

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ**

Διπλωματική εργασία

**«Μοντέλα βάσης της Bayesienne Ανάλυσης
και
Σύμφυτες Απώλειες»**

της Θεοδώρας Καραμπατάκη

Επιβλέπων καθηγητής : Δούκισσα Κρητικού

**Εξεταστική επιτροπή
Δ. Κρητικού – Γ. Κοσιώρης – Ε. Τσολάκη**

Ηράκλειο, Μάιος 2004

MONTELA ΒΑΣΗΣ ΤΗΣ BAYESIENNE ΑΝΑΛΥΣΗΣ

και

ΣΥΜΦΥΤΕΣ ΑΠΩΛΕΙΕΣ

Εισαγωγή

Στηριζόμενοι στις παρατηρήσεις ενός τυχαίου φαινομένου, αντικείμενο της Στατιστικής Ανάλυσης είναι η εξαγωγή ενός συμπεράσματος στην αιτία του νόμου που γεννά τις παρατηρήσεις ώστε, είτε να αναλύσουμε ένα περασμένο φαινόμενο, είτε να προβλέψουμε ένα μελλοντικό γεγονός το οποίο συνδέεται εξ ολοκλήρου με τις αποφασιστικές όψεις αυτού του συμπεράσματος.

Η bayesienne μέθοδος χορηγεί ένα φυσικό πλαίσιο για την επίλυση των προβλημάτων της Στατιστικής Συμπερασματολογίας. Μέσα σ' αυτό το πλαίσιο η bayesienne λογική είναι συλληπτικά πολύ απλή με τον τύπο του Bayes να συνδυάζει την εκ των προτέρων πληροφορία (καθορισμένη, ακαθόριστη ή ακόμα και ανύπαρκτη) με την πληροφορία η οποία προέρχεται από το δείγμα.

Το Κεφάλαιο I, βασισμένο στον J-P. Florens (1998) σκοπό έχει να περιγράψει δύο τρόπους bayesienne σκέψης. Ο πρώτος είναι υποκειμενικός και αφορά τον καθορισμό των εκ των προτέρων πληροφορικών νόμων ενώ ο δεύτερος είναι αντικειμενικός και αφορά τον καθορισμό των μη-πληροφορικών εκ των προτέρων νόμων.

Το Κεφάλαιο II, βασισμένο στον L.Simar (1998) περιγράφει το πώς οι γενικές αρχές της bayesienne συμπερασματολογίας εφαρμόζονται σε κάποια μοντέλα βάσης χρήσιμα στη στατιστική ανάλυση. Στα παραμετρικά μοντέλα που εκτίθενται, οι μέθοδοι του Κεφαλαίου I χαρακτηρίζουν τους εκ των προτέρων νόμους και για όλα τα μοντέλα βάσης δίδονται αναλυτικές εκφράσεις των περιθωριακών (predictives) και των εκ των υστέρων νόμων.

Στη Στατιστική Θεωρία Αποφάσεων πέρα από τη δειγματική κατανομή σημαντικό ρόλο παίζει και η συνάρτηση απώλειας. Επειδή η επιλογή μιας συγκεκριμένης συνάρτησης απώλειας επηρεάζει ισχυρά τη συμπερασματολογία το Κεφάλαιο III, βασισμένο στον C.Robert (1996), σκοπό έχει να παρουσιάσει ως απαραίτητες τις *σύμφυτες απώλειες* όταν καμία πληροφορία δεν είναι διαθέσιμη σχετικά με τη συνάρτηση ωφέλειας του αποφασιστή, από το να επικαλεστούμε κλασσικές απώλειες, όπως η τετραγωνική απώλεια. Επειδή αυτή η τοποθέτηση είναι σχεδόν όμοια με την παραγωγή μη-πληροφορικών εκ των προτέρων νόμων στην Bayesienne Ανάλυση, ανακαλούμε αρχικά τις συνθήκες αυτής της παραγωγής και με βάση αυτές καταλήγουμε σε ορισμένες απαιτήσεις σχετικά με τις σύμφυτες απώλειες. Φαίνεται τότε ότι αυτές οι συναρτήσεις απώλειας εξαρτώνται μόνο από την κατανομή του δείγματος και πρέπει να είναι ανεξάρτητες από την παραμετροποίηση της κατανομής. Οι εκτιμητές που προκύπτουν είναι τότε *equivariant*. Μελετάμε τις ιδιότητες δύο σύμφυτων απωλειών δηλαδή τις απώλειες της entropie και του Hellinger και αποδεικνύουμε ότι μπορούν να εκφραστούν με συγκεκριμένο τύπο για τις εκθετικές οικογένειες. Επιπλέον η entropie απώλεια παρέχει αναλυτική έκφραση των εκτιμητών του Bayes, υπό συζυγείς εκ των προτέρων κατανομές. Η παραγωγή των εκτιμητών του Bayes που συνδέονται με την Hellinger

απώλεια είναι περισσότερο «φορτική» όπως φαίνεται από τις περιπτώσεις Γάμμα και Poisson, ενώ και οι δύο απώλειες οδηγούν σε παρόμοιους εκτιμητές.

Επειδή πολλές και σύνθετες είναι οι κατανομές που εμφανίζονται στην ανάλυση των παραμετρικών μοντέλων ιδιαίτερα του Κεφαλαίου II, το Παράρτημα περιλαμβάνει τις απαραίτητες κατανομές και κάποιες χαρακτηριστικές ιδιότητες τους χρήσιμες στη στατιστική μελέτη μας.

Θεωρούμε απαραίτητο πριν αναπτύξουμε το θέμα μας, να δώσουμε ένα γενικό πλαίσιο περιγραφής της bayesianne μεταχείρισης (βλέπε π.χ. C. Robert (1992)) στο οποίο στηρίζεται η ανάλυση των παρακάτω κεφαλαίων.

Το στατιστικό μοντέλο απαιτεί τρεις χώρους

- το χώρο των παρατηρήσεων, X
- τον παραμετρικό χώρο, Θ
- το χώρο των αποφάσεων, D

Το συμπέρασμα συνίσταται στο να πάρουμε μια απόφαση $\delta \in D$ που αφορά το $\theta (\in \Theta)$, στη θέα μιας παρατήρησης $x (\in X)$, τα x και θ πράγματι δεδομένα υπό το νόμο της τ.μ. x , $f(x|\theta)$.

Στην πράξη μια θεμελιώδης αρχή της bayesianne μεθόδου είναι :

Κάθε «στατιστικό συμπέρασμα» οφείλει να θεμελιώνεται στον αυστηρό καθορισμό των παρακάτω παραγόντων

- (1) της οικογένειας των νόμων των παρατηρήσεων, $f(x|\theta)$, ως προς ένα κυρίαρχο μέτρο ν
- (2) της εκ των προτέρων κατανομής των παραμέτρων, $\pi(\theta)$
- (3) της συνάρτησης απώλειας που συνδέεται με τις αποφάσεις δ , $L(\theta, \delta)$
($L : \Theta \times D \rightarrow R_+$).

Τα (1) και (2) επιτρέπουν να κατασκευάσουμε

α) τον από κοινού νόμο των x, θ

$$f(x, \theta) = f(x|\theta)\pi(\theta)$$

β) τον περιθωριακό (ή predictive) νόμο του x

$$m(x) = \int_{\Theta} f(x, \theta) d\theta$$

γ) τον εκ των υστέρων νόμο του θ δεδομένου του x

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

Εφαρμογές αυτών περιλαμβάνει το Κεφάλαιο II.

Το (2) μελετάται εν μέρει στο Κεφάλαιο I καθορίζοντας μεθόδους (υποκειμενικές ή αντικειμενικές) προσδιορισμού του π .

Στο (3), το $L(\theta, \delta)$ εκτιμά την «ποινή» που προκύπτει από τη χρήση της απόφασης δ όταν η παράμετρος είναι το θ . Με άλλα λόγια, το $L(\theta, \delta)$ παριστά την απώλεια η οποία εκτίθεται από την κακή εκτίμηση του θ .

Τα (2) και (3) (εκτός από το (1)) μπορούν γενικά να καθορισθούν και από εκτιμήσεις μερικώς υποκειμενικές. Όπως θα φανεί στο Κεφάλαιο III, το πρόβλημα κατασκευής του (3) είναι συχνά πιο σύνθετο από το (2) και ο καθορισμός της απώλειας είναι εξίσου ισχυρά συνδεδεμένος με τις εκ των προτέρων πληροφορίες επί του μοντέλου παρά με μια bayesienne ανάλυση.

Επειδή είναι δυνατό να ελαχιστοποιήσουμε την απώλεια $L(\theta, \delta)$ όταν το θ είναι άγνωστο, για να συγκαλύψουμε αυτό το πρόβλημα, θεωρούμε τη μέση απώλεια ή τον κίνδυνο

$$R(\theta, \delta) = \int_x L(\theta, \delta(x)) f(x | \theta) \nu(dx)$$

όπου $\delta(x)$ η συνάρτηση απόφασης (ή ο εκτιμητής του θ)

Βεβαίως μπορούμε να αναζητήσουμε, αν υπάρχει, μια απόφαση δ_0 η οποία ελαχιστοποιεί τον $R(\theta, \delta)$ ομοιόμορφα ως προς θ . Αυτό όμως αποτελεί σπάνια περίπτωση.

Η bayesienne μέθοδος αντί της ολοκλήρωσης επί του X , ολοκληρώνει επί του Θ επειδή το θ είναι η άγνωστη παράμετρος όταν το x είναι γνωστό. Ορίζουμε έτσι την εκ των υστέρων απώλεια

$$\rho(\pi, \delta | x) = \int_{\Theta} L(\theta, \delta) \pi(\theta | x) d\theta \quad (= E^{\pi}(L(\theta, \delta) | x))$$

η οποία αποτελεί το μέσο σφάλμα ακολουθώντας τον εκ των υστέρων νόμο του θ , δεδομένης της παρατηρούμενης τιμής x . Η εκ των υστέρων απώλεια είναι μια συνάρτηση του x , αυτή όμως η εξάρτηση δεν έχει τα ίδια αποτελέσματα με εκείνη η οποία συνδέει κίνδυνο και παράμετρο (επειδή το x , εν αντιθέσει με το θ , είναι γνωστό). Δυνάμεθα εξίσου να ορίσουμε τον κίνδυνο του Bayes (ή bayesienne κίνδυνο) :

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta$$

ο οποίος δίνει μια πραγματική τιμή και όχι πλέον μια συνάρτηση του θ . Ο κίνδυνος του Bayes αντισταθμίζει τους κινδύνους ακολουθώντας την πιθανοφάνεια της παραμέτρου θ . Οι παραπάνω δύο έννοιες εκτίμησης είναι ισοδύναμες με την ακόλουθη έννοια :

Ένας εκτιμητής ελαχιστοποιώντας τον κίνδυνο του Bayes $r(\pi, \delta)$ δύναται να προκύψει εκλέγοντας, για κάθε x , την τιμή $\delta(x)$ η οποία ελαχιστοποιεί την εκ των υστέρων απώλεια $\rho(\pi, \delta | x)$, επειδή

$$r(\pi, \delta) = \int_X \rho(\pi, \delta | x) m(x) \nu(dx)$$

Δυνάμεθα τότε να ορίσουμε την έννοια του εκτιμητή του Bayes συνδεδεμένη με μια απώλεια και με μια εκ των προτέρων κατανομή ως εξής :

Ορισμός : Ονομάζουμε εκτιμητή του Bayes συνδεδεμένο με ένα εκ των προτέρων νόμο π και με μια απώλεια L , κάθε εκτιμητή δ^π ο οποίος ελαχιστοποιεί τον $r(\pi, \delta)$.

Για κάθε $x \in X$, αυτός δίδεται από το $\delta^\pi(x)$ που συμπεραίνεται από την :

$$\min_{\delta} \rho(\pi, \delta | x) = \min_{\delta} E^\pi(L(\theta, \delta) | x)$$

Έτσι προκύπτει μια κατασκευαστική μέθοδος καθορισμού των εκτιμητών του Bayes. Σημειώνουμε ότι, από απόψεως καθαρά bayesienne, μόνο η εκ των υστέρων απώλεια είναι ενδιαφέρουσα επειδή το bayesien παράδειγμα θεμελιώνεται επί της δεσμευτικής μεθόδου.

Το παραπάνω αποτέλεσμα είναι έγκυρο για τους propriés $\left(\int_{\Theta} \pi(\theta) d\theta = 1 \right)$ ή τους impropriés $\left(\int_{\Theta} \pi(\theta) d\theta = +\infty \right)$ νόμους, λοιπόν το ίδιο αν ο bayesien κίνδυνος είναι άπειρο.

Στην περίπτωση ενός improprie νόμου, ορίζουμε τον συνδεδεμένο με αυτόν εκτιμητή του Bayes, καλούμενο γενικευμένο εκτιμητή του Bayes, σαν εκείνο ο οποίος ελαχιστοποιεί, για κάθε x , την εκ των υστέρων απώλεια. Γενικά ο εκτιμητής του Bayes είναι μοναδικός.

Οι δύο έννοιες της optimalite, “minimaxite” και “admissibilite” της Θεωρίας Αποφάσεων, εξεταζόμενες υπό την bayesienne οπτική φαίνεται να ικανοποιούνται από τους εκτιμητές του Bayes. Δυστυχώς, σε ορισμένες περιπτώσεις η χρήση των σύμφυτων απωλειών υπό τις εκ των προτέρων συζυγείς που χρησιμοποιούμε στο Κεφάλαιο III, μας οδηγούν σε εκτιμητές του Bayes με αναλυτικές εκφράσεις πολύπλοκες καθιστώντας δύσκολη τη μελέτη της optimalite αυτών των εκτιμητών.

Παρατήρηση : Γενικά στη Θεωρία των Πιθανοτήτων, χρησιμοποιούμε το κεφαλαίο X για να σημειώσουμε μια τυχαία μεταβλητή (τ.μ.) και το μικρό x για να παραστήσουμε μια δυνατή τιμή της X . Στην εργασία αυτή, σε όλα τα Κεφάλαια εκτός από το Παράρτημα, για να ελαφρώσουμε τις διάφορες έννοιες που παρουσιάζονται, το x θα σημειώνει την τ.μ. ή μια τιμή της.

Οι συμβολισμοί των παραπάνω εννοιών δεν τηρούνται αυστηρά στο παρακάτω κείμενο (Κεφάλαια I, II) από το γεγονός ότι υπάρχει μεγάλη ποικιλία παραμέτρων που εμπλέκονται στα διάφορα μοντέλα που μελετάμε. Έτσι η επιλογή των συμβόλων για τις διάφορες έννοιες γίνεται κατά τον καλύτερο δυνατό τρόπο.

Επίσης, πολλές φορές εξαιτίας της σύνθετης μορφής μιας πυκνότητας πιθανότητας (π.π) που εμφανίζεται στον λογισμό, κάνουμε χρήση του πυρήνα της δηλαδή του βασικού μέρους της συνάρτησης πυκνότητας χωρίς τη σταθερά.

I. Καθορισμός της εκ των προτέρων κατανομής

Το bayesienne στατιστικό μοντέλο αναλύεται γενικά σε δύο μέρη : στην εκλογή της οικογένειας των πιθανοτήτων του δείγματος και εκείνη της εκ των προτέρων κατανομής των παραμέτρων. Ο καθορισμός του μοντέλου του δείγματος δεν είναι αντικείμενο της bayesienne ανάλυσης και το σύνολο των τεχνικών εκτίμησης του (tests ή κριτήρια επιλογής του μοντέλου, robustes μέθοδοι κ.α.) μπορούν να χρησιμοποιηθούν. Ο καθορισμός της εκ των προτέρων κατανομής είναι αντιθέτως η ουσία της bayesienne ανάλυσης και απαιτεί ένα σύνολο αρχικών συλλογισμών ως προς τη συμπερασματική προσέγγιση η οποία θεμελιώνεται μοναδικά επί του δείγματος.

Πολύ σχηματικά μπορούμε να εκθέσουμε δύο τρόπους bayesienne σκέψης. Ο πρώτος είναι υποκειμενικός και θεωρεί ότι η εκ των προτέρων κατανομή εκφράζει τις γνώσεις προ της παρατήρησης των δεδομένων. Η θεωρία αποφάσεως προτείνει μια αξιωματική καθορισμού των εκ των προτέρων γνώσεων οδηγώντας στην κατασκευή των εκ των προτέρων πιθανοτήτων. Εμείς δε θα αναπτύξουμε αυτή την αξιωματική της οποίας η δικαίωση είναι περισσότερο θεωρητική παρά πρακτική. Π.χ. προκειμένου περί ευαισθησίας και ιδιαιτέρως στη μελέτη των κινδύνων με ασθενή πιθανότητα, είναι θεμελιώδες να συσσωματώσουμε στη στατιστική μελέτη τη γνώμη των ειδικών. Το αντικείμενο είναι τότε να συμβάλλουμε στο να εκφραστούν αυτές οι γνώμες υπό τη μορφή εκ των προτέρων πιθανοτήτων. Μακροπρόθεσμα όμως αυτός ο εκ των προτέρων ορισμός οφείλει να ικανοποιήσει την ανάγκη απλότητας των υπολογισμών και αυτός είναι ιδιαίτερα ο λόγος της ανάπτυξης των «φυσικών συζυγών» (“naturelles conjuguées”) κατανομών σ’ένα δειγματοληπτικό μοντέλο που θα μελετήσουμε στο κεφάλαιο II . Η ανάπτυξη των αριθμητικών μεθόδων κερδίζει το ενδιαφέρον σ’αυτήν την απαίτηση απλότητας των υπολογισμών σε όφελος της ορθότητας της εκ των προτέρων κατανομής.

Ο δεύτερος τρόπος είναι πιο αντικειμενικός και δεν παροχετεύει τις εκ των προτέρων γνώσεις πέραν του χρήστη. Μπορούμε τότε να παραμείνουμε bayesiens απουσία της εκ των προτέρων πληροφορίας. Αναζητούμε λοιπόν τις εκ των προτέρων « μη – πληροφορικές », εκφράζοντας την εκ των προτέρων άγνοια, αλλά χρησιμοποιούμε τις παραμέτρους σαν τυχαίες μεταβλητές και κατά συνέπεια χρησιμοποιούμε το σύνολο των bayesiens υπολογισμών. Ο ορισμός της εκ των προτέρων « μη – πληροφορικής » δεν είναι μοναδικός και υπάρχουν διάφοροι τύποι αυτής της έννοιας. Αντικειμενική προσέγγιση δε σημαίνει αναγκαστικά εκ των προτέρων μη – πληροφορική και θα επικαλεσθούμε τότε την άποψη καλούμενη “ bayesienne empirique”.

1. Εκ των προτέρων μη – πληροφορικές

Στην παράγραφο αυτή θα φανεί ότι η εκλογή μιας εκ των προτέρων κατανομής μη – πληροφορικής οδηγεί συχνά στον καθορισμό ενός μέτρου και όχι μιας πιθανότητας και θα εξετασθούν συνοπτικά τα μαθηματικά συμπεράσματα μιας τέτοιας τροποποίησης. Θα θεωρήσουμε δύο από τις πιο γνωστές μεθόδους :

- την αναζήτηση μιας εκ των προτέρων αναλλοίωτης και
- τη μέθοδο του Jeffrey

από τις οποίες τη δεύτερη θα επικαλεσθούμε κυρίως στα επόμενα κεφάλαια.

Για οποιαδήποτε μέθοδο, ο καθορισμός ενός εκ των προτέρων μέτρου « μη – πληροφορικού » συνίσταται πάντοτε στο να ορίσουμε ένα μέτρο επί του παραμετρικού χώρου Θ δια μέσου του δειγματικού μηχανισμού ο οποίος περιγράφεται από το δείγμα x και τη δειγματική πιθανότητα $p(x|\theta)$. Αυτή η δειγματική πιθανότητα είναι δεσμευτική γενικά με το μέγεθος του δείγματος και με ένα σύνολο επεξηγηματικών μεταβλητών που περιγράφουν τις υποθέσεις της παρατήρησης. Η εκ των προτέρων πιθανότητα που θα προκύψει δια μέσου των μεθόδων που θα παρουσιάσουμε θα εξαρτάται λοιπόν γενικά από το μέγεθος του δείγματος και τις επεξηγηματικές μεταβλητές.

Το γεγονός ότι το εκ των προτέρων μέτρο εξαρτάται από το μέγεθος του δείγματος μπορεί να είναι ενοχλητικό σε ό,τι αφορά τις συνήθεις ασυμπτωτικές ιδιότητες του μοντέλου και τη συνέπεια των εκτιμητών του Bayes. Η παρουσία του μεγέθους του δείγματος στην εκ των προτέρων συμβαίνει στα ανεξάρτητα και ισόνομα μοντέλα αυτό το οποίο τονίζει τη δυσκολία της ασυμπτωτικής ανάλυσης.

Το γεγονός ότι η εκ των προτέρων κατανομή εξαρτάται από τις επεξηγηματικές μεταβλητές είναι πιο φυσικό. Σ'ένα μοντέλο παλινδρόμησης (βλέπε κεφάλαιο II) η ερμηνεία των παραμέτρων είναι συνάρτηση των υπό όρων μεταβλητών και δεν αποτελεί λοιπόν έκπληξη το ότι η εκ των προτέρων περιγράφει αυτή τη σχέση. Από το άλλο μέρος μπορούμε να θεωρήσουμε ότι η παρατήρηση των επεξηγηματικών μεταβλητών έχει ήδη πραγματοποιηθεί σε έναν πρώτο χρόνο και ότι το μέτρο επί των παραμέτρων συσσωματώνει τις πληροφορίες που προέρχονται από αυτές τις μεταβλητές.

1.1 Εκ των προτέρων μέτρα

Θεωρούμε κατ' αρχήν την αυτονόητη σύσταση : Αν το δειγματοληπτικό μοντέλο ορίζεται από μια πυκνότητα $p(x|\theta)$ παραμέτρου θ , πεπερασμένης διάστασης, ένα εκ των προτέρων μέτρο θα χαρακτηρίζεται πολύ συχνά από την πυκνότητα του $m(\theta)$, αναφορικά με το μέτρο του Lebesgue. Αυτή η πυκνότητα είναι μια πραγματική συνάρτηση θετική ή μηδέν αλλά το ολοκλήρωμά της δεν είναι αναγκαία πεπερασμένο. Στην πράξη, αν το ολοκλήρωμα είναι πεπερασμένο, μπορούμε να οδηγηθούμε στην περίπτωση μιας πιθανότητας χωρίς απώλεια της γενικότητας διότι ο υπολογισμός της εκ των υστέρων είναι δυνατός. Το πρόβλημα λοιπόν τροποποιείται όταν $\int_{\Theta} m(\theta)d\theta = +\infty$.

Για να χρησιμοποιήσουμε σ'αυτήν την περίπτωση το θεώρημα του Bayes, πρέπει να επαληθεύσουμε ότι ο παρονομαστής δηλαδή η περιθωριακή πυκνότητα (ή ο “predictive” του δείγματος)

$$p(x) = \int_{\Theta} p(x|\theta)m(\theta)d\theta$$

δεν είναι άπειρο. Αν $p(x) = +\infty$, ο τύπος του Bayes δεν μπορεί πλέον να δικαιολογείται σαν τύπος υπολογισμού ενός δεσμευτικού νόμου.

Παράδειγμα 1 : Έστω $x|\theta \sim N(\theta_1, \theta_2)$ όπου $\theta = (\theta_1, \theta_2) \in \Theta (= R \times R_+^*)$ και θ ακολουθεί εκ των προτέρων το μέτρο του Lebesgue επί του R^2 . Τότε

$$p(x|\theta) = (2\pi)^{-1/2} \theta_2^{-1/2} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\}, \forall x \in R \text{ και } m(\theta) = 1.$$

Λοιπόν ,

$$p(x, \theta) = p(x|\theta)m(\theta) = (2\pi)^{-1/2} \theta_2^{-1/2} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\}, \forall (x, \theta) \in R \times \Theta$$

και

$$\begin{aligned} p(x) &= \int_{\Theta} p(x, \theta) d\theta = \int_{R \times R_+^*} (2\pi)^{-1/2} \theta_2^{-1/2} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\} d\theta_1 d\theta_2 \\ &= \int_{R_+^*} \underbrace{\left(\int_R (2\pi)^{-1/2} \theta_2^{-1/2} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\} d\theta_1 \right)}_{\text{ολοκλήρω. } N(x, \theta_2)=1} d\theta_2 \\ &= \int_{R_+^*} d\theta_2 = +\infty \end{aligned} \quad \blacktriangle$$

Μαθηματικότερα, η μεταχείριση ενός εκ των προτέρων μέτρου δύναται να παρουσιασθεί κατά τον ακόλουθο τρόπο :

Έστω μ ένα μέτρο σ -πεπερασμένο επί ενός χώρου γινόμενο $(A \times B, \mathcal{A} \otimes \mathcal{B})$. Για να αποσυντεθεί το μ σ' ένα περιθωριακό μέτρο επί του A και ένα δεσμευμένο μέτρο επί του B δεδομένου ενός στοιχείου του A , είναι αναγκαίο το περιθωριακό μέτρο επί του A να είναι σ -πεπερασμένο (δηλαδή να υπάρχει μια αριθμήσιμη οικογένεια $(A_n)_{n \in N}$, $A_n \in \mathcal{A} : A = \bigcup_n A_n$ με $\mu(A_n) < +\infty, \forall n$). Υπό αυτήν την υπόθεση το θεώρημα των Radon – Nikodym εξασφαλίζει την ύπαρξη δεσμευμένης πιθανότητας (μάζας 1 για σχεδόν κάθε τιμή της μεταβλητής δέσμευσης).

Εφαρμόζοντας αυτό το αποτέλεσμα στην bayesienne ανάλυση επιτρέπει να κατασκευάσουμε το ακόλουθο :

Θεωρούμε μια οικογένεια δειγματοληπτικών πιθανοτήτων P^θ επί του X (χώρος παρατηρήσεων) και ένα εκ των προτέρων μέτρο m σ -πεπερασμένο επί του Θ . Ορίζουμε τότε ένα μέτρο σ -πεπερασμένο επί του γινομένου $\Theta \times X$ και ο δεσμευμένος νόμος του Θ ως προς X θα είναι ορισμένος μόνο όταν ο περιθωριακός νόμος επί του X είναι σ -πεπερασμένος. Σ' αυτήν την περίπτωση ο εκ των υστέρων νόμος θα είναι σ.β. μια πιθανότητα.

Αυτός ο συλλογισμός δύναται επίσης να εφαρμοστεί για τον καθορισμό μιας εκ των προτέρων επί ενός διανύσματος : αν $\theta = {}^t(\theta_1, \theta_2)$, μπορούμε να ορίσουμε ένα από κοινού μέτρο επί του ${}^t(\theta_1, \theta_2)$ καθορίζοντας μια πιθανότητα επί του θ_2 δεδομένου του θ_1 και ένα μέτρο σ -πεπερασμένο επί του θ_1 . Αντιθέτως, αν μ_1, μ_2 είναι δύο μέτρα επί των Θ_1, Θ_2 αντιστοίχως με πυκνότητες $m_1(\theta_1)$ και $m_2(\theta_2)$ αντιστοίχως :

$$\int_{\Theta_1} m_1(\theta_1) d\theta_1 = +\infty \text{ και } \int_{\Theta_2} m_2(\theta_2) d\theta_2 = +\infty$$

τότε το γινόμενο $m(\theta) = m_1(\theta_1)m_2(\theta_2)$ με $\theta = {}^t(\theta_1, \theta_2)$, ορίζει καλώς ένα μέτρο μ επί του $\Theta_1 \times \Theta_2$ με πυκνότητα m και

$$\mu(\Theta_1 \times \Theta_2) = \int_{\Theta_1 \times \Theta_2} m(\theta) d\theta = \int_{\Theta_1 \times \Theta_2} m_1(\theta_1) m_2(\theta_2) d\theta_1 d\theta_2 = \int_{\Theta_1} m_1(\theta_1) d\theta_1 \int_{\Theta_2} m_2(\theta_2) d\theta_2 = +\infty$$

Η ερμηνεία των μ_1, μ_2 ως περιθωριακά μέτρα δεν έχει καμία βάση αφού π.χ.

$$\int_{\Theta_2} m(\theta) d\theta_2 = \int_{\Theta_2} m_1(\theta_1) m_2(\theta_2) d\theta_2 = m_1(\theta_1) \int_{\Theta_2} m_2(\theta_2) d\theta_2 = m_1(\theta_1)(+\infty)$$

Σημειώνουμε τέλος για να κλείσουμε αυτήν την παράγραφο ότι : αν $m(\theta)$ είναι τέτοια ώστε η predictive $p(x) < +\infty$, είναι το ίδιο για την $cm(\theta)$, όπου c οποιαδήποτε σταθερά. Επιπλέον η $m(\theta | x)$ θα είναι αμετάβλητη περνώντας από την m στην cm .

1.2 Αναλλοίωτα εκ των προτέρων μέτρα

Θεωρείται φυσικό το γεγονός να τυποποιούμε την απουσία εκ των προτέρων πληροφορίας από μια ιδιότητα του αναλλοίωτου : αν η παράμετρος $\theta (\in R)$ παριστά τη μάζα ενός αντικειμένου και αν το αντικείμενο κατέχει μια πληροφορία α για το θ , οι εκ των προτέρων κατανομές των θ και $\theta + \alpha$ αντιστοίχως δεν είναι βεβαίως ταυτόσημες. Αντιθέτως θα λέμε ότι η κατανομή του θ και εκείνη του $\theta + \alpha$ είναι ταυτόσημες για κάθε α , εκφράζοντας σαφώς την άγνοια επί της τιμής του θ .

Λοιπόν, ένα μέτρο M επί του Θ είναι *αναλλοίωτο* (*invariante*) υπό έναν μετασχηματισμό $\varphi : \Theta \rightarrow \Theta$, αν το M συμπίπτει με την εικόνα του μέσω του φ , $\varphi(M)$:

$$\forall \Theta_1, \varphi(M)(\Theta_1) = M(\varphi^{-1}(\Theta_1))$$

Αν το M χαρακτηρίζεται από την πυκνότητα του m και αν φ είναι 1-1 και επί και αυτή καθώς και η αντίστροφη της είναι παραγωγίσιμες, η πυκνότητα του $\varphi(M)$ είναι ίση με :

$$\|\partial\varphi^{-1}\| m \circ \varphi^{-1}$$

όπου $\|\partial\varphi^{-1}\|$ σημειώνει την απόλυτη τιμή της ορίζουσας του πίνακα των μερικών παραγώγων της φ^{-1} και η υπόθεση του αναλλοίωτου είναι λοιπόν η ισότητα

$$m = \|\partial\varphi^{-1}\| m \circ \varphi^{-1} \quad \text{σ.β.}$$

Παράδειγμα 2 : Το μέτρο του Lebesgue στον R^k ($m(\theta) = 1$) είναι αναλλοίωτο για όλες τις μεταφορές $\varphi(\theta) = \theta + \alpha$, $\alpha \in R^k$.

Πράγματι, θέτοντας $y = \theta + \alpha$ λαμβάνουμε $\theta = y - \alpha$ και $(\partial\varphi^{-1}) = \left(\frac{\partial\theta_i}{\partial y_j} \right)_{i,j=1,\dots,k} = I_k$,

όπου I_k ο μοναδιαίος πίνακας τάξεως k . Λοιπόν, $\|\partial\varphi^{-1}\| = 1$ και $m(y) = 1 = \|\partial\varphi^{-1}\| m \circ \varphi^{-1}(y)$. ▲

Θα λέμε ότι το M είναι *σχετικά αναλλοίωτο* υπό τον φ , αν

$$\varphi(M) = \Delta(\varphi)M$$

όπου ο πολλαπλασιαστικός συντελεστής $\Delta(\varphi)$ είναι συνάρτηση μόνο του μετασχηματισμού φ .

Παράδειγμα 3 : Το μέτρο του Lebesgue επί του R^k είναι σχετικά αναλλοίωτο υπό τους μετασχηματισμούς

$$\varphi(\theta) = A\theta + \alpha$$

όπου $A (k, k)$ - πίνακας αντιστρέψιμος και $\alpha \in R^k$.

Πράγματι, θέτοντας $y = A\theta + \alpha$ έχουμε $\theta = A^{-1}(y - \alpha)$. Αν λοιπόν $\theta = {}^t(\theta_1, \dots, \theta_k)$, $\alpha = {}^t(\alpha_1, \dots, \alpha_k)$ και $A^{-1} = (a^1 : a^2 : \dots : a^k)$, όπου $a^j = {}^t(a_{1j}, \dots, a_{kj})$, $j = 1, \dots, k$, λαμβάνουμε $\theta_i = a_{i1}(y_1 - \alpha_1) + \dots + a_{ik}(y_k - \alpha_k)$, $i = 1, \dots, k$.

Συνεπώς,

$$\left(\partial \varphi^{-1} \right) = \left(\frac{\partial \theta_i}{\partial y_j} \right)_{i,j=1,\dots,k} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix} = A^{-1}.$$

Σημειώνοντας με M το μέτρο του Lebesgue επί του R^k και m την πυκνότητα του (εδώ $m(\theta) = 1$) έχουμε ότι το $\varphi(M)$ δέχεται π.π. την : $\| \partial \varphi^{-1} \| m \circ \varphi^{-1}(y) = \| A^{-1} \| m(y)$, δηλαδή

$$\varphi(M) = \Delta(\varphi)M \text{ με } \Delta(\varphi) = \| A^{-1} \|.$$

Αν $A (k, k)$ - πίνακας ορθογώνιος, υπό τον ίδιο μετασχηματισμό, $\varphi(\theta) = A\theta + \alpha$, $\| A^{-1} \| = 1$ και το M είναι πάλι αναλλοίωτο. ▲

Η θεωρία των μέτρων του Haar (βλέπε π.χ. στο [12]) και πιο γενικά των ομογενών χώρων ενδιαφέρεται για μέτρα αναλλοίωτα και σχετικά αναλλοίωτα για ομάδες μετασχηματισμών και θεωρεί ιδιαίτερος το πρόβλημα της *μοναδικότητας* (υπό έναν πολλαπλασιαστικό παράγοντα) ενός αναλλοίωτου μέτρου. Βασιζόμενοι σ' αυτήν την θεωρία, σ' ένα bayesian μοντέλο μπορούμε να ορίσουμε την εκ των προτέρων άγνοια δια του αναλλοίωτου αναφορικά με μια οικογένεια Φ , μετασχηματισμών του Θ επί του εαυτού του και αν η Φ είναι αρκετά μεγάλη θα εκλέξουμε τότε σαν εκ των προτέρων μέτρο μη-πληροφορικό το μοναδικό αναλλοίωτο μέτρο αναφορικά με την Φ . Το γεγονός ότι το μέτρο αυτό ορίζεται υπό έναν πολλαπλασιαστικό παράγοντα δεν αποτελεί πρόβλημα διότι η εκ των υστέρων πιθανότητα δε θα τροποποιηθεί από τον παράγοντα. Για τον ίδιο λόγο ένα μέτρο *σχετικά αναλλοίωτο* είναι μια κατάλληλη εκλογή διότι η εκ των υστέρων πιθανότητα θα παραμείνει η ίδια αν αντικαταστήσουμε το M με το $\varphi(M)$. Πράγματι ο πολλαπλασιαστικός παράγοντας $\Delta(\varphi)$ εξαφανίζεται στον λογιισμό της εκ των υστέρων. Ενδιαφέρον αποτελεί η στατιστική κατασκευή της οικογένειας Φ (βλέπε

στο [3]) από το δειγματοληπτικό στατιστικό μοντέλο, η οποία είναι ιδιαίτερα χρήσιμη στους ελέγχους στατιστικών υποθέσεων.

1.3 Εκ των προτέρων μέτρο του Jeffrey

Ο τρόπος καθορισμού μιας εκ των προτέρων μη-πληροφορικής, γνωστή υπό το όνομα *εκ των προτέρων μέτρο του Jeffrey*, συνίσταται στο να προσδιοριστεί σ'ένα δειγματοληπτικό μοντέλο το οποίο χαρακτηρίζεται από την πιθανοφάνεια $p(x|\theta)$, το εκ των προτέρων μέτρο, πυκνότητας

$$m_J(\theta) = [\det I(\theta)]^{1/2}$$

αναφορικά με το μέτρο του Lebesgue. Ο πίνακας $I(\theta)$ είναι ο πίνακας πληροφοριών του Fisher :

$$I(\theta) = (I_{ij}(\theta))_{i,j} \text{ όπου } I_{ij}(\theta) = -E_{\theta} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(x|\theta) \right).$$

Αυτή η κατασκευή έχει έννοια μόνο όταν το θ είναι ένα διάνυσμα πραγματικών συντεταγμένων και υπό τις συνήθεις υποθέσεις της κανονικότητας. Η πυκνότητα $m_J(\theta)$ δεν χαρακτηρίζει γενικώς μια πιθανότητα, αλλά ένα μέτρο όπως το δείχνει το παρακάτω παράδειγμα.

Παράδειγμα 4 : Έστω $x = (x_1, \dots, x_n)$ τυχαίο δείγμα (x_i ανεξάρτητες και ισόνομες τ.μ.) μιας $N(\mu, \sigma^2)$. Θέτοντας $\theta = (\theta_1, \theta_2) = (\mu, \sigma) \in \Theta = R \times R_+^*$, έχουμε :

$$p(x|\theta) = \prod_{i=1}^n p(x_i|\theta) = (2\pi)^{-n/2} \theta_2^{-n} \exp \left\{ -\frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2 \right\}$$

και

$$\log p(x|\theta) = \log(2\pi)^{-n/2} - n \log \theta_2 - \frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2.$$

Απ' όπου, λαμβάνοντας τις μερικές παραγώγους μέχρι 2ας τάξεως ως προς τις συντεταγμένες του θ , από τον ορισμό των $I_{ij}(\theta)$ και του γεγονότος ότι

$$\sum_{j=1}^n \left(\frac{x_j - \theta_1}{\theta_2} \right)^2 \sim \chi_n^2, \text{ προκύπτει :}$$

$$I(\theta) = \begin{pmatrix} \frac{n}{\theta_2^2} & 0 \\ 0 & \frac{2n}{\theta_2^2} \end{pmatrix} = n \begin{pmatrix} \frac{1}{\theta_2^2} & 0 \\ 0 & \frac{2}{\theta_2^2} \end{pmatrix}.$$

Λοιπόν,

$$[\det I(\theta)]^{1/2} = \frac{n\sqrt{2}}{\theta_2^2} \text{ και } m_J(y) \propto \frac{1}{\theta_2^2}. \quad \blacktriangle$$

Το εκ των προτέρων μέτρο του Jeffrey ικανοποιεί ορισμένες ιδιότητες συνδεδεμένες με τις χαρακτηριστικές του υλικού πληροφορίας :

- Σε μια δειγματοληψία από ανεξάρτητες και ισόνομες δοκιμές ($p(x|\theta) = \prod_{i=1}^n p(x_i|\theta)$) το μέτρο του Jeffrey εξαρτάται από το μέγεθος του δείγματος μόνο διαμέσου ενός πολλαπλασιαστικού παράγοντα τον οποίο μπορούμε να παραλείψουμε.
- Το εκ των προτέρων μέτρο του Jeffrey δεν κατασκευάζεται δια μέσου μιας επαρκούς στατιστικής συνάρτησης του αρχικού δείγματος (διότι η $I(\theta)$ δεν τροποποιείται)
- Το μέτρο του Jeffrey δεν είναι αναλλοίωτο (γενικά) με την έννοια του αναλλοίωτου για μια οικογένεια μετασχηματισμών όπως είδαμε στην προηγούμενη παράγραφο. Όμως λέμε συχνά ότι αυτό το μέτρο είναι αναλλοίωτο χρησιμοποιώντας τον όρο υπό την εξής έννοια : Θεωρούμε ένα μετασχηματισμό $\varphi: \Theta \rightarrow \Lambda$, 1-1 και επί, ο οποίος μετασχηματίζει την παράμετρο θ στην $\lambda = \varphi(\theta)$. Δυνάμεθα τότε να ορίσουμε ένα καινούριο στατιστικό μοντέλο, δείγματος x και πιθανοφάνειας $p(x|\lambda) = p(x|\varphi^{-1}(\lambda))$. Αυτό το καινούριο μοντέλο έχει ένα εκ των προτέρων μέτρο του Jeffrey πυκνότητας $\tilde{m}_J(\lambda)$ και επαληθεύεται άμεσα ότι η $\tilde{m}_J(\lambda)$ είναι η εικόνα της $m_J(\theta)$ μέσω της φ .

Θεωρώντας χάριν απλότητας τη μονοδιάστατη περίπτωση έχουμε :

$$\tilde{m}_J(\lambda) = \left[-E_{\varphi^{-1}(\lambda)} \left[\left(\log p(x|\varphi^{-1}(\lambda)) \right)'' \right] \right]^{1/2}$$

(όπου η παράγωγος είναι ως προς λ)

Όμως,

$$\begin{aligned} \left(\log p(x|\varphi^{-1}(\lambda)) \right)'' &= \frac{d}{d\lambda} \left[\frac{d}{d\lambda} \log p(x|\varphi^{-1}(\lambda)) \right] \\ &= \frac{d}{d\varphi^{-1}(\lambda)} \left[\frac{d}{d\varphi^{-1}(\lambda)} \log p(x|\varphi^{-1}(\lambda)) \frac{d\varphi^{-1}(\lambda)}{d\lambda} \right] \frac{d\varphi^{-1}(\lambda)}{d\lambda} \\ &= \frac{d^2}{(d\varphi^{-1}(\lambda))^2} \log p(x|\varphi^{-1}(\lambda)) \left(\frac{d\varphi^{-1}(\lambda)}{d\lambda} \right)^2 + \frac{d}{d\varphi^{-1}(\lambda)} \log p(x|\varphi^{-1}(\lambda)) \frac{d^2}{d\lambda^2} \varphi^{-1}(\lambda) \end{aligned}$$

και

$$E_{\varphi^{-1}(\lambda)} \left[\left(\log p(x|\varphi^{-1}(\lambda)) \right)'' \right] = \frac{d^2}{d\lambda^2} \varphi^{-1}(\lambda) E_{\varphi^{-1}(\lambda)} \left[\frac{d}{d\varphi^{-1}(\lambda)} \log p(x|\varphi^{-1}(\lambda)) \right]$$

$$+ \left(\frac{d\varphi^{-1}(\lambda)}{d\lambda} \right)^2 E_{\varphi^{-1}(\lambda)} \left[\frac{d^2}{d(\varphi^{-1}(\lambda))^2} \log p(x | \varphi^{-1}(\lambda)) \right]$$

Όπου, λαμβάνοντας υπόψη τις υποθέσεις κανονικότητας (βλέπε π.χ. στο [14])

$$\begin{aligned} E_{\varphi^{-1}(\lambda)} \left[\frac{d}{d\varphi^{-1}(\lambda)} \log p(x | \varphi^{-1}(\lambda)) \right] &= \int_x \frac{d}{d\varphi^{-1}(\lambda)} \log p(x | \varphi^{-1}(\lambda)) p(x | \varphi^{-1}(\lambda)) dx \\ &= \int_x \frac{1}{p(x | \varphi^{-1}(\lambda))} \frac{d}{d\varphi^{-1}(\lambda)} p(x | \varphi^{-1}(\lambda)) p(x | \varphi^{-1}(\lambda)) dx \\ &= \int_x \frac{d}{d\varphi^{-1}(\lambda)} p(x | \varphi^{-1}(\lambda)) dx \\ &= \frac{d}{d\varphi^{-1}(\lambda)} \underbrace{\int_x p(x | \varphi^{-1}(\lambda)) dx}_{1} = 0. \end{aligned}$$

Συνεπώς,

$$\begin{aligned} \tilde{m}_J(\lambda) &= \left| - E_{\varphi^{-1}(\lambda)} \left[\frac{d^2}{d(\varphi^{-1}(\lambda))^2} \log p(x | \varphi^{-1}(\lambda)) \right] \right|^{1/2} \left| \frac{d\varphi^{-1}(\lambda)}{d\lambda} \right| \\ &= \left| \frac{d\varphi^{-1}(\lambda)}{d\lambda} \right| \left| \mathbf{I}(\varphi^{-1}(\lambda)) \right|^{1/2} \\ &= \left| \frac{d\varphi^{-1}(\lambda)}{d\lambda} \right| m_J(\varphi^{-1}(\lambda)). \end{aligned}$$

Λοιπόν, $\tilde{M}_J = \varphi(M_J)$ για τα αντίστοιχα μέτρα, πυκνοτήτων \tilde{m}_J και m_J αντιστοίχως.

Το αναλλοίωτο του μέτρου του Jeffrey οφείλει λοιπόν να εννοείται σαν αναλλοίωτο αναφορικά με την εκλογή της παραμετροποίησης.

2. Κατανομές εκ των προτέρων πληροφορικές – Οικογένειες « φυσικές συζυγείς » με μια οικογένεια δειγματοληπτικών πιθανοτήτων

Έστω $P^\theta, \theta \in \Theta$, μια οικογένεια δειγματοληπτικών πιθανοτήτων επί του χώρου των παρατηρήσεων X . Η παράμετρος θ δεν είναι αναγκαία πεπερασμένης διάστασης. Έστω \mathcal{M} μια οικογένεια πιθανοτήτων επί του Θ .

1. Θα λέμε ότι η \mathcal{M} είναι *κλειστή* αν για κάθε πιθανότητα της \mathcal{M} εκλεγόμενη *εκ των προτέρων* και κάθε παρατηρούμενο δείγμα, ο εξ αυτών συμπεραινόμενος *εκ των υστέρων νόμος*, είναι επίσης ένα στοιχείο της \mathcal{M} .
2. Υποθέτουμε επιπλέον ότι τα στοιχεία της \mathcal{M} υποδεικνύονται από μια «υπερπαράμετρο» $a \in A$. Αυτή η γραφή επιτρέπει να παραστήσουμε το bayesian συμπέρασμα δηλαδή το μετασχηματισμό εκ των προτέρων – εκ των υστέρων, σαν

μια συνάρτηση του $A \times X$ μέσα στο A , η οποία στην τιμή (a^*, x) , όπου a^* χαρακτηρίζει την εκ των προτέρων και x το δείγμα, συνδέει την a χαρακτηρίζοντας την εκ των υστέρων πιθανότητα μέσα στην \mathcal{M} . Αυτή η παράσταση έχει ενδιαφέρον μόνο όταν η ένδειξη είναι «απλή». Στα μοντέλα πεπερασμένης διάστασης (βλέπε κεφάλαιο II) θα αναζητήσουμε την οικογένεια \mathcal{M} έτσι ώστε το A να είναι επίσης πεπερασμένης διάστασης.

Παραδείγματα οικογενειών \mathcal{M} αναφέρονται αναλυτικά στο κεφάλαιο II επαληθεύοντας την ιδιότητα της «κλειστότητας». Ενδεικτικά και περιληπτικά αναφέρουμε τα εξής :

Παράδειγμα 5 : Στο διωνυμικό μοντέλο, $x | \theta \sim Bin(n, \theta)$, $x \in \{0, 1, \dots, n\}$, $\theta \in (0, 1)$, η οικογένεια των νόμων Βήτα είναι κλειστή. Ένας νόμος Βήτα υποδεικνύεται από δύο παραμέτρους (a, b) και αν $\theta \sim Beta(a_0, b_0)$ - εκ των προτέρων, έχουμε : $\theta | x \sim Beta(a^*, b^*)$ -εκ των υστέρων με $a^* = a_0 + x$, $b^* = b_0 + n - x$ ▲

Παράδειγμα 6 : Αν $x = {}^t(x_1, \dots, x_n)$ τυχαίο δείγμα από την $N_p(\theta, I_p)$, η οικογένεια των κανονικών νόμων επί του R^p είναι κλειστή : αν $\theta \sim N_p(a_0, \Sigma_0)$ - εκ των προτέρων, τότε $\theta | x \sim N_p(a_*, \Sigma_*)$ - εκ των υστέρων με $a_* = (\Sigma_0^{-1} + nI_p)^{-1}(\Sigma_0^{-1}a_0 + n\bar{x})$, $\Sigma_* = (\Sigma_0^{-1} + nI_p)^{-1}$ και $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Πράγματι, σημειώνοντας με $\|\cdot\|$ τη συνήθη ευκλείδεια πορμε, έχουμε :

$$\begin{aligned} p(x, \theta) &= p(x | \theta)m(\theta) = \prod_{j=1}^n p(x_j | \theta)m(\theta) \\ &= (2\pi)^{-pn/2} \exp\left\{-\frac{1}{2} \sum_{j=1}^n \|x_j - \theta\|^2\right\} (2\pi)^{-p/2} |\Sigma_0|^{-1/2} \exp\left\{-\frac{1}{2} {}^t(\theta - a_0)\Sigma_0^{-1}(\theta - a_0)\right\} \\ &= (2\pi)^{-(n+1)p/2} |\Sigma_0|^{-1/2} \exp\left\{-\frac{1}{2} \left(\sum_{j=1}^n \|x_j - \theta\|^2 + {}^t(\theta - a_0)\Sigma_0^{-1}(\theta - a_0)\right)\right\} \end{aligned}$$

όπου

$$\begin{aligned} &\sum_{j=1}^n \|x_j - \theta\|^2 + {}^t(\theta - a_0)\Sigma_0^{-1}(\theta - a_0) \\ &= \sum_{j=1}^n \|x_j\|^2 + n\|\theta\|^2 + {}^t\theta\Sigma_0^{-1}\theta - 2n{}^t\theta\bar{x} - 2{}^t\theta\Sigma_0^{-1}a_0 + {}^t a_0\Sigma_0^{-1}a_0 \\ &= \sum_{j=1}^n \|x_j\|^2 + {}^t a_0\Sigma_0^{-1}a_0 + {}^t\theta(nI_p + \Sigma_0^{-1})\theta - 2{}^t\theta(n\bar{x} + \Sigma_0^{-1}a_0) \\ &= \sum_{j=1}^n \|x_j\|^2 + {}^t a_0\Sigma_0^{-1}a_0 + {}^t\theta(nI_p + \Sigma_0^{-1})\theta - 2{}^t\theta(nI_p + \Sigma_0^{-1})(nI_p + \Sigma_0^{-1})^{-1}(n\bar{x} + \Sigma_0^{-1}a_0) \\ &\quad \pm {}^t(n\bar{x} + \Sigma_0^{-1}a_0)(nI_p + \Sigma_0^{-1})^{-1}(nI_p + \Sigma_0^{-1})(nI_p + \Sigma_0^{-1})^{-1}(n\bar{x} + \Sigma_0^{-1}a_0) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n \|x_j\|^2 + {}^t a_0 \Sigma_0^{-1} a_0 - {}^t (n\bar{x} + \Sigma_0^{-1} a_0) (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0) + \\
&\quad + {}^t \theta (nI_p + \Sigma_0^{-1}) \theta - 2 {}^t \theta (nI_p + \Sigma_0^{-1}) (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0) + \\
&\quad + {}^t (n\bar{x} + \Sigma_0^{-1} a_0) (nI_p + \Sigma_0^{-1})^{-1} (nI_p + \Sigma_0^{-1}) (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0) \\
&= \sum_{j=1}^n \|x_j\|^2 + {}^t a_0 \Sigma_0^{-1} a_0 - {}^t (n\bar{x} + \Sigma_0^{-1} a_0) (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0) + \\
&+ {}^t \underbrace{(\theta - (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0))}_{a_*} \underbrace{(nI_p + \Sigma_0^{-1})}_{\Sigma_*^{-1}} (\theta - (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0)).
\end{aligned}$$

Άρα,

$$\begin{aligned}
p(x) &= \int_{R^p} p(x, \theta) d\theta = (2\pi)^{-(n+1)p/2} |\Sigma_0|^{-1/2} \int_{R^p} \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n \|x_j - \theta\|^2 + {}^t (\theta - a_0) \Sigma_0^{-1} (\theta - a_0) \right] \right\} d\theta \\
&= (2\pi)^{-(n+1)p/2} |\Sigma_0|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n \|x_j\|^2 + {}^t a_0 \Sigma_0^{-1} a_0 - {}^t (n\bar{x} + \Sigma_0^{-1} a_0) (nI_p + \Sigma_0^{-1})^{-1} (n\bar{x} + \Sigma_0^{-1} a_0) \right] \right\} \\
&\quad \times \underbrace{\int_{R^p} \exp \left\{ -\frac{1}{2} {}^t (\theta - a_*) \Sigma_*^{-1} (\theta - a_*) \right\} d\theta}_{\text{ολικλ. } N_p(a_*, \Sigma_*) = (2\pi)^{p/2} |\Sigma_*|^{1/2}}
\end{aligned}$$

Συνεπώς,

$$\begin{aligned}
m(\theta | x) &= \frac{p(x | \theta) m(\theta)}{p(x)} \\
&= (2\pi)^{-p/2} |\Sigma_*|^{1/2} \exp \left\{ -\frac{1}{2} {}^t (\theta - a_*) \Sigma_*^{-1} (\theta - a_*) \right\}, \text{π.π. της } N_p(a_*, \Sigma_*). \quad \blacktriangle
\end{aligned}$$

Η έννοια της «φυσικής συζυγής» περιέχει περισσότερα από την απλή υπόθεση της κλειστότητας. Αυτή περιέχει την διαισθητική ιδέα ότι μια εκ των προτέρων πιθανότητα, προερχόμενη από τη φυσική συζυγή οικογένεια, μπορεί να θεωρηθεί σαν μια *εκ των υστέρων* πηγάζουσα από μια *εκ των προτέρων μη-πληροφορική* και από ένα *κατάλληλο δείγμα*.

Αυτό συμβαίνει θεωρώντας τις εκ των προτέρων πιθανότητες οι οποίες περιέχουν μια πληροφορία ισοδύναμη με την παρατήρηση ενός προγενέστερου δείγματος συνδεδεμένου με μια μη-πληροφορική κατανομή. Για να δηλώσουμε αυτό το σημείο θεωρούμε μια δειγματοληψία ανεξάρτητη και ισόνομη από την εκθετική. Στην πράξη η χρήσιμη ιδιότητα είναι :

για κάθε μέγεθος n του δείγματος, υπάρχει μια επί συνάρτηση $s(n, x)$ με τιμές στο $S (\subset R^l, l$ -σταθερό) έτσι ώστε η πυκνότητα των παρατηρήσεων να παραγοντοποιείται στην :

$$p(x | \theta) = u(n, x)v(s(n, x), \theta)$$

Λοιπόν η v είναι μια πραγματική συνάρτηση επί του $S \times \Theta (\subset R^l \times R^k)$ με $\theta \in \Theta (\subset R^k)$.

Υποθέτουμε δεδομένο ένα μέτρο μ επί του R^k τέτοιο ώστε :

$$\int_{\Theta} v(s, \theta) \mu(d\theta) < +\infty, \forall s.$$

Θεωρούμε τότε την οικογένεια πιθανοτήτων,

$$\mathcal{M}_{\mu} = \left\{ M : \exists a \in S, \frac{dM}{d\mu} = \frac{v(a, \theta)}{\int_{\Theta} v(a, \theta) \mu(d\theta)}, \mu - \sigma\beta \right\}$$

Κάθε στοιχείο M της \mathcal{M}_{μ} μπορεί να θεωρηθεί σαν μια εκ των υστέρων : αν $a \in S, \exists n_0$ δειγματικό μέγεθος και ένα δείγμα $x_0 : a = s(n_0, x_0)$. Θεωρούμε το μοντέλο που γεννάται από το δείγμα x_0 εφοδιασμένο με την εκ των προτέρων πιθανότητα μ . Η εκ των υστέρων πιθανότητα είναι τότε ακριβώς η M , διότι :

$$\frac{dM}{d\mu} = \frac{u(n_0, x_0)v(s(n_0, x_0), \theta)}{\int_{\Theta} u(n_0, x_0)v(s(n_0, x_0), \theta)\mu(d\theta)}.$$

Η οικογένεια \mathcal{M}_{μ} είναι κλειστή, διότι αν πάρουμε τώρα M , σαν εκ των προτέρων και αν παρατηρήσουμε ένα δείγμα μεγέθους n , η εκ των υστέρων πιθανότητα θα είναι ισοδύναμη με την πιθανότητα προκύπτουσα από την εκ των προτέρων μ και την ταυτόχρονη παρατήρηση του δείγματος μεγέθους $n_* = n + n_0$ και πραγματοποίησεων $(x_0, x) = x_*$ (διότι η εκ των υστέρων πυκνότητα ως προς το μέτρο του Lebesgue λ , θα

$$\text{ήταν : } m(\theta | x) \propto p(x | \theta) \frac{dM}{d\mu}(\theta) \frac{d\mu}{d\lambda}(\theta) \propto p(x | \theta) p(x_0 | \theta) m(\theta).$$

Αυτή η πιθανότητα ανήκει βεβαίως στην \mathcal{M}_{μ} . Τέλος η οικογένεια \mathcal{M}_{μ} δεικνύεται κανονικά από το $a \in S$ και ο τύπος περάσματος από το a_0 στο a_* δίδεται από τη συνάρτηση s ($a_* = s(n_*, x_*), n_* = n + n_0, x_* = (x_0, x), a_0 = s(n_0, x_0)$).

Αυτή η κατασκευή εξαρτάται από την εκλογή του μέτρου βάσης μ . Αυτό μπορεί να μην είναι ένα χαρακτηριστικό μέτρο απουσίας των εκ των προτέρων πληροφοριών αλλά να είναι μια πιθανότητα «πληροφορική». Λαμβάνουμε τότε μια αντίληψη της «γενικευμένης» φυσικής συζυγής οικογένειας που χρησιμοποιείται π.χ. στην οικονομετρία. Μπορούμε τέλος να επεκτείνουμε ελαφρά την οικογένεια των νόμων πιθανότητας κατασκευασμένη από τον νόμο μ και την συνάρτηση v όπως το δείχνει το παρακάτω παράδειγμα :

Παράδειγμα 7 : Θεωρούμε το τυχαίο δείγμα $x = (x_1, \dots, x_n)$ από την $Bin(1, \theta)$ με $\theta \in [0, 1]$ δηλαδή $x_i \in \{0, 1\}$ και $P_\theta(x_i = 1) = \theta, i = 1, \dots, n$. Έστω μ ένα μέτρο επί του $[0, 1]$ πυκνότητας $\theta^{-1}(1-\theta)^{-1}$ αναφορικά με το μέτρο του Lebesgue. Το μοντέλο είναι εκθετικό και η δειγματοληπτική πιθανότητα παραγοντοποιείται ως εξής :

$$\begin{aligned} p(x | \theta) &= \prod_{i=1}^n p(x_i | \theta) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} 1_{\{0,1\}}(x_i) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} 1_{\{0,1\}^n}(x) \\ &= \nu(s(n, x), \theta) u(n, x) \end{aligned}$$

με

$$s(n, x) = \left(n, \sum_{i=1}^n x_i \right), \nu(s, \theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}, u(n, x) = 1_{\{0,1\}^n}(x).$$

Η οικογένεια των νόμων πιθανότητας που λαμβάνεται κανονικοποιώντας αυτή την έκφραση της $\nu(s, \theta)$ στο θ είναι τότε η οικογένεια των νόμων Βήτα , ακεραίων παραμέτρων. Θεωρούμε τότε ότι η οικογένεια όλων των νόμων Βήτα είναι η «φυσική συζυγής» οικογένεια ακόμη αν μόνο οι Βήτα νόμοι των ακεραίων παραμέτρων ερμηνεύονται αυστηρά σαν δυνάμενοι να προκύψουν από ένα προγενέστερο δείγμα. ▲

Η θεωρία των εκ των προτέρων φυσικών συζυγών έχει γίνει αντικείμενο συστηματικής μεταχείρισης στις εκθετικές οικογένειες πυκνοτήτων (βλέπε πχ κεφάλαιο II και III) καθώς στα μη-παραμετρικά ανεξάρτητα και ισόνομα μοντέλα.

II. Μοντέλα Βάσης της Bayesienne Ανάλυσης

Σ' αυτό το κεφάλαιο, βασιζόμενοι στο [16], δείχνουμε πώς οι γενικές αρχές της bayesienne συμπερασματολογίας εφαρμόζονται σε κάποια μοντέλα βάσης, χρήσιμα στη στατιστική ανάλυση. Σε έκαστο των μοντέλων η διαδικασία η οποία γεννά τα δεδομένα, χαρακτηρίζεται από ένα πεπερασμένο αριθμό άγνωστων παραμέτρων : θα μιλάμε λοιπόν περί παραμετρικών μοντέλων. Θα εκλέξουμε τη μέθοδο «φυσική συζυγής» για να χαρακτηρίσουμε τους *εκ των προτέρων νόμους*, δείχνοντας κάθε φορά πώς να παρίσταται σε κάθε οικογένεια νόμων, μια εκ των προτέρων «λίγο» ή «μη-πληροφορική». Για κάθε παρουσιαζόμενο μοντέλο θα δόσουμε την έκφραση των εκ των υστέρων νόμων καθώς και των predictives (περιθωριακών) νόμων. Για όλα τα μοντέλα βάσης, από τις διαδικασίες Bernoulli μέχρι τις διαδικασίες της Κανονικής παλινδρόμησης, λαμβάνουμε τις αναλυτικές εκφράσεις αυτών των νόμων.

1. Διαδικασίες Bernoulli και Binomial

1.1 Ορισμοί

Μια *διαδικασία Bernoulli* μπορεί να ορισθεί σαν μια ακολουθία ταυτοτικών και ανεξάρτητων δοκιμών όπου έκαστη δοκιμή δύναται να δώσει δύο δυνατές καταστάσεις (τις οποίες θα ονομάσουμε απλά «επιτυχία» και «αποτυχία»). Αν π είναι η πιθανότητα επιτυχίας σε κάθε δοκιμή, ο διωνυμικός νόμος $p(x | m, \pi)$ επιτρέπει να εκτιμήσουμε την πιθανότητα του αριθμού των επιτυχιών x σε μια ακολουθία m δοκιμών Bernoulli, όπου το m είναι σταθερό. Λοιπόν,

$$x \sim Bin(m, \pi)$$

$$p(x | m, \pi) = \binom{m}{x} \pi^x (1 - \pi)^{m-x} 1_{\{0,1,\dots,m\}}(x) \quad (1)$$

Μια *διωνυμική διαδικασία* είναι μια ακολουθία από ανεξάρτητες τ.μ. (x_1, \dots, x_i, \dots) διωνυμικού νόμου παραμέτρων (m_i, π) όπου τα m_i είναι σταθερά. Αυτός ο τύπος διαδικασίας συναντάται π.χ. στον ποιοτικό έλεγχο, όπου κάποιος παίρνει, κατά τη διάρκεια της παραγωγικής διαδικασίας, σε διαφορετικές στιγμές, m_i αντικείμενα και λαμβάνει υπόψη τον αριθμό x_i των μη-ελαττωματικών αντικειμένων.

Η διαδικασία Bernoulli αποτελεί λοιπόν μερική περίπτωση της διωνυμικής διαδικασίας, όταν $m_i = 1$, για κάθε i . Αυτή αποτελεί ιδιαίτερα περίπτωση πολύ χρήσιμη, π.χ. στις δειγματοληψίες γνώμης. Σε μια διαδικασία Bernoulli, κάθε x_i είναι μια δείτρια συνάρτηση ($x_i = 1$ όταν έχουμε «επιτυχία» και $x_i = 0$ αν έχουμε «αποτυχία») και βεβαίως, π είναι η πιθανότητα επιτυχίας. Εδώ,

$$p(x_i | \pi) = \pi^{x_i} (1 - \pi)^{1-x_i} 1_{\{0,1\}}(x_i) \quad (2)$$

Ένα τυχαίο δείγμα μεγέθους n μιας διωνυμικής διαδικασίας είναι λοιπόν η ακολουθία των ανεξάρτητων μεταβλητών (x_1, x_2, \dots, x_n) της οποίας η πιθανοφάνεια δίδεται από τη σχέση :

$$\begin{aligned} L(\pi, m_1, x_1, \dots, m_n, x_n) &= \prod_{j=1}^n p(x_j | m_j, \pi) \\ &= \prod_{j=1}^n \binom{m_j}{x_j} \pi^{x_j} (1-\pi)^{m_j-x_j} 1_{\{0,1,\dots,m_j\}}(x_j) \\ &= \left(\prod_{j=1}^n \binom{m_j}{x_j} \right) \pi^{\sum_{j=1}^n x_j} (1-\pi)^{\sum_{j=1}^n m_j - \sum_{j=1}^n x_j} 1_{\prod_{j=1}^n \{0,1,\dots,m_j\}}(x_1, \dots, x_n). \end{aligned}$$

Θέτουμε $s = \sum_{j=1}^n x_j$ και $n_b = \sum_{j=1}^n m_j$, οπότε η τελευταία σχέση γράφεται πιο απλά :

$$L(\pi, m_1, x_1, \dots, m_n, x_n) = \left(\prod_{j=1}^n \binom{m_j}{x_j} \right) \pi^s (1-\pi)^{n_b-s} 1_{\prod_{j=1}^n \{0,1,\dots,m_j\}}(x_1, \dots, x_n) \quad (3)$$

Επειδή ο συνολικός αριθμός Bernoulli, n_b , είναι δεδομένος, αν στη σχέση (3) θέσουμε

$$g(s, \pi) = \pi^s (1-\pi)^{n_b-s} \quad \text{και} \quad h(x_1, \dots, x_n) = \left(\prod_{j=1}^n \binom{m_j}{x_j} \right) 1_{\prod_{j=1}^n \{0,1,\dots,m_j\}}(x_1, \dots, x_n)$$

τότε από το Παραγοντικό Θεώρημα των Fisher και Neymann, η στατιστική συνάρτηση s είναι επαρκής για το π . Διαφορετικά, λαμβάνοντας υπόψη μόνο τον παράγοντα που εξαρτάται από την παράμετρο π , παίρνουμε :

$$L(\pi, m_1, x_1, \dots, m_n, x_n) \propto \pi^s (1-\pi)^{n_b-s} \quad (4)$$

Από τη Στατιστική Συμπερασματολογία (βλέπε π.χ. στο [15]), κάνοντας χρήση π.χ της μεθόδου μεγίστης πιθανοφάνειας, ο κλασσικός εκτιμητής του π είναι ο $\hat{\pi} = \frac{s}{n_b}$.

Επίσης, είναι γνωστό ότι η μέση τιμή του $\hat{\pi}$ και η διασπορά του, δίνονται από τις σχέσεις

$$\begin{aligned} E(\hat{\pi} | n_b, \pi) &= \pi \\ \text{Var}(\hat{\pi} | n_b, \pi) &= \frac{\pi(1-\pi)}{n_b} \end{aligned} \quad (5)$$

(αφού $s = \sum_{j=1}^n x_j \sim \text{Bin}(n_b, \pi)$).

Στη μερική περίπτωση της διαδικασίας Bernoulli ($m_i = 1, \forall i$) s είναι ο αριθμός των επιτυχιών σε $n_b = n$ ανεξάρτητες δοκιμές και ξαναβρίσκουμε τις γνωστές ιδιότητες του $\hat{\pi}$, αναλογία των παρατηρούμενων επιτυχιών. Στην πιο γενική περίπτωση ($m_i \geq 1$), δεδομένης της ανεξαρτησίας των m_i δοκιμών σε έκαστο των διωνυμικών πειραμάτων και της ανεξαρτησίας μεταξύ των διαφόρων παρατηρήσεων x_i , ξαναβρίσκουμε το ίδιο αποτέλεσμα επειδή στην πράξη θα έχουμε $n_b = \sum_{i=1}^n m_i$ ανεξάρτητες παρατηρήσεις μιας διαδικασίας Bernoulli παραμέτρου π και $s = \sum_{i=1}^n x_i$ είναι ο συνολικός αριθμός των παρατηρούμενων επιτυχιών.

1.2 Φυσική συζυγής

Η έννοια της «φυσικής συζυγούς» οικογένειας σε μια δειγματοληπτική διαδικασία έχει ορισθεί στο κεφάλαιο I. Σημειώνουμε εδώ ότι μια οικογένεια «φυσική συζυγής» είναι μια οικογένεια νόμων συνδεδεμένη με τη μελετούμενη διαδικασία και η οποία είναι ιδιαίτερος ενδιαφέρουσα στα παραμετρικά μοντέλα όπου επαρκείς στατιστικές συναρτήσεις μπορούν να ορισθούν (βλέπε I, §2). Πρόκειται για ένα σύνολο νόμων το οποίο από το ένα μέρος είναι αρκετά πλούσιο και εύκαμπτο για να παραστήσει την *εκ των προτέρων πληροφορία* την οποία κάποιος έχει επί των παραμέτρων του μοντέλου και από το άλλο μέρος αυτοί οι νόμοι συνδυάζονται αξιόλογα με τη συνάρτηση πιθανοφάνειας στον τύπο του Bayes για να δώσει ένα *εκ των υστέρων νόμο* επί των παραμέτρων του μοντέλου, ο οποίος ανήκει στην ίδια οικογένεια. Θα δούμε ότι η εκλογή αυτής της οικογένειας συμπεραίνεται από τον πυρήνα της πιθανοφάνειας θεωρώντας τον σαν συνάρτηση των παραμέτρων.

Στη συγκεκριμένη περίπτωση, από τη σχέση (4), παρατηρούμε τον πυρήνα της πιθανοφάνειας συναρτήσει του π . Αναγνωρίζουμε σαν πυρήνα της φυσικής συζυγούς οικογένειας, τον πυρήνα μιας Βήτα πυκνότητας.

α) Εκ των προτέρων

Ένας νόμος Βήτα υποδεικνύεται από δύο παραμέτρους, έτσι αν εκλέξουμε *εκ των προτέρων* μια πυκνότητα Βήτα, θα έχουμε :

$$p(\pi) \propto \pi^{s_0-1} (1-\pi)^{n_0-s_0-1} \quad , \quad s_0 < n_0, \quad s_0 > 0 \quad (6)$$

$$\pi \sim \text{Beta}(s_0, n_0 - s_0)$$

Έχουμε λοιπόν (Παράρτημα 3) τις ακόλουθες *εκ των προτέρων* ροπές

$$E(\pi) = \frac{s_0}{n_0} \quad (7)$$

$$Var(\pi) = \frac{s_0(n_0 - s_0)}{n_0^2(n_0 + 1)}.$$

Η επιλογή ιδιαιτέρως των τιμών των n_0 και s_0 εξαρτάται από την πληροφορία την οποία εκ των προτέρων έχουμε επί της παραμέτρου π . Όπως αναφέραμε στο κεφάλαιο I, §2 μπορούμε να ερμηνεύσουμε τις τιμές του (n_0, s_0) σαν προερχόμενες από ένα υποθετικό ισοδύναμο δείγμα από n_0 δοκιμές Bernoulli χορηγώντας s_0 «επιτυχίες». Θα δούμε δε στη συνέχεια ότι η τιμή n_0 παριστά το βάρος της *εκ των προτέρων πληροφορίας* και το s_0 ισοσταθμίζει την *εκ των προτέρων μαθηματική ελπίδα* του π .

Αξίζει να σημειώσουμε ότι η επιλογή του $s_0 = 1$ και $n_0 = 2$ αντιστοιχεί σε μια ομοιόμορφη κατανομή επί του $[0,1]$, η οποία αποτελεί μια εκ των προτέρων μη-πληροφορική (περίπτωση της αναλλοιώτου). Ενώ η περίπτωση $s_0 = 1/2$ και $n_0 = 1$ είναι μια ακόμη δυνατή εκλογή μιας εκ των προτέρων μη – πληροφορικής προκύπτουσα από τη μέθοδο του Jeffrey.

Στο παρακάτω γράφημα 1.1 θα παραστήσουμε τις πυκνότητες $Beta(s_0, n_0 - s_0)$ για διαφορετικές τιμές των n_0 και s_0 κάθε φορά. Συγκεκριμένα :

- για $n_0 = 2s_0$ και $s_0 = 1, 10$ προκύπτουν δύο καμπύλες των $Beta(1,1) = U(0,1)$ (μη – πληροφορική) και $Beta(10,10)$ με

$$p(\pi) \propto \pi^{s_0-1}(1-\pi)^{s_0-1} \text{ και } E(\pi) = 1/2,$$

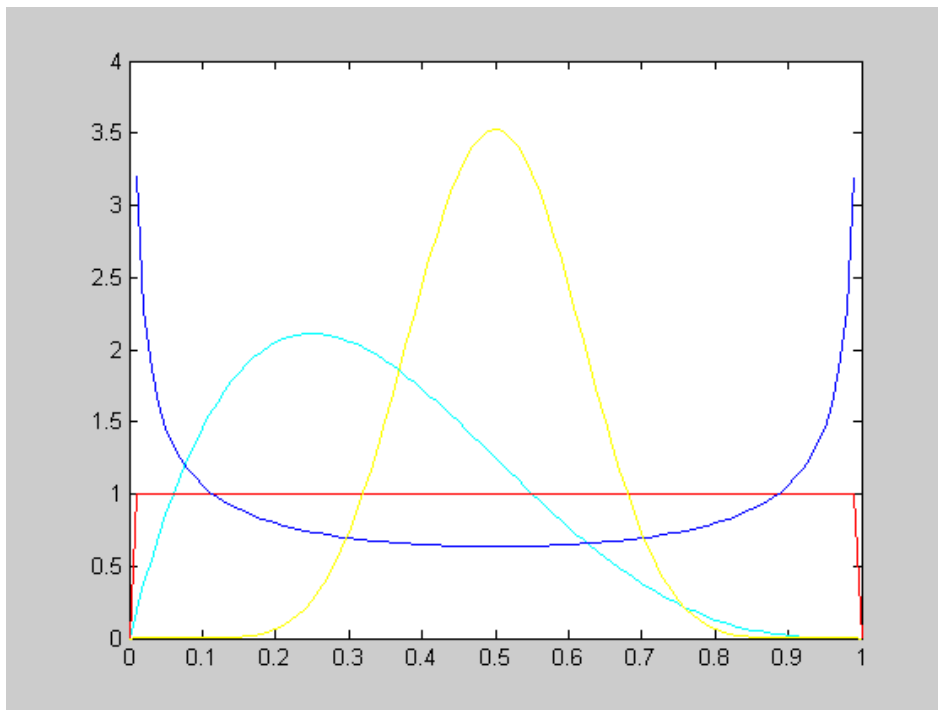
- για $n_0 = 3s_0$ και $s_0 = 2$ προκύπτει μια καμπύλη της $Beta(2,4)$ με

$$p(\pi) \propto \pi^{s_0-1}(1-\pi)^{2s_0-1} \text{ και } E(\pi) = 1/3,$$

- για $s_0 = 1/2$ και $n_0 = 1$ προκύπτει ακόμη ένα γράφημα της $Beta(1/2,1/2)$ (που είναι μη – πληροφορική) με

$$p(\pi) \propto \pi^{1/2-1}(1-\pi)^{1/2-1} \text{ και } E(\pi) = 1/2.$$

Γράφημα 1.1



Οι πυκνότητες $Beta(2,4)$ (γαλάζιο) , $Beta(1,1)$ (κόκκινο) , $Beta(1/2,1/2)$ (κυανούν) και $Beta(10,10)$ (κίτρινο).

b) Εκ των υστέρων

Για να πάρουμε τον εκ των υστέρων νόμο συνδυάζουμε τις σχέσεις (4) και (6) από τις οποίες έχουμε

$$p(\pi | s, n_b) \propto \pi^{s_0+s-1} (1-\pi)^{n_b+n_0-(s+s_0)-1} \quad (8)$$

Πράγματι, έχουμε :

$$\begin{aligned} p(\pi | s, n_b) &\propto L(\pi, m_1, x_1, \dots, m_n, x_n) p(\pi) \propto \pi^s (1-\pi)^{n_b-s} \pi^{s_0-1} (1-\pi)^{n_0-s_0-1} \\ &= \pi^{s_0+s-1} (1-\pi)^{n_b+n_0-(s+s_0)-1} \end{aligned}$$

Αναγνωρίζουμε τον πυρήνα μιας Βήτα πυκνότητας :

$$\pi | x_1, \dots, x_n, n_b \sim Beta(s^*, n^* - s^*), \text{ με } s^* = s + s_0 \text{ και } n^* = n + n_0 \quad (9)$$

Βλέπουμε λοιπόν το ενδιαφέρον της φυσικής συζυγούς οικογένειας, ο εκ των υστέρων νόμος είναι της ίδιας οικογένειας με τον εκ των προτέρων νόμο, μόνο οι παράμετροι του νόμου έχουν αλλάξει. Δηλαδή από τις (n_0, s_0) περνάμε στις (n^*, s^*) μέσω των απλών γραμμικών σχέσεων που φαίνονται στην (9). Αυτό εδώ διευκολύνει τη σύγκριση των εκ των προτέρων και εκ των υστέρων νόμων. Είναι π.χ. ενδιαφέρον το να αποσυνθέσουμε την εκ των υστέρων μέση τιμή του π για να δείξουμε το βάρος που αντιστοιχεί στην εκ των προτέρων πληροφορία και στην πληροφορία που προέρχεται από το δείγμα :

$$\begin{aligned} E(\pi | x_1, \dots, x_n, n_b) &= \frac{s^*}{n^*} = \frac{s + s_0}{n_b + n_0} = \frac{n_b \left(\frac{s}{n_b} \right) + n_0 \left(\frac{s_0}{n_0} \right)}{n_b + n_0} \\ &= \frac{n_b}{n_b + n_0} \hat{\pi} + \frac{n_0}{n_b + n_0} E(\pi) \end{aligned} \quad (10)$$

Η (10) επιτρέπει να ερμηνεύσουμε καλύτερα την εκλογή των τιμών (n_0, s_0) σε όρους ισοδύναμου υποθετικού δείγματος. Παρατηρούμε επίσης ότι, όταν το μέγεθος του δείγματος αυξάνει, $n \rightarrow +\infty$, και λοιπόν $n_b \rightarrow +\infty$, ξαναβρίσκουμε τον συνήθη κλασσικό εκτιμητή $\hat{\pi}$ σαν την εκ των υστέρων μέση τιμή του π .

1.3 Predictive νόμος

Ο predictive (εκ των προτέρων) νόμος λαμβάνεται αναζητώντας τον περιθωριακό νόμο του x , αριθμός επιτυχιών επί m δοκιμών (όπου m σταθερό). Αυτός λαμβάνεται ολοκληρώνοντας τον από κοινού νόμο του (x, π) επί όλων των δυνατών τιμών του π . Είναι φανερό από τη σχέση (6) ότι, ο εκ των προτέρων νόμος του π δεν εξαρτάται από το m , τυπικά λοιπόν μπορούμε να γράψουμε ότι $p(\pi | m) = p(\pi)$ αφού το m δεν θεωρείται, όπως εδώ, τυχαίο. Οπότε :

$$\begin{aligned} p(x | m) &= \int_0^1 p(x | m, \pi) p(\pi | m) d\pi \\ &= \int_0^1 p(x | m, \pi) p(\pi) d\pi = \int_0^1 f_b(x | m, \pi) f_\beta(\pi | s_0, n_0 - s_0) d\pi \\ &= \int_0^1 \binom{m}{x} \pi^x (1 - \pi)^{m-x} \frac{1}{B(s_0, n_0 - s_0)} \pi^{s_0-1} (1 - \pi)^{n_0-s_0-1} d\pi \\ &= \frac{\binom{m}{x}}{B(s_0, n_0 - s_0)} \int_0^1 \pi^{x+s_0-1} (1 - \pi)^{m-x+n_0-s_0-1} d\pi \end{aligned}$$

$$\begin{aligned}
&= \frac{\binom{m}{x}}{B(s_0, n_0 - s_0)} B(x + s_0, m - x + n_0 - s_0) \\
&= \frac{m!}{x!(m-x)!} \frac{\Gamma(n_0)}{\Gamma(s_0)\Gamma(n_0 - s_0)} \frac{\Gamma(s_0 + x)\Gamma(m - x + n_0 - s_0)}{\Gamma(m + n_0)} \\
&= \frac{\Gamma(m+1)}{\Gamma(x+1)\Gamma(m-x+1)} \frac{\Gamma(n_0)}{\Gamma(s_0)\Gamma(n_0 - s_0)} \frac{\Gamma(s_0 + x)\Gamma(m - x + n_0 - s_0)}{\Gamma(m + n_0)} \\
&= \frac{\Gamma(x + s_0)\Gamma(m - x + n_0 - s_0)}{\Gamma(x+1)\Gamma(m-x+1)} \frac{\Gamma(m+1)\Gamma(n_0)}{\Gamma(m+n_0)\Gamma(s_0)\Gamma(n_0 - s_0)},
\end{aligned}$$

όπου $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ για $a, b > 0, n_0 > s_0$ και $s_0 > 0$ για να υπάρχει το ολοκλήρωμα της συνάρτησης Γάμμα. Άρα παίρνουμε ένα νόμο Beta – Binomial (βλέπε Παράρτημα 3.1) :

$$\begin{aligned}
p(x | m) &= f_{\beta b}(x | m, s_0, n_0 - s_0) \\
x | m &\sim BeBin(m, s_0, n_0 - s_0)
\end{aligned} \tag{11}$$

Ο εκ των υστέρων *predictive* λαμβάνεται κατά τον ίδιο τρόπο, αρκεί στη θέση του ζεύγους (n_0, s_0) να βάλουμε το ζεύγος (n^*, s^*) με τον περιορισμό φυσικά το $s^* > 0$ και $n^* > s^*$ για να υπάρχουν τα επιμέρους ολοκληρώματα. Δηλαδή

$$p(x | s, n_b, m) = \int_0^1 p(x | m, \pi) p(\pi | s, n_b) d\pi$$

και ακολουθώντας πιστά τα παραπάνω βήματα καταλήγουμε στο :

$$x | s, n_b, m \sim BeBin(m, s^*, n^* - s^*).$$

Μπορούμε επίσης να υπολογίσουμε την *predictive* μέση τιμή του x έχοντας ήδη παρατηρήσει ένα δείγμα (x_1, \dots, x_n) . Αυτή λοιπόν θα είναι η μέση τιμή του αριθμού των επιτυχιών μέσα σε ένα μελλοντικό διωνυμικό πείραμα παραμέτρων (m, π) :

$$E(x | s, n_b, m) = m \frac{s^*}{n^*}.$$

Αν η εκ των προτέρων επί του π είναι η ομοιόμορφη στο $[0,1]$, τότε η αναμενόμενη τιμή του αριθμού των επιτυχιών x επί m καινούριων δοκιμών σε ένα δείγμα Bernoulli ($m_i = 1$, για κάθε i) μεγέθους n , είναι

$$E(x | s, n, m) = m \frac{s+1}{n+2},$$

(εδώ $s_0 = 1$ και $n_0 = 2$, οπότε από τη σχέση (9) : $s^* = s + s_0 = s + 1$ και $n^* = n + n_0 = n + 2$).

Εφαρμογή: Θεωρούμε ένα πρόβλημα ελέγχου της ποιότητας όπου ένα μέρος μεγάλου αριθμού τεμαχίων (ή αντικειμένων) είναι για ανάλυση: ενδιαφερόμαστε για την αναλογία π των ελαττωματικών αντικειμένων μέσα σ' αυτό το μέρος. Γνωρίζουμε τον προμηθευτή ο οποίος κρίνει ότι το π δεν είναι μακριά από την προτεινόμενη στάθμη 0.10.

Για να προσαρμόσουμε τη γνώμη του στο προτεινόμενο «μέρος», παίρνουμε ένα τυχαίο δείγμα n αντικειμένων από αυτό το μέρος και παρατηρούμε το συνολικό αριθμό y των ελαττωματικών αντικειμένων. Είναι γνωστό από την Θεωρία των Πιθανοτήτων, ότι η y είναι μια διωνυμική τ.μ. με παραμέτρους (n, π) . Λοιπόν δέχεται π.π.

$$p(y | n, \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} 1_{\{0,1,\dots,n\}}(y)$$

Θεωρώντας αυτή σαν μια συνάρτηση του π , η μέθοδος «φυσική συζυγής» επιτρέπει η αβεβαιότητα επί του π να μοντελοποιηθεί από μία $Beta(y_0, n_0 - y_0)$ με $y_0 = 1$ και $n_0 = 11$ ($E(\pi) = \frac{y_0}{n_0} \approx 0,09$). Ο εκ των υστέρων νόμος του π δεδομένου του y , θα είναι:

$$\pi | y \sim Beta(y^*, n^* - y^*)$$

όπου $n^* = n + n_0$ και $y^* = y + y_0$, και

$$E(\pi | y) = \frac{y^*}{n^*} = \frac{y + y_0}{n + n_0} = \frac{n_0}{n + n_0} E(\pi) + \frac{n}{n + n_0} \hat{\pi}, \quad \text{όπου } \hat{\pi} = \frac{y}{n}$$

Σ' αυτόν τον κυρτό συνδιασμό βλέπουμε τα βάρη σχετικά με την εκ των προτέρων πληροφορία που μετριέται από το n_0 και με την πληροφορία του δείγματος που μετριέται από το n . Ιδιαίτερος αν $n \rightarrow +\infty$, $E(\pi | y) \rightarrow \hat{\pi}$, δηλαδή ξαναβρίσκουμε το συνήθη εκτιμητή του π .

Το γράφημα 1.2 δείχνει την επιρροή του μεγέθους του δείγματος επί της εκ των υστέρων πυκνότητας για τρεις διαφορετικές τιμές του n : $n = 5, 10, 50$ και αυτό για μια παρατηρούμενη αναλογία ελαττωματικών αντικειμένων σταθερά στο $y/n = 0.20$ και για την εκ των προτέρων $Beta(y_0, n_0 - y_0)$ εκλεγόμενη ανωτέρω με $n_0 = 11$ και $y_0 = 1$.

Διαπιστώνουμε ότι : όταν το n αυξάνει η εκ των υστέρων κατανομή του π συγκεντρώνεται προοδευτικά γύρω από τον κλασσικό εκτιμητή $\hat{\pi} = \frac{y}{n} = 0,20$. Αν το

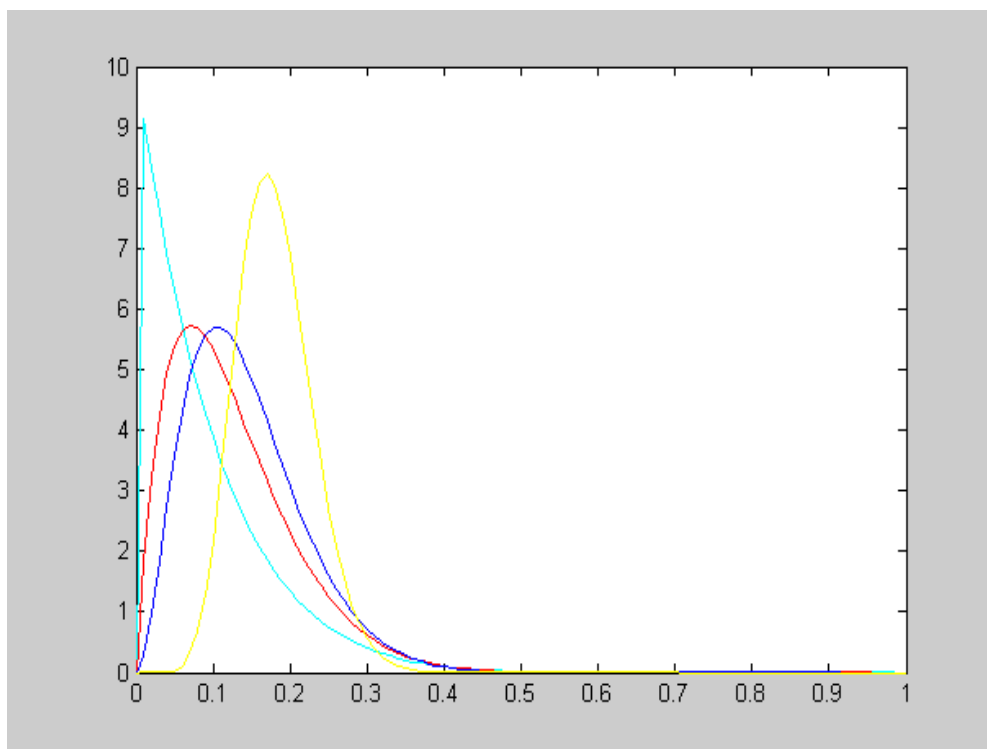
μέγεθος του δείγματος είναι υψηλό, ο ρόλος της εκ των προτέρων κατανομής γίνεται αμελητέος, καθώς η $E(\pi|y)$ εκφράζει τον τρόπο δια του οποίου οι δύο τύποι πληροφορίας συνδυάζονται.

Για το γράφημα 1.2 :

για $y_0 = 1$ και $n_0 = 11$, έχουμε την εκ των προτέρων πυκνότητα του π να είναι η $Beta(1,10)$, ενώ η εκ των υστέρων για $y/n = 0.20$, είναι :

- για $n = 5$: η $Beta(2, 14)$
- για $n = 10$: η $Beta(3, 18)$
- για $n = 50$: η $Beta(11, 50)$

Γράφημα 1.2



Εκ των προτέρων πυκνότητα του π (γαλάζιο) για $y_0 = 1$ και $n_0 = 11$,
 η εκ των υστέρων για $y/n = 0.20$ όταν $n = 5$ (κόκκινο),
 $n = 10$ (κυανού) και $n = 50$ (κίτρινο)

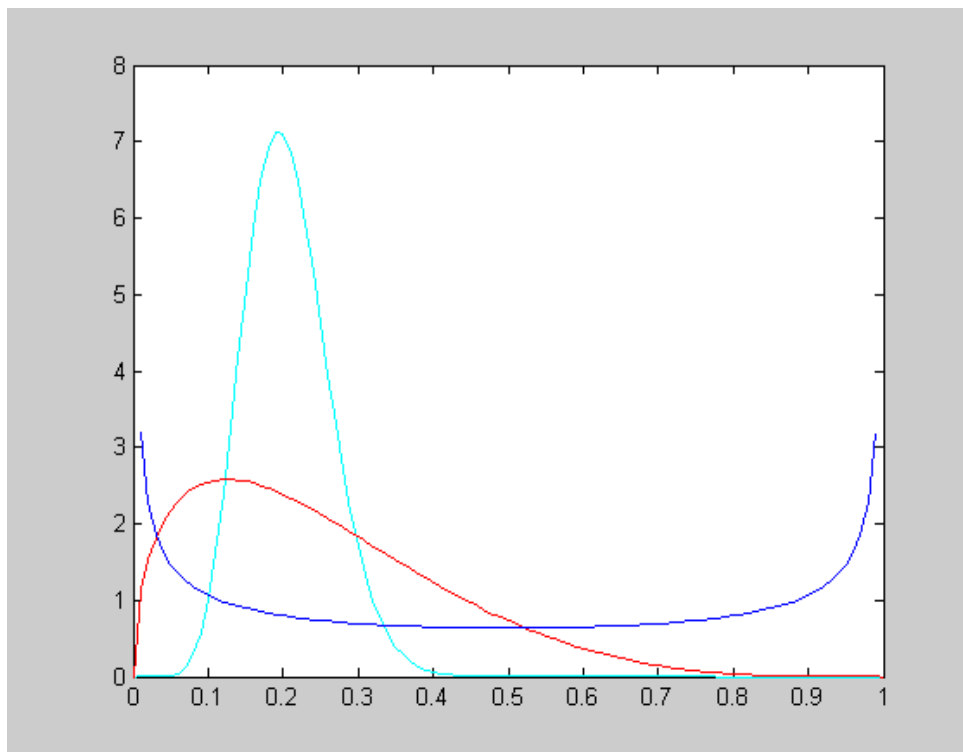
Τα επόμενα γραφήματα 1.3 ,1.4 και 1.5 επιτρέπουν ακόμα το να εκτιμήσουμε την ευαισθησία των αποτελεσμάτων στις διάφορες επιλογές των y_0 , n_0 για την εκ των προτέρων καθώς και το ρόλο του δειγματικού μεγέθους.

Για το γράφημα 1.3 :

για $y_0 = 1/2$ και $n_0 = 1$, έχουμε την εκ των προτέρων πυκνότητα του π να είναι η $Beta(1/2, 1/2)$, ενώ η εκ των υστέρων για $y/n = 0.20$, είναι :

- για $n = 5$: η $Beta(3/2, 9/2)$
- για $n = 50$: η $Beta(21/2, 81/2)$.

Γράφημα 1.3



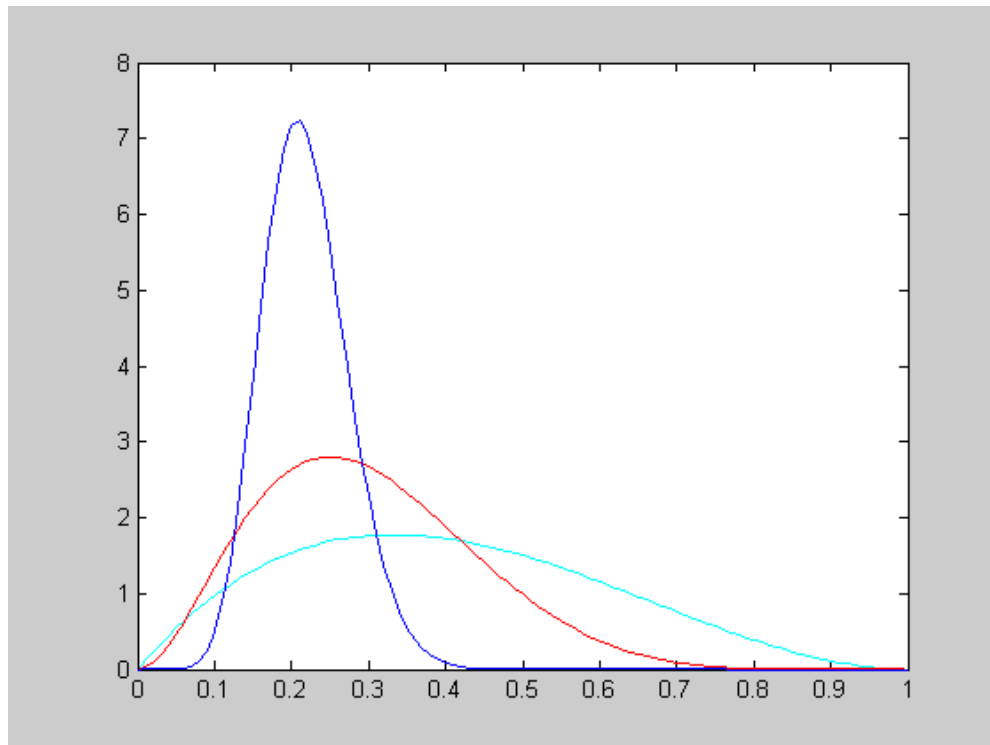
*Εκ των προτέρων πυκνότητα του π (κυανούν) για $y_0 = 1/2$ και $n_0 = 1$,
η εκ των υστέρων για $y/n = 0.20$ όταν $n = 5$ (κόκκινο) και $n = 50$ (γαλάζιο)*

Για το γράφημα 1.4 :

για $y_0 = 2$ και $n_0 = 5$, έχουμε την εκ των προτέρων πυκνότητα του π να είναι η $Beta(2, 3)$, ενώ η εκ των υστέρων για $y/n = 0.20$ είναι :

- για $n = 5$: η $Beta(3, 7)$
- για $n = 50$: η $Beta(12, 43)$.

Γράφημα 1.4



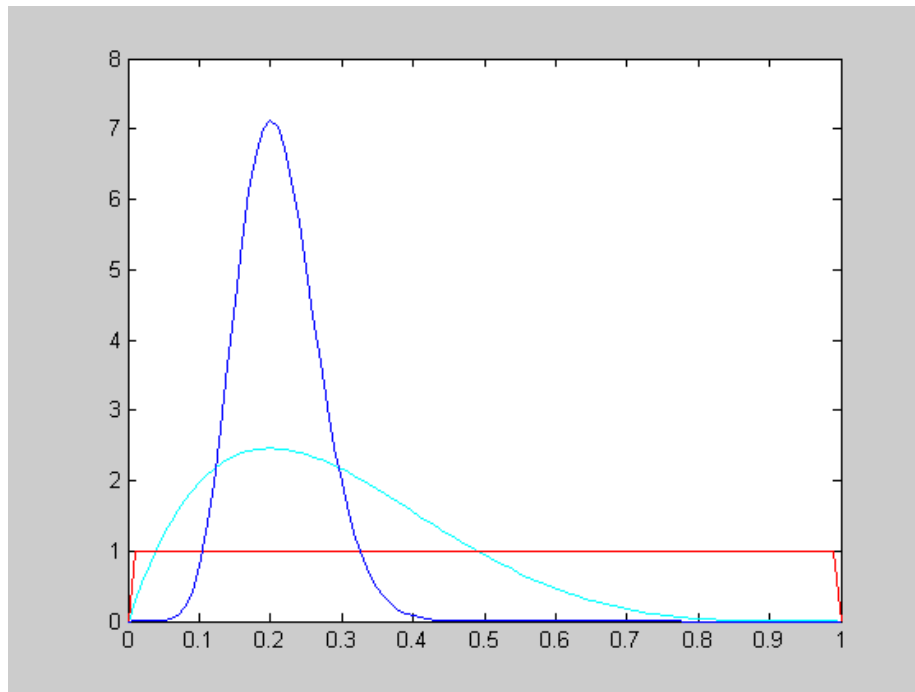
Εκ των προτέρων πυκνότητα του π (γαλάζιο) για $y_0 = 2$ και $n_0 = 5$,
 η εκ των υστέρων για $y/n = 0.20$ όταν $n = 5$ (κόκκινο) και $n = 50$ (κυανό)

Για το γράφημα 1.5 :

για $y_0 = 1$ και $n_0 = 2$, έχουμε την εκ των προτέρων πυκνότητα του π να είναι η $Beta(1,1)$, ενώ η εκ των υστέρων για $y/n = 0.20$ είναι :

- για $n = 5$: η $Beta(2, 5)$
- για $n = 50$: η $Beta(11, 41)$

Γράφημα 1.5



*Εκ των προτέρων πυκνότητα του π (κόκκινο) για $y_0 = 1$ και $n_0 = 2$,
η εκ των υστέρων για $y/n = 0.20$ όταν $n = 5$ (γαλάζιο) και $n = 50$ (κυανούν)*

Αυτά επιβεβαιώνουν ότι για ένα ίδιο μέτριο δειγματικό μέγεθος ($n = 50$) ο ρόλος των εκ των προτέρων είναι αμελητέος. Σε όλα αυτά τα γραφήματα και για τις διάφορες προτεινόμενες επιλογές για τα y_0 , n_0 , η εκ των υστέρων πυκνότητα συγκεντρώνεται γύρω από το y/n όταν το n αυξάνει.

2. Πολυωνυμική δειγματοληψία

Στη στατιστική ανάλυση υπάρχουν πολλά φαινόμενα όπου η ενδιαφέρουσα χαρακτηριστική μπορεί να πάρει ένα πεπερασμένο και σταθερό αριθμό διαφορετικών τιμών. Αυτό καλύπτει την περίπτωση των *ονομαστικών χαρακτηριστικών*, όπου υπάρχει ένας σταθερός αριθμός από δυνατές τιμές (modalites). Παράδειγμα, σε μια δειγματοληψία γνώμης, για μια τιθέμενη ερώτηση, υπάρχουν τρεις δυνατές απαντήσεις: «υπέρ», «κατά», «αδιάφορη». Σε άλλες περιπτώσεις, δυνάμεθα να μελετήσουμε τη συμπεριφορά μιας πραγματικής διακριτής μεταβλητής, όπως τα δυνατά αποτελέσματα σε μια ρίψη ενός ζαριού (6 δυνατά αποτελέσματα). Συμβαίνει επίσης στη μελέτη ενός φαινομένου μοντελοποιημένου υπό μια συνεχή τυχαία μεταβλητή x , να ενδιαφερόμεθα μοναδικά με τις πιθανότητες ώστε η ενδιαφέρουσα μεταβλητή να ανήκει σε μια των κλάσεων μιας δεδομένης διαμέρισης των δυνατών τιμών της x (τάξεις μισθών,...).

Όλα αυτά τα παραδείγματα εμπίπτουν στο πλαίσιο του πολυωνυμικού μοντέλου. Είναι λοιπόν ένα μοντέλο ιδιαίτερα χρήσιμο. Να σημειώσουμε ακόμη ότι αυτό το

μοντέλο αποτελεί το βασικό εργαλείο για την bayesienne ανάλυση μη – παραμετρικών μοντέλων.

2.1 Ορισμός

Σε όλες τις περιπτώσεις επικαλούμενες ανωτέρω, έχουμε λοιπόν μια ακολουθία γεγονότων A_1, \dots, A_{k+1} παριστάνοντας τα δυνατά αποτελέσματα ενός πειράματος τύχης. Οι παράμετροι του μοντέλου είναι οι πιθανότητες π_j πραγματοποίησης των A_j :

$$\begin{aligned} \pi_j &= \Pr(A_j) : j = 1, 2, \dots, k+1 \\ \pi_j &\geq 0 : \sum_{j=1}^{k+1} \pi_j = 1 \end{aligned} \quad (12)$$

Υπάρχουν λοιπόν k -παράμετροι $\pi_j, j = 1, 2, \dots, k$, επειδή $\pi_{k+1} = 1 - \sum_{j=1}^k \pi_j$, αυτοί οι k – παράμετροι ανήκουν στο χωρίο $S_k \subset R^k$:

$$S_k = \left\{ (\pi_1, \dots, \pi_k) \in R^k : \pi_j \geq 0, j = 1, \dots, k, \sum_{j=1}^k \pi_j \leq 1 \right\}$$

Το πολυωνυμικό δείγμα συνίσταται από μια ακολουθία n ανεξάρτητων παρατηρήσεων αυτής της διαδικασίας, στην οποία λαμβάνεται υπόψη ο αριθμός n_j των φορών πραγματοποίησης εκάστου γεγονότος $A_j, j = 1, \dots, k$. Έχουμε λοιπόν :

$$\begin{aligned} n_j &= \text{αριθμός πραγματοποιήσεων του } A_j, j = 1, \dots, k, k+1 \\ n_j &\geq 0 \text{ και } \sum_{j=1}^{k+1} n_j = n \end{aligned} \quad (13)$$

Οι στοχαστικές ιδιότητες αυτού του μοντέλου δύναται να χαρακτηρίζονται από τις k τυχαίες μεταβλητές $n_j, j = 1, \dots, k$, επειδή

$$n_{k+1} = n - \sum_{j=1}^k n_j .$$

Η πιθανοφάνεια του δείγματος δίδεται από το διακριτό πολυωνυμικό νόμο :

$$p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = \frac{n!}{n_1! n_2! \dots n_k! (n - n_1 - \dots - n_k)!} \pi_1^{n_1} \dots \pi_k^{n_k} (1 - \sum_{j=1}^k \pi_j)^{n - \sum_{j=1}^k n_j} \quad (14)$$

$$\forall (n_1, \dots, n_k) \in \left\{ (n_1, \dots, n_k) \in \{0, 1, \dots, n\}^k : \sum_{j=1}^k n_j \leq n \right\}$$

Παρατηρούμε ότι στη μερική περίπτωση όπου $k=1$, η διαδικασία βάσης είναι μια διαδικασία Bernoulli και ξαναβρίσκουμε το διωνυμικό δείγμα της προηγούμενης παραγράφου (με $m_i = 1, \forall i$). Σ' αυτή την περίπτωση ο διακριτός νόμος (14) είναι ο διωνυμικός $f_b(n_1 | n, \pi_1)$.

Κάνοντας χρήση της μεθόδου μέγιστης πιθανοφάνειας καταλήγουμε στο συμπέρασμα ότι ο κλασσικός εκτιμητής του π_j είναι ο $\hat{\pi}_j = \frac{n_j}{n}, j = 1, \dots, k$.

Πράγματι, έχουμε

$$\begin{aligned} \log p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) &= \log \frac{n!}{n_1! n_2! \dots n_k! (n - n_1 - \dots - n_k)!} + \\ &+ \sum_{j=1}^k n_j \log \pi_j + (n - \sum_{j=1}^k n_j) \log(1 - \sum_{j=1}^k \pi_j) \end{aligned}$$

Έτσι, η λογαριθμική εξίσωση πιθανοφάνειας γράφεται

$$\frac{\partial}{\partial \pi_j} \log p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = \frac{n_j}{\pi_j} - \frac{n_{k+1}}{\pi_{k+1}} = 0, \quad \forall j \in \{1, 2, \dots, k\}$$

απ' όπου προκύπτει ότι $\frac{n_1}{\pi_1} = \dots = \frac{n_{k+1}}{\pi_{k+1}} = n$ και λαμβάνουμε $\hat{\pi}_j = \frac{n_j}{n}$.

Ακολούθως, η λύση αυτή μεγιστοποιεί τη συνάρτηση πιθανοφάνειας, αφού

$$\frac{\partial}{\partial \pi_j} \log p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = \frac{n_j}{\pi_j} - \frac{n_{k+1}}{\pi_{k+1}}, \quad \forall j \in \{1, 2, \dots, k\}$$

$$\frac{\partial^2}{\partial \pi_j^2} \log p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = -\frac{n_j}{\pi_j^2}, \quad \forall j \in \{1, 2, \dots, k\}$$

$$\frac{\partial^2}{\partial \pi_j \partial \pi_i} \log p(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = 0, \quad \forall i \neq j \text{ και } i, j \in \{1, 2, \dots, k\}$$

και ο πίνακας των δευτέρων μερικών παραγώγων στο $(\hat{\pi}_1, \dots, \hat{\pi}_k)$ καταλήγει στον εξής :

$$A = \begin{pmatrix} \frac{-n^2}{n_1} & 0 & \dots & 0 \\ 0 & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \frac{-n^2}{n_k} \end{pmatrix} = -n^2 \begin{pmatrix} \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{n_k} \end{pmatrix},$$

ο οποίος είναι αρνητικά ορισμένος πίνακας τύπου (k, k) . Άρα το ${}^t(\hat{\pi}_1, \dots, \hat{\pi}_k)$ μεγιστοποιεί τη συνάρτηση πιθανοφάνειας και επομένως ο εκτιμητής μεγίστης πιθανοφάνειας του ${}^t(\pi_1, \dots, \pi_k)$ είναι το ${}^t(\hat{\pi}_1, \dots, \hat{\pi}_k) = \left(\frac{n_1}{n}, \dots, \frac{n_k}{n}\right)$.

2.1 Φυσική συζυγής

Όπως και στην προηγούμενη παράγραφο, σημειώνουμε την πιθανοφάνεια ως συνάρτηση των παραμέτρων :

$$L(\pi_1, \dots, \pi_k, n_1, \dots, n_k, n) \propto \pi_1^{n_1} \dots \pi_k^{n_k} (1 - \pi_1 - \dots - \pi_k)^{n - \sum_{j=1}^k n_j} \quad (15)$$

Από τον παραπάνω τύπο αναγνωρίζουμε ως συνάρτηση του ${}^t(\pi_1, \dots, \pi_k)$ τον πυρήνα μιας πυκνότητας Dirichlet που δίνεται από τον τύπο :

$$f_{dir}(x_1, \dots, x_k | a_1, \dots, a_{k+1}) = \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1) \dots \Gamma(a_{k+1})} x_1^{a_1-1} \dots x_k^{a_k-1} (1 - x_1 - \dots - x_k)^{a_{k+1}-1} 1_{S_k}(x_1, \dots, x_k)$$

όπου $S_k = \left\{ {}^t(x_1, \dots, x_k) \in R^k : x_j \geq 0, j = 1, \dots, k, \sum_{j=1}^k x_j \leq 1 \right\}$ και $a_j > 0, j = 1, \dots, k+1$

παραμέτροι.

Συνεπώς η φυσική συζυγής οικογένεια μέσα σ' αυτό το μοντέλο είναι μια οικογένεια πυκνοτήτων Dirichlet. Περισσότερες λεπτομέρειες για την κατανομή Dirichlet αναφέρονται στο Παράρτημα 5.2.

α) Εκ των προτέρων

Αν επιλέξουμε ως *εκ των προτέρων* μια πυκνότητα Dirichlet $f_{dir}(\pi_1, \dots, \pi_k | a_1, \dots, a_{k+1})$ έχουμε :

$$p(\pi_1, \dots, \pi_k) \propto \pi_1^{a_1-1} \dots \pi_k^{a_k-1} (1 - \pi_1 - \dots - \pi_k)^{a_{k+1}-1} \quad (16)$$

δηλαδή

$$(\pi_1, \dots, \pi_k) \sim \text{Dir}(\alpha_1, \dots, \alpha_{k+1}).$$

Οι σημαντικές ροπές της, δηλαδή η μέση τιμή, η διασπορά και η συνδιασπορά, είναι οι εξής :

$$E(\pi_i) = \frac{a_i}{a_1 + \dots + a_{k+1}}, \quad i = 1, \dots, k+1$$

$$\text{Var}(\pi_i) = \frac{a_i(a_1 + \dots + a_{k+1} - a_i)}{(a_1 + \dots + a_{k+1})^2(a_1 + \dots + a_{k+1} + 1)}, \quad i = 1, \dots, k+1$$

$$\text{Cov}(\pi_i, \pi_j) = \frac{a_i a_j}{(a_1 + \dots + a_{k+1})^2(a_1 + \dots + a_{k+1} + 1)}, \quad \forall i \neq j.$$

b) Εκ των υστέρων

Συνδιάζοντας στη συνέχεια τις σχέσεις (15) και (16) λαμβάνουμε τον εκ των υστέρων πυρήνα :

$$p(\pi_1, \dots, \pi_k | n_1, \dots, n_k, n) \propto \pi_1^{n_1 + a_1 - 1} \dots \pi_k^{n_k + a_k - 1} (1 - \pi_1 - \dots - \pi_k)^{a_{k+1} + n_{k+1} - 1} \quad (17)$$

όπου $n_{k+1} = n - \sum_{j=1}^k n_j$. Οπότε οδηγούμαστε πάλι σε μια πυκνότητα Dirichlet

$$(\pi_1, \dots, \pi_k) | n, n_1, \dots, n_k \sim \text{Dir}(a_1^*, \dots, a_{k+1}^*),$$

όπου $a_j^* = a_j + n_j$, $j = 1, \dots, k+1$ και $n_{k+1} = n - \sum_{j=1}^k n_j$.

Ιδιαίτερως μπορούμε να υπολογίσουμε τη μέση τιμή του εκ των υστέρων νόμου :

$$E(\pi_i | n, n_1, \dots, n_k) = \frac{a_i^*}{a_1^* + \dots + a_{k+1}^*} = \frac{a_i + n_i}{n + \sum_{j=1}^{k+1} a_j} = \frac{n \binom{n_i}{n} + \sum_{j=1}^{k+1} a_j \frac{a_i}{\sum_{j=1}^{k+1} a_j}}{n + \sum_{j=1}^{k+1} a_j}$$

$$= \frac{n \hat{\pi}_i + \sum_{j=1}^{k+1} a_j E(\pi_i)}{n + \sum_{j=1}^{k+1} a_j}$$

$$= \frac{n}{n + \sum_{j=1}^{k+1} a_j} \hat{\pi}_i + \frac{\sum_{j=1}^{k+1} a_j}{n + \sum_{j=1}^{k+1} a_j} E(\pi_i) \quad (18)$$

Παρατηρούμε λοιπόν και σε αυτή την περίπτωση ότι, η μέση τιμή του εκ των υστέρων νόμου είναι ένας αντισταθμιζόμενος μέσος του κλασσικού εκτιμητή $\hat{\pi}_i = \frac{n_i}{n}$ και της εκ των προτέρων μέσης τιμής του π_i , με βάρη αντιστοίχως το μέγεθος του δείγματος n και το $\sum_{j=1}^{k+1} a_j$. Αυτό μας επιτρέπει να ερμηνεύσουμε τις εκ των προτέρων «παραμέτρους» a_j σε όρους ισοδύναμου υποθετικού δείγματος μεγέθους $\sum_{j=1}^{k+1} a_j$, οι ατομικές τιμές a_j ισοσταθμίζοντας τις εκ των προτέρων μέσες τιμές των π_i , $i = 1, \dots, k+1$.

3. Διαδικασία Poisson

Οι διαδικασίες Poisson είναι ιδιαίτερες χρήσιμες στους περισσότερους τομείς της στατιστικής (π.χ στις ασφαλίσσεις, στη βιοϊατρική και αλλού) κάθε φορά που ενδιαφερόμαστε για τον αριθμό των φορών που ένα «γεγονός» πραγματοποιείται σ' ένα διάστημα χρόνου, μήκους, επιφάνειας ή όγκου κ.α. Υπό ειδικές υποθέσεις μπορούμε πράγματι να αποδείξουμε ότι αυτή η διακριτή μεταβλητή του μετρήματος ακολουθεί ένα νόμο Poisson. Για να απλοποιήσουμε την παρουσίαση αυτής της διαδικασίας θα θεωρήσουμε εδώ ότι ενδιαφερόμαστε για το μέτρημα γεγονότων σε χρονικά διαστήματα.

3.1 Ορισμός

Θεωρούμε ένα διάστημα χρόνου μήκους $h > 0$. Υποθέτουμε ότι :

H₁ : το h είναι αρκετά μικρό, η πιθανότητα να παρατηρήσουμε ένα «γεγονός» επί διαστήματος χρόνου h είναι ανάλογη του h :

$$P(\text{“ ένα γεγονός κατά τη διάρκεια χρόνου } h \text{ ”}) = \lambda h + o(h),$$

όπου το $o(h)$ παριστάνει μια ποσότητα η οποία τείνει προς το μηδέν πιο γρήγορα από το h : $\frac{o(h)}{h} \rightarrow 0$ του $h \rightarrow 0$. Η παράμετρος λ παριστάνει την ένταση της ανέλιξης.

H₂ : Η πιθανότητα να παρατηρήσουμε περισσότερα του ενός «γεγονότα» επί του διαστήματος χρόνου h είναι αμελητέα αν το h είναι μικρό :

P (“δύο ή περισσότερα γεγονότα κατά τη διάρκεια χρόνου h ”) = $o(h)$.

H_3 : Υπάρχει ανεξαρτησία μέσα στο χρόνο δηλαδή τα γεγονότα τα οποία πραγματοποιούνται σε δύο ξένα διαστήματα χρόνου είναι ανεξάρτητα.

Υπό αυτές τις υποθέσεις η τ.μ. x , αριθμός γεγονότων που πραγματοποιούνται σ'ένα διάστημα δεδομένου χρόνου t , ακολουθεί ένα νόμο Poisson παραμέτρου λt :

$$\begin{aligned} x | \lambda, t &\sim Poi(\lambda t) \\ p(x | \lambda, t) &= e^{-\lambda t} \frac{(\lambda t)^x}{x!} 1_{\{0,1,\dots\}}(x) \end{aligned} \quad (19)$$

Η *δειγματοληψία Poisson* συνίσταται, ορισμένες φορές απλά στην παρατήρηση του αριθμού των γεγονότων τα οποία πραγματοποιούνται σ'ένα δεδομένο διάστημα χρόνου t . Σ'αυτή την περίπτωση η πιθανοφάνεια δίνεται αμέσως από την (19), ενώ αν θεωρήσουμε ότι αυτή είναι συνάρτηση μόνο του λ μπορεί να γραφεί :

$$L(\lambda, x, t) \propto e^{-\lambda t} \lambda^x \quad (20)$$

Γενικότερα, δυνάμεθα να παρατηρήσουμε ένα δείγμα μεγέθους n από πραγματοποιήσεις ανεξάρτητων μεταβλητών x_i , που εκφράζουν τον αριθμό των γεγονότων τα οποία πραγματοποιούνται σε σταθερά διαστήματα δεδομένων χρόνων u_i , $i = 1, \dots, n$:

$$x_i | \lambda_i, u_i \sim Poi(\lambda u_i)$$

Στη συγκεκριμένη περίπτωση η πιθανοφάνεια γράφεται:

$$L(\lambda, x_1, \dots, x_n, u_1, \dots, u_n) = \frac{e^{-\lambda \sum_{i=1}^n u_i} \prod_{i=1}^n \lambda^{x_i} u_i^{x_i}}{\prod_{i=1}^n x_i!} \propto \lambda^s e^{-\lambda t} \quad (21)$$

όπου $s = \sum_{i=1}^n x_i$ και $t = \sum_{i=1}^n u_i$.

Παρατηρούμε ότι οι δύο πιθανοφάνειες (20) και (21) είναι «ισοδύναμες». Πράγματι και στις δύο περιπτώσεις η επαρκής στατιστική συνάρτηση είναι ο συνολικός αριθμός των γεγονότων (x ή s) τα οποία πραγματοποιούνται σε συνολικό χρόνο t και ακόμη το s ακολουθεί τον ίδιο δεσμευμένο νόμο με το x , δηλαδή

$$s | \lambda, t \sim Poi(\lambda t),$$

$$\left(\sum_{i=1}^n x_i \sim Poi\left(\lambda \sum_{i=1}^n u_i\right) \right).$$

Κάνοντας χρήση της μεθόδου μέγιστης πιθανοφάνειας καταλήγουμε στο συμπέρασμα ότι ο κλαστικός εκτιμητής του λ είναι ο $\hat{\lambda} = \frac{S}{t}$.

Στην πιο μερική περίπτωση όπου οι παρατηρήσεις x_i γίνονται επί διαστημάτων χρόνου ίσου με την μονάδα χρόνου (π.χ. το έτος, η ημέρα κ.α) έχουμε $u_i = 1, i = 1, \dots, n$ και λοιπόν $t = n$.

Παρατήρηση 1 : Είδαμε παραπάνω ότι στη δειγματοληψία Poisson έχουμε σταθεροποιήσει το διάστημα χρόνου t και παρατηρούμε την πραγματοποίηση του αριθμού (τυχαίου) των γεγονότων x . Θα μπορούσαμε, επίσης να αντιστρέψουμε τους ρόλους : να σταθεροποιήσουμε έναν αριθμό γεγονότων $x = k$ και να παρατηρούμε το χρόνο t (τυχαία) ο οποίος προκύπτει μέχρι της πραγματοποίησης του k -οστού γεγονότος. Μπορούμε πολύ εύκολα να δούμε ότι : $t | \lambda, k \sim \Gamma(k, 1/\lambda)$ και μιλάμε τότε για δειγματοληψία Gamma της ανέλιξης Poisson, δειγματοληψία που θα μελετήσουμε παρακάτω.

3.2 Φυσική συζυγής

Αν θεωρήσουμε τον πυρήνα της πιθανοφάνειας ως συνάρτηση του λ στην σχέση (21) αναγνωρίζουμε αμέσως τον πυρήνα μιας Gamma πυκνότητας. Συνεπώς, η φυσική συζυγής οικογένειας μιας δειγματοληψίας Poisson είναι η οικογένεια της Gamma πυκνότητας.

a) Εκ των προτέρων

Η εκ των προτέρων πυκνότητα του λ μπορεί να γραφεί :

$$p(\lambda) \propto \lambda^{s_0-1} e^{-\lambda t_0},$$

δηλαδή

$$\lambda \sim \Gamma\left(s_0, \frac{1}{t_0}\right) \quad (22)$$

Οι εκ των προτέρων ροπές δίνονται από τις σχέσεις :

$$E(\lambda) = \frac{s_0}{t_0} \quad \text{και} \quad Var(\lambda) = \frac{s_0}{t_0^2}. \quad (23)$$

Θα δούμε ότι η εκ των προτέρων πληροφορία μπορεί να θεωρηθεί ότι προέρχεται από ένα ισοδύναμο υποθετικό δείγμα όπου έχουν παρατηρηθεί s_0 γεγονότα Poisson κατά τη διάρκεια ενός διαστήματος χρόνου t_0 σταθερού.

Ένας τρόπος να παραστήσουμε μια εκ των προτέρων μη-πληροφορική είναι να εκλέξουμε την οριακή περίπτωση $s_0 \rightarrow 0$ και $t_0 \rightarrow 0$ αυτό το οποίο δίνει μια πυκνότητα “improgre” (η οποία δεν έχει ολοκλήρωμα 1) :

$$p(\lambda) \propto \lambda^{-1} \quad (24)$$

Οφείλουμε να πούμε ότι αυτή συνιστάται συχνά στην bayesienne ανάλυση για τις θετικές παραμέτρους ($\lambda > 0$). Παρατηρούμε ότι αυτή η εκλογή συνεπάγεται ένα ομοιόμορφο νόμο «improgre» για τον $\log \lambda$:

$$p(\log \lambda) \propto c^{ste}, \quad \text{με} \quad \log \lambda \in R.$$

b) Εκ των υστέρων

Συνδιάζοντας την πιθανοφάνεια (21) και τον πυρήνα της εκ των προτέρων (22) λαμβάνουμε :

$$p(\lambda | s, t) \propto \lambda^{s+s_0-1} e^{-\lambda(t+t_0)},$$

όπου αναγνωρίζουμε όπως το περιμέναμε (ιδιότητες της φυσικής συζυγούς οικογένειας) τον πυρήνα μιας Gamma πυκνότητας

$$\lambda | s, t \sim \Gamma(s^*, \frac{1}{t^*}) \quad (25)$$

όπου $t^* = t + t_0$ και $s^* = s + s_0$.

Αν τώρα συγκρίνουμε για ακόμη μια φορά την εκ των προτέρων πληροφορία με εκείνη την οποία διαθέτουμε εκ των υστέρων μέσω της εκ των υστέρων μέσης τιμής του λ , βλέπουμε τα βάρη της εκ των προτέρων και του δείγματος :

$$E(\lambda | s, t) = \frac{s + s_0}{t + t_0} = \frac{s^*}{t^*} = \frac{t_0 \frac{s_0}{t_0} + t \frac{s}{t}}{t + t_0} = \frac{t_0}{t + t_0} E(\lambda) + \frac{t}{t + t_0} \hat{\lambda} \quad (26)$$

όπου $\hat{\lambda}$ ο συνήθης εκτιμητής του λ . Αυτή η αποσύνθεση δικαιολογεί την ερμηνεία των «παραμέτρων» της εκ των προτέρων σε όρους ισοδύναμου υποθετικού δείγματος.

Παρατηρούμε επίσης ότι στην περίπτωση ενός εκ των προτέρων νόμου μη-πληροφοριακού, δηλαδή αν $t_0, s_0 \rightarrow 0$, από τη σχέση (26) παίρνουμε $E(\lambda | s, t) = \hat{\lambda}$. Το ίδιο αποτέλεσμα θα πάρουμε αν το μέγεθος του δείγματος n (ή ο συνολικός χρόνος παρατήρησης $t = \sum_{i=1}^n u_i$) τείνει προς το $+\infty$: το βάρος της εκ των προτέρων ελαττώνεται αν το μέγεθος του δείγματος αυξάνει.

3.3 Predictive νόμος

Ο predictive (εκ των προτέρων) νόμος του x , δηλαδή ο αριθμός των γεγονότων τα οποία θα πραγματοποιηθούν σε ένα διάστημα δεδομένου χρόνου u , λαμβάνεται ολοκληρώνοντας τον από κοινού νόμο του $(x, \lambda | u)$ ως προς λ . Όπως έχουμε προαναφέρει το λ είναι η ένταση της ανέλιξης Poisson και η εκ των προτέρων πυκνότητα της που δίνεται από τη σχέση (22) δεν εξαρτάται από το u (το μήκος του θεωρούμενου χρονικού διαστήματος). Τυπικά λοιπόν $p(\lambda | u) = p(\lambda)$ και από το άλλο μέρος θα πρέπει $t_0, s_0 > 0$, ώστε το παρακάτω ολοκλήρωμα να συγκλίνει :

$$\begin{aligned}
 p(x | u) &= \int_0^{+\infty} p(x, \lambda | u) d\lambda = \int_0^{+\infty} p(x | \lambda, u) p(\lambda) d\lambda = \int_0^{+\infty} e^{-\lambda u} \frac{(\lambda u)^x}{x!} \frac{t_0^{s_0}}{\Gamma(s_0)} \lambda^{s_0-1} e^{-\lambda t_0} d\lambda \\
 &= \frac{t_0^{s_0} u^x}{(s_0 - 1)! x!} \int_0^{+\infty} \lambda^{x+s_0-1} e^{-\lambda(u+t_0)} d\lambda = \frac{t_0^{s_0} u^x}{x! (s_0 - 1)! (u+t_0)^{x+s_0}} \Gamma(x+s_0) \\
 &= \frac{t_0^{s_0} u^x}{x! (s_0 - 1)! (u+t_0)^{x+s_0}} \frac{(x+s_0-1)!}{x! (s_0 - 1)!} \left(\frac{t_0}{u+t_0} \right)^{s_0} \left(\frac{u}{u+t_0} \right)^x \\
 &= \binom{x+s_0-1}{s_0-1} \left(\frac{t_0}{u+t_0} \right)^{s_0} \left(1 - \frac{t_0}{u+t_0} \right)^x \tag{27}
 \end{aligned}$$

Έχουμε λοιπόν για αυτό τον predictive, ένα νόμο Αρνητικό – Διωνυμικό (βλέπε Παράρτημα 4)

$$\begin{aligned}
 p(x | u) &= f_{nb}(x | s_0, \frac{t_0}{u+t_0}) \\
 x | u &\sim \text{NegBin}(s_0, \frac{t_0}{u+t_0}) \tag{28}
 \end{aligned}$$

Ο εκ των υστέρων predictive νόμος δηλαδή ο μετά νόμος, έχοντας παρατηρήσει ένα δείγμα περιγραφόμενο από τα (t, s) , λαμβάνεται κατά τον ίδιο τρόπο, αρκεί να αντικαταστήσουμε το ζεύγος (t_0, s_0) από το ζεύγος (t^*, s^*) . Και πάλι εδώ για να υπάρχει το ολοκλήρωμα θα πρέπει $t^*, s^* > 0$:

$$x | u, s, t \sim \text{NegBin}(s^*, \frac{t^*}{u+t^*})$$

Ιδιαίτερος, μπορούμε να υπολογίσουμε τον αναμενόμενο αριθμό γεγονότων επί του διαστήματος δεδομένου χρόνου u : $E(x | u, s, t) = \frac{s^*}{t^*} u$, και στην περίπτωση που έχουμε

μια εκ των προτέρων μη-πληροφορική ($t_0, s_0 \rightarrow 0$) ξαναβρίσκουμε ένα πολύ φυσικό

$$\text{αποτέλεσμα : } E(x | u, s, t) = \frac{s}{t} u = \hat{\lambda} u .$$

4. Διαδικασία Gamma : παράμετρος μορφής σταθερά

Οι Gamme νόμοι επιτρέπουν τη μοντελοποίηση θετικών συνεχών μεταβλητών και λοιπόν μπορούν να είναι χρήσιμοι στη μοντελοποίηση μιας μεταβλητής όπως ο χρόνος (αναμονής ή διάρκειας) που χάνεται προ της πραγματοποίησης ενός ιδιαίτερου φαινομένου. Ήδη, στην προηγούμενη παράγραφο, παρατηρήσαμε ότι ο νόμος Gamma υπεισέρχεται εξίσου μέσα σε μια διαδικασία Poisson σαν χρόνος αναμονής για την πραγματοποίηση του k -οστού γεγονότος του Poisson.

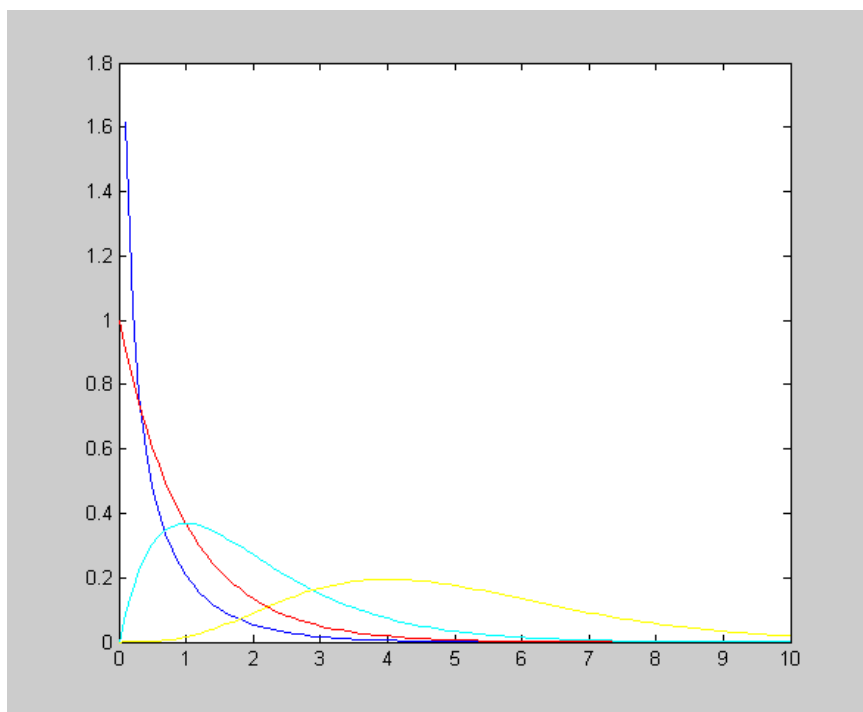
4.1 Ορισμός

Μια μεταβλητή x είναι μια *Gamma μεταβλητή*, αν η πυκνότητα πιθανότητάς της δίδεται από τη σχέση

$$p(x | a, \lambda) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad \forall x > 0 \quad (29)$$

Οι Gamma νόμοι χαρακτηρίζονται, λοιπόν, από δύο παραμέτρους : μια βαθμωτή παράμετρο $\lambda (> 0)$ και μια παράμετρο μορφής $a > 0$. Θα υποθέσουμε εδώ ότι η παράμετρος μορφής a είναι σταθερή. Η πλήρης μεταχείριση των Gamma νόμων είναι δυνατή αλλά καθιστά αναγκαίες τις αριθμητικές μεθόδους (αριθμητική ολοκλήρωση για την παράμετρο μορφής).

Γράφημα 4.1



Μερικές Gamma πυκνότητες παραμέτρων $\lambda = 1$, $a = 0.5$ (γαλάζιο),
 $a = 1$ (κόκκινο), $a = 2$ (κυανούν) και $a = 5$ (κίτρινο)

Το Γράφημα 4.1 δίνει τη μορφή αυτών των πυκνοτήτων για τις διάφορες τιμές του a , ενώ το λ διατηρείται ίσο με 1. Όταν το $a = 1$ λαμβάνουμε τον εκθετικό νόμο.

Η διαδικασία Gamma θα ορίζεται λοιπόν ως μια ακολουθία από τυχαίες μεταβλητές (x_1, \dots, x_i, \dots) ανεξάρτητες και ισόνομες κατά το νόμο (29). Ένα δείγμα μεγέθους n θα χαρακτηρίζεται τότε από την ακολουθία (x_1, \dots, x_n) ανεξάρτητων και ισόνομων τ.μ. Η πιθανοφάνεια θα γράφεται τότε :

$$L(\lambda, a, x_1, \dots, x_n) = \frac{\lambda^{na}}{(\Gamma(a))^n} \left(\prod_{i=1}^n x_i \right)^{a-1} e^{-\lambda \sum_{i=1}^n x_i} = \frac{\lambda^{na}}{(\Gamma(a))^n} t^{a-1} e^{-\lambda s} \quad (30)$$

όπου $(t, s) = \left(\prod_{i=1}^n x_i, \sum_{i=1}^n x_i \right)$ είναι μια επαρκής σ.σ. για το (λ, a) .

Επειδή εδώ το a θεωρείται σταθερό και γνωστό, η s είναι επαρκής για το λ και ο κλασσικός εκτιμητής του λ είναι ο $\hat{\lambda} = \frac{na}{s}$ ($\hat{\lambda}$ εκτιμητής μέγιστης πιθανοφάνειας του λ). Ο πυρήνας πιθανοφάνειας σε αυτή την περίπτωση, όπου το a είναι γνωστό, είναι :

$$L(\lambda, a, s) \propto \lambda^{na} e^{-\lambda s} \quad (31)$$

4.2 Φυσική συζυγής

Η ανάλυση της σχέσης (31), ως συνάρτηση του λ επιτρέπει να ταυτίσουμε την οικογένεια των Gamma πυκνοτήτων με τη φυσική συζυγή της Gamma διαδικασίας όπου το a είναι γνωστό.

a) Εκ των προτέρων

Η εκ των προτέρων πυκνότητα του λ γράφεται ως εξής :

$$p(\lambda) \propto \lambda^{n_0-1} e^{-s_0 \lambda}, \quad \text{δηλαδή } \lambda \sim \Gamma(n_0, \frac{1}{s_0}) \quad (32)$$

Ως συνήθως, οι τιμές (s_0, n_0) εκλέγονται κατά τρόπο ώστε να εκφράζουν την εκ των προτέρων πληροφορία την οποία έχουμε επί του λ . Έτσι έχουμε :

$$\begin{aligned} E(x) &= \frac{n_0}{s_0} \\ \text{Var}(x) &= \frac{n_0}{s_0^2} \end{aligned} \quad (33)$$

Θα δούμε στη συνέχεια ότι οι τιμές (s_0, n_0) μπορούν να ερμηνευθούν σε όρους ενός ισοδύναμου υποθετικού δείγματος. Εδώ επίσης, όπως και για την παράμετρο της ανέλιξης Poisson, μια εκ των προτέρων μη – πληροφορική δύναται να προκύψει από την οριακή περίπτωση (πυκνότητα improper) όπου $s_0, n_0 \rightarrow 0$ και λοιπόν

$$p(\lambda) \propto \lambda^{-1}$$

b) Εκ των υστέρων

Είναι φανερό ότι συνδιάζοντας την εκ των προτέρων (32) και την πιθανοφάνεια (31) λαμβάνουμε για πυρήνα της εκ των υστέρων, έναν Gamma πυρήνα:

$$\begin{aligned} p(\lambda | s, a) &\propto p(s | \lambda, a) p(\lambda) \\ &\propto \lambda^{n_0+na-1} e^{-\lambda(s+s_0)} \end{aligned}$$

ή ακόμη $\lambda | s, a \sim \Gamma(n^*, \frac{1}{s^*}) \quad (34)$

όπου $s^* = s + s_0$ και $n^* = n_0 + na$.

Όπως και στις προηγούμενες περιπτώσεις τα αντίστοιχα βάρη της εκ των προτέρων και του δείγματος φαίνονται μέσα στην παρακάτω ανάλυση του εκ των υστέρων μέσου

$$E(\lambda | s, \alpha) = \frac{n^*}{s^*} = \frac{n_0 + na}{s + s_0} = \frac{s_0 \frac{n_0}{s_0} + s \frac{na}{s}}{s + s_0} = \frac{s_0}{s + s_0} E(\lambda) + \frac{s}{s + s_0} \hat{\lambda} \quad (35)$$

και στην περίπτωση μιας εκ των προτέρων μη – πληροφορικής ξαναβρίσκουμε, για την εκ των υστέρων μέση τιμή, τον κλασσικό εκτιμητή $\hat{\lambda}$.

Πράγματι, αν θεωρήσουμε ότι $p(\lambda) \propto \lambda^{-1}$ έχουμε $p(\lambda | s, a) \propto \lambda^{na-1} e^{-\lambda s}$ δηλαδή $\lambda | s, a \sim \Gamma(na, \frac{1}{s})$. Επομένως $E(\lambda | s, a) = \frac{na}{s} = \hat{\lambda}$.

ς) Συμπέρασμα επί του μέσου του x .

Σε πολλές περιπτώσεις εφαρμογών, η ενδιαφέρουσα παράμετρος δεν είναι αναγκαία το λ , αλλά περισσότερο το μ , ο μέσος του x :

$$\mu = E(x | \lambda, a) = \frac{a}{\lambda}$$

Επειδή το a είναι γνωστό πρόκειται για ένα απλό μετασχηματισμό του λ . Αν $\lambda \sim \Gamma(n_0, \frac{1}{s_0})$ αποδεικνύεται ότι το μ ακολουθεί μια Gamma – αντίστροφη κατανομή:

$$\mu \sim \Gamma^{-1}\left(n_0, \frac{1}{as_0}\right) \quad (\text{βλέπε Παράρτημα 2}).$$

Έχουμε, προφανώς εκ των υστέρων ένα ισοδύναμο αποτέλεσμα :

$$\mu | s, a \sim \Gamma^{-1}\left(n^*, \frac{1}{as^*}\right)$$

Πράγματι, η εκ των υστέρων πυκνότητα του μ δεδομένου του a θα είναι

$$\begin{aligned} p(\mu | s, a) &\propto p(s | \lambda, a) p(\mu) \\ &\propto \lambda^{na} e^{-\lambda s} \mu^{-(n_0+1)} e^{-\frac{as_0}{\mu}} = \frac{a^{na}}{\mu^{na}} e^{-\frac{a}{\mu}s} \mu^{-(n_0+1)} e^{-\frac{as_0}{\mu}} \\ &\propto \mu^{-(na+n_0+1)} e^{-\frac{a(s+s_0)}{\mu}}, \end{aligned}$$

Δηλαδή, η π.π. της $\Gamma^{-1}\left(na + n_0, \frac{1}{a(s + s_0)}\right)$.

Ιδιαίτέρως

$$E(\mu | s, a) = a \frac{s^*}{n^* - 1}, \quad \text{αν } n^* > 1 \quad (36)$$

Γνωρίζουμε ότι, αν $x \sim \Gamma^{-1}\left(a, \frac{1}{\beta}\right)$, τότε η μέση τιμή της x υπολογίζεται ως εξής :

$$\begin{aligned} E(x) &= \int_0^{+\infty} x \frac{\beta^a}{\Gamma(a)} x^{-(a+1)} e^{-\frac{\beta}{x}} dx = \frac{\beta^a}{\Gamma(a)} \int_0^{+\infty} x^{-a} e^{-\frac{\beta}{x}} dx = \frac{\beta^a}{\Gamma(a)} \underbrace{\int_0^{+\infty} x^{-[(a-1)+1]} e^{-\frac{\beta}{x}} dx}_{\text{ολοκλ. } \Gamma^{-1}\left(a-1, \frac{1}{\beta}\right)} \\ &= \frac{\beta^a}{\Gamma(a)} \frac{\Gamma(a-1)}{\beta^{a-1}} = \frac{\Gamma(a-1)}{\Gamma(a)} \beta. \end{aligned}$$

Λοιπόν,

$$E(\mu | s, a) = \frac{\Gamma(na + n_0 - 1)}{\Gamma(na + n_0)} a(s + s_0) = \frac{1}{na + n_0 - 1} a(s + s_0) = \frac{as^*}{n^* - 1}, \quad \text{αν } n^* > 1.$$

Ακόμη για μια εκ των προτέρων μη – πληροφορική βρίσκουμε ότι

$$E(\mu | s, a) = a \frac{s}{na - 1}, \quad \text{αν } na > 1 \quad (37)$$

Πράγματι, για μια εκ των προτέρων μη – πληροφορική με $p(\lambda) \propto \lambda^{-1}$, το $\mu = \frac{a}{\lambda}$ ακολουθεί :

$$f_\mu(\mu) = p\left(\frac{a}{\mu}\right) \frac{a}{\mu^2} \propto \frac{\mu}{a} \frac{a}{\mu^2} = \frac{1}{\mu}, \quad \forall \mu > 0.$$

Άρα,

$$p(\mu | s, a) \propto \frac{a^{na}}{\mu^{na}} e^{-\frac{a}{\mu}s} \frac{1}{\mu} \propto \mu^{-(na+1)} e^{-\frac{as}{\mu}}, \quad \text{δηλαδή } \mu | s, a \sim \Gamma^{-1}\left(na, \frac{1}{as}\right).$$

‘Επομένως,

$$E(\mu | s, a) = \frac{\Gamma(na - 1)}{\Gamma(na)} = a \frac{s}{na - 1}, \quad \text{αν } na > 1.$$

Στην περίπτωση ενός εκθετικού νόμου, ο οποίος λαμβάνεται για $a=1$, ο συνήθης συμπερασματικός εκτιμητής (εκτιμητής μεγίστης πιθανοφάνειας) του μ είναι ο δειγματικός μέσος $\hat{\mu} = \bar{x} = \frac{S}{n}$ και η εκ των υστέρων μέση τιμή του μ είναι :

$$E(\mu | s, a) = \frac{s}{n-1} = \frac{n}{n-1} \hat{\mu}, \quad \text{αν } n > 1.$$

4.3 Predictive νόμος

Δυνάμεθα επίσης εδώ να λάβουμε μια αναλυτική έκφραση της predictive (εκ των προτέρων) πυκνότητας του x , ολοκληρώνοντας την από κοινού πυκνότητα του (x, λ) επί όλων των δυνατών τιμών του λ , με την προϋπόθεση ότι τα $s_0, n_0 > 0$ για να εξασφαλίζεται η ύπαρξη του παρακάτω ολοκληρώματος. Η εκ των προτέρων επί του λ δίδεται από τη σχέση (32),

$$\lambda \sim \Gamma(n_0, \frac{1}{s_0})$$

και δεν εξαρτάται από την τιμή του a , τυπικά μπορούμε λοιπόν να γράψουμε $p(\lambda | a) = p(\lambda)$. Τότε

$$\begin{aligned} p(x | a) &= \int_0^{+\infty} p(x | \lambda, a) p(\lambda) d\lambda = \int_0^{+\infty} f_\gamma(x | \lambda, a) f_\gamma(\lambda | s_0, n_0) d\lambda \\ &= \int_0^{+\infty} \frac{\lambda^\alpha x^{a-1} e^{-\lambda x}}{\Gamma(\alpha)} \frac{s_0^{n_0} \lambda^{n_0-1} e^{-\lambda s_0}}{\Gamma(n_0)} d\lambda = \frac{s_0^{n_0} x^{\alpha-1}}{\Gamma(\alpha)\Gamma(n_0)} \underbrace{\int_0^{+\infty} \lambda^{\alpha+n_0-1} e^{-\lambda(x+s_0)} d\lambda}_{\text{ολοκλ. } \Gamma(\alpha+n_0, \frac{1}{s_0+x})} \\ &= \frac{s_0^{n_0} x^{\alpha-1}}{\Gamma(\alpha)\Gamma(n_0)} \frac{\Gamma(\alpha+n_0)}{(s_0+x)^{\alpha+n_0}} = \frac{\Gamma(\alpha+n_0)}{\Gamma(\alpha)\Gamma(n_0)} \frac{s_0^{n_0} x^{\alpha-1}}{(s_0+x)^{\alpha+n_0}}. \end{aligned} \quad (38)$$

Λαμβάνουμε λοιπόν, μια πυκνότητα της αντιστρόφου Βήτα -2 (βλέπε Παράρτημα 3.3) :

$$p(x | a) = f_{i\beta 2}(x | a, n_0, s_0).$$

Εκ των υστέρων, η predictive μιας καινούριας παρατήρησης x δεδομένου του ήδη παρατηρούμενου δείγματος x_1, \dots, x_n , μεγέθους n , δίδεται από τη σχέση :

$$p(x | x_1, \dots, x_n, a) = f_{i\beta 2}(x | a, n^*, s^*),$$

εδώ αρκεί $n^*, s^* > 0$ για την ύπαρξη του ολοκληρώματος.

Λαμβάνουμε επίσης

$$E(x | x_1, \dots, x_n, a) = a \frac{s^*}{n^* - 1}, \quad \text{αν } n^* > 1.$$

Στην περίπτωση μιας εκ των προτέρων μη – πληροφορικής η αναμενόμενη τιμή ενός μελλοντικού x θα είναι :

$$E(x | x_1, \dots, x_n, a) = a \frac{s}{na - 1}, \quad \text{αν } na > 1,$$

όπου $s = \sum_{i=1}^n x_i$ είναι το επαρκές συμπέρασμα του παρατηρούμενου δείγματος.

Στην περίπτωση μιας εκ των προτέρων μη-πληροφορικής,

$$p(x | x_1, \dots, x_n, a) = \int_0^{+\infty} p(x | \lambda, a) p(\lambda | s, a) d\lambda,$$

όπου η εκ των υστέρων μη-πληροφορική π.π. $p(\lambda | s, a)$ δίδεται από τη σχέση :

$$p(\lambda | s, a) = \frac{s^{na}}{\Gamma(na)} \lambda^{na-1} e^{-\lambda s}, \quad \text{δηλαδή } \lambda | s, a \sim \Gamma(na, \frac{1}{s})$$

Λοιπόν,

$$\begin{aligned} p(x | x_1, \dots, x_n, a) &= \int_0^{+\infty} \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)} \frac{s^{na}}{\Gamma(na)} \lambda^{na-1} e^{-\lambda s} d\lambda \\ &= \frac{x^{a-1} s^{na}}{\Gamma(a)\Gamma(na)} \underbrace{\int_0^{+\infty} \lambda^{(n+1)a-1} e^{-\lambda(x+s)} d\lambda}_{\text{ολοκλ.}\Gamma((n+1)a, \frac{1}{x+s})} = \frac{x^{a-1} s^{na}}{\Gamma(a)\Gamma(na)} \frac{\Gamma((n+1)a)}{(x+s)^{(n+1)a}} \\ &= \frac{\Gamma((n+1)a)}{\Gamma(a)\Gamma(na)} \frac{x^{a-1} s^{na}}{(x+s)^{(n+1)a}} = \frac{1}{B(a, na)} \frac{x^{a-1} s^{na}}{(x+s)^{(n+1)a}} \end{aligned}$$

Συνεπώς, $x | x_1, \dots, x_n, a \sim Beinv.2(a, na, s)$, οπότε η μέση τιμή της είναι

$$E(x | x_1, \dots, x_n, a) = a \frac{s}{na - 1} \quad \text{αν } na > 1.$$

Σημειώνουμε ότι αυτό το αποτέλεσμα είναι σύμφωνο με τα λαμβανόμενα αποτελέσματα πιο πάνω για τον εκ των υστέρων μέσο $\mu = E(x | \lambda, a)$ (βλέπε (36) και (37)).

5. Διαδικασία της Κανονικής Παλινδρόμησης

Τα μοντέλα της παλινδρόμησης μας επιτρέπουν να δούμε πώς οι μεταβλητές μιας ενδιαφέρουσας μεταβλητής y μπορούν να εκφραστούν δια μέσου ενός συνόλου επεξηγηματικών μεταβλητών x_1, \dots, x_p . Σκοπός μας είναι η εκτίμηση των αποτελεσμάτων των μεταβολών από έκαστη των επεξηγηματικών μεταβλητών, επί του y , αλλά επίσης να προβλέψουμε την τιμή του y για τις νέες τιμές των x_1, \dots, x_p . Η κανονική παλινδρόμηση υποθέτει ένα γραμμικό μοντέλο και μια κανονική κατανομή επί του y .

5.1 Το μοντέλο

Η διαδικασία της p -διάστατης κανονικής παλινδρόμησης δύναται να θεωρηθεί σαν μια διαδικασία η οποία γεννά την ακολουθία των τ.μ. y_1, \dots, y_i, \dots μονοδιάστατες και ανεξάρτητες, ακολουθώντας το μοντέλο :

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i \quad (39)$$

όπου $\varepsilon_i \sim N(0, \sigma^2)$ είναι ανεξάρτητες και της ίδιας κατανομής. Οι παράμετροι του μοντέλου είναι οι $(\beta_1, \beta_2, \dots, \beta_p, \sigma^2)$. Τα x_{ij} θεωρούνται μη – τυχαία, συνεπώς είναι δεδομένοι αριθμοί.

Παρατηρούμε ότι αυτό το μοντέλο είναι γραμμικό στα β , ενώ τα x δύναται να είναι μετασχηματισμός των προτύπων μεταβλητών. Έτσι τα μοντέλα

$$\text{ή } \begin{cases} y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \beta_4 x_i^3 + \varepsilon_i, & i = 1, 2, \dots, p \\ y_i = \beta_1 \log x_{i1} + \beta_2 \log x_{i2} + \varepsilon_i \end{cases}$$

ανήκουν στην οικογένεια των μοντέλων που θεωρούμε εδώ.

Χρησιμοποιώντας την διανυσματική έννοια, η σχέση (39) γράφεται επίσης υπό τη μορφή :

$$y_i = {}^t x_i \beta + \varepsilon_i \quad (40)$$

όπου $x_i = {}^t(x_{i1}, \dots, x_{ip})$ και $\beta = {}^t(\beta_1, \dots, \beta_p)$ διανύσματα του \mathbb{R}^p . Η διαδικασία της κανονικής παλινδρόμησης γεννά λοιπόν τις ανεξάρτητες κανονικές μεταβλητές y_i :

$$y_i | \beta, \sigma^2 \sim N({}^t x_i \beta, \sigma^2).$$

Σε ορισμένες εφαρμογές όπου τα x είναι τυχαία, εργαζόμαστε δεσμευτικά με τις παρατηρούμενες τιμές των x . Γράφουμε τότε:

$$y_i | \beta, \sigma^2, x_i \sim N(x_i \beta, \sigma^2) \quad (41)$$

Αυτή είναι η γραφή την οποία θα ακολουθήσουμε σε αυτή την παράγραφο όπου τα x θα είναι τυχαία ή σταθερά.

Ένα τυχαίο δείγμα μεγέθους n θα είναι λοιπόν μια ακολουθία των μεταβλητών (y_1, \dots, y_n) που ακολουθούν την κατανομή (41). Μπορούμε τότε να χρησιμοποιήσουμε την matricielle έννοια :

$$y = X\beta + \varepsilon,$$

$$\text{όπου } y = {}^t(y_1, \dots, y_n), \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad (n, p) - \text{πίνακας}$$

$$\varepsilon = {}^t(\varepsilon_1, \dots, \varepsilon_n) \text{ με } \varepsilon \sim N_n(0, \sigma^2 I_n) \text{ και } I_n = \text{o ταυτοτικός πίνακας τάξεως } n \quad (42)$$

Όταν το μοντέλο περιλαμβάνει ένα σταθερό όρο, η πρώτη στήλη του πίνακα X θα είναι μια στήλη από 1.

Να παρατηρήσουμε ότι θα είναι πιο εύχρηστο στη συνέχεια να σκεφτόμαστε σε όρους της ακρίβειας $\eta = \sigma^{-2}$ περισσότερο παρά σε όρους της διασποράς. Τότε το μοντέλο της γραμμικής παλινδρόμησης μπορεί να γραφεί ως εξής :

$$y | \beta, \eta, X \sim N_n(X\beta, \eta^{-1} I_n) \quad (43)$$

Γνωρίζουμε ότι ο συνήθης εκτιμητής του β , που προκύπτει με τη μέθοδο ελαχίστων τετραγώνων (βλέπε π.χ. στο [15]) είναι η στατιστική συνάρτηση,

$$b = ({}^t X X)^{-1} {}^t X y \quad (44)$$

Αυτός ο εκτιμητής (ε.ε.τ.) έχει τις ακόλουθες ιδιότητες:

$$E(b | \beta, \eta, X) = ({}^t X X)^{-1} {}^t X E(y | \beta, \eta, X) = ({}^t X X)^{-1} {}^t X X \beta = \beta$$

$$\begin{aligned} \text{Var}(b | \beta, \eta, X) &= ({}^t X X)^{-1} {}^t X \text{Var}(y) X ({}^t X X)^{-1} \\ &= ({}^t X X)^{-1} {}^t X (\sigma^2 I_n) X ({}^t X X)^{-1} \\ &= \sigma^2 ({}^t X X)^{-1} {}^t X X ({}^t X X)^{-1} \\ &= \sigma^2 ({}^t X X)^{-1} \\ &= \eta^{-1} ({}^t X X)^{-1} \end{aligned}$$

(Η τελευταία σχέση προκύπτει από τον ορισμό της διασποράς μιας δ.τ.μ.:

$$\text{Var}(b | \beta, \eta, X) = E[(b - Eb)'(b - Eb) | \beta, \eta, X]$$

$$\text{και } b | \beta, \eta, X \sim N_p(\beta, \eta^{-1}({}^tXX)^{-1}) \quad (45)$$

(αυτή επαληθεύεται από τις ιδιότητες της πολυδιάστατης Κανονικής κατανομής, βλέπε Παράρτημα, Θεώρημα 6.2.3).

Η διασπορά σ^2 εκτιμάται γενικώς από τη διασπορά των υπολοίπων (residus) μέσω της συνήθους μεθόδου των ελαχίστων τετραγώνων :

$$u = \frac{1}{n-p} \sum_{i=1}^n (y_i - {}^t x_i b)^2 = \frac{1}{n-p} {}^t (y - Xb)(y - Xb) \quad (46)$$

Το κανονικό μοντέλο παλινδρόμησης (43) περιλαμβάνει πολλές ενδιαφέρουσες μερικές περιπτώσεις :

1. Μονοδιάστατη κανονική διαδικασία :

Αν $p = 1$ και $X = i_n = {}^t(1, \dots, 1)$, ξαναβρίσκουμε τη μονοδιάστατη κανονική διαδικασία

$$y_i \sim N(\beta, \sigma^2), \quad i = 1, \dots, n \quad (47)$$

ανεξάρτητες και ισόνομες.

Οι εκτιμητές ελαχίστων τετραγώνων (44) και (46) γίνονται :

$$\hat{\beta} = b = \bar{y}, \quad \text{Var}(b | \beta, \sigma^2) = \frac{\sigma^2}{n} \quad \text{και} \quad \hat{\sigma}^2 = u = s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

(εδώ έχουμε ${}^tXX = n$).

2. Μοντέλο Ανάλυσης Διασποράς (ANOVA) :

Αν ο πίνακας X έχει την παρακάτω κατασκευή,

$$X = \begin{pmatrix} i_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & i_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_k} & 0_{n_k} & \cdots & i_{n_k} \end{pmatrix} \quad (48)$$

όπου $n_1 + n_2 + \dots + n_k = n$ και $p = k$, ξαναβρίσκουμε με την (43) το μοντέλο ανάλυσης της διασποράς με ένα παράγοντα :

$$y_{ij} | \beta_j, \sigma^2 \sim N(\beta_j, \sigma^2), \quad i = 1, \dots, n_j, \quad j = 1, \dots, k$$

ανεξάρτητα, τα οποία επιτρέπουν να συγκρίνουμε μέσω των k ανεξάρτητων δειγμάτων τους μέσους $\beta_j, j = 1, \dots, k$, k – κανονικών πληθυσμών ίδιας διασποράς σ^2 .

Εδώ παρατηρούμε ότι

$${}^tXX = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_k \end{pmatrix} = \text{diag}(n_1, \dots, n_k)$$

και $({}^tXX)^{-1} = \text{diag}(n_1^{-1}, \dots, n_k^{-1})$.

Οι εκτιμητές ελαχίστων τετραγώνων (44) και (46) γίνονται τότε :

$$\hat{\beta}_j = b_j = \bar{y}_j \left(= \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \right), \quad \hat{\sigma}^2 = u = \frac{(n_1 - 1)s_1^2 + \dots + (n_k - 1)s_k^2}{n - k},$$

όπου $s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$, $j = 1, \dots, k$.

Οι παρακάτω παράγραφοι προτείνουν την bayesienne ανάλυση του μοντέλου κανονικής παλινδρόμησης υπό τη γενική μορφή του (43) και θα περιορίσουμε τα λαμβανόμενα αποτελέσματα στο μονοδιάστατο κανονικό μοντέλο (47).

5.2 Η πιθανοφάνεια

Επειδή $y | \beta, \eta, X \sim N_n(X\beta, \eta^{-1}I_n)$, η συνάρτηση πιθανοφάνειας γράφεται :

$$L(\beta, \eta, y, X) = (2\pi)^{-\frac{n}{2}} \eta^{\frac{n}{2}} \exp \left\{ -\frac{\eta}{2} {}^t(y - X\beta)(y - X\beta) \right\} \quad (49)$$

(βλέπε Παράρτημα 6).

Είναι χρήσιμο να εισάγουμε σε αυτό το σημείο ορισμένες έννοιες. Το διάνυσμα e των υπολοίπων (residus) της μεθόδου ελαχίστων τετραγώνων (MET) γράφεται:

$$e = y - Xb = y - X({}^tXX)^{-1}{}^tXy = M_X y,$$

όπου $M_X y = I_n - X({}^tXX)^{-1}{}^tX$ είναι ο πίνακας προβολής επί του γραμμικού χώρου ο οποίος είναι ορθογώνιος με τον χώρο ο οποίος γεννάται από τις στήλες του πίνακα X . Οι παρακάτω ιδιότητες του M_X είναι ενδιαφέρουσες :

$$\begin{aligned} M_X &= {}^tM_X = M_X^2 \\ M_X X &= 0 \end{aligned}$$

Το άθροισμα των τετραγώνων των υπολοίπων της MET γίνεται τότε

$${}^t e e = {}^t (y - Xb)(y - Xb) = {}^t y' M_X M_X y = {}^t y M_X^2 y = {}^t y M_X y \quad (50)$$

Η τετραγωνική μορφή που εμφανίζεται στο εκθετικό μέρος της συνάρτησης πιθανοφάνειας (49) δύναται να ξαναγραφεί κατά τον ακόλουθο τρόπο :

$$\begin{aligned} {}^t (y - Xb)(y - Xb) &= {}^t (y - Xb + Xb - Xb)(y - Xb + Xb - Xb) \\ &= {}^t [(y - Xb) + (Xb - Xb)] [(y - Xb)(Xb - Xb)] \\ &= {}^t (y - Xb)(y - Xb) + {}^t (y - Xb)(Xb - Xb) + {}^t (Xb - Xb)(y - Xb) + {}^t (Xb - Xb)(Xb - Xb) \\ &= {}^t (y - Xb)(y - Xb) + {}^t (Xb - Xb)(Xb - Xb) + 2 {}^t (y - Xb)(Xb - Xb) \\ &= {}^t y M_X y + {}^t (b - \beta)' X X (b - \beta), \end{aligned}$$

η τελευταία ισότητα ακολουθεί από τις ιδιότητες του M_X , αφού ${}^t (y - Xb)(Xb - Xb) = {}^t y' M_X X (b - \beta)$ και ${}^t M_X X = 0$.

Μπορούμε τώρα να ξαναγράψουμε την πιθανοφάνεια :

$$\begin{aligned} L(\beta, \eta, y, X) &= (2\pi)^{-\frac{n}{2}} \eta^{\frac{n}{2}} \exp\left\{-\frac{\eta}{2} ({}^t y M_X y + {}^t (b - \beta)' X X (b - \beta))\right\} \\ &= (2\pi)^{-\frac{v}{2}} \eta^{\frac{v}{2}} \exp\left\{-\frac{\eta}{2} v u\right\} \times (2\pi)^{-\frac{p}{2}} \eta^{\frac{p}{2}} \exp\left\{-\frac{\eta}{2} {}^t (b - \beta)' X X (b - \beta)\right\} \quad (51) \end{aligned}$$

όπου $v = n - p$ παριστά τους «βαθμούς ελευθερίας» του μοντέλου, ενώ η έννοια $u = \frac{1}{v} {}^t y M_X y$ έχει ήδη ορισθεί παραπάνω (αυτός είναι ο συνήθης εκτιμητής ελαχίστων τετραγώνων του σ^2). Η τελευταία παραγοντοποίηση της πιθανοφάνειας θα απλοποιήσει την ανάλυση η οποία ακολουθεί. Ιδιαίτερως αυτή επιτρέπει να δούμε ότι η σ.σ. (b, u) είναι επαρκής για την παράμετρο (β, η) .

Η bayesianne ανάλυση αυτού του μοντέλου είναι σχετικά άμεση αν η παράμετρος σ^2 (ή η ακρίβεια η) είναι γνωστή. Θα αρχίσουμε, λοιπόν, την παρουσίαση από αυτή την απλή περίπτωση έτσι ώστε να ετοιμάσουμε την ανάλυση του πιο γενικού μοντέλου όπου το σ^2 είναι άγνωστο.

5.3 Διασπορά γνωστή

α) Φυσική συζυγής

Όταν η διασπορά $\sigma^2 (= \eta^{-1})$ είναι γνωστή, το μέρος της πιθανοφάνειας όπου φαίνεται η μοναδική παράμετρος β , δίδεται από το δεύτερο εκθετικό παράγοντα της

πιθανοφάνειας (51). Παρατηρούμε τότε ότι η b είναι επαρκής σ.σ. για το β . Βλέποντας την πιθανοφάνεια σαν συνάρτηση του β μόνο, λαμβάνουμε :

$$L(\beta, \eta, y, X) \propto \exp\left\{-\frac{\eta'}{2}(b-\beta)'XX(b-\beta)\right\} \quad (52)$$

Αναγνωρίζουμε σαν συνάρτηση του β , τον πυρήνα μιας p -διάστατης κανονικής. Η φυσική συζυγής οικογένεια θα είναι λοιπόν, σ' αυτό το μοντέλο, η οικογένεια των p -διάστατων κανονικών πυκνοτήτων.

b) Εκ των προτέρων

Αν εκλέξουμε μια εκ των προτέρων μέσα στη φυσική συζυγή οικογένεια, έχουμε :

$$\beta \sim N_p(b_0, \Omega_0^{-1})$$

όπου $b_0 = E\beta$ είναι το διάνυσμα των εκ των προτέρων μέσων των στοιχείων του β και ο Ω_0 είναι ο πίνακας των εκ των προτέρων ακριβειών : $Var(\beta) = \Omega_0^{-1}$.

Στην πράξη, επειδή το η είναι γνωστό, για να απλοποιήσουμε τις έννοιες οι οποίες θα ακολουθήσουν, θα ορίσουμε την εκ των προτέρων ακρίβεια από τον πίνακα B_0 , ο οποίος θα είναι ο πίνακας των σχετικών ακριβειών του β , «βάρους η » :

$$\Omega_0 = \eta B_0.$$

Έχουμε τότε για την εκ των προτέρων πυκνότητα του β τον ακόλουθο ορισμό :

$$p(\beta | \eta) \propto \exp\left\{-\frac{\eta'}{2}(\beta - b_0)B_0(\beta - b_0)\right\}$$

δηλαδή $\beta | \eta \sim N_p(b_0, (\eta B_0^{-1}))$ (53)

Αν οι εκ των προτέρων διασπορές του β τείνουν προς το $+\infty$, θα έχουμε μια εκ των προτέρων μη-πληροφορική επί του β . Μπορούμε να θεωρήσουμε ότι αυτή η περίπτωση αντιστοιχεί στην οριακή τιμή των σχετικών ακριβειών $B_0 \rightarrow 0$. Οι παράμετροι $\beta_j, j = 1, \dots, p$ είναι τότε εκ των προτέρων ανεξάρτητες και κατανομημένες κατά ένα (impropre) ομοιόμορφο νόμο :

$$p(\beta | \eta) \propto cte \quad (54)$$

c) Εκ των υστέρων

Μπορούμε να πάρουμε τον πυρήνα της εκ των υστέρων συνδυάζοντας τις σχέσεις (52) και (53), αλλά σε αυτή την περίπτωση θα ήταν πιο απλό να χρησιμοποιήσουμε τις ιδιότητες της Πολυδιάστατης Κανονικής κατανομής. Είναι φανερό ότι η εκ των προτέρων πυκνότητα για το β δεν εξαρτάται από τον πίνακα X , θεωρούμενος σαν πίνακας των σταθερών· τυπικά μπορούμε να γράψουμε

$$p(\beta | \eta) = p(\beta | \eta, X)$$

Ανακαλώντας τις ιδιότητες του b που δίδονται στην (45), έχουμε

$$\begin{aligned}\beta | \eta, X &\sim N_p(b_0, (\eta B_0)^{-1}) \\ b | \beta, \eta, X &\sim N_p(\beta, \eta^{-1}({}^t X X)^{-1})\end{aligned}$$

Χρησιμοποιώντας τότε τις ιδιότητες της πολυδιάστατης κανονικής (βλέπε Παράρτημα, Θεώρημα 6.5.7 ii) :

$$\begin{pmatrix} \beta \\ b \end{pmatrix} | \eta, X \sim N_{2p} \left(\begin{pmatrix} b_0 \\ b_0 \end{pmatrix}, \begin{pmatrix} \eta^{-1} B_0^{-1} & \eta^{-1} B_0^{-1} \\ \eta^{-1} B_0^{-1} & \eta^{-1} (B_0^{-1} - ({}^t X X)^{-1}) \end{pmatrix} \right) \quad (55)$$

Από αυτή την από κοινού, χρησιμοποιώντας τώρα την αντίστροφη ανάλυση της πολυδιάστατης κανονικής (βλέπε Παράρτημα, Θεώρημα 6.5.6 i)) λαμβάνουμε:

$$\beta | b, \eta, X \sim N_p(\mu'_1, \Gamma_{11.2})$$

όπου

$$\mu'_1 = \eta^{-1} B_0^{-1} [\eta^{-1} (B_0^{-1} + ({}^t X X)^{-1})]^{-1} (b - b_0) + b_0 = B_0^{-1} [B_0^{-1} + ({}^t X X)^{-1}]^{-1} (b - b_0) + b_0$$

$$\begin{aligned}\Gamma_{11.2} &= \eta^{-1} B_0^{-1} - \eta^{-1} B_0^{-1} [\eta^{-1} (B_0^{-1} + ({}^t X X)^{-1})]^{-1} \eta^{-1} B_0^{-1} \\ &= \eta^{-1} \left\{ B_0^{-1} - B_0^{-1} [B_0^{-1} + ({}^t X X)^{-1}]^{-1} B_0^{-1} \right\}\end{aligned}$$

δηλαδή

$$\begin{aligned}\beta | b, \eta, X &\sim \beta | y, \eta, X \sim \\ &\sim N_p \left(B_0^{-1} [B_0^{-1} + ({}^t X X)^{-1}]^{-1} (b - b_0) + b_0, \eta^{-1} \left\{ B_0^{-1} - B_0^{-1} [B_0^{-1} + ({}^t X X)^{-1}]^{-1} B_0^{-1} \right\} \right) \quad (56)\end{aligned}$$

Χάρη σε κάποιους αλγεβρικούς χειρισμούς βασιζόμενους στις ιδιότητες των αντιστρόφων διαχωρισμένων πινάκων μπορούμε να αποδείξουμε τις παρακάτω ιδιότητες.

Από την άλγεβρα των πινάκων ανακαλούμε το διωνυμικό αντίστροφο θεώρημα :
Έστω A και B δύο (m, m) – αντιστρέψιμοι πίνακες. Τότε:

$$(A + B)^{-1} = A^{-1}(A^{-1} + B^{-1})B^{-1} \quad (57)$$

$$(A + B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1} \quad (58)$$

Αυτές οι δύο σχέσεις δίνουν διαδοχικά τις σχέσεις :

$$B_0^{-1}[B_0^{-1} + ({}^tXX)^{-1}]^{-1} = (B_0 + {}^tXX)^{-1}{}^tXX = W$$

$$B_0^{-1} - B_0^{-1}[B_0^{-1} + ({}^tXX)^{-1}]^{-1}B_0^{-1} = (B_0 + {}^tXX)^{-1}$$

Έτσι προκύπτει αμέσως ότι :

$$E(\beta | b, n, X) = b_0 + W(b - b_0) = (I_p - W)b_0 + Wb = (B_0 + {}^tXX)^{-1}(B_0b_0 + {}^tXXb)$$

$$Var(\beta | b, n, X) = \eta^{-1}(B_0 + {}^tXX)^{-1} \quad (59)$$

Μπορούμε λοιπόν τελικά να γράψουμε :

$$\beta | b, \eta, X \sim N_p(b^*, (\eta B^*)^{-1})$$

όπου

$$b^* = (B_0 + {}^tXX)^{-1}(B_0b_0 + {}^tXXb)$$

$$B^* = B_0 + {}^tXX \quad (60)$$

Οι εκφράσεις των εκ των υστέρων ροπών του β δεδομένες από τις (59) και (60) επιτρέπουν να καταλάβουμε καλύτερα πως οι πληροφορίες εκ των προτέρων και του δείγματος συνδυάζονται :

1. Πρόσθεση ακριβειών

Ο πίνακας των εκ των προτέρων ακριβειών δίδεται από τον ηB_0 και η δειγματική ακρίβεια του b , ο ε.ε.τ., δίδεται από την $\eta {}^tXX$. Βεβαιώνεται ότι η εκ των υστέρων ακρίβεια, δεδομένη από την ηB^* , είναι το άθροισμα της εκ των προτέρων και του δείγματος.

2. Εκ των υστέρων μέσος

Ο εκ των υστέρων μέσος b^* είναι ο βεβαρυμένος μέσος του b_0 , ο εκ των προτέρων μέσος, και του b , ο ε.ε.τ., τα αντίστοιχα βάρη όντως οι ακρίβειες εκ των προτέρων και η δειγματική

$$b^* = (B_0 + {}^tXX)^{-1}B_0b_0 + (B_0 + {}^tXX)^{-1}{}^tXXb.$$

Αυτό εδώ επιτρέπει να ερμηνεύσουμε εκ νέου τις τιμές (b_0, B_0) της εκ των προτέρων σε όρους ισοδύναμου υποθετικού δείγματος.

3. Εκ των προτέρων μη – πληροφορική

Παρατηρούμε ότι, αν $B_0 \rightarrow 0, W \rightarrow I_p$ και τότε

$$\begin{aligned} E(\beta | b, \eta, X) &\rightarrow b \\ \text{Var}(\beta | b, \eta, X) &\rightarrow \eta^{-1} ({}^t X X)^{-1} \end{aligned}$$

και ξαναβρίσκουμε τα συνήθη αποτελέσματα της Στατιστικής Συμπερασματολογίας. Λοιπόν, ένας εκ των υστέρων μη – πληροφορικός νόμος για το β θα ήταν ο $N_p(b, \eta^{-1} ({}^t X X)^{-1})$.

4. Ασυμπτωτικά αποτελέσματα

Υπό τις συνήθειες υποθέσεις της κανονικότητας, αν $n \rightarrow +\infty, ({}^t X X)^{-1} \rightarrow 0$ και $W \rightarrow I_p$. Έχουμε τότε :

$$\begin{aligned} E(\beta | b, \eta, X) &\rightarrow b \\ \text{Var}(\beta | b, \eta, X) &\rightarrow 0 \end{aligned}$$

Αν το μέγεθος του δείγματος αυξάνει, το εκ των προτέρων βάρος ελαττώνεται και η εκ των υστέρων συγκεντρώνεται γύρω από το b , ο ε.ε.τ.

d) Predictive νόμος

Η εκ των προτέρων predictive του y προκύπτει εύκολα αν συνδυάσουμε την εκ των προτέρων και την πιθανοφάνεια του y για ένα πίνακα τιμών, X , των επεξηγηματικών μεταβλητών, δεδομένο.

Δεδομένου, λοιπόν, ότι

$$\begin{aligned} \beta | \eta, X &\sim N_p(b_0, (\eta B_0)^{-1}) \\ y | \beta, \eta, X &\sim N_n(X\beta, \eta^{-1} I_n) \end{aligned}$$

χρησιμοποιώντας όπως και προηγουμένως τις ιδιότητες της πολυδιάστατης κανονικής (βλέπε Παράρτημα ,Θεώρημα 6.5.7 ii)) λαμβάνουμε :

$$\begin{pmatrix} \beta \\ y \end{pmatrix} | \eta, X \sim N_{p+n} \left(\begin{pmatrix} b_0 \\ Xb_0 \end{pmatrix}, \begin{pmatrix} \eta^{-1} B_0^{-1} & \eta^{-1} B_0^{-1} {}^t X \\ \eta^{-1} X B_0^{-1} & \eta^{-1} (I_p + X B_0^{-1} {}^t X) \end{pmatrix} \right)$$

Από αυτή την από κοινού κατανομή λαμβάνουμε αμέσως τον *predictive* (ή περιθωριακό) νόμο του y :

$$y | \eta, X \sim N_n(Xb_0, \eta^{-1}(I_p + XB_0^{-1}X)) \quad (61)$$

Πρέπει εδώ η *εκ των προτέρων* να έχει μια πυκνότητα, δηλαδή ο B_0 να είναι αντιστρέψιμος.

Για να πάρουμε την *predictive* πυκνότητα του y_f , για n_f καινούριες τιμές που παράγονται από την ίδια διαδικασία της κανονικής παλινδρόμησης (39) και για ένα δεδομένο πίνακα X_f , τύπου (n_f, p) , επεξηγηματικών μεταβλητών, αρκεί να αντικαταστήσουμε το b_0 με το b^* , τον B_0 με τον B^* και τον X με τον X_f μέσα στην σχέση (61) :

$$y_f | \eta, y, X, X_f \sim N_{n_f}(X_f b^*, \eta^{-1}(I_{n_f} + X_f (B^*)^{-1} X_f)) \quad (62)$$

Σ' αυτήν την περίπτωση, για να υπάρχει η *predictive πυκνότητα* αρκεί ο $B^* = B_0 + {}^t X X$ να είναι αντιστρέψιμος. Αυτή υπάρχει λοιπόν για μια *εκ των προτέρων improper* (μη - πληροφορική : $B_0 \rightarrow 0$) αν ο X είναι πλήρης τάξεως ($rg(X) = p$).

ε) Μερική περίπτωση: Η διαδικασία της μονοδιάστατης Κανονικής με διασπορά γνωστή

Σ' αυτήν τη μερική περίπτωση, $p = 1, X = i_n, {}^t X X = n$. Λοιπόν το β είναι βαθμωτό και $b = \bar{y} = (\frac{1}{n} \sum_{i=1}^n y_i, \text{δειγματικός μέσος})$. Σημειώνοντας $y = {}^t (y_1, \dots, y_n)$, διαδοχικά λαμβάνουμε :

$$y | \beta, \eta \sim N_n(i_n \beta, \eta^{-1} I_n), \quad \eta = \sigma^{-2}$$

δηλαδή οι y_i είναι ανεξάρτητες και ισόνομες τ.μ. με κατανομή $N(\beta, \sigma^2)$, $\bar{y} \sim N_1(\beta, (\eta n)^{-1})$,

$$\text{την εκ των προτέρων επί του } \beta : \beta | \eta \sim N_1(b_0, \frac{1}{\eta B_0})$$

$$\text{και την εκ των υστέρων επί του } \beta : \beta | \bar{y}, \eta \sim N_1(b^*, \frac{1}{\eta B^*}),$$

$$\text{όπου } b^* = \frac{B_0}{B_0 + n} b_0 + \frac{n}{B_0 + n} \bar{y}, \quad B^* = B_0 + n.$$

Σχολιάζοντας, όπως και στην παράγραφο c), τα παραπάνω αποτελέσματα παρατηρούμε ότι :

1°. Ο εκ των υστέρων μέσος b^* γράφεται :

$$b^* = \frac{\eta B_0}{\eta B_0 + n\eta} b_0 + \frac{n\eta}{\eta B_0 + n\eta} \bar{y} \quad \text{και} \quad \eta B^* = \eta B_0 + n\eta$$

Λοιπόν ο b^* είναι ένας αντισταθμιζόμενος μέσος του εκ των προτέρων μέσου b_0 και του δειγματικού μέσου \bar{y} . Τα σχετικά βάρη είναι οι αντίστοιχες ακρίβειες.

2°. Η εκ των υστέρων διασπορά $(\eta B^*)^{-1}$ είναι : $(\eta B^*)^{-1} = \frac{1}{\eta B_0 + n\eta}$

Λοιπόν , η εκ των υστέρων ακρίβεια θα είναι : $\eta B^* = \eta B_0 + n\eta$, δηλαδή το άθροισμα της εκ των προτέρων ακρίβειας ηB_0 και της δειγματικής ακρίβειας $n\eta$.

3°. Με βάση το 1° και το 2° μπορούμε να δούμε και κάποιες οριακές καταστάσεις

• Αν $n \rightarrow +\infty$, διατηρώντας το B_0 σταθερό, έχουμε : $b^* \rightarrow \bar{y}$ και $(\eta B^*)^{-1} \rightarrow 0$.

Όλη λοιπόν η εκ των υστέρων κατανομή του β συγκεντρώνει τη μάζα της στο \bar{y} (δειγματικό μέσο) για οποιαδήποτε εκ των προτέρων ακρίβεια ηB_0 : ξαναβρίσκουμε δηλαδή τον κλασσικό εκτιμητή του β .

• Αν το $B_0^{-1} \rightarrow +\infty$ και επομένως $B_0 \rightarrow 0$, το εκ των προτέρων βάρος γίνεται αμελητέο και όλη η πληροφορία προέρχεται μόνο από το δείγμα. Εδώ θα μπορούσαμε να πούμε ότι υπάρχει εκ των προτέρων λίγη ή καθόλου πληροφορία και ξαναβρίσκουμε επίσης τον κλασσικό εκτιμητή του β :

$$b^* = E(\beta | \bar{y}, \eta) \rightarrow \bar{y}$$

$$(\eta B^*)^{-1} \rightarrow \frac{\sigma^2}{n}$$

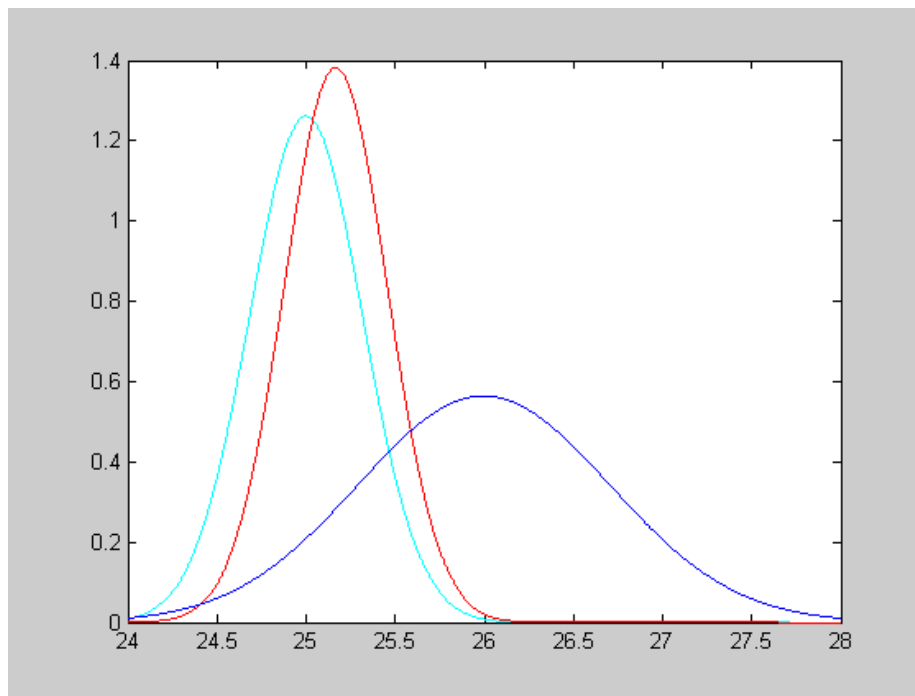
Μια εκ των υστέρων μη – πληροφορική θα είναι τότε η $N(\bar{y}, \frac{\sigma^2}{n})$

• Αντίστροφα, αν $B_0^{-1} \rightarrow 0$ και επομένως $B_0 \rightarrow +\infty$, έχουμε μια εκ των προτέρων «δογματική» (μάζα του Dirac στο σημείο b_0). Εδώ γνωρίζουμε με βεβαιότητα την τιμή του β και το δείγμα δεν μπορεί να τροποποιήσει αυτή την εκ των προτέρων πληροφορία

Για παράδειγμα, μπορούμε να παραστήσουμε γραφικά την εκ των προτέρων πυκνότητα του β , την εκ των υστέρων πυκνότητα του β και την εκ των υστέρων πυκνότητα του β για μια εκ των προτέρων μη – πληροφορική. Στα γραφήματα που ακολουθούν το n είναι το μόνο που μεταβάλλεται για να φανεί ο ρόλος του βάρους της

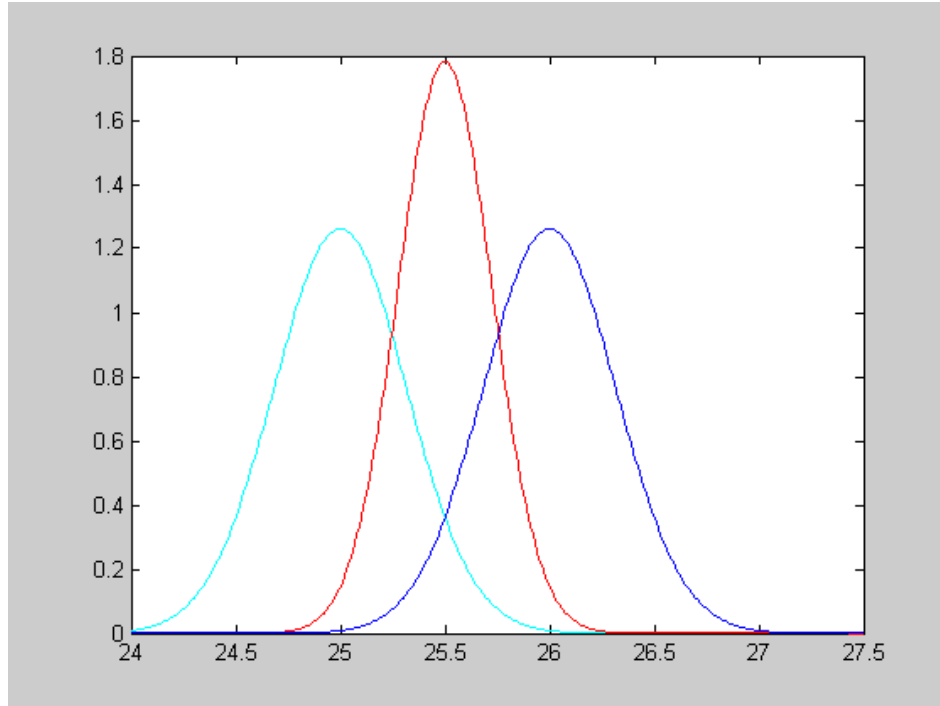
πληροφορίας που προέρχεται από το δείγμα. Βεβαιώνεται ότι, όταν το n αυξάνει, η εκ των υστέρων πληροφορία επί του β , συγκεντρώνεται γύρω από την παρατηρούμενη τιμή \bar{y} και ότι το εκ των προτέρων βάρος είναι ήδη ελαττωμένο στην περίπτωση που $n = 50$.

Γράφημα 5.1



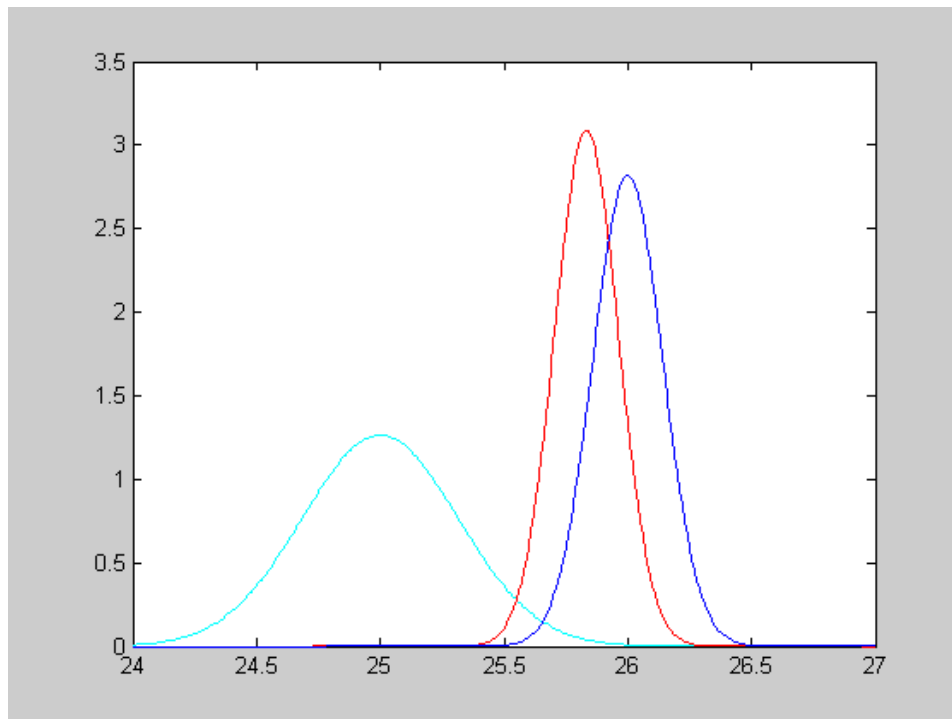
Η εκ των προτέρων πυκνότητα του β (γαλάζιο), η εκ των υστέρων πυκνότητα του β (κόκκινο) και η εκ των υστέρων πυκνότητα του β για μια εκ των προτέρων μη – πληροφορική (κυανούν) με $\sigma^2 = 1$, $b_0 = 25$, $B_0 = 10$, $\bar{y} = 26$, $n = 2$

Γράφημα 5.2



Η εκ των προτέρων πυκνότητα του β (γαλάζιο), η εκ των υστέρων πυκνότητα του β (κόκκινο) και η εκ των υστέρων πυκνότητα του β για μια εκ των προτέρων μη – πληροφορική (κυανούν) με $\sigma^2 = 1, b_0 = 25, B_0 = 10, \bar{y} = 26, n = 10$

Γράφημα 5.3



Η εκ των προτέρων πυκνότητα του β (γαλάζιο), η εκ των υστέρων πυκνότητα του β (κόκκινο) και η εκ των υστέρων πυκνότητα του β για μια εκ των προτέρων μη-πληροφορική (κυανούν) με $\sigma^2 = 1$, $b_0 = 25$, $B_0 = 10$, $\bar{y} = 26$, $n = 50$.

Τελικά όμως, αυτό που ενδιαφέρει τους περισσότερους μηχανικούς στο πλαίσιο του ελέγχου ποιότητας, είναι η προβλεπόμενη τιμή (κόστους, βάρους κ.ά.) ενός βιομηχανικού προϊόντος, διότι αυτή είναι εκείνη η οποία οφείλει να σεβαστεί ορισμένους κανόνες.

Αυτή η τιμή είναι μια τ.μ. την οποία σημειώνουμε με y_f , ανεξάρτητη του ήδη εξεταζόμενου αρχικού δείγματος $y = (y_1, \dots, y_n)$.

Γνωρίζουμε ότι ο νόμος πιθανότητας του y_f , αν το β είναι σταθερό, είναι :

$$y_f | \beta, \bar{y} \sim N(\beta, \sigma^2) = N(\beta, \eta^{-1})$$

Συνδυάζοντας αυτή την κατανομή με την εκ των προτέρων κατανομή του β και με την κατανομή του \bar{y} , και χρησιμοποιώντας εκ νέου τις ιδιότητες της πολυδιάστατης κανονικής λαμβάνουμε όπως στην παράγραφο d) :

- την εκ των προτέρων predictive κατανομή του y_f :

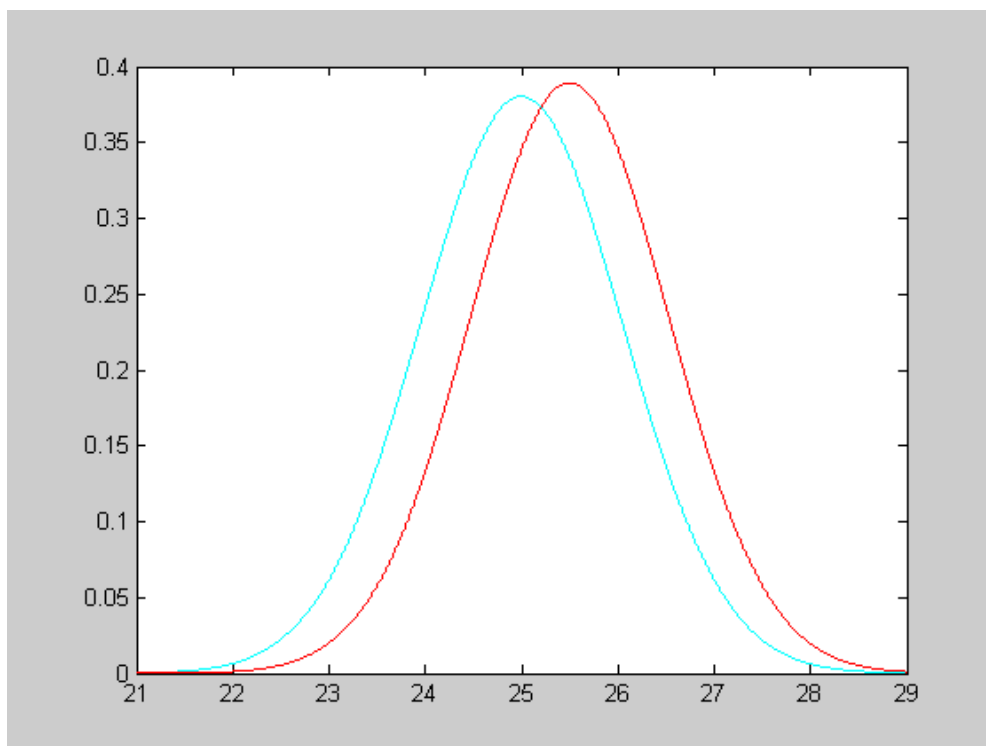
$$y_f \sim N(b_0, \eta^{-1} + (\eta B_0)^{-1})$$

- την εκ των υστέρων predictive κατανομή του y_f :

$$y_f | \bar{y}, \eta \sim N(b^*, \eta^{-1} (1 + \frac{1}{B^*}))$$

Μπορούμε να συγκρίνουμε τις ανωτέρω δύο κατανομές για να εκτιμήσουμε καλύτερα αυτό το οποίο έχουμε πληροφορηθεί επί της διαδικασίας παρατηρώντας το \bar{y} . Αυτή η σύγκριση φαίνεται στο παρακάτω γράφημα για $n = 10$.

Γράφημα 5.4



Η εκ των προτέρων predictive του y_f (γαλάζιο), και η εκ των υστέρων predictive του y_f (κόκκινο) με $\sigma^2 = 1, b_0 = 25, B_0 = 10, \bar{y} = 26, n = 10$.

Μέσω αυτής της predictive κατανομής μπορούμε να δούμε αν η παραγωγική διαδικασία είναι ακόμη υπό έλεγχο ή υπάρχει ανάγκη ρύθμισης (ή αναπροσαρμογής) αυτής.

5.4 Διασπορά άγνωστη

Αυτή η περίπτωση είναι λίγο περισσότερο περίπλοκη από την προηγούμενη, αλλά πρόκειται να χρησιμοποιήσουμε τα λαμβανόμενα αποτελέσματα αποσυνθέτοντας τον από κοινού νόμο του (β, η) σ' ένα δεσμευμένο του $\beta | \eta$, τον οποίο θα μεταχειρισθούμε όπως στην προηγούμενη παράγραφο και ένα περιθωριακό του η . Η πιθανοφάνεια σ' αυτή την γενική περίπτωση έχει δοθεί από την (51). Την γράφουμε :

$$L(\beta, \eta, y, X) = (2\pi)^{-\frac{v}{2}} \eta^{\frac{v}{2}} \exp\left\{-\frac{\eta}{2} \nu u\right\} \times (2\pi)^{-\frac{p}{2}} \eta^{\frac{p}{2}} \exp\left\{-\frac{\eta}{2} (b - \beta)' X X (b - \beta)\right\},$$

όπου υπενθυμίζουμε $v = n - p$ παριστά τους «βαθμούς ελευθερίας» και $u = 'y M_x y / v$ είναι ο αμερόληπτος εκτιμητής του σ^2 .

Αν η πιθανοφάνεια θεωρηθεί σαν συνάρτηση των β και η , έχουμε :

$$L(\beta, \eta, y, X) \propto \eta^{\frac{v}{2}} \exp\left\{-\frac{\eta}{2} \nu u\right\} \times \eta^{\frac{p}{2}} \exp\left\{-\frac{\eta}{2} (b - \beta)' X X (b - \beta)\right\} \quad (63)$$

a) Φυσική συζυγής

Παρατηρώντας τη σχέση (63) μπορούμε να αναγνωρίσουμε σε αυτήν, ως συνάρτηση των (β, η) , τον πυρήνα μιας πυκνότητας Κανονικής – Γάμμα πολυδιάστατης (βλέπε Παράρτημα 7), για το (β, η) .

b) Εκ των προτέρων

Επιλέγουμε λοιπόν ως *εκ των προτέρων* για το (β, η) έναν νόμο Κανονικό – Γάμμα. Γνωρίζοντας ότι ο X θεωρείται μη – τυχαίος ή σταθερός με την παρατηρούμενη τιμή του και σημειώνοντας ότι ο X δεν υπεισέρχεται μέσα στους εκ των προτέρων νόμους, έχουμε τυπικά : $(\beta, \eta) \sim (\beta, \eta) | X$

Θα γράφουμε τότε για την *εκ των προτέρων* ότι : $(\beta, \eta) | X \sim N_p \gamma(b_0, B_0^{-1}, u_0, v_0)$.

Η εκ των προτέρων πυκνότητα είναι τότε :

$$p(\beta, \eta | X) \propto \eta^{\frac{p_0}{2}} \exp\left\{-\frac{\eta}{2} (\beta - b_0)' B_0 (\beta - b_0)\right\} \times \eta^{\frac{v_0}{2}-1} e^{-\frac{1}{2} \eta u_0} \quad (64)$$

όπου p_0 είναι μια έννοια που εισάγουμε για να μας βοηθήσει να παραστήσουμε το βαθμό του πίνακα B_0 . Γνωρίζουμε ότι (βλέπε Παράρτημα 7) για να πάρουμε μια

«proper» πυκνότητα, πρέπει ο B_0 να είναι αντιστρέψιμος, άρα $p_0 = p$. Αλλά θα διατηρήσουμε αυτή την έννοια p_0 για να μπορέσουμε να μεταχειρισθούμε την οριακή περίπτωση της εκ των προτέρων μη – πληροφορικής όπου $B_0 \rightarrow 0$ και συνεπώς $p_0 \rightarrow 0$.

Αυτή η εκ των προτέρων πυκνότητα, όταν $p_0 = p$, αποσυντίθεται σε δύο μέρη :

$$\begin{aligned} & \beta | \eta, X \sim N_p(b_0, (\eta B_0)^{-1}) \\ \text{και} & \eta | X \sim \Gamma\left(\frac{\nu_0}{2}, \frac{2}{u_0 \nu_0}\right) \end{aligned} \quad (65)$$

Παρατηρώντας τη σχέση (65) είναι φανερό ότι η δεσμευμένη πυκνότητα του β όταν το η είναι δεδομένο, συμπίπτει με την εκ των προτέρων (53) που έχουμε επιλέξει στην προηγούμενη παράγραφο, όταν το η υποτίθετο γνωστό, με b_0 ως εκ των προτέρων μέση τιμή και B_0 ως σχετική εκ των προτέρων ακρίβεια.

Εδώ επειδή το η είναι άγνωστο, έχουμε επιπλέον ένα εκ των προτέρων νόμο επί του η με :

$$E(\eta | X) = \frac{\nu_0}{u_0 \nu_0} = \frac{1}{u_0} \quad \text{και} \quad \text{Var}(\eta | X) = \frac{\nu_0}{2} \frac{4}{u_0^2 \nu_0^2} = \frac{2}{u_0^2 \nu_0}.$$

Η εκ των προτέρων περιθωριακή του β είναι τότε ένας νόμος Student πολυδιάστατος :

$$\beta | X \sim t_p(b_0, u_0 B_0^{-1}, \nu_0)$$

με τις δύο πρώτες ροπές :

$$\begin{aligned} E(\beta | X) &= b_0 \\ \text{Var}(\beta | X) &= \frac{\nu_0}{\nu_0 - 2} u_0 B_0^{-1}, \quad \nu_0 > 2 \end{aligned}$$

(βλέπε Παράρτημα 7,8)

Μια εκ των προτέρων μη – πληροφορική αντιστοιχεί στην περίπτωση όπου $B_0 \rightarrow 0$, συνεπώς $p_0 \rightarrow 0$ και από το άλλο μέρος $u_0, \nu_0 \rightarrow 0$. Από τη σχέση (64) προκύπτει τότε ότι :

$$p(\beta, \eta) = p(\beta, \eta | X) \propto \frac{1}{\eta}.$$

c) Εκ των υστέρων

Γνωρίζουμε ότι η χρήση της φυσικής συζυγής οφείλει να χορηγεί εκ των υστέρων μια νέα Κανονική – Γάμμα. Συνδυάζοντας την εκ των προτέρων (64) και εκείνη της πιθανοφάνειας (63), βρίσκουμε :

$$p(\beta, \eta | y, X) = p(\beta, \eta | b, u, X)$$

$$\begin{aligned} & \propto \eta^{\frac{p_0 + v_0}{2} - 1} \exp\left\{-\frac{\eta'}{2}(\beta - b_0)B_0(\beta - b_0) - \frac{\eta}{2}u_0v_0\right\} \\ & \quad \times \eta^{\frac{p}{2}} \exp\left\{-\frac{\eta'}{2}(b - \beta)'(XX)(b - \beta)\right\} \eta^{\frac{v}{2}} e^{-\frac{\eta}{2}vu} \\ & = \eta^{\frac{v+p+p_0+v_0}{2} - 1} \exp\left\{-\frac{\eta}{2}\left[{}^t(\beta - b_0)B_0(\beta - b_0) + {}^t(b - \beta)'(XX)(b - \beta) + u_0v_0 + uv\right]\right\} \\ & = \eta^{\frac{v+p+p_0+v_0}{2} - 1} \exp\left\{-\frac{\eta}{2}[Q + u_0v_0 + uv]\right\}, \end{aligned}$$

$$\text{όπου } Q = {}^t(\beta - b_0)B_0(\beta - b_0) + {}^t(b - \beta)'(XX)(b - \beta).$$

Αναλύοντας τις τετραγωνικές μορφές που εμφανίζονται στο Q και ορίζοντας $B^* = B_0 + {}^tXX$ και $b^* = B^{*-1}(B_0b_0 + {}^tXXb)$, παίρνουμε διαδοχικά :

$$\begin{aligned} Q &= {}^t\beta B_0\beta - 2{}^t\beta B_0b_0 + {}^t b_0 B_0 b_0 + {}^t b' (XX) b - 2{}^t \beta' (XX) b + {}^t \beta' (XX) \beta + {}^t b^* B^* b^* - {}^t b^* B^* b^* \\ & \quad - 2{}^t b^* B^* \beta + 2{}^t b^* B^* \beta \\ &= {}^t \beta B^* \beta - 2{}^t b_0 B_0 \beta + {}^t b_0 B_0 b_0 + {}^t b' (XX) b - 2{}^t b' (XX) \beta + {}^t b^* B^* b^* - {}^t b^* B^* b^* - 2{}^t b^* B^* \beta + 2{}^t b^* B^* \beta \\ &= {}^t (\beta - b^*) B^* (\beta - b^*) - {}^t b^* B^* b^* + {}^t b_0 B_0 b_0 + {}^t b' (XX) b - \underbrace{2{}^t b_0 B_0 \beta - 2{}^t b' (XX) \beta + 2{}^t b^* B^* \beta}_0 \\ &= {}^t (\beta - b^*) B^* (\beta - b^*) - {}^t b^* B^* b^* + {}^t b_0 B_0 b_0 + {}^t b' (XX) b. \end{aligned}$$

αφού

$$\begin{aligned} -2{}^t b_0 B_0 \beta - 2{}^t b' (XX) \beta + 2{}^t b^* B^* \beta &= -2{}^t b_0 B_0 \beta - 2{}^t b' (XX) \beta + 2{}^t (B_0 b_0 + {}^t XX b)' B^{*-1} B^* \beta \\ &= -2{}^t b_0 B_0 \beta - 2{}^t b' (XX) \beta + 2{}^t b_0 B_0 \beta + 2{}^t b' (XX) \beta = 0 \end{aligned}$$

Τότε μπορούμε να γράψουμε :

$$\begin{aligned} Q + vu + v_0 u_0 &= {}^t (\beta - b^*) B^* (\beta - b^*) - {}^t b^* B^* b^* + {}^t b_0 B_0 b_0 + {}^t b' (XX) b + vu + v_0 u_0 \\ &= {}^t (\beta - b^*) B^* (\beta - b^*) + v^* u^*, \end{aligned}$$

$$\text{όπου ορίζουμε } v^* u^* = -{}^t b^* B^* b^* + {}^t b_0 B_0 b_0 + {}^t b' (XX) b + vu + v_0 u_0.$$

Τέλος, ορίζοντας $v^* = v + v_0 + p_0$, μπορούμε να γράψουμε τον πυρήνα της εκ των υστέρων κατά τον ακόλουθο τρόπο

$$p(\beta, \eta | b, u, X) \propto \eta^{\frac{p}{2}} \exp\left\{-\frac{\eta}{2} (\beta - b^*)' B^* (\beta - b^*)\right\} \times \eta^{\frac{v^*}{2}-1} \exp\left\{-\frac{\eta}{2} v^* u^*\right\} \quad (66)$$

όπου αναγνωρίζουμε τον πυρήνα μιας Κανονικής – Γάμμα. Έχουμε, λοιπόν, πάρει :

$$\beta, \eta | b, u, X \sim N_p \gamma(b^*, B^{*-1}, u^*, v^*) \quad (67)$$

όπου $b^* = (B_0 + {}^tXX)^{-1}(B_0 b_0 + {}^tXXb)$, $B^* = B_0 + {}^tXX$, $v^* = v + v_0 + p_0$

και

$$u^* = \frac{1}{v^*} \left\{ {}^t b_0 B_0 b_0 + {}^t b ({}^tXX) b - {}^t b^* B^* b^* + v u + v_0 u_0 \right\}.$$

Υπενθυμίζουμε ότι $p_0 = rg(B_0) = p$ αν η εκ των προτέρων είναι μια «propre» πυκνότητα.

Οι εκ των προτέρων περιθωριακές είναι λοιπόν για το β μια p – διάστατη Student και για το η μια Γάμμα κατανομή :

$$\beta | b, u, X \sim t_p(b^*, u^* B^{*-1}, v^*) \quad (68)$$

$$\eta | b, u, X \sim \Gamma\left(\frac{v^*}{2}, \frac{2}{v^* u^*}\right) \quad (69)$$

Ιδιαίτερος,

$$E(\beta | b, u, X) = b^*, \quad v^* > 1$$

$$Var(\beta | b, u, X) = \frac{v^*}{v^* - 2} u^* B^{*-1}, \quad v^* > 2.$$

Αν β_1, β_2 είναι συντεταγμένες του β , η περιθωριακή του β_1 ή η δεσμευμένη του $\beta_1 | \beta_2$ προκύπτουν αμέσως από τις ιδιότητες της p – διάστατης Student (βλέπε Παράρτημα 8).

Παραδείγματος χάριν, αν ενδιαφερόμαστε για την περιθωριακή μιας συντεταγμένης β_j του β , έχουμε :

$$\beta_j \sim t(b_j^*, u^* b^{j*}, v^*) \quad (70)$$

όπου b^{j*} είναι το (j, j) στοιχείο του πίνακα B^{*-1} .

Στην περίπτωση μιας εκ των προτέρων μη – πληροφορικής $(B_0, p_0, u_0, v_0 \rightarrow 0)$ επαληθεύεται ότι :

$$B^* \rightarrow {}^tXX, \quad b^* \rightarrow b, \quad u^* \rightarrow u, \quad v^* \rightarrow v.$$

Μπορούμε, επίσης, να επαληθεύσουμε ότι σ' αυτήν την περίπτωση οι εκ των υστέρων νόμοι γίνονται :

$$\begin{aligned} \beta, \eta | b, u, X &\sim N_p \gamma(b, ({}^tXX)^{-1}, u, v) \\ \beta | b, u, X &\sim t_p(b, u({}^tXX)^{-1}, v) \\ \eta | b, u, X &\sim \Gamma\left(\frac{v}{2}, \frac{2}{uv}\right). \end{aligned}$$

d) Predictive νόμος

Η εκ των προτέρων *predictive* του y λαμβάνεται συνδυάζοντας την εκ των προτέρων επί του (β, η) και την πιθανοφάνεια του y για έναν πίνακα X δεδομένο. Πράγματι, λόγω της ιδιαίτερης κατασκευής της Κανονικής – Γάμμα, μπορούμε να εκμεταλλευτούμε τα αποτελέσματα που έχουμε ήδη πάρει για την περίπτωση που το η είναι γνωστό. Έχουμε:

$$\begin{aligned} \beta | \eta, X &\sim N_p(b_0, (\eta B_0)^{-1}) \\ y | \beta, \eta, X &\sim N_n(X\beta, \eta^{-1}I_n). \end{aligned}$$

Λαμβάνουμε τότε, όπως ανωτέρω στην (61) την κατανομή του $y | \eta, X$:

$$y | \eta, X \sim N_n(Xb_0, \eta^{-1}(I_n + XB_0^{-1}{}^tX))$$

Επειδή $\eta \sim \eta | X \sim \Gamma\left(\frac{v_0}{2}, \frac{2}{u_0 v_0}\right)$, έχουμε (βλέπε Παράρτημα 7,8) ότι η από κοινού κατανομή για το $(y, \eta) | X$ είναι η Κανονική – Γάμμα:

$$(y, \eta) | X \sim N_n \gamma(Xb_0, (I_n + XB_0^{-1}{}^tX), u_0, v_0)$$

Η εκ των προτέρων *predictive* για το y είναι η αντίστοιχη περιθωριακή, δηλαδή μια n -διάστατη Student :

$$y | X \sim t_n(Xb_0, u_0(I_n + XB_0^{-1}{}^tX), v_0) \quad (71)$$

Πρέπει και εδώ η εκ των προτέρων να έχει μια πυκνότητα, δηλαδή B_0 αντιστρέψιμη και $u_0, v_0 > 0$, ώστε αυτή η εκ των προτέρων *predictive* να έχει μια «priorre» πυκνότητα.

Έχουμε τότε τις εκ των προτέρων predictive ροπές του y :

$$E(y | X) = XB_0$$

$$Var(y | X) = \frac{v_0}{v_0 - 2} u_0 (I_n + XB_0^{-1} X), \quad v_0 > 2.$$

Για να προκύψει η εκ των υστέρων predictive πυκνότητα του y_f όταν έχουμε n_f νέες τιμές που γεννιούνται από την ίδια διαδικασία γραμμικής παλινδρόμησης (39) και για ένα δεδομένο πίνακα X_f , (n_f, p) επεξηγηματικών μεταβλητών, αρκεί να αντικαταστήσουμε το (b_0, B_0, u_0, v_0) με το (b^*, B^*, u^*, v^*) και τον X με τον X_f στη σχέση (71):

$$y_f | y, X, X_f \sim t_{n_f} (X_f b^*, u^* [I_{n_f} + X_f B^{*-1} X_f], v^*) \quad (72)$$

όπου εδώ ο B^* υποτίθεται αντιστρέψιμος.

Παράδειγμα : Αν η εκ των προτέρων είναι μη – πληροφορική η εκ των υστέρων predictive του y_f χρησιμοποιεί μόνο την πληροφορία του δείγματος, αν ο X είναι πλήρης τάξεως ($rg(X) = p$) αυτή δίνεται από την

$$y_f | y, X, X_f \sim t_{n_f} (X_f b, u [I_{n_f} + X_f ({}^t X X)^{-1} X_f], v) \quad (73)$$

λαμβάνοντας υπόψη ότι

$$\beta, \eta | y, X \sim \beta, \eta | b, u, X \sim N_p \gamma(b, ({}^t X X)^{-1}, u, v)$$

και

$$y_f | \beta, \eta, X_f \sim N_{n_f} (X_f \beta, n^{-1} I_{n_f}).$$

ε) Μερική περίπτωση: Η μονοδιάστατη Κανονική διαδικασία με άγνωστη διασπορά

Σ' αυτήν τη μερική περίπτωση όπου $p = 1$, $X = i_n = {}^t(1, \dots, 1)$, ${}^t X X = n$, παίρνουμε όλα τα επιθυμητά αποτελέσματα από αυτά που προηγήθησαν. Εδώ το β είναι βαθμωτό και $b = \bar{y}$. Σημειώνοντας με $y = {}^t(y_1, \dots, y_n)$ έχουμε την πιθανοφάνεια :

$$y | \beta, \eta \sim N_n(i_n \beta, \eta^{-1} I_n)$$

Οι εκ των προτέρων χαρακτηριστικές είναι οι ακόλουθες :

$$\left. \begin{array}{l} \beta | \eta \sim N_1 \left(b_0, \frac{1}{\eta B_0} \right) \\ \eta \sim \Gamma \left(\frac{\nu_0}{2}, \frac{2}{u_0 \nu_0} \right) \end{array} \right\} \leftrightarrow \begin{cases} (\beta, \eta) \sim N_1 \gamma(b_0, B_0^{-1}, u_0, \nu_0) \\ \beta \sim t_p(b_0, u_0 B_0^{-1}, \nu_0) \end{cases}$$

Για την εκ των υστέρων με $\nu = n - 1$ και $u = s^2 = \frac{1}{\nu} \sum_{i=1}^n (y_i - \bar{y})^2$ έχουμε :

$$\begin{aligned} (\beta, \eta) | y &\sim N_1 \gamma(b^*, B^{*-1}, u^*, \nu^*) \\ \eta | y &\sim \Gamma \left(\frac{\nu^* u^*}{2}, \frac{\nu^*}{2} \right) \\ \beta | y &\sim t_p(b^*, u^* B^{*-1}, \nu^*), \end{aligned}$$

όπου

$$b^* = \frac{B_0 b_0 + n \bar{y}}{n + B_0}, B^* = B_0 + n, \nu^* = \nu + \nu_0 + p_0, u^* = \frac{1}{\nu^*} \{B_0 b_0^2 + n \bar{y}^2 - B^* b^{*2} + \nu u + \nu_0 u_0\}$$

με $p_0 = 1$ αν η εκ των προτέρων είναι «propre» και $p_0 = 0$ για μια εκ των προτέρων μη – πληροφορική ($B_0, u_0, \nu_0 \rightarrow 0$).

Αν ενδιαφερόμαστε για μια εκ των υστέρων predictive μιας «μελλοντικής» τιμής y_f της διαδικασίας, έχουμε :

$$y_f | y \sim t \left(b^*, u^* \left(1 + \frac{1}{B^*} \right), \nu^* \right)$$

Στην ιδιαίτερη περίπτωση μιας εκ των προτέρων μη – πληροφορικής θα είχαμε :

$$\begin{aligned} y_f | y &\sim t \left(\bar{y}, s^2 \left(1 + \frac{1}{n} \right), \nu \right) \\ E(y_f | y) &= \bar{y} \\ \text{Var}(y_f | y) &= \frac{\nu}{\nu - 2} s^2 \left(1 + \frac{1}{n} \right). \end{aligned}$$

III. Σύμφυτες απώλειες

Στην bayesienne ανάλυση έχει αναπτυχθεί, σχετικά με τα μη-πληροφορικά μοντέλα, ένας συγκεκριμένος τύπος εκ των προτέρων νόμων. Αυτοί οι εκ των προτέρων νόμοι λέγονται, όπως αναφέραμε στο Κεφάλαιο I, μη – πληροφορικοί και μπορούν να καθοριστούν αυτόματα από την κατανομή του δείγματος. Οι ιδιότητες τους είναι αρκετά γνωστές και συχνά ανταποκρίνονται στις κλασσικές διαδικασίες (εκτιμητής μεγίστης πιθανοφάνειας, καλύτερος αμερόληπτος εκτιμητής,...). Επιπλέον, παρέχουν μια βάση για περισσότερες υποκειμενικές προσεγγίσεις ενός δοθέντος προβλήματος συμπερασματολογίας με την έννοια ότι υπάρχει μια πλατιά γνώμη μεταξύ των στατιστικών σχετικά με το συνολικό κύρος τέτοιων εκ των προτέρων αναφορών.

Είναι εκπληκτικό ότι πολύ λίγα έχουν γίνει σχετικά με την ισότιμη παραγωγή βασικών απωλειών που μπορούν να χρησιμοποιηθούν όταν η συνάρτηση ωφέλειας που σχετίζεται με το πρόβλημα συμπερασματολογίας δεν μπορεί να δοθεί από στατιστικούς. Για αυτό το λόγο ο C.Robert (1996) προτείνει, όπως θα αναπτύξουμε στη συνέχεια, την κατασκευή ορισμένων εναλλακτικών των κλασσικών απωλειών, έτσι ώστε να προκύψουν απώλειες που μπορούν να θεωρηθούν μη-πληροφορικές όπως οι αντίστοιχες εκ των προτέρων κατανομές. Αυτές οι απώλειες ονομάζονται «σύμφυτες» γιατί παράγονται αυτόματα από την κατανομή του δείγματος χωρίς υποκειμενική εισαγωγή δεδομένων.

1. Συνήθης συνάρτηση απώλειας και επικρίσεις

Εδώ αναφερόμαστε στη συνηθισμένη πρακτική της Στατιστικής Θεωρίας Αποφάσεων και στα σχετικά μειονεκτήματα χρήσης κλασσικών απωλειών μαζί με την έννοια των μη-πληροφορικών εκ των προτέρων νόμων.

Η Στατιστική Θεωρία Αποφάσεως βασίζεται σε δύο κύριες έννοιες: *την κατανομή του δείγματος* $f(x|\theta)$, $x \in X$, $\theta \in \Theta$ και *τη συνάρτηση απώλειας* $L(\theta, d)$, $d \in D$ (D ο χώρος των αποφάσεων). Οι στατιστικές διαδικασίες τότε εξετάζονται και συγκρίνονται σε όρους της *συναρτήσεως κινδύνου* :

$$R(\theta, \delta) = E_{\theta}[L(\theta, \delta)] = \int_x L(\theta, \delta(x))f(x|\theta)v(dx)$$

και οι έννοιες της optimalite , όπως η «admissibilite» και « minimaxite» ορίζονται σε σχέση με μια συγκεκριμένη συνάρτηση απώλειας (βλέπε π.χ. στο [12]). Η bayesienne μέθοδος απαιτεί επιπρόσθετα μια *εκ των προτέρων κατανομή* π επί του Θ , η οποία παρέχει ένα καλά ορισμένο (και συνήθως μοναδικό) *εκτιμητή αποφάσεως* (ή *εκτιμητή του Bayes*) δ^{π} , όπου $\delta^{\pi}(x)$ είναι η λύση της:

$$\min_d E^{\pi}[L(\theta, d) | x] = \min_d \int_{\Theta} L(\theta, d)\pi(\theta | x)d\theta$$

όπου $\pi(\theta | x)$ σημειώνει την *εκ των υστέρων πυκνότητα* του θ ($\pi(\theta | x) \propto f(x|\theta)\pi(\theta)$).

Ένα κλειδί παράμετρος της Στατιστικής Θεωρίας Αποφάσεων είναι ως εκ τούτου η συνάρτηση απώλειας $L(\theta, d)$ και αυτό το στατιστικό παράδειγμα συχνά δέχεται επίθεση λόγω του αυθαιρέτου της συνάρτησης απώλειας στις περισσότερες μελέτες της Θεωρίας

Αποφάσεων. Πράγματι, αρχικά με τον Gauss (1810) η εισαγωγή της τετραγωνικής απώλειας γίνεται για λόγους ευκολίας παρά πραγματικά βασιζόμενη σε μια προκαταρκτική χρήσιμη ανάλυση. Αυτό το στοιχείο της αυθαιρεσίας στις συμπερασματικές μεθόδους εμφανίζεται ακόμη στο μοντέλο της bayesienne ανάλυσης.

Μια συνάρτηση απώλειας που χρησιμοποιείται συχνά είναι βεβαίως η τετραγωνική απώλεια $L(\theta, d) = (h(\theta) - d)^2$ για τη μονοδιάστατη περίπτωση και $L(\theta, d) = \|h(\theta) - d\|^2$ για τη διανυσματική περίπτωση. Ο κύριος λόγος για τη δημοτικότητα της είναι ότι οι τετραγωνικές απώλειες οδηγούν σε εύλογους υπολογισμούς για πολλά προβλήματα σημειοεκτιμητικής. Συγκεκριμένα, ο εκτιμητής του Bayes που συνδέεται με την εκ των προτέρων κατανομή π και με μια τετραγωνική συνάρτηση απώλειας είναι η εκ των υστέρων μαθηματική ελπίδα : $E^\pi(h(\theta) | x)$ που φαίνεται ένας φυσικός εκτιμητής για να εμποδίσει στην πράξη την παραγωγή μιας αληθινής συνάρτησης ωφέλειας. Ένας δεύτερος λόγος για τη χρησιμοποίηση των τετραγωνικών απωλειών είναι η αρχική σύγκυση μεταξύ του *σφάλματος* με την έννοια της μεταβλητότητας που συνυπάρχει σε διάφορα φαινόμενα, συνήθως παρουσιάζεται από τη διασπορά $E_\theta(h(\theta) - \delta(x))^2$ όταν ο δ είναι αμερόληπτος εκτιμητής για το $h(\theta)$ και του *σφάλματος* με την έννοια του συμπεράσματος μιας εσφαλμένης απόφασης δηλαδή της προσέγγισης της αληθινής τιμής του θ από το δ .

Η συστηματική προσφυγή στις τετραγωνικές απώλειες έχει επικριθεί κατά πολύ στην βιβλιογραφία (βλέπε Lehmann και Casella, 1995) ενώ μερικοί έχουν προτείνει πιο ισχυρά υποκατάστατα απωλειών, όπως την «κολοβή» απώλεια του Huber:

$$L(\theta, d) = \begin{cases} \|\theta - d\|^2 & \text{αν } \|\theta - d\|^2 \leq c^2 \\ 2c\|\theta - d\| - c^2 & \text{αλλιώς} \end{cases}$$

Άλλοι περισσότερο ριζοσπαστικοί, απορρίπτουν εντελώς την Θεωρία Αποφάσεων με βάση το γεγονός ότι η συνάρτηση απώλειας δεν μπορεί να εκτιμηθεί στην πράξη. (Σημειώνουμε την ομοιότητα με την απόρριψη της Bayesienne προσέγγισης επειδή οι εκ των προτέρων κατανομές καθορίζονται συνήθως δύσκολα). Για παράδειγμα, αυτό είναι ένα μεγάλο επιχείρημα των υποστηρικτών της «εγγύτητας του Pitman» ένα κριτήριο που παρουσιάστηκε από τον Pitman (1937), το οποίο συγκρίνει δύο εκτιμητές δ_1 και δ_2 μέσω της πιθανοφάνειας :

$$P_\theta(\|\theta - \delta_1(x)\| \leq \|\theta - \delta_2(x)\|)$$

Οι υποστηρικτές αυτής της προσέγγισης θεωρούν ότι η *εγγύτητα του Pitman* είναι περισσότερο «σύμφυτη» από μια συγκεκριμένη συνάρτηση απώλειας επειδή εμπλέκει την κατανομή του εκτιμητή.

Βεβαίως πολλοί είναι εκείνοι που ενώ συμφωνούν στην απόρριψη μιας αυτόματης απώλειας όπως η τετραγωνική απώλεια, εναντιώνονται ισχυρά στην απόρριψη της Θεωρίας Αποφάσεων επειδή αυτή αποτελεί το κατάλληλο αξιωματικό σύστημα για την ανάλυση προβλημάτων συμπεραματολογίας και τη συναφή σύγκριση των διαθέσιμων μεθόδων. Μια συνάρτηση απώλειας είναι τότε αναγκαία.

Πολλά αξιωματικά συστήματα είναι διαθέσιμα στη βιβλιογραφία που υποστηρίζουν την παραγωγή μιας συναρτήσεως ωφέλειας/απώλειας με διάφορους βαθμούς γενικότητας. Ο Rubin (1987) προτείνει ένα γενικό αξιωματικό σύστημα στο οποίο η ωφέλεια και η εκ των προτέρων κατανομή δε διαχωρίζονται, αλλά τα οποία χορηγούνται

για μια απόφαση. Έτσι τέτοιες διαδικασίες εξασφαλίζουν την ύπαρξη μιας συγκεκριμένης απώλειας, αλλά η πραγματική παραγωγή της μπορεί να είναι αρκετά δυσκίνητη και χρονοβόρα. Σε άλλες περιπτώσεις, τα αξιώματα αυτών των θεωριών ωφέλειας δεν ικανοποιούνται απαραίτητα ή είναι δύσκολο να ελεγχθούν. Υπάρχει λοιπόν πραγματική ανάγκη για *εναλλακτικές συναρτήσεις απώλειας* που μπορούν να δουλέψουν σε τέτοια μοντέλα. Αυτή η ανάγκη περιλαμβάνει πολύ συγκεκριμένες μελέτες όπου ο στατιστικός δεν είναι ο αποφασιστής και δεν μπορεί να εξασφαλίσει μια ακριβή συνάρτηση απώλειας είτε εξαιτίας ελλείψεως χρόνου και χρήματος είτε εξαιτίας προβλημάτων επικοινωνίας με τον αποφασιστή.

Σ' αυτό το σημείο παρατηρείται η ισχυρή ομοιότητα με την bayesienne παραγωγή μη-πληροφορικών εκ των προτέρων νόμων που χρησιμοποιούνται όταν καμία εκ των προτέρων πληροφορία δεν είναι διαθέσιμη και οι οποίοι νόμοι παράγονται από την κατανομή του δείγματος $f(x|\theta)$ όπως στη μέθοδο Jeffrey του Κεφαλαίου I. Η επέκταση του bayesian παραδείγματος είναι πράγματι απαραίτητη για να υποδείξουμε μια ολοκληρωμένη συμπερασματική μέθοδο που να οδηγεί σε μια σαφή απόφαση (με την έννοια ότι δεν υπάρχει αυθαιρεσία στην επιλογή της διαδικασίας απόφασης). Τα τελευταία χρόνια υπάρχει αξιοσημείωτο ενδιαφέρον για τις μη-πληροφορικές εκ των προτέρων κατανομές και νέες έρευνες έχουν δείξει την καλή εκτέλεση κάποιων μη-πληροφορικών εκ των προτέρων κατανομών, τόσο από τη συμπερασματική άποψη όσο και από την bayesienne άποψη. Η στάνταρ μέθοδος σ' αυτή την περιοχή καλούμενη «μέθοδος της εκ των προτέρων πληροφορίας» που αναπτύχθηκε από τον Bernardo(1979), συνδέεται με τη γνώση της συνάρτησης απώλειας αφού η εκ των προτέρων πληροφορία επιλέγεται έτσι ώστε να μεγιστοποιεί την αναμενόμενη ωφέλεια του πειράματος. Η τελευταία μη-πληροφορική μέθοδος στην Bayesianne Θεωρία Αποφάσεων, καθιστά αναγκαία την ταυτόχρονη παραγωγή μιας εκ των προτέρων κατανομής π και μιας συνάρτησης απώλειας L που βασίζεται στη μόνη διαθέσιμη πληροφορία καλούμενη δειγματική κατανομή $f(\cdot|\theta)$.

2. Απαιτήσεις για μια μη-πληροφορική συνάρτηση απώλειας

Θα αποκαλούμε *σύμφυτες απώλειες* τις συναρτήσεις απώλειας που παράγονται από την κατανομή του δείγματος με αυτόματο τρόπο. Μπορούν να χρησιμοποιηθούν με ακριβώς τον ίδιο τρόπο που χρησιμοποιούνται οι συνηθισμένες συναρτήσεις απώλειας και ιδιαιτέρως βοηθούν στον καθορισμό μη-πληροφορικών εκ των προτέρων κατανομών όπως οι εκ των προτέρων αναφορές. Εφόσον ολόκληρος ο σκοπός αυτής της μεθόδου είναι να αποφύγει πιθανή εισαγωγή δεδομένων από τον αποφασιστή (έτσι ώστε να υπάρχει περισσότερη αντικειμενικότητα) η αυτόματη παραγωγή της L πρέπει να ακολουθείται από απρόσβλητες για τους στατιστικούς αρχές επί των οποίων υπάρχει ισχυρή συμφωνία.

Αν αναλογιστούμε τη μακρά ιστορία των μη-πληροφορικών εκ των προτέρων νόμων μπορούμε να δούμε ότι τα πρώτα επιχειρήματα του Laplace στις αρχές του 19^{ου} αιώνα, αργότερα απορρίφθηκαν από τους Venn και Bertrand βασιζόμενοι σε δικές τους αυθαιρεσίες. Μια από τις κύριες επικρίσεις της μη-πληροφορικής προσέγγισης του Laplace είναι η απουσία του αναλλοίωτου υπό την παραμετρικοποίηση. Για παράδειγμα, ο Laplace πρότεινε τη χρήση ομοιομόρφου εκ των προτέρων κατανομής για φραγμένους παραμετρικούς χώρους. Έτσι, αν $x \sim Bin(n, p)$ με $p \in [0,1]$, η εκ των προτέρων του Laplace στο p θα είναι $\pi(p) = 1$ (δηλαδή $p \sim \pi = U(0,1)$). Ενώ μια αλλαγή μεταβλητής από το p στο $q(=\sqrt{p})$ συνεπάγεται ότι $x \sim Bin(n, q^2)$ με $q \in [0,1]$ και μια άμεση

εφαρμογή των κανόνων του Laplace συνεπάγει ότι η εκ των προτέρων μετασχηματισμένη πρέπει να είναι $\pi(q) = 2q$. Άρα η προσέγγιση του Laplace δεν μπορεί να διατηρηθεί αν δεν υπάρχει προτιμητή παραμετρικοποίηση του μοντέλου. Μετέπειτα εξελίξεις των μη-πληροφορικών εκ των προτέρων νόμων, όπως οι εκ των προτέρων του Jeffrey, σκόπευαν να καταβάλουν αυτό το μεγάλο μειονέκτημα.

Είναι λοιπόν λογικό να επιβάλλουμε τη συνθήκη του αναλλοίωτου (invariante) υπό την παραμετρικοποίηση ώστε να ισχύει επίσης για τη συνάρτηση απώλειας : αν $x \sim f(x|\theta)$ και το θ εκτιμάται από το δ , υπό τη συνάρτηση απώλειας $L(\theta, \delta)$, μια 1-1 αλλαγή της παραμέτρου θ σε $\eta = \psi(\theta)$ έτσι ώστε $x \sim g(x|\eta)$ πρέπει να οδηγεί στην ίδια απώλεια $L(\theta, \delta)$ αν το η εκτιμάται από το $\psi(\delta)$, δηλαδή $L(\theta, \delta) = L(\eta, \psi(\delta))$, διότι $f(x|\theta) = g(x|\eta)$ και $f(x|\delta) = g(x|\psi(\delta))$.

Συνεπώς οι σύμφυτες απώλειες θα είναι *ελεύθερης παραμετρικοποίησης*. Αυτός ο περιορισμός είναι λογικός αν κάποιος θεωρήσει ότι τέτοιες τοποθετήσεις οφείλουν να μην επιτρέπονται για ειδικές παραμετρικοποιήσεις : σ' αυτή την περίπτωση ο κύριος σκοπός της συμπερασματικής ανάλυσης μπορεί να είναι μόνο ο οραματισμός διαμέσου μιας «σφαιρικής» συνάρτησης ωφέλειας η οποία να ερμηνεύει την εκτίμηση υπό την έννοια της πρόβλεψης δηλαδή μέσω των συμπερασμάτων, σαν μια προσέγγιση της κατανομής μελλοντικών παρατηρήσεων. Επομένως η πυκνότητα $f(x|\theta)$ είναι τότε ταυτόσημη με την $g(x|\eta)$ όσον αφορά τις πιθανοθεωρητικές ιδιότητές τους.

Η ανάγκη της ελεύθερης παραμετρικοποίησης υποδηλώνει ότι ο εκτιμητής πρέπει να είναι *equivariant* : για κάθε 1-1 μετασχηματισμό $h(\theta)$ της παραμέτρου θ , ο εκτιμητής του $h(\theta)$ θα πρέπει να είναι ο μετασχηματισμός υπό την h του εκτιμητή του θ . Ιδιαίτερως αυτή θα είναι η περίπτωση των εκτιμητών του Bayes (§6). Όμως, ο περιορισμός του αναλλοίωτου επί της συναρτήσεως απώλειας περιορίζει ισχυρά την επιλογή της L , επειδή οι συνήθεις απώλειες οδηγούν συνήθως σε εξαρτώμενους από την παραμετρικοποίηση εκτιμητές του Bayes. (Σημειώνουμε ότι η ιδιότητα του αναλλοίωτου ικανοποιείται από τους εκτιμητές μέγιστης πιθανοφάνειας, παρόλο που αυτοί δε συνδέονται με τις συναρτήσεις απώλειας).

Το αποτέλεσμα της εκτίμησης του θ από το δ πράγματι εκτιμάται μέσω της διαφοράς μεταξύ $f(.|\theta)$ και $f(.|\delta)$, για αυτό οι σύμφυτες απώλειες πρέπει να εκφράζονται από τη μέση απόσταση μεταξύ των $f(.|\theta)$ και $f(.|\delta)$, $d(f(.|\theta), f(.|\delta))$. Αυτός ο περιορισμός δεν υποβαθμίζει την επιλογή της L σε μια μόνο δυνατότητα, αλλά θα δούμε στις επόμενες παραγράφους ότι οι δύο φυσικές υποψήφια για την d , δηλαδή αποστάσεις Hellinger και Kullback-Leibler, οδηγούν σε παρόμοιες παραστάσεις των εκτιμητών του Bayes.

3. Αποστάσεις μεταξύ κατανομών

Παρόλο που το πρόβλημα αξιολόγησης των εκτιμητών στις μη-πληροφορικές συναρτήσεις ωφέλειας έχει πράγματι μετατραπεί στον καθορισμό μιας απόστασης μεταξύ των δύο κατανομών $f(.|\theta)$ και $f(.|\delta)$, υπάρχει ακόμη κάποιο υπόλειμμα αμφιβολίας γύρω από την εκλογή μιας απόστασης $d(f(.|\theta), f(.|\delta))$. Όμως, αν και πολλές εκλογές είναι πράγματι δυνατές, μια φυσική υποψήφια είναι η Hellinger-απόσταση η οποία χρησιμοποιείται στη Θεωρία Συνάφειας του Le Cam (1986).

Δοθέντων δύο πυκνοτήτων $f(.|\theta)$ και $f(.|\delta)$ αναφορικά με ένα κυρίαρχο μέτρο ν επί του χώρου των παρατηρήσεων $X(\subseteq R^n, n \geq 1)$, η Hellinger-απόσταση $H(f(.|\theta), f(.|\delta))$, ορίζεται από τη σχέση:

$$H^2(f(\cdot|\theta), f(\cdot|\delta)) = \frac{1}{2} \int_X \left(\sqrt{f(x|\theta)} - \sqrt{f(x|\delta)} \right)^2 \nu(dx) \quad (1)$$

και είναι πάντοτε καλά ορισμένη επειδή ο $\sqrt{f(x|\theta)}$ και $\sqrt{f(x|\delta)}$ είναι στον $L^2(\nu)$. Πράγματι,

$$\begin{aligned} H^2(f(\cdot|\theta), f(\cdot|\delta)) &= \frac{1}{2} \left\{ \int_X f(x|\theta) \nu(dx) + \int_X f(x|\delta) \nu(dx) - 2 \int_X \sqrt{f(x|\theta)} \sqrt{f(x|\delta)} \nu(dx) \right\} \\ &= 1 - \int_X \sqrt{f(x|\theta)} \sqrt{f(x|\delta)} \nu(dx) \end{aligned} \quad (2)$$

Επειδή $\sqrt{f(\cdot|\theta)}, \sqrt{f(\cdot|\delta)} \in L^2(\nu)$, από την ανισότητα του Schwarz έχουμε,

$$\int_X \sqrt{f(x|\theta)} \sqrt{f(x|\delta)} \nu(dx) \leq \left(\int_X f(x|\theta) \nu(dx) \right)^{1/2} \left(\int_X f(x|\delta) \nu(dx) \right)^{1/2} = 1.$$

Λοιπόν το ολοκλήρωμα της (1) υπάρχει πάντοτε.

Από την (2) προκύπτει ότι η H φράσσεται από την 1. Αυτή είναι επίσης ισοδύναμη με την L^1 -απόσταση, αφού

$$\begin{aligned} H^2(f(\cdot|\theta), f(\cdot|\delta)) &\leq \frac{1}{2} \int_X |f(x|\theta) - f(x|\delta)| \nu(dx) \left(= \frac{1}{2} \|f(\cdot|\theta) - f(\cdot|\delta)\|_1 \right) \\ &\leq \sqrt{2} H(f(\cdot|\theta), f(\cdot|\delta)) \end{aligned} \quad (3)$$

Πράγματι, θέτοντας $f(\cdot|\theta) = f$ και $f(\cdot|\delta) = g$, έχουμε :

$$\begin{aligned} H^2(f, g) &= \frac{1}{2} \int_X \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 \nu(dx) = \frac{1}{2} \int_X \left| \sqrt{f(x)} - \sqrt{g(x)} \right| \left| \sqrt{f(x)} - \sqrt{g(x)} \right| \nu(dx) \\ &\leq \frac{1}{2} \int_X \left| \sqrt{f(x)} - \sqrt{g(x)} \right| \left| \sqrt{f(x)} + \sqrt{g(x)} \right| \nu(dx) = \frac{1}{2} \int_X |f(x) - g(x)| \nu(dx) \end{aligned}$$

Όμως από την ανισότητα του Schwarz,

$$\begin{aligned} \int_X \left| \sqrt{f(x)} - \sqrt{g(x)} \right| \left| \sqrt{f(x)} + \sqrt{g(x)} \right| \nu(dx) &\leq \left(\int_X \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 \nu(dx) \right)^{1/2} \left(\int_X \left(\sqrt{f(x)} + \sqrt{g(x)} \right)^2 \nu(dx) \right)^{1/2} \\ &= \sqrt{2} H(f, g) \left(\int_X \left(\sqrt{f(x)} + \sqrt{g(x)} \right)^2 \nu(dx) \right)^{1/2}. \end{aligned}$$

Λοιπόν,

$$H^2(f, g) \leq \frac{1}{2} \int_X |f(x) - g(x)| \nu(dx) \leq \frac{\sqrt{2}}{2} H(f, g) \left(\int_X \left(\sqrt{f(x)} + \sqrt{g(x)} \right)^2 \nu(dx) \right)^{1/2}$$

$$= H(f, g) \left(\frac{1}{2} \int_X (\sqrt{f(x)} + \sqrt{g(x)})^2 \nu(dx) \right)^{1/2}.$$

Λαμβάνοντας υπόψη τη σχέση (2) έχουμε ότι :

$$\frac{1}{2} \int_X (\sqrt{f(x)} + \sqrt{g(x)})^2 \nu(dx) = 1 + \int_X \sqrt{f(x)} \sqrt{g(x)} \nu(dx) = 2 - H^2(f, g) \leq 2.$$

Συνεπώς, από τις δύο τελευταίες ανισότητες προκύπτει η σχέση (3).

Η Hellinger-απόσταση σχετίζεται με την K^2 -απόσταση που ορίζεται από τη σχέση :

$$K^2(f(\cdot|\theta), f(\cdot|\delta)) = \frac{1}{2} \int_X \frac{(f(x|\theta) - f(x|\delta))^2}{f(x|\theta) + f(x|\delta)} \nu(dx) \quad (4)$$

επειδή

$$H^2(f(\cdot|\theta), f(\cdot|\delta)) \leq K^2(f(\cdot|\theta), f(\cdot|\delta)) \leq 2H^2(f(\cdot|\theta), f(\cdot|\delta)) \quad (5)$$

Πράγματι, σύμφωνα με τον Le Cam (1986, σελ.47-48), χρησιμοποιώντας τις παραπάνω έννοιες f και g , έχουμε :

$$f(x) + g(x) \leq (\sqrt{f(x)} + \sqrt{g(x)})^2 \leq 2(f(x) + g(x)), \quad \forall x \in X.$$

Πολλαπλασιάζοντας όλα τα μέλη της ανισότητας αυτής με το $\frac{(\sqrt{f(x)} - \sqrt{g(x)})^2}{f(x) + g(x)}$, λαμβάνουμε

$$(\sqrt{f(x)} - \sqrt{g(x)})^2 \leq \frac{(f(x) - g(x))^2}{f(x) + g(x)} \leq 2(\sqrt{f(x)} - \sqrt{g(x)})^2, \quad \forall x \in X$$

Ολοκληρώνοντας στη συνέχεια τη σχέση αυτή αναφορικά με το μέτρο ν , προκύπτει :

$$\int_X (\sqrt{f(x)} - \sqrt{g(x)})^2 \nu(dx) \leq \int_X \frac{(f(x) - g(x))^2}{f(x) + g(x)} \nu(dx) \leq 2 \int_X (\sqrt{f(x)} - \sqrt{g(x)})^2 \nu(dx)$$

δηλαδή

$$2H^2(f, g) \leq 2K^2(f, g) \leq 4H^2(f, g)$$

απ' όπου η σχέση (5).

Να σημειώσουμε ότι η K^2 - απόσταση έχει χρησιμοποιηθεί εκτός από τον Le Cam (1982, 1986) και από άλλους συγγραφείς.

Μια δεύτερη δυνατότητα η οποία είναι ένας σοβαρός ανταγωνιστής της Hellinger απόστασης είναι η *εντροπία* (*entropie*) ή *Kullback-Leibler-απόσταση* που προτείνεται

στην επεξεργασία σήματος, στη στατιστική μηχανική και στην οικονομετρία για την εκλογή του μοντέλου. Εδώ αυτή η απόσταση ορίζεται ως :

$$E(f(\cdot|\theta), f(\cdot|\delta)) = \int_x \log\left(\frac{f(x|\theta)}{f(x|\delta)}\right) f(x|\theta) \nu(dx) \quad (6)$$

όπου το δεξιό μέλος της (6) είναι πάντοτε θετικό. Πράγματι, επειδή η συνάρτηση $x \mapsto -\log x$ είναι κυρτή, από την ανισότητα του Jensen έχουμε :

$$\begin{aligned} E(f, g) &= E_\theta \left[\log\left(\frac{f(x)}{g(x)}\right) \right] = E_\theta \left[-\log\left(\frac{g(x)}{f(x)}\right) \right] \geq -\log E_\theta \left(\frac{g(x)}{f(x)} \right) \\ &= -\log \int_x \frac{g(x)}{f(x)} f(x) \nu(dx) \\ &= -\log \underbrace{\int_x g(x) \nu(dx)}_1 = 0 \end{aligned}$$

Να σημειώσουμε ότι, εν αντιθέσει με τις (1) και (4), η εντροπία, E , μπορεί να είναι $+\infty$ ακόμη όταν οι κατανομές στην οικογένεια $F_\theta = \{f(\cdot|\theta); \theta \in \Theta\}$ έχουν τον ίδιο φορέα. (βλέπε Παράδειγμα 1, κατωτέρω). Επιπλέον η E δεν είναι απόσταση (με τη μαθηματική έννοια) επειδή αυτή δεν είναι συμμετρική: αρκεί να παρατηρήσουμε ότι,

$$E(f, g) = \int_x \log\left(\frac{f(x)}{g(x)}\right) f(x) \nu(dx) \quad \text{ενώ} \quad E(g, f) = \int_x \log\left(\frac{g(x)}{f(x)}\right) g(x) \nu(dx)$$

Ένα κίνητρο για την E , είναι ότι αυτή είναι η *οριακή απώλεια* που σχετίζεται με την εκτίμηση μεγίστης πιθανοφάνειας επειδή, δεδομένου ενός τυχαίο δείγματος x_1, \dots, x_n από την $f(x|\theta)$, ο μέσος όρος της λογαριθμικής πιθανοφάνειας, $\frac{1}{n} \sum_{j=1}^n \log f(x_j|\delta)$ συγκλίνει σ.β. του $n \rightarrow +\infty$, στην $E_\theta(\log(f(x|\delta))) = \int_x \log f(x|\delta) f(x|\theta) dx$ δηλαδή στο αντίθετο του ελάχιστον μέρους της (6).

Παράδειγμα 1 : Θεωρούμε το γεωμετρικό μείγμα μιας Cauchy $\mathcal{C}(0,1)$ και μιας κανονικής $N(0,1)$ -κατανομής :

$$f_a(x) = K_a e^{-\frac{ax^2}{2}} (1+x^2)^{-(1-a)}, \quad 0 \leq a \leq 1, \quad x \in R$$

Παρατηρούμε ότι :

για $a=0$, $f_0(x) = K_0(1+x^2)^{-1}$ με $K_0 = \frac{1}{\pi}$ (περίπτωση της $\mathcal{C}(0,1)$)

για $a=1$, $f_1(x) = K_1 e^{-\frac{x^2}{2}}$ με $K_1 = \frac{1}{\sqrt{2\pi}}$ (περίπτωση της $N(0,1)$).

Τότε, για $a > 0$,

$$E(f_0, f_a) = \int_{-\infty}^{+\infty} \log\left(\frac{f_0(x)}{f_a(x)}\right) f_0(x) dx,$$

όπου

$$\frac{f_0(x)}{f_a(x)} = \frac{K_0(1+x^2)^{-1}}{K_a e^{-\frac{ax^2}{2}}(1+x^2)^{-a}} = \frac{K_0}{K_a} e^{\frac{ax^2}{2}} (1+x^2)^{-a}$$

και

$$\log\left(\frac{f_0(x)}{f_a(x)}\right) = \log\frac{K_0}{K_a} + \frac{ax^2}{2} - a \log(1+x^2).$$

Λοιπόν,

$$\begin{aligned} E(f_0, f_a) &= \int_{-\infty}^{+\infty} \left(-\log(\pi K_a) - a \log(1+x^2) + \frac{ax^2}{2}\right) \frac{1}{\pi} (1+x^2)^{-1} dx \\ &= -\log(\pi K_a) - \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{\log(1+x^2)}{1+x^2} dx + \frac{a}{2\pi} \int_{-\infty}^{+\infty} \frac{x^2}{1+x^2} dx \\ &= -\log(\pi K_a) - \frac{2a}{\pi} \int_0^{+\infty} \frac{\log(1+x^2)}{1+x^2} dx + \frac{a}{\pi} \int_0^{+\infty} \frac{x^2}{1+x^2} dx \end{aligned}$$

όπου τα δύο τελευταία ολοκληρώματα δεν συγκλίνουν στη γειτονιά του $+\infty$. Παρά του γεγονότος ότι η $E(f_0, f_a)$ δεν ορίζεται για $a > 0$, η οικογένεια αυτών των κατανομών μπορεί να χρησιμοποιηθεί για να αποτιμηθεί η βαρύτητα των ουρών της κατανομής. ▲

Υπάρχει επίσης μια σχέση μεταξύ της E και της H , αφού

$$\min\{E(f(\cdot|\theta), f(\cdot|\delta)), E(f(\cdot|\delta), f(\cdot|\theta))\} \geq -2 \log[1 - H^2(f(\cdot|\theta), f(\cdot|\delta))] \quad (7)$$

Πράγματι, θέτοντας όπως ανωτέρω $f(\cdot|\theta) = f$ και $f(\cdot|\delta) = g$, έχουμε :

$$\begin{aligned} 1 - H^2(f, g) &= 1 - \frac{1}{2} \int_x \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 \nu(dx) = 1 - \frac{1}{2} \int_x (f(x) + g(x) - 2\sqrt{f(x)g(x)}) \nu(dx) \\ &= \int_x \sqrt{f(x)g(x)} \nu(dx) \\ &= \int_x \frac{\sqrt{f(x)}}{\sqrt{g(x)}} g(x) \nu(dx) = E_g \left[\frac{\sqrt{f(x)}}{\sqrt{g(x)}} \right]. \end{aligned}$$

Χρησιμοποιώντας την ανισότητα του Jensen για την κυρτή συνάρτηση $x \mapsto -\log x$, λαμβάνουμε :

$$\begin{aligned}
-2\log(1-H^2(f,g)) &= -2\log E_g \left[\frac{\sqrt{f(x)}}{\sqrt{g(x)}} \right] \leq E_g \left[-2\log \sqrt{\frac{f(x)}{g(x)}} \right] = \int_x -2\log \sqrt{\frac{f(x)}{g(x)}} g(x) \nu(dx) \\
&= \int_x \log \left(\frac{g(x)}{f(x)} \right) g(x) dx = E(f,g).
\end{aligned}$$

Επίσης, επειδή η $H^2(f,g)$ είναι συμμετρική έχουμε ότι :

$$-2\log(1-H^2(f,g)) = -2\log(1-H^2(g,f)) \leq E(f,g),$$

εργαζόμενοι όπως παραπάνω. Άρα,

$$-2\log(1-H^2(f,g)) \leq \min\{E(f,g), E(g,f)\}.$$

Οι δύο αποστάσεις δε συγκρίνονται επιπλέον και δεν υπάρχει αντίστροφη της (7) δίδοντας ένα άνω φράγμα της $E(f,g)$ σε όρους της άλλης συνάρτησης $H(f,g)$. Όμως οι δύο αντίστοιχες συναρτήσεις απώλειας $L_H(\theta,\delta)$ και $L_E(\theta,\delta)$ είναι κλειστές υπό την έννοια ότι αυτές οδηγούν στην ίδια δευτέρας τάξεως προσέγγιση, μέχρι μια σταθερά, όπως φαίνεται στο παρακάτω αποτέλεσμα.

Λήμμα 3.1

Το ανάπτυγμα 2^{as} τάξεως των συναρτήσεων απώλειας L_H και L_E είναι :

$$L_H(\theta,\delta) \cong \frac{1}{8} {}^t(\theta - \delta) I(\theta) (\theta - \delta)$$

και

$$L_E(\theta,\delta) \cong \frac{1}{2} {}^t(\theta - \delta) I(\theta) (\theta - \delta)$$

όπου $I(\theta)$ είναι ο πίνακας πληροφοριών του Fisher:

$$I(\theta) = (I_{ij}(\theta))_{i,j} \text{ με } I_{ij}(\theta) = -E_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta) \right].$$

Απόδειξη : Για λόγους απλότητας παρουσιάζουμε μόνο τη μονοδιάστατη περίπτωση. Η επέκταση στον R^n είναι άμεση.

α) Θεωρούμε πρώτα την απόσταση Hellinger. Τότε, υπό ορισμένες συνθήκες κανονικότητας, λαμβάνουμε :

$$\begin{aligned}
\frac{\partial}{\partial \delta} H^2(f(\cdot|\theta), f(\cdot|\delta)) &= \frac{\partial}{\partial \delta} \left(1 - \int_x \sqrt{f(x|\theta)f(x|\delta)} \nu(dx) \right) \\
&= - \int_x \frac{1}{2\sqrt{f(x|\delta)}} \frac{\partial}{\partial \delta} f(x|\delta) \sqrt{f(x|\theta)} \nu(dx)
\end{aligned}$$

και

$$\frac{\partial^2}{\partial \delta^2} H^2(f(\cdot|\theta), f(\cdot|\delta)) = -\frac{1}{2} \frac{\partial}{\partial \delta} \int_x \frac{1}{\sqrt{f(x|\delta)}} \frac{\partial}{\partial \delta} f(x|\delta) \sqrt{f(x|\theta)} \nu(dx)$$

$$\begin{aligned}
&= -\frac{1}{2} \int_x \frac{1}{f(x|\delta)} \left[\frac{\partial^2}{\partial \delta^2} f(x|\delta) \sqrt{f(x|\delta)} - \left(\frac{\partial}{\partial \delta} f(x|\delta) \right)^2 \frac{1}{2\sqrt{f(x|\delta)}} \right] \sqrt{f(x|\theta)} v(dx) \\
&= -\frac{1}{2} \int_x \left[\frac{\partial^2}{\partial \delta^2} f(x|\delta) - \left(\frac{\partial}{\partial \delta} f(x|\delta) \right)^2 \frac{1}{2f(x|\delta)} \right] \sqrt{\frac{f(x|\theta)}{f(x|\delta)}} v(dx)
\end{aligned}$$

Επίσης,

$$\frac{\partial}{\partial \delta} H^2(f(\cdot|\theta), f(\cdot|\delta)) \Big|_{\delta=\theta} = -\frac{1}{2} \int_x \frac{\partial}{\partial \delta} f(x|\delta) v(dx) = -\frac{1}{2} \frac{\partial}{\partial \delta} \underbrace{\int_x f(x|\delta) v(dx)}_1 = 0$$

και

$$\begin{aligned}
\frac{\partial^2}{\partial \delta^2} H^2(f(\cdot|\theta), f(\cdot|\delta)) \Big|_{\delta=\theta} &= -\frac{1}{2} \int_x \left[\frac{\partial^2}{\partial \delta^2} f(x|\delta) - \frac{1}{2f(x|\delta)} \left(\frac{\partial}{\partial \delta} f(x|\delta) \right)^2 \right] v(dx) \\
&= -\frac{1}{2} \int_x \frac{\partial^2}{\partial \delta^2} f(x|\delta) v(dx) + \frac{1}{4} \int_x \frac{1}{f(x|\delta)} \left(\frac{\partial}{\partial \delta} f(x|\delta) \right)^2 v(dx) \\
&= -\frac{1}{2} \frac{\partial^2}{\partial \delta^2} \underbrace{\int_x f(x|\delta) v(dx)}_1 + \frac{1}{4} \int_x \frac{1}{f^2(x|\delta)} \left(\frac{\partial}{\partial \delta} f(x|\delta) \right)^2 f(x|\delta) v(dx) \\
&= \frac{1}{4} \int_x \left(\frac{\partial}{\partial \delta} \log f(x|\delta) \right)^2 f(x|\delta) v(dx) = \frac{1}{4} I(\delta) = \frac{1}{4} I(\theta).
\end{aligned}$$

Γνωρίζουμε ότι το ανάπτυγμα Taylor μιας συνάρτησης $f(x, y)$ σε μια γειτονιά του σημείου (a, b) είναι:

$$\begin{aligned}
f(x, y) &= f(a, b) + (x-a)f_x(a, b) + (y-b)f_y(a, b) \\
&\quad + \frac{1}{2!} \left[(x-a)^2 f_{xx}(a, b) + 2(x-a)(y-b)f_{xy}(a, b) + (y-b)^2 f_{yy}(a, b) \right] + \dots
\end{aligned}$$

Εφαρμόζοντας τον τύπο αυτό για την $L_H(\theta, \delta) = H^2(f(\cdot|\theta), f(\cdot|\delta))$ όταν $\begin{cases} x = \theta, y = \delta \\ a = \theta, b = \theta \end{cases}$ λαμβάνουμε:

$$L_H(\theta, \delta) = \frac{1}{2} (\delta - \theta)^2 \frac{\partial^2}{\partial \delta^2} L_H(\theta, \theta) + \dots = \frac{1}{2} (\delta - \theta)^2 \frac{\partial^2}{\partial \delta^2} L_H(\theta, \delta) \Big|_{\delta=\theta} + \dots = \frac{1}{8} (\delta - \theta)^2 I(\theta) + \dots$$

απ'όπου

$$L_H(\theta, \delta) \cong \frac{1}{8}(\delta - \theta)^2 I(\theta).$$

β) Για την entropie απώλεια έχουμε :

$$\begin{aligned} \frac{\partial}{\partial \delta} E(f(\cdot | \theta), f(\cdot | \delta)) &= \frac{\partial}{\partial \delta} \int_x [\log f(x | \theta) - \log f(x | \delta)] f(x | \theta) v(dx) \\ &= -\frac{\partial}{\partial \delta} \int_x \log f(x | \delta) f(x | \theta) v(dx) \\ &= -\int_x \frac{\partial}{\partial \delta} \log f(x | \delta) f(x | \theta) v(dx) \\ &= -\int_x \frac{\partial}{\partial \delta} f(x | \delta) \frac{f(x | \theta)}{f(x | \delta)} v(dx) \end{aligned}$$

και

$$\begin{aligned} \frac{\partial^2}{\partial \delta^2} E(f(\cdot | \theta), f(\cdot | \delta)) &= -\frac{\partial}{\partial \delta} \int_x \frac{\partial}{\partial \delta} f(x | \delta) \frac{f(x | \theta)}{f(x | \delta)} v(dx) \\ &= -\int_x \frac{1}{f^2(x | \delta)} \left[\frac{\partial^2}{\partial \delta^2} f(x | \delta) f(x | \delta) - \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 \right] f(x | \theta) v(dx) \\ &= -\int_x \left[\frac{\partial^2}{\partial \delta^2} f(x | \delta) - \frac{1}{f(x | \delta)} \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 \right] \frac{f(x | \theta)}{f(x | \delta)} v(dx). \end{aligned}$$

Επίσης,

$$\frac{\partial}{\partial \delta} E(f(\cdot | \theta), f(\cdot | \delta)) \Big|_{\delta=\theta} = -\int_x \frac{\partial}{\partial \delta} f(x | \delta) v(dx) = 0$$

και

$$\begin{aligned} \frac{\partial^2}{\partial \delta^2} E(f(\cdot | \theta), f(\cdot | \delta)) \Big|_{\delta=\theta} &= -\int_x \left[\frac{\partial^2}{\partial \delta^2} f(x | \delta) - \frac{1}{f(x | \delta)} \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 \right] v(dx) \\ &= \underbrace{-\int_x \frac{\partial^2}{\partial \delta^2} f(x | \delta) v(dx)}_0 + \int_x \frac{1}{f(x | \delta)} \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 v(dx) \\ &= \int_x \frac{1}{f(x | \delta)} \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 v(dx) \\ &= \int_x \frac{1}{f^2(x | \delta)} \left(\frac{\partial}{\partial \delta} f(x | \delta) \right)^2 f(x | \delta) v(dx) \\ &= \int_x \left(\frac{\partial}{\partial \delta} \log f(x | \delta) \right)^2 f(x | \delta) v(dx) \end{aligned}$$

$$= I(\delta) = I(\theta).$$

Επομένως, όπως στο α), λαμβάνοντας το ανάπτυγμα Taylor της $L_E(\theta, \delta) = E(f(\cdot | \theta), f(\cdot | \delta))$ σε μια γειτονιά του σημείου (θ, θ) , έχουμε :

$$L_E(\theta, \delta) = \frac{1}{2}(\delta - \theta)^2 I(\theta) + \dots$$

Λοιπόν,

$$E_E(\theta, \delta) \cong \frac{1}{2}(\delta - \theta)^2 I(\theta) \quad \blacktriangle$$

Από το παραπάνω Λήμμα προκύπτει ότι οι δύο συναρτήσεις απώλειας συμπεριφέρονται παρόμοια σε μια γειτονιά του θ , με την έννοια ότι είναι περίπου συμμετρικές και πλησίον μιας τετραγωνικής συνάρτησης απώλειας. Αυτό το αποτέλεσμα είναι ενθαρρυντικό επειδή αυτό σημαίνει ότι και οι δύο απώλειες μπορούν να χρησιμοποιηθούν αδιάφορα σε μια γειτονιά του θ . Είναι επίσης εκπληκτικό ότι το αποτέλεσμα του Λήμματος διατηρείται για κάθε παραμετροποίηση της κατανομής.

Θα δούμε στις παρακάτω παραγράφους ότι οι L_E και L_H παράγουν παρόμοιους εκτιμητές του Bayes σε ορισμένες ειδικές περιπτώσεις ακόμη και αν η μορφή της συνάρτησης απώλειας είναι εντελώς διαφορετική. Το γεγονός αυτό μαζί με το Λήμμα 3.1 υποδηλώνει ότι η επιλογή μεταξύ των L_E και L_H είναι ελάχιστης σημασίας. Όταν η L_E οδηγεί σε μια απλή παραγωγή εκτιμητών του Bayes (βλέπε §5), η απώλεια της εντροπίας θα είναι προτιμότερη.

4. Σύμφυτες απώλειες για τις εκθετικές οικογένειες

Γενικά, εκτός των εκθετικών οικογενειών, ο υπολογισμός Hellinger ή ακόμη και της entropie απώλειας είναι δύσκολος και συχνά αποτυγχάνει στην παραγωγή εκφράσεων αναλυτικού τύπου. Όμως αυτό δεν συμβαίνει για τις εκθετικές οικογένειες πυκνοτήτων (αναφορικά με το μέτρο ν) της μορφής :

$$f(x | \theta) = e^{\theta \bullet x - \psi(\theta)} \quad (8)$$

όπου $\theta \bullet x$ σημειώνει το σύννηθες εσωτερικό γινόμενο : $\theta \bullet x = \langle \theta, x \rangle = \theta^t x$. Στην πραγματικότητα παίρνουμε το παρακάτω αποτέλεσμα :

Πρόταση 4.1

Για τις εκθετικές οικογένειες της μορφής (8), έχουμε:

$$L_H(\theta, \delta) = 1 - \exp\left\{\psi\left(\frac{\theta + \delta}{2}\right) - \frac{\psi(\theta) + \psi(\delta)}{2}\right\} \quad (9)$$

και υπό ορισμένες συνθήκες κανονικότητας,

$$L_E(\theta, \delta) = (\theta - \delta) \bullet \nabla \psi(\theta) + \psi(\delta) - \psi(\theta) \quad (10)$$

όπου $\nabla \psi$ σημειώνει το ανάδελα της ψ .

Απόδειξη : α) Για την Hellinger απώλεια,

$$\begin{aligned}
 L_H(\theta, \delta) &= H^2(f(\cdot | \theta), f(\cdot | \delta)) = 1 - \int_X \sqrt{f(x | \theta) f(x | \delta)} v(dx) \\
 &= 1 - \int_X \{\exp\{\theta \bullet x - \psi(\theta)\} \exp\{\delta \bullet x - \psi(\delta)\}\}^{\frac{1}{2}} v(dx) \\
 &= 1 - \int_X \{\exp\{(\theta + \delta) \bullet x - (\psi(\theta) + \psi(\delta))\}\}^{\frac{1}{2}} v(dx) \\
 &= 1 - \exp\left\{-\frac{\psi(\theta) + \psi(\delta)}{2}\right\} \int_X \exp\left\{\frac{\theta + \delta}{2} \bullet x\right\} v(dx) \\
 &= 1 - \exp\left\{\psi\left(\frac{\theta + \delta}{2}\right) - \frac{\psi(\theta) + \psi(\delta)}{2}\right\} \underbrace{\int_X \exp\left\{\frac{\theta + \delta}{2} \bullet x - \psi\left(\frac{\theta + \delta}{2}\right)\right\} v(dx)}_1 \\
 &= 1 - \exp\left\{\psi\left(\frac{\theta + \delta}{2}\right) - \frac{\psi(\theta) + \psi(\delta)}{2}\right\}
 \end{aligned}$$

β) Για την entropie απώλεια,

$$\begin{aligned}
 L_E(\theta, \delta) &= E(f(\cdot | \theta), f(\cdot | \delta)) = \int_X \log\left(\frac{f(x | \theta)}{f(x | \delta)}\right) f(x | \theta) v(dx) \\
 &= \int_X \log(\exp\{\theta \bullet x - \psi(\theta)\} \exp\{-\delta \bullet x + \psi(\delta)\}) \exp\{\theta \bullet x - \psi(\theta)\} v(dx) \\
 &= \int_X \log(\exp\{(\theta - \delta) \bullet x - (\psi(\theta) - \psi(\delta))\}) \exp\{\theta \bullet x - \psi(\theta)\} v(dx) \\
 &= \int_X \{(\theta - \delta) \bullet x - (\psi(\theta) - \psi(\delta))\} \exp\{\theta \bullet x - \psi(\theta)\} v(dx) \\
 &= \int_X (\theta - \delta) \exp\{\theta \bullet x - \psi(\theta)\} v(dx) - (\psi(\theta) - \psi(\delta)) \underbrace{\int_X \exp\{\theta \bullet x - \psi(\theta)\} v(dx)}_1 \\
 &= (\theta - \delta) \bullet \int_X x \exp\{\theta \bullet x - \psi(\theta)\} v(dx) - (\psi(\theta) - \psi(\delta)) \\
 &= (\theta - \delta) \bullet E_\theta(x) - (\psi(\theta) - \psi(\delta))
 \end{aligned}$$

όμως, $E_\theta(x) = \nabla \psi(\theta)$.

Πράγματι, για τη μονοδιάστατη περίπτωση έχουμε :

$$\begin{aligned}
 \frac{\partial}{\partial \theta} \exp\{\theta x - \psi(\theta)\} &= (x - \psi'(\theta)) \exp\{\theta x - \psi(\theta)\} \\
 &= x \exp\{\theta x - \psi(\theta)\} - \psi'(\theta) \exp\{\theta x - \psi(\theta)\}
 \end{aligned}$$

κατά συνέπεια και υπό ορισμένες συνθήκες κανονικότητας, έχουμε :

$$\begin{aligned}
E_{\theta}(x) &= \int_{\mathcal{X}} x \exp\{\theta x - \psi(\theta)\} \nu(dx) \\
&= \int_{\mathcal{X}} \frac{\partial}{\partial \theta} \exp\{\theta x - \psi(\theta)\} \nu(dx) + \psi'(\theta) \underbrace{\int_{\mathcal{X}} \exp\{\theta x - \psi(\theta)\} \nu(dx)}_1 \\
&= \frac{\partial}{\partial \theta} \underbrace{\int_{\mathcal{X}} \exp\{\theta x - \psi(\theta)\} \nu(dx)}_1 + \psi'(\theta) \\
&= \psi'(\theta).
\end{aligned}$$

Η γενίκευση στη διανυσματική περίπτωση είναι άμεση λαμβάνοντας υπόψη ότι :

$$E_{\theta}(x) = {}^t(E_{\theta}(x_1), \dots, E_{\theta}(x_n)) \quad \text{όταν } x = {}^t(x_1, \dots, x_n) \text{ τ.μ.}$$

και

$$\nabla \psi(\theta) = \left(\frac{\partial}{\partial \theta_1} \psi(\theta), \dots, \frac{\partial}{\partial \theta_n} \psi(\theta) \right) \quad \text{όταν } \theta = {}^t(\theta_1, \dots, \theta_n) \quad \blacktriangle$$

Ως εκ τούτου, εφόσον το $\psi(\theta)$ μπορεί να εξασφαλισθεί σε συγκεκριμένη μορφή, οι δύο απώλειες είναι διαθέσιμες. Ο πίνακας 4.1 δίνει αυτές τις δύο απώλειες για τις συνήθεις εκθετικές οικογένειες. Λόγω της ιδιότητας του αναλλοίωτου που εγκλείουν αυτές οι απώλειες, χρησιμοποιούνται επίσης για τους μετασχηματισμούς των συνήθων εκθετικών οικογενειών όπως είναι οι κατανομές : log-Normale, \mathcal{F} ή Weibbul.. Επιπρόσθετα, κάποια μη-εκθετικά μοντέλα, όπως φαίνεται στα παρακάτω παραδείγματα, επιτρέπουν συγκεκριμένου τύπου εκφράσεις των Hellinger και entropie απωλειών.

Παράδειγμα 2 : Θεωρούμε την οικογένεια των κατανομών Pareto , $\mathcal{Pa}(a, \theta)$ με

$$f(x | \theta) = a \theta^a x^{-a-1} 1_{[\theta, +\infty)}(x), \quad a > 0, \theta > 0 \quad (11)$$

η οποία περιέχει μια ειδική περίπτωση της ομοιόμορφης κατανομής. Συγκεκριμένα, αν $x \sim U([0, \theta])$, $\frac{1}{x} \sim \mathcal{Pa}\left(1, \frac{1}{\theta}\right)$.

Επειδή ο φορέας (support) της $f(x | \theta)$ εξαρτάται από το θ , η L_E δεν μπορεί να χρησιμοποιηθεί. Αντίθετα η Hellinger απώλεια ορίζεται και δίδεται από τη σχέση :

$$L_H(\theta, \delta) = 1 - \left(\frac{\sqrt{\theta \delta}}{\max(\theta, \delta)} \right)^a \quad (12)$$

Πράγματι,

$$\begin{aligned}
L_H(\theta, \delta) &= 1 - \int_{\mathcal{X}} \sqrt{f(x | \theta) f(x | \delta)} dx \\
&= 1 - \int_R \sqrt{a \theta^a x^{-a-1} a \delta^a x^{-a-1}} 1_{[\theta, +\infty)}(x) 1_{[\delta, +\infty)}(x) dx
\end{aligned}$$

$$\begin{aligned}
&= 1 - a(\theta\delta)^{\frac{a}{2}} \int_R x^{-a-1} 1_{[\theta, +\infty) \cap [\delta, +\infty)}(x) dx \\
&= 1 - a(\theta\delta)^{\frac{a}{2}} \int_{\max(\theta, \delta)}^{+\infty} x^{-a-1} dx \\
&= 1 - \frac{(\theta\delta)^{\frac{a}{2}}}{(\max(\theta, \delta))^a} \\
&= 1 - \left(\frac{\sqrt{\delta\theta}}{\max(\theta, \delta)} \right)^a
\end{aligned}$$

▲

Παράδειγμα 3: Παρά το όνομά της, η διπλή εκθετική κατανομή $\mathcal{L}(\theta)$, στερείται της δομής των εκθετικών οικογενειών, αφού

$$f(x|\theta) = \frac{1}{2} \exp\{-|x-\theta|\}, \quad \forall x \in R, \theta \in R \quad (13)$$

Όμως άμεσα αποδεικνύεται ότι :

$$L_E(\theta, \delta) = |\delta - \theta| + \exp\{-|\delta - \theta|\} - 1 \quad (14)$$

και

$$L_H(\theta, \delta) = 1 - \exp\left\{-\frac{|\delta - \theta|}{2}\right\} \left(1 + \frac{1}{2}|\delta - \theta|\right) \quad (15)$$

Πράγματι,

α) για την entropie απώλεια,

$$\begin{aligned}
L_E(\theta, \delta) &= \int_{X=R} \log\left(\frac{f(x|\theta)}{f(x|\delta)}\right) f(x|\theta) dx \\
&= \int_R \log\{\exp(-|x-\theta| + |x-\delta|)\} \frac{1}{2} \exp\{-|x-\theta|\} dx \\
&= \frac{1}{2} \int_R (|x-\delta| - |x-\theta|) \exp\{-|x-\theta|\} dx
\end{aligned} \quad (16)$$

όπου μέσω απλών παραγοντικών ολοκληρώσεων λαμβάνουμε :

$$\begin{aligned}
I &= \int_X |x-\theta| \exp\{-|x-\theta|\} dx = \int_{-\infty}^{\theta} |x-\theta| \exp\{-|x-\theta|\} dx + \int_{\theta}^{+\infty} |x-\theta| \exp\{-|x-\theta|\} dx \\
&= \int_{-\infty}^{\theta} (\theta-x) \exp(x-\theta) dx + \int_{\theta}^{+\infty} (x-\theta) \exp\{-(x-\theta)\} dx \\
&= e^{-\theta} \int_{-\infty}^{\theta} (\theta-x) e^x dx + e^{\theta} \int_{\theta}^{+\infty} (x-\theta) e^{-x} dx
\end{aligned}$$

$$\begin{aligned}
&= e^{-\theta} \left(\theta \int_{-\infty}^{\theta} e^x dx - \int_{-\infty}^{\theta} x e^x dx \right) + e^{\theta} \left(\int_{\theta}^{+\infty} x e^{-x} dx - \theta \int_{\theta}^{+\infty} e^{-x} dx \right) \\
&= e^{-\theta} e^{\theta} + e^{\theta} e^{-\theta} = e^0 + e^0 = 1 + 1 = 2.
\end{aligned}$$

Υποθέτοντας ότι $\delta < \theta$, εργαζόμενοι όπως ανωτέρω έχουμε :

$$\begin{aligned}
II &= \int_R |x - \delta| \exp\{-|x - \theta|\} dx \\
&= \int_{-\infty}^{\delta} |x - \delta| \exp\{-|x - \theta|\} dx + \int_{\delta}^{\theta} |x - \delta| \exp\{-|x - \theta|\} dx + \int_{\theta}^{+\infty} |x - \delta| \exp\{-|x - \theta|\} dx \\
&= \int_{-\infty}^{\delta} (\delta - x) \exp(x - \theta) dx + \int_{\delta}^{\theta} (x - \delta) \exp(x - \theta) dx + \int_{\theta}^{+\infty} (x - \delta) \exp\{-(x - \theta)\} dx \\
&= e^{-\theta} \int_{-\infty}^{\delta} (\delta - x) e^x dx + e^{-\theta} \int_{\delta}^{\theta} (x - \delta) e^x dx + e^{\theta} \int_{\theta}^{+\infty} (x - \delta) e^{-x} dx \\
&= e^{-\theta} \left(\delta \int_{-\infty}^{\delta} e^x dx - \int_{-\infty}^{\delta} x e^x dx \right) + e^{-\theta} \left(\int_{\delta}^{\theta} x e^x dx - \delta \int_{\delta}^{\theta} e^x dx \right) + e^{\theta} \left(\int_{\theta}^{+\infty} x e^{-x} dx - \delta \int_{\theta}^{+\infty} e^{-x} dx \right) \\
&= e^{\delta - \theta} + (\theta - 1 + e^{\delta - \theta} - \delta) + (\theta + 1 - \delta) \\
&= 2(\theta - \delta) + 2e^{\delta - \theta}.
\end{aligned}$$

Λοιπόν από τις τιμές των I και II για $\delta < \theta$, η σχέση (16) δίνει :

$$L_E(\theta, \delta) = \frac{1}{2} \{2(\theta - \delta) + 2e^{\delta - \theta} - 2\} = (\theta - \delta) + e^{\delta - \theta} - 1 = |\delta - \theta| + \exp\{-|\delta - \theta|\} - 1.$$

Ομοίως, για $\delta > \theta$, αρκεί να παρατηρήσουμε ότι :

$$\begin{aligned}
II &= \int_{-\infty}^{\theta} |x - \delta| \exp\{-|x - \theta|\} dx + \int_{\theta}^{\delta} |x - \delta| \exp\{-|x - \theta|\} dx + \int_{\delta}^{+\infty} |x - \delta| \exp\{-|x - \theta|\} dx \\
&= 2(\delta - \theta) + 2e^{\theta - \delta}
\end{aligned}$$

β) Για την Hellinger απώλεια,

$$L_H(\theta, \delta) = 1 - \int_{X=R} \sqrt{f(x|\theta)f(x|\delta)} dx = 1 - \frac{1}{2} \int_R \exp\left\{-\frac{1}{2}(|x - \theta| + |x - \delta|)\right\} dx$$

Λοιπόν, για $\delta < \theta$, έχουμε :

$$\begin{aligned}
L_H(\theta, \delta) &= 1 - \frac{1}{2} \int_{-\infty}^{\delta} \exp\left\{-\frac{1}{2}(|x - \theta| + |x - \delta|)\right\} dx - \frac{1}{2} \int_{\delta}^{\theta} \exp\left\{-\frac{1}{2}(|x - \theta| + |x - \delta|)\right\} dx \\
&\quad - \frac{1}{2} \int_{\theta}^{+\infty} \exp\left\{-\frac{1}{2}(|x - \theta| + |x - \delta|)\right\} dx
\end{aligned}$$

$$\begin{aligned}
&= 1 - \frac{1}{2} \int_{-\infty}^{\delta} \exp\left\{-\frac{1}{2}(\theta + \delta - 2x)\right\} dx - \frac{1}{2} \int_{\delta}^{\theta} \exp\left\{-\frac{1}{2}(\theta - \delta)\right\} dx \\
&\quad - \frac{1}{2} \int_{\theta}^{+\infty} \exp\left\{-\frac{1}{2}(2x - \theta - \delta)\right\} dx \\
&= 1 - \frac{1}{2} \exp\left(-\frac{\theta + \delta}{2}\right) \int_{-\infty}^{\delta} e^x dx - \frac{1}{2} \exp\left\{-\frac{1}{2}(\theta - \delta)\right\} (\theta - \delta) \\
&\quad - \frac{1}{2} \exp\left(\frac{\theta + \delta}{2}\right) \int_{\theta}^{+\infty} e^{-x} dx \\
&= 1 - \frac{1}{2} \exp\left(-\frac{\theta - \delta}{2}\right) - \frac{1}{2} (\theta - \delta) \exp\left(-\frac{\theta - \delta}{2}\right) - \frac{1}{2} \exp\left(-\frac{\theta - \delta}{2}\right) \\
&= 1 - \exp\left(-\frac{\theta - \delta}{2}\right) \left(1 + \frac{\theta - \delta}{2}\right) \\
&= 1 - \exp\left(-\frac{|\delta - \theta|}{2}\right) \left(1 + \frac{|\delta - \theta|}{2}\right)
\end{aligned}$$

Εργαζόμενοι ομοίως για $\delta > \theta$, καταλήγουμε στη σχέση (15)



Πίνακας 4.1

Entropie και Hellinger απώλειες για τις φυσικές παραμετροποιήσεις των συνηθών εκθετικών οικογενειών

Κατανομή $f(x \theta)$	Φυσική παράμετρος θ	$\psi(\theta)$	Entropie απώλεια $L_E(\theta, \delta)$	Hellinger Απώλεια $L_H(\theta, \delta)$
Κανονική $N(\mu, 1)$	μ	$\frac{\theta^2}{2}$	$\frac{1}{2}(\theta - \delta)^2$	$1 - \exp\left\{-\frac{(\theta - \delta)^2}{8}\right\}$
Κανονική $N(0, \sigma^2)$	$\frac{1}{\sigma^2}$	$-\frac{\log \theta}{2}$	$\frac{1}{2}\left(\frac{\delta}{\theta} - \log \frac{\delta}{\theta} - 1\right)$	$1 - \left(\frac{2\sqrt{\delta\theta}}{\theta + \delta}\right)^{\frac{1}{2}}$
Κανονική $N(\mu, \sigma^2)$	$\left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)$	$\frac{1}{2}\left(\frac{\theta_1^2}{\theta_2} - \log \theta_2\right)$	$\frac{1}{2}\left\{\delta_2\left(\frac{\theta_1}{\theta_2} - \frac{\delta_1}{\delta_2}\right)^2 + \frac{\delta_2}{\theta_2} - \log \frac{\delta_2}{\theta_2} - 1\right\}$	$1 - \exp \frac{G(\theta)}{2}$ *
Γάμμα $\Gamma\left(a, \frac{1}{\beta}\right)$	$\frac{1}{\beta}$	$-a \log \theta$	$a\left(\frac{\delta}{\theta} - \log \frac{\delta}{\theta} - 1\right)$	$1 - \left(\frac{2\sqrt{\delta\theta}}{\theta + \delta}\right)^a$
Poisson $P(\lambda)$	$\log \lambda$	$\exp \theta$	$e^\theta(\theta - \delta) + e^\delta - e^\theta$	$1 - \exp\left\{-\frac{1}{2}\left(e^{\frac{\theta}{2}} - e^{\frac{\delta}{2}}\right)^2\right\}$
Διωνυμική $Bin(n, p)$	$\log \frac{p}{1-p}$	$n \log(1 + e^\theta)$	$n\left\{\frac{e^\theta(\theta - \delta)}{1 + e^\theta} + \log \frac{1 + e^\delta}{1 + e^\theta}\right\}$	$1 - \left(\frac{1 + e^{\frac{\theta + \delta}{2}}}{\sqrt{(1 + e^\theta)(1 + e^\delta)}}\right)^n$
Αρνητική Διωνυμική $NegBin(r, p)$	$\log(1 - p)$	$-r(\log(1 - e^\theta))$	$r\left\{\frac{e^\theta(\theta - \delta)}{1 + e^\theta} + \log \frac{1 + e^\theta}{1 + e^\delta}\right\}$	$1 - \left(\frac{\sqrt{(1 - e^\theta)(1 - e^\delta)}}{1 - e^{\frac{\theta + \delta}{2}}}\right)^r$

* όπου στην περίπτωση της Κανονικής $N(\mu, \sigma^2)$, σημειώνουμε :

$\theta = {}^t(\theta_1, \theta_2)$, $\theta_1 = \frac{\mu}{\sigma^2}$, $\theta_2 = \frac{1}{\sigma^2}$, $\delta_1 = \frac{\hat{\mu}}{\hat{\sigma}^2}$, $\delta_2 = \frac{1}{\hat{\sigma}^2}$ με $\hat{\mu}, \hat{\sigma}^2$ τους εκτιμητές των μ, σ^2 αντιστοίχως και

$$G(\theta) = \frac{(\theta_1 + \delta_1)^2}{2(\theta_2 + \delta_2)} - \frac{\theta_1^2}{2\theta_2} - \frac{\delta_1^2}{2\delta_2} + \log \frac{2\sqrt{\delta_2\theta_2}}{\delta_2 + \theta_2}.$$

Πίνακας 4.2

Hellinger και Entropie απώλειες για τις αρχικές παραμέτρους ορισμένων εκθετικών οικογενειών

(οι εκτιμητές των σ, μ, λ και p σημειώνονται με $\hat{\sigma}, \hat{\mu}, \hat{\lambda}$ και \hat{p} αντιστοίχως)

Κατανομή	$1 - L_H(\theta, \hat{\theta})$	$L_E(\theta, \hat{\theta})$
$N(0, \sigma^2)$	$\sqrt{\frac{2\hat{\sigma}\sigma}{\sigma^2 + \hat{\sigma}^2}}$	$\frac{1}{2} \left[\frac{\sigma^2}{\hat{\sigma}^2} - \log \frac{\sigma^2}{\hat{\sigma}^2} - 1 \right]$
$N(\mu, \sigma^2)$	$\sqrt{\frac{2\hat{\sigma}\sigma}{\sigma^2 + \hat{\sigma}^2}} e^{-\frac{(\mu - \hat{\mu})^2}{4(\sigma^2 + \hat{\sigma}^2)}}$	$\frac{1}{2} \left[\frac{(\mu - \hat{\mu})^2}{\hat{\sigma}^2} + \frac{\sigma^2}{\hat{\sigma}^2} - \log \frac{\sigma^2}{\hat{\sigma}^2} - 1 \right]$
$P(\lambda)$	$e^{-\frac{(\sqrt{\lambda} - \sqrt{\hat{\lambda}})^2}{2}}$	$\lambda \left[\frac{\hat{\lambda}}{\lambda} + \log \frac{\lambda}{\hat{\lambda}} - 1 \right]$
$Bin(n, p)$	$(\sqrt{(1-p)(1-\hat{p})} + \sqrt{p\hat{p}})^n$	$n \left\{ p \log \frac{p}{\hat{p}} + (1+p) \log \frac{1-\hat{p}}{1-p} \right\}$
$NegBin(r, p)$	$\left(\frac{\sqrt{p\hat{p}}}{1 - \sqrt{(1-p)(1-\hat{p})}} \right)^r$	$r \left\{ \frac{1-p}{p} \log \frac{1-p}{1-\hat{p}} + \log \frac{p}{\hat{p}} \right\}$

Σημειώνουμε ότι ο Πίνακας 4.1 δεν περιλαμβάνει την $Beta(a, \beta)$ - κατανομή επειδή σ' αυτή την περίπτωση και οι δύο απώλειες είναι αρκετά πολύπλοκες. Αν ξαναγράψουμε τις Hellinger και entropie απώλειες του Πίνακα 4.1 για τις αρχικές παραμετροποιήσεις λαμβάνουμε τις μετασχηματισμένες απώλειες του Πίνακα 4.2. Παρατηρούμε ότι και στους δύο πίνακες οι δύο απώλειες είναι εντελώς διαφορετικές. Ο Πίνακας 4.1 περιλαμβάνει μεταξύ των άλλων την τόσο συχνά καλούμενη «entropie απώλεια»

$$L_E(\theta, \delta) = \frac{\delta}{\theta} - \log\left(\frac{\delta}{\theta}\right) - 1$$

η οποία φαίνεται στις κατανομές $N(0, \sigma^2)$ και $\Gamma(a, \beta)$.

Τέλος να σημειώσουμε το εκπληκτικό χαρακτηριστικό της entropie απώλειας στην $N(\mu, \sigma^2)$ - περίπτωση. Αυτή μπορεί να ξαναγραφεί ως εξής :

$$L_E(t(\mu, \sigma^2), t(\hat{\mu}, \hat{\sigma}^2)) = \frac{1}{2} \left[\frac{(\mu - \hat{\mu})^2}{\hat{\sigma}^2} + \frac{\sigma^2}{\hat{\sigma}^2} - \log\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) - 1 \right]$$

και ως εκ τούτου η τετραγωνική απώλεια βαρύνεται από το $\frac{1}{\hat{\sigma}^2}$ αντί του $\frac{1}{\sigma^2}$.

5. Εκτιμητές του Bayes υπό σύμφυτες απώλειες

Από bayesiennne άποψη, κάποιες απώλειες είναι περισσότερο ενδιαφέρουσες από άλλες γιατί οδηγούν σε αναλυτικές εκφράσεις των εκτιμητών του Bayes (τουλάχιστον για ορισμένες εκ των προτέρων οικογένειες). Σε αυτήν την παράγραφο θα δούμε ότι μια τέτοια περίπτωση είναι εκείνη της entropie απώλειας εν αντιθέσει με την Hellinger απώλεια.

Πράγματι, οι εκτιμητές του Bayes υπό την entropie απώλεια, για συζυγείς εκ των προτέρων, δίδονται από το παρακάτω αποτέλεσμα.

Πρόταση 5.1

Αν $x \sim f(x | \theta) = e^{\theta \cdot x - \psi(\theta)}$ και $\theta \sim \pi(\theta) \propto e^{\theta \cdot x_0 - \lambda \psi(\theta)}$ ο εκτιμητής του Bayes, δ^π , για το θ , υπό την L_E δίδεται από τη σχέση :

$$\nabla \psi(\delta^\pi(x)) = \frac{x + x_0}{\lambda + 1}$$

Ιδιαίτερος ο εκτιμητής του Bayes για το $\nabla \psi(\theta)$ είναι $\frac{x + x_0}{\lambda + 1}$ και συμπίπτει με τον εκτιμητή του Bayes για το $\nabla \psi(\theta)$ υπό την τετραγωνική απώλεια.

Απόδειξη: Από τον ορισμό του εκτιμητή του Bayes (βλέπε §2) δεδομένων της εκ των προτέρων συζυγούς $\pi(\theta)$ και της απώλειας L_E , ο $\delta^\pi(x)$ αποτελεί λύση της

$$\min_{\delta} E^{\pi} [L_E(\theta, \delta) | x] = \min_{\delta} \int_{\Theta} L_E(\theta, \delta) \pi(\theta | x) d\theta$$

όπου από την Πρόταση 4.1,

$$L_E(\theta, \delta) = (\theta - \delta) \cdot \nabla \psi(\theta) + \psi(\delta) - \psi(\theta),$$

ενώ

$$\pi(\theta | x) \propto f(x | \theta) \pi(\theta) \propto e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)}.$$

Λαμβάνουμε λοιπόν τον εκτιμητή του Bayes, συνδέοντας με το x το στοιχείο $\delta^{\pi}(x)$ του $D(\subseteq R^n, n \geq 1)$ στο οποίο η συνάρτηση

$$\Phi_x(\delta) = \int_{\Theta} \{(\theta - \delta) \nabla \psi(\theta) + \psi(\delta) - \psi(\theta)\} e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta$$

λαμβάνει την ελάχιστη τιμή της (υπό την προϋπόθεση ότι ένα τέτοιο σημείο υπάρχει για κάθε x).

Επειδή

$$\begin{aligned} \Phi_x(\delta) &= \int_{\Theta} \theta \cdot \nabla \psi(\theta) e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta - \delta \cdot \int_{\Theta} \nabla \psi(\theta) e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta \\ &\quad + \psi(\delta) \int_{\Theta} e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta - \int_{\Theta} \psi(\theta) e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta, \end{aligned}$$

παραγωγίζοντας ως προς δ , έχουμε:

$$\nabla \Phi_x(\delta) = - \int_{\Theta} \nabla \psi(\theta) e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta + \nabla \psi(\delta) \int_{\Theta} e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta.$$

Συνεπώς,

$$\begin{aligned} \nabla \Phi_x(\delta) = 0 &\Rightarrow \nabla \psi(\delta) = \frac{\int_{\Theta} \nabla \psi(\theta) e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta}{\int_{\Theta} e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta} \\ &= \int_{\Theta} \nabla \psi(\theta) \frac{e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)}}{\int_{\Theta} e^{\theta \cdot (x+x_0) - (\lambda+1)\psi(\theta)} d\theta} \\ &= \int_{\Theta} \nabla \psi(\theta) \pi(\theta | x) d\theta \\ & (= E^{\pi}(\nabla \psi(\theta) | x) \text{ εκ των υστέρων μέσος του } \nabla \psi(\theta)) \end{aligned}$$

και ο $\delta^{\pi}(x)$ θα ικανοποιεί την τελευταία εξίσωση μόνο όταν ο πίνακας των δευτέρων μερικών παραγώγων ως προς δ του $\psi(\delta)$ είναι θετικά ορισμένος.

Υπολογισμός της $E^{\pi}(\nabla \psi(\theta) | x)$.

Περιοριζόμενοι στη μονοδιάστατη περίπτωση, υπό μια απλή παραγοντική ολοκλήρωση, έχουμε :

$$\begin{aligned}
 E^\pi(\nabla \psi(\theta) | x) &= \int_{\Theta} \psi'(\theta) \pi(\theta | x) d\theta = \int_R \psi'(\theta) c e^{\theta(x+x_0) - (\lambda+1)\psi(\theta)} d\theta \\
 &= c \int_R e^{\theta(x+x_0)} \psi'(\theta) e^{-(\lambda+1)\psi(\theta)} d\theta \\
 &= - \underbrace{\frac{c}{\lambda+1} e^{\theta(x+x_0)} e^{-(\lambda+1)\psi(\theta)}}_0 \Big|_{-\infty}^{+\infty} + \frac{c}{\lambda+1} \int_{-\infty}^{+\infty} (x+x_0) e^{\theta(x+x_0) - (\lambda+1)\psi(\theta)} d\theta \\
 &= \frac{x+x_0}{\lambda+1} \underbrace{\int_R c e^{\theta(x+x_0) - (\lambda+1)\psi(\theta)} d\theta}_1.
 \end{aligned}$$

Για την πολυδιάστατη περίπτωση αρκεί να παρατηρήσουμε :

$$\nabla \psi(\theta) = \left(\frac{\partial}{\partial \theta_1} \psi(\theta), \dots, \frac{\partial}{\partial \theta_n} \psi(\theta) \right) \text{ και } x = {}^t(x_1, \dots, x_n)$$

Λοιπόν,

$$E^\pi(\nabla \psi(\theta) | x) = \frac{x+x_0}{\lambda+1}.$$

Παρατηρούμε επίσης ότι, ο εκ των υστέρων μέσος $E^\pi(\nabla \psi(\theta) | x)$ αποτελεί τον εκτιμητή του Bayes για το $\nabla \psi(\theta)$ υπό την τετραγωνική απώλεια.

Πράγματι, υπό την τετραγωνική απώλεια,

$$\begin{aligned}
 \Phi_x(\delta) &= \int_{\Theta} \|\nabla \psi(\theta) - \delta\|^2 \pi(\theta | x) d\theta \\
 &= \int_{\Theta} \|\nabla \psi(\theta)\|^2 \pi(\theta | x) d\theta + \underbrace{\|\delta\|^2 \int_{\Theta} \pi(\theta | x) d\theta}_1 - 2\delta \bullet \int_{\Theta} \nabla \psi(\theta) \pi(\theta | x) d\theta
 \end{aligned}$$

και παραγωγίζοντας ως προς δ , έχουμε:

$$\nabla \Phi_x(\delta) = 2\delta - 2 \int_{\Theta} \nabla \psi(\theta) \pi(\theta | x) d\theta.$$

Συνεπώς,

$$\Phi_x(\delta) = 0 \Rightarrow \hat{\delta}(x) = \int_{\Theta} \nabla \psi(\theta) \pi(\theta | x) d\theta.$$

Επειδή ο πίνακας των δευτέρων μερικών παραγώγων του $\Phi_x(\delta)$ ως προς δ είναι $2I_n$ (θετικά ορισμένος), ο $\hat{\delta}$ είναι ο εκτιμητής του Bayes του $\nabla \psi(\theta)$ υπό την τετραγωνική απώλεια. ▲

Αυτό το σημειακό αποτέλεσμα υποδηλώνει ότι η τετραγωνική απώλεια είναι πράγματι επαρκής ως μια μη-πληροφορική συνάρτηση απώλειας μόνο για τη μέση παραμετρικοποίηση μιας εκθετικής οικογένειας ($E_\theta(x) = \nabla \psi(\theta)$). Επιπλέον, εξουσιοδοτεί την ανάκτηση του εκτιμητή μεγίστης πιθανοφάνειας για οποιαδήποτε 1-1 μετασχηματισμό του θ σαν ένα εκτιμητή του Bayes υπό μια εκ των προτέρων σταθερά ($\lambda = 0, x_0 = 0$) επειδή ο εκτιμητής μεγίστης πιθανοφάνειας του θ , $\hat{\theta}$, είναι λύση της $\nabla \psi(\theta) = x$. Είναι ενδιαφέρον ότι το γεγονός αυτό υπαινίσσεται την αποκατάσταση της εκ των προτέρων σταθεράς του Laplace ως εκ των προτέρων μη-πληροφορικής όταν χρησιμοποιείται η entropie απώλεια.

Σημειώνουμε την περιπλοκή των συμπερασμάτων στις εκ των προτέρων του Jeffrey στη διαδικασία προσέγγισης που δίδεται από το Λήμμα 3.1. Πράγματι, θεωρώντας τη μονοδιάστατη περίπτωση, επειδή η entropie απώλεια συμπεριφέρεται περίπου όπως το $(\theta - \delta)^2 I(\theta)$, χρησιμοποιώντας μια εκ των προτέρων σταθερά επί του θ υπό αυτή την απώλεια είναι ισοδύναμο με το να χρησιμοποιήσουμε την εκ των προτέρων $I(\theta)$ υπό την τετραγωνική απώλεια αντί του $(I(\theta))^{1/2}$ που αποτελεί την εκ των προτέρων του Jeffrey. Υπάρχει ταυτότητα μεταξύ των δύο προσεγγίσεων όταν $I(\theta) = 1$ δηλαδή για μια «ομοιόμορφη παραμετρικοποίηση». Αλλά αυτό συμβαίνει μόνο για τη μέση παραμετρικοποίηση δηλαδή όταν η παράμετρος είναι η $E_\theta(x) (= \nabla \psi(\theta))$, επειδή τότε $I(\nabla \psi(\theta)) = 1$ και το θ είναι η φυσική παράμετρος. Παρά του ότι δίδεται έμφαση στην έλλειψη αλληλεπίδρασης μεταξύ απώλειας και εκ των προτέρων μη-πληροφορικής του Jeffrey, η φυσική παραμετρικοποίηση των εκθετικών οικογενειών είναι η προτιμότερη παραμετρικοποίηση για αμφότερες τις εκ των προτέρων μη-πληροφορικές των Laplace και Jeffrey.

Όπως αναφέρθηκε και προηγουμένως, οι δύο σύμφυτες απώλειες οδηγούν σε equivariant εκτιμητές του Bayes όταν η εκ των προτέρων κατανομή δεν εξαρτάται από την παραμετρικοποίηση :

Λήμμα 5.2

Για κάθε 1-1 μετασχηματισμό $h(\theta)$ και υπό τις σύμφυτες απώλειες L_E και L_H , ο εκτιμητής του Bayes του $h(\theta)$ είναι ο $h(\delta^\pi)$ όπου δ^π ο εκτιμητής του Bayes για το θ .

Απόδειξη: Εξ' ορισμού, ο εκτιμητής του Bayes για το $h(\theta)$ είναι η λύση της

$$\begin{aligned} \min_d \int_{\Theta} L(h(\theta), d) \pi(\theta | x) d\theta &= \min_{t: h(t)=d} \int_{\Theta} L(h(\theta), h(t)) \pi(\theta | x) d\theta \\ &= \min_{t: h(t)=d} \int_{\Theta} L(\theta, t) \pi(\theta | x) d\theta \\ &= \min_{t: h(t)=d} E^\pi(L(\theta, t) | x) \end{aligned}$$

όπου η απώλεια L είναι είτε η L_E είτε η L_H .

Λοιπόν, αν ο $\delta^\pi(x)$ ελαχιστοποιεί (στο t) τον εκ των υστέρων κίνδυνο $E^\pi(L(\theta, t) | x)$, ο $h(\delta^\pi(x))$ ελαχιστοποιεί (στο d) τον εκ των υστέρων κίνδυνο $E^\pi(L(h(\theta), d) | x)$. ▲

Συνεπώς, κάθε μη-πληροφορική μέθοδος που βασίζει την επιλογή μιας εκ των προτέρων κατανομής επί της δειγματικής κατανομής, θα οδηγεί σε αναλλοίωτους υπό την παραμετρικοποίηση εκτιμητές, αν χρησιμοποιούνται οι σύμφυτες απώλειες.

Επιστρέφοντας στην Hellinger απώλεια, θα δούμε μέσω παραδειγμάτων ότι δεν μπορούμε να έχουμε σ' αυτήν την περίπτωση ισοδύναμη της Πρότασης 5.1.

Παράδειγμα 4 : Αν $x \sim N(\theta, 1)$, από τον πίνακα 4.1, η entropie απώλεια είναι :

$$L_E(\theta, d) = \frac{1}{2}(\theta - d)^2 \quad (= \text{τετραγωνική απώλεια})$$

όταν η φυσική παράμετρος είναι το θ και $\psi(\theta) = \frac{\theta^2}{2}$.

Θεωρώντας ως εκ των προτέρων π επί του θ την $N(\mu, \tau^2)$, από την Πρόταση 5.1 για $x_0 = \frac{\mu}{\tau^2}$ και $\lambda = \frac{1}{\tau^2}$, έχουμε ότι ο εκτιμητής του Bayes δ^π , του θ υπό την L_E δίδεται από τη σχέση :

$$\delta^\pi(x) = \frac{x + \frac{\mu}{\tau^2}}{1 + \frac{1}{\tau^2}} = \frac{\mu + x\tau^2}{1 + \tau^2}$$

Επειδή $\delta^\pi(x) = E(\theta | x)$, αυτός αποτελεί και τον εκτιμητή του Bayes για το θ υπό την τετραγωνική απώλεια.

Αναζητούμε στη συνέχεια τον εκτιμητή του Bayes $\tilde{\delta}^\pi$, του θ υπό την Hellinger απώλεια η οποία, από τον Πίνακα 4.1, είναι :

$$L_H(\theta, d) = 1 - \exp\left\{-\frac{(\theta - d)^2}{8}\right\}$$

Ως γνωστό, για $x \in X = R$, ο $\tilde{\delta}^\pi$ συμπεραίνεται από την

$$\begin{aligned} \min_d E^\pi(L_H(\theta, d) | x) &= \min_d \int_{\Theta=R} L_H(\theta, d) \pi(\theta | x) d\theta \\ &= \min_d \int_{\Theta} \left(1 - \exp\left\{-\frac{(\theta - d)^2}{8}\right\}\right) \pi(\theta | x) d\theta \end{aligned}$$

ισοδύναμα, από την

$$\begin{aligned}
\max_d \int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2\right\} \pi(\theta|x) d\theta &\Leftrightarrow \max_d \int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2\right\} f(x|\theta) \pi(\theta) d\theta \\
&\Leftrightarrow \max_d \int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2\right\} \exp\left\{-\frac{1}{2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
&= \max_d \int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2 - \frac{1}{2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta.
\end{aligned}$$

Όμως,

$$\begin{aligned}
-\frac{1}{8}(\theta-d)^2 - \frac{1}{2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2 &= -\frac{d^2}{8} - \frac{\theta^2}{8} \left(5 + \frac{4}{\tau^2}\right) + \theta \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right) - \frac{x^2}{2} - \frac{\mu^2}{2\tau^2} \\
&= -\frac{d^2}{8} - \frac{x^2}{2} - \frac{\mu^2}{2\tau^2} - \frac{1}{8} \left(5 + \frac{4}{\tau^2}\right) \left[\theta^2 - 8\theta \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right) \left(5 + \frac{4}{\tau^2}\right)^{-1} \right] \\
&= -\frac{d^2}{8} - \frac{x^2}{2} - \frac{\mu^2}{2\tau^2} - \frac{1}{8} \left(5 + \frac{4}{\tau^2}\right) \left[\theta - 4 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right) \left(5 + \frac{4}{\tau^2}\right)^{-1} \right]^2 + 2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \left(5 + \frac{4}{\tau^2}\right)^{-1}
\end{aligned}$$

Λοιπόν,

$$\begin{aligned}
&\int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2 - \frac{1}{2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
&= \exp\left\{-\frac{d^2}{8} - \frac{x^2}{2} - \frac{\mu^2}{2\tau^2}\right\} \exp\left\{2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \left(5 + \frac{4}{\tau^2}\right)^{-1}\right\} \\
&\quad \times \underbrace{\int_R \exp\left\{-\frac{1}{8} \left(5 + \frac{4}{\tau^2}\right) \left[\theta - 4 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right) \left(5 + \frac{4}{\tau^2}\right)^{-1} \right]^2\right\} d\theta}_{\text{ολοκλ. } N\left(4 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right) \left(5 + \frac{4}{\tau^2}\right)^{-1}, 4 \left(5 + \frac{4}{\tau^2}\right)^{-1}\right)} \\
&= (2\pi)^{\frac{1}{2}} 2 \left(5 + \frac{\mu}{\tau^2}\right)^{-\frac{1}{2}} \exp\left\{-\frac{d^2}{8} + 2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \left(5 + \frac{4}{\tau^2}\right)^{-1}\right\} \exp\left\{-\frac{x^2}{2} - \frac{\mu^2}{2\tau^2}\right\}
\end{aligned}$$

και

$$\begin{aligned}
&\max_d \int_{\Theta} \exp\left\{-\frac{1}{8}(\theta-d)^2 - \frac{1}{2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} d\theta \\
&\Leftrightarrow \max_d \exp\left\{-\frac{d^2}{8} + 2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \left(5 + \frac{4}{\tau^2}\right)^{-1}\right\}
\end{aligned}$$

Επειδή

$$-\frac{d^2}{8} + 2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \left(5 + \frac{4}{\tau^2}\right)^{-1} = -\frac{d^2}{8} + 2 \left(\frac{d}{4} + x + \frac{\mu}{\tau^2}\right)^2 \frac{\tau^2}{4 + 5\tau^2}$$

$$\begin{aligned}
&= -\frac{d^2}{8} + 2\frac{\tau^2}{4+5\tau^2} \left(\frac{d^2}{16} + \left(x + \frac{\mu}{\tau^2} \right)^2 + 2\frac{d}{4} \left(x + \frac{\mu}{\tau^2} \right) \right) \\
&= -\frac{d^2}{8} + \frac{\tau^2 d^2}{8(4+5\tau^2)} + \frac{2(\mu+x\tau^2)^2}{\tau^2(4+5\tau^2)} + \frac{\mu+x\tau^2}{4+5\tau^2} d \\
&= -\frac{1+\tau^2}{2(4+5\tau^2)} d^2 + \frac{\mu+x\tau^2}{4+5\tau^2} d + \frac{2(\mu+x\tau^2)^2}{\tau^2(4+5\tau^2)} \\
&= -\frac{1+\tau^2}{2(4+5\tau^2)} \left(d^2 - 2\frac{\mu+x\tau^2}{1+\tau^2} d \right) + \frac{2(\mu+x\tau^2)^2}{\tau^2(4+5\tau^2)} \\
&= -\frac{1+\tau^2}{2(4+5\tau^2)} \left(d - 2\frac{\mu+x\tau^2}{1+\tau^2} \right)^2 + \frac{(\mu+x\tau^2)^2}{2(1+\tau^2)(4+5\tau^2)} + \frac{2(\mu+x\tau^2)^2}{\tau^2(4+5\tau^2)} \\
&= -\frac{1+\tau^2}{2(4+5\tau^2)} \left(d - \frac{\mu+x\tau^2}{1+\tau^2} \right)^2 - \frac{(\mu+x\tau^2)^2}{2\tau^2(1+\tau^2)}, \\
\max_d \exp \left\{ -\frac{d^2}{8} + 2\left(\frac{d}{4} + x + \frac{\mu}{\tau^2} \right)^2 \left(5 + \frac{4}{\tau^2} \right)^{-1} \right\} &\Leftrightarrow \max_d \exp \left\{ -\frac{1+\tau^2}{2(4+5\tau^2)} \left(d - \frac{\mu+x\tau^2}{1+\tau^2} \right)^2 \right\} \\
&\Leftrightarrow \max_d \exp \left\{ \frac{1+\tau^2}{2(4+5\tau^2)} \left(d - \frac{\mu+x\tau^2}{1+\tau^2} \right)^2 \right\},
\end{aligned}$$

το οποίο επιτυγχάνεται όταν $d = \frac{\mu+x\tau^2}{1+\tau^2}$.

Συνεπώς,

$$\tilde{\delta}^\pi(x) = \frac{\mu+x\tau^2}{1+\tau^2} = \delta^\pi(x) \quad \blacktriangle$$

Δυστυχώς αυτή η ισότητα των εκτιμητών μεταξύ L_E και L_H σπανίως συμβαίνει και η παραγωγή ενός εκτιμητή του Bayes γίνεται όπως στα παρακάτω παραδείγματα.

Παράδειγμα 5: Θεωρούμε τη Γάμμα περίπτωση : $x \sim \Gamma\left(a, \frac{1}{\theta}\right)$, $\theta \sim \Gamma\left(\nu, \frac{1}{x_0}\right)$. Η εκ των

υστέρων κατανομή του $\theta | x$ είναι τότε η $\Gamma\left(a+\nu, \frac{1}{x+x_0}\right)$ (βλέπε II, §4) και η Hellinger

απώλεια, από τον Πίνακα 4.1, είναι :

$$L_H(\theta, d) = 1 - \left(\frac{2\sqrt{\theta d}}{\theta + d} \right)^a$$

Ως γνωστό, για $x \in X = R_+^*$, ο εκτιμητής του Bayes $\delta^\pi(x)$ του θ λαμβάνεται από τη λύση του

$$\min_d E^\pi(L_H(\theta, d) | x) = \min_{\Theta=R_+^*} \int L_H(\theta, d) \pi(\theta | x) d\theta.$$

Όμως,

$$\begin{aligned} \int_{\Theta} L_H(\theta, d) \pi(\theta | x) d\theta &= \int_0^{+\infty} \left(1 - \frac{2^a \theta^{\frac{a}{2}} d^{\frac{a}{2}}}{(\theta + d)^a} \right) \pi(\theta | x) d\theta \\ &= 1 - 2^a d^{\frac{a}{2}} \int_0^{+\infty} \frac{\theta^{\frac{a}{2}}}{(\theta + d)^a} \pi(\theta | x) d\theta \\ &= 1 - 2^a d^{\frac{a}{2}} \frac{(x + x_0)^{a+v}}{\Gamma(a+v)} \int_0^{+\infty} \frac{\theta^{\frac{3a}{2}+v-1}}{(\theta + d)^a} e^{-(x+x_0)\theta} d\theta \end{aligned}$$

και θέτοντας στο τελευταίο ολοκλήρωμα, $u = \frac{\theta}{d} (\Rightarrow \theta = ud, d\theta = d \cdot du)$ έχουμε:

$$\int_{\Theta} L_H(\theta, d) \pi(\theta | x) d\theta = 1 - 2^a \frac{(x + x_0)^{a+v}}{\Gamma(a+v)} d^{v+a} \int_0^{+\infty} u^{\frac{3a}{2}+v-1} (u+1)^{-a} e^{-(x+x_0)du} du.$$

Λοιπόν, ο $\delta^\pi(x)$ θα λαμβάνεται ισοδύναμα από την

$$\max_d d^{v+a} \int_0^{+\infty} u^{\frac{3a}{2}+v-1} (u+1)^{-a} e^{-(x+x_0)du} du$$

ισοδύναμα, θέτοντας $K = (x + x_0)d$, από την

$$\max_K K^{v+a} \int_0^{+\infty} u^{\frac{3a}{2}+v-1} (u+1)^{-a} e^{-Ku} du$$

δηλαδή ο $\delta^\pi(x)$ θα έχει τη μορφή :

$$\delta^\pi(x) = \frac{K_{a,v}}{x + x_0} \quad (17)$$

όπου το $K_{a,v}$ θα αποτελεί λύση της

$$\max_K K^{a+v} \int_0^{+\infty} u^{\frac{3a}{2}+v-1} (u+1)^{-a} e^{-Ku} du \quad (18)$$

Από τη φόρμουλα (13.2.5) των Abramowitz και Stegun (1964),

$$\int_0^{+\infty} u^{\frac{3a}{2}+\nu-1} (u+1)^{-a} e^{-Ku} du = \Gamma\left(\frac{3a}{2}+\nu\right) {}_1F_1\left(\frac{3a}{2}+\nu, \frac{a}{2}+\nu+1, K\right)$$

όπου ${}_1F_1(a, b, z)$ είναι η confluent hyper geometric function. Λοιπόν, η σταθερά $K_{a,\nu}$ θα είναι λύση της

$$\max_K K^{a+\nu} {}_1F_1\left(\frac{3a}{2}+\nu, \frac{a}{2}+\nu+1, K\right)$$

ή (κατά τον Olver, 1974) της

$$\max_K K^{\frac{a}{2}} {}_1F_1\left(a, 1-\frac{a}{2}-\nu, K\right)$$

Θέτοντας

$$\Phi(K) = K^{\frac{a}{2}} {}_1F_1\left(a, 1-\frac{a}{2}-\nu, K\right),$$

από τη φόρμουλα (13.4.22) των Abramowitz και Stegun (1964), έχουμε:

$$\begin{aligned} \Phi'(K) &= \frac{a}{2} K^{\frac{a}{2}-1} {}_1F_1\left(a, 1-\frac{a}{2}-\nu, K\right) - K^{\frac{a}{2}} (a) {}_1F_1\left(a+1, 2-\frac{a}{2}-\nu, K\right) \\ &= \frac{a}{2} K^{\frac{a}{2}-1} {}_1F_1\left(a, 1-\frac{a}{2}-\nu, K\right) - a K^{\frac{a}{2}} (a) {}_1F_1\left(a+1, 2-\frac{a}{2}-\nu, K\right) \end{aligned}$$

και ο $\Phi'(K) = 0$ οδηγεί στην συμπερασματική εξίσωση υπολογισμού του K :

$$2K = \frac{{}_1F_1\left(a, 1-\frac{a}{2}-\nu, K\right)}{{}_1F_1\left(a+1, 2-\frac{a}{2}-\nu, K\right)} \quad (19)$$

Φαίνεται δια μέσου αριθμητικών υπολογισμών ότι η $K_{a,\nu}$ είναι πλησίον του $a+\nu-0,5$. Το γράφημα 5.1 δίνει τη διαφορά μεταξύ της λύσεως της (19) και της $a+\nu-0,5$ για τις διάφορες τιμές των a και ν . Αυτή η διαφορά είναι ασήμαντη στις περισσότερες περιπτώσεις αν και φαίνεται να είναι σημαντική για μεγάλες τιμές των a και ν .

Γράφημα 5.1

Τιμές της σταθεράς $K_{a,v}$ για τον εκτιμητή του Bayes $\delta^\pi(x) = \frac{K_{a,v}}{x+x_0}$ του θ στην

$$\Gamma\left(a, \frac{1}{\theta}\right) \div \Gamma\left(v, \frac{1}{x_0}\right)$$

περίπτωση, για $v = 1, 5, 10$ και σύγκριση με τη συνήθη σταθερά

$$a + v - 0.5$$

Αναζητούμε στη συνέχεια τον εκτιμητή του Bayes $\tilde{\delta}^\pi$ του θ υπό την entropic απώλεια η οποία, από τον Πίνακα 4.1, είναι :

$$L_E(\theta, d) = a \left(\frac{d}{\theta} - \log \frac{d}{\theta} - 1 \right)$$

Επειδή η δειγματική πυκνότητα $f(x|\theta) \left(\sim \Gamma \left(a, \frac{1}{\theta} \right) \right)$ και η εκ των προτέρων πυκνότητα

$\pi(\theta) \left(\sim \Gamma \left(\nu, \frac{1}{x_0} \right) \right)$ είναι της εκθετικής μορφής, εφαρμόζοντας την Πρόταση 5.1 για

$$\psi(\theta) = -a \log \theta \text{ και } \lambda = \frac{\nu-1}{a},$$

λαμβάνουμε

$$\tilde{\delta}^\pi(x) = \frac{a + \nu - 1}{x + x_0}.$$

Παρατηρούμε ότι ο $\tilde{\delta}^\pi$ έχει την ίδια μορφή με τον δ^π και η διαφορά τους είναι μόνο ως προς την σταθερά. ▲

Παράδειγμα 6: Θεωρούμε την Poisson περίπτωση : $x | \lambda \sim P(\lambda)$.

Βασιζόμενοι στο Π,§3, λαμβάνουμε ως εκ των προτέρων επί του λ την

$\Gamma \left(a, \frac{1}{\beta} \right) \left(\lambda \sim \Gamma \left(a, \frac{1}{\beta} \right) \right)$ οπότε η εκ των υστέρων επί του λ δεδομένου του x θα είναι η

$$\Gamma \left(x + a, \frac{1}{\beta + 1} \right) \left(\lambda | x \sim \Gamma \left(x + a, \frac{1}{\beta + 1} \right) \right).$$

Στην Poisson περίπτωση, η Hellinger απώλεια, όπως φαίνεται στον Πίνακα 4.2 είναι

$$L_H(\lambda, d) = 1 - \exp \left\{ -\frac{1}{2} (\sqrt{\lambda} - \sqrt{d})^2 \right\}.$$

Συνεπώς, ο εκτιμητής του Bayes δ^π του λ υπό την L_H απώλεια, θα αποτελεί λύση της

$$\begin{aligned} \min_d E^\pi(L_H(\lambda, d) | x) &= \min_d \int_{R_+^*} L_H(\lambda, d) \pi(\lambda | x) d\lambda \\ &= \min_d \int_0^{+\infty} \left(1 - \exp \left\{ -\frac{1}{2} (\sqrt{\lambda} - \sqrt{d})^2 \right\} \right) \pi(\lambda | x) d\lambda \\ &\Leftrightarrow \max_d \int_0^{+\infty} \exp \left\{ -\frac{1}{2} (\sqrt{\lambda} - \sqrt{d})^2 \right\} \pi(\lambda | x) d\lambda \end{aligned}$$

$$\Leftrightarrow \max_d \int_0^{+\infty} \lambda^{x+a-1} \exp\left\{-\frac{1}{2}(\sqrt{\lambda} - \sqrt{d})^2 - (\beta+1)\lambda\right\} d\lambda.$$

Αν θέσουμε $d = \varphi^2$ και $\lambda = \theta^2 (\Rightarrow d\lambda = 2d\theta)$ το τελευταίο πρόβλημα μεγιστοποίησης, γίνεται :

$$\max_{\varphi} \int_0^{+\infty} \theta^{2(x+a)-1} \exp\left\{-\frac{1}{2}(\theta - \varphi)^2 - (\beta+1)\theta^2\right\} d\theta \quad (20)$$

Επειδή

$$\int_0^{+\infty} \theta^{2(x+a)-1} \exp\left\{-\frac{1}{2}(\theta - \varphi)^2 - (\beta+1)\theta^2\right\} d\theta = e^{-\frac{\varphi^2}{2}} \int_0^{+\infty} \exp\left\{-\frac{\theta^2}{2}(2\beta+3) + \theta\varphi\right\} d\theta,$$

θέτοντας στο τελευταίο ολοκλήρωμα

$$\theta\sqrt{2\beta+3} = \eta \quad \left(\Rightarrow \theta = \frac{\eta}{\sqrt{2\beta+3}}, d\theta = \frac{d\eta}{\sqrt{2\beta+3}}\right) \text{ και } \gamma = \frac{1}{\sqrt{2\beta+3}}$$

έχουμε

$$\int_0^{+\infty} \theta^{2(x+a)-1} \exp\left\{-\frac{1}{2}(\theta - \varphi)^2 - (\beta+1)\theta^2\right\} d\theta = e^{-\frac{\varphi^2}{2}} \int_0^{+\infty} \gamma^{2(x+a)} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta.$$

Έτσι το πρόβλημα μεγιστοποίησης (20), ισοδυναμεί με το :

$$\max_{\varphi} e^{-\frac{\varphi^2}{2}} \int_0^{+\infty} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta \quad (21)$$

Για τη λύση του προβλήματος (21), θέτουμε

$$\Phi(\varphi) = e^{-\frac{\varphi^2}{2}} \int_0^{+\infty} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta.$$

Τότε, με τη δυνατότητα εναλλαγής παραγώγου και ολοκληρώματος, έχουμε :

$$\Phi'(\varphi) = -e^{-\frac{\varphi^2}{2}} \varphi \int_0^{+\infty} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta + e^{-\frac{\varphi^2}{2}} \gamma \int_0^{+\infty} \eta^{2(x+a)} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta$$

και η $\Phi'(\varphi) = 0$ οδηγεί στην συμπερασματική εξίσωση του φ ,

$$-e^{-\frac{\varphi^2}{2}} \varphi \int_0^{+\infty} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta + e^{-\frac{\varphi^2}{2}} \gamma \int_0^{+\infty} \eta^{2(x+a)} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta = 0 \quad (22)$$

Λοιπόν, ο εκτιμητής του Bayes φ του $\sqrt{\lambda}(=\theta)$, θα αποτελεί λύση της (22). Χρησιμοποιώντας τη φόρμουλα (19.5.3) των Abramowitz και Stegun (1964) :

$$U(\alpha, z) = \frac{1}{\Gamma\left(\frac{1}{2} + \alpha\right)} e^{-\frac{z^2}{2}} \int_0^{+\infty} e^{-zs - \frac{1}{2}s^2} s^{\alpha - \frac{1}{2}} ds$$

(με U σημειώνει την Whittaker συνάρτηση) αφενός για $\alpha = 2(x+a) - \frac{1}{2}$, $z = -\gamma\varphi$ και αφετέρου για $\alpha = 2(x+a) + \frac{1}{2}$, $z = -\gamma\varphi$ με $2(x+a) \in N$, στην (22), λαμβάνουμε :

$$\begin{aligned} \varphi &= \gamma \frac{\int_0^{+\infty} \eta^{2(x+a)} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta}{\int_0^{+\infty} \eta^{2(x+a)-1} \exp\left\{\eta\gamma\varphi - \frac{\eta^2}{2}\right\} d\eta} = \gamma \frac{\Gamma(2(x+a)+1)U\left(2(x+a) + \frac{1}{2}, -\varphi\gamma\right)}{\Gamma(2(x+a))U\left(2(x+a) - \frac{1}{2}, -\varphi\gamma\right)} \\ &= 2\gamma(x+a) \frac{U\left(2(x+a) + \frac{1}{2}, -\varphi\gamma\right)}{U\left(2(x+a) - \frac{1}{2}, -\varphi\gamma\right)} \end{aligned} \quad (23)$$

Εφαρμόζοντας στην (23) τη φόρμουλα (19.12.3) των Abramowitz και Stegun (1964) :

$$\begin{aligned} U(\alpha, \pm x) &= \frac{\sqrt{\pi} 2^{\frac{1}{4} \frac{\alpha}{2}} e^{-\frac{x^2}{4}} {}_1F_1\left(\frac{1}{2}\alpha + \frac{1}{4}, \frac{1}{2}, \frac{1}{2}x^2\right)}{\Gamma\left(\frac{3}{4} + \frac{\alpha}{2}\right)} \\ &\mp \frac{\sqrt{\pi} 2^{\frac{1}{4} \frac{\alpha}{2}} x e^{-\frac{x^2}{4}} {}_1F_1\left(\frac{1}{2}\alpha + \frac{3}{4}, \frac{3}{2}, \frac{1}{2}x^2\right)}{\Gamma\left(\frac{1}{4} + \frac{\alpha}{2}\right)} \end{aligned}$$

με τις προφανείς αντικαταστάσεις των α και x , λαμβάνουμε μετά από στοιχειώδεις πράξεις τη σχέση (24):

$$\varphi = \gamma(a+x) \frac{\frac{\sqrt{2}}{x+a} {}_1F_1\left(x+a+\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\varphi^2\gamma^2\right) + 2\varphi\gamma \frac{\Gamma(x+a)}{\Gamma\left(x+a+\frac{1}{2}\right)} {}_1F_1\left(x+a+1, \frac{3}{2}, \frac{1}{2}\varphi^2\gamma^2\right)}{\frac{\Gamma(x+a)}{\Gamma\left(x+a+\frac{1}{2}\right)} {}_1F_1\left(x+a, \frac{1}{2}, \frac{1}{2}\varphi^2\gamma^2\right) + \sqrt{2}\varphi\gamma {}_1F_1\left(x+a+\frac{1}{2}, \frac{3}{2}, \frac{1}{2}\varphi^2\gamma^2\right)}$$

Ακριβώς, όπως για τη Γάμμα περίπτωση (Παράδειγμα 5) είναι αρκετά δύσκολο να μελετήσουμε τις ιδιότητες της συμπερασματικής εξίσωσης (24).

Υπολογισμός του εκτιμητή του Bayes $\hat{\lambda}$ του λ , υπό την entropie απώλεια
Θεωρώντας όπως ανωτέρω

$$\left. \begin{array}{l} x | \lambda \sim P(\lambda) \\ \lambda \sim \Gamma\left(a, \frac{1}{\beta}\right) \end{array} \right\} \Rightarrow \lambda | x \sim \Gamma\left(x+a, \frac{1}{\beta+1}\right)$$

υπό την entropie απώλεια $L_E(\lambda, d) = \lambda(\log \lambda - \log d) + d - \lambda$ (Πίνακας 4.2) ο εκτιμητής του Bayes $\hat{\lambda}$ του λ , θα αποτελεί λύση της :

$$\begin{aligned} \min_d E^\pi(L_E(\lambda, d) | x) &= \min_d \int_{R_+^*} L_E(\lambda, d) \pi(\lambda | x) d\lambda \\ \Leftrightarrow \min_d \int_0^{+\infty} [\lambda(\log \lambda - \log d) + d - \lambda] \lambda^{x+a-1} e^{-(\beta+1)\lambda} d\lambda \end{aligned} \quad (25)$$

Για την επίλυση του προβλήματος ελαχιστοποίησης (25), θέτοντας

$$\Phi(d) = \int_0^{+\infty} [\lambda(\log \lambda - \log d) + d - \lambda] \lambda^{x+a-1} e^{-(\beta+1)\lambda} d\lambda$$

έχουμε

$$\Phi'(d) = -\frac{1}{d} \int_0^{+\infty} \lambda^{x+a} e^{-(\beta+1)\lambda} d\lambda + \int_0^{+\infty} \lambda^{x+a-1} e^{-(\beta+1)\lambda} d\lambda$$

και η $\Phi'(d) = 0$ δίνει τον ζητούμενο εκτιμητή του Bayes $\hat{\lambda}$

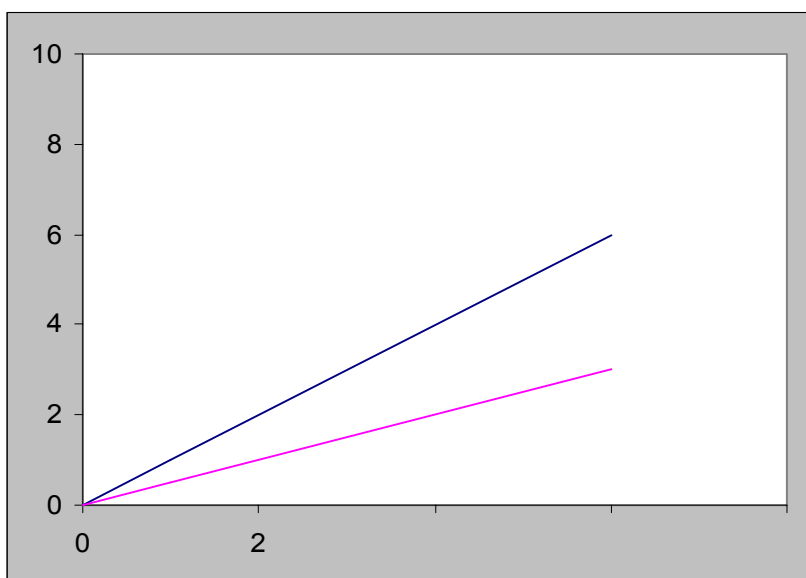
$$\hat{\lambda} = \frac{\int_0^{+\infty} \lambda^{x+a} e^{-(\beta+1)\lambda} d\lambda}{\int_0^{+\infty} \lambda^{x+a-1} e^{-(\beta+1)\lambda} d\lambda} = \frac{\Gamma(x+a+1)(\beta+1)^{-(x+a+1)}}{\Gamma(x+a)(\beta+1)^{-(x+a)}} = \frac{x+a}{\beta+1} \quad (26)$$

Όπως φαίνεται στο γράφημα 5.2, ο C. Robert παρουσιάζει κάποιες συγκρίσεις του προκύπτοντος από την (24), εκτιμητή του λ , φ^2 με τον $\hat{\lambda}$, εκτιμητή του Bayes υπό την

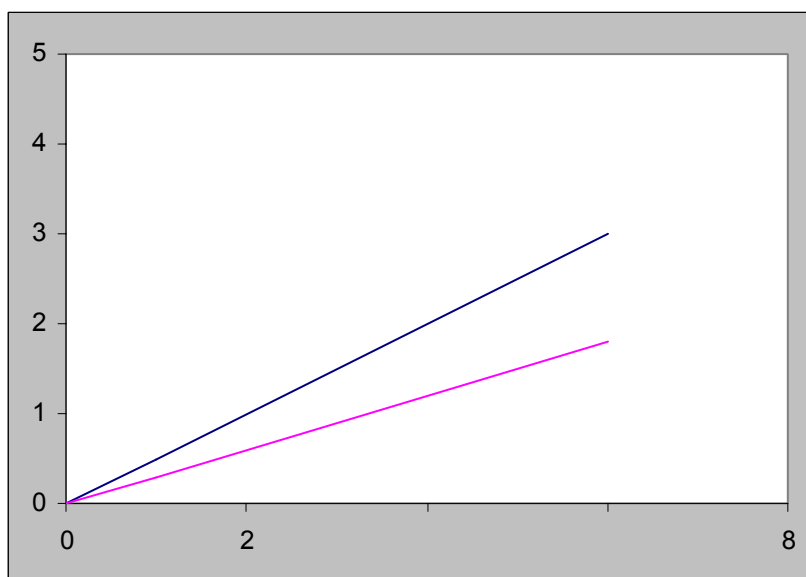
entropie απώλεια. Η διαφορά των δύο εκτιμητών φ^2 και $\hat{\lambda}$ είναι περισσότερο ενδιαφέρουσα από τη γάμμα περίπτωση. Όμως αμφότεροι οι εκτιμητές έχουν όμοια συμπεριφορά για τη μη-πληροφορική περίπτωση ($a = 0, \beta = 0$).

Γράφημα 5.2

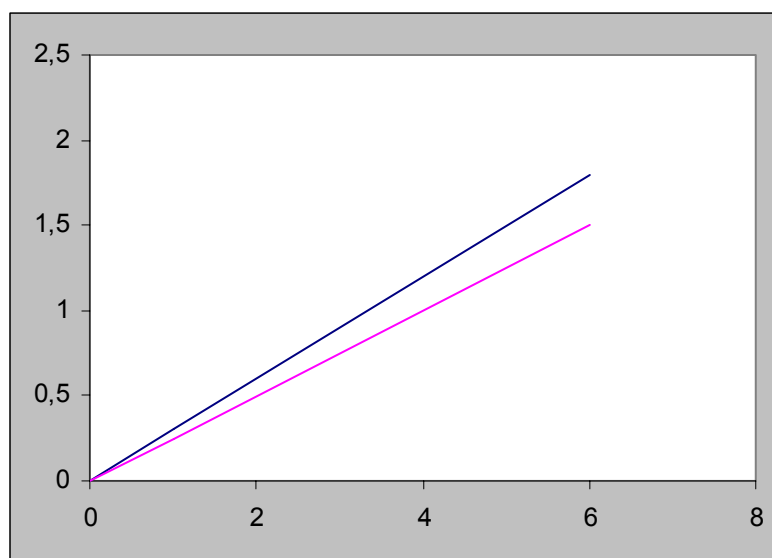
Τιμές του εκτιμητή του Bayes φ^2 του λ υπό την L_H -απώλεια (ροζ) και σύγκριση με τον εκτιμητή του Bayes υπό την L_E -απώλεια (μπλε) για $x \sim P(\lambda)$, $\lambda \sim \Gamma\left(a, \frac{1}{\beta}\right)$ και $\beta = 0, 1, 2$



$\beta = 0$



$\beta = 1$



$$\beta = 2$$

6. Συμπέρασμα

Σ' αυτό το κεφάλαιο έχουν δοθεί οι βασικές ιδιότητες των σύμφυτων απωλειών, δηλαδή ότι μπορούν να εκφραστούν αναλυτικά για τις περισσότερες εκθετικές οικογένειες και ότι η entropie απώλεια παρέχει πάντα μια αναλυτικού τύπου έκφραση για συζυγείς εκ των προτέρων όταν η εξίσωση $\nabla \psi(\hat{\theta}) = \frac{x + x_0}{\lambda + 1}$ μπορεί να αντιστραφεί. Τα παραδείγματα της Γάμμα και της Poisson στην §5 έδειξαν επίσης ότι η απώλεια Hellinger είναι λιγότερο εύχρηστη, παρόλο που είναι πιο φυσική (ως μια πραγματική απόσταση κατανομής, καθώς αυτή είναι μια φραγμένη απώλεια ισοδύναμη της L_1 -norme και της K^2 -απόστασης). Ωστόσο, η αντικατάσταση της Hellinger απώλειας από την entropie απώλεια δεν οδηγεί σε μεγάλες τροποποιήσεις στη συμπεριφορά του αντίστοιχου εκτιμητή του Bayes και συνεπώς μπορούμε να εισηγηθούμε τη χρήση της entropie απώλειας στην πράξη. Στην πραγματικότητα, είναι αδύνατον να κατασκευαστούν επαρκείς κλάσεις εκ των προτέρων κατανομών, όμοιες με τις συζυγείς κλάσεις, έτσι ώστε η εκ των υστέρων Hellinger απώλεια να είναι εύχρηστη.

Σημειώνουμε επίσης ότι η επίκληση της entropie απώλειας έναντι της Hellinger απώλειας, περιορίζεται στις εκ των προτέρων συζυγείς επειδή αυτές εμφανίζονται εξίσου επιβλητικές όπως οι άλλες εκ των προτέρων κατανομές. Μια πιο γενική συγκριτική μελέτη είναι λοιπόν απαραίτητη για να αποτιμηθούν οι διαφορές μεταξύ των δύο απωλειών και η ευπειθειά τους. Όπως αναφέρθηκε στην αρχή της §5, φαίνεται επίσης ότι οι σύμφυτες απώλειες απαιτούν διάφορες μη-πληροφορικές εκ των προτέρων που προσαρμόζονται στη μορφή τους. Συγκεκριμένα, οι εκ των προτέρων του Jeffrey φαίνονται ανεπαρκείς σε αυτή την τοποθέτηση επειδή μιας δευτέρας τάξεως προσέγγιση των απωλειών της entropie και του Hellinger, δείχνει ότι αυτές οι απώλειες ήδη ενσωματώνουν την πληροφορία κατά Fisher. Πράγματι, αυτή η απαίτηση μπορεί να επανατοποθετηθεί σε μια πιο γενική προοπτική όπου οι απώλειες και οι εκ των προτέρων κατανομές να καθορίζονται ταυτόχρονα σε όρους μιας πέρα για πέρα ωφέλειας. Εδώ λοιπόν έχουμε δώσει πολύ λίγα σ' ένα τόσο μεγάλο θέμα (βλέπε Rubin, 1987).

Παράρτημα

Σ' αυτό το παράρτημα θα δώσουμε κάποια μαθηματικά αποτελέσματα και βασικές ιδιότητες ορισμένων κατανομών που συναντάμε συχνά στην bayesienne ανάλυση και τις οποίες επικαλούμαστε στην εργασία.

1. Οι συναρτήσεις Γάμμα και Βήτα

1.1 Η συνάρτηση Γάμμα

Για κάθε $a > 0$ η συνάρτηση Γάμμα, $\Gamma(a)$, ορίζεται από την εξής σχέση :

$$\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} dt .$$

Ολοκληρώνοντας κατά μέρη παίρνουμε: $\Gamma(a+1) = a\Gamma(a)$

Εφαρμόζοντας αυτό τον τύπο διαδοχικά για ένα ακέραιο $a = n$ έχουμε :

$$\Gamma(n) = (n-1)!$$

Μερικές φορές χρησιμοποιούμε την τιμή: $\Gamma(1/2) = \sqrt{\pi}$.

1.2 Η συνάρτηση Βήτα

Για κάθε $a, \beta > 0$, η συνάρτηση Βήτα ορίζεται από το ολοκλήρωμα :

$$B(a, \beta) = \int_0^1 t^{a-1} (1-t)^{\beta-1} dt ,$$

αυτή συνδέεται με τη συνάρτηση Γάμμα μέσω της σχέσης : $B(a, \beta) = \frac{\Gamma(a)\Gamma(\beta)}{\Gamma(a+\beta)}$.

2. Γάμμα κατανομή και σχετικές με αυτήν κατανομές

Μια τ.μ. X θα λέμε ότι ακολουθεί την $\Gamma(a, \beta)$ – κατανομή, αν δέχεται π.π. την

$$f(x|a, \beta) = \frac{1}{\Gamma(a)\beta^a} x^{a-1} e^{-\frac{x}{\beta}} 1_{(0, +\infty)}(x) , \quad a, \beta > 0$$

όπου $\Gamma(a)$ η συνάρτηση Γάμμα στο σημείο a .

Οι πρώτες ροπές της $\Gamma(a, \beta)$ – κατανομής είναι :

$$E(X) = a\beta$$

$$Var(X) = a\beta^2$$

Ειδικές περιπτώσεις :

(i) Εκθετική κατανομή $\mathcal{E}(\lambda) = \Gamma(a = 1, \beta = 1/\lambda)$, $\lambda > 0$.

(ii) χ^2 – κατανομή με ν βαθμούς ελευθερίας : $\chi^2_\nu = \Gamma(a = \nu/2, \beta = 2)$, $\nu \in \mathbb{N}$

(iii) Γάμμα αντίστροφη κατανομή : $\Gamma^{-1}(a, \beta)$, $a, \beta > 0$

Αν η τ.μ. $X \sim \Gamma(a, \beta)$ τότε η τ.μ. $Y = 1/X \sim \Gamma^{-1}(a, \beta)$ με π.π. την

$$f_{iy}(y | a, \beta) = \frac{1}{\Gamma(a)\beta^a} y^{-(a+1)} e^{-\frac{1}{y\beta}} 1_{(0,+\infty)}(y)$$

και οι πρώτες ροπές της δίνονται από τις σχέσεις :

$$E(Y) = \frac{1}{(a-1)\beta}, \quad \text{για } a > 1$$

$$Var(Y) = \frac{1}{(a-1)(a-2)\beta^2} - \frac{1}{(a-1)^2\beta^2} = \frac{1}{(a-1)^2(a-2)\beta^2}, \quad \text{για } a > 2.$$

(βλέπε π.χ. στο [17]).

(iv) χ^2 - αντίστροφη : Η χ^2 – αντίστροφη είναι η κατανομή της αντιστρόφου μιας χ^2 – τ.μ. συνεπώς είναι μια μερική περίπτωση της Γάμμα αντιστρόφου :

$$X \sim \chi^2_\nu{}^{-2} = \Gamma^{-1}\left(\frac{\nu}{2}, 2\right)$$

λοιπόν δέχεται π.π. την

$$f_X(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}} x^{-\left(\frac{\nu}{2}+1\right)} e^{-\frac{1}{2x}} 1_{(0,+\infty)}(x), \quad \nu > 0$$

Για τις πρώτες ροπές έχουμε :

$$E(X) = \frac{1}{\nu-2}, \quad \nu > 2$$

$$Var(X) = \frac{2}{(\nu-2)^2(\nu-4)}, \quad \nu > 4.$$

(v) \mathcal{F} -κατανομή (ή κατανομή Fischer-Snedecor):

1. Έστω X_1, X_2 ανεξάρτητες τ.μ. τέτοιες ώστε $X_1 \sim \chi_{\nu_1}^2, X_2 \sim \chi_{\nu_2}^2$ τότε η τ.μ.

$$Y = \frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\nu_2}}$$

ακολουθεί την \mathcal{F} -κατανομή με ν_1, ν_2 βαθμούς ελευθερίας και γράφουμε

$$Y \sim \mathcal{F}_{\nu_1, \nu_2}, \text{ με π.π.}$$

$$f_Y(y) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} y^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2} y\right)^{-\frac{\nu_1+\nu_2}{2}} 1_{(0,+\infty)}(y)$$

Για τις πρώτες ροπές έχουμε :

$$E(Y) = \frac{\nu_2}{\nu_2 - 2}, \quad \nu_2 > 2$$

$$Var(Y) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \quad \nu_2 > 4.$$

(βλέπε π.χ στο [17]).

2. Θεωρώντας τον μετασχηματισμό $Z = \frac{1}{Y}$ λαμβάνουμε ότι η τ.μ Z δέχεται π.π. την :

$$f_Z(z) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} z^{-\left(\frac{\nu_1}{2}-1\right)} \left(1 + \frac{\nu_1}{\nu_2} \frac{1}{z}\right)^{-\frac{\nu_1+\nu_2}{2}} 1_{(0,+\infty)}(z)$$

Σ' αυτήν την περίπτωση λέμε ότι η Z ακολουθεί την *αντίστροφη* \mathcal{F} - κατανομή με ν_1, ν_2 βαθμούς ελευθερίας και σημειώνουμε $Z \sim \mathcal{F}_{\nu_1, \nu_2}^{-1}$. Παρατηρούμε ότι η π.π. της Z μπορεί να γραφτεί ως εξής :

$$f_Z(z) = \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_2}{2}} z^{\frac{\nu_2}{2}-1} \left(1 + \frac{\nu_2}{\nu_1} z\right)^{-\frac{\nu_1+\nu_2}{2}} 1_{(0,+\infty)}(z)$$

που αποτελεί την π.π. μιας $\mathcal{F}_{\nu_2, \nu_1}$ κατανομής. Δηλαδή, $\mathcal{F}_{\nu_1, \nu_2}^{-1} = \mathcal{F}_{\nu_2, \nu_1}$ (βλέπε π.χ. στο [17]).

3. Βήτα κατανομή και σχετικές με αυτήν κατανομές

Μια τ.μ. X θα λέμε ότι ακολουθεί την *Βήτα κατανομή με παραμέτρους* $\alpha, \beta > 0$ και σημειώνουμε : $X \sim \text{Beta}(\alpha, \beta)$ αν $X(\Omega) = [0, 1]$ και δέχεται π.π την :

$$f_{\beta}(x | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} 1_{[0,1]}(x)$$

όπου $B(\alpha, \beta)$ η Βήτα συνάρτηση.

Οι πρώτες ροπές της $\text{Beta}(\alpha, \beta)$ – κατανομής είναι :

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

Πράγματι, έχουμε

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_{\beta}(x | \alpha, \beta) dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+1} (1-x)^{\beta-1} dx = \frac{1}{B(\alpha, \beta)} \underbrace{\int_0^1 x^{\alpha} (1-x)^{\beta-1} dx}_{\text{συνάρτηση } B(\alpha+1, \beta)} \\ &= \frac{1}{B(\alpha, \beta)} B(\alpha+1, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \\ &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1)} = \frac{\alpha}{\alpha + \beta} \end{aligned}$$

Και για τη διασπορά έχουμε :

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f_{\beta}(x | \alpha, \beta) dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+2} (1-x)^{\beta-1} dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+1} (1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha, \beta)} B(\alpha+2, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2+\beta)} \\ &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 2)} = \frac{(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{(\alpha + \beta + 1)\Gamma(\alpha + \beta + 1)} \\ &= (\alpha+1)\alpha \frac{1}{(\alpha + \beta + 1)(\alpha + \beta)} \end{aligned}$$

Οπότε

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 = \frac{a(a+1)}{(a+\beta)(a+\beta+1)} - \left(\frac{a}{a+\beta}\right)^2 \\ &= \frac{a(a+1)(a+\beta) - a^2(a+\beta+1)}{(a+\beta+1)(a+\beta)^2} = \frac{a^3 + a^2\beta + \alpha^2 + a\beta - a^3 - a^2\beta - \alpha^2}{(\alpha+\beta+1)(a+\beta)^2} \\ &= \frac{a\beta}{(\alpha+\beta+1)(a+\beta)^2} \end{aligned}$$

Μια ενδιαφέρουσα μερική περίπτωση αυτής της κατανομής είναι η ομοιόμορφη στο $[0,1]$ ($U(0,1)$) η οποία προκύπτει για $a = \beta = 1$.

3.1 Βήτα – Διωνυμική διακριτή κατανομή

Η κατανομή Βήτα – Διωνυμική εμφανίζεται σαν ένα μείγμα των νόμων $Bin(n, \pi)$, όπου π κατανέμεται ακολουθώντας έναν νόμο $Beta(a, \beta)$ (με $Bin(n, \pi)$ σημειώνουμε τη διωνυμική κατανομή παραμέτρων n, π και με f_b θα σημειώνουμε την π.π. της). Συγκεκριμένα έχουμε: $X | \pi \sim Bin(n, \pi)$ και $\pi \sim Beta(a, \beta)$. Τότε η προκύπτουσα κατανομή για την X θα εξαρτάται από τρεις παραμέτρους (n, a, β) και η πυκνότητά της δίδεται από τη σχέση:

$$\begin{aligned} f_{\beta b}(x | n, a, \beta) &= \int_0^1 f_b(x | n, \pi) f_{\beta}(\pi | a, \beta) d\pi \\ &= \binom{n}{x} \frac{1}{B(a, \beta)} \underbrace{\int_0^1 \pi^{x+a-1} (1-\pi)^{n-x+\beta-1} d\pi}_{\text{ολοκλ. Beta}(x+a, n-x+\beta)\text{-κατανομής}} \\ &= \binom{n}{x} \frac{1}{B(a, \beta)} B(x+a, n-x+\beta) \\ &= \frac{n!}{x!(n-x)!} \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{\Gamma(x+a)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)} \\ &= \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{\Gamma(x+a)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)} \\ &= \frac{\Gamma(x+a)\Gamma(n-x+\beta)}{\Gamma(x+1)\Gamma(n-x+1)} \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(a+\beta+n)}, \quad \forall x \in \{0, 1, 2, \dots, n\}, \end{aligned}$$

με $n \in \mathbb{N}$ και $a, \beta > 0$.

Το γεγονός ότι η τ.μ. X ακολουθεί μια τέτοια κατανομή σημειώνεται:

$$X \sim BeBin(n, a, \beta).$$

Για τις πρώτες ροπές έχουμε :

$$E(X) = n \frac{a}{a + \beta}$$

$$Var(X) = n(n + a + \beta) \frac{a\beta}{(a + \beta)^2 (a + \beta + 1)}$$

Πράγματι έχουμε :

$$E(X) = \sum_{x=0}^n x f_X(x) = \sum_{x=1}^n x \frac{\Gamma(a+x)\Gamma(n-x+\beta)\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(a)\Gamma(\beta)\Gamma(a+\beta+n)}$$

$$= \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \sum_{x=1}^n \frac{\Gamma(a+x)\Gamma(n-x+\beta)}{\Gamma(x)\Gamma(n-x+1)},$$

επειδή $\Gamma(x+1) = x\Gamma(x)$.

Θέτοντας $x-1 = j (\Rightarrow x = j+1)$, η παραπάνω σχέση γράφεται :

$$E(X) = \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \sum_{j=0}^{n-1} \frac{\Gamma(a+j+1)\Gamma(n-j-1+\beta)}{\Gamma(j+1)\Gamma(n-j-1+1)}$$

$$= \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \frac{\Gamma(a+1)\Gamma(\beta)\Gamma(a+1+\beta+n-1)}{\Gamma(n)\Gamma(a+1+\beta)}$$

$$= \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \frac{\Gamma(a+1)\Gamma(\beta)\Gamma(a+\beta+n)}{\Gamma(n)\Gamma(a+\beta+1)}$$

$$= \frac{\Gamma(n+1)}{\Gamma(n)} \frac{\Gamma(a+\beta)}{\Gamma(a+\beta+1)} \frac{\Gamma(a+1)}{\Gamma(a)}$$

$$= \frac{na}{a+\beta}.$$

Για να υπολογίσουμε τη διασπορά $Var(X)$, μπορούμε να κάνουμε χρήση του τύπου :

$$Var(X) = E[X(X-1)] + E(X) - (E(X))^2$$

όπου :

$$E[X(X-1)] = \sum_{x=0}^n x(x-1) f_X(x) = \sum_{x=2}^n x(x-1) \frac{\Gamma(a+x)\Gamma(n-x+\beta)\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(a)\Gamma(\beta)\Gamma(a+\beta+n)}$$

$$= \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \sum_{j=0}^{n-2} \frac{\Gamma(a+j+2)\Gamma(n-j-2+\beta)}{\Gamma(j+2+1)\Gamma(n-j-2+1)}$$

$$\begin{aligned}
&= \frac{\Gamma(n+1)\Gamma(a+\beta)}{\Gamma(a)\Gamma(a+\beta+n)} \frac{\Gamma(a+2)\Gamma(\beta)\Gamma(a+2+\beta+n-2)}{\Gamma(n-1)\Gamma(a+2+\beta)} \\
&= \frac{n(n-1)a(a+1)}{(a+\beta)(a+\beta+1)} \\
&= \frac{(na)^2 + an^2 - na^2 - na}{(a+\beta)(a+\beta+1)}
\end{aligned}$$

Οπότε έχουμε :

$$\begin{aligned}
Var(X) &= \frac{(na)^2 + an^2 - na^2 - na}{(a+\beta)(a+\beta+1)} + \frac{na}{a+\beta} - \frac{(na)^2}{(a+\beta)^2} \\
&= \frac{a\beta n^2 - na^2\beta + 2na^2\beta + na\beta^2}{(a+\beta)^2(a+\beta+1)} = \frac{a\beta n^2 + na^2\beta + na\beta^2}{(a+\beta)^2(a+\beta+1)} \\
&= \frac{na\beta(n+a+\beta)}{(a+\beta)^2(a+\beta+1)}.
\end{aligned}$$

Παρατηρούμε ότι η μερική περίπτωση $a = \beta = 1$, δίνει για την X μια διακριτή ομοιόμορφη κατανομή επί του $\{0, 1, \dots, n\}$. Πράγματι :

$$\begin{aligned}
f_{\beta\beta}(x | n, 1, 1) &= \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(x+1)\Gamma(n-x+1)} \frac{\Gamma(n+1)\Gamma(2)}{\Gamma(1)\Gamma(1)\Gamma(n+2)} = \frac{\Gamma(n+1)}{(n+1)\Gamma(n+1)} 1_{\{0,1,\dots,n\}}(x) \\
&= \frac{1}{n+1} 1_{\{0,1,\dots,n\}}(x).
\end{aligned}$$

3.2 Πυκνότητα της αντιστρόφου Βήτα – 1

Αυτή την πυκνότητα την λαμβάνουμε μετασχηματίζοντας την τ.μ. $Y \sim Beta(a, \beta)$ στην $X = \frac{c}{Y}$, όπου $c > 0$. Η πυκνότητα της X είναι τότε η

$$f(x | a, \beta, c) = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{c^a (x-c)^{\beta-1}}{x^{a+\beta}} 1_{[c, +\infty)}(x)$$

με $a, \beta > 0$ και $c > 0$.

Πράγματι, θέτοντας $x = \frac{c}{y}$ ($\Rightarrow y = \frac{c}{x}$ και $\frac{dy}{dx} = -\frac{c}{x^2}$), έχουμε :

$$\begin{aligned} f(x | a, \beta, c) &= f_Y\left(\frac{c}{x}\right) \left| -\frac{c}{x^2} \right| = \frac{1}{B(a, \beta)} \left(\frac{c}{x}\right)^{a-1} \left(1 - \frac{c}{x}\right)^{\beta-1} \frac{c}{x^2} 1_{[c, +\infty)}(x) \\ &= \frac{1}{B(a, \beta)} \left(\frac{c}{x}\right)^{a-1} \left(\frac{x-c}{x}\right)^{\beta-1} \frac{c}{x^2} 1_{[c, +\infty)}(x) \\ &= \frac{1}{B(a, \beta)} \frac{c^a (x-c)^{\beta-1}}{x^{a+\beta}} 1_{[c, +\infty)}(x) \end{aligned}$$

Αυτή θα τη σημειώνουμε με $f_{i\beta 1}(x | a, \beta, c)$ και το γεγονός ότι η X ακολουθεί μια τέτοια κατανομή θα το σημειώνουμε

$$X \sim \text{Beinv.1}(a, \beta, c)$$

Οι πρώτες ροπές της δίνονται από τις σχέσεις :

$$\begin{aligned} E(X) &= c \frac{a + \beta - 1}{a - 1}, \quad a > 1 \\ \text{Var}(X) &= c^2 \frac{(a + \beta - 1)\beta}{(a - 1)^2(a - 2)}. \end{aligned}$$

Πράγματι, έχουμε

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_{i\beta 1}(x | a, \beta, c) dx = \int_c^{+\infty} x \frac{1}{B(a, \beta)} \frac{c^a (x-c)^{\beta-1}}{x^{a+\beta}} dx = \frac{c^a}{B(a, \beta)} \underbrace{\int_c^{+\infty} \frac{(x-c)^{\beta-1}}{x^{a+\beta-1}} dx}_{\text{οολκλ. Beinv.1}(a-1, \beta, c)} \\ &= \frac{c^a}{B(a, \beta)} \cdot \frac{B(a-1, \beta)}{c^{a-1}} = c \frac{\Gamma(a+\beta) \Gamma(a-1)\Gamma(\beta)}{\Gamma(a)\Gamma(\beta) \Gamma(a+\beta-1)} = c \frac{\Gamma(a-1)}{\Gamma(a)} \frac{\Gamma(a+\beta)}{\Gamma(a+\beta-1)} \\ &= c \frac{a + \beta - 1}{a - 1}. \end{aligned}$$

Επίσης,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_{i\beta 1}(x | a, \beta, c) dx = \int_c^{+\infty} x^2 \frac{1}{B(a, \beta)} \frac{c^a (x-c)^{\beta-1}}{x^{a+\beta}} dx$$

$$\begin{aligned}
&= \frac{c^a}{B(a, \beta)} \underbrace{\int_c^{+\infty} \frac{(x-c)^{\beta-1}}{x^{a+\beta-2}} dx}_{\text{οολκλ. } Be.inv.1(a-2, \beta, c)} = \frac{c^a}{B(a, \beta)} \cdot \frac{B(a-2, \beta)}{c^{a-2}} = c^2 \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{\Gamma(a-2)\Gamma(\beta)}{\Gamma(a+\beta-2)} \\
&= c^2 \frac{\Gamma(a-2)}{\Gamma(a)} \frac{\Gamma(a+\beta)}{\Gamma(a+\beta-2)} = c^2 \frac{(a+\beta-1)(a+\beta-2)}{(a-1)(a-2)}
\end{aligned}$$

Οπότε προκύπτει ότι

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 = c^2 \frac{(a+\beta-1)(a+\beta-2)}{(a-1)(a-2)} - \frac{c^2(a+\beta-1)^2}{(a-1)^2} \\
&= c^2 \frac{(a+\beta-1)(a+\beta-2)(a-1) - (a+\beta-1)^2(a-2)}{(a-1)^2(a-2)} \\
&= \frac{c^2(a+\beta-1)}{(a-1)^2(a-2)} ((a+\beta-2)(a-1) - (a+\beta-1)(a-2)) \\
&= \frac{c^2(a+\beta-1)}{(a-1)^2(a-2)} (a^2 - a + \beta a - \beta - 2a + 2 - a^2 + 2a - \beta a + 2\beta + a - 2) \\
&= \frac{c^2 \beta(a+\beta-1)}{(a-1)^2(a-2)}
\end{aligned}$$

3.3 Πυκνότητα της αντιστρόφου Βήτα – 2

Αυτή η πυκνότητα λαμβάνεται όταν μετασχηματίσουμε την τ.μ. $Y \sim Beta(a, \beta)$ στην $X = c \frac{Y}{1-Y}$, όπου $c > 0$. Η πυκνότητα της X είναι τότε

$$f(x | a, \beta, c) = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{c^\beta x^{a-1}}{(x+c)^{a+\beta}} 1_{[0, +\infty)}(x),$$

με $a, \beta, c > 0$. Αυτή θα τη σημειώνουμε με $f_{i\beta 2}(x | a, \beta, c)$.

Πράγματι, αν $x = c \frac{y}{1-y}$ ($= g(y)$) έχουμε $y = \frac{x}{x+c}$ με $x \in [0, +\infty)$ αφού $y \in [0, 1]$.

Επειδή ο μετασχηματισμός $x = g(y)$ είναι μονότονος (\uparrow) και διαφορίσιμος με $\frac{dy}{dx} = \frac{d}{dx} g^{-1}(y) = \frac{c}{(c+x)^2}$, έχουμε :

$$\begin{aligned} f(x | a, \beta, c) &= f_Y \left(\frac{x}{x+c} \right) \frac{c}{(c+x)^2} 1_{[0, +\infty)}(x) \\ &= \frac{1}{B(a, \beta)} \left(\frac{x}{c+x} \right)^{a-1} \left(1 - \frac{x}{x+c} \right)^{\beta-1} \frac{c}{(c+x)^2} 1_{[0, +\infty)}(x) \\ &= \frac{1}{B(a, \beta)} \left(\frac{x}{c+x} \right)^{a-1} \left(\frac{c}{x+c} \right)^{\beta-1} \frac{c}{(c+x)^2} 1_{[0, +\infty)}(x) \\ &= \frac{1}{B(a, \beta)} x^{a-1} c^{\beta-1+1} (x+c)^{-(a-1)-(\beta-1)-2} 1_{[0, +\infty)}(x) \\ &= \frac{c^\beta}{B(a, \beta)} x^{a-1} (x+c)^{-a-\beta} 1_{[0, +\infty)}(x). \end{aligned}$$

Το γεγονός ότι η X ακολουθεί μια τέτοια κατανομή θα το σημειώνουμε ως εξής

$$X \sim Be.inv.2(a, \beta, c)$$

Οι πρώτες ροπές της δίδονται από τις παρακάτω σχέσεις

$$E(X) = c \frac{a}{\beta-1}, \quad \beta > 1$$

$$Var(X) = c^2 \frac{a(a+\beta-1)}{(\beta-1)^2(\beta-2)}, \quad \beta > 2.$$

Πράγματι ,

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x^2 f(x | a, \beta, c) dx = \frac{c^\beta}{B(a, \beta)} \int_0^{+\infty} \frac{1}{(x+c)^\beta} \left(\frac{x}{x+c} \right)^a dx \\ &= \frac{c^\beta}{B(a, \beta)} \int_0^{+\infty} \frac{x^{a+1-1}}{(x+c)^{(a+1)+(\beta-1)}} dx = \frac{c^\beta}{B(a, \beta)} \cdot \frac{B(a+1, \beta-1)}{c^{\beta-1}} \\ &\quad \underbrace{\int_0^{+\infty} \frac{x^{a+1-1}}{(x+c)^{(a+1)+(\beta-1)}} dx}_{ολοκλ. Be.inv.2(a+1, \beta-1, c)} \end{aligned}$$

$$= c \frac{\Gamma(a+1)}{\Gamma(a)} \cdot \frac{\Gamma(\beta-1)}{\Gamma(\beta)} = c \frac{a}{\beta-1}, \quad \beta > 1$$

Για τον υπολογισμό της διασποράς έχουμε:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x | a, \beta, c) dx = \int_0^{+\infty} \frac{c^\beta}{B(a, \beta)} \frac{x^{a+1}}{(x+c)^{a+\beta}} dx \\ &= \frac{c^\beta}{B(a, \beta)} \int_0^{+\infty} \frac{x^{a+2-1}}{\underbrace{(x+c)^{(a+2)+(\beta-2)}}_{\text{ολοκλ. Be.inv.2}(a+2, \beta-2, c)}} dx = \frac{c^\beta}{B(a, \beta)} \cdot \frac{B(a+2, \beta-2)}{c^{\beta-2}} \\ &= \frac{c^2}{B(a, \beta)} B(a+2, \beta-2) = c^2 \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \frac{\Gamma(a+2)\Gamma(\beta-2)}{\Gamma(a+\beta)} = c^2 \frac{\Gamma(a+2)}{\Gamma(a)} \frac{\Gamma(\beta-2)}{\Gamma(\beta)} \\ &= c^2 \frac{a(a+1)}{(\beta-1)(\beta-2)}, \quad \beta > 2. \end{aligned}$$

Άρα, όταν $\beta > 2$,

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 = c^2 \frac{a(a+1)}{(\beta-1)(\beta-2)} - c^2 \frac{a^2}{(\beta-1)^2} \\ &= \frac{c^2}{(\beta-1)^2(\beta-2)} (a^2\beta - a^2 + a\beta - a - a^2\beta + 2a^2) \\ &= \frac{c^2}{(\beta-1)^2(\beta-2)} (a^2 + a\beta - a) = \frac{c^2 a(a+\beta-1)}{(\beta-1)^2(\beta-2)}. \end{aligned}$$

Μερική περίπτωση : Πυκνότητα της Fisher- Snedecor

Αν $X \sim \text{Be.inv.2}(a, \beta, c)$ και λάβουμε $a = \frac{\nu_1}{2}$, $\beta = \frac{\nu_2}{2}$ και $c = \frac{\nu_2}{\nu_1}$ με $\nu_1, \nu_2 \in \mathbb{N}$, προκύπτει η Fisher – Snedecor κατανομή με ν_1, ν_2 βαθμούς ελευθερίας: $\mathcal{F}_{\nu_1, \nu_2}$ (βλέπε §2).

4. Αρνητική – Διωνυμική κατανομή

Θα λέμε ότι μια τ.μ. X ακολουθεί αρνητική διωνυμική κατανομή και γράφουμε $X \sim \text{NegBin}(r, p)$, αν έχει συνάρτηση πυκνότητας την :

$$f(x | r, p) = \binom{r+x-1}{x} p^r (1-p)^x 1_{\{0,1,\dots\}}(x).$$

Αυτή θα τη σημειώνουμε με $f_{nb}(x | r, p)$

Οι πρώτες ροπές της αρνητικής διωνυμικής κατανομής είναι :

$$E(X) = \frac{r(1-p)}{p}$$

$$Var(X) = \frac{r(1-p)}{p^2}$$

Πράγματι, γνωρίζοντας ότι : $\sum_{x=0}^{+\infty} \binom{r+x-1}{x} q^x = (1-q)^{-r}$, όταν $|q| < 1$, έχουμε :

$$\begin{aligned} E(X) &= \sum_{x=0}^{+\infty} x \binom{r+x-1}{x} p^r (1-p)^x = \sum_{x=1}^{+\infty} x \frac{(r+x-1)!}{x!(r-1)!} p^r (1-p)^x \\ &= p^r \sum_{x=1}^{+\infty} \frac{(r+x-1)!}{(x-1)!(r-1)!} (1-p)^x = p^r \sum_{y=0}^{+\infty} \frac{(r+y)!}{y!(r-1)!} (1-p)^{y+1} \\ &= rp^r (1-p) \sum_{y=0}^{+\infty} \frac{(r+y)!}{y!r!} (1-p)^y = rp^r (1-p) \sum_{y=0}^{+\infty} \binom{r+y}{y} (1-p)^y \\ &= rp^r (1-p) p^{-(r+1)} = r(1-p)p^{-1}. \end{aligned}$$

Αντίστοιχα για τη διασπορά παίρνουμε :

$$\begin{aligned} E(X(X-1)) &= \sum_{x=0}^{+\infty} x(x-1) \binom{r+x-1}{x} p^r (1-p)^x = p^r \sum_{x=2}^{+\infty} \frac{(r+x-1)!}{(x-2)!(r-1)!} (1-p)^x \\ &= p^r \sum_{y=0}^{+\infty} \frac{(r+y+1)!}{y!(r-1)!} (1-p)^{y+2} = p^r (1-p)^2 r(r+1) \sum_{y=0}^{+\infty} \binom{r+y+1}{y} (1-p)^y \\ &= r(r+1)p^r (1-p)^2 p^{-(r+2)} = r(r+1) \frac{(1-p)^2}{p^2}. \end{aligned}$$

Άρα,

$$\begin{aligned} Var(X) &= E(X(X-1)) + E(X) - (E(X))^2 = r(r+1) \frac{(1-p)^2}{p^2} + \frac{r(1-p)}{p} - \frac{r^2(1-p)^2}{p^2} \\ &= \frac{r(r+1)(1-p)^2 + rp(1-p) - r^2(1-p)^2}{p^2} \\ &= \frac{r(1-p)((r+1)(1-p) + p - r(1-p))}{p^2} \\ &= \frac{r(1-p)(r - rp + 1 - p + p - r + rp)}{p^2} = \frac{r(1-p)}{p^2}. \end{aligned}$$

5. Πολυωνυμική κατανομή και κατανομή Dirichlet

5.1 Πολυωνυμική κατανομή

Η πολυωνυμική κατανομή παραμέτρων (n, π_1, \dots, π_k) χαρακτηρίζεται από τη συνάρτηση πιθανότητας k – μεταβλητών:

$$f(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) = \frac{n!}{n_1! \dots n_k! (n - n_1 - \dots - n_k)!} \pi_1^{n_1} \dots \pi_k^{n_k} (1 - \pi_1 - \dots - \pi_k)^{n - n_1 - \dots - n_k}$$

Οι παράμετροι (π_1, \dots, π_k) ανήκουν στο χωρίο $S_k (\subset R^k)$:

$$S_k = \left\{ {}^t(\pi_1, \dots, \pi_k) \in R^k : \pi_j \geq 0, j = 1, 2, \dots, k, \sum_{j=1}^k \pi_j \leq 1 \right\}$$

Αυτός ο νόμος εμφανίζεται σε μια ακολουθία n ανεξάρτητων παρατηρήσεων ενός πειράματος τύχης, όπου λαμβάνεται υπόψη ο αριθμός n_j των φορών όπου ένα γεγονός A_j $j = 1, 2, \dots, k+1$ πραγματοποιείται : τα A_1, \dots, A_{k+1} σχηματίζουν μια διαμέριση του συνόλου των δυνατών αποτελεσμάτων του πειράματος και $\pi_j = P(A_j)$, $j = 1, 2, \dots, k+1$ με $\sum_{j=1}^{k+1} \pi_j = 1$. Έχουμε λοιπόν, $n_j \in \{0, 1, \dots, n\}$, $j = 1, 2, \dots, k+1$ και $\sum_{j=1}^{k+1} n_j = n$. Το γεγονός ότι μια k -διάστατη τ.μ. ${}^t(X_1, \dots, X_k)$ ακολουθεί μια τέτοια κατανομή παραμέτρων n, π_1, \dots, π_k με $\pi_j \geq 0$ και $\sum_{j=1}^{k+1} \pi_j = 1$, σημειώνεται :

$${}^t(X_1, \dots, X_k) \sim Mult(n, \pi_1, \dots, \pi_k).$$

Αποδεικνύεται ότι :

$$\begin{aligned} E(X_i) &= n\pi_i \\ Var(X_i) &= n\pi_i(1 - \pi_i) \\ Cov(X_i, X_j) &= -n\pi_i\pi_j, \quad \forall i \neq j, i, j \in \{1, \dots, k\} \end{aligned}$$

Πράγματι, κάνοντας χρήση του πολυωνυμικού αναπτύγματος, έχουμε : $\forall i \in \{1, 2, \dots, k\}$,

$$f_{X_i}(n_i | n, \pi_i) = \sum_{n_j, j=1, \dots, k, j \neq i} f(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k)$$

$$\begin{aligned}
&= \sum_{\substack{n_j, j=1, \dots, k, j \neq i \\ \sum_{j \neq i} n_j \leq n - n_i}} \frac{n!}{n_1! \dots n_k! (n - n_1 - \dots - n_k)!} \cdot \pi_1^{n_1} \dots \pi_k^{n_k} \left(1 - \sum_{j=1}^k \pi_j\right)^{n - \sum_{j=1}^k n_j} \\
&= \frac{n!}{n_i! (n - n_i)!} \pi_i^{n_i} \sum_{\substack{n_j, j=1, \dots, k+1, j \neq i \\ \sum_{j \neq i} n_j \leq n - n_i}} \frac{(n - n_i)!}{\prod_{j \neq i} n_j! (n - n_i - \sum_{j \neq i} n_j)!} \prod_{j \neq i} \pi_j^{n_j} \left(1 - \pi_i - \sum_{j \neq i} \pi_j\right)^{n - n_i - \sum_{j \neq i} n_j} \\
&= \frac{n!}{n_i! (n - n_i)!} \pi_i^{n_i} \left(\sum_{j \neq i} \pi_j + 1 - \pi_i - \sum_{j \neq i} \pi_j\right)^{n - n_i} = \frac{n!}{n_i! (n - n_i)!} \pi_i^{n_i} (1 - \pi_i)^{n - n_i}.
\end{aligned}$$

Άρα, $X_i \sim \text{Bin}(n, \pi_i)$ και επομένως

$$E(X_i) = n\pi_i \text{ και } \text{Var}(X_i) = n\pi_i(1 - \pi_i), \quad \forall i \in \{1, 2, \dots, k\}$$

Για να υπολογίσουμε τη συνδιασπορά θα κάνουμε χρήση της σχέσης:

$$\text{Cov}(X_i, X_j) = E(X_i \cdot X_j) - E(X_i)E(X_j)$$

Παρατηρούμε ότι: $\forall i, j \in \{1, 2, \dots, k\}$ και $i \neq j$,

$$\begin{aligned}
f_{X_i, X_j}(n_i, n_j | n, \pi_i, \pi_j) &= \sum_{\substack{n_l, l \neq i, j \\ \sum_{l \neq i, j} n_l \leq n - n_i - n_j}} f(n_1, \dots, n_k | n, \pi_1, \dots, \pi_k) \\
&= \sum_{\substack{n_l, l \neq i, j \\ \sum_{l \neq i, j} n_l \leq n - n_i - n_j}} \frac{n!}{n_1! \dots n_k! (n - n_1 - \dots - n_k)!} \cdot \pi_1^{n_1} \dots \pi_k^{n_k} \left(1 - \sum_{l=1}^k \pi_l\right)^{n - \sum_{l=1}^k n_l} \\
&= \frac{n!}{n_i! n_j! (n - n_i - n_j)!} \pi_i^{n_i} \pi_j^{n_j} \sum_{\substack{n_l, l \neq i, j \\ \sum_{l \neq i, j} n_l \leq n - n_i - n_j}} \frac{(n - n_i - n_j)!}{\prod_{l \neq i, j} n_l! (n - n_i - n_j - \sum_{l \neq i, j} n_l)!} \times \\
&\quad \times \prod_{l \neq i, j} \pi_l^{n_l} \left(1 - \pi_i - \pi_j - \sum_{l \neq i, j} \pi_l\right)^{n - n_i - n_j - \sum_{l \neq i, j} n_l} \\
&= \frac{n!}{n_i! n_j! (n - n_i - n_j)!} \pi_i^{n_i} \pi_j^{n_j} \left(\sum_{l \neq i, j} \pi_l + 1 - \pi_i - \pi_j - \sum_{l \neq i, j} \pi_l\right)^{n - n_i - n_j} \\
&= \frac{n!}{n_i! n_j! (n - n_i - n_j)!} \pi_i^{n_i} \pi_j^{n_j} (1 - \pi_i - \pi_j)^{n - n_i - n_j},
\end{aligned}$$

δηλαδή $(X_i, X_j) \sim \text{Mult}(n, \pi_i, \pi_j)$, $\forall i, j \in \{1, 2, \dots, k\}$ και $i \neq j$.

Επομένως,

$$\begin{aligned}
E(X_i X_j) &= \sum_{\substack{n_i, n_j \in \{0, \dots, n\} \\ n_i + n_j \leq n}} n_i n_j f_{X_i, X_j}(n_i, n_j | n, \pi_i, \pi_j) \\
&= \pi_i \pi_j \sum_{\substack{n_i, n_j \in \{1, \dots, n\} \\ n_i + n_j \leq n}} \frac{n!}{(n_i - 1)! (n_j - 1)! (n - n_i - n_j)!} \pi_i^{n_i} \pi_j^{n_j} (1 - \pi_i - \pi_j)^{n - n_i - n_j} \\
&= \frac{n!}{(n-2)!} \pi_i \pi_j \sum_{\substack{n_i, n_j \in \{0, \dots, n-1\} \\ n_i + n_j \leq n-2}} \frac{(n-2)!}{n_i! n_j! (n-2 - n_i - n_j)!} \pi_i^{n_i} \pi_j^{n_j} (1 - \pi_i - \pi_j)^{n-2 - n_i - n_j} \\
&= \frac{n!}{(n-2)!} \pi_i \pi_j (\pi_i + \pi_j + 1 - \pi_i - \pi_j)^{n-2} = \frac{n!}{(n-2)!} \pi_i \pi_j = n(n-1) \pi_i \pi_j.
\end{aligned}$$

Λοιπόν,

$$Cov(X_i, X_j) = n(n-1) \pi_i \pi_j - n^2 \pi_i \pi_j = -n \pi_i \pi_j, \quad \forall i, j \in \{1, 2, \dots, k\} \text{ και } i \neq j.$$

5.2 Κατανομή Dirichlet

Η πυκνότητα Dirichlet είναι μια πολυδιάστατη γενίκευση της Βήτα κατανομής. Εδώ η τ.μ $X = {}^t(X_1, \dots, X_k)$ με μια κατανομή Dirichlet είναι συνεχής και με π.π.

$$f(x_1, \dots, x_k | a_1, \dots, a_{k+1}) = \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1) \dots \Gamma(a_{k+1})} x_1^{a_1-1} \dots x_k^{a_k-1} (1 - x_1 - \dots - x_k)^{a_{k+1}-1} \mathbf{1}_{S_k}(x_1, \dots, x_k),$$

$$\text{όπου } S_k = \left\{ (x_1, \dots, x_k) \in \mathbb{R}^k : x_i \geq 0, i = 1, \dots, k, \sum_{i=1}^k x_i \leq 1 \right\} \text{ και } a_j > 0, j = 1, \dots, k+1,$$

παράμετροι.

Αυτή την πυκνότητα θα την σημειώνουμε μερικές φορές : $f_{dir}(x_1, \dots, x_k | a_1, \dots, a_{k+1})$ και το γεγονός ότι η τ.μ. $X = {}^t(X_1, \dots, X_k)$, ακολουθεί μια τέτοια κατανομή θα το σημειώνουμε :

$$X \sim Dir(a_1, \dots, a_{k+1})$$

Οι κύριες ροπές της είναι :

$$E(X_i) = \frac{a_i}{a_1 + \dots + a_{k+1}}$$

$$Var(X_i) = \frac{a_i(a_1 + \dots + a_{k+1} - a_i)}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)}$$

$$\text{Cov}(X_i, X_j) = -\frac{a_i a_j}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)}$$

Πράγματι, για να υπολογίσουμε τις παραπάνω ροπές αρκεί να υπολογίσουμε την παραγοντική ροπή (r_1, \dots, r_k) -τάξης της $X = (X_1, \dots, X_k)$, τότε :

$$\begin{aligned} E(X_1^{r_1} \dots X_k^{r_k}) &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1) \dots \Gamma(a_{k+1})} \int_0^1 \dots \int_0^1 x_1^{a_1+r_1-1} \dots x_k^{a_k+r_k-1} (1-x_1-\dots-x_k)^{a_{k+1}-1} dx_1 \dots dx_k \\ &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1) \dots \Gamma(a_{k+1})} \frac{\Gamma(a_{k+1}) \Gamma(a_1 + r_1) \dots \Gamma(a_k + r_k)}{\Gamma(a_{k+1} + a_1 + \dots + a_k + r_1 + \dots + r_k)} \\ &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + r_1 + \dots + r_k)} \frac{\Gamma(a_1 + r_1)}{\Gamma(a_1)} \dots \frac{\Gamma(a_k + r_k)}{\Gamma(a_k)} \end{aligned}$$

Για $r_i = 1, r_s = 0$ για $s \neq i$, παίρνουμε :

$$E(X_i) = \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + 1)} \frac{\Gamma(a_i + 1)}{\Gamma(a_i)} = \frac{a_i}{a_1 + \dots + a_{k+1}}.$$

Ομοίως από την παραπάνω σχέση, για $r_i = 2, r_s = 0, \forall s \neq i$, έχουμε :

$$\begin{aligned} E(X_i^2) &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + 2)} \frac{\Gamma(a_i + 2)}{\Gamma(a_i)} \\ &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{(a_1 + \dots + a_{k+1} + 1) \Gamma(a_1 + \dots + a_{k+1} + 1)} \cdot \frac{(a_i + 1) \Gamma(a_i + 1)}{\Gamma(a_i)} \\ &= \frac{a_i + 1}{a_1 + \dots + a_{k+1} + 1} \cdot \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + 1)} \cdot \frac{\Gamma(a_i + 1)}{\Gamma(a_i)} \\ &= \frac{a_i + 1}{a_1 + \dots + a_{k+1} + 1} \cdot \frac{a_i}{a_1 + \dots + a_{k+1}} \end{aligned}$$

Άρα,

$$\begin{aligned} \text{Var}(X_i) &= E(X_i^2) - (E(X_i))^2 = \frac{a_i(a_i + 1)}{(a_1 + \dots + a_{k+1})(a_1 + \dots + a_{k+1} + 1)} - \frac{a_i^2}{(a_1 + \dots + a_{k+1})^2} \\ &= \frac{a_i(a_i + 1)(a_1 + \dots + a_{k+1}) - a_i^2(a_1 + \dots + a_{k+1} + 1)}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)} \\ &= \frac{a_i^2(a_1 + \dots + a_{k+1} - a_1 - \dots - a_{k+1} - 1) + a_i(a_1 + \dots + a_{k+1})}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)} \end{aligned}$$

$$= \frac{a_i(a_1 + \dots + a_{k+1} - a_i)}{(a_1 + \dots + a_{k+1})^2(a_1 + \dots + a_{k+1} + 1)}.$$

Τέλος, αν θέσουμε $r_i = 1, r_j = 1, r_s = 0, \forall s \neq i \neq j$ παίρνουμε :

$$\begin{aligned} E(X_i X_j) &= \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + 1 + 1)} \frac{\Gamma(a_i + 1)}{\Gamma(a_i)} \frac{\Gamma(a_j + 1)}{\Gamma(a_j)} \\ &= \frac{a_i a_j}{a_1 + \dots + a_{k+1} + 1} \frac{\Gamma(a_1 + \dots + a_{k+1})}{\Gamma(a_1 + \dots + a_{k+1} + 1)} \\ &= \frac{a_i a_j}{(a_1 + \dots + a_{k+1})(a_1 + \dots + a_{k+1} + 1)} \end{aligned}$$

Τότε,

$$\begin{aligned} Cov(X_i, X_j) &= \frac{a_i a_j}{(a_1 + \dots + a_{k+1})(a_1 + \dots + a_{k+1} + 1)} - \frac{a_i a_j}{(a_1 + \dots + a_{k+1})^2} \\ &= \frac{a_i a_j (a_1 + \dots + a_{k+1}) - a_i a_j (a_1 + \dots + a_{k+1} + 1)}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)} \\ &= - \frac{a_i a_j}{(a_1 + \dots + a_{k+1})^2 (a_1 + \dots + a_{k+1} + 1)} \end{aligned}$$

Θα πρέπει να σημειώσουμε ότι για $k = 1$ ξαναβρίσκουμε την $Beta(a_1, a_2)$ - κατανομή.

6. k – διάστατη Κανονική Κατανομή (ή κατανομή των Laplace-Gauss στον \mathbb{R}^k)

Η πολυδιάστατη Κανονική κατανομή παίζει θεμελιώδη ρόλο στη Στατιστική Ανάλυση. Αυτή περιγράφεται μέσα σ' όλα τα βασικά συγγράμματα της Στατιστικής (βλέπε π.χ. στο [4]).

Είναι γνωστό ότι την κατανομή μιας τ.μ. μπορούμε να την ορίσουμε από την συνάρτηση κατανομής της ή από την χαρακτηριστική της συνάρτηση (χ.σ.). Στην περίπτωση της L-G κατανομής φαίνεται πιο κατάλληλη η χρήση της χ.σ.

6.1. Ορισμοί

6.1.1. Έστω Γ ένας (k, k) - πίνακας συμμετρικός και θετικός επί του \mathbb{R} . Θα λέμε ότι η διανυσματική τ.μ $M = (X_1, \dots, X_k)$ ακολουθεί την *κεντρική k - διάστατη Κανονική* (ή την κεντρική κατανομή των L-G μέσα στον \mathbb{R}^k) αν δέχεται χ.σ. την

$$\varphi_M(u) = E(e^{i'uM}) = \exp\left\{-\frac{1}{2}{}^t u \Gamma u\right\}, \forall u = {}^t(u_1, \dots, u_k) \in R^k$$

Αν ο πίνακας Γ είναι αντιστρέψιμος, ο νόμος του M λέγεται *μη – εκφυλισμένος*, ενώ θα λέγεται *εκφυλισμένος* στην αντίθετη περίπτωση.

Παρατήρηση : Το ανάπτυγμα της $\varphi_M(u)$ μέχρι 2ας τάξεως, ερμηνεύει τον Γ σαν τον πίνακα των διασπορών – συνδιασπορών του M :

$$\Gamma = E(M^t M) \Leftrightarrow \gamma_{ij} = E(X_i X_j) \quad i, j = 1, \dots, k$$

Πράγματι, $\exp\left\{-\frac{1}{2}{}^t u \Gamma u\right\} = E(e^{i'uM}) = 1 + iE({}^t u M) + \frac{i^2}{2!} E(({}^t u M)^2) + \dots$

όπου $E[({}^t u M)^2] = E({}^t u M^t u M) = E({}^t u M^t M u) = {}^t u E(M^t M) u$, αφού ${}^t u M = {}^t M u$.

Τότε,

$$\exp\left\{-\frac{1}{2}{}^t u \Gamma u\right\} = 1 + i{}^t u E(M) + \frac{i^2}{2!} {}^t u E(M^t M) u + \dots,$$

ισοδύναμα

$$1 - \frac{1}{2}{}^t u \Gamma u + \dots = 1 + i{}^t u E(M) + \frac{i^2}{2} {}^t u E(M^t M) u + \dots$$

Απ' όπου

$$E(M) = 0, {}^t u \Gamma u = {}^t u E(M^t M) u, \forall u = {}^t(u_1, \dots, u_k) \in R^k. \text{ Άρα, } \Gamma = E(M^t M) \text{ και } EM = 0.$$

6.1.2 Έστω $\mu = {}^t(\mu_1, \dots, \mu_k) \in R^k$. Θα λέμε ότι η διανυσματική τ.μ. $M = {}^t(X_1, \dots, X_k)$ ακολουθεί την *k-διάστατη Κανονική* αν η δ.τ.μ. $M - \mu$ ακολουθεί την *κεντρική κατανομή των Laplace – Gauss* :

$$M \sim N_k(\mu, \Gamma) \Leftrightarrow M - \mu \sim N_k(0, \Gamma).$$

Η χαρακτηριστική της συνάρτηση δίνεται από τη σχέση :

$$\varphi_M(u) = \exp\{i{}^t u \mu\} \exp\left\{-\frac{1}{2}{}^t u \Gamma u\right\}, \forall u = {}^t(u_1, \dots, u_k) \in R^k.$$

Έχουμε τότε : $\mu = E(M)$ και $\Gamma = E[(M - \mu)^t (M - \mu)]$.

6.2 Ιδιότητες σχετικές με ένα διαχωρισμό του M

Έστω $M = {}^t(X_1, \dots, X_k) \sim N_k(0, \Gamma), k > 1$. Χωρίζουμε τις συντεταγμένες του M σε δύο ομάδες M_1, M_2 με p και $k - p$ στοιχεία αντίστοιχα ($0 < p < k$).

Έτσι , $M = \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix}$, όπου $M_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$, $M_2 = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_k \end{pmatrix}$ και ο πίνακας διασπορών – συνδιασπορών Γ γράφεται τότε :

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{12} \\ \dots & \vdots & \dots \\ \Gamma_{21} & \vdots & \Gamma_{22} \end{pmatrix}, \text{ με } \begin{matrix} \Gamma_{11} = E(M_1 {}^t M_1) & \Gamma_{22} = E(M_2 {}^t M_2) \\ \Gamma_{12} = E(M_1 {}^t M_2) = {}^t \Gamma_{21} & \Gamma_{21} = E(M_2 {}^t M_1) \end{matrix}.$$

Θεώρημα 6.2.1

Κάθε περιθωριακή κατανομή του νόμου των L – G είναι επίσης L – G.

$$\text{π.χ. } M_1 \sim N_p(0, \Gamma_{11})$$

$$\begin{aligned} \text{Απόδειξη : } \forall u \in R^k, \varphi_{M_1}(u_1) &= E[\exp(i {}^t u_1 M_1)] = E \left[\exp \left\{ i \begin{pmatrix} {}^t u_1 & \vdots & 0 \end{pmatrix} \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix} \right\} \right] \\ &= \exp \left\{ -\frac{1}{2} \begin{pmatrix} {}^t u_1 & \vdots & 0 \end{pmatrix} \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{12} \\ \dots & \vdots & \dots \\ \Gamma_{21} & \vdots & \Gamma_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ \dots \\ 0 \end{pmatrix} \right\} \\ &= \exp \left\{ -\frac{1}{2} {}^t u_1 \Gamma_{11} u_1 \right\}. \end{aligned}$$

Θεώρημα 6.2.2

Έστω $M \sim N_k(0, \Gamma)$. Οι δ.τ.μ. M_1, M_2 είναι ανεξάρτητες αν

$$\Gamma_{21} = {}^t \Gamma_{21} = E(M_2 {}^t M_1) = 0$$

Απόδειξη : Έστω M_1, M_2 ανεξάρτητες. Τότε από το Θεώρημα 6.2.1 έχουμε :

$$\begin{aligned} \varphi_M(u) = \varphi_{M_1, M_2}(u_1, u_2) &= \varphi_{M_1}(u_1) \varphi_{M_2}(u_2) = \exp \left\{ -\frac{1}{2} {}^t u_1 \Gamma_{11} u_1 \right\} \exp \left\{ -\frac{1}{2} {}^t u_2 \Gamma_{22} u_2 \right\} \\ &= \exp \left\{ -\frac{1}{2} ({}^t u_1 \Gamma_{11} u_1 + {}^t u_2 \Gamma_{22} u_2) \right\} \end{aligned}$$

Οπότε

$$\exp \left\{ -\frac{1}{2} {}^t u \Gamma u \right\} = \exp \left\{ -\frac{1}{2} ({}^t u_1 \Gamma_{11} u_1 + {}^t u_2 \Gamma_{22} u_2) \right\}$$

$$\Leftrightarrow {}^t u \Gamma u = {}^t u_1 \Gamma_{11} u_1 + {}^t u_2 \Gamma_{22} u_2 \Leftrightarrow {}^t u_1 \Gamma_{11} u_1 + {}^t u_2 \Gamma_{22} u_2 + {}^t u_2 \Gamma_{21} u_1 + {}^t u_1 \Gamma_{12} u_2 = {}^t u_1 \Gamma_{11} u_1 + {}^t u_2 \Gamma_{22} u_2$$

$$\Leftrightarrow {}^t u_2 \Gamma_{21} u_1 + {}^t u_1 \Gamma_{12} u_2 = 0 \Leftrightarrow {}^t u_1 \Gamma_{12} u_2 + {}^t u_1 \Gamma_{12} u_2 = 0 \Leftrightarrow 2({}^t u_1 \Gamma_{12} u_2) = 0.$$

Απ'όπου $\Gamma_{12} = 0 = \Gamma_{21}$.

6.3 Ιδιότητες σχετικές με τους γραμμικούς μετασχηματισμούς

Θεώρημα 6.3.3

i) Έστω $M \sim N_k(0, \Gamma)$ (Γ αντιστρέψιμος ή όχι) και A ένας (l, k) – πίνακας (μη – τυχαίος) επί του \mathbb{R} . Τότε

$$M' = AM = {}^t (X'_1, \dots, X'_l) \sim N_l(0, \Gamma') \text{ όπου } \Gamma' = A\Gamma^t A$$

ii) Έστω $M \sim N_k(\mu, \Gamma)$ με $\mu \in \mathbb{R}^k$ και Γ ως ανωτέρω. Έστω επίσης A (l, k) γνωστός πίνακας επί του \mathbb{R} και $b \in \mathbb{R}^l$ σταθερό διάνυσμα. Τότε

$$M' = AM + b = {}^t (X'_1, \dots, X'_l) \sim N_l(A\mu + b, A\Gamma^t A).$$

Απόδειξη : i) $\varphi_{M'}(u) = E(e^{i^t u M'}) = E(e^{i^t u AM}) = E(e^{i^t (Au) M}) = \varphi_M({}^t Au)$

$$= \exp\left\{-\frac{1}{2}({}^t Au)\Gamma({}^t Au)\right\} = \exp\left\{-\frac{1}{2}u A\Gamma^t Au\right\}, \forall u \in \mathbb{R}^l.$$

ii) $\varphi_{M'}(u) = E(e^{i^t u M'}) = E(e^{i^t u (AM+b)}) = \exp\{i^t u b\} E[\exp\{i^t u AM\}]$

$$= \exp\{i^t u b\} E[\exp\{i^t ({}^t Au) M\}] = \exp\{i^t u b\} \varphi_M({}^t Au)$$

$$= \exp\{i^t u b\} \exp\{i^t ({}^t Au)\mu\} \exp\left\{-\frac{1}{2}({}^t Au)\Gamma({}^t Au)\right\}$$

$$= \exp\{i^t u (AM + b)\} \exp\left\{-\frac{1}{2}u A\Gamma^t Au\right\}$$

Πόρισμα 6.3.1

Η δ.τ.μ. $M' = {}^t (X'_1, \dots, X'_l) \sim N_k(0, \Gamma')$ σχηματίζεται από ανεξάρτητες τ.μ. αν ο Γ' είναι ένας διαγώνιος πίνακας.

Απόδειξη : Έστω $\Gamma' = \begin{pmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_l \end{pmatrix}$.

$$\text{Τότε, } \varphi_{M'}(u) = \exp\left\{-\frac{1}{2}u\Gamma'u\right\} = \exp\left\{-\frac{1}{2}\sum_{j=1}^l \gamma_j u_j^2\right\} = \prod_{j=1}^l \exp\left\{-\frac{1}{2}\gamma_j u_j^2\right\} = \prod_{j=1}^l \varphi_{X'_j}(u_j),$$

απ' όπου βάσει του κριτηρίου ανεξαρτησίας τ.μ., οι τ.μ. X'_1, \dots, X'_l είναι ανεξάρτητες.

Αντιστρόφως : Αν $X'_j, j = 1, \dots, l$ ανεξάρτητες, τότε $E(X'_i X'_j) = 0, \forall i \neq j$ · αυτό σημαίνει ότι ο Γ' καταλήγει σε ένα διαγώνιο πίνακα.

Πόρισμα 6.3.2

Υποθέτουμε ότι ο Γ είναι αντιστρέψιμος. Τότε υπάρχει ένας (k, k) -πίνακας A αντιστρέψιμος τέτοιος ώστε η δ.τ.μ. $M' = AM = (X'_1, \dots, X'_k)$ να σχηματίζεται από ανεξάρτητες τ.μ. με κατανομή $N(0, 1)$, δηλαδή $M' \sim N_k(0, I_k)$.

Για την απόδειξη αυτού του Πορίσματος θα χρειαστούμε το παρακάτω Λήμμα

Λήμμα : Έστω Γ (k, k) -πίνακας, συμμετρικός, αντιστρέψιμος και θετικός. Τότε υπάρχει (k, k) -πίνακας A , αντιστρέψιμος : $A\Gamma'A = I_k$.

Απόδειξη: Εφόσον ο Γ είναι αντιστρέψιμος, υπάρχει S (k, k) -πίνακας ορθογώνιος :

$$S^{-1}\Gamma S = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix} \text{ με } \lambda_j > 0, \forall j \in \{1, \dots, k\}.$$

Θέτουμε, $D^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_k} \end{pmatrix}$. Ο $D^{\frac{1}{2}}$ είναι διαγώνιος και αντιστρέψιμος με αντίστροφο τον

$$\left(D^{\frac{1}{2}}\right)^{-1} = D^{-\frac{1}{2}} = \begin{pmatrix} 1/\sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\lambda_k} \end{pmatrix}. \text{ Θέτοντας } A = D^{-\frac{1}{2}}S^{-1}, \text{ έχουμε ότι ο } A \text{ είναι}$$

αντιστρέψιμος και τότε

$$\Gamma' = A\Gamma'A = D^{-\frac{1}{2}}S^{-1}\Gamma'(S^{-1})D^{-\frac{1}{2}} = D^{-\frac{1}{2}}S^{-1}\Gamma S D^{-\frac{1}{2}} = D^{-\frac{1}{2}}DD^{-\frac{1}{2}} = I_k.$$

Απόδειξη Πορίσματος 6.3.2. : Εφόσον ο Γ είναι ένας (k, k) -πίνακας, αντιστρέψιμος και θετικός, σύμφωνα με το Λήμμα υπάρχει ένας (k, k) -πίνακας A αντιστρέψιμος : $A\Gamma'A = I_k$. Τότε από το Θεώρημα 6.3.3 έχουμε, $M' = AM \sim N_k(0, A\Gamma'A) = N_k(0, I_k)$.

Θεώρημα 6.3.4

Μια αναγκαία και ικανή συνθήκη ώστε η δ.τ.μ. $M = {}^t(X_1, \dots, X_k)$ να ακολουθεί την κατανομή L-G μέσα στον R^k , είναι ότι κάθε γραμμικός συνδυασμός των συντεταγμένων της να ακολουθεί την L-G στον R .

Απόδειξη : Αρκεί να υποθέσουμε ότι η δ.τ.μ. M είναι κεντρική : $E(M) = 0$. Το ότι η υπόθεση είναι αναγκαία είναι προφανές, αρκεί να πάρουμε $l = 1$ στο Θεώρημα 6.3.3.

Θα αποδείξουμε ότι αυτή είναι ικανή.

Έστω $L_u = {}^t u M$ ένας γραμμικός συνδυασμός των X_1, X_2, \dots, X_k με $u = {}^t(u_1, \dots, u_k) \in R^k$ και έστω ότι :

$$\forall u = {}^t(u_1, \dots, u_k) \in R^k, \varphi_{L_u}(t) = E\left(e^{it^t u M}\right) = e^{-\frac{1}{2} Q(u) t^2}, \text{ με } Q(u) > 0, t \in R.$$

Η Q είναι μια τετραγωνική μορφή, θετική στο u διότι :

$$Q(u) = \sigma^2({}^t u M) = E\left(\left({}^t u M\right)^2\right) = E\left(\left({}^t u M {}^t u M\right)\right) = E\left({}^t u M {}^t M u\right) = {}^t u E(M {}^t M) u.$$

Λαμβάνοντας $t = 1$, έχουμε : $\exp\left\{-\frac{1}{2} Q(u)\right\} = E\left[\left[e^{it^t u M}\right]\right] = \varphi_M(u)$, δηλαδή η M ακολουθεί την L-G μέσα στον R^k .

6.4 Μη – εκφυλισμένος νόμος των Laplace – Gauss στον R^k **Θεώρημα 6.4.5**

Έστω $M \sim N_k(0, \Gamma)$, όπου Γ αντιστρέψιμος πίνακας. Τότε η δ.τ.μ M δέχεται πυκνότητα πιθανότητας (π.π) της μορφής :

$$f(x) = (2\pi)^{-\frac{k}{2}} |\Gamma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t x \Gamma^{-1} x\right\}, \forall x = {}^t(x_1, \dots, x_k) \in R^k,$$

όπου $|\Gamma|$ σημειώνει την ορίζουσα του πίνακα Γ ($|\Gamma| = \det(\Gamma)$).

Απόδειξη : Επειδή ο πίνακας Γ είναι αντιστρέψιμος από το Πόρισμα 6.3.2, έχουμε ότι υπάρχει ένας (k, k) – πίνακας A αντιστρέψιμος : $A \Gamma^{-1} A = I_k$ και $M' = A M \sim N_k(0, I_k)$, απ' όπου, $|\Gamma| = |A|^{-2}$. Έτσι η δ.τ.μ. M' δέχεται π.π. την :

$$\tilde{f}(x') = \prod_{i=1}^k (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (x'_i)^2\right\} = (2\pi)^{-\frac{k}{2}} \exp\left\{-\frac{1}{2} {}^t x' x'\right\}, \text{ όπου } x' = {}^t(x'_1, \dots, x'_k) \in R^k.$$

Επειδή ο A είναι αντιστρέψιμος, το M δέχεται εξίσου μια π.π $f(x)$ λαμβανόμενη από την ανωτέρω π.π. θεωρώντας το μετασχηματισμό : $x' = Ax$.

Επομένως,

$$\begin{aligned} f(x) &= \tilde{f}(Ax) |\det A| = (2\pi)^{-\frac{k}{2}} |\Gamma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t x' A A x\right\} \\ &= (2\pi)^{-\frac{k}{2}} |\Gamma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t x \Gamma^{-1} x\right\}. \end{aligned}$$

Παρατήρηση: Αν $M \sim N_k(\mu, \Gamma)$ και Γ είναι αντιστρέψιμος, τότε η M δέχεται π.π. την

$$f(x) = (2\pi)^{-\frac{k}{2}} |\Gamma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t (x - \mu) \Gamma^{-1} (x - \mu)\right\}, \forall x = {}^t (x_1, \dots, x_k) \in R^k.$$

6.5 Δεσμευμένες Ροπές

Στις ιδιότητες που αναφέραμε παραπάνω ο νόμος των L-G δεν ήταν αναγκαίως μη – εκφυλισμένος (δηλαδή ο Γ αντιστρέψιμος). Όμως για τη μελέτη των δεσμευμένων κατανομών τούτο είναι αναγκαίο και θα το δεχθούμε σ' αυτή την παράγραφο. Θεωρούμε τις έννοιες της παραγράφου 6.2.

Έστω $M \sim N_k(0, \Gamma)$ και Γ αντιστρέψιμος. Χωρίζουμε τις συντεταγμένες του M σε δύο ομάδες M_1, M_2 με p και $k - p$ στοιχεία αντίστοιχα ($0 < p < k$).

Έτσι, $M = \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix}$, όπου $M_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}, M_2 = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_k \end{pmatrix}$ και ο πίνακας διασπορών –

συνδιασπορών Γ γράφεται τότε :

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{12} \\ \dots & \vdots & \dots \\ \Gamma_{21} & \vdots & \Gamma_{22} \end{pmatrix}, \text{ με } \begin{cases} \Gamma_{11} = E(M_1 {}^t M_1) & \Gamma_{22} = E(M_2 {}^t M_2) \\ \Gamma_{12} = E(M_1 {}^t M_2) = {}^t \Gamma_{21} & \Gamma_{21} = E(M_2 {}^t M_1) \end{cases}.$$

i) Θεωρούμε τη δ.τ.μ. M_1 δοθέντος ότι $M_2 = x_2 (\in R^{k-p})$. Αυτό θα το σημειώνουμε με $M_1 | x_2$ και αναζητούμε το νόμο αυτής της τ.μ.

1° Θεωρούμε την αλλαγή μεταβλητής

$$\left. \begin{matrix} M'_1 = M_1 - A M_2 \\ M'_2 = M_2 \end{matrix} \right\} \Leftrightarrow \begin{pmatrix} M'_1 \\ \dots \\ M'_2 \end{pmatrix} = \begin{pmatrix} I_p & \vdots & -A \\ \dots & \vdots & \dots \\ 0 & \vdots & I_{k-p} \end{pmatrix} \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix},$$

όπου ο A είναι ένας $(p, k-p)$ -πίνακας τον οποίο ζητούμε να καθορίσουμε έτσι ώστε το ζεύγος (M'_1, M'_2) να είναι μη - συσχετισμένο (δηλαδή M'_1, M'_2 ανεξάρτητες δ.τ.μ. επειδή το εν λόγω ζεύγος είναι των L-G).

Γράφοντας ότι $E(M_1' M_2') = 0$, έχουμε αμέσως ότι $A = \Gamma_{12} \Gamma_{22}^{-1}$.

2° Οι δ.τ.μ. $\begin{cases} M'_1 = M_1 - \Gamma_{11} \Gamma_{22}^{-1} M_2 \\ M'_2 = M_2 \end{cases}$ είναι ανεξάρτητες και η M'_1 είναι μια p -διάστατη τ.μ. των L-G, κεντρική με πίνακα διασπορών - συνδιασπορών τον

$$\Gamma_{11 \bullet 2} = E(M_1' M_1') = \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}$$

3° Από το 2° προκύπτει ότι η δ.τ.μ. M_1 είναι δυνατό να εκφρασθεί σαν το άθροισμα δύο ανεξάρτητων όρων ως εξής :

$$M_1 = \Gamma_{12} \Gamma_{22}^{-1} M_2 + M'_1, \quad (M_2 = M'_2).$$

Δεσμευτικά, όταν $M_2 = x_2$, αυτή η έκφραση γίνεται :

$$M_1 | x_2 = \Gamma_{12} \Gamma_{22}^{-1} x_2 + M'_1.$$

ii) Θεωρούμε τη M_2 δ.τ.μ. δοθέντος $M_1 = x_1 (\in R^p)$ και αναζητούμε όπως στο i) το νόμο της $M_2 | x_1$. Θεωρώντας το μετασχηματισμό

$$\left. \begin{matrix} M'_1 = M_1 \\ M'_2 = M_2 - B M_1 \end{matrix} \right\} \Leftrightarrow \begin{pmatrix} M'_1 \\ \dots \\ M'_2 \end{pmatrix} = \begin{pmatrix} I_p & \vdots & 0 \\ \dots & \vdots & \dots \\ -B & \vdots & I_{k-p} \end{pmatrix} \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix},$$

όπου ο πίνακας B τύπου $(k-p, p)$ επιλέγεται έτσι ώστε οι M'_1, M'_2 να είναι μη συσχετισμένες δ.τ.μ. (δηλαδή $E(M_2' M_1') = 0$) παίρνοντας, όπως παραπάνω :

$$B = \Gamma_{22} \Gamma_{11}^{-1} \text{ και } M_2 | x_1 = B x_1 + M'_2 = \Gamma_{22} \Gamma_{11}^{-1} x_1 + M'_2.$$

Θεώρημα 6.5.6

$$\text{Αν } M = \begin{pmatrix} M_1 \\ \dots \\ M_2 \end{pmatrix} \sim N_k(0, \Gamma), \text{ όπου } M_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}, M_2 = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_k \end{pmatrix} \text{ και } \Gamma = \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{12} \\ \dots & \vdots & \dots \\ \Gamma_{21} & \vdots & \Gamma_{22} \end{pmatrix}$$

αντιστρέψιμος, τότε :

$$\text{i) } M_1 | x_2 \sim N_p(\mu'_1, \Gamma_{11 \bullet 2}) \text{ όπου } \begin{cases} \mu'_1 = \Gamma_{12} \Gamma_{22}^{-1} x_2 \\ \Gamma_{11 \bullet 2} = \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \end{cases}.$$

$$\text{ii) } M_2 | x_1 \sim N_{k-p}(\mu'_2, \Gamma_{22 \bullet 1}) \text{ όπου } \begin{cases} \mu'_2 = \Gamma_{21} \Gamma_{11}^{-1} x_1 \\ \Gamma_{22 \bullet 1} = \Gamma_{22} - \Gamma_{21} \Gamma_{11}^{-1} \Gamma_{12} \end{cases}.$$

Ιδιαίτερως, αν $M \sim N_k(\mu, \Gamma)$, όπου $\mu = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix}$, $\mu_1 \in R^p$, $\mu_2 \in R^{k-p}$, τότε

$$i') M_1 | x_2 \sim N_p(\mu_{1 \bullet 2}, \Gamma_{11 \bullet 2}), \text{ όπου } \mu_{1 \bullet 2} = \mu_1 + \Gamma_{12} \Gamma_{22}^{-1} (x_2 - \mu_2)$$

$$ii') M_2 | x_1 \sim N_{k-p}(\mu_{2 \bullet 1}, \Gamma_{22 \bullet 1}), \text{ όπου } \mu_{2 \bullet 1} = \mu_2 + \Gamma_{21} \Gamma_{11}^{-1} (x_1 - \mu_1).$$

Θεώρημα 6.5.7 (Ανασύνθεση της από κοινού Κανονικής)

i) Έστω $M = \begin{pmatrix} M_1 \\ \cdots \\ M_2 \end{pmatrix}$. Αν, $M_1 | x_2 \sim N_p(Ax_2' + b, \Omega)$ και $M_2 \sim N_{k-p}(\mu_2, \Gamma_{22})$, με A, b, Ω

γνωστούς πίνακες τύπων $(p, k-p), (p, 1)$ και (p, p) αντιστοίχως, τότε :

$$M = \begin{pmatrix} M_1 \\ \cdots \\ M_2 \end{pmatrix} \sim N_k \left(\begin{pmatrix} A\mu_2 + b \\ \cdots \\ \mu_2 \end{pmatrix}, \begin{pmatrix} A\Gamma_{22}'A + \Omega & \vdots & A\Gamma_{22} \\ \cdots & \vdots & \cdots \\ \Gamma_{22}'A & \vdots & \Gamma_{22} \end{pmatrix} \right)$$

ii) Αν $M_1 \sim N_p(\mu_1, \Gamma_{11})$ και $M_2 | x_1 \sim N_{k-p}(A'x_1 + b', \Omega')$, όπου A', b', Ω' γνωστοί πίνακες τύπων $(k-p, p), (k-p, 1)$ και $(k-p, k-p)$ αντιστοίχως, τότε

$$M = \begin{pmatrix} M_1 \\ \cdots \\ M_2 \end{pmatrix} \sim N_k \left(\begin{pmatrix} \mu_1 \\ \cdots \\ A'\mu_1 + b' \end{pmatrix}, \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{11}'A' \\ \cdots & \vdots & \cdots \\ A'\Gamma_{11} & \vdots & A'\Gamma_{11}'A' + \Omega' \end{pmatrix} \right).$$

Απόδειξη i) Έστω $x_1 \in R^p$ και $x_2 \in R^{k-p}$. Τότε

$$\begin{aligned} f_{M_1 M_2}(x_1, x_2) &= f_{M_1 | x_2}(x_1 | x_2) f_{M_2}(x_2) \\ &= (2\pi)^{-\frac{p}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_1 - (Ax_2 + b)) \Omega^{-1} (x_1 - (Ax_2 + b)) \right\} \\ &\quad \times (2\pi)^{-\frac{k-p}{2}} |\Gamma_{22}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_2 - \mu_2) \Gamma_{22}^{-1} (x_2 - \mu_2) \right\} \\ &= (2\pi)^{-\frac{k}{2}} |\Omega|^{-\frac{1}{2}} |\Gamma_{22}|^{-\frac{1}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} (x_1 - (Ax_2 + b)) \Omega^{-1} (x_1 - (Ax_2 + b)) + (x_2 - \mu_2) \Gamma_{22}^{-1} (x_2 - \mu_2) \right\} \end{aligned}$$

όπου

$$\begin{aligned}
& {}^t(x_1 - (Ax_2 + b))\Omega^{-1}(x_1 - (Ax_2 + b)) + {}^t(x_2 - \mu_2)\Gamma_{22}^{-1}(x_2 - \mu_2) \\
&= {}^t(x_1 - (A(x_2 - \mu_2) + A\mu_2 + b))\Omega^{-1}(x_1 - (A(x_2 - \mu_2) + A\mu_2 + b)) + {}^t(x_2 - \mu_2)\Gamma_{22}^{-1}(x_2 - \mu_2) \\
&= {}^t((x_1 - (A\mu_2 + b)) - A(x_2 - \mu_2))\Omega^{-1}((x_1 - (A\mu_2 + b)) - A(x_2 - \mu_2)) + {}^t(x_2 - \mu_2)\Gamma_{22}^{-1}(x_2 - \mu_2) \\
&= {}^t(x_1 - (A\mu_2 + b))\Omega^{-1}(x_1 - (A\mu_2 + b)) - {}^t(x_1 - (A\mu_2 + b))\Omega^{-1}A(x_2 - \mu_2) \\
&\quad - {}^t(x_2 - \mu_2)A\Omega^{-1}(x_1 - (A\mu_2 + b)) + {}^t(x_2 - \mu_2)(A\Omega^{-1}A + \Gamma_{22}^{-1})(x_2 - \mu_2) \\
&= ({}^t(x_1 - (A\mu_2 + b)) : {}^t(x_2 - \mu_2)) \begin{pmatrix} \Omega^{-1} & \vdots & -\Omega^{-1}A \\ \dots & \vdots & \dots \\ {}^tA\Omega^{-1} & \vdots & {}^tA\Omega^{-1}A + \Gamma_{22}^{-1} \end{pmatrix} \begin{pmatrix} x_1 - (A\mu_2 + b) \\ \dots \\ x_2 - \mu_2 \end{pmatrix}.
\end{aligned}$$

Αλλά (βλέπε Rao, σελ.32,33):

$$\begin{pmatrix} \Omega^{-1} & \vdots & -\Omega^{-1}A \\ \dots & \vdots & \dots \\ {}^tA\Omega^{-1} & \vdots & {}^tA\Omega^{-1}A + \Gamma_{22}^{-1} \end{pmatrix} = \begin{pmatrix} A\Gamma_{22} {}^tA + \Omega & \vdots & A\Gamma_{22} \\ \dots & \vdots & \dots \\ \Gamma_{22} {}^tA & \vdots & \Gamma_{22} \end{pmatrix}^{-1}$$

και

$$\begin{vmatrix} A\Gamma_{22} {}^tA + \Omega & \vdots & A\Gamma_{22} \\ \dots & \vdots & \dots \\ \Gamma_{22} {}^tA & \vdots & \Gamma_{22} \end{vmatrix} = \begin{vmatrix} I_p & \vdots & -A \\ \dots & \vdots & \dots \\ 0 & \vdots & I_{k-p} \end{vmatrix} \begin{vmatrix} A\Gamma_{22} {}^tA + \Omega & \vdots & A\Gamma_{22} \\ \dots & \vdots & \dots \\ \Gamma_{22} {}^tA & \vdots & \Gamma_{22} \end{vmatrix} = \begin{vmatrix} \Omega & \vdots & 0 \\ \dots & \vdots & \dots \\ \Gamma_{22} {}^tA & \vdots & \Gamma_{22} \end{vmatrix} = |\Omega| |\Gamma_{22}|$$

λαμβάνοντας υπόψη τις ιδιότητες των οριζουσών.

ii) Εργαζόμενοι κατά τον ίδιο τρόπο, παίρνουμε :

$$\begin{aligned}
f_{M_1, M_2}(x_1, x_2) &= f_{M_2|x_1}(x_2 | x_1) f_{M_1}(x_1) \\
&= (2\pi)^{-\frac{k+p}{2}} |\Omega'|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t(x_2 - (A'x_1 + b'))\Omega'^{-1}(x_2 - (A'x_1 + b'))\right\} \\
&\quad \times (2\pi)^{-\frac{p}{2}} |\Gamma_{11}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} {}^t(x_1 - \mu_1)\Gamma_{11}^{-1}(x_1 - \mu_1)\right\} \\
&= (2\pi)^{-\frac{k}{2}} |\Omega'|^{-\frac{1}{2}} |\Gamma_{11}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2} ({}^t(x_2 - (A'x_1 + b'))\Omega'^{-1}(x_2 - (A'x_1 + b')) + {}^t(x_1 - \mu_1)\Gamma_{11}^{-1}(x_1 - \mu_1))\right\}
\end{aligned}$$

όπου

$$\begin{aligned}
& {}^t(x_2 - (A'x_1 + b'))\Omega'^{-1}(x_2 - (A'x_1 + b')) + {}^t(x_1 - \mu_1)\Gamma_{11}^{-1}(x_1 - \mu_1) \\
&= {}^t(x_2 - (A'(x_1 - \mu_1) + A'\mu_1 + b'))\Omega'^{-1}(x_2 - (A'(x_1 - \mu_1) + A'\mu_1 + b')) + {}^t(x_1 - \mu_1)\Gamma_{11}^{-1}(x_1 - \mu_1) \\
&= {}^t((x_2 - (A'\mu_1 + b')) - A'(x_1 - \mu_1))\Omega'^{-1}((x_2 - (A'\mu_1 + b')) - A'(x_1 - \mu_1)) + {}^t(x_1 - \mu_1)\Gamma_{11}^{-1}(x_1 - \mu_1) \\
&= {}^t(x_2 - (A'\mu_1 + b'))\Omega'^{-1}(x_2 - (A'\mu_1 + b')) - {}^t(x_2 - (A'\mu_1 + b'))\Omega'^{-1}A'(x_1 - \mu_1)
\end{aligned}$$

$$\begin{aligned}
& -{}^t(x_1 - \mu_1)' A' \Omega'^{-1} (x_2 - (A' \mu_1 + b')) + {}^t(x_1 - \mu_1) ({}^t A' \Omega'^{-1} A' + \Gamma_{11}^{-1}) (x_1 - \mu_1) \\
& = ({}^t(x_2 - \mu_2) : {}^t(x_2 - (A' \mu_1 + b'))) \begin{pmatrix} {}^t A' \Omega'^{-1} A' + \Gamma_{11}^{-1} & \vdots & -{}^t A' \Omega'^{-1} \\ \dots & \vdots & \dots \\ \Omega'^{-1} A' & \vdots & \Omega'^{-1} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ \dots \\ x_2 - (A' \mu_1 + b') \end{pmatrix}.
\end{aligned}$$

Παρατηρούμε όμως ότι :

$$\begin{pmatrix} {}^t A' \Omega'^{-1} A' + \Gamma_{11}^{-1} & \vdots & -{}^t A' \Omega'^{-1} \\ \dots & \vdots & \dots \\ \Omega'^{-1} A' & \vdots & \Omega'^{-1} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{11} {}^t A' \\ \dots & \vdots & \dots \\ A' \Gamma_{11} & \vdots & A' \Gamma_{11} {}^t A' + \Omega' \end{pmatrix}^{-1}.$$

(Από τον Rao, σελ.32,33, έχουμε ότι :

$$\begin{pmatrix} \Gamma_{11} & \vdots & \Gamma_{11} {}^t A' \\ \dots & \vdots & \dots \\ A' \Gamma_{11} & \vdots & A' \Gamma_{11} {}^t A' + \Omega' \end{pmatrix}^{-1} = \begin{pmatrix} \Gamma_{11}^{-1} + F E^{-1} {}^t F & \vdots & -F E^{-1} \\ \dots & \vdots & \dots \\ -E^{-1} {}^t F & \vdots & E^{-1} \end{pmatrix},$$

όπου, $E = A' \Gamma_{11} {}^t A' + \Omega' - A' \Gamma_{11} \Gamma_{11}^{-1} \Gamma_{11} {}^t A' = \Omega'$ και $F = \Gamma_{11}^{-1} \Gamma_{11} {}^t A' = {}^t A'$.

Συνεπώς ,

$$\begin{pmatrix} \Gamma_{11}^{-1} + F E^{-1} {}^t F & \vdots & -F E^{-1} \\ \dots & \vdots & \dots \\ -E^{-1} {}^t F & \vdots & E^{-1} \end{pmatrix} = \begin{pmatrix} {}^t A' \Omega'^{-1} A' + \Gamma_{11}^{-1} & \vdots & -{}^t A' \Omega'^{-1} \\ \dots & \vdots & \dots \\ -\Omega'^{-1} A' & \vdots & \Omega'^{-1} \end{pmatrix}$$

και

$$\begin{vmatrix} \Gamma_{11} & \vdots & \Gamma_{11} {}^t A' \\ \dots & \dots & \dots \\ A' \Gamma_{11} & \vdots & A' \Gamma_{11} {}^t A' + \Omega' \end{vmatrix} = \begin{vmatrix} I_p & \vdots & \mathbf{0} \\ \dots & \vdots & \dots \\ -A' & \vdots & I_{k-p} \end{vmatrix} \begin{vmatrix} \Gamma_{11} & \vdots & \Gamma_{11} {}^t A' \\ \dots & \vdots & \dots \\ A' \Gamma_{11} & \vdots & A' \Gamma_{11} {}^t A' + \Omega' \end{vmatrix} = \begin{vmatrix} \Gamma_{11} & \vdots & \Gamma_{11} {}^t A' \\ \dots & \dots & \dots \\ \mathbf{0} & \vdots & \Omega' \end{vmatrix} = |\Gamma_{11}| |\Omega'|$$

από τις ιδιότητες των οριζουσών. Απ' όπου το αποτέλεσμα.

7. Κανονική – Γάμμα Πολυδιάστατη κατανομή

Έστω $X = (X_1, \dots, X_p)$ ($p \geq 1$) διανυσματική τ.μ. και Y μια τ.μ. θετική. Η Κανονική – Γάμμα κατανομή προκύπτει συνδυάζοντας την περιθωριακή της Y (μια κατάλληλη Γάμμα κατανομή) και την δεσμευμένη της $X | Y = y$ (μια p - διάστατη Κανονική κατανομή)

Λοιπόν, αν έχουμε :

$$X | y \sim N_p(\mu, y^{-1} B), \mu \in R^p, B : (p, p) \text{ πίνακας συμμετρικός και θετικά ορισμένος}$$

και

$$Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{u\nu}\right), \quad u, \nu > 0$$

τότε λέμε ότι

$$(X, Y) \sim N_p\gamma(\mu, B, u, \nu)$$

με πυκνότητα πιθανότητας

$$\begin{aligned} f(x, y) &= f_{N\gamma}(x, y | \mu, B, u, \nu) = f(x | y)f(y) = f_N(x | \mu, y^{-1}B)f_Y(y | \frac{\nu}{2}, \frac{2}{u\nu}) \\ &= (2\pi)^{-\frac{p}{2}} y^{\frac{p}{2}} |B|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}y^t(x-\mu)B^{-1}(x-\mu)\right\} \times \frac{\left(\frac{u\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} y^{\frac{\nu}{2}-1} e^{-y\frac{u\nu}{2}} 1_{R^p}(x) 1_{R^+}(y) \\ &= (2\pi)^{-\frac{p}{2}} |B|^{-\frac{1}{2}} \frac{\left(\frac{u\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} y^{\frac{\nu+p}{2}-1} \exp\left\{-\frac{y^t}{2}(x-\mu)B^{-1}(x-\mu) - \frac{1}{2}y u \nu\right\} 1_{R^p \times R^+}(y) \end{aligned}$$

Έχουμε τις παρακάτω ροπές :

$$\begin{aligned} E(X | y) &= \mu \\ \text{Var}(X | y) &= y^{-1}B \\ E(Y) &= \frac{1}{u} \\ \text{Var}(Y) &= \frac{2}{\nu u^2}. \end{aligned}$$

Παρατηρούμε ότι η B^{-1} δύναται να ερμηνευθεί σαν μια σχετική ακρίβεια της X σε μονάδες y .

Με δεδομένο ότι $(X, Y) \sim N_p\gamma(\mu, B, u, \nu)$ μπορούμε να υπολογίσουμε τις περιθωριακές ροπές του διανύσματος X :

$$E(X) = E(E(X | Y)) = \mu$$

$$\text{Var}(X) = E(\text{Var}(X | Y)) + \text{Var}(E(X | Y)) = E(Y^{-1}B) = E(Y^{-1})B = \frac{\nu}{\nu-2}uB, \quad \text{όταν } \nu > 2$$

η τελευταία ισότητα προκύπτει από το γεγονός ότι η $Y^{-1} \sim \Gamma^{-1}\left(\frac{\nu}{2}, \frac{2}{u\nu}\right)$, ενώ ο χρησιμοποιούμενος τύπος της διασποράς αποδεικνύεται ως εξής :

$$\begin{aligned}
\text{Var}(X) &= E[(X - EX)^2] = E[E[(X - EX)^2 | Y]] \\
&= E[E[(X - E(X | Y) + E(X | Y) - EX)^2 | Y]] \\
&= E[E[(X - E(X | Y))^2 | Y]] + E[E[(E(X | Y) - EX)^2 | Y]] \\
&\quad + 2E[E[(X - E(X | Y))(E(X | Y) - EX) | Y]] \\
&= E[\text{Var}(X | Y)] + E[(E(X | Y) - EX)^2] \\
&\quad + 2E[(E(X | Y) - EX)E[(X - E(X | Y)) | Y]] \\
&= E[\text{Var}(X | Y)] + \text{Var}(E(X | Y)) + \underbrace{2E[(E(X | Y) - EX)(E(X | Y) - E(X | Y))]}_0 \\
&= E[\text{Var}(X | Y)] + \text{Var}(E(X | Y))
\end{aligned}$$

Να σημειώσουμε ότι η πολυδιάστατη Κανονική – Γάμμα κατανομή ορίζεται γενικά όταν $rg(B) = r (< p)$ Δέχεται την π.π. της παραπάνω μορφής μόνο όταν $rg(B) = p$, δηλαδή όταν ο B είναι θετικά ορισμένος.

Παρατήρηση : Μερικές φορές στην bayssienne ανάλυση είναι χρήσιμες οι οριακές περιπτώσεις όπου ο νόμος της (X, Y) τείνει προς έναν impropre (μη – ολοκληρώσιμο) νόμο :

$$f(x, y) \propto \frac{1}{y}.$$

Προκύπτει έτσι για το $(X, \log Y)$ ένας ομοιόμορφος νόμος επί του \mathbb{R}^{p+1} , με όλες τις συντεταγμένες ανεξάρτητες. Ο παραπάνω νόμος αντιστοιχεί στην πράξη στην περίπτωση όπου $B^{-1}, u, v \rightarrow 0$, σ' αυτήν την περίπτωση πρέπει προφανώς και το $r \rightarrow 0$. Εκτός της οριακής αυτής περίπτωσης έχουμε γενικά $r = p$ και οι πυκνότητες υπάρχουν και είναι propres (το ολοκλήρωμά τους είναι 1). Ο περιθωριακός νόμος της X είναι ένας νόμος *Student- πολυδιάστατος*. Συγκεκριμένα θα δείξουμε ότι :

$$(X, Y) \sim N_p \gamma(\mu, B, u, v) \Rightarrow X \sim t_p(\mu, uB, v)$$

Πράγματι,

$$\begin{aligned}
f(x) &= \int_R f(x, y) dy \\
&= (2\pi)^{-\frac{p}{2}} |B|^{-\frac{1}{2}} \left(\frac{uv}{2}\right)^{\frac{v}{2}} \frac{1}{\Gamma\left(\frac{v}{2}\right)} 1_{R^p}(x) \int_0^{+\infty} y^{\frac{v+p}{2}-1} \exp\left\{-\frac{y}{2}[(x - \mu)B^{-1}(x - \mu) + uv]\right\} dy
\end{aligned}$$

$$\begin{aligned}
&= (2\pi)^{-\frac{p}{2}} |B|^{-\frac{1}{2}} \frac{(uv)^{\frac{v}{2}}}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} \Gamma\left(\frac{v+p}{2}\right) 2^{\frac{v+p}{2}} \left[{}^t(x-\mu)B^{-1}(x-\mu)+uv\right]^{\frac{v+p}{2}} 1_{R^p}(x) \\
&= \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \pi^{\frac{p}{2}}} |B|^{-\frac{1}{2}} (uv)^{\frac{v}{2}} u^{-\frac{v+p}{2}} \left[{}^t(x-\mu)u^{-1}B^{-1}(x-\mu)+v\right]^{\frac{v+p}{2}} 1_{R^p}(x) \\
&= \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \pi^{\frac{p}{2}}} |B|^{-\frac{1}{2}} v^{\frac{v}{2}} u^{-\frac{p}{2}} \left[{}^t(x-\mu)u^{-1}B^{-1}(x-\mu)+v\right]^{\frac{v+p}{2}} 1_{R^p}(x)
\end{aligned}$$

όπου η τελευταία παράσταση είναι η π.π. μιας p -διάστατης *Student* με παραμέτρους μ, uB και το v , δηλαδή $X \sim t_p(\mu, uB, v)$ (βλέπε κατωτέρω).

Όταν $p=1$, για το ζεύγος (X, Y) λαμβάνουμε την *Κανονική – Γάμμα* και τότε $X \sim t_1(\mu, uB, v)$.

8. Πολυδιάστατη Student κατανομή

8.1 Ορισμός

Η p -διάστατη *Student* κατανομή ορίζεται σαν ένα μίγμα του p -διάστατου κανονικού νόμου και ενός κατάλληλου Γάμμα νόμου.

Συγκεκριμένα αν, δεδομένου ότι $Y = y$,

$$X | y \sim N_p(\mu, y^{-1}V) \text{ και } Y \sim \Gamma\left(\frac{v}{2}, \frac{2}{v}\right) \text{ με } v > 0,$$

τότε ως γνωστό από την παράγραφο 7, $(X, Y) \sim N_p \gamma(\mu, V, u, v)$ με $u=1$.

Συνεπώς η τ.μ. X , δέχεται π.π την

$$\begin{aligned}
f(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{+\infty} f(x | y) f(y) dy = \int_{-\infty}^{+\infty} f_N(x | \mu, y^{-1}V) f_\gamma(y | \frac{v}{2}, \frac{2}{v}) dy \\
&= \int_{-\infty}^{+\infty} f_{N\gamma}(x, y | \mu, V, u=1, v) dy
\end{aligned}$$

$$= \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{\frac{p}{2}}} |V|^{-\frac{1}{2}} \nu^{\frac{\nu}{2}} \left[{}^t(x-\mu)V^{-1}(x-\mu) + \nu\right]^{-\frac{\nu+p}{2}} 1_{R^p}(x)$$

Το γεγονός ότι η τ.μ. X ακολουθεί μια τέτοια κατανομή το σημειώνουμε με

$$X \sim t_p(\mu, V, \nu).$$

Αποδεικνύεται, όπως στην παράγραφο 7, ότι οι ροπές της X είναι :

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \frac{\nu}{\nu-2} V, \quad \nu > 2 \end{aligned}$$

Σημειώνουμε ότι η μερική περίπτωση $p=1$, αντιστοιχεί στη μονοδιάστατη *Student* με π.π.

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{\frac{1}{2}}} V^{-\frac{1}{2}} \nu^{\frac{\nu}{2}} \left[\frac{(x-\mu)^2}{V} + \nu\right]^{-\frac{\nu+1}{2}} 1_R(x).$$

Λαμβάνοντας υπόψη ότι : $\Gamma(1/2) = \sqrt{\pi}$, η παραπάνω πυκνότητα γράφεται :

$$f(x) = \frac{1}{\mathbf{B}\left(\frac{\nu}{2}, \frac{1}{2}\right)} V^{-\frac{1}{2}} \nu^{\frac{\nu}{2}} \left[\frac{(x-\mu)^2}{V} + \nu\right]^{-\frac{\nu+1}{2}} 1_R(x),$$

δηλαδή $X \sim t_1(\mu, V, \nu)$ όπου μ, V βαθμωτά.

Τέλος, αν $\mu=0$, $V=1$ και $X \sim t_1(\mu, V, \nu)$ λέμε ότι η X ακολουθεί την κεντρική t -κατανομή με ν -βαθμούς ελευθερίας και σημειώνουμε $X \sim t_\nu (= t_1(0,1,\nu))$. Σε αυτή την περίπτωση η π.π. της X γίνεται :

$$f(x) = \frac{1}{\mathbf{B}\left(\frac{\nu}{2}, \frac{1}{2}\right)} \nu^{\frac{\nu}{2}} (x^2 + \nu)^{-\frac{\nu+1}{2}} 1_R(x) = \frac{1}{\mathbf{B}\left(\frac{\nu}{2}, \frac{1}{2}\right)} \nu^{\frac{1}{2}} \left(\frac{x^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}} 1_R(x).$$

8.2 Περιθωριακές και Δεσμευμένες

Από τις ιδιότητες της p -διάστατης Κανονικής, βρίσκουμε αμέσως τις ιδιότητες των περιθωριακών και δεσμευμένων μιας p -διάστατης Student. Αν $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ p -διάστατη τ.μ. με X_i, p_i -διάστατη τ.μ., $p_1 + p_2 = p$, $\mu = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_2 \end{pmatrix}$ και $V = \begin{pmatrix} V_{11} & \dots & V_{12} \\ \dots & \dots & \dots \\ V_{21} & \dots & V_{22} \end{pmatrix}$

τότε, αν $X \sim t_p(\mu, V, \nu)$, λαμβάνουμε :

$$X_1 \sim t_{p_1}(\mu_1, V_{11}, \nu)$$

και δοθέντος ότι $X_1 = x_1$,

$$X_2 | x_1 \sim t_{p_2}(\mu_{2\cdot 1}, \gamma \cdot V_{22\cdot 1}, \nu + p_1)$$

όπου

$$\begin{aligned} \mu_{2\cdot 1} &= \mu_2 + V_{21}V_{11}^{-1}(x_1 - \mu_1) \\ V_{22\cdot 1} &= V_{22} - V_{21}V_{11}^{-1}V_{12} \\ \gamma &= \frac{{}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1)}{\nu + p_1}. \end{aligned}$$

Επειδή $X \sim t_p(\mu, V, \nu)$, από τον ορισμό της p -διάστατης Student, αυτή προκύπτει όταν :

$$X | Y = y \sim N_p(\mu, y^{-1}V), \quad Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

ή

$$\begin{pmatrix} X_1 \\ \dots \\ X_2 \end{pmatrix} | Y = y \sim N_p\left(\begin{pmatrix} \mu_1 \\ \dots \\ \mu_2 \end{pmatrix}, y^{-1}\begin{pmatrix} V_{11} & \dots & V_{12} \\ \dots & \dots & \dots \\ V_{21} & \dots & V_{22} \end{pmatrix}\right), \quad Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

- Από το Θεώρημα 6.2.1 λαμβάνουμε

$$X_1 | Y = y \sim N_{p_1}(\mu_1, y^{-1}V_{11})$$

Λοιπόν, δεδομένου ότι :

$$X_1 | Y = y \sim N_{p_1}(\mu_1, y^{-1}V_{11}) \quad \text{και} \quad Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

από τον ορισμό της πολυδιάστατης Student

$$X_1 \sim t_{p_1}(\mu_1, V_{11}, \nu)$$

- Για τον καθορισμό της δεσμευμένης κατανομής της τ.μ. $X_2 | X_1 = x_1$ εργαζόμαστε ως εξής :
Από το Θεώρημα 6.5.6 ii') λαμβάνουμε :

$$X_2 | X_1 = x_1, Y = y \sim N_{p_2}(\mu_{2\bullet 1}, y^{-1}V_{22\bullet 1})$$

$$\text{όπου } \mu_{2\bullet 1} = \mu_2 + V_{21}V_{11}^{-1}(x_1 - \mu_1) \text{ και } V_{22\bullet 1} = V_{22} - V_{21}V_{11}^{-1}V_{12}.$$

Με δεδομένο λοιπόν ότι :

$$X_2 | X_1 = x_1, Y = y \sim N_{p_2}(\mu_{2\bullet 1}, y^{-1}V_{22\bullet 1}) \text{ και } Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

έχουμε :

$$f(x_2 | x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{1}{f(x_1)} \int_R f(x_1, x_2, y) dy = \frac{1}{f(x_1)} \int_R f(x_2 | x_1, y) f(x_1, y) dy$$

Λοιπόν,

$$\begin{aligned} \int_R f(x_2 | x_1, y) f(x_1, y) dy &= \int_R f(x_2 | x_1, y) f(x_1 | y) f(y) dy \\ &= \int_R f_{N_{p_2}}(x_2 | \mu_{2\bullet 1}, y^{-1}V_{22\bullet 1}) f_{N_{p_1}}(x_1 | \mu_1, y^{-1}V_{11}) f_Y(y | \frac{\nu}{2}, \frac{2}{\nu}) dy \\ &= \int_R (2\pi)^{-\frac{p_2}{2}} y^{\frac{p_2}{2}} |V_{22\bullet 1}|^{-\frac{1}{2}} \exp\left\{-\frac{y^{-1}}{2}(x_2 - \mu_{2\bullet 1})V_{22\bullet 1}^{-1}(x_2 - \mu_{2\bullet 1})\right\} \\ &\quad \times (2\pi)^{-\frac{p_1}{2}} y^{\frac{p_1}{2}} |V_{11}|^{-\frac{1}{2}} \exp\left\{-\frac{y^{-1}}{2}(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1)\right\} \\ &\quad \times \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} y^{\frac{\nu}{2}-1} e^{-\frac{y\nu}{2}} 1_{R^{p_1}}(x_1) 1_{R^{p_2}}(x_2) 1_{R^+}(y) dy \\ &= (2\pi)^{-\frac{p}{2}} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} |V_{22\bullet 1}|^{-\frac{1}{2}} |V_{11}|^{-\frac{1}{2}} \int_0^{+\infty} y^{\frac{p+\nu}{2}-1} \exp\left\{-\frac{y}{2}[\nu(x_2 - \mu_{2\bullet 1})V_{22\bullet 1}^{-1}(x_2 - \mu_{2\bullet 1})\right. \\ &\quad \left.+ (x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + \nu]\right\} dy \end{aligned}$$

όπου το τελευταίο ολοκλήρωμα είναι εκείνο μιας Γάμμα κατανομής :

$$\Gamma\left(\frac{p+v}{2}, 2\left[{}^t(x_2 - \mu_{2\bullet})V_{22\bullet}^{-1}(x_2 - \mu_{2\bullet}) + {}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + v\right]^{-1}\right)$$

Επομένως,

$$\begin{aligned} \int_R f(x_2 | x_1, y)f(x_1, y)dy &= \\ &= (2\pi)^{-\frac{p}{2}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet}^{-1}|^{-\frac{1}{2}} |V_{11}^{-1}|^{-\frac{1}{2}} \Gamma\left(\frac{p+v}{2}\right) 2^{\frac{p+v}{2}} \left[{}^t(x_2 - \mu_{2\bullet})V_{22\bullet}^{-1}(x_2 - \mu_{2\bullet})\right. \\ &\quad \left. + {}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + v\right]^{\frac{p+v}{2}} \\ &= \pi^{-\frac{p}{2}} v^{\frac{v}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet}^{-1}|^{-\frac{1}{2}} |V_{11}^{-1}|^{-\frac{1}{2}} \left[{}^t(x_2 - \mu_{2\bullet})V_{22\bullet}^{-1}(x_2 - \mu_{2\bullet})\right. \\ &\quad \left. + \frac{{}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + v}{v + p_1} (v + p_1)\right]^{\frac{p+v}{2}}. \end{aligned}$$

Θέτοντας

$$\gamma = \frac{{}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + v}{v + p_1}$$

στην τελευταία παράσταση, έχουμε :

$$\begin{aligned} \int_R f(x_2 | x_1, y)f(x_1, y)dy &= \\ &= \pi^{-\frac{p}{2}} v^{\frac{v}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet}^{-1}|^{-\frac{1}{2}} |V_{11}^{-1}|^{-\frac{1}{2}} \gamma^{-\frac{p+v}{2}} \left[{}^t(x_2 - \mu_{2\bullet})(\gamma V_{22\bullet}^{-1})^{-1}(x_2 - \mu_{2\bullet}) + (v + p_1)\right]^{\frac{p+v}{2}} \end{aligned}$$

Επίσης,

$$\begin{aligned} f(x_1) &= \int_{R^{p_2}} f(x_1, x_2)dx_2 \\ &= \pi^{-\frac{p}{2}} v^{\frac{v}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet}^{-1}|^{-\frac{1}{2}} |V_{11}^{-1}|^{-\frac{1}{2}} \gamma^{-\frac{p+v}{2}} \end{aligned}$$

$$\begin{aligned}
& \times \underbrace{\int_{R^{p_2}} [{}^t(x_2 - \mu_{2\bullet 1})(\gamma \mathcal{W}_{22\bullet 1})^{-1}(x_2 - \mu_{2\bullet 1}) + (v + p_1)]^{\frac{-(p_2 + (v + p_1))}{2}} dx_2}_{\text{ολοκλ. } t_{p_2}(\mu_{2\bullet 1}, \gamma \mathcal{W}_{22\bullet 1}, v + p_1)} \\
& = \pi^{-\frac{p}{2}} \nu^{\frac{v}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet 1}^{-1}|^{\frac{1}{2}} |V_{11}|^{\frac{1}{2}} \gamma^{-\frac{p+v}{2}} \pi^{\frac{p_2}{2}} \frac{\Gamma\left(\frac{v+p_1}{2}\right)}{\Gamma\left(\frac{v+p}{2}\right)} |\gamma \mathcal{W}_{22\bullet 1}|^{\frac{1}{2}} (v+p_1)^{\frac{v+p_1}{2}} \\
& = \pi^{-\frac{p_1}{2}} \nu^{\frac{v}{2}} \frac{\Gamma\left(\frac{p_1+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{11}|^{\frac{1}{2}} \gamma^{-\frac{p_1+v}{2}} (v+p_1)^{\frac{v+p_1}{2}} \\
& = \pi^{-\frac{p_1}{2}} \nu^{\frac{v}{2}} \frac{\Gamma\left(\frac{p_1+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{11}|^{\frac{1}{2}} [{}^t(x_1 - \mu_1)V_{11}^{-1}(x_1 - \mu_1) + v]^{\frac{v+p_1}{2}},
\end{aligned}$$

δηλαδή

$$X_1 \sim t_{p_1}(\mu_1, V_{11}, v)$$

Επομένως,

$$\begin{aligned}
f(x_2 | x_1) & = \left\{ \pi^{-\frac{p}{2}} \nu^{\frac{v}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{22\bullet 1}|^{\frac{1}{2}} |V_{11}|^{\frac{1}{2}} \gamma^{-\frac{p+v}{2}} \right. \\
& \quad \left. \times [{}^t(x_2 - \mu_{2\bullet 1})(\gamma \mathcal{W}_{22\bullet 1})^{-1}(x_2 - \mu_{2\bullet 1}) + (v + p_1)]^{\frac{-(p_2 + (v + p_1))}{2}} \right\} \\
& \div \left\{ \pi^{-\frac{p_1}{2}} \nu^{\frac{v}{2}} \frac{\Gamma\left(\frac{p_1+v}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} |V_{11}|^{\frac{1}{2}} \gamma^{-\frac{v+p_1}{2}} (v+p_1)^{\frac{v+p_1}{2}} \right\} \\
& = \pi^{-\frac{p_2}{2}} \frac{\Gamma\left(\frac{p+v}{2}\right)}{\Gamma\left(\frac{v+p_1}{2}\right)} |\gamma \mathcal{W}_{22\bullet 1}|^{\frac{1}{2}} (v+p_1)^{\frac{v+p_1}{2}} \\
& \quad \times [{}^t(x_2 - \mu_{2\bullet 1})(\gamma \mathcal{W}_{22\bullet 1})^{-1}(x_2 - \mu_{2\bullet 1}) + (v + p_1)]^{\frac{-(p_2 + (v + p_1))}{2}}
\end{aligned}$$

η οποία αποτελεί την π.π. της $t_{p_2}(\mu_{2\bullet 1}, \gamma \cdot V_{22\bullet 1}, v + p_1)$.

Λοιπόν,

$$X_2 | X_1 = x_1 \sim t_{p_2}(\mu_{2.1}, \gamma \cdot V_{22.1}, \nu + p_1)$$

Περιθωριακά, κάθε συντεταγμένη $X_i, i=1, \dots, p$ του X ακολουθεί μια μονοδιάστατη Student:

$$X_i \sim t(\mu_i, u_{ii}, \nu),$$

όπου u_{ii} είναι το (i, i) -στοιχείο του πίνακα V .

8.3 Γραμμικοί μετασχηματισμοί

Οι ιδιότητες των γραμμικών μετασχηματισμών των πολυδιάστατων κανονικών κατανομών μεταφέρονται, μέσω του ορισμού της Student, στις πολυδιάστατες Student μεταβλητές.

Έστω $X \sim t_p(\mu, V, \nu)$ και $Z = AX + b$, όπου A (r, p) πίνακας γνωστός και b ($r, 1$) διάνυσμα γνωστό με $r \leq p$. Τότε

$$Z \sim t_r(A\mu + b, AV'A, \nu).$$

Επειδή, $X \sim t_p(\mu, V, \nu)$ αυτή θα προκύπτει όταν

$$X | Y = y \sim N_p(\mu, y^{-1}V) \text{ και } Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

Από το Θεώρημα 6.3.3 ii),

$$Z | Y = y \sim N_r(A\mu + b, y^{-1}AV'A)$$

Με δεδομένο λοιπόν ότι :

$$Z | Y = y \sim N_r(A\mu + b, y^{-1}AV'A) \text{ και } Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu}\right)$$

έχουμε

$$Z \sim t_r(A\mu + b, AV'A, \nu).$$

Βιβλιογραφία

- [1] M.Abramowitz and I.Stegun (1964). Handbook of Mathematical Functions, Dover.
- [2] J.M. Bernardo (1979). Reference posterior distribution for Bayesian inference (with discussion). J.Ryal Statist.Soc.(Ser.B)41,113-147.
- [3] J.P. Florens(1998). La specification de la distribution a priori. Journées d' Etude en Statistique, C.I.R.M.
- [4] G.Fourgeau, A.Fuschs (1972). Statistique. Dunod
- [5] C.F.Gauss (1810). Methode des moindres carres.Memoire sur lu combination des opservations, Trad.J.Bertrand.Mallet-Bachelier, Paris(1985)
- [6] L.Le Cam (1982). On the risk of Bayes estimates, in Statistical Decision Theory and Related Topics III, Vol.2, J.L.Berger and S.S.Gupta (Eds). Academic Press, New York.
- [7] L.Le Cam (1986). Asymptotic Method, in Statistical Decision Theory. Springer – Verlag, New York.
- [8] E.L. Lehmann and G. Casella (1995). Theory of Point Estimation. Wadsworth, Pacific Grove, California.
- [9] F.W.J. Olver(1974). Asymptotics and Special Functions. Academic Press, New York.
- [10] E. Pitman (1937). The closest estimates of statistical parameter. Proc.Cambridge Phil.Soc.33,212-222.
- [11] C.R.Rao (1973). Linear Statistical Inference and its Applications. Wiley.
- [12] C.Robert (1992). L'Analyse Statistique Bayesienne. Economica.
- [13] C. Robert (1996). Intrinsic Losses. Theory and Decision, 40, 191-214, INSEE.
- [14] H.Rubin (1987). A weak system of axioms for “rational” behavior and the nonseparability of utility from prior. Statist.Decisions 5,47-58.
- [15] G.Roussas (1973). A first course in mathematical statistics. Addison Wesley.
- [16] L.Simar (1998). Methods Bayesienne en Statistique. Journées d' Etude en Statistique, C.I.R.M.
- [17] Γ. Χριστοδούλου (2003). Το Bayesian Παράδειγμα και Minimax Bayes Εκτιμήσεις. Διπλωματική εργασία. Τμήμα Μαθηματικών, Πανεπιστήμιο Κρήτης.