}_{ji}^{N^*} \text{BF}_{ij}^{N^*}(\mathbf{x}(l)) \right)^{1/L} \\ &= \text{BF}_{ji}^{N^*} \left(\prod_{l=1}^L \text{BF}_{ij}^{N^*}(\mathbf{x}(l)) \right)^{1/L}. \quad (\text{K3.6}) \end{aligned}$$

Όπως είναι γνωστό ο γεωμετρικός μέσος θα είναι μικρότερος ή ίσος από τον αριθμητικό μέσο αυτών, οπότε κατά αντιστοιχία ο γεωμετρικός ενδογενής παράγοντας Μπεϋες θα είναι μικρότερος ή ίσος από τον αριθμητικό ενδογενή παράγοντα Μπεϋες, χαρακτηριστικό το οποίο σε περιπτώσεις σύγκρισης εμφωλευμένων μοντέλων, οδηγεί τον γεωμετρικό ενδογενή παράγοντα Μπεϋες να στηρίζει περισσότερο το απλούστερο μοντέλο συγκριτικά με τον αριθμητικό ενδογενή παράγοντα Μπεϋες.

3.5.2 Αναμενόμενος αριθμητικός και γεωμετρικός ενδογενής παράγοντας Μπεϋες

Σε περιπτώσεις όπου το δείγμα το οποίο εξετάζουμε είναι μικρό, ο αριθμητικός και ο γεωμετρικός ενδογενής παράγοντας Μπεϋες μπορούν να εμφανίσουν μεγάλες διακυμάνσεις, συνεπώς οι ενδογενείς παράγοντες Μπεϋες εμφανίζουν αρκετά μεγάλες αστάθειες. Σε περιπτώσεις αρκετά μεγάλων δειγμάτων το σύνολο των ελαχίστων δειγμάτων εκπαίδευσης θα είναι και αυτό αντίστοιχα μεγάλο, έχοντας ως αποτέλεσμα οι υπολογισμοί να γίνονται αρκετά δύσκολοι και χρονοβόροι. Για την αντιμετώπιση αυτών των προβλημάτων οι Berger & Perrichi (1996) πρότειναν τον υπολογισμό των αναμενόμενων τιμών των ενδογενών παραγόντων Μπεϋες των σχέσεων (Κ3.5) και (Κ3.6), υπολογισμένους στα σημεία στα οποία μεγιστοποιείται η συνάρτηση πιθανοφάνειας. Η ιδέα αυτή οδήγησε στην δημιουργία του *αναμενόμενου αριθμητικού ενδογενή παράγοντα Μπεϋες (expected arithmetic intrinsic Bayes factor)* και αντίστοιχα του *αναμενόμενου γεωμετρικού ενδογενή παράγοντα Μπεϋες (expected geometric intrinsic Bayes factor)*

Δοθέντος ενός ελαχίστου δείγματος εκπαίδευσης ο αναμενόμενος αριθμητικός ενδογενής παράγοντας Μπεϋες υπέρ του μοντέλου M_j ορίζεται ως

$$EAIBF_{ji} = BF_{ji}^{N^*} \frac{1}{L} \sum_{l=1}^L E_{\hat{\theta}_j}^{M_j} [BF_{ij}^{N^*}(\mathbf{X}(l))] . (K3.7)$$

Αντίστοιχα, δοθέντος ενός ελαχίστου δείγματος εκπαίδευσης ο αναμενόμενος γεωμετρικός ενδογενής παράγοντας Μπεϋες υπέρ του μοντέλου M_j ορίζεται ως

$$EGIBF_{ji} = BF_{ji}^{N^*} \exp \left\{ \frac{1}{L} \sum_{l=1}^L E_{\hat{\theta}_j}^{M_j} [\log (BF_{ij}^{N^*}(\mathbf{X}(l)))] \right\} .$$

Σε περιπτώσεις όπου το σύνολο \mathbf{X}_L είναι αρκετά μεγάλο, οι αναμενόμενοι ενδογενείς παράγοντες Μπεϋες χρησιμοποιούνται για την προσέγγιση των αντίστοιχων αριθμητικών ενδογενών παραγόντων Μπεϋες. Επίσης, όταν ισχύει η ανταλλαξιμότητα στο δείγμα παρατηρήσεων το οποίο εξετάζεται, απαλείφονται οι όροι του αθροίσματος και της διαίρεσης και υπολογίζονται μόνο οι αναμενόμενες τιμές.

3.5.3 Πλεονεκτήματα και μειονεκτήματα του ενδογενή παράγοντα Μπεϋες

Ο ενδογενής παράγοντας Μπεϋες και οι παραλλαγές του αποτελούν ένα σημαντικό βήμα για την εξέλιξη της Μπεϋζιανής Στατιστικής καθώς ανταποκρίνονται σημαντικά στην χρήση μη κατάλληλων εκ των προτέρων κατανομών, όμως συναντάμε συχνά ιδιαίτερες υπολογιστικές δυσκολίες κατά την χρήση τους. Οι Berger & Perrichi (1996) αναφέρουν χαρακτηριστικά ότι:

- Αρχικά συναντάμε τις τυπικές δυσκολίες τις οποίες εμφανίζουν και ο κλασικός παράγοντας Μπεϋες, όπως αναφέρθηκε και στο **Κεφάλαιο 2**.
- Οι αναγκαίοι υπολογισμοί του όρου $BF_{ij}^{N*}(x(l))$ της σχέσης (Κ3.4) μπορούν να γίνουν αρκετά δύσκολοι σε ορισμένες περιπτώσεις αν και συχνά είναι διαθέσιμοι σε κλειστή μορφή.
- Στην περίπτωση του αριθμητικού ενδογενούς παράγοντα Μπεϋες, ο υπολογισμός του αθροίσματος στο σύνολο \mathcal{X}_L μπορεί να αποβεί χρονοβόρος σε περιπτώσεις μεγάλων δειγμάτων. Ένας τρόπος αντιμετώπισης αυτού του προβλήματος είναι η δημιουργία ενός υποσυνόλου \mathcal{X}_{L_T} με $\mathcal{X}_{L_T} \subset \mathcal{X}_L$ το οποίο μπορεί να οριστεί είτε από μία τυχαία είτε από μία συστηματική επιλογή στοιχείων του συνόλου \mathcal{X}_L .
- Όταν δεν ισχύει η ανταλλαξιμότητα στο εξεταζόμενο δείγμα x , οι υπολογισμοί του αριθμητικού και του γεωμετρικού ενδογενούς παράγοντα Μπεϋες μπορούν να γίνουν αρκετά περίπλοκοι, εκτός εάν η αναμενόμενη τιμή των παραγόντων κάτω από το μοντέλο M_j δίνεται σε κλειστή μορφή.

Οι ενδογενείς παράγοντες Μπεϋες και οι παραλλαγές τους εμφανίζουν αρκετά πλεονεκτήματα κατά την εφαρμογή τους. Όπως αναφέρουν οι Berger & Perrichi (1996):

- Αποτελούν πλήρως αυτοματοποιημένες διαδικασίες οι οποίες βασίζονται στα παρατηρηθέντα δεδομένα και σε μη κατάλληλες εκ των προτέρων κατανομές.
- Βρίσκουν άμεση εφαρμογή σε εμφωλευμένες και μη υποθέσεις χωρίς να υπάρχει κάποιος περιορισμός ως προς την χρήση κάποιας κατανομής.
- Εφαρμόζονται άμεσα σε αυτοματοποιημένους ελέγχους υποθέσεων και σύγκριση μοντέλων.
- Εφαρμόζονται αποτελεσματικότερα σε περιπτώσεις όπου ασυμπτωτικά κριτήρια (BIC κλπ.) δεν είναι δυνατόν να εφαρμοστούν.
- Ανταποκρίνονται επαρκώς σε περιπτώσεις πολλαπλών συγκρίσεων μοντέλων και προβλέψεων λαμβάνοντας υπόψιν την αβεβαιότητα γύρω από την επιλογή ενός μοντέλου.
- Παραμένουν αναλλοίωτοι ως προς γραμμικούς μετασχηματισμούς των δεδομένων.
- Όπως θα δούμε και στην συνέχεια, προσεγγίζουν κανονικούς παράγοντες Μπεϋες μέσω ενδογενών εκ των προτέρων κατανομών.

Οι ενδογενείς παράγοντες Μπεϋες και οι παραλλαγές τους εμφανίζουν όμως ορισμένα μειονεκτήματα κατά την εφαρμογή τους. Όπως αναφέρουν οι Berger & Perrichi (1996):

- Εμφανίζονται συχνά διάφορες υπολογιστικές δυσκολίες όπως αναφερθήκαν και στην αρχή της ενότητας αυτής, όμως μπορούν να περιοριστούν καθώς οι παράγοντες οι οποίοι καθορίζουν την επιλογή των δειγμάτων εκπαίδευσης είναι συγκεκριμένοι, συνεπώς μπορούμε να επιλέξουμε δειγματοληπτικά στοιχεία από το σύνολο \mathcal{X}_L , περιορίζοντας τους υπολογισμούς.
- Σε ορισμένες περιπτώσεις είναι αναγκαία η τροποποίηση του αριθμητικού ενδογενούς παράγοντα Μπεϋες προκειμένου να εξασφαλιστεί η συνάφεια του κατά την πολλαπλή σύγκριση μοντέλων.
- Σε περιπτώσεις όπου το εξεταζόμενο δείγμα x είναι αρκετά μικρό εμφανίζονται αστάθειες στον ενδογενή παράγοντα Μπεϋες και τις αριθμητικές παραλλαγές του. Μάλιστα στην ακραία περίπτωση όπου το ελάχιστο δείγμα εκπαίδευσης είναι λίγο μικρότερο από το μέγεθος του δείγματος x είναι προτιμότερο να χρησιμοποιηθεί κάποια μορφή του αναμενόμενου ενδογενούς παράγοντα Μπεϋες.
- Ο ενδογενής παράγοντας Μπεϋες και οι παραλλαγές του δεν παραμένουν αναλλοίωτοι ως προς μη γραμμικούς μετασχηματισμούς των δεδομένων οι οποίοι μεταφέρονται και στην δομή των δειγμάτων εκπαίδευσης.

3.6 Επίλογος 3^{ου} Κεφαλαίου

Στο κεφάλαιο αυτό έγινε μία εκτενής ανάλυση στον υπολογισμό παραγόντων Μπεϋες για την σύγκριση δύο ανταγωνιστικών μοντέλων M_i και M_j κάνοντας χρήση μη κατάλληλων εκ των προτέρων κατανομών. Στην πραγματικότητα όταν καλούμαστε να συγκρίνουμε ένα σεβαστό πλήθος μοντέλων, καλούμαστε να υπολογίσουμε όλους τους διαθέσιμους παράγοντες Μπεϋες, καθιστώντας την διαδικασία επιλογής μοντέλων σημαντικά χρονοβόρα. Το πρόβλημα αυτό οδήγησε στην δημιουργία αυτοματοποιημένων διαδικασιών καθορισμού εκ των προτέρων κατανομών και ειδικά σε περιπτώσεις όπου μελετάμε μη πληροφοριακές κατανομές οι οποίες κρίνονται και μη κατάλληλες. Στο επόμενο κεφάλαιο, το οποίο αποτελεί τον βασικό πυρήνα αυτής της εργασίας, θα μελετήσουμε αναλυτικά δύο βασικές μεθοδολογίες παραγωγής εκ των προτέρων κατανομών οι οποίες ανταποκρίνονται σε πραγματικούς παράγοντες Μπεϋες.

Κεφάλαιο 4

Ενδογενείς εκ των προτέρων κατανομές

Στο προηγούμενο κεφάλαιο μελετήσαμε αναλυτικά τις δυσκολίες που προκύπτουν κατά την χρήση μη κατάλληλων εκ των προτέρων κατανομών στην μπεϋζιανή συμπερασματολογία καθώς δεν είναι εφικτός ο καθορισμός παραγόντων Μπεϋες. Μία έξυπνη λύση αποτέλεσε η χρήση των μερικών παραγόντων Μπεϋες όπως είδαμε, όπου ο ενδογενής παράγοντας Μπεϋες αποτέλεσε την σημαντικότερη εκδοχή αυτών. Το βασικό ερώτημα το οποίο ανακύπτει σχετικά με την χρήση του ενδογενούς παράγοντα Μπεϋες είναι η ανταπόκρισή του σε πραγματικούς παράγοντες Μπεϋες.

Στο παρόν κεφάλαιο θα μελετήσουμε αναλυτικά δύο βασικές διαδικασίες παραγωγής εκ των προτέρων κατανομών οι οποίες ανταποκρίνονται σε πραγματικούς παράγοντες Μπεϋες και ταυτόχρονα δεν επηρεάζονται από αυθαίρετες σταθερές κανονικοποίησης μη κατάλληλων κατανομών, όπως και επίσης μία ειδική περίπτωση ταύτισης των δύο μεθοδολογιών.

4.1 Ενδογενείς εκ των προτέρων κατανομές

Οι Berger & Perrichi (1996) αφού προχώρησαν στην δημιουργία των ενδογενών παραγόντων Μπεϋες, θέλησαν να εξετάσουν όπως αναφέραμε, κατά πόσο ο αριθμητικός ενδογενής παράγοντας Μπεϋες ισοδυναμεί ασυμπτωτικά με έναν πραγματικό παράγοντα Μπεϋες για συγκεκριμένα ορισμένες εκ των προτέρων κατανομές. Οι κατανομές αυτές εάν υπάρχουν θα καλούνται *ενδογενείς εκ των προτέρων κατανομές (intrinsic priors)*, έχοντας ως βασικό πλεονέκτημα την καταξίωση των ενδογενών παραγόντων Μπεϋες ως μία αμιγώς μπεϋζιανή μεθοδολογία. Βασικό πλεονέκτημα των Ενδογενών εκ των προτέρων κατανομών είναι η άμεση χρήση τους στην θέση των αρχικών μη κατάλληλων εκ των προτέρων κατανομών, απαλλάσσοντάς μας από τους υπολογισμούς που απαιτούνται για τον καθορισμό ενός δείγματος εκπαίδευσης και από ζητήματα ευστάθειας που προκύπτουν στους ενδογενείς παράγοντες Μπεϋες.

Έστω δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και έστω ένα σύνολο προς εξέταση μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$. Κάτω από τα αντίστοιχα μοντέλα M_i και M_j ορίζονται οι αντίστοιχες συναρτήσεις πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ και $f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)$, με τις αντίστοιχες αρχικές μη-κατάλληλες εκ των προτέρων κατανομές $\pi^N(\boldsymbol{\theta}_i|M_i)$ και $\pi^N(\boldsymbol{\theta}_j|M_j)$ των παραμέτρων $\boldsymbol{\theta}_i$ και $\boldsymbol{\theta}_j$. Έστω επίσης $\pi^I(\boldsymbol{\theta}_i|M_i)$ και $\pi^I(\boldsymbol{\theta}_j|M_j)$ οι ενδογενείς εκ των προτέρων κατανομές των μοντέλων M_i και M_j αντίστοιχα.

Κάνοντας χρήση των παραπάνω ενδογενών εκ των προτέρων κατανομών ορίζουμε τον ενδογενή παράγοντα Μπεϋες υπέρ του μοντέλου M_j ως

$$BF_{ji}^I = \frac{m_j^I(\mathbf{x})}{m_i^I(\mathbf{x})} = \frac{\int_{\theta_j} f(\mathbf{x}|\theta_j, M_j) \pi^I(\theta_j|M_j) d\theta_j}{\int_{\theta_i} f(\mathbf{x}|\theta_i, M_i) \pi^I(\theta_i|M_i) d\theta_i}. \quad (K4.1)$$

Ο αριθμητικός ενδογενής παράγοντας Μπεϋες υπέρ του μοντέλου M_j όπως ορίστηκε και στην σχέση (K3.5) περιγράφεται από την

$$AIBF_{ji} = BF_{ji}^N \frac{1}{L} \sum_{l=1}^L BF_{ij}^N(\mathbf{x}(l)). \quad (K4.2)$$

Οι ενδογενείς εκ των προτέρων κατανομές εξασφαλίζουν την ασυμπτωτική ισοδυναμία των παραγόντων Μπεϋες όπως αυτοί περιγράφονται από τις σχέσεις (K4.1) και (K4.2). Στην συνέχεια θα αποδείξουμε πώς προκύπτει ένα σύστημα εξισώσεων το οποίο καθορίζει τις κατανομές αυτές.

4.1.1 Ενδογενείς εξισώσεις

Ο καθορισμός των ενδογενών εκ των προτέρων κατανομών πραγματοποιείται σε δύο στάδια με την βοήθεια ενός φανταστικού δείγματος εκπαίδευσης $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_{n^*}^*)$ το οποίο θα αποτελεί και τυχαίο δείγμα, όπου το μέγεθός του n^* θα τείνει στο άπειρο. Στο πρώτο στάδιο θα υπολογίσουμε τις ασυμπτωτικές προσεγγίσεις του αριθμητικού ενδογενούς παράγοντα Μπεϋες και του ενδογενούς παράγοντα Μπεϋες που προκύπτει μέσω της χρήσης των ενδογενών εκ των προτέρων κατανομών και θα προχωρήσουμε στην εξίσωση των αποτελεσμάτων τους. Στο δεύτερο στάδιο, αφού θέσουμε ορισμένους περιορισμούς, θα προχωρήσουμε στην εξίσωση των ορίων των αποτελεσμάτων, οδηγώντας στο σύστημα των ενδογενών εξισώσεων (*intrinsic equations*), από την επίλυση του οποίου προκύπτουν οι ενδογενείς εκ των προτέρων κατανομές.

Στάδιο 1

1) Ασυμπτωτική προσέγγιση του BF_{ji}^I

Για την ασυμπτωτική προσέγγιση του BF_{ji}^I θα χρησιμοποιήσουμε την ειδική περίπτωση της ασυμπτωτικής προσέγγισης Laplace του παράγοντα Μπεϋες όπως περιγράφεται στην σχέση (K2.3) στην **Ενότητα 2.6.1**. Επομένως κάτω από το αντίστοιχο μοντέλο M_k , όσο το μέγεθος του φανταστικού δείγματος εκπαίδευσης τείνει στο άπειρο, η αντίστοιχη περιθώρια πιθανοφάνεια του μοντέλου M_k βάσει των δεδομένων \mathbf{x}^* θα προσεγγίζεται από την ποσότητα

$$m_k^I(\mathbf{x}^*) = \hat{m}_k^I(\mathbf{x}^*)(1 + O(n^{-1}))$$

όπου

$$\hat{m}_k^l(\mathbf{x}^*) = (2\pi)^{\frac{d_k}{2}} |\hat{\Sigma}_k|^{-\frac{1}{2}} f(\mathbf{x}^* | \hat{\boldsymbol{\theta}}_k, M_k) \pi^l(\hat{\boldsymbol{\theta}}_k | M_k).$$

Παρατηρούμε ότι η ασυμπτωτική προσέγγιση της πιθανοφάνειας $m_k^l(\mathbf{x}^*)$ εμπεριέχει μία **σχετική απόκλιση (relative error)** της τάξης $1 + O(n^{-1})$. Αφού προχωρήσουμε στην εφαρμογή αυτής της ασυμπτωτικής προσέγγισης στις περιθώριες πιθανοφάνειες που αντιστοιχούν στον αριθμητή και αντίστοιχα στον παρονομαστή του ενδογενούς Παράγοντα Μπεϋες BF_{ji}^l , παρατηρούμε ότι η κάθε μία θα εμπεριέχει μία σχετική απόκλιση της τάξης $(1 + O(n^{-1}))$, οδηγώντας την ασυμπτωτική προσέγγιση του BF_{ji}^l να εμπεριέχει σχετική απόκλιση της τάξης $O(n^{-1})$. Επομένως καθώς το μέγεθος n^* του φανταστικού δείγματος εκπαίδευσης \mathbf{x}^* θα τείνει στο άπειρο, έχουμε ότι

$$\begin{aligned} BF_{ji}^l &= \frac{m_j^l(\mathbf{x}^*)}{m_i^l(\mathbf{x}^*)} \\ &= \frac{\hat{m}_j^l(\mathbf{x}^*)(1 + O(n^{-1}))}{\hat{m}_i^l(\mathbf{x}^*)(1 + O(n^{-1}))} \\ &= \frac{(2\pi)^{\frac{d_j}{2}} |\hat{\Sigma}_j|^{-\frac{1}{2}} f(\mathbf{x}^* | \hat{\boldsymbol{\theta}}_j, M_j) \pi^l(\hat{\boldsymbol{\theta}}_j | M_j) (1 + O(n^{-1}))}{(2\pi)^{\frac{d_i}{2}} |\hat{\Sigma}_i|^{-\frac{1}{2}} f(\mathbf{x}^* | \hat{\boldsymbol{\theta}}_i, M_i) \pi^l(\hat{\boldsymbol{\theta}}_i | M_i) (1 + O(n^{-1}))}. \end{aligned} \quad (K4.3)$$

2) Ασυμπτωτική προσέγγιση του AIBF_{ji}

Όπως αναφέρθηκε στην **Ενότητα 3.5.2** όταν το μέγεθος του δείγματος το οποίο εξετάζεται είναι πάρα πολύ μεγάλο, ο αριθμητικός ενδογενής παράγοντας Μπεϋες προσεγγίζεται από τον αναμενόμενο αριθμητικό ενδογενή παράγοντα Μπεϋες, ο οποίος περιγράφεται από την σχέση

$$EAIBF_{ji} = BF_{ji}^N \frac{1}{L} \sum_{l=1}^L E_{\theta_j^{M_j}} [BF_{ij}^N(\mathbf{x}(l))].$$

Επομένως για την ασυμπτωτική προσέγγιση του αναμενόμενου ενδογενή παράγοντα Μπεϋες θα προχωρήσουμε στην προσέγγιση των δύο στοιχείων από τα οποία αποτελείται.

➤ Ασυμπτωτική προσέγγιση του όρου BF_{ji}^N

Για την ασυμπτωτική προσέγγιση του όρου BF_{ji}^N θα χρησιμοποιήσουμε αντίστοιχα την ειδική περίπτωση της ασυμπτωτικής προσέγγισης του Laplace. Συνεπώς προκύπτει ότι

$$BF_{ji}^N = \frac{m_j^N(\mathbf{x}^*)}{m_i^N(\mathbf{x}^*)}$$

$$\begin{aligned}
&= \frac{\widehat{m}^N_j(\mathbf{x}^*)(1 + O(n^{-1}))}{\widehat{m}^N_i(\mathbf{x}^*)(1 + O(n^{-1}))} \\
&= \frac{(2\pi)^{\frac{d_j}{2}} |\widehat{\Sigma}_j|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_j, M_j) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j) (1 + O(n^{-1}))}{(2\pi)^{\frac{d_i}{2}} |\widehat{\Sigma}_i|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_i, M_i) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i) (1 + O(n^{-1}))}. \quad (K4.4)
\end{aligned}$$

Ο δεύτερος όρος του αθροίσματος των αναμενομένων τιμών των παραγόντων Μπεϋες οι οποίοι θα υπολογίζονται από τα δείγματα εκπαίδευσης θα χρησιμοποιηθούν αυτούσια στην συνέχεια, συνεπώς θα παραμείνουν στην μορφή αυτή.

Συνεπώς με την βοήθεια των σχέσεων (K4.3) και (K4.4), εξισώνοντας τις ασυμπτωτικές προσεγγίσεις των εμπλεκόμενων Παραγόντων Μπεϋες προκύπτει ότι

$$\begin{aligned}
\frac{(2\pi)^{\frac{d_j}{2}} |\widehat{\Sigma}_j|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_j, M_j) \pi^l(\widehat{\boldsymbol{\theta}}_j | M_j) (1 + O(n^{-1}))}{(2\pi)^{\frac{d_i}{2}} |\widehat{\Sigma}_i|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_i, M_i) \pi^l(\widehat{\boldsymbol{\theta}}_i | M_i) (1 + O(n^{-1}))} &= \frac{(2\pi)^{\frac{d_j}{2}} |\widehat{\Sigma}_j|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_j, M_j) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j) (1 + O(n^{-1}))}{(2\pi)^{\frac{d_i}{2}} |\widehat{\Sigma}_i|^{-\frac{1}{2}} f(\mathbf{x}^* | \widehat{\boldsymbol{\theta}}_i, M_i) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i) (1 + O(n^{-1}))} \frac{1}{L} \sum_{l=1}^L E_{\widehat{\boldsymbol{\theta}}_j}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \\
&\Leftrightarrow \frac{\pi^l(\widehat{\boldsymbol{\theta}}_j | M_j) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i)}{\pi^l(\widehat{\boldsymbol{\theta}}_i | M_i) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j)} (1 + o(1)) = \widetilde{BF}_{ij}^N \quad (K4.5)
\end{aligned}$$

όπου \widetilde{BF}_{ij}^N είναι είτε ο αριθμητικός είτε ο γεωμετρικός μέσος των όρων $BF_{ij}^N(\mathbf{x}(l))$.

Επομένως ολοκληρώνεται το πρώτο στάδιο καθορισμού των ενδογενών εκ των προτέρων κατανομών, δίνοντας μας την σχέση (K4.5) μέσω της οποίας θα προχωρήσουμε στην δημιουργία των ενδογενών εξισώσεων.

Στάδιο 2

Πριν προχωρήσουμε στην εξίσωση των ορίων των παραγόντων Μπεϋες θα εισαγάγουμε ορισμένους περιορισμούς σχετικά με την ασυμπτωτική συμπεριφορά των εκτιμητών μέγιστης πιθανοφάνειας των παραμέτρων του εκάστοτε μοντέλου, αλλά και του παράγοντα Μπεϋες που καθορίζεται από τα δείγματα εκπαίδευσης στο δεύτερο σκέλος της εξίσωσης (K4.5). Το 1994 ο Dmochowski σε μία ενδιαφέρουσα εργασία του, χρησιμοποιεί μία παρατήρηση του Huber (1967) σχετικά με την ασυμπτωτική συμπεριφορά των εκτιμητών μέγιστης πιθανοφάνειας, που μελέτησε ο Wald (1949), οδηγώντας στις εξής υποθέσεις:

- i) Κάτω από το μοντέλο $M_j: \widehat{\boldsymbol{\theta}}_j \rightarrow \boldsymbol{\theta}_j, \widehat{\boldsymbol{\theta}}_i \rightarrow \boldsymbol{\psi}_i(\boldsymbol{\theta}_j)$, και $\widetilde{BF}_{ij}^N \rightarrow B_j^*(\boldsymbol{\theta}_j)$
- ii) Κάτω από το μοντέλο $M_i: \widehat{\boldsymbol{\theta}}_i \rightarrow \boldsymbol{\theta}_i, \widehat{\boldsymbol{\theta}}_j \rightarrow \boldsymbol{\psi}_j(\boldsymbol{\theta}_i)$, και $\widetilde{BF}_{ij}^N \rightarrow B_i^*(\boldsymbol{\theta}_i)$

Ο όρος $\boldsymbol{\psi}_j(\boldsymbol{\theta}_i)$, ονομάζεται και *ψ-εικόνα (ψ-image)* του $\widehat{\boldsymbol{\theta}}_j$ στο $\boldsymbol{\theta}_i$, αντιπροσωπεύει το όριο του εκτιμητή μέγιστης πιθανοφάνειας $\widehat{\boldsymbol{\theta}}_j(\mathbf{x}^*)$ κάτω από το μοντέλο M_i στο σημείο $\boldsymbol{\theta}_i$ και αντίστοιχα ο όρος $\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)$ αντιπροσωπεύει το όριο του εκτιμητή μέγιστης

πιθανοφάνειας $\widehat{\boldsymbol{\theta}}_i(\mathbf{x}^*)$ κάτω από το μοντέλο M_j στο σημείο $\boldsymbol{\theta}_j$. Ο όρος $B_k^*(\boldsymbol{\theta}_k)$ για $k = i, j$ αντιπροσωπεύει

$$B_k^*(\boldsymbol{\theta}_k) = \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_k}^{M_k} \left[\frac{1}{L} \sum_{l=1}^L B_{F_{ij}^N}(\mathbf{x}(l)) \right]$$

όταν χρησιμοποιούμε τον αναμενόμενο αριθμητικό ενδογενή παράγοντα Μπεύες, και

$$B_k^*(\boldsymbol{\theta}_k) = \lim_{L \rightarrow \infty} \exp \left\{ E_{\boldsymbol{\theta}_k}^{M_k} \left[\frac{1}{L} \sum_{l=1}^L \log B_{F_{ij}^N}(\mathbf{x}(l)) \right] \right\}$$

όταν χρησιμοποιούμε τον αναμενόμενο γεωμετρικό ενδογενή παράγοντα Μπεύες. Όπως αναφέρουν και οι Berger & Perrichi (1996), στην περίπτωση όπου ισχύει η ανταλλαξιμότητα στο δείγμα \mathbf{x}^* , τότε οι όροι του ορίου και της αριθμητικής στάθμησης απαλείφονται.

Επομένως εξισώνοντας τα όρια (όταν το μέγεθος n^* του δείγματος \mathbf{x}^* τείνει στο άπειρο) των μελών της σχέσης (K4.5) προκύπτει ότι

$$\lim_{n \rightarrow \infty} \frac{\pi^I(\widehat{\boldsymbol{\theta}}_j | M_j) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i)}{\pi^I(\widehat{\boldsymbol{\theta}}_i | M_i) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j)} (1 + o(1)) = \lim_{n \rightarrow \infty} \widetilde{B}_{ij}^N. \quad (K4.6)$$

Χρησιμοποιώντας τις παραπάνω υποθέσεις προχωράμε στην μελέτη των παρακάτω περιπτώσεων:

i) Κάτω από το μοντέλο M_j η σχέση (K4.6) διαμορφώνεται,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\pi^I(\widehat{\boldsymbol{\theta}}_j | M_j) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i)}{\pi^I(\widehat{\boldsymbol{\theta}}_i | M_i) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j)} (1 + o(1)) = \lim_{n \rightarrow \infty} \widetilde{B}_{ij}^N \\ \Leftrightarrow & \frac{\pi^I(\boldsymbol{\theta}_j | M_j) \pi^N(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j) | M_i)}{\pi^I(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j) | M_i) \pi^N(\boldsymbol{\theta}_j | M_j)} = B_j^*(\boldsymbol{\theta}_j) \quad (K4.7) \end{aligned}$$

ii) Κάτω από το μοντέλο M_i η σχέση (K4.6) διαμορφώνεται,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\pi^I(\widehat{\boldsymbol{\theta}}_j | M_j) \pi^N(\widehat{\boldsymbol{\theta}}_i | M_i)}{\pi^I(\widehat{\boldsymbol{\theta}}_i | M_i) \pi^N(\widehat{\boldsymbol{\theta}}_j | M_j)} (1 + o(1)) = \lim_{n \rightarrow \infty} \widetilde{B}_{ij}^N \\ \Leftrightarrow & \frac{\pi^I(\boldsymbol{\psi}_j(\boldsymbol{\theta}_i) | M_j) \pi^N(\boldsymbol{\theta}_i | M_i)}{\pi^I(\boldsymbol{\theta}_i | M_i) \pi^N(\boldsymbol{\psi}_j(\boldsymbol{\theta}_i) | M_j)} = B_i^*(\boldsymbol{\theta}_i) \quad (K4.8) \end{aligned}$$

Το όριο $\lim_{n \rightarrow \infty} \widetilde{B}_{ij}^N$ ισοδυναμεί με το όριο $\lim_{L \rightarrow \infty} \widetilde{B}_{ij}^N$ καθώς όσο το μέγεθος n^* του δείγματος \mathbf{x}^* τείνει στο άπειρο, το μέγεθος L του συνόλου των δειγμάτων εκπαίδευσης

\mathcal{X}_L^* θα τείνει και αυτό στο άπειρο. Επίσης όσο το μέγεθος n^* τείνει στο άπειρο, ο όρος της σχετικής απόκλισης $1 + o(1)$ θα τείνει στην μονάδα.

Επομένως η επίλυση του συστήματος των ενδογενών εξισώσεων που περιγράφονται από τις σχέσεις (K4.7) και (K4.8) οδηγεί στον προσδιορισμό των ενδογενών εκ των προτέρων κατανομών. Όπως αναφέρουν χαρακτηριστικά οι Berger & Perrichi (1996), οι κατανομές οι οποίες προκύπτουν δεν είναι απαραίτητο να είναι μοναδικές, όπως επίσης δεν είναι απαραίτητο να είναι και κατάλληλες εκ των προτέρων κατανομές. Στην περίπτωση όπου προκύπτουν μη κατάλληλες εκ των προτέρων κατανομές, υπάρχει η δυνατότητα να προχωρήσουμε στον υπολογισμό παραγόντων Μπεϋες καθώς οι ενδογενείς εκ των προτέρων κατανομές δεν εξαρτιούνται από αυθαίρετες σταθερές κανονικοποίησης.

Στην συνέχεια θα μελετήσουμε την μορφή που λαμβάνουν οι ενδογενείς εκ των προτέρων κατανομές όταν μελετάμε περιπτώσεις εμφωλευμένων μοντέλων και μη.

4.2 Σύγκριση εμφωλευμένων μοντέλων

Όταν καλούμαστε να βρούμε τις ενδογενείς εκ των προτέρων κατανομές κατά την σύγκριση εμφωλευμένων μοντέλων, μπορούμε να εκμεταλλευτούμε αυτήν την ιδιότητα προκειμένου να απλοποιηθούν οι υπολογισμοί εύρεσης των κατανομών. Οι Moreno & Bertolino & Racugno (2014) ορίσανε κλάσεις για τις ενδογενείς εκ των προτέρων κατανομές προκειμένου να καταστεί ευκολότερος ο προσδιορισμός τους όταν συγκρίνουμε εμφωλευμένα μοντέλα.

Έστω δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και έστω ένα σύνολο προς εξέταση μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$. Κάτω από τα αντίστοιχα μοντέλα M_i και M_j ορίζονται οι αντίστοιχες συναρτήσεις πιθανοφάνειας $f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)$ και $f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)$, με τις αντίστοιχες αρχικές μη κατάλληλες εκ των προτέρων κατανομές $\pi^N(\boldsymbol{\theta}_i|M_i)$ και $\pi^N(\boldsymbol{\theta}_j|M_j)$ των παραμέτρων $\boldsymbol{\theta}_i$ και $\boldsymbol{\theta}_j$.

Θεωρώντας το μοντέλο M_i εμφωλευμένο του μοντέλου M_j , θεωρούμε στην ουσία ότι

$$\boldsymbol{\theta}_i \subset \boldsymbol{\theta}_j$$

και

$$f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \equiv f(\mathbf{x}|\boldsymbol{\theta}_i, M_i), \text{ αν } \boldsymbol{\theta}_i = \boldsymbol{\theta}_j.$$

Επομένως λαμβάνοντας υπόψιν αυτήν την χαρακτηριστική ιδιότητα λύνοντας την ενδογενή εξίσωση (K4.7) ως προς την ενδογενή εκ των προτέρων κατανομή $\pi^I(\boldsymbol{\theta}_j|M_j)$ προκύπτει ότι

$$\pi^I(\boldsymbol{\theta}_j|M_j) = B_j^*(\boldsymbol{\theta}_j) \frac{\pi^N(\boldsymbol{\theta}_j|M_j)}{\pi^N(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i)} \pi^I(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i).$$

Είναι εμφανές ότι η ενδογενής εκ των προτέρων κατανομή του μοντέλου M_j καθορίζεται εξ ολοκλήρου από την ενδογενή εκ των προτέρων κατανομή του μοντέλου M_i χωρίς να παρουσιάζονται περιορισμοί ως προς την επιλογή της. Επειδή οι Ενδογενείς εκ των προτέρων κατανομές προορίζονται για τον υπολογισμό παραγόντων Μπεϋές είναι σημαντικό να ολοκληρώνουν στην μονάδα. Συνεπώς είναι αναγκαίο να θέσουμε τους περιορισμούς

$$\int_{\boldsymbol{\theta}_i} \pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = 1$$

και

$$\int_{\boldsymbol{\theta}_j} T(\boldsymbol{\theta}_j)\pi^I(\psi_i(\boldsymbol{\theta}_j)|M_i) d\boldsymbol{\theta}_j = 1 \quad (K4.9)$$

με

$$T(\boldsymbol{\theta}_j) = B_j^*(\boldsymbol{\theta}_j) \frac{\pi^N(\boldsymbol{\theta}_j|M_j)}{\pi^N(\psi_i(\boldsymbol{\theta}_j)|M_i)}.$$

Επειδή το μοντέλο M_i είναι εμφωλευμένο του M_j ο παραμετρικός χώρος για το M_j μοντέλο μπορεί να περιγραφεί ως

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_i \cup (\boldsymbol{\theta}_j \setminus \boldsymbol{\theta}_i).$$

Συνεπώς η σχέση (K4.9) αναπροσαρμόζεται ως

$$\int_{\boldsymbol{\theta}_i} V(\boldsymbol{\theta}_i)\pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = 1$$

όπου

$$V(\boldsymbol{\theta}_i) = T(\boldsymbol{\theta}_i) + \left\{ \int_{\boldsymbol{\theta}_j(\boldsymbol{\theta}_i)} T(\boldsymbol{\theta}_j) d\boldsymbol{\theta}_j \right\} I_{\boldsymbol{\theta}_i^*}(\boldsymbol{\theta}_i)$$

με $\boldsymbol{\theta}_j(\boldsymbol{\theta}_i) = \{\boldsymbol{\theta}_j \in (\boldsymbol{\theta}_j \setminus \boldsymbol{\theta}_i) : \psi_i(\boldsymbol{\theta}_j) = \boldsymbol{\theta}_i\}$ και το σύνολο $\boldsymbol{\theta}_i^*$ εκπροσωπεί την ψ_i εικόνα του συνόλου $\boldsymbol{\theta}_j \setminus \boldsymbol{\theta}_i$. Επομένως ο καθορισμός των ενδογενών εκ των προτέρων κατανομών και για τα δύο μοντέλα, καθορίζεται από την επιλογή της ενδογενούς εκ των προτέρων κατανομής $\pi^I(\boldsymbol{\theta}_i|M_i)$ για το εμφωλευμένο μοντέλο M_i , η οποία θα ανήκει στην κλάση

$$\Gamma = \left\{ \pi^I(\boldsymbol{\theta}_i|M_i) : \int_{\boldsymbol{\theta}_i} \pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = 1, \int_{\boldsymbol{\theta}_i} V(\boldsymbol{\theta}_i) \pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = 1 \right\}. \quad (K4.10)$$

Συνεπώς για οποιαδήποτε ενδογενή εκ των προτέρων κατανομή $\pi^I(\boldsymbol{\theta}_i|M_i)$ η οποία ανήκει στην κλάση Γ , είναι εφικτό να προσδιορίσουμε την ενδογενή εκ των προτέρων κατανομή $\pi^I(\boldsymbol{\theta}_j|M_j)$.

Οι Berger & Perrichi (1996) πρότειναν έναν πιο άμεσο τρόπο για την εξεύρεση των ενδογενών εκ των προτέρων κατανομών στην περίπτωση σύγκρισης εμφωλευμένων μοντέλων μέσω της απλούστευσης του συστήματος των ενδογενών εξισώσεων. Κατέληξαν ότι η απλούστερη επιλογή ενδογενών εκ των προτέρων κατανομών προκύπτει επιλέγοντας ως ενδογενή εκ των προτέρων κατανομή του εμφωλευμένου μοντέλου, την αρχική μη κατάλληλη εκ των προτέρων κατανομή, δηλαδή

$$\pi^I(\boldsymbol{\theta}_i|M_i) = \pi^N(\boldsymbol{\theta}_i|M_i).$$

Για αυτήν την επιλογή η ενδογενής εξίσωση η οποία περιγράφεται από την σχέση (K4.7) διαμορφώνεται ως

$$\begin{aligned} \frac{\pi^I(\boldsymbol{\theta}_j|M_j)\pi^N(\psi_i(\boldsymbol{\theta}_j)|M_i)}{\pi^N(\psi_i(\boldsymbol{\theta}_j)|M_i)\pi^N(\boldsymbol{\theta}_j|M_j)} &= B_j^*(\boldsymbol{\theta}_j) \\ \Leftrightarrow \pi^I(\boldsymbol{\theta}_j|M_j) &= \pi^N(\boldsymbol{\theta}_j|M_j)B_j^*(\boldsymbol{\theta}_j). \end{aligned}$$

Συνεπώς οι ενδογενείς εξισώσεις λαμβάνουν την μορφή

$$\pi^I(\boldsymbol{\theta}_i|M_i) = \pi^N(\boldsymbol{\theta}_i|M_i), \quad \pi^I(\boldsymbol{\theta}_j|M_j) = \pi^N(\boldsymbol{\theta}_j|M_j)B_j^*(\boldsymbol{\theta}_j). \quad (K4.11)$$

Με αυτήν την επιλογή παρατηρούμε ότι η ενδογενής εκ των προτέρων κατανομή του μοντέλου M_i θα εξαρτάται από την αντίστοιχη αυθαίρετη σταθερά κανονικοποίησης c_i , όμως δεν θα αποτελέσει πρόβλημα κατά τον υπολογισμό του ενδογενή παράγοντα Μπεϋες. Η ενδογενής εκ των προτέρων κατανομή για το μοντέλο M_j διαμορφώνεται ως

$$\begin{aligned} \pi^I(\boldsymbol{\theta}_j|M_j) &= \pi^N(\boldsymbol{\theta}_j|M_j)B_j^*(\boldsymbol{\theta}_j) \\ &= \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L BF_{ij}^N(\mathbf{x}(l)) \right] \\ &= \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L \frac{m_i^N(\mathbf{x}(l))}{m_j^N(\mathbf{x}(l))} \right] \\ &= \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i)\pi^N(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(l)|\boldsymbol{\theta}_j, M_j)\pi^N(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i) \frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i} d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(l)|\boldsymbol{\theta}_j, M_j) \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_j} d\boldsymbol{\theta}_j} \right] \\
&= \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_i} \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i) \pi^{N^*}(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}(l)|\boldsymbol{\theta}_j, M_j) \pi^{N^*}(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j} \right].
\end{aligned}$$

Άρα ο ενδογενής παράγοντας Μπεϋές μέσω της χρήσης των ενδογενών εκ των προτέρων κατανομών θα υπολογίζεται

$$\begin{aligned}
BF_{ji}^I &= \frac{m_j^I(\mathbf{x})}{m_i^I(\mathbf{x})} \\
&= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \frac{\pi^{N^*}(\boldsymbol{\theta}_j|M_j)}{c_i} CF_{ij} d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \frac{\pi^{N^*}(\boldsymbol{\theta}_i|M_i)}{c_i} d\boldsymbol{\theta}_i} \\
&= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \pi^{N^*}(\boldsymbol{\theta}_j|M_j) CF_{ij} d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \pi^{N^*}(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}
\end{aligned}$$

όπου

$$CF_{ij} = \lim_{L \rightarrow \infty} E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{1}{L} \sum_{l=1}^L \frac{\int_{\boldsymbol{\theta}_i} f_i(\mathbf{x}(l)|\boldsymbol{\theta}_i, M_i) \pi^{N^*}(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f_j(\mathbf{x}(l)|\boldsymbol{\theta}_j, M_j) \pi^{N^*}(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j} \right].$$

Άρα η επιλογή μίας μη κατάλληλης Ενδογενούς εκ των προτέρων κατανομής δεν περιορίζει την χρήση της για τον καθορισμό ενός ενδογενή παράγοντα Μπεϋές. Γενικά μπορούν να υπάρξουν αρκετές εναλλακτικές επιλογές για τον καθορισμό των ενδογενών εκ των προτέρων κατανομών, καθώς αρκεί να προέρχονται από την κλάση Γ όπως αυτή καθορίζεται στην σχέση (K4.10), όμως συνήθως αυτή η επιλογή που αναλύσαμε σε αυτήν την ενότητα είναι η απλούστερη επιλογή.

4.3 Σύγκριση μη εμφωλευμένων μοντέλων

Όταν προχωράμε σε συγκρίσεις μη εμφωλευμένων μοντέλων ο καθορισμός των ενδογενών εκ των προτέρων κατανομών πραγματοποιείται πλήρως από το αρχικό σύστημα των ενδογενών εξισώσεων (K4.7) και (K4.8), χωρίς να καταστεί εφικτή η μετατροπή του συστήματος εξισώσεων στην επίλυση μίας απλής εξίσωσης όπως παρουσιάστηκε στην σύγκριση εμφωλευμένων μοντέλων. Μέχρι στιγμής δεν είναι γνωστό ποια είναι η γενική λύση του συστήματος των ενδογενών εξισώσεων στην περίπτωση μη εμφωλευμένων μοντέλων αλλά και ούτε είναι γνωστό εάν υπάρχει κάποια γενική λύση.

Οι Moreno & Cano & Kessler (2004) παρουσίασαν έναν γενικό τρόπο επίλυσης των ενδογενών εξισώσεων στην περίπτωση όπου εξετάζουμε μη εμφωλευμένα μοντέλα, όπως επίσης εξέτασαν πώς αυτές διαμορφώνονται σε διάφορες περιπτώσεις. Επίσης κατά την πορεία της περιγραφής της μεθοδολογίας θα θεωρήσουμε ότι για το δείγμα το οποίο εξετάζουμε ισχύει η ανταλλαξιμότητα προκειμένου να απλοποιηθούν οι συμβολισμοί.

Η ενδογενής εξίσωση (K4.7) μπορεί να γραφεί στην μορφή

$$\frac{\pi^I(\boldsymbol{\theta}_j|M_j)}{\pi^N(\boldsymbol{\theta}_j|M_j)} = B_j^*(\boldsymbol{\theta}_j) \frac{\pi^I(\psi_i(\boldsymbol{\theta}_j)|M_i)}{\pi^N(\psi_i(\boldsymbol{\theta}_j)|M_i)}$$

όπου για τιμές $\boldsymbol{\theta}_j = \psi_j(\boldsymbol{\theta}_i)$ παίρνουμε ότι

$$\frac{\pi^I(\psi_j(\boldsymbol{\theta}_i)|M_j)}{\pi^N(\psi_j(\boldsymbol{\theta}_i)|M_j)} = E_{\psi_j(\boldsymbol{\theta}_i)}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \frac{\pi^I(h_i(\boldsymbol{\theta}_i)|M_i)}{\pi^N(h_i(\boldsymbol{\theta}_i)|M_i)} \quad (K4.12)$$

όπου $h_i(\boldsymbol{\theta}_i) = \psi_i(\psi_j(\boldsymbol{\theta}_i))$. Στην συνέχεια γράφουμε την ενδογενή εξίσωση η οποία περιγράφεται από την (K4.8) ως

$$\frac{\pi^I(\psi_j(\boldsymbol{\theta}_i)|M_j)}{\pi^N(\psi_j(\boldsymbol{\theta}_i)|M_j)} = E_{\boldsymbol{\theta}_i}^{M_i} [BF_{ij}^N(\mathbf{x}(l))] \frac{\pi^I(\boldsymbol{\theta}_i|M_i)}{\pi^N(\boldsymbol{\theta}_i|M_i)} \quad (K4.13).$$

Αν παρατηρήσουμε τις σχέσεις (K4.12) και (K4.13) θα δούμε ότι τα πρώτα μέλη τους είναι ίσα, επομένως εξισώνοντας τα δεύτερα μέλη παίρνουμε ότι

$$\begin{aligned} E_{\psi_j(\boldsymbol{\theta}_i)}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \frac{\pi^I(h_i(\boldsymbol{\theta}_i)|M_i)}{\pi^N(h_i(\boldsymbol{\theta}_i)|M_i)} &= E_{\boldsymbol{\theta}_i}^{M_i} [BF_{ij}^N(\mathbf{x}(l))] \frac{\pi^I(\boldsymbol{\theta}_i|M_i)}{\pi^N(\boldsymbol{\theta}_i|M_i)} \\ \Leftrightarrow \pi_i^I(h_i(\boldsymbol{\theta}_i)) &= \frac{E_{\boldsymbol{\theta}_i}^{M_i} [BF_{ij}^N(\mathbf{x}(l))] \pi^N(h_i(\boldsymbol{\theta}_i)|M_i)}{E_{\psi_j(\boldsymbol{\theta}_i)}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \pi^N(\boldsymbol{\theta}_i|M_i)} \pi^I(\boldsymbol{\theta}_i|M_i) \\ &\Leftrightarrow \pi^I(h_i(\boldsymbol{\theta}_i)|M_i) = g_i(\boldsymbol{\theta}_i) \pi^I(\boldsymbol{\theta}_i|M_i) \end{aligned}$$

με

$$g_i(\boldsymbol{\theta}_i) = \frac{E_{\boldsymbol{\theta}_i}^{M_i} [BF_{ij}^N(\mathbf{x}(l))] \pi^N(h_i(\boldsymbol{\theta}_i)|M_i)}{E_{\psi_j(\boldsymbol{\theta}_i)}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \pi^N(\boldsymbol{\theta}_i|M_i)}.$$

Στην συνέχεια λύνουμε την σχέση (K4.7) ως προς $\pi^I(\boldsymbol{\theta}_j|M_j)$ και παίρνουμε ότι

$$\begin{aligned} \pi^I(\boldsymbol{\theta}_j|M_j) &= E_{\boldsymbol{\theta}_j}^{M_j} [BF_{ij}^N(\mathbf{x}(l))] \frac{\pi^N(\boldsymbol{\theta}_j|M_j)}{\pi^N(\psi_i(\boldsymbol{\theta}_j)|M_i)} \pi^I(\psi_i(\boldsymbol{\theta}_j)|M_i) \\ &\Leftrightarrow \pi^I(\boldsymbol{\theta}_j|M_j) = g_j(\boldsymbol{\theta}_j) \pi^I(\psi_i(\boldsymbol{\theta}_j)|M_i) \end{aligned}$$

με

$$g_j(\boldsymbol{\theta}_j) = E_{\boldsymbol{\theta}_j}^{M_j} [BF_{ij}^N(\mathbf{X}(l))] \frac{\pi^N(\boldsymbol{\theta}_j|M_j)}{\pi^N(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i)}.$$

Συνεπώς μετά από αυτές τις μετατροπές δημιουργούμε ένα ισοδύναμο σύστημα ενδογενών εξισώσεων το οποίο περιγράφεται από

$$\pi^l(h_i(\boldsymbol{\theta}_i)|M_i) = g_i(\boldsymbol{\theta}_i)\pi^l(\boldsymbol{\theta}_i|M_i)$$

$$\pi^l(\boldsymbol{\theta}_j|M_j) = g_j(\boldsymbol{\theta}_j)\pi^l(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i) .$$

Επομένως όταν προχωράμε στην σύγκριση μη εμφωλευμένων μοντέλων οι ενδογενείς εκ των προτέρων κατανομές που αντιστοιχούν στα μοντέλα M_i και M_j καθορίζονται επιλέγοντας την ενδογενή εκ των προτέρων κατανομή του μοντέλου M_i μέσα από την κλάση

$$\Gamma^l = \{\pi_i^l: \pi^l(h_i(\boldsymbol{\theta}_i)|M_i) = g_i(\boldsymbol{\theta}_i)\pi^l(\boldsymbol{\theta}_i|M_i)\} \quad (K4.14)$$

και στην συνέχεια προχωράμε στον καθορισμό της Ενδογενούς εκ των προτέρων κατανομής του μοντέλου M_j μέσω της σχέσης

$$\pi^l(\boldsymbol{\theta}_j|M_j) = g_j(\boldsymbol{\theta}_j)\pi^l(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i) . \quad (K4.15)$$

Στην προκειμένη περίπτωση επιλέξαμε να καθορίσουμε αρχικά την κλάση των κατανομών για την ενδογενή εκ των προτέρων κατανομή $\pi_i^l(\boldsymbol{\theta}_i|M_i)$. Αντίστοιχα μπορούμε να καθορίσουμε την κλάση των κατανομών της ενδογενούς εκ των προτέρων κατανομής $\pi^l(\boldsymbol{\theta}_j|M_j)$ και στην συνέχεια να υπολογίσουμε την αντίστοιχη του μοντέλου M_i , καταλήγοντας σε αντίστοιχα όμοια αποτελέσματα.

Για ένα σετ δεδομένων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και για ένα ζεύγος ενδογενών εκ των προτέρων κατανομών $\pi^l(\boldsymbol{\theta}_i|M_i), \pi^l(\boldsymbol{\theta}_j|M_j)$, όπως αυτές προκύπτουν από το επαναπροσδιορισμένο σύστημα των ενδογενών εξισώσεων όπως αυτό περιγράφεται από τις σχέσεις (K4.14) και (K4.15), ο ενδογενής παράγοντας Μπεϋες υπέρ του μοντέλου M_j υπολογίζεται ως

$$\begin{aligned} BF_{ji}^l &= \frac{m_j^l(\mathbf{x})}{m_i^l(\mathbf{x})} = \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^l(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)\pi^l(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j} \\ &= \frac{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^l(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i}{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)g_j(\boldsymbol{\theta}_j)\pi^l(\boldsymbol{\psi}_i(\boldsymbol{\theta}_j)|M_i)d\boldsymbol{\theta}_j}. \end{aligned}$$

Όπως αναφέρουν και οι Moreno & Cano & Kessler (2004) ένας τέτοιος παράγοντας Μπεϋες δεν θα λειτουργεί φυσιολογικά ακόμα και όταν η $\pi_i^l(\boldsymbol{\theta}_i)$ είναι κατάλληλη.

Χαρακτηριστικά επισημαίνουν ότι αν για μία ενδογενή εκ των προτέρων κατανομή $\pi^I(\boldsymbol{\theta}_i|M_i)$ από την κλάση Γ^I έχουμε ότι

$$\int_{\boldsymbol{\theta}_i} \pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = k < \infty$$

τότε προκειμένου να μην ευνοηθεί εκ των προτέρων κάποιο μοντέλο, πρέπει να ισχύει η συνθήκη

$$\int_{\boldsymbol{\theta}_i} \pi^I(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i = \int_{\boldsymbol{\theta}_j} g_j(\boldsymbol{\theta}_j)\pi^I(\psi_i(\boldsymbol{\theta}_j)|M_i)d\boldsymbol{\theta}_j.$$

4.4 Αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές

Οι Perez & Berger (2002) ανέπτυξαν μία αυτοματοποιημένη μεθοδολογία καθορισμού εκ των προτέρων κατανομών η οποία βασίζεται σε φανταστικά δείγματα εκπαίδευσης. Αυτές οι εκ των προτέρων κατανομές θα καλούνται *αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές (expected-posterior prior distribution)* και για συντομία θα τις αναφέρουμε EPPs. Η βασική ιδέα στην οποία στηρίζεται η μεθοδολογία των EPPs είναι ο καθορισμός εκ των προτέρων κατανομών με την βοήθεια αλγορίθμων MCMC μέσω μίας βασικής προβλεπτικής κατανομής.

Έστω δείγμα παρατηρήσεων $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ και έστω ένα σύνολο προς εξέταση μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$. Κάτω από τα αντίστοιχα μοντέλα M_i και M_j ορίζονται οι αντίστοιχες συναρτήσεις πιθανοφάνειας $f(\boldsymbol{x}|\boldsymbol{\theta}_i, M_i)$ και $f(\boldsymbol{x}|\boldsymbol{\theta}_j, M_j)$, με τις αντίστοιχες αρχικές μη κατάλληλες εκ των προτέρων κατανομές $\pi^N(\boldsymbol{\theta}_i|M_i)$ και $\pi^N(\boldsymbol{\theta}_j|M_j)$ των παραμέτρων $\boldsymbol{\theta}_i$ και $\boldsymbol{\theta}_j$.

Έστω επίσης ένα φανταστικό δείγμα εκπαίδευσης $\boldsymbol{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ το οποίο θα χρησιμοποιηθεί αποκλειστικά για τον καθορισμό των EPPs προκειμένου να προχωρήσουμε σε σύγκριση μοντέλων. Για κάθε μοντέλο M_k με την αντίστοιχη μη κατάλληλη εκ των προτέρων κατανομή $\pi^N(\boldsymbol{\theta}_k|M_k)$, δοθέντος του φανταστικού δείγματος εκπαίδευσης \boldsymbol{x}^* η εκ των υστέρων κατανομή θα υπολογίζεται ως

$$\pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{x}^*) = \frac{f(\boldsymbol{x}^*|\boldsymbol{\theta}_k, M_k)\pi^N(\boldsymbol{\theta}_k|M_k)}{\int_{\boldsymbol{\theta}_k} f(\boldsymbol{x}^*|\boldsymbol{\theta}_k, M_k)\pi^N(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}.$$

Για μία δεδομένη κατανομή πρόβλεψης $m^*(\boldsymbol{x}^*)$, η οποία θα εξεταστεί στην συνέχεια, ορίζουμε την αναμενόμενη μεταγενέστερη εκ των προτέρων κατανομή της παραμέτρου $\boldsymbol{\theta}_k$ κάτω από την m^* ως

$$\pi^*(\boldsymbol{\theta}_k|M_k) = \int \pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{x}^*)m^*(\boldsymbol{x}^*)d\boldsymbol{x}^*.$$

Όπως και στον ενδογενή παράγοντα Μπεύες, αντίστοιχα για να εξασφαλίσουμε ότι οι EPPs θα είναι καλά ορισμένες έχουμε ότι για το φανταστικό δείγμα εκπαίδευσης ισχύει για κάθε μοντέλο M_k από το σύνολο \mathcal{M} ότι

$$0 < m_k^N(\mathbf{x}^*) < \infty, 0 < E_{\theta_k}^{M_k} \left[\frac{m^*(\mathbf{x}^*)}{m_k^N(\mathbf{x}^*)} \middle| \theta_k \right] < \infty.$$

Όπως αναφέρουν οι Perez & Berger (2002) δεν είναι απαραίτητο η κατανομή m^* να είναι κατάλληλη, οδηγώντας και σε μη-κατάλληλη εκ των προτέρων κατανομή π^* , όμως δεν παρουσιάζονται αυθαίρετες σταθερές κανονικοποίησης κατά τον καθορισμό του παράγοντα Μπεύες όπως θα αποδείξουμε και στην συνέχεια.

4.4.1 Επιλογή της προβλεπτικής κατανομής

Η επιλογή της προβλεπτικής κατανομής m^* μπορεί να βασιστεί σε υποκειμενικά και αντικειμενικά κριτήρια, καθώς αντιπροσωπεύει τις πεποιθήσεις μας σχετικά με την συμπεριφορά του φανταστικού δείγματος εκπαίδευσης \mathbf{x}^* . Στα πλαίσια της παρούσας διπλωματικής εργασίας θα μελετήσουμε μεθοδολογίες οι οποίες βασίζονται σε αντικειμενικά κριτήρια.

4.4.1.1 Επιλογή της μη πληροφοριακής περιθώριας κατανομής του μοντέλου βάσης

Μία συνηθισμένη και πρακτική επιλογή αποτελεί η κατανομή πρόβλεψης η οποία θα βασίζεται στην εκ των προτέρων κατανομή την οποία έχουμε αναθέσει στο μοντέλο βάσης. Με τον όρο **μοντέλο βάσης (base model)** αναφερόμαστε στο απλούστερο μοντέλο M_* του συνόλου \mathcal{M} , το οποίο προκύπτει ως προς την αξιολόγηση της πολυπλοκότητάς του είτε είναι εμφωλευμένο στα υπόλοιπα μοντέλα του συνόλου \mathcal{M} .

Έστω M_1 το μοντέλο το οποίο είναι εμφωλευμένο στα υπόλοιπα μοντέλα του συνόλου \mathcal{M} , με συνάρτηση πυκνότητας πιθανότητας των φανταστικών δεδομένων \mathbf{x}^* ως $f(\mathbf{x}^*|\theta_1, M_1)$ και με αντίστοιχη εκ των προτέρων κατανομή $\pi^N(\theta_1|M_1)$. Η κατανομή πρόβλεψης των δεδομένων \mathbf{x}^* θα δίνεται ως

$$m^*(\mathbf{x}^*) = m_1^N(\mathbf{x}^*) = \int_{\theta_1} f(\mathbf{x}^*|\theta_1, M_1) \pi^N(\theta_1|M_1) d\theta_1.$$

Συνεπώς η EPP για την παράμετρο θ_1 θα δίνεται από

$$\begin{aligned} \pi^*(\theta_1|M_1) &= \int \pi_1^N(\theta_1|\mathbf{x}^*) m^*(\mathbf{x}^*) d\mathbf{x}^* \\ &= \int \frac{f(\mathbf{x}^*|\theta_1, M_1) \pi^N(\theta_1|M_1)}{m^*(\mathbf{x}^*)} m^*(\mathbf{x}^*) d\mathbf{x}^* \\ &= \pi^N(\theta_1|M_1) \int f_1(\mathbf{x}^*|\theta_1, M_1) d\mathbf{x}^* \end{aligned}$$

$$= \pi^N(\boldsymbol{\theta}_1|M_1).$$

Άρα παρατηρούμε ότι η EPP για την παράμετρο $\boldsymbol{\theta}_1$ ισοδυναμεί με την αρχική μη πληροφοριακή κατανομή $\pi^N(\boldsymbol{\theta}_1|M_1)$ που είχαμε επιλέξει. Στην **Ενότητα 6** του κεφαλαίου αυτού θα παρουσιάσουμε μία ειδική περίπτωση ισοδυναμίας των EPPs και των ενδογενών εκ των προτέρων κατανομών στην περίπτωση σύγκρισης εμφωλευμένων μοντέλων.

Για οποιοδήποτε μοντέλο M_k του συνόλου \mathcal{M} η EPP που θα αντιστοιχεί στην παράμετρο $\boldsymbol{\theta}_k$ θα δίνεται από

$$\pi^*(\boldsymbol{\theta}_k|M_k) = \int \pi_k^N(\boldsymbol{\theta}_k|\mathbf{x}^*)m_1^N(\mathbf{x}^*)d\mathbf{x}^*. \quad (K4.16)$$

Εάν η εκ των προτέρων κατανομή $\pi^N(\boldsymbol{\theta}_1|M_1)$ είναι μη κατάλληλη της μορφής (K3.10) τότε δεν δημιουργείται πρόβλημα στον καθορισμό ενός παράγοντα Μπεϋες, καθώς οι αυθαίρετες σταθερές κανονικοποίησης απαλείφονται. Έστω ένα σετ δεδομένων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ δύο μοντέλα προς σύγκριση M_i και M_j του συνόλου \mathcal{M} με αντίστοιχες EPPs $\pi^*(\boldsymbol{\theta}_i|M_i)$ και $\pi^*(\boldsymbol{\theta}_j|M_j)$ οι οποίες βασίζονται στο μοντέλο βάσης M_1 όπως περιγράφονται από την σχέση (K4.16), έχοντας ως βάση μία μη κατάλληλη εκ των προτέρων κατανομή $\pi^N(\boldsymbol{\theta}_1|M_1)$. Τότε ο παράγοντας Μπεϋες του μοντέλου M_j εναντίον του μοντέλου M_i δίνεται από

$$\begin{aligned} BF_{ji}^{EPP} &= \frac{m_j^{EPP}(\mathbf{x})}{m_i^{EPP}(\mathbf{x})} \\ &= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j)\pi^*(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i)\pi^*(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i} \\ &= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \int \pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}^*)m_1^N(\mathbf{x}^*)d\mathbf{x}^* d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \int \pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}^*)m_1^N(\mathbf{x}^*)d\mathbf{x}^* d\boldsymbol{\theta}_i} \\ &= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \int \pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}^*) \int_{\boldsymbol{\theta}_1} f(\mathbf{x}^*|\boldsymbol{\theta}_1, M_1) \frac{\pi^{N^*}(\boldsymbol{\theta}_1|M_1)}{c_1} d\boldsymbol{\theta}_1 d\mathbf{x}^* d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \int \pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}^*) \int_{\boldsymbol{\theta}_1} f(\mathbf{x}^*|\boldsymbol{\theta}_1, M_1) \frac{\pi^{N^*}(\boldsymbol{\theta}_1|M_1)}{c_1} d\boldsymbol{\theta}_1 d\mathbf{x}^* d\boldsymbol{\theta}_i} \\ &= \frac{\int_{\boldsymbol{\theta}_j} f(\mathbf{x}|\boldsymbol{\theta}_j, M_j) \int \pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}^*) \int_{\boldsymbol{\theta}_1} f(\mathbf{x}^*|\boldsymbol{\theta}_1, M_1)\pi^{N^*}(\boldsymbol{\theta}_1|M_1) d\boldsymbol{\theta}_1 d\mathbf{x}^* d\boldsymbol{\theta}_j}{\int_{\boldsymbol{\theta}_i} f(\mathbf{x}|\boldsymbol{\theta}_i, M_i) \int \pi_i^N(\boldsymbol{\theta}_i|\mathbf{x}^*) \int_{\boldsymbol{\theta}_1} f(\mathbf{x}^*|\boldsymbol{\theta}_1, M_1)\pi^{N^*}(\boldsymbol{\theta}_1|M_1) d\boldsymbol{\theta}_1 d\mathbf{x}^* d\boldsymbol{\theta}_i}. \end{aligned}$$

Επομένως καταλήγουμε στον υπολογισμό ενός Παράγοντα Μπεϋες ο οποίος δεν εξαρτάται από αυθαίρετες σταθερές κανονικοποίησης.

4.4.1.2 Χρήση εμπειρικής κατανομής για προβλεπτική κατανομή

Σε αρκετές περιπτώσεις δεν είναι εφικτός ο προσδιορισμός ενός μοντέλου βάσης προκειμένου να προχωρήσουμε στον καθορισμό της προβλεπτικής κατανομής m^* , ειδικά όταν εξετάζουμε μη εμφωλευμένα μοντέλα. Μία ενδιαφέρουσα εναλλακτική επιλογή αποτελεί η χρήση της εμπειρικής κατανομής των πραγματικών δεδομένων \mathbf{x} .

Για ένα σετ δεδομένων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και για ένα φανταστικό δείγμα εκπαίδευσης $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ ορίζουμε την εμπειρική κατανομή

$$m^*(\mathbf{x}^*) = \frac{1}{L} \sum_l I_{\{\mathbf{x}(l)\}}(\mathbf{x}^*)$$

όπου I είναι η δείκτρια συνάρτηση, $\mathbf{x}(l) = (x_{l_1}, x_{l_2}, \dots, x_{l_m})$ είναι ένα υποδείγμα μεγέθους m του δείγματος παρατηρήσεων \mathbf{x} για το οποίο η εκ των υστέρων κατανομή $\pi_k^N(\boldsymbol{\theta}_k | \mathbf{x}(l))$ υπάρχει και ορίζεται για κάθε μοντέλο M_k του συνόλου \mathcal{M} και L είναι το σύνολο όλων των υποδειγμάτων μεγέθους m . Η m^* στην προκειμένη περίπτωση αντιπροσωπεύει την εκτίμηση της προβλεπτικής κατανομής του φανταστικού δείγματος εκπαίδευσης κάτω από το πραγματικό μοντέλο το οποίο μπορεί να παρήγαγε τα δεδομένα.

Η *εμπειρική αναμενόμενη μεταγενέστερη εκ των προτέρων κατανομή (empirical expected pPosterior prior)* ενός μοντέλου M_k ορίζεται ως

$$\pi^*(\boldsymbol{\theta}_k | M_k) = \frac{1}{L} \sum_l \pi_k^N(\boldsymbol{\theta}_k | \mathbf{x}(l)).$$

Επομένως η εμπειρική EPP προκύπτει από την επαναδειγματοληψία του δείγματος εκπαίδευσης μεγέθους m από το σετ των πραγματικών δεδομένων \mathbf{x} .

Επομένως ο παράγοντας Μπεϋες υπέρ του μοντέλου M_j θα δίνεται από

$$BF_{ji}^{EEP} = \frac{\sum_l m_j^N(\mathbf{x} | \mathbf{x}(l))}{\sum_l m_i^N(\mathbf{x} | \mathbf{x}(l))}.$$

Όπως αναφέρουν και οι Perez & Berger (2002), καθώς το μέγεθος δείγματος n των παρατηρήσεων \mathbf{x} τείνει στο άπειρο και κάτω από ορισμένες συνθήκες ομαλότητας, η Εμπειρική EPP ενός μοντέλου M_k του συνόλου \mathcal{M} συγκλίνει στην κατανομή

$$\pi_G^*(\boldsymbol{\theta}_k) = E_G[\pi_k^N(\boldsymbol{\theta}_k | \mathbf{X}^*)]$$

όπου $G(\mathbf{x})$ η κατανομή από την οποία προέρχονται τα δεδομένα \mathbf{x} . Εφόσον η κατανομή G είναι άγνωστη κατά συνέπεια και η Εμπειρική EPP π_G^* θα είναι και αυτή άγνωστη.

4.5 Πλεονεκτήματα EPPs

Οι EPPs είναι σε θέση να μας λύσουν αρκετές δυσκολίες κατά την μπεϋζιανή επιλογή μοντέλου επιτρέποντας μας την χρήση μη-πληροφοριακών και μη κατάλληλων κατανομών, όπως επίσης την αυτόματη παραγωγή εκ των προτέρων κατανομών. Τα πλεονεκτήματα χρήσης τους είναι τα εξής:

- i) Είναι αναγκαίος ο καθορισμός μόνο μίας κατανομής πρόβλεψης m^* προκειμένου να υπολογιστούν όλες οι υπόλοιπες εκ των προτέρων κατανομές των υπό εξέταση μοντέλων.
- ii) Οι απαραίτητοι υπολογισμοί για τον καθορισμό των EPPs είναι εφικτοί μέσω χρήσης αλγορίθμων MCMC.
- iii) Δεν υπάρχουν περιορισμοί ως προς την χρήση μη κατάλληλων εκ των προτέρων κατανομών, καθώς οι οποιεσδήποτε αυθαίρετες σταθερές κανονικοποίησης απαλείφονται είτε κατά την διαδικασία καθορισμού των κατανομών είτε κατά τον υπολογισμό παραγόντων Μπεϋς.
- iv) Οι EPPs προσαρμόζονται αρκετά καλά στην ερμηνεία των παραμέτρων καθώς προχωράμε από μοντέλο σε μοντέλο.
- v) Γίνεται εφικτή η πολλαπλή σύγκριση μοντέλων, δηλαδή $BF_{ji}/BF_{ik} = BF_{jk}$, καθώς οι EPPs είναι συγκεκριμένες εκ των προτέρων κατανομές, χαρακτηριστικό το οποίο δεν συναντάται σε άλλες μεθοδολογίες.
- vi) Όπως αναφέρει και ο Consonni (2008) οι EPPs είναι συμβατές μεταξύ τους καθώς κινούμαστε από το ένα μοντέλο στο άλλο.

4.6 Ειδική περίπτωση ισοδυναμίας EPP και ενδογενών εκ των προτέρων κατανομών

Σε αυτήν την ενότητα θα μελετήσουμε μία ειδική περίπτωση ισοδυναμίας αποτελεσμάτων των ενδογενών εκ των προτέρων κατανομών με των αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών στην περίπτωση εξέτασης εμφωλευμένων μοντέλων.

Αρχικά έστω το σετ φανταστικών δεδομένων $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ για το οποίο ισχύει η ανταλλαξιμότητα και επίσης έστω ένα σύνολο προς εξέταση μοντέλων $\mathcal{M} = \{M_1, M_2, \dots, M_k\}$ με $k \geq 2$.

Οι ενδογενείς εκ των προτέρων κατανομές δύο μοντέλων M_i και M_j όπου το M_i είναι εμφωλευμένο του μοντέλου M_j , δίνονται από τις σχέσεις

$$\pi^I(\boldsymbol{\theta}_i|M_i) = \pi^N(\boldsymbol{\theta}_i|M_i), \quad \pi^I(\boldsymbol{\theta}_j|M_j) = \pi^N(\boldsymbol{\theta}_j|M_j)B_j^*(\boldsymbol{\theta}_j).$$

όπως παρουσιάστηκαν στην (K4.11). Μία βασική ομοιότητα των δύο μεθοδολογιών είναι ότι βασίζονται σε φανταστικά δείγματα εκπαίδευσης. Επομένως για ένα φανταστικό σετ δεδομένων $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$, για το οποίο θα ισχύει η ανταλλαξιμότητα, η ενδογενής εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_j μπορεί να γραφεί ως εξής

$$\begin{aligned}
\pi^l(\boldsymbol{\theta}_j|M_j) &= \pi^N(\boldsymbol{\theta}_j|M_j)B_j^*(\boldsymbol{\theta}_j) \\
&= \pi^N(\boldsymbol{\theta}_j|M_j)E_{\boldsymbol{\theta}_j}^{M_j} \left[\frac{m_i^N(\mathbf{x}^*)}{m_j^N(\mathbf{x}^*)} \right] \\
&= \pi^N(\boldsymbol{\theta}_j|M_j) \int \frac{m_i^N(\mathbf{x}^*)}{m_j^N(\mathbf{x}^*)} f(\mathbf{x}^*|\boldsymbol{\theta}_j, M_j) d\mathbf{x}^* \\
&= \int \pi^N(\boldsymbol{\theta}_j|M_j) \frac{m_i^N(\mathbf{x}^*)}{m_j^N(\mathbf{x}^*)} f(\mathbf{x}^*|\boldsymbol{\theta}_j, M_j) d\mathbf{x}^* \\
&= \int \frac{f(\mathbf{x}^*|\boldsymbol{\theta}_j, M_j)\pi^N(\boldsymbol{\theta}_j|M_j)}{m_j^N(\mathbf{x}^*)} m_i^N(\mathbf{x}^*) d\mathbf{x}^* \\
&= \int \pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}^*) m_i^N(\mathbf{x}^*) d\mathbf{x}^*.
\end{aligned}$$

Αφού έχουμε προχωρήσει στην σύγκριση των μοντέλων M_i και M_j έχοντας το μοντέλο M_i εμφωλευμένο του μοντέλου M_j , τότε το μοντέλο M_i αποτελεί το μοντέλο βάσης. Επομένως η αναμενόμενη μεταγενέστερη εκ των προτέρων κατανομή του μοντέλου M_i θα είναι η αρχική μη πληροφοριακή κατανομή $\pi^N(\boldsymbol{\theta}_i|M_i)$ και η αντίστοιχη του μοντέλου M_j θα δίνεται από την σχέση (Κ4.16) και θα περιγράφεται ως

$$\int \pi_j^N(\boldsymbol{\theta}_j|\mathbf{x}^*) m_i^N(\mathbf{x}^*) d\mathbf{x}^* .$$

Άρα παρατηρούμε ότι η λύση του συστήματος των ενδογενών εξισώσεων όταν εξετάζουμε εμφωλευμένα μοντέλα, ισοδυναμεί με τις EPPs όταν πραγματοποιούμε την προσέγγιση μέσω του μοντέλου βάσης. Αυτή είναι και η μοναδική περίπτωση όπου οι δύο αυτές μέθοδοι παράγουν ισοδύναμα αποτελέσματα, καθώς σε όλες τις άλλες περιπτώσεις σύγκρισης μοντέλων τα αποτελέσματα είναι ασύνδετα.

4.7 Επίλογος 4^ο Κεφαλαίου

Σε αυτό το κεφάλαιο εξετάσαμε δύο βασικές μεθοδολογίες παραγωγής εκ των προτέρων κατανομών οι οποίες βασίζονται σε φανταστικά δείγματα και επιλύουν τα προβλήματα που ανακύπτουν από την χρήση μη κατάλληλων και μη πληροφοριακών κατανομών. Οι ενδογενείς εκ των προτέρων κατανομές προχωράνε στον καθορισμό εκ των προτέρων κατανομών κατά την σύγκριση δύο μοντέλων, μέσω του συστήματος των ενδογενών εξισώσεων, ενώ οι EPPs αποτελούν μία πλήρως αυτοματοποιημένη διαδικασία για την παραγωγή εκ των προτέρων κατανομών σε όλο το σύνολο των μοντέλων. Το κεφάλαιο αυτό αποτελεί το τελευταίο θεωρητικό κομμάτι της παρούσας διπλωματικής εργασίας, καθώς στην συνέχεια θα προχωρήσουμε σε εφαρμογές των ενδογενών παραγόντων Μπεϋές και των ενδογενών εκ των προτέρων κατανομών.

Κεφάλαιο 5

Εφαρμογές

Στο κεφάλαιο αυτό θα προχωρήσουμε στην παρουσίαση των αποτελεσμάτων τεσσάρων εφαρμογών που προκύπτουν με την βοήθεια της γλώσσας προγραμματισμού R. Ο κώδικας ο οποίος χρησιμοποιήθηκε για την διενέργεια των εφαρμογών παρουσιάζεται αναλυτικά στο *Παράρτημα 1*.

5.1 Υπολογισμός ενδογενών παραγόντων Μπεϋες σε πραγματικά δεδομένα

Σε αυτήν την ενότητα θα παρουσιάσουμε αναλυτικά το *Παράδειγμα 6* της *Ενότητας 3.1* της εργασίας των Berger & Perrichì (1996), όπου πραγματοποιήθηκε η σύγκριση μη εμφολευμένων μοντέλων χρησιμοποιώντας τον αριθμητικό και γεωμετρικό ενδογενή Παράγοντα Μπεϋες. Η εφαρμογή αυτή αφορά σε δεδομένα τα οποία συνέλεξε ο Proschan (1963) τα οποία μελετούν ανεξάρτητους χρόνους αποτυχίας λειτουργίας κλιματιστικών αεροπλάνων. Τα μοντέλα τα οποία θα εξετάσουμε για ένα δείγμα ανεξάρτητων παρατηρήσεων χρόνων αποτυχίας $\mathbf{x} = (x_1, x_2, \dots, x_n)$ προερχόμενο από την τυχαία μεταβλητή X είναι:

- $M_1: f(x_i|\theta_1, M_1) = \theta_1 e^{-x_i\theta_1}, x_i > 0, \theta_1 > 0$ με $\pi^N(\theta_1|M_1) = \frac{c_1}{\theta_1}$.
- $M_2: f(x_i|\mu, \sigma, M_2) = \frac{1}{\sqrt{2\pi\sigma x_i}} e^{-\frac{1}{2\sigma^2}(\log(x_i)-\mu)^2}$ με $\pi^N(\mu, \sigma|M_2) = \frac{c_2}{\sigma}$.

με c_1 και c_2 οι αυθαίρετες σταθερές κανονικοποίησης.

Η περιθώρια κατανομή των δεδομένων \mathbf{x} κάτω από το μοντέλο M_1 θα δίνεται από

$$\begin{aligned} m_1^N(\mathbf{x}) &= \int_0^\infty f(\mathbf{x}|\theta, M_1)\pi^N(\theta|M_1)d\theta \\ &= \int_0^\infty \theta^n e^{-\theta \sum x_i} \frac{c_1}{\theta} d\theta \\ &= c_1 \int_0^\infty \theta^{n-1} e^{-\theta \sum x_i} d\theta. \end{aligned}$$

Παρατηρούμε στο παραπάνω ολοκλήρωμα ότι σχηματίζεται ο πυρήνας μίας κατανομής γάμμα με παραμέτρους $\alpha = n$ και $\beta = \sum x_i$. Συνεπώς το παραπάνω ολοκλήρωμα μας δίνει

$$m_1^N(\mathbf{x}) = c_1 \frac{\Gamma(n)}{(\sum x_i)^n}. \quad (E1.1)$$

Η περιθώρια κατανομή των δεδομένων \mathbf{x} κάτω από το μοντέλο M_2 θα δίνεται από

$$\begin{aligned}
m_2^N(\mathbf{x}) &= \int_{-\infty}^{\infty} \int_0^{\infty} f(\mathbf{x}|\theta, M_1) \pi^N(\mu, \sigma|M_2) d\sigma d\mu \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sigma^n \prod x_i} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2\right] \frac{c_2}{\sigma} d\sigma d\mu \\
&= \frac{c_2}{\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{-n-1} \exp\left[-\frac{\sum_{i=1}^n (\log x_i - \mu)^2 / 2}{\sigma^2}\right] d\sigma d\mu.
\end{aligned}$$

Θέτοντας $\sigma = \sqrt{k}$ έχουμε ότι $d\sigma = 1/2\sqrt{k} dk$ και $0 < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα μας δίνει ότι

$$\begin{aligned}
m_2^N(\mathbf{x}) &= \frac{c_2}{\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} \int_0^{\infty} k^{-\frac{n}{2}-\frac{1}{2}} \exp\left[-\frac{\sum_{i=1}^n (\log x_i - \mu)^2 / 2}{k}\right] \frac{1}{2\sqrt{k}} dk d\mu \\
&= \frac{c_2}{2\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} \int_0^{\infty} k^{-\frac{n}{2}-1} \exp\left[-\frac{\sum_{i=1}^n (\log x_i - \mu)^2 / 2}{k}\right] dk d\mu.
\end{aligned}$$

Παρατηρούμε ότι ο πυρήνας του παραπάνω ολοκληρώματος ως προς k αντιπροσωπεύει μία αντίστροφη Γάμμα κατανομή με παραμέτρους $\alpha = n/2$ και $\beta = \sum_{i=1}^n (\log x_i - \mu)^2 / 2$, επομένως παίρνουμε ότι

$$\begin{aligned}
m_2^N(\mathbf{x}) &= \frac{c_2}{2\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\sum_{i=1}^n (\log x_i - \mu)^2 / 2\right)^{\frac{n}{2}}} d\mu \\
&= \frac{c_2 \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}{2\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} \frac{1}{\left(\sum_{i=1}^n (\log x_i - \mu)^2\right)^{\frac{n}{2}}} d\mu. \quad (E1.2)
\end{aligned}$$

Για την απλοποίηση των υπολογισμών θέτουμε $y_i = \log x_i$, επομένως ο παρονομαστής μας δίνει ότι

$$\sum_{i=1}^n (y_i - \mu)^2 = n(\mu - \bar{y})^2 + S_y^2$$

όπου $S_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2$. Επομένως η σχέση (E1.2) μας δίνει ότι

$$m_2^N(\mathbf{x}) = \frac{c_2 \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}{2\sqrt{2\pi}^n \prod x_i} \int_{-\infty}^{\infty} (n(\mu - \bar{y})^2 + S_y^2)^{-\frac{n}{2}} d\mu. \quad (E1.3)$$

Η μη τυποποιημένη Student t κατανομή δίνεται από τον τύπο

$$f(x|v, \mu, \sigma) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{\pi v} \sigma} \left\{ 1 + \frac{1}{v} \frac{(x - \mu)^2}{\sigma^2} \right\}^{-\frac{v+1}{2}} \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0, v > 0.$$

Στην συνέχεια θα επεξεργαστούμε το ολοκλήρωμα της σχέσης (E1.3) προκειμένου να το φέρουμε στην μορφή μίας μη τυποποιημένης Student t κατανομής. Επομένως έχουμε ότι

$$\begin{aligned} (n(\mu - \bar{y})^2 + S_y^2)^{-\frac{n}{2}} &= \left(S_y^2 \left(\frac{n(\mu - \bar{y})^2}{S_y^2} + 1 \right) \right)^{-\frac{(n-1)+1}{2}} \\ &= (S_y^2)^{-\frac{(n-1)+1}{2}} \left(\frac{(\mu - \bar{y})^2}{\frac{S_y^2}{n}} + 1 \right)^{-\frac{(n-1)+1}{2}} \\ &= S_y^{-n} \left(1 + \frac{(\mu - \bar{y})^2}{\frac{S_y^2}{n(n-1)}} \right)^{-\frac{(n-1)+1}{2}}. \end{aligned}$$

Παρατηρούμε ότι στην παραπάνω σχέση εμφανίζεται ο πυρήνας μίας μη τυποποιημένης Student t κατανομής με παραμέτρους $\mu = \bar{y}$, $v = n - 1$, $\sigma^2 = S_y^2 / \{n(n - 1)\}$. Επομένως η σχέση (E1.3) διαμορφώνεται ως

$$\begin{aligned} m_2^N(\mathbf{x}) &= \frac{c_2 \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}} \frac{1}{S_y^n} \Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi} \sqrt{n-1} \sqrt{\frac{S_y^2}{n(n-1)}}}{2\sqrt{2\pi}^n \Gamma\left(\frac{n-1+1}{2}\right) \prod x_i} \\ &= \frac{c_2 2^{\frac{n}{2}-1-\frac{n}{2}} \Gamma\left(\frac{n-1}{2}\right) S_y}{\sqrt{\pi}^{n-1} S_y^n \sqrt{n} \prod x_i} \\ &= \frac{c_2 \Gamma\left(\frac{n-1}{2}\right)}{\prod x_i \sqrt{\pi}^{n-1} S_y^{n-1} 2\sqrt{n}}. \quad (E1.4) \end{aligned}$$

Συνεπώς οι περιθώριες συναρτήσεις πιθανοφάνειας των δύο μοντέλων θα δίνονται από τις σχέσεις (E1.1) και (E1.4), χωρίς τις αυθαίρετες σταθερές κανονικοποίησης καθώς απλοποιούνται κατά τον υπολογισμό των ενδογενών παραγόντων Μπεύες, δηλαδή

$$m_1^N(\mathbf{x}) = \frac{\Gamma(n)}{(\sum x_i)^n}$$

και

$$m_2^N(\mathbf{x}) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\prod x_i \sqrt{\pi}^{n-1} S_y^{n-1} 2\sqrt{n}}.$$

Το ελάχιστο δείγμα εκπαίδευσης θα αποτελείται από δύο παρατηρήσεις αφού έχουμε δύο άγνωστες παραμέτρους, συνεπώς θα είναι της μορφής

$$\mathbf{x}(l) = \{x_i, x_j\}, \quad x_i \neq x_j.$$

Επομένως δοθέντος ενός ελαχίστου δείγματος εκπαίδευσης $\mathbf{x}(l)$ οι περιθώριες συναρτήσεις πιθανοφάνειας γίνονται

$$m_1^N(\mathbf{x}(l)) = \frac{1}{(x_i + x_j)^2}$$

και

$$m_2^N(\mathbf{x}(l)) = \frac{1}{2x_i x_j \left| \log\left(\frac{x_i}{x_j}\right) \right|}.$$

Χρησιμοποιώντας τις σχέσεις (Κ3.5) και (Κ3.6) υπολογίζουμε τον αριθμητικό και αντίστοιχα τον γεωμετρικό ενδογενή παράγοντα Μπεύες υπέρ του μοντέλου M_2 ως

$$\begin{aligned} \text{AIBF}_{21} &= BF_{21}^{N*} \frac{1}{L} \sum_{l=1}^L BF_{12}^{N*}(\mathbf{x}(l)) \\ &= \frac{m_2^N(\mathbf{x})}{m_1^N(\mathbf{x})} \frac{1}{L} \sum_{l=1}^L \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} \\ &= \frac{\frac{\Gamma\left(\frac{n-1}{2}\right)}{\prod x_i \sqrt{\pi}^{n-1} S_y^{n-1} 2\sqrt{n}}}{\frac{\Gamma(n)}{(\sum x_i)^n}} \frac{1}{L} \sum_{l=1}^L \frac{\frac{1}{(x_i(l) + x_j(l))^2}}{\frac{1}{2x_i(l)x_j(l) \left| \log\left(\frac{x_i(l)}{x_j(l)}\right) \right|}} \\ &= \frac{\Gamma\left(\frac{n-1}{2}\right) (\sum x_i)^n}{\Gamma(n) \prod x_i \sqrt{\pi}^{n-1} S_y^{n-1} 2\sqrt{n}} \frac{1}{L} \sum_{l=1}^L \frac{2x_i(l)x_j(l) \left| \log\left(\frac{x_i(l)}{x_j(l)}\right) \right|}{(x_i(l) + x_j(l))^2} \quad (E1.5) \end{aligned}$$

και

$$\begin{aligned}
GIBF_{21} &= BF_{21}^{N*} \left(\prod_{l=1}^L BF_{12}^{N*}(\mathbf{x}(l)) \right)^{1/L} \\
&= \frac{m_2^N(\mathbf{x})}{m_1^N(\mathbf{x})} \left(\prod_{l=1}^L \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} \right)^{1/L} \\
&= \frac{\Gamma\left(\frac{n-1}{2}\right) (\sum x_i)^n}{\Gamma(n) \prod x_i \sqrt{\pi}^{n-1} S_y^{n-1} 2\sqrt{n}} \left(\prod_{l=1}^L \frac{2x_i(l)x_j(l) \left| \log\left(\frac{x_i(l)}{x_j(l)}\right) \right|}{(x_i(l) + x_j(l))^2} \right)^{1/L}. \quad (E1.6)
\end{aligned}$$

5.1.1 Εφαρμογή Δεδομένων του Proshan (1963)

Ο Proshan συνέλεξε το ακόλουθο δείγμα $n = 30$ παρατηρήσεων χρόνων αποτυχίας από ένα συγκεκριμένο αεροπλάνο: 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95. Το σύνολο \mathcal{X}_L των ελαχίστων δειγμάτων εκπαίδευσης θα αποτελείται από $\binom{30}{2} = 435$ διαθέσιμα ζευγάρια παρατηρήσεων. Όμως στο παραπάνω δείγμα υπάρχουν παρατηρήσεις οι οποίες είναι ίσες. Συνεπώς χρησιμοποιώντας την R, κατασκευάζουμε το σύνολο \mathcal{X}_L αφήνοντας εκτός 9 ζευγάρια ίσων παρατηρήσεων, δηλαδή έχουμε ότι $L = 426$. Υποθέτουμε για τις εκ των προτέρων πιθανότητες των μοντέλων $\pi(M_1)$ και $\pi(M_2)$ ότι

$$\pi(M_1) = \pi(M_2) = \frac{1}{2}$$

επομένως ο εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων του μοντέλου M_2 εναντίον του μοντέλου M_1 θα ισούται με τον αντίστοιχο παράγοντα Μπεϋες του μοντέλου υπέρ του M_2 . Επομένως έχουμε ότι

$$\begin{aligned}
PO_{21} &= AIBF_{21} \\
\Rightarrow \frac{\pi^{AI}(M_2|\mathbf{x})}{\pi^{AI}(M_1|\mathbf{x})} &= AIBF_{21} \\
\Rightarrow \frac{\pi^{AI}(M_2|\mathbf{x})}{1 - \pi^{AI}(M_2|\mathbf{x})} &= AIBF_{21} \\
\Rightarrow \pi^{AI}(M_2|\mathbf{x}) &= \frac{AIBF_{21}}{1 + AIBF_{21}}
\end{aligned}$$

και αντίστοιχα για τον γεωμετρικό ενδογενή παράγοντα Μπεϋες έχουμε ότι

$$\pi^{GI}(M_2|\mathbf{x}) = \frac{GIBF_{21}}{1 + GIBF_{21}}.$$

Στον **Πίνακα 5.1** παρουσιάζονται οι ενδογενείς παράγοντες Μπεϋες υπέρ του μοντέλου M_2 καθώς και οι εκ των υστέρων πιθανότητες των μοντέλων M_1 και M_2 .

Πίνακας 5.1 Τιμές ενδογενών παραγόντων Μπεϋες και εκ των υστέρων πιθανοτήτων

$AIBF_{21}$	$\pi^{AI}(M_2 \mathbf{x})$	$\pi^{AI}(M_1 \mathbf{x})$	$GIBF_{21}$	$\pi^{GI}(M_2 \mathbf{x})$	$\pi^{GI}(M_1 \mathbf{x})$
0.3749037	0.2726763	0.7273237	0.3262656	0.2460032	0.7539968

Παρατηρούμε ότι το μοντέλο M_1 το οποίο περιγράφει το εκθετικό μοντέλο έχει εκ των υστέρων πιθανότητα $\pi^{AI}(M_1|\mathbf{x}) = 0.7273237$, συνεπώς προτιμάται αρκετά περισσότερο από το μοντέλο M_2 το οποίο περιγράφει το λογαριθμοκανονικό μοντέλο, άρα ευνοείται το απλούστερο μοντέλο.

5.2 Έλεγχος παραμέτρου Εκθετικής κατανομής

Σε αυτήν την εφαρμογή θα υπολογίσουμε τις ενδογενείς εκ των προτέρων κατανομές όταν εξετάζουμε δύο μοντέλα όπου ελέγχουμε την παράμετρο λ μίας Εκθετικής κατανομής χρησιμοποιώντας μία μη κατάλληλη εκ των προτέρων κατανομή και στην συνέχεια βασιζόμενοι στις ενδογενείς εκ των προτέρων κατανομές θα προχωρήσουμε στον υπολογισμό του ενδογενή παράγοντα Μπεϋες.

Έστω τυχαίο δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ για το οποίο ισχύει η ανταλλαξιμότητα, και τα υπό εξέταση μοντέλα

$$M_1: \mathbf{X} \sim \text{Exp}(\lambda_0) \text{ με } \lambda_0 \in \mathbb{R} \text{ γνωστό}$$

$$M_2: \mathbf{X} \sim \text{Exp}(\lambda) \text{ με } \lambda \neq \lambda_0, \pi^N(\lambda | M_2) \propto \frac{1}{\lambda}.$$

Είναι προφανές ότι τα δύο εξεταζόμενα μοντέλα είναι εμφωλευμένα, συνεπώς θα καταφύγουμε στο σύστημα ενδογενών εξισώσεων όπως διαμορφώνεται στην περίπτωση σύγκρισης εμφωλευμένων μοντέλων, όπου περιγράφεται από τις εξισώσεις όπως παρουσιάστηκαν στην (K4.11)

$$\pi^I(\lambda_0 | M_1) = \pi^N(\lambda_0 | M_1) = I_{(0, \infty)}(\lambda_0), \quad \pi^I(\lambda | M_2) = \pi^N(\lambda | M_2) B_2^*(\lambda).$$

Σκοπός μας είναι ο υπολογισμός του ενδογενή παράγοντα Μπεϋες

$$BF_{21}^I = \frac{m_2^I(\mathbf{x})}{m_1^I(\mathbf{x})}. \quad (E2.1)$$

Επειδή για το μοντέλο M_1 η τιμή λ_0 είναι μία γνωστή σταθερά, αυτόματα ισχύει για την περιθώρια πιθανοφάνεια του ότι

$$m_1^I(\mathbf{x}) = \lambda_0^n e^{-\lambda_0 n \bar{x}}. (E2.2)$$

Για να προχωρήσουμε στον υπολογισμό της περιθώριας πιθανοφάνειας του μοντέλου M_2 κάνοντας χρήση της ενδογενούς εκ των προτέρων κατανομής $\pi_2^I(\lambda)$, είναι αναγκαίος ο υπολογισμός αυτής.

Συνεπώς έχουμε αρχικά ότι

$$\pi^N(\lambda|M_2) \propto \frac{1}{\lambda} = \frac{c}{\lambda}$$

όπου c η αυθαίρετη σταθερά κανονικοποίησης. Άρα

$$\begin{aligned} \pi^I(\lambda|M_2) &= \pi^N(\lambda|M_2)B_2^*(\lambda) \\ &= \frac{c}{\lambda} E^{M_2}[BF_{12}^N(\mathbf{x}(l))]. (E2.3) \end{aligned}$$

Επειδή το πρόβλημα το οποίο εξετάζουμε έχει μία άγνωστη μονοδιάστατη παράμετρο, το ελάχιστο δείγμα εκπαίδευσης θα αποτελείται από μία παρατήρηση. Συνεπώς

$$\mathbf{x}(l) = \{x\}.$$

Για να προχωρήσουμε στον καθορισμό της ενδογενούς εκ των προτέρων κατανομής για το μοντέλο M_2 , χρειάζεται να υπολογίσουμε τον παράγοντα Μπεϋες ο οποίος θα βασίζεται στο ελάχιστο δείγμα εκπαίδευσης,

$$BF_{12}^N(\mathbf{x}(l)) = \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} = \frac{m_1^N(x)}{m_2^N(x)}. (E2.4)$$

Επειδή για το μοντέλο M_1 η τιμή λ_0 είναι μία γνωστή σταθερά, αυτόματα ισχύει για την περιθώρια πιθανοφάνεια του ότι

$$m_1^N(x) = m_1(x) = \lambda_0 e^{-\lambda_0 x}. (E2.5)$$

Για την περιθώρια πιθανοφάνεια του μοντέλου M_2 έχουμε ότι

$$\begin{aligned} m_2^N(x) &= \int_0^{+\infty} f(x|\lambda, M_2) \pi^N(\lambda|M_2) d\lambda \\ &= \int_0^{+\infty} \lambda e^{-\lambda x} \frac{c}{\lambda} d\lambda \\ &= c \int_0^{+\infty} e^{-\lambda x} d\lambda \end{aligned}$$

$$= \frac{c}{x}$$

Επομένως έχουμε ότι

$$m_2^N(x) = \frac{c}{x}. \quad (E2.6)$$

Άρα από τις σχέσεις (E2.5) και (E2.6) ο παράγοντας Μπεϋές που περιγράφεται από την σχέση (E2.4) λαμβάνει την μορφή

$$BF_{12}^N(x(l)) = \frac{\lambda_0 e^{-\lambda_0 x}}{\frac{c}{x}} = \frac{x \lambda_0 e^{-\lambda_0 x}}{c}. \quad (E2.7)$$

Μετά τον καθορισμό του παράγοντα Μπεϋές της σχέσης (E2.7) προχωράμε στον υπολογισμό της μέσης τιμής της σχέσης (E2.3), επομένως

$$\begin{aligned} E^{M_2}[BF_{12}^N(x)] &= \int_0^{+\infty} BF_{12}^N(x) f(x|\lambda, M_2) dx \\ &= \int_0^{+\infty} \frac{x \lambda_0 e^{-\lambda_0 x}}{c} \lambda e^{-\lambda x} dx \\ &= \frac{\lambda_0 \lambda}{c} \int_0^{\infty} x e^{-(\lambda + \lambda_0)x} dx. \end{aligned}$$

Παρατηρούμε στο παραπάνω ολοκλήρωμα ότι εμφανίζεται ο πυρήνας μίας κατανομής γάμμα με παραμέτρους $a = 2$ και $\beta = \lambda + \lambda_0$. Συνεπώς παίρνουμε ότι

$$\begin{aligned} E^{M_2}[BF_{12}^N(x)] &= \frac{\lambda_0 \lambda}{c} \frac{\Gamma(2)}{(\lambda + \lambda_0)^2} \int_0^{\infty} \frac{x(\lambda + \lambda_0)^2 e^{-(\lambda + \lambda_0)x}}{\Gamma(2)} dx \\ &= \frac{1}{c} \frac{\lambda_0 \lambda}{(\lambda + \lambda_0)^2}. \end{aligned}$$

Επομένως η σχέση (E2.3) διαμορφώνεται

$$\begin{aligned} \pi^I(\lambda|M_2) &= \frac{c}{\lambda} E^{M_2}[BF_{12}^N(x(l))] \\ &= \frac{c}{\lambda} \frac{\lambda_0 \lambda}{(\lambda + \lambda_0)^2} \\ &= \frac{\lambda_0}{(\lambda + \lambda_0)^2}. \end{aligned}$$

Άρα η ενδογενής εκ των προτέρων κατανομή του μοντέλου M_2 θα δίνεται από

$$\begin{aligned}\pi^l(\lambda|M_2) &= \frac{\lambda_0}{(\lambda + \lambda_0)^2} \cdot (E2.8) \\ &= \frac{\lambda_0}{\lambda_0^2 \left(\frac{\lambda}{\lambda_0} + 1\right)^2} \\ &= \frac{1}{\lambda_0} \frac{1}{\left(\frac{\lambda}{\lambda_0} + 1\right)^2}.\end{aligned}$$

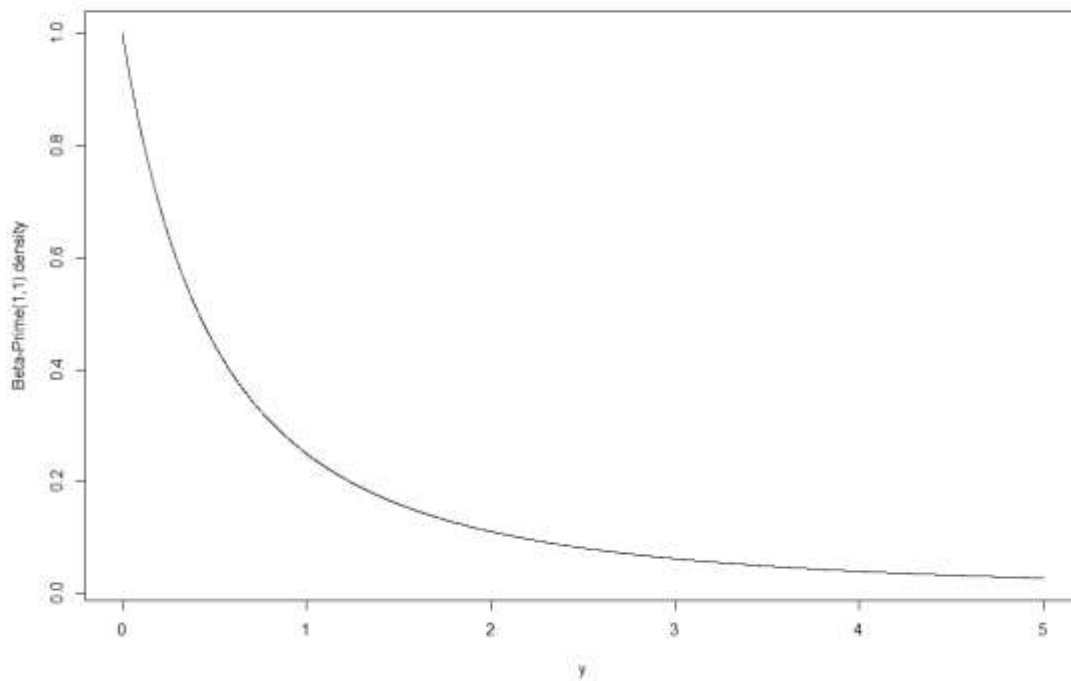
Παρατηρούμε ότι για

$$y = \frac{\lambda}{\lambda_0} = f(\lambda) \Rightarrow f^{-1}(y) = \lambda_0 y \Rightarrow \frac{df^{-1}(y)}{dy} = \lambda_0.$$

Άρα η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y θα είναι

$$f_Y(y) = \frac{1}{\lambda_0} \frac{1}{(y+1)^2} \left| \frac{df^{-1}(y)}{dy} \right| = \frac{1}{\lambda_0} \frac{1}{(y+1)^2} \lambda_0 = \frac{1}{(y+1)^2}.$$

Παρατηρούμε ότι η τυχαία μεταβλητή λ/λ_0 δοθέντος λ_0 ακολουθεί την *Beta-Prime* κατανομή με παραμέτρους $\alpha = 1$ και $b = 1$ και απεικονίζεται στο **Διάγραμμα 5.1**.



Διάγραμμα 5.1 : Διάγραμμα Beta-Prime κατανομής με $\alpha = 1$ και $b = 1$

Επομένως μετά τον καθορισμό της $\pi^I(\lambda|M_2)$ προχωράμε στον υπολογισμό της περιθώριας πιθανοφάνειας του μοντέλου M_2 κάνοντας χρήση της παραπάνω εκ των προτέρων κατανομής. Συνεπώς έχουμε ότι

$$\begin{aligned} m_2^I(\mathbf{x}) &= \int_0^{+\infty} f(\mathbf{x}|\lambda, M_2)\pi^I(\lambda|M_2) d\lambda \\ &= \int_0^{+\infty} \lambda^n \exp\left[-\lambda \sum_{i=1}^n x_i\right] \frac{\lambda_0}{(\lambda + \lambda_0)^2} d\lambda. \end{aligned} \quad (E1.9)$$

Στην συνέχεια θέτουμε $\lambda + \lambda_0 = y$, άρα $\lambda = y - \lambda_0$, $d\lambda = dy$ και $0 < \lambda < \infty \Rightarrow \lambda_0 < y < \infty$. Επομένως η σχέση (E2.9) διαμορφώνεται ως εξής

$$m_2^I(\mathbf{x}) = \lambda_0 \int_{\lambda_0}^{+\infty} \frac{1}{y^2} (y - \lambda_0)^n e^{-(y-\lambda_0)n\bar{x}} dy. \quad (E2.10)$$

Στο παραπάνω ολοκλήρωμα περιγράφεται μία Γενικευμένη κατανομή γάμμα τεσσάρων παραμέτρων, η οποία στην γενική της μορφή όπως αυτή περιγράφεται από τον Tsionas (2007), δίνεται από

$$f(x|\alpha, \theta, \gamma, c) = \theta^\alpha c \frac{1}{\Gamma(\alpha)} (x - \gamma)^{\alpha-1} e^{-\theta(x-\gamma)^c}, \text{ με } x > \gamma.$$

Επομένως για $x = y$, $\gamma = \lambda_0$, $\alpha = n + 1$, $\theta = n\bar{x}$, $c = 1$ παίρνουμε ότι

$$\begin{aligned} f(y|n + 1, n\bar{x}, \lambda_0, 1) &= (n\bar{x})^{n+1} \frac{1}{\Gamma(n + 1)} (y - \lambda_0)^{1(n+1)-1} e^{-n\bar{x}(y-\lambda_0)^1} \\ \Rightarrow f(y|n + 1, n\bar{x}, \lambda_0, 1) &= (n\bar{x})^{n+1} \frac{1}{\Gamma(n + 1)} (y - \lambda_0)^n e^{-n\bar{x}(y-\lambda_0)} \text{ με } y > \lambda_0. \end{aligned} \quad (E2.11)$$

Συνεπώς η σχέση (E2.10) διαμορφώνεται ως

$$m_2^I(\mathbf{x}) = \lambda_0 \frac{\Gamma(n + 1)}{(n\bar{x})^{n+1}} \int_{\lambda_0}^{+\infty} \frac{1}{y^2} \frac{(n\bar{x})^{n+1}}{\Gamma(n + 1)} (y - \lambda_0)^n e^{-(y-\lambda_0)n\bar{x}} dy. \quad (E2.12)$$

Το ολοκλήρωμα της σχέσης (E2.12) αναπαριστά την μέση τιμή $E(Y^{-2})$ όταν η τυχαία μεταβλητή Y ακολουθεί μία γενικευμένη κατανομή γάμμα τεσσάρων παραμέτρων, όπως αυτή περιγράφεται από την σχέση (E2.11). Η μέση τιμή αυτή εκφράζει την αρνητική ροπή δεύτερης τάξης της τυχαίας μεταβλητής Y , όπου οι Cressie & Davis et. al. (2013) έδειξαν ότι δίνεται από

$$E[Y^{-r}] = \frac{1}{\Gamma(r)} \int_0^\infty t^{r-1} M_Y(-t) dt. \quad (E2.13)$$

Συνεπώς χρειάζεται αρχικά να υπολογίσουμε την ροπογεννήτρια της τυχαίας μεταβλητής Y όταν αυτή ακολουθεί μία γενικευμένη κατανομή γάμμα τεσσάρων παραμέτρων με παραμέτρους $\gamma = \lambda_0, \alpha = n + 1, \theta = n\bar{x}, c = 1$. Επομένως έχουμε ότι

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \int_{\lambda_0}^{+\infty} e^{ty} \frac{(n\bar{x})^{n+1}}{\Gamma(n+1)} (y - \lambda_0)^n e^{-(y-\lambda_0)n\bar{x}} dy \\ &= \frac{(n\bar{x})^{n+1}}{\Gamma(n+1)} \int_{\lambda_0}^{+\infty} (y - \lambda_0)^n e^{-yn\bar{x} + \lambda_0 n\bar{x} + ty} dy \\ &= \frac{(n\bar{x})^{n+1}}{\Gamma(n+1)} \int_{\lambda_0}^{+\infty} (y - \lambda_0)^n e^{-yn\bar{x} + \lambda_0 n\bar{x} + ty + t\lambda_0 - t\lambda_0} dy \\ &= \frac{(n\bar{x})^{n+1}}{\Gamma(n+1)} e^{t\lambda_0} \int_{\lambda_0}^{+\infty} (y - \lambda_0)^n e^{-(n\bar{x}-t)(y-\lambda_0)} dy. \end{aligned}$$

Παρατηρούμε ότι στο παραπάνω ολοκλήρωμα σχηματίζεται μία καινούρια Γενικευμένη Γάμμα κατανομή τεσσάρων παραμέτρων με καινούρια παράμετρο $\theta' = (n\bar{x} - t)$, επομένως παίρνουμε ότι

$$\begin{aligned} M_Y(t) &= \frac{(n\bar{x})^{n+1}}{(n\bar{x} - t)^{n+1}} e^{t\lambda_0} \int_{\lambda_0}^{+\infty} \frac{(n\bar{x} - t)^{n+1}}{\Gamma(n+1)} (y - \lambda_0)^n e^{-(n\bar{x}-t)(y-\lambda_0)} dy \\ &= \frac{(n\bar{x})^{n+1}}{(n\bar{x} - t)^{n+1}} e^{t\lambda_0}, \quad n\bar{x} < t. \end{aligned}$$

Επομένως η ροπογεννήτρια συνάρτηση μίας τυχαίας μεταβλητής Y όταν αυτή ακολουθεί μία Γενικευμένη Γάμμα κατανομή τεσσάρων παραμέτρων με παραμέτρους $\gamma = \lambda_0, \alpha = n + 1, \theta = n\bar{x}, c = 1$ δίνεται από

$$M_Y(t) = \frac{(n\bar{x})^{n+1}}{(n\bar{x} - t)^{n+1}} e^{t\lambda_0}. \quad (E2.14)$$

Άρα για $r = 2$ και χρησιμοποιώντας την ροπογεννήτρια συνάρτηση της σχέσης (E2.14) η σχέση (E2.13) μας δίνει

$$\begin{aligned} E[Y^{-2}] &= \frac{1}{\Gamma(2)} \int_0^{\infty} t^{2-1} M_Y(-t) dt \\ &= \int_0^{\infty} t \frac{(n\bar{x})^{n+1}}{(n\bar{x} + t)^{n+1}} e^{-t\lambda_0} dt \\ &= (n\bar{x})^{n+1} \int_0^{\infty} \frac{t}{(n\bar{x} + t)^{n+1}} e^{-t\lambda_0} dt. \end{aligned}$$

Στην συνέχεια θέτουμε $n\bar{x} + t = k$, άρα $t = k - n\bar{x}$ και επίσης $dt = dk$. Επίσης για $0 < t < \infty \Rightarrow n\bar{x} < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα διαμορφώνεται ως

$$\begin{aligned} E[Y^{-2}] &= (n\bar{x})^{n+1} \int_{n\bar{x}}^{\infty} \frac{k - n\bar{x}}{k^{n+1}} e^{-(k-n\bar{x})\lambda_0} dk \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \int_{n\bar{x}}^{\infty} \frac{k - n\bar{x}}{k^{n+1}} e^{-k\lambda_0} dk \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\int_{n\bar{x}}^{\infty} k^{-n} e^{-k\lambda_0} dk - n\bar{x} \int_{n\bar{x}}^{\infty} k^{-(n+1)} e^{-k\lambda_0} dk \right). \end{aligned} \quad (E2.15)$$

Στην συνέχεια θα πραγματοποιήσουμε την ίδια αλλαγή μεταβλητής και στα δύο ολοκληρώματα, θέτοντας $k\lambda_0 = w$, άρα $k = w/\lambda_0$ και $dk = 1/\lambda_0 dw$. Επίσης για $n\bar{x} < k < \infty \Rightarrow \lambda_0 n\bar{x} < w < \infty$.

Άρα η σχέση (E2.15) διαμορφώνεται ως

$$\begin{aligned} E[Y^{-2}] &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\int_{\lambda_0 n\bar{x}}^{\infty} \left(\frac{w}{\lambda_0}\right)^{-n} e^{-w} \frac{1}{\lambda_0} dw - n\bar{x} \int_{\lambda_0 n\bar{x}}^{\infty} \left(\frac{w}{\lambda_0}\right)^{-(n+1)} e^{-w} \frac{1}{\lambda_0} dw \right) \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\frac{1}{\lambda_0^{-n+1}} \int_{\lambda_0 n\bar{x}}^{\infty} w^{-n} e^{-w} dw - n\bar{x} \frac{1}{\lambda_0^{-(n+1)+1}} \int_{\lambda_0 n\bar{x}}^{\infty} w^{-(n+1)} e^{-w} dw \right) \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\lambda_0^{n-1} \int_{\lambda_0 n\bar{x}}^{\infty} w^{-n} e^{-w} dw - n\bar{x} \lambda_0^n \int_{\lambda_0 n\bar{x}}^{\infty} w^{-(n+1)} e^{-w} dw \right). \end{aligned} \quad (E2.16)$$

Στην συνέχεια πραγματοποιούμε έναν τελευταίο μετασχηματισμό, θέτοντας $w = z\lambda_0 n\bar{x}$, άρα $dw = \lambda_0 n\bar{x} dz$ και για $\lambda_0 n\bar{x} < w < \infty \Rightarrow 1 < z < \infty$. Συνεπώς η σχέση (E2.16) μας δίνει

$$\begin{aligned} E[Y^{-2}] &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\lambda_0^{n-1} \int_1^{\infty} (z\lambda_0 n\bar{x})^{-n} e^{-z\lambda_0 n\bar{x}} \lambda_0 n\bar{x} dz - n\bar{x} \lambda_0^n \int_1^{\infty} (z\lambda_0 n\bar{x})^{-(n+1)} e^{-z\lambda_0 n\bar{x}} \lambda_0 n\bar{x} dz \right) \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left(\lambda_0^{n-1} (\lambda_0 n\bar{x})^{-n} \lambda_0 n\bar{x} \int_1^{\infty} z^{-n} e^{-z\lambda_0 n\bar{x}} dz - n\bar{x} \lambda_0^n \lambda_0 n\bar{x} (\lambda_0 n\bar{x})^{-(n+1)} \int_1^{\infty} z^{-(n+1)} e^{-z\lambda_0 n\bar{x}} dz \right) \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} \left((n\bar{x})^{-n+1} \int_1^{\infty} z^{-n} e^{-z\lambda_0 n\bar{x}} dz - (n\bar{x})^{-n+1} \int_1^{\infty} z^{-(n+1)} e^{-z\lambda_0 n\bar{x}} dz \right) \\ &= (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} (n\bar{x})^{-n+1} [E_n[\lambda_0 n\bar{x}] - E_{n+1}[\lambda_0 n\bar{x}]] \end{aligned} \quad (E2.17)$$

όπου

$$E_n[x] = \int_1^{\infty} t^{-n} e^{-tx} dt$$

είναι η *συνάρτηση του εκθετικού ολοκληρώματος (exponential integral function)* ενός αριθμού x (πραγματικού ή μιγαδικού) όπως αυτή ορίζεται στο **Κεφάλαιο 5** του βιβλίου των Abramowitz & Stegun (1964). Μία χαρακτηριστική ιδιότητα αυτής της συνάρτησης, όπως αναφέρεται στο ίδιο κεφάλαιο, είναι ότι

$$E_{n+1}[x] = \frac{1}{n} [e^{-x} - xE_n[x]], n = 1, 2, 3, \dots \quad (E2.18).$$

Συνεπώς χρησιμοποιώντας την σχέση (E2.18) η σχέση (E2.17) μας δίνει

$$E[Y^{-2}] = (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} (n\bar{x})^{-n+1} \left[E_n[\lambda_0 n\bar{x}] - \frac{1}{n} [e^{-\lambda_0 n\bar{x}} - \lambda_0 n\bar{x} E_n[\lambda_0 n\bar{x}]] \right]. \quad (E2.19)$$

Χρησιμοποιώντας την σχέση (E2.19) η σχέση (E2.12) μας δίνει

$$m_2^I(x) = \lambda_0 \frac{\Gamma(n+1)}{(n\bar{x})^{n+1}} (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} (n\bar{x})^{-n+1} \left[E_n[\lambda_0 n\bar{x}] - \frac{1}{n} [e^{-\lambda_0 n\bar{x}} - \lambda_0 n\bar{x} E_n[\lambda_0 n\bar{x}]] \right]. \quad (E2.20)$$

Επομένως χρησιμοποιώντας τις σχέσεις (E2.2) και (E2.20) ο ενδογενής παράγοντας Μπεύες της σχέσης (E2.1) διαμορφώνεται

$$BF_{21}^I = \frac{\lambda_0 \frac{\Gamma(n+1)}{(n\bar{x})^{n+1}} (n\bar{x})^{n+1} e^{n\bar{x}\lambda_0} (n\bar{x})^{-n+1} \left[E_n[\lambda_0 n\bar{x}] - \frac{1}{n} [e^{-\lambda_0 n\bar{x}} - \lambda_0 n\bar{x} E_n[\lambda_0 n\bar{x}]] \right]}{\lambda_0^n e^{-\lambda_0 n\bar{x}}}$$

$$= \lambda_0^{1-n} e^{2\lambda_0 n\bar{x}} \Gamma(n+1) (n\bar{x})^{1-n} \left[E_n[\lambda_0 n\bar{x}] - \frac{1}{n} [e^{-\lambda_0 n\bar{x}} - \lambda_0 n\bar{x} E_n[\lambda_0 n\bar{x}]] \right],$$

όπου ο όρος $E_n[\lambda_0 n\bar{x}]$ υπολογίζεται με την βοήθεια του πακέτου *gsl* της R μέσω της εντολής

$$\text{expint_En}(n, lo * n * \text{mean}(x), \text{give} = \text{FALSE}, \text{strict} = \text{FALSE}).$$

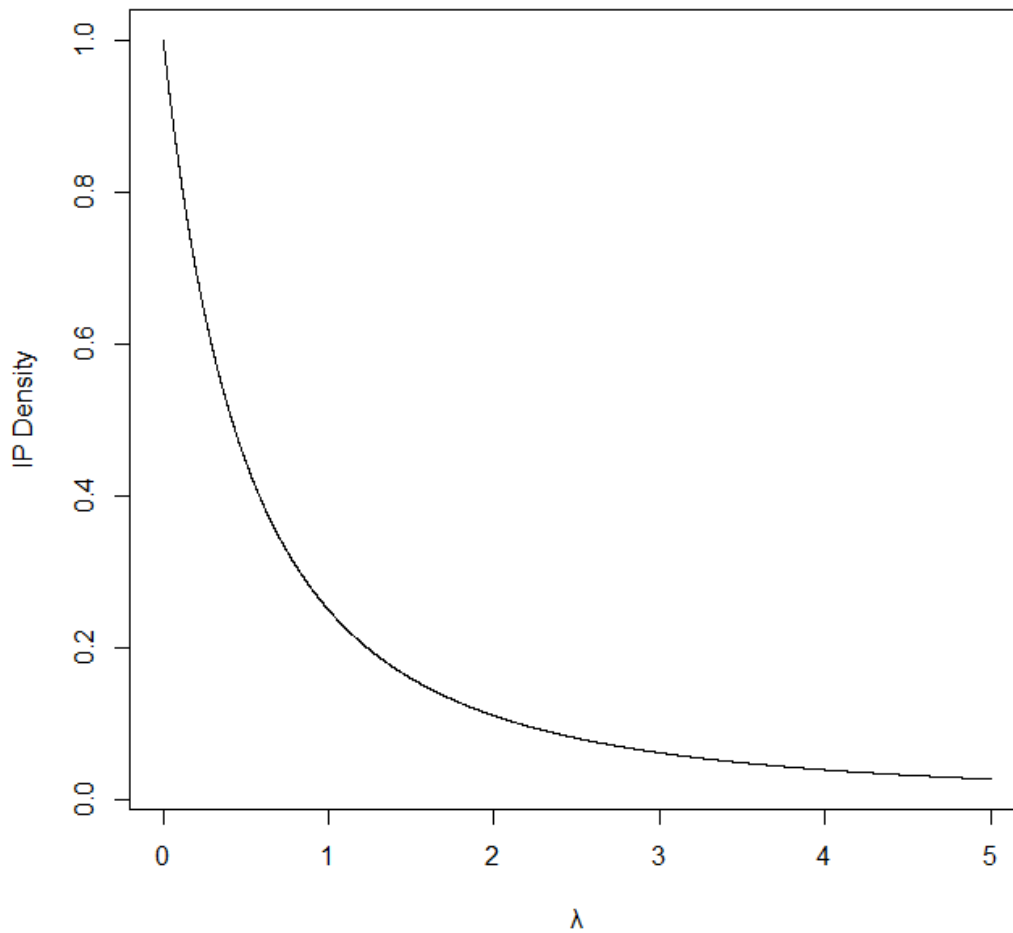
5.2.1 Παράδειγμα με προσομοιωμένα δεδομένα

Έστω ότι θέλουμε να συγκρίνουμε τα μοντέλα

$$M_1: X \sim \text{Exp}(1)$$

$$M_2: X \sim \text{Exp}(\lambda) \text{ με } \lambda > 0, \pi^N(\lambda | M_2) \propto \frac{1}{\lambda}.$$

Σύμφωνα με τα αποτελέσματα της **Ενότητας 5.2** η Ενδογενής εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 , η οποία περιγράφεται από την σχέση (E2.8) απεικονίζεται στο **Διάγραμμα 5.2**.



Διάγραμμα 5.2 : Διάγραμμα Ενδογενούς εκ των προτέρων κατανομής $\pi^I(\lambda|M_2)$

Υποθέτουμε ότι για τις εκ των προτέρων πιθανότητες των μοντέλων M_1 και M_2 ισχύει

$$\pi(M_1) = \pi(M_2) = \frac{1}{2}$$

επομένως ο εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων υπέρ του μοντέλου M_2 θα ισούται με τον παράγοντα Μπεϋες υπέρ του μοντέλου M_2 . Επομένως έχουμε ότι

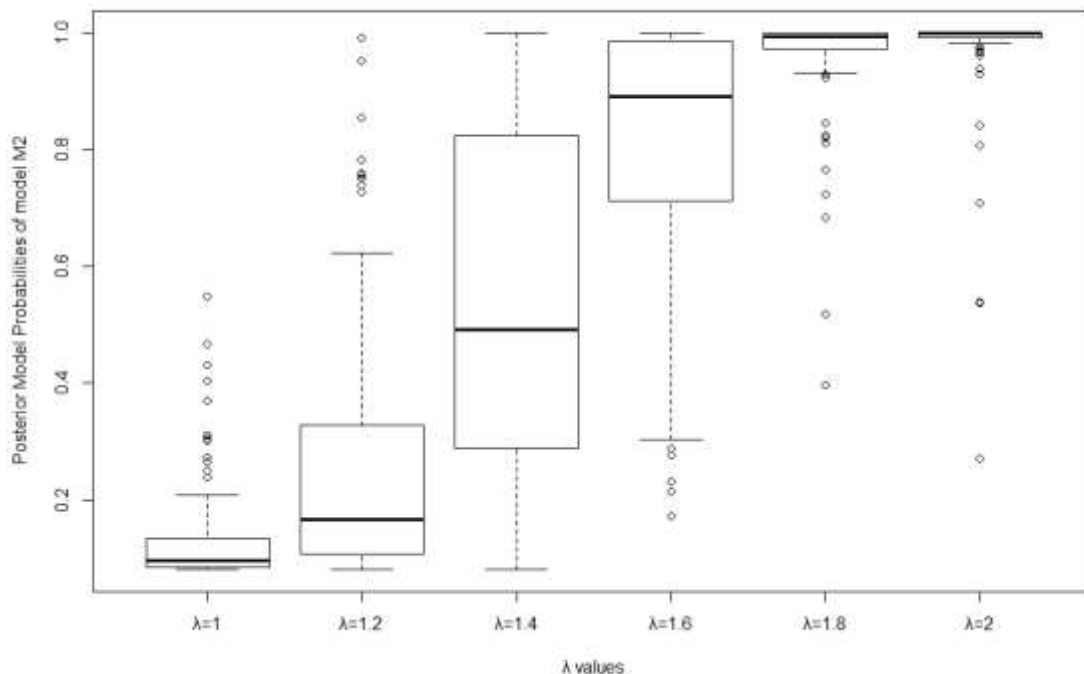
$$\begin{aligned} PO_{21} &= BF_{21} \\ \Rightarrow \frac{\pi(M_2|\mathbf{x})}{\pi(M_1|\mathbf{x})} &= BF_{21} \\ \Rightarrow \frac{\pi(M_2|\mathbf{x})}{1 - \pi(M_2|\mathbf{x})} &= BF_{21} \end{aligned}$$

$$\Rightarrow \pi(M_2|\mathbf{x}) = \frac{BF_{21}}{1 + BF_{21}} . (E2.21)$$

Εν συνεχεία μπορούμε να προσομοιώσουμε τυχαία δείγματα από την εκθετική κατανομή για διάφορες τιμές της παραμέτρου λ που θέλουμε να εξετάσουμε, προκειμένου να μελετήσουμε την συμπεριφορά της εκ των υστέρων πιθανότητας του μοντέλου M_2 . Θα πραγματοποιήσουμε 2 κύκλους προσομοιώσεων όπου θα παράγουμε τυχαία δείγματα προερχόμενα από την Εκθετική κατανομή σύμφωνα με τις τιμές της παραμέτρου λ όπου έχουμε επιλέξει και στην συνέχεια θα προχωράμε στον υπολογισμό και την αποθήκευση των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 όπως αυτές προκύπτουν από την σχέση (E2.21).

1^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 100 τυχαία δείγματα τα οποία θα αποτελούνται από 50 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την Εκθετική κατανομή με παράμετρο $\lambda = 1, 1.2, 1.4, 1.6, 1.8, 2$. Στο **Διάγραμμα 5.3** παρουσιάζονται τα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου λ για 100 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 50$.



Διάγραμμα 5.3: Θηκογράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για $n = 50$

Στον **Πίνακα 5.2** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|\mathbf{x}) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου λ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

Πίνακας 5.2 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου λ

Ποσοστά επιλογής μοντέλου M_2					
$\lambda = 1$	$\lambda = 1.2$	$\lambda = 1.4$	$\lambda = 1.6$	$\lambda = 1.8$	$\lambda = 2$
1%	16%	49%	83%	99%	99%

Όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την Εκθετική κατανομή με παράμετρο $\lambda = 1$, στις 99% των περιπτώσεων προτιμάται το μοντέλο M_1 ως το μοντέλο το οποίο γέννησε τα προσομοιωμένα δεδομένα. Η παρουσία έκτροπων τιμών οφείλεται στο γεγονός ότι μελετάμε προσομοιωμένα δεδομένα, συνεπώς υπάρχει και το αντίστοιχο σφάλμα Monte Carlo, όπως επίσης χρησιμοποιούμε τυχαία δείγματα τα οποία αποτελούνται από αρκετά μικρό αριθμό παρατηρήσεων.

Αντίστοιχα όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την εκθετική κατανομή με παράμετρο $\lambda = 1.2$ παρατηρούμε ξανά ότι επιλέγεται το μοντέλο M_1 ως το μοντέλο που το οποίο γέννησε τα προσομοιωμένα δεδομένα, αλλά στο 84% των περιπτώσεων. Επίσης από το θηκοδιάγραμμα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 παρατηρούμε μία μεγάλη μεταβλητότητα των εκ των υστέρων πιθανοτήτων του, γεγονός που οφείλεται στους ίδιους λόγους που αναφέραμε στην προηγούμενη περίπτωση, καθώς και στην τιμή της παραμέτρου λ την οποία μελετάμε, η οποία βρίσκεται αρκετά κοντά στην τιμή που υποστηρίζει το μοντέλο M_1 .

Στην συνέχεια όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την εκθετική κατανομή με παράμετρο $\lambda = 1.6$ δεν είμαστε σε θέση να απαντήσουμε με βεβαιότητα, καθώς στο 51% των περιπτώσεων επιλέγεται το μοντέλο M_1 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα και στις υπόλοιπες επιλέγεται το M_2 μοντέλο. Για να μπορέσουμε να λάβουμε μία πιο ξεκάθαρη εικόνα αλλά και για να μειωθεί η μεταβλητότητα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 , θα χρειαστεί να παραγάγουμε περισσότερες παρατηρήσεις.

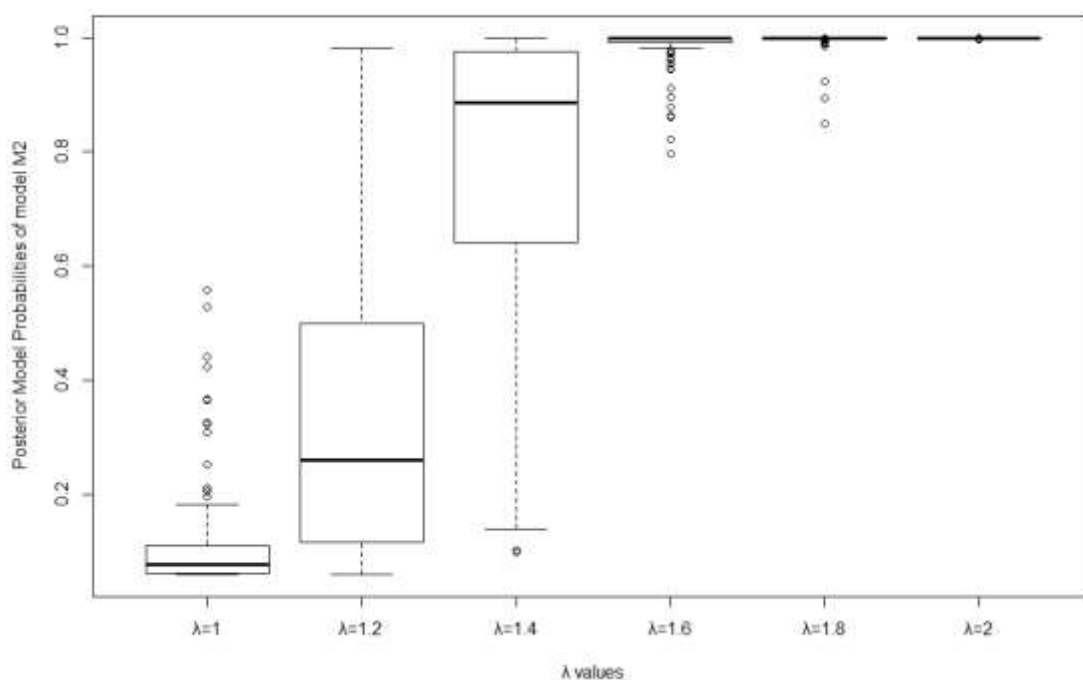
Όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την εκθετική κατανομή με παράμετρο $\lambda = 1.8$, επιλέγεται στο 83% των περιπτώσεων το μοντέλο M_2 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα. Αντίστοιχα όπως όταν είχαμε παρατηρήσεις προερχόμενες από την εκθετική κατανομή με παράμετρο $\lambda = 1.2$, αντίστοιχα και εδώ εμφανίζονται έκτροπες παρατηρήσεις όπως επίσης και μεγάλη

μεταβλητότητα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για τους ίδιους ακριβώς λόγους που αναφέραμε παραπάνω.

Στις δύο τελευταίες περιπτώσεις όπου τα προσομοιωμένα τυχαία δείγματα προέρχονται από την εκθετική κατανομή με παράμετρο $\lambda = 1.8$ και $\lambda = 2$ αντίστοιχα, επιλέγεται το μοντέλο M_2 ως αυτό το οποίο παρήγαγε τα προσομοιωμένα δεδομένα στο 99% των περιπτώσεων. Αντίστοιχα και σε αυτές τις περιπτώσεις εμφανίζονται έκτροπες τιμές για τους ίδιους λόγους, όμως γίνεται ξεκάθαρη η υπεροχή του μοντέλου M_2 επομένως είμαστε σε θέση να επιλέξουμε με βεβαιότητα το μοντέλο M_2 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα.

2^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 100 τυχαία δείγματα τα οποία θα αποτελούνται από 100 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την Εκθετική κατανομή με παράμετρο $\lambda = 1, 1.2, 1.4, 1.6, 1.8, 2$. Στο **Διάγραμμα 5.4** παρουσιάζονται τα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου λ για 100 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 100$.



Διάγραμμα 5.4: Θηκογράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για $n = 100$

Στον **Πίνακα 5.3** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|x) > 0.5$), ως το μοντέλο το οποίο

περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου λ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

Πίνακας 5.3 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου λ

Ποσοστά επιλογής μοντέλου M_2					
$\lambda = 1$	$\lambda = 1.2$	$\lambda = 1.4$	$\lambda = 1.6$	$\lambda = 1.8$	$\lambda = 2$
2%	25%	83%	100%	100%	100%

Σε αυτόν τον κύκλο προσομοιώσεων λαμβάνουμε περισσότερο ξεκάθαρα αποτελέσματα συγκριτικά με τον πρώτο κύκλο προσομοιώσεων. Όταν εξετάζουμε προσομοιωμένα τυχαία δείγματα προερχόμενα από την εκθετική κατανομή με παράμετρο $\lambda = 1$ λαμβάνουμε ακριβώς τα ίδια αποτελέσματα με ελάχιστα μικρότερη ακρίβεια, καθώς δεχόμαστε το μοντέλο M_1 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα στο 98% των περιπτώσεων, αλλά όπως και στον πρώτο κύκλο προσομοιώσεων, έτσι και εδώ αυτή η παρουσία έκτροπων τιμών οφείλεται στο σφάλμα Monte Carlo της προσομοίωσης.

Στην συνέχεια παρατηρούμε ότι όταν τα προσομοιωμένα τυχαία δείγματα προερχόμενα από την εκθετική κατανομή με παράμετρο $\lambda = 1.2$, οι εκ των υστέρων πιθανότητες του μοντέλου M_2 εμφανίζουν μεγαλύτερη μεταβλητότητα συγκριτικά με τον πρώτο κύκλο προσομοιώσεων, όμως αυξήθηκε το ποσοστό επιλογής του μοντέλου M_2 ως το μοντέλο το οποίο παρήγαγε τα δεδομένα κατά 9%.

Όταν τα προσομοιωμένα τυχαία δείγματα προερχόμενα από την εκθετική κατανομή με παράμετρο $\lambda = 1.4$ παρατηρούμε μία σημαντική μείωση της μεταβλητότητας των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 . Επίσης είναι σημαντικό το γεγονός, πως ενώ στον πρώτο κύκλο προσομοιώσεων επιλέγαμε το μοντέλο M_2 στο 49% των περιπτώσεων και δυσκολευόμασταν να αποφασίσουμε με βεβαιότητα, στον δεύτερο κύκλο παρατηρούμε μία αύξηση του ποσοστού επιλογής του μοντέλου M_2 της τάξης του 34%, όπως επίσης επιτυγχάνεται μία σημαντική μείωση της μεταβλητότητας των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 .

Στις τελευταίες τρεις περιπτώσεις, όπου τα προσομοιωμένα τυχαία δείγματα προέρχονται από την εκθετική κατανομή με παραμέτρους $\lambda = 1.6, 1.8, 2$ αντίστοιχα, παρατηρούμε επίσης μία πολύ σημαντική μείωση της μεταβλητότητας των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 , καθώς επιλέγεται το μοντέλο M_2 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα στο 100% των περιπτώσεων

Συνεπώς για να επιτύχουμε την μείωση της μεταβλητότητας των εκ των υστέρων πιθανοτήτων, όπως επίσης και τον περιορισμό της εμφάνισης έκτροπων τιμών, αυξάνουμε το πλήθος των παρατηρήσεων που παράγουμε από την εκάστοτε εκθετική κατανομή. Όμως στην συγκεκριμένη εφαρμογή μπορούμε να υπολογίσουμε τις εκ των υστέρων πιθανότητες του μοντέλου M_2 για δείγματα μεγέθους το πολύ 120 παρατηρήσεων. Αυτός ο περιορισμός αφορά καθαρά τεχνικούς λόγους καθώς η συνάρτηση `expint_En` στην **R**, δεν είναι εφικτό να επιστρέψει κάποια τιμή για πολύ μεγάλες τιμές της τιμής του n . Για την αποφυγή τέτοιων προβλημάτων καταφεύγουμε σε μεθόδους ολοκλήρωσης Monte Carlo, καθώς επιτυγχάνουμε την απλοποίηση των υπολογισμών των Παραγόντων Μπεϋες όπως επίσης και μείωση του χρόνου των υπολογισμών αυτών.

5.3 Έλεγχος μέσης τιμής κανονικής κατανομής με γνωστή διακύμανση

Σε αυτήν την εφαρμογή θα υπολογίσουμε τις ενδογενείς εκ των προτέρων κατανομές όταν εξετάζουμε δύο μοντέλα, όπου ελέγχουμε την μέση τιμή μιας κανονικής κατανομής με γνωστή διακύμανση, χρησιμοποιώντας μία μη κατάλληλη εκ των προτέρων κατανομή.

Έστω τυχαίο δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ για το οποίο ισχύει η ανταλλαξιμότητα, και τα υπό εξέταση μοντέλα

$$M_1: \mathbf{X} \sim N(\mu_0, \sigma^2) \text{ με } \mu_0 \in \mathbb{R} \text{ και } \sigma > 0 \text{ γνωστές σταθερές}$$

$$M_2: \mathbf{X} \sim N(\mu, \sigma^2), \pi^N(\mu|M_2) \propto 1 \text{ και } \sigma > 0 \text{ γνωστή σταθερά.}$$

Είναι προφανές ότι τα δύο εξεταζόμενα μοντέλα είναι εμφωλευμένα, συνεπώς θα καταφύγουμε στο σύστημα ενδογενών εξισώσεων όπως διαμορφώνεται στην περίπτωση σύγκρισης εμφωλευμένων μοντέλων, όπου περιγράφεται από τις εξισώσεις όπως παρουσιάστηκαν στην (K4.11)

$$\pi^I(\mu_0|M_1) = \pi^N(\mu_0|M_1) = I_{\mathbb{R}}(\mu_0), \quad \pi^I(\mu|M_2) = \pi^N(\mu|M_2)B_2^*(\mu).$$

Σκοπός μας είναι ο υπολογισμός του παράγοντα Μπεϋες

$$BF_{21}^I = \frac{m_2^I(\mathbf{x})}{m_1^I(\mathbf{x})}.$$

Οι Didelot & Everitt et al. (2011) έδειξαν ότι όταν εξετάζουμε εμφωλευμένα μοντέλα, ο παράγοντας Μπεϋες χρησιμοποιώντας τις παρατηρήσεις \mathbf{x} , ισοδυναμεί με τον παράγοντα Μπεϋες χρησιμοποιώντας την επαρκή στατιστική συνάρτηση του πολυπλοκότερου μοντέλου. Ο αριθμητικός μέσος \bar{x} είναι επαρκής στατιστική συνάρτηση της παραμέτρου μ κάτω από το μοντέλο M_2 , οπότε ο παράγοντας Μπεϋες είναι

$$BF_{21}^I = \frac{m_2^I(\bar{x})}{m_1^I(\bar{x})}. \quad (E2.1)$$

Όταν μία τυχαία μεταβλητή X ακολουθεί κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 τότε ο αριθμητικός της μέσος \bar{X} θα ακολουθεί και αυτός κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2/n .

Επομένως έχουμε ότι

$$m_1^I(\bar{x}) = \sqrt{n}(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\right]. \quad (E2.2)$$

Για να προχωρήσουμε στον υπολογισμό της περιθώριας πιθανοφάνειας του μοντέλου M_2 κάνοντας χρήση της ενδογενούς εκ των προτέρων κατανομής $\pi_2^I(\mu)$, είναι αναγκαίος ο υπολογισμός αυτής.

Συνεπώς έχουμε αρχικά ότι

$$\pi^N(\mu|M_2) \propto 1 = c$$

όπου c η αυθαίρετη σταθερά κανονικοποίησης. Άρα

$$\begin{aligned} \pi^I(\mu|M_2) &= \pi^N(\mu|M_2)B_2^*(\mu) \\ &= cE^{M_2}[BF_{12}^N(\mathbf{x}(l))]. \quad (E2.3) \end{aligned}$$

Επειδή το πρόβλημα το οποίο εξετάζουμε έχει μία άγνωστη μονοδιάστατη παράμετρο, το ελάχιστο δείγμα εκπαίδευσης θα αποτελείται από μία παρατήρηση. Συνεπώς

$$\mathbf{x}(l) = \{x\}.$$

Για να προχωρήσουμε στον καθορισμό της ενδογενούς εκ των προτέρων κατανομής για το μοντέλο M_2 , υπολογίζουμε τον παράγοντα Μπεϋες ο οποίος βασίζεται στο ελάχιστο δείγμα εκπαίδευσης

$$BF_{12}^N(\mathbf{x}(l)) = \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} = \frac{m_1^N(x)}{m_2^N(x)}. \quad (E2.4)$$

Επειδή για το μοντέλο M_1 η τιμή μ_0 είναι μία γνωστή σταθερά, αυτόματα ισχύει για την περιθώρια πιθανοφάνεια $m_1^N(x)$ ότι

$$m_1^N(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu_0)^2\right]. \quad (E2.5)$$

Για την περιθώρια πιθανοφάνεια του μοντέλου M_2 έχουμε ότι

$$m_2^N(x) = \int_{-\infty}^{+\infty} f(x|\mu, M_2)\pi^N(\mu|M_2) d\mu$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] c \, d\mu \\
&= c \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\mu-x)^2\right] d\mu.
\end{aligned}$$

Παρατηρούμε ότι το ολοκλήρωμα περιγράφει μία συνάρτηση πυκνότητας πιθανότητας μίας κανονικής κατανομής ως προς μ , με μέση τιμή x και διακύμανση σ^2 . Συνεπώς έχουμε ότι

$$m_2^N(x) = c. \quad (E2.6)$$

Άρα από τις σχέσεις (E2.5) και (E2.6) ο παράγοντας Μπεύες που περιγράφεται από την σχέση (E2.4) λαμβάνει την μορφή

$$BF_{12}^N(x) = \frac{m_1^N(x)}{m_2^N(x)} = \frac{1}{c} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu_0)^2\right]. \quad (E2.7)$$

Μετά τον καθορισμό του παράγοντα Μπεύες της σχέσης (E2.7) προχωράμε στον υπολογισμό της μέσης τιμής της σχέσης (E2.3), επομένως

$$\begin{aligned}
E^{M_2}[BF_{12}^N(x)] &= \int_{-\infty}^{+\infty} BF_{12}^N(x) f(x|\mu, M_2) dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{c} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu_0)^2\right] \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx \\
&= \frac{1}{c} \frac{1}{\sigma^2 2\pi} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2}(x^2 - 2x\mu_0 + \mu_0^2 + x^2 - 2x\mu + \mu^2)\right] dx \\
&= \frac{1}{c} \frac{1}{\sigma^2 2\pi} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2}(2x^2 - 2x(\mu_0 + \mu) + \mu_0^2 + \mu^2)\right] dx \\
&= \frac{1}{c} \frac{\exp\left[-\frac{1}{2\sigma^2}(\mu_0^2 + \mu^2)\right]}{\sigma^2 2\pi} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2} \left(x^2 - 2x\frac{(\mu_0 + \mu)}{2} + \left(\frac{(\mu_0 + \mu)}{2}\right)^2\right)\right] dx \\
&= \frac{1}{c} \frac{1}{\sigma^2 2\pi} \exp\left[-\frac{1}{2\sigma^2}(\mu_0^2 + \mu^2) + \frac{1}{4\sigma^2}(\mu_0 + \mu)^2\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2} \left(x - \frac{\mu_0 + \mu}{2}\right)^2\right] dx.
\end{aligned}$$

Παρατηρούμε ότι το ολοκλήρωμα απεικονίζει τον πυρήνα μίας κανονικής κατανομής με μέση τιμή $(\mu_0 + \mu)/2$ και διακύμανση $\sigma^2/2$, επομένως παίρνουμε

$$E^{M_2}[BF_{12}^N(x)] = \frac{1}{c} \frac{1}{\sigma^2 2\pi} \exp\left[-\frac{1}{2\sigma^2}(\mu_0^2 + \mu^2) + \frac{1}{4\sigma^2}(\mu_0 + \mu)^2\right] \frac{\sigma}{\sqrt{2}} \sqrt{2\pi}$$

Επομένως η σχέση (E2.3) διαμορφώνεται ως

$$\begin{aligned}
 \pi^I(\mu|M_2) &= c \frac{1}{\sigma^2 2\pi} \exp\left[-\frac{1}{2\sigma^2}(\mu_0^2 + \mu^2) + \frac{1}{4\sigma^2}(\mu_0 + \mu)^2\right] \frac{\sigma}{\sqrt{2}} \sqrt{2\pi} \\
 &= \frac{\sqrt{\pi}}{\sigma 2\pi} \exp\left[-\frac{1}{2\sigma^2}\left(\mu_0^2 + \mu^2 - \frac{1}{2}(\mu_0 + \mu)^2\right)\right] \\
 &= \frac{1}{\sigma\sqrt{2}\sqrt{2}\sqrt{\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(\mu_0^2 + \mu^2 - \frac{1}{2}\mu_0^2 - \frac{1}{2}2\mu_0\mu - \frac{1}{2}\mu^2\right)\right] \\
 &= \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{1}{2}\mu_0^2 - 2\frac{\mu_0}{\sqrt{2}}\frac{\mu}{\sqrt{2}} + \frac{1}{2}\mu^2\right)\right] \\
 &= \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{\mu}{\sqrt{2}} - \frac{\mu_0}{\sqrt{2}}\right)^2\right] \\
 &= \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp\left[-\frac{1}{2(2\sigma^2)}(\mu - \mu_0)^2\right]. \quad (E2.8)
 \end{aligned}$$

Άρα η ενδογενής εκ των προτέρων κατανομή της παραμέτρου μ είναι μία κανονική κατανομή με μέση τιμή μ_0 και διακύμανση $2\sigma^2$.

Επομένως μετά τον καθορισμό της $\pi^I(\mu|M_2)$ προχωράμε στον υπολογισμό της περιθώριας πιθανοφάνειας $m_2^I(\bar{x})$ του μοντέλου M_2 κάνοντας χρήση της παραπάνω κατανομής. Συνεπώς έχουμε ότι

$$m_2^I(\bar{x}) = \int_{-\infty}^{\infty} f(\bar{x}|\mu, M_2) \pi^I(\mu|M_2) d\mu.$$

Γνωρίζουμε ότι όταν έχουμε ένα τυχαίο δείγμα μεγέθους n προερχόμενο από μία κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 , τότε η στατιστική συνάρτηση του δειγματικού μέσου του τυχαίου δείγματος θα είναι επίσης μία κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2/n .

Άρα έχουμε

$$\bar{x}|\mu \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \mu \sim N(\mu_0, 2\sigma^2).$$

Επομένως

$$\begin{aligned}
 \bar{x} &= \mu + \varepsilon_1, & \varepsilon_1 &\sim N\left(0, \frac{\sigma^2}{n}\right) \\
 \mu &= \mu_0 + \varepsilon_2, & \varepsilon_2 &\sim N(0, 2\sigma^2),
 \end{aligned}$$

με $\varepsilon_1, \varepsilon_2$ ανεξάρτητες τυχαίες μεταβλητές. Άρα

$$\bar{x} = \mu + \varepsilon_1 = \mu_0 + \varepsilon_1 + \varepsilon_2,$$

δηλαδή

$$\bar{X} \sim N\left(\mu_0, 2\sigma^2 + \frac{\sigma^2}{n}\right) \Rightarrow \bar{X} \sim N\left(\mu_0, \sigma^2 \frac{2n+1}{n}\right). \quad (E2.9)$$

Συνεπώς η περιθώρια πιθανοφάνεια $m_2^I(\bar{x})$ του μοντέλου M_2 είναι μία κανονική κατανομή με μέση τιμή μ_0 και διακύμανση $\sigma^2 (2n+1)/n$.

Άρα από τις σχέσεις (E2.9) και (E2.2) η σχέση (E2.1) μας δίνει ότι

$$BF_{21}^I = \frac{m_2^I(\bar{x})}{m_1^I(\bar{x})} = \frac{N\left(\mu_0, \sigma^2 \frac{2n+1}{n}\right)}{N\left(\mu_0, \frac{\sigma^2}{n}\right)}. \quad (E2.10)$$

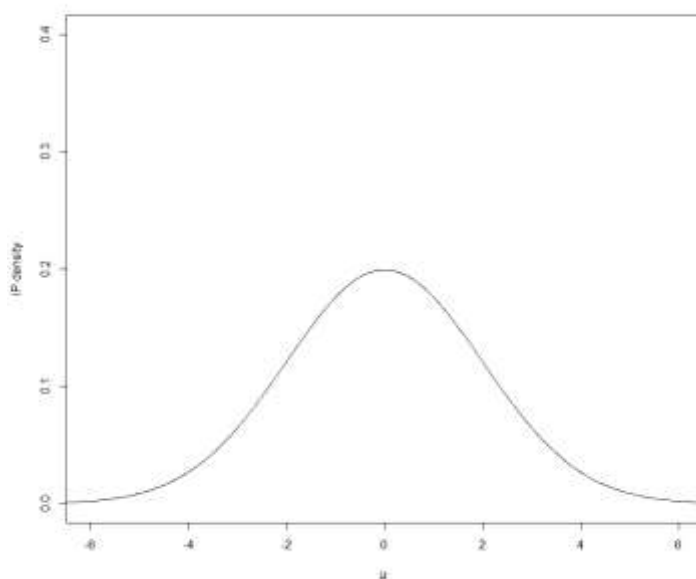
5.3.1 Παράδειγμα με προσομοιωμένα δεδομένα

Έστω ότι θέλουμε να συγκρίνουμε τα μοντέλα

$$M_1: X \sim N(0, 2^2)$$

$$M_2: X \sim N(\mu, 2^2), \pi^N(\mu|M_2) \propto 1.$$

Σύμφωνα με τα αποτελέσματα της **Ενότητας 5.3** η Ενδογενής εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 , η οποία περιγράφεται από την σχέση (E2.8), είναι μία κανονική κατανομή $N(0, 2^2)$ και απεικονίζεται στο **Διάγραμμα 5.5**.



Διάγραμμα 5.5 : Διάγραμμα ενδογενούς εκ των προτέρων κατανομής $\pi^I(\mu|M_2)$

Υποθέτουμε ότι για τις εκ των προτέρων πιθανότητες των μοντέλων M_1 και M_2 ισχύει

$$\pi(M_1) = \pi(M_2) = \frac{1}{2}$$

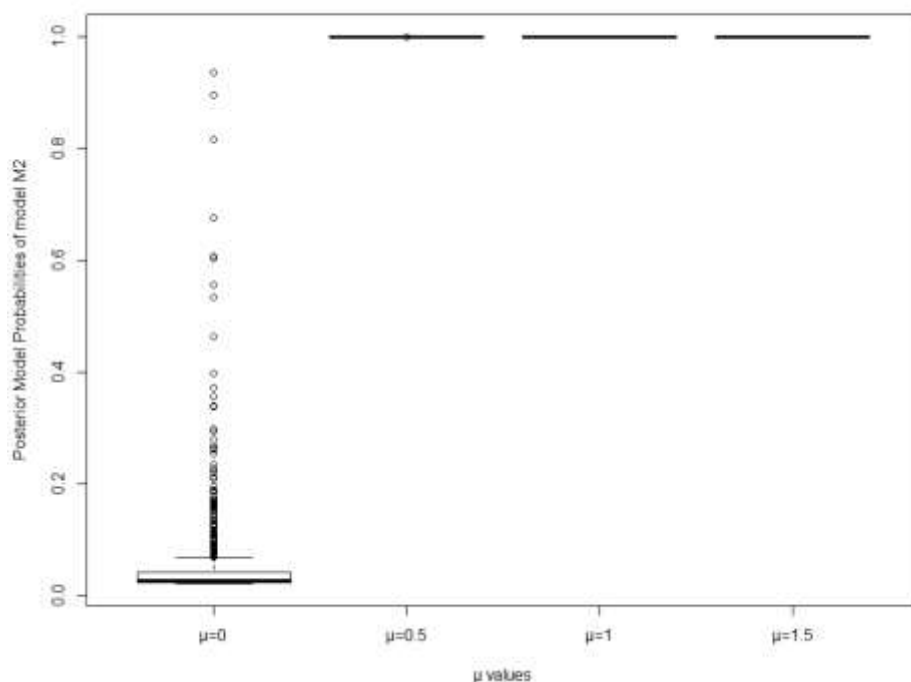
επομένως ο εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων υπέρ του μοντέλου M_2 θα ισούται με τον παράγοντα Μπεϋες υπέρ του μοντέλου M_2 . Επομένως έχουμε ότι

$$\begin{aligned} PO_{21} &= BF_{21} \\ \Rightarrow \frac{\pi(M_2|\mathbf{x})}{\pi(M_1|\mathbf{x})} &= BF_{21} \\ \Rightarrow \frac{\pi(M_2|\mathbf{x})}{1 - \pi(M_2|\mathbf{x})} &= BF_{21} \\ \Rightarrow \pi(M_2|\mathbf{x}) &= \frac{BF_{21}}{1 + BF_{21}}. \quad (E2.11) \end{aligned}$$

Εν συνεχεία μπορούμε να προσομοιώσουμε τυχαία δείγματα από την κανονική κατανομή για διάφορες τιμές των παραμέτρων μ, σ που θέλουμε να εξετάσουμε, προκειμένου να μελετήσουμε την συμπεριφορά της εκ των υστέρων πιθανότητας του μοντέλου M_2 . Θα πραγματοποιήσουμε 4 κύκλους προσομοιώσεων όπου θα παράγουμε τυχαία δείγματα προερχόμενα από την κανονική κατανομή σύμφωνα με τις τιμές των παραμέτρων μ και σ όπου έχουμε επιλέξει και στην συνέχεια θα προχωράμε στον υπολογισμό και την αποθήκευση των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 όπως αυτές προκύπτουν από την σχέση (E2.11).

1^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 1000 τυχαία δείγματα τα οποία θα αποτελούνται από 1000 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την κανονική κατανομή με διακύμανση $\sigma^2 = 2^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.6** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 1000 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 1000$.



Διάγραμμα 5.6: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 1000 τυχαία δείγματα μεγέθους $n = 1000$, $\sigma^2 = 2^2$

Στον **Πίνακα 5.4** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|\mathbf{x}) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

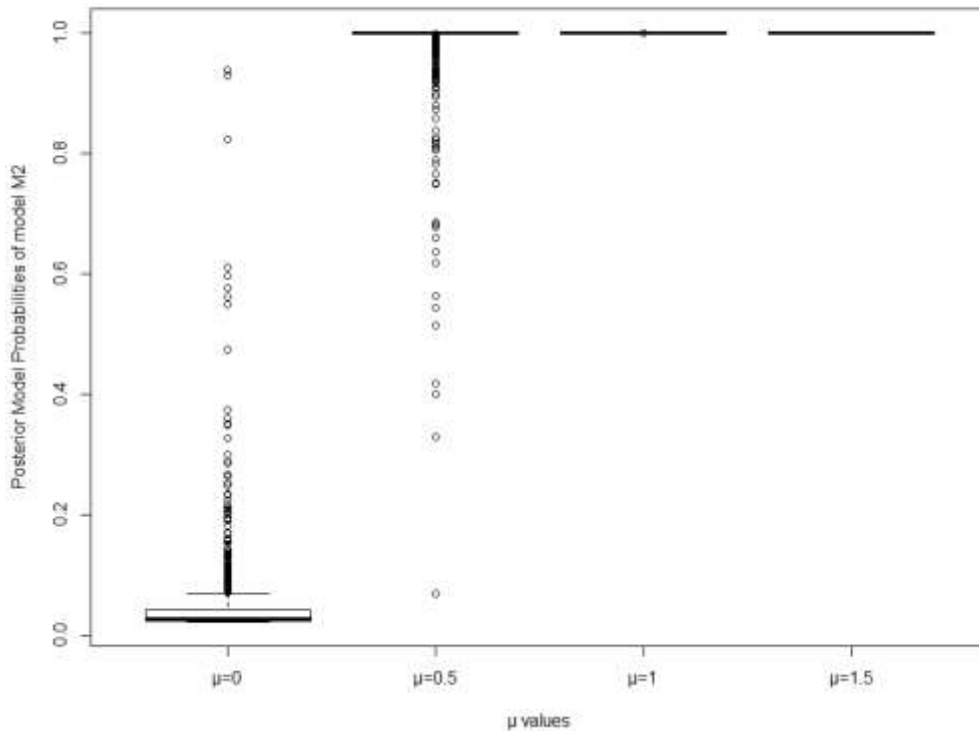
Πίνακας 5.4 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
0.8%	100%	100%	100%

Όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την κανονική κατανομή με παραμέτρους $\mu = 0$ και $\sigma^2 = 2^2$, παρατηρούμε ότι το μοντέλο M_1 που υποδεικνύει ότι τα δεδομένα προέρχονται από την $N(0,2^2)$, υποστηρίζεται στο 99.2% των περιπτώσεων. Η εμφάνιση έκτροπων τιμών και μάλιστα μεγαλύτερων του 0.5, οφείλεται στο γεγονός ότι μελετάμε προσομοιωμένα τυχαία δείγματα, συνεπώς θα υπάρξουν και μεμονωμένες περιπτώσεις εμφάνισης ακραίων τιμών. Για τις υπόλοιπες τιμές της παραμέτρου μ παρατηρούμε ότι υποστηρίζεται πλήρως το μοντέλο M_2 για κάθε τυχαίο δείγμα που παράγεται χωρίς την εμφάνιση έκτροπων τιμών.

2^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 1000 τυχαία δείγματα τα οποία θα αποτελούνται από 1000 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την κανονική κατανομή με διακύμανση $\sigma^2 = 3^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.7** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 1000 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 1000$.



Διάγραμμα 5.7: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 1000 τυχαία δείγματα μεγέθους $n = 1000$, $\sigma^2 = 3^2$

Στον **Πίνακα 5.5** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|\mathbf{x}) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

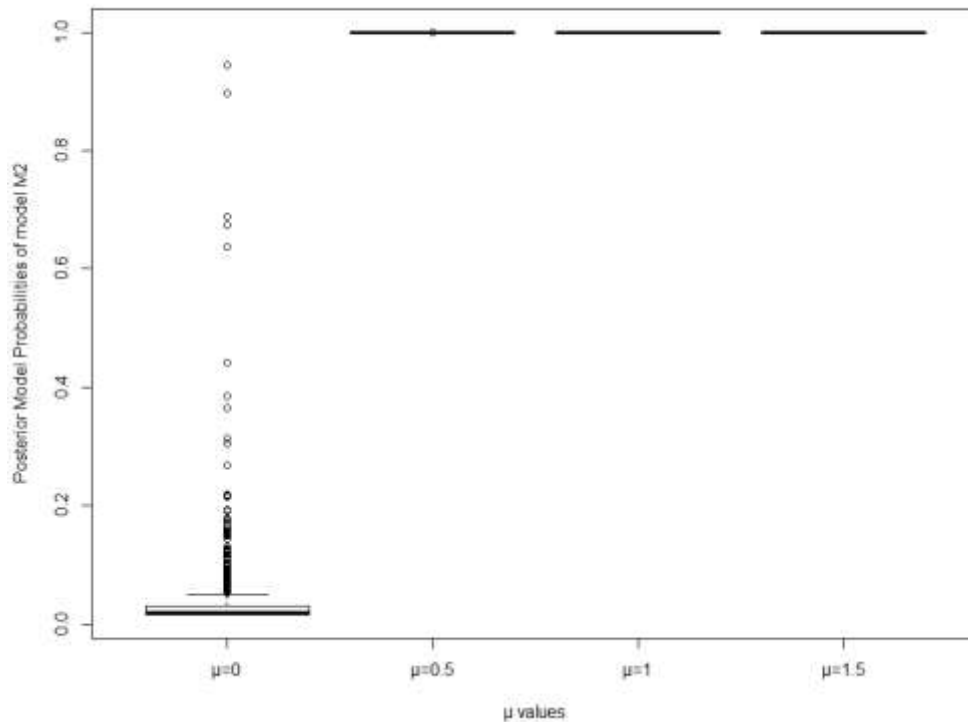
Πίνακας 5.5 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
0.8%	99.6%	100%	100%

Σε αυτόν τον κύκλο προσομοιώσεων αυξήσαμε την διακύμανση της κανονικής κατανομής η οποία παράγει τα τυχαία δείγματα των παρατηρήσεων x και συγκριτικά με τον πρώτο κύκλο προσομοιώσεων λαμβάνουμε τα ίδια αποτελέσματα ως προς την επιλογή του κατάλληλου μοντέλου για την περιγραφή των δεδομένων. Η μόνη διαφορά που παρουσιάζεται συγκριτικά με τον πρώτο κύκλο προσομοιώσεων, είναι η εμφάνιση περισσότερων έκτροπων τιμών λόγω της μεγαλύτερης διασποράς για τις δύο πρώτες τιμές της παραμέτρου μ που εξετάστηκαν. Είναι φυσιολογικό αυτό το φαινόμενο καθώς δημιουργούμε δείγματα με αριθμητικό μέσο διάφορο του 0.

3^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 1000 τυχαία δείγματα τα οποία θα αποτελούνται από 1000 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την κανονική κατανομή με διακύμανση $\sigma^2 = 2^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.8** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 1000 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 2000$.



Διάγραμμα 5.8: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 1000 τυχαία δείγματα μεγέθους $n = 2000$, $\sigma^2 = 2^2$

Στον **Πίνακα 5.6** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|x) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

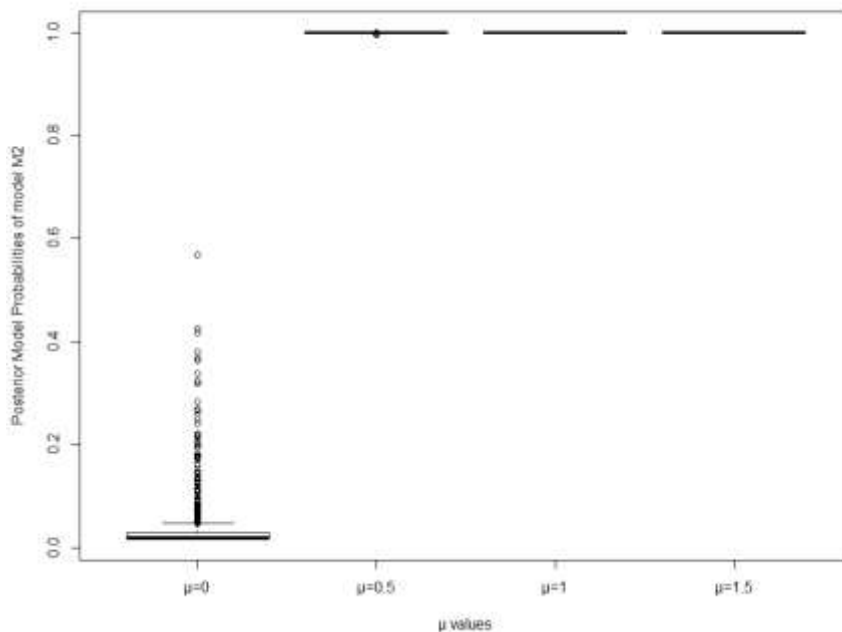
Πίνακας 5.6 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
0.5%	100%	100%	100%

Σε αυτόν τον κύκλο προσομοιώσεων διπλασιάσαμε το πλήθος των παρατηρήσεων των τυχαίων δειγμάτων που προέρχονται από την κανονική κατανομή για τις αντίστοιχες τιμές της παραμέτρου μ που εξετάζουμε με διακύμανση $\sigma^2 = 2^2$. Συγκριτικά με τον πρώτο κύκλο προσομοιώσεων λαμβάνουμε τα ίδια αποτελέσματα ως προς την επιλογή του κατάλληλου μοντέλου για την περιγραφή των δεδομένων, εμφανίζοντας λιγότερες ακραίες τιμές.

4^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 1000 τυχαία δείγματα τα οποία θα αποτελούνται από 1000 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την κανονική κατανομή με διακύμανση $\sigma^2 = 3^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.9** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 1000 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 2000$.



Διάγραμμα 5.9: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 1000 τυχαία δείγματα μεγέθους $n = 2000$, $\sigma^2 = 3^2$

Στον **Πίνακα 5.7** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|\mathbf{x}) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

Πίνακας 5.7 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
0.1%	100%	100%	100%

Σε αυτόν τον κύκλο παρατηρήσεων αφού διπλασιάσαμε το πλήθος των παρατηρήσεων από τις οποίες αποτελούνται τα τυχαία δείγματα, αυξήσαμε και την διακύμανση της κανονικής κατανομής η οποία γεννάει τα δεδομένα. Αντίστοιχα και σε αυτόν τον κύκλο παρατηρήσεων παρατηρούμε τα ίδια αποτελέσματα όπως και με τους προηγούμενους, όπου επίσης παρατηρούμε την αύξηση των έκτροπων τιμών συγκριτικά με τον τρίτο κύκλο προσομοιώσεων, όπως αντίστοιχα συνέβη και στον δεύτερο κύκλο προσομοιώσεων όπου αυξήθηκε η διακύμανση της κανονικής κατανομής που γεννάει τα δεδομένα.

5.4 Έλεγχος μέσης τιμής Κανονικής κατανομής με άγνωστη διακύμανση

Σε αυτήν την εφαρμογή θα υπολογίσουμε τις ενδογενείς εκ των προτέρων κατανομές καθώς και τις αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές, όταν εξετάζουμε δύο μοντέλα όπου ελέγχουμε την μέση τιμή μίας κανονικής κατανομής με άγνωστη διακύμανση, χρησιμοποιώντας μη-κατάλληλες εκ των προτέρων κατανομές.

Έστω τυχαίο δείγμα παρατηρήσεων $\mathbf{x} = (x_1, x_2, \dots, x_n)$ όπου για το οποίο ισχύει η ανταλλαξιμότητα, και τα υπό εξέταση μοντέλα

$$M_1: \mathbf{X} \sim N(0, \sigma^2) \text{ με } \pi^N(\sigma | M_1) = \frac{c_1}{\sigma} \propto \frac{1}{\sigma}.$$

$$M_2: \mathbf{X} \sim N(\mu, \sigma^2) \text{ με } \pi^N(\mu, \sigma | M_2) = \frac{c_2}{\sigma^2} \propto \frac{1}{\sigma^2}.$$

με c_1, c_2 οι αυθαίρετες σταθερές κανονικοποίησης.

5.4.1 Υπολογισμός ενδογενών εκ των προτέρων κατανομών

Είναι προφανές ότι τα δύο εξεταζόμενα μοντέλα είναι εμφωλευμένα, συνεπώς θα καταφύγουμε στο σύστημα ενδογενών εξισώσεων όπως διαμορφώνεται στην

περίπτωση σύγκρισης εμφωλευμένων μοντέλων, όπου περιγράφεται από τις εξισώσεις όπως παρουσιάστηκαν στην (K4.11)

$$\pi^I(\sigma|M_1) = \pi^N(\sigma|M_1), \quad \pi^I(\mu, \sigma|M_2) = \pi^N(\mu, \sigma|M_2)B_2^*(\mu, \sigma).$$

Σκοπός μας είναι ο υπολογισμός του ενδογενή παράγοντα Μπεϋες

$$BF_{21}^I = \frac{m_2^I(\mathbf{x})}{m_1^I(\mathbf{x})}. \quad (E4.1)$$

Για τον υπολογισμό της ενδογενούς εκ των προτέρων κατανομής του μοντέλου M_2 έχουμε ότι

$$\begin{aligned} \pi^I(\mu, \sigma|M_2) &= \pi^N(\mu, \sigma|M_2)B_2^*(\mu, \sigma) \\ &= \frac{1}{\sigma^2} E^{M_2}[BF_{12}^N(\mathbf{x}(l))] \quad (E4.2) \end{aligned}$$

όπου

$$BF_{12}^N(\mathbf{x}(l)) = \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))}. \quad (E4.3)$$

Επειδή το πρόβλημα το οποίο εξετάζουμε περιέχει δύο άγνωστες μονοδιάστατες παραμέτρους το ελάχιστο δείγμα εκπαίδευσης θα αποτελείται από δύο παρατηρήσεις.

Συνεπώς ορίζουμε το ελάχιστο δείγμα εκπαίδευσης ως

$$\mathbf{x}(l) = \{x_1, x_2\}.$$

Επομένως για την περιθώρια πιθανοφάνεια του μοντέλου M_1 κάνοντας χρήση της μη-κατάλληλης εκ των προτέρων κατανομής $\pi^N(\sigma|M_1)$, έχουμε ότι

$$\begin{aligned} m_1^N(\mathbf{x}(l)) &= m_1^N(x_1, x_2) = \int_0^\infty f(x_1, x_2|\sigma, M_1)\pi^N(\sigma|M_1)d\sigma \\ &= \int_0^\infty \frac{1}{\sigma^2 2\pi} \exp\left[-\frac{1}{2\sigma^2}x_1^2\right] \exp\left[-\frac{1}{2\sigma^2}x_2^2\right] \frac{1}{\sigma} d\sigma \\ &= \frac{1}{2\pi} \int_0^\infty \frac{1}{\sigma^3} \exp\left[-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)\right] d\sigma \end{aligned}$$

Θέτοντας $\sigma = \sqrt{k}$ έχουμε ότι $d\sigma = 1/2\sqrt{k} dk$ και $0 < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα μας δίνει ότι

$$m_1^N(\mathbf{x}(l)) = \frac{1}{2\pi} \int_0^\infty \frac{1}{k^{3/2}} \exp\left[-\frac{(x_1^2 + x_2^2)/2}{k}\right] \frac{1}{2\sqrt{k}} dk$$

$$= \frac{1}{4\pi} \int_0^{\infty} k^{-1-1} \exp\left[-\frac{(x_1^2 + x_2^2)/2}{k}\right] dk.$$

Παρατηρούμε ότι ο πυρήνας του παραπάνω ολοκληρώματος αντιπροσωπεύει μία αντίστροφη κατανομή γάμμα με παραμέτρους $\alpha = 1$ και $\beta = (x_1^2 + x_2^2)/2$, επομένως παίρνουμε ότι

$$\begin{aligned} m_1^N(\mathbf{x}(l)) &= \frac{1}{4\pi} \frac{\Gamma(1)}{\frac{x_1^2 + x_2^2}{2}} \\ &= \frac{1}{2\pi} \frac{1}{x_1^2 + x_2^2}. \end{aligned} \quad (E4.4)$$

Για την περιθώρια πιθανοφάνεια του μοντέλου M_2 κάνοντας χρήση της μη κατάλληλης εκ των προτέρων κατανομής $\pi^N(\mu, \sigma|M_2)$, έχουμε ότι

$$\begin{aligned} m_2^N(\mathbf{x}(l)) &= m_2^N(x_1, x_2) = \int_0^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2|\mu, \sigma, M_2) \pi^N(\mu, \sigma|M_2) d\mu d\sigma \\ &= \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma^2 2\pi} \exp\left[-\frac{1}{2\sigma^2}(x_1 - \mu)^2\right] \exp\left[-\frac{1}{2\sigma^2}(x_2 - \mu)^2\right] \frac{1}{\sigma^2} d\mu d\sigma \\ &= \frac{1}{2\pi} \int_0^{\infty} \frac{1}{\sigma^4} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}[(x_1 - \mu)^2 + (x_2 - \mu)^2]\right] d\mu d\sigma. \end{aligned} \quad (E4.5)$$

Στην συνέχεια για να προχωρήσουμε στον υπολογισμό του ολοκληρώματος ως προς μ θα τροποποιήσουμε την ποσότητα $(x_1 - \mu)^2 + (x_2 - \mu)^2$.

Συνεπώς έχουμε ότι

$$\begin{aligned} (x_1 - \mu)^2 + (x_2 - \mu)^2 &= (\mu - x_1)^2 + (\mu - x_2)^2 \\ &= \mu^2 - 2\mu x_1 + x_1^2 + \mu^2 - 2\mu x_2 + x_2^2 \\ &= 2\mu^2 - 2(x_1 + x_2)\mu + x_1^2 + x_2^2 \\ &= 2\left[\mu^2 + 2\frac{x_1 + x_2}{2}\mu\right] + x_1^2 + x_2^2 \\ &= 2\left[\mu^2 + 2\frac{x_1 + x_2}{2}\mu \pm \left(\frac{x_1 + x_2}{2}\right)^2\right] + x_1^2 + x_2^2 \end{aligned}$$

$$\begin{aligned}
&= 2 \left[\mu^2 + 2 \frac{x_1 + x_2}{2} \mu + \left(\frac{x_1 + x_2}{2} \right)^2 \right] + x_1^2 + x_2^2 \\
&\quad - 2 \left(\frac{x_1 + x_2}{2} \right)^2 \\
&= 2 \left(\mu - \frac{x_1 + x_2}{2} \right)^2 + \frac{2x_1^2 + 2x_2^2 - x_1^2 - 2x_1x_2 - x_2^2}{2} \\
&= 2 \left(\mu - \frac{x_1 + x_2}{2} \right)^2 + \frac{(x_1 - x_2)^2}{2}.
\end{aligned}$$

Επομένως το ολοκλήρωμα που περιγράφεται από την σχέση (E4.5) διαμορφώνεται ως εξής,

$$\begin{aligned}
m_2^N(\mathbf{x}(l)) &= \frac{1}{2\pi} \int_0^\infty \frac{1}{\sigma^4} \int_{-\infty}^\infty \exp \left[-\frac{1}{2\sigma^2} \left[2 \left(\mu - \frac{x_1 + x_2}{2} \right)^2 + \frac{(x_1 - x_2)^2}{2} \right] \right] d\mu d\sigma \\
&= \frac{1}{2\pi} \int_0^\infty \frac{1}{\sigma^4} \exp \left[-\frac{1}{4\sigma^2} (x_1 - x_2)^2 \right] \int_{-\infty}^\infty \exp \left[-\frac{1}{2 \left(\frac{\sigma}{\sqrt{2}} \right)^2} \left[2 \left(\mu - \frac{x_1 + x_2}{2} \right)^2 \right] \right] d\mu d\sigma.
\end{aligned}$$

Παρατηρούμε ότι στο παραπάνω ολοκλήρωμα ως προς μ εμφανίζεται ο πυρήνας μίας κανονικής κατανομής με μέση τιμή $(x_1 + x_2)/2$ και διακύμανση $(\sigma/\sqrt{2})^2$. Επομένως η σχέση (E4.5) διαμορφώνεται ως εξής,

$$\begin{aligned}
m_2^N(\mathbf{x}(l)) &= \frac{1}{2\pi} \int_0^\infty \frac{1}{\sigma^4} \exp \left[-\frac{1}{4\sigma^2} (x_1 - x_2)^2 \right] \frac{\sigma}{\sqrt{2}} \sqrt{2\pi} d\sigma \\
&= \frac{\sqrt{\pi}}{2\pi} \int_0^\infty \frac{1}{\sigma^3} \exp \left[-\frac{1}{4\sigma^2} (x_1 - x_2)^2 \right] d\sigma.
\end{aligned}$$

Θέτοντας $\sigma = \sqrt{k}$ έχουμε ότι $d\sigma = 1/2\sqrt{k} dk$ και $0 < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα μας δίνει ότι

$$\begin{aligned}
m_2^N(\mathbf{x}(l)) &= \frac{1}{2\sqrt{\pi}} \int_0^\infty \frac{1}{k^{\frac{3}{2}}} \exp \left[-\frac{(x_1 - x_2)^2/4}{k} \right] \frac{1}{2\sqrt{k}} dk \\
&= \frac{1}{4\sqrt{\pi}} \int_0^\infty k^{-1-1} \exp \left[-\frac{(x_1 - x_2)^2/4}{k} \right] dk.
\end{aligned}$$

Παρατηρούμε ότι ο πυρήνας του παραπάνω ολοκληρώματος αντιπροσωπεύει μία αντίστροφη κατανομή γάμμα με παραμέτρους $\alpha = 1$ και $\beta = (x_1 - x_2)^2/4$, επομένως παίρνουμε ότι

$$\begin{aligned} m_2^N(\mathbf{x}(l)) &= \frac{1}{4\sqrt{\pi}} \frac{\Gamma(1)}{\frac{(x_1 - x_2)^2}{4}} \\ &= \frac{1}{\sqrt{\pi}} \frac{1}{(x_1 - x_2)^2} \cdot (E4.6) \end{aligned}$$

Επομένως από τις σχέσεις (E4.4) και (E4.6) η σχέση (E4.3) μας δίνει ότι

$$\begin{aligned} BF_{12}^N(\mathbf{x}(l)) &= \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} \\ &= \frac{\frac{1}{2\pi} \frac{1}{x_1^2 + x_2^2}}{\frac{1}{\sqrt{\pi}} \frac{1}{(x_1 - x_2)^2}} \\ &= \frac{\sqrt{\pi} (x_1 - x_2)^2}{2\pi (x_1^2 + x_2^2)} \cdot (E4.7) \end{aligned}$$

Για να προχωρήσουμε στον υπολογισμό της ενδογενούς εκ των προτέρων κατανομής $\pi^l(\mu, \sigma | M_2)$ θα τροποποιήσουμε τον αριθμητή και τον παρονομαστή της σχέσης (E4.7). Επομένως για τον αριθμητή της σχέσης (E4.7) έχουμε ότι

$$\begin{aligned} \frac{(x_1 - x_2)^2}{2} &= \frac{(x_1 - x_2)^2}{4} + \frac{(x_1 - x_2)^2}{4} \\ &= \left(\frac{2x_1 - x_1 - x_2}{2}\right)^2 + \left(\frac{2x_2 - x_2 - x_1}{2}\right)^2 \\ &= \left(x_1 - \frac{x_1 + x_2}{2}\right)^2 + \left(x_2 - \frac{x_1 + x_2}{2}\right)^2 \\ &= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 \\ &= S_{\bar{x}}^2. (E4.8) \end{aligned}$$

Αντίστοιχα για τον παρονομαστή της σχέσης (E4.7) έχουμε ότι

$$\begin{aligned} x_1^2 + x_2^2 &= x_1^2 + x_2^2 \pm 2x_1\bar{x} \pm 2x_2\bar{x} \pm 2\bar{x}^2 \pm 2\bar{x}^2 \\ &= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2\bar{x}^2 + 2\bar{x}(x_1 - \bar{x}) \\ &\quad + 2\bar{x}(x_2 - \bar{x}) \\ &= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2\bar{x}[x_1 - \bar{x} + x_2 - \bar{x} + \bar{x}] \end{aligned}$$

$$\begin{aligned}
&= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2\bar{x} \left[\frac{2x_1 + 2x_2 - x_1 - x_2}{2} \right] \\
&= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2\bar{x} \left[\frac{x_1 + x_2}{2} \right] \\
&= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2\bar{x}^2 \\
&= S_x^2 + 2\bar{x}^2. \quad (E4.9)
\end{aligned}$$

Επομένως από τις σχέσεις (E4.8) και (E4.9) η σχέση (E4.7) μας δίνει

$$\begin{aligned}
BF_{12}^N(\mathbf{x}(l)) &= \frac{\sqrt{\pi} (x_1 - x_2)^2}{2\pi x_1^2 + x_2^2} \\
&= \frac{\sqrt{\pi} (x_1 - x_2)^2 / 2}{\pi x_1^2 + x_2^2} \\
&= \frac{\sqrt{\pi} S_x^2}{\pi S_x^2 + 2\bar{x}^2} \quad (E4.10)
\end{aligned}$$

Συνεπώς χρησιμοποιώντας την σχέση (E4.10) η ενδογενής εκ των προτέρων κατανομή $\pi^l(\mu, \sigma | M_2)$ γίνεται

$$\begin{aligned}
\pi_2^l(\mu, \sigma) &= \frac{1}{\sigma^2} E^{M_2} [BF_{12}^N(\mathbf{x}(l))] \\
&= \frac{1}{\sigma^2} E_{x_1, x_2} \left[\frac{\sqrt{\pi} S_x^2}{\pi S_x^2 + 2\bar{x}^2} \right] \\
&= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} E_{x_1, x_2} \left[\frac{S_x^2 / \sigma^2}{S_x^2 / \sigma^2 + 2\bar{x}^2 / \sigma^2} \right]. \quad (E4.11)
\end{aligned}$$

Είναι γνωστό ότι όταν έχουμε X_1, X_2 ανεξάρτητες και ισόνομες τυχαίες μεταβλητές οι οποίες ακολουθούν Κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 , τότε ισχύει ότι η ποσότητα

$$\tilde{S} = \frac{S_x^2}{\sigma^2}$$

θα ακολουθεί X_1^2 κατανομή. Αντίστοιχα η ποσότητα

$$Z = \frac{2\bar{X}^2}{\sigma^2} = \frac{\bar{X}^2 + \bar{X}^2}{\sigma^2}$$

ακολουθεί την *non-Central* $X_2^2(\lambda)$ κατανομή με παράμετρο θέσης $\lambda = 2\mu^2 / \sigma^2$ και 1 βαθμό ελευθερίας.

Επομένως η σχέση (E4.11) μπορεί να γραφεί ως

$$\begin{aligned}\pi^I(\mu, \sigma|M_2) &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} E_{x_1, x_2} \left[\frac{S_x^2/\sigma^2}{S_x^2/\sigma^2 + 2\bar{x}^2/\sigma^2} \right] \\ &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} E_{\tilde{S}, Z} \left[\frac{\tilde{S}}{\tilde{S} + Z} \right].\end{aligned}$$

Στην συνέχεια από το παράρτημα της εργασίας των Berger & Perrichi (1994), έχουμε ότι

$$\begin{aligned}\pi^I(\mu, \sigma|M_2) &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} E_{\tilde{S}, Z} \left[\frac{\tilde{S}}{\tilde{S} + Z} \right] \\ &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} e^{-\frac{\lambda}{2}} E_{\tilde{S}, T} \left[\frac{\tilde{S}}{\tilde{S} + T} \right] \quad (E4.12)\end{aligned}$$

όπου τώρα η τυχαία μεταβλητή T ακολουθεί την X_{1+2j}^2 κατανομή. Ο σκοπός της χρήσης αυτής της ιδιότητας είναι να μεταβούμε από την *non-Central* X^2 κατανομή σε μία απλή X^2 κατανομή.

Επίσης από το ίδιο παράρτημα της εργασίας των Berger & Perrichi (1994) έχουμε ότι, εάν μία τυχαία μεταβλητή X η οποία ακολουθεί μία X_c^2 κατανομή και μία τυχαία μεταβλητή Y η οποία ακολουθεί X_d^2 κατανομή και ταυτόχρονα ισχύει

$$\frac{c+d}{2} + a > \beta \text{ με } \alpha, \beta > 0$$

τότε παίρνουμε ότι

$$E \left[\frac{X^\alpha}{(X+Y)^\beta} \right] = \frac{2^\alpha \Gamma\left(\frac{c+d}{2} + a - \beta\right) \Gamma\left(\frac{c}{2} + a\right)}{2^\beta \Gamma\left(\frac{c+d}{2} + a\right) \Gamma\left(\frac{c}{2}\right)}. \quad (E4.13)$$

Άρα για $\alpha = \beta = 1, c = 1, d = 1 + 2j$, η σχέση (E4.12) μέσω της σχέσης (E4.13) διαμορφώνεται ως

$$\begin{aligned}\pi^I(\mu, \sigma|M_2) &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} e^{-\frac{\lambda}{2}} \frac{2^1 \Gamma\left(\frac{1+1+2j}{2} + 1 - 1\right) \Gamma\left(\frac{1}{2} + 1\right)}{2^1 \Gamma\left(\frac{1+1+2j}{2} + 1\right) \Gamma\left(\frac{1}{2}\right)} \\ &= \frac{1}{\sigma^2} \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} e^{-\frac{\lambda}{2}} \frac{\Gamma(1+1j) \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{\Gamma(2+j) \Gamma\left(\frac{1}{2}\right)}\end{aligned}$$

$$= \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} e^{-\frac{\lambda}{2}} M(1,2, \lambda/2), \quad (E4.14)$$

όπου η ποσότητα $M(1,2, \lambda/2)$, είναι η *υπεργεωμετρική συνάρτηση του Kummer (Kummer hypergeometric function)* όπως αυτή ορίζεται στην *Ενότητα 15* του βιβλίου των Abramowitz & Stegun (1964).

Επίσης από το ίδιο κεφάλαιο του βιβλίου των Abramowitz & Stegun (1964) χρησιμοποιούμε την ιδιότητα της υπεργεωμετρική συνάρτησης του Kummer

$$e^{-z} M(a, b, z) = M(b - a, b, -z)$$

επομένως

$$\pi^I(\mu, \sigma | M_2) = \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} M(1,2, -\lambda/2).$$

Επίσης ισχύει ότι

$$M(a, b, z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1} dt. \quad (E4.15)$$

Επομένως χρησιμοποιώντας την σχέση (E4.15) και για $a = 1, b = 2, z = -\lambda/2$ η σχέση (E4.14) μας δίνει ότι

$$\begin{aligned} \pi^I(\mu, \sigma | M_2) &= \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \int_0^1 e^{zt} t^0 (1-t)^0 dt \\ &= \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} \int_0^1 e^{-\frac{\lambda}{2}t} dt \\ &= \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} \int_0^1 e^{-\frac{2\mu^2}{2\sigma^2}t} dt \\ &= -\frac{1}{\sigma} \frac{1}{2} \frac{1}{\sqrt{\pi}} \frac{2\sigma^2}{2\mu^2} \left[e^{-\frac{\mu^2}{\sigma^2}t} \right]_0^1 \\ &= \frac{1}{2\sqrt{\pi}\mu^2} \left[e^{-\frac{\mu^2}{\sigma^2}} - 1 \right] \\ &= \frac{1 - e^{-\frac{\mu^2}{\sigma^2}}}{2\sqrt{\pi}\mu^2}. \quad (E4.16) \end{aligned}$$

Η Ενδογενής εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 μπορεί να υπολογιστεί και ως

$$\begin{aligned}
 \pi^I(\mu, \sigma | M_2) &= \frac{1}{\sigma^2} \frac{1}{2} \frac{1}{\sqrt{\pi}} \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \int_0^1 e^{zt} t^0 (1-t)^0 dt \\
 &= \frac{1}{\sigma^2} \frac{1}{2} \int_0^1 \frac{1}{\sqrt{2\pi\sigma}} t^{1/2} \sqrt{2} e^{-\frac{2\mu^2}{2\sigma^2}t} t^0 (1-t)^0 t^{-1/2} dt \\
 &= \frac{1}{\sigma^2} \frac{1}{2} \int_0^1 N\left(\mu \middle| 0, \frac{2\sigma^2}{t}\right) t^{-1/2} (1-t)^{1-1} dt \\
 &= \frac{1}{\sigma^2} \frac{1}{2} \int_0^1 N\left(\mu \middle| 0, \frac{2\sigma^2}{t}\right) \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(1)\Gamma\left(\frac{1}{2}\right)} t^{-1/2} (1-t)^{1-1} dt \\
 &= \frac{1}{\sigma} \int_0^1 N\left(\mu \middle| 0, \frac{2\sigma^2}{t}\right) \text{Beta}\left(t \middle| \frac{1}{2}, 1\right) dt.
 \end{aligned}$$

Επομένως παρατηρούμε ότι η ενδογενής εκ των προτέρων κατανομή των παραμέτρων του M_2 μοντέλου μπορεί να γραφεί στην μορφή

$$\pi^I(\mu, \sigma | M_2) = \pi^*(\sigma | M_2) \pi^*(\mu | \sigma, M_2)$$

όπου

$$\pi^*(\sigma | M_2) = \frac{1}{\sigma}$$

και

$$\pi_2^*(\mu | \sigma, M_2) = \int_0^1 N\left(\mu \middle| 0, \frac{2\sigma^2}{t}\right) \text{Beta}\left(t \middle| \frac{1}{2}, 1\right) dt.$$

Παρατηρούμε ότι η κατανομή $\pi^*(\mu | \sigma, M_2)$ αποτελεί μία μίξη Κανονικής με Βήτα κατανομής και προκύπτει ότι

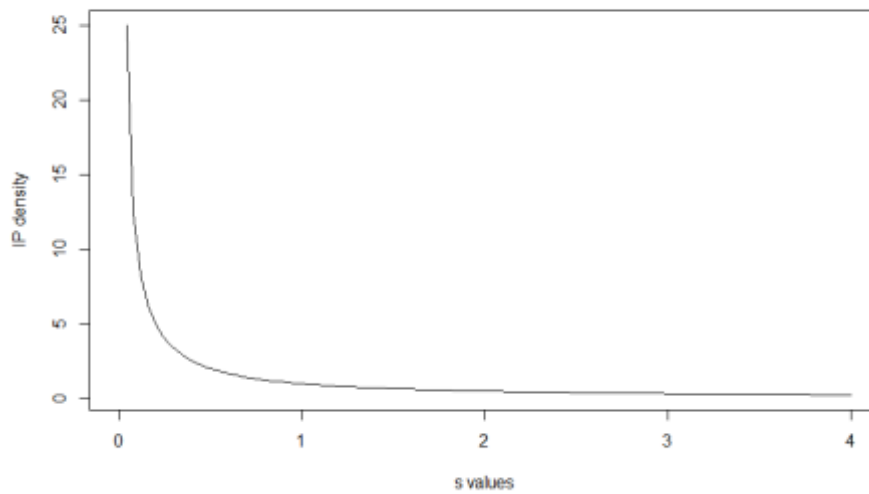
$$\pi^*(\mu | \sigma, M_2) = \int_0^1 N\left(\mu \middle| 0, \frac{2\sigma^2}{t}\right) \text{Beta}\left(t \middle| \frac{1}{2}, 1\right) dt = \frac{1 - e^{-\mu^2/\sigma^2}}{2\sqrt{\pi}\mu^2/\sigma}.$$

Επομένως οι ενδογενείς εκ των προτέρων κατανομές των παραμέτρων των μοντέλων M_1 και M_2 αντίστοιχα, χωρίς την αυθαίρετη σταθερά c_1 καθώς θα απλοποιηθεί κατά

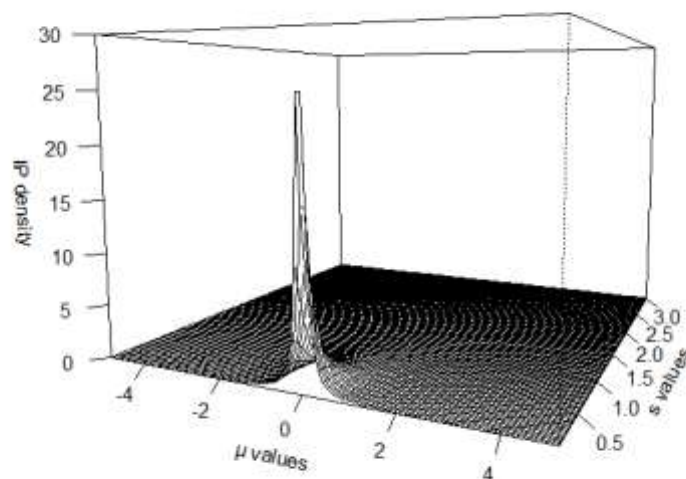
την διαδικασία υπολογισμού του ενδογενή παράγοντα Μπεϊές, δίνονται από τις σχέσεις

$$\pi^I(\sigma|M_1) = \frac{1}{\sigma}, \quad \pi^I(\mu, \sigma|M_2) = \frac{1 - e^{-\mu^2/\sigma^2}}{2\sqrt{\pi}\mu^2}.$$

Στο **Διάγραμμα 5.10** απεικονίζεται η ενδογενής εκ των προτέρων κατανομή της παραμέτρου του μοντέλου M_1 και αντίστοιχα στο **Διάγραμμα 5.11** απεικονίζεται η ενδογενής εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 .



Διάγραμμα 5.10 : Διάγραμμα ενδογενούς εκ των προτέρων κατανομής $\pi^I(\sigma|M_1)$



Διάγραμμα 5.11 : Διάγραμμα ενδογενούς εκ των προτέρων κατανομής $\pi^I(\mu, \sigma|M_2)$

Στην συνέχεια θα υπολογίσουμε τον παράγοντα Μπεϋες υπέρ του μοντέλου M_2 υπολογίζοντας τις αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές των μοντέλων M_1 και M_2 αντίστοιχα.

5.4.2 Υπολογισμός αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών

Σκοπός μας σε αυτήν την ενότητα είναι ο υπολογισμός του παράγοντα Μπεϋες υπέρ του μοντέλου M_2 , αφού πρώτα υπολογίσουμε τις αναμενόμενες μεταγενέστερες εκ των προτέρων κατανομές των μοντέλων αυτών. Εφόσον το πρόβλημα το οποίο εξετάζουμε αφορά εμφωλευμένα μοντέλα, καθώς το μοντέλο M_1 είναι εμφωλευμένο στο μοντέλο M_2 θα ακολουθήσουμε την διαδικασία την οποία περιγράψαμε στην **Ενότητα 4.4.1.1**, χρησιμοποιώντας ως μοντέλο βάσης το μοντέλο M_1 . Για την διενέργεια των υπολογισμών των αναμενόμενων μεταγενέστερων εκ των προτέρων κατανομών θα χρησιμοποιηθεί ένα φανταστικό δείγμα εκπαίδευσης, το οποίο θα αποτελείται από τις τιμές δύο ανεξάρτητων και ισόνομων τυχαίων μεταβλητών και το ορίζουμε ως

$$\mathbf{x}^* = \{x_1^*, x_2^*\}.$$

Επιπλέον θεωρούμε ότι οι τυχαίες μεταβλητές \mathbf{X}^* και \mathbf{X} είναι ανεξάρτητες.

Επομένως σκοπός μας είναι ο υπολογισμός του παράγοντα Μπεϋες ο οποίος θα περιγράφεται από

$$BF_{21}^{EPP}(\mathbf{x}) = \frac{m_1^N(\mathbf{x})}{m_2^{EPP}(\mathbf{x})}. \quad (E4.17)$$

Συνεπώς έχουμε ότι

$$\begin{aligned} m_1^N(\mathbf{x}) &= \int_0^{\infty} f(\mathbf{x}|\sigma, M_1) \pi^N(\sigma|M_1) d\sigma \\ &= \int_0^{\infty} \frac{1}{\sigma^n \sqrt{2\pi}^n} \exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^n x_i^2\right] \frac{1}{\sigma} d\sigma \\ &= \sqrt{2\pi}^{-n} \int_0^{\infty} \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right] d\sigma \end{aligned}$$

Θέτοντας $\sigma = \sqrt{k}$ έχουμε ότι $d\sigma = 1/2\sqrt{k} dk$ και $0 < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα μας δίνει ότι

$$m_1^N(\mathbf{x}) = \sqrt{2\pi}^{-n} \int_0^{\infty} k^{-\frac{n}{2}-\frac{1}{2}} \exp\left[-\frac{\sum_{i=1}^n x_i^2/2}{k}\right] \frac{1}{2\sqrt{k}} dk$$

$$= 2^{-1}\sqrt{2\pi}^{-n} \int_0^{\infty} k^{-\frac{n}{2}-1} \exp\left[-\frac{\sum_{i=1}^n x_i^2/2}{k}\right] dk.$$

Παρατηρούμε ότι ο πυρήνας του παραπάνω ολοκληρώματος αντιπροσωπεύει μία αντίστροφη κατανομή γάμμα με παραμέτρους $\alpha = n/2$ και $\beta = \sum_{i=1}^n x_i^2/2$, επομένως παίρνουμε ότι

$$m_1^N(\mathbf{x}) = 2^{-1}\sqrt{2\pi}^{-n} \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{\sum_{i=1}^n x_i^2}{2}\right)^{n/2}}. \quad (E4.18)$$

Στην συνέχεια θα προχωρήσουμε στον υπολογισμό της περιθώριας πιθανοφάνειας χρησιμοποιώντας την αναμενόμενη μεταγενέστερη εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 . Συνεπώς έχουμε ότι

$$m_2^{EPP}(\mathbf{x}) = \int_{\Theta_2} f(\mathbf{x}|\boldsymbol{\theta}_2, M_2) \pi^{EPP}(\boldsymbol{\theta}_2|M_2) d\boldsymbol{\theta}_2,$$

όπου

$$\boldsymbol{\theta}_2 = (\mu, \sigma_2) \in \Theta_2 = \mathbb{R} \times (0, \infty) \equiv \Theta_2.$$

Συνεπώς

$$\begin{aligned} m_2^{EPP}(\mathbf{x}) &= \int_{-\infty}^{\infty} \int_{\Theta_2} f(\mathbf{x}|\boldsymbol{\theta}_2, M_2) \pi_2^N(\boldsymbol{\theta}_2|\mathbf{x}^*) m_1^N(\mathbf{x}^*) d\boldsymbol{\theta}_2 d\mathbf{x}^* \\ &= \int_{-\infty}^{\infty} \int_{\Theta_2} f(\mathbf{x}|\boldsymbol{\theta}_2, M_2) \frac{f(\mathbf{x}^*|\boldsymbol{\theta}_2, M_2) \pi^N(\boldsymbol{\theta}_2|M_2)}{m_2^N(\mathbf{x}^*)} m_1^N(\mathbf{x}^*) d\boldsymbol{\theta}_2 d\mathbf{x}^* \\ &= \int_{-\infty}^{\infty} \int_{\Theta_2} f(\mathbf{x}, \mathbf{x}^*|\boldsymbol{\theta}_2, M_2) \pi^N(\boldsymbol{\theta}_2|M_2) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} d\boldsymbol{\theta}_2 d\mathbf{x}^* \\ &= \int_{-\infty}^{\infty} \int_{\Theta_2} \pi_2^N(\boldsymbol{\theta}_2|\mathbf{x}, \mathbf{x}^*) m_2^N(\mathbf{x}, \mathbf{x}^*) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} d\boldsymbol{\theta}_2 d\mathbf{x}^* \\ &= \int_{-\infty}^{\infty} m_2^N(\mathbf{x}, \mathbf{x}^*) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} \int_{\Theta_2} \pi_2^N(\boldsymbol{\theta}_2|\mathbf{x}, \mathbf{x}^*) d\boldsymbol{\theta}_2 d\mathbf{x}^* \\ &= \int_{-\infty}^{\infty} m_2^N(\mathbf{x}, \mathbf{x}^*) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} d\mathbf{x}^* \end{aligned}$$

$$= m_2^N(\mathbf{x}) \int_{-\infty}^{\infty} m_2^N(\mathbf{x}^*|\mathbf{x}) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} d\mathbf{x}^*. \quad (E4.19)$$

Επομένως για τον υπολογισμό της περιθώριας κατανομής του μοντέλου M_2 χρησιμοποιώντας την αναμενόμενη μεταγενέστερη εκ των προτέρων κατανομή των παραμέτρων του μοντέλου M_2 , χρειάζεται να προχωρήσουμε στον υπολογισμό των περιθώριων κατανομών της σχέσης (E4.19). Χρησιμοποιώντας τα αποτελέσματα της **Ενότητας 5.3.1.**, από τις σχέσεις (E4.4) και (E4.6) παίρνουμε ότι

$$m_1^N(\mathbf{x}^*) = \frac{1}{2\pi} \frac{1}{x_1^{*2} + x_2^{*2}} \quad (E4.20)$$

και

$$m_2^N(\mathbf{x}^*) = \frac{1}{\sqrt{\pi}} \frac{1}{(x_1^* - x_2^*)^2}. \quad (E4.21)$$

Στην συνέχεια προχωράμε στον υπολογισμό της περιθώριας κατανομής των δεδομένων \mathbf{x} κάτω από το μοντέλο M_2 κάνοντας χρήση της αρχικής μη κατάλληλης εκ των προτέρων κατανομής των παραμέτρων του μοντέλου M_2 , δηλαδή

$$\begin{aligned} m_2^N(\mathbf{x}) &= \int_0^{\infty} \int_{-\infty}^{\infty} f(\mathbf{x}|\mu, \sigma, M_2) \pi^N(\mu, \sigma|M_2) d\mu d\sigma \\ &= \int_0^{\infty} \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma} \sum_{i=1}^n (x_i - \mu)^2\right] \frac{1}{\sigma^2} d\mu d\sigma \\ &= (2\pi)^{-n/2} \int_0^{\infty} \int_{-\infty}^{\infty} (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma} \sum_{i=1}^n (x_i - \bar{x})^2\right] \exp\left[-\frac{1}{2\sigma} \sum_{i=1}^n (\bar{x} - \mu)^2\right] \frac{1}{\sigma^2} d\mu d\sigma \\ &= (2\pi)^{-n/2} \int_0^{\infty} (\sigma^2)^{-\frac{n+2}{2}} \exp\left[-\frac{1}{2\sigma} \sum_{i=1}^n (x_i - \bar{x})^2\right] \int_{-\infty}^{\infty} \exp\left[-\frac{n}{2\sigma^2} (\mu - \bar{x})^2\right] d\mu d\sigma. \end{aligned}$$

Παρατηρούμε ότι στο ολοκλήρωμα ως προς την μεταβλητή μ εμφανίζεται ο πυρήνας μίας κανονικής κατανομής με μέση τιμή \bar{x} και διακύμανση σ^2 , επομένως έχουμε ότι

$$\begin{aligned} m_2^N(\mathbf{x}) &= (2\pi)^{-n/2} \int_0^{\infty} (\sigma^2)^{-\frac{n+2}{2}} \exp\left[-\frac{1}{2\sigma} \sum_{i=1}^n (x_i - \bar{x})^2\right] \frac{(2\pi)^{1/2} (\sigma^2)^{1/2}}{n^{1/2}} d\sigma \\ &= (2\pi)^{-n/2} (2\pi)^{1/2} n^{-1/2} \int_0^{\infty} (\sigma^2)^{-\frac{n-1}{2}-1} \exp\left[-\frac{(n-1)S^2}{2\sigma}\right] d\sigma, \end{aligned}$$

όπου

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Θέτοντας $\sigma = \sqrt{k}$ έχουμε ότι $d\sigma = 1/2\sqrt{k} dk$ και $0 < k < \infty$. Επομένως το παραπάνω ολοκλήρωμα μας δίνει ότι

$$\begin{aligned} m_2^N(\mathbf{x}) &= 2^{-1}(2\pi)^{-n/2}(2\pi)^{1/2}n^{-1/2} \int_0^\infty k^{-\frac{n-1}{2}-1} \exp\left[-\frac{(n-1)S^2}{2k}\right] k^{-1/2} dk \\ &= 2^{-1}(2\pi)^{-n/2}(2\pi)^{1/2}n^{-1/2} \int_0^\infty k^{-\frac{n}{2}-1} \exp\left[-\frac{(n-1)S^2}{2k}\right] dk. \end{aligned}$$

Παρατηρούμε ότι στο παραπάνω ολοκλήρωμα ως προς k σχηματίζεται ο πυρήνας μίας αντίστροφης κατανομή γάμμα με παραμέτρους $\alpha = n/2$ και $\beta = (n-1)S^2/2$, επομένως έχουμε ότι

$$\begin{aligned} m_2^N(\mathbf{x}) &= 2^{-1}(2\pi)^{-n/2}(2\pi)^{1/2}n^{-1/2} \frac{\Gamma(n/2)}{((n-1)S^2/2)^{n/2}} \\ &= \frac{\Gamma(n/2)}{\pi^{\frac{n-1}{2}} \sqrt{2n} ((n-1)S^2/2)^{n/2}}. \quad (E4.22) \end{aligned}$$

Στην συνέχεια προχωράμε στον υπολογισμό της εκ των υστέρων προβλεπτικής κατανομής των δεδομένων \mathbf{x} δοθέντος των φανταστικών δειγμάτων εκπαίδευσης \mathbf{x}^* του μοντέλου M_2 κάνοντας χρήση της αρχικής μη-κατάλληλης εκ των προτέρων κατανομής $\pi^N(\mu, \sigma)$. Αρχικά θα χρειαστεί να υπολογίσουμε την εκ των υστέρων κατανομή των παραμέτρων του μοντέλου M_2 . Συνεπώς έχουμε ότι

$$\begin{aligned} \pi_2^N(\mu, \sigma | \mathbf{x}) &\propto f(\mathbf{x} | \mu, \sigma, M_2) \pi^N(\mu, \sigma) \\ &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{(n-1)S^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right] \frac{1}{\sigma^2} \\ &\propto (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(\mu - \bar{x})^2}{2\sigma^2/n}\right] (\sigma^2)^{-\left(\frac{n-1}{2}+1\right)} \exp\left[-\frac{(n-1)S^2}{2\sigma^2}\right]. \end{aligned}$$

Επομένως παρατηρούμε ότι η εκ των υστέρων κατανομή των παραμέτρων μ και σ^2 του μοντέλου M_2 δοθέντος των δεδομένων \mathbf{x} είναι μία *Normal-Scaled-Inverse X^2* κατανομή με παραμέτρους $\mu = \bar{x}$, $\sigma^2 = \sigma^2/n$, $k = n-1$, $\tau^2 = S^2$.

Συνεπώς η εκ των υστέρων προβλεπτική κατανομή $m_2^N(\mathbf{x}^*|\mathbf{x})$ θα είναι

$$m_2^N(\mathbf{x}^*|\mathbf{x}) = \int_0^\infty \int_{-\infty}^\infty f(\mathbf{x}^*|\mu, \sigma) \pi_2^N(\mu, \sigma|\mathbf{x}) d\mu d\sigma$$

$$\propto \int_0^\infty \int_{-\infty}^\infty (\sigma^2)^{-\frac{n^*}{2}} \exp\left[-\frac{n^*}{2\sigma^2}(\bar{x}^* - \mu)^2\right] \exp\left[-\frac{(n^* - 1)S^{*2}}{2\sigma^2}\right]$$

$$\times (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right] (\sigma^2)^{-\left(\frac{n-1}{2}+1\right)} \exp\left[-\frac{(n-1)S^2}{2\sigma^2}\right] d\mu d\sigma.$$

όπου

$$S^{*2} = \frac{1}{n^* - 1} \sum_{i=1}^{n^*} (x_i^* - \bar{x}^*)^2.$$

Όμως ισχύει ότι

$$n^*(\bar{x}^* - \mu)^2 + (n^* - 1)S^{*2} + n(\bar{x} - \mu)^2 + (n - 1)S^2$$

$$= (n^* + n)(\mu - \mu^*)^2 + (n^* - 1)S^{*2} + (n - 1)S^2 + \frac{nn^*}{n + n^*}(\bar{x}^* - \bar{x})^2$$

με

$$\mu^* = \frac{n\bar{x} + n^*\bar{x}^*}{n + n^*}.$$

Επομένως έχουμε

$$m_2^N(\mathbf{x}^*|\mathbf{x}) = \int_0^\infty (\sigma^2)^{-\left[\frac{n^*+n+1}{2}+1\right]} \exp\left[-\frac{1}{2\sigma^2}\left((n^* - 1)S^{*2} + (n - 1)S^2\right.\right.$$

$$\left.\left. + \frac{nn^*}{n + n^*}(\bar{x}^* - \bar{x})^2\right)\right] \int_{-\infty}^\infty \exp\left[-\frac{1}{2\sigma^2}(n^* + n)(\mu - \mu^*)^2\right] d\mu d\sigma$$

$$\propto \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\frac{n^*+n+1}{2}+1} \sigma \exp\left[-\frac{1}{2\sigma^2}\left((n^* - 1)S^{*2} + (n - 1)S^2 + \frac{nn^*}{n + n^*}(\bar{x}^* - \bar{x})^2\right)\right] d\sigma$$

$$\propto \left[(n^* - 1)S^{*2} + (n - 1)S^2 + \frac{nn^*}{n + n^*}(\bar{x}^* - \bar{x})^2\right]^{-\frac{n+n^*}{2}}.$$

Επομένως παρατηρούμε ότι η εκ των υστέρων προβλεπτική κατανομή $m_2^N(\mathbf{x}^*|\mathbf{x})$ είναι μία δισδιάστατη t Student κατανομή με n βαθμούς ελευθερίας, $\bar{x}1_2$ παράμετρο θέσης και S^2M παράμετρο κλίμακας, όπου

$$M = I_2 + \frac{1}{n} 1_2 1_2^T,$$

όπου I_2 είναι ο 2×2 ταυτοτικός πίνακας και $1_2 = (1 \ 1)^T$.

Επομένως για να προχωρήσουμε στον υπολογισμό της περιθώριας πιθανοφάνειας $m_2^{EPP}(\mathbf{x})$ θα πραγματοποιήσουμε με την βοήθεια της R μία ολοκλήρωση *Monte Carlo* για την εκτίμηση της ποσότητας

$$m_2^{EPP}(\mathbf{x}) = m_2^N(\mathbf{x}) \int_{-\infty}^{\infty} m_2^N(\mathbf{x}^*|\mathbf{x}) \frac{m_1^N(\mathbf{x}^*)}{m_2^N(\mathbf{x}^*)} d\mathbf{x}^*.$$

Επομένως η διαδικασία που θα ακολουθήσουμε είναι

- Δημιουργία T φανταστικών δειγμάτων εκπαίδευσης $\mathbf{x}^{*(t)}, t = (1, 2, \dots, T)$ μεγέθους $n^* = 2$ από την εκ των υστέρων προβλεπτική κατανομή $m_2^N(\mathbf{x}^*|\mathbf{x})$.
- Εκτίμηση της ποσότητας $m_2^{EPP}(\mathbf{x})$ από την

$$\hat{m}_2^{EPP}(\mathbf{x}) = m_2^N(\mathbf{x}) \frac{1}{T} \sum_{t=1}^T \frac{m_1^N(\mathbf{x}^{*(t)})}{m_2^N(\mathbf{x}^{*(t)})}.$$

Συνεπώς ο εκτιμώμενος παράγοντας Μπεϋες της σχέσης (E4.17) θα δίνεται από

$$\widehat{BF}_{21}^{EPP}(\mathbf{x}) = \frac{m_1^N(\mathbf{x})}{\hat{m}_2^{EPP}(\mathbf{x})}. \quad (E4.23)$$

5.4.3 Παράδειγμα με προσομοιωμένα δεδομένα

Έστω ότι θέλουμε να συγκρίνουμε τα μοντέλα

$$M_1: \mathbf{X} \sim N(0, \sigma^2) \text{ με } \pi_1^N(\sigma) \propto \frac{1}{\sigma}$$

$$M_2: \mathbf{X} \sim N(\mu, \sigma^2) \text{ με } \pi_2^N(\mu, \sigma) \propto \frac{1}{\sigma^2}.$$

Υποθέτουμε ότι για τις εκ των προτέρων πιθανότητες των μοντέλων M_1 και M_2 ισχύει

$$\pi(M_1) = \pi(M_2) = \frac{1}{2}$$

επομένως ο εκ των υστέρων λόγος συμπληρωματικών πιθανοτήτων υπέρ του μοντέλου M_2 θα ισούται με παράγοντα Μπεϋες υπέρ του μοντέλου M_2 . Επομένως έχουμε ότι

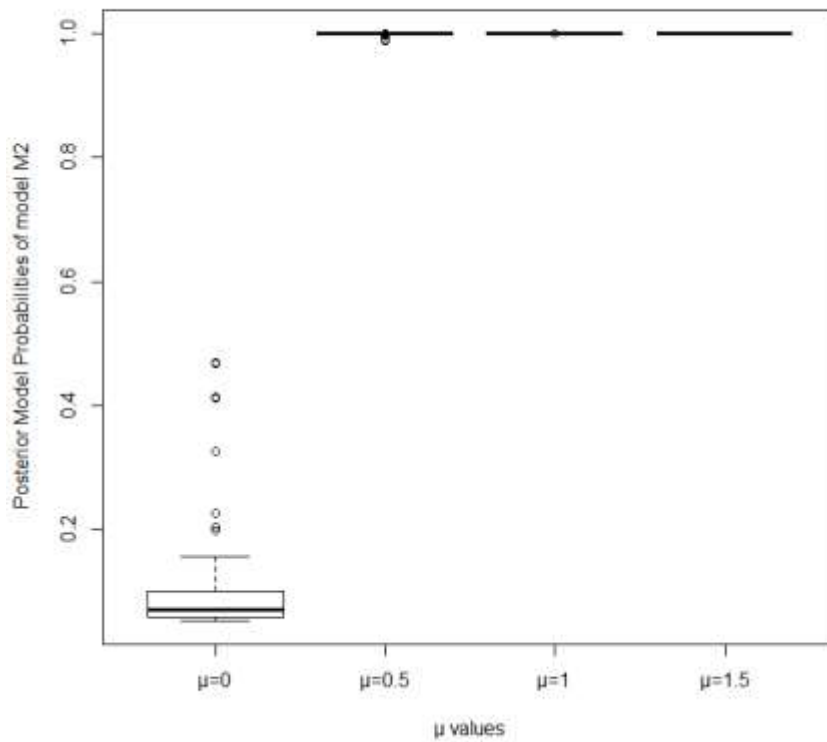
$$PO_{21} = \widehat{BF}_{21}^{EPP}(\mathbf{x})$$

$$\begin{aligned} \Rightarrow \frac{\pi(M_2|\mathbf{x})}{\pi(M_1|\mathbf{x})} &= \widehat{BF}_{21}^{EPP}(\mathbf{x}) \Rightarrow \frac{\pi(M_2|\mathbf{x})}{1 - \pi(M_2|\mathbf{x})} = \widehat{BF}_{21}^{EPP}(\mathbf{x}) \\ \Rightarrow \pi(M_2|\mathbf{x}) &= \frac{\widehat{BF}_{21}^{EPP}(\mathbf{x})}{1 + \widehat{BF}_{21}^{EPP}(\mathbf{x})}. \end{aligned} \quad (E4.24)$$

Εν συνεχεία μπορούμε να προσομοιώσουμε τυχαία δείγματα από την κανονική κατανομή για διάφορες τιμές των παραμέτρων μ, σ που θέλουμε να εξετάσουμε, προκειμένου να μελετήσουμε την συμπεριφορά της εκ των υστέρων πιθανότητας του μοντέλου M_2 . Θα πραγματοποιήσουμε 2 κύκλους προσομοιώσεων όπου θα παράγουμε τυχαία δείγματα προερχόμενα από την κανονική κατανομή σύμφωνα με τις τιμές των παραμέτρων μ και σ όπου έχουμε επιλέξει και στην συνέχεια θα προχωράμε στον υπολογισμό και την αποθήκευση των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 όπως αυτές προκύπτουν από την σχέση (E4.24).

1^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 100 τυχαία δείγματα τα οποία θα αποτελούνται από 150 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την κανονική κατανομή με διακύμανση $\sigma^2 = 1^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.10** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 100 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 150$.



Διάγραμμα 5.10: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 100 τυχαία δείγματα μεγέθους $n = 150$, $\sigma^2 = 1^2$

Στον **Πίνακα 5.8** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|\mathbf{x}) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

Πίνακας 5.8 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

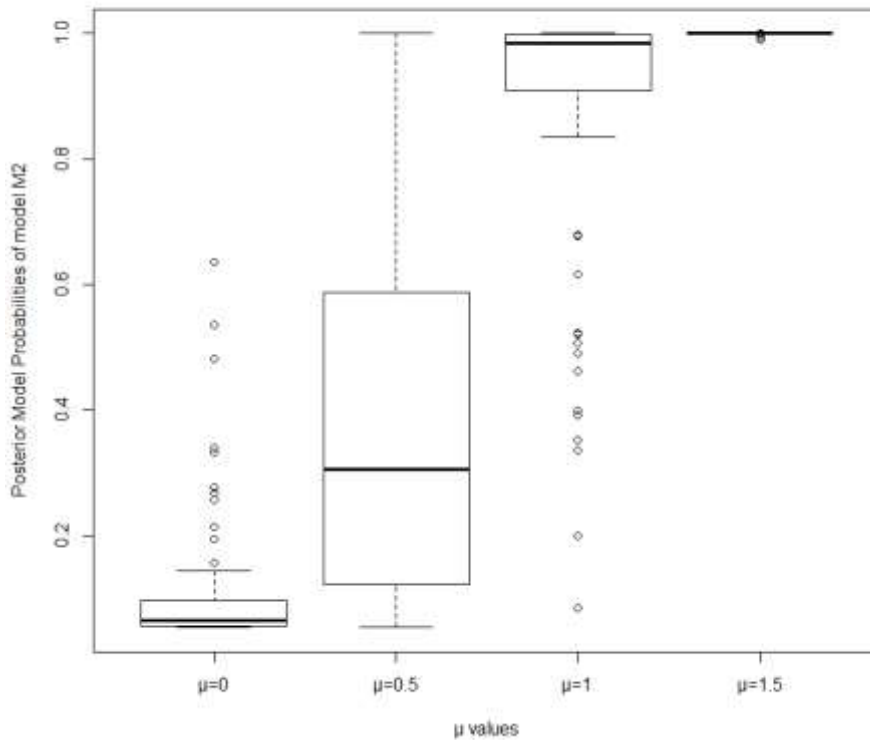
Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
0%	100%	100%	100%

Παρατηρούμε ότι όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την Κανονική κατανομή με παραμέτρους $\mu = 0$ και $\sigma^2 = 1$, επιλέγεται το μοντέλο M_1 ως το μοντέλο το οποίο παρήγαγε τα δεδομένα στο 100% των επαναλήψεων. Παρατηρούμε επίσης μία μικρή μεταβλητότητα στις εκ των υστέρων πιθανότητες του μοντέλου M_2 καθώς και την ύπαρξη έκτροπων τιμών, όπου οφείλονται στο γεγονός ότι χρησιμοποιούμε προσομοιωμένες τιμές.

Όσο απομακρυνόμαστε από την τιμή της παραμέτρου μ την οποία υποστηρίζει το μοντέλο M_1 , υποστηρίζεται πλήρως το μοντέλο M_2 στο 100% των επαναλήψεων, όπως γίνεται εμφανές όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την Κανονική κατανομή με μέση τιμή $\mu = 0.5, 1, 1.5$ αντίστοιχα και διακύμανση $\sigma^2 = 1$.

2^{ος} κύκλος προσομοιώσεων

Χρησιμοποιώντας την R θα προσομοιώσουμε 100 τυχαία δείγματα τα οποία θα αποτελούνται από 150 παρατηρήσεις το καθένα, οι οποίες θα προέρχονται από την Κανονική κατανομή με διακύμανση $\sigma^2 = 3^2$ και μέση τιμή $\mu = 0, 0.5, 1, 1.5$. Στο **Διάγραμμα 5.11** παρουσιάζονται τα αντίστοιχα θηκοδιαγράμματα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 για την κάθε τιμή της παραμέτρου μ για 100 προσομοιωμένα τυχαία δείγματα μεγέθους $n = 150$.



Διάγραμμα 5.11: Θηκοδιαγράμματα εκ των υστέρων πιθανοτήτων μοντέλου M_2 , για 100 τυχαία δείγματα μεγέθους $n = 150$, $\sigma^2 = 3^2$

Στον **Πίνακα 5.9** παρουσιάζεται επίσης το ποσοστό που επιλέχθηκε το μοντέλο M_2 βάση των εκ των υστέρων πιθανοτήτων του ($P(M_2|x) > 0.5$), ως το μοντέλο το οποίο περιγράφει τα προσομοιωμένα δεδομένα, για την κάθε τιμή της παραμέτρου μ από την οποία προήλθαν τα εκάστοτε τυχαία δείγματα.

Πίνακας 5.8 Επιλογές Μοντέλου M_2 ως αληθινό μοντέλο για τις διάφορες τιμές της παραμέτρου μ .

Επιλογή μοντέλου M_2			
$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
2%	29%	92%	100%

Σε αυτόν τον κύκλο προσομοιώσεων επιχειρήσαμε να τριπλασιάσουμε την τυπική απόκλιση που χρησιμοποιήσαμε στον πρώτο κύκλο προσομοιώσεων. Παρατηρούμε εμφανώς μία σημαντική αύξηση της μεταβλητότητας των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 . Όταν εξετάζουμε παρατηρήσεις οι οποίες προήλθαν από την κανονική κατανομή με μέση τιμή $\mu = 0$ και διακύμανση $\sigma^2 = 3^2$, παρατηρούμε ότι υποστηρίζεται το M_1 μοντέλο στο 98% των περιπτώσεων. Η εμφάνιση περισσότερων

έκτροπων τιμών οφείλεται στο γεγονός ότι μελετάμε προσομοιωμένα τυχαία δείγματα τα οποία έχουν αρκετά μεγάλη διακύμανση.

Στην συνέχεια παρατηρούμε ότι όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την κανονική κατανομή με μέση τιμή $\mu = 0.5$ και διακύμανση $\sigma^2 = 9$, το μοντέλο M_2 υποστηρίζεται στο 29% των περιπτώσεων, ποσοστό συντριπτικά μικρότερο του αντίστοιχου του πρώτου κύκλου προσομοιώσεων, γεγονός που οφείλεται στην υψηλή τιμή της διακύμανσης που έχουμε ορίσει, καθώς και στην απόσταση από την τιμή της παραμέτρου μ που ορίζει το μοντέλο M_1 .

Στην συνέχεια όταν τα προσομοιωμένα τυχαία δείγματα προέρχονται από την κανονική κατανομή με μέση τιμή $\mu = 1$ και διακύμανση $\sigma^2 = 9$, επιλέγεται το μοντέλο M_2 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα δεδομένα στο 92% των περιπτώσεων. Συγκριτικά με τον πρώτο κύκλο προσομοιώσεων παρατηρούμε ότι και πάλι υποστηρίζεται ισχυρά το μοντέλο M_2 , αλλά εμφανίζονται περισσότερες έκτροπες τιμές όπως επίσης και μεγαλύτερη μεταβλητότητα των εκ των υστέρων πιθανοτήτων του μοντέλου M_2 , εξαιτίας της μεγάλης τιμής της διακύμανσης που έχουμε ορίσει.

Στην συνέχεια όσο απομακρυνόμαστε από την τιμή της παραμέτρου μ την οποία ορίζει το μοντέλο M_1 , επιλέγεται το μοντέλο M_2 ως το μοντέλο το οποίο παρήγαγε τα προσομοιωμένα τυχαία δείγματα, παρά την υψηλή διακύμανση την οποία έχουμε ορίσει.

Επίλογος

Στη παρούσα διπλωματική εργασία μελετήθηκε η αντιμετώπιση του προβλήματος επιλογής μοντέλου με χρήση της Μπεϋζιανής Στατιστικής, χρησιμοποιώντας ενδογενείς και αναμενόμενα μεταγενέστερες εκ των προτέρων κατανομές.

Η Μπεϋζιανή Στατιστική προσεγγίζει το πρόβλημα επιλογής μοντέλου πολύ πρακτικότερα από την Κλασική Στατιστική, δίνοντας απαντήσεις ακόμα και σε αρκετά πολύπλοκα προβλήματα. Η μόνη δυσκολία ως προς τη διενέργεια των μεθόδων που έχουν προταθεί είναι ο υπολογισμός των παραγόντων Μπεϋες, ειδικά σε πολύπλοκα προβλήματα, καθώς επίσης και η ισχυρή εξάρτηση αυτών από τις πρότερες κατανομές που χρησιμοποιούνται.

Οι ενδογενείς παράγοντες Μπεϋες των Berger & Perrichi (1996) κατέστησαν δυνατό τον υπολογισμό παραγόντων Μπεϋες όταν χρησιμοποιούμε μη κατάλληλες, μη πληροφοριακές εκ των προτέρων κατανομές για τις παραμέτρους των μοντέλων που συγκρίνουμε. Άρα η συγκεκριμένη μεθοδολογία μπορούμε να πούμε ότι αντιμετωπίζει το μπεϋζιανό πρόβλημα επιλογής μοντέλου «αντικειμενικά». Επιπλέον, οι ενδογενείς παράγοντες Μπεϋες αντιστοιχούν ασυμπτωτικά σε πραγματικούς παράγοντες Μπεϋες με χρήση πρότερων κατανομών, οι οποίες καλούνται ενδογενείς εκ των προτέρων κατανομές. Βέβαια στην πράξη αρκετά συχνά προκύπτει ένα σημαντικό ζήτημα υπολογιστικού κόστους, καθώς όταν καλούμαστε να εξετάσουμε σχετικά πολύπλοκα μοντέλα, οι υπολογισμοί μπορούν να γίνουν αρκετά χρονοβόροι και πολύπλοκοι.

Οι αναμενόμενα μεταγενέστερες εκ των προτέρων κατανομές των Perez & Berger (2002) αποτελούν μία αρκετά πρακτικότερη προσέγγιση συγκριτικά με τις ενδογενείς εκ των προτέρων κατανομές. Σημαντικό προτέρημα τους αποτελεί η φιλοσοφία τους η οποία βασίζεται στον καθορισμό εκ των προτέρων κατανομών μέσω μίας αυτοματοποιημένης διαδικασίας παραγωγής φανταστικών δεδομένων από κατανομές πρόβλεψης. Επίσης πολύ σημαντικό προτέρημα της μεθόδου είναι ότι οι αναμενόμενα μεταγενέστερες εκ των προτέρων κατανομές είναι συμβατές μεταξύ τους. Το γεγονός ότι κατά την σύγκριση εμφωλευμένων μοντέλων, και για συγκεκριμένη επιλογή της προβλεπτικής κατανομής από όπου προέρχονται τα φανταστικά δεδομένα, οι ενδογενείς εκ των προτέρων κατανομές ισοδυναμούν με τις αναμενόμενα μεταγενέστερες εκ των προτέρων κατανομές, προσδίδει μία μπεϋζιανή ‘επικύρωση’ στην μεθοδολογία των αναμενόμενα μεταγενέστερων πρότερων κατανομών. Όπως και στις ενδογενείς έτσι και στις αναμενόμενα μεταγενέστερες εκ των προτέρων κατανομές, συναντάμε αρκετές δυσκολίες στους υπολογισμούς, με τη διαφορά όμως ότι τώρα είμαστε σε πιο εύκολη θέση να περιορίσουμε το υπολογιστικό κόστος μέσω αλγορίθμων *Monte Carlo* και μεθόδων MCMC.

Παράρτημα 1

Σε αυτό το παράρτημα θα παρουσιαστούν οι κώδικες της R που χρησιμοποιήθηκαν για την δημιουργία διαγραμμάτων, υλοποίηση υπολογισμών και την διενέργεια των εφαρμογών του *Κεφαλαίου 5*.

1) Διάγραμμα Παραδείγματος 1.1

```
sb<-function(n, a, b, theta)
{
  x<-rexp(n,theta)
  sx<-sum(x)
  theta<-seq(0,2*theta,0.0001)
  prior<-dgamma(sx,a,b)
  posterior<-rgamma(n,n+a,n*mean(x)+b)
  plot(theta, dgamma(theta, a+n, b+sx), type='l',ylab="density",xlab="theta")
  lines(theta, dgamma(theta, a,b), col=2)
  lines(theta, dgamma(theta, n-1, sx), col=3)
  legend("topright", c("posterior", "prior", "likelihood"), lty=1, col=1:4)
}
sb(1000,5,2,6)
```

2) Εφαρμογή 5.1

```
x<-c(23,261,87,7,120,14,62,47,225,71,246,21,42,20,5,12,120,11,3,14,
71,11,14,11,16,90,1,16,52,95)
n<-length(x)
L<-t(combn(x,2))
nn<-nrow(L)
Lproblem<-c()
j<-0
for(i in 1:nn)
{
  if(L[i,1]==L[i,2])
  {
    j<-j+1
    Lproblem[j]<-i
  }
}
L<-L[-Lproblem,]
S2y<-(n-1)*var(log(x))
m1N<-gamma(n)/(sum(x))^n
m2N<-gamma((n-1)/2)/((prod(x))*((pi)^((n-1)/2))*(sqrt(S2y))^(n-1)*2
*sqrt(n))

#Arithmetic Bayes Factor
CF12_arithmetic<-0
for(i in (1:length(L[,1])))
{
  CF12_arithmetic<-CF12_arithmetic+(2*L[i,1]*L[i,2]*abs(log(L[i,1]
]/L[i,2]))) / ((L[i,1]+L[i,2])^2)
}
IBF21arithmetic<-(m2N/m1N)*CF12_arithmetic/(length(L[,1]))
IBF21arithmetic
aprob_2<-IBF21arithmetic/(1+IBF21arithmetic)
aprob_2

#Geometric Bayes Factor
log_CF12_geometric<-0
```

```

for(i in (1:length(L[,1]))) {
  log_CF12_geometric<-log_CF12_geometric+log(2)+log(L[i,1])+log(L
[i,2])+log(abs(log(L[i,1])-log(L[i,2]))) - 2*log((L[i,1]+L[i,2])))
}
logIBF21geometric<-log(m2N)-log(m1N)+log_CF12_geometric/length(L[,1
])
IBF21geometric<-exp(logIBF21geometric)
IBF21geometric
gprob_2<-IBF21geometric/(1+IBF21geometric)
gprob_2

```

3) Εφαρμογή 5.2

Διάγραμμα 5.1

```

x<-seq(0,5,0.01)
f<-x^(1-1)/(beta(1,1)*(1+x)^(1+1))
plot(x,f,type="l",xlab="y",ylab="Beta-Prime(1,1) density")

```

Διάγραμμα 5.2

```

lo<-1
l<-seq(0.01,5,0.001)
f<-1/(lo*(1/lo+1)^2)
plot(l,f,type="l",xlab="λ",ylab="IP Density")

```

Εφαρμογή Εκθετικής κατανομής

```

x<-seq(0,5,0.01)
f<-x^(1-1)/(beta(1,1)*(1+x)^(1+1))
plot(x,f,type="l",xlab="y",ylab="Beta-Prime(1,1) density")
n<-100
lo<-1
M2acceptance<-mat.or.vec(1,6)
P0model2<-mat.or.vec(100,6)
val<-c(1,1.2, 1.4, 1.6, 1.8, 2)
for(i in (1:length(val))) {
  M2acceptance[1,i]<-0
  for(j in (1:n)) {
    x<-rexp(n,val[i])
    a<-expint_En(n, lo*n*mean(x), give=FALSE, strict=FALSE)
    logBF21<-(1-n)*log(lo) + 2*lo*n*mean(x) + lgamma(n+1) + (1-n
)*log(n*mean(x)) + log(a-(1/n)*(exp(-lo*n*mean(x))-lo*n*mean(x)*a))
    BF21<-exp(logBF21)
    P0model2[j,i]<-BF21/(1+BF21)
    if(P0model2[j,i]>0.5){
      M2acceptance[1,i]<-M2acceptance[1,i]+1
    }
  }
}
}
boxplot(P0model2[,1],P0model2[,2],P0model2[,3],P0model2[,4],P0model2
[,5],P0model2[,6],names=c("λ=1", "λ=1.2", "λ=1.4", "λ=1.6", "λ=1.8",
"λ=2"),xlab="λ values", ylab="Posterior Model Probabilities of model
M2")
M2acceptance<-M2acceptance/100
M2acceptance

```

Για την υλοποίηση του παραπάνω κώδικα χρειάζεται πρώτα να εγκαταστήσουμε και να φορτώσουμε την βιβλιοθήκη `gs1`. Ο παραπάνω κώδικας αφορά τον δεύτερο κύκλο προσομοιώσεων, επομένως για την υλοποίηση του πρώτου κύκλου προσομοιώσεων αρκεί να δώσουμε στην μεταβλητή `n` την τιμή 50.

4) Εφαρμογή 5.3

```

iterations<-1000
M2acceptance<-mat.or.vec(3,4)
POmodel2<-mat.or.vec(iterations,4)
mvalues<-c(0,0.5,1,1.5)
sigma2<-4
n<-2000
for(i in (1:length(mvalues))) {
  M2acceptance[2,i]<-0
  for(j in (1:iterations)) {
    x<-rnorm(n,mvalues[i],sqrt(sigma2))
    BF21<-dnorm(mean(x),0,sqrt((2*n+1)*sigma2/n))/dnorm(mean(x)
, 0, sqrt(sigma2/n))
    POmodel2[j,i]<-BF21/(1+BF21)
    if(POmodel2[j,i] > 0.5) {
      M2acceptance[3,i]<-M2acceptance[3,i]+1
    }
    else
    {
      M2acceptance[3,i]<-M2acceptance[3,i]
    }
  }
}
boxplot(POmodel2[,1],POmodel2[,2],POmodel2[,3],POmodel2[,4],names=c
("μ=0","μ=0.5", "μ=1", "μ=1.5"),xlab="μ values", ylab="Posterior Model
Probabilities of model M2")
M2acceptance[2,1]=M2acceptance[2,2]=M2acceptance[2,3]=M2acceptance[
2,4]= "times M2 true"
M2acceptance[1,1]="μ=0"
M2acceptance[1,2]="μ=0.5"
M2acceptance[1,3]="μ=1"
M2acceptance[1,4]="μ=1.5"
M2acceptance

```

Ο παραπάνω κώδικας αφορά τον πρώτο κύκλο προσομοιώσεων. Για την διενέργεια των υπολοίπων δίνουμε τις αντίστοιχες τιμές στις μεταβλητές `sigma2` και `n`.

5) Εφαρμογή 5.4

Διάγραμμα 5.10

```
curve(1/x,0,4,xlab="s values",ylab="IP density")
```

Διάγραμμα 5.11

```
f<-function(m,s){
  (1 - exp(-(m^2/s^2)))/(2*sqrt(pi)*m^2)
}
m<-seq(-5,5,length=100)
s<-seq(0.1,3,length=100)
z<-f(m,s)
persp(m,s,z = outer(m, s, f),theta=45,phi=5,expand=0.6,ticktype="detailed",xlab="μ values",ylab="s values",zlim=c(0,50),zlab="IP density")
```

Εφαρμογή Κανονικής κατανομής

```
iterations<-100
n<-150
POModel2<-mat.or.vec(iterations,4)
mvalues<-c(0,0.5,1,1.5)
sigma2<-9
M2acceptance<-mat.or.vec(1,4)
for(i in (1:length(mvalues))){
  for(j in (1:iterations)){
    x<-rnorm(n,mvalues[i],sqrt(sigma2))
    m2Nx<-(1/2)*(2*pi)^(-n/2)*sqrt(2*pi)*sqrt(1/n)/((n-1)*var(x)
/2)^(n/2)
    m1Nx<-(1/2)*(sqrt(2*pi))^(-n)/(sum(x^2)/2)^(n/2)
    MCsum<-0
    for(k in (1:1000)){
      I2<-diag(2)
      I<-c(1,1)
      I<-as.matrix(I)
      M<-I2 + (1/n)*I%*%t(I)
      I<-as.vector(I)
      z<-rmvt(1, sigma = M*var(x), df = n) + I*mean(x)
      MCsum<-MCsum + sqrt(pi)*(z[1]-z[2])^2/(2*pi*(z[1]^2 + z[
2]^2))
    }
    m2epp<-m2Nx*(MCsum/1000)
    BF21<-m2epp/m1Nx
    POModel2[j,i]<-BF21/(1+BF21)
    if(POModel2[j,i] > 0.5){
      M2acceptance[1,i]<-M2acceptance[1,i]+1
    }
    else
    {
      M2acceptance[1,i]<-M2acceptance[1,i]
    }
  }
}
M2acceptance<-M2acceptance/iterations
M2acceptance
boxplot(POModel2,names=c("μ=0","μ=0.5", "μ=1", "μ=1.5"),xlab="μ values", ylab="Posterior Model Probabilities of model M2")
```

Για την υλοποίηση του παραπάνω κώδικα χρειάζεται πρώτα να εγκαταστήσουμε και να φορτώσουμε την βιβλιοθήκη `rmvtnorm`. Ο παραπάνω κώδικας αφορά τον δεύτερο κύκλο προσομοιώσεων, επομένως για την υλοποίηση του πρώτου κύκλου προσομοιώσεων αρκεί να δώσουμε στην μεταβλητή `sigma2` την τιμή 1.

Βιβλιογραφία

A) Διεθνής Βιβλιογραφία

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. National Bureau of Standard Applied Mathematics Series 55.

Aitkin, M. (1991). Posterior Bayes Factors. *Journal of the Royal Statistical Society Series B*, **53**, 111-142.

Barnard, G.A. (1963). New Methods of Quality Control. *Journal of the Royal Statistical Society Series A*, **126**, 126-255.

Berger, J.O. and Perez, J.M. (2002). Expected-Posterior Prior Distributions for Model Selection. *Biometrika*, **89**, 491-511.

Berger, J.O., Bernardo, J. and Dongchu, S. (2009). The Formal Definition of Reference Priors. *The Annals of Statistics*, **37**, 905-938.

Berger, J.O. and Pericchi, L. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, **91**, 109-122.

Berger, J. and Pericchi, L. (1994). The Intrinsic Bayes Factor for Linear Models. *Technical Report #94-10C*, Purdue University Department of Statistics.

Bernardo, J. (1979). Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society Series B*, **41**, 113-147.

Cano, J.A. and Kessler, M. and Moreno, E. (2004). On Intrinsic Priors for Nonnested Models. *Sociedad de Estadística e Investigación Operativa*, **13**, 445-463.

Carlin, B.P., and Chib, S. (1993). Bayesian Model Choice via Markov Chain Monte Carlo. *Research Report 93-006*, University of Minnesota, Division of Biostatistics.

Consonni, G. (2008). Compatibility of Prior Specifications Across Linear Models. *Statistical Science*, **23**, 332-353.

Cressie, N., Davis, A.S., Folks J.L. and Policello G.E. (2013). The Moment-Generating Function and Negative Integer Moments. *The American Statistician*, **35**, 148-150.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Didelot, X., Everitt, R.G., Johansen, A.M. and Lawson, D.J. (2011). Likelihood-Free Estimation of Model Evidence. *Bayesian Analysis*, **6**, 49-76.

Dmochowski, J. (1994). Intrinsic Priors Via Kullback-Liebler Geometry. *Technical Report 94-15*, Purdue University, Department of Statistics.

Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society Series B*, **57**, 45–97.

Garthwaite, P.H. and Dickey, J.M. (1996). Quantifying and Using Expert Opinion for Variable-Selection Problems in Regression. *Chemometrics and Intelligent Laboratory Systems*, **35**, 1-34.

Gelman, A., Carling, J.B., Stern H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis Second Edition*. Chapman & Hall/CRC, Boca Raton USA.

George, E.I. (2000). Methods and Criteria for Model Selection. *Journal of the American Statistical Association*, **95**, 1304-1308.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integrations. *Econometrica*, **57**, 1317-1340.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, **14**, 382–417.

Huber, P. (1967). The Behavior of Maximum Likelihood Estimators under Nonstandard Conditions. *Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, **1**, 221-233.

Ibrahim, G.J. and Chen, M. (2000). Power Prior Distributions for Regression Models. *Statistical Science*, **15**, 46-60.

Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.

Kass, R.E., Tierney, L. and Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's method. *Essays in Honor of George Bernard*, eds. S. Geisser, J.S. Hodges, S.J. Press, and A. Zellner, Amsterdam: North Holland, 473-488.

Kass, R.E., and Vaidyanathan, S. (1992). Approximate Bayes Factors and Orthogonal Parameters, with Applications to Testing Equality of Two Binomial Proportions. *Journal of the Royal Statistical Society Series B*, **54**, 129-144.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.

Lewis, S. and Raftery, A. (1997). Estimating Bayes factor via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, **92**, 648-655.

Lindley, D. V. (1968). The Choice of Variables in Multiple Regression. *Journal of the Royal Statistical Society Series B*, **30**, 31-66.

McCulloch, R.E. and Rossi, P.E. (1991). A Bayesian Approach to Testing the Arbitrage Pricing Theory. *Journal of Econometrics*, **49**, 141-168.

Moreno, E., Bertolino, F. and Racugno, W. (2014). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing. *Journal of the American Statistical Association*, **93**, 1451-1460.

Newton, M.A. and Raftery, A. E. (1994). Approximate Bayesian Inference by the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society Series B*, **56**, 3-48.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBugs*. Wiley New York.

O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society Series B*, **57**, 99-138.

Raftery, A. (1996b). "Hypothesis testing and model selection", in W. Gilks, S. Richardson, and D. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, Suffolk, UK.

Roberts, H.V. (1965). Probabilistic Prediction. *Journal of the American Statistical Association*, **60**, 50-62.

Tierney, L. and Kadane, J.B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, **81**, 82-86

Tsionas. E.G. (2007). Exact Inference in Four-Parameter Generalized Gamma Distributions. *Communications in Statistics - Theory and Methods*, **30**, 747-756.

Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *Annals of Mathematics and Statistics*, **20**, 595-601.

B) Ελληνική Βιβλιογραφία

Ηλιόπουλος, Γ. (2004). *Σημειώσεις για το Μάθημα Μπεϋζιανή Στατιστική*. Πανεπιστήμιο Πειραιά, Πειραιάς.

Κούτρας, Μ. και Δαμιανού, Χ. (2003). *Εισαγωγή στη Στατιστική Μέρος 1*, Εκδόσεις Συμμετρία, Αθήνα.

Φουσκάκης, Δ. (2013). *Ανάλυση Δεδομένων με Χρήση της R*. Εκδόσεις Τσότρας, Αθήνα.