ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Πέτρος Δελλαπόρτας – Παναγιώτης Τσιαμυρτζής

ΣΗΜΕΙΩΣΕΙΣ ΜΑΘΗΜΑΤΟΣ

"Στατιστική κατά Bayes"

ΑΘΗΝΑ 2004

Κεφάλαιο 1ο

Εισαγωγή

1.1 Αβεβαιότητα και πιθανότητα

Θεωρία και πρακτική πάνε χέρι-χέρι. Αν η πρώτη δεν έχει πρακτικές εφαρμογές, δεν έχει αναπτυχθεί σωστά. Αν η δεύτερη δεν υποστηρίζεται από μία καλά θεμελιωμένη θεωρία είναι υποκειμενική, παραπλανητική και ασυνεπής.

Η στατιστική κατά Bayes βασίζεται σε μία απλή ιδέα: ***η μόνη ικανοποιητική περιγραφή της αβεβαιότητας μας επιτυγχάνεται μέσω της πιθανότητας.*** Η αβεβαιότητα υπάρχει καθημερινά στην ζωή μας, δίνει ομορφιά και ενδιαφέρον στην ζωή μας. Η Μπευζιανή προσέγγιση μας δίνει, μέσω του υπολογισμού πιθανοτήτων, ένα ισχυρό εργαλείο να καταλάβουμε, να χειριστούμε και να ελέγξουμε την αβεβαιότητα. Ο χρυσός κανόνας έπεται αυτόματα: όλες οι άγνωστες ποσότητες πρέπει να περιγράφονται δια μέσου πιθανοτήτων.

Γιατί δεν υπάρχει άλλος συνεπής τρόπος να περιγράψουμε την αβεβαιότητα; Γιατί ο μόνος ικανοποιητικός τρόπος να συνδέσουμε διάφορες γνώμες που αφορούν την αβεβαιότητα είναι να το κάνουμε με τον ίδιο ακριβώς τρόπο όπως χειριζόμαστε τις πιθανότητες. Εάν προχωρήσουμε πέρα από μία απλή πρόταση για την αβεβαιότητα και θεωρήσουμε σύνολα προτάσεων, που δεν επιθυμούμε να αντικρούονται μεταξύ τους, τότε καταλήγουμε στην πιθανότητα. Η βασική ιδέα ονομάζεται συνέπεια (coherence), και δύο προσεγγίσεις βρίσκονται στους DeGroot (1970, Optimal Statistical Decisions, McGraw-Hill) και De Finetti (1974, Theory of Probability, Vol 1, Wiley). Σημειώστε ότι η συνέπεια δεν συζητείται στην κλασική στατιστική.

Οι τρεις βασικοί κανόνες με τους οποίους ο λογισμός πιθανοτήτων καταφέρνει να είναι συνεπής είναι η πιθανότητα (βρίσκεται στο (0,1) με το 0 να εκφράζει το αδύνατο), ο αθροιστικός και ο πολλαπλασιαστικός κανόνας. Με βάση αυτούς, όλοι οι άλλοι κανόνες έπονται. Αυτά είναι τα μόνα εργαλεία μας. Πιθανότατα είναι το μόνο εργαλείο που χρειαζόμαστε για την μελέτη της αβεβαιότητας μας.

Όπως θα περιγράψουμε στις σημειώσεις αυτές, η συνταγή είναι απλή: τί είναι αβέβαιο και σε ενδιαφέρει; ονόμασε το θ. Τί ξέρεις; ονόμασε το x. Στην συνέχεια υπολόγισε το P(θ|x). Πώς; με τους κανόνες του λογισμού πιθανοτήτων. Τίποτα παραπάνω, τίποτα παρακάτω! Θέλεις να πάρεις αποφάσεις; μεγιστοποίησε την αναμενόμενη χρησιμότητα. Θέλεις να ερμηνεύσεις την πιθανότητα; αυτό είναι προσωπική Σου υπόθεση. Δεν υπάρχει αληθινή πιθανότητα αλλά απλά μία έκφραση της σχέσης Σου με τον κόσμο. Θέλεις να κάνεις προβλέψεις; χρησιμοποίησε μόνο λογισμό πιθανοτήτων. Θέλεις να επιλέξεις ένα μοντέλο από ένα σύνολο μοντέλων M; Χρησιμοποίησε μόνο λογισμό πιθανοτήτων!

1.2 Στατιστική συμπερασματολογία

Πριν καθορίσουμε τί είναι η συμπερασματολογία κατά Bayes, θα πρέπει να δώσουμε απάντηση σε μία γενικότερη και ευρύτερη ερώτηση: «Τί είναι η στατιστική συμπερασματολογία». Πολλές απαντήσεις θα μπορούσαν να δοθούν στο συγκεκριμένο ερώτημα, αλλά οι περισσότερες από αυτές καταλήγουν στην θέση ότι *στατιστική συμπερασματολογία είναι η επιστήμη η οποία έχει στόχο να εξάγει συμπεράσματα για έναν πληθυσμό μελετώντας ένα δείγμα που προέρχεται από τον πληθυσμό αυτό*. Μία τέτοια ερμηνεία της στατιστικής συμπερασματολογίας, οδηγεί σε μία πληθώρα ερωτημάτων όπως τί εννοούμε λέγοντας πληθυσμός, ποια είναι η σχέση πληθυσμού-δείγματος, πώς θα επιλέξουμε το κατάλληλο δειγματοληπτικό σχήμα, κ.α. Ας επικεντρώσουμε την προσοχή μας σε ένα απλό παράδειγμα, για να γίνουν κατανοητά τα παραπάνω.

Ας υποθέσουμε ότι η Δασική Υπηρεσία θέλει να εκτιμήσει την αναλογία των δέντρων που παρουσιάζουν κάποια συγκεκριμένη ασθένεια σε ένα μεγάλο δάσος. Είναι εκ των πραγμάτων αδύνατο να ελεγχθεί κάθε ένα από τα δέντρα του δάσους για να βρεθεί η αναλογία τους, και για αυτόν τον λόγο κρίνεται σκόπιμο να ληφθεί και να μελετηθεί ένα δείγμα από τα δέντρα για την εξαγωγή συμπερασμάτων. Με βάση την υπόθεσή ότι ακολουθούμε τυχαία δειγματοληψία για την επιλογή του δείγματος, θέτουμε θ την αναλογία των δέντρων που έχουν την ασθένεια έτσι ώστε κάθε δέντρο που έχει την ασθένεια στο δείγμα μας συλλέγεται ανεξάρτητα από όλα τα άλλα στο δείγμα, με πιθανότητα θ. Θεωρώντας X ως την τυχαία μεταβλητή η οποία αντιστοιχεί στον αριθμό των δέντρων που έχουν την ασθένεια στο δείγμα, η Δασική Υπηρεσία θα χρησιμοποιήσει την παρατηρηθείσα τιμή X=x, για να μπορέσει να εξάγει συμπεράσματα για την παράμετρο θ του πληθυσμού. Η συμπερασματολογία θα μπορούσε να προκύψει είτε από σημειακή εκτίμηση, είτε από την κατασκευή ενός διαστήματος (με βεβαιότητα 95% το θ βρίσκεται στο διάστημα [0.08,0.12]), είτε από τον έλεγχο κάποιας υπόθεσης (απορρίπτουμε την υπόθεση ότι το θ<0.07 σε 5% επίπεδο σημαντικότητας), είτε κάνοντας προβλέψεις (προβλέπουμε ότι το 15% των δέντρων θα προβληθούν από την ασθένεια το επόμενο έτος), είτε με την λήψη απόφασης (αποφασίζουμε να εντοπίσουμε και να αφαιρέσουμε όλα τα μολυσμένα δέντρα).

Στην περίπτωση αυτή γίνεται εμφανές πως για την εξαγωγή συμπερασμάτων για την παράμετρο θ του πληθυσμού, είναι απαραίτητο να παρατηρηθεί η τιμή της μεταβλητής X=x, του δείγματος. Επιπλέον, τα συμπεράσματα προκύπτουν εφόσον καθοριστεί κάποιο μοντέλο πιθανότητας, $f(x|\theta)$, το οποίο καθορίζει πως για μια δεδομένη τιμή του θ κατανέμονται οι πιθανότητες για διαφορετικές τιμές της μεταβλητής X.

Στο συγκεκριμένο παράδειγμα, με τις υποθέσεις που κάναμε για την τυχαία δειγματοληψία, το μοντέλο θα είναι της μορφής:

$$X|\theta \sim \mathrm{Bin}(n,\theta).$$

Η στατιστική συμπερασματολογία οδηγεί σε συμπεράσματα για την παράμετρο θ του πληθυσμού μέσω της παρατήρησης της μεταβλητής X, και τα βασικά συμπεράσματα βασίζονται στο ότι οι τιμές του θ που δίνουν μεγάλη πιθανότητα στην τιμή του x που παρατηρήθηκε, είναι πιο πιθανές απ'ότι εκείνες που δίνουν στο x μικρή πιθανότητα (αρχή της μέγιστης πιθανοφάνειας).

Πριν κάνουμε λόγο για την συμπερασματολογία κατά Bayes, υπάρχουν ορισμένα σημεία που αφορούν την κλασική στατιστική τα οποία θα πρέπει να αναφερθούν. Το πιο σημαντικό από τα σημεία αυτά είναι ότι η παράμετρος θ (η οποία δεν είναι γνωστή), χρησιμοποιείται περισσότερο σαν να είναι μία σταθερά και όχι σαν τυχαία μεταβλητή. Αυτό αποτελεί και τον θεμέλιο λίθο της κλασικής στατιστικής, όμως οδηγεί σε πολλά προβλήματα ερμηνείας. Για παράδειγμα, όταν λέμε ότι θέλουμε το 95% διάστημα εμπιστοσύνης του [0.08,0.12], εννοούμε πως υπάρχει 95% πιθανότητα η παράμετρος θ να βρίσκεται μεταξύ 0.08 και 0.12. Όμως σε αυτήν την περίπτωση υπάρχει πρόβλημα ερμηνείας, διότι το θ δεν είναι τυχαίο: είτε θα ανήκει στο διάστημα, είτε δεν θα ανήκει σε αυτό, και άρα η πιθανότητα δεν μπορεί και δεν πρέπει να υπεισέρχεται σαν παράγοντας στην ερμηνεία του. Το μόνο τυχαίο στοιχείο στο μοντέλο πιθανότητας είναι τα δεδομένα, οπότε η σωστή ερμηνεία του διαστήματος είναι πως αν επαναλάβουμε την διαδικασία πολλές φορές, τότε στο προσεχές μέλλον τα διαστήματα που θα κατασκευάσουμε θα περιλαμβάνουν την παράμετρο θ στο 95% των περιπτώσεων. Όλες οι συμπερασματολογίες που βασίζονται στην κλασική θεωρία, οδηγούνται στο να έχουν αυτήν την μακροχρόνια ερμηνεία, παρ' όλο που στην πράξη έχουμε ένα διάστημα να ερμηνεύσουμε.

1.3 Συμπερασματολογία κατά Bayes

Το πλαίσιο στο οποίο κινείται η συμπερασματολογία κατά Bayes είναι παρόμοιο με αυτό της κλασικής στατιστικής: υπάρχει η παράμετρος θ του πληθυσμού την οποία θέλουμε να εκτιμήσουμε, καθώς και η πιθανότητα $f(x|\theta)$ η οποία καθορίζει την πιθανότητα παρατήρησης διαφορετικών x, κάτω από διαφορετικές τιμές της παραμέτρου θ. Όμως ***η θεμελιώδης διαφορά είναι ότι το θ χρησιμοποιείται σαν τυχαία ποσότητα***. Αν και η διαφορά αυτή μπορεί να φανεί όχι και τόσο ουσιαστική, οδηγεί σε μία τελείως διαφορετική προσέγγιση, ως προς την ερμηνεία, από αυτήν την κλασικής στατιστικής.

Στην ουσία, η συμπερασματολογία μας θα βασιστεί στην $f(\theta|x)$ και όχι στην $f(x|\theta)$, δηλαδή στην πιθανότητα της κατανομής της παραμέτρου δεδομένης της x (δεδομένα) και όχι της x δεδομένης της παραμέτρου. Σε πολλές περιπτώσεις αυτό οδηγεί σε περισσότερο φυσικά συμπεράσματα σε σχέση με την κλασική στατιστική, για να μπορέσει όμως να επιτευχθεί αυτό θα πρέπει να καθορίσουμε την a-priori κατανομή $f(\theta)$ (prior probability distribution), η οποία αντιπροσωπεύει «τις πεποιθήσεις» μας για την κατανομή του θ προτού αποκτήσουμε οποιαδήποτε πληροφορία για τα δεδομένα μας.

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

1.4 Η a-priori κατανομή (Prior Distribution)

Σχεδόν σε όλες τις περιπτώσεις, όταν προσπαθούμε να εκτιμήσουμε την παράμετρο θ θα πρέπει να έχουμε κάποια γνώση ή κάποια πεποίθηση σχετικά με την τιμή της θ προτού λάβουμε υπόψη μας τα δεδομένα. Στο παράδειγμα (O'Hagan 1994) που ακολουθεί, γίνονται κατανοητά τα παραπάνω.

Ας υποθέσουμε ότι κάποιος κοιτάει έξω από το παράθυρο του σπιτιού του και βλέπει ένα ξύλινο αντικείμενο με πράσινα φύλλα. Υπάρχουν δύο πιθανές υποθέσεις: η μία είναι πως αυτό το οποίο βλέπει το άτομο είναι ένα δέντρο, και η άλλη είναι πως είναι ένας ταχυδρόμος. Φυσικά αποκλείουμε την περίπτωση να είναι ταχυδρόμος γιατί οι ταχυδρόμοι δεν έχουν αυτήν την μορφή, ενώ την έχουν τα δέντρα. Σε γλώσσα πιθανοτήτων αν θεωρήσουμε A το ενδεχόμενο να δει κανείς το ξύλινο αντικείμενο με τα πράσινα φύλλα, B_1 το ενδεχόμενο να είναι δέντρο και B_2 το ενδεχόμενο να είναι ταχυδρόμος, αποκλείουμε το B_2 γιατί $f(A|B_1)>f(A|B_2)$. Εδώ χρησιμοποιείται η αρχή της μέγιστης πιθανοφάνειας.

Όμως μπορούμε να θεωρήσουμε πως υπάρχει και μία τρίτη πιθανότητα B_3, ότι το αντικείμενο που βλέπουμε είναι αντίγραφο δέντρου. Σε αυτήν την περίπτωση μπορούμε να πούμε με σιγουριά ότι $f(A|B_1)=f(A|B_3)$ αλλά πάλι θα απορρίπταμε την B_3 έναντι της B_1. Αυτό το οποία προκύπτει από το παράδειγμα, είναι πως παρ'όλο που η πιθανότητα αυτό που είδες να είναι δέντρο ή ένα αντίγραφό του είναι η ίδια, η a-priori πεποίθηση μας είναι ότι το αντικείμενο που βλέπουμε είναι πιθανότερο να είναι δέντρο και όχι αντίγραφο του, και για αυτόν τον σκοπό περιλαμβάνουμε την πληροφορία αυτή όταν παίρνουμε την απόφαση μας.

Ένα άλλο παράδειγμα είναι αυτό που ακολουθεί όπου σε κάθε μία από τις περιπτώσεις το μοντέλο είναι $X|\theta \sim Bin(10,\theta)$ και $x=10$, και στόχος μας είναι να ελέγξουμε εάν η υπόθεση H_0:$\theta \leq 0.5$ απορρίπτεται έναντι της H_1:$\theta > 0.5$ σε κάθε μία από τις παραπάνω περιπτώσεις:

1. Μία γυναίκα η οποία πίνει συχνά τσάι, ισχυρίζεται πως δοκιμάζοντας ένα φλυτζάνι τσάι είναι σε θέση να πει αν το γάλα που περιέχεται σε αυτό, έχει προστεθεί πριν ή μετά από το τσάι. Και στις 10 δοκιμές απαντά σωστά.
2. Ένας έμπειρος μουσικός ισχυρίζεται πως μπορεί να ξεχωρίσει από ένα απόσπασμα ενός κομματιού, αν το κομμάτι ανήκει σε έργο του Hayden ή του Mozart. Απαντά σωστά και τις 10 φορές.

3. Ένας μεθυσμένος άντρας λέει πως είναι σε θέση να προβλέψει το αποτέλεσμα από την ρίψη ενός κέρματος (κεφάλι ή γράμματα), και πραγματικά προβλέπει σωστά και τις 10 φορές.

Ας προσπαθήσουμε τώρα χρησιμοποιώντας τα παραπάνω παραδείγματα, να βγάλουμε τα ίδια συμπεράσματα και για τις τρεις περιπτώσεις. Όμως οι a-priori πεποιθήσεις μας για κάθε μία από τις παραπάνω περιπτώσεις είναι διαφορετικές: αντιμετωπίζουμε με αρκετή διστακτικότητα την περίπτωση του μεθυσμένου άντρα, με σχετική έκπληξη αυτήν της γυναίκας με το τσάι, και θεωρούμε σχεδόν αναμενόμενη την περίπτωση του μουσικού.

Το βασικό σημείο που προκύπτει από τα παραπάνω είναι το εξής: τα πειράματα δεν είναι αφηρημένα τεχνάσματα. Αντίθετα, έχουμε κάποια γνώση για την διαδικασία την οποία μελετάμε προτού να συλλέξουμε τα δεδομένα. Είναι λογικό, πολλοί θα έλεγαν απαραίτητο, ότι τα συμπεράσματα θα πρέπει να βασίζονται στον συνδυασμό της a-priori γνώσης με τα δεδομένα. Η συμπερασματολογία κατά Bayes είναι ο μηχανισμός ο οποίος εξάγει συμπεράσματα από τον συνδυασμό αυτόν.

1.5 Χαρακτηριστικά της Μπευζιανής προσέγγισης

Μπορούμε να εντοπίσουμε τέσσερα βασικά σημεία τα οποία χαρακτηρίζουν την Μπευζιανή θεωρία σε σχέση με την κλασική στατιστική.

- **A-priori Πληροφορία (Prior Information)**: Κάθε πρόβλημα είναι μοναδικό και έχει το δικό του περιεχόμενο. Από αυτό ακριβώς το περιεχόμενο εξάγονται a-priori πληροφορίες και είναι η διατύπωση και η εκμετάλλευση της προηγούμενης γνώσης που διαχωρίζουν την Μπευζιανή θεωρία από αυτήν της κλασικής στατιστικής.
- **Υποκειμενική Πιθανότητα (Subjective Probability)**: Η κλασική στατιστική εξαρτάται από μία μακροχρόνια συχνότητα καθορισμού των πιθανοτήτων. Αν και αυτό είναι επιθυμητό, οδηγεί σε «δυσκίνητα» συμπεράσματα. Αντίθετα, η Μπευζιανή στατιστική θέτει με σαφήνεια την ιδέα ότι όλες οι πιθανότητες είναι υποκειμενικές και εξαρτώνται από τις πεποιθήσεις του κάθε ατόμου και τις γνώσεις που μπορεί να έχει καθένας από μας για μια δεδομένη «κατάσταση». Η συμπερασματολογία της

βασίζεται στην ***a-posteriori κατανομή (posterior distribution) f(θ|x),*** η μορφή της οποίας εξαρτάται (μέσω του θεωρήματος του Bayes) από τον τρόπο καθορισμού της a-priori κατανομής *f(θ)*.

- **Συνέπεια (Self-Consistency)**: Χρησιμοποιώντας την παράμετρο θ σαν τυχαία, όλη η ανάπτυξη της Μπευζιανής συμπερασματολογίας πηγάζει και εξαρτάται μόνο από την θεωρία πιθανοτήτων. Αυτό έχει πολλά πλεονεκτήματα και σημαίνει πως όλα τα συμπεράσματα μπορούν να παρουσιαστούν με την μορφή πιθανοτήτων για την παράμετρο θ, πράγματι προκύπτουν άμεσα από την a-posteriori κατανομή.
- **Μη προσκόλληση σε «συνταγές»**: Επειδή η κλασική στατιστική δεν είναι σε θέση να χρησιμοποιήσει όρους πιθανοτήτων για την παράμετρο θ, έχουν αναπτυχθεί αρκετά κριτήρια με σκοπό να καθορίσουν πότε ένας συγκεκριμένος εκτιμητής θα μπορούσε να χαρακτηριστεί ως «καλός».

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

  - Επιλέγουμε μία στατιστική συνάρτηση δεδομένων t=t(x) η οποία συνδέεται με την παράμετρο θ.
  - Βρίσκουμε τη δειγματική κατανομή του t, p(t|x).
  - Μετράμε την πιθανότητα κάθε τιμής του θ, ελέγχοντας τη δεδομένη τιμή του t σε σχέση με την αναμενόμενη (σε πολλαπλές δοκιμές) συμπεριφορά του, δεδομένης της θ. Για μία συγκεκριμένη τιμή της $\theta=\theta_0$, αν η τιμή t βρίσκεται μέσα σε μία περιοχή που ανήκει η περισσότερη πυκνότητα πιθανότητας της $p(t|\theta_0)$, λέμε ότι η θ_0 είναι συμβατή με τα δεδομένα. Αλλιώς ή κάτι σπάνιο έχει συμβεί, ή η θ_0 δεν είναι η αληθινή τιμή της θ.

  Αυτό έχει οδηγήσει σε έναν πολλαπλασιασμό διαδικασιών οι οποίες συχνά έρχονται σε αντίθεση η μία με την άλλη. Η Μπευζιανή θεωρία έχει καταφέρει να ξεπεράσει το πρόβλημα αυτό, δημιουργώντας μια σειρά από κριτήρια τα οποία κρίνουν και συγκρίνουν τους εκτιμητές βασιζόμενα στην a-posteriori κατανομή.

1.6 Αντίθετες Απόψεις για την Μπευζιανή Θεωρία

Η βασική αντίρρηση για την Μπευζιανή θεωρία, εντοπίζεται στο γεγονός ότι τα συμπεράσματα εξαρτώνται από την επιλογή της a-priori κατανομής. Όμως όπως τονίστηκε και προηγουμένως, κάποιοι υποστηρίζουν πως ακριβώς αυτό το σημείο κρύβει όλη την «ομορφιά» της Μπευζιανής θεωρίας. Δυστυχώς η αντιπαράθεση αυτή θα μπορούσε να οδηγήσει σε μία ατέλειωτη συζήτηση πάνω σε αυτό το ζήτημα. Πριν όμως ολοκληρώσουμε το μέρος αυτό, θα πρέπει να τονίσουμε ότι ακόμα και η κλασική συμπερασματολογία χρησιμοποιεί κάποιες προηγούμενες γνώσεις. Η προηγούμενη γνώση χρησιμοποιείται για να κατασκευαστεί το κατάλληλο μοντέλο πιθανοφάνειας. Στον έλεγχο υποθέσεων, οι a-priori πεποιθήσεις μας για την αληθοφάνεια της υπόθεσης πολύ συχνά προσαρμόζονται, συνήθως όχι φανερά, μέσω της αλλαγής του επιπέδου σημαντικότητας του ελέγχου μας. Αν πιστεύουμε ότι τα δεδομένα μας πρέπει να οδηγήσουν στην απόρριψη της υπόθεσης, μπορούμε να πετύχουμε την απόρριψή αυτή επιλέγοντας ένα επίπεδο σημαντικότητας αρκετά υψηλό. Η Μπευζιανή συμπερασματολογία κατά κάποιον τρόπο τυποποιεί την ενσωμάτωση των a-priori πληροφοριών, κάτι που στην κλασική στατιστική συνήθως γίνεται όχι φανερά.

Υπάρχει και ένας ακόμα τρόπος σκέψης που αφορά την Μπευζιανή συμπερασματολογία, ο οποίος φαίνεται να ξεπερνά τις όποιες φιλοσοφικές και εννοιολογικές δυσκολίες που σχετίζονται με την a-priori γνώση. Στην κλασική στατιστική, οι εκτιμητές μεγίστης πιθανοφάνειας προκύπτουν από την επιλογή εκείνης της τιμής που μεγιστοποιεί την πιθανοφάνεια. Αντίθετα όσον αφορά την Μπευζιανή συμπερασματολογία, θα μπορούσαμε να θεωρήσουμε ότι χρησιμοποιεί έναν μέσο όρο της πιθανοφάνειας, παρά την μεγιστοποίηση της. Αυτό φαίνεται αρκετά λογικό. Η αμφισβήτηση για την όλη συμπερασματολογία του Bayes, προκαλείται εξαιτίας του ότι ο μέσος όρος αυτός σταθμίζεται με βάση την a-priori κατανομή. Όμως ακόμα και στην κλασική στατιστική, είναι αρκετά σύνηθες να δίνονται διαφορετικά βάρη σε διαφορετικά κομμάτια πληροφορίας, όπως γίνεται για παράδειγμα στην σταθμισμένη παλινδρόμηση. Από όλα τα παραπάνω φαίνεται πως όλα όσα προσπαθούν να δείξουν οι αντίθετοι με την Μπευζιανή θεωρία σαν αρνητικά της σημεία, είναι τελικά τεχνικές που χρησιμοποιούνται και στην κλασική στατιστική.

1.7 Το Θεώρημα του Bayes

Στην βασική του μορφή το θεώρημα του Bayes είναι απλό και αφορά υπό συνθήκη πιθανότητες:
Αν A και B είναι δύο ενδεχόμενα με P(A)>0, τότε

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

Η χρησιμότητα του θεωρήματος του Bayes σε εφαρμογές πιθανοτήτων είναι ότι παρέχει την δυνατότητα αντιστροφής της «θέσης» των ενδεχομένων. Έτσι, γίνεται εμφανές πώς η πιθανότητα του B|A σχετίζεται με την πιθανότητα του A|B.

Μια μικρή προέκταση του θεωρήματος του Bayes μπορεί να γίνει, αν θεωρήσουμε τα ενδεχόμενα $C_1,\ldots,C_k$, τα οποία διαμερίζουν ένα δειγματικό χώρο Ω, έτσι ώστε τα $C_i \cap C_j = \varnothing$ για κάθε $i \neq j$ και $C_1 \cup \ldots \cup C_k = \Omega$. Σε αυτήν την περίπτωση θα έχουμε

$$P(C_i \mid A) = \frac{P(A \mid C_i)P(C_i)}{\sum_{j=1}^{k} P(A \mid C_J)P(C_j)} \qquad ,i = 1,\ldots,k.$$

✍ **Παράδειγμα 1.1** Σε έναν πληθυσμό ο οποίος βρίσκεται σε μεγάλο κίνδυνο να προσβληθεί από τον ιό HIV, εφαρμόζεται μια μέθοδος εντοπισμού του ιού. Το 10% από τον πληθυσμό αυτό πιστεύεται πως έχει προσβληθεί από τον ιό. Από τα αποτελέσματα του τεστ προκύπτει ότι το τεστ είναι θετικό για το 90% των ατόμων τα οποία είναι στην πραγματικότητα προσβεβλημένα από τον ιό, και αρνητικό για το 85% των ατόμων τα οποία έχουν αρνητικό HIV. Ποιες είναι οι πιθανότητες να πάρουμε λανθασμένα θετικά και αρνητικά αποτελέσματα από το τεστ;

Λύση Αν θεωρήσουμε A το ενδεχόμενο ένα άτομο να έχει τον ιό HIV, και B το ενδεχόμενο το τεστ να δείξει αποτέλεσμα θετικό, τότε θα είναι: P(A)=0.1, P(B|A)=0.9 και $P(B^c \mid A^c) = 0.85$. Οπότε

P(το τεστ να είναι λανθασμένα θετικό)=

$$= P(A^c \mid B)$$

$$= \frac{P(B \mid A^c)P(A^c)}{P(B)}$$

$$= \frac{0.15 * 0.90}{(0.15 * 0.90) + (0.90 * 0.10)} = 0.6$$

Αντίστοιχα έχουμε

P(το τεστ να είναι λανθασμένα αρνητικό)=

$$= P(A \mid B^c)$$

$$= \frac{P(B^c \mid A)P(A)}{P(B^c)}$$

$$= \frac{0.1 * 0.1}{(0.1 * 0.1) + (0.85 * 0.90)} = 0.0129$$

✍ **Παράδειγμα 1.2** Σε ένα δοχείο υπάρχουν 6 μπάλες με άγνωστα χρώματα. Τρεις από τις μπάλες αυτές λαμβάνονται χωρίς επανατοποθέτηση και το χρώμα τους είναι μαύρο. Να βρεθεί η πιθανότητα μέσα στο βάζο να μην έχει μείνει καμία μαύρη μπάλα.

Λύση Ας θεωρήσουμε A το ενδεχόμενο να τραβήξουμε 3 μαύρες μπάλες από το βάζο και C_i το ενδεχόμενο να υπάρχουν i μαύρες μπάλες μέσα στο βάζο. Τότε με βάση το θεώρημα του Bayes θα είναι:

$$P(C_3 \mid A) = \frac{P(A \mid C_3)P(C_3)}{\sum_{j=0}^{6} P(A \mid C_j)P(C_j)}$$

Το πρόβλημα όμως που προκύπτει σε αυτό ακριβώς το σημείο είναι τί τιμές θα δώσουμε στις πιθανότητες $P(C_0),\ldots,P(C_6)$; Αυτές είναι οι πιθανότητες των διαφορετικών αριθμών από τις μαύρες μπάλες που βρίσκονται μέσα στο βάζο, *πριν* να δούμε τις μπάλες μας, δηλαδή τα δεδομένα μας. Χωρίς να έχουμε κάποια πληροφορία για τα δεδομένα μας, θα μπορούσαμε πολύ εύκολα να υποθέσουμε ότι όλα τα νούμερα έχουν την ίδια πιθανότητα, οπότε $P(C_0)=\ldots=P(C_6)=1/7$. Η ερώτηση όμως η οποία προκύπτει από την παραπάνω ισότητα είναι αν είναι λογικό να θέσουμε την ισότητα αυτή. Αν σκεφτούμε για παράδειγμα ότι είναι πιο πιθανό όλες οι μπάλες μέσα στο βάζο να είναι του ιδίου χρώματος, τότε θα πρέπει να δώσουμε μεγαλύτερες a-priori πιθανότητες στις $P(C_0)$ και $P(C_6)$. Ή θα μπορούσαμε ακόμα να γνωρίζουμε πως το εργοστάσιο το οποίο κατασκευάζει τις μπάλες, παράγει μπάλες 10 διαφορετικώς χρωμάτων. Έχοντας αυτήν την «a-priori» πληροφορία, θα μπορούσαμε να θεωρήσουμε πως κάθε ένα από τα μπαλάκια έχει πιθανότητα ίση με 1/10 και με βάση την πληροφορία αυτή θα μπορούσαμε να υπολογίσουμε και όλες τις άλλες a-priori πιθανότητες. Το σημείο το οποίο θίγεται στα παραπάνω, είναι πως θα πρέπει να σκεφτούμε αρκετά πως θέλουμε να εκφράσουμε τις a-priori πεποιθήσεις μας, και η λύση σε αυτόν τον προβληματισμό εξαρτάται από το τί πιστεύουμε πως θα πρέπει να χρησιμοποιήσουμε σαν αρχή των προβληματισμών μας. Στο σημείο αυτό θα αναφερθούμε αργότερα. Ας χρησιμοποιήσουμε λοιπόν το θεώρημα του Bayes για το παράδειγμά μας. Θα είναι:

$$P(C_3 \mid A) = \frac{P(C_3)P(A \mid C_3)}{\sum_{j=0}^{6} P(A \mid C_j)P(C_j)}$$

$$= \frac{(1/7)*[(3/6)*(2/5)*(1/4)]}{(1/7)\{0+0+0+[(3/6)(2/5)(1/4)]+[(4/6)(3/5)(2/4)]+[(5/6)(4/5)(3/4)]+[(6/6)(5/5)(4/4)]\}}$$

$$=1/35$$

Όπως είναι φανερό, τα δεδομένα μας έχουν αλλάξει τις a-priori πεποιθήσεις μας από $P(C_3)=1/7$ σε μία a-posteriori πιθανότητα $P(C_3|A)=1/35$.

1.8 Ασκήσεις

Άσκηση 1.1 Σε μία περιοχή τα γεωλογικά πετρώματα Α και Β είναι δύσκολο να διαχωριστούν. Μετά από προσεκτικές μελέτες στα εργαστήρια, βρέθηκε ότι υπάρχει ένα μόνο χαρακτηριστικό το οποίο μπορεί να καθορίσει τα δύο πετρώματα, και αυτό είναι η παρουσία ή η απουσία ενός συγκεκριμένου απολιθώματος. Η πιθανότητα να υπάρχει το απολίθωμα φαίνεται στον πίνακα 1. Είναι επίσης γνωστό, ότι το πέτρωμα τύπου Α συναντάται 4 φορές περισσότερο από ότι το πέτρωμα Β στην περιοχή η οποία μελετάται. Αν ληφθεί ένα δείγμα και βρεθεί να υπάρχει απολίθωμα σε αυτό, υπολογίστε την a-posteriori κατανομή των πετρωμάτων.

	Απολίθωμα	
Στρώματα	ΝΑΙ	ΟΧΙ
Α	0.9	0.1
Β	0.2	0.8

Εάν ο γεωλόγος κατηγοριοποιεί πάντα το πέτρωμα σαν τύπου Α, όταν το απολίθωμα υπάρχει, και ως Β όταν δεν υπάρχει, ποια είναι η πιθανότητα να κάνει σωστή κατηγοριοποίηση και στο μέλλον;

Άσκηση 1.2 Επαναλάβετε το Παράδειγμα 1.2 χρησιμοποιώντας μια διαφορετική a-priori κατανομή. Πώς η αλλαγή της a-priori κατανομής επηρεάζει την a-posteriori πιθανότητα να μην έχουν μείνει μαύρες μπάλες στο βάζο;

Άσκηση 1.3 Έστω ότι μία παρατήρηση x ακολουθεί την εκθετική κατανομή

$$p(x\,|\,\theta) = \theta e^{-\theta x}, \qquad (\theta \geq 0, \qquad x > 0)$$

Αν η a-priori κατανομή του θ ορίζεται από τις διακριτές πιθανότητες

$$\rho(\theta=1) = \rho(\theta=2) = \tfrac{1}{2}$$

βρείτε την a-posteriori κατανομή του θ.

Κεφάλαιο 2ο

Σύγχρονη Θεωρία κατά Bayes

2.1 Εισαγωγή

Όπως αναφέρθηκε στο κεφάλαιο 1, τα βασικά χαρακτηριστικά της Μπευζιανής προσέγγισης είναι η αντιμετώπιση και χρήση της άγνωστης παραμέτρου θ ως τυχαία μεταβλητή, ο καθορισμός της a-priori κατανομής για το θ (η οποία αντιπροσωπεύει τις πεποιθήσεις μας σχετικά με το θ προτού να έχουμε οποιαδήποτε πληροφορία για τα δεδομένα μας) η χρήση του θεωρήματος του Bayes για τον «εκσυγχρονισμό» των a-priori πεποιθήσεων μας σε a-posteriori πιθανότητες, και η εξαγωγή της κατάλληλης συμπερασματολογίας. Για τον σκοπό αυτό υπάρχουν τέσσερα βήματα-κλειδιά για την Μπευζιανή προσέγγιση:

1. Καθορισμός του μοντέλου πιθανοφάνειας $f(x|\theta)$.
2. Καθορισμός της a-priori κατανομής $f(\theta)$.
3. Υπολογισμός της a-posteriori κατανομής $f(\theta|x)$, από το θεώρημα του Bayes.
4. Εξαγωγή συμπερασμάτων από την a-posteriori πληροφορία.

Σε αυτό το κεφάλαιο θα διατυπώσουμε ξανά το θεώρημα του Bayes σε μία μορφή κατάλληλη για τυχαίες μεταβλητές αντί για ενδεχόμενα, και θα εξετάσουμε κάποια θέματα τα οποία προκύπτουν σαν αποτέλεσμα της προσπάθειας μας να χρησιμοποιήσουμε το αποτέλεσμα αυτό στην όλη συμπερασματολογία που αναπτύσουμε για την παράμετρο θ. Κάθε ένα από τα θέματα αυτά θα αναπτυχθεί στο σχετικό κεφάλαιο. Στο συγκεκριμένο κεφάλαιο θα ασχοληθούμε με αρκετά παραδείγματα στα οποία συγκεκριμένοι συνδυασμοί των a-priori κατανομών και του μοντέλου πιθανοφάνειας οδηγούν σε μαθηματικούς τύπους για την a-posteriori κατανομή.

2.2 Το Θεώρημα του Bayes

Το θεώρημα του Bayes σε όρους τυχαίων μεταβλητών με πυκνότητες που συμβολίζονται γενικά με f, παίρνει την εξής μορφή:

$$f(\theta \mid x) = \frac{f(\theta)f(x \mid \theta)}{\int f(\theta)f(x \mid \theta)d\theta}.$$

Θα χρησιμοποιήσουμε αυτόν τον συμβολισμό για τις περιπτώσεις που το x είναι είτε συνεχής είτε διακριτή μεταβλητή, με την σημείωση ότι στην συνεχή περίπτωση η f είναι η a-priori συνάρτηση πυκνότητας πιθανότητας, ενώ στην διακριτή περίπτωση η f είναι η συνάρτηση πιθανότητας (μάζας) του x. Αντίστοιχα, το θ μπορεί να είναι διακριτό ή συνεχές, αλλά στην διακριτή περίπτωση ο παρανομαστής της σχέσης δηλ. το $\int f(\theta)f(x|\theta)d\theta$ θα ερμηνευθεί αντίστοιχα όπως το $\Sigma f(\theta_j)f(x|\theta_j)$.
Θα πρέπει να προσέξουμε ιδιαίτερα το γεγονός ότι από την στιγμή που ολοκληρώσαμε ως προς θ, ο παρανομαστής στο θεώρημα του Bayes είναι συνάρτηση μόνο ως προς x. Συνεπώς, για δεδομένες παρατηρήσεις x, ο παρανομαστής είναι σταθερά και ονομάζεται σταθερά κανονικοποίησης. Με βάσει αυτά ένας εναλλακτικός τρόπος παρουσίασης του θεωρήματος του Bayes είναι ο εξής:

$$f(\theta|x) \propto f(\theta)f(x|\theta)$$

ή χρησιμοποιώντας λόγια θα λέγαμε ότι **η a-posteriori κατανομή (posterior distribution) είναι ανάλογη της a-priori κατανομής (prior distribution) πολλαπλασιαζόμενης με την συνάρτηση πιθανοφάνειας (likelihood function).**
Προσέξτε ακόμα ότι η f(θ|x) είναι συνάρτηση πυκνότητας πιθανότητας και αυτό το εγγυάται η σταθερά κανονικοποίησης. Επίσης, η a-priori και η a-posteriori κατανομές είναι έννοιες σχετικές: η σημερινή a-posteriori είναι η αυριανή a-priori. Απλά, όταν καθορίσουμε την a-priori ανανεώνουμε τα πιστεύω μας (ΜΟΝΟ!) μέσω του θεωρήματος του Bayes.

2.3 Βασικά Βήματα

2.3.1 Επιλογή του κατάλληλου μοντέλου πιθανοφάνειας

Η επιλογή του κατάλληλου μοντέλου πιθανοφάνειας εξαρτάται από τον μηχανισμό με βάση τον οποίο θα αντιμετωπίσουμε το κάθε πρόβλημα, και είναι ίδια η περίπτωση αυτή με εκείνη που αντιμετωπίζουμε στην κλασσική στατιστική δηλ. ποιο είναι το μοντέλο των δεδομένων μας. Πολύ συχνά η γνώση της δομής των δεδομένων μας βοηθά στην επιλογή του κατάλληλου μοντέλου (για παράδειγμα διωνυμικό μοντέλο δειγματοληψίας ή Poisson), αλλά συχνότερα *υποθέτουμε τί κατανομή ακολουθεί το μοντέλο μας* (για παράδειγμα το Y σχετίζεται γραμμικά με το X και τα λάθη είναι ανεξάρτητα, ισόνομα και ακολουθούν την κανονική κατανομή) και η αληθοφάνεια των υποθέσεων αυτών θα εκτιμηθεί αργότερα μέσα από την εξέταση της φύσης των δεδομένων (επιλογή μοντέλου).

2.3.2 Επιλογή της a-priori κατανομής

Η επιλογή της a-priori κατανομής αποτελεί βασικό θέμα στην Μπευζιανή θεωρία και για αυτόν τον σκοπό θα μελετηθεί εκτενώς στο 3ο κεφάλαιο. Παρόλα αυτά κάποια σημεία κρίνεται απαραίτητο να αναφερθούν από τώρα.

1. Από την στιγμή που η a-priori κατανομή αντιπροσωπεύει τις πεποιθήσεις σου για την παράμετρο θ προτού μελετηθούν τα δεδομένα σου, είναι φυσικό ότι η μεταγενέστερη ανάλυση είναι μοναδική για σένα. Αυτό που εννοούμε με τα παραπάνω είναι πως η a-priori κατανομή που θέτει κάποιος άλλος, θα οδηγήσει σε διαφορετική μεταγενέστερη (a-posteriori) ανάλυση. Με αυτήν την έννοια η ανάλυση είναι καθαρά υποκειμενική.
2. Θα δούμε αργότερα ότι στην περίπτωση που η a-priori δεν είναι «εντελώς παράλογη», η επιρροή της γίνεται ολοένα μικρότερη καθώς προστίθενται νέα δεδομένα. Σε αυτήν την περίπτωση ο λάθος προσδιορισμός της a-priori είναι άνευ σημασίας.
3. Πολύ συχνά έχουμε μια «γενική» ιδέα για το ποια θα πρέπει να είναι η a-priori (πιθανότατα να μπορούμε να πούμε ποιος είναι ο μέσος και η διακύμανσή της), χωρίς όμως να μπορούμε να είμαστε πιο συγκεκριμένοι για την μορφή της. Σε αυτές τις περιπτώσεις μπορούμε να χρησιμοποιήσουμε μια «βολική» μορφή της a-priori η οποία θα είναι το αποτέλεσμα των πεποιθήσεων μας και

ταυτόχρονα θα κάνει τους μαθηματικούς υπολογισμούς σχετικά εύκολους. Θα δούμε κάποια τέτοια παραδείγματα στην πορεία.

4. Μερικές φορές νομίζουμε ότι δεν έχουμε a-priori πληροφορίες σχετικά με την παράμετρο θ. Σε τέτοιες περιπτώσεις είναι αρκετά συνηθισμένο να χρησιμοποιούμε μια a-priori η οποία να ανακλά την άγνοια μας για την παράμετρο. Αυτό είναι συχνά δυνατό να γίνει, αλλά υπάρχουν αρκετές δυσκολίες που σχετίζονται με αυτό. Με το θέμα αυτό θα ασχοληθούμε στο κεφάλαιο 3.

2.3.3 Υπολογισμός

Η εφαρμογή του θεωρήματος του Bayes στην πράξη είναι αρκετά δύσκολη όσον αφορά τους μαθηματικούς υπολογισμούς και η δυσκολία αυτή οφείλεται κυρίως στο ολοκλήρωμα το οποίο υπάρχει στον παρανομαστή. Στην συνέχεια θα δούμε πως με ορισμένες επιλογές των a-priori κατανομών, το ολοκλήρωμα αυτό μπορεί να παραληφθεί, αλλά γενικά ειδικές τεχνικές χρειάζονται για να απλοποιηθούν οι υπολογισμοί (βλ. Κεφάλαιο 8).

2.3.4 Συμπερασματολογία

Η Μπευζιανή ανάλυση παρέχει μία περισσότερο πλήρη συμπερασματολογία υπό την έννοια ότι όλη η γνώση σχετικά με την παράμετρο θ η οποία είναι διαθέσιμη λόγω της a-priori κατανομής και των δεδομένων, αντιπροσωπεύεται από την a-posteriori κατανομή. Αυτό σημαίνει πως **η $f(\theta|x)$ «είναι» η συμπερασματολογία**. Παρόλα αυτά, είναι συχνά επιθυμητό να συνοψίσουμε τα συμπεράσματα μας μέσω σημειακής εκτίμησης ή μέσω ενός διαστήματος εκτίμησης. Περισσότερες λεπτομέρειες σχετικά, θα δοθούν στο κεφάλαιο 5.

2.4 Παραδείγματα

Παράδειγμα 2.1 Όταν ένα συγκεκριμένο μηχάνημα παρουσιάζει κάποιο πρόβλημα αυτό μπορεί να οφείλεται είτε στην μηχανή, είτε στην ύπαρξη λάθους στην μεταφορά. Η φύση του προβλήματος μπορεί να προσδιοριστεί με ακρίβεια μόνο στην περίπτωση που «λυθεί» η μηχανή. Παρόλα αυτά το πρόβλημα έχει τρεις ενδείξεις οι οποίες μπορούν να μελετηθούν: υπερθέρμανση (ΥΘ), ανώμαλη κίνηση (ΑΚ), ή και τα δύο. Οι πιθανότητες του Πίνακα 2.1 καταρτίστηκαν με βάση προηγούμενα στοιχεία. Επίσης, το 60% των προβλημάτων που παρουσιάζουν τα μηχανήματα οφείλονται στην μεταφορά, οπότε $f(\theta_2)=0.6$. Βρείτε την a-posteriori κατανομή και ερμηνεύστε την κατάλληλα.

Περιοχή προβλήματος	ΥΘ(x1)	ΑΚ(x2)	και τα δύο(x3)
Μηχανή (θ1)	0.1	0.4	0.5
Μεταφορά (θ2)	0.5	0.3	0.2

Πίνακας 2.1: Πίνακας Πιθανοτήτων

Για αυτό το απλό διακριτό παράδειγμα μπορούμε να συνοψίσουμε την ανάλυση μας σε πίνακες όπως στον πίνακα 2.2, όπου έχουμε χρησιμοποιήσει την $f(x,\theta)=f(x|\theta)f(\theta)$ καθώς και την:

$$f(x) = \sum_{i=1}^{2} f(x,\theta_i)$$

f(θ)	f(x|θ)	x1	x2	x3
0.4	θ1	0.1	0.4	0.5
0.6	θ2	0.5	0.3	0.2
f(x,θ)	θ1	0.04	0.16	0.20
	θ2	0.30	0.18	0.12
	f(x)	0.34	0.34	0.32
f(θ|x)	θ1	4/34	16/34	20/32
	θ2	30/34	18/34	12/32

Πίνακας 2.2: Πίνακας Ανάλυσης

Ένας τρόπος για να ερμηνεύσουμε αυτήν την ανάλυση είναι να μελετήσουμε τα "odds" για κάθε έναν από τους δύο τύπους προβλήματος. Πριν από οποιαδήποτε πληροφορία σχετικά με το x, τα odds ήταν 3:2 για το θ_2 (από την στιγμή που το 60% των προβλημάτων οφείλεται στην μεταφορά). Αλλά έχοντας μελετήσει το x, αυτά τα odds αλλάζουν σε 15:2 για το θ_2 όταν παρατηρήσουμε το x_1, 9:8 για το θ_2 αν παρατηρήσουμε το x_2, και 5:3 για το θ_1 αν παρατηρήσουμε το x_3. Κατά συνέπεια, εάν το κριτήριο βάσει του οποίου θα αποφασίσουμε να επιλέξουμε την περισσότερο αληθοφανή αιτία, μελετώντας είτε το x_1 είτε το x_2 θα μας οδηγήσει να αποφασίσουμε ότι το πρόβλημα οφείλεται στην μεταφορά, ενώ αν μελετήσουμε το x_3 τότε θα οδηγηθούμε στην απόφαση ότι το πρόβλημα οφείλεται στην μηχανή.

✍ **Παράδειγμα 2.2** (Διωνυμικό δείγμα) Ας υποθέσουμε ότι το μοντέλο πιθανοφάνειας μας είναι x~Bin(n,θ), και αυτό που επιδιώκουμε είναι να εξάγουμε συμπεράσματα για την παράμετρο *θ*. Οπότε θα είναι:

$$f(x \mid \theta) = {}^{n}C_{x}\theta^{x}(1-\theta)^{n-x}, x = 0,...,n.$$

Γενικά, η επιλογή της a-priori περιγραφής της παραμέτρου *θ* μπορεί να ποικίλει από πρόβλημα σε πρόβλημα, και εξ ορισμού θα εξαρτάται από την έκταση της a-priori γνώσης μας σχετικά με την κατάσταση. Ωστόσο, η διαδικασία την οποία θα ακολουθήσουμε εδώ, θα βασιστεί σε μία πιθανή οικογένεια a-priori κατανομών οι οποίες, όπως και θα δούμε, θα οδηγήσει σε απλούς μαθηματικούς υπολογισμούς. Το σημείο το οποίο θίγεται εδώ, είναι πως δεδομένου ότι η οικογένεια είναι αρκετά μεγάλη και κατά συνέπεια καλύπτει επαρκώς ένα μεγάλο εύρος από πιθανές μορφές κατανομών, μπορούμε να χρησιμοποιήσουμε την a-priori μέσα από αυτήν την οικογένεια η οποία είναι αρκετά κοντά στις a-priori πεποιθήσεις μας. Εάν, παρόλα αυτά, δεν υπάρχει a-priori κατανομή μέσα στην οικογένεια αυτή η οποία ανακλά τί πραγματικά πιστεύουμε, τότε θα πρέπει να αποφύγουμε την προσέγγιση αυτή.
Έτσι, σε αυτήν την περίπτωση, ας υποθέσουμε ότι μπορούμε να αντιπροσωπεύσουμε τις πεποιθήσεις μας σχετικά με την παράμετρο *θ* μέσω της κατανομής Beta:

$$\theta \sim Be(p,q)$$

έτσι ώστε

$$f(\theta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\theta^{p-1}(1-\theta)^{q-1} \propto \theta^{p-1}(1-\theta)^{q-1}, \qquad (0 \le \theta \le 1).$$

Χρησιμοποιώντας το θεώρημα του Bayes θα έχουμε:

$$
\begin{aligned}
f(\theta \mid x) &\propto f(\theta)f(x \mid \theta) \\
&\propto \theta^{x}(1-\theta)^{n-x} \times \theta^{p-1}(1-\theta)^{q-1} = \\
&= \theta^{p+x-1}(1-\theta)^{q+n-x-1}
\end{aligned}
$$

Τώρα, από την στιγμή που γνωρίζουμε ότι η f(θ|x) είναι κατάλληλη συνάρτηση πυκνότητας, βρισκόμαστε στην περίπτωση όπου:

$$\theta|x \sim Be(p+x,q+n-x).$$

Κατ' αυτόν τον τρόπο, με προσεκτική επιλογή, έχουμε βρει την a-posteriori κατανομή η οποία ανήκει στην ίδια οικογένεια όπως και η a-priori κατανομή, και μέσω αυτής της διαδικασίας έχουμε αποφύγει να κάνουμε υπολογισμούς ολοκληρωμάτων. Η επίδραση των δεδομένων είναι η αναπροσαρμογή των παραμέτρων της κατανομής Beta από τις αρχικές τους τιμές (p,q), στις a-posteriori τιμές των (p+x,q+n-x).

Θα δώσουμε ένα αριθμητικό παράδειγμα, χρησιμοποιώντας τα δεδομένα «Καρκίνου». Από τους 70 ασθενείς στους οποίους εφαρμοζόταν μια καινούργια θεραπεία για ένα συγκεκριμένο τύπο καρκίνου, οι 34 από αυτούς έζησαν πέρα από το αναμενόμενο όριο. Ας θεωρήσουμε θ την πιθανότητα ένας ασθενής να επιβιώσει. Οι ειδικοί σε ιατρικά θέματα, οι οποίοι είναι εξοικειωμένοι με κλινικές μελέτες (trials), εκφράζουν πεποιθήσεις ότι θα είναι E(θ)=0.4 και Var(θ)=0.02. Η κατανομή Beta είναι κατάλληλη για να εκφράσει τις a-priori πεποιθήσεις τους και για αυτό θα πρέπει να επιλέξουμε μια a-priori κατανομή θ~Be(p,q) τέτοια ώστε E(θ)=0.4 και Var(θ)=0.02. Δηλαδή θέλουμε:

$$\frac{p}{p+q} = 0.4 \qquad \text{και} \qquad \frac{pq}{(p+q)^2(p+q+1)} = 0.02$$

Μπορεί εύκολα να αποδειχθεί ότι αν θέσουμε m=E(θ) και u=Var(θ) οι παραπάνω εξισώσεις θα έχουν την εξής μορφή:

$$p = \frac{(1-m)m^2}{u} - m \qquad \text{και} \quad q = \frac{(1-m)^2 m}{u} - (1-m)$$

Για το παράδειγμα μας οι τιμές των p και q θα είναι αντίστοιχα 4.4 και 6.6. Οι τιμές αυτές καθορίζουν την a-priori κατανομή για το *θ*. Στην πραγματικότητα, είναι απαραίτητο να επιβεβαιώσουμε ότι η όλη κατανομή είναι το αποτέλεσμα των a-priori πεποιθήσεων των ειδικών.

Με την προϋπόθεση ότι πραγματικά αντιπροσωπεύονται οι πεποιθήσεις των ειδικών στην a-priori κατανομή, μπορούμε να βρούμε την a-posteriori κατανομή για το *θ* η οποία είναι:

$$\theta|x \sim Be(38,4\ 42,6)$$

H a-posteriori αυτή κατανομή αντιπροσωπεύει όλη την πληροφορία σχετικά με την παράμετρο *θ* και αποτελεί ολοκληρωμένη συμπερασματολογία για την *θ*. Θα μελετήσουμε τώρα πως τα δεδομένα μας αλλάζουν τις a-priori πεποιθήσεις μας, συγκρίνοντας τις αναμενόμενες τιμές των a-priori και a-posteriori κατανομών:

$$E(\theta) = \frac{p}{p+q} \qquad E(\theta\,|\,x) = \frac{p+x}{p+q+n}$$

Στην περίπτωση μας θα είναι:

$$E(\theta)=0.4 \text{ και } E(\theta|x)=0.474$$

Οπότε αυτό το οποίο προκύπτει είναι ότι όταν μελετήσουμε και τα δεδομένα μας η a-priori εκτίμηση για την παράμετρο *θ* αυξάνεται από 0.4 σε 0.474. Από την άλλη πλευρά, ένας φυσικός εκτιμητής για το *θ* βασιζόμενος μόνο στα δεδομένα μας είναι ο x/n=0.486, ο οποίος είναι ο εκτιμητής μέγιστης πιθανοφάνειας. Επομένως συνοψίζοντας τα παραπάνω θα λέγαμε ότι ο a-posteriori εκτιμητής είναι ένας συνδυασμός των δεδομένων με τις a-priori πεποιθήσεις μας.

Γενικά, αν τα x και n είναι μεγάλα, τότε η αναμενόμενη a-posteriori θα είναι κατά προσέγγιση η x/n, δηλαδή ο εκτιμητής μέγιστης πιθανοφάνειας.

Παράδειγμα 2.3 (Poisson δείγμα). Ας υποθέσουμε ότι τα $x_1,\ldots,x_n$ είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν κατανομή Poisson με παράμετρο *θ*. Τότε θα έχουμε:

$$L(\theta;x)=\prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!}\propto e^{-n\theta}\theta^{\sum x_i}.$$

Όπως και στο παράδειγμα με την διωνυμική κατανομή, οι a-priori πεποιθήσεις μας σχετικά με την παράμετρο *θ* αλλάζουν από πρόβλημα σε πρόβλημα, αλλά αυτό που θα προσπαθήσουμε να κάνουμε εδώ είναι να βρούμε μία σχέση η οποία να δίνει ένα σύνολο από διαφορετικές πιθανότητες και η οποία να είναι ταυτόχρονα μαθηματικά αποδεκτή. Στην συγκεκριμένη περίπτωση θα υποθέσουμε ότι οι a-priori πεποιθήσεις μας μπορούν να αντιπροσωπευτούν μέσα από μία Γάμμα κατανομή, δηλ. $\theta \sim Ga(p,q)$, έτσι ώστε:

$$f(\theta)=\frac{q^p}{\Gamma(p)}\theta^{p-1}\exp\{-q\theta\} \qquad (\theta>0)$$

Χρησιμοποιώντας το θεώρημα του Bayes, θα έχουμε:

$$\begin{aligned} f(\theta\,|\,x) &\propto \frac{q^p}{\Gamma(p)}\theta^{p-1}\exp\{-q\theta\}\times\exp\{-n\theta\}\theta^{\sum x_i} \\ &\propto \theta^{(p+\sum x_i-1)}\exp\{-(q+n)\theta\}, \end{aligned}$$

και επομένως:

$$\theta\,|\,x \sim Ga(p+\sum_{i=1}^{n}x_i,\, q+n).$$

η οποία είναι μία νέα Γάμμα κατανομή της οποίας οι παράμετροι έχουν καθοριστεί με βάσει τα δεδομένα μέσω των $\sum x_i$ και του n.

Ένα αριθμητικό παράδειγμα είναι το ακόλουθο. Ας συμβολίσουμε με *θ* τον μέσο αριθμό ενός είδους πτηνών σε μία συγκεκριμένη περιοχή. Λεπτομερείς φωτογραφίες από 45 ομάδες του συγκεκριμένου είδους πτηνών, έδωσαν $\sum x_i=4019$. Θα υποθέσουμε ότι αν η $\theta \sim Ga(p,q)$, με γνωστή μέση τιμή και διασπορά, τότε θα ήταν $E(\theta)=p/q=100$ και $Var(\theta)=p/(q^2)=20$, και λύνοντας ως προς p και q θα πάρουμε p=500 και q=5. Έτσι προκύπτει ότι η a-posteriori κατανομή μας θ|x είναι Ga(4519,50).

Παράδειγμα 2.4 (Κανονικός μέσος). Ας θεωρήσουμε $x_1,\ldots,x_n$ ένα σύνολο ανεξάρτητων μεταβλητών από μία κανονική κατανομή με μέσο *θ* και διακύμανση σ^2, με σ^2 γνωστό. Τότε θα είναι:

$$f(x_i \mid \theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{(x_i - \theta)^2}{2\sigma^2}\}$$

με πιθανοφάνεια

$$L(\theta; x) \propto \exp\{-\frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\sigma^2}\}.$$

Τώρα ας υποθέσουμε ότι οι a-priori πεποιθήσεις μας σχετικά με την παράμετρο *θ*, μπορούν να αντιπροσωπευθούν μέσα από την Κανονική κατανομή, δηλ.$\theta \sim N(b,d^2)$. (Και εδώ η επιλογή της συγκεκριμένης κατανομής έγινε για να χρησιμοποιήσουμε απλά μαθηματικά, αλλά πρέπει να τονίσουμε πως η κατανομή θα πρέπει να επιλέγεται με βάση το αν είναι ικανή να αποτελέσει μια καλή προσέγγιση των a-priori πεποιθήσεων μας σχετικά με την παράμετρο *θ*). Με βάση το θεώρημα του Bayes έχουμε:

$$f(\theta|x) \propto f(x|\theta)\, f(\theta)$$

και έτσι:

$$f(\theta \mid x) \propto \exp\{-\frac{(\theta - b)^2}{2d^2}\} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{(\theta - b)^2}{2d^2} - \frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\frac{(\theta - b)^2}{d^2} - \frac{\sum_{i=1}^{n}(x_i - \theta)^2}{\sigma^2}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\frac{(\theta - b)^2}{d^2} - \frac{\sum_{i=1}^{n}(\theta - x_i)^2}{\sigma^2}\right]\right\}$$

Όμως ισχύει

$$\sum A_i (x-\alpha_i)^2 = \sum A_i \left(x - \frac{\sum \alpha_i A_i}{\sum A_i} \right)^2 + \sigma\tau\alpha\theta\varepsilon\rho\acute{\alpha}$$

οπότε από την τελευταία σχέση, η παράσταση

$$\frac{(\theta - b)^2}{d^2} - \frac{\sum_{i=1}^{n} (\theta - x_i)^2}{\sigma^2}$$

μπορεί να γραφεί ως:

$$\left[\frac{1}{d^2} + \sum_{i=1}^{n} \frac{1}{\sigma^2}\right] \left[\theta - \frac{\frac{b}{d^2} + \frac{x_1}{\sigma^2} + \ldots + \frac{x_n}{\sigma^2}}{\frac{1}{d^2} + \sum_{i=1}^{n} \frac{1}{\sigma^2}} \right]^2 =$$

$$\left[\frac{1}{d^2} + \frac{n}{\sigma^2}\right] \left[\theta - \frac{\frac{b}{d^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2}}{\frac{1}{d^2} + \frac{n}{\sigma^2}} \right]^2 =$$

$$\left[\frac{1}{d^2} + \frac{n}{\sigma^2}\right] \left[\theta - \frac{\frac{b}{d^2} + \frac{\bar{x} n}{\sigma^2}}{\frac{1}{d^2} + \frac{n}{\sigma^2}} \right]^2$$

Οπότε η a-posteriori θα έχει την ακόλουθη μορφή:

$$f(\theta \mid x) \propto \exp\left\{ -\frac{1}{2} \left[\frac{1}{d^2} + \frac{n}{\sigma^2}\right] \left[\theta - \frac{\frac{b}{d^2} + \frac{\bar{x} n}{\sigma^2}}{\frac{1}{d^2} + \frac{n}{\sigma^2}} \right]^2 \right\}$$

$$\theta \mid x \sim N\left(\frac{\frac{b}{d^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{d^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{d^2} + \frac{n}{\sigma^2}} \right)$$

Το αποτέλεσμα μπορεί να δοθεί με πιο σύντομο τρόπο, εάν καθορίσουμε την «ακρίβεια» να είναι το αντίστοιχο της διακύμανσης: πχ. ας θεωρήσουμε $\tau=1/\sigma^2$ και $c=1/d^2$. Τότε η $\theta|x$ μπορεί να γραφεί:

$$\theta \mid x \sim N\left(\frac{cb + n\tau\bar{x}}{c + n\tau}, \frac{1}{c + n\tau} \right).$$

Προτού δώσουμε ένα αριθμητικό παράδειγμα, μπορούμε να κάνουμε μία σειρά από παρατηρήσεις:

1. Παρατηρούμε ότι

$$E(\theta \mid x) = \gamma_n b + (1 - \gamma_n)\bar{x}$$

όπου

$$\gamma_n = \frac{c}{c + n\tau}$$

Συνεπώς ο μέσος της a-posteriori κατανομής είναι ένας σταθμισμένος μέσος του μέσου της a-priori κατανομής και του $\bar{x}$. Επίσης, η στάθμιση γ_n καθορίζεται με βάση τη σχετική βαρύτητα της πληροφορίας στην a-priori κατανομή και στην δομή των δεδομένων. Σε αυτήν την περίπτωση, αν το $n\tau$ είναι μεγάλο σχετικά με το c, τότε το $\gamma_n \approx 0$ και ο μέσος της a-posteriori κατανομής είναι κοντά στον $\bar{x}$.

2. Παρατηρούμε ότι:
ακρίβεια της a-posteriori = ακρίβεια της a-priori+n_την ακρίβεια κάθε παρατηρήσεως.

3. Καθώς το $n \to \infty$,

$$\theta \mid x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

έτσι ώστε η a-priori δεν επηρεάζει την α posteriori στο όριο.

4. Καθώς το $d \to \infty$, ή αντίστοιχα το $c \to 0$, θα πάρουμε και πάλι:

$$\theta \mid x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

5. Παρατηρούμε ότι η a-posteriori κατανομή εξαρτάται από τα δεδομένα μόνο μέσω του $\bar{x}$ και όχι μέσω των μεμονωμένων τιμών των x_i. Μπορούμε να πούμε και πάλι ότι η $\bar{x}$ είναι επαρκής για την παράμετρο *θ*.

Οι παρατηρήσεις 3 και 4 είναι πολύ σημαντικές και θα εξεταστούν λεπτομερώς παρακάτω.

Στο αριθμητικό παράδειγμα που ακολουθεί, τα δεδομένα είναι καταγραμμένα από τον Henry Cavendish τον 18ο αιώνα, ο οποίος έκανε 23 μετρήσεις για την πυκνότητα της γης. Για αυτά τα δεδομένα υπολογίστηκε ότι ο $\bar{x}$ είναι 5,48 και θα υποθέσουμε ότι η διακύμανση και τα σφάλματα μετρήσεως είναι ίση με 0,04. Τώρα ας υποθέσουμε ότι από προηγούμενα πειράματα η a-priori πληροφορία μας για την παράμετρο *θ* δηλ. την πυκνότητα της γης, ακολουθεί κανονική κατανομή με N(5,4 0,01). Σε αυτήν την περίπτωση η a-posteriori κατανομή θα είναι: $\theta|x \sim N(5.46, 0.00303)$.

2.5 Γενικά Θέματα

Οι αρχές και οι λεπτομέρειες των παραδειγμάτων που προηγήθηκαν, δημιούργησαν μία πληθώρα από θέματα τα οποία χρήζουν ιδιαιτέρας ανάλυσης.

2.5.1 Συνεχής αναθεώρηση

Είδαμε ότι το θεώρημα του Bayes προσφέρει τον μηχανισμό με βάση τον οποίο οι α priori πληροφορίες μας αναθεωρούνται από τα δεδομένα και δίνουν την α posteriori πληροφορία. Αυτή η α posteriori πληροφορία, θα αποτελέσει την

καινούργια α priori πληροφόρηση, πριν προστεθούν και νέα δεδομένα. Έτσι προκύπτει η εξής ερώτηση: εάν πάρουμε μια σειρά από δεδομένα και αναθεωρήσουμε τις α priori πεποιθήσεις μας την στιγμή που λαμβάνουμε κάθε ένα από τα δεδομένα μας, θα πάρουμε διαφορετικό αποτέλεσμα από ότι θα παίρναμε αν περιμέναμε να φτάσουν στα χέρια μας όλα τα δεδομένα μαζί και μετά να αναθεωρήσουμε την a-priori πληροφόρηση μας;

Ας δούμε μία απλή περίπτωση στην οποία έχουμε δύο ανεξάρτητες τυχαίες μεταβλητές x_1 και x_2, κάθε μία από τις οποίες έχει συνάρτηση πυκνότητας πιθανότητας $f(x|\theta)$. Τώρα ας υποθέσουμε ότι μελετάμε την x_1 και αναθεωρούμε την a-priori πληροφόρηση μας, μέσα από το θεώρημα του Bayes και έτσι έχουμε:

$$f(\theta \mid x_1) \propto f(\theta) \times f(x_1 \mid \theta).$$

Η συνάρτηση αυτή θα γίνει η καινούργια μας a-priori προτού μελετήσουμε την μεταβλητή x_2. Τότε θα έχουμε:

$$\begin{aligned} f(\theta \mid x_1, x_2) &\propto f(\theta) \times f(x_1 \mid \theta) \times f(x_2 \mid \theta) \\ &= f(\theta) \times f(x_1, x_2 \mid \theta) \end{aligned}$$

το οποίο είναι το ίδιο αποτέλεσμα το οποίο θα παίρναμε αν αναθεωρούσαμε την πληροφόρηση μας με βάση τη συνολική πληροφορία που θα λαμβάναμε από τα x_1 και x_2 μαζί. Η διαπίστωση αυτή μπορεί να γενικευθεί για οποιονδήποτε αριθμό δεδομένων.

2.5.2 Επάρκεια

Στην κλασική στατιστική συμπερασματολογία, η επάρκεια παίζει βασικό ρόλο τόσο στην θεωρητική ανάπτυξη, όσο και στις πρακτικές εφαρμογές. Το ίδιο ακριβώς συμβαίνει και στην Μπευζιανή ανάλυση, και έχουμε ήδη αντιμετωπίσει μία σειρά παραδειγμάτων όπου η a-posteriori κατανομή εξαρτάται από τα δεδομένα μόνο μέσω μίας επαρκούς στατιστικής παραμέτρου.

Κατ' αναλογία με την κλασική στατιστική, ορίζεται και στην στατιστική κατά Bayes η έννοια της επάρκειας. Μία στατιστική συνάρτηση δεδομένων $t(x)$ ονομάζεται επαρκής στατιστική συνάρτηση για την παράμετρο θ εάν

$$f(\theta \mid t(x)) = f(\theta|x)$$

για κάθε a-priori κατανομή $f(\theta)$.

2.5.3. Η Αρχή της Πιθανοφάνειας

Η αρχή της πιθανοφάνειας αναφέρει ότι αν 2 πειράματα «μοιράζονται» την ίδια πιθανοφάνεια (αναλογικά), τότε η συμπερασματολογία μας σχετικά με την παράμετρο θ, θα πρέπει να είναι η ίδια σε κάθε περίπτωση. Με άλλα λόγια, όλα τα στοιχεία της συμπερασματολογίας μας θα πρέπει να είναι βασισμένα μονάχα στην συνάρτηση πιθανοφάνειας. Ένα σημαντικό πλεονέκτημα της Μπευζιανής θεωρίας είναι ότι οι τεχνικές οι οποίες χρησιμοποιούνται είναι άρρηκτα συνδεδεμένες με την αρχή της πιθανοφάνειας. Αντίθετα πολλές απλές τεχνικές της κλασικής στατιστικής παραβιάζουν την αρχή αυτή.

✍ **Παράδειγμα 2.5** Έστω ότι ένα διωνυμικό πείραμα σε κάποιο εργαστήριο είχε σαν αποτέλεσμα μία επιτυχία σε 10 δοκιμές. Ο/Η στατιστικός ελέγχει την υπόθεση ότι η πιθανότητα επιτυχίας θ είναι ίση με ½ υπολογίζοντας την πιθανότητα ουράς της διωνυμικής κατανομής. Αυτή η πιθανότητα είναι:

$$P(x=0|\theta=1/2)+P(x=1|\theta=1/2)=2^{-10}+10(2)^{-1}(2)^{-10}=11(2)^{-10}$$

Παρόλα αυτά όταν παρουσιάζει τα αποτελέσματα του/της, πληροφορείται ότι το πείραμα δεν ήταν διωνυμικό: οι δοκιμές έπρεπε να σταματήσουν μόλις υπήρχε η πρώτη επιτυχία. Συνεπώς ενώ τα δεδομένα είναι τα ίδια, η στατιστική ανάλυση πρέπει να επαναληφθεί. Η νέα πιθανότητα ουράς, σε αύτη την περίπτωση δίνεται από το γεωμετρικό μοντέλο, και είναι:

$$P[Y\geq 10|\theta=1/2]=2^{-10}+2^{-11}+\ldots=2^{-10}(1+1/2+\ldots)=2^{-10}(1/(1-1/2))=2^{-9}.$$

Από το συγκεκριμένο παράδειγμα συμπεραίνουμε ότι τα ίδια δεδομένα, μέσω της κλασικής προσέγγισης, μπορεί να μας δώσουν διαφορετικά αποτελέσματα.

2.6 Ασκήσεις

Άσκηση 2.1 Για κάθε μία από τις ακόλουθες περιπτώσεις, υπολόγισε την a-posteriori κατανομή:

A)$x_1,\ldots,x_n$ είναι τυχαίο δείγμα από κατανομή με συνάρτηση πιθανότητας

$$f(x \mid \theta) = \theta^{x-1}(1-\theta) \qquad x = 1,2,...$$

με την Beta(p,q) a-priori κατανομή:

$$f(\theta) = \frac{\theta^{p-1}(1-\theta)^{q-1}}{Be(p,q)} \qquad 0 < \theta < 1$$

B) $x_1,\ldots,x_n$ είναι τυχαίο δείγμα από την κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(x \mid \theta) = \frac{e^{-\theta}\theta^{x}}{x!} \qquad x = 0,1,...$$

και με a-priori κατανομή την:

$$f(\theta) = e^{-\theta} \qquad 0 < \theta$$

Άσκηση 2.2 Η αναλογία θ των ελαττωματικών αντικειμένων σε ένα μεγάλο φορτίο είναι άγνωστη, αλλά οι ειδικοί πιστεύουν πως το θ ακολουθεί Beta(2, 200) κατανομή. Αν 100 αντικείμενα επιλεγούν τυχαία από το φορτίο και 3 από αυτά βρεθούν να είναι ελαττωματικά, ποια είναι η a-posteriori κατανομή του θ;
Εάν κάποιος άλλος στατιστικός έχοντας μελετήσει τα 3 ελαττωματικά, υπολόγιζε πως η a-posteriori κατανομή ήταν Beta με μέσο 4/102 και διακύμανση 0.0003658, ποια a-priori κατανομή είχε χρησιμοποιήσει;

Άσκηση 2.3 Η διάμετρος ενός προϊόντος σε μία μεγάλη γραμμή παραγωγής, ποικίλει βάσει της N(θ,1) κατανομής. Ένας μηχανικός υπολογίζει ότι η a-priori κατανομή του θ είναι N(10, 0.25). Σε μία γραμμή παραγωγής 12 προϊόντα ελέγχονται και βρίσκεται ότι έχουν δειγματικό μέσο 31/3. Χρησιμοποιώντας αυτήν την πληροφορία, υπολογίστε την πιθανότητα ο μέσος της διαμέτρου του προϊόντος να είναι τουλάχιστον 10 μονάδες.

Άσκηση 2.4 Ο αριθμός των ελαττωμάτων σε 1200 μέτρα ρολού μαγνητικής ταινίας ακολουθεί κατανομή Poisson(θ). Η a-priori κατανομή για το θ είναι Γάμμα(3,1).

Όταν 5 ρολά μαγνητικής ταινίας επιλεγούν τυχαία, ο αριθμός των ελαττωμάτων είναι 2,2,6,0 και 3 αντίστοιχα. Καθορίστε την a-posteriori κατανομή για το θ.

Άσκηση 2.5 Ας υποθέσουμε ότι ο χρόνος σε λεπτά που απαιτείται για να εξυπηρετηθεί ένας πελάτης σε μία τράπεζα ακολουθεί εκθετική κατανομή με παράμετρο θ. Η a-priori κατανομή για το θ είναι Γάμμα με μέσο 0,2 και τυπική απόκλιση 1. Εάν ο μέσος χρόνος για να εξυπηρετηθούν 20 πελάτες έχει παρατηρηθεί ότι είναι 3.8 λεπτά, καθορίστε ποια είναι η a-posteriori κατανομή για την παράμετρο θ.

Άσκηση 2.6 Έστω p η πιθανότητα πρόσληψης ενός ιού εάν κάποιος έρθει σε επαφή με αυτόν. Ένας γιατρός πιστεύει ότι αυτή η πιθανότητα είναι μία από τις τιμές

0.10, 0.12, 0.14, 0.16, 0.18 ή 0.20.

Εκτιμά τις a-priori πιθανότητες να είναι Pr(p=0.10)=0.05, Pr(p=0.12)=0.08, Pr(p=0.14)=0.13, Pr(p=0.16)=0.30, Pr(p=0.18)=0.34 και Pr(p=0.20)=0.10.
Από αυτήν την a-priori κατανομή, βρείτε τα E(p) και V(p). Ο γιατρός εκτελεί ένα πείραμα στο οποίο 20 άτομα εκτίθενται στον ιο, και 2 προσβάλλονται από αυτόν. Με βάση την καινούργια πληροφορία, βρείτε την a-posteriori κατανομή του ιατρού για το p.

Άσκηση 2.7 Ένα τυχαίο δείγμα $x1,\ldots,x_n$ προέρχεται από κατανομή Poisson με μέσο θ. Η a-priori κατανομή για το θ ακολουθεί Γάμμα κατανομή με μέσο μ_0. Εάν ο δειγματικός μέσος είναι $\overline{x}_n$, δείξτε ότι ο μέσος της a-posteriori κατανομής του θ θα είναι ένας σταθμισμένος μέσος της μορφής:

$$\gamma_n \overline{x}_n + (1-\gamma_n)\mu_0.$$

Δείξτε ότι το $\gamma n \rightarrow 1$ καθώς το $n \rightarrow \infty$.

Άσκηση 2.8 Γιατί στο παράδειγμα 2.5 φαίνεται ότι η κλασική στατιστική δεν υπακούει στην αρχή της πιθανοφάνειας;

Κεφάλαιο 3ο

Καθορισμός των A-priori Κατανομών

3.1 Εισαγωγή

Έχουμε ήδη δει ότι η βασική διαφορά μεταξύ την Μπευζιανής θεωρίας και της κλασικής στατιστικής είναι ότι σύμφωνα με την Μπευζιανή στατιστική οι άγνωστες παράμετροι χρησιμοποιούνται σαν τυχαίες μεταβλητές και για αυτόν τον λόγο η χρησιμοποίηση του θεωρήματος του Bayes απαιτεί τον καθορισμό a-priori κατανομών για τις μεταβλητές αυτές. Παρόλο που αυτό διευκολύνει τον συνυπολογισμό των a-priori πεποιθήσεων μας σχετικά με τις παραμέτρους, η επιλογή της a-priori κατανομής δεν πρέπει να γίνεται «τυφλά». Αντίθετα, χρειάζεται πολλή προσοχή καθώς και γνώση γύρω από ορισμένα βασικά στοιχεία τα οποία σχετίζονται με την επιλογή της. Στο κεφάλαιο αυτό θα ασχοληθούμε με ορισμένα από τα βασικά αυτά στοιχεία.

3.2 Συζυγείς A-priori Κατανομές

Η χρησιμοποίηση του θεωρήματος του Bayes συνεπάγεται αρκετές υπολογιστικές δυσκολίες, και για αυτό είναι απαραίτητος ο υπολογισμός της σταθεράς κανονικοποίησης στο ολοκλήρωμα,

$$\int f(\theta) f(x \mid \theta) d\theta.$$

Για παράδειγμα, ας υποθέσουμε ότι οι $x_1,\ldots,x_n$ είναι ανεξάρτητες μεταβλητές που ακολουθούν κατανομή Poisson $Po(\theta)$, και οι a-priori πεποιθήσεις μας σχετικά με την τιμή της παραμέτρου θ βρίσκεται σίγουρα μέσα στο διάστημα [0,1], αλλά όλες οι τιμές στο διάστημα αυτό έχουν την ίδια πιθανότητα να είναι η τιμή της παραμέτρου θ. Έτσι $f(\theta)=1$ με $0<\theta<1$. Έτσι η σταθερά κανονικότητας είναι:

$$\int_0^1 \exp(-n\theta)\theta^{\sum x_i} d\theta$$

και το ολοκλήρωμα αυτό, το οποίο είναι μία ημιτελής Γάμμα συνάρτηση, μπορεί μόνο να υπολογιστεί αριθμητικά.

Αυτό το οποίο προκύπτει είναι ότι ακόμα και οι απλές επιλογές των a-priori κατανομών μπορούν να οδηγήσουν σε περίεργα αριθμητικά προβλήματα. Παρόλα αυτά, στο προηγούμενο κεφάλαιο αντιμετωπίσαμε τρία παραδείγματα όπου η σωστή επιλογή της a-priori κατανομής οδήγησε σε υπολογισμούς της a-posteriori κατανομής χωρίς την εμπλοκή κάποιου ολοκληρώματος. Σε κάθε μία από αυτές τις περιπτώσεις, ήμασταν σε θέση να καθορίσουμε μια a-priori κατανομή για την οποία η a-posteriori κατανομή που προέκυπτε ανήκε στην ίδια οικογένεια κατανομών με την a-priori. Τέτοιες κατανομές ονομάζονται ***συζυγείς a-priori κατανομές***. Ας δούμε ένα ακόμα παράδειγμα, στο οποίο η συζυγής a-priori μπορεί να υπολογιστεί.

Παράδειγμα 3.1 (Γάμμα δείγμα) Ας υποθέσουμε ότι οι μεταβλητές $x_1,\ldots,x_n$ είναι ανεξάρτητες και ακολουθούν κατανομή Γάμμα(κ,θ), όπου το κ είναι γνωστό. (Σημειώνουμε ότι στην ειδική περίπτωση που κ=1 τότε βρισκόμαστε στην περίπτωση της εκθετικής κατανομής). Έχουμε ότι:

$$L(\theta;x) \propto \theta^{nk} \exp(-\theta \sum x_i)$$

Μελετώντας αυτήν την μορφή την οποία θεωρούμε συνάρτηση του θ, παρατηρούμε ότι μπορούμε να πάρουμε μία a-priori κατανομή της μορφής:

$$f(\theta) \propto \theta^{p-1} \exp(-q\theta)$$

οπότε η παράμετρος $\theta \sim Ga(p,q)$, και με βάση το θεώρημα του Bayes θα είναι:

$$f(\theta \mid x) \propto \theta^{p+nk-1} \exp\{-(q+\sum x_i)\theta\}$$

και άρα η $\theta|x \sim Ga(p+nk, q+\sum x_i)$.

3.2.1 Χρήση των συζυγών a-priori κατανομών

Η χρήση των συζυγών a-priori πρέπει να αντιμετωπίζεται ακριβώς όπως είναι: ένας βολικός μαθηματικός μετασχηματισμός. Παρόλα αυτά, η έκφραση των a-priori πεποιθήσεων μέσω κάποιας παραμετρικής κατανομής είναι πάντα μία προσέγγιση. Σε πολλές περιπτώσεις ο πλούτος της οικογένειας των συζυγών κατανομών είναι αρκετά μεγάλος ώστε να μπορέσει να προσδιοριστεί η συζυγής a-priori η οποία είναι αρκετά κοντά στις a-priori πεποιθήσεις μας. Αυτό όμως το οποία θα πρέπει να τονίσουμε είναι πως οι συζυγείς a-priori δεν θα πρέπει να χρησιμοποιούνται απλά επειδή κάνουν τους μαθηματικούς υπολογισμούς ευκολότερους.

3.2.2 Ύπαρξη συζυγών a-priori κατανομών

Με την προϋπόθεση ότι δεν έρχονται σε αντίθεση με τις a-priori πεποιθήσεις μας και δεδομένου ότι μία τέτοια οικογένεια κατανομών μπορεί να βρεθεί, η απλοποίηση των υπολογισμών που προσφέρουν οι συζυγείς a-priori, είναι εντυπωσιακή. Όμως σε ποιες περιπτώσεις μπορούμε να πάρουμε μία οικογένεια συζυγών κατανομών;

Προκύπτει ότι η μόνη περίπτωση στη οποία οι συζυγείς κατανομές μπορούν να προκύψουν εύκολα, είναι για τα μοντέλα δεδομένων τα οποία ανήκουν στην ***εκθετική οικογένεια κατανομών***. Αυτή είναι:

$$f(x \mid \theta) = h(x)g(\theta)\exp\{t(x)c(\theta)\}$$

για τις συναρτήσεις h,g,t και c έτσι ώστε:

$$\int f(x \mid \theta)dx = g(\theta)\int h(x)\exp\{t(x)c(\theta)\}dx = 1.$$

Αυτό αν και μπορεί να φαίνεται περιοριστικό, στην πραγματικότητα περιλαμβάνει την εκθετική κατανομή, την κατανομή Poisson, την Γάμμα με μία παράμετρο, την Διωνυμική και την Κανονική κατανομή με γνωστή διακύμανση.

Χρησιμοποιώντας μία a-priori f(θ), έχουμε:

$$\begin{aligned} f(\theta \mid x) &\propto f(\theta)L(\theta; x) \\ &= f(\theta)\prod_{i=1}^{n}\{h(x_i)\}g(\theta)^n \exp\{\sum_{i=1}^{n} t(x_i)c(\theta)\} \\ &\propto f(\theta)g(\theta)^n \exp\{\sum_{i=1}^{n} t(x_i)c(\theta)\}. \end{aligned}$$

Εάν επιλέξουμε την

$$f(\theta) \propto g(\theta)^d \exp\{bc(\theta)\}$$

θα πάρουμε την a-posteriori:

$$f(\theta \mid x) \propto g(\theta)^{n+d} \exp\{c(\theta)[\sum_{i=1}^{n} t(x_i) + b]\}$$
$$= g(\theta)^{\tilde{d}} \exp\{\tilde{b}c(\theta)\},$$

η οποία ανήκει στην ίδια οικογένεια κατανομών με την a-priori, αλλά με προσαρμοσμένες παραμέτρους.

Μπορεί εύκολα να ελεγχθεί ότι σε όλα τα παραδείγματα των συζυγών a-priori τα οποία έχουμε συναντήσει ως τώρα, έχουν προκύψει κατ' αυτόν τον τρόπο. Για παράδειγμα, στην περίπτωση της Διωνυμικής κατανομής ήταν

$$f(x \mid \theta) = {}^nC_x \theta^x (1-\theta)^{n-x}$$
$$= {}^nC_x \exp\{x \log(\frac{\theta}{1-\theta})\}.$$

Στην εκθετική οικογένεια κατανομών έχουμε

$$h(x) = {}^nC_x$$
$$g(\theta) = (1-\theta)^n$$
$$t(x) = x \qquad \text{και}$$
$$c(\theta) = \log(\frac{\theta}{1-\theta}).$$

Έτσι μπορούμε να κατασκευάσουμε μία συζυγή a-priori με την μορφή:

$$f(\theta) \propto [(1-\theta)^n]^d \exp\{b \log(\frac{\theta}{1-\theta})\}$$
$$= (1-\theta)^{nd-b} \theta^b$$

η οποία είναι μέλος της οικογένειας κατανομών Βήτα (Beta).

3.2.3 Ανάλυση συνηθισμένων συζυγών κατανομών

Ο Πίνακας 3.1 παρουσιάζει πολλές από τις συνηθισμένες συζυγείς a-priori κατανομές. Κάποιες από αυτές έχουν ήδη παρουσιαστεί σε παραδείγματα.

Πιθανοφάνεια	A-priori	A-posteriori
$x \sim B(n,\theta)$	$Be(p,q)$	$Be(p+x,q+n-x)$
$x_1,\dots,x_n \sim Po(\theta)$	$Ga(p,q)$	$Ga(p+\Sigma x_i,q+n)$
$x_1,\dots,x_n \sim Ga(k,\theta)$ (k γνωστό)	$Ga(p,q)$	$Ga(p+nk,q+\Sigma x_i)$
$x_1,\dots,x_n \sim Ge(\theta)$	$Be(p,q)$	$Be(p+n,q+\Sigma x_i-n)$
$x \sim NeB(r,\theta)$	$Be(p,q)$	$Be(p+r,q+x-r)$
$x_1,\dots,x_n \sim N(\theta,1/\tau)$ (τ γνωστό)	$N(b,1/c)$	$N(\frac{cb+n\tau\bar{x}}{c+n\tau},\frac{1}{c+n\tau})$

Πίνακας 3.1

3.3 Ακατάλληλες A-Priori

Ας ξαναθυμηθούμε την a-posteriori ανάλυση στην οποία καταλήξαμε όταν εκτιμήσαμε τον μέσο κανονικής κατανομής με γνωστή διακύμανση, χρησιμοποιώντας κανονική a-priori κατανομή. Τώρα η βαρύτητα των a-priori πεποιθήσεων μας σχετικά με την παράμετρο θ, καθορίζονται από την διακύμανση ή ισοδύναμα από το c της κανονικής a-priori κατανομής. Μεγάλη τιμή του c αντιστοιχεί σε a-priori πεποιθήσεις με μεγάλη βαρύτητα, ενώ από την άλλη πλευρά, μικρές τιμές του c δείχνουν πολύ «λίγη» a-priori πληροφόρηση. Ας υποθέσουμε τώρα, ότι οι a-priori πληροφορίες μας σχετικά με την παράμετρο θ είναι τόσο λίγες, ώστε το $c \rightarrow 0$. Σε αυτήν την περίπτωση η a-posteriori κατανομή είναι $N(\bar{x},1/(n\tau)$ ή στην πιο γνωστή της μορφή $N(\bar{x},\frac{\sigma^2}{n})$. Κατ΄ αυτόν τον τρόπο βρήκαμε μία a-posteriori κατανομή η οποία έχει ισχύ, μέσα από αυτήν την διαδικασία που περιγράψαμε.

Όμως υπάρχει ένα σοβαρό ζήτημα. Ας σκεφτούμε τι παθαίνει η a-priori κατανομή καθώς το $c \to 0$. Αυτό που λαμβάνουμε είναι μία κανονική κατανομή $N(0,\infty)$, η οποία όμως δεν είναι αυθεντική κατανομή. Στην πραγματικότητα, αυτό το οποίο συμβαίνει είναι ότι καθώς το $c \to 0$, η κατανομή $N(b,1/c)$ γίνεται επίπεδη ενώ αυξάνει, με συνέπεια στο όριο $f(\theta) \propto 1$, $\theta \in R$, το οποίο όμως δεν μπορεί να είναι ισχύουσα συνάρτηση πυκνότητας πιθανότητας από την στιγμή που $\int_R f(\theta)d\theta = \infty$.

Αυτό που προκύπτει από τα παραπάνω είναι ότι η a-posteriori $N(\bar{x}, \frac{\sigma^2}{n})$ η οποία προήλθε καθώς το $c \to 0$ στην συνηθισμένη ανάλυση των συζυγών κατανομών, δεν μπορεί να προκύψει από την χρήση κάποιας κατάλληλης a-priori κατανομής. Αντίθετα μπορεί να προκύψει από την χρήση του $f(\theta) \propto 1$, το οποίο αποτελεί παράδειγμα του τι ακριβώς εννοούμε όταν λέμε ***ακατάλληλη a-priori κατανομή***.

Το ερώτημα το οποίο προκύπτει από τα όσα έχουν αναφερθεί παραπάνω, είναι εάν είναι επιτρεπτό να χρησιμοποιήσουμε μια a-posteriori κατανομή η οποία έχει προκύψει από τον καθορισμό μίας ακατάλληλης a-priori, για να δείξουμε ασαφή γνώση για κάποιο φαινόμενο; Παρόλο που υπάρχουν κάποιες ακόμα δυσκολίες (θα τις δούμε παρακάτω), γενικά η χρήση ακατάλληλων a-priori κατανομών θεωρείται αποδεκτή. Ένα σημείο στο οποίο στηρίζεται αυτή η θέση είναι ότι εάν επιλέξουμε c τέτοιο ώστε να παίρνει τιμή οποιαδήποτε εκτός από το 0, θα είχαμε καταλήξει σε μία αποδεκτή a-priori κατανομή και δεν θα υπήρχαν ενδοιασμοί για την υπόλοιπη ανάλυση μας. Θα μπορούσαμε αντίθετα, να είχαμε επιλέξει c με τιμή πολύ κοντά στο 0, και έτσι να υπολογίσουμε την a-posteriori κατανομή η οποία θα ήταν σχεδόν όμοια με αυτήν που βρήκαμε χρησιμοποιώντας την ακατάλληλη a-priori $f(\theta) \propto 1$.

3.4 Περιπτώσεις Άγνοιας

Στο προηγούμενο μέρος, είδαμε ότι εάν προσπαθήσουμε να αντιπροσωπεύσουμε την άγνοια μας μέσω των συνηθισμένων συζυγών κατανομών με κανονικό μέσο, τότε οδηγούμαστε σε ακατάλληλες a-priori κατανομές. Όμως υπάρχουν και άλλα σοβαρά προβλήματα εκτός από αυτό. Αν για παράδειγμα, καθορίσουμε μία a-priori για την παράμετρο θ της μορφής $f(\theta)\propto 1$, και θεωρήσουμε την παράμετρο $\varphi=\theta^2$, τότε

$$f(\phi) = f(\theta) \times \left|\frac{d\theta}{d\phi}\right| \propto \left(\sqrt{\phi}\right)^{-1}$$

Από την άλλη πλευρά, εάν δεν γνωρίζαμε τίποτα σχετικά με την παράμετρο θ, θα είχαμε άγνοια και για την παράμετρο φ αντίστοιχα, και έτσι μπορεί να κάναμε την υπόθεση ότι $f(\varphi)\propto 1$.

Μία συγκεκριμένη οπτική γωνία, είναι ότι ο καθορισμός της a-priori άγνοιας μας πρέπει να συνεπής στους 1-1 μετασχηματισμούς των παραμέτρων. Αυτό οδηγεί στην ιδέα των 'Jeffreys' priors' οι οποίες είναι βασισμένες στο σκεπτικό της πληροφορίας του Fisher

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

Η a-priori του Jeffreys ορίζεται ως

$$f_0(\theta) \propto |I(\theta)|^{1/2}.$$

Πρόταση: Η a-priori του Jeffreys είναι αμετάβλητη σε μετασχηματισμούς.

Απόδειξη: Αν $\varphi=g(\theta)$, πρέπει να αποδείξω ότι $f_0(\theta) = f_0(\phi)\left|\frac{\partial\phi}{\partial\theta}\right|$

Είναι: : $$\frac{\partial^2}{\partial\phi^2}[\log f] = \frac{\partial}{\partial\phi}\left[\frac{\partial \log f}{\partial\theta}\frac{\partial\theta}{\partial\phi}\right] = \frac{\partial^2 \log f}{\partial\theta^2}\left(\frac{\partial\theta}{\partial\phi}\right)^2 + \frac{\partial \log f}{\partial\theta}\frac{\partial^2\theta}{\partial\phi^2}.$$

Όμως, $$E_x\left[\frac{\partial \log f}{\partial\theta}\frac{\partial^2\theta}{\partial\phi^2}\right] = \frac{\partial^2\theta}{\partial\phi^2}E_x\left[\frac{\partial \log f}{\partial\theta}\right] = 0,$$

διότι, $$\int \left(\frac{\partial \log f}{\partial\theta} f\right)dx = \int \frac{1}{f}\frac{\partial f}{\partial\theta} f dx = \int \frac{\partial f}{\partial\theta}dx = \frac{\partial}{\partial\theta}1 = 0$$

Άρα $$f_0(\phi) = f_0(\theta)\cdot\left|\frac{\partial\theta}{\partial\phi}\right|$$

3.4.1 Παραδείγματα

✍ **Παράδειγμα 3.2** (Κανονικός μέσος) Ας υποθέσουμε ότι $x_1,\ldots x_n$ είναι ανεξάρτητες τυχαίες μεταβλητές οι οποίες κατανέμονται ως $N(\theta, \sigma^2)$ με σ^2 γνωστό. Τότε

$$f(x \mid \theta) \propto \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right\}$$

και

$$\log(f(x \mid \theta)) = -\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}$$

Ακόμα:

$$\begin{aligned} I(\theta) &= -E\left\{\frac{d^2 \log f(x \mid \theta)}{d\theta^2}\right\} \\ &= E\left(\frac{n}{\sigma^2}\right) \\ &= \frac{n}{\sigma^2} \end{aligned}$$

Συνεπώς $f_0(\theta) \propto 1$. Ας παρατηρήσουμε ότι χρησιμοποιήσαμε την συνολική πιθανοφάνεια εδώ. Αν ωστόσο είχαμε χρησιμοποιήσει την πιθανοφάνεια για *μία μόνο* παρατήρηση x, και λόγω της ανεξαρτησίας των x είχαμε $I_n(\theta)=nI_1(\theta)$, όπου τα I_1 και I_n είναι η πληροφορία από την 1η και την n-οστή ανεξάρτητη τιμή των x, αντίστοιχα. Θα μπορούσαμε επομένως να πάρουμε τις ίδιες a-priori του Jeffreys ανεξάρτητα από το πόσες παρατηρήσεις θα είχαμε.

✍ **Παράδειγμα 3.3** (Διωνυμικό δείγμα) Ας υποθέσουμε ότι $x|\theta \sim Bin(n,\theta)$. Τότε:

$$\log(f(x \mid \theta)) = x\log(\theta) + (n-x)\log(1-\theta), \qquad \text{έτσι} \qquad \text{ώστε}$$

$$\frac{d^2 \log f(x \mid \theta)}{d\theta^2} = \frac{-x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}$$

και επειδή E(x)=nθ,

$$I(\theta)=\frac{n\theta}{\theta^2}+\frac{(n-n\theta)}{(1-\theta)^2}$$

$$=n\theta^{-1}(1-\theta)^{-1},$$ η οποία οδηγεί στην

$$f_0(\theta)\propto\theta^{-1/2}(1-\theta)^{-1/2}$$

η οποία στην περίπτωση μας είναι κατάλληλη Βήτα(1/2,1/2) κατανομή.

3.5 Μείγματα A-Priori Κατανομών

Οι διαφορετικές απόψεις σχετικά με την χρήση των συζυγών οικογενειών των a-priori κατανομών, αναφέρθηκαν στην παράγραφο 3.1. Παρόλα αυτά θα δώσουμε και πάλι έμφαση στο γεγονός ότι τέτοιες οικογένειες κατανομών θα πρέπει να χρησιμοποιούνται μονάχα εάν ένα κατάλληλο μέλος της οικογένειας μπορεί να βρεθεί και το οποίο να είναι σύμφωνο με τις πραγματικές a-priori πεποιθήσεις μας. Σε μερικές περιπτώσεις η φυσική οικογένεια των συζυγών κατανομών μπορεί να είναι πολύ «αυστηρή» για να μπορεί να είναι δυνατά όσα αναφέρθηκαν.

Ας δούμε το επόμενο παράδειγμα. Όταν στρίβουμε ένα νόμισμα, τότε υπάρχει 0,5 πιθανότητα το αποτέλεσμα να είναι κεφάλι. Όμως, εάν ρίξουμε το νόμισμα επάνω σε ένα τραπέζι, τότε εξαιτίας κάποιων ατελιών που έχει το νόμισμα, υπάρχει μία τάση να έρχονται είτε κεφάλι είτε γράμματα. Λαμβάνοντας αυτήν την παράμετρο υπόψη μας, τότε μπορεί να επιλέξουμε να δώσουμε την πιθανότητα θ το νόμισμα να φέρει κεφάλι μία a-priori κατανομή η οποία έχει τάση να παίρνει τιμές κοντά στο 0.3 ή στο 0.7. Στην περίπτωση αυτή οι a-priori πεποιθήσεις μας μπορούν να αντιπροσωπευτούν από μία Διωνυμική κατανομή και η οποία μετά από n ρίψεις θα έχει την εξής μορφή: $X|\theta\sim Bin(n,\theta)$ και η συζυγής a-priori οικογένεια είναι η οικογένεια Βήτα. Παρόλα αυτά κανένα μέλος της οικογένειας δεν είναι πολυκόρυφο.

Μία λύση είναι να χρησιμοποιήσουμε μείγματα από συζυγείς κατανομές. Αυτή η μεγάλη οικογένεια κατανομών είναι μία συζυγής οικογένεια κατανομών για τον εξής λόγο.

Ας υποθέσουμε ότι $f_1(\theta),\dots,f_k(\theta)$ είναι όλες συζυγείς κατανομές του *θ*, και έχουν αντίστοιχες a-posteriori κατανομές $f_1(\theta|x),\dots,f_k(\theta|x)$. Τώρα ας θεωρήσουμε την οικογένεια του μείγματος των κατανομών: $f(\theta)=\sum_{i=1}^{k} p_i f_i(\theta)$.

Σε αυτήν την περίπτωση θα έχουμε:

$$\begin{aligned} f(\theta \mid x) &\propto f_i(\theta) f(x \mid \theta) \\ &= \sum_{i=1}^{k} p_i f_i(\theta) f(x \mid \theta) \\ &\propto \sum_{i=1}^{k} p_i^* f_i(\theta \mid x) \end{aligned}$$

και διαπιστώνουμε ότι η a-posteriori ανήκει στο ίδιο μείγμα-οικογένεια. Παρατηρούμε ωστόσο, ότι το μείγμα των αναλογιών στην a-posteriori p_i^* είναι γενικά διαφορετικό από αυτό που υπάρχει στην a-priori κατανομή.

Μπορεί να αποδειχθεί ότι το τελικό μείγμα από τις συζυγείς a-priori μπορεί να βρίσκεται πολύ κοντά σε οποιαδήποτε a-priori κατανομή. Ωστόσο, ο αριθμός των όρων στο μείγμα μπορεί να είναι μεγάλος και μπορεί να είναι δυνατόν να αντιπροσωπεύσει τις a-priori πεποιθήσεις κάποιου με περισσότερη συντομία χρησιμοποιώντας άλλες μη-συζυγείς οικογένειες μοντέλων.

3.6 Επιλογή των A-priori

Δεν υπάρχουν εύκολες απαντήσεις σε ερωτήματα όπως πόσο καλά μπορούμε να εξάγουμε a-priori πληροφορίες, αλλά ο λόγος για το οποίο θίγεται το συγκεκριμένο ερώτημα εδώ είναι γιατί αποτελεί ένα πολύ σημαντικό ζήτημα. Θα πρέπει να θυμηθούμε ωστόσο, ότι κανένας δεν είναι σε θέση να κάνει αρκετές ακριβείς κριτικές στις πιθανότητες. Για αυτόν τον λόγο, ο καθορισμός μίας a-priori κατανομής πρέπει να αντιμετωπίζεται σαν μία προσπάθεια για να αντιπροσωπευθούν οι πεποιθήσεις που έχει κάποιος, μέσα από μία ομοιόμορφη μαθηματική διαδικασία. Όπως έχουμε δει μέχρι τώρα, μία προσέγγιση είναι να πάρουμε μια συγκεκριμένη οικογένεια κατανομών (την οικογένεια των συζυγών κατανομών για παράδειγμα) να «ζητήσουμε» a-priori πληροφόρηση για μερικά αντιπροσωπευτικά στοιχεία της (πχ. για τον μέσο και για την διακύμανση), και έπειτα να επιλέξουμε σαν a-priori το μέλος εκείνο της συζυγούς οικογένειας με τις συγκεκριμένες τιμές (για τον μέσο και την διακύμανση). Γενικά πάντως, μπορεί να γίνει εξαιρετικά δύσκολο να προσαρμόσουμε τις a-priori πεποιθήσεις ενός ή περισσοτέρων ειδικών, σε μία a-priori κατανομή.

3.7 Ασκήσεις

Άσκηση 3.1 Αποδείξτε κάθε μία από τις συζυγείς κατανομές που δίνονται στον Πίνακα 3.1.

Άσκηση 3.2 Βρείτε την a-priori του Jeffreys για την παράμετρο θ, στο γεωμετρικό μοντέλο:

$$f(x \mid \theta) = (1-\theta)^{x-1}\theta \qquad x = 1,2,...$$

για κάθε ένα από τα τυχαία δείγματα $x_1,...,x_n$.

Άσκηση 3.3 Ας υποθέσουμε ότι το x~Pa(a,b) (κατανομή Pareto), όπου το a είναι γνωστό και το b άγνωστο. Τότε:

$$f(x \mid b) = ba^b x^{-b-1}, \qquad x > a.$$

Να βρείτε την a-priori του Jeffreys και την αντίστοιχη a-posteriori κατανομή για το b

Άσκηση 3.4 Επαναλάβετε το παράδειγμα 2.5 χρησιμοποιώντας κάποια a-priori κατανομή για το θ. Αλλάζουν οι a-posteriori κατανομές στις δύο περιπτώσεις; Υπακούει η στατιστική κατά Bayes στην αρχή της πιθανοφάνειας;

Άσκηση 3.5 Ένας διευθυντής μιας τράπεζας ανησυχεί για τον ρυθμό με τον οποίο οι υπάλληλοι στο ταμείο εξυπηρετούν τους πελάτες. Νομίζει ότι όλοι οι υπάλληλοι δουλεύουν με την ίδια περίπου ταχύτητα, που είναι 30, 40 ή 50 πελάτες την ώρα. Επιπλέον, 40 πελάτες την ώρα είναι δύο φορές πιο πιθανό από τις άλλες δύο τιμές, οι οποίες θεωρούνται ισοπίθανες. Για να αποκτήσει περισσότερη πληροφορία, ο διευθυντής παρατηρεί ότι σε δύο ώρες και οι πέντε υπάλληλοι εξυπηρετούν 350 πελάτες συνολικά. Χρησιμοποιείστε αυτή την νέα πληροφορία για να αναθεωρήσετε την a-priori κατανομή του διευθυντή για τον ρυθμό εξυπηρέτησης πελατών από τους υπαλλήλους του ταμείου. (Υπόδειξη: Θα βρείτε χρήσιμο να κάνετε πράξεις δουλεύοντας με το πηλίκο L(θ)/L(30) αντί απευθείας με το L(θ), όπου L είναι η πιθανοφάνεια.)

Άσκηση 3.6 Έστω ότι *θ* είναι ο ετήσιος τόκος μίας επένδυσης, εκφραζόμενος σε δεκαδική αντί της % μορφή. Σημειώστε ότι το *θ* μπορεί να πάρει αρνητικές τιμές εάν η επένδυση δεν είναι καλή. Έστω ακόμα ότι η αβεβαιότητα σας για το *θ* μπορεί να εκφραστεί με την μορφή της παρακάτω σ.π.π:

$$\Pi(\theta)=\begin{cases}100(\theta+0.1)/3 & \text{εάν} \quad -0.1\le\theta\le 0.1\\ 200(0.2-\theta)/3 & \text{εάν} \quad 0.1\le\theta\le 0.20\\ 0 & \text{αλλού}\end{cases}$$

(α) Σχεδιάστε την παραπάνω a-priori κατανομή και ερμηνεύστε τί συνεπάγεται για τις κρίσεις σας όσον αφορά το *θ*.

(β) Ένας αναλυτής επενδύσεων προσπαθεί να σας πείσει ότι πρόκειται για καλή επένδυση, και ισχυρίζεται ότι ο τόκος της επένδυσης θα είναι 0.15. Μεταχειριζόμενοι αυτόν τον ισχυρισμό σαν νέα πληροφορία, και συμβολίζοντας την με y, αποφασίζεται ότι η συνάρτηση πιθανοφάνειας (σαν συνάρτηση του *θ*) είναι

$$f(y\mid\theta)=\begin{cases}5 & \text{εάν} \quad 0\le\theta\ge 0.2\\ 0 & \text{αλλού}\end{cases}$$

Σχεδιάστε την συνάρτηση πιθανοφάνειας και σχολιάστε την επίδραση της στην αβεβαιότητα σας για το *θ*.

(γ) Με βάση την νέα πληροφορία y, αναθεωρήστε την κατανομή του *θ* και σχολιάστε την αλλαγή της.

Κεφάλαιο 4ο

Πολυπαραμετρικά Προβλήματα

Όλα τα παραδείγματα τα οποία εξετάσαμε μέχρι αυτό το σημείο, είχανε μία μονάχα παράμετρο, τον μέσο ή την διακύμανση του πληθυσμού. Τα περισσότερα στατιστικά προβλήματα περιλαμβάνουν στατιστικά μοντέλα τα οποία περιέχουν περισσότερες από μία άγνωστες παραμέτρους. Μπορεί να υπάρξει η περίπτωση όπου μονάχα μία από τις παραμέτρους να έχει ενδιαφέρον, αλλά συνήθως θα υπάρχουν και άλλες παράμετροι των οποίων οι τιμές θα είναι άγνωστες.

Η μέθοδος ανάλυσης των πολυπαραμετρικών προβλημάτων στην Μπευζιανή στατιστική είναι πολύ πιο άμεση (τουλάχιστον ως προς τις αρχές της), σε σχέση με αυτήν που χρησιμοποιείται στην κλασική στατιστική. Ωστόσο, δεν είναι απαραίτητη καμία επιπλέον γνώση εκτός από όσα έχουμε αναφέρει μέχρι στιγμής, για την κατανόηση της συγκεκριμένης μεθόδου. Στην περίπτωση των πολυπαραμετρικών προβλημάτων, έχουμε ένα διάνυσμα $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_d)$ από παραμέτρους για το οποίο θέλουμε να εξάγουμε κάποια συμπεράσματα. Καθορίζουμε μία a-priori (πολυμεταβλητή) κατανομή $f(\boldsymbol{\theta})$ για το διάνυσμα $\boldsymbol{\theta}$, και συνδυαζόμενο με την πιθανοφάνεια $f(x|\boldsymbol{\theta})$ μέσω του θεωρήματος του Bayes, παίρνουμε όπως και προηγουμένως:

$$f(\boldsymbol{\theta} \mid x) = \frac{f(\boldsymbol{\theta})f(x \mid \boldsymbol{\theta})}{\int f(\boldsymbol{\theta})f(x \mid \boldsymbol{\theta})d\boldsymbol{\theta}}$$

Φυσικά, η a-posteriori κατανομή θα είναι και αυτή τώρα μία πολυμεταβλητή κατανομή. Ωστόσο, η απλότητα της Μπευζιανής προσέγγισης τώρα, σημαίνει ότι η συμπερασματολογία για οποιαδήποτε υποομάδα παραμέτρων του διανύσματος $\boldsymbol{\theta}$ μπορεί να υπολογιστεί άμεσα με υπολογισμούς πιθανοτήτων στην μεταβλητή κατανομή. Για παράδειγμα, η περιθώρια a-posteriori κατανομή για το θ_1 μπορεί εύκολα να υπολογιστεί ολοκληρώνοντας ως προς όλες τις άλλες παραμέτρους του διανύσματος $\boldsymbol{\theta}$.

$$f(\theta_1 \mid x) = \int_{\theta_2} \ldots \int_{\theta_d} f(\theta \mid x) d\theta_2 \ldots d\theta_d$$

Παρόλο που δεν χρειάζεται καινούργια θεωρία για να γενικευτεί το πρόβλημα στις d-διαστάσεις, μία σειρά από πρακτικά προβλήματα δημιουργούνται:

1. **Καθορισμός των a-priori Κατανομών (A-priori Specification).** Οι a-priori κατανομές τώρα είναι πολυδιάστατες κατανομές. Αυτό σημαίνει ότι ο καθορισμός των a-priori κατανομών πρέπει να αντιπροσωπεύει τις a-priori πεποιθήσεις μας όχι απλά για κάθε μία παράμετρο ξεχωριστά, αλλά θα πρέπει να αντιπροσωπεύει και τις πεποιθήσεις μας σχετικά με την ανεξαρτησία ανάμεσα σε διαφορετικούς συνδυασμούς παραμέτρων (εάν μία παράμετρος θεωρηθεί μεγάλη, είναι δυνατόν κάποια άλλη παράμετρος να θεωρηθεί μικρή;). Η επιλογή κατάλληλων οικογενειών από a-priori κατανομές και ο συνοψισμός των a-priori πληροφοριών των ειδικών με αυτόν τον τρόπο, είναι αισθητά πιο πολύπλοκο πρόβλημα.
2. **Υπολογισμός.** Ακόμα και στα προβλήματα της μίας-διάστασης, είδαμε την χρησιμότητα των συζυγών οικογενειών κατανομών στην απλούστευση της a-posteriori ανάλυσης μέσα από το θεώρημα του Bayes. Με τα πολυδιάστατα προβλήματα, τα ολοκληρώματα γίνονται ακόμα δυσκολότερα για υπολογισμό. Αυτό κάνει την χρήση των συζυγών a-priori ακόμα πιο απαραίτητη, και φανερώνει την ανάγκη για υπολογιστικές τεχνικές, με σκοπό να βγάλουμε συμπεράσματα για το πότε η χρήση των συζυγών κατανομών είναι κατάλληλη και πότε είναι ακατάλληλη.
3. **Ερμηνεία.** Ολόκληρη η a-posteriori συμπερασματολογία, περιλαμβάνεται στην a-posteriori κατανομή, η οποία έχει ***τόσες διαστάσεις όσες και η παράμετρος θ***. Η δομή της a-posteriori κατανομής μπορεί να είναι ιδιαίτερα πολύπλοκη, και μπορεί να απαιτεί συγκεκριμένη υποδομή (για παράδειγμα έναν υπολογιστή με δυνατότητες παροχής καλών γραφικών) για να μπορέσει να δοθεί έμφαση στις πιο σημαντικές σχέσεις τις οποίες περιλαμβάνει.

Παρόλα τα πρακτικά προβλήματα, είναι πολύ σημαντικό να τονίσουμε το γεγονός για μία ακόμα φορά, ότι για τα πολυπαραμετρικά προβλήματα χρησιμοποιείται η ίδια θεωρία όπως και για τα προβλήματα μίας διάστασης. Το πλαίσιο της Μπευζιανής θεωρίας τονίζει ότι όλα τα συμπεράσματα πηγάζουν από τους βασικούς κανόνες των πιθανοτήτων.

4.1 Παραδείγματα

Παράδειγμα 4.1 Ας υποθέσουμε ότι ένα μηχάνημα μπορεί να χαρακτηριστεί είτε ως ικανοποιητικό ($x=1$) είτε ως μη ικανοποιητικό ($x=2$). Η πιθανότητα το μηχάνημα να χαρακτηριστεί ικανοποιητικό εξαρτάται από την θερμοκρασία δωματίου ($\theta_1=0$:κρύο, $\theta_1=1$:ζεστό) και την υγρασία ($\theta_2=0$:ξηρότητα, $\theta_2=1$: υγρασία). Οι πιθανότητες όταν $x=1$ δίνονται στον Πίνακα 4.1. Επιπλέον, η από κοινού a-priori κατανομή των (θ_1, θ_2), δίνεται στον Πίνακα 4.2.

$Pr(x=1 \mid \theta_1, \theta_2)$	$\theta_1=0$	$\theta_1=1$
$\theta_2=0$	0,6	0,8
$\theta_2=1$	0,7	0,6

Πίνακας 4.1

$Pr(\theta_1, \theta_2)$	$\theta_1=0$	$\theta_1=1$
$\theta_2=0$	0,3	0,2
$\theta_2=1$	0,2	0,3

Πίνακας 4.2

Η από κοινού a-posteriori κατανομή μπορεί να υπολογιστεί όπως στον Πίνακα 4.3.

		$\theta_1=0$	$\theta_1=1$
$Pr(x=1\mid \theta_1, \theta_2)\times Pr(\theta_1, \theta_2)$	$\theta_2=0$	0,18	0,16
	$\theta_2=1$	0,14	0,18
$Pr(\theta_1, \theta_2\mid x=1)$	$\theta_2=0$	18/66	16/66
	$\theta_2=1$	14/66	18/66

Πίνακας 4.3

Όμως, εάν αθροίσουμε κατά μήκος τις περιθώριες, τότε θα έχουμε τις περιθώριες a-posteriori κατανομές:

$Pr(\theta_1=0)=32/66$ $\qquad$ $Pr(\theta_1=1)=34/66$

και

$Pr(\theta_2=0)=34/66$ $\qquad$ $Pr(\theta_2=1)=32/66$.

Παράδειγμα 4.2 Ας υποθέσουμε ότι $Y_1 \sim Po(\alpha\beta)$ και $Y_2 \sim Po((1-\alpha)\beta)$ με τις Y_1 και Y_2 να είναι ανεξάρτητες και τα α,β να δίνονται. Τώρα, ας υποθέσουμε ότι οι a-priori πληροφορίες μας για τα α και β μπορούν να εκφραστούν ως: $\alpha \sim Be(p,q)$ και $\beta \sim Ga(p+q,1)$ με α,β ανεξάρτητα και p,q προκαθορισμένες παραμέτρους. (Σημειώστε ότι ο καθορισμός της a-priori ανεξαρτησίας, είναι ένα πολύ σημαντικό μέρος του a-priori καθορισμού των a-priori). Έτσι έχουμε την ακόλουθη πιθανοφάνεια:

$$f(y_1, y_2 \mid \alpha,\beta) = \frac{\exp(-\alpha\beta)(\alpha\beta)^{y_1}}{y_1!} \times \frac{\exp(-(1-\alpha)\beta)[(1-\alpha)\beta]^{y_2}}{y_2!}$$

και η a-priori

$$f(\alpha,\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \alpha^{p-1}(1-\alpha)^{q-1} \times \frac{e^{-\beta}\beta^{p+q-1}}{\Gamma(p+q)}.$$

Με βάση το θεώρημα του Bayes θα είναι:

$$f(\alpha,\beta \mid y_1, y_2) \propto e^{-\beta}\beta^{y_1+y_2}\alpha^{y_1}(1-\alpha)^{y_2}\alpha^{p-1}(1-\alpha)^{q-1}e^{-\beta}\beta^{q+p-1}$$
$$= \beta^{y_1+y_2+p+q-1}e^{-2\beta}\alpha^{y_1+p-1}(1-\alpha)^{y_2+q-1}$$

Αυτή είναι η από κοινού κατανομή για τα α και β η οποία περιλαμβάνει όλη την πληροφορία από την a-priori κατανομή και τα δεδομένα μας. Σε αυτήν την ειδική περίπτωση, η a-posteriori κατανομή παραγοντοποιείται σε συναρτήσεις των α και β. Οι περιθώριες a-posteriori κατανομές δίνονται από τους τύπους

$$f(\alpha \mid y_1, y_2) = \int_0^\infty f(\alpha,\beta \mid y_1, y_2)d\beta \propto \alpha^{y_1+p-1}(1-\alpha)^{y_2+q-1} \qquad \text{και}$$

$$f(\beta \mid y_1, y_2) = \int_0^1 f(\alpha,\beta \mid y_1, y_2)d\alpha \propto \beta^{y_1+y_2+p+q-1}e^{-2\beta}.$$

Από τις περιθώριες προκύπτει ότι η $\alpha|Y_1,Y_2 \sim Be(Y_1+p,\ Y_2+q)$ και $\beta|Y_1,Y_2 \sim Ga(Y_1+Y_2+p+q,2)$.

✍ **Παράδειγμα 4.3** (Κανονικός μέσος, άγνωστη διακύμανση) Έστω $X_1,\dots,X_n$ ανεξάρτητες τυχαίες μεταβλητές από κανονική κατανομή $N(\theta,\varphi)$, όπου και ο μέσος και η διακύμανση (μ,φ) αντίστοιχα, είναι άγνωστα.

Τότε:

$$f(x_i \mid \theta,\varphi) = \frac{1}{\sqrt{2\pi\varphi}}\exp\left\{-\frac{(x_i-\theta)^2}{2\varphi}\right\}$$

:

η οποία δίνει αντίστοιχη συνάρτηση πιθανοφάνειας:

$$L((\theta,\varphi);x) \propto \varphi^{-n/2}\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\varphi}\right\}$$

$$=\varphi^{-n/2}\exp\left\{-\frac{1}{2\varphi}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta)^2\right)\right\}$$

$$=\varphi^{-n/2}\exp\left\{-\frac{1}{2\varphi}(S^2+n(\bar{x}-\theta)^2)\right\} \qquad \text{όπου} \qquad S^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Τώρα είναι δυνατόν να κάνουμε μία πλήρη συζυγή ανάλυση στο μοντέλο αυτό, χρησιμοποιώντας μία συζυγή a-priori κατανομή η οποία είναι γνωστή ως: Κανονική-X^2 κατανομή. Οι αλγεβρικοί μετασχηματισμοί είναι σχετικά πολύπλοκοι, και για αυτό θα εξετάσουμε την πιο απλή από τις περιπτώσεις όπου η αναφορά a-priori αντιστοιχεί στην a-priori άγνοια μας.

Συγκεκριμένα έχουμε: $\quad p(\theta,\varphi) \propto \frac{1}{\varphi} \qquad -\infty<\theta<\infty, \quad 0<\varphi.$

Αυτή όμως είναι μία ακατάλληλη a-priori από την στιγμή που το ολοκλήρωμα αποτυγχάνει στον υπολογισμό. Τώρα, απλά χρησιμοποιώντας αυτήν την a-priori μέσα από το θεώρημα του Bayes παίρνουμε:

$$p(\theta,\varphi\,|\,x) \propto \varphi^{-n/2-1}\exp\left\{-\frac{1}{2\varphi}(S^2+n(\bar{x}-\theta)^2)\right\}$$

Ας προσέξουμε ότι από την στιγμή που δεν μπορούμε να παραγοντοποιήσουμε τηη παραπάνω παράσταση, οι a-posteriori περιθώριες κατανομές του θ και του φ δεν είναι ανεξάρτητες. Μπορούν να υπολογιστούν ως εξής. Για να απλουστεύσουμε τον τύπο μας θα θέσουμε:

$$A = S^2 + n(\bar{x}-\theta)^2$$

Τότε θα είναι:

$$p(\theta\,|\,x) \propto \int_0^{\infty}\varphi^{-\frac{n}{2}-1}\exp\{-\frac{A}{2\varphi}\}d\varphi.$$

Αλλάζοντας τις μεταβλητές σε u=A/(2φ), θα έχουμε:

$$p(\theta \mid x) \propto \int_0^\infty \varphi^{-\frac{n}{2}-1} e^{-u} \frac{2\varphi^2}{A} du$$

$$= \left(\frac{2}{A}\right)^{n/2} \int_0^\infty u^{n/2-1} e^{-u} du$$

$$= \left(\frac{2}{A}\right)^{n/2} \Gamma(n/2)$$

Έτσι έχουμε: $p(\theta \mid x) \propto A^{-n/2} = \{S^2 + n(\bar{x}-\theta)^2\}^{-n/2}$.

Αυτή δεν είναι η συνηθισμένη μορφή, αλλά αν θέσουμε: $t = \dfrac{\theta - \bar{x}}{s/\sqrt{n}}$ όπου $s^2 = \dfrac{S^2}{n-1}$

τότε:

$$p(t \mid x) \propto \{(n-1)s^2 + (st)^2\}^{-n/2}$$

$$\propto \{1 + \frac{t^2}{n-1}\}^{-n/2}.$$

Αυτή είναι η συνάρτηση πυκνότητας πιθανότητας της t-κατανομής με n-1 βαθμούς ελευθερίας.

Κατ΄αντιστοιχία:

$$p(\phi \mid x) \propto \int_{-\infty}^{\infty} \phi^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\phi}(S^2 + n(\bar{x}-\theta)^2)\right\} d\theta$$

$$\propto \phi^{-(n+1)/2} \exp\{-S^2/(2\phi)\} \int_{-\infty}^{\infty} (2\pi\phi/n)^{-1/2} \exp\{-\frac{n}{2\phi}(\theta-\bar{x})^2\} d\theta$$

$$= \phi^{-(n+1)/2} \exp\{-S^2/(2\phi)\}$$

και $S^2/\varphi \sim X^2_{n-1}$.

Κάνοντας μία περίληψη των όσων αναφέραμε μέχρι εδώ, για ένα μοντέλο που ακολουθεί Κανονική κατανομή με άγνωστο μέσο και διακύμανση, λαμβάνοντας την a-priori $p(\theta,\varphi) \propto 1/\varphi$, έχουμε :

$$t = \frac{\theta - \bar{x}}{s/\sqrt{n}} \sim t_{n-1} \quad \text{και}$$

$$S^2/\varphi \sim X^2_{n-1}$$

4.2 Ασκήσεις

Άσκηση 4.1 Η ποιότητα ενός ηλεκτρικού λαμπτήρα μπορεί να χαρακτηριστεί ως άριστη (x=1), ικανοποιητική (x=2) ή κακή (x=3). Η πιθανότητα ο ηλεκτρικός λαμπτήρας να ανήκει σε μία από τις παραπάνω κατηγορίες εξαρτάται από το εργοστάσιο κατασκευής του (θ1=0: εργοστάσιο A, θ1=1: εργοστάσιο B) και από τον τύπο του μηχανήματος που τους παράγει (θ2=0: μηχάνημα 1, θ2=1: μηχάνημα 2 και θ2=2: μηχάνημα 3). Οι πιθανότητες για x=3 δίνονται στον Πίνακα 4.4. Επιπλέον, η από κοινού a-priori κατανομή των (θ1,θ2) δίνεται στον Πίνακα 4.5. Βρείτε την από κοινού a-posteriori κατανομή του θ1,θ2|x=3 και κάθε μία από τις περιθώριες κατανομές. Έχοντας παρατηρήσει την περίπτωση x=3 είσαστε σε θέση να πείτε ποιός συνδυασμός εργοστασίου/μηχανήματος είναι ο πιο πιθανός να έχει παράγει τον λαμπτήρα;

Pr(x=3|θ1,θ2)	θ1=0	θ1=1
θ2=0	0,2	0,3
θ2=1	0,4	0,1
θ2=2	0,5	0,2

Πίνακας 4.4 Υπό συνθήκη Πιθανότητες όταν x=3

Pr(x=3|θ1,θ2)	θ1=0	θ1=1
θ2=0	0,1	0,2
θ2=1	0,2	0,3
θ2=2	0,1	0,1

Πίνακας 4.5 A-priori πιθανότητες των συνδυασμών μηχανήματος/εργοστασίου

Άσκηση 4.2 Ένα τυχαίο δείγμα $x_1,x_2,\ldots,x_n$ συλλέγεται από μία κανονική κατανομή $N(\mu,\tau^{-1})$, όπου τα μ και τ είναι άγνωστα και δεν έχουμε καθόλου a-priori πληροφορία για αυτά. Αποδείξτε ότι a-posteriori τα $\sqrt{n}(\mu-\bar{x})/s$ και $(n-1)s^2\tau$ έχουν περιθώριες κατανομές t_{n-1} και X^2_{n-1} αντίστοιχα. Τα $\bar{x}$ και s^2 δίνονται από τις σχέσεις

$$\bar{x}=\frac{1}{n}\sum_i x_i, \qquad s^2=\sum_i (x_i-\bar{x})^2/n-1.$$

Κεφάλαιο 5

Συνοψίζοντας την a-posteriori Πληροφόρηση

5.1 Θεωρία Αποφάσεων

Τα πάντα στη ζωή μας είναι θέμα αποφάσεων. Από τα πιο απλά καθημερινά πράγματα, όπως το ποια ρούχα θα φορέσουμε πριν φύγουμε από το σπίτι μας, έως τα πλέον σοβαρά όπως το αν θα κάνουμε μια χειρουργική επέμβαση (που εγκυμονεί κινδύνους), για ένα πρόβλημα υγείας που αντιμετωπίζουμε, στηρίζονται σε προσωπικές αποφάσεις. Τις περισσότερες φορές αποφασίζουμε υποσυνείδητα (τουλάχιστον όσον αφορά τα τετριμμένα). Σε αρκετές όμως περιπτώσεις είμαστε ιδιαίτερα προσεκτικοί αναλογιζόμενοι τις συνέπειες που μπορεί να υπάρξουν από μια "λάθος" απόφαση.

Στη στατιστική, η θεωρία αποφάσεων αποτελεί μια από τις πλέον σημαντικές ενότητες. Στο κεφάλαιο αυτό, αφού παραθέσουμε τις βασικές αρχές της, θα ασχοληθούμε με τη συμπερασματολογία (σημειακή εκτίμηση, εκτίμηση σε διάστημα και ελέγχους υποθέσεων) για την παράμετρο $\boldsymbol{\theta}$ που μας ενδιαφέρει, υπό το πρίσμα της θεωρίας αποφάσεων. Υπάρχουν πολλές προσεγγίσεις στην θεωρία αποφάσεων, αλλά η βασικότερη στηρίζεται στην ανάλυση κατά Bayes. Επιπροσθέτως αποδεικνύεται ότι αν για ένα συγκεκριμένο πρόβλημα υπάρχει "βέλτιστη" απόφαση (βάση κάποιων κριτηρίων που θα έχουν οριστεί εκ των προτέρων) τότε η Μπεϋζιανή προσέγγιση σίγουρα μπορεί να οδηγήσει στη λύση αυτή.

Τα βασικά στοιχεία που απαιτούνται για να κατασκευαστεί ένα πρόβλημα απόφασης είναι τα ακόλουθα:

1. Ο παραμετρικός χώρος Θ, ο οποίος περιέχει όλες τις πιθανές τιμές της παραμέτρου $\boldsymbol{\theta}$ (την πραγματική τιμή της οποίας δεν γνωρίζουμε) για την οποία ενδιαφερόμαστε να λάβουμε κάποια απόφαση.
2. Ένα σύνολο $\mathcal{A}$, από δυνατές ενέργειες/πράξεις (actions) a, οι οποίες είναι διαθέσιμες στο άτομο το οποίο θα λάβει την απόφαση.
3. Μία **συνάρτηση απώλειας** (Loss function), $L(\boldsymbol{\theta}, a) : \Theta \times \mathcal{A} \rightarrow \Re$, που εκφράζει την απώλεια που έχουμε όταν επιλέγουμε την πράξη a, την στιγμή που η πραγματική τιμή της αγνώστου παραμέτρου είναι $\boldsymbol{\theta}$.

Σημείωση: Όσον αφορά τα σύνολα Θ και $\mathcal{A}$ μπορεί να είναι διακριτά ή συνεχή, πεπερασμένα ή άπειρα. Σχετικά με τη συνάρτηση απώλειας μπορεί να πάρει και αρνητικές τιμές υποδηλώνοντας όφελος παρά απώλεια. Αρκετοί συγγραφείς (DeGroot(1970), Berger(1985) κ.α.) χρησιμοποιούν τη συνάρτηση χρησιμότητας (Utility function) αντί της συνάρτησης απώλειας, η οποία μπορεί να θεωρηθεί ως $-L(\boldsymbol{\theta}, a)$, και συναντάται πολύ συχνά στην οικονομική θεωρία.

Πριν προχωρήσουμε όμως στη λήψη οποιασδήποτε απόφασης έχουμε στη διάθεσή μας το αποτέλεσμα ενός πειράματος $\mathbf{X}$, το οποίο προέρχεται από την κατανομή $f(\mathbf{X}|\boldsymbol{\theta})$.

Ορισμός 1 *Η τριάδα (Θ, $\mathcal{A}$, $L(\boldsymbol{\theta}, a)$) συνοδευόμενη από την τυχαία παρατήρηση $\mathbf{X}$, η οποία προέρχεται από την κατανομή $f(\mathbf{X}|\boldsymbol{\theta})$, όπου $\boldsymbol{\theta} \in \Theta$ είναι η άγνωστη τιμή της παραμέτρου που μας ενδιαφέρει, ορίζουν ένα "***Στατιστικό Πρόβλημα Απόφασης***".*

Έχοντας λοιπόν παρατηρήσει $\mathbf{X} = \mathbf{x}$, καλούμαι να πάρω μια απόφαση για την ενέργεια, a, στην οποία θα προβώ. Αν με $\mathcal{X}$ συμβολίσουμε το δειγματικό χώρο όλων των πιθανών αποτελεσμάτων του πειράματος τότε ορίζουμε:

Ορισμός 2 **Κανόνα απόφασης** *(Decision rule) την συνάρτηση $\delta(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{A}$, η οποία μου υποδεικνύει ποια ενέργεια $\delta(\mathbf{x}) \in \mathcal{A}$ θα επιλέξω όταν παρατηρώ $\mathbf{X} = \mathbf{x}$, ως αποτέλεσμα του τυχαίου πειράματος.*

Ο τρόπος ορισμού του κανόνα απόφασης είναι ντετερμινιστικός, με την έννοια οτι γνωρίζουμε σε ποια ακριβώς ενέργεια θα προβούμε όταν παρατηρήσουμε ένα συγκεκριμένο αποτέλεσμα του πειράματος. Για τον λόγο αυτό οι συναρτήσεις $\delta(\mathbf{x})$ στην βιβλιογραφία (για παράδειγμα Ferguson(1967)) συχνά αναφέρονται ως "μή-τυχαιοποιημένοι κανόνες απόφασης" (non-randomised decision rules). Θα συμβολίζουμε με $\mathcal{D}$ το σύνολο όλων των δυνατών κανόνων απόφασης $\delta(\mathbf{x})$. Δανειζόμενοι ορολογία από την θεωρία παιγνίων, το σύνολο $\mathcal{D}$ δεν είναι τίποτε άλλο από ένα σύνολο διαφορετικών 'στρατηγικών'. Όπως αντιλαμβανόμαστε εύκολα, κύριος σκοπός μας είναι να εντοπίσουμε εκείνη τη στρατηγική η οποία θα έχει το ελάχιστο δυνατό κόστος. Επειδή όμως η συνάρτηση απώλειας $L(\boldsymbol{\theta},\delta(\mathbf{x}))$ είναι τυχαία ποσότητα, προκειμένου να συγκρινουμε διαφορετικές στρατηγικές θα χρησιμοποιήσουμε την μέση απώλεια. Οδηγούμαστε λοιπόν στους ακόλουθους ορισμούς συναρτήσεων κινδύνου:

Ορισμός 3 *Ορίζουμε ως* **Κλασσικό Κίνδυνο** *(Frequentist Risk) ενός δεδομένου κανόνα αποφάσεως δ, τη συνάρτηση* $FR(\ .\ ,\delta) : \Theta \to \Re$, *η οποία εκφράζει τη μέση (αναμενόμενη) απώλεια ως προς την πιθανοφάνεια* $f(\mathbf{x}|\boldsymbol{\theta})$*:*

$$FR(\boldsymbol{\theta},\delta(\mathbf{x})) = E_{\mathbf{x}|\boldsymbol{\theta}}\left[L(\boldsymbol{\theta},\delta(\mathbf{x}))\right] = \int L(\boldsymbol{\theta},\delta(\mathbf{x}))f(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$$

Η λέξη 'κλασικός' στον παραπάνω ορισμό παραπέμπει στην κλασική προσέγγιση στη Στατιστική (γι' αυτό το λόγο άλλωστε η μέση τιμή έχει ληφθεί ως προς την συνάρτηση πιθανοφάνειας). Με βάση όμως την κατά Bayes προσέγγιση, η άγνωστη παράμετρος $\boldsymbol{\theta}$ θα ακολουθεί μια prior κατανομή $\pi(\boldsymbol{\theta})$. Έχοντας παρατηρήσει το αποτέλεσμα του πειράματος $f(\mathbf{x}|\boldsymbol{\theta})$ το θεώρημα του Bayes μας εξασφαλίζει την a-posteriori κατανομή $p(\boldsymbol{\theta}|\mathbf{x})$ η οποία ως γνωστό περιλαμβάνει όλη τη δυνατή πληροφορία για την άγνωστη παράμετρο $\boldsymbol{\theta}$. Με βάση την κατά Bayes προσέγγιση λοιπόν:

Ορισμός 4 *Ορίζουμε ως* **Εκ των Υστέρων Κίνδυνο** *(Posterior Risk) ενός δεδομένου κανόνα αποφάσεως δ και μιας συγκεκριμένης prior κατανομής* $\pi(\boldsymbol{\theta})$*, τη συνάρτηση* $PR(\pi(\boldsymbol{\theta}),\delta(.)) : \mathcal{X} \to \Re$, *η οποία εκφράζει τη μέση (αναμενόμενη) a-posteriori απώλεια:*

$$PR(\pi(\boldsymbol{\theta}),\delta(\mathbf{x})) = E_{\boldsymbol{\theta}|\mathbf{x}}\left[L(\boldsymbol{\theta},\delta(\mathbf{x}))\right] = \int L(\boldsymbol{\theta},\delta(\mathbf{x}))p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$$

Η σύγκριση δύο κανόνων απόφασης $\delta_1(.)$ και $\delta_2(.)$ για το ποιος είναι προτιμότερος (κοινώς ποιος εκ των δύο έχει το μικρότερο ρίσκο), μπορεί να γίνει μέσω των συναρτήσεων κινδύνων που δώσαμε στους δύο τελευταίους ορισμούς.

Έτσι με βάση την κλασική προσέγγιση αν $\boldsymbol{\theta} = \boldsymbol{\theta_1}$ τότε η $\delta_1(.)$ είναι προτιμότερη της $\delta_2(.)$ εφόσον: $FR(\boldsymbol{\theta_1},\delta_1(\mathbf{x})) \leq FR(\boldsymbol{\theta_1},\delta_2(\mathbf{x}))$. Φυσικά αν υπάρχει κανόνας απόφασης

$\delta^*(.)$, ο οποίος είναι ομοιόμορφα προτιμότερος από κάθε άλλο κανόνα $\delta(.)$, κοινώς: $FR(\boldsymbol{\theta}, \delta^*(\mathrm{x})) \leq FR(\boldsymbol{\theta}, \delta(\mathrm{x}))$, $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ και $\forall\ \delta(.) \in \mathcal{D}$ τότε ο κανόνας $\delta^*(.)$ αποκαλείται **παραδεκτός** (admissible).

Από την άλλη πλευρά βάση της κατά Bayes προσέγγισης όταν παρατηρούμε το $\mathbf{X} = \mathbf{x}$ τότε η $\delta_1(.)$ είναι προτιμότερη της $\delta_2(.)$ εφόσον: $PR(\pi(\boldsymbol{\theta}),\delta_1(.)) \leq PR(\pi(\boldsymbol{\theta}),\delta_2(.))$

Σχόλιο: Αξίζει στο σημείο αυτό να σχολιάσουμε τον τρόπο με τον οποίο οι δύο προσεγγίσεις 'αξιολογούν' διαφορετικούς κανόνες απόφασης. Από την πλευρά της κλασικής στατιστικής προκειμένου να υπολογίσουμε τη συνάρτηση κινδύνου μιας απόφασης είναι απαραίτητο να γνωρίζουμε την τιμή του $\boldsymbol{\theta}$ (!). Από την πλευρά της κατά Bayes προσέγγισης απαιτείται το αποτέλεσμα του τυχαίου πειράματος $\mathbf{x}$.

Διαισθητικά αντιλαμβανόμαστε ότι ένας παραδεκτός κανόνας απόφασης πρέπει να είναι (πάντα) ο καλύτερος σε όλες τις δυνατές περιπτώσεις (καλύτερος για όλα τα δυνατά $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Κάτι τέτοιο όμως συμβαίνει συνήθως μόνον σε τετριμμένα προβλήματα. Το πλέον πιθανό (και για τις δύο διαφορετικές προσεγγίσεις) είναι ότι ένας κανόνας απόφασης θα έχει μικρότερη τιμή της συνάρτησης κινδύνου (συγκρινόμενος με κάποιον άλλο κανόνα απόφασης) για κάποιες τιμές του πεδίου ορισμού και μεγαλύτερη για κάποιες άλλες.

Είναι απαραίτητο λοιπόν να ορίσουμε κάποια κριτήρια κατάταξης των κανόνων απόφασης, με βάση τα οποία θα μπορούμε να αξιολογούμε διαφορετικούς κανόνες απόφασης, ανεξαρτήτως $\boldsymbol{\theta}$ (για την κλασική προσέγγιση) και ανεξαρτήτως $\mathbf{x}$ (για την κατά Bayes προσέγγιση). Θα ορίσουμε δύο τέτοια κριτήρια: τον κίνδυνο κατά Bayes και το κριτήριο Minimax.

5.1.1 Κίνδυνος κατά Bayes

Αν και με βάση την κλασική προσέγγιση η άγνωστη παράμετρος δεν ακολουθεί κατανομή, προκειμένου να καταστεί δυνατή η σύγκριση διαφορετικών αποφάσεων, οι κλασικοί στατιστικοί ορίζουν την συνάρτηση $\pi(\boldsymbol{\theta})$, των πιθανών τιμών της αγνώστου παραμέτρου $\boldsymbol{\theta}$. Έτσι:

Ορισμός 5 *Ορίζουμε ως κίνδυνο κατά Bayes $BR(\pi(\boldsymbol{\theta}), \delta(\mathrm{x}))$ για τον κανόνα απόφασης $\delta(.)$ με βάση την κατανομή $\pi(\boldsymbol{\theta})$, τον μέσο (αναμενόμενο) κλασικό κίνδυνο:*

$$BR(\pi(\boldsymbol{\theta}), \delta(\mathrm{x})) = E_{\boldsymbol{\theta}}\left[FR(\boldsymbol{\theta}, \delta(\mathrm{x}))\right] = \int FR(\boldsymbol{\theta}, \delta(\mathrm{x}))\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Λήμμα 6 *Ο κίνδυνος κατά Bayes $BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))$ είναι ο μέσος (αναμενόμενος) εκ των υστέρων κίνδυνος ως προς την μή-δεσμευμένη κατανομή του* $\mathbf{X}$*:*

$$BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x})) = E_{\mathbf{X}}\left[PR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))\right] = \int PR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))f(\mathbf{x})d\mathbf{x}$$

Απόδειξη:

$$\begin{aligned}
BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x})) &= E_{\boldsymbol{\theta}}\left[FR(\boldsymbol{\theta}, \delta(\mathbf{x}))\right] \\
&= E_{\boldsymbol{\theta}}\left[E_{\mathbf{X}|\boldsymbol{\theta}}\left[L(\boldsymbol{\theta}, \delta(\mathbf{x}))\right]\right] \\
&= E_{\boldsymbol{\theta}}\left[\int_{\mathcal{X}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}\right] \\
&= \int_{\boldsymbol{\Theta}}\left[\int_{\mathcal{X}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}\right]\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\Theta}}\int_{\mathcal{X}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\mathbf{x}d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\Theta}}\int_{\mathcal{X}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))f(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x}d\boldsymbol{\theta} \\
&= \int_{\mathcal{X}}\int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))p(\boldsymbol{\theta}|\mathbf{x})f(\mathbf{x})d\boldsymbol{\theta}d\mathbf{x} \\
&= \int_{\mathcal{X}}\left[\int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}, \delta(\mathbf{x}))p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}\right]f(\mathbf{x})d\mathbf{x} \\
&= \int_{\mathcal{X}}\left[E_{\boldsymbol{\theta}|\mathbf{X}}\left[L(\boldsymbol{\theta}, \delta(\mathbf{x}))\right]\right]f(\mathbf{x})d\mathbf{x} \\
&= \int_{\mathcal{X}} PR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))f(\mathbf{x})d\mathbf{x}
\end{aligned}$$

Η αλλαγή στη σειρά ολοκλήρωσης (7 ισότητα) καθίσταται δυνατή κάτω από (χαλαρές) υποθέσεις του θεωρήματος Fubini (Chung (1974))

Παρατηρούμε λοιπόν ότι ο κίνδυνος κατά Bayes κατά κάποιο τρόπο 'ενοποιεί' τις δύο διαφορετικές προσεγγίσεις συμπηκνώνοντας όλη την πληροφορία που έχουμε για το κόστος ενός κανόνα απόφασης σε έναν αριθμό (τον $BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))$). Βάση λοιπόν του κατά Bayes κινδύνου κριτηρίου θα πρέπει να επιλέγουμε τον κανόνα απόφασης, ο οποίος τον ελαχιστοποιεί.

Ορισμός 7 *Ο κανόνας $\delta_\pi(.)$, ο οποίος ελαχιστοποιεί την ποσότητα $BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))$, δηλαδή:*

$$\delta_\pi(.) = \min_{\delta \in \mathcal{D}} BR(\pi(\boldsymbol{\theta}), \delta(\mathbf{x}))$$

αποκαλείται **κανόνας του Bayes** *(Bayes rule), ως προς την prior $\pi(\boldsymbol{\theta})$.*

Αξίζει στο σημείο αυτό να αναφέρουμε ότι είναι δυνατόν ο κανόνας του Bayes είτε να μην υπάρχει, είτε να μην είναι μοναδικός (όπως μπορεί να συμβεί με το ελάχιστο μιας οποιασδήποτε πραγματικής συνάρτησης).

5.1.2 Κανόνας Minimax

Το κριτήριο minimax ιεραρχεί τους κανόνες απόφασης $\delta(.)$ βάση του χειρότερου κόστους που μπορούν να μας προκαλέσουν. Έτσι βάση αυτού του κριτηρίου, μεταξύ των κανόνων απόφασης $\delta_1(.)$ και $\delta_2(.)$ προτιμότερος είναι ο $\delta_1(.)$ εάν: $\sup_{\boldsymbol{\theta}} FR(\boldsymbol{\theta}, \delta_1(\mathbf{x})) < \sup_{\boldsymbol{\theta}} FR(\boldsymbol{\theta}, \delta_2(\mathbf{x}))$. Έχουμε λοιπόν:

Ορισμός 8 *Ο κανόνας απόφασης $\delta^*(.)$ καλείται* **κανόνας minimax** *εάν:*

$$\sup_{\boldsymbol{\theta} \in \Theta} FR(\boldsymbol{\theta}, \delta^*(.)) = \inf_{\delta \in \mathcal{D}} \left[\sup_{\boldsymbol{\theta} \in \Theta} FR(\boldsymbol{\theta}, \delta(.)) \right]$$

Αν στον παραπάνω ορισμό αντικαταστήσουμε το $\inf$ με $\min$ και το $\sup$ με $\max$ τότε καταλαβαίνουμε πως προέκυψε το όνομα minimax.

Στο σημείο αυτό ας δούμε πως εφαρμόζονται οι θεωρίες που μας έχουν απασχολήσει στην πράξη.

Παράδειγμα 5.1

Ένας φορέας της δημόσιας υγείας, ψάχνει για έναν τρόπο εμβολιασμού μίας ήπιας ασθένειας, η οποία όμως προκαλεί σοβαρές απουσίες στον εργασιακό χώρο. Οι έρευνες που έγιναν για το θέμα αυτό, έδειξαν ότι το 40% του πληθυσμού είναι ήδη κάτω από σοβαρό κίνδυνο, αλλά τα τεστ τα οποία μπορούν να δείξουν πόσο ευάλωτος είναι κάποιος στην ασθένεια αυτή, στοιχίζουν πάρα πολύ για να χρησιμοποιηθούν στο σύνολο του πληθυσμού. Ένα απλό δερματικό τεστ δημιουργήθηκε, αλλά δεν είναι εντελώς αξιόπιστο. Οι πιθανότητες αντίδρασης φαίνονται στον ακόλουθο πίνακα.

	ΗΠΙΑ ΑΝΤΙΔΡΑΣΗ	ΣΟΒΑΡΗ ΑΝΤΙΔΡΑΣΗ
ΕΧΕΙ ΑΝΟΣΙΑ	0.65	0.35
ΔΕΝ ΕΧΕΙ ΑΝΟΣΙΑ	0.25	0.75

Έχει εκτιμηθεί ότι η αντίστοιχη αμοιβή των χαμένων ανθρωποωρών από την αποτυχία να εμβολιάσουν ένα ευάλωτο άτομο είναι 20, το περιττό κόστος του εμβολιασμού ενός ατόμου που δεν το χρειάζεται είναι 8, και δεν υπάρχει κόστος στην περίπτωση που εμβολιάζεται ένα ευάλωτο άτομο, ή αποτυγχάνει να εμβολιαστεί ένα άτομο που δεν είναι ευάλωτο στην ασθένεια αυτή. Ποια πολιτική θα πρέπει να ακολουθήσουμε σχετικά με τον εμβολιασμό ή όχι του πληθυσμού;

Λύση: Ας ξεκινήσουμε ορίζοντας ένα στατιστικό πρόβλημα απόφασης. Ο παραμετρικός χώρος ορίζεται ως $\Theta = \{\theta_1, \theta_2\}$, όπου θ_1 εκφράζει την περίπτωση ανοσίας και θ_2 την έλλειψη ανοσίας. Οι a-priori πιθανότητες των δύο διαφορετικών καταστάσεων του παραμετρικού χώρου έχουν υπολογιστεί και είναι: $\pi(\theta_1) = 0.6$ και $\pi(\theta_2) = 0.4$. Οι δυνατές ενέργειες είναι a_1=εμβολιάζω και a_2=δεν εμβολιάζω, άρα $\mathcal{A} = \{a_1, a_2\}$. Η συνάρτηση απώλειας $L(\theta, a)$ έχει υπολογιστεί:

$L(\theta,a)$	a_1	a_2
θ_1	8	0
θ_2	0	20

Πριν όμως αποφασίσουμε για τον εάν θα προβούμε σε εμβολιασμό ή όχι τα αποτελέσματα ενός δερματικού τεστ μας είναι διαθέσιμα. Αν ονομάσουμε x_1 την περίπτωση της ήπιας αντίδρασης και x_2 της σοβαρής αντίδρασης στο δερματικό τεστ τότε έχουμε $\mathcal{X} = \{x_1, x_2\}$ και η δεσμευμένες πιθανότητες του πειράματος δίνονται από:

$f(x\|\theta)$	x_1	x_2
θ_1	0.65	0.35
θ_2	0.25	0.75

Το σύνολο όλων των δυνατών αποφάσεων, $\delta(x) : \mathcal{X} \to \mathcal{A}$, στο συγκεκριμένο πρόβλημα είναι, $\mathcal{D} = \{\delta_1, \delta_2, \delta_3, \delta_4\}$, όπου:

$\delta_1(x_1) = a_1$, $\delta_1(x_2) = a_1$ αγνοούμε το τεστ και εμβολιάζουμε όλους

$\delta_2(x_1) = a_1$, $\delta_2(x_2) = a_2$ εμβολιάζουμε μόνον όσους έχουν ήπια αντίδραση στο τεστ

$\delta_3(x_1) = a_2$, $\delta_3(x_2) = a_1$ εμβολιάζουμε μόνον όσους έχουν σοβαρή αντίδραση στο τεστ

$\delta_4(x_1) = a_2$, $\delta_4(x_2) = a_2$ αγνοούμε το τεστ και δεν εμβολιάζουμε κανέναν

Πρέπει λοιπόν να βρούμε ποιος είναι ο προτιμότερος κανόνας απόφασης $\delta_i(x)$, κοινώς ποια από τις προαναφερθέντες 4 πολιτικές έχει το μικρότερο αναμενόμενο κόστος.

Κλασικός κίνδυνος: $FR(\ .\ , \delta_i(x)) : \mathbf{\Theta} \rightarrow \Re$. Για τον κανόνα απόφασης $\delta_1(x)$ έχουμε:

$$\begin{aligned} FR(\theta_1, \delta_1(x)) &= E_{x|\theta_1}\left[L(\theta_1, \delta_1(x))\right] = \sum_{i=1}^{2} L(\theta_1, \delta_1(x_i))f(x_i|\theta_1) \\ &= L(\theta_1, \delta_1(x_1))f(x_1|\theta_1) + L(\theta_1, \delta_1(x_2))f(x_2|\theta_1) \\ &= 8 \times .65 + 8 \times .35 = 8 \end{aligned}$$

και

$$\begin{aligned} FR(\theta_2, \delta_1(x)) &= E_{x|\theta_2}\left[L(\theta_2, \delta_1(x))\right] = \sum_{i=1}^{2} L(\theta_2, \delta_1(x_i))f(x_i|\theta_2) \\ &= L(\theta_2, \delta_1(x_1))f(x_1|\theta_2) + L(\theta_2, \delta_1(x_2))f(x_2|\theta_2) \\ &= 0 \times .25 + 0 \times .75 = 0 \end{aligned}$$

Ομοίως μπορούμε να υπολογίσουμε τον αναμενόμενο κλασικό κίνδυνο και για τις άλλες τρεις πολιτικές. Έχουμε λοιπόν:

$FR(\theta_i, \delta_j(x))$	$\delta_1(x)$	$\delta_2(x)$	$\delta_3(x)$	$\delta_4(x)$
$\theta = \theta_1$	8	5.2	2.8	0
$\theta = \theta_2$	0	15	5	20

Από τον παραπάνω πίνακα προκύπτει ότι βάση κλασικού κινδύνου δεν υπάρχει παραδεκτή (admissible) πολιτική, αφού εάν $\theta = \theta_1$ η βέλτιστη πολιτική είναι η $\delta_1(x)$, ενώ όταν $\theta = \theta_2$ βέλτιστη είναι η $\delta_4(x)$. Εφόσον δεν υπάρχει παραδεκτός κανόνας απόφασης προκειμένου να αποφασίσουμε ποια είναι η πιο συμφέρουσα πολιτική θα βρούμε ποια πολιτική είναι κανόνας του Bayes (κοινώς ποια πολιτική ελαχιστοποιεί τον κίνδυνο κατά Bayes).

Κίνδυνος κατά Bayes: Δοθέντος ότι $\pi(\theta_1) = 0.6$ και $\pi(\theta_2) = 0.4$ έχουμε:

$$\begin{aligned} BR(\pi(\theta), \delta_1(x)) &= E_\theta[FR(\theta, \delta_1(x))] &&= \textstyle\sum_{i=1}^{2} FR(\theta_i, \delta_1(x))\pi(\theta_i) \\ &= 8 \times 0.6 + 0 \times 0.4 &&= 4.8 \\ BR(\pi(\theta), \delta_2(x)) &= 5.2 \times 0.6 + 15 \times 0.4 &&= 9.12 \\ BR(\pi(\theta), \delta_3(x)) &= 2.8 \times 0.6 + 5 \times 0.4 &&= 3.68 \\ BR(\pi(\theta), \delta_4(x)) &= 0 \times 0.6 + 20 \times 0.4 &&= 8 \end{aligned}$$

Παρατηρούμε λοιπόν ότι ως προς την συγκεκριμένη prior $\pi(.)$ έχουμε:

$$\min_{\delta \in \mathcal{D}} BR(\pi(\theta), \delta(x)) = 3.68 = BR(\pi(\theta), \delta_3(x))$$

Άρα ο κανόνας $\delta_3(x)$ είναι ο κανόνας του Bayes. Επομένως βάση του κριτηρίου του "κατά Bayes κινδύνου" θα ακολουθήσουμε την πολιτική που ορίζει ο κανόνας απόφασης $\delta_3(x)$ ($\delta_3(x_1) = a_2$, $\delta_3(x_2) = a_1$) και θα εμβολιάσω μόνον εκείνους οι οποίοι έχουν σοβαρή αντίδραση στο δερματικό τεστ.

Εκ των υστέρων κίνδυνος: $PR(\pi(\theta), \delta_i(x)) : \mathcal{X} \to \Re$. Αρχικά πρέπει να υπολογίσουμε την a-posteriori κατανομή των πιθανοτήτων. Από το θεώρημα του Bayes έχουμε:

$$p(\theta_1|x_1) = \frac{f(x_1|\theta_1)\pi(\theta_1)}{f(x_1|\theta_1)\pi(\theta_1) + f(x_1|\theta_2)\pi(\theta_2)} = \frac{0.65 \times 0.6}{0.65 \times 0.6 + 0.25 \times 0.4} = \frac{39}{49}$$

Ομοίως υπολογίζουμε και τις υπόλοιπες a-posteriori πιθανότητες:

$p(\theta\|x)$	θ_1	θ_2
x_1	39/49	10/49
x_2	21/51	30/51

Έτσι για τον κανόνα απόφασης $\delta_1(x)$ έχουμε:

$$\begin{aligned}
PR(\pi(\theta), \delta_1(x_1)) &= E_{\theta|X=x_1}\left[L(\theta, \delta_1(x_1))\right] = \sum_{i=1}^{2} L(\theta_i, \delta_1(x_1))p(\theta_i|x_1) \\
&= L(\theta_1, \delta_1(x_1))p(\theta_1|x_1) + L(\theta_2, \delta_1(x_1))p(\theta_2|x_1) \\
&= 8 \times \frac{39}{49} + 0 \times \frac{10}{49} = \frac{312}{49} \\
\text{και} \\
PR(\pi(\theta), \delta_1(x_2)) &= E_{\theta|X=x_2}\left[L(\theta, \delta_1(x_2))\right] = \sum_{i=1}^{2} L(\theta_i, \delta_1(x_2))p(\theta_i|x_2) \\
&= L(\theta_1, \delta_1(x_2))p(\theta_1|x_2) + L(\theta_2, \delta_1(x_2))p(\theta_2|x_2) \\
&= 8 \times \frac{21}{51} + 0 \times \frac{30}{51} = \frac{168}{51}
\end{aligned}$$

Ομοίως υπολογίζουμε τον αναμενόμενο εκ των υστέρων κίνδυνο και για τις άλλες τρεις πολιτικές:

$PR(\pi(\theta), \delta_j(x_i))$	$\delta_1(x)$	$\delta_2(x)$	$\delta_3(x)$	$\delta_4(x)$
$x_i = x_1$	312/49	312/49	200/49	200/49
$x_i = x_2$	168/51	600/51	168/51	600/51

Από τον παραπάνω πίνακα παρατηρούμε ότι ο κανόνας απόφασης $\delta_3(x)$ ελαχιστοποιεί τον εκ των υστέρων κίνδυνο ομοιόμορφα ως προς όλες τις πιθανές τιμές του $x \in \mathcal{X}$. Με άλλα λόγια ο κανόνας $\delta_3(x)$ είναι κανόνας του Bayes. Αυτό ήταν αναμενόμενο αφού ο κανόνας του Bayes προκύπτει από την ελαχιστοποίηση του κινδύνου κατά Bayes, ο οποίος βάση του Λήμματος 6 μπορεί να υπολογιστεί είτε ως η μέση τιμή του κλασικού κινδύνου με βάση την κατανομή $\pi(\theta)$ (όπως τον υπολογίσαμε παραπάνω) είτε ως η μέση τιμή του εκ των υστέρων κινδύνου με βάση την κατανομή $f(x)$. Για λόγους απλής επαλήθευσης θα υπολογίσουμε τον κίνδυνο κατά Bayes χρησιμοποιώντας τον εκ των υστέρων κίνδυνο για τον κανόνα απόφασης $\delta_3(x)$. Απαιτείται να υπολογίσουμε τις περιθώρειες πιθανότητες $f(x_1)$ και $f(x_2)$. Έχουμε:

$$\begin{aligned} f(x_1) &= \sum_{i=1}^{2} f(x_1|\theta_i)\pi(\theta_i) = 0.65 \times 0.6 + 0.25 \times 0.4 = 0.49 \\ f(x_2) &= 1 - f(x_1) = 0.51 \end{aligned}$$

Άρα για τον κανόνα απόφασης $\delta_3(x)$ έχουμε:

$$\begin{aligned} BR(\pi(\theta), \delta_3(x)) &= E_x\left[PR(\pi(\theta), \delta_3(x))\right] = \sum_{i=1}^{2} PR(\pi(\theta), \delta_3(x_i))f(x_i) \\ &= \frac{200}{49} \times 0.49 + \frac{168}{51} \times 0.51 = 3.68 \end{aligned}$$

ότι ακριβώς είχαμε και με βάση την κλασική προσέγγιση.

Σχόλιο: Στο παραπάνω παράδειγμα, για καθαρά εκπαιδευτικούς λόγους, εφαρμόσαμε όλες τις μεθόδους τις οποίες αναπτύξαμε στο συγκεκριμένο κεφάλαιο. Στην πράξη όμως αυτό που συνιστούμε είναι το εξής:

- Αρχικά αναγνωρίστε στο πρόβλημά σας όλα τα στοιχεία ενός στατιστικού προβλήματος απόφασης ($\mathbf{\Theta}$, $\mathcal{A}$, συνάρτηση απώλειας, $\mathcal{D}$ κ.τ.λ.)
- Βρείτε την a-posteriori κατανομή $p(\theta|x)$
- Ορίστε την συνάρτηση του εκ των υστέρων κινδύνου $PR(\pi(\theta), \delta(.)) : \mathcal{X} \to \Re$
- Δοθέντος των δεδομένων x, βρείτε τον κανόνα απόφασης $\delta(x)$, ο οποίος ελαχιστοποιεί τον εκ των υστέρων κίνδυνο $PR(\pi(\theta), \delta(.))$. Αυτός θα είναι ο κανόνας του Bayes.

5.2 Κατά Bayes συμπερασματολογία

Από όλα όσα έχουμε αναφέρει έως τώρα προκύπτει ότι ο καλύτερος τρόπος πλήρους περιγραφής της αγνώστου παραμέτρου $\boldsymbol{\theta}$ μπορεί να επιτευχθεί μόνον μέσω της a-posteriori κατανομής. Αποτελεί επομένως η a-posteriori κατανομή μία άρτια “περίληψη” της όλης συμπερασματολογίας σχετικά με την παράμετρο που μας ενδιαφέρει. Με άλλα λόγια, **η a-posteriori $p(\boldsymbol{\theta}|\mathrm{x})$ είναι η συμπερασματολογία**. Σε πολλές περιπτώσεις όμως είναι συχνά επιθυμητό ή απαραίτητο να συνοψίσουμε όλη την πληροφορία αυτή με κάποιον τρόπο. Στο υπόλοιπο αυτού του κεφαλαίου θα αναπτύξουμε τις κοινές μεθόδους στατιστικής συμπερασματολογίας (σημειακή εκτίμηση, εκτίμηση σε διάστημα και ελέγχους υποθέσεων) στο πλαίσιο της κατά Bayes προσέγγισης και υπό το πρίσμα της θεωρίας αποφάσεων.

5.2.1 Σημειακή εκτίμηση

Όταν μιλάμε για εκτίμηση σε σημείο, αυτό το οποίο επιθυμούμε να κάνουμε, είναι να βρούμε την "καλύτερη" εκτιμήτρια για την άγνωστη μας παράμετρο.

Στην κλασική στατιστική οι σημειακές εκτιμήσεις των παραμέτρων είναι η λογική συνέπεια της συμπερασματολογίας, αφού έχουμε υποθέσει ότι η παράμετρος δεν έχει κατανομή αλλά λαμβάνει μια σταθερή (άγνωστη σ' εμάς) τιμή, που προσπαθούμε να υπολογίσουμε. Στις σημειακές αυτές εκτιμήσεις ανακλάται η αβεβαιότητα η οποία δημιουργεί αρκετά προβλήματα.

Από την πλευρά της Μπευζιανής θεωρίας το ερώτημα το οποίο προκύπτει, είναι πως θα μειώσουμε την πληροφορία στην a-posteriori κατανομή, ώστε να έχουμε μία μοναδική άριστη εκτιμήτρια; Στην πραγματικότητα, η απάντηση στο ερώτημα αυτό εξαρτάται από το τί ακριβώς εννοούμε όταν λέμε άριστη εκτιμήτρια, και ο καθορισμός αυτός οδηγεί στην μετάλλαξη του αρχικού προβλήματος σε ένα πρόβλημα απόφασης. Έχουμε λοιπόν ένα πρόβλημα απόφασης με $\mathcal{A} = (-\infty, +\infty)$. Για την λύση του προβλήματος αυτού απαιτείται ο καθορισμός και μίας συνάρτησης απώλειας $L(\boldsymbol{\theta}, \alpha)$, η οποία θα μετρά την "ποινή" μας όταν εκτιμούμε την παράμετρο $\boldsymbol{\theta}$ μέσω του α. Υπάρχει μία πληθώρα από συναρτήσεις απωλειών τις οποίες θα μπορούσαμε να χρησιμοποιήσουμε, και η συγκεκριμένη επιλογή για κάθε πρόβλημα εξαρτάται από το περιεχόμενο του προβλήματος αυτού. Οι πιο συνηθισμένες συναρτήσεις είναι οι εξής:

- Δευτεροβάθμια Συνάρτηση Απώλειας (Quadratic Error Loss)

$$L(\theta, \alpha) = (\theta - \alpha)^2$$

- Συνάρτηση Απώλειας Απολύτου Σφάλματος (Absolute Error Loss)

$$L(\theta, \alpha) = |\theta - \alpha|$$

- Γραμμική Συνάρτηση Απώλειας (Linear Loss)

$$L(\theta,\alpha) = \left\{ \begin{array}{lll} K_0(\alpha-\theta) & \text{όταν} & \alpha > \theta \\ K_1(\theta-\alpha) & \text{όταν} & \alpha < \theta \end{array} \right\}, \text{όπου } K_0 > 0 \text{ και } K_1 > 0 \text{ σταθερές}$$

- $0-1$ Συνάρτηση Απώλειας (0-1 Loss)

$$L(\theta,\alpha) = \left\{ \begin{array}{lll} 0 & \text{όταν} & |\alpha-\theta| \le \varepsilon \\ 1 & \text{όταν} & |\alpha-\theta| > \varepsilon \end{array} \right\}, \text{με } \varepsilon > 0$$

Σε κάθε μία από τις παραπάνω περιπτώσεις , με την ελαχιστοποίηση της a-posteriori αναμενόμενης απώλειας, μπορούμε να βρούμε απλές μορφές για τους κανόνες αποφάσεων του Bayes , οι οποίες μπορούν να θεωρηθούν ως οι σημειακές εκτιμήσεις για την παράμετρο θ για την συγκεκριμένη επιλογή της συνάρτησης απώλειας.

Δευτεροβάθμια Συνάρτηση Απώλειας

Η δευτεροβάθμια συνάρτηση απώλειας χρησιμοποιείται όταν επιθυμούμε η "ποινή" που επιβάλεται να μεγαλώνει (το τετράγωνο της αποκλίσεως απο την πραγματική τιμή) καθώς απομακρυνόμαστε από την πραγματική τιμή. Πρόκειται για συμμετρική συνάρτηση υπό την έννοια ότι τιμωρεί υπό-εκτίμηση και υπέρ-εκτίμηση ομοίως.

Ο κανόνας του Bayes θα προκύψει από την ελαχιστοποίηση της συναρτήσεως του εκ των υστέρων κινδύνου (posterior risk) ως προς α. Πρέπει λοιπόν να βρούμε το ελάχιστο (ως προς α) της συναρτήσεως:

$$PR(\pi(\theta),\alpha) = \int L(\theta,\alpha)p(\theta|\mathbf{x})d\theta = \int (\theta-\alpha)^2 p(\theta|\mathbf{x})d\theta$$

Έχουμε:

$$\begin{aligned} \frac{\partial}{\partial\alpha}\left[PR(\pi(\theta),\alpha)\right] = 0 \quad &\Rightarrow \quad \int \frac{\partial}{\partial\alpha}(\theta-\alpha)^2 p(\theta|\mathbf{x})d\theta = 0 \\ &\Rightarrow \quad \int 2(\theta-\alpha)(-1)p(\theta|\mathbf{x})d\theta = 0 \\ &\Rightarrow \quad \alpha\int p(\theta|\mathbf{x})d\theta = \int \theta p(\theta|\mathbf{x})d\theta \\ &\Rightarrow \quad \alpha = E[\theta|\mathbf{x}] \end{aligned}$$

και

$$\frac{\partial^2}{\partial\alpha^2}\left[PR(\pi(\theta),\alpha)\right] = 2\int p(\theta|\mathbf{x})d\theta = 2 > 0$$

το οποίο επιβεβαιώνει ότι το ακρότατο $E[\theta|\mathrm{x}]$ που προέκυψε εξισώνοντας την πρώτη παράγωγο με το μηδέν είναι ελάχιστο. Άρα όταν χρησιμοποιούμε την δευτεροβάθμια συνάρτηση απώλειας ο κανόνας του Bayes (και κατά συνέπεια ο σημειακός εκτιμητής που προτείνουμε) είναι η **μέση τιμή της a-posteriori κατανομής** $E[\theta|\mathrm{x}]$.

Γραμμική Συνάρτηση Απώλειας

Η γραμμική συνάρτηση απώλειας έχει ως κύριο χαρακτηριστικό το ότι η "ποινή" αυξάνεται γραμμικά καθώς απομακρυνόμαστε από την πραγματική τιμή. Επίσης τιμωρεί υπό-εκτίμηση και υπέρ-εκτίμηση διαφορετικά για $K_0 \neq K_1$. Συγκρινόμενη με την δευτεροβάθμια συνάρτηση απώλειας τιμωρεί περισσότερο τις "μικρές" αποκλίσεις και λιγότερο τις "μεγάλες".

Ως γνωστόν απαιτείται να βρούμε που ελαχιστοποιείται η συνάρτηση του εκ των υστέρων κινδύνου. Καλούμε q το $K_1/(K_0+K_1)$ ποσοστιαίο σημείο της a-posteriori κατανομής, δηλαδή:

$$\frac{K_1}{K_0+K_1} = \int_{-\infty}^{q} p(\theta|\mathbf{x})d\theta$$

και ας υποθέσουμε, χωρίς βλάβη της γενικότητας, ότι $\alpha > q$. Τότε:

$$L(\theta,q) - L(\theta,\alpha) = \left\{ \begin{array}{lll} K_0(q-\alpha) & \text{όταν} & \theta \leq q < \alpha \\ K_1(\theta-q) + K_0(\theta-\alpha) & \text{όταν} & q < \theta < \alpha \\ K_1(\alpha-q) & \text{όταν} & q < \alpha < \theta \end{array} \right\}$$

Όμως στην περίπτωση που $q < \theta < \alpha$ έχουμε:

$$K_1(\theta-q) + K_0(\theta-\alpha) < K_1\theta - K_1 q < K_1\alpha - K_1 q = K_1(\alpha-q)$$

Άρα:

$$L(\theta,q) - L(\theta,\alpha) \leq \left\{ \begin{array}{lll} K_0(q-\alpha) & \text{όταν} & \theta \leq q \\ K_1(\alpha-q) & \text{όταν} & \theta > q \end{array} \right\}$$

Συνεπώς έχουμε:

$$\begin{aligned} E\left[L(\theta,q) - L(\theta,\alpha)\right] &\leq K_0(q-\alpha)P(\theta \leq q|\mathbf{x}) + K_1(\alpha-q)P(\theta > q|\mathbf{x}) \\ &= K_0(q-\alpha)\left(\frac{K_1}{K_0+K_1}\right) + K_1(\alpha-q)\left(1-\frac{K_1}{K_0+K_1}\right) = 0 \end{aligned}$$

Επομένως $\forall\ \alpha \in \mathcal{A}$ έχουμε $E[L(\theta,q)] \leq E[L(\theta,\alpha)]$. Άρα όταν χρησιμοποιούμε γραμμική συνάρτηση απώλειας, ο κανόνας απόφασης $\delta(\mathbf{x}) = q$ δηλαδή το $K_1/(K_0+K_1)$ **ποσοστιαίο σημείο της a-posteriori κατανομής** είναι ο κανόνας του Bayes.

Συνάρτηση Απώλειας Απολύτου Σφάλματος

Η περίπτωση της συνάρτησης απώλειας απόλυτου σφάλματος, είναι μία ειδική περίπτωση της γραμμικής συνάρτησης απώλειας όταν $K_0 = K_1 = 1$. Εφαρμόζωντας λοιπόν αυτά που βρήκαμε παραπάνω προκύπτει ότι στην περίπτωση αυτή ο κανόνας του Bayes είναι **η διάμεσος (το 0.5 ποσοστιαίο σημείο) της a-posteriori κατανομής**.

$0-1$ Συνάρτηση Απώλειας

Η συνάρτηση απώλειας $0-1$ χρησιμοποιείται όταν επιθυμούμε να μην τιμωρούμε την "σωστή" (ακριβείας ε) εκτίμηση, ενώ η ποινή της "λανθασμενης" εκτίμησης να είναι σταθερή, ανεξαρτήτως μεγέθους αποκλίσεως. Η συνάρτηση απώλειας $0-1$ εύκολα γενικεύεται σε $0-K$ συνάρτηση απώλειας, όπου K θετική σταθερά.

Προκειμένου να βρούμε τον κανόνα του Bayes, απαιτείται να βρούμε που ελαχιστοποιήται η συνάρτηση του εκ των υστέρων κινδύνου (ως συνάρτηση του α). Έχουμε:

$$PR(\pi(\theta), \alpha) = E_{\theta|\mathbf{x}}[L(\theta, \alpha)] = p(|\theta - \alpha| > \varepsilon|\mathbf{x}) = 1 - p(|\theta - \alpha| \leq \varepsilon|\mathbf{x})$$

Άρα απαιτείτε να βρούμε την τιμή του α για την οποία η πιθανότητα $p(|\theta - \alpha| \leq \varepsilon|\mathbf{x})$ μεγιστοποιείται. Εάν καθορίσουμε ένα διάστημα μήκους 2ε, όπως για παράδειγμα το διάστημα $[\alpha - \varepsilon, \alpha + \varepsilon]$, το οποίο έχει την μέγιστη δυνατή πιθανότητα, τότε ο κανόνας του Bayes θα είναι το μέσο σημείο του διαστήματος (με την μεγαλύτερη πιθανότητα). Επιλέγοντας ένα ε αρκετά μικρό, η διαδικασία θα οδηγήσει να επιλεχθεί ως σημειακή εκτιμήτρια (κανόνας του Bayes) η **κορυφή (mode) της a-posteriori κατανομής** μιας και για σταθερό ε, το διάστημα $[\text{mode} - \varepsilon, \text{mode} + \varepsilon]$ έχει την μεγαλύτερη πιθανότητα από οποιοδήποτε άλλο διάστημα.

Περίληψη

Ανακεφαλαιώνοντας την κατά Bayes προσέγγιση στη σημειακή εκτίμηση μπορούμε να πούμε ότι η σημειακή εκτίμηση μίας παραμέτρου είναι απλώς ένα μοναδικό στατιστικό το οποίο δεν περιλαμβάνει όλη την πληροφορία της a-posteriori κατανομής. Καθορίζοντας την ποιότητα ενός εκτιμητή μέσα από μία συνάρτηση απώλειας, η μεθοδολογία της θεωρίας αποφάσεων οδηγεί στις καλύτερες επιλογές των σημειακών εκτιμητών. Ειδικότερα, οι πιο φυσικές (συνήθης) επιλογές των συναρτήσεων απωλειών οδηγούν σε αντίστοιχους a-posteriori μέσους, διάμεσους ή κορυφές σαν τους ιδανικούς σημειακούς εκτιμητές. Ας κλείσουμε το κομάτι της σημειακής εκτίμησης μελετώντας ένα παράδειγμα συναρτήσεως απώλειας που συναντάται συχνά στην οικονομική θεωρία.

Παράδειγμα 5.2

Έστω ότι έχουμε τυχαίο δείγμα $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ από την κατανομή $f(x_i|\theta) \sim Poisson(\theta)$. Επίσης η a-priori πληροφόρηση που έχουμε για την παράμετρο θ είναι ότι $\pi(\theta) \sim Gamma(c, d)$. Αν χρησιμοποιήσουμε την LINEX (LINear-EXponential) συνάρτηση απώλειας (Zellner (1986)):

$$L(\theta, \alpha) = e^{k(\alpha-\theta)} - k(\alpha - \theta) - 1$$

με k θετική σταθερά, ποιό σημειακό εκτιμητή προτείνετε για την παράμετρο θ;

Λύση: Καταρχήν θα υπολογίσουμε την a-posteriori κατανομή $p(\theta|\mathbf{x})$. Είναι εύκολο να δούμε ότι:

$$\left.\begin{array}{rcl} f(x_i|\theta) & \sim & Poisson(\theta) \\ \pi(\theta) & \sim & Gamma(c, d) \end{array}\right\} \Rightarrow p(\theta|\mathbf{x}) \sim Gamma\left(c + \sum_{i=1}^{n} x_i,\ d + n\right)$$

Εν συνεχεία θα βρούμε την τιμή του α για την οποία η συνάρτηση του εκ των υστέρων κινδύνου ελαχιστοποιείται:

$$PR(\pi(\theta), \alpha) = \int L(\theta, \alpha) p(\theta|\mathbf{x}) d\theta = \int \left[e^{k(\alpha-\theta)} - k(\alpha - \theta) - 1\right] p(\theta|\mathbf{x}) d\theta$$

Έχουμε:

$$\begin{aligned} \frac{\partial}{\partial \alpha}\left[PR(\pi(\theta), \alpha)\right] = 0 \quad & \Rightarrow \quad \int \frac{\partial}{\partial \alpha}\left[e^{k(\alpha-\theta)} - k(\alpha - \theta) - 1\right] p(\theta|\mathbf{x}) d\theta = 0 \\ & \Rightarrow \quad \int \left[k e^{k(\alpha-\theta)} - k\right] p(\theta|\mathbf{x}) d\theta = 0 \\ & \Rightarrow \quad e^{k\alpha} \int e^{-k\theta} p(\theta|\mathbf{x}) d\theta = \int p(\theta|\mathbf{x}) d\theta \\ & \Rightarrow \quad e^{-k\alpha} = E\left[e^{-k\theta}|\mathbf{x}\right] \\ & \Rightarrow \quad \alpha = -\frac{log E\left[e^{-k\theta}|\mathbf{x}\right]}{k} \end{aligned}$$

και

$$\frac{\partial^2}{\partial \alpha^2}\left[PR(\pi(\theta), \alpha)\right] = \int k^2 e^{k(\alpha-\theta)} p(\theta|\mathbf{x}) d\theta = k^2 e^{k\alpha} E\left[e^{-k\theta}|\mathbf{x}\right] > 0$$

διότι $e^{-k\theta}$ είναι μή-αρνητική συνάρτηση. Άρα το τοπικό ακρότατο που βρήκαμε παραπάνω είναι ελάχιστο. Αυτό που απομένει, είναι να υπολογίσουμε την $E\left[e^{-k\theta}|\mathbf{x}\right]$ στην περίπτωση που η a-posteriori κατανομή $p(\theta|\mathbf{x})$ είναι $Gamma\left(c+\sum_{i=1}^{n} x_i, d+n\right) \equiv Gamma(A,B)$:

$$E\left[e^{-k\theta}|\mathbf{x}\right] = \int_0^{+\infty} e^{-k\theta}\frac{B^A}{\Gamma(A)}\theta^{A-1}e^{-B\theta}d\theta = \frac{B^A}{\Gamma(A)}\int_0^{+\infty}\theta^{A-1}e^{-(B+k)\theta}d\theta = \frac{B^A}{(B+k)^A}$$

Άρα ο προτεινόμενος σημειακός εκτιμητής του θ (κανόνας του Bayes) όταν χρησιμοποιούμε την LINEX συνάρτηση απώλειας είναι:

$$\delta_\pi(\mathbf{x}) = -\frac{logE\left[e^{-k\theta}|\mathbf{x}\right]}{k} = \frac{\left(c+\sum_{i=1}^{n} x_i\right)\left[log(d+n+k)-log(d+n)\right]}{k}$$

5.2.2 Εκτίμηση σε Διάστημα

Η εκτίμηση σε διάστημα έχει ως κύριο σκοπό να μας δώσει ένα διάστημα (των πλέον) πιθανών τιμών της αγνώστου παραμέτρου θ. Ο λόγος ύπαρξής τους είναι ότι οι σημειακοί εκτιμητές δεν μπορούν να μετρηθούν με ακρίβεια, και για αυτό είναι προτιμότερο να κατασκευάσουμε ένα διάστημα μέσα στο οποίο είναι πιθανό να βρίσκεται η τιμή της παραμέτρου μας.

Στην κλασική στατιστική, εκτίμηση σε διάστημα γίνεται με την χρήση των διαστημάτων εμπιστοσύνης. Τα διαστήματα εμπιστοσύνης όμως, δεν είναι συμβατά με την παραπάνω φιλοσοφία, με αποτέλεσμα να έχουν ιδιόμορφη ερμηνεία. Το πρόβλημα ξεκινάει από το ότι στην κλασική στατιστική η άγνωστη παράμετρος θ δεν ακολουθεί κατανομή αλλά θεωρείται σταθερός αριθμός. Αυτό έχει ως αποτέλεσμα η τυχαία ποσότητα να είναι το διάστημα εμπιστοσύνης και όχι η άγνωστη παράμετρος (θ). Άρα είναι **λάθος** να λέμε ότι το θ ανήκει στο διάστημα εμπιστοσύνης με κάποια πιθανότητα που καθορίζεται από το επίπεδο εμπιστοσύνης (όπως θα ευχόμασταν να ίσχυε), μιας και το θ ως σταθερά θα ανήκει στο διάστημα εμπιστοσύνης με πιθανότητα 1 ή 0 (δηλαδή είτε θα βρισκεται εντός του διαστήματος είτε όχι). Η ορθή ερμηνεία ενός διαστήματος εμπιστοσύνης είναι ότι αν το πείραμα επαναλαμβανόταν πολλές φορές, υπάρχει μια συγκεκριμένη πιθανότητα (που καθορίζεται από το επίπεδο εμπιστοσύνης) ότι το διάστημα το οποίο έχει κατασκευαστεί θα περιλαμβάνει την (σταθερή) τιμή της παραμέτρου θ.

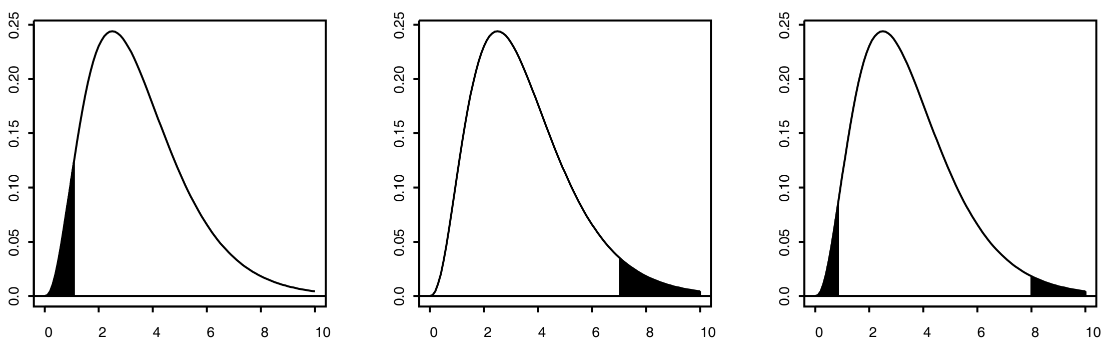
Στον αντίποδα με βάση την κατά Bayes προσέγγιση δεν υπάρχει καμία δυσκολία διότι οι παράμετροι είναι τυχαίες μεταβλητές. Οδηγούμαστε λοιπόν πολύ εύκολα στην κατασκευή των επονομαζόμενων "διαστημάτων αξιοπιστίας (credibility intervals)" που μπορούν να χρησιμοποιηθούν για εκτίμηση σε διάστημα. Συγκεκριμένα:

Ορισμός 9 *Ορίζουμε μια περιοχή* $C_\alpha(\mathbf{x})$ *ως* $100(1-\alpha)\%$ **περιοχή αξιοπιστίας** *για την* *παράμετρο θ εάν:*

$$\int_{C_\alpha(\mathbf{x})} p(\theta|\mathbf{x})d\theta = 1 - \alpha$$

Η ερμηνεία της περιοχής αξιοπιστίας είναι η ακόλουθη: βάση της a-posteriori κατανομής η παράμετρος θ βρίσκεται στο $C_\alpha(\mathbf{x})$ διάστημα με πιθανότητα του $1 - \alpha$. Μερικοί υποστηρικτές της θεωρίας του Bayes πιστεύουν ότι τα διαστήματα αξιοπιστίας έχουν μικρή σχετικά αξία, από την στιγμή που ολόκληρη η a-posteriori κατανομή είναι αυτή η οποία περιλαμβάνει όλη την πληροφορία για τα συμπεράσματα ενώ τα διαστήματα αξιοπιστίας έχουν απλά δημιουργηθεί με σκοπό να παρουσιάσουν το αντίστοιχο των διαστημάτων εμπιστοσύνης.

Μία δυσκολία η οποία αφορά τα διαστήματα αξιοπιστίας (κοινή με τα διαστήματα εμπιστοσύνης) είναι το γεγονός ότι δεν είναι μοναδικά ορισμένα. Συγκεκριμένα κάθε περιοχή με πιθανότητα $1 - \alpha$ ορίζει ένα διάστημα. Στο Σχήμα 5.1 μπορούμε να δούμε τρία τελείως διαφορετικά 95% διαστήματα αξιοπιστίας για την ίδια a-posteriori κατανομή. Συγκεκριμένα το πρώτο (το οποίο έχει και άπειρο μήκος) έχει κατασκευαστεί έτσι ώστε να αφήνει 5% πιθανότητα εκτός στην αριστερή ουρά της κατανομής. Το δεύτερο αφήνει 5% πιθανότητα εκτός στη δεξιά ουρά της κατανομής και το τελευταίο αφήνει 2.5% πιθανότητα σε κάθε ουρά.



Σχήμα 5.1: Τρία 95% διαστήματα αξιοπιστίας για την ίδια a-posteriori κατανομή

Από την στιγμή όμως που εμείς θέλουμε το διάστημα μας να περιλαμβάνει μονάχα τις πιο "πειστικές" τιμές (με την έννοια των πιο πιθανών τιμών) της παραμέτρου θ, είναι σύνηθες να προσθέτουμε και έναν ακόμα περιορισμό με βάση τον οποίο θα πρέπει το διάστημα να είναι όσο το δυνατόν μικρότερο. Με βάση αυτόν τον περιορισμό οδηγούμαστε στις μοναδικά ορισμένες "περιοχές υψίστης a-posteriori πυκνότητας (highest posterior density regions)":

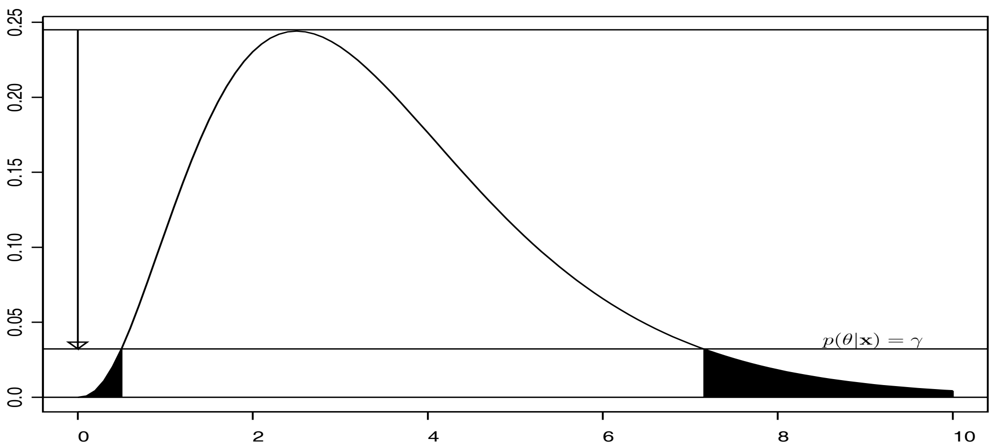
Ορισμός 10 *Ορίζουμε μια περιοχή* $HPD_\alpha(\mathbf{x})$ *ως* $100(1-\alpha)\%$ **περιοχή υψίστης a-posteriori πυκνότητας (Highest Posterior Density)** *για την παράμετρο* θ *εάν:*

$$HPD_\alpha(\mathbf{x}) = \{\theta : p(\theta|\mathbf{x}) \geq \gamma\}$$

όπου γ *είναι σταθερά η οποία ικανοποιεί την σχέση:*

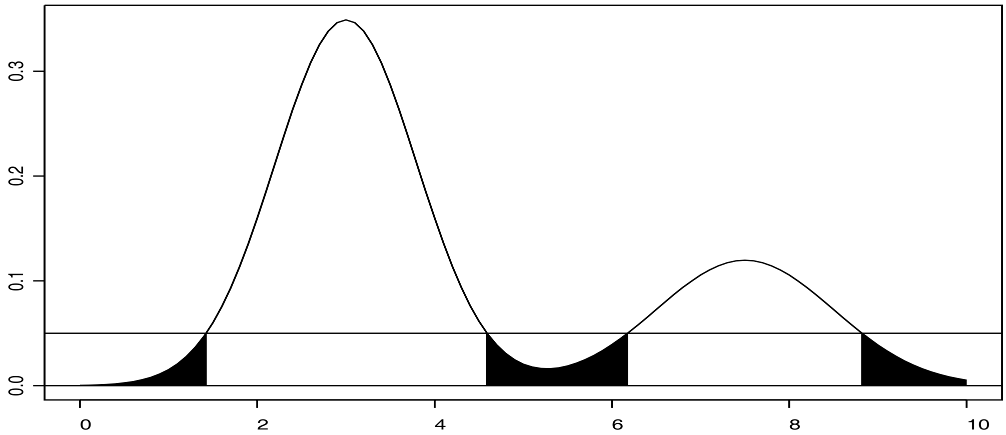
$$\int_{HPD_\alpha(\mathbf{x})} p(\theta|\mathbf{x}) d\theta = 1 - \alpha$$

Όπως φαίνεται και στο Σχήμα 5.2 προκειμένου να κατασκευάσουμε ένα διάστημα υψίστης a-posteriori πυκνότητας φέρουμαι μια ευθεία παράλληλη με τον άξονα των θ στην κορυφή της κατανομής την οποία μετακινούμε προς τα κάτω και σταματάμε σε ύψος τέτοιο ώστε η πιθανότητα εκτός του διαστήματος να είναι α (το ύψος αυτό αποτελεί το γ στον παραπάνω ορισμό).



Σχήμα 5.2: 95% διάστημα υψίστης a-posteriori πυκνότητας

Γενικά, τα διαστήματα αυτά θα πρέπει να υπολογιστούν αριθμητικά, παρόλο που για τις περισσότερες μονοδιάστατες a-posteriori κατανομές οι τιμές συγκεντρώνονται σε πίνακες ανάλογα με το εύρος των τιμών του α. Στο σημείο αυτό ας σημειώσουμε ότι είναι σημαντικό να κάνουμε σωστή επιλογή του α: μικρές τιμές του α θα δώσουν μεγάλα διαστήματα, ενώ αντίθετα μεγάλες τιμές δίνουν διαστήματα στα οποία η παράμετρος έχει μικρές πιθανότητες να βρίσκεται στο διάστημα. Στην περίπτωση συμμετρικής, μονόκορφης κατανομής το διάστημα υψίστης a-posteriori πυκνότητας ταυτίζεται με το διάστημα αξιοπιστίας που αφήνει $\alpha/2$ πιθανότητα εκτός σε κάθε άκρο. Τέλος, όπως μπορούμε να δούμε και στο Σχήμα 5.3, στην περίπτωση που η a-posteriori κατανομή είναι πολύκορφη η περιοχή υψίστης a-posteriori πυκνότητας μπορεί να είναι ένωση διαστημάτων.



Σχήμα 5.3: $100(1-\alpha)\%$ περιοχή υψίστης a-posteriori πυκνότητας σε πολύκορφη a-posteriori κατανομή

Παράδειγμα 5.3

Έστω ότι έχουμε τυχαίο δείγμα $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ από την κανονική κατανομή $N(\theta, \sigma^2)$, με σ^2 γνωστό και έστω η a-priori κατανομή του θ είναι επίσης κανονική: $\pi(\theta) \sim N(\mu, \tau^2)$ (με μ, τ^2 γνωστά). Να βρεθεί η $100(1-\alpha)\%$ περιοχή υψίστης a-posteriori πυκνότητας για την παράμετρο θ.

<u>**Λύση:**</u> Καταρχήν θα υπολογίσουμε την a-posteriori κατανομή $p(\theta|\mathbf{x})$. Όπως έχουμε ήδη δείξει:

$$\left.\begin{array}{rcl} f(x_i|\theta) & \sim & N(\theta, \sigma^2) \\ \pi(\theta) & \sim & N(\mu, \tau^2) \end{array}\right\} \Rightarrow p(\theta|\mathbf{x}) \sim N\left(\frac{\tau^2\overline{x} + \mu\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}, \frac{\tau^2\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}\right)$$

Από τη στιγμή που η a-posteriori κατανομή της παραμέτρου θ είναι συμμετρική η $100(1-\alpha)\%$ περιοχή υψίστης a-posteriori πυκνότητας ταυτίζεται με το διάστημα αξιοπιστίας το οποίο "αφήνει εκτός" $\alpha/2$ από κάθε ουρά της κατανομής. Έτσι η $100(1-\alpha)\%$ περοιοχή υψίστης a-posteriori κατανομής για την παράμετρο θ είναι:

$$\frac{\tau^2\overline{x} + \mu\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \pm z_{\alpha/2}\sqrt{\frac{\tau^2\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}}$$

όπου το $z_{\alpha/2}$ είναι το κατάλληλο ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής $N(0,1)$.

<u>Σχόλιο:</u>Αξίζει στο σημείο αυτό να προσέξουμε ότι καθώς το n μεγαλώνει η παραπάνω περιοχή τείνει προς το

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

διάστημα, το οποίο δεν είναι άλλο από το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης το οποίο χρησιμοποιείται στην κλασική στατιστική. Παρά το γεγονός ότι τα διαστήματα σε αυτή την ειδική περίπτωση ταυτίζονται οι ερμηνείες τους είναι τελείως διαφορετικές.

Παράδειγμα 5.4

Να βρεθεί η $100(1-\alpha)\%$ περιοχή υψίστης a-posteriori πυκνότητας για την παράμετρο θ όταν η πιθανοθάνεια είναι Διωνυμική: $f(x|\theta) \sim B(n,\theta)$ με n γνωστό και η prior κατανομή είναι κατανομή Βήτα: $\pi(\theta) \sim Beta(p,q)$ με α, β επίσης γνωστά.

<u>**Λύση:**</u> Η a-posteriori κατανομή είναι:

$$\left.\begin{array}{rcl} f(x|\theta) & \sim & B(n,\theta) \\ \pi(\theta) & \sim & Beta(p,q) \end{array}\right\} \Rightarrow p(\theta|x) \sim Beta\left(p+x, q+n-x\right)$$

Στην περίπτωση αυτή η $100(1-\alpha)\%$ περιοχή υψίστης a-posteriori πυκνότητας $[L,U]$ πρέπει να ικανοποιεί βάση ορισμού τις ακόλουθες εξισώσεις:

$$\frac{1}{Be(p+x,q+n-x)}\int_L^U \theta^{p+x-1}(1-\theta)^{q+n-x-1}d\theta = 1-\alpha$$

και

$$\frac{1}{Be(p+x,q+n-x)}L^{p+x-1}(1-L)^{q+n-x-1} = \frac{1}{Be(p+x,q+n-x)}U^{p+x-1}(1-U)^{q+n-x-1} = \gamma$$

Το παραπάνω σύστημα εξισώσεων μπορεί να επιλυθεί μόνον αριθμητικά.

5.2.3 Έλεγχοι Υποθέσεων

Πολλές φορές ενδιαφερόμαστε όχι για την εκτίμηση αυτή καθεαυτή (σε σημείο ή διάστημα) της αγνώστου παραμέτρου θ αλλά για το αν και κατά πόσο είναι αλήθής κάποια δήλωση που αφορά την άγνωστη αυτή παράμετρο του πληθυσμού.

Οι έλεγχοι υποθέσεων λοιπόν είναι απλά αποφάσεις με την μορφή της επιλογής μεταξύ δύο διαφορετικών υποθέσεων:

$$\left\{\begin{array}{ll} H_0: & \theta \in \Omega_0 \\ H_1: & \theta \in \Omega_1 \end{array}\right\}$$

όπου $\Omega_i, i=0,1$ αποτελούν υποσύνολα (συνήθως συμπληρωματικά) του παραμετρικού χώρου Θ. Στην ενότητα αυτή θα ασχοληθούμε με την απλή περίπτωση που τόσο η μεδενική όσο και η εναλλακτική υπόθεση αφορούν μεμονομένα σημεία, δηλαδή:

$$\left\{ \begin{array}{ll} H_0 : & \theta = \theta_0 \\ H_1 : & \theta = \theta_1 \end{array} \right\}$$

Αν το δούμε σαν πρόβλημα της θεωρίας αποφάσεων έχουμε $\mathcal{A} = \{\alpha_0, \alpha_1\}$. Υπάρχουν δηλαδή δύο ενέργειες στις οποίες μπορούμε να προβούμε: α_0 =αποδεχόμαστε την H_0 και α_1 =αποδεχόμαστε την H_1.

Με βάση την κλασική προσέγγιση ο συνήθης τρόπος επίλυσης του συγκεκριμένου προβλήματος είναι να βασίσουμε το τέστ στον λόγο των πιθανοφανειών (likelihood ratio test):

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta_0} L(\theta|\mathbf{x})}{\sup_{\theta_1} L(\theta|\mathbf{x})}$$

με περιοχή απορρίψεως της H_0 την $\{\mathbf{x} : \lambda(\mathbf{x}) < c\}$, κοινώς, μικρές τιμές για το $\lambda(\mathbf{x})$, υποδεικνύουν ότι τα δεδομένα μας είναι πιο πιθανό να έχουν προέλθει εάν η θ_1 είναι η πραγματική τιμή της παραμέτρου θ παρά εάν ήταν η θ_0. Σχεδόν πάντα η απόφαση μας συνοδεύεται και από την καταγραφή του p-value (που αναφέρεται στην πιθανότητα η ελεγχοσυνάρτηση να λάβει ακόμη πιο ακραίες τιμές από αυτή που παρατηρήθηκε για τα συγκεκριμένα δεδομένα, υπό την προϋπόθεση φυσικά ότι η μηδενική υπόθεση είναι η σωστή). Σωρεία προβλημάτων και παρερμηνειών όμως ελοχεύουν κάτω από την συγκεκριμένη προσέγγιση. Αναφορικά το p-value συγκρινόμενο με την στάθμη σημαντικότητας που έχουμε ορίσει, μας δίνει σαν συμπέρασμα την απόρριψη (ή αποτυχία απόρριψης) της μηδενικής υπόθεσης (μιας και η όλη διαδικασία έχει στηριχθεί στην υπόθεση ότι η H_0 είναι αληθής), δεν επιτρέπεται λοιπόν να διατυπώσουμε σχόλια αναφορικά με την αποδοχή ή όχι της H_1. Επίσης δεν πρέπει να ξεχνάμε ότι πολλές φορές η αρχή της πιθανοφάνειας παραβιάζεται κάτω από την συγκεκριμένη προσέγγιση (παράδειγμα 1ου κεφαλαίου).

Με βάση την κατά Bayes προσέγγιση μπορούμε να χρησιμοποιήσουμε την a-priori πληροφόρηση που έχουμε για τις H_0 και H_1. Συγκεκριμένα αν $\pi(\theta_0)$ και $\pi(\theta_1)$ είναι οι a-priori κατανομές για τις H_0 και H_1 αντίστοιχα, τότε μπορούμε απλά να υπολογίσουμε τις a-posteriori πιθανότητες των δύο υποθέσεων: $p(\theta_0|\mathbf{x})$ και $p(\theta_1|\mathbf{x})$ και να αποφασίσουμε συγκρίνοντάς τες (αποδεχόμαστε την H_1 αν $p(\theta_0|\mathbf{x}) < p(\theta_1|\mathbf{x})$). Εναλλακτικά μπορούμε να στηρίξουμε το συμπέρασμά μας στον λόγο των a-posteriori πιθανοτήτων γνωστό και ως posterior-odds:

$$\lambda_B = \frac{p(\theta_0|\mathbf{x})}{p(\theta_1|\mathbf{x})} = \frac{f(\mathbf{x}|\theta_0)\pi(\theta_0)}{f(\mathbf{x}|\theta_1)\pi(\theta_1)}$$

Αξίζει να παρατηρήσουμε ότι δεν υπάρχει λόγος να υπολογίσουμε τη σταθερά κανονικοποίησης, από την στιγμή που ο ίδιος παράγοντας εμφανίζεται και στον αριθμητή και στον παρανομαστή. Ομοίως με τα προηγούμενα, μικρές τιμές για το λ_B είναι υπέρ της H_1.

Μια άλλη ενδιαφέρουσα ποσότητα είναι ο παράγοντας του Bayes (Bayes Factor), ο οποίος ορίστηκε από τον Jeffreys (1948) ως ο λόγος των a-posteriori odds προς a-priori odds. Συγκεκριμένα στην περίπτωση που η H_0 και H_1 αφορούν μεμονωμένα σημεία τότε έχουμε:

$$B = \frac{p(\theta_0|\mathbf{x})/p(\theta_1|\mathbf{x})}{\pi(\theta_0)/\pi(\theta_1)} = \frac{[f(\mathbf{x}|\theta_0)\pi(\theta_0)] \,/\, [f(\mathbf{x}|\theta_1)\pi(\theta_1)]}{\pi(\theta_0)/\pi(\theta_1)} = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)}$$

Ο λόγος των πιθανοφανειών στην περίπτωση αυτή λοιπόν καθορίζει τον παράγοντα Bayes. Η αιτία για την οποία πρέπει να επικεντρώσουμε την προσοχή μας στον παράγοντα Bayes είναι ότι ο παράγοντας αυτός είναι ένα μέτρο για την βαρύτητα της πληροφορίας η οποία περιλαμβάνεται στα δεδομένα υπέρ της H_1 έναντι της H_0. Αν ο παράγοντας Bayes είναι αρκετά μικρός, τότε θα ξεπεράσουμε την αρχική προτίμηση μας για την H_0, οπότε η a-posteriori επιλογή μας θα είναι η H_1.

Ασκήσεις

Άσκηση 5.1 Δύο ειδικοί σε θέματα τέχνης, ο A και ο B, δίνουν κάποιο τεστ με σκοπό να ελεγχθεί η αξιοπιστία τους βάσει της οποίας κρίνουν ένα έργο τέχνης ως αυθεντικό ή πλαστό. Τα έργα τέχνης είναι γνωστής προέλευσης. Από τα τεστ προκύπτει ότι ο A έχει πιθανότητα 0.8 να ξεχωρίσει ένα πλαστό έργο, και 0.7 ένα αυθεντικό. Ο B έχει μεγαλύτερη πιθανότητα, 0.9, να ξεχωρίσει ένα πλαστό έργο, αλλά δυστυχώς έχει και 0.4 πιθανότητα να θεωρήσει λανθασμένα ένα αυθεντικό έργο ως πλαστό. Μου έχει προσφερθεί κάποιο έργο τέχνης το οποίο είναι στην τιμή ευκαιρίας των 100 λιρών. Αν δεν είναι αυθεντικό, τότε δεν αξίζει τίποτα. Αλλά αν είναι, τότε θα άξιζε σίγουρα 300 λίρες. Πιστεύω ότι υπάρχει 0.5 πιθανότητα το έργο να είναι αυθεντικό. Οι ειδικοί A και B χρεώνουν 30 και 40 λίρες αντίστοιχα τις υπηρεσίες τους. Θα έχω όφελος να πληρώσω κάποιον από τους δύο ειδικούς για να αποφασίσω για την επιλογή μου;

Άσκηση 5.2 Για το πρόβλημα με τον κανονικό μέσο που περιγράφηκε στο παράδειγμα 2.4, βρείτε τις σημειακές εκτιμήσεις για το θ χρησιμοποιώντας κάθε μία από τις συναρτήσεις απώλειας που περιγράφηκαν σε αυτό το κεφάλαιο.

Άσκηση 5.3 Επαναλάβετε την άσκηση 5.2 για την συζυγή ανάλυση της Διωνυμικής κατανομής με άγνωστη παράμετρο θ.

Άσκηση 5.4 Ένα δείγμα από 6 σοδειές έδωσε τα εξής αποτελέσματα σχετικά με το ύψος κάθε μίας από τις σοδειές.

5,3 5,6 5,9 6,1 6,2 6,5.

Αν υποθέσουμε ότι το μοντέλο μας είναι κανονικό με ομοιόμορφη μη κατάλληλη a-priori, υπολογίστε τις 90% HDR περιοχές για τον μέσο του πληθυσμού υποθέτοντας ότι: α) η διακύμανση του πληθυσμού είναι 1, β)η διακύμανση του πληθυσμού είναι άγνωστη.

Άσκηση 5.5 Έστω τυχαίο δείγμα $x_1,\ldots,x_n$ από πληθυσμό

$$f(x \mid \theta) = \theta^x (1-\theta)^{1-x}, \qquad (0 < \theta < 1, \qquad x = 0,1).$$

Αν η a-priori κατανομή του *θ* είναι ομοιόμορφη στο διάστημα (0,1), και η συνάρτηση απώλειας είναι τετραγωνική, να βρεθεί ο εκτιμητής Bayes της *θ*.

Άσκηση 5.6 Στην προηγούμενη άσκηση να βρεθεί ο εκτιμητής Bayes του *θ,* αν η a-priori κατανομή είναι Βήτα με παραμέτρους α και β.

Άσκηση 5.7 Έστω ότι έχουμε μία παρατήρηση x από διωνυμική κατανομή με παραμέτρους n, θ και ότι η a-priori κατανομή του θ είναι ομοιόμορφη στο διάστημα (0,1). Να δειχθεί ότι

$$p(\theta \mid x) = \binom{n}{x}(n+1)\theta^{x}(1-\theta)^{n-x}, \qquad (x = 0,1,\ldots,n)$$

και ότι για τετραγωνική συνάρτηση απώλειας ο εκτιμητής Bayes είναι

$$\hat{\theta} = \frac{x+1}{n+2}.$$

Άσκηση 5.8 Ένα τυχαίο δείγμα από n αντικείμενα συλλέγεται από μία παρτίδα προιόντων, και Γ από αυτά βρίσκονται ελαττωματικά. Εάν το ποσοστό των ελαττωματικών *θ* έχει a-priori κατανομή ανάλογη του $\theta^{\alpha}(1-\theta)^{\beta}$, καθορίστε την a-posteriori κατανομή του *θ*.

Έστω α=β=1 και σε δείγμα 13 αντικειμένων, βρίσκουμε 3 ελαττωματικά. Ορίσατε ένα 95% υψηλότατο διάστημα εμπιστοσύνης κατά Bayes για το *θ*.

Άσκηση 5.9 Τί έλεγχο θα χρησιμοποιούσατε για τον έλεγχο *θ*=0 κατά του *θ*=10 δεδομένης μίας παρατήρησης από (α) $N(\theta,10^2)$, (β) $N(\theta,1^2)$;

Κεφάλαιο 6ο

Πρόγνωση

6.1 Η Κατανομή Πρόγνωσης

Ως τώρα έχουμε επικεντρώσει την προσοχή μας στην εκτίμηση των παραμέτρων. Έχουμε δηλαδή, καθορίσει ένα μοντέλο πιθανότητας με σκοπό να περιγράψει την τυχαία διαδικασία την οποία έχει προκληθεί από τα δεδομένα μας, και έχουμε ακόμα δείξει πώς η Μπευζιανή θεωρία συνδυάζει την πληροφορία την οποία παρέχει το δείγμα και την a-priori πληροφόρηση η οποία χρησιμοποιείται στην εκτίμηση των παραμέτρων με την μορφή της a-posteriori κατανομής. Συνήθως, ο λόγος για τον οποίο «προτείνονται» τα στατιστικά μοντέλα, είναι η ***διεξαγωγή προβλέψεων*** σχετικά με τις μελλοντικές τιμές της διαδικασίας. Την διαδικασία αυτή μπορούμε να την χειριστούμε πολύ καλύτερα στην Μπευζιανή στατιστική απ' ότι στην κλασική θεωρία. Το βασικό αυτό σημείο, είναι ότι στην διεξαγωγή προβλέψεων για μελλοντικές τιμές με βάση την εκτίμηση μοντέλων, υπάρχουν δύο πηγές αβεβαιότητας:

- Αβεβαιότητα στις τιμές της παραμέτρου οι οποίες έχουν εκτιμηθεί με βάση προηγούμενα δεδομένα, και
- Αβεβαιότητα η οποία οφείλεται στο γεγονός ότι κάθε μελλοντική τιμή είναι από μόνη της ένα τυχαίο γεγονός.

Στην κλασική στατιστική, είναι σύνηθες να προσαρμόζουμε ένα μοντέλο (fit a model) στα ήδη υπάρχοντα δεδομένα, και να κάνουμε προβλέψεις με την υπόθεση ότι το μοντέλο μας είναι σωστό. Η διαδικασία αυτή ονομάζεται ***εκτιμητική προσέγγιση (estimative approach).*** Αυτή η προσέγγιση όμως στηρίζεται μόνο στην δεύτερη πηγή αβεβαιότητας, και οδηγεί σε εκτιμητές οι οποίοι πιστεύεται ότι είναι περισσότερο ακριβείς απ' ότι είναι στην πραγματικότητα. Δεν υπάρχει ένας ολοκληρωτικά ικανοποιητικός τρόπος για το πρόβλημα αυτό στο πλαίσιο της κλασικής στατιστικής, από την στιγμή που οι παράμετροι δεν θεωρούνται και δεν συμπεριφέρονται σαν τυχαίες μεταβλητές.

Αντίθετα, στο πλαίσιο της θεωρίας του Bayes οι δύο πηγές αβεβαιότητας καταμερίζονται στην αβεβαιότητα της εκτίμησης των παραμέτρων (η πληροφορία των οποίων περιλαμβάνεται στην a-posteriori κατανομή).

Ας υποθέσουμε λοιπόν ότι έχουμε τα δεδομένα $X=(x_1,\ldots x_n)$ μίας μεταβλητής με συνάρτηση πυκνότητας πιθανότητας (ή συνάρτηση πιθανοφάνειας) $p(x|\theta)$, και θέλουμε να εξάγουμε συμπεράσματα σχετικά με την κατανομή της μελλοντικής τιμής y μέσα από την διαδικασία αυτή. Με μία a-priori κατανομή $f(\theta)$, το θεώρημα του Bayes οδηγεί σε μία a-posteriori κατανομή $f(\theta|x)$. Η συνάρτηση πρόβλεψης για το y δεδομένου του x θα είναι:

$$f(y \mid x) = \int f(y \mid \theta) f(\theta \mid x) d\theta.$$

Επομένως η συνάρτηση πρόβλεψης είναι ένα ολοκλήρωμα της πιθανοφάνειας (για μία και μόνο παρατήρηση) πολλαπλασιαζόμενο με την a-posteriori κατανομή. Για μία ακόμα φορά είναι αναγκαίο να τονίσουμε ότι ο καθορισμός αυτός έχει απλά κατασκευαστεί από τους κλασικούς κανόνες του χειρισμού των πιθανοτήτων, και ο καθορισμός αυτός από μόνος του αποτελεί μία άμεση ερμηνεία με όρους πιθανοτήτων.

Η προσέγγιση στην κλασική στατιστική θα ήταν για παράδειγμα, η εύρεση του εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου θ και η συμπερασματολογία θα βασιζόταν στην κατανομή $f(y \mid \hat{\theta})$, δηλαδή την κατανομή εκτίμησης. Ας δώσουμε έμφαση και πάλι στο γεγονός ότι αυτό ακριβώς είναι που δεν επιτρέπει την ύπαρξη μεταβλητότητας σαν αποτέλεσμα της εκτίμησης του θ, και έτσι παρέχει μία λανθασμένη εικόνα ακρίβειας.

6.2 Παραδείγματα

Αν και η θεωρία στην οποία στηρίζεται η διεξαγωγή προβλέψεων είναι απλή, οι αριθμητικοί υπολογισμοί μπορούν να είναι ιδιαίτερα δύσκολοι. Παρόλα αυτά, πολλές συζυγείς οικογένειες κατανομών με μορφή a-priori-πιθανοφάνειας οδηγούν σε «βολικές» μορφές για τις συναρτήσεις πρόγνωσης.

✍ **Παράδειγμα 6.1** (Διωνυμικό δείγμα) Ας υποθέσουμε ότι έχουμε κάνει κάποια παρατήρηση x~Bin(n,θ) και η συζυγής a-priori για το *θ* είναι *θ*~Be(p,q). Έχουμε ήδη δείξει ότι η a-posteriori κατανομή για το *θ* δίνεται από την σχέση:

$$\theta|x \sim \text{Beta}(\, p+x,\ q+n-x\,).$$

Τώρα ας υποθέσουμε ότι έχουμε σκοπό να κάνουμε ακόμα N παρατηρήσεις στο μέλλον και για αυτό ας θεωρήσουμε z τον αριθμό των επιτυχιών στις N μελέτες έτσι ώστε z|θ~Bin(N,θ).

Για την μελλοντική μας παρατήρηση θα έχουμε την πιθανοφάνεια:

$$f(z\mid\theta) = {}^{N}C_z \theta^z (1-\theta)^{N-z}$$

Οπότε για τις z=0,1,…,N, θα είναι:

$$f(z\mid x) = \int_0^1 {}^{N}C_z \theta^z (1-\theta)^{N-z} \times \frac{\theta^{p+x-1}(1-\theta)^{q+n-x-1}}{B(p+x,q+n-x)}\, d\theta$$

$$= {}^{N}C_z \frac{B(z+p+x, N-z+q+n-x)}{B(p+x,q+n-x)}.$$

Αυτή η κατανομή είναι γνωστή σαν Βήτα-Διωνυμική κατανομή.

✍ **Παράδειγμα 6.2** (Γάμμα δείγμα) Όπως και στο κεφάλαιο 2, ας υποθέσουμε ότι $x_1,\ldots x_n$ είναι ανεξάρτητες μεταβλητές που ακολουθούν Γάμμα κατανομή Ga(κ,θ), όπου το κ είναι γνωστό, και έχουμε την συζυγή a-priori για την παράμετρο *θ* όπου *θ*~ Ga(p,q)

$$f(\theta) \propto \theta^{p-1} \exp\{-q\theta)$$

η οποία οδηγεί μέσα από το θεώρημα του Bayes στην θ|x~Ga(p+nκ,q+Σx_i)=Ga(G,H).

Η πιθανοφάνεια για την μελλοντική παρατήρηση y είναι

$$f(y\,|\,\theta)=\frac{\theta^{k}y^{k-1}\exp\{-\theta y\}}{\Gamma(k)}$$

και έτσι

$$f(y\,|\,x)=\int_{0}^{\infty}\frac{\theta^{k}y^{k-1}\exp\{-\theta y\}}{\Gamma(k)}\times\frac{H^{G}\theta^{G-1}\exp(-H\theta)}{\Gamma(G)}d\theta$$

$$=\frac{H^{G}y^{k-1}}{Be(k,G)(H+y)^{G+k}},\qquad (y>0)$$

6.3 Ασκήσεις

Άσκηση 6.1 Ένα τυχαίο δείγμα $x_1,\ldots x_n$ ακολουθεί κατανομή Poisson(θ). Η a-priori κατανομή του *θ* είναι Γάμμα Ga(g,h). Δείξτε ότι η κατανομή πρόγνωσης για την μελλοντική παρατήρηση y, από κατανομή Poisson(θ) είναι

$$\Pr(y\,|\,x)=\binom{y+G-1}{G-1}\left(\frac{1}{1+H}\right)^{y}\left(1-\frac{1}{1+H}\right)^{G},\qquad y=0,1,....,$$

για κάποια τιμή των G και H. Ποια είναι αυτή η κατανομή;

Άσκηση 6.2 Τα βάρη των προϊόντων κάποιας συγκεκριμένης παραγωγικής διαδικασίας είναι ανεξάρτητα και κατανεμημένα έτσι ώστε κάθε ένα από αυτά να ακολουθεί N(θ,4) κατανομή. Ο διευθυντής παραγωγής πιστεύει ότι το *θ* διαφέρει από παρτίδα σε παρτίδα με N(110,0.4) κατανομή. Ένα δείγμα από 5 προϊόντα επιλέγεται τυχαία, και για το δείγμα αυτό οι μετρήσεις έδωσαν

108 109 107,4 109,6 112.

Προσδιορίστε την a-posteriori κατανομή για την παράμετρο *θ*. Βρείτε ακόμα την κατανομή πρόγνωσης για α) το βάρος για ένα ακόμα προϊόν από την παρτίδα β) τον δειγματικό μέσο των βαρών για m προϊόντα ακόμα από την παρτίδα. Τί συμβαίνει στην περίπτωση β) όταν $m \to \infty$;

Άσκηση 6.3 Η κατανομή των ελαττωμάτων κατά μήκος ενός συνθετικού υφάσματος είναι Poisson, έτσι ώστε ο αριθμός των ελαττωμάτων σε λ μήκος του υφάσματος να είναι Poisson(λθ). Πολύ λίγα είναι γνωστά για την παράμετρο θ. Ο αριθμός των ελαττωμάτων σε 5 υφάσματα μήκους 10 15 25 30 και 40 μέτρα αντίστοιχα ήταν 3 2 7 6 και 10. Βρείτε την κατανομή πρόγνωσης για τον αριθμό των ελαττωμάτων σε ένα άλλο ύφασμα μήκους 60 μέτρων.

Άσκηση 6.4 Έστω ότι x είναι μία παρατήρηση από διωνυμική κατανομή με παραμέτρους n και π. Υποθέτουμε ότι η a-priori κατανομή της π είναι Βήτα με παραμέτρους α, β. Έστω ότι y είναι η επόμενη παρατήρηση και προέρχεται από διωνυμική κατανομή με παραμέτρους m και π. Τί μπορούμε να ξέρουμε για την y όταν συνδυάσουμε την a-priori γνώση μας με την μοναδική παρατήρηση x;

Άσκηση 6.5 Έστω ένα ενδεχόμενο που σε n δοκιμές συνέβη n φορές και δεν απέτυχε καμία. O Laplace ισχυρίστηκε ότι η πιθανότητα να συμβεί το ενδεχόμενο στην επόμενη δοκιμή είναι (n+1)/(n+2). Δικαιολογείστε τον ισχυρισμό αυτό υποθέτοντας ότι η a-priori κατανομή για την πιθανότητα του ενδεχομένου είναι ομοιόμορφη στο διάστημα (0,1).

Άσκηση 6.6 Ένα τυχαίο δείγμα μεγέθους m από την $N(\theta,\sigma^2)$, με σ^2 γνωστή και θ να έχει a-priori U(0,1), έχει δειγματικό μέσο $\overline{x}$. Αποδείξτε ότι η κατανομή του μέσου $\overline{y}$ από ένα δεύτερο ανεξάρτητο τυχαίο δείγμα μεγέθους n από την ίδια κατανομή, δεδομένης της τιμής του $\overline{x}$, (δηλ. η $P(\overline{y}\,|\,\overline{x},\sigma^2)$ είναι

$$N(\overline{x},\sigma^2(m^{-1}+n^{-1})).$$

Ένας επιστήμονας χρησιμοποιεί ένα όργανο με τυπική απόκλιση 0.12 και παίρνει 9 ανεξάρτητες μετρήσεις της ίδιας ποσότητας. Εάν ο μέσος των μετρήσεων είναι 17.653, αποκτείστε τα όρια μέσα στα οποία μία δέκατη μέτρηση θα είναι με πιθανότητα 99%.

Κεφάλαιο 7ο

Ασυμπτωτική Θεωρία

7.1 Εισαγωγή

Ας θυμηθούμε την ανάλυση της συζυγούς κανονικής κατανομής με μέσο θ και $x_1,\ldots x_n \sim N(\theta,1/\tau)$, η οποία για την a-priori $\theta \sim N(b,1/c)$ έδωσε:

$$\theta \mid x \sim N\left(\frac{cb + n\tau\bar{x}}{c+n\tau}, \frac{1}{c+n\tau}\right)$$

Καθώς όμως το $n\to\infty$, η παραπάνω μορφή γίνεται:

$$\theta \mid x \sim N(\bar{x}, 1/(n\tau)) = N(\bar{x}, \sigma^2/n)$$

Καθώς το n αυξάνει, η επιρροή της a-priori «εξαφανίζεται», και η a-posteriori καθορίζεται αποκλειστικά από τα δεδομένα και μόνο. Επίσης, η a-posteriori κατανομή συγκεντρώνεται γύρω από τον μέσο, ο οποίος σύμφωνα με τον κανόνα των μεγάλων αριθμών, πλησιάζει την πραγματική τιμή της παραμέτρου θ. Τα επιχειρήματα αυτά, αναλύονται και γενικεύονται στις επόμενες παραγράφους.

7.2 Ασυμπτωτική Κανονικότητα

Όταν η θ είναι συνεχής, όσα αναφέραμε παραπάνω μπορούν να επεκταθούν με σκοπό να βρεθεί κάποια «φόρμα» προσέγγισης για την περίπτωση που το n είναι μεγάλο. Με βάση όσα αναφέρθηκαν στην προηγούμενη παράγραφο, καθώς το n αυξάνει, ο λόγος $\exp\{nL_n(\theta o,x)\}/\exp\{nL_n(\theta,x)\}$ γίνεται αρκετά μικρός ιδιαίτερα για μία μικρή περιοχή του θo. Το $f(\theta)$ μπορεί να θεωρηθεί σαν σταθερά στο διάστημα αυτό και έτσι:

$$f(\theta \mid x_1,...,x_n) \propto \exp(n\overline{L}(\theta))$$

Επιπλέον, αν αναπτύξουμε το $\overline{L}(\theta)$ σαν σειρά Taylor για το *θο* έχουμε:

$$\overline{L}(\theta) \approx \overline{L}(\theta_o) - (\theta-\theta_0)^2/(2u) \qquad \text{όπου}$$

$$u = -1/\overline{L}''(\theta_0) = [I_n(\theta_0)]^{-1}$$

το αντίστροφο της πληροφορίας για το *θο*. Τελικά βρίσκουμε την προσέγγιση

$$f(\theta \mid x_1,...,x_n) \propto \exp\{-(\theta-\theta_0)^2\}/(2u)\}$$

$$\text{δηλαδή} \qquad \theta \mid x \sim N(\theta_0, [I_n(\theta_0)]^{-1})$$

Δηλαδή, καθώς το n→∞, η κατανομή της a-posteriori είναι κατά προσέγγιση κανονική για την πραγματική τιμή της παραμέτρου *θ*, με διακύμανση η οποία δίνεται από το $1/[I_n(\theta o)]$. Παρατηρήστε και πάλι ότι το αποτέλεσμα αυτό ισχύει ανεξάρτητα από τους a-priori καθορισμούς, με την προϋπόθεση ότι η a-priori δεν είναι μηδέν στην πραγματική της τιμή.

Το αποτέλεσμα αυτό είναι χρήσιμο για πολλούς λόγους. Πρώτα απ' όλα, μπορεί να χρησιμοποιηθεί απευθείας για να προσεγγίσουμε τις a-posteriori πιθανότητες σε περιπτώσεις όπου οι υπολογισμοί για την a-posteriori κατανομή είναι πολύπλοκοι. Έπειτα, μπορεί να δώσει χρήσιμές αρχικές τιμές για αριθμητικούς υπολογισμούς όπου οι αναλυτικές λύσεις είναι δύσκολες. Πιο βασικό ωστόσο είναι ότι αποδεικνύει ότι όταν αποκτήσουμε αρκετά δεδομένα, σκέψεις που αφορούν την συγκεκριμένη a-priori που έχουμε επιλέξει, γίνονται άνευ σημασίας. Δύο άτομα μπορεί να έχουν καθορίσει διαφορετικές a-priori για να αντιπροσωπεύσουν τις πεποιθήσεις τους, όμως από την στιγμή που είναι διαθέσιμα αρκετά δεδομένα, τα συμπεράσματα που αφορούν την a-posteriori θα συμφωνούν.

7.4 Παραδείγματα

✍ **Παράδειγμα 7.1** (Κανονικός μέσος) Έστω $x_1,\ldots x_n$ ανεξάρτητες μεταβλητές από $N(\theta,\sigma^2)$ όπου το σ^2 είναι γνωστό. Ως γνωστόν, η πιθανοφάνεια θα έχει την μορφή:

$$f(x\,|\,\theta) \propto \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right\}$$

Έχουμε ότι:

$$\log(f(x\,|\,\theta)) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2 \quad \text{έτσι} \quad \text{ώστε}$$

$$\frac{d\log(f(x\,|\,\theta))}{d\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) \quad \text{και}$$

$$\frac{d^2\log(f(x\,|\,\theta))}{d^2\theta} = -n/\sigma^2.$$

Κατά συνέπεια ο εκτιμητής μέγιστης πιθανοφάνειας θα είναι:

$$\hat{\theta} = \overline{x} \quad \text{και} \quad I_n(\theta) = n/\sigma^2.$$

Ασυμπτωτικά, καθώς το n→∞, $\theta\,|\,x \sim N(\overline{x}, \sigma^2/n)$.

Αυτή είναι η πραγματική κατανομή για *κάθε* a-priori κατανομή η οποία χρησιμοποιεί μη μηδενική πιθανότητα γύρω από την πραγματική τιμή της *θ*.

Παράδειγμα 7.2 (Διωνυμικό δείγμα) Ας θεωρήσουμε και πάλι ένα μοντέλο πιθανοφάνειας x~Bin(n,θ). Ως γνωστόν θα είναι:

$$f(x\,|\,\theta) = {}^{N}C_x\theta^x(1-\theta)^{N-z} \qquad x = 0,\ldots,n$$

Έτσι:

$$\log(f(x|\theta)) = x\log\theta + (n-x)\log(1-\theta).$$

και συνεπώς:

$$\frac{d\log(f(x\,|\,\theta))}{d\theta} = \frac{x}{\theta} - \frac{(n-x)}{1-\theta} \quad \text{και}$$

$$\frac{d^2\log l(\theta)}{d^2\theta} = -\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}.$$

Κατά συνέπεια θα είναι: $\hat{\theta} = x / n$

και:

$$I_n(\hat{\theta}) = \frac{n\theta}{\theta^2} - \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.$$

Και καθώς το $n \to \infty$,

$$\theta \mid x \sim N\left(\frac{x}{n}, \frac{\frac{x}{n}(1-\frac{x}{n})}{n} \right).$$

7.3 Ασκήσεις

Άσκηση 7.1 Βρείτε την ασυμπτωτική a-posteriori κατανομή για την θ για κάθε ένα από τα δύο μοντέλα στην άσκηση 2.1.

Άσκηση 7.2 Βρείτε την ασυμπτωτική a-posteriori κατανομή για την παράμετρο b στο μοντέλο Pareto της άσκησης 3.3.

Κεφάλαιο 8ο

Μεθοδολογία MCMC

8.1 Markov Chain Monte Carlo (MCMC)

Η ολοκλήρωση είναι πολύ σημαντική στη στατιστική κατά Bayes. Για δεδομένα $\underline{\mathbf{y}}$ και παραμέτρους $\underline{\boldsymbol{\theta}}$ χρειάζεται να ολοκληρώσουμε σε διάσταση dim($\underline{\boldsymbol{\theta}}$) για να υπολογίσουμε αναμενόμενες τιμές της μορφής ,

$$E[g(\underline{\boldsymbol{\theta}}|\}y)] = \int g(\theta) f(\underline{\boldsymbol{\theta}} \mid y) d\theta$$

Η ιδέα του MCMC βασίζεται στο εξής: Έστω ότι έχουμε κάποια κατανομή $\pi(x),\ \ x \in E \subseteq R^d$ από την οποία θέλουμε να πάρουμε δείγμα και η οποία είναι γνωστή χωρίς να ξέρουμε την σταθερά κανονικοποίησης. Μπορούμε να κατασκευάσουμε μια αλυσίδα Markov με χώρο καταστάσεων E, της οποίας η κατανομή ισορροπίας είναι η $\pi(x)$. Αν προσομοιώσουμε τιμές $\{x^t : t = 0,1,2,...\}$ από αυτή την αλυσίδα, ασυμπτωτικά υπάρχουν αποτελέσματα που περιλαμβάνουν την σύγκλιση (κατά κατανομή) των X^t :

$$X^t \xrightarrow[t\to\infty]{d} \pi(x)$$

και την συνέπεια του εργοδικού μέσου όρου για οποιαδήποτε συνάρτηση $g(.)$:

$$\frac{1}{n}\sum_{t=1}^{n} g(x^t) \xrightarrow{n\to\infty} E_n[g(x)] \quad \text{σχεδόν σίγουρα (a.s.)}$$

8.2 Αλυσίδες Markov

Ενδιαφερόμαστε για αλυσίδες Markov ορισμένες σε συνεχή χώρο καταστάσεων. Μια αλυσίδα Markov προσομοιώνεται ως εξής: στο χρόνο t προσομοιώνουμε την νέα κατάσταση της αλυσίδας από μια σ.π.π που εξαρτάται μόνο από το x^t :

$$x^{t+1} \sim K(x^t, x) \qquad (= K(x \mid x^t))$$

Η K ονομάζεται " transition kernel " της αλυσίδας.

Παράδειγμα 8.1 Έστω μία αλυσίδα Markov με transition Kernel

$$X^{t+1} \sim N(\frac{1}{2}x^t, 1)$$

Δοκιμάστε να προσομοιώσετε από αυτή την αλυσίδα με αρχικές τιμές $x^0 = 20$, $x^0 = 0$, $x^0 = -20$. Παρατηρείστε ότι τα μονοπάτια (paths) της αλυσίδας έχουν μετά από μερικές επαναλήψεις, τον ίδιο σχεδιασμό (pattern).
Τι γίνεται όταν $t \to \infty$;
Κάτω από γενικές συνθήκες, η αλυσίδα συγκλίνει στην κατανομή ισορροπίας:

$$P(x^t \in A) \to \pi(\mathrm{A}) \quad \forall \mathrm{A} \in \mathrm{E}.$$

Στο παραπάνω παράδειγμα, η κατανομή ισορροπίας π, είναι $x^t \sim N(0, 1.33)$, η οποία είναι ανεξάρτητη της αρχικής τιμής x^0. Η κατανομή ισορροπίας είναι μοναδική.

8.3 Δειγματολήπτης Gibbs (Gibbs sampler)

Ένας απλός τρόπος κατασκευής μιας transition kernel είναι ο δειγματολήπτης Gibbs. Έστω μια πολυμεταβλητή κατανομή $\pi(\underline{\mathbf{x}})$, $\underline{\mathbf{x}} = (x_1, ..., x_p)$. Μια αλυσίδα Markov που συγκλίνει στην κατανομή ισορροπίας $\pi(\underline{x})$ είναι η εξής:

Έστω $x^0 = (x_1^0, x_2^0, ..., x_p^0)$ μία οποιαδήποτε αρχική τιμή.

Μία επανάληψη του δειγματολήπτη Gibbs προσομοιώνει με τον παρακάτω τρόπο:

$$\begin{gathered}
x_1^1 \sim \pi(x_1 \mid x_1^0, x_2^0, ..., x_p^0) \\
x_2^1 \sim \pi(x_2 / x_1^1, x_3^0, ..., x_p^0) \\
. \\
. \\
. \\
x_i^1 \sim \pi(x_i \mid x_j^1, j < i, x_j^0, j > i) \\
. \\
. \\
. \\
x_p^1 \sim \pi(x_p \mid x_1^1, x_2^0, ..., x_{p-1}^0)
\end{gathered}$$

Παράδειγμα 8.2 Έστω $y_1, ..., y_n$ ισόνομες παρατηρήσεις από την κατανομή $N(\mu, \tau^{-1})$. Η πιθανοφάνεια δίνεται από την

$$f(\mathbf{y} \mid \mu, \tau) \propto \tau^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right).$$

Έστω a-priori κατανομές για το μ και τ της μορφής

$$f(\mu,\tau) = f(\mu)f(\tau) \propto \exp\left[\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right]\cdot e^{-\beta_0\tau}\tau^{\alpha_0-1}, \quad \mu\in\mathbb{R}, \tau>0.$$

Τότε έχουμε ότι

$$f(\mu,\tau\,|\,\mathbf{y}) \propto \tau^{\frac{n}{2}+\alpha_0-1}\exp\left[-\beta_0\tau-\frac{\tau S_n}{2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right]$$

$$\text{όπου } \; S_n=\sum_{i=1}^{n}(y_i-\mu)^2.$$

Για να προσομοιώσουμε από την $f(\mu,\tau/\mathbf{y})$ αρκεί να πάρουμε διαδοχικά δείγμα από τις

$$\mu\,|\,\tau,\underline{\mathbf{y}} \sim N\left(\frac{n\overline{y}\tau+\mu_0\tau_0}{n\tau+\tau_0},(n\tau+\tau_0)^{-1}\right), \quad \tau_0=\sigma_0^{-2}$$

$$\tau\,|\,\mu,\underline{\mathbf{y}} \sim Gamma\left(a_0+\frac{n}{2},\beta_0+\frac{S_n}{2}\right).$$

8.4 Ο Αλγόριθμος Metropolis

Αυτός ο αλγόριθμος είναι πιο γενικός από τον δειγματολήπτη Gibbs (στην πραγματικότητα ο δειγματολήπτης Gibbs είναι ειδική περίπτωση του αλγόριθμου Metropolis). Στην γενική του μορφή γράφεται ως εξής:

1. Έστω ότι στον χρόνο t η αλυσίδα είναι στην κατάσταση $x^t = x.$ Προσομοιώνουμε μια νέα τιμή y από την κατανομή $q(x,y)$ (κατανομή πρότασης).

2. Με πιθανότητα

$$a(x,y)=\min\left(1,\frac{\pi(y)q(y,x^t)}{\pi(x^t)q(x^t,y)}\right)$$

   θέτουμε $x^{t+1}=y.$ Αλλιώς θέτουμε $x^{t+1}=x.$

3. Go to 1.

Παράδειγμα 8.3 Έστω ότι θέλουμε να προσομοιώσουμε από τη $N(0,1)$ και επιλέγουμε

$$q(x, y) \equiv N(x, \sigma^2)$$

για κάποια τιμή του σ^2. Τότε

$$a(x, y) = \exp\left(-\frac{1}{2}(y^2 - x^2)\right).$$

8.4.1 Σημείωση

Αν η $q(x, y)$ είναι συμμετρική, δηλ. $q(x, y) = q(y, x)$, τότε

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

8.5 Ασκήσεις

Άσκηση 8.1

(α) Προσομοιώστε από την $\pi(x) \propto e^{-\frac{x^2}{2}}$ χρησιμοποιώντας την

$$q(x, y) \propto \exp\left(-\frac{(y-x)^2}{0.5}\right).$$

(β) Προσομοιώστε από την $\pi(x) \propto \left(1 + \frac{x^2}{3}\right)^{-2}$ χρησιμοποιώντας την

$$q(x, y) \propto \exp\left(-\frac{y^2}{0.5}\right)$$

(γ) Σχολιάστε τα αποτελέσματα της προσομοίωσης. Συγκλίνουν γρήγορα οι αλυσίδες;
Μπορείτε να υποπτευθείτε πως οι επιλογές της q επηρεάζουν το ρυθμό σύγκλισης;

Άσκηση 8.2 Να περιγράψετε πως θα προσομοιώνατε από την

$$\pi(x_1, x_2) = \exp\left[-\frac{1}{2}(x_1 - 1)^2 (x_2 - 2)^2\right]$$

με το δειγματολήπτη Gibbs και τον Metropolis. Μετά να κατασκευάσετε ένα αλγόριθμο που είναι μίξη Gibbs και Metropolis ως εξής:
Για την κατανομή $f(x_1 \mid x_2)$ προσομοιώνετε κανονικά όπως θα κάνατε στον δειγματολήπτη Gibbs. Υποθέστε όμως ότι δεν ξέρετε την κατανομή $f(x_2 \mid x_1)$ και πάρτε δείγμα από αυτήν χρησιμοποιώντας τον αλγόριθμο Metropolis.