Εθνικο και Καποδιστριακο Πανεπιστημιο
Αθηνων

Μπεϋζιανή Συμπερασματολογία

Διδάσκουσα: Λουκία Μελιγκοτσίδου

Σχολή Θετικών Επιστημών
Τμήμα Μαθηματικών
2020

Περιεχόμενα

Πρόλογος **5**

1 Εισαγωγή **7**

1.1 Τι είναι η στατιστική συμπερασματολογία; . . . 7
1.2 Τι είναι η Μπεϋζιανή συμπερασματολογία; . . . 8
1.3 Η prior κατανομή . . . 10
1.4 Χαρακτηριστικά της Μπεϋζιανής προσέγγισης . . . 12
1.5 Ενστάσεις κατά της Μπεϋζιανής προσέγγισης . . . 13
1.6 Επανάληψη στο θεώρημα του Bayes . . . 13
1.7 Ασκήσεις . . . 16

2 Μπεϋζιανή Ανανέωση Πληροφορίας **19**

2.1 Εισαγωγή . . . 19
2.2 Το θεώρημα του Bayes . . . 20
2.3 Ζητήματα . . . 21
2.3.1 Επιλογή μοντέλου πιθανοφάνειας . . . 21
2.3.2 Επιλογή prior . . . 21
2.3.3 Υπολογισμός . . . 22
2.3.4 Συμπερασματολογία . . . 22
2.4 Παραδείγματα . . . 22
2.5 Γενικά ζητήματα . . . 29
2.5.1 Συνεχής ανανέωση πληροφορίας . . . 30
2.5.2 Επάρκεια . . . 30
2.5.3 Η αρχή της πιθανοφάνειας . . . 31
2.6 Ασκήσεις . . . 31

3 Προσδιορισμός Prior Κατανομών **35**

3.1 Εισαγωγή . . . 35
3.2 Συζυγείς prior κατανομές . . . 35
3.2.1 Χρήση συζυγών prior . . . 36
3.2.2 Προσδιορισμός συζυγών prior . . . 36

3.2.3 Τυπικές συζυγείς αναλύσεις . . . 38
3.3 Μίξεις prior κατανομών . . . 38
3.4 Ακατάλληλες prior κατανομές . . . 40
3.5 Η prior του Jeffreys . . . 41
3.5.1 Παραδείγματα . . . 43
3.6 Εξαγωγή της prior κατανομής . . . 44
3.7 Ασκήσεις . . . 45

4 Πολυπαραμετρικά Προβλήματα **47**
4.1 Εισαγωγή . . . 47
4.2 Παραδείγματα . . . 49
4.3 Ασκήσεις . . . 51

5 Σύνοψη της Posterior Πληροφορίας **53**
5.1 Θεωρία αποφάσεων . . . 53
5.2 Σημειακή εκτίμηση . . . 56
5.2.1 Τετραγωνική συνάρτηση απώλειας . . . 57
5.2.2 Συνάρτηση απώλειας απολύτου σφάλματος . . . 58
5.2.3 Συνάρτηση απώλειας 0-1 . . . 59
5.2.4 Συμπέρασμα . . . 59
5.3 Διαστήματα και περιοχές αξιοπιστίας . . . 59
5.4 Έλεγχοι υποθέσεων . . . 63
5.5 Μπεϋζιανή σύγκριση μοντέλων . . . 64
5.6 Ασκήσεις . . . 69

6 Πρόβλεψη **71**
6.1 Η κατανομή πρόβλεψης . . . 71
6.2 Ασκήσεις . . . 74

7 Ασυμπτωτική Ανάλυση **77**
7.1 Εισαγωγή . . . 77
7.2 Συνέπεια . . . 77
7.3 Ασυμπτωτική κανονικότητα . . . 78
7.4 Ασκήσεις . . . 80

Πρόλογος

Οι παρούσες σημειώσεις αποτελούν ελεύθερη απόδοση των σημειώσεων της καθηγήτριας Λουκίας Μελιγκοτσίδου με τίτλο "Bayesian Inference". Έχουν προσαρμοστεί και εμπλουτιστεί με βάση τις παραδόσεις της καθηγήτριας στο προπτυχιακό μάθημα "Μπεϋζιανή Στατιστική και Εφαρμογές". Την ευχαριστούμε για τη σημαντική προσφορά της στον τομέα Στατιστικής και Επιχειρησιακής Έρευνας του τμήματος όλα αυτά τα χρόνια.

Οι σημειώσεις περιέχουν ένα εισαγωγικό κεφάλαιο με μία επισκόπηση των πυλώνων της Μπεϋζιανής συμπερασματολογίας, καθώς και των σημείων που τη διαφοροποιούν από την κλασική συμπερασματολογία, η οποία μελετάται στο μάθημα της μαθηματικής στατιστικής. Προτείνεται κάποιος να έχει κατανοήσει σε ικανοποιητικό βαθμό τα βασικά στοιχεία των πιθανοτήτων και να έχει μία πρότερη επαφή με τη μαθηματική στατιστική προτού ασχοληθεί περαιτέρω με το μάθημα.

Σε κάθε κεφάλαιο παρατίθεται η απαιτούμενη θεωρία, συνοδευόμενη από παραδείγματα που καλύπτουν ολόκληρο το φάσμα των εφαρμογών του μαθήματος. Ενδιάμεσα, παρεμβάλλονται αρκετές παρατηρήσεις που υποδεικνύουν συχνά λάθη των φοιτητών στην εφαρμογή των μεθόδων που παρουσιάζονται. Στο τέλος κάθε κεφαλαίου, παρατίθεται μία σειρά ενδεικτικών ασκήσεων. Προτείνεται οι φοιτητές που παρακολουθούν το μάθημα να ασχολούνται τακτικά με την επίλυσή τους κατά τη διάρκεια του ακαδημαϊκού εξαμήνου.

Ανδρεόυ Πάνος
Κατσιάνος Βασίλης

Κεφάλαιο 1

Εισαγωγή

1.1 Τι είναι η στατιστική συμπερασματολογία;

Προτού ορίσουμε την Μπεϋζιανή συμπερασματολογία θα πρέπει να θεωρήσουμε την ευρύτερη ερώτηση - ”τι είναι η στατιστική συμπερασματολογία;”. Πολλοί ορισμοί είναι δυνατό να δοθούν, αλλά οι περισσότεροι ανάγονται στη βασική αρχή ότι η στατιστική συμπερασματολογία είναι η επιστήμη της εξαγωγής συμπερασμάτων για έναν ”πληθυσμό” από ένα ”δείγμα”, δηλαδή από μονάδες που επιλέγονται από αυτόν τον πληθυσμό. Αυτό επισύρει αρκετές ερωτήσεις σχετικά με το τι εννοούμε με την έννοια του πληθυσμού, πώς το δείγμα σχετίζεται με τον πληθυσμό, πώς θα έπρεπε να εξάγουμε το δείγμα αν όλες οι επιλογές είναι διαθέσιμες και ούτω καθεξής. Θα αφήσουμε όμως όλα αυτά τα θέματα στην άκρη και θα εστιάσουμε τη συζήτησή μας σε ένα απλό παράδειγμα.

Ας υποθέσουμε ότι η δασική επιτροπή θέλει να εκτιμήσει το ποσοστό των δέντρων σε ένα μεγάλο δάσος που πάσχουν από μία συγκεκριμένη ασθένεια. Είναι μη-πρακτικό να ελέγξει κάθε δέντρο, οπότε επιλέγει ένα δείγμα από n δέντρα. Για άλλη μία φορά, δε συζητάμε εδώ πώς μπορούν να διαλέξουν το δείγμα τους, αλλά υποθέτουμε ότι η δειγματοληψία είναι τυχούσα, υπό την έννοια ότι αν θ είναι το ποσοστό των δέντρων του δάσους που έχουν την ασθένεια, τότε κάθε δέντρο στο δείγμα θα έχει την ασθένεια με πιθανότητα θ, ανεξάρτητα από όλα τα άλλα δέντρα στο δείγμα. Συμβολίζοντας με X την τυχαία μεταβλητή που αντιστοιχεί στο πλήθος των άρρωστων δέντρων του δείγματος, η επιτροπή θα χρησιμοποιήσει την παρατηρούμενη τιμή $X = x$ για να εξάγει μία συμπερασματολογία σχετικά με την παράμετρο θ του πληθυσμού. Αυτή η συμπερασματολογία θα μπορούσε να έχει τη μορφή μίας *σημειακής εκτίμησης* ($\hat{\theta} = 0.1$), ενός *διαστήματος εμπιστοσύνης* (95% εμπιστοσύνη ότι το θ βρίσκεται στο εύρος $[0.08, 0.12]$), ενός *ελέγχου υποθέσεων* (απορρίπτεται η υπόθεση $\theta < 0.07$ σε επίπεδο σημαντικότητας 5%) ή μίας *πρόβλεψης* (πρόβλεψη ότι 15% των δέντρων θα έχουν προσβληθεί μέχρι την επόμενη χρονιά).

Σε κάθε περίπτωση, η γνώση της παρατηρούμενης τιμής $X = x$ χρησιμοποιείται για να εξάγουμε συμπεράσματα για το χαρακτηριστικό θ του πληθυσμού. Επιπλέον, αυτά τα συμπεράσματα εξάγονται καθορίζοντας ένα μοντέλο πιθανότητας $f(x \mid \theta)$, το οποίο καθορίζει πώς είναι κατανεμημένες οι πιθανότητες των διαφορετικών τιμών της X, για μία δεδομένη τιμή του θ.

Εδώ κάτω από τις υποθέσεις περί τυχαίας δειγματοληψίας, το μοντέλο μας θα ήταν $X \mid \theta \sim \text{Bin}(n, \theta)$, δηλαδή:

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

Τότε, η στατιστική συμπερασματολογία αναφέρεται σε συμπερασματολογία σχετικά με την παράμετρο θ του πληθυσμού, έχοντας παρατηρήσει $X = x$, και στην ουσία θα συμπεραίναμε ότι τιμές του θ που δίνουν υψηλή πιθανότητα στην τιμή x που παρατηρήσαμε είναι ”πιο πιθανές” από αυτές που αποδίδουν στο x μικρή πιθανότητα - η αρχή της μέγιστης πιθανοφάνειας. (Παρατηρούμε ότι στο ευρύτερο πλαίσιό της, η στατιστική συμπερασματολογία εμπεριέχει επίσης τα ζητήματα της επιλογής μοντέλου, της επαλήθευσης μοντέλου και άλλα, αλλά θα περιορίσουμε την προσοχή μας στη συμπερασματολογία παραμέτρων μέσα σε μία παραμετρική οικογένεια μοντέλων.)

Πριν μεταβούμε ειδικά στην Μπεϋζιανή συμπερασματολογία, υπάρχουν ορισμένα σημεία που πρέπει να διασαφηνιστούν σχετικά με την κλασική προσέγγιση στη συμπερασματολογία. Το πιο θεμελιώδες είναι ότι η παράμετρος θ, αν και άγνωστη, αντιμετωπίζεται ως *σταθερά* και όχι ως *τυχαία* ποσότητα. Αυτός είναι ο ακρογωνιαίος λίθος της κλασικής θεωρίας, αλλά οδηγεί σε προβλήματα ερμηνείας. Θα θέλαμε το 95% διάστημα εμπιστοσύνης $[0.08, 0.12]$ για το θ να σημαίνει ότι υπάρχει μία 95% πιθανότητα το θ να είναι μεταξύ 0.08 και 0.12. *Δεν μπορεί* να σημαίνει κάτι τέτοιο, από τη στιγμή που το θ δεν είναι τυχαίο: είτε *βρίσκεται* στο διάστημα είτε *δε βρίσκεται* - πιθανότητα δεν υπεισέρχεται ούτε μπορεί να υπεισέλθει εδώ. Το μόνο τυχαίο στοιχείο σε αυτό το μοντέλο πιθανότητας είναι τα δεδομένα, οπότε η σωστή ερμηνεία του διαστήματος είναι ότι αν εφαρμόζαμε τη διαδικασία ”πολλές φορές”, τότε τα διαστήματα που θα κατασκευάζαμε ”μακροπρόθεσμα” θα περιείχαν το θ στο 95% των περιπτώσεων. Όλα τα συμπεράσματα που βασίζονται στην κλασική θεωρία είναι αναγκασμένα να έχουν αυτή τη μακροπρόθεσμη συχνοτιστική ερμηνεία, παρόλο που, για παράδειγμα, έχουμε μόνο το διάστημα $[0.08, 0.12]$ για να ερμηνεύσουμε.

1.2 Τι είναι η Μπεϋζιανή συμπερασματολογία;

Το γενικό πλαίσιο μέσα στο οποίο λειτουργεί η Μπεϋζιανή συμπερασματολογία είναι ταυτόσημο με το παραπάνω: υπάρχει μία πληθυσμιακή παράμετρος θ για την οποία θέλουμε να εξάγουμε συμπεράσματα και ένας πιθανοτικός μηχανισμός $f(x \mid \theta)$ που καθορίζει την πιθανότητα να παρατηρήσουμε διαφορετικά δεδομένα x, κάτω από διαφορετικές τιμές του θ. Η θεμελιώδης διαφορά ωστόσο είναι ότι μεταχειριζόμαστε το θ ως *τυχαία* ποσότητα.

Αυτό μπορεί να φαίνεται αρκετά αθώο, αλλά στην πραγματικότητα οδηγεί σε ουσιωδώς διαφορετική προσέγγιση στη στατιστική μοντελοποίηση και συμπερασματολογία.

Στην ουσία, η συμπερασματολογία μας θα βασιστεί στην $f(\theta \mid x)$ αντί της $f(x \mid \theta)$, δηλαδή στην κατανομή πιθανότητας της παραμέτρου δεδομένου του δείγματος αντί της κατανομής πιθανότητας του δείγματος δεδομένης της παραμέτρου. Με πολλούς τρόπους, αυτό οδηγεί σε πολύ πιο φυσική συμπερασματολογία, αλλά για να το πετύχουμε αυτό, θα δούμε ότι είναι απαραίτητο να καθορίσουμε μία *prior (εκ των προτέρων) κατανομή πιθανότητας* $f(\theta)$, η οποία αντιπροσωπεύει πεποιθήσεις σχετικά με την κατανομή του θ *προτού* αποκτήσουμε οποιαδήποτε πληροφορία για τα δεδομένα.

Αυτή η ιδέα μίας prior κατανομής για την παράμετρο θ βρίσκεται στην καρδιά της Μπεϋζιανής σκέψης και, αναλόγως με το αν μιλάει κάποιος σε έναν συνήγορο ή έναν αντίπαλο αυτής της μεθοδολογίας, είναι είτε το μεγαλύτερό της πλεονέκτημα έναντι της κλασικής θεωρίας είτε η μεγαλύτερή της παγίδα.

Το παράδειγμα της δειγματοληψίας νομισμάτων.

Αυτό το παράδειγμα είναι παρόμοιο σε ιδέα με το παράδειγμα της δασικής επιτροπής, αλλά επαρκώς απλό για να καταδείξει την Μπεϋζιανή προσέγγιση στη συμπερασματολογία. Σε αυτό το παράδειγμα, η χρήση της prior κατανομής είναι μη-αμφισβητίσιμη. Σημειώνουμε πως η πιθανοφάνεια εξακολουθεί να έχει κεντρικό ρόλο στην Μπεϋζιανή μέθοδο.

Πέντε νομίσματα έχουν τοποθετηθεί στο τραπέζι. Ζητείται από εσάς να εκτιμήσετε το ποσοστό θ αυτών των νομισμάτων που δείχνουν "Γ", κοιτάζοντας ένα δείγμα μόνο δύο νομισμάτων. Τώρα, υπάρχουν μόνο 6 πιθανές τιμές του θ, δηλαδή $\theta = \frac{M}{5}$ για $M = 0, 1, \ldots, 5$. Έστω X το πλήθος από "Γ" στο δείγμα των δύο νομισμάτων και ας υποθέσουμε ότι παρατηρούμε $x = 1$. Στη *δεύτερη* γραμμή του πίνακα 1.1, βρίσκονται οι πιθανότητες να παρατηρήσουμε αυτό το αποτέλεσμα, δεδομένης της (άγνωστης) τιμής του θ.

θ	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1
$P(X=1 \mid \theta)$	0	0.4	0.6	0.6	0.4	0
$f(\theta)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$P(X=1 \mid \theta) \cdot f(\theta) = f(X=1,\theta)$	0	$\frac{2}{32}$	$\frac{6}{32}$	$\frac{6}{32}$	$\frac{2}{32}$	0
$f(\theta \mid X=1)$	0	$\frac{4}{32}$	$\frac{12}{32}$	$\frac{12}{32}$	$\frac{4}{32}$	0

Πινακας 1.1: Posterior Κατανομή Πιθανότητας του θ

Για να βεβαιωθείτε ότι μπορείτε να υπολογίσετε αυτές τις πιθανότητες, θεωρήστε για παράδειγμα την $P\left(X = 1 \mid \theta = \frac{3}{5}\right)$.

Εφόσον $\theta = \frac{3}{5}$, έχουμε 3 "Γ" και 2 "Κ" σε ένα σύνολο 5 νομισμάτων. Βρίσκουμε 1 "Γ" και 1 "Κ" σε ένα δείγμα 2 νομισμάτων.

Το πλήθος τρόπων με τους οποίους μπορούμε να επιλέξουμε 1 ”Γ” και 1 ”Κ” από 3 ”Γ” και 2 ”Κ” είναι $\binom{3}{1} \cdot \binom{2}{1} = 3 \cdot 2 = 6$, ενώ το συνολικό πλήθος τρόπων με τους οποίους μπορούμε να επιλέξουμε 2 νομίσματα είναι $\binom{5}{2} = 10$. Επομένως, η πιθανότητα είναι $\frac{6}{10} = 0.6$.

Η δεύτερη γραμμή του πίνακα είναι η *πιθανοφάνεια* του θ. *Δεν* είναι κατανομή πιθανότητας - δεν αθροίζει στη μονάδα. Οι πιο πιθανές τιμές του θ είναι $\frac{2}{5}$ και $\frac{3}{5}$, ενώ οι τιμές 0 και 1 αποκλείονται τελείως.

Η Μπεϋζιανή προσέγγιση χρησιμοποιεί τη συνάρτηση πιθανοφάνειας, αλλά τη συνδυάζει με *prior* γνώση. Η τρίτη γραμμή του πίνακα δείχνει την κατανομή του θ που εύλογα θα υποθέταμε ακόμα και αν *δεν* επιτρεπόταν να δούμε το δείγμα των 2 νομισμάτων. Με άλλα λόγια, είναι η κατανομή που θα περιέγραφε εύλογα τη γνώση μας *προτού* παρατηρήσουμε το δείγμα. Δηλαδή, θα θεωρούσαμε ότι τα 2 νομίσματα είναι δίκαια και θα υπολογίζαμε τις εξής διωνυμικές πιθανότητες:

$$P(\theta = 0) = P(\theta = 1) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32},$$

$$P\left(\theta = \frac{1}{5}\right) = P\left(\theta = \frac{4}{5}\right) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32},$$

$$P\left(\theta = \frac{2}{5}\right) = P\left(\theta = \frac{3}{5}\right) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32},$$

Πολλαπλασιάζουμε τη δεύτερη γραμμή με την τρίτη για να πάρουμε μέρος της *από κοινού* κατανομής του X και του θ στην τέταρτη γραμμή. Αυτή δεν αθροίζει στη μονάδα, αφού είναι μόνο ένα μέρος της από κοινού κατανομής. Για την ακρίβεια, αθροίζει στην $P(X = 1)$.

Το σημαντικό βήμα είναι να βρούμε τη δεσμευμένη κατανομή του θ δεδομένου ότι $X = 1$, διαιρώντας την τέταρτη γραμμή με το άθροισμά της - την $P(X = 1)$. Το αποτέλεσμα δίνεται στην πέμπτη γραμμή. *Αθροίζει* στη μονάδα - το πετύχαμε αυτό κάνοντας τη διαίρεση, μία διαδικασία που ονομάζουμε *κανονικοποίηση* της κατανομής.

Η τελευταία γραμμή περιέχει όλη την πληροφορία που χρειαζόμαστε για να κάνουμε οποιοδήποτε από τα είδη συμπερασματολογίας για το θ τα οποία συζητήσαμε. Ονομάζεται (Μπεϋζιανή) posterior (εκ των υστέρων) κατανομή του θ. Μας λέει ότι οι τιμές $\frac{2}{5}$ και $\frac{3}{5}$ είναι ισοπίθανες, οπότε τυχαίνει να μας δίνει μία παρόμοια εικόνα με τη συνάρτηση πιθανοφάνειας σε αυτή την περίπτωση.

1.3 Η prior κατανομή

Σχεδόν πάντα, όταν προσπαθούμε να εκτιμήσουμε μία παράμετρο θ, έχουμε κάποια γνώση ή κάποια πεποίθηση σχετικά με την τιμή του θ προτού λάβουμε υπόψη τα δεδομένα. Ένα παράδειγμα από τον O'Hagan (1994) το καθιστά αυτό σαφές από ποιοτικής άποψης.

Κοιτάζετε έξω από το παράθυρό σας και βλέπετε ένα μεγάλο ξύλινο πράγμα με κλαδιά, καλυμμένο από μικρά πράσινα πράγματα. Σας περνάνε από το μυαλό δύο πιθανές υποθέσεις: η μία είναι ότι είναι ένα δέντρο, η άλλη ότι είναι ο ταχυδρόμος. Φυσικά, απορρίπτετε την υπόθεση ότι είναι ο ταχυδρόμος, καθώς οι ταχυδρόμοι συνήθως δε δείχνουν έτσι, ενώ τα δέντρα δείχνουν έτσι. Έτσι, σε επίσημη γλώσσα, συμβολίζοντας με A το ενδεχόμενο να δείτε ένα ξύλινο πράγμα με πράσινα στοιχεία, B_1 το ενδεχόμενο να είναι δέντρο και B_2 το ενδεχόμενο να είναι ο ταχυδρόμος, απορρίπτετε τη B_2 έναντι της B_1 διότι $P(A \mid B_1) > P(A \mid B_2)$. Εδώ, χρησιμοποιείτε την αρχή της μέγιστης πιθανοφάνειας.

Ωστόσο, μπορείτε να σας περάσει από το μυαλό και ένα τρίτο ενδεχόμενο B_3 ότι το πράγμα που βλέπετε είναι μία απομίμηση ενός δέντρου. Σε αυτή την περίπτωση, παρόλο που ισχύει ότι $P(A \mid B_1) = P(A \mid B_3)$, θα απορρίπτατε την υπόθεση B_3 έναντι της B_1. Αυτό συμβαίνει επειδή, παρόλο που η πιθανότητα να δείτε αυτό που παρατηρήσατε είναι ίδια είτε πρόκειται για δέντρο είτε για απομίμησή του, η *prior* πεποίθησή σας λέει ότι είναι πιθανότερο να είναι δέντρο παρά απομίμησή του, οπότε συμπεριλαμβάνετε αυτή την πληροφορία όταν παίρνετε την απόφασή σας.

Ας πάρουμε ένα άλλο παράδειγμα, όπου σε καθεμία από τις επόμενες περιπτώσεις το μοντέλο που ακολουθούν τα δεδομένα είναι $X \mid \theta \sim \text{Bin}(10, \theta)$ και παρατηρούμε $x = 10$, οπότε η υπόθεση $H_0 : \theta \leq 0.5$ απορρίπτεται έναντι της $H_1 : \theta > 0.5$ κάθε φορά:

1. Μία γυναίκα που πίνει τσάι ισχυρίζεται ότι μπορεί να ανιχνεύσει σε ένα φλιτζάνι με τσάι αν το γάλα προστέθηκε πριν ή μετά το τσάι. Το κάνει αυτό ορθά για δέκα φλιτζάνια.

2. Μία εμπειρογνωμόνων του μουσικού χώρου ισχυρίζεται ότι μπορεί να διακρίνει μεταξύ μίας σελίδας έργου του Hayden και μίας του Mozart. Κατηγοριοποιεί σωστά 10 κομμάτια.

3. Μία μεθυσμένη φίλη ισχυρίζεται ότι μπορεί να προβλέψει το αποτέλεσμα της ρίψης ενός δίκαιου νομίσματος και το καταφέρνει για 10 ρίψεις.

Τώρα, σε επίπεδο δεδομένων, θα ήμασταν αναγκασμένοι να εξάγουμε τα ίδια συμπεράσματα σε κάθε περίπτωση. Ωστόσο, οι prior πεποιθήσεις μας συνηγορούν στο ότι θα ήμασταν αρκετά σκεπτικοί στην περίπτωση της μεθυσμένης φίλης μας, σχετικά εντυπωσιασμένοι στην περίπτωση της γυναίκας που πίνει τσάι και καθόλου έκπληκτοι με την εμπειρογνωμόνων του μουσικού χώρου.

Το βασικό επιχείρημα είναι το εξής: τα πειράματα δεν είναι αφηρημένοι μηχανισμοί. Κατά κανόνα, έχουμε κάποια γνώση ως προς τη διαδικασία που θέλουμε να μελετήσουμε πριν συλλέξουμε τα δεδομένα. Είναι θεμιτό (αρκετοί θα έλεγαν απαραίτητο) τα συμπεράσματα να βασίζονται στη συνδυασμένη πληροφορία που αντιπροσωπεύουν η εκ των προτέρων γνώση *και* τα δεδομένα. Η Μπεϋζιανή συμπερασματολογία είναι ο μηχανισμός που βγάζει συμπεράσματα από αυτή τη συνδυασμένη γνώση.

Απλώς για να αναφέρουμε και την εναλλακτική άποψη, είναι ακριβώς αυτή η εξάρτηση από prior πεποιθήσεις στην οποία αντιτίθενται οι αντίπαλοι της Μπεϋζιανής οπτικής. Διαφορετικές prior πεποιθήσεις θα οδηγήσουν σε διαφορετικές συμπερασματολογίες με την Μπεϋζιανή θεώρηση των πραγμάτων και η αποδοχή του Μπεϋζιανού πλαισίου λειτουργίας καθορίζεται από το αν θεωρούμε καλό ή κακό αυτό το γεγονός.

1.4 Χαρακτηριστικά της Μπεϋζιανής προσέγγισης

Ακολουθώντας τον O'Hagan (1994), μπορούμε να αναγνωρίσουμε 4 θεμελιώδεις απόψεις που χαρακτηρίζουν την Μπεϋζιανή προσέγγιση στη στατιστική συμπερασματολογία:

- **Prior Πληροφορία.** Όλα τα προβλήματα είναι μοναδικά και έχουν τις δικές τους συνθήκες. Αυτές οι συνθήκες δημιουργούν prior πληροφορία και είναι η διατύπωση και εκμετάλλευση αυτής της prior πληροφορίας που κάνει την Μπεϋζιανή συμπερασματολογία να ξεχωρίζει από την κλασική στατιστική.

- **Υποκειμενική Πιθανότητα.** Η κλασική στατιστική βασίζεται σε έναν αντικειμενικό "μακροπρόθεσμο συχνοτιστικό" ορισμό των πιθανοτήτων. Ακόμα κι αν αυτό είναι επιθυμητό, πράγμα που είναι εύλογο, οδηγεί σε πολύπλοκες συμπερασματολογίες. Αντιθέτως, η Μπεϋζιανή στατιστική επισημοποιεί την αντίληψη ότι όλες οι πιθανότητες είναι υποκειμενικές - εξαρτώμενες από τις πεποιθήσεις και τη γνώση του καθενός. Έτσι, η Μπεϋζιανή ανάλυση είναι προσωποκεντρική - μοναδική ως προς τις prior πεποιθήσεις κάθε ατόμου. Η συμπερασματολογία βασίζεται στην *posterior* κατανομή $f(\theta \mid x)$, η μορφή της οποίας θα δούμε ότι εξαρτάται (μέσω του θεωρήματος του Bayes) από τον προσδιορισμό της prior $f(\theta)$.

- **Εσωτερική Συνοχή.** Χειριζόμενοι την παράμετρο θ ως τυχαία, προκύπτει ότι όλη η ανάπτυξη της Μπεϋζιανής συμπερασμτολογίας απορρέει αρκετά φυσικά μόνο από τη θεωρία πιθανοτήτων. Αυτό έχει πολλά πλεονεκτήματα και σημαίνει ότι όλα τα ζητήματα συμπερασματολογίας μπορούν να τεθούν ως πιθανοτικές εκφράσεις για το θ, οι οποίες τότε απορρέουν απευθείας από την posterior κατανομή. Θα δούμε ένα τέτοιο πλεονέκτημα όταν θα αναλογιστούμε τις προβλέψεις μελλοντικών παρατηρήσεων στο κεφάλαιο 6.

- **Απουσία Αυθαίρετων Κριτηρίων Απόφασης.** Επειδή η κλασική συμπερασματολογία δεν μπορεί να κάνει πιθανοτικές διατυπώσεις για το θ, διάφορα κριτήρια έχουν αναπτυχθεί για να κρίνουν αν μία συγκεκριμένη εκτιμήτρια είναι κατά κάποια έννοια "καλή". Αυτό έχει οδηγήσει σε μία πληθώρα διαδικασιών οι οποίες είναι συχνά αντικρουόμενες. Η Μπεϋζιανή συμπερασματολογία αποφεύγει αυτή την τάση εφεύρεσης *αυθαίρετων* κριτηρίων για την αξιολόγηση και σύγκριση εκτιμητριών, στηριζόμενη στην posterior κατανομή για να εκφράσει με σαφείς πιθανοτικούς όρους ολόκληρη τη συμπερασματολογία για την άγνωστη παράμετρο θ.

1.5 Ενστάσεις κατά της Μπεϋζιανής προσέγγισης

Οι κύριες ενστάσεις κατά της Μπεϋζιανής προσέγγισης, όπως περιγράφηκε παραπάνω, είναι ότι τα συμπεράσματα εξαρτώνται από τη συγκεκριμένη επιλογή της prior. Αυτό είναι ιδιαίτερα σημαντικό για επιστημονικές εργασίες, οι οποίες πρέπει να είναι αντικειμενικές και να προσπαθούν να παρουσιάζουν την πληροφορία για τα δεδομένα και όχι τις πεποιθήσεις του συγγραφέα.

Ωστόσο, υπάρχει συχνά prior πληροφορία η οποία μπορεί λογικά να διατυπωθεί σε μία prior κατανομή και σε τέτοιες περιπτώσεις η Μπεϋζιανή συμπερασματολογία είναι ξεκάθαρα η καλύτερη προσέγγιση.

Ακόμα και όταν δεν υπάρχει prior πληροφορία, είναι δυνατόν να κατασκευάσουμε prior κατανομές που αντανακλούν αυτή την έλλειψη πληροφορίας (τουλάχιστον σε κάποιον βαθμό - βλέπε κεφάλαιο 3). Δεδομένου ότι χρησιμοποιούνται τέτοιες μη-πληροφοριακές priors και υπάρχουν επαρκή δεδομένα, τότε η posterior θα μοιάζει πολύ με την πιθανοφάνεια και η επιλογή της prior θα έχει μικρή επίδραση στα συμπεράσματα που βγαίνουν.

1.6 Επανάληψη στο θεώρημα του Bayes

Στην απλή μορφή του, το θεώρημα του Bayes είναι ένα απλό αποτέλεσμα που αφορά δεσμευμένες πιθανότητες.

Αν A και B είναι δύο ενδεχόμενα με $P(A) > 0$, τότε:

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

Απόδειξη.

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

□

Η χρησιμότητα του θεωρήματος του Bayes σε εφαρμογές των πιθανοτήτων έγκειται στην αντιστροφή της δέσμευσης των γεγονότων. Δηλαδή, μας δείχνει πώς η πιθανότητα του ενδεχομένου $B \mid A$ συνδέεται με την πιθανότητα του $A \mid B$.

Μία μικρή επέκταση του θεωρήματος του Bayes προκύπτει θεωρώντας τα ενδεχόμενα

$C_1, C_2 \ldots, C_k$ που διαμερίζουν τον δειγματικό χώρο Ω, ώστε $C_i \cap C_j = \varnothing$ για $i \neq j$ και $C_1 \cup C_2 \cup \cdots \cup C_k = \Omega$. Τότε,

$$P(C_i \mid A) = \frac{P(C_i) \cdot P(A \mid C_i)}{\sum\limits_{j=1}^{k} P(C_j) \cdot P(A \mid C_j)}, \quad i = 1, 2, \ldots, k.$$

Απόδειξη.

$$P(C_i \mid A) = \frac{P(A \cap C_i)}{P(A)} = \frac{P(C_i) \cdot P(A \mid C_i)}{P(A)}, \ i = 1, 2, \ldots, k, \ \text{όπου:}$$

$$P(A) = \sum_{i=1}^{k} P(A \cap C_i) = \sum_{i=1}^{k} P(C_i) \cdot P(A \mid C_i),$$

αφού τα C_i είναι ξένα και η ένωσή τους είναι ο Ω. □

Μία περαιτέρω επέκταση είναι για συνεχείς τυχαίες μεταβλητές:

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

Σημείωση 1. *Όταν κανονικοποιούμε μία posterior κατανομή, προσέχουμε τα όρια στις παραμέτρους.*

Παράδειγμα 1.1. *Μία διαδικασία ανίχνευσης του ιού HIV εφαρμόζεται σε έναν πληθυσμό που έχει υψηλό ρίσκο προσβολής από τον HIV - 10% του πληθυσμού πιστεύεται ότι είναι φορείς του HIV. Η διαδικασία ανίχνευσης δίνει θετικό αποτέλεσμα για το 90% των ατόμων που είναι πραγματικά φορείς του HIV και αρνητικό αποτέλεσμα για το 85% των ατόμων που δεν είναι φορείς του HIV. Ποιες είναι οι πιθανότητες για ψευδώς θετικά και ψευδώς αρνητικά αποτελέσματα;*

Συμβολίζοντας με A το ενδεχόμενο το άτομο να είναι φορέας του HIV και B το ενδεχόμενο το αποτέλεσμα του τεστ να είναι θετικό, έχουμε ότι $P(A) = 0.1$, $P(B \mid A) = 0.9$ και $P(B^c \mid A^c) = 0.85$.

Πρέπει να βρούμε τις $P(\text{ψευδώς θετικό}) = P(A^c \mid B)$ και $P(\text{ψευδώς αρνητικό}) = P(A \mid B^c)$. Έχουμε:

$P(A) = 0.1$ και $P(A^c) = 1 - P(A) = 0.9$,

$P(B \mid A) = 0.9$ και $P(B^c \mid A) = 1 - P(B \mid A) = 0.1$,

$P(B^c \mid A^c) = 0.85$ και $P(B \mid A^c) = 1 - P(B^c \mid A^c) = 0.15$,

$P(B) = P(A)P(B \mid A) + P(A^c)P(B \mid A^c) = 0.1 \cdot 0.9 + 0.9 \cdot 0.15 = 0.09 + 0.135 = 0.225$ και

$P(B^c) = 1 - P(B) = 0.775$. Επομένως,

$$P(A^c \mid B) = \frac{P(A^c) \cdot P(B \mid A^c)}{P(B)} = \frac{0.9 \cdot 0.15}{0.225} = 0.6 \quad \text{και}$$

$$P(A \mid B^c) = \frac{P(A) \cdot P(B^c \mid A)}{P(B^c)} = \frac{0.1 \cdot 0.1}{0.775} = 0.0129.$$

Παράδειγμα 1.2. *Σε μία τσάντα υπάρχουν 6 μπάλες αγνώστων χρωμάτων. Εξάγουμε 3 μπάλες χωρίς επανάθεση και βλέπουμε ότι είναι μαύρες. Να βρεθεί η πιθανότητα να μην έχει απομείνει μαύρη μπάλα στην τσάντα.*

Έστω A το ενδεχόμενο να εξάγουμε 3 μαύρες μπάλες και C_i το ενδεχόμενο να υπήρχαν i το πλήθος μαύρες μπάλες μέσα στην τσάντα. Τότε, από το θεώρημα του Bayes παίρνουμε:

$$P(C_i \mid A) = \frac{P(C_i) \cdot P(A \mid C_i)}{\sum_{j=0}^{6} P(C_j) \cdot P(A \mid C_j)}, \quad i = 0, 1, \ldots, 6.$$

Όμως, εδώ βρίσκεται το βασικό θέμα: ποιες τιμές να δώσουμε στις $P(C_0), P(C_1), \ldots, P(C_6)$; Πρόκειται για τις πιθανότητες να έχουμε διαφορετικά πλήθη μαύρων μπαλών μέσα στην τσάντα, *προτού* να δούμε τα δεδομένα. Χωρίς κάποια πληροφορία για το αντίθετο, μπορούμε εύλογα να υποθέσουμε ότι όλα τα δυνατά πλήθη είναι εξίσου πιθανά, οπότε παίρνουμε $P(C_0) = P(C_1) = \cdots = P(C_6) = \frac{1}{7}$. Για την ακρίβεια, θα χρησιμοποιήσουμε αυτήν την επιλογή της prior κατανομής για το υπόλοιπο του προβλήματος. Όμως, είναι αυτό το πιο λογικό; Θα μπορούσαμε να υιοθετήσουμε την άποψη ότι είναι αρκετά πιθανό όλες οι μπάλες μέσα στην τσάντα να είναι του ίδιου χρώματος και, κατά συνέπεια, να δώσουμε μεγαλύτερες prior πιθανότητες στα ενδεχόμενα C_0 και C_6. Ή μπορεί να μαθαίναμε τελικά από τους κατασκευαστές των μπαλών ότι παράγουν μπάλες 10 διαφορετικών χρωμάτων. Τότε, θα μπορούσαμε να υιοθετήσουμε την prior πεποίθηση ότι κάθε μπάλα είναι μαύρη με πιθανότητα $\frac{1}{10}$ και να χρησιμοποιήσουμε αυτό ως βάση για τον υπολογισμό των prior πιθανοτήτων μέσω της διωνυμικής κατανομής, όπως στο παράδειγμα της δειγματοληψίας νομισμάτων όπου είχαμε υποθέσει ότι κάθε νόμισμα είναι δίκαιο, οπότε φέρνει ”Γ” με πιθανότητα $\frac{1}{2}$. Το νόημα είναι ότι πρέπει να *σκεφτούμε καλά* πώς να εκφράσουμε τις prior πεποιθήσεις μας και η απάντηση που θα πάρουμε για το πρόβλημα θα εξαρτάται από το τι πιστεύαμε εξαρχής.

Θα επανέλθουμε σε αυτό το επιχείρημα αργότερα. Για την ώρα πάντως, χρησιμοποιώντας την prior που προσδιορίσαμε παραπάνω, εφαρμόζουμε απλώς το θεώρημα του Bayes και παίρνουμε ότι:

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

$$= \frac{\frac{1}{7}\cdot\left(\frac{3}{6}\cdot\frac{2}{5}\cdot\frac{1}{4}\right)}{\frac{1}{7}\cdot\left[0+0+0+\left(\frac{3}{6}\cdot\frac{2}{5}\cdot\frac{1}{4}\right)+\left(\frac{4}{6}\cdot\frac{3}{5}\cdot\frac{2}{4}\right)+\left(\frac{5}{6}\cdot\frac{4}{5}\cdot\frac{3}{4}\right)+1\right]}$$
$$= \frac{1}{35}$$

Έτσι, τα δεδομένα ανανέωσαν την prior πεποίθηση ότι $P(C_3) = \frac{1}{7}$ και έδωσαν την posterior πιθανότητα $P(C_3 \mid A) = \frac{1}{35}$ στη θέση της. Δηλαδή, το ενδεχόμενο είναι πολύ λιγότερο πιθανό από ότι ήταν προτού παρατηρήσουμε τα δεδομένα.

1.7 Ασκήσεις

Άσκηση 1.1. *Τα βραχώδη στρώματα A και B είναι δύσκολο να διακριθούν μεταξύ τους στο πεδίο. Μέσα από προσεκτικές εργαστηριακές μελέτες έχει προσδιοριστεί ότι το μόνο χαρακτηριστικό που ίσως είναι χρήσιμο για να βοηθήσει στη διάκρισή τους είναι η παρουσία ή απουσία ενός συγκεκριμένου απολιθώματος βραχιόποδα. Σε βραχώδη εξάρματα του μεγέθους που συνήθως συναντώνται, οι πιθανότητες παρουσίας του απολιθώματος βρίσκονται στον πίνακα 1.2. Είναι επίσης γνωστό ότι το βραχώδες στρώμα A συναντάται 4 φορές πιο συχνά από το στρώμα B σε αυτό το πεδίο μελέτης. Αν ληφθεί ένα δείγμα στο οποίο υπάρχει απολίθωμα, υπολογίστε τις posterior πιθανότητες των 2 τύπων στρωμάτων.*

Στρώμα	Απολίθωμα	Όχι απολίθωμα
A	0.9	0.1
B	0.2	0.8

Πινακας 1.2: Πιθανότητες τύπων στρωμάτων υπό την παρουσία ή απουσία απολιθώματος

Αν η γεωλόγος κατηγοριοποιεί πάντα το στρώμα ως A όταν υπάρχει το απολίθωμα και ως B όταν δεν υπάρχει, ποια είναι η πιθανότητα να είναι σωστή σε μία μελλοντική κατηγοριοποίηση;

Άσκηση 1.2. *Επαναλάβετε το παράδειγμα 1.2, χρησιμοποιώντας μία διαφορετική επιλογή prior κατανομής. Μπορείτε να χρησιμοποιήσετε οποιαδήποτε prior κατανομή, αλλά πρέπει να δώσετε κάποιον πειστικό (αλλά σύντομο) λόγο για τη χρήση της. Με ποια έννοια επηρεάζει αυτή η αλλαγή της prior την posterior πιθανότητα να μην έχει απομείνει καμία μαύρη μπάλα μέσα στην τσάντα;*

Άσκηση 1.3. *Επαναλάβετε το παράδειγμα της δειγματοληψίας νομισμάτων, γράφοντας τον πίνακα για την περίπτωση όπου τα 2 νομίσματα του δείγματος παρατηρούνται να είναι "Γ", δηλαδή $X = 2$. Μόνο οι γραμμές 2, 4 και 5 θα αλλάξουν. Εξηγήστε σύντομα πώς η εικόνα που δίνεται από την posterior κατανομή διαφέρει από αυτήν της πιθανοφάνειας, σε αυτή την περίπτωση.*

Άσκηση 1.4. *Μία συλλέκτρια σπόρων η οποία έχει αποκτήσει έναν μικρό πλήθος σπόρων από ένα φυτό, έχει μία prior πεποίθηση ότι η πιθανότητα θ βλάστησης κάθε σπόρου* *είναι ομοιόμορφη στο εύρος* $[0, 1]$. *Πειραματίζεται σπέρνοντας δύο σπόρους και βλέπει ότι βλαστίζουν.*

i. *Γράψτε τη συνάρτηση πιθανοφάνειας του* θ *η οποία απορρέει από αυτή την παρατήρηση και βρείτε την posterior κατανομή του* θ *σύμφωνα με τη συλλέκτρια.*

ii. *Υπολογίστε την posterior πιθανότητα το* θ *να είναι μικρότερο από* $\frac{1}{2}$.

Άσκηση 1.5. *Μία posterior κατανομή υπολογίζεται ότι είναι ανάλογη στη συνάρτηση* θ^{-3}, *δηλαδή:*

$$f(\theta \mid x) \propto \theta^{-3}, \quad \theta > 1.$$

Με άλλα λόγια, $f(\theta \mid x) = c\theta^{-3}$ *για* $\theta > 1$, *όπου* c *άγνωστη σταθερά κανονικοποίησης. Υπολογίστε τη σταθερά κανονικοποίησης αυτής της posterior και την posterior πιθανότητα του* $\theta < 2$.

Άσκηση 1.6. *Ένας μελισσοκόμος έχει* 4 *κυψέλες και, αργά τον χειμώνα, η prior κατανομή του πλήθους* θ *των κυψελών στις οποίες υπάρχουν ακόμα ζωντανές μέλισσες, δίνεται στον πίνακα:*

θ	0	1	2	3	4
$f(\theta)$	0.3	0.1	0.1	0.2	0.3

Διαλέγει 2 *κυψέλες στην τύχη και μετά από προσεκτική επιθεώρηση βρίσκει ότι οι μέλισσες είναι ζωντανές και στις* 2.

i. *Γράψτε την πιθανοφάνεια του* θ *δεδομένης αυτής της παρατήρησης.*

ii. *Υπολογίστε την posterior πιθανότητα οι μέλισσες να είναι ζωντανές και στις* 4 *κυψέλες.*

Κεφάλαιο 2

Μπεϋζιανή Ανανέωση Πληροφορίας

2.1 Εισαγωγή

Όπως τονίστηκε στο κεφάλαιο 1, η ουσία της Μπεϋζιανής προσέγγισης είναι να χειριζόμαστε την άγνωστη παράμετρο θ ως τυχαία μεταβλητή, να προσδιορίζουμε μία prior κατανομή για το θ, η οποία να αντιπροσωπεύει τις πεποιθήσεις μας για το θ προτού δούμε τα δεδομένα, να χρησιμοποιούμε το θεώρημα του Bayes για την ανανέωση των prior πεποιθήσεών μας, οι οποίες αντικαθίστανται από τις posterior πιθανότητες, και να εξάγουμε κατάλληλα συμπεράσματα. Έτσι, υπάρχουν 4 βασικά βήματα στην Μπεϋζιανή προσέγγιση:

1. Καθορισμός ενός μοντέλου πιθανοφάνειας $f(x \mid \theta)$,
2. Καθορισμός μίας prior κατανομής $f(\theta)$,
3. Υπολογισμός της posterior κατανομής $f(\theta \mid x)$ από το θεώρημα του Bayes,
4. Εξαγωγή συμπερασμάτων από την posterior πληροφορία.

Σε αυτό το κεφάλαιο, θα επαναδιατυπώσουμε το θεώρημα του Bayes σε μία μορφή κατάλληλη για τυχαίες μεταβλητές αντί για ενδεχόμενα και θα αναλογιστούμε ορισμένα θέματα που προκύπτουν όταν προσπαθούμε να χρησιμοποιήσουμε αυτό το αποτέλεσμα στο πλαίσιο της συμπερασματολογίας για μία παράμετρο θ. Αυτά τα θέματα θα συζητηθούν σε επακόλουθα κεφάλαια. Επιπλέον, θα δούμε εδώ μία σειρά παραδειγμάτων όπου συγκεκριμένοι συνδυασμοί prior κατανομών και πιθανοφανειών δίνουν υπολογιστικά βολικές εκφράσεις για την posterior κατανομή.

2.2 Το θεώρημα του Bayes

Διατυπωμένο σε επίπεδο τυχαίων μεταβλητών με συναρτήσεις (πυκνότητας) πιθανότητας που συμβολίζονται γενικά με f, το θεώρημα του Bayes παίρνει τη μορφή:

$$f(\theta \mid x) = \frac{f(\theta) \cdot f(x \mid \theta)}{\int_\Theta f(\theta) \cdot f(x \mid \theta) d\theta}, \quad \theta \in \Theta.$$

Θα χρησιμοποιούμε αυτόν τον συμβολισμό για να καλύψουμε την περίπτωση όπου το X είναι συνεχές με σ.π.π. (συνάρτηση πυκνότητας πιθανότητας) f και την περίπτωση όπου είναι διακριτό με συνάρτηση μάζας πιθανότητας f. Παρομοίως, το θ μπορεί να είναι διακριτό ή συνεχές, αλλά στη διακριτή περίπτωση το $\int_\Theta f(\theta) \cdot f(x \mid \theta) d\theta$ θα μεταφράζεται ως $\sum\limits_{j \in \Theta} f(\theta_j) \cdot f(x \mid \theta_j)$.

Παρατηρούμε ότι ο παρονομαστής στο θεώρημα του Bayes είναι συνάρτηση μόνο του x - το θ "έχει φύγει με το ολοκλήρωμα". Έτσι, ένας άλλος τρόπος να γράψουμε το Θεώρημα του Bayes είναι:

$$\begin{aligned} f(\theta \mid x) &= c \cdot f(\theta) \cdot f(x \mid \theta) \\ &\propto f(\theta) \cdot f(x \mid \theta) \\ &= h(\theta). \end{aligned}$$

Με άλλα λόγια, "η posterior είναι ανάλογη με το γινόμενο της prior επί την πιθανοφάνεια". Η σταθερά αναλογίας c, η οποία μπορεί να εξαρτάται από το x αλλά όχι από το θ, είναι μία σταθερά κανονικοποίησης, ορισμένη έτσι ώστε η posterior κατανομή να ολοκληρώνει στη μονάδα.

Υπάρχει μία μοναδική σ.π.π., έστω $g(\theta)$, που είναι ανάλογη σε κάποια δοσμένη συνάρτηση $h(\theta)$ για $\theta \in \Theta$, αφού η $g(\theta)$ μπορεί να καθοριστεί μονοσήμαντα ως $g(\theta) = c \cdot h(\theta)$, όπου:

$$c = \frac{1}{\int_\Theta h(\theta) d\theta}.$$

Θα χρησιμοποιήσουμε αυτό το γεγονός σύντομα για να αναγνωρίσουμε γνωστές σ.π.π. που είναι ανάλογες σε συγκεκριμένες posterior κατανομές.

Το γεγονός αυτό μας επιτρέπει επίσης να απομακρύνουμε όποιον παράγοντα της συνάρτησης $h(\theta) = f(\theta) \cdot f(x \mid \theta)$ ΔΕΝ εξαρτάται από το θ, πριν πραγματοποιήσουμε την κανονικοποίηση. Αυτό θα επεξηγηθεί μέσω παραδειγμάτων.

Σημείωση 2. *Η posterior κατανομή ΔΕΝ είναι η $h(\theta)$ - αυτό είναι ιδιαίτερα σημαντικό στην πραγματοποίηση πιθανοτικών διατυπώσεων για το θ (βλέπε κεφάλαιο 5) και στον υπολογισμό κατανομών πρόβλεψης (βλέπε κεφάλαιο 6).*

2.3 Ζητήματα

2.3.1 Επιλογή μοντέλου πιθανοφάνειας

Αυτή η επιλογή εξαρτάται από τον μηχανισμό του προβλήματος που έχουμε στα χέρια μας και είναι το ίδιο ζήτημα που αντιμετωπίζεται όταν κάνουμε χρήση κλασικής συμπερασματολογίας - ποιο είναι το καταλληλότερο μοντέλο για τα δεδομένα μας; Συχνά, η γνώστη της δομής μέσω της οποίας έχουμε συλλέξει τα δεδομένα μπορεί να υποδεικνύει κατάλληλα μοντέλα. Για παράδειγμα, η διωνυμική κατανομή χρησιμοποιείται για δεδομένα από ανεξάρτητα επαναλαμβανόμενα πειράματα με 2 πιθανές εκβάσεις και η κατανομή Poisson για δεδομένα από καταμετρήσεις γεγονότων. Συχνά, όμως, ένα μοντέλο είναι τελείως υποθετικό και η αληθοφάνειά του αξιολογείται αργότερα μέσω των δεδομένων. Για παράδειγμα, στο γραμμικό μοντέλο, υποθέτουμε ότι η Y είναι γραμμικά εξαρτημένη από τη X με ανεξάρτητα τυχαία σφάλματα που ακολουθούν την κανονική κατανομή.

2.3.2 Επιλογή prior

Αυτό το ζήτημα είναι θεμελιώδες στο Μπεϋζιανό πλαίσιο και θα συζητηθεί εκτενώς στο κεφάλαιο 3. Ωστόσο, κάποια σημεία πρέπει να σημειωθούν τώρα:

1. Επειδή η prior αντιπροσωπεύει τις πεποιθήσεις μας για το θ προτού παρατηρήσουμε τα δεδομένα, έπεται ότι η επακόλουθη ανάλυση είναι μοναδική για εμάς. Οι prior πεποιθήσεις κάποιου άλλου θα οδηγούσαν σε διαφορετική posterior ανάλυση. Υπό αυτή την έννοια, η ανάλυση είναι υποκειμενική.

2. Θα δούμε αργότερα ότι, με την προϋπόθεση ότι η prior που θα επιλέξουμε δεν είναι "εντελώς παράλογη", τότε η επίδρασή της θα γίνεται όλο και λιγότερο σημαντική όσο γίνονται διαθέσιμα περισσότερα δεδομένα. Έτσι, κατά μία έννοια, ο λανθασμένος προσδιορισμός της prior είναι ασήμαντος, αρκεί να έχουμε αρκετά διαθέσιμα δεδομένα.

3. Συχνά, μπορεί να έχουμε μία "πρόχειρη ιδέα" για το πώς θα πρέπει να είναι η prior (θα μπορούσαμε ίσως να δώσουμε τη μέση τιμή και τη διασπορά της), χωρίς να μπορούμε να γίνουμε πιο ακριβείς. Σε τέτοιες περιπτώσεις, θα μπορούσαμε να χρησιμοποιήσουμε μία "βολική" μορφή για την prior που να είναι σύμφωνη με τις πεποιθήσεις μας, αλλά και να κάνει τους μαθηματικούς υπολογισμούς σχετικά απλούς. Θα δούμε σύντομα μερικά παραδείγματα αυτού του τύπου ανάλυσης.

4. Κάποιες φορές, μπορεί να έχουμε την αίσθηση ότι δεν υπάρχει κάποια prior πληροφορία για μία παράμετρο. Σε τέτοιες περιπτώσεις, θα θέλαμε να χρησιμοποιήσουμε μία prior που να αντανακλά την άγνοιά μας για την παράμετρο. Αυτό είναι συχνά εφικτό, αλλά ενέχει ορισμένες δυσκολίες. Αυτά θα συζητηθούν στο κεφάλαιο 3.

2.3.3 Υπολογισμός

Αν και θεωρητικά απλή, η εφαρμογή του θεωρήματος του Bayes μπορεί στην πράξη να είναι υπολογιστικά δύσκολη, κυρίως ως αποτέλεσμα του ολοκληρώματος, δηλαδή της σταθεράς κανονικοποίησης, στον παρονομαστή. Θα δούμε ότι για ορισμένες επιλογές συνδυασμών prior και πιθανοφάνειας, ο υπολογισμός αυτού του ολοκληρώματος μπορεί να αποφευχθεί, αλλά, γενικά, απαιτούνται εξειδικευμένες τεχνικές για να απλοποιηθεί αυτός ο υπολογισμός.

2.3.4 Συμπερασματολογία

Η Μπεϋζιανή ανάλυση δίνει μία πιο πλήρη συμπερασματολογία, υπό την έννοια ότι όλη η γνώση για το θ, η οποία είναι διαθέσιμη από την prior και τα δεδομένα, απεικονίζεται στην posterior κατανομή. Με άλλα λόγια, η $f(\theta \mid x)$ *είναι* όλη η συμπερασματολογία. Παρόλα αυτά, είναι συχνά επιθυμητό να συνοψίσουμε αυτή τη συμπερασματολογία υπό τη μορφή μίας σημειακής εκτίμησης ή ενός διαστήματος εκτίμησης. Αυτά θα συζητηθούν στο κεφάλαιο 5. Επιπλέον, κάποιες από τις επιθυμητές ιδιότητες των στατιστικών συναρτήσεων, δηλαδή των συναρτήσεων των δεδομένων που χρησιμοποιούνται για τους σκοπούς της συμπερασματολογίας, υπάρχουν και στη Μπεϋζιανή ανάλυση. Για παράδειγμα, στην κλασική συμπερασματολογία, ο ρόλος των επαρκών στατιστικών συναρτήσεων συζητείται εκτενέστατα. Αυτή η έννοια έχει ανάλογο ρόλο στην Μπεϋζιανή συμπερασματολογία, αλλά είναι πιο διαισθητικά ελκυστική. Για παράδειγμα, στην Μπεϋζιανή στατιστική, η επάρκεια μπορεί να χαρακτηριστεί λέγοντας ότι αν πάρουμε μία συνάρτηση $h(x)$ των δεδομένων μας, τότε η $h(x)$ είναι επαρκής για το θ αν η posterior $f(\theta \mid x)$ εξαρτάται από τα δεδομένα μόνο μέσω της $h(x)$ και όχι από τις επιμέρους τιμές των x_i.

2.4 Παραδείγματα

Παράδειγμα 2.1. *(Διωνυμικό δείγμα). Υποθέτουμε ότι έχουμε το μοντέλο πιθανοφάνειας $X \sim Bin(n, \theta)$ και θέλουμε να βγάλουμε συμπεράσματα για το θ από 1 παρατήρηση x.*

Αφού έχουμε μόνο 1 παρατήρηση, η πιθανοφάνεια του μοντέλου είναι η εξής:

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

Προφανώς, ο προσδιορισμός της prior του θ θα διαφέρει γενικά από πρόβλημα σε πρόβλημα και, εξ ορισμού, θα εξαρτάται από το εύρος της prior γνώσης μας για την κατάσταση. Ωστόσο, εδώ θα προχωρήσουμε θεωρώντας μία πιθανή οικογένεια prior κατανομών η οποία, όπως θα δούμε, οδηγεί σε απλούς υπολογισμούς. Το νόημα αυτού είναι πως, με την προϋπόθεση ότι η οικογένεια κατανομών είναι αρκετά μεγάλη και καλύπτει ένα σημαντικό εύρος διαφορετικών μορφών, μπορούμε να χρησιμοποιήσουμε μία prior από αυτή την οικογένεια η οποία να συμφωνεί σε μεγάλο βαθμό με τις prior πεποιθήσεις μας. Αν αυτό συμβαίνει, τότε παίρνουμε και απλές απαντήσεις. Αν, όμως, δεν υπάρχει prior μέσα σε αυτήν την οικογένεια η οποία να συμφωνεί με αυτό που πραγματικά πιστεύουμε, τότε θα πρέπει να αποφύγουμε αυτήν την προσέγγιση.

Έτσι, σε αυτή την περίπτωση, ας υποθέσουμε ότι μπορούμε να αναπαραστήσουμε τις prior πεποιθήσεις μας για το θ μέσω της κατανομής Βήτα, δηλαδή $\theta \sim \text{Beta}(p, q)$. Οπότε:

$$
\begin{aligned}
f(\theta) &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot \theta^{p-1}(1-\theta)^{q-1} \\
&\propto \theta^{p-1}(1-\theta)^{q-1}, \quad \theta \in (0,1).
\end{aligned}
$$

Οι παράμετροι αυτής της κατανομής είναι τα $p > 0$ και $q > 0$. (ΔΕΝ είναι πιθανότητες και μπορούν να πάρουν οποιαδήποτε θετική τιμή.) Η μέση τιμή και η διασπορά αυτής της κατανομής είναι:

$$
\mu = E(\theta) = \frac{p}{p+q} \quad \text{και} \quad \sigma^2 = \text{Var}(\theta) = \frac{pq}{(p+q)^2(p+q+1)}. \tag{2.1}
$$

Η κατανομή Βήτα γράφεται επίσης ως:

$$
f(\theta) = \frac{\theta^{p-1}(1-\theta)^{q-1}}{B(p,q)}, \quad \text{όπου} \quad B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 \theta^{p-1}(1-\theta)^{q-1} d\theta.
$$

Ονομάζουμε το $B(p, q)$ συνάρτηση Βήτα.

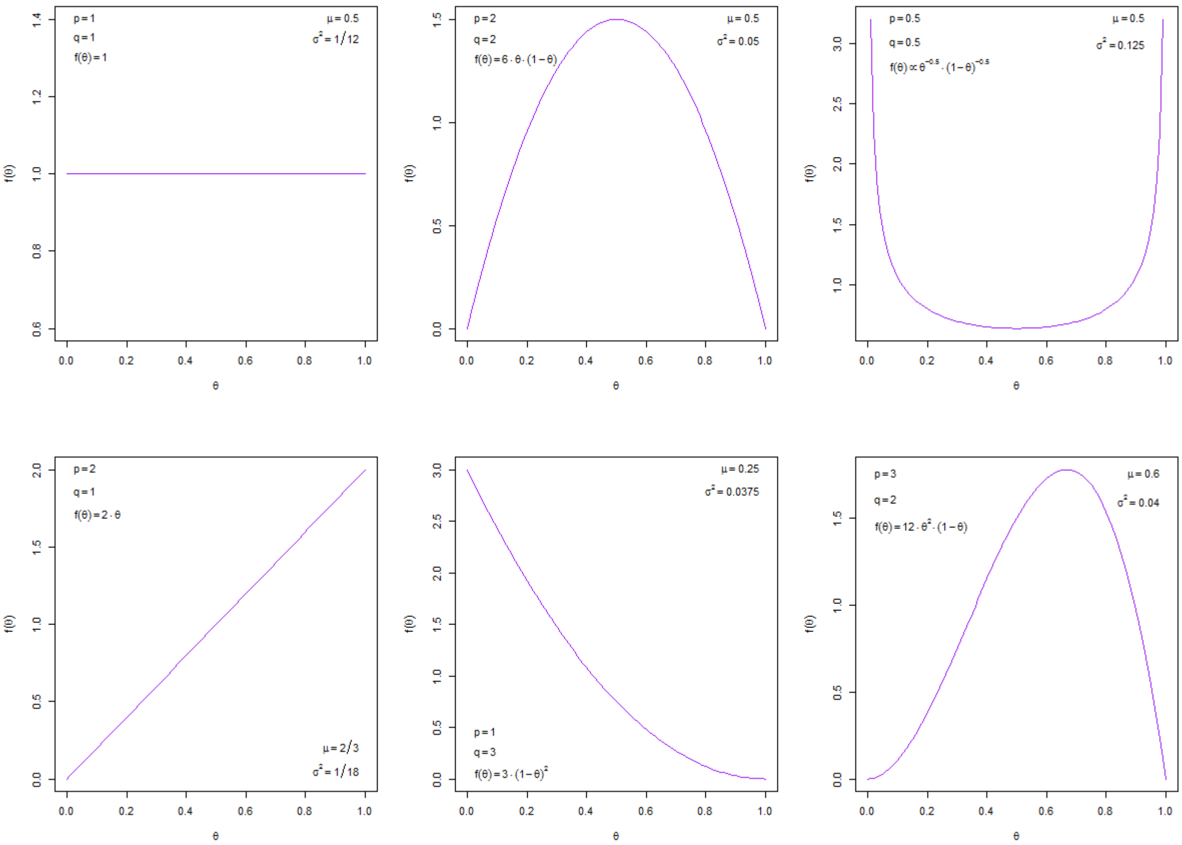
Στο σχήμα 2.1, φαίνονται ορισμένες απλές περιπτώσεις της κατανομής βήτα.

Τώρα, εφαρμόζουμε το θεώρημα του Bayes χρησιμοποιώντας αυτήν την prior κατανομή:

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

όπου $P = p + x$ και $Q = q + n - x$. Υπάρχει μόνο μία συνάρτηση πυκνότητας πιθανότητας ανάλογη με τη συνάρτηση $\theta^{P-1} \cdot (1-\theta)^{Q-1}$, οπότε θα πρέπει να ισχύει ότι $\theta \mid x \sim \text{Beta}(P, Q)$. Έτσι, μετά από προσεκτική επιλογή, έχουμε αποκτήσει μία posterior κατανομή η οποία ανήκει στην ίδια οικογένεια κατανομών με την prior και, με αυτόν τον τρόπο, αποφύγαμε τον αναλυτικό υπολογισμό του ολοκληρώματος για τη σταθερά κανονικοποίησης. Αυτό έχει ήδη γίνει στον γενικό τύπο της κατανομής Βήτα.

Η επίδραση των δεδομένων συνοψίζεται στην τροποποίηση των παραμέτρων της κατανομής βήτα από τις prior τιμές (p, q) στις posterior τιμές $(p + x, q + n - x)$.



Σχημα 2.1: Περιπτώσεις της κατανομής Βήτα

Ως αριθμητικό παράδειγμα, θεωρούμε το σύνολο δεδομένων "CANCER" από το First Bayes[1]. Από 70 ασθενείς στους οποίους δόθηκε μία νέα θεραπεία για μία συγκεκριμένη μορφή καρκίνου, 34 επιβίωσαν μετά το πέρας μίας συγκεκριμένης περιόδου. Συμβολίζουμε με θ την πιθανότητα επιβίωσης ενός ασθενούς. Έπειτα από συζήτηση με ειδικούς στον χώρο της υγείας, οι οποίοι είναι εξοικειωμένοι με παρόμοιες κλινικές δοκιμές, οδηγούμαστε στην έκφραση της prior πεποίθησης ότι $E(\theta) = 0.4$ και $\mathrm{Var}(\theta) = 0.02$. Τώρα, αν μία βήτα κατανομή είναι εύλογη για τις prior πεποιθήσεις των ειδικών, τότε θα πρέπει να επιλέξουμε $\theta \sim \mathrm{Beta}(p, q)$, ώστε $E(\theta) = 0.4$ και $\mathrm{Var}(\theta) = 0.02$. Έτσι, απαιτούμε:

$$\mu = \frac{p}{p+q} = 0.4 \quad \text{και} \quad \sigma^2 = \frac{pq}{(p+q)^2(p+q+1)} = 0.02.$$

Το σύστημα αυτών των εξισώσεων οδηγεί στην εξής λύση:

$$p = \frac{(1-\mu)\mu^2}{\sigma^2} - \mu \quad \text{και} \quad q = \frac{(1-\mu)^2\mu}{\sigma^2} - (1-\mu),$$

από όπου παίρνουμε $p = 4.4$ και $q = 6.6$, τα οποία προσδιορίζουν την prior κατανομή για το θ. Αυτό ήταν ένα απλό παράδειγμα αυτού που ονομάζουμε *εξαγωγή* της prior κατανομής.

[1]Το First Bayes είναι ένα υπολογιστικό πακέτο, γραμμένο από τον Tony O'Hagan, το οποίο περιέχει διάφορες Μπεϋζιανές αναλύσεις.

Στην πράξη, θα ήταν απαραίτητο να βεβαιωθούμε ότι ολόκληρη η prior κατανομή είναι συμβατή με τις prior πεποιθήσεις των ειδικών.

Υποθέτοντας ότι είναι συμβατή, παίρνουμε $P = 4.4 + 34 = 38.4$ και $Q = 6.6 + 70 - 34 = 42.6$, οπότε $\theta \mid x \sim \text{Beta}(38.4, 42.6)$. Αυτή η posterior κατανομή συνοψίζει όλη τη διαθέσιμη πληροφορία για το θ και αντιπροσωπεύει όλη τη συμπερασματολογία για το θ. Θα συζητήσουμε αργότερα πώς, αν χρειαστεί, θα μπορούσαμε να συνοψίσουμε αυτή τη συμπερασματολογία, αλλά προς το παρόν μπορούμε να δούμε πώς τα δεδομένα τροποποιούν τις prior πεποιθήσεις μας, συγκρίνοντας την prior και την posterior μέση τιμή του θ:

$$E(\theta) = \frac{p}{p+q} = 0.4 \quad \text{και} \quad E(\theta \mid x) = \frac{P}{P+Q} = \frac{p+x}{p+q+n} = 0.474.$$

Τα παρατηρούμενα δεδομένα οδήγησαν στην αύξηση της prior εκτίμησης του θ από 0.4 σε 0.474. Από την άλλη, μία λογική εκτίμηση του θ, βασισμένη αποκλειστικά στα δεδομένα, είναι η $\frac{x}{n} = 0.486$, η οποία είναι η εκτίμηση μέγιστης πιθανοφάνειας. Δηλαδή, η posterior εκτίμηση δίνει μία ισορροπία ανάμεσα στις prior πεποιθήσεις μας και στην πληροφορία που παρέχεται από τα δεδομένα.

Γενικότερα, αν τα x και n είναι αρκετά μεγάλα συγκριτικά με τα p και q, τότε η posterior μέση τιμή είναι περίπου $\frac{x}{n}$, δηλαδή ίση με την εκτίμηση μέγιστης πιθανοφάνειας. Από την άλλη, αν τα p και q είναι σχετικά μεγάλα, τότε θα έχουν σημαντική επιρροή στην posterior μέση τιμή. Μπορεί ακόμα να ελεγχθεί ότι όσο τα x και n γίνονται μεγαλύτερα - ή, ακόμα, αν τα p και q επιλεγούν μεγαλύτερα - τότε η posterior διασπορά θα είναι μικρότερη.

Παράδειγμα 2.2. *(Δείγμα Poisson). Έστω ότι έχουμε ένα τυχαίο δείγμα $x = (x_1, x_2, \ldots, x_n)$ (δηλαδή ανεξάρτητες παρατηρήσεις) μεγέθους n από μία τυχαία μεταβλητή X της οποίας η κατανομή είναι Poisson(θ), δηλαδή:*

$$f(x_i \mid \theta) = e^{-\theta} \cdot \frac{\theta^{x_i}}{x_i!}, \quad \theta > 0, \quad i = 1, 2, \ldots, n.$$

Υπολογίζουμε τη συνάρτηση πιθανοφάνειας του θ:

$$\begin{aligned} f(x \mid \theta) &= \prod_{i=1}^{n} \left(e^{-\theta} \cdot \frac{\theta^{x_i}}{x_i!} \right) \\ &\propto e^{-n\theta} \theta^{n\overline{x}}, \quad \text{όπου} \quad \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \end{aligned}$$

Όπως και στο παράδειγμα του διωνυμικού δείγματος, οι prior πεποιθήσεις για το θ θα διαφέρουν από πρόβλημα σε πρόβλημα. Θα αναζητήσουμε μία μορφή που δίνει ένα εύρος διαφορετικών δυνατοτήτων, αλλά είναι και μαθηματικά εύκολη στον χειρισμό.

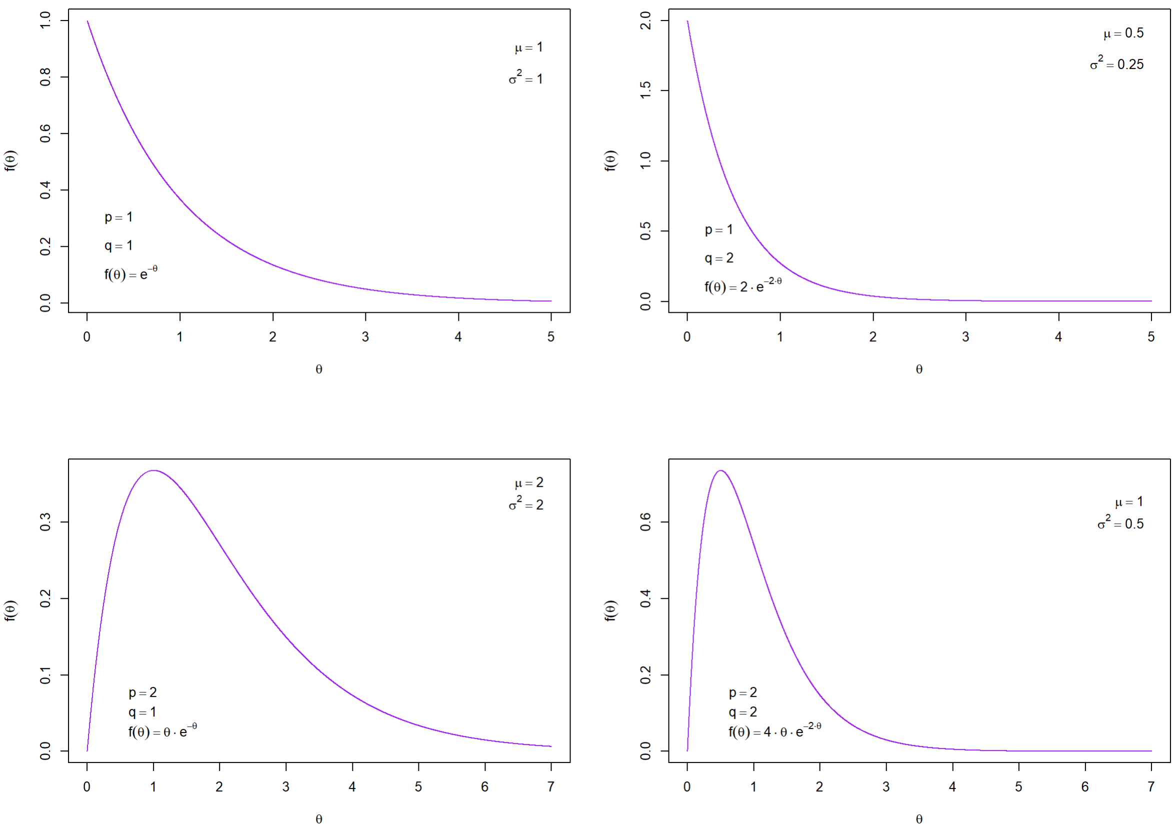
Σε αυτήν την περίπτωση, υποθέτουμε ότι οι prior πεποιθήσεις μας μπορούν να αναπαρασταθούν μέσω της κατανομής Γάμμα, δηλαδή $\theta \sim \text{Gamma}(p, q)$, οπότε:

$$f(\theta) = \frac{q^p}{\Gamma(p)} \cdot \theta^{p-1} e^{-q\theta}, \quad \theta > 0.$$

Η παράμετρος $p > 0$ καλείται παράμετρος σχήματος και η $q > 0$ καλείται παράμετρος κλίμακας. Η μέση τιμή και η διασπορά της κατανομής είναι:

$$\mu = E(\theta) = \frac{p}{q} \quad \text{και} \quad \sigma^2 = \text{Var}(\theta) = \frac{p}{q^2}. \tag{2.2}$$

Στο σχήμα 2.2, φαίνονται ορισμένα παραδείγματα της κατανομής Γάμμα.



Σχημα 2.2: Παραδείγματα της κατανομής Γάμμα

Εφαρμόζοντας το θεώρημα του Bayes με αυτή την prior κατανομή, παίρνουμε ότι:

$$\begin{aligned} f(\theta \mid x) &\propto \theta^{p-1} e^{-q\theta} \cdot e^{-n\theta} \theta^{n\overline{x}} \\ &= \theta^{p+n\overline{x}-1} e^{-(q+n)\theta} \\ &= \theta^{P-1} e^{-Q\theta}, \end{aligned}$$

όπου $P = p+n\overline{x}$ και $Q = q+n$. Υπάρχει πάλι μοναδική συνάρτηση πυκνότητας πιθανότητας ανάλογη με τη συνάρτηση $\theta^{P-1}e^{-Q\theta}$, οπότε θα πρέπει $\theta \mid x \sim \text{Gamma}(P, Q)$. Δηλαδή, παίρνουμε μία άλλη κατανομή Γάμμα, της οποίας οι παράμετροι έχουν τροποποιηθεί μέσω του δειγματικού μέσου των παρατηρήσεων και του μεγέθους του δείγματος[2].

Σημείωση 3. *Όταν αναγνωρίζουμε την posterior κατανομή, θυμόμαστε ότι είναι κατανομή για την παράμετρο θ, όχι για τα δεδομένα x_i. Για παράδειγμα, είναι συνηθισμένο λάθος να αναγνωρίζουμε την $f(\theta \mid x) \propto \theta^{p+n\overline{x}-1}e^{-(q+n)\theta}$ ως κατανομή Poisson.*

Ως αριθμητικό παράδειγμα, πάλι από το First Bayes, θεωρούμε θ το μέσο πλήθος από χήνες μέσα σε ένα κοπάδι σε μία συγκεκριμένη περιοχή. Λεπτομερείς αεροφωτογραφίες από 45 κοπάδια έδωσαν $\overline{x} = \frac{4019}{45}$. Υποθέτουμε ότι η prior μέση τιμή και η prior διασπορά του θ είναι $\mu = E(\theta) = \frac{p}{q} = 100$ και $\sigma^2 = \mathrm{Var}(\theta) = \frac{p}{q^2} = 20$ αντίστοιχα. Μπορούμε να λύσουμε ως προς $q = \frac{\mu}{\sigma^2} = 5$ και $p = q \cdot \mu = 500$. Επομένως, $P = 500 + 4019 = 4519$ και $Q = 5+45 = 50$, δηλαδή $\theta \mid x \sim \mathrm{Gamma}(4519, 50)$. Η μέση τιμή και η διασπορά της posterior είναι:

$$E(\theta \mid x) = \frac{P}{Q} = \frac{4519}{50} = 90.4 \quad \text{και} \quad \mathrm{Var}(\theta \mid x) = \frac{E(\theta \mid x)}{Q} = \frac{90.4}{50} = 1.81.$$

Παράδειγμα 2.3. *(Μέση τιμή κανονικής). Έστω $x = (x_1, x_2, \ldots, x_n)$ ένα τυχαίο δείγμα μεγέθους n από μία τυχαία μεταβλητή X που ακολουθεί την κατανομή $N(\theta, \sigma^2)$ με σ^2 γνωστό, δηλαδή:*

$$f(x_i \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\theta)^2}{2\sigma^2}\right\}, \quad \theta \in \mathbb{R}.$$

Δεδομένης μίας μόνο παρατήρησης x_i από την κατανομή $N(\theta, \sigma^2)$, η πιθανοφάνεια του θ δίνεται από τη σχέση:

$$\begin{aligned}
f(x_i \mid \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\theta)^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{x_i^2 - 2x_i\theta + \theta^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\} \exp\left\{-\frac{\theta^2 - 2x_i\theta}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2} \cdot \theta^2 + \frac{x_i}{\sigma^2} \cdot \theta\right\}.
\end{aligned}$$

Τότε, η πιθανοφάνεια ολόκληρου του δείγματος είναι:

$$\begin{aligned}
f(x \mid \theta) &= \prod_{i=1}^{n} f(x_i \mid \theta) \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\sigma^2} \cdot \theta^2 + \frac{x_i}{\sigma^2} \cdot \theta\right\}
\end{aligned}$$

[2]Να σημειωθεί ότι δε χρειάζεται γνώση των επιμέρους τιμών x_i, παρά μόνο του δειγματικού μέσου τους. Λέμε ότι ο δειγματικός μέσος των x_i είναι επαρκής για το θ.

$$
\begin{aligned}
&= \exp\left\{\sum_{i=1}^{n}\left(-\frac{1}{2\sigma^2}\cdot\theta^2+\frac{x_i}{\sigma^2}\cdot\theta\right)\right\} \\
&= \exp\left\{-\frac{n}{2\sigma^2}\cdot\theta^2+\frac{n\overline{x}}{\sigma^2}\cdot\theta\right\}.
\end{aligned}
$$

Υποθέτουμε ότι οι prior πεποιθήσεις μας για το θ μπορούν κι αυτές να αναπαρασταθούν μέσω μίας κανονικής κατανομής, δηλαδή $\theta \sim N(b, d^2)$. Για άλλη μία φορά, αυτό το κάνουμε για να επιτύχουμε απλή μαθηματική ανάλυση, αλλά θα πρέπει να χρησιμοποιείται μόνο αν μία τέτοια επιλογή *είναι* μία καλή προσέγγιση των prior πεποιθήσεών μας για το θ.

Η μέση τιμή και η διασπορά αυτής της prior κατανομής είναι $E(\theta) = b$ και $\mathrm{Var}(\theta) = d^2$. Τώρα, μπορούμε να γράψουμε την prior κατανομή του θ ως εξής:

$$
\begin{aligned}
f(\theta) &= \frac{1}{\sqrt{2\pi d^2}}\exp\left\{-\frac{(\theta-b)^2}{2d^2}\right\} \\
&\propto \exp\left\{-\frac{\theta^2}{2d^2}+\frac{b\theta}{d^2}-\frac{b^2}{2d^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2d^2}\cdot\theta^2+\frac{b}{d^2}\cdot\theta\right\}.
\end{aligned}
$$

Παίρνουμε τώρα την posterior κατανομή του θ ως εξής:

$$
\begin{aligned}
f(\theta \mid x) &\propto f(\theta)\cdot f(x \mid \theta) \\
&\propto \exp\left\{-\frac{1}{2d^2}\cdot\theta^2+\frac{b}{d^2}\cdot\theta\right\}\exp\left\{-\frac{n}{2\sigma^2}\cdot\theta^2+\frac{n\overline{x}}{\sigma^2}\cdot\theta\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{d^2}+\frac{n}{\sigma^2}\right)\theta^2+\left(\frac{b}{d^2}+\frac{n\overline{x}}{\sigma^2}\right)\theta\right\} \\
&= \exp\left\{-\frac{1}{2D^2}\cdot\theta^2+\frac{B}{D^2}\cdot\theta\right\}.
\end{aligned}
$$

Επομένως, μπορούμε να συμπεράνουμε ότι $\theta \mid x \sim N(B, D^2)$, όπου:

$$
D^2 = \mathrm{Var}(\theta \mid x) = \left(\frac{1}{d^2}+\frac{n}{\sigma^2}\right)^{-1} \quad \text{και}
$$

$$
B = \mathrm{E}(\theta \mid x) = \left(\frac{b}{d^2}+\frac{n\overline{x}}{\sigma^2}\right)\cdot D^2 = \left(\frac{b}{d^2}+\frac{n\overline{x}}{\sigma^2}\right)\left(\frac{1}{d^2}+\frac{n}{\sigma^2}\right)^{-1}.
$$

Το αποτέλεσμα αυτό εκφράζεται πιο συνοπτικά αν ορίσουμε ως "ακρίβεια" την αντίστροφη της διασποράς, δηλαδή ορίσουμε $\tau = \frac{1}{\sigma^2}$ και $c = \frac{1}{d^2}$. Τότε,

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

Πριν δούμε ένα αριθμητικό παράδειγμα, μπορούμε να κάνουμε ορισμένες παρατηρήσεις:

1. Παρατηρούμε ότι $E(\theta \mid x) = \gamma_n b + (1-\gamma_n)\overline{x}$, όπου:

$$\gamma_n = \frac{c}{c+n\tau} \quad \text{και} \quad 1-\gamma_n = \frac{n\tau}{c+n\tau}.$$

   Έτσι, η posterior μέση τιμή είναι απλώς ένας σταθμισμένος μέσος όρος της prior μέσης τιμής και του $\overline{x}$. Επιπλέον, η παράμετρος στάθμισης γ_n καθορίζεται από τη σχετική ακρίβεια της prior και των δεδομένων. Δηλαδή, αν το $n\tau$ είναι μεγάλο σε σχέση με το c, τότε $\gamma \approx 0$ και η posterior μέση τιμή προσεγγίζει το $\overline{x}$.

2. Παρατηρούμε ότι ”posterior ακρίβεια” = ”prior ακρίβεια” + $n \cdot$ ”ακρίβεια δεδομένων”.

3. Καθώς $n \to \infty$, τότε (προσεγγιστικά),

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

   οπότε η prior δεν επιδρά πάνω στην οριακή κατανομή.

4. Καθώς $d \to \infty$ ή, ισοδύναμα, $c \to 0$, παίρνουμε πάλι ότι:

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

5. Παρατηρούμε ότι η posterior κατανομή εξαρτάται από τα δεδομένα μόνο μέσω του $\overline{x}$ και όχι από τις επιμέρους τιμές x_i. Λέμε πάλι ότι το $\overline{x}$ είναι επαρκές για το θ.

Οι παρατηρήσεις 3 και 4 είναι εύστοχες και θα μελετηθούν εκτενέστερα παρακάτω.

Ως αριθμητικό παράδειγμα, πάλι από το First Bayes, κοιτάμε ένα ιστορικό σετ δεδομένων το οποίο καταγράφηκε από τον Henry Cavendish κατά τον $18^{\text{ο}}$ αιώνα. Έκανε 23 μετρήσεις για την πυκνότητα της γης. Για αυτά δεδομένα, είναι $\overline{x} = 5.48$ και θα υποθέσουμε ότι η διασπορά των σφαλμάτων μέτρησης είναι 0.04. Τώρα, από προηγούμενα πειράματα παίρνουμε την prior $\theta \sim N(5.4, 0.01)$, όπου θ η μέση πυκνότητα της γης. Τότε,

$$D^2 = \left(\frac{1}{0.01} + \frac{23}{0.04}\right)^{-1} \approx 0.00148 \quad \text{και} \quad B = \left(\frac{5.4}{0.01} + \frac{23 \cdot 5.48}{0.04}\right) \cdot D^2 \approx 5.47,$$

οπότε $\theta \mid x \sim N(5.47, 0.00148)$.

2.5 Γενικά ζητήματα

Οι αρχές και οι λεπτομέρειες που συναντήθηκαν στα παραπάνω παραδείγματα γεννάνε μία σειρά ζητημάτων που είναι χρήσιμο να συζητηθούν τώρα.

2.5.1 Συνεχής ανανέωση πληροφορίας

Έχουμε δει ότι το θεώρημα του Bayes μάς παρέχει τον μηχανισμό με τον οποίο η prior πληροφορία μας ενημερώνεται από τα δεδομένα για να μας δώσει την posterior πληροφορία. Αυτή η posterior πληροφορία χρησιμεύει ως prior πληροφορία προτού γίνουν περισσότερα δεδομένα διαθέσιμα. Γεννάται λοιπόν το εξής ερώτημα: αν λαμβάνουμε μία ακολουθία από δεδομένα και ανανεώνουμε τις πεποιθήσεις μας κάθε φορά που έρχεται ένα καινούργιο μέρος των δεδομένων, θα πάρουμε διαφορετικό αποτέλεσμα σε σχέση με αυτό που θα παίρναμε αν περιμέναμε να έρθουν πρώτα όλα τα δεδομένα και μετά ανανεώναμε την prior; Αναλογιστείτε την απλή περίπτωση 2 ανεξάρτητων τυχαίων μεταβλητών X_1 και X_2, καθεμία από τις οποίες έχει πυκνότητα $f(x \mid \theta)$. Τώρα, υποθέστε ότι παρατηρούμε τη x_1 και ανανεώνουμε την prior μέσω του θεωρήματος του Bayes ως εξής:

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

Αυτή γίνεται η νέα μας prior προτού παρατηρήσουμε τη x_2. Έτσι,

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

το οποίο είναι το ίδιο αποτέλεσμα με αυτό που θα αποκτούσαμε ανανεώνοντας απευθείας την prior με ολόκληρη την πληροφορία των (x_1, x_2). Σύμφωνα με την αρχή της επαγωγής, αυτό το επιχείρημα γενικεύεται σε ακολουθίες οποιουδήποτε πλήθους παρατηρήσεων.

2.5.2 Επάρκεια

Στην κλασική συμπερασματολογία, η επάρκεια παίζει κεντρικό ρόλο τόσο στη θεωρητική ανάπτυξη όσο και στις πρακτικές εφαρμογές. Το κλασικό αποτέλεσμα βάσει του οποίου αναγνωρίζουμε ότι μία στατιστική συνάρτηση $s(x)$ είναι επαρκής για μία παράμετρο θ, είναι να γράφεται η πιθανοφάνεια ως $f(x \mid \theta) = g(x) \cdot h\,(s(x), \theta)$, όπου η $g(x)$ δεν εμπλέκει το θ, παρά μόνο τα δεδομένα (αν και ούτε απαραιτήτως τα δεδομένα - θα μπορούσε να είναι κάποια σταθερά).

Αν ισχύει αυτό, τότε στην Μπεϋζιανή ανάλυση θα έχουμε $f(x \mid \theta) \propto h\,(s(x), \theta)$, δηλαδή η πιθανοφάνεια θα εξαρτάται από τα δεδομένα μόνο μέσω της επαρκούς στατιστικής συνάρτησης $s(x)$, όπως έχουμε ήδη δει σε αρκετά παραδείγματα. Σε αυτή την περίπτωση, η posterior κατανομή $f(\theta \mid x)$ θα εξαρτάται κι αυτή από τα δεδομένα μόνο μέσω της επαρκούς στατιστικής συνάρτησης $s(x)$.

Ισχύει επίσης και το αντίστροφο, δηλαδή αν η posterior κατανομή $f(\theta \mid x)$ εξαρτάται από τα δεδομένα μόνο μέσω της συνάρτησης $s(x)$, τότε η $s(x)$ είναι μία επαρκής στατιστική συνάρτηση για το θ.

2.5.3 Η αρχή της πιθανοφάνειας

Η αρχή της πιθανοφάνειας λέει ότι αν δύο πειράματα αποδίδουν την ίδια πιθανοφάνεια (κατά αναλογία), τότε η συμπερασματολογία που αντλούμε για το θ θα πρέπει να είναι η ίδια σε κάθε περίπτωση. Με άλλα λόγια, όλα τα μέρη της συμπερασματολογίας θα πρέπει να βασίζονται μόνο στη συνάρτηση πιθανοφάνειας. Μία βασική αρετή του Μπεϋζιανού πλαισίου είναι ότι οι Μπεϋζιανές τεχνικές είναι έμφυτα συμβατές με την αρχή της πιθανοφάνειας, ενώ πολλές απλές διαδικασίες από την κλασική στατιστική την παραβιάζουν.

Παράδειγμα 2.4. *Θεωρούμε 2 πειράματα που αφορούν την εκτίμηση της πιθανότητας επιτυχίας θ σε ανεξάρτητες δοκιμές. Στο πρώτο πείραμα, καταγράφεται το πλήθος x των επιτυχιών σε n δοκιμές. Στο δεύτερο, καταγράφεται ο αριθμός y των δοκιμών που απαιτούνται ως την m-οστή επιτυχία.*

Οι κατανομές των τυχαίων μεταβλητών X, Y που περιγράφουν τα αποτελέσματα αυτών των 2 πειραμάτων διαφέρουν. Είναι η διωνυμική και η αρνητική διωνυμική κατανομή αντίστοιχα, δηλαδή:

$$f(x \mid \theta) = P_\theta(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0, 1, \dots, n \text{ και}$$

$$f(y \mid \theta) = P_\theta(Y = y) = \binom{y-1}{m-1}\theta^m(1-\theta)^{y-m}, \quad y = m, m+1, \dots.$$

Οι αντίστοιχες εκτιμήτριες μέγιστης πιθανοφάνειας είναι οι $\widehat{\theta} = \frac{x}{n}$ και $\widehat{\theta} = \frac{m}{y}$. Οι κατανομές αυτών των εκτιμητριών είναι αρκετά διαφορετικές. Για $n = 2$, το κλάσμα $\frac{x}{n}$ μπορεί να πάρει τις τιμές $0, \frac{1}{2}$ και 1. Για $m = 1$, το κλάσμα $\frac{m}{y}$ μπορεί να πάρει τις τιμές $1, \frac{1}{2}, \frac{1}{3}, \dots$. Όμως, αν επιπλέον συνέβαινε να παρατηρήσουμε $x = 1$ και $y = 2$, τότε:

$$f(x \mid \theta) = 2\theta(1-\theta) \quad \text{και} \quad f(y \mid \theta) = \theta(1-\theta),$$

δηλαδή οι συναρτήσεις πιθανοφάνειας θα ήταν ανάλογες και η Μπεϋζιανή συμπερασματολογία θα ήταν η ίδια και στις δύο περιπτώσεις. *Δε θα εξαρτώταν από την πιθανότητα οποιασδήποτε άλλης πιθανής έκβασης του πειράματος, δηλαδή από την κατανομή της X ή της Y.*

2.6 Ασκήσεις

Άσκηση 2.1. *Σε καθεμία από τις επόμενες περιπτώσεις, εξάγετε την posterior κατανομή:*

i. *Έστω $x_1, x_2, \dots, x_n$ τυχαίο δείγμα από κατανομή με συνάρτηση μάζας πιθανότητας $f(x_i \mid \theta) = \theta^{x_i - 1}(1-\theta)$ για $x_i = 1, 2, \dots$. Χρησιμοποιήστε την prior κατανομή Beta(p, q),* *δηλαδή:*

$$f(\theta) = \frac{\theta^{p-1}(1-\theta)^{q-1}}{B(p,q)}, \quad \theta \in (0,1).$$

ii. *Έστω $x_1, x_2, \ldots, x_n$ τυχαίο δείγμα από κατανομή με συνάρτηση μάζας πιθανότητας:*

$$f(x_i \mid \theta) = e^{-\theta} \cdot \frac{\theta^{x_i}}{x_i!}, \quad x_i = 0, 1, \ldots.$$

Χρησιμοποιήστε την prior κατανομή $f(\theta) = e^{-\theta}$ για $\theta > 0$.

Άσκηση 2.2. *Το ποσοστό θ των ελαττωματικών προϊόντων σε ένα φορτίο είναι άγνωστο, αλλά έμπειρη αξιολόγηση εκχωρεί στο θ την Beta$(2, 200)$ prior κατανομή. Αν διαλέξουμε 100 προϊόντα στην τύχη από το φορτίο και 3 από αυτά βρεθούν να είναι ελαττωματικά, ποια είναι η posterior κατανομή του θ;*

Αν μία άλλη στατιστικός, έχοντας παρατηρήσει 3 ελαττωματικά προϊόντα, υπολόγισε την posterior να είναι μία κατανομή Βήτα με μέση τιμή $\frac{4}{102}$ και διασπορά 0.0003658, τότε ποια prior κατανομή είχε χρησιμοποιήσει;

Άσκηση 2.3. *Η διάμετρος ενός εξαρτήματος από μία μεγάλη σειρά παραγωγής ποικίλει σύμφωνα με μία κατανομή $N(\theta, 1)$. Ένας μηχανικός προσδιορίζει ότι η prior κατανομή για το θ είναι $N(10, 0.25)$. Σε μία σειρά παραγωγής, 12 εξαρτήματα εξετάζονται και βρίσκεται ότι έχουν δειγματική μέση διάμετρο ίση με $\frac{31}{3}$. Χρησιμοποιήστε αυτή την πληροφορία για να βρείτε την posterior κατανομή της μέσης διαμέτρου ενός εξαρτήματος. Έπειτα, υπολογίστε την πιθανότητα η διάμετρος να είναι πάνω από 10 μονάδες μήκους.*

Άσκηση 2.4. *Το πλήθος των ατελειών σε ένα ρολό μαγνητικής ταινίας ακολουθεί κατανομή Poisson(θ). Η prior κατανομή για το θ είναι η Gamma$(3, 1)$. Διαλέγουμε στην τύχη 5 τέτοια ρολά μαγνητικής ταινίας και βρίσκουμε ότι το πλήθος των ατελειών σε καθένα από αυτά είναι 2, 2, 6, 0 και 3 αντίστοιχα. Προσδιορίστε την posterior κατανομή του θ.*

Άσκηση 2.5. *Έστω ότι ο χρόνος εξυπηρέτησης x_i σε λεπτά ενός πελάτη σε μία τράπεζα ακολουθεί την κατανομή Exp(θ) με σ.π.π. $f(x_i \mid \theta) = \theta e^{-\theta x_i}$ για $x_i > 0$. Η prior για το θ καθορίζεται ότι είναι μία κατανομή Γάμμα με μέση τιμή 0.2 και τυπική απόκλιση 1. Παρατηρούνται οι χρόνοι $x_1, x_2, \ldots, x_{20}$ για την εξυπηρέτηση 20 πελατών και ο δειγματικός μέσος $\overline{x}$ είναι 3.8 λεπτά. Προσδιορίστε την posterior κατανομή του θ.*

Άσκηση 2.6. *Ένα τυχαίο δείγμα $x_1, x_2, \ldots, x_n$ λαμβάνεται από μία κατανομή Poisson με μέση τιμή θ. Η prior κατανομή του θ είναι μία κατανομή Γάμμα με μέση τιμή μ_0. Αν ο δειγματικός μέσος είναι $\overline{x}_n$, δείξτε ότι η μέση τιμή της posterior κατανομής του θ θα είναι ένας σταθμισμένος μέσος όρος της μορφής $\gamma_n \overline{x}_n + (1-\gamma_n)\mu_0$ και δείξτε ότι $\gamma_n \to 1$ καθώς $n \to \infty$.*

Άσκηση 2.7. *Μία μηχανικός χρησιμοποιεί μία κατανομή Gamma(p, q) ως prior για μία παράμετρο θ που αντιπροσωπεύει το μέσο πλήθος x των ελαττωμάτων σε ένα πλαστικό καλούπι.*

i. *Δεδομένου ότι η μέση τιμή και η διασπορά της κατανομής Γάμμα είναι $\frac{p}{q}$ και $\frac{p}{q^2}$ αντίστοιχα, τι τιμές θα πρέπει να επιλέξει για τα p και q, ώστε το θ να έχει μέση τιμή 12 και τυπική απόκλιση 4;*

ii. *Η μηχανικός έπειτα εξακριβώνει ότι υπάρχουν συνολικά 37 ελαττώματα σε 5 από τα πλαστικά καλούπια. Υποθέτοντας ότι τα πλήθη των ελαττωμάτων σε κάθε καλούπι είναι ανεξάρτητα και ακολουθούν μία κατανομή Poisson, βρείτε την posterior κατανομή του θ δεδομένης αυτής της πληροφορίας.*

iii. *Ποια είναι η posterior μέση τιμή του θ;*

Άσκηση 2.8. *Λαμβάνονται παρατηρήσεις $y_1, y_2, \ldots, y_n$ από ανεξάρτητες και κανονικά κατανεμημένες τυχαίες μεταβλητές καθεμία με κοινή (γνωστή) διασπορά σ^2, αλλά με αντίστοιχες μέσες τιμές $x_1\theta, x_2\theta, \ldots, x_n\theta$. Οι τιμές των $x_1, x_2, \ldots, x_n$ είναι γνωστές, αλλά το θ είναι άγνωστο.*

i. *Δείξτε ότι η πιθανοφάνεια, δεδομένης μόνο μίας μόνο παρατήρησης y_i, είναι της μορφής:*
$$f(y_i \mid \theta) \propto \exp\left\{-\frac{x_i^2}{2\sigma^2} \cdot \theta^2 + \frac{x_i y_i}{\sigma^2} \cdot \theta\right\}.$$

ii. *Δεδομένου ότι η prior κατανομή της άγνωστης παραμέτρου θ μπορεί να περιγραφεί ως κανονική με μέση τιμή b και διασπορά $\frac{\sigma^2}{a^2}$, δείξτε ότι η posterior κατανομή του θ είναι ανάλογη με τη συνάρτηση:*
$$\exp\left\{-\frac{1}{2\sigma^2}\left(a^2 + \sum_{i=1}^n x_i^2\right)\theta^2 + \frac{1}{\sigma^2}\left(a^2 b + \sum_{i=1}^n x_i y_i\right)\theta\right\}.$$

iii. *Χρησιμοποιήστε αυτό για να γράψετε την posterior μέση τιμή του θ. Δείξτε ότι μπορεί να γραφτεί ως:*
$$E(\theta \mid y) = wb + (1-w)\left(\sum_{i=1}^n x_i y_i\right)\left(\sum_{i=1}^n x_i^2\right)^{-1},$$
και βρείτε μία έκφραση για το w.

Κεφάλαιο 3

Προσδιορισμός Prior Κατανομών

3.1 Εισαγωγή

Είδαμε ότι η θεμελιώδης διαφορά μεταξύ της Μπεϋζιανής και της κλασικής στατιστικής είναι ότι στην Μπεϋζιανή στατιστική οι άγνωστες παράμετροι αντιμετωπίζονται ως τυχαίες μεταβλητές και ότι η χρήση του θεωρήματος του Bayes απαιτεί τον προσδιορισμό prior κατανομών για αυτές τις παραμέτρους. Ενώ αυτό διευκολύνει την ενσωμάτωση γνήσιων prior πεποιθήσεων για τις παραμέτρους, η επιλογή της prior κατανομής δεν μπορεί να γίνει στα τυφλά - χρειάζεται μεγάλη προσοχή και εμπλέκονται ορισμένα πολύ ουσιώδη ζητήματα. Σε αυτό το κεφάλαιο θα δούμε ορισμένα από αυτά τα ζητήματα.

3.2 Συζυγείς prior κατανομές

Οι υπολογιστικές δυσκολίες ανακύπτουν στη χρήση του θεωρήματος του Bayes όταν είναι απαραίτητο να υπολογίσουμε τη σταθερά κανονικοποίησης στον παρονομαστή:

$$\int_{\Theta} f(\theta) \cdot f(x \mid \theta) d\theta.$$

Για παράδειγμα, έστω ότι οι $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές από την κατανομή Poisson(θ) και η πεποίθησή μας για το θ είναι ότι βρίσκεται *σίγουρα* στο διάστημα $(0,1)$, αλλά όλες οι τιμές μέσα σε αυτό το εύρος είναι ισοπίθανες, δηλαδή $f(\theta) = 1$ για $\theta \in (0,1)$ και $f(\theta \mid x) \propto e^{-n\theta}\theta^{n\overline{x}}$. Τότε, η σταθερά κανονικοποίησης είναι:

$$\int_0^1 e^{-n\theta}\theta^{n\overline{x}} d\theta,$$

δηλαδή μία μη-πλήρης συνάρτηση Γάμμα η οποία μπορεί να υπολογιστεί μόνο μέσω αριθμητικών μεθόδων.

Επομένως, ακόμα και απλές επιλογές prior κατανομών μπορούν να οδηγήσουν σε δύσκολα αριθμητικά προβλήματα. Ωστόσο, είδαμε 3 παραδείγματα στο προηγούμενο κεφάλαιο όπου συνετές επιλογές prior οδήγησαν σε υπολογισμούς posterior που δεν απαιτούσαν κάποια ολοκλήρωση. Σε καθεμία από αυτές τις περιπτώσεις, μπορέσαμε να αναγνωρίσουμε μία prior κατανομή για την οποία η posterior κατανομή άνηκε στην ίδια οικογένεια κατανομών με την prior - τέτοιες prior κατανομές λέγονται *συζυγείς prior*. Ας δούμε ένα ακόμα παράδειγμα όπου μπορεί να βρεθεί μία συζυγής prior.

Παράδειγμα 3.1. *(Δείγμα Γάμμα). Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές από την κατανομή Gamma(k, θ), όπου το k είναι γνωστό. Να σημειωθεί ότι η περίπτωση $k = 1$ αντιστοιχεί στην εκθετική κατανομή.*

Τότε,

$$\begin{aligned} f(x_i \mid \theta) &= \frac{\theta^k}{\Gamma(k)} x_i^{k-1} e^{-\theta x_i} \\ &\propto \theta^k e^{-\theta x_i}, \quad \text{οπότε:} \\ f(x \mid \theta) &\propto \prod_{i=1}^{n} \theta^k e^{-\theta x_i} \\ &= \theta^{nk} e^{-n\overline{x}\theta}, \quad \theta > 0. \end{aligned}$$

Τώρα, η μελέτη της $\theta^{nk} e^{-n\overline{x}\theta}$ ως συνάρτησης του θ, υποδεικνύει ότι θα μπορούσαμε να πάρουμε μία prior της μορφής $f(\theta) \propto \theta^{p-1} e^{-q\theta}$, δηλαδή $\theta \sim \text{Gamma}(p, q)$, αφού τότε από το θεώρημα του Bayes θα παίρναμε:

$$f(\theta \mid x) \propto \theta^{p+nk-1} e^{-(q+n\overline{x})\theta},$$

οπότε $\theta \mid x \sim \text{Gamma}(p + nk, q + n\overline{x})$.

3.2.1 Χρήση συζυγών prior

Δεν πρέπει να θεωρούμε τη χρήση των συζυγών prior ως κάτι παραπάνω από αυτό που είναι: μία βολική μαθηματική τεχνική. Ωστόσο, η έκφραση των prior πεποιθήσεων κάποιου μέσω μίας παραμετρικής κατανομής είναι πάντα μία προσέγγιση. Σε πολλές περιπτώσεις, η συζυγής οικογένεια κατανομών είναι αρκετά πλούσια, ώστε να μπορεί να βρεθεί μία συζυγής prior που να είναι αρκετά κοντά στις prior πεποιθήσεις κάποιου και, επομένως, να είναι αποδεκτή αυτή η προσέγγιση. Όμως, αν αυτό δεν ισχύει, δεν πρέπει να χρησιμοποιηθούν απλώς επειδή κάνουν τους μαθηματικούς υπολογισμούς ευκολότερους.

3.2.2 Προσδιορισμός συζυγών prior

Υπό την προϋπόθεση ότι δε συγκρούεται άμεσα με τις prior πεποιθήσεις μας και ότι μία τέτοια οικογένεια μπορεί να βρεθεί, η απλότητα που επιφέρει η χρήση μίας συζυγούς prior είναι ελκυστική. Όμως, σε ποιες καταστάσεις μπορεί να βρεθεί μία συζυγής οικογένεια κατανομών;

Προκύπτει ότι η μόνη περίπτωση όπου συζυγείς κατανομές μπορούν να βρεθούν εύκολα είναι για μοντέλα δεδομένων μέσα από την *εκθετική οικογένεια κατανομών*. Δηλαδή για μοντέλα για τα οποία $f(x \mid \theta) = h(x)g(\theta)e^{t(x)c(\theta)}$, όπου h, g, t και c συναρτήσεις τέτοιες, ώστε:

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

Αυτό μπορεί να φαίνεται περιοριστικό, αλλά στην πραγματικότητα περιλαμβάνει την εκθετική κατανομή, την κατανομή Poisson, την κατανομή Γάμμα με γνωστή παράμετρο σχήματος, τη διωνυμική κατανομή με γνωστό πλήθος δοκιμών και την κανονική κατανομή.

Δεδομένου ενός τυχαίου δείγματος $x = (x_1, x_2, \ldots, x_n)$ από αυτήν τη γενική κατανομή, η πιθανοφάνεια για το θ είναι:

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

Έτσι, αν διαλέξουμε μία prior της μορφής $f(\theta) \propto [g(\theta)]^d \, e^{bc(\theta)}$, παίρνουμε:

$$\begin{aligned} f(\theta \mid x) &\propto f(\theta) \cdot f(x \mid \theta) \\ &\propto [g(\theta)]^d \, e^{bc(\theta)} \cdot [g(\theta)]^n \exp\left\{\sum_{i=1}^n t(x_i)c(\theta)\right\} \\ &= [g(\theta)]^{d+n} \exp\left\{\left[b + \sum_{i=1}^n t(x_i)\right] c(\theta)\right\} \\ &= [g(\theta)]^D \, e^{Bc(\theta)}, \end{aligned}$$

όπου $D = d + n$ και $B = b + \sum_{i=1}^{n} t(x_i)$. Αυτή η επιλογή οδηγεί σε μία posterior που ανήκει στην ίδια οικογένεια κατανομών με την prior, αλλά με τροποποιημένες παραμέτρους.

Μπορεί εύκολα να ελεγχθεί ότι όλα τα παραδείγματα συζυγών prior κατανομών που συναντήσαμε έως τώρα μπορούν να προκύψουν με αυτόν τον τρόπο. Για παράδειγμα, μία διωνυμική τυχαία μεταβλητή έχει συνάρτηση πιθανότητας:

$$\begin{aligned} f(x \mid \theta) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \binom{n}{x} (1-\theta)^n \left(\frac{\theta}{1-\theta}\right)^x \\ &= \binom{n}{x} (1-\theta)^n \exp\left\{x \log \frac{\theta}{1-\theta}\right\}. \end{aligned}$$

Οπότε, σύμφωνα με τον συμβολισμό της εκθετικής οικογένειας κατανομών, έχουμε ότι $h(x) = \binom{n}{x}$, $g(\theta) = (1-\theta)^n$, $t(x) = x$ και $c(\theta) = \log \frac{\theta}{1-\theta}$. Έτσι, κατασκευάζουμε μία συζυγή prior της μορφής:

$$\begin{aligned} f(\theta) &\propto (1-\theta)^{nd} \exp\left\{ b \log \frac{\theta}{1-\theta} \right\} \\ &= \theta^b (1-\theta)^{nd-b}, \end{aligned}$$

η οποία είναι μέλος της Βήτα οικογένειας κατανομών.

3.2.3 Τυπικές συζυγείς αναλύσεις

Ο πίνακας 3.1 απαριθμεί πολλές από τις τυπικές συζυγείς αναλύσεις ζευγών prior και πιθανοφανειών.

Πιθανοφάνεια	Prior	Posterior
$x_i \sim \text{Bernoulli}(\theta)$	$\text{Beta}(p,q)$	$\text{Beta}\left(p+n\overline{x}, q+n-n\overline{x}\right)$
$x_i \sim \text{Bin}(k_i,\theta)$, k_i γνωστά	$\text{Beta}(p,q)$	$\text{Beta}\left(p+n\overline{x}, q+\sum_{i=1}^{n} k_i - n\overline{x}\right)$
$x_i \sim \text{Geom}(\theta)$	$\text{Beta}(p,q)$	$\text{Beta}\left(p+n, q+n\overline{x}-n\right)$
$x_i \sim \text{NegBin}(k_i,\theta)$, k_i γνωστά	$\text{Beta}(p,q)$	$\text{Beta}\left(p+\sum_{i=1}^{n} k_i, q+n\overline{x}-\sum_{i=1}^{n} k_i\right)$
$x_i \sim \text{Poisson}(\theta)$	$\text{Gamma}(p,q)$	$\text{Gamma}\left(p+n\overline{x}, q+n\right)$
$x_i \sim \text{Exp}(\theta)$	$\text{Gamma}(p,q)$	$\text{Gamma}\left(p+n, q+n\overline{x}\right)$
$x_i \sim \text{Gamma}(k_i,\theta)$, k_i γνωστά	$\text{Gamma}(p,q)$	$\text{Gamma}\left(p+\sum_{i=1}^{n} k_i, q+n\overline{x}\right)$
$x_i \sim N\left(\theta,\tau^{-1}\right)$, τ γνωστό	$N\left(b,c^{-1}\right)$	$N\left(\frac{cb+n\tau\overline{x}}{c+n\tau}, \frac{1}{c+n\tau}\right)$
$x_i \sim N\left(\mu_i,\theta^{-1}\right)$, μ_i γνωστά	$\text{Gamma}\left(p,q\right)$	$\text{Gamma}\left(p+\frac{n}{2}, q+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_i)^2\right)$

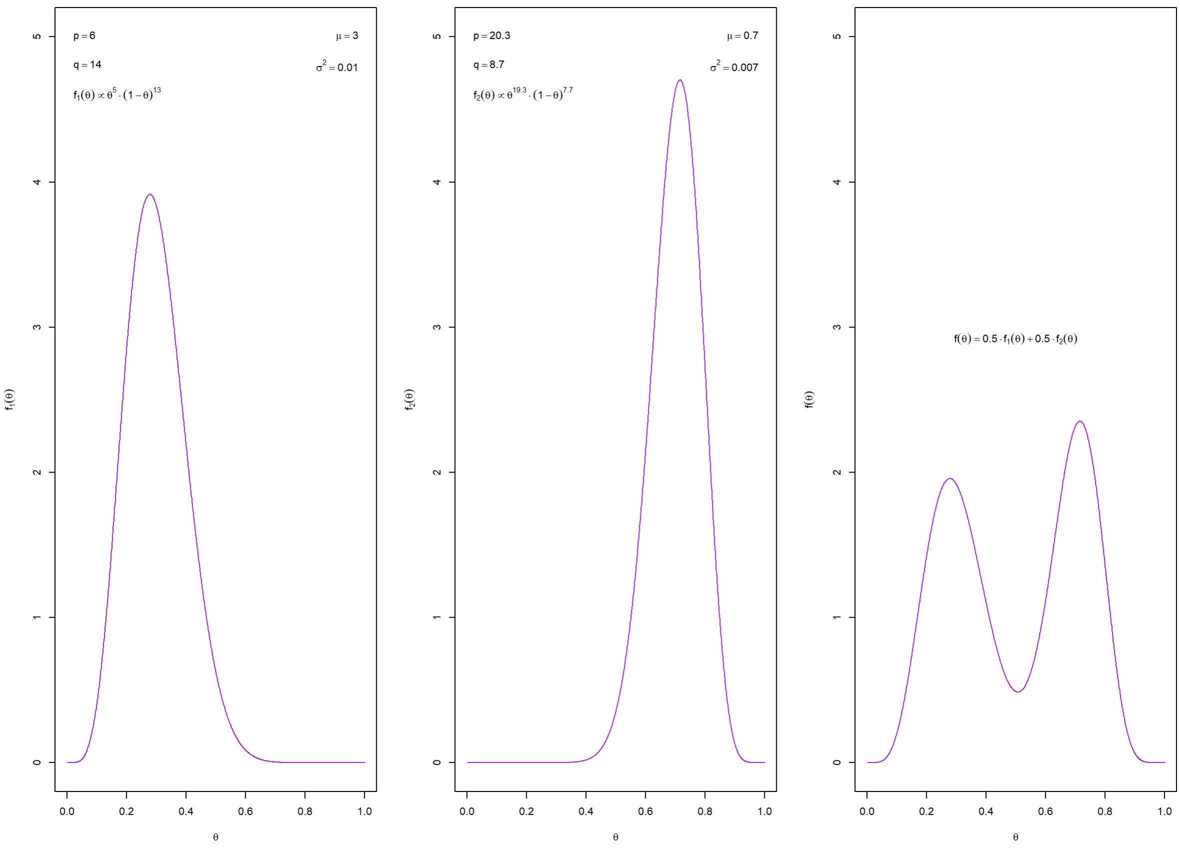
Πινακας 3.1: Τυπικές συζυγείς αναλύσεις

Ορισμένες από αυτές έχουν δοθεί ως παραδείγματα - τις υπόλοιπες καλό θα ήταν να τις επιβεβαιώσετε.

3.3 Μίξεις prior κατανομών

Τα επιχειρήματα για τη χρήση συζυγών οικογενειών prior κατανομών εκτέθηκαν στην παράγραφο 3.1. Ωστόσο, τονίζουμε ξανά ότι τέτοιες οικογένειες πρέπει να χρησιμοποιηθούν μόνο όταν μπορεί να βρεθεί ένα κατάλληλο μέλος της οικογένειας που να είναι σύμφωνο με τις πραγματικές prior πεποιθήσεις μας. Σε ορισμένες περιπτώσεις, η φυσιολογική συζυγής οικογένεια μπορεί να είναι πολύ περιοριστική για να γίνει αυτό εφικτό.

Ας αναλογιστούμε το ακόλουθο παράδειγμα (το οποίο επικαλείται συχνά). Όταν ρίχνουμε ένα νόμισμα, τότε υπάρχει σχεδόν πάντα πιθανότητα 0.5 να φέρει ”Κ”. Ωστόσο, αν το νόμισμα *περιστραφεί* πάνω σε ένα τραπέζι, είναι σύνηθες οι μικρές ατέλειες στην άκρη του νομίσματος να προκαλούν μία τάση να φέρνει περισσότερο είτε "Κ" είτε "Γ". Λαμβάνοντας υπόψη αυτήν την πληροφορία, θα θέλαμε ίσως να δώσουμε στην πιθανότητα θ το νόμισμα να φέρει "Κ" μία prior κατανομή που να ευνοεί τιμές κοντά είτε στο 0.3 είτε στο 0.7. Δηλαδή, οι prior πεποιθήσεις μας μπορούν εύλογα να αναπαρασταθούν μέσω μίας *δικόρυφης* κατανομής (ή ακόμα και *τρικόρυφης* αν θέλουμε να δώσουμε επιπλέον βάρος και στην αμερόληπτη πιθανότητα $\theta = 0.5$). Το μοντέλο πιθανοφάνειας για το πλήθος των "Κ" σε n περιστροφές του νομίσματος θα είναι διωνυμικό: $X \mid \theta \sim \text{Bin}(n, \theta)$, οπότε η συζυγής prior οικογένεια είναι η οικογένεια κατανομών Βήτα. Ωστόσο, καμία κατανομή μέλος αυτής της οικογένειας δεν είναι δικόρυφη. Μία λύση είναι να χρησιμοποιήσουμε *μίξεις* συζυγών κατανομών (βλέπε σχήμα 3.1).



Σχημα 3.1: Μίξη prior κατανομών

Αυτή η εκτεταμένη οικογένεια θα είναι επίσης συζυγής prior οικογένεια για τον ακόλουθο λόγο. Έστω $f_1(\theta), f_2(\theta), \ldots, f_k(\theta)$ συζυγείς κατανομές για το θ οι οποίες οδηγούν στις posterior κατανομές $f_1(\theta|x), f_2(\theta|x), \ldots, f_k(\theta|x)$. Τώρα, θεωρούμε την οικογένεια μίξεων κατανομών:

$$f(\theta) = \sum_{i=1}^{k} p_i f_i(\theta),$$

όπου $0 \leqslant p_i \leqslant 1$ για $i = 1, 2, \ldots, k$ και $\sum\limits_{i=1}^{k} p_i = 1$. Τότε,

$$\begin{aligned} f(\theta \mid x) &\propto f(\theta) \cdot f(x \mid \theta) \\ &= \sum_{i=1}^{k} p_i f_i(\theta) f(x \mid \theta) \\ &= \sum_{i=1}^{k} p_i f_i(x) f_i(\theta \mid x), \quad \text{όπου:} \end{aligned}$$

$$f_i(\theta \mid x) = \frac{f_i(\theta) \cdot f(x \mid \theta)}{f_i(x)} \quad \text{και} \quad f_i(x) = \int_{\Theta} f_i(\theta) \cdot f(x \mid \theta) d\theta.$$

Οπότε,

$$f(\theta \mid x) = \sum_{i=1}^{k} p_i^* f_i(\theta \mid x), \text{ όπου } p_i^* = \frac{p_i f_i(x)}{\sum_{j=1}^{k} p_j f_j(x)}.$$

Συνεπώς, η posterior ανήκει στην ίδια οικογένεια μίξεων κατανομών. Παρατηρούμε πάντως ότι τα βάρη p_i^* της μίξης στην posterior θα είναι γενικά διαφορετικά από τα αντίστοιχα βάρη p_i στην prior.

Μπορεί να αποδειχθεί ότι οι πεπερασμένες μίξεις συζυγών priors μπορούν να κατασκευαστούν έτσι ώστε να είναι οσοδήποτε κοντά σε *οποιαδήποτε* prior κατανομή. Ωστόσο, το πλήθος των όρων στη μίξη ενδέχεται να είναι μεγάλο και μπορεί να είναι πιθανό να αναπαραστήσουμε τις prior πεποιθήσεις κάποιου με μεγαλύτερη σαφήνεια χρησιμοποιώντας άλλες μη-συζυγείς οικογένειες μοντέλων.

3.4 Ακατάλληλες prior κατανομές

Θεωρούμε πάλι την posterior ανάλυση που προκύπτει όταν εκτιμούμε τη μέση τιμή της κανονικής κατανομής με γνωστή διασπορά, χρησιμοποιώντας μία κανονική prior κατανομή. Δηλαδή, $X_1, X_2, \ldots, X_n \sim N\left(\theta, \tau^{-1}\right)$ και $\theta \sim N\left(b, c^{-1}\right)$, τα οποία οδηγούν στο αποτέλεσμα $\theta \mid x \sim \left(\frac{cb+n\tau\overline{x}}{c+n\tau}, \frac{1}{c+n\tau}\right)$. Η ισχύς των prior πεποιθήσεών μας για το θ καθορίζεται από τη διασπορά ή ισοδύναμα από την ακρίβεια c της κανονικής prior κατανομής. Μεγάλες τιμές του c αντιστοιχούν σε πολύ ισχυρές prior πεποιθήσεις - από την άλλη μεριά, μικρές τιμές του c αντικατοπτρίζουν πολύ ασθενή prior πληροφορία. Τώρα, υποθέτουμε ότι οι prior πεποιθήσεις μας για το θ είναι τόσο ασθενείς ώστε $c \to 0$. Τότε, η posterior γίνεται απλά $N\left(\overline{x}, \frac{1}{n\tau}\right)$ ή, χρησιμοποιώντας πιο οικείο συμβολισμό, $N\left(\overline{x}, \frac{\sigma^2}{n}\right)$. Οπότε, παίρνουμε φαινομενικά μία απολύτως έγκυρη posterior κατανομή μέσω αυτής της οριακής διαδικασίας.

Όμως, υπάρχει ένα πρόβλημα. Ας σκεφτούμε τι συμβαίνει στην prior καθώς $c \to 0$. Στην πράξη, παίρνουμε μία κατανομή $N(b, \infty)$, η οποία δεν είναι γνήσια κατανομή. Στην πραγματικότητα, καθώς $c \to 0$, η κατανομή $N(b, c^{-1})$ γίνεται ολοένα και πιο επίπεδη, τόσο που σε κάθε διάστημα της μορφής $-K \leq \theta \leq K$, δεδομένου ότι το c είναι επαρκώς κοντά στο

0, έχουμε προσεγγιστικά $f(\theta) \propto 1$. Όμως, αυτό δεν μπορεί να είναι έγκυρο, καθώς $c \to 0$, πάνω σε ολόκληρη την πραγματική ευθεία, αφού:

$$\int_{\mathbb{R}} f(\theta) d\theta = \infty.$$

Έτσι, η posterior $N\left(\overline{x}, \frac{\sigma^2}{n}\right)$ που παίρνουμε επιτρέποντας $c \to 0$ στην τυπική συζυγή ανάλυση, δεν μπορεί να προκύψει μέσω της χρήσης οποιασδήποτε ορθής prior κατανομής. Προκύπτει ωστόσο μέσω της επίσημης χρήσης της prior $f(\theta) \propto 1$, η οποία είναι ένα παράδειγμα αυτού που ονομάζουμε *ακατάλληλη* prior κατανομή.

Επομένως, είναι ορθό να χρησιμοποιούμε μία posterior κατανομή η οποία προκύπτει προσδιορίζοντας μία ακατάλληλη prior για να αντικατοπτρίσει την ανεπαρκή γνώση μας; Παρόλο που εμπλέκονται κάποιες περαιτέρω δυσκολίες, η χρήση ακατάλληλων prior κατανομών θεωρείται αποδεκτή. Το βασικό επιχείρημα είναι ότι αν διαλέγαμε το c να έχει οποιαδήποτε άλλη τιμή εκτός από το 0, θα παίρναμε μία απολύτως ορθή prior κατανομή και δε θα υπήρχαν καθόλου ενδοιασμοί για την επακόλουθη posterior ανάλυση. Έτσι, θα μπορούσαμε να διαλέξουμε το c να είναι οσοδήποτε κοντά στο 0 και να πάρουμε μία posterior οσοδήποτε κοντά σε αυτήν που πραγματικά πήραμε χρησιμοποιώντας την ακατάλληλη prior $f(\theta) \propto 1$.

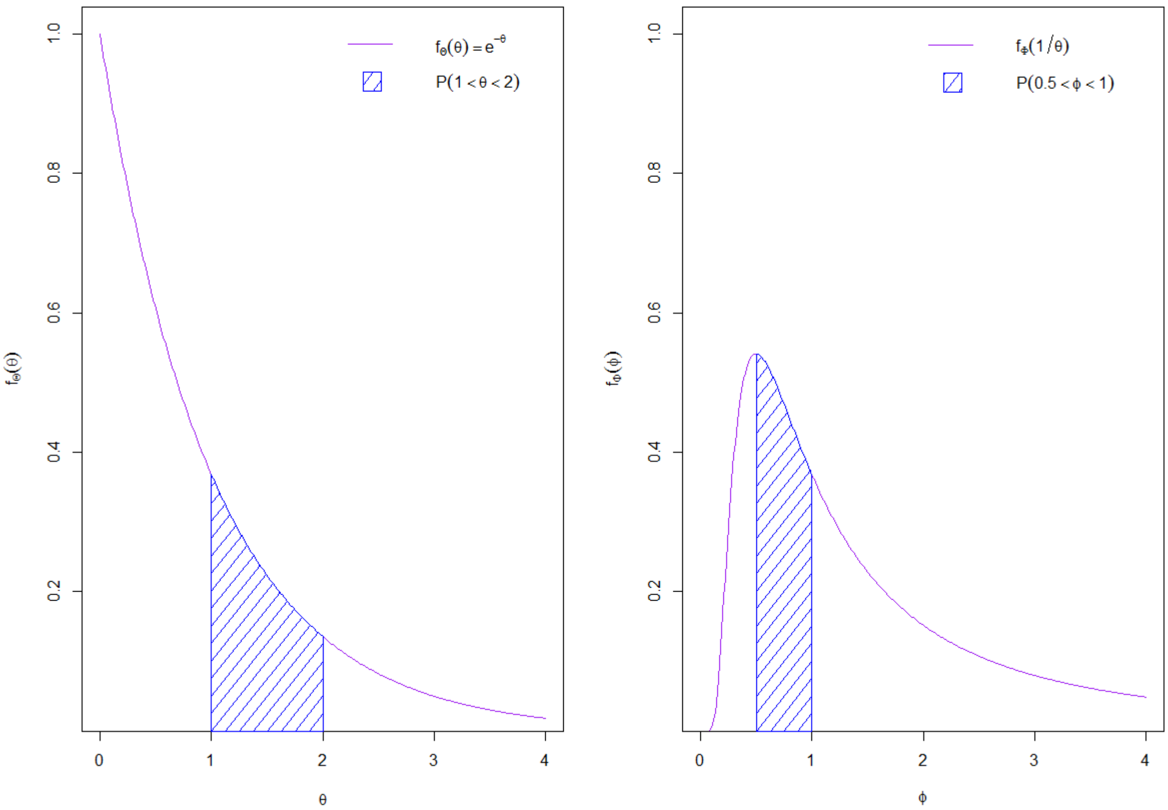
3.5 Η prior του Jeffreys

Στην προηγούμενη παράγραφο, είδαμε ότι η προσπάθεια να αναπαραστήσουμε την άγνοια μας στα πλαίσια της τυπικής συζυγούς ανάλυσης για τη μέση τιμή της κανονικής κατανομής οδήγησε στην έννοια της ακατάλληλης prior. Όμως, υπάρχουν και πιο θεμελιώδη ζητήματα. Έστω ότι έχουμε προσδιορίσει μία prior $f_\Theta(\theta)$ για μία παράμετρο θ σε ένα μοντέλο. Είναι αρκετά λογικό να αποφασίσουμε μετά να χρησιμοποιήσουμε στη θέση της την παράμετρο $\phi = \frac{1}{\theta}$. Για παράδειγμα, το θ μπορεί να είναι η παράμετρος της εκθετικής κατανομής των ενδιάμεσων χρόνων αφίξεων σε μία ουρά και να αντιπροσωπεύει τον ρυθμό αφίξεων. Τότε, το ϕ αντιπροσωπεύει τον μέσο ενδιάμεσο χρόνο αφίξεων. Σύμφωνα με τη θεωρία πιθανοτήτων, ξέρουμε ότι η αντίστοιχη prior πυκνότητα για το ϕ δίνεται από τον τύπο αλλαγής μεταβλητών:

$$\begin{aligned} f_\Phi(\phi) &= f_\Theta(\theta) \cdot \left|\frac{d\theta}{d\phi}\right| \\ &= f_\Theta\left(\frac{1}{\phi}\right) \cdot \frac{1}{\phi^2}. \end{aligned}$$

Στο σχήμα 3.2, απεικονίζονται οι γραφικές παραστάσεις των prior κατανομών μίας παραμέτρου θ και της $\phi = \frac{1}{\theta}$. Οι γραμμοσκιασμένες περιοχές δείχνουν το προφανές, δηλαδή ότι $P(1 < \theta < 2) = P\left(\frac{1}{2} < \phi < 1\right)$.

Υπάρχει ένας τρόπος να χρησιμοποιήσουμε την πιθανοφάνεια $f(x \mid \theta)$ ή ακριβέστερα τη



Σχημα 3.2: Prior για το θ και το $\phi = \frac{1}{\theta}$

λογαριθμοπιθανοφάνεια $\ell(\theta) = \log f(x \mid \theta)$ για να προσδιορίσουμε μία "μη-πληροφοριακή" prior που να είναι *συνεπής* ανάμεσα σε όλους τους 1-1 μετασχηματισμούς παραμέτρων. Αυτή είναι η prior του Jeffreys και βασίζεται στην έννοια της πληροφορίας του Fisher:

$$I(\theta) = -E\left[\frac{d^2\ell(\theta)}{d\theta^2}\right] = E\left[\left(\frac{d\ell(\theta)}{d\theta}\right)^2\right] \geqslant 0.$$

Τότε, η prior του Jeffreys ορίζεται ως:

$$J_\Theta(\theta) \propto \sqrt{I(\theta)}.$$

Η "συνέπεια" αυτής της prior κάτω από έναν 1-1 μετασχηματισμό παραμέτρου $\phi = \phi(\theta)$ μπορεί να διατυπωθεί ως:

$$f_\Phi(\phi) = J_\Theta(\theta) \cdot \left|\frac{d\theta}{d\phi}\right| \Rightarrow f_\Phi(\phi) = J_\Phi(\phi) \propto \sqrt{I(\phi)},$$

δηλαδή αν χρησιμοποιήσουμε την prior του Jeffreys για μία παράμετρο θ και έπειτα τη μετασχηματίσουμε σε μία παράμετρο ϕ, τότε η αντίστοιχη prior για το ϕ, η οποία δίνεται μέσω του τύπου αλλαγής μεταβλητών, δε θα είναι άλλη από την prior του Jeffreys για το ϕ. Με άλλα λόγια, η ασθενής πληροφόρηση που δίνει η prior του Jeffreys για το θ παραμένει ασθενής για όλους τους 1-1 μετασχηματισμούς του θ.

Αντικαθιστώντας από τον ορισμό της prior του Jeffreys και παίρνοντας τετράγωνα, αρκεί να επαληθεύσουμε ότι:

$$I(\phi) = I(\theta) \cdot \left| \frac{d\theta}{d\phi} \right|^2 .$$

Μπορούμε να το κάνουμε αυτό χρησιμοποιώντας τον κανόνα της αλυσίδας για αλλαγή μεταβλητής στην παραγώγιση. Έχουμε:

$$\ell_\Phi(\phi) = \ell_\Theta\left(\theta(\phi)\right) \Rightarrow$$

$$\frac{d\ell_\Phi(\phi)}{d\phi} = \frac{d\ell_\Theta(\theta)}{d\theta} \cdot \frac{d\theta(\phi)}{d\phi} \Rightarrow$$

$$I(\phi) = E\left[\left(\frac{d\ell(\phi)}{d\phi}\right)^2\right] = E\left[\left(\frac{d\ell(\theta)}{d\theta} \cdot \frac{d\theta}{d\phi}\right)^2\right] = E\left[\left(\frac{d\ell(\theta)}{d\theta}\right)^2\right]\left(\frac{d\theta}{d\phi}\right)^2 = I_\Theta(\theta) \cdot \left(\frac{d\theta}{d\phi}\right)^2 .$$

3.5.1 Παραδείγματα

Παράδειγμα 3.2. *(Μέση τιμή κανονικής). Έστω $X_1, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές από την κατανομή $N(\theta, \sigma^2)$ με σ^2 γνωστό.*

Τότε,

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

Έτσι,

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

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

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

Επομένως,

$$I(\theta) = -E\left[\frac{d^2\ell(\theta)}{d\theta^2}\right] = -E\left(-\frac{n}{\sigma^2}\right) = \frac{n}{\sigma^2} \propto 1.$$

Εφόσον σ^2 γνωστό και σταθερό, συμπεραίνουμε ότι $J(\theta) \propto \sqrt{I(\theta)} \propto 1$ για $\theta \in \mathbb{R}$, δηλαδή η prior του Jeffreys για την μέση τιμή της κανονικής κατανομής ταυτίζεται με την ακατάλληλη prior στην οποία καταλήξαμε στην προηγούμενη παράγραφο και είναι ομοιόμορφη σε όλη την πραγματική ευθεία.

Παρατηρούμε ότι εδώ δουλέψαμε με την πιθανοφάνεια όλων των δεδομένων. Ωστόσο, θα μπορούσαμε να είχαμε δουλέψει με την πιθανοφάνεια από *μία* μόνο παρατήρηση x και να χρησιμοποιήσουμε την ιδιότητα $I_n(\theta) = nI_1(\theta)$, λόγω ανεξαρτησίας, όπου $I_1(\theta)$ και $I_n(\theta)$

είναι η πληροφορία από 1 παρατήρηση και από ολόκληρο το δείγμα αντίστοιχα. Επομένως, θα παίρναμε (όπως άλλωστε πρέπει) την ίδια prior του Jeffreys ανεξαρτήτως του πόσες παρατηρήσεις συλλέγαμε στη συνέχεια.

Παράδειγμα 3.3. *(Διωνυμικό δείγμα). Έστω ότι $X \mid \theta \sim Bin(n, \theta)$.*

Τότε,

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \Rightarrow$$

$$\ell(\theta) = \log f(x \mid \theta) = x \log(\theta) + (n-x)\log(1-\theta) + c.$$

Επομένως,

$$\begin{aligned} \frac{\ell(\theta)}{d\theta} &= \frac{x}{\theta} - \frac{n-x}{1-\theta} \Rightarrow \\ \frac{d^2\ell(\theta)}{d\theta^2} &= -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}. \end{aligned}$$

Εφόσον $E(X) = n\theta$, παίρνουμε:

$$\begin{aligned} I(\theta) &= -E\left[\frac{d^2\ell(\theta)}{d\theta^2}\right] = \frac{E(X)}{n\theta} + \frac{n - E(X)}{(1-\theta)^2} = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2} \\ &= \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)} = n\theta^{-1}(1-\theta)^{-1} \propto \theta^{-1}(1-\theta)^{-1}, \end{aligned}$$

το οποίο οδηγεί στην $J(\theta) \propto \sqrt{I(\theta)} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$, η οποία σε αυτήν την περίπτωση είναι η ορθή κατανομή Beta $\left(\frac{1}{2}, \frac{1}{2}\right)$.

3.6 Εξαγωγή της prior κατανομής

Δεν υπάρχουν εύκολες απαντήσεις στο ερώτημα πώς εξάγουμε με τον καλύτερο δυνατό τρόπο prior πληροφορία, αλλά ο λόγος για τον οποίο το αναφέρουμε είναι ότι πρόκειται για ένα σημαντικό ζήτημα. Πρέπει να θυμόμαστε όμως ότι κανένας δεν μπορεί να πάρει άπειρες (ή ακόμα και πολλές) ακριβείς πιθανοτικές αποφάσεις. Έτσι, ο προσδιορισμός μίας prior κατανομής πρέπει να θεωρείται ως μία προσπάθεια συμφιλίωσης των πεποιθήσεων που έχει ένα άτομο σε μία ενιαία μορφή. Όπως έχουμε δει, μία προσέγγιση είναι να πάρουμε μία συγκεκριμένη οικογένεια κατανομών (τη συζυγή οικογένεια για παράδειγμα), να ζητήσουμε prior πληροφορία για κάποια περιγραφικά χαρακτηριστικά, όπως η prior μέση τιμή και η διασπορά, και μετά να διαλέξουμε ως prior την κατανομή μέλος της συζυγούς οικογένειας με αυτά τα συγκεκριμένα χαρακτηριστικά. Γενικά όμως, μπορεί να είναι υπερβολικά δύσκολο να συμφιλιώσουμε τις prior πεποιθήσεις ενός ειδικού (ή ακόμα παραπάνω πολλών ειδικών) σε μία prior κατανομή.

3.7 Ασκήσεις

Άσκηση 3.1. *Ποιες από τις παρακάτω πυκνότητες ανήκουν στην εκθετική οικογένεια κατανομών (εξηγήστε τον τρόπο λειτουργίας σας);*

$$f_1(x \mid \theta) = \theta 2^{\theta} x^{-(\theta+1)}, \quad x > 2,$$

$$f_2(x \mid \theta) = \theta x^{\theta-1} e^{-x^{\theta}}, \quad x > 0.$$

Σε κάθε περίπτωση, υπολογίστε τη συζυγή prior αν η πυκνότητα ανήκει στην εκθετική οικογένεια κατανομών.

Άσκηση 3.2. *Να βρεθεί η prior του Jeffreys για το θ στο γεωμετρικό μοντέλο:*

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

Άσκηση 3.3. *Έστω ότι η X ακολουθεί την κατανομή Pareto(a, b), όπου το a είναι γνωστό, αλλά το $b > 0$ είναι άγνωστο. Δηλαδή,*

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

Να βρεθεί η prior του Jeffreys και η αντίστοιχη posterior κατανομή για το b.

Άσκηση 3.4. *Παίρνουμε μία παρατήρηση x από τη συνάρτηση πυκνότητας πιθανότητας $f(x \mid \theta)$.*

i. *Να ορίσετε την prior $J_{\Theta}(\theta)$ του Jeffreys για την παράμετρο θ σε αυτό το πλαίσιο. Να διατυπώσετε την ιδιότητα του αναλλοίωτου αυτής της prior κατανομής ως προς έναν μετασχηματισμό παραμέτρου $\phi = \phi(\theta)$.*

ii. *Έστω $f(x \mid \theta) = \theta e^{-\theta x}$ για $\theta > 0$. Να βρείτε την prior $J_{\Theta}(\theta)$ του Jeffreys για αυτό το μοντέλο.*

iii. *Ως εκ τούτου, εξάγετε την prior $J_{\Phi}(\phi)$ του Jeffreys για αυτό το μοντέλο όταν $\theta = e^{\phi}$.*

Άσκηση 3.5. *Ενδιαφερόμαστε να εκτιμήσουμε την πιθανότητα θ μία πινέζα να προσγειωθεί με τη μύτη προς τα πάνω. Η prior πεποίθησή μας μπορεί να περιγραφεί μέσω μίας μίξης κατανομών Βήτα:*

$$f(\theta) = \frac{\Gamma(a+b)}{2\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} + \frac{\Gamma(p+q)}{2\Gamma(p)\Gamma(q)}\theta^{p-1}(1-\theta)^{q-1}.$$

Ρίχνουμε μία πινέζα n το πλήθος ανεξάρτητες φορές και παρατηρούμε x περιπτώσεις όπου η καρφίτσα προσγειώνεται με τη μύτη προς τα πάνω. Να υπολογίσετε την posterior κατανομή για το θ.

Κεφάλαιο 4

Πολυπαραμετρικά Προβλήματα

4.1 Εισαγωγή

Όλα τα παραδείγματα που είδαμε μέχρι τώρα εμπλέκουν μόνο μία παράμετρο - συνήθως τη μέση τιμή ή τη διασπορά ενός πληθυσμού. Τα περισσότερα στατιστικά προβλήματα εμπλέκουν ένα στατιστικό μοντέλο που περιέχει περισσότερες από μία άγνωστες παραμέτρους. Μπορεί να υπάρχει μόνο μία παράμετρος ιδιαίτερου ενδιαφέροντος, αλλά συνήθως θα υπάρχουν και άλλες παράμετροι των οποίων οι τιμές είναι άγνωστες.

Η μέθοδος ανάλυσης πολυπαραμετρικών προβλημάτων στην Μπεϋζιανή στατιστική είναι πολύ πιο άμεση (τουλάχιστον στη θεωρία) σε σχέση με το αντίστοιχο πεδίο της κλασικής στατιστικής. Πράγματι, δεν υπάρχει απολύτως καμία καινούρια απαιτούμενη θεωρία πέρα από αυτήν που έχουμε ήδη δει. Έχουμε τώρα ένα διάνυσμα $\theta = (\theta_1, \ldots, \theta_d)$ από παραμέτρους για τις οποίες θέλουμε να εξάγουμε συμπεράσματα. Προσδιορίζουμε μία prior (πολυδιάστατη) κατανομή $f(\theta)$ για το θ και τη συνδυάζουμε με μία πιθανοφάνεια $f(x \mid \theta)$ μέσω του θεωρήματος του Bayes, για να πάρουμε:

$$f(\theta \mid x) = \frac{f(\theta) \cdot f(x \mid \theta)}{\int_\Theta f(\theta) \cdot f(x \mid \theta) d\theta}, \quad \theta \in \Theta,$$

όπως πριν. Φυσικά, η posterior κατανομή θα είναι τώρα κι αυτή μία *πολυδιάστατη* κατανομή. Ωστόσο, η απλότητα της Μπεϋζιανής προσέγγισης σημαίνει τώρα ότι η συμπερασματολογία για οποιοδήποτε υποσύνολο παραμέτρων του θ γίνεται με απλούς πιθανοτικούς υπολογισμούς πάνω σε αυτήν την από κοινού κατανομή.

Η *δεσμευμένη posterior κατανομή* μίας συνιστώσας του θ, έστω θ_i, δεδομένων των τιμών των εναπομειναςών συνιστωσών θ_{-i}, δίνεται ως:

$$f_i(\theta_i \mid x, \theta_{-i}) \propto f(\theta \mid x),$$

όπου οι τιμές του θ_i κρατούνται σταθερές. Δηλαδή, η δεσμευμένη posterior κατανομή του θ_i δίνεται από την από κοινού posterior κατανομή $f(\theta \mid x)$ του θ αν τη θεωρήσουμε ως μία συνάρτηση μόνο του θ_i, κρατώντας τις υπόλοιπες συνιστώσες θ_{-i} του θ σταθερές, κανονικοποιημένη ώστε να είναι μία συνάρτηση πυκνότητας πιθανότητας όπως είναι ορθό.

Ωστόσο, *ακριβής Μπεϋζιανή συμπερασματολογία* για την παράμετρο ενδιαφέροντος θ_i μπορεί μόνο να γίνει μόνο από την posterior κατανομή, ολοκληρώνοντάς την ως προς θ_{-i}:

$$f(\theta_i \mid x) = \int f(\theta \mid x) d\theta_{-i}.$$

Αυτή η *περιθώρια posterior* μίας δεδομένης παραμέτρου ενδιαφέροντος θ_i, η οποία προκύπτει έχοντας εξαλείψει τις οχληρές παραμέτρους θ_{-i} μέσω ολοκλήρωσης, μπορεί να χρησιμοποιηθεί για να βγάλουμε συμπεράσματα για αυτήν την παράμετρο.

Αν αυτή η περιθωριοποίηση δεν είναι εφικτή, μία άλλη προσέγγιση που μπορεί να χρησιμοποιηθεί για την εξάλειψη των οχληρών παραμέτρων είναι να υπολογίσουμε την posterior κατανομή της παραμέτρου ενδιαφέροντος δεσμεύοντας στις εκτιμήσεις μέγιστης πιθανοφάνειας των άλλων συνιστωσών του διανύσματος των παραμέτρων. Αυτή η τεχνική, η οποία δεν είναι πλήρως Μπεϋζιανή, καλείται *εμπειρική Μπεϋζιανή* μέθοδος, ώστε να διαχωρίζεται από τις πλήρως Μπεϋζιανές μεθόδους συμπερασματολογίας.

Παρόλο που δεν απαιτείται νέα θεωρία στα πολυπαραμετρικά προβλήματα, η αύξηση της διάστασης δημιουργεί ορισμένα πρακτικά προβλήματα:

1. **Προσδιορισμός prior.** Οι prior είναι πολυδιάστατες κατανομές. Αυτό σημαίνει ότι η επιλογή prior πρέπει να αντανακλά prior πεποιθήσεις όχι μόνο για κάθε παράμετρο ξεχωριστά, αλλά και για την εξάρτηση μεταξύ διαφορετικών συνδυασμών παραμέτρων. Η επιλογή κατάλληλων οικογενειών prior κατανομών και η σύνοψη της prior πληροφορίας των ειδικών κατά αυτόν τον τρόπο είναι ουσιωδώς πιο πολύπλοκη.

2. **Υπολογισμός.** Ακόμα και σε μονοδιάστατα προβλήματα είδαμε το πλεονέκτημα της χρήσης συζυγών οικογενειών για την απλοποίηση της posterior ανάλυσης που προκύπτει από τη χρήση του θεωρήματος του Bayes. Στα πολυπαραμετρικά προβλήματα τα ολοκληρώματα γίνονται ακόμα πιο δύσκολα στον υπολογισμό. Το γεγονός αυτό καθιστά τη χρήση των συζυγών οικογενειών prior ακόμα πιο πολύτιμη και δημιουργεί την ανάγκη για αριθμητικές τεχνικές ώστε να εξάγουμε επιχειρήματα όταν οι συζυγείς οικογένειες είναι είτε μη-διαθέσιμες είτε ακατάλληλες.

3. **Ερμηνεία.** Όλη η posterior συμπερασματολογία περιέχεται στην posterior κατανομή, η οποία θα έχει τόση διάσταση όση και η μεταβλητή θ. Η δομή της posterior κατανομής ενδέχεται να είναι εξαιρετικά πολύπλοκη και να απαιτεί αξιόλογη δεξιοτεχνία (καθώς και έναν υπολογιστή με καλές γραφικές λειτουργίες) για να αναγνωρίσουμε τις πιο

σημαντικές σχέσεις που εμπεριέχει.

Παρόλα αυτά τα πρακτικά ζητήματα, είναι σημαντικό να τονίσουμε για ακόμα μία φορά ότι απαιτείται ακριβώς η ίδια θεωρία τόσο στα πολυπαραμετρικά όσο και στα μονοπαραμετρικά προβλήματα. Το Μπεύζιανό πλαίσιο σημαίνει ότι όλη η συμπερασματολογία προκύπτει από στοιχειώδεις κανόνες των πιθανοτήτων.

4.2 Παραδείγματα

Παράδειγμα 4.1. *Υποθέτουμε ότι ένα μηχάνημα είναι είτε ικανοποιητικό* ($x = 1$) *είτε μη-ικανοποιητικό* ($x = 2$). *Η πιθανότητα το μηχάνημα να είναι ικανοποιητικό εξαρτάται από τη θερμοκρασία του δωματίου* ($\theta_1 = 0$*: κρύο,* $\theta_1 = 1$*: ζεστό) και την υγρασία του δωματίου* ($\theta_2 = 0$*: ξηρό,* $\theta_2 = 1$*: υγρό). Οι πιθανότητες του* $x = 1$ *δίνονται στον πίνακα 4.1. Επιπροσθέτως, η από κοινού prior κατανομή του* (θ_1, θ_2) *δίνεται στον πίνακα 4.2.*

$P(X = 1 \mid \theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0.6	0.8
$\theta_2 = 1$	0.7	0.6

Πινακας 4.1: Δεσμευμένες πιθανότητες το μηχάνημα να είναι ικανοποιητικό

$f(\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0.3	0.2
$\theta_2 = 1$	0.2	0.3

Πινακας 4.2: Prior πιθανότητες των συνθηκών του δωματίου

Ο πίνακας 4.1 αναπαριστά την πιθανοφάνεια του $x = 1$, οπότε *δεν* είναι κατανομή πιθανότητας και *δεν* αθροίζει στη μονάδα. Αντιθέτως, ο πίνακας 4.2 αναπαριστά την prior του διανύσματος παραμέτρων (θ_1, θ_2), οπότε *είναι* κατανομή πιθανότητας και *αθροίζει* στη μονάδα.

		$\theta_1 = 0$	$\theta_1 = 1$
$f(X = 1, \theta_1, \theta_2)$	$\theta_2 = 0$	0.18	0.16
	$\theta_2 = 1$	0.14	0.18
$P(X = 1)$		0.66	
$f(\theta_1, \theta_2 \mid X = 1)$	$\theta_2 = 0$	$\frac{18}{66}$	$\frac{16}{66}$
	$\theta_2 = 1$	$\frac{14}{66}$	$\frac{18}{66}$

Πινακας 4.3: Posterior πιθανότητες των συνθηκών του δωματίου

Η από κοινού posterior κατανομή μπορεί να υπολογιστεί όπως στον πίνακα 4.3, όπου:

$$f(X = 1, \theta_1, \theta_2) = f(\theta_1, \theta_2) \cdot P(X = 1 \mid \theta_1, \theta_2),$$

$$P(X = 1) = \sum_{i=0}^{1} \sum_{j=0}^{1} P(X = 1, \theta_1 = i, \theta_2 = j) \quad \text{και}$$

$$f(\theta_1, \theta_2 \mid X = 1) = \frac{f(\theta_1, \theta_2) \cdot P(X = 1 \mid \theta_1, \theta_2)}{P(X = 1)} = \frac{f(X = 1, \theta_1, \theta_2)}{P(X = 1)}.$$

Έτσι, αθροίζοντας κατά γραμμές και κατά στήλες, παίρνουμε τις περιθώριες posterior κατανομές:

$$P(\theta_1 = 0 \mid X = 1) = \sum_{j=0}^{1} P(\theta_1 = 0, \theta_2 = j \mid x = 1) = \frac{32}{66},$$

$$P(\theta_1 = 1 \mid X = 1) = \sum_{j=0}^{1} P(\theta_1 = 1, \theta_2 = j \mid x = 1) = \frac{34}{66},$$

$$P(\theta_2 = 0 \mid X = 1) = \sum_{i=0}^{1} P(\theta_1 = i, \theta_2 = 0 \mid x = 1) = \frac{34}{66},$$

$$P(\theta_2 = 1 \mid X = 1) = \sum_{i=0}^{1} P(\theta_1 = i, \theta_2 = 1 \mid x = 1) = \frac{32}{66}.$$

Παράδειγμα 4.2. *Έστω ότι $Y_1 \sim$ Poisson$(a\beta)$ και $Y_2 \sim$ Poisson$((1-a)\beta)$ ανεξάρτητες δεδομένων των a και β. Τώρα, υποθέτουμε ότι η prior πληροφορία μας για τα a και β μπορεί να εκφραστεί ως: $a \sim$ Beta(p, q) και $\beta \sim$ Gamma$(p + q, 1)$, όπου τα a και β είναι ανεξάρτητα για συγκεκριμένες υπερπαραμέτρους p και q.*

Να σημειωθεί ότι είναι απαραίτητο να προσδιορίσουμε την από κοινού prior κατανομή των a και β. Αυτό γίνεται απλούστερο αν μπορούμε να υποθέσουμε prior ανεξαρτησία μεταξύ αυτών των παραμέτρων, αφού τότε αρκεί να προσδιορίσουμε μόνο τις περιθώριες κατανομές τους. Όμως, αυτό θα πρέπει να γίνει μόνο αν είναι μία ρεαλιστική υπόθεση. (Για παράδειγμα, μπορεί να έχουμε $\beta =$ αναμενόμενος αριθμός μοτοσυκλετιστικών ατυχημάτων και $a =$ αναμενόμενο ποσοστό αυτών που συνδέονται με την κατανάλωση αλκοόλ).

Έχουμε την ακόλουθη πιθανοφάνεια:

$$f(y_1, y_2 \mid a, \beta) = e^{-a\beta} \frac{(a\beta)^{y_1}}{y_1!} \cdot e^{-(1-a)\beta} \frac{[(1-a)\beta]^{y_2}}{y_2!} \propto a^{y_1}(1-a)^{y_2} \cdot \beta^{y_1+y_2} e^{-\beta},$$

και την prior

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

Επομένως, από το θεώρημα του Bayes:

$$\begin{aligned} f(a, \beta \mid y_1, y_2) &\propto a^{y_1}(1-a)^{y_2} \cdot \beta^{y_1+y_2} e^{-\beta} \cdot a^{p-1}(1-a)^{q-1} \cdot \beta^{p+q-1} e^{-\beta} \\ &= a^{p+y_1-1}(1-a)^{q+y_2-1} \cdot \beta^{p+q+y_1+y_2-1} e^{-2\beta}, \quad a \in (0,1) \quad \beta > 0. \end{aligned}$$

Αυτή είναι η (από κοινού) posterior κατανομή των a και β και περιλαμβάνει όλη την πληροφορία από την prior και τα δεδομένα. Στην προκειμένη περίπτωση, η posterior παραγοντοποιείται σε συναρτήσεις του a και του β. Δηλαδή, $f(a, \beta \mid y_1, y_2) \propto g(a)h(\beta)$, όπου:

$$g(a) = a^{p+y_1-1}(1-a)^{q+y_2-1} \quad \text{και} \quad h(\beta) = \beta^{p+q+y_1+y_2-1}e^{-2\beta}.$$

Έπεται λοιπόν ότι οι περιθώριες posterior κατανομές δίνονται από τις σχέσεις:

$$f(a \mid y_1, y_2) = \int_0^\infty f(a, \beta \mid y_1, y_2) d\beta \propto g(a) \qquad \text{και}$$

$$f(\beta \mid y_1, y_2) = \int_0^1 f(a, \beta \mid y_1, y_2) da \propto h(\beta).$$

Δηλαδή, $a \mid y_1, y_2 \sim \text{Beta}(p + y_1, q + y_2)$ και $\beta \mid y_1, y_2 \sim \text{Gamma}(p + q + y_1 + y_2, 2)$.

4.3 Ασκήσεις

Άσκηση 4.1. *Η ποιότητα ενός ηλεκτρικού εξαρτήματος είναι άριστη ($x = 1$), καλή ($x = 2$) ή κακή ($x = 3$). Η πιθανότητα των διαφόρων επιπέδων ποιότητας εξαρτάται από το εργοστάσιο παραγωγής ($\theta_1 = 0$: εργοστάσιο Α, $\theta_1 = 1$: εργοστάσιο Β) και τον τύπο του μηχανήματος ($\theta_2 = 0$: μηχάνημα Ι, $\theta_2 = 1$: μηχάνημα ΙΙ, $\theta_2 = 2$: μηχάνημα ΙΙΙ). Οι πιθανότητες του $x = 3$ δίνονται στον πίνακα 4.4. Η από κοινού prior κατανομή του (θ_1, θ_2) δίνεται στον πίνακα 4.5. Να βρεθεί η από κοινού posterior κατανομή του (θ_1, θ_2), δεδομένου ότι $x = 3$, και οι περιθώριες posterior κατανομές των θ_1 και θ_2. Έχοντας παρατηρήσει ότι $x = 3$, επηρέασε αυτό ποιος συνδυασμός εργοστασίου και μηχανήματος είναι πιθανότερο να έχει παραγάγει το εξάρτημα;*

$P(X = 3 \mid \theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0.2	0.3
$\theta_2 = 1$	0.4	0.1
$\theta_2 = 2$	0.5	0.2

Πινακας 4.4: Δεσμευμένες πιθανότητες του $x = 3$

$f(\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0.1	0.2
$\theta_2 = 1$	0.2	0.3
$\theta_2 = 2$	0.1	0.1

Πινακας 4.5: Prior πιθανότητες των συνδυασμών εργοστασίου και μηχανήματος

Άσκηση 4.2. *Η τυχαία μεταβλητή X παίρνει είτε την τιμή 0 είτε την τιμή 1 με πιθανότητες που εξαρτώνται από τις παραμέτρους a και β:*

$$P(X = 0 \mid a, \beta) = \begin{cases} 0.2, & a = a_1, \ \beta = \beta_1 \\ 0.8, & a = a_1, \ \beta = \beta_2 \\ 0.9, & a = a_2, \ \beta = \beta_1 \\ 0.3, & a = a_2, \ \beta = \beta_2 \end{cases}.$$

Η από κοινού prior είναι:

$$f(a,\beta) = \begin{cases} 0.1, & a = a_1,\ \beta = \beta_1 \\ 0.5, & a = a_1,\ \beta = \beta_2 \\ 0.3, & a = a_2,\ \beta = \beta_1 \\ 0.1, & a = a_2,\ \beta = \beta_2 \end{cases}.$$

Έστω ότι παίρνουμε δύο ανεξάρτητες τιμές της X, καθεμία από τις οποίες τυχαίνει να είναι 0.

i. *Γράψτε τον πίνακα της πιθανοφάνειας.*

ii. *Βρείτε την από κοινού posterior κατανομή.*

iii. *Βρείτε την περιθώρια posterior κατανομή για το a.*

Άσκηση 4.3. *Παίρνουμε παρατηρήσεις x_1 και x_2 από τις τυχαίες μεταβλητές X_1 και X_2 που ακολουθούν κατανομές Poisson με μέσες τιμές θ και $\phi\theta$ αντίστοιχα, όπου ϕ είναι μία γνωστή θετική σταθερά.*

i. *Να δείξετε, υπολογίζοντας την posterior πυκνότητα του θ, ότι η οικογένεια των Γάμμα κατανομών είναι συζυγής για αυτό το μοντέλο. Ποια είναι η posterior μέση τιμή του θ στην περίπτωση όπου $p = 1$ και $q = 1$;*

ii. *Τώρα, υποθέτουμε ότι το ϕ είναι επίσης άγνωστη παράμετρος, ανεξάρτητη από το θ και με prior πυκνότητα $f(\phi) = \frac{1}{(1+\phi)^2}$. Να βρείτε την από κοινού posterior κατανομή των θ και ϕ και να δείξετε ότι η περιθώρια posterior κατανομή του ϕ είναι ανάλογη με τη συνάρτηση:*

$$\frac{\phi^{x_2}}{(1+\phi)^2(1+\phi+q)^{x_1+x_2+p}}.$$

Άσκηση 4.4. *Έστω $x_1, x_2, \ldots, x_n$ ένα σύνολο ανεξάρτητων παρατηρήσεων από την κανονική κατανομή με μέση τιμή θ και διασπορά ϕ^{-1}. Να δείξετε ότι η πιθανοφάνεια των παρατηρήσεων $x_1, x_2, \ldots, x_n$ είναι:*

$$L(\theta,\phi) \propto \phi^{\frac{n}{2}} \exp\left\{-\frac{\phi}{2}\left[s^2 + n(\overline{x}-\theta)^2\right]\right\}, \quad \text{όπου :}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{και} \quad s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Υποθέτουμε μία Γάμμα prior για το ϕ με παραμέτρους p και q και μία κανονική prior για το θ με μέση τιμή 0 και διασπορά ϕ^{-1}. Να βρείτε την από κοινού posterior για τα θ και ϕ (κατά αναλογία) - ποια είναι η περιθώρια posterior για το ϕ και ποια η δεσμευμένη posterior για το θ δεδομένου του ϕ;

Κεφάλαιο 5

Σύνοψη της Posterior Πληροφορίας

5.1 Θεωρία αποφάσεων

Αυτή είναι από μόνη της μία εξαιρετικά σημαντική περιοχή. Μέσα στο μάθημα θα αγγίξουμε μόνο τα σημαντικά ζητήματα.

Πολλά προβλήματα στον πραγματικό κόσμο είναι αυτά της λήψης αποφάσεων ενώπιον της αβεβαιότητας: "ποιο πολιτικό κόμμα είναι το καλύτερο να ψηφίσω;", "να δεχτώ μία προσφορά εργασίας ή να περιμένω με την ελπίδα ότι θα μου προσφερθεί μία καλύτερη δουλειά;". Όλη η στατιστική συμπερασματολογία μπορεί να θεωρηθεί ως λήψη αποφάσεων: έχοντας παρατηρήσει ένα συγκεκριμένο σύνολο δεδομένων, με ποια τιμή πρέπει να αποφασίσουμε να εκτιμήσουμε μία παράμετρο; Υπάρχουν πολλές προσεγγίσεις στη θεωρία αποφάσεων, αλλά μακράν μεγαλύτερη συνοχή έχει η προσέγγιση που βασίζεται στην Μπεϋζιανή ανάλυση. Πράγματι, μπορεί να δειχθεί ότι αν ακολουθούνται ορισμένα αξιώματα (δηλαδή αν ακολουθούνται ένα πλήθος από λογικούς κανόνες αποφάσεων), τότε η Μπεϋζιανή ανάλυση είναι η *μόνη* λογική προσέγγιση στη λήψη αποφάσεων. Αυτό χρησιμοποιείται συχνά ως επιχείρημα για να δικαιολογήσει την προτίμηση της Μπεϋζιανής συμπερασματολογίας έναντι της κλασικής.

Τα στοιχεία τα οποία απαιτούνται για την κατασκευή ενός προβλήματος αποφάσεων είναι τα εξής:

1. Ένας *παραμετρικός χώρος* Θ που περιέχει τις πιθανές καταστάσεις του προβλήματος.

2. Ένα σύνολο A των *πράξεων* που είναι διαθέσιμες στον υπεύθυνο για τη λήψη των αποφάσεων.

3. Μία *συνάρτηση απώλειας* L, όπου $L(\theta, a)$ είναι η απώλεια επιβάλλεται υιοθετώντας την πράξη a όταν η πραγματική κατάσταση του προβλήματος είναι η θ.

Θα δούμε αυτά τα στοιχεία στα πλαίσια ενός συγκεκριμένου παραδείγματος.

Παράδειγμα 5.1. *(Πρώτο μέρος). Ένας υπεύθυνος δημόσιας υγείας αναζητά μία λογική πολιτική εμβολιασμού ενάντια σε μία σχετικά ήπια ασθένεια που προκαλεί απουσίες από τον εργασιακό χώρο. Έρευνες υποδηλώνουν ότι το 60% του πληθυσμού έχει ήδη ανοσία. Εκτιμάται ότι το χρηματικό ισοδύναμο των εργατοωρών που χάνονται αν δεν εμβολιαστεί ένα ευάλωτο άτομο είναι 20 μονάδες, ότι το περιττό κόστος εμβολιασμού ενός ατόμου που έχει ανοσία είναι 8 μονάδες και ότι δεν επιβάλλεται κάποιο κόστος από τον εμβολιασμό ενός ευάλωτου ατόμου ή τον μη-εμβολιασμό ενός ατόμου που έχει ανοσία.*

Έτσι, για τη συγκεκριμένη περίπτωση έχουμε:

1. Τον παραμετρικό χώρο $\Theta = \{\theta_1, \theta_2\}$, όπου τα θ_1 και θ_2 αντιστοιχούν στο άτομο να έχει ανοσία και να είναι ευάλωτο αντίστοιχα.

2. Το σύνολο των πράξεων $A = \{a_1, a_2\}$, όπου τα a_1 και a_2 αντιστοιχούν στον εμβολιασμό και τον μη-εμβολιασμό του ατόμου αντίστοιχα.

3. Η συνάρτηση απώλειας ορίζεται στον πίνακα 5.1:

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

Πινακας 5.1: Συνάρτηση απώλειας

Η στρατηγική λήψης της απόφασης είναι τότε να αξιολογήσουμε τη μέση απώλεια για κάθε πράξη και να διαλέξουμε την πράξη που έχει το ελάχιστο δυνατό μέσο κόστος. Αυτό θα είναι το ίδιο για κάθε άτομο σε αυτήν την περίπτωση. Ο πίνακας 5.2 δείχνει τον υπολογισμό της μέσης απώλειας για κάθε πράξη, βασισμένο πάνω στην prior κατανομή του θ. Το συμπέρασμα είναι ότι συμφέρει να εμβολιάσουμε όλα τα άτομα. Το μέσο κόστος (ή η μέση απώλεια) είναι 4.8 μονάδες ανά άτομο.

$f(\theta)$	0.6	0.4	
$L(\theta, a)$	θ_1	θ_2	$E_\theta\left[L(\theta, a)\right]$
a_1	8	0	$0.6 \cdot 8 + 0.4 \cdot 0 = 4.8$
a_2	0	20	$0.6 \cdot 0 + 0.4 \cdot 20 = 8$

Πινακας 5.2: Μέση απώλεια

Η υπόθεση της μηδενικής απώλειας στην περίπτωση της ”σωστής” απόφασης δεν είναι περιοριστική. Μπορούμε να προσθέσουμε ή να αφαιρέσουμε μία ποσότητα στην πρώτη στήλη της συνάρτησης απώλειας και μία άλλη ποσότητα στη δεύτερη στήλη χωρίς να επηρεάσουμε την καλύτερη απόφαση. Έτσι, αν η συνάρτηση απώλειας δεν είχε μηδενικά, θα μπορούσαμε να την προσαρμόσουμε, ώστε η μικρότερη τιμή σε κάθε στήλη να είναι μηδέν. Το μόνο που έχει σημασία είναι το σχετικό κόστος των πράξεων για κάθε "κατάσταση του προβλήματος" θ.

Υποθέτουμε τώρα ότι έχουμε περαιτέρω διαθέσιμη πληροφορία, δηλαδή δεδομένα x, η οποία αντικατοπτρίζει στην τιμή του θ. Για να είμαστε ακριβείς, υποθέτουμε ότι έχουμε παρατηρήσει δεδομένα x από την $f(x \mid \theta)$. Τότε, μπορούμε να αντικαταστήσουμε την $f(\theta)$ από την posterior $f(\theta \mid x)$ στον υπολογισμό της μέσης απώλειας. Η καλύτερη απόφαση θα εξαρτάται τότε από τη συγκεκριμένη έκβαση x.

Παράδειγμα 5.2. *(Συνέχεια). Μία απλή δερματική εξέταση έχει αναπτυχθεί, η οποία, αν και όχι πλήρως αξιόπιστη, τείνει να υποδεικνύει την ευπάθεια του ατόμου ως προς την ασθένεια. Οι πιθανότητες αντίδρασης συνοψίζονται στον πίνακα 5.3.*

		Ανοσία (θ_1)	Ευπάθεια (θ_2)
Αντίδραση	Αμελητέα (x_1)	0.35	0.09
	Ήπια (x_2)	0.3	0.17
	Μέτρια (x_3)	0.21	0.25
	Ισχυρή (x_4)	0.14	0.49

Πινακας 5.3: Πιθανότητες αντίδρασης δεδομένης της ευπάθειας του ατόμου

Η γενική διαδικασία είναι να χρησιμοποιήσουμε το θεώρημα του Bayes για να μας δώσει την posterior κατανομή $f(\theta \mid x)$. Τότε, για μία οποιαδήποτε πράξη a, η *posterior μέση απώλεια* είναι:

$$\rho(a, x) = E_\theta \left[L(\theta, a) \mid x\right] = \int_\Theta f(\theta \mid x) L(\theta, a) d\theta.$$

Έχοντας παρατηρήσει μία συγκεκριμένη τιμή x, επιλέγουμε την πράξη a που οδηγεί στην ελάχιστη τιμή του ρ. Γράφοντας $d(x) = \underset{a \in A}{\operatorname{argmin}}\, \rho(a, x)$, καλούμε το $d(x)$ *κανόνα απόφασης κατά Bayes*.

Για το παράδειγμά μας, ο πίνακας 5.4 δείχνει τον τρόπο υπολογισμού του κανόνα απόφασης. Θεωρούμε όλες τις πιθανές εκβάσεις x, υπολογίζοντας για καθεμία από αυτές την αντίστοιχη posterior $f(\theta \mid x)$. Για καθεμία από αυτές, υπολογίζουμε μετά τη μέση posterior απώλεια για κάθε πράξη. Τέλος, επιλέγουμε την καλύτερη απόφαση, δηλαδή αυτή με την ελάχιστη μέση posterior απώλεια για αυτήν την έκβαση.

Οπότε, συνοψίζοντας, αν παρατηρηθεί μία αμελητέα ή μία ήπια αντίδραση σε ένα άτομο, τότε η απόφαση κατά Bayes είναι να μην εμβολιάσουμε το άτομο, ενώ αν παρατηρηθεί μία μέτρια ή μία ισχυρή αντίδραση, η απόφαση είναι να το εμβολιάσουμε.

Μπορούμε να προχωρήσουμε ένα βήμα παραπέρα και να υπολογίσουμε το *ρίσκο* που συνδέεται με αυτήν την πολιτική, σταθμίζοντας σύμφωνα με την αβεβαιότητα των παρα-

	θ_1	θ_2				
$f(\theta)$	0.6	0.4				
$f(x_1 \mid \theta)$	0.35	0.09				
$f(x_2 \mid \theta)$	0.3	0.17				
$f(x_3 \mid \theta)$	0.21	0.25				
$f(x_4 \mid \theta)$	0.14	0.49				
			$f(x)$			
$f(x_1, \theta)$	0.21	0.036	0.246			
$f(x_2, \theta)$	0.18	0.068	0.248			
$f(x_3, \theta)$	0.126	0.1	0.226			
$f(x_4, \theta)$	0.084	0.196	0.28			
			$\rho(a, x)$			
			a_1	a_2	$d(x)$	$\rho(d(x), x)$
$f(\theta \mid x_1)$	0.854	0.146	6.829	2.927	a_2	2.927
$f(\theta \mid x_2)$	0.726	0.274	5.806	5.484	a_2	5.484
$f(\theta \mid x_3)$	0.558	0.442	4.46	8.847	a_1	4.46
$f(\theta \mid x_4)$	0.3	0.7	2.4	14	a_1	2.4

Πινακας 5.4: Πινακοποίηση της Μπεϋζιανής ανάλυσης αποφάσεων

τηρήσεων x. Έτσι, ορίζουμε το ρίσκο κατά Bayes ως:

$$\mathrm{BR}(d) = \int \rho(d(x), x) f(x) dx.$$

Για το παράδειγμά μας, αυτό γίνεται:

$$\mathrm{BR}(d) = \sum_{i=1}^{4} \rho(d(x_i), x_i) f(x_i) = 2.927 \cdot 0.246 + 5.484 \cdot 0.248 + 4.46 \cdot 0.226 + 2.4 \cdot 0.28 = 3.76.$$

Αυτό είναι μικρότερο από το ελάχιστο μέσο κόστος των 4.8 μονάδων ανά άτομο που πήραμε χρησιμοποιώντας μόνο την prior πληροφορία, χωρίς γνώση του x. Αυτό πρέπει να ισχύει πάντα. Αυτό δε σημαίνει ότι η εφαρμογή της δερματικής εξέτασης αξίζει πάντα, επειδή υπάρχει συνήθως ένα σταθερό κόστος ανά άτομο συνδεδεμένο με την εξέταση. Σε αυτή την περίπτωση, υπάρχει ένα καθαρό κέρδος να πραγματοποιήσουμε τη δερματική εξέταση μόνο αν κοστίζει σε μονάδες ανά άτομο λιγότερο από τη διαφορά $4.8 - 3.76 = 1.04$.

5.2 Σημειακή εκτίμηση

Έχουμε τονίσει επανειλημμένα ότι η posterior κατανομή είναι μία πλήρης σύνοψη της συμπερασματολογίας για μία παράμετρο θ. Στην ουσία, η posterior κατανομή *είναι* η συμπερασματολογία. Ωστόσο, για κάποιες εφαρμογές είναι επιθυμητό (ή και απαραίτητο) να συνοψίσουμε περαιτέρω αυτήν την πληροφορία με κάποιον τρόπο. Συγκεκριμένα, μπορεί να θέλουμε να δώσουμε μία μόνο ”βέλτιστη” εκτίμηση της άγνωστης παραμέτρου. (Σημειώνουμε τη διάκριση με την κλασική στατιστική όπου οι σημειακές εκτιμήσεις των παραμέτρων είναι το φυσικό επακόλουθο μίας συμπερασματολογίας και το σημείο που είναι πιο προβληματικό είναι η αντανάκλαση της αβεβαιότητας για αυτήν την εκτίμηση.)

Επομένως, στο Μπεϋζιανό πλαίσιο, πώς συνοψίζουμε την πληροφορία σε μία posterior κατανομή ώστε να δώσουμε μία μόνο ”βέλτιστη” εκτίμηση; Η απάντηση εξαρτάται από το τι εννοούμε με τον όρο ”βέλτιστη” και αυτό με τη σειρά του προσδιορίζεται μετατρέποντας το πρόβλημα σε ένα πρόβλημα αποφάσεων. Με άλλα λόγια, προσδιορίζουμε μία συνάρτηση απώλειας $L(\theta, a)$ που εκφράζει την εκλαμβανόμενη ποινή για την εκτίμηση του θ από το a. Υπάρχει ένα εύρος από φυσικές συναρτήσεις απώλειας που θα μπορούσαμε να χρησιμοποιήσουμε και η συγκεκριμένη επιλογή για οποιοδήποτε πρόβλημα θα εξαρτάται από το γενικό πλαίσιο. Οι πιο συχνά χρησιμοποιούμενες είναι:

1. Τετραγωνική ή δευτεροβάθμια συνάρτηση απώλειας: $L(\theta, a) = (\theta - a)^2$.

2. Συνάρτηση απώλειας απολύτου σφάλματος: $L(\theta, a) = |\theta - a|$.

3. Συνάρτηση απώλειας 0-1:

$$L(\theta, a) = \begin{cases} 0, & |\theta - a| \leqslant \varepsilon \\ 1, & |\theta - a| > \varepsilon \end{cases}.$$

Σε καθεμία από αυτές τις περιπτώσεις, ελαχιστοποιώντας την posterior μέση απώλεια, παίρνουμε απλές μορφές για τον κανόνα απόφασης κατά Bayes, ο οποίος λαμβάνεται ως η σημειακή εκτίμηση του θ για αυτήν τη συγκεκριμένη επιλογή συνάρτησης απώλειας.

5.2.1 Τετραγωνική συνάρτηση απώλειας

Σε αυτήν την περίπτωση, μπορούμε να απλοποιήσουμε το $\rho(a, x) = E\left[(\theta - a)^2 \mid x\right]$ προσθαφαιρώντας την posterior μέση τιμή $\mu = E(\theta \mid x)$ και αναπτύσσοντας το τετράγωνο:

$$\begin{aligned} E\left[(\theta - a)^2 \mid x\right] &= E\left[(\theta - \mu + \mu - a)^2 \mid x\right] \\ &= E\left[(\theta - \mu)^2 + (\mu - a)^2 + 2(\theta - \mu)(\mu - a) \mid x\right] \\ &= E\left[(\theta - E(\theta \mid x))^2 \mid x\right] + (\mu - a)^2 + 2(\mu - a)\cancelto{0}{E\left[\theta - E(\theta \mid x) \mid x\right]} \\ &= \mathrm{Var}(\theta \mid x) + \left[E(\theta \mid x) - a\right]^2 . \end{aligned}$$

Ο όρος $\mathrm{Var}(\theta \mid x)$ δεν εξαρτάται πλέον από το a και ο όρος $[E(\theta \mid x) - a]^2$ επιτυγχάνει την ελάχιστη τιμή 0 παίρνοντας $a = E(\theta \mid x)$. Συνοψίζοντας, η posterior μέση τετραγωνική συνάρτηση απώλειας παίρνει την ελάχιστη τιμή $\mathrm{Var}(\theta \mid x)$ όταν το a ισούται με την posterior μέση τιμή του θ.

5.2.2 Συνάρτηση απώλειας απολύτου σφάλματος

Δείχνουμε ότι σε αυτήν την περίπτωση η ελάχιστη posterior μέση απώλεια επιτυγχάνεται παίρνοντας $a = m$, όπου m η διάμεσος της posterior κατανομής $f(\theta \mid x)$. Υποθέτουμε ότι αυτή είναι μοναδική και ορίζεται ως $P(\theta < m \mid x) = P(\theta > m \mid x) = \frac{1}{2}$. Για να αποδείξουμε το αποτέλεσμα, παρατηρούμε πρώτα ότι η συνάρτηση $s(\theta) = -I_{\{\theta<m\}} + I_{\{\theta>m\}}$ έχει την ιδιότητα:

$$E\left[s(\theta) \mid x\right] = -P(\theta < m \mid x) + P(\theta > m \mid x) = -\frac{1}{2} + \frac{1}{2} = 0.$$

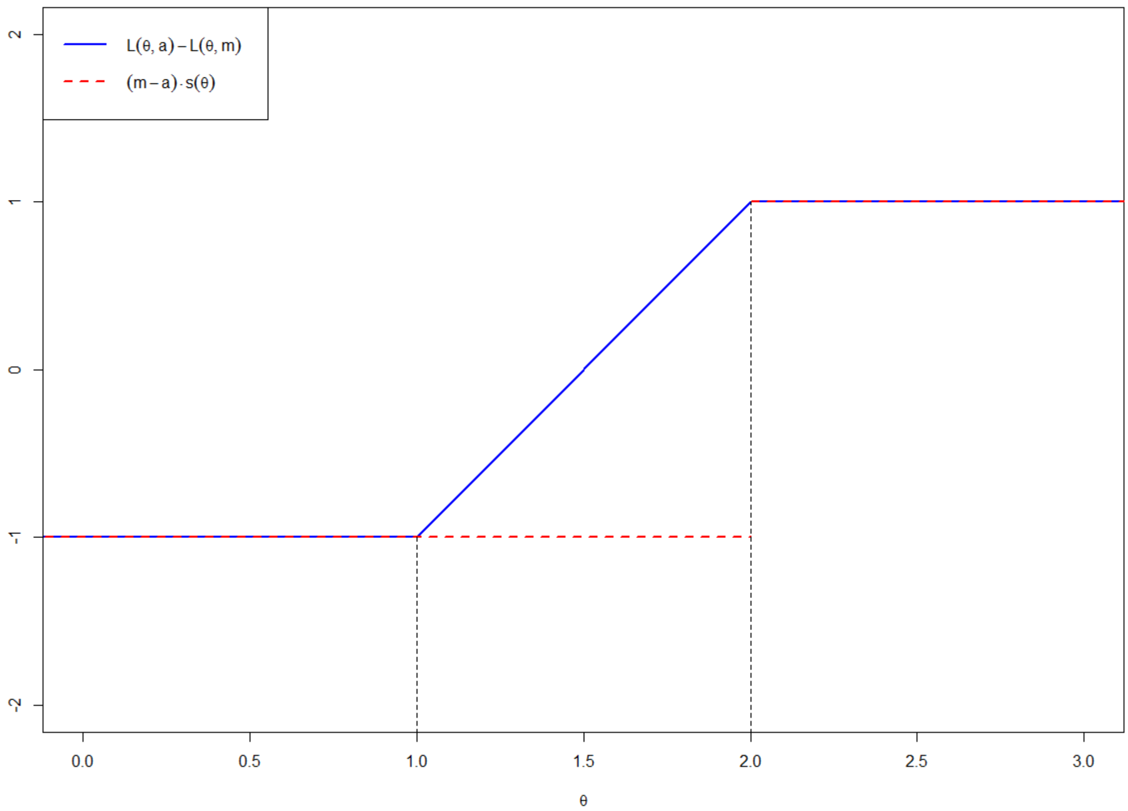
Τώρα, θεωρούμε τη διαφορά $L(\theta, a)-L(\theta, m) = |\theta-a|-|\theta-m|$ για κάποιο $a < m$. Διακρίνουμε τις εξής περιπτώσεις για το θ:

i. Αν $\theta < a < m$, τότε $L(\theta, a) - L(\theta, m) = -(\theta - a) + (\theta - m) = -(m - a) = (m - a)s(\theta)$.

ii. Αν $a < \theta < m$, τότε:

$$L(\theta, a) - L(\theta, m) = (\theta - a) + (\theta - m) = 2\theta - a - m > 2a - a - m = (m - a)s(\theta).$$

iii. Αν $a < m < \theta$, τότε $L(\theta, a) - L(\theta, m) = (\theta - a) - (\theta - m) = m - a = (m - a)s(\theta)$.



Σχημα 5.1: Γράφημα των $L(\theta, a) - L(\theta, m)$ και $(m - a)s(\theta)$ για $m = 2$ και $a = 1$

Μπορούμε να δούμε ότι $E\left[L(\theta, a) - L(\theta, m) \mid x\right] > (m - a)E\left[s(\theta) \mid x\right] = 0$. Επομένως, ισχύει ότι $E[L(\theta,a) \mid x] > E[L(\theta,m) \mid x]$. Με παρόμοιο τρόπο, μπορούμε να δείξουμε ότι ισχύει το ίδιο για $a > m$, οπότε η συνάρτηση $E[L(\theta,a) \mid x]$ ελαχιστοποιείται όταν $a = m$.

5.2.3 Συνάρτηση απώλειας 0-1

Προφανώς, σε αυτήν την περίπτωση, $P(|\theta - a| > \varepsilon \mid x) = 1 - P(|\theta - a| \leqslant \varepsilon \mid x)$. Συνεπώς, η εκτίμηση κατά Bayes είναι το μέσο του διαστήματος $[\theta - \varepsilon, \theta + \varepsilon]$ με τη μεγαλύτερη posterior πιθανότητα. Επιλέγοντας ε αυθαίρετα μικρό, αυτή η διαδικασία θα οδηγήσει στην posterior κορυφή ως την εκτίμηση κατά Bayes κάτω από αυτήν τη συγκεκριμένη συνάρτηση απώλειας.

5.2.4 Συμπέρασμα

Συμπερασματικά, το σημαντικό σημείο είναι ότι στο Μπεϋζιανό πλαίσιο μία σημειακή εκτίμηση μίας παραμέτρου είναι ένα μόνο συνοπτικό στατιστικό στοιχείο της posterior κατανομής. Ορίζοντας την ποιότητα μίας εκτιμήτριας μέσω μίας συνάρτησης απώλειας, η μεθοδολογία της θεωρίας αποφάσεων οδηγεί σε βέλτιστες επιλογές σημειακών εκτιμήσεων. Συγκεκριμένα, οι πιο φυσικές επιλογές συναρτήσεων απώλειας οδηγούν αντίστοιχα στην posterior μέση τιμή, στην posterior διάμεσο και στην posterior κορυφή ως βέλτιστες σημειακές εκτιμήτριες.

Παράδειγμα 5.3. *Αν η posterior κατανομή για το θ είναι η $f(\theta \mid x) = 1$ για $\theta \in (0,1)$, θα υπολογίσουμε τη βέλτιστη εκτιμήτρια του $\phi = \theta^2$ σύμφωνα με την τετραγωνική συνάρτηση απώλειας.*

Η βέλτιστη εκτιμήτρια του ϕ σύμφωνα με την τετραγωνική συνάρτηση απώλειας είναι η posterior μέση τιμή του ϕ, δηλαδή:

$$\widehat{\phi} = E(\phi \mid x) = E\left(\theta^2 \mid x\right) = \int_0^1 \theta^2 f(\theta \mid x) d\theta = \int_0^1 \theta^2 d\theta = \frac{1}{3}.$$

5.3 Διαστήματα και περιοχές αξιοπιστίας

Η ιδέα ενός διαστήματος αξιοπιστίας είναι να δώσουμε ένα ανάλογο του διαστήματος εμπιστοσύνης στην κλασική στατιστική. Ο συλλογισμός είναι ότι οι σημειακές εκτιμήσεις δε δίνουν ένα μέτρο της ακρίβειας, οπότε είναι προτιμότερο να δώσουμε ένα διάστημα μέσα στο οποίο είναι "πιθανό" να ανήκει η παράμετρος. Αυτό προκαλεί προβλήματα στην κλασική στατιστική αφού οι παράμετροι δε θεωρούνται τυχαίες, οπότε δεν είναι δυνατό να δώσουμε ένα διάστημα με την ερμηνεία ότι υπάρχει μία συγκεκριμένη πιθανότητα η παράμετρος να ανήκει στο διάστημα. (Αντιθέτως, τα διαστήματα εμπιστοσύνης έχουν την ερμηνεία ότι αν η δειγματοληψία επαναλαμβανόταν, θα υπήρχε μία συγκεκριμένη πιθανότητα το διάστημα που θα προέκυπτε να περιείχε την παράμετρο - είναι το διάστημα που είναι τυχαίο και όχι η παράμετρος.)

Δεν υπάρχει τέτοια δυσκολία στην Μπεϋζιανή προσέγγιση επειδή οι παράμετροι θεωρούνται τυχαίες. Έτσι, μία περιοχή $C_\alpha(x)$ είναι μία $100(1-\alpha)\%$ *περιοχή αξιοπιστίας* για το θ αν:

$$\int_{C_\alpha(x)} f(\theta \mid x) d\theta = 1 - \alpha.$$

Δηλαδή, υπάρχει posterior πιθανότητα $1-\alpha$ το θ να ανήκει στην περιοχή $C_\alpha(x)$. Ακόμα και στην περίπτωση μονοδιάστατης παραμέτρου θ, καλούνται περιοχές και όχι διαστήματα αξιοπιστίας γιατί δεν είναι απαραίτητο η περιοχή $C_\alpha(x)$ να ορίζει ένα διάστημα στον $\mathbb{R}$ - θα μπορούσε κάλλιστα να ορίζει μία ένωση πεπερασμένων διαστημάτων.

Κάποιοι Μπεϋζιανοί επιχειρηματολογούν ότι οι περιοχές αξιοπιστίας έχουν μικρή αξία, αφού ολόκληρη η posterior κατανομή περιέχει την πληροφορία για τη συμπερασματολογία, και ότι έχουν προταθεί μόνο προκειμένου να δώσουν κάτι συγκρίσιμο με τα διαστήματα εμπιστοσύνης.

Μία δυσκολία με τις περιοχές αξιοπιστίας (κοινή με τα διαστήματα εμπιστοσύνης) είναι ότι δεν ορίζονται με μοναδικό τρόπο. Οποιαδήποτε περιοχή με posterior πιθανότητα $1-\alpha$ αρκεί. Αφού θέλουμε η περιοχή να περιέχει μόνο τις ”πιο πιθανές” τιμές της παραμέτρου, συνηθίζεται να επιβάλλουμε έναν επιπλέον περιορισμό, ο οποίος είναι το μήκος της περιοχής να είναι όσο το δυνατόν μικρότερο. Αυτό ισοδυναμεί με μία περιοχή της μορφής $C_\alpha(x) = \{\theta \in \Theta : f(\theta \mid x) \geqslant \gamma\}$, όπου το γ επιλέγεται έτσι ώστε να εξασφαλίσει ότι:

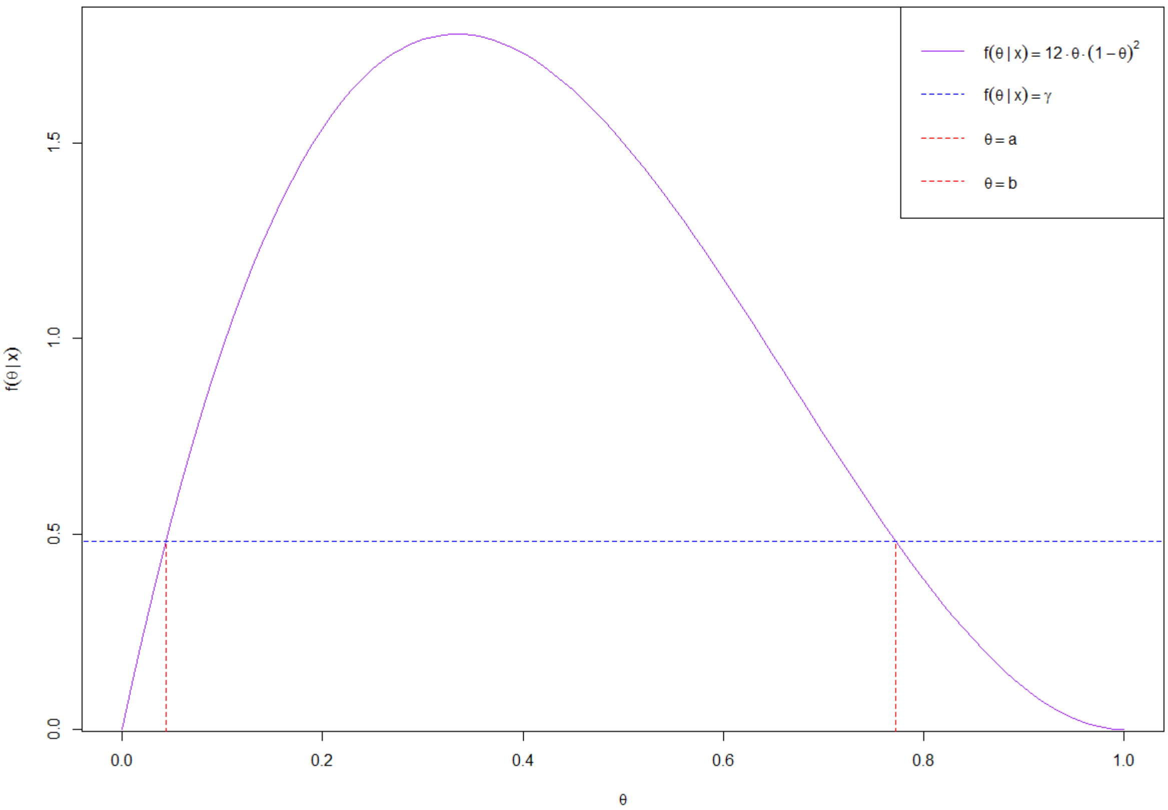
$$\int_{C_\alpha(x)} f(\theta \mid x) d\theta = 1 - \alpha.$$

Τέτοιες περιοχές καλούνται *περιοχές αξιοπιστίας υψηλότερης πυκνότητας*. Αν η posterior πυκνότητα είναι μονοκόρυφη, τότε οι περιοχές αξιοπιστίας υψηλότερης πυκνότητας είναι διαστήματα της μορφής (a, b) (βλέπε σχήμα 5.2).

Γενικά, αυτές οι περιοχές μπορούν να υπολογιστούν μόνο αριθμητικά, όμως στις περισσότερες τυπικές μονοδιάστατες posterior κατανομές πινακοποιούνται για ένα εύρος τιμών του α. (Παρεμπιπτόντως, παρατηρούμε ότι υπάρχει ένας συμβιβασμός στην επιλογή ενός κατάλληλου α: μικρές τιμές του α θα δώσουν μεγάλες περιοχές, ενώ μεγάλες τιμές του α δίνουν περιοχές στις οποίες η παράμετρος έχει μικρή πιθανότητα να εμπεριέχεται.)

Σημείωση 4. *Πρέπει ΠΑΝΤΑ να σχεδιάζουμε την posterior πυκνότητα προτού να προσπαθήσουμε να υπολογίσουμε την περιοχή αξιοπιστίας υψηλότερης πυκνότητας.*

Παράδειγμα 5.4. *(Μέση τιμή κανονικής). Έστω $X_1, X_2, \ldots, X_n$ τυχαίες μεταβλητές από την κατανομή $N(\theta, \sigma^2)$ με σ^2 γνωστό και μία prior για το θ της μορφής $\theta \sim N(b, d^2)$.*



Σχημα 5.2: Περιοχή αξιοπιστίας υψηλότερης πυκνότητας μονοκόρυφης posterior κατανομής. Η περιοχή ορίζει ένα διάστημα της μορφής (a, b).

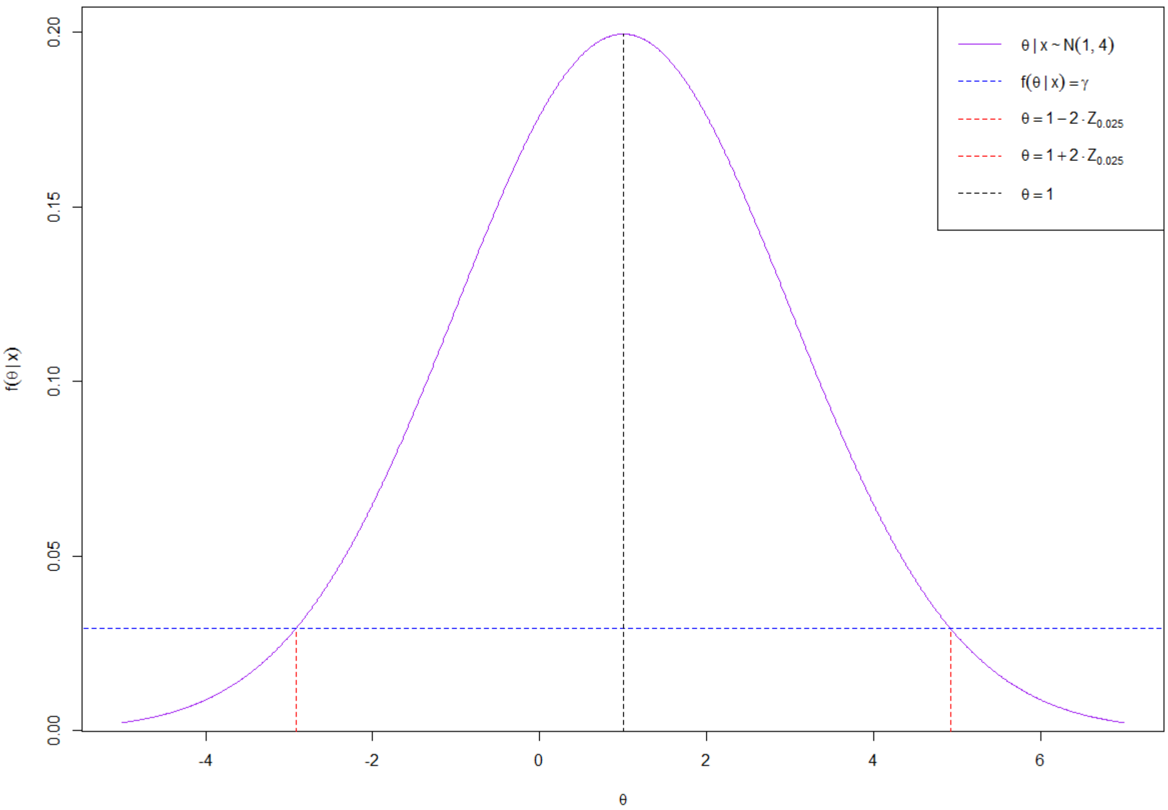
Με αυτήν την κατασκευή πήραμε μία posterior για το θ της μορφής $\theta \mid x \sim N(\mu, s^2)$, όπου:

$$\mu = \left(\frac{b}{d^2} + \frac{n\overline{x}}{\sigma^2}\right)\left(\frac{1}{d^2} + \frac{n}{\sigma^2}\right)^{-1} \quad \text{και} \quad s^2 = \left(\frac{1}{d^2} + \frac{n}{\sigma^2}\right)^{-1}.$$

Τώρα, αφού η κανονική κατανομή είναι μονοκόρυφη και συμμετρική γύρω από το μ, η περιοχή αξιοπιστίας υψηλότερης πυκνότητας θα είναι ένα συμμετρικό διάστημα της μορφής $(\mu - c, \mu + c)$ (βλέπε σχήμα 5.3). Έπεται ότι το $100(1-\alpha)\%$ διάστημα αξιοπιστίας υψηλότερης πυκνότητας για το θ είναι το $\left(\mu - Z_{\frac{\alpha}{2}} \cdot s, \mu + Z_{\frac{\alpha}{2}} \cdot s\right)$, όπου $Z_{\frac{\alpha}{2}}$ είναι το κατάλληλο άνω ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής $N(0, 1)$.

Επιπλέον, παρατηρούμε ότι, καθώς $n \to \infty$, το διάστημα γίνεται $\overline{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, το οποίο είναι επακριβώς το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για το θ που παίρνουμε στην κλασική συμπερασματολογία. Σε αυτήν την ειδική περίπτωση λοιπόν, το Μπεϋζιανό διάστημα αξιοπιστίας και το κλασικό διάστημα εμπιστοσύνης ταυτίζονται, παρόλο που οι *ερμηνείες* τους είναι αρκετά διαφορετικές.

Παράδειγμα 5.5. *(Άγνωστη διασπορά). Έστω $X_1, X_2, \ldots, X_n$ τυχαίες μεταβλητές από την κατανομή $N(\theta, \phi)$ με ϕ άγνωστο και μία "μη-πληροφοριακή" prior της μορφής $f(\theta, \phi) \propto \phi^{-1}$*



Σχημα 5.3: Περιοχή αξιοπιστίας υψηλότερης πυκνότητας μίας $N(1,4)$ posterior κατανομής. Η περιοχή ορίζει ένα διάστημα της μορφής $\left(1-2\cdot Z_{\frac{\alpha}{2}}, 1+2\cdot Z_{\frac{\alpha}{2}}\right)$.

για $\theta \in \mathbb{R}$.

Αυτό οδηγεί στις περιθώριες posterior κατανομές:

$$\frac{\theta-\overline{x}}{s}\sqrt{n} \sim t_{n-1} \quad \text{και} \quad \frac{(n-1)s^2}{\phi} \sim \chi^2_{n-1}, \quad \text{όπου } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

Οπότε, λόγω της συμμετρίας της κατανομής t του Student, το $100(1-\alpha)\%$ διάστημα αξιοπιστίας υψηλότερης πυκνότητας για το θ είναι το $\overline{x} \pm t_{n-1;\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$, όπου $t_{n-1;\frac{\alpha}{2}}$ είναι το κατάλληλο άνω ποσοστιαίο σημείο της κατανομής t του Student με $n-1$ βαθμούς ελευθερίας.

Από την άλλη, το διάστημα αξιοπιστίας για το ϕ είναι πιο δύσκολο. Αφού γνωρίζουμε ότι $\frac{(n-1)s^2}{\phi} \sim \chi^2_{n-1}$, προκύπτει ότι $\frac{\phi}{(n-1)s^2} \sim \text{Inv-}\chi^2_{n-1}$, δηλαδή μία λεγόμενη αντίστροφη κατανομή χ^2. Τα ζητούμενα ποσοστιαία σημεία των διαστημάτων αξιοπιστίας υψηλότερης πυκνότητας για μία ποικιλία τιμών του α είναι διαθέσιμα σε πίνακες.

Παράδειγμα 5.6. *(Διωνυμικό δείγμα). Θεωρούμε ότι έχουμε $x \mid \theta \sim Bin(n,\theta)$ και μία prior κατανομή $\theta \sim Beta(p,q)$ με $p>1$ και $q>1$.*

Αυτό δίνει την posterior κατανομή $\theta \mid x \sim \text{Beta}(p+x, q+n-x)$. Σύμφωνα με το σχήμα 2.1, η κατανομή αυτή είναι μονοκόρυφη σε αυτήν την περίπτωση. Επομένως, το $100(1-\alpha)\%$ διάστημα αξιοπιστίας υψηλότερης πυκνότητας της μορφής (a, b) ικανοποιεί τις σχέσεις:

$$\frac{1}{B(p+x, q+n-x)} \int_a^b \theta^{p+x-1}(1-\theta)^{q+n-x-1} d\theta = 1-\alpha \quad \text{και}$$

$$\frac{1}{B(p+x, q+n-x)} a^{p+x-1}(1-a)^{q+n-x-1} = \frac{1}{B(p+x, q+n-x)} b^{p+x-1}(1-b)^{q+n-x-1} = \gamma.$$

Γενικά, αυτό το πρόβλημα χρειάζεται αριθμητική επίλυση.

5.4 Έλεγχοι υποθέσεων

Οι έλεγχοι υποθέσεων είναι προβλήματα αποφάσεων για την επιλογή μεταξύ 2 (ή και περισσότερων) υποθέσεων $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$. Στην απλούστερη περίπτωση όπου τα Ω_0 και Ω_1 είναι μονοσύνολα, ο έλεγχος παίρνει τη μορφή $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$. Η κλασική προσέγγιση για αυτό το πρόβλημα είναι συνήθως να βασίσουμε τον έλεγχο σε κάποιον *λόγο πιθανοφανειών* της μορφής:

$$\lambda = \frac{f(x \mid \theta_1)}{f(x \mid \theta_0)}.$$

Μεγάλες τιμές του λ υποδεικνύουν ότι τα παρατηρούμενα δεδομένα είναι πιο πιθανό να έχουν προκύψει αν το θ_1 είναι η πραγματική τιμή του θ και όχι το θ_0. Στην Μπεϋζιανή θεώρηση των πραγμάτων, θα πρέπει να χρησιμοποιήσουμε και την prior πληροφορία που έχουμε για το θ. Επομένως, μπορούμε να υπολογίσουμε τις posterior πιθανότητες των τιμών θ_0 και θ_1:

$$f(\theta_1 \mid x) = \frac{f(\theta_1) \cdot f(x \mid \theta_1)}{f(\theta_0) \cdot f(x \mid \theta_0) + f(\theta_1) \cdot f(x \mid \theta_1)} \quad \text{και} \quad f(\theta_0 \mid x) = 1 - f(\theta_1 \mid x).$$

Έτσι, είναι φυσικό να βασίσουμε τις αποφάσεις του ελέγχου στις σχετικές posterior πιθανότητες των 2 υποτιθέμενων τιμών. Δηλαδή, κοιτάζουμε τον λόγο:

$$\lambda_B = \frac{f(\theta_1 \mid x)}{f(\theta_0 \mid x)} = \frac{f(\theta_1) \cdot f(x \mid \theta_1)}{f(\theta_0) \cdot f(x \mid \theta_0)}. \tag{5.1}$$

Αυτός καλείται συνήθως *posterior λόγος πιθανοτήτων*. Παρατηρούμε συγκεκριμένα ότι δεν υπάρχει ανάγκη υπολογισμού σταθερών κανονικοποίησης αφού απλοποιούνται. Πάλι, μεγάλες τιμές του λ_B συνηγορούν ότι η πιο πιθανή υπόθεση είναι η H_1.

Υπάρχει μία συναφής έννοια που καλείται *παράγοντας Bayes*. Μπορούμε να δούμε από την εξίσωση 5.1 ότι ο posterior λόγος πιθανοτήτων είναι το γινόμενο του prior λόγου πιθανοτήτων επί τον λόγο πιθανοφανειών. Σε αυτό το πλαίσιο, ο λόγος πιθανοφανειών καλείται παράγοντας Bayes (Bayes Factor), δηλαδή:

$$\mathrm{BF} = \frac{f(x \mid \theta_1)}{f(x \mid \theta_0)} = \frac{f(\theta_0)}{f(\theta_1)} \cdot \frac{f(\theta_1 \mid x)}{f(\theta_0 \mid x)}.$$

Ο σκοπός του να εστιάσουμε στον παράγοντα Bayes είναι ότι μετράει το βάρος της πληροφορίας που περιέχεται στα δεδομένα υπέρ της H_1. Αν ο παράγοντας Bayes είναι επαρκώς μεγάλος, τότε θα υπερνικήσει οποιαδήποτε prior προτίμηση μπορεί να είχαμε για την H_0, ώστε η posterior προτίμησή μας να είναι για την H_1.

Στη γενική περίπτωση όπου ενδιαφερόμαστε να ελέγξουμε 2 υποθέσεις $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$ μπορούμε και πάλι να υπολογίσουμε τις posterior πιθανότητες των 2 υποθέσεων, έχοντας πρώτα προσδιορίσει prior πιθανότητες $P(\theta \in \Omega_0)$ και $P(\theta \in \Omega_1)$ για τις υποθέσεις. Τότε, έχουμε:

$$P(\theta \in \Omega_1 \mid x) = \frac{P(\theta \in \Omega_1) \cdot f(x \mid \theta \in \Omega_1)}{P(\theta \in \Omega_0) \cdot f(x \mid \theta \in \Omega_0) + P(\theta \in \Omega_1) \cdot f(x \mid \theta \in \Omega_1)}, \quad \text{όπου:}$$

$$f(x \mid \theta \in \Omega_i) = \int_{\Omega_i} f(\theta) \cdot f(x \mid \theta) d\theta, \quad i = 0, 1.$$

Προφανώς, είναι απλό να γενικεύσουμε την παραπάνω προσέγγιση του ελέγχου στην περίπτωση όπου θέλουμε να ελέγξουμε περισσότερες από 2 υποθέσεις.

5.5 Μπεϋζιανή σύγκριση μοντέλων

Η σύγκριση ενός αριθμού από ανταγωνιζόμενα μοντέλα για ένα δεδομένο σύνολο παρατηρούμενων δεδομένων είναι μία σημαντική περιοχή της Μπεϋζιανής θεωρίας αποφάσεων. Προτού παρουσιάσουμε το πρόβλημα της σύγκρισης μοντέλου, ας ορίσουμε την περιθώρια πιθανοφάνεια ενός δοσμένου μοντέλου.

Η *περιθώρια πιθανοφάνεια* (ή *evidence*) $f(x)$ ενός δοσμένου μοντέλου $f(x \mid \theta)$ είναι η περιθώρια κατανομή των δεδομένων κάτω από αυτό το μοντέλο. Υπολογίζεται ολοκληρώνοντας πάνω στον παραμετρικό χώρο Θ το γινόμενο της prior κατανομής $f(\theta)$ επί την πιθανοφάνεια, δηλαδή:

$$f(x) = \int_{\Theta} f(\theta) \cdot f(x \mid \theta) d\theta = E_\theta \left[f(x \mid \theta) \right].$$

Με άλλα λόγια, η περιθώρια πιθανοφάνεια $f(x)$ είναι η σταθερά κανονικοποίησης της posterior κατανομής του θ, η οποία δίνεται ως:

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

Παρατηρούμε ότι όλες οι σταθερές από την prior και την πιθανοφάνεια περιλαμβάνονται στον υπολογισμό της $f(x)$.

Ισοδύναμα, η περιθώρια πιθανοφάνεια ορίζεται ως η μέση τιμή της πιθανοφάνειας κάτω από την prior κατανομή $f(\theta)$. Προκειμένου να είμαστε σε θέση να υπολογίσουμε το ολοκλήρωμα που εμπλέκεται στον υπολογισμό της $f(x)$ χρειάζεται να επιλέξουμε προσεκτικά μία κατάλληλη (προτιμότερα συζυγή) prior για το θ.

Οι περιθώριες πιθανοφάνειες παίζουν σημαντικό ρόλο στην *Μπεϋζιανή σύγκριση μοντέλων*. Θεωρούμε ένα πλήθος από ανταγωνιζόμενα μοντέλα $M_1, M_2, \ldots, M_k$ που παραμετροποιούνται αντίστοιχα από τις παραμέτρους $\theta_1, \theta_2, \ldots, \theta_k$ για ένα παρατηρούμενο σύνολο δεδομένων. Υπό την παρουσία αβεβαιότητας για το σωστό μοντέλο, η Μπεϋζιανή συμπερασματολογία περιλαμβάνει:

1. Υπολογισμό των posterior πιθανοτήτων $P(M_j \mid x)$ για κάθε μοντέλο M_j, $j = 1, 2, \ldots, k$.

2. Υπολογισμό των posteior κατανομών $f(\theta_j \mid x, M_j)$ των παραμέτρων θ_j του μοντέλου M_j, $j = 1, 2, \ldots, k$.

Έχοντας προσδιορίσει τις prior πιθανότητες $P(M_j)$ για όλα τα ανταγωνιζόμενα μοντέλα και επιλέξει προσεκτικά κατάλληλες prior κατανομές $f(\theta_j \mid M_j)$ για τις παραμέτρους του εκάστοτε μοντέλου, τα posterior συμπεράσματα λαμβάνονται ως ακολούθως:

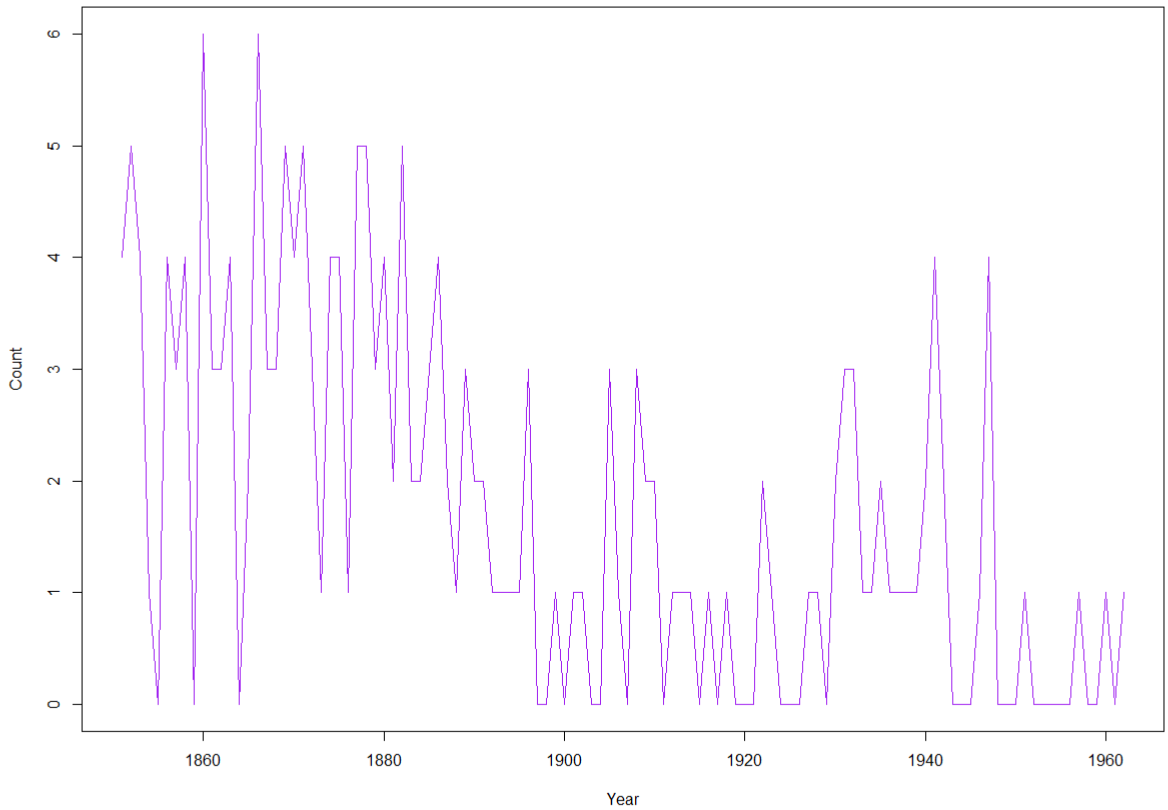
1. Η posterior πιθανότητα του μοντέλου M_j υπολογίζεται χρησιμοποιώντας το θεώρημα του Bayes:
$$P(M_j \mid x) \propto P(M_j) \cdot f(x \mid M_j), \quad j = 1, 2, \ldots, k.$$

2. Η posterior κατανομή των παραμέτρων θ_j του μοντέλου M_j υπολογίζεται χρησιμοποιώντας το θεώρημα του Bayes:
$$f(\theta_j \mid x, M_j) = \frac{f(\theta_j \mid M_j) \cdot f(x \mid \theta_j, M_j)}{f(x \mid M_j)}, \quad j = 1, 2, \ldots, k.$$

Παράδειγμα 5.7. *Για να επιδείξουμε την Μπεϋζιανή σύγκριση μοντέλων, θα δούμε ένα αρκετά περίπλοκο παράδειγμα που αφορά σημεία δομικής αλλαγής σε διαδικασίες Poisson. Τα δεδομένα αποτελούνται από το πλήθος των ατυχημάτων ανά έτος σε Βρετανικά ανθρακωρυχεία την περίοδο* 1851-1962*. Ένα γράφημα των δεδομένων δίνεται στο σχήμα 5.4. Από αυτό το γράφημα φαίνεται ότι είχε όντως υπάρξει μία μείωση στον ρυθμό των ατυχημάτων μέσα σε αυτήν την περίοδο.*

Για τα δεδομένα των ατυχημάτων στα ανθρακωρυχεία μπορούμε λοιπόν να θεωρήσουμε 2 *μοντέλα:*

M_1 : *Κάθε παρατήρηση x_i είναι μία ανεξάρτητη πραγματοποίηση μίας τυχαίας μεταβλη-*



Σχημα 5.4: Χρονοσειρά του πλήθους των ατυχημάτων στα ανθρακορυχεία

τής Poisson με μέση τιμή θ.

M_2 : *Για $i = 1, 2, \ldots, t$, η παρατήρηση x_i είναι μία ανεξάρτητη πραγματοποίηση μίας τυχαίας μεταβλητής Poisson με μέση τιμή θ_1.*

Για $i = t+1, t+2, \ldots, n$, η παρατήρηση x_i είναι μία ανεξάρτητη πραγματοποίηση μίας τυχαίας μεταβλητής Poisson με μέση τιμή θ_2.

Ορίζουμε prior πιθανότητες $P(M_1) = P(M_2) = \frac{1}{2}$ για τα μοντέλα και prior κατανομές $\theta \sim \text{Exp}\left(\frac{1}{2}\right)$, $\theta_1 \sim \text{Exp}\left(\frac{1}{2}\right)$, $\theta_2 \sim \text{Exp}\left(\frac{1}{2}\right)$, $t \sim U\{1, 2, \ldots, n-1\}$ για τις παραμέτρους των μοντέλων. Παρατηρούμε ότι $\text{Exp}\left(\frac{1}{2}\right) \equiv \text{Gamma}\left(1, \frac{1}{2}\right)$, δηλαδή έχουμε επιλέξει τη συζυγή $\text{Gamma}(p, q)$ prior για τα θ, θ_1, θ_2 με $p = 1$ και $q = \frac{1}{2}$.

Μοντέλο $\mathbf{M_1}$

Prior του θ:

$$f(\theta) = \frac{1}{2} e^{-\frac{\theta}{2}}, \quad \theta > 0.$$

Πιθανοφάνεια του M_1:

$$f(x \mid \theta) = e^{-n\theta} \theta^{n\bar{x}} \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Posterior του θ:

$$\begin{aligned} f(\theta \mid x) &\propto f(\theta) \cdot f(x \mid \theta) \\ &= \frac{1}{2} e^{-\frac{\theta}{2}} \cdot e^{-n\theta} \theta^{n\overline{x}} \prod_{i=1}^{n} \frac{1}{x_i!} \\ &= \theta^{n\overline{x}} e^{-\left(n+\frac{1}{2}\right)\theta} \cdot \frac{1}{2} \prod_{i=1}^{n} \frac{1}{x_i!}. \end{aligned}$$

Περιθώρια πιθανοφάνεια του M_1:

$$\begin{aligned} f(x \mid M_1) &= \int_0^{\infty} f(\theta) \cdot f(x \mid \theta) d\theta \\ &= \int_0^{\infty} \theta^{n\overline{x}} e^{-\left(n+\frac{1}{2}\right)\theta} d\theta \cdot \frac{1}{2} \prod_{i=1}^{n} \frac{1}{x_i!} \\ &= \frac{\Gamma(n\overline{x}+1)}{\left(n+\frac{1}{2}\right)^{n\overline{x}+1}} \cdot \frac{1}{2} \prod_{i=1}^{n} \frac{1}{x_i!}, \quad \text{αφού:} \end{aligned}$$

$$\int_0^{\infty} \underbrace{\frac{\left(n+\frac{1}{2}\right)^{n\overline{x}+1}}{\Gamma(n\overline{x}+1)} \theta^{n\overline{x}} e^{-\left(n+\frac{1}{2}\right)\theta}}_{\text{σ.π.π. της κατανομής Gamma}\left(n\overline{x}+1,\, n+\frac{1}{2}\right)} d\theta = 1.$$

Μοντέλο $\mathbf{M_2}$

Από κοινού prior των θ_1, θ_2 και t:

$$\begin{aligned} f(\theta_1, \theta_2, t) &= f(\theta_1) \cdot f(\theta_2) \cdot f(t) \\ &= \frac{1}{2} e^{-\frac{\theta_1}{2}} \cdot \frac{1}{2} e^{-\frac{\theta_2}{2}} \cdot \frac{1}{n-1} \\ &= \frac{1}{4(n-1)} \cdot e^{-\frac{\theta_1}{2}} \cdot e^{-\frac{\theta_2}{2}}, \ \theta_1 > 0, \ \theta_2 > 0, \ t \in \{1, 2, \ldots, n-1\}. \end{aligned}$$

Πιθανοφάνεια του M_2:

$$\begin{aligned} f(x \mid \theta_1, \theta_2, t) &= \prod_{i=1}^{t} f(x_i \mid \theta_1) \cdot \prod_{i=t+1}^{n} f(x_i \mid \theta_2) \\ &= e^{-t\theta_1} \theta_1^{s(t)} \prod_{i=1}^{t} \frac{1}{x_i!} \cdot e^{-(n-t)\theta_2} \theta_2^{s(n)-s(t)} \prod_{i=t+1}^{n} \frac{1}{x_i!} \\ &= e^{-t\theta_1} \theta_1^{s(t)} \cdot e^{-(n-t)\theta_2} \theta_2^{s(n)-s(t)} \cdot \prod_{i=1}^{n} \frac{1}{x_i!}, \quad \text{όπου:} \end{aligned}$$

$$s(t) = \sum_{i=1}^{t} x_i \quad \text{και} \quad s(n) - s(t) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{t} x_i = \sum_{i=t+1}^{n} x_i.$$

Από κοινού posterior των θ_1, θ_2 και t:

$$f(\theta_1, \theta_2, t \mid x) \propto f(\theta_1, \theta_2, t) \cdot f(x \mid \theta_1, \theta_2, t)$$

$$
\begin{aligned}
&= \frac{1}{4(n-1)} \cdot e^{-\frac{\theta_1}{2}} \cdot e^{-\frac{\theta_2}{2}} \cdot e^{-t\theta_1}\theta_1^{s(t)} \cdot e^{-(n-t)\theta_2}\theta_2^{s(n)-s(t)} \cdot \prod_{i=1}^{n} \frac{1}{x_i!} \\
&= \theta_1^{s(t)} e^{-\left(t+\frac{1}{2}\right)\theta_1} \cdot \theta_2^{s(n)-s(t)} e^{-\left(n-t+\frac{1}{2}\right)\theta_2} \cdot \frac{1}{4(n-1)} \prod_{i=1}^{n} \frac{1}{x_i!}.
\end{aligned}
$$

Περιθώρια posterior του t:

$$
\begin{aligned}
f(t \mid x) &= \int_0^\infty \int_0^\infty f(\theta_1, \theta_2, t \mid x) d\theta_1 d\theta_2 \\
&\propto \int_0^\infty \int_0^\infty f(\theta_1, \theta_2, t) \cdot f(x \mid \theta_1, \theta_2, t) d\theta_1 d\theta_2 \\
&= \int_0^\infty \int_0^\infty \theta_1^{s(t)} e^{-\left(t+\frac{1}{2}\right)\theta_1} \cdot \theta_2^{s(n)-s(t)} e^{-\left(n-t+\frac{1}{2}\right)\theta_2} d\theta_1 d\theta_2 \cdot \frac{1}{4(n-1)} \prod_{i=1}^{n} \frac{1}{x_i!} \\
&= \int_0^\infty \theta_1^{s(t)} e^{-\left(t+\frac{1}{2}\right)\theta_1} d\theta_1 \cdot \int_0^\infty \theta_2^{s(n)-s(t)} e^{-\left(n-t+\frac{1}{2}\right)\theta_2} d\theta_2 \cdot \frac{1}{4(n-1)} \prod_{i=1}^{n} \frac{1}{x_i!} \\
&= \frac{\Gamma(s(t)+1)}{\left(t+\frac{1}{2}\right)^{s(t)+1}} \cdot \frac{\Gamma(s(n)-s(t)+1)}{\left(n-t+\frac{1}{2}\right)^{s(n)-s(t)+1}} \cdot \frac{1}{4(n-1)} \prod_{i=1}^{n} \frac{1}{x_i!} = p(t) \Rightarrow
\end{aligned}
$$

$$
f(t \mid x) = \frac{p(t)}{\sum_{i=1}^{n-1} p(i)}, \quad t = 1, 2, \ldots, n-1.
$$

Περιθώρια πιθανοφάνεια του M_2:

$$
\begin{aligned}
f(x \mid M_2) &= \sum_{t=1}^{n-1} \int_0^\infty \int_0^\infty f(\theta_1, \theta_2, t) \cdot f(x \mid \theta_1, \theta_2, t) d\theta_1 d\theta_2 \\
&= \sum_{t=1}^{n-1} \left[\frac{\Gamma(s(t)+1)}{\left(t+\frac{1}{2}\right)^{s(t)+1}} \cdot \frac{\Gamma(s(n)-s(t)+1)}{\left(n-t+\frac{1}{2}\right)^{s(n)-s(t)+1}} \right] \cdot \frac{1}{4(n-1)} \prod_{i=1}^{n} \frac{1}{x_i!} = \sum_{t=1}^{n-1} p(t).
\end{aligned}
$$

Posterior πιθανότητα του M_2:

$$
\begin{aligned}
P(M_2 \mid x) &= \frac{P(M_2) \cdot f(x \mid M_2)}{P(M_1) \cdot f(x \mid M_1) + P(M_2) \cdot f(x \mid M_2)} \\
&= \frac{f(x \mid M_2)}{f(x \mid M_1) + f(x \mid M_2)}.
\end{aligned}
$$

Posterior πιθανότητα του M_1:

$$
P(M_1 \mid x) = 1 - P(M_2 \mid x).1
$$

Δεσμευμένη posterior του θ_1:

$$
\begin{aligned}
f(\theta_1 \mid x, \theta_2, t) &\propto f(\theta_1, \theta_2, t \mid x) \\
&\propto \theta_1^{s(t)} e^{-\left(t+\frac{1}{2}\right)\theta_1}.
\end{aligned}
$$

Επομένως, το θ_1 είναι ανεξάρτητο από το θ_2 και $\theta_1 \mid x, t \sim \text{Gamma}\left(s(t)+1, t+\frac{1}{2}\right)$.

Δεσμευμένη posterior του θ_2:

$$\begin{aligned} f(\theta_2 \mid x, \theta_1, t) &\propto f(\theta_1, \theta_2, t \mid x) \\ &\propto \theta_2^{s(n)-s(t)} e^{-\left(n-t+\frac{1}{2}\right)\theta_2}. \end{aligned}$$

Συνεπώς, το θ_2 είναι ανεξάρτητο από το θ_1 και $\theta_2 \mid x, t \sim \text{Gamma}\left(s(n) - s(t) + 1, n - t + \frac{1}{2}\right)$. Συμπεραίνουμε ότι η από κοινού posterior των παραμέτρων θ_1, θ_2 και t παραγοντοποιείται ως $f(\theta_1, \theta_2, t \mid x) = f(t \mid x) \cdot f(\theta_1 \mid x, t) \cdot f(\theta_2 \mid x, t)$.

5.6 Ασκήσεις

Άσκηση 5.1. *Μία ειδικός στον χώρο της τέχνης έχει υποβληθεί σε έναν έλεγχο της αξιοπιστίας της στον οποίο έχει κρίνει ξεχωριστά ως "γνήσια" ή "πλαστά" ένα μεγάλο πλήθος από έργα τέχνης γνωστής προέλευσης. Από αυτά, φαίνεται ότι έχει πιθανότητα 0.8 να εντοπίσει σωστά ένα πλαστό έργο και πιθανότητα 0.7 να αναγνωρίσει σωστά ένα γνήσιο έργο. Μας έχει προσφερθεί ένα έργο τέχνης στη φαινομενική τιμή ευκαιρία των 100€. Αν δεν είναι γνήσιο, τότε δεν αξίζει τίποτα. Αν είναι γνήσιο, τότε πιστεύουμε ότι μπορούμε να το πουλήσουμε άμεσα για 300€. Πιστεύουμε ότι υπάρχει πιθανότητα 0.5 το έργο να είναι γνήσιο. Η ειδικός χρεώνει 30€ για τις υπηρεσίες της. Μας συμφέρει να την προσλάβουμε για την αξιολόγηση του έργου;*

Άσκηση 5.2. *Θεωρούμε ένα πρόβλημα αποφάσεων με 2 πράξεις a_1 και a_2 και μία συνάρτηση απώλειας που εξαρτάται από μία παράμετρο $\theta \in (0,1)$. Η συνάρτηση απώλειας είναι:*

$$L(\theta, a) = \begin{cases} 0, & a = a_1 \\ 2 - 3\theta, & a = a_2 \end{cases}.$$

Θεωρούμε ότι έχουμε μία Beta$(1,1)$ prior για το θ και μία παρατήρηση $x \mid \theta \sim$ Bin(n, θ). Η posterior κατανομή είναι Beta$(x+1, n-x+1)$.

Υπολογίστε την posterior μέση απώλεια για κάθε πράξη και τον κανόνα απόφασης κατά Bayes. (Μπορείτε να θεωρήσετε γνωστή τη μέση τιμή της κατανομής Beta(p,q), η οποία δίνεται από την εξίσωση 2.1.)

Άσκηση 5.3. i. *Για μία παράμετρο θ με posterior κατανομή Beta(P,Q) υπολογίστε την posterior κορυφή συναρτήσει των P και Q και συγκρίνετέ την με την posterior μέση τιμή.*

ii. *Επαναλάβετε το μέρος i. για μία παράμετρο θ με posterior κατανομή Gamma(P,Q).*

Άσκηση 5.4. *Η παράμετρος θ έχει posterior κατανομή Beta$(3,2)$. Δείξτε ότι το διάστημα $\left(\frac{5}{21}, \frac{20}{21}\right)$ είναι η 94.3% περιοχή αξιοπιστίας υψηλότερης πυκνότητας για το θ.*

Άσκηση 5.5. *Μία παράμετρος θ έχει posterior μέση τιμή μ και posterior διασπορά v. Θέτουμε $\phi = \theta^2$. Δείξτε ότι η εκτίμηση του ϕ που έχει ελάχιστο μέσο posterior τετραγωνικό σφάλμα δίνεται ως $a = \mu + v^2$.*

Άσκηση 5.6. *Μία παράμετρος θ έχει posterior κατανομή Gamma$(1, 1)$. Υπολογίστε την 95% περιοχή αξιοπιστίας υψηλότερης πυκνότητας για το θ. Τώρα, θεωρήστε τον μετασχηματισμό $\phi = \sqrt{2\theta}$. Βρείτε την posterior πυκνότητα του ϕ και εξηγείστε γιατί η περιοχή αξιοπιστίας υψηλότερης πυκνότητας για το ϕ δε λαμβάνεται μετασχηματίζοντας την περιοχή για το θ κατά τον ίδιο τρόπο.*

Άσκηση 5.7. *Σε ένα πλαίσιο θεωρίας αποφάσεων, έστω $L(\theta, a)$ η απώλεια που επιφέρεται παίρνοντας την απόφαση a όταν η πραγματική τιμή της άγνωστης παραμέτρου είναι θ και έστω $f(\theta \mid x)$ η posterior κατανομή του θ δεδομένης της μέτρησης μίας σχετικής μεταβλητής x.*

i. *Προσδιορίστε τον κανόνα απόφασης $d(x)$ κατά Bayes και τις αντίστοιχες απώλειες στο διακριτό παράδειγμα όπου η συνάρτηση απώλειας και η posterior κατανομή δίνονται στους παρακάτω πίνακες:*

$L(\theta, a)$	θ_1	θ_2	θ_3
a_1	5	10	25
a_2	15	15	5

$f(\theta \mid x)$	θ_1	θ_2	θ_3
x_1	0.5	0.3	0.2
x_2	0.1	0.5	0.4

ii. *Ποια επιπλέον πληροφορία θα χρειαζόταν για να προσδιορίσουμε το ρίσκο κατά Bayes σε αυτό το πρόβλημα και να αποφασίσουμε αν άξιζε το κόστος της μέτρησης του x;*

Άσκηση 5.8. *Θεωρούμε ένα δείγμα $x_1, x_2 \ldots, x_n$ ανεξάρτητων παρατηρήσεων από μία τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με μέση τιμή θ. Θεωρούμε τον έλεγχο των υποθέσεων $H_0 : \theta = 1$ vs. $H_1 : \theta \neq 1$. Υποθέτουμε prior πιθανότητα 0.95 για την H_0, 0.05 για την H_1 και μία Gamma(p, q) για το θ κάτω από την H_1, δηλαδή:*

$$f(\theta) = \frac{q^p}{\Gamma(p)} \theta^{p-1} e^{-q\theta}, \quad \theta \neq 1.$$

i. *Υπολογίστε την posterior πιθανότητα της H_0.*

ii. *Υποθέτουμε ότι $n = 10$, $\sum_{i=1}^{10} x_i = 20$ και $p = 2q$. Ποια είναι η posterior πιθανότητα της H_0 για $p = 2, 1, 0.5, 0.1$; Τι συμβαίνει σε αυτήν την posterior πιθανότητα καθώς $p \to 0$;*

Κεφάλαιο 6

Πρόβλεψη

6.1 Η κατανομή πρόβλεψης

Μέχρι τώρα, εστιάσαμε στην εκτίμηση παραμέτρων. Δηλαδή, καθορίσαμε ένα μοντέλο πιθανότητας για να περιγράψουμε την τυχαία διαδικασία που παρήγαγε ένα σετ δεδομένων και δείξαμε πώς το Μπεϋζιανό πλαίσιο συνδυάζει τη δειγματική πληροφορία με την prior πληροφορία για να δώσει εκτιμήσεις παραμέτρων με τη μορφή μίας posterior κατανομής. Συνήθως, ο λόγος για τον οποίο φτιάχνουμε ένα στατιστικό μοντέλο είναι για να κάνουμε *προβλέψεις* σχετικά με μελλοντικές τιμές της διαδικασίας. Αυτό γίνεται αρκετά πιο κομψά στην Μπεϋζιανή στατιστική σε σχέση με την αντίστοιχη κλασική θεωρία. Το βασικό σημείο είναι ότι στην πραγματοποίηση προβλέψεων για μελλοντικές τιμές βάσει ενός εκτιμημένου μοντέλου, υπάρχουν δύο πηγές αβεβαιότητας:

- Αβεβαιότητα για τις τιμές των παραμέτρων οι οποίες έχουν εκτιμηθεί βάσει των προηγούμενων δεδομένων.

- Αβεβαιότητα εξαιτίας του γεγονότος ότι κάθε μελλοντική τιμή είναι από μόνη της ένα τυχαίο γεγονός.

Στην κλασική στατιστική είναι σύνηθες να αποδίδουμε ένα μοντέλο στα δεδομένα και μετά να κάνουμε προβλέψεις μελλοντικών τιμών κάτω από την υπόθεση ότι αυτό το μοντέλο είναι σωστό - η λεγόμενη εκτιμητική προσέγγιση. Δηλαδή, μόνο η δεύτερη πηγή αβεβαιότητας εμπεριέχεται στην ανάλυση, οδηγώντας σε εκτιμήσεις που θεωρούνται πιο ακριβείς από ό,τι είναι στην πραγματικότητα. Δεν υπάρχει απολύτως ικανοποιητική προσέγγιση σε αυτό το πρόβλημα κάτω από το κλασικό πλαίσιο, καθώς οι παράμετροι δε θεωρούνται τυχαίες μεταβλητές.

Μέσα στην Μπεϋζιανή συμπερασματολογία, είναι ευθές να επιτρέψουμε και τις δύο πηγές αβεβαιότητας, σταθμίζοντας απλώς σύμφωνα με την αβεβαιότητα των εκτιμήσεων των παραμέτρων, της οποίας η πληροφορία περιέχεται εξ ολοκλήρου στην posterior κατανομή.

Έτσι, υποθέτουμε ότι έχουμε παλαιότερες παρατηρήσεις $x = (x_1, \dots, x_n)$ μίας τυχαίας μεταβλητής με συνάρτηση πυκνότητας πιθανότητας $f(x \mid \theta)$ και θέλουμε να βγάλουμε συμπεράσματα για την κατανομή μίας μελλοντικής τιμής μίας τυχαίας μεταβλητής Y από το ίδιο μοντέλο. Με μία prior κατανομή $f(\theta)$, το θεώρημα του Bayes οδηγεί σε μία posterior κατανομή $f(\theta \mid x)$. Τότε, η ***προβλεπτική συνάρτηση πυκνότητας πιθανότητας*** του y δεδομένου του x είναι:

$$f(y \mid x) = \int_{\Theta} f(y \mid \theta) \cdot f(\theta \mid x) d\theta. \tag{6.1}$$

Συνεπώς, η προβλεπτική πυκνότητα, υπολογισμένη σε μία συγκεκριμένη τιμή y, ισούται με το ολοκλήρωμα της πιθανοφάνειας του y επί την posterior του θ. Το αποτέλεσμα μπορεί επίσης να γραφτεί ως η μέση τιμή της πιθανοφάνειας του y ως προς την posterior κατανομή του θ:

$$f(y \mid x) = E_{\theta} \left[f(y \mid \theta) \mid x \right].$$

Για άλλη μία φορά, είναι σημαντικό να παρατηρήσουμε ότι αυτό το αποτέλεσμα κατασκευάστηκε απλά από τους συνήθεις νόμους χειρισμού των πιθανοτήτων και έχει μία απλή ερμηνεία με όρους πιθανοτήτων. Η κατασκευή πρέπει ωστόσο να γίνεται προσεκτικά και με πρέπουσα έμφαση στις ακόλουθες υποθέσεις. Στα ακόλουθα, η y δεν είναι απαραίτητο να προέρχεται από την ίδια κατανομή με τις παρατηρήσεις x. Είναι ωστόσο σημαντικό ότι, *δεδομένου του* θ, υποθέτουμε ότι η Y είναι ανεξάρτητη του X. Αυτό μας επιτρέπει να γράψουμε την από κοινού πυκνότητα των y και x, δεδομένου του θ, ως το γινόμενο των πυκνοτήτων τους, δηλαδή $f(y, x \mid \theta) = f(y \mid \theta) \cdot f(x \mid \theta)$. Από αυτήν τη σχέση παίρνουμε την από κοινού πυκνότητα των y, x και θ:

$$\begin{aligned} f(y, x, \theta) = f(\theta) \cdot f(y, x \mid \theta) = f(\theta) \cdot f(y \mid \theta) \cdot f(x \mid \theta) \Rightarrow \\ f(y, \theta \mid x) = \frac{f(y, x, \theta)}{f(x)} = \frac{f(\theta) \cdot f(y \mid \theta) \cdot f(x \mid \theta)}{f(x)} \\ = f(y \mid \theta) \cdot \frac{f(\theta) \cdot f(x \mid \theta)}{f(x)} = f(y \mid \theta) \cdot f(\theta \mid x). \end{aligned}$$

Τελικά, ολοκληρώνοντας ως προς θ, παίρνουμε το αποτέλεσμα:

$$f(y \mid x) = \int_{\Theta} f(y, \theta \mid x) d\theta = \int_{\Theta} f(y \mid \theta) \cdot f(\theta \mid x) d\theta.$$

Η αντίστοιχη προσέγγιση στην κλασική στατιστική θα ήταν, για παράδειγμα, να βρούμε την εκτίμηση μέγιστης πιθανοφάνειας $\widehat{\theta}$ του θ και να βασίσουμε τη συμπερασματολογία στην κατανομή $f(y \mid \widehat{\theta})$, δηλαδή την *εκτιμητική* κατανομή. Τονίζουμε ξανά ότι αυτό δεν αφήνει περιθώριο για τη μεταβλητότητα που δημιουργείται ως αποτέλεσμα της εκτίμησης του θ και έτσι δίνει μία εσφαλμένη αίσθηση ακρίβειας.

Σημείωση 5. *ΔΕΝ μπορούμε να απομακρύνουμε μία σταθερά αναλογίας από την* $f(y \mid \theta)$

ή την $f(\theta \mid x)$. Αν χρησιμοποιήσουμε την posterior πυκνότητα ή την πιθανοφάνεια κατά αναλογία, τότε η απάντηση για την $f(y \mid x)$ θα είναι κι αυτή κατά αναλογία.

Αν και θεωρητικά απλός, ο υπολογισμός μπορεί να είναι δύσκολος. Ωστόσο, αρκετές από τις τυπικές συζυγείς οικογένειες prior οδηγούν σε εύκολα διαχειρίσιμες μορφές για την κατανομή πρόβλεψης.

Παράδειγμα 6.1. *(Διωνυμικό δείγμα). Έστω ότι έχουμε μία παρατήρηση $x \sim$ Bin(n, θ) και ως prior για το θ παίρνουμε τη συζυγή κατανομή Beta(p, q). Τότε, έχουμε δείξει ότι η posterior για το θ είναι η κατανομή Beta$(p+x, q+n-x)$. Τώρα, υποθέτουμε ότι σκοπεύουμε να πάρουμε N ακόμα παρατηρήσεις στο μέλλον και έστω z το πλήθος των επιτυχιών σε αυτές τις N δοκιμές, οπότε $z \mid \theta \sim$ Bin(N, θ).*

Τότε, έχουμε την πιθανοφάνεια της μελλοντικής παρατήρησης:

$$f(z \mid \theta) = \binom{N}{z} \theta^z (1-\theta)^{N-z}, \quad \theta \in (0, 1).$$

Προτού συνεχίσουμε με τη γενική περίπτωση, ας δούμε την απλή ειδική περίπτωση όπου $N = 1$, δηλαδή z είναι το αποτέλεσμα μίας δοκιμής Bernoulli. Θέτοντας, όπως προηγουμένως, $P = p + x$ και $Q = q + n - x$, έχουμε ότι $E(\theta \mid x) = \frac{P}{P+Q}$, οπότε η κατανομή πρόβλεψης του z δεδομένου του x είναι:

z	0	1
$f(z \mid \theta)$	$1-\theta$	θ
$E_\theta\left[f(z \mid \theta) \mid x\right]$	$\frac{Q}{P+Q}$	$\frac{P}{P+Q}$

Επομένως, η κατανομή του $z \mid x$ είναι Bernoulli με πιθανότητα επιτυχίας $\frac{P}{P+Q}$, δηλαδή την posterior μέση τιμή του θ.

Στη γενική περίπτωση, για $z = 0, 1, \ldots, N$, παίρνουμε:

$$\begin{aligned} f(z \mid x) &= \int_0^1 \binom{N}{z} \theta^z (1-\theta)^{N-z} \cdot \frac{\theta^{P-1}(1-\theta)^{Q-1}}{B(P, Q)} d\theta \\ &= \binom{N}{z} \frac{1}{B(P, Q)} \int_0^1 \theta^{P+z-1}(1-\theta)^{Q+N-z-1} d\theta \\ &= \binom{N}{z} \frac{B(P+z, Q+N-z)}{B(P, Q)}. \end{aligned}$$

Αυτή είναι στην πραγματικότητα γνωστή ως η Βήτα-διωνυμική κατανομή.

Παράδειγμα 6.2. *(Δείγμα Γάμμα). Όπως και στο κεφάλαιο 2, θεωρούμε $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές από την κατανομή Gamma(k, θ), όπου το k είναι γνωστό, και χρησιμοποιούμε τη συζυγή prior $\theta \sim$ Gamma(p, q), δηλαδή $f(\theta) \propto \theta^{p-1} e^{-q\theta}$, το οποίο μας οδηγεί, μέσω της εφαρμογής του θεωρήματος του Bayes, στην posterior κατανομή*

$\theta \mid x \sim \textit{Gamma}(p+nk, q+n\overline{x}) \equiv \textit{Gamma}(P,Q)$.

Η πιθανοφάνεια για μία μελλοντική παρατήρηση y είναι:

$$\begin{aligned}
f(y \mid \theta) &= \frac{\theta^k}{\Gamma(k)} y^{k-1} e^{-\theta y}, \quad \theta > 0 \Rightarrow \\
f(y \mid x) &= \int_0^\infty \frac{\theta^k}{\Gamma(k)} y^{k-1} e^{-\theta y} \cdot \frac{Q^P}{\Gamma(P)} \theta^{P-1} e^{-Q\theta} d\theta \\
&= \frac{Q^P y^{k-1}}{\Gamma(P)\Gamma(k)} \int_0^\infty \theta^{P+k-1} e^{-(Q+y)\theta} d\theta \\
&= \frac{Q^P y^{k-1}}{\Gamma(P)\Gamma(k)} \cdot \frac{\Gamma(P+k)}{(Q+y)^{P+k}} = \frac{Q^P y^{k-1}}{B(P,k)(Q+y)^{P+k}}, \quad \text{αφού:}
\end{aligned}$$

$$\int_0^\infty \underbrace{\frac{(Q+y)^{P+k}}{\Gamma(P+k)} \theta^{P+k-1} e^{-(Q+y)\theta}}_{\text{σ.π.π. της κατανομής Gamma}(P+k, Q+y)} d\theta = 1.$$

Σημείωση 6. *Γενικά, ο υπολογισμός αυτών των κατανομών πρόβλεψης απαιτεί την ολοκλήρωση μίας συνάρτησης του θ η οποία είναι ανάλογη σε μία τυπική συνάρτηση πυκνότητας πιθανότητας. Μπορούμε να χρησιμοποιήσουμε το γεγονός ότι γνωρίζουμε τη σταθερά κανονικοποίησης της συνάρτησης πυκνότητας πιθανότητας προκειμένου να μας βοηθήσει να υπολογίσουμε το ολοκλήρωμα. Όταν αναγνωρίζουμε τη συνάρτηση πυκνότητας πιθανότητας, πρέπει να θυμόμαστε ότι είναι πυκνότητα για το θ (αντί το θ να είναι μία γνωστή παράμετρος στην πυκνότητα).*

6.2 Ασκήσεις

Άσκηση 6.1. *Ένα τυχαίο δείγμα $x_1, x_2, \ldots, x_n$ παρατηρείται από την κατανομή Poisson(θ). Η prior για το θ είναι η Gamma(g,h) (τα g και h είναι ακέραιοι). Να δείξετε ότι η κατανομή πρόβλεψης για μία μελλοντική παρατήρηση y από την κατανομή Poisson(θ) είναι:*

$$f(y \mid x) = \binom{G+y-1}{G-1} \left(\frac{1}{1+H}\right)^y \left(1 - \frac{1}{1+H}\right)^G, \quad y = 0, 1, \ldots,$$

για κάποιες τιμές των G και H. Ποια είναι αυτή η κατανομή; (Να χρησιμοποιήσετε το γεγονός ότι $\Gamma(n+1) = n!$ αν το n είναι ακέραιος.)

Άσκηση 6.2. *Η κατανομή των ψεγαδιών κατά μήκος μίας τεχνητής ίνας ακολουθεί μία διαδικασία Poisson, ώστε το πλήθος των ψεγαδιών σε μήκος ℓ της ίνας να ακολουθεί την κατανομή Poisson$(\ell\theta)$. Πολύ λίγα είναι γνωστά για το θ, οπότε χρησιμοποιούμε την prior του Jeffreys για αυτό. Το πλήθος των ψεγαδιών που πήραμε σε 5 ίνες μήκους 10, 15, 25, 30 και 40 μέτρων αντίστοιχα ήταν 3, 2, 7, 6 και 10. Να βρείτε την κατανομή πρόβλεψης για το πλήθος των ψεγαδιών σε μία άλλη ίνα μήκους 60 μέτρων.*

Άσκηση 6.3. *Παρατηρούμε ένα τυχαίο δείγμα $x_1, x_2, \ldots, x_n$ από την κατανομή $N(\theta, \sigma^2)$ με σ^2 γνωστό και υποθέτουμε μία κανονική prior για το θ, η οποία οδηγεί σε μία posterior κατανομή $N(B, D^2)$ για το θ. Να δείξετε ότι η κατανομή πρόβλεψης για μία περαιτέρω παρατήρηση y από την κατανομή $N(\theta, \sigma^2)$ είναι $N(B, D^2 + \sigma^2)$.*

Άσκηση 6.4. *Τα βάρη των προϊόντων από μία συγκεκριμένη διαδικασία παραγωγής είναι ανεξάρτητα και ισόνομα κατανεμημένα, καθένα με κατανομή $N(\theta, 4)$. Ο υπεύθυνος παραγωγής πιστεύει ότι το θ ποικίλει από παρτίδα σε παρτίδα σύμφωνα με μία κατανομή $N(110, 0.4)$. Ένα δείγμα από 5 προϊόντα επιλέγεται τυχαία, αποφέροντας τις μετρήσεις* 108, 109, 107.4, 109.6 *και* 112. *Να εξάγετε την posterior κατανομή για το θ. Να βρείτε επίσης την κατανομή πρόβλεψης για:*

i. *το βάρος ενός περαιτέρω προϊόντος από την παρτίδα και*

ii. *τον δειγματικό μέσο των βαρών m περαιτέρω αντικειμένων από την παρτίδα.*

Άσκηση 6.5. *Γίνονται παρατηρήσεις $x = (x_1, x_2, \ldots, x_n)$ από ανεξάρτητες τυχαίες μεταβλητές $X = (X_1, X_2, \ldots, X_n)$ με τη X_i να ακολουθεί την ομοιόμορφη κατανομή στο $(0, \theta)$, δηλαδή $f(x_i \mid \theta) = \frac{1}{\theta}$ για $x_i \in (0, \theta)$. Υποθέτουμε ότι το θ έχει μία ακατάλληλη prior κατανομή $f(\theta) = \frac{1}{\theta}$ για $\theta > 0$.*

i. *Να δείξετε ότι η posterior κατανομή του θ δίνεται ως:*

$$f(\theta \mid x) = \frac{nM^n}{\theta^{n+1}}, \quad \theta > M,$$

όπου $M = \max\{x_1, x_2, \ldots, x_n\}$.

ii. *Να δείξετε ότι το θ έχει posterior μέση τιμή:*

$$\mathrm{E}(\theta \mid x) = \frac{nM}{n-1}.$$

iii. *Να επαληθεύσετε ότι:*

$$P(\theta > tM \mid x) = \frac{1}{t^n}, \quad \forall t \geqslant 1.$$

iv. *Μία περαιτέρω ανεξάρτητη, παρατήρηση y γίνεται από την ίδια κατανομή με τα x_i. Να δείξετε ότι η κατανομή πρόβλεψης της y έχει πυκνότητα:*

$$f(y \mid x) = \frac{1}{M} \cdot \frac{n}{n+1} \cdot \left[\max\left\{1, \frac{y}{M}\right\}\right]^{-n-1}, \quad y > 0.$$

v. *Να σχεδιάσετε αυτήν την πυκνότητα στην περίπτωση όπου $n = 1$ και $M = 1$.*

Κεφάλαιο 7

Ασυμπτωτική Ανάλυση

7.1 Εισαγωγή

Θυμόμαστε ότι στη συζυγή ανάλυση για το θ, όπου $X_1, X_2, \ldots, X_n \sim N(\theta, \tau^{-1})$, με prior $\theta \sim N(b, c^{-1})$ πήραμε:

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

Τώρα, καθώς $n \to \infty$, αυτή γίνεται:

$$\theta \mid x \sim N\left(\overline{x}, \frac{1}{n\tau}\right) \equiv N\left(\overline{x}, \frac{\sigma^2}{n}\right).$$

Έτσι, καθώς το n μεγαλώνει, η επίδραση της prior εξαφανίζεται και η posterior καθορίζεται αποκλειστικά από τα δεδομένα. Επιπλέον, η posterior κατανομή συγκεντρώνεται όλο και περισσότερο γύρω από το $\overline{x}$, το οποίο συγκλίνει με πιθανότητα 1 στην πραγματική τιμή του θ σύμφωνα με τον ισχυρό νόμο των μεγάλων αριθμών. Τα επιχειρήματα αυτά επισημοποιούνται και γενικεύονται παρακάτω.

7.2 Συνέπεια

Αν η πραγματική τιμή του θ είναι θ_0 και η prior πιθανότητα του θ_0 (ή στη συνεχή περίπτωση μίας αυθαίρετης γειτονιάς του θ_0) δεν είναι μηδέν, τότε, με αυξανόμενο πλήθος δεδομένων x, η posterior πιθανότητα ότι $\theta = \theta_0$ (ή ότι θ βρίσκεται σε μία γειτονιά του θ_0 αντίστοιχα) τείνει στη μονάδα. Αυτό αποδεικνύεται ακολούθως.

Έστω $x_1, x_2, \ldots, x_n$ ανεξάρτητες και ισόνομες παρατηρήσεις, καθεμία με κατανομή $g(x \mid \theta)$. Τότε, η posterior πυκνότητα είναι:

$$f(\theta \mid x) \propto f(\theta) \prod_{i=1}^{n} g(x_i \mid \theta)$$

$$
\begin{aligned}
&= f(\theta)\exp\left\{\sum_{i=1}^{n}\log g(x_i \mid \theta)\right\} \\
&= f(\theta)e^{n\bar{\ell}_n(\theta)} \\
&\propto f(\theta)e^{n\left[\bar{\ell}_n(\theta)-\bar{\ell}_n(\theta_0)\right]}, \quad \text{όπου:} \\
\bar{\ell}_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\log g(x_i \mid \theta).
\end{aligned}
$$

Για δεδομένο θ, έστω $h(x_i) = \log g(x_i \mid \theta) - \log g(x_i \mid \theta_0)$. Τότε, η συνάρτηση $\bar{\ell}_n(\theta) - \bar{\ell}_n(\theta_0)$ είναι ο δειγματικός μέσος των $h(x_i)$ και έτσι συγκλίνει με πιθανότητα 1 στη μέση τιμή της $h(X)$ κάτω από την πραγματική τιμή θ_0 του θ, δηλαδή στην:

$$
E_{\theta_0}\left[h(X)\right] = \int h(x)g(x \mid \theta_0)dx = \int \left[\log g(x \mid \theta) - \log g(x \mid \theta_0)\right]g(x \mid \theta_0)dx.
$$

Μπορεί να δειχθεί ότι $E_{\theta_0}\left[h(X)\right] \leqslant 0$ και $E_{\theta_0}\left[h(X)\right] = 0$ για $\theta = \theta_0$. Έτσι, για $\theta \neq \theta_0$, έπεται ότι η ποσότητα $e^{n\left[\bar{\ell}_n(\theta)-\bar{\ell}_n(\theta_0)\right]}$ συγκλίνει στο 0 με πιθανότητα 1, καθώς $n \to \infty$, αλλά παραμένει 1 για $\theta = \theta_0$. Αυτό αρκεί, δεδομένου ότι $f(\theta_0) \neq 0$, για να αποδείξουμε τον ισχυρισμό.

Επομένως, υπό την προϋπόθεση ότι η prior κατανομή δίνει μη-μηδενικό βάρος στην πραγματική τιμή του θ, τότε η posterior πιθανότητα θα συγκεντρωθεί τελικά στην πραγματική τιμή θ_0.

7.3 Ασυμπτωτική κανονικότητα

Όταν η παράμετρος θ είναι συνεχής, το επιχείρημα αυτό μπορεί να επεκταθεί προκειμένου να πάρουμε την προσεγγιστική μορφή της posterior κατανομής όταν το n είναι μεγάλο. Σύμφωνα με το επιχείρημα της προηγούμενης παραγράφου, όσο το n μεγαλώνει, η ποσότητα $e^{n\left[\bar{\ell}_n(\theta)-\bar{\ell}_n(\theta_0)\right]}$ είναι αμελητέα παντού εκτός από μία συρρικνούμενη περιοχή του θ_0. Έτσι, η $f(\theta)$ μπορεί να θεωρηθεί σταθερή σε αυτή τη γειτονιά, οπότε παίρνουμε:

$$
\begin{aligned}
f(\theta \mid x) &\propto e^{n\bar{\ell}_n(\theta)} \\
&\propto e^{n\left[\bar{\ell}_n(\theta)-\bar{\ell}_n(\widehat{\theta})\right]},
\end{aligned}
$$

όπου $\widehat{\theta}$ είναι η εκτίμηση μέγιστης πιθανοφάνειας του θ. Αναπτύσσουμε τώρα τον εκθέτη $\phi(\theta) = n\left[\bar{\ell}_n(\theta) - \bar{\ell}_n(\widehat{\theta})\right]$ σε αυτή την έκφραση σύμφωνα με μία σειρά Taylor γύρω από το $\widehat{\theta}$:

$$
\begin{aligned}
\phi(\theta) &= \phi(\widehat{\theta}) + (\theta-\widehat{\theta})\phi'(\widehat{\theta}) + \frac{1}{2}(\theta-\widehat{\theta})^2\phi''(\widehat{\theta}) + \cdots \\
&= n\cancelto{0}{\left[\bar{\ell}_n(\widehat{\theta}) - \bar{\ell}_n(\widehat{\theta})\right]} + n(\theta-\widehat{\theta})\cancelto{0}{\bar{\ell}_n'(\widehat{\theta})} + \frac{n}{2}(\theta-\widehat{\theta})^2\bar{\ell}_n''(\widehat{\theta}) + \cdots \\
&\approx \frac{n}{2}(\theta-\widehat{\theta})^2\bar{\ell}_n''(\widehat{\theta}),
\end{aligned}
$$

εφόσον το $\widehat{\theta}$ είναι η εκτίμηση μέγιστης πιθανοφάνειας του θ, οπότε μηδενίζει τη συνάρτηση $\overline{\ell}'_n(\theta)$. Παρατηρούμε επίσης ότι θα πρέπει $\overline{\ell}''_n(\widehat{\theta}) \leqslant 0$, αφού το $\widehat{\theta}$ είναι σημείο μεγίστου της συνάρτησης $\overline{\ell}_n(\theta)$. Τότε,

$$f(\theta \mid x) \propto \exp\left\{\frac{n(\theta-\widehat{\theta})^2\overline{\ell}''_n(\widehat{\theta})}{2}\right\} = \exp\left\{-\frac{-n\overline{\ell}''_n(\widehat{\theta})}{2}\cdot(\theta-\widehat{\theta})^2\right\} \Rightarrow$$

$$\theta \mid x \sim N\left(\widehat{\theta}, -\frac{1}{n\overline{\ell}''_n(\widehat{\theta})}\right).$$

Έτσι, καθώς $n \to \infty$, η posterior είναι κατά προσέγγιση κανονικά κατανεμημένη γύρω από την εκτίμηση μέγιστης πιθανοφάνειας $\widehat{\theta}$. Παρατηρούμε ξανά ότι αυτό το αποτέλεσμα ισχύει ανεξαρτήτως της prior που έχει προσδιοριστεί, υπό την προϋπόθεση ότι η prior δε μηδενίζεται γύρω από την πραγματική τιμή του θ.

Το αποτέλεσμα αυτό έχει αρκετές χρήσεις. Κατ' αρχάς, μπορεί να χρησιμοποιηθεί απευθείας για να πάρουμε προσεγγιστικές posterior πιθανότητες σε καταστάσεις όπου οι υπολογισμοί για να βρούμε την posterior είναι δύσκολοι. Έπειτα, μπορεί να δώσει χρήσιμες τιμές εκκίνησης για αριθμητικούς υπολογισμούς σε περιπτώσεις όπου οι αναλυτικές λύσεις δεν είναι εύκολες στον χειρισμό. Το πιο σημαντικό όμως είναι πως καταδεικνύει ότι μόλις πάρουμε αρκετά δεδομένα, η επιλογή μίας συγκεκριμένης prior δεν ενέχει λόγους ανησυχίας. Δύο άτομα μπορεί να θεωρήσουν αρκετά διαφορετικές prior για να απεικονίσουν τις πρότερες πεποιθήσεις τους και τελικά μόλις γίνουν διαθέσιμα αρκετά δεδομένα, οι posterior συμπερασματολογίες τους να συμφωνούν.

Παράδειγμα 7.1. *(Μέση τιμή κανονικής). Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές από την κατανομή $N(\theta, \sigma^2)$, όπου το σ^2 είναι γνωστό.*

Ως συνήθως, αυτό δίνει την πιθανοφάνεια:

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

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

$$\ell'(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) \Rightarrow$$

$$\ell''(\theta) = -\frac{n}{\sigma^2} \Rightarrow$$

$$-n\overline{\ell}''(\widehat{\theta}) = -\ell''(\widehat{\theta}) = \frac{n}{\sigma^2}.$$

Επομένως, καθώς $n \to \infty$,

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

Αυτό ισχύει *για κάθε* prior κατανομή που αποδίδει μη-μηδενική πιθανότητα γύρω από την πραγματική τιμή του θ.

Παράδειγμα 7.2. *(Διωνυμικό δείγμα). Θεωρούμε το μοντέλο πιθανοφάνειας $x \sim Bin(n, \theta)$.*

Τότε,

$$f(x \mid \theta) \propto \theta^x (1-\theta)^{n-x} \Rightarrow$$

$$\ell(\theta) = \log f(x \mid \theta) = x \log \theta + (n-x) \log(1-\theta) + c \Rightarrow$$

$$\ell'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} \Rightarrow$$

$$\ell''(\theta) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \Rightarrow$$

$$-\ell''(\widehat{\theta}) = -\ell''\left(\frac{x}{n}\right) = \frac{n^2 x}{x^2} + \frac{n^2(n-x)}{(n-x)^2} = \frac{n^3}{x(n-x)} = \frac{n}{\widehat{\theta}(1-\widehat{\theta})}.$$

Έτσι, καθώς $n \to \infty$,

$$\theta \mid x \sim N\left(\frac{x}{n}, \frac{x(n-x)}{n^3}\right).$$

7.4 Ασκήσεις

Άσκηση 7.1. *Να βρείτε την ασυμπτωτική posterior κατανομή του θ για καθένα από τα δύο μοντέλα στην άσκηση 2.1.*

Άσκηση 7.2. *Να βρείτε την ασυμπτωτική posterior κατανομή του b για το μοντέλο Pareto της άσκησης 3.3.*